

## Тема 2. Парная и множественная регрессия и корреляция в эконометрических исследованиях

### 1. ОСНОВНЫЕ ПОЛОЖЕНИЯ

*Корреляционно-регрессионный анализ – один из наиболее широко распространенных и гибких приемов обработки статистических данных. Данный метод начинает свой отсчет с 1795 г., когда английский исследователь Фрэнсис Гальтон предложил теоретические основы регрессионного метода, а в 1801 г. рассчитал с его помощью траекторию полета планеты Церера. Им же введен в статистику термин «корреляция». Можно также назвать французского кристаллографа Огюста Браве, немецкого физика Густава Теодора Фехнера, английского экономиста и статистика Фрэнсиса Эджуорта, впервые высказывавших в середине-конце XIX в. идеи о количественном измерении связей явлений. В разное время над теорией анализа работали известные в области теоретической статистики ученые Карл Фридрих Гаусс (Германия), Адриен Мари Лежандр (Франция), Карл Пирсон (Англия) и др.*

*Корреляционно-регрессионный анализ состоит в построении и анализе экономико-математической модели в виде уравнения регрессии (корреляционной связи), характеризующего зависимость признака от определяющих его факторов.*

В природе одни признаки обуславливают влияние других (первые называют факторными или независимыми, вторые - результативными или зависимыми). Зависимость может быть либо функциональной, либо статистической (корреляционно-регрессионной).

Функциональная связь характеризуется полным соответствием между влиянием факторного и результативного признаков, то есть одному значению фактора всегда соответствует одно единственное значение результативного признака. Зная признак-фактор можно найти признак-результат (например, зная валовой сбор и площадь посева зерновых культур, можно абсолютно точно найти их урожайность).

В подавляющем числе случаев между экономическими переменными нет строгой (функциональной зависимости). Это связано с:

- **не учётом целого ряда других факторов;**
- **не прямого воздействия ряда факторов;**
- **многие воздействия носят случайный характер**

Если при изменении одного из признаков происходит изменение распределения другого, то такая зависимость называется статистической. А если статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой, то такую статистическую зависимость называют корреляционно-регрессионной.

В корреляционных связях между влиянием факторного и результативного признаков нет полного соответствия, воздействие факторов проявляется лишь при наблюдении за большим количеством фактических данных. Это связано с воздействием на результативный признак большого числа факторных (например, доза внесения удобрений оказывает влияние на урожайность, но, зная её, невозможно абсолютно точно определить какой будет урожайность, и выявить эту взаимосвязь можно только при большом числе наблюдений).

Регрессия (зависимость) – это усредненная (т. е. свободная от случайных мелкоаппаратных колебаний) связь между объясняемой переменной (переменными) и объясняющей переменной (переменными).

Регрессионный анализ – статистический метод исследования взаимосвязи переменных. Его цель – разработать статистическую модель, позволяющую предсказывать значения зависимой переменной, или отклика, по значениям, по крайней мере одной, независимой, или объясняющей, переменной.

**По количеству переменных регрессия бывает простая и множественная.**

**Парная регрессия** устанавливает зависимость одной случайной переменной от другой:

$$Y=f(x) + \varepsilon$$

где  $y$  – зависимая переменная (результативный признак);

$x$  – независимая, объясняющая переменная (факторный признак).

$\varepsilon$  – случайная величина, характеризующая отклонения фактического значения результативного признака ( $y_i$ ) от его расчетного значения ( $y_x$ ), найденного по уравнению парной регрессии.

**Множественная регрессия** – регрессия с двумя и более факторными переменными.

Уравнение множественной регрессии:

$$Y=f(x_1, x_2, \dots, x_m) + \varepsilon$$

где  $y$  – результативный признак,  $x_1, x_2, \dots, x_m$  – признаки-факторы.

**В зависимости от вида зависимости регрессии подразделяются на линейные и нелинейные.**

**Линейная регрессия** – это связь (регрессия), которая представлена уравнением прямой линии и выражает простейшую линейную зависимость.

