

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ПУТЕЙ СООБЩЕНИЯ (МИИТ)

---

Кафедра “Математика”

А.С.МИЛЕВСКИЙ

**ВЫСШАЯ МАТЕМАТИКА**

ЧАСТЬ 7.

**ЭКОНОМЕТРИКА**

*Конспект лекций*

МОСКВА – 2011

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ПУТЕЙ СООБЩЕНИЯ (МИИТ)

---

Кафедра “Математика”

А.С.МИЛЕВСКИЙ

**ВЫСШАЯ МАТЕМАТИКА**

ЧАСТЬ 7.

**ЭКОНОМЕТРИКА**

*Конспект лекций*

Рекомендовано редакционно-издательским  
советом университета в качестве учебного  
пособия для студентов ИЭФ

МОСКВА – 2011

УДК-330.43

М-60

Милевский А.С. Высшая математика. Часть 7.  
Эконометрика: конспект лекций. – М.: МИИТ, 2011. –  
123 с.

Конспект лекций предназначен для студентов,  
изучающих курс эконометрики в институте ИЭФ.  
Включает в себя материал по классической линейной  
модели множественной регрессии, нарушениям  
предположений классической модели, временным  
рядам, динамическим моделям и системам  
эконометрических уравнений.

Рецензенты:

Соболева Е.С., к.ф.-м.н., доцент кафедры  
“Математический анализ” МГУ им. М.В.Ломоносова,  
Деснянский В.Н., к.ф.-м.н., зав. кафедрой  
“Математический анализ” МИИТ.

© Московский государственный  
университет путей сообщения  
(МИИТ), 2011

Св. план 2011 г.; поз. 155

Милевский Александр Станиславович

ВЫСШАЯ МАТЕМАТИКА.

Часть 7. Эконометрика.  
Конспект лекций.

---

Подписано в печать

Формат 60x84 / 16

Заказ №

Усл. печ. л. –

Тираж – 200 экз.

---

127994, ГСП-4, ул. Образцова, д.9, стр.9,  
типография МИИТа

## 1. Введение.

### 1.1. Предмет эконометрики.

**Эконометрика – это наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов при помощи математических методов и моделей.**

История становления эконометрики как самостоятельной научной дисциплины насчитывает не так уж много лет, но за последние десятилетия эконометрика получила значительное развитие.

Норвежский ученый Р. Фриш определил ее как «...единство экономической теории, статистики и математики». Попробуем в этом разобраться.



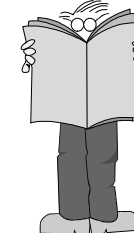
Основные результаты *экономической теории* носят *качественный* характер, например, даются ответы на вопросы вида:

- Какие факторы влияют на совокупный объём инвестиций в частном бизнесе?
- Как зависят темпы экономического роста от размеров дефицита государственного бюджета, объёма инвестиций, внешнего долга, уровня инфляции?
- Каким образом изменит структуру потребления снижение ставок налогов?

Однако для принятия обоснованных решений часто требуется отвечать на вопросы, начинающиеся словами «**на сколько**». Например,

- Какого изменения объёма потребления электроэнергии можно ожидать, если предположить, что цены на неё возрастут на 30%?
  - На сколько увеличится спрос на данную продукцию, если удвоить текущие расходы на её рекламу?
- Это **количественное содержание** и является предметом изучения эконометрики.

Следует отметить, что зачастую не удаётся достичь желаемой степени точности предсказаний. Модели и уравнения, дающие хорошее согласование для одного периода времени, могут оказаться неудовлетворительными для другого. Это связано с большой сложностью экономических процессов.



Но в любом случае анализ в эконометрике производится на основе широкого применения *самых разнообразных математических методов и моделей*.

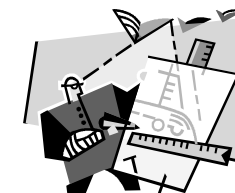
И, конечно, значимые результаты можно получить лишь путем обработки *реальных статистических данных*.

### 1.2. Модели и статистические данные

Закономерности в экономике проявляются как взаимосвязи между экономическими показателями. В большинстве случаев их стремятся выразить в сравнительно простой математической форме.

Надо иметь в виду, что всякая математическая модель является лишь упрощённым формализованным представлением реальности.

Искусство построения модели состоит в совмещении “желаемого с возможным”.



**Пример.** Изучая стоимость квадратного метра квартиры в Москве  $Y$ , естественно предполагать, что она зависит от

- количества комнат  $N$ ,
- жилой площади квартиры  $P$ ,
- нежилой площади  $Q$ ,
- расстояния от центра города  $R$ ,
- расстояния от ближайшей станции метро  $S$ ,
- вида дома (кирпичный, панельный и т.п.)  $T$ ,
- этажа  $U$ ,...



$$Y = f(N, P, Q, R, S, T, U, \dots)$$

А какие ещё факторы вы бы включили в эту модель?

В любом случае влияние неучтённых (или неизвестных) факторов будет представлено как некая случайная добавка  $\varepsilon$  к  $Y$ .

Уже в этом примере видна различная роль рассматриваемых переменных  $Y, N, P, Q, R, S, T, U$  (или, как их ещё называют, **факторов**).

В пределах данной модели одни переменные являются *независимыми* (или *объясняющими*, *экзогенными* –  $N, P, Q, R, S, T, U$ ), а другие *зависимыми* (или *эндогенными* –  $Y$ ).

В общем случае экономическая модель с одной *эндогенной*  $Y$  и  $m$  *объясняющими* переменными  $x_1, x_2, \dots, x_m$  имеет вид:

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

Конкретный вид функциональной зависимости может быть различным. Наиболее часто строятся **линейные эконометрические модели**:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \quad (1.1)$$

разумеется, из-за их большей простоты.

Здесь  $\beta_0, \beta_1, \dots, \beta_m$  – **параметры модели**, которые оцениваются на основе статистических данных.

Наличие *случайной составляющей*  $\varepsilon$  в модели обусловлено причинами двоякой природы.

- Во-первых, она отражает влияние на формирование значения  $Y$  *факторов*, не учтённых в перечне *объясняющих переменных*  $X$
- Во-вторых, она может включать в себя *погрешность в измерении*  $Y$ .

В реальности возможны также ошибки в выборе вида функциональной зависимости и т.п.

Зависимости другого вида тоже часто удаётся свести к линейным. Так, например, зависимость

$$Y = Ax^B \cdot u$$

сводится к линейной так:  $\ln Y = \ln A + B \ln x + \ln u$

Выбор вида функциональной зависимости называется **спецификацией модели**, а определение состава объясняющих переменных – **спецификацией переменных**.

Если модель (1.1) содержит только одну объясняющую переменную ( $m=1$ ), она называется **парной регрессией**.

При  $m > 1$  мы имеем дело с **множественной регрессией**

### 1.3. Статистические данные

Основу эконометрического моделирования составляют статистические данные. Их различают по типам.

*Пространственные данные* собираются по какому-либо экономическому показателю для разных объектов в один момент времени или в разные моменты в случае, когда время несущественно.

Например, данные по курсам покупки/продажи валюты сегодня в различных обменных пунктах Москвы.



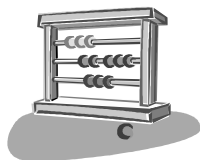
*Временные ряды* – данные для одного объекта в различные моменты времени.

Например, ежеквартальные данные об инфляции, данные о денежной эмиссии за последние годы, ежедневный курс доллара.



Промежуточное положение занимают *панельные данные*, которые отражают наблюдения по большому числу объектов за относительно небольшой период времени.

Например, данные об объемах выпуска и капитальных вложениях российских предприятий топливно-энергетического комплекса за 1995–2000 г.

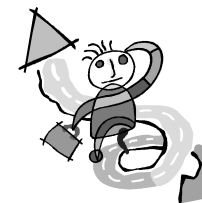


### 1.4. Основные этапы эконометрического моделирования

В процессе эконометрического исследования можно выделить следующие этапы:

- *1-й этап (постановочный)*. Определение конечных целей моделирования, формирование набора участвующих в модели факторов и их ролей.
- *2-й этап (спецификация модели)*. Выбор общего вида модели, состава и вида входящих в неё функциональных взаимосвязей.
- *3-й этап (информационный)*. Сбор необходимой статистической информации, т.е. значений участвующих в модели факторов.
- *4-й этап (идентификация модели)*. Анализ модели, статистическое оценивание параметров модели.
- *5-й этап (проверка модели)*. Сопоставление реальных и оценённых данных, проверка адекватности модели.

В результате проверки модели нередко приходится неоднократно возвращаться к предыдущим этапам и вносить коррективы.



## 2. Классическая модель множественной линейной регрессии

### 2.1. Описание модели

Модель множественной линейной регрессии имеет вид

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \quad (2.1)$$

где  $y$  – зависимая переменная,  $x_1, \dots, x_m$  – объясняющие переменные (также называемые *регрессорами*),  $\beta_0, \beta_1, \dots, \beta_m$  – коэффициенты регрессии,  $\varepsilon$  – случайный член (или “ошибки измерений”).

Коэффициенты  $\beta_0, \beta_1, \dots, \beta_m$  оцениваются на основе статистических данных.

Для нахождения оценок проводятся  $n$  наблюдений ( $n > m$ ), в результате чего получается  $n$  равенств

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

(индекс  $i$  показывает номер наблюдения).

Это удобно записать в *матричной форме*:

$$Y = X\beta + \varepsilon, \quad (2.2)$$

где

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.3)$$

## 2.2. Основные гипотезы

Гипотезы, лежащие в основе *классической модели множественной регрессии*, следующие:

- (A) “Истинная” зависимость  $Y$  от регрессоров  $X$  имеет вид (2.2).
- (B) Величины  $X$  – детерминированные (не случайные).
- (C) Столбцы матрицы  $X$  линейно независимы (другими словами, *ранг матрицы  $X$  равен  $m+1$* ).
- (D)  $M\varepsilon_i = 0$ ,  $D\varepsilon_i = \sigma^2$  не зависит от  $i$
- (E)  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  при  $i \neq j$  (*некоррелированность ошибок для разных наблюдений*).

**Замечание.** Вместо (D) часто добавляют условие

- (F) Ошибки  $\varepsilon_i$  имеют *нормальное* распределение  $N(0, \sigma^2)$ .

В этом случае модель называется *нормальной*.