где  $a_0$  и  $a_1$  – коэффициенты регрессии;  $y_x = a_0 + a_1x + \varepsilon$

$a_0$  – показывает влияние на результативный признак  $y$  неучтенных в модели факторов;

$a_1$  – указывает величину изменений зависимой переменной  $y$  при изменении независимой переменной  $x$  на одну единицу.

**Нелинейные регрессии делятся на:**

а) регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам:

полином второй степени –

$$y_x = a_0 + a_1x + a_2x^2 + \varepsilon$$

гипербола –  $y_x = a_0 + \frac{a_1}{x} + \varepsilon$

обратная

$$y_x = \frac{1}{a_0 + a_1x + \varepsilon}$$

б) регрессии, нелинейные по оцениваемым параметрам:

показательная –

$$y_x = a_0 \cdot a_1^x \cdot \varepsilon$$

степенная –

$$y_x = a_0x^{a_1} \cdot \varepsilon$$

## 2. ЗНАЧЕНИЕ, РОЛЬ, СУЩНОСТЬ И ВОЗМОЖНОСТИ КМ.

Корреляционно-регрессионный анализ предназначен для изучения корреляционных связей.

**Корреляционная модель (КМ)** – это математическое выражение типа уравнения, которое показывает, на сколько единиц изменится результативный показатель при изменении факторного на единицу.

Иными словами КМ есть количественный аналог формирования показателей.

*КМ имеет результативный показатель, который изменяется при изменении одного или нескольких факторных показателей. КМ дает усредненную оценку эффективности включенных в нее факторов.*

*В обобщенном виде КМ означает, что любой результативный показатель есть функция от каких-то факторов.*

**Объясняемая (результативная) переменная** — переменная, которая статистически зависит от факторной переменной, или объясняющей (регрессора).

Методика подбора результативного и факторных показателей.

Основные этапы построения КМ:

1. Выбор результативного и факторных показателей.
2. Сбор информации и проверка ее на достоверность.

3. Выбор вида модели.

4. Расчет параметров и характеристик КМ.

1). Выполняя подбор результативного и факторных показателей КМ необходимо руководствоваться следующими положениями.

**1. Результативный показатель в цепочке причинно-следственных связей всегда находится на более высоком уровне.** И этот уровень определяют на основе логичных рассуждений, на основе знаний о том, какие из рассматриваемых показателей являются первичными и вторичными.

**2. В корреляционную модель включают факторы, которые логически определяют все или почти все содержание результативного показателя.**

**3. Если результативный показатель синтетический, сложный, то и факторные показатели должны быть такими же.**

**4. Если результативный показатель относительный, то и факторные должны иметь один и тот же знаменатель.**

## **2) сбор информации и проверка ее на достоверность**

Корреляционные зависимости проявляются только при массовых наблюдениях. Поэтому для построения КМ выборка должна иметь не менее 20 наблюдений (организаций), если КМ многофакторная то – в 2,5 или 3 раза больше числа факторов КМ.

В выборку следует включать однотипные объекты, то есть нельзя в одну выборку включать данные о деятельности крупных агрохолдингов и небольших организаций. Собранные информация должна быть однородной, то есть соответствовать закону нормального распределения: чем ближе значение к средней арифметической, тем выше вероятность его появления.

## **3) Определение вида КМ.**

По виду КМ бывают линейные и нелинейные. По количеству учтенных факторов делятся на однофакторные и многофакторные.

Способы определения вида КМ:

а) логический – используется опыт, знания о существующих взаимосвязях;

б) аналитический – основан на выполнении определенных расчетов;

в) графический – строится точечный график зависимости результата от фактора (поле корреляции) и с помощью линии тренда определяется вид КМ.

## **4) Расчет параметров и характеристик КМ**

Параметры чаще всего исчисляются с помощью метода наименьших квадратов.

*Метод наименьших квадратов (МНК) — способ приближенного нахождения (оценки) неизвестных коэффициентов (параметров) регрессии. Этот метод основан на требовании минимизации суммы квадратов отклонений значений результата, рассчитанных по уравнению регрессии, и истинных (наблюденных) значений результата.*

**Характеристики КМ** определяют устойчивость модели, возможность ее использования в анализе экономики и планировании показателей на перспективу.

Если модель однофакторная линейная, то важнейшей характеристикой является коэффициент парной корреляции – определяет силу связи между двумя показателями.

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

где  $-1 \leq r_{yx} \leq 1$ ,  $r_{yx}$  - парный коэффициент корреляции между x и y.

В том случае, если КМ многофакторная или нелинейная, то для определения тесноты связи используется **коэффициент множественной корреляции R**;

$$R = \sqrt{1 - \frac{\sum (y_x - y_i)^2}{\sum (y_i - y)^2}}$$

В таком случае, если  $t_{R(n)} \geq 2,48$ , то модель устойчива и на ее основе можем рассчитать другие характеристики.

Одной из таких характеристик является коэффициент эластичности  $\mathcal{E}_j$

$$\mathcal{E}_j = a_j \frac{\bar{x}_j}{\bar{y}}$$

где  $a_j$  - коэффициент регрессии при факторе  $j$ ;  $\bar{x}_j$  - среднеарифметическая фактора  $j$ .

Коэффициент эластичности позволяет сравнить роль отдельных факторов в формировании результативного показателя. Он показывает, на сколько единиц изменяется (уменьшается или увеличивается) результативный показатель, если факторный показатель изменится на единицу. Если в сумме коэффициент эластичности больше 1, т.е.  $\sum_j \mathcal{E}_j > 1$ ,

то это означает, что результативный показатель прирастает более быстрыми темпами чем факторные и наоборот.

Коэффициенты эластичности имеют один недостаток, который состоит в том, что они приемлимы, если вариация факторов одинакова или почти одинакова.

В том случае если вариация факторов существенно отличается, то для объяснения роли отдельных факторов в формировании результативного показателя используются  $\beta_j$  - коэффициенты.

$$\beta_j = a_j \frac{\sigma_{x_j}}{\sigma_y}$$

Коэффициент  $\beta_j$  показывает, на сколько сравнимых или стандартных единиц возрастает результативный показатель при изменении (увеличении) факторного на 1.

В целом коэффициенты  $\mathcal{E}_j$  и  $\beta_j$  позволяют сравнить степень влияния на результативный показатель факторов, имеющих несопоставимые единицы измерения.

В многофакторной КМ роль отдельных факторов в формировании результативного показателя не одинакова. Различна и теснота связи отдельных факторных показателей с результативным. Отсюда следует, что устойчивость характеристик различных факторов может быть разной.

В связи с этим требуется проверить на существенность отдельные коэффициенты регрессии.

Существенность коэффициента регрессии определяется по формуле

$$t_{aj} = a_j / M_{aj},$$

где  $M_{aj}$  - ошибка коэффициента регрессии,

$$M_{aj} = \frac{\sigma_{y_x y_i}}{\sigma_{x_j} \sqrt{n}},$$

где  $\sigma_{y_x y_i}$  - стандартная ошибка по модели или необъяснимая вариация,

$$\sigma_{y_x y_i} = \sqrt{\frac{\sum (y_x - y_i)^2}{n}}$$

Существенным является тот коэффициент регрессии, для которого выполняются условия, что фактическое  $t_{aj}$  больше табличного  $t_{aj}^e \approx 1,96$  при  $n \rightarrow \infty$ .

В том случае, если коэффициенты существенности отдельных факторов ниже табличных, то эти факторы из КМ исключаются и производится перерасчет параметров и характеристик модели. При этом коэффициенты регрессии изменяются.