Замечание. Условия (D) – (F) удобно записать в *матричной форме*

- (G)  $M\varepsilon = 0$ ,  $M(\varepsilon\varepsilon^T) = \sigma^2 E$  ( $E$  – единичная матрица)
- (F)  $\varepsilon \sim N(0, \sigma^2 E)$

## 2.3. Оценки параметров регрессии

Будем искать оценки  $\hat{\beta}$  коэффициентов  $\beta$  в задаче (2.1) *методом наименьших квадратов*. Другими словами, оценки должны удовлетворять условию

$$\sum (y_i - \hat{y}_i)^2 \rightarrow \min; \quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots$$

или, в матричной форме,  $(Y - X\hat{\beta})^T \cdot (Y - X\hat{\beta}) \rightarrow \min$

Если продифференцировать это равенство по  $\beta$  и приравнять производные нулю, получим

$$X^T X \hat{\beta} = X^T Y$$

Отсюда

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad (2.4)$$

**Пример 1.** В таблице представлены объём импорта (Y) и ВВП (X<sub>1</sub>) и индекс потребительских цен (X<sub>2</sub>) в США

Годы	1964	1965	1966	1967	1968
X <sub>1</sub>	636	689	753	796	868
X <sub>2</sub>	93	95	97	100	104
Y	28	32	38	41	53

Построим оценки параметров уравнения линейной регрессии Y на X<sub>1</sub>, X<sub>2</sub>:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

$$Y = \begin{pmatrix} 28 \\ 32 \\ 38 \\ 41 \\ 53 \end{pmatrix}, X = \begin{pmatrix} 1 & 636 & 93 \\ 1 & 689 & 95 \\ 1 & 753 & 97 \\ 1 & 796 & 100 \\ 1 & 868 & 104 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}.$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 636 & 689 & 753 & 796 & 868 \\ 93 & 95 & 97 & 100 & 104 \end{pmatrix} \cdot \begin{pmatrix} 1 & 636 & 93 \\ 1 & 689 & 95 \\ 1 & 753 & 97 \\ 1 & 796 & 100 \\ 1 & 868 & 104 \end{pmatrix} = \begin{pmatrix} 5 & 3742 & 489 \\ 3742 & 2833266 & 367516 \\ 489 & 367516 & 47899 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 636 & 689 & 753 & 796 & 868 \\ 93 & 95 & 97 & 100 & 104 \end{pmatrix} \cdot \begin{pmatrix} 28 \\ 32 \\ 38 \\ 41 \\ 53 \end{pmatrix} = \begin{pmatrix} 192 \\ 147110 \\ 18942 \end{pmatrix}$$

$$\begin{cases} 5\hat{\beta}_0 + 3742\hat{\beta}_1 + 489\hat{\beta}_2 = 192 \\ 3742\hat{\beta}_0 + 2833266\hat{\beta}_1 + 367516\hat{\beta}_2 = 147110 \\ 489\hat{\beta}_0 + 367516\hat{\beta}_1 + 47899\hat{\beta}_2 = 18942 \end{cases}$$

$$(X^T X)^{-1} = \begin{pmatrix} 2452,8 & 1,822 & -39,02 \\ 1,822 & 0,0014 & -0,03 \\ -39,02 & -0,03 & 0,625 \end{pmatrix} \quad \hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} -150,99 \\ 0,02 \\ 1,78 \end{pmatrix}$$

$$y = -150,99 + 0,02x_1 + 1,78x_2$$

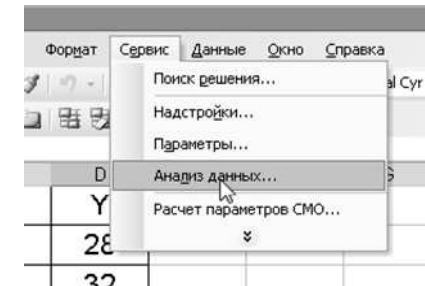
## 2.4. Нахождение параметров регрессии при помощи MS Excel

Для нахождения параметров регрессии при помощи MS Excel необходимо, чтобы был установлен инструмент "Анализ данных", который не устанавливается при стандартной установке MS Office, а только при выборочной.

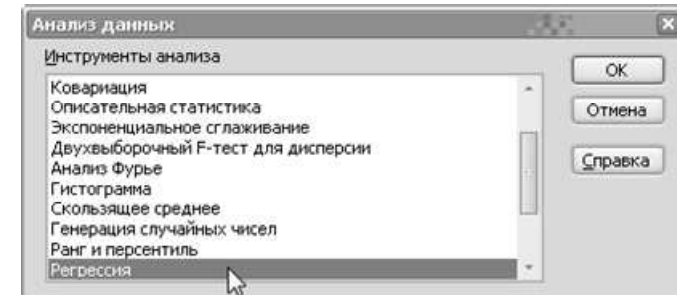
**Пример.** Возьмём тот же пример и внесём данные в таблицу MS Excel по столбцам:

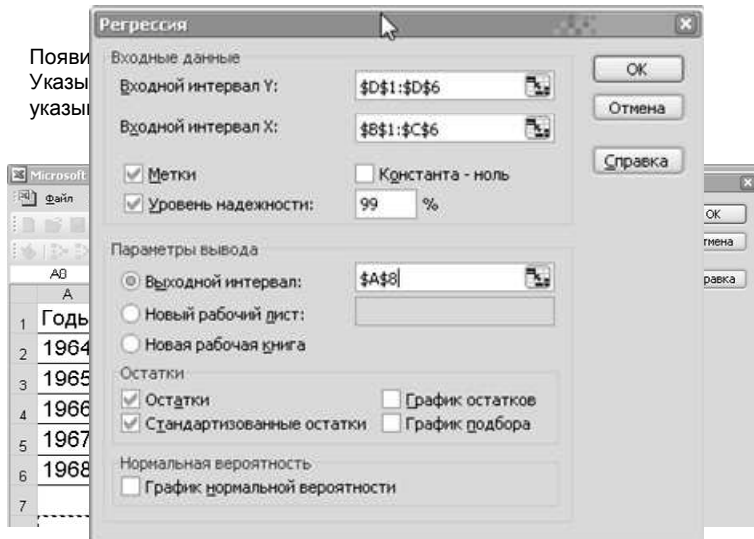
	A	B	C	D
1	Годы	X1	X2	Y
2	1964	636	93	28
3	1965	689	95	32
4	1966	753	97	38
5	1967	796	100	41
6	1968	868	104	53

Выберем пункт меню *Сервис-Анализ данных*



Укажем пункт *Регрессия*:





Программа выдаст множество данных по регрессии, в том числе и оценки параметров регрессии:

Дисперсионный анализ			
	df	SS	F
Регрессия	2,0000	361,6094	
Остаток	2,0000	7,5906	
Итого	4,0000	369,2000	

	Коэффициенты	Стандартная ошибка	t-статистика
Y-пересечение	-150,9880	96,4838	
X1	0,0200	0,0736	
X2	1,7832	1,5403	

В частности, уравнение регрессии:

$$y = -150,99 + 0,02x_1 + 1,78x_2$$

(96,48)            (0,07)            (1,54)

## 2.4. Теорема Гаусса-Маркова

**Теорема.** В предположениях (A) – (E) оценки, полученные методом наименьших квадратов по формуле (2.4), являются *несмещёнными* и обладают *наименьшей дисперсией среди всех линейных несмещённых оценок* параметров  $\beta$ .

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad (2.4)$$

**Доказательство.** Докажем только несмещённость.

$$\begin{aligned} M\hat{\beta} &= M(X^T X)^{-1} X^T Y = M(X^T X)^{-1} X^T (X\beta + \varepsilon) = \\ &= (X^T X)^{-1} X^T M(X\beta + \varepsilon) = (X^T X)^{-1} X^T X\beta = \beta, \\ &\text{так как } M\varepsilon = 0 \end{aligned}$$

## 2.5. Коэффициент детерминации

Дисперсия зависимой переменной (или *полная сумма квадратов*) обозначается TSS:

$$TSS = \sum (y_i - \bar{y})^2 = nD_y^*$$

Сумма

$$RSS = \sum (\hat{y}_i - \bar{y})^2 = nD_{\hat{y}}^*$$

определяет разброс оценки относительно среднего, или *“объяснённую часть дисперсии”*. И, наконец, сумма квадратов *“остатков”*

$$ESS = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2 = nD_e^*$$

представляет собой *“необъяснённую часть дисперсии”* зависимой переменной.



Нетрудно проверить, что

$$RSS + ESS = TSS$$

Величина

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

называется *коэффициентом детерминации* и представляет собой долю дисперсии зависимой переменной  $y$ , объясненную при помощи оцененного уравнения регрессии.

$$0 \leq R^2 \leq 1$$

Чем ближе коэффициент детерминации к 1, тем лучше приближаются исходные данные уравнением линейной регрессии.

**Замечание.** В модели *множественной линейной регрессии*, кроме  $\beta$ , есть ещё один параметр –  $\sigma^2$ . Его оценку можно найти по формуле

$$\hat{\sigma}^2 = S^2 = \frac{ESS}{n - m - 1}$$

**Пример.** В рассмотренном примере 1  $y = -150,99 + 0,02x_1 + 1,78x_2$

$$Y = \begin{pmatrix} 28 \\ 32 \\ 38 \\ 41 \\ 53 \end{pmatrix}, \hat{Y} = X\hat{\beta} = \begin{pmatrix} 1 & 636 & 93 \\ 1 & 689 & 95 \\ 1 & 753 & 97 \\ 1 & 796 & 100 \\ 1 & 868 & 104 \end{pmatrix} \cdot \begin{pmatrix} 150,99 \\ 0,03 \\ 1,78 \end{pmatrix} = \begin{pmatrix} 27,59 \\ 32,22 \\ 37,07 \\ 43,28 \\ 51,85 \end{pmatrix}$$

$$\bar{Y} = \frac{28 + 32 + \dots}{5} = 38,4; \quad RSS = (27,59 - 38,4)^2 + \dots = 361,609;$$

$$TSS = (28 - 38,4)^2 + \dots = 369,24; \quad ESS = (28 - 27,59)^2 + \dots = 7,5906;$$

$$S^2 = \frac{ESS}{5 - 2 - 1} = 3,7953; \quad R^2 = \frac{361,609}{369,24} = 0,9794$$

**Замечание.** На MS Excel эти параметры выглядят следующим образом:

Вывод остатка

Наблюдение	Предсказанное Y
1	27,5891
2	32,2172
3	37,0656
4	43,2765
5	51,8516

$$\hat{Y} = \begin{pmatrix} 27,59 \\ 32,22 \\ 37,07 \\ 43,28 \\ 51,85 \end{pmatrix}$$

Дисперсионный анализ

	df	SS
Регрессия	2,0000	361,6094
Остаток	2,0000	7,5906
Итого	4,0000	369,2000

$$RSS = 361,609;$$

$$TSS = 369,24;$$

$$ESS = 7,5906;$$

Регрессионная статистика

Множественный R	0,9897
<b>R-квадрат</b>	<b>0,9794</b>
Нормированный R-квадрат	0,9589
Стандартная ошибка	1,9481

$$S^2 = 3,7953;$$

$$R^2 = 0,9794$$

**Замечание.** Важное свойство коэффициента детерминации  $R^2$  состоит в том, что при добавлении в модель новых объясняющих переменных  $x$  он не может уменьшиться. Поэтому при сравнении двух моделей с различным числом переменных не всегда ясно, за счёт чего возрос показатель детерминации: за счёт реального влияния дополнительного фактора на результат или просто из-за возрастания числа переменных.

Для того чтобы можно было *сравнивать различные модели*, вводят так называемый **скорректированный коэффициент детерминации**:

$$R_{корр}^2 = 1 - \frac{ESS/(n - m - 1)}{TSS/(n - 1)}$$

Это число также не превосходит 1, но в некоторых случаях может оказаться и отрицательным. В рассматриваемом примере

$$R_{корр}^2 = 0,9589$$

## 2.6. Квантили

Формулы, описывающие статистические свойства оценок, содержат табличные значения, называемые квантилями.

Квантиль стандартного нормального закона  $N(0;1)$  обозначается  $U_p$ .

Свойство:

$$U_{1-\alpha} = -U_\alpha$$

p	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
$U_p$	1.282	1.645	1.960	2.325	2.576	3.090	3.291

Вопрос.

Чему равно  $U_{0,01}$ ?