Наряду с перечисленными выше, важным является критерий Фишера (F), он характеризует эффективность КМ в целом, точнее определяет насколько полно построенная модель выражает изучаемую закономерность.

Если  $F > 1,5$ , то КМ количественно выражает ту закономерность, которую изучают. F-критерий измеряется как отношение общей дисперсии к остаточной.

$$F = \frac{D_{y_i}}{D_{y_x y_i}} = \frac{\sum (y_i - \bar{y})^2}{\sum (y_x - y_i)^2},$$

где

$$D_{y_i} = \sum (y_i - \bar{y})^2, \quad D_{y_x y_i} = \sum (y_x - y_i)^2.$$

Для характеристики степени связи или мультиколлинеарности факторов используются коэффициенты внутренней меры определенности  $d_{вн}$ ,  $j \leq 0,5$  факторы слабо коррелируют между собой. По мере увеличения  $d_{вн}$ ,  $j > 0,5$  степень взаимосвязи факторов КМ возрастает.

$R^2$ ,  $r^2$ ,  $\eta^2$  выраженные в процентах, показывают на сколько процентов факторы, учтенные в КМ объясняют изменение резульативного показателя. (коэффициент детерминации).

### **3. МЕТОДИКА РАСЧЕТА ПАРАМЕТРОВ КМ С КОЛИЧЕСТВЕННОЙ ОЦЕНКОЙ ВЛИЯНИЯ КАЧЕСТВЕННЫХ ПРИЗНАКОВ**

Все качественные признаки можно разделить на две группы: альтернативные и нарастающие.

Нарастающие- изменяются от какой-то величины и возрастают. Специфика этих признаков в том, что они формируются под влиянием нескольких элементов, взаимосвязанных между собой. Пример: квалификация специалиста. Ее элементами являются: образование, стаж работы, черты характера и др. Применительно к каждому специалисту влияние этих параметров на квалификацию неодинаково.

Качественные альтернативные признаки характеризуются тем, что они или присутствуют или нет. Пример качественного альтернативного признака - сорт растений, который в посевах или есть, или нет; породы животных и т.д.

Для количественной характеристики качественных признаков используется два параметра 1 и 0. Если количественный признак присутствует в наблюдении (опыте) ставим 1, если отсутствует – 0. В КМ вводим столько дискретных величин, сколько качественных признаков.

Исходя из вышеуказанного методика расчета параметров КМ с количественной оценкой влияния качественных признаков следующая:

1. выбор резульативного и факторных показателей, в том числе качественных признаков;
2. сбор информации, в том числе количественное выражение качественных признаков, и проверка ее на достоверность;
3. выбор вида КМ одним из способов: аналитическим, графическим, логическим;
4. расчет параметров КМ используя метод наименьших квадратов и определение характеристик КМ;

5. экономическая интерпретация влияния качественного признака и количественных факторов на результат
6. сравнение фактических и расчетных значений результативного показателя, анализ использования ресурсов на основе КМ и планирование показателей.

#### 4. ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИ НАРУШЕНИИ КЛАССИЧЕСКИХ ПРЕДПОЛОЖЕНИЙ (4 ч)

##### Классические условия регрессионного анализа (условия Гаусса-Маркова):

1. Регрессионная модель линейна по параметрам, корректно специфицирована, и содержит аддитивный случайный член.
  2. Случайный член имеет нулевое среднее.
  3. Все объясняющие переменные не коррелированы со случайным членом (случайными отклонениями).
  4. Наблюдаемые значения случайного члена (случайных отклонений) не коррелированы друг с другом.
  5. Дисперсия случайных отклонений постоянна для любых наблюдений.
- Ни одна из объясняющих переменных не является строгой линейной функцией других объясняющих переменных

##### Суть, последствия и определение мультиколлинеарности.

Термин «коллинеарность» описывает линейную связь между двумя независимыми переменными, тогда как мультиколлинеарность – между более, чем двумя переменными. На практике обычно используется один термин.

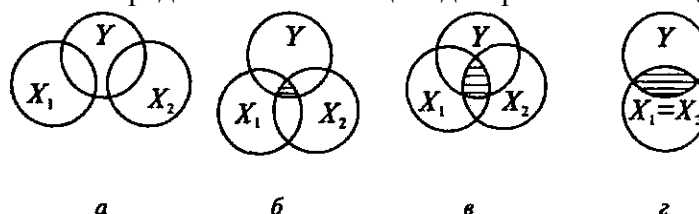
Термин «мультиколлинеарность» введен Рагнарсом Фришем.

Если объясняющие переменные связаны строгой функциональной зависимостью, то говорят о совершенной мультиколлинеарности.

Мультиколлинеарность может быть проблемой лишь в случае множественной регрессии.

Совершенная мультиколлинеарность не позволяет однозначно определить коэффициенты регрессии и разделить вклады объясняющих переменных ( $X_1, X_2, \dots, X_n$ ) в их влиянии на зависимую переменную  $Y$ . Следовательно, в этом случае выводы по коэффициентам и по самому уравнению регрессии будут ненадежными.

Совершенная мультиколлинеарность является скорее теоретическим примером. Реальна же ситуация, когда между объясняющими переменными существует довольно сильная корреляционная зависимость, а не строгая функциональная. Такая зависимость называется несовершенной мультиколлинеарностью. Она характеризуется высоким коэффициентом корреляции между соответствующими объясняющими переменными. Наглядно мультиколлинеарность можно представить помощью диаграммы Венна (рис.)



На рис., а коррелированность между объясняющими переменными отсутствует и влияние каждой из них на  $Y$  находит отражение в наложении кругов  $X_1$  и  $X_2$  на круг  $Y$ . По мере усиления линейной зависимости между  $X_1$  и  $X_2$  соответствующие круги все больше накладываются друг на друга. Заштрихованная область отражает совпадающие части влияния  $X_1$  и  $X_2$  на  $Y$ . На рис. г при совершенной мультиколлинеарности невозможно разграничить степени индивидуального влияния объясняющих переменных на зависимую переменную  $Y$ .

Существует несколько признаков, по которым может быть установлено наличие мультиколлинеарности.

1. Значение коэффициента детерминации  $R^2$  высокое, но некоторые из коэффициентов регрессии статистически незначимы, т.е. они имеют низкие t-статистики.

2. Парная корреляция между малозначимыми объясняющими переменными достаточно высокая.

*Причины возникновения мультиколлинеарности:*

1. Ошибочное включение в уравнение двух или более линейно зависимых переменных

2. Две или более объясняющие переменные, в нормальной ситуации слабо коррелированные, становятся в конкретных условиях выборки сильно коррелированными

3. В модель включается переменная, сильно коррелирующая с зависимой переменной (такая независимая переменная называется доминантой)

В большинстве случаев мультиколлинеарность не является серьезной проблемой, чтобы прилагать существенные усилия по ее выявлению и устранению. В основном все зависит от целей исследования.

Если основная задача модели — прогноз будущих значений зависимой переменной, то при достаточно большом коэффициенте детерминации  $R^2 (\geq 0,9)$  наличие мультиколлинеарности обычно не сказывается на прогнозных качествах модели. Но это утверждение будет обоснованным лишь в том случае, если и в будущем между коррелированными переменными будут сохраняться те же отношения, что и ранее.

Если же целью исследования является определение степени влияния каждой из объясняющих переменных на зависимую переменную, то наличие мультиколлинеарности, приводящее к увеличению стандартных ошибок, исказит истинные зависимости между переменными. В этой ситуации мультиколлинеарность является серьезной проблемой.

*Последствия мультиколлинеарности*

1. Оценки коэффициентов остаются несмещенными  
2. Стандартные ошибки коэффициентов увеличиваются  
3. Вычисленные t-статистики занижены.  
4. Оценки становятся очень чувствительными к изменению отдельных наблюдений.