Квантили закона распределения хи-квадрат:  $\chi_p^2(k)$ .

k \ p	0,005	0,001	0,05	0,1	0,9	0,95	0,99	0,995
1	0,0	0,0	0,0	0,0	2,7	3,8	6,6	7,9
2	0,0	0,0	0,1	0,2	4,6	6,0	9,2	10,6
3	0,1	0,0	0,4	0,6	6,3	7,8	11,3	12,8
4	0,2	0,1	0,7	1,1	7,8	9,5	13,3	14,9
9	1,7	1,2	3,3	4,2	14,7	16,9	21,7	23,6
19	6,8	5,4	10,1	11,7	27,2	30,1	36,2	38,6
29	13,1	11,0	17,7	19,8	39,1	42,6	49,6	52,3
100	67,3	61,9	77,9	82,4	118,5	124,3	135,8	140,2

$$\chi_p^2(k) \approx k \cdot \left( 1 - \frac{2}{9k} + u_p \sqrt{\frac{2}{9k}} \right)^3$$

На MS Excel можно найти квантиль так:

A	B	A	B
1	$\alpha$ 0,1	1	$\alpha$ 0,1
2	k 4	2	k 4
3	$\chi_{1-\alpha}^2(k)$ =ХИ2ОБР(В1;В2)	3	$\chi_{1-\alpha}^2(k)$ 7,78

Квантили закона Стьюдента:  $t_p(k)$ .

$$t_{1-\alpha}(k) = -t_\alpha(k)$$

k \ p	0,9	0,95	0,975	0,99	0,995	0,999
1	3,08	6,31	12,71	31,82	63,66	318,31
2	1,89	2,92	4,30	6,96	9,92	22,33
3	1,64	2,35	3,18	4,54	5,84	10,21
4	1,53	2,13	2,78	3,75	4,60	7,17
9	1,38	1,83	2,26	2,82	3,25	4,30
19	1,33	1,73	2,09	2,54	2,86	3,58
29	1,31	1,70	2,05	2,46	2,76	3,40
100	1,29	1,66	1,98	2,36	2,63	3,17

$t_p(k) \approx u_p$  при больших k

На MS Excel можно найти квантиль так:

$$t_p(k) \approx u_p \cdot \left( 1 - \frac{u_p^2 + 1}{4k} \right)^{-1}$$

A	B	A	B
1	$\alpha$ 0,2	1	$\alpha$ 0,2
2	k 4	2	k 4
3	$t_{1-\alpha/2}(k)$ =СТЮДРАСПОБР(В1;В2)	3	$t_{1-\alpha/2}(k)$ 1,533

Квантили закона Фишера:  $F_p(k_1, k_2)$ .

Свойство:

$$F_{1-\alpha}(k_1, k_2) = \frac{1}{F_\alpha(k_2, k_1)}$$

Квантили распределения  $F_{0,9}(k_1, k_2)$

$k_1 \setminus k_2$	1	2	3	4	9	19	29	39	50	100
1	40	8,5	5,5	4,5	3,4	3,0	2,9	2,8	2,8	2,8
2	49	9,0	5,5	4,3	3,0	2,6	2,5	2,4	2,4	2,4
3	54	9,2	5,4	4,2	2,8	2,4	2,3	2,2	2,2	2,1
4	56	9,2	5,3	4,1	2,7	2,3	2,1	2,1	2,1	2,0
9	60	9,4	5,2	3,9	2,4	2,0	1,9	1,8	1,8	1,7
19	62	9,4	5,2	3,8	2,3	1,8	1,7	1,6	1,6	1,5
29	62	9,5	5,2	3,8	2,3	1,8	1,6	1,6	1,5	1,4
100	63	9,5	5,1	3,8	2,2	1,7	1,5	1,4	1,4	1,3

Есть несколько приближённых формул для  $F_p(k_1, k_2)$ .  
Например, при  $k_1, k_2 \geq 10$  такая:

$$F_p(k_1, k_2) \approx \left( \frac{C \cdot D + u_p(\sqrt{A \cdot E + B \cdot C})}{E} \right)^3,$$

$$A = \frac{2}{9k_1}, B = \frac{2}{9k_2}, C = 1 - A, D = 1 - B, E = D - Bu_p^2$$

На MS Excel можно найти квантиль так:

	A	B
1	$\alpha$	0,1
2	$k_1$	4
3	$k_2$	2
4	$F_{1-\alpha}(k_1, k_2)$	=ФРАСПОБР(В1;В2;В3)

	A	B
1	$\alpha$	0,1
2	$k_1$	4
3	$k_2$	2
4	$F_{1-\alpha}(k_1, k_2)$	9,24

## 2.7. Статистические свойства оценок параметров регрессии

Теорема. В предположениях (A) – (E) и (F) теоремы Гаусса-Маркова оценки, полученные методом наименьших квадратов по формуле (2.4), обладают следующими свойствами:

- $\frac{(n-m-1) \cdot S^2}{\sigma^2} \sim \chi^2(n-m-1);$
- $S^2$  и  $\hat{\beta}$  независимы.

## 2.8. Доверительные интервалы

Будем считать, что условие нормальности (F) выполнено. Тогда можно построить *доверительные интервалы* для параметров регрессии, т.е. интервалы, *содержащие точные значения этих параметров с заданной вероятностью  $1-\alpha$* .

Вот соответствующие формулы:

$$\hat{\beta}_i - t_{1-\alpha/2}(n-m-1) \cdot S_{\beta_i} < \beta_i < \hat{\beta}_i + t_{1-\alpha/2}(n-m-1) \cdot S_{\beta_i}$$

$$\frac{ESS}{\chi_{1-\alpha/2}^2(n-m-1)} < \sigma^2 < \frac{ESS}{\chi_{\alpha/2}^2(n-m-1)}$$

Здесь  $\alpha$  – заданный “уровень значимости”, т.е. *вероятность того, что построенный интервал не содержит истинного значения параметра*.

Далее,  $S_{\beta}$  – выборочные среднеквадратичные отклонения оценок параметров (или *стандартные ошибки параметров*), они находятся по формулам:

$$S_{\beta_i}^2 = S^2 \cdot (X^T \cdot X)^{-1}_{ii}$$

Как правило, стандартные ошибки параметров вычисляются в эконометрических программах вместе с оценками самих параметров.

**Пример.** В рассмотренном примере 1

$$(X^T X)^{-1} = \begin{pmatrix} 2452,8 & 1,822 & -39,02 \\ 1,822 & 0,0014 & -0,03 \\ -39,02 & -0,03 & 0,625 \end{pmatrix}, y = -150,99 + 0,02x_1 + 1,78x_2, S^2 = 3,7953$$

$$S_{\beta_0}^2 = S^2 \cdot 2452,8 = 9309,16; S_{\beta_0} = 96,48$$

$$S_{\beta_1}^2 = S^2 \cdot 0,0014 = 0,0053; S_{\beta_1} = 0,07$$

$$S_{\beta_2}^2 = S^2 \cdot 0,625 = 2,372; S_{\beta_2} = 1,54$$

Таким образом,  $y = -150,99 + 0,02x_1 + 1,78x_2$   
 (96,48) (0,07) (1,54)

Далее, на уровне значимости  $\alpha=0,1$ :

$t_{1-\alpha/2}(n-m-1) = t_{0,95}(2) = 2,92$ ;  
 $-150,99 - 2,92 \cdot 96,48 = -432,72 < \beta_0 < 130,74 = -150,99 + 2,92 \cdot 96,48$

Аналогично,

$-0,19 < \beta_1 < 0,23$ ;  
 $-2,71 < \beta_2 < 6,28$ ;

$\chi^2_{1-\alpha/2}(n-m-1) = \chi^2_{0,95}(2) = 5,99$ ;  $\chi^2_{\alpha/2}(n-m-1) = \chi^2_{0,05}(2) = 0,1$ ;  
 $\frac{ESS}{5,99} = \frac{7,5906}{5,99} = 1,27 < \sigma^2 < 75,9 = \frac{7,5906}{0,1}$



**2.9. Доверительный интервал для прогнозного значения**

Рассмотрим модель(2.2):  $Y = X\beta + \varepsilon$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Напомним, что X – матрица, каждая из n строк которой соответствует какому-то набору значений независимых переменных.

Предположим, что у нас есть ещё один набор значений независимых переменных

$x^{(n+1)} = (1, x_1^{(n+1)}, x_2^{(n+1)}, \dots, x_m^{(n+1)})$ .

Требуется найти доверительный интервал для соответствующего значения зависимой переменной  $y^{(n+1)}$ .

Тогда

$$\hat{y}^{(n+1)} - t_{1-\alpha/2}(n-m-1) \cdot S_y < y^{(n+1)} < \hat{y}^{(n+1)} + t_{1-\alpha/2}(n-m-1) \cdot S_y$$

где

$$\hat{y}^{(n+1)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(n+1)} + \dots + \hat{\beta}_m x_m^{(n+1)} = x^{(n+1)} \cdot \hat{\beta}$$

$$S_y = S \cdot \sqrt{1 + x^{(n+1)} \cdot (X^T X)^{-1} \cdot (x^{(n+1)})^T}$$

**Пример 2.** В таблице приведены значения двух факторов.

X	1	2	3	6	8
Y	1	4	5	9	11

Построить уравнение линейной регрессии и найти доверительный интервал для Y при X=7. Взять  $\alpha=0,1$ .

**Решение.**

$Y = \begin{pmatrix} 1 \\ 4 \\ 5 \\ 9 \\ 11 \end{pmatrix}, X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 6 \\ 1 & 8 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$

Тогда

$(X^T X)^{-1} = \begin{pmatrix} 0,671 & -0,118 \\ -0,118 & 0,029 \end{pmatrix}, y = 0,59 + 1,35x$   
 $S^2 = 0,5882$

и

$\hat{y}^{(n+1)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(n+1)} = (1 \ 7) \cdot \begin{pmatrix} 0,59 \\ 1,35 \end{pmatrix} = 10,04$

$S_y = \sqrt{0,5882} \cdot \sqrt{1 + (1 \ 7) \cdot \begin{pmatrix} 0,671 & -0,118 \\ -0,118 & 0,029 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 7 \end{pmatrix}}$

$S_y \approx 0,921$ ;  $t_{1-\alpha/2}(n-m-1) = t_{0,95}(5-1-1) = 2,35$

Окончательно,

$10,04 - 2,35 \cdot 0,921 < y^{(n+1)} < 10,04 + 2,35 \cdot 0,921$

## 2.10. Проверка гипотез о параметрах регрессии

В этом параграфе также будем считать, что условие нормальности **(F)** выполнено:

(F) Ошибки  $\varepsilon_i$  имеют *нормальное* распределение  $N(0, \sigma^2)$ .

При исследовании модели приходится проверять гипотезы о равенстве истинных значений параметров регрессии заданным числам. Для этого **достаточно построить доверительный интервал и посмотреть, попадает ли в него это заданное число.**

Рассмотрим, например, гипотезу  $H_0: \beta_i = A$   
при альтернативной гипотезе  $H_1: \beta_i \neq A$ .

Тогда гипотеза  $H_0$  принимается в случае, *если число A лежит в интервале*

$$\left( \hat{\beta}_i - t_{1-\alpha/2}(n-m-1) \cdot S_{\beta_i}, \hat{\beta}_i + t_{1-\alpha/2}(n-m-1) \cdot S_{\beta_i} \right)$$

В противном случае принимается гипотеза  $H_1$

Особенно часто проверяют гипотезу  $H_0: \beta_i = 0$ .  
Дело в том, что если принимается альтернативная гипотеза

$$H_1: \beta_i \neq 0$$

то говорят, что коэффициент  $\beta_i$  **значим**.

Таким образом, получаем следующий **критерий значимости параметра  $\beta_i$** :

$$\frac{|\hat{\beta}_i|}{S_{\beta_i}} \geq t_{1-\alpha/2}(n-m-1)$$

**Пример.** В рассмотренном примере 1 было:  
 $-0,19 < \beta_1 < 0,23;$   
 $-2,71 < \beta_2 < 6,28;$

Доверительный интервал для коэффициента  $\beta_1$  содержит 0, так что коэффициент  $\beta_1$  *не значим*.

Коэффициент  $\beta_2$  также не значим.

**Задача.** По выборке объема 8 получено следующее уравнение линейной регрессии:

$$y = 5,27 - 0,89x_1 + 1,69x_2 - 0,41x_3$$

(2,97)      (0,53)      (0,58)      (0,32)

Какие из его коэффициентов значимы на уровне значимости  $\alpha = 0,1$ ?

**Решение.** Условие значимости:

$$\frac{|\hat{\beta}_i|}{S_{\beta_i}} \geq t_{1-\alpha/2}(n-m-1)$$

$$n = \dots$$

$$m = \dots$$

$$t_{1-\alpha/2}(n-m-1) = \dots \quad \frac{0,89}{0,53} \approx 1,68 < 2,13 \Rightarrow \dots$$

$$n = 8$$

$$m = 3$$

$$\frac{1,69}{0,59} \approx \dots \Rightarrow \dots$$

$$t_{1-\alpha/2}(n-m-1) = t_{0,95}(4) = 2,13$$

## 2.11. Проверка гипотезы о значимости уравнения регрессии

При построении эконометрической модели обязательно возникает вопрос: в какой мере можно ей доверять, насколько полученное уравнение *значимо*?

Для проверки *значимости уравнения* используется **F-статистика Фишера**:

$$F = \frac{RSS/m}{ESS/(n-m-1)}$$

Разделив числитель и знаменатель на TSS, получим ещё одну формулу для F:

$$F = \frac{R^2/m}{(1-R^2)/(n-m-1)}$$

Пусть выбрано некоторое значение *уровня значимости  $\alpha$* .

Если вычисленное значение F-статистики оказывается больше  $F_{1-\alpha}(m, n - m - 1)$ , то оцененное уравнение регрессии считается значимым при выбранном  $\alpha$ .

**Задача.** По выборке получено уравнение линейной регрессии:

$$y = 4,8 - 2,15x_1 + 2,18x_2$$

Значимо ли это уравнение на уровне 0,1?

**Решение.** 1) Вычислим расчётные значения  $\hat{Y}$ :

y	x <sub>1</sub>	x <sub>2</sub>	$\hat{y}$
6	2	3	
10	2	4	
10	5	7	
11	7	10	
3	7	6	

$\hat{y}$
7,04
9,22
9,31
11,55
2,83

2) Вычислим RSS и ESS ...

$$3) F = \frac{RSS/m}{ESS/(n-m-1)} = \frac{43,5/2}{2,5/(5-2-1)} = 17,4$$

Квантиль  $F_{0,9}(2;5-2-1) = F_{0,9}(2;2)$  по таблице равен 9.

Так как  $17,4 > 9$ , то... .. уравнение значимо.



## 2.12. Матрица корреляций

Наличие линейной связи между факторами А и В показывает коэффициент корреляции

$$r_{AB} = \frac{\overline{AB} - \overline{A} \cdot \overline{B}}{\sqrt{D_A \cdot D_B}}, \quad D_A = \overline{A^2} - \overline{A}^2, \quad D_B = \overline{B^2} - \overline{B}^2$$

Известно, что

$$-1 \leq r_{AB} \leq 1$$

и линейная зависимость между А и В тем сильнее, чем ближе модуль коэффициента корреляции к 1.

В случае, когда рассматривается несколько факторов, попарные коэффициенты корреляции записывают в матрицу

**Пример 3.** Для следующих табличных данных составить матрицу попарных корреляций

y	x <sub>1</sub>	x <sub>2</sub>
7	2	3
14	2	4
11	5	7
12	7	10

**Решение.** 1) Составляем матрицу

$$M = \begin{pmatrix} 7 & 2 & 3 \\ 14 & 2 & 4 \\ 11 & 5 & 7 \\ 12 & 7 & 10 \end{pmatrix}$$

2) Из каждого столбца вычитаем его среднее значение (например, из второго вычитаем 4):

$$M_1 = \begin{pmatrix} -4 & -2 & -3 \\ 3 & -2 & -2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix}$$

3) Умножаем  $M_1^T$  на  $M_1$ , получаем матрицу ковариаций:

$$M_1^T \cdot M_1 = \begin{pmatrix} -4 & 3 & 0 & 1 \\ -2 & -2 & 1 & 3 \\ -3 & -2 & 1 & 4 \end{pmatrix} \cdot \begin{pmatrix} -4 & -2 & -3 \\ 3 & -2 & -2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 26 & 5 & 10 \\ 5 & 18 & 23 \\ 10 & 23 & 30 \end{pmatrix}$$

На диагонали в матрице стоят дисперсии, например  $D_y = 26$ .

4) Делим каждое число на корень из произведения соответствующих дисперсий (например, 5 на корень из  $26 \cdot 18$ , 10 на корень из  $26 \cdot 30$ , получаем матрицу корреляций:

$$r_{yx} = \begin{pmatrix} 1 & 0,2311 & 0,3581 \\ 0,2311 & 1 & 0,9898 \\ 0,3581 & 0,9898 & 1 \end{pmatrix}$$

## 2.13. Частная корреляция

Корреляция между двумя переменными, вычисленная после устранения влияния всех других переменных, называется частной корреляцией.

Например, **длина волос** может коррелировать с **ростом** человека (чем выше человек, тем короче волосы), однако эта зависимость становится слабой или совсем исчезает, если устранить влияние **пола** наблюдаемых людей, поскольку женщины обычно ниже ростом и чаще имеют более длинные волосы, чем мужчины.

**Частный коэффициент корреляции** между переменными X и Y при исключении влияния переменной Z вычисляется по формуле :

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

**Пример.** Для предыдущего примера

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

$$r_{YX} = \begin{pmatrix} 1 & 0,2311 & 0,3581 \\ 0,2311 & 1 & 0,9898 \\ 0,3581 & 0,9898 & 1 \end{pmatrix}$$

$$r_{yx_1|x_2} = \frac{0,2311 - 0,3581 \cdot 0,9898}{\sqrt{(1 - 0,3581^2) \cdot (1 - 0,9898^2)}} = -0,925$$

$$r_{yx_2|x_1} = \frac{0,3581 - 0,9898 \cdot 0,3581}{\sqrt{(1 - 0,3581^2) \cdot (1 - 0,9898^2)}} = 0,9311$$

И так далее...

$$r_{YX}^{частн} = \begin{pmatrix} 1 & -0,925 & 0,9311 \\ -0,925 & 1 & 0,9984 \\ 0,9311 & 0,9984 & 1 \end{pmatrix}$$

## 2.14. Спецификация модели

Спецификация модели включает следующие этапы:

1. Определение списка объясняющих и зависимых переменных.
2. Выбор функциональной формы модели.

Рассмотрим подробнее первый пункт

### 2.14.1. Отбор объясняющих переменных

При определении состава факторов (переменных), включаемых в уравнение регрессии, прежде всего руководствуются **теоретическими представлениями** о взаимосвязях этих факторов.

Однако часто встречается ситуация, когда имеется большое число факторов, **но нет априорной модели изучаемого явления**, так что неясно, какие именно переменные включать в модель. В этом случае используются различные **эвристические процедуры**.

Приведём простой пример такой процедуры

#### Процедура пошагового отбора независимых переменных.

1. Из исходного набора переменных отбирается (включается в модель) переменная, имеющая **наибольший по модулю коэффициент корреляции с зависимой переменной Y**.
2. На каждом последующем шаге в модель **добавляется та** из переменных, добавление которой **максимально увеличивает скорректированный коэффициент детерминации**.
3. Если добавление новых переменных не увеличивает этот коэффициент, процедура считается завершённой

При отборе переменных также часто используется **матрица частных корреляций**.

### 2.14.2. Влияние ошибок спецификации

Неправильная спецификация переменных может произойти из-за

- отбрасывания переменной, которая должна быть в составе объясняющих переменных или
- включения “лишней” переменной, которой не должно быть в составе объясняющих переменных.

Каковы  
последствия  
подобных ошибок?

#### 2.14.2.1. Отбрасывание значимой переменной

Пусть вместо “истинного” уравнения  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$  из-за невключения переменной  $x_2$  взято уравнение  $y = \beta_0 + \beta_1 x_1 + \varepsilon$

В такой ситуации:

- ❖ МНК-оценка параметров  $\beta$  в общем случае оказывается смещённой;
- ❖ Стандартные ошибки коэффициентов  $\beta$  имеют неотрицательное смещение;
- ❖ Оценка дисперсии  $\sigma^2$  имеет неотрицательное смещение

Таким образом, **последствия** достаточно **серьёзны**.

#### 2.14.2.2. Добавление незначимой переменной

Пусть вместо “истинного” уравнения  $y = \beta_0 + \beta_1 x_1 + \varepsilon$  из-за добавления лишней переменной  $x_2$  взято уравнение  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

В такой ситуации:

- ❖ МНК-оценки параметров  $\beta$  несмещённые;
- ❖ Стандартные ошибки коэффициентов  $\beta$  имеют неотрицательное смещение;
- ❖ Оценка дисперсии  $\sigma^2$  имеет неотрицательное смещение

Последствия такой ошибки спецификации являются не столь серьёзными, как в предыдущем случае.

Однако увеличение стандартных ошибок может привести к неверным выводам при проверке гипотез и ухудшению интервальных оценок. В частности, может возникнуть незначимость коэффициентов.

### 2.15. Фиктивные переменные

Как правило, независимые переменные принимают непрерывный ряд значений (цена товара, уровень инфляции и т.д.). Однако нередко приходится принимать во внимание и какие-либо *качественные* признаки.

Например, можно изучать влияние пола работника на уровень его зарплаты (переменная “пол” принимает всего два значения), или спросом на прохладительные напитки в зависимости от времени года (сколько различных значений принимает эта переменная?).

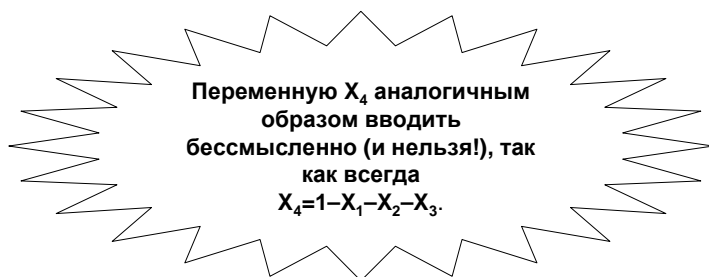
В подобных ситуациях “качественному” значению условно присваивается числовая метка, как правило, это 0 или 1.

Если качественный признак принимает два значения, то это присвоение можно сделать произвольно (например, “наличие детей”=1, “отсутствие детей”=0). Если же качественный признак принимает  $k$  значений ( $k > 2$ ), то вводят  $k-1$  переменную, каждая из которых принимает значения 0 и 1.



**Пример.** Качественный признак – время года, значения: зима, весна, лето, осень.

**Фиктивные переменные:**  
 $X_1=1$ , если зима, иначе  $X_1=0$ ,  
 $X_2=1$ , если весна, иначе  $X_2=0$ ,  
 $X_3=1$ , если лето, иначе  $X_3=0$ .



### 2.16. Тест Чоу

Иногда выборка состоит из нескольких частей и требуется выяснить, следует ли их объединить в одну, или рассматривать каждую подвыборку отдельно.

Например, следует ли рассматривать статистические данные для каждого времени года отдельно, или их можно объединить в общую модель?

Обозначим регрессию по первой выборке (объёма  $n_1$ ) через А, по второй (объёма  $n_2$ ) – через В, а объединённую (объёма  $n=n_1+n_2$ ) – через U. Ясно, что

$$ESS_A + ESS_B \leq ESS_U.$$

Если левая часть неравенства близка к правой, то это говорит о том, что регрессия по объединённой выборке почти так же хороша, как "составная" их двух регрессий по двум выборкам. В этом случае объединение выборок в одну допустимо.

Проверка этой близости составляет суть критерия Чоу.

1) Составляется статистика

$$F = \frac{(ESS_U - ESS_A - ESS_B)/(m+1)}{(ESS_A + ESS_B)/(n-2m-2)}$$

2) Если вычисленное значение F-статистики оказывается меньше  $F_{1-\alpha}(m+1, n-2m-2)$ , то объединение выборок возможно при данном  $\alpha$ .