5. Общее качество уравнения, а также оценки переменных, не связанных мультиколлинеарностью, остаются незатронутыми.

6. Чем ближе мультиколлинеарность к совершенной (строгой), тем серьезнее ее последствия.

*Выделяют следующие методы устранения мультиколлинеарности.*

1. Исключение одной или ряда коррелированных переменных из модели - простейший метод устранения мультиколлинеарности. Но необходима осмотрительность при применении данного метода, т.к. возможны ошибки спецификации. Например, при исследовании спроса на некоторое благо в качестве объясняющих переменных можно использовать цену данного блага и цены заменителей данного блага, которые зачастую коррелируют друг с другом. Исключив из модели цены заменителей, будет допущена ошибка спецификации. Вследствие этого можно получить смещенные оценки и сделать необоснованные выводы.

Таким образом, в прикладных эконометрических моделях желательно не исключать объясняющие переменные до тех пор, пока коллинеарность не станет серьезной проблемой.

2. Получение дополнительных данных или новой выборки

Поскольку мультиколлинеарность напрямую зависит от выборки, то, возможно, при другой выборке мультиколлинеарности не будет либо она не будет столь серьезной.

Иногда для уменьшения мультиколлинеарности достаточно увеличить объем выборки. Например, при использовании ежегодных данных можно перейти к поквартальным

данным. Увеличение количества данных сокращает дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость.

3. Изменение спецификации модели – т.е. изменяется форма модели, или добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную.

Использование данного метода уменьшает сумму квадратов отклонений, приводит к уменьшению стандартных ошибок коэффициентов регрессии.

4. Использование предварительной информации о некоторых параметрах

Например, известными значениями некоторых коэффициентов регрессии. Значения коэффициентов, рассчитанные для каких-либо предварительных (обычно более простых) моделей либо для аналогичной модели по ранее полученной выборке, иногда могут быть использованы для разрабатываемой в данный момент модели.

Недостатки данного метода - получение предварительной информации зачастую затруднительно, а во-вторых, вероятность того, что выделенный коэффициент регрессии будет одним и тем же для различных моделей, невысокая.

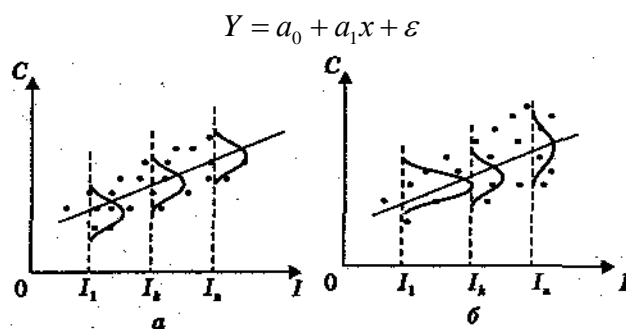
5. Преобразование переменных. Например, каскадный корреляционный анализ.

### Суть, последствия и обнаружение гетероскедастичности.

Постоянство дисперсии случайных отклонений называется **гомоскедастичностью**, а непостоянство – **гетероскедастичностью**.

Условие равенства дисперсий отклонений, подразумевает, что несмотря на то что при каждом конкретном наблюдении случайное отклонение может быть большим или малым, положительным или отрицательным, не должно быть причины, вызывающей большее отклонение (ошибку) при одних наблюдениях и меньшее - при других.

Рассмотрим два примера линейной регрессии зависимости потребления( $y$ ) от дохода ( $x$ ):



В обоих случаях с ростом дохода растет среднее значение потребления. Но если в первом случае дисперсия потребления остается одной и той же для различных уровней дохода, то во втором случае при аналогичной зависимости среднего потребления от дохода дисперсия потребления не постоянна, а увеличивается с ростом дохода. Фактически это означает, что во 2 сл. Субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, и разброс в их потреблении более существенен для большего уровня дохода. Т.е. разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, что и определяет дисперсию случайных отклонений.

*Виды гетероскедастичности:* истинная (вызываемая непостоянством дисперсии свободного члена, ее зависимостью от различных факторов) и ложная (вызывается ошибочной спецификацией модели регрессии).

Источники истинной гетероскедастичности:

1. Возникает в пространственных выборках при зависимости масштаба изменений зависимой переменной от некоторой, называемой фактором пропорциональности( $Z$ )

2. Возникает во временных рядах когда зависимая переменная имеет большой интервал качественно неоднородных значений или высокий темп изменения.

3. Возникает в любой модели, в случае если качество данных варьирует внутри выборки.

*Последствия применения МНК при гетероскедастичности:*

1. Оценки коэффициентов останутся несмещенными и линейными.

2. Увеличение дисперсии распределения оценок снижает вероятность получения максимально точных оценок. *Оценки не будут эффективными (т.е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра).*

3. Недооценка стандартных ошибок т.к. если стандартные ошибки коэффициентов регрессии будут занижены, следовательно t-статистики будут завышены. Это может привести к признанию статистически значимых коэффициентов, которые таковыми не являются.

В ряде случаев, зная характер данных, проявление проблемы гетероскедастичности можно предвидеть и устранить на этапе спецификации. Ответив на вопросы: нет ли очевидных ошибок спецификации, можно ли содержательно предполагать какой-то вид гетероскедастичности, рассмотреть график остатков.

Однако чаще эту проблему приходится решать после построения уравнения регрессии.

*Методы обнаружения гетероскедастичности:*

1. Графический анализ остатков

2. Тест ранговой корреляции Спирмена

3. Тест Парка.

*Методы смягчения проблемы гетероскедастичности.*

1. Метод взвешенных наименьших квадратов (ВНК) – применяется при известных для каждого наблюдения значениях (дисперсии). В этом случае можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение среднего квадратического отклонения.

2. Переопределение переменных.

3. Вычисление стандартных ошибок с поправкой на гетероскедастичность (метод Уайта)

### **Суть, последствия и обнаружение автокорреляции.**

В силу ряда причин (ошибок спецификации, инерционности рассматриваемых зависимостей и др.) в регрессионных моделях может иметь место корреляционная зависимость между соседними случайными отклонениями. Это нарушает одну из фундаментальных предпосылок МНК. Вследствие этого оценки, полученные на основе МНК, перестают быть эффективными. Это делает ненадежными выводы по значимости коэффициентов регрессии и по качеству самого уравнения. Поэтому достаточно важным является умение определить наличие автокорреляции и устранить это нежелательное явление.

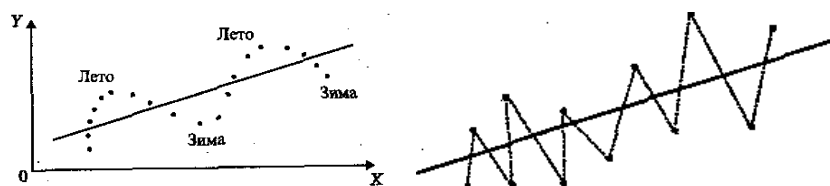
*Автокорреляция (последовательная корреляция)* определяется как корреляция между наблюдаемыми показателями; упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные). Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов. При использовании перекрестных данных наличие автокорреляции (пространственной корреляции) крайне редко.

Поэтому в дальнейших выкладках вместо символа  $i$  порядкового номера наблюдения будем использовать символ  $t$ , отражающий момент наблюдения. Объем выборки при этом будем обозначать символом  $T$  вместо  $n$ . В экономических задачах значительно чаще встречается так называемая *положительная автокорреляция* ( $\sigma(\varepsilon_{t-1}, \varepsilon_t) > 0$ ), нежели *отрицательная автокорреляция* ( $\sigma(\varepsilon_{t-1}, \varepsilon_t) < 0$ ).