**Замечание.** Если одна из выборок слишком мала, то применяется другая формула. Пусть, например,  $n_2 \leq m+1$ , тогда

$$F = \frac{(ESS_U - ESS_A)/(n-n_1)}{ESS_A/(n_1-m-1)}$$

Если вычисленное значение F-статистики оказывается меньше  $F_{1-\alpha}(m+1, n_1-m-1)$ , то объединение выборок возможно при данном  $\alpha$ .

**Замечание.** Другой способ проверки состоит в использовании фиктивных переменных.

Добавим в модель фиктивную переменную z, равную 0 для первой выборки и 1 для второй, а также переменные  $zx_1, zx_2, \dots$ . Оценим соответствующее уравнение регрессии. Если эти переменные окажутся незначимыми, то объединение выборок возможно.

**Пример.** Проверить возможность объединения выборок при  $\alpha=0,1$ .

y	$x_1$	$x_2$	y	$x_1$	$x_2$
3	2	3	9	4	7
5	3	4	9	6	8
5	3	6	10	6	9
6	4	1	12	7	9
7	5	7			

**Решение.** 1 способ (критерий Чоу).

Для первой выборки:

$$y = 0,86 + 1,26x_1 + 0,01x_2, ESS = 0,42$$

Для второй выборки:

$$y = 3 + 0,5x_1 + 0,5x_2, ESS = 2,5$$

Для объединённой выборки:  $y = 0,11 + 1,32x_1 + 0,23x_2, ESS = 7,38$

$$F = \frac{(7,38 - 0,42 - 2,5)/(2+1)}{(0,42 + 2,5)/(9 - 2 \cdot 2 - 2)} = 1,527$$

$$F_{0,9}(2+1; 9 - 2 \cdot 2 - 2) = F_{0,9}(3; 3) = 5,4$$

Так как  $1,527 < 5,4$ , то объединение возможно.

### 2 способ (фиктивная переменная).

Добавляем новые переменные  $z, zx_1, zx_2$ .

Проверяем их значимость

$$y = 0,86 + 1,26x_1 + 0,01x_2 + 2,14z + 0,76zx_1 + 0,49zx_2$$

(1,59)
(0,46)
(0,22)  
(6,63)
(1,13)
(0,38)

Все добавленные незначимы

y	x <sub>1</sub>	x <sub>2</sub>	z	zx <sub>1</sub>	zx <sub>2</sub>
3	2	3	0	0	0
5	3	4	0	0	0
5	3	6	0	0	0
6	4	1	0	0	0
7	5	7	0	0	0
9	4	7	1	4	7
9	6	8	1	6	8
10	6	9	1	6	9
12	7	9	1	7	9

Σ Γ ◁

### 3. Нарушения допущений классической линейной модели

Вспомним предположения, лежащие в основе классической модели множественной регрессии:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + \varepsilon_i \quad (2.2)$$

- (A) "Истинная" зависимость Y от регрессоров X имеет вид (2.2).
- (B) Величины X – детерминированные (не случайные).
- (C) Столбцы матрицы X линейно независимы (другими словами, *ранг матрицы X равен m+1*).
- (D)  $M\varepsilon_i = 0$ ,  $D\varepsilon_i = \sigma^2$  не зависит от i
- (E)  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  при  $i \neq j$  (*некоррелированность ошибок для разных наблюдений*).

### 3.1. Мультиколлинеарность

#### 3.1.1. Определение

Одним из условий классической регрессионной модели является предположение (C) о линейной независимости столбцов матрицы X, т.е. о линейной независимости значений объясняющих переменных (включая константу).

Нарушение этого условия – линейная зависимость двух или более объясняющих переменных. В этой ситуации нельзя построить оценку параметров регрессии, что формально видно из обращения в нуль определителя матрицы

$$X^T X,$$

в силу чего оказывается бесполезной формула

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad (2.4)$$

Точная линейная зависимость (например,  $x_1$  – расстояние в метрах,  $x_2$  – в километрах, так что  $x_1 = 1000 \cdot x_2$ ) является скорее теоретическим примером.

На практике более реальна ситуация, когда между объясняющими переменными существует не точная линейная, а сильная корреляционная зависимость.

Например, если в состав объясняющих переменных входят располагаемый доход и потребление, то обе эти переменные будут сильно коррелированными.

В ситуации такой нестрогой мультиколлинеарности коэффициенты уравнения, как правило, оказываются незначимыми и могут даже иметь неверные знаки.

### 3.1.2. Способы обнаружения

Вот некоторые наиболее **характерные признаки мультиколлинеарности**:

- ❖ Небольшое изменение исходных данных приводит к существенному изменению оценок параметров модели.
- ❖ Оценки коэффициентов имеют малую значимость по t-критерию, в то время как модель в целом является значимой по F-критерию.
- ❖ Оценки коэффициентов имеют неправильные с точки зрения теории знаки.

На практике о наличии мультиколлинеарности судят по **матрице парных корреляций** факторов модели, либо по **матрице частных корреляций**. В частности, **близкое к нулю значение определителя** говорит о наличии мультиколлинеарности.

$$\begin{pmatrix} 1 & r_{01} & r_{02} & \dots & r_{0m} \\ r_{10} & 1 & r_{12} & \dots & r_{1m} \\ r_{20} & r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m0} & r_{m1} & r_{m2} & \dots & 1 \end{pmatrix}$$

### 3.1.3. Методы устранения

Если задачей исследования является прогноз будущих значений зависимой переменной, то мультиколлинеарность не является серьезной проблемой, так как она не ухудшает значимость уравнения.

## 3.2. Гетероскедастичность

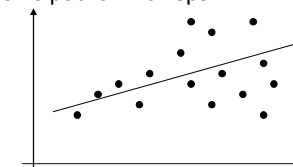
### 3.2.1. Определение

Условие **(D)** теоремы Гаусса-Маркова требует постоянства дисперсии случайного члена  $\varepsilon_i$  во всех наблюдениях:

$$D\varepsilon_i = \sigma^2 \text{ не зависит от } i$$

Это случай **гомоскедастичности**, при которой разброс точек наблюдений по вертикали явно не меняется с ростом номера наблюдений.

**Если же дисперсия случайного члена меняется от наблюдения к наблюдению, то мы имеем дело с гетероскедастичностью**



При **гетероскедастичности** оценки коэффициентов регрессии будут **несмещенными, но неэффективными**. Вследствие этого окажутся завышенными t-статистики оценок коэффициентов, и будут сделаны **ложные выводы о значимости коэффициентов регрессии**.

### 3.2.2. Способы обнаружения

Существуют различные способы выявления гетероскедастичности.

#### 3.2.2.1. Графический метод.

- 1) При помощи метода наименьших квадратов оценивается регрессия  $y=f(x_1, x_2, \dots, x_m)$ .
- 2) Вычисляются остатки  $e_i$ .
- 3) Строится график зависимости остатков  $e_i$  от номера наблюдения.
- 4) Далее анализируется внешний вид этого графика

#### 3.2.2.2. Тест Уайта.

- 1) К исходной модели (объём выборки =  $n$ ) применяется обычный МНК, и находятся остатки  $e_i$ .
- 2) Осуществляется регрессия **квадратов** этих остатков **на все исходные переменные, их квадраты, попарные произведения и константу** (пусть всего получится  $N$  переменных, не считая константы).
- 3) Если для второй регрессии  $n \cdot R^2$  больше  $\chi^2_{1-\alpha}(N)$ , то гетероскедастичность есть.

**Пример.** Проверить на гетероскедастичность при  $\alpha=0,1$ , используя тест Уайта

**Решение.** 1) Построим регрессию и найдём расчётные значения и остатки  $e_i$ .

$$y = 0,77 + 1,15x_1 + 0,22x_2$$

2) Построим регрессию по выборке, описанной в тесте Уайта:

$e_i$	$e_i^2$	$x_1$	$x_2$	$x_1^2$	$x_2^2$	$x_1x_2$
-0,72	0,52	2	3	4	9	6
0,13	0,017	3	3	9	9	9
-0,52	0,27	3	6	9	36	18
0,41	0,17	4	1	16	1	4
-1,03	1,06	5	7	25	49	35
2,12	4,5	4	7	16	49	28
-0,4	0,16	6	8	36	64	48

$\hat{y}$	$e = y - \hat{y}$	$y$	$x_1$	$x_2$
3,72	-0,72	3	2	3
4,87	0,13	5	3	4
5,52	-0,52	5	3	6
5,59	0,41	6	4	1
8,03	-1,03	7	5	7
6,99	2,12	9	4	7
9,4	-0,4	9	6	8

3) Для этой регрессии вычислим:

$$nR^2 = 7 \cdot 0,4835 = 3,3845,$$

$$\chi^2_{1-\alpha}(N) = \chi^2_{0,9}(5) = 9,236$$

$3,3845 < 9,236$ , так что гетероскедастичности нет.

#### 3.2.2.3. Тест Голдфелда-Квандта.

В этом тесте предполагается, что стандартное отклонение  $\sigma_i$  пропорционально значению некоторой независимой переменной  $X_i$ .

- 1) Наблюдения упорядочиваются по возрастанию значений этой переменной  $x_i$ .
- 2) Берутся первые  $n'$  и последние  $n'$  наблюдений, и для них по отдельности строятся регрессии, и оцениваются суммы квадратов остатков  $ESS_1$  и  $ESS_2$  соответственно. Число  $n'$  можно взять примерно равным  $n/3$ .
- 3) Составляется статистика
 
$$\frac{ESS_2}{ESS_1} \sim F(n' - m - 1; n' - m - 1),$$
 где  $m$  – число объясняющих переменных в исходном уравнении регрессии.
- 4) Большое значение этой статистики (больше соответствующего квантиля  $F_{1-\alpha}(n'-m-1, n'-m-1)$ ) означает **наличие гетероскедастичности**.

**Пример.** Проверить на гетероскедастичность по переменной  $x_1$  при  $\alpha=0,1$ , используя тест Голдфелда-Квандта.

**Решение.** 1) Объём выборки=12, так что выберем  $n'=12/3=4$ . Отсортируем выборку по возрастанию  $x_1$  и построим две регрессии (по первым и последним четырём строкам для  $x_1$ ).

$$y = -3,37 + 2,79x_1 + 0,11x_2, \quad ESS_1 = 0,95$$

$$y = -0,67 - 13,67x_1 + 32x_2, \quad ESS_2 = 112,7$$

2)

$$\frac{ESS_2}{ESS_1} = \frac{112,7}{0,95} \approx 118$$

$$F_{1-\alpha}(n' - m - 1; n' - m - 1) = F_{0,9}(4 - 2 - 1; 4 - 2 - 1) = 40,$$

$118 > 40$ , так что гетероскедастичность в виде пропорциональности  $\sigma$  и  $x_1$  есть.

y	$x_1$	$x_2$
3	2	3
5	3	4
5	3	6
6	6	1
7	5	7
9	4	7
9	6	8
11	6	7
18	8	4
32	9	5
68	12	7
78	15	9

### 3.2.2.4. Тест Глейзера.

- 1) Оценивается регрессионная зависимость  $|e_i| = a + bx_i^k + u_i$
- 2) Параметр  $k$  изменяется с некоторым шагом, и для каждого его значения строится регрессия.
- 3) Статистическая значимость коэффициента  $b$  при каком-либо значении означает наличие гетероскедастичности

### 3.2.3. Методы устранения

Рассмотрим случай, когда стандартное отклонение  $\sigma_i$  пропорционально значению некоторой независимой переменной  $X_k$ . Разделив уравнение регрессии

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon$$

на  $X_k$ , запишем его в виде

$$\frac{y}{x_k} = \frac{\beta_0}{x_k} + \frac{\beta_1 x_1}{x_k} + \dots + 1 + \dots + \frac{\beta_m x_m}{x_k} + \frac{\varepsilon}{x_k}$$

Если теперь перейти к новым переменным

$$\tilde{y} = \frac{y}{x_k}; \quad \tilde{\varepsilon} = \frac{\varepsilon}{x_k}; \quad \tilde{x}_i = \dots; \quad \tilde{\beta}_i = \dots$$

то оно примет вид

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_1 + \dots + \tilde{\beta}_m \tilde{x}_m + \tilde{\varepsilon}$$

причем

$$D(\tilde{\varepsilon}) = const$$

Аналогичные преобразования можно применять и при других видах зависимости  $\sigma_i$  от переменных задачи.