В большинстве случаев положительная автокорреляция вызывается направленным постоянным воздействием некоторых не учтенных в модели факторов.

Суть автокорреляции рассмотрим на примере. Пусть исследуется спрос  $Y$  на прохладительные напитки в зависимости от дохода  $X$  по ежемесячно данным. Трендовая зависи-

мость, отражающая увеличение спроса с ростом дохода, может быть представлена линейной функцией  $y_i = b_0 + b_1 x_i$ , изображенной на рис.



Однако фактические точки наблюдений обычно будут превышать трендовую линию в летние периоды и будут ниже в зимние.

*Аналогичная картина может иметь место в макроэкономическом анализе с учетом циклов деловой активности.*

Отрицательная автокорреляция фактически означает, что за положительным отклонением следует отрицательное и наоборот. Возможная схема рассеивания точек в этом случае представлена на рис.

Такая ситуация может иметь место, например, если ту же зависимость между спросом на прохладительные напитки и доходами рассматривать по сезонным данным (зима — лето).

Среди *основных причин, вызывающих появление автокорреляции*, можно выделить ошибки спецификации, инерцию в изменении экономических показателей, эффект паутины, сглаживание данных.

*Ошибки спецификации.* Неучет в модели какой-либо важной объясняющей переменной либо неправильный выбор формы зависимости обычно приводит к системным отклонениям точек наблюдений от линии регрессии, что может обусловить автокорреляцию.

*Инерция.* Многие экономические показатели (например инфляция, безработица, ВВП и т. п.) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Действительно, экономический подъем приводит к росту занятости, сокращению инфляции, увеличению ВВП и т.д. Это продолжается до тех пор, пока изменение конъюнктуры рынка и ряда экономических характеристик не приведет к замедлению роста, затем остановке и движению вспять рассматриваемых показателей. В любом случае эта трансформация происходит не мгновенно, а обладает определенной инертностью.

*Эффект паутины.* Во многих производственных и других сферах экономические показатели реагируют на изменение экономических условий с запаздыванием (временным лагом) пример, предложение сельскохозяйственной продукции реагирует на изменение цены с запаздыванием (равным периоду созревания урожая). Большая цена сельскохозяйственной продукции в прошедшем году вызовет (скорее всего) ее перепроизводство в текущем году, а следовательно, цена на нее снизится и т.д.

*Сглаживание данных.* Зачастую данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подынтервалам. Это может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода, что в свою очередь может послужить причиной автокорреляции.

**Последствия автокорреляции** в определенной степени сходны с последствиями гетероскедастичности. Среди них при применении МНК обычно выделяются следующие.

1. Оценки параметров, оставаясь линейными и несмещенными, перестают быть эффективными. *Следовательно, они перестают обладать свойствами наилучших линейных несмещенных оценок (BLUE-оценок).*

2. Дисперсии оценок являются смещенными. Часто дисперсии, вычисляемые по стандартным формулам, являются заниженными, что влечет за собой увеличение  $t$ -

статистик. Это может привести к признанию статистически значимыми объясняющие переменные, которые в действительности таковыми могут и не являться.

4. Выводы по  $t$ -статистикам, определяющим значимость коэффициентов регрессии и коэффициента детерминации, возможно, будут неверными. Вследствие этого ухудшаются прогнозные качества модели.

*Методы определения автокорреляции:* графический, метод рядов, критерий Дарбина-Уотсона.

*Методы устранения автокорреляции.* При установлении автокорреляции необходимо в первую очередь проанализировать правильность спецификации модели. Если после ряда возможных усовершенствований регрессии (уточнения состава объясняющих переменных либо изменения формы зависимости) автокорреляция по-прежнему имеет место, то, возможно, это связано с внутренними свойствами ряда отклонений. В этом случае возможны определенные преобразования, устраняющие автокорреляцию. Среди них выделяется авторегрессионная схема первого порядка АН(1). Для применения которой необходимо оценить коэффициент корреляции между отклонениями.