**Пример.** В предыдущем примере мы выяснили наличие гетероскедастичности в виде пропорциональности  $\sigma$  переменной  $x_1$

Если использовать обычный МНК, то получится

$$y = -23,76 + 6x_1 + 0,92x_2$$

Но если учесть гетероскедастичность указанного вида, то более правильно, **разделив всё уравнение на  $x_1$** , ввести новые переменные

$$\tilde{y} = \frac{y}{x_1}; \quad \tilde{x}_1 = \frac{1}{x_1}; \quad \tilde{x}_2 = \frac{x_2}{x_1}$$

Получим таблицу

$\tilde{y}$	$\tilde{x}_1$	$\tilde{x}_2$
3/2	1/2	3/2
5/3	1/3	4/3
5/3	1/3	6/3
6/6	1/6	1/6
...	...	...

Из которой

$$\tilde{y} = 3,99 - 6,51\tilde{x}_1 - 0,16\tilde{x}_2,$$

так что

$$y = -6,51 + 3,99x_1 - 0,16x_2$$

y	$x_1$	$x_2$
3	2	3
5	3	4
5	3	6
6	6	1
7	5	7
9	4	7
9	6	8
11	6	7
18	8	4
32	9	5
68	12	7
78	15	9

### 3.3. Автокорреляция

#### 3.3.1. Определение

Условие **(E)** теоремы Гаусса-Маркова требует отсутствия корреляции между значениями случайного фактора  $\varepsilon_i$  во всех наблюдениях:

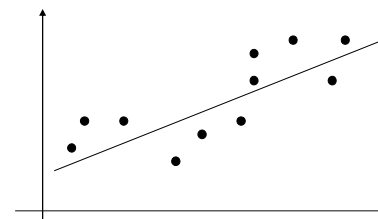
$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ при } i \neq j$$

Фактически это условие соответствует независимости случайных членов в разных наблюдениях.

При нарушении этого условия, т.е. **при наличии связи между случайными членами** возникает явление **автокорреляции**.

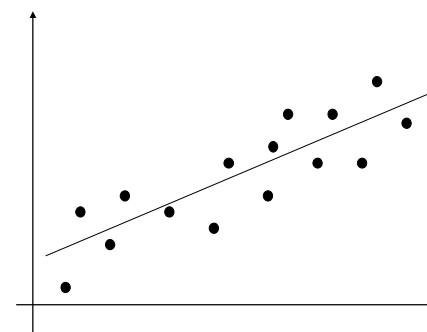
Обычно автокорреляция рассматривается при изучении **временных рядов** (т.е. когда номер наблюдения соответствует моменту времени).

В случае **положительной автокорреляции** реализации случайного члена  $\varepsilon_i$  для ряда последовательных наблюдений смещают значения зависимой переменной в одном направлении, затем для нескольких последовательных наблюдений – в противоположном направлении, потом снова в первоначальном направлении и т.д:



В экономике положительная автокорреляция может быть связана с циклами деловой активности, с сезонными колебаниями, влияние которых и отражается в случайном члене уравнения регрессии

При **отрицательной автокорреляции** каждая реализация случайного члена  $\varepsilon_i$ , как правило, сменяется реализацией случайного члена  $\varepsilon_{i+1}$  противоположного знака:



### 3.3.2. Способы обнаружения

Существуют различные способы выявления автокорреляции.

#### 3.3.2.1. Графический метод.

- 1) При помощи метода наименьших квадратов оценивается регрессия  $y=f(x_1, x_2, \dots, x_m)$ .
- 2) Вычисляются остатки  $e_i$ .
- 3) Строится график зависимости остатков  $e_i$  от номера наблюдения.
- 4) Далее анализируется внешний вид этого графика

#### 3.3.2.2. Метод рядов.

В этом методе рассматриваются **знаки отклонений  $e_i$**   
“Ряд” это **непрерывная последовательность одинаковых знаков.**

При достаточно большом числе наблюдений для определения автокорреляции применяется следующая процедура.

- 1) Рассчитываются статистики.

$$M = \frac{2n_+n_-}{n_+ + n_-} + 1, \quad D = \frac{2n_+n_-}{(n_+ + n_-)^2} (2n_+n_- - n_+ - n_-)$$

где  $n_+$  - количество положительных отклонений,  
 $n_-$  - количество отрицательных.

- 2) Если

$$M - u_{1-\alpha/2} \cdot D < k < M + u_{1-\alpha/2} \cdot D$$

где  $k$  – количество рядов, то принимается гипотеза об отсутствии автокорреляции .

Для **малых** объемов выборки есть таблицы критических значений  $k$  (пример для  $\alpha=0,05$ ):

$n_+ \setminus n_-$	5	6	7	8	...
3	–	2–	2–	2–	...
4	2–9	2–9	2–	3–	...
5	2–10	3–10	3–11	3–11	...
6		3–11	3–12	3–12	...
7			3–13	4–13	...
8				4–14	
...	...	...	...	...	...

**Пример.** Рассмотрим ситуацию, когда знаки отклонений имеют вид:  
(---)(+++)(---)(+++)  
Здесь  $k=4$ ,  $n_+=7$ ,  $n_-=8$ . По таблице находим интервал 4–13.  
Число  $k$  попадает в эти пределы, так что на уровне  $\alpha=0,05$  гипотеза об отсутствии автокорреляции принимается.

#### 3.3.2.3. Критерий Дарбина-Уотсона.

Этот метод применяется для обнаружения автокорреляции вида

$$\varepsilon_i = \rho \cdot \varepsilon_{i-1} + u_i, \quad (3.1)$$

где  $u_i$ – случайные попарно независимые величины,  
 $\rho$  – некоторый параметр, называемый **коэффициентом авторегрессии** ( $|\rho|<1$ ).

Такая ситуация называется **авторегрессионным процессом первого порядка.**

Случайный член в данном наблюдении прямо связан лишь со случайным членом в предыдущем наблюдении.

Если  $\rho>0$ , то автокорреляция **положительная**,  
если  $\rho<0$ , то **отрицательная.**

В критерии Дарбина-Уотсона проверяется гипотеза  $\rho=0$  при альтернативной  $\rho \neq 0$ . Для обнаружения автокорреляции используется статистика Дарбина-Уотсона:

$$DW = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}, \quad e_i = y_i - \hat{y}_i \quad (3.2)$$

Для больших выборок  $DW \approx 2 \cdot (1 - \rho)$ .  
 При значительной положительной автокорреляции  $\rho \approx 1$  и  $DW \approx 0$ .  
 При отрицательной автокорреляции  $\rho \approx -1$  и  $DW \approx 4$ .  
 При отсутствии автокорреляции автокорреляции  $\rho \approx 0$  и  $DW \approx 2$ .

Критические значения статистики Дарбина-Уотсона зависят не только от числа объясняющих переменных, но и от значений, которые они принимают в выборке. Поэтому в отличие от t-статистики и от F-статистики **невозможно составить таблицы критических значений DW** статистики Дарбина-Уотсона.

Но можно указать верхнюю  $d_U$  и нижнюю  $d_L$  границы для  $d$ . Они определяются в зависимости от  $n$  и числа оцениваемых параметров  $m$ .

0	+	$d_L$	?	$d_U$	нет	2	нет	$4 - d_u$	?	$4 - d_L$	-	4
---	---	-------	---	-------	-----	---	-----	-----------	---	-----------	---	---

При  $0 \leq DW \leq d_L$  принимается гипотеза  $\rho > 0$  (положительная автокорреляция),  
 При  $4 - d_U \leq DW \leq 4$  принимается гипотеза  $\rho < 0$  (отрицательная автокорреляция).  
 Если  $DW$  попадает в интервал  $(d_U, 4 - d_U)$ , то принимается гипотеза об **отсутствии** авторегрессии первого порядка.

Критические точки  $d_L, d_U$  распределения Дарбина-Уотсона при  $\alpha=0,05$

$m$ $\downarrow$	$n \rightarrow$	7	8	9	10	20	30
1	$d_L$	0,7	0,76	0,82	0,88	1,2	1,35
1	$d_U$	1,356	1,33	1,32	1,32	1,41	1,49
2	$d_L$	0,47	0,36	0,63	0,7	1,1	1,29
2	$d_U$	1,9	1,78	1,7	1,64	1,54	1,57
3	$d_L$		0,37	0,44	0,53	1	1,21
3	$d_U$		2,29	2,13	2,02	1,68	1,65
4	$d_L$			0,3	0,38	0,9	1,14
4	$d_U$			2,39	2,41	1,83	1,74

$\Sigma \Gamma \Gamma \Sigma$

**Пример.** Построена регрессия с  $n=8, m=2$ . Используя тест Дарбина-Уотсона, проверить гипотезу о наличии автокорреляции первого порядка при  $\alpha=0,05$

**Решение.** Заполним таблицу

$e_i = y_i - \hat{y}_i$	$e_{i-1}$	$(e_i - e_{i-1})^2$	$e_i^2$
-0,09	-	-	0,0081
-0,32	-0,09	0,17	0,1024
-0,79	-0,32	1,23	0,6241
-0,13	-0,79	0,44	0,0169
-1,43	-0,13	1,69	2,0449
1,61	-1,43	9,24	2,5921
-1,02	1,61	6,92	1,0404
1,54	-1,02	6,55	2,3716
$\Sigma$		= 26,2	= 8,8

$y$	$\hat{y}$
3	3,09
5	4,68
5	5,79
6	6,13
7	8,43
9	7,39
9	10,02
11	9,46

$$DW = \frac{26,2}{8,8} \approx 3$$

По таблице  $d_L=0,36, d_U=1,78$

Число  $DW=3$  попадает в зону  $(4-d_U; 4-d_L)=(2,22; 3,64)$

Поэтому критерий Дарбина-Уотсона в этом примере ответа не даёт



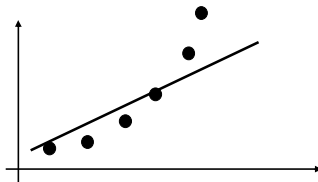
### 3.3.3. Методы устранения

#### 3.3.3.1. Исправление спецификации модели

Основной причиной наличия случайного фактора является несовершенство знаний о причинах и взаимосвязях, определяющих то или иное значение зависимой переменной. Поэтому свойства случайного члена зависят от выбора формы зависимости и состава объясняющих переменных.

Другими словами, автокорреляция чаще всего вызывается ошибками в спецификации модели.

Например, при выборе линейной формы зависимости в ситуации, когда имеет место экспоненциальная, возникает положительная автокорреляция:



#### 3.3.3.2. Авторегрессионное преобразование

Рассмотрим случай **парной** линейной регрессии.

Если имеет место зависимость (3.1)

$$\varepsilon_i = \rho \cdot \varepsilon_{i-1} + u_i, \quad (3.1)$$

и коэффициент  $\rho$  известен, то для исходного уравнения регрессии

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

можно выполнить **авторегрессионное преобразование**.

Запишем это уравнение для предыдущего номера:

$$y_{i-1} = \beta_0 + \beta_1 x_{i-1} + \varepsilon_{i-1},$$

умножим обе его части на  $\rho$  и вычтем из исходного уравнения:

$$y_i - \rho y_{i-1} = \beta_0 \cdot (1 - \rho) + \beta_1 \cdot (x_i - \rho \cdot x_{i-1}) + u_i$$

Если ввести новые переменные

$$y_i^* = y_i - \rho \cdot y_{i-1}, \quad x_i^* = x_i - \rho \cdot x_{i-1}, \quad \beta_0^* = \beta_0 \cdot (1 - \rho),$$

то получим **уравнение без автокорреляции**:

$$y_i^* = \beta_0^* + \beta_1 x_i^* + u_i.$$

Найдя оценки коэффициентов этого уравнения, мы тем самым сможем вычислить и оценки коэффициентов исходного.

Однако **проблема** заключается в том, что **величина  $\rho$  неизвестна**.

Поэтому используют разные методы получения **оценки для  $\rho$**

❖ В качестве оценки для  $\rho$  часто используют  $1 - DW/2$ .

Другой метод оценки используется в **процедуре Кохрейна-Оркатта**. Она представляет собой итеративный процесс определения параметров  $\beta_0, \beta_1, \rho$ :

- 1) Методом наименьших квадратов находится оценка  $\rho_1$  параметра  $\rho$  регрессии  $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + u_i$ .
- 2) Проводится авторегрессионное преобразование с этим значением параметра  $\rho$  и находятся МНК-оценки параметров  $\beta_i$ .
- 3) Строится новый вектор остатков  $e_i$  и процедура повторяется.
- 4) Процесс заканчивается, когда очередное приближение  $\rho$  мало отличается от предыдущего

**Метод Хилдрета-Лу** использует перебор по значениям  $\rho$  из интервала  $(-1, 1)$  с некоторыми шагом при оценивании  $\beta_0, \beta_1$  по преобразованному уравнению.

В процессе перебора определяется то значение  $\rho$ , при котором сумма квадратов отклонений ESS для преобразованного уравнения минимальна. Вблизи найденного минимума процедура может повторяться с более мелкими шагами до достижения заданной точности.

### 3.4. Стохастические объясняющие переменные

До сих пор мы предполагали, что объясняющие переменные  $x_1, x_2, \dots, x_m$  являются неслучайными, детерминированными (в отличие от  $y$ ).

Это предположение позволяло упростить анализ модели.

Но в некоторых случаях значения  $X$  приходится считать случайными. Выясним, к чему это может привести.

Пусть, как и раньше,

$$Y = X \cdot \beta + \varepsilon$$

Будем считать элементы матрицы  $X$  случайными величинами. Потребуем, чтобы были выполнены условия

- (C<sub>1</sub>) Ранг матрицы  $X$  равен  $m+1$  с вероятностью единица.  
 (D<sub>1</sub>)  $M(\varepsilon|X) = 0, M(\varepsilon \varepsilon^T|X) = \sigma^2 E$

Здесь  $M(\dots|X)$  – условные математические ожидания.

**Теорема.** В этих предположениях

- ✓ МНК-оценки параметров регрессии являются несмещёнными.
- ✓ Среди всех линейных условно несмещённых оценок МНК-оценка обладает наименьшей условной дисперсией.
- ✓ Если в каждом наблюдении значения объясняющих переменных (регрессоров) выбираются из одной и той же генеральной совокупности, а ошибки независимы, одинаково распределены и не зависят от регрессоров, то МНК-оценки состоятельны.

Если же объясняющие переменные  $X$  и ошибки  $\varepsilon$  коррелированы (т.е.  $M(\varepsilon|X) \neq 0$ ), то в общем случае МНК-оценка будет смещённой и несостоятельной. В этом случае можно попробовать применить метод инструментальных переменных.

### 3.5. Метод инструментальных переменных

Суть метода состоит в замене объясняющей переменной  $X$ , коррелированной с  $\varepsilon$  (т.е.  $M(\varepsilon|X) \neq 0$ ), на другую переменную, которая не коррелирована с  $\varepsilon$ .

Рассмотрим пример

$$Y = X \cdot \beta + \varepsilon$$

Допустим, что есть переменная  $Z$ , тесно связанная (коррелированная) с  $X$ , но не коррелированная с  $\varepsilon$ . Обозначим

$$\hat{\beta}_{IP} = (Z^T X)^{-1} Z^T Y$$

Можно показать, что с при “хорошей” корреляции между  $X$  и  $Z$  такая оценка является состоятельной, т.е. с ростом объема выборки эта оценка параметров  $\beta$  стремится к истинным значениям параметров.

### 3.6. Нелинейные формы зависимости

Использование лишь линейных зависимостей для описания экономических взаимосвязей часто оказывается недостаточным. Необходимо использовать и нелинейные соотношения.

Например, производственная функция Кобба-Дугласа:

$$Y = A \cdot K^\alpha \cdot L^\beta$$

где Y – выпуск, K – затраты капитала, L – затраты труда, A, α, β, – параметры.

Или производственная функция с постоянной эластичностью замещения CES:

$$Y = a \cdot [\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{-\frac{\eta}{\rho}}$$

с параметрами α, ρ, δ, η.

(Она более реалистична, чем функция Кобба-Дугласа, так как не позволяет сколь угодно заменять капитал трудом и наоборот.)

В некоторых случаях **нелинейные** зависимости путем замены переменных можно преобразовать **к линейному виду**.

Пример (гиперболическая зависимость).

По исходным данным оценить коэффициенты в формуле

$$y = a + \frac{b}{x} + \varepsilon$$

Решение. Заменой  $z = 1/x$  зависимость приводится к линейному виду  $y = a + bz + \varepsilon$ .

y	z = 1/x
5	0,5
3	0,3333
3	0,2
2	0,2
2	0,14286
1	0,125

Далее как обычно:

$$Y = \begin{pmatrix} 5 \\ 3 \\ 3 \\ 2 \\ 2 \\ 1 \end{pmatrix}, Z = \begin{pmatrix} 1 & 0,5 \\ 1 & 0,33.. \\ 1 & 0,2 \\ \vdots & \vdots \end{pmatrix}, Z^T Z = \dots, \dots$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = (Z^T Z)^{-1} Z^T Y = \begin{pmatrix} 0,43 \\ 8,94 \end{pmatrix}$$

Ответ.  
 $y = 0,43 + \frac{8,94}{x} + \varepsilon$

y	x
5	2
3	3
3	5
2	5
2	7
1	8

Пример (полиномиальная зависимость). По исходным данным оценить коэффициенты в формуле

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

Решение. Заменой  $z = x^2$  зависимость приводится к линейному виду  $y = a + bx + cz + \varepsilon$ .

y	x	z = x <sup>2</sup>
5	2	4
3	3	9
3	5	25
2	5	25
2	7	49
1	8	64

Далее как обычно:

$$Y = \begin{pmatrix} 5 \\ 3 \\ 3 \\ 2 \\ 2 \\ 1 \end{pmatrix}, X = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ \vdots & \vdots & \vdots \end{pmatrix}, X^T X = \dots, \dots$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 6,54 \\ -1,11 \\ 0,06 \end{pmatrix}$$

Ответ.  
 $y = 6,54 - 1,11x + 0,06x^2 + \varepsilon$

y	x
5	2
3	3
3	5
2	5
2	7
1	8

Пример (логистическая кривая). По исходным данным оценить коэффициенты в формуле

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

Решение. Заменой  $\tilde{y} = 1/y$ ,  $\tilde{x} = e^{-x}$  зависимость приводится к линейному виду:  $\tilde{y} = a + b\tilde{x} + \varepsilon$

Далее как обычно:

$\tilde{y} = 1/y$	$\tilde{x} = e^{-x}$
0,62	0,37
0,56	0,14
0,52	0,14
0,52	0,14
0,52	0,05
0,49	0,02

$$Y = \begin{pmatrix} 0,62 \\ 0,56 \\ 0,52 \\ \vdots \end{pmatrix}, X = \begin{pmatrix} 1 & 0,37 \\ 1 & 0,14 \\ 1 & 0,14 \\ \vdots & \vdots \end{pmatrix}, X^T X = \dots, \dots$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 0,49 \\ 0,35 \end{pmatrix}$$

Ответ.  $y = \frac{1}{0,49 + 0,35e^{-x} + \varepsilon}$

y	x
1,61	1
1,78	2
1,92	2
1,90	3
2,06	4

Пример (степенная зависимость). К линейному виду можно привести степенную зависимость:

$$y = a \cdot x_1^{b_1} x_2^{b_2} \cdot \dots \cdot x_m^{b_m} \cdot (1 + \varepsilon)$$

где  $a, b_1, \dots, b_m$  – параметры, а  $\varepsilon$  – случайный множитель. Прологарифмируем это соотношение

$$\ln y = \ln a + b_1 \ln x_1 + \dots + b_m \ln x_m + \ln(1 + \varepsilon)$$

Если теперь обозначить

$$\tilde{y} = \ln y; \beta_0 = \ln a; \tilde{x}_i = \ln x_i; \tilde{\varepsilon} = \ln(1 + \varepsilon)$$

то оно примет вид линейной регрессионной модели

$$\tilde{y} = \beta_0 + \beta_1 \tilde{x}_1 + \dots + \tilde{\varepsilon}$$

**Пример.** По исходным данным оценить коэффициенты  $A, \alpha, \beta$  в формуле Кобба=Дугласа

$$Y = A \cdot K^\alpha \cdot L^\beta \cdot (1 + \varepsilon)$$

**Решение.** Логарифмированием зависимость приводится к линейному виду

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln(1 + \varepsilon).$$

Далее как обычно:

$\ln Y$	$\ln K$	$\ln L$
3,53	0,69	1,39
3,96	1,1	1,79
4,39	1,61	1,95
4,39	1,61	1,95
4,73	1,95	2,2
4,57	2,2	1,39

$$\ln Y = \begin{pmatrix} 3,53 \\ 3,96 \\ 4,39 \\ \vdots \end{pmatrix}, X = \begin{pmatrix} 1 & 0,69 & 1,39 \\ 1 & 1,1 & 1,79 \\ 1 & 1,61 & 1,95 \\ \vdots & \vdots & \vdots \end{pmatrix}, \dots$$

Y	K	L
34	2	4
53	3	6
81	5	7
81	5	7
113	7	9
97	9	4

$$\begin{pmatrix} \ln A \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2,48 \\ 0,69 \\ 0,41 \end{pmatrix} \quad A = e^{2,48} \approx 12$$

**Ответ.**  
 $Y = 12 \cdot K^{0,69} \cdot L^{0,41} \cdot (1 + \varepsilon)$

### 3.7. Нелинейные модели, не приводимые к линейному виду

Всё же далеко не всякую зависимость заменой переменных можно привести к линейному виду.

**Например,** если искать логистическую зависимость в виде

$$y = \frac{1}{a + b e^{-x}} + \varepsilon$$

вместо

$$y = \frac{1}{a + b e^{-x} + \varepsilon}$$

то никакими переобозначениями **к линейному виду её не привести.**

Напрашивающаяся замена

$$\tilde{y} = \frac{1}{y - \varepsilon}, \tilde{x} = e^{-x}$$

ничего не даст, так как  $\varepsilon$  неизвестна и мы не сможем заполнить таблицу значений для нового  $y$

**Другой пример:** степенная зависимость вида

$$y = a x^b \cdot (1 + \varepsilon)$$

легко приводится к линейной логарифмированием:

$$\ln y = \ln a + b \ln x + \ln(1 + \varepsilon),$$

а степенную зависимость вида

$$y = a x^b + \varepsilon$$

к линейной привести **не удастся.**

**Вот ещё примеры нелинеаризуемых зависимостей:**

$$y = a + b x^c + \varepsilon, \quad y = a e^{bx} + \varepsilon, \dots$$

В общем случае уравнение нелинейной регрессии с аддитивным случайным членом  $\varepsilon$  имеет вид

$$y = f(x_1, \dots, x_m, \beta_1, \dots, \beta_m) + \varepsilon,$$

где  $y$  – зависимая переменная,  $x_1, \dots, x_m$  – объясняющие переменные,  $\beta_1, \dots, \beta_m$  – параметры модели.

Для нахождения оценок этих параметров можно использовать, как и в линейном случае, **метод наименьших квадратов**

$$ESS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

ESS является функцией  $m$  переменных  $\beta_1, \dots, \beta_m$ .  
Задача минимизации решается при помощи итеративных методов оптимизации, например, методом Ньютона.

**Пример.** По исходным данным оценить коэффициенты  $a, b$  в формуле

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

y	x
2	1
9	2
47	3

**Решение.** Задача не сводится к линейной. Выпишем формулу для ESS

$$ESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (2 - ae^b)^2 + (9 - ae^{2b})^2 + (47 - ae^{3b})^2.$$

Требуется тем или иным способом найти минимум этой функции

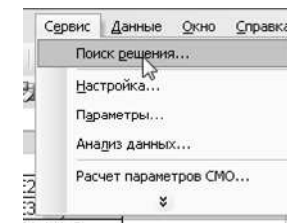
Используем надстройку **Поиск решения** из MS Excel

Внесём данные и формулы в таблицу:

	A	B	C	D	E	F
1	Y	X	a	b	F(x)	
2	2	1	0	0	=\$C\$2*EXP(\$D\$2*B2)	=(E2-A2)*2
3	9	2			=\$C\$2*EXP(\$D\$2*B3)	=(E3-A3)*2
4	47	3			=\$C\$2*EXP(\$D\$2*B4)	=(E4-A4)*2
5					ESS=CYMM(F2:F4)	

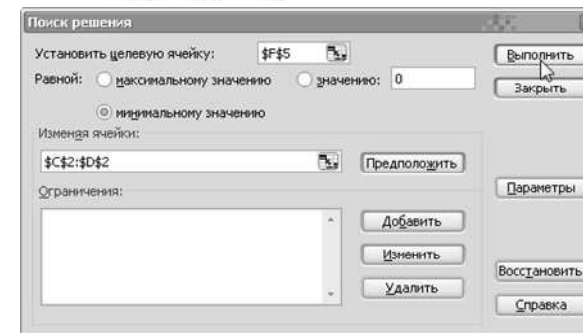
Начальные значения  $a$  и  $b$  положены равными нулю

Вызовем надстройку **Поиск решения**



и введём данные в диалоге

	A	B	C	D	E	F
1	Y	X	a	b	F(x)	
2	2	1	0	0	=\$C\$2*EXP(\$D\$2*B2)	=(E2-A2)*2
3	9	2			=\$C\$2*EXP(\$D\$2*B3)	=(E3-A3)*2
4	47	3			=\$C\$2*EXP(\$D\$2*B4)	=(E4-A4)*2
5					ESS=CYMM(F2:F4)	



Получим результат:

	A	B	C	D	E	F
1	Y	X	a	b	F(x)	
2	2	1	0,341	1,642	1,76	0,06
3	9	2			9,09	0,01
4	47	3			46,99	0,00
5					ESS	0,07

Таким образом,  $a=0,341, b=1,642,$

$$y = 0,341 \cdot e^{1,642x} + \varepsilon$$



## 4. Временные ряды.

### 4.1. Введение.

Временной ряд – это набор наблюдений какой-либо случайной величины, произведённых в последовательные моменты времени.



Это могут быть, например, цены на батон хлеба в соседнем магазине, курс обмена доллара на рубли в ближайшем обменном пункте или годовые объёмы добычи нефти странами ОПЕК.

Анализ временных рядов – это раздел эконометрики, идеей которого является попытка представить будущие значения какой либо величины только на основе её значения в прошлом (конечно, с учётом случайных ошибок).

Этот раздел эконометрики *существенно отличается* от изучаемых нами ранее.

В предыдущих разделах значения *результатирующей переменной* выводились из значений одной или нескольких *объясняющих переменных*. Тем самым считалось, что объясняющая переменная может быть выведена из нескольких значимых факторов и у нас есть возможность выделить их и учесть влияние каждого.



Анализ временных рядов основан на другой идее: результирующая переменная складывается под влиянием большого числа факторов, многие из которых не поддаются непосредственному измерению. Поэтому лучшим источником информации о совокупности этих факторов являются значения самой исследуемой переменной в предыдущие моменты времени.

### 4.2. Одномерные и многомерные ряды.

- ✓ Одномерный временной ряд состоит из наблюдений  $y(t_1), y(t_2), \dots, y(t_N)$  анализируемой случайной величины, проведенных в последовательные моменты времени  $t_1, t_2, \dots, t_N$ .
- ✓ Если нас интересует эволюция значений сразу нескольких признаков данного объекта (т.е. одна строка матрицы исходных статистических данных), то мы имеем дело с многомерным временным рядом.

### 4.3. Зависимость от времени.

Чаще всего рассматривают одномерные временные ряды с равноотстоящими моментами наблюдений

$$t_2 - t_1 = t_3 - t_2 = \dots = t_N - t_{N-1} = \Delta.$$

где  $\Delta$ -заданный временной интервал (минута, час, сутки, неделя, месяц, квартал, год и т.п.)



В таком случае бывает удобнее писать  $y(1), y(2) \dots$  вместо  $y(t_1), y(t_2), \dots$ , где  $y(k)$  - значение анализируемого показателя, зарегистрированное в момент  $t_k$

Определение временного ряда опирается на понятие случайной величины  $y(t)$ , зависящей от параметра  $t$  (время).

Это значит, что её закон распределения вероятностей также может зависеть от времени  $t$ .

#### 4.4. Отличия временного ряда от случайной выборки.

Временной ряд имеет особенности, отличающие его от перекрестных данных, поэтому временные ряды требуют методов анализа, существенно отличающихся от рассмотренных ранее.

- 1. В отличие от элементов случайной выборки члены временного ряда не являются статистически независимыми.
- 2. Члены временного ряда не обязательно являются одинаково распределенными.

Это представляет некоторые трудности, но одновременно позволяет делать прогноз!

#### Пример.

Рассмотрим простейшую модель, в которой значение переменной  $y$  в некоторый момент времени  $t$  выражается через её значение в предыдущий момент плюс случайный фактор:

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

Здесь роль объясняющей переменной играет  $y_{t-1}$ . Но тогда все  $y$  выражаются через  $y_0$  и случайные факторы:

$$y_1 = \beta y_0 + \varepsilon_1, \quad y_2 = \beta y_1 + \varepsilon_2 = \beta^2 y_0 + \beta \varepsilon_1 + \varepsilon_2, \\ y_3 = \beta y_2 + \varepsilon_3 = \beta^3 y_0 + \beta^2 \varepsilon_1 + \beta \varepsilon_2 + \varepsilon_3, \dots$$

Отсюда видно, что **нарушается** важнейшее **условие независимости объясняющей переменной от случайного фактора**.

Поэтому к линейной модели временного ряда нельзя применять обычный метод наименьших квадратов, иначе оценки коэффициентов получатся смещёнными



#### 4.5. Области практического применения временных рядов

Основное прикладное значение временных рядов состоит в прогнозировании значений экономических показателей.

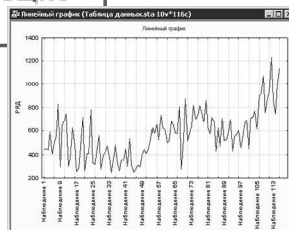


Соответствующие задачи возникают в таких областях, как

- Планирование в производстве и торговле;
- Управление и оптимизация социально-экономических процессов в обществе;
- Частичное управление демографическими процессами;
- Принятие решений в бизнесе и т.п.

#### 4.6. Основные факторы, формирующие временной ряд

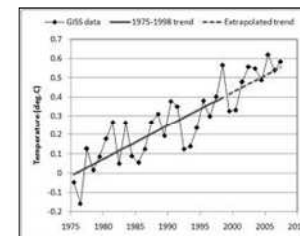
Обычно выделяют четыре типа факторов, под влиянием которых формируются значения временного ряда



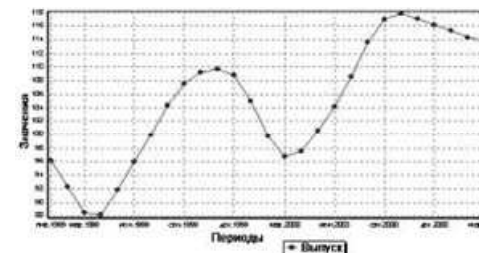
1. **Долговременные** – формируют общую тенденцию изменений в длительной перспективе. Функция, описывающая эту тенденцию, называется **трендом**.
2. **Сезонные** – формируют периодические, повторяющиеся в определённое время года колебания изменяемого показателя.
3. **Циклические** – формируют изменения показателя под влиянием действия циклов экономической, демографической и другой природы.
4. **Случайные** – эти факторы не поддаются учёту.

Рассмотрим немного подробнее соответствующие функции.

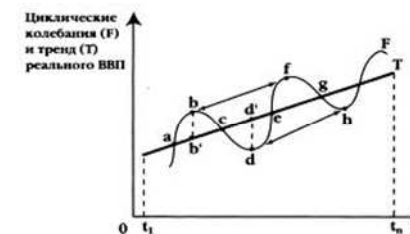
Функция тренда  $f(t)$ , как правило, монотонная.



Результат действия сезонных факторов описывает периодическая функция  $\varphi(t)$ . Поскольку она должна быть периодической, в ее аналитическом выражении участвуют гармоники (тригонометрические функции).



Результат действия циклических факторов также описывает периодическая функция  $\psi(t)$ .



Результат воздействия случайных факторов обозначается с помощью случайных величин  $\epsilon(t)$ .

Не обязательно, чтобы в процессе формирования значений временного ряда участвовали одновременно факторы всех четырех типов. Но будем считать, что случайные факторы присутствуют всегда (иначе неинтересно).



- Таким образом, на временной ряд могут оказывать влияние долговременные, сезонные, циклические и случайные факторы.
- Случайные факторы присутствуют всегда, остальные - не во всяком временном ряду.
- Выводы о том, участвуют или нет факторы данного типа в формировании значений  $y(t)$ , могут базироваться как на анализе содержательной сущности задачи, так и на специальном статистическом анализе исследуемого временного ряда.

**Пример.** В таблице приведены данные о суммарных месячных расстояниях  $y(t)$  (в тысячах миль), пройденных британскими авиалайнерами за 96 месяцев с января 1963 г. по декабрь 1970 г. (т. е.  $t=1, 2, \dots, 96$ ), временной интервал  $\Delta$  равен месяцу

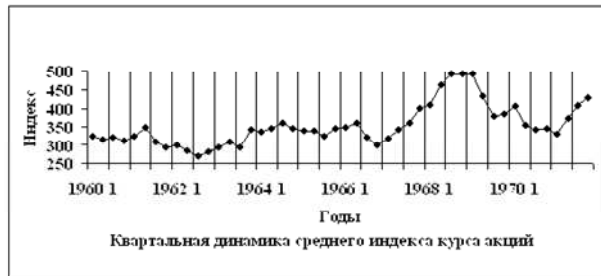
Расстояния, пройденные британскими авиалайнерами за месяц (тыс. миль)								
Месяц \ Год	1963	1964	1965	1966	1967	1968	1969	1970
Январь	6827	7269	8350	8186	8334	8639	9491	10840
Февраль	6178	6775	7829	7444	7899	8772	8919	10436
Март	7084	7819	8829	8484	9994	10894	11607	13589
Апрель	8162	8371	9948	9864	10078	10455	8852	13402
Май	8462	9069	10638	10252	10801	11179	12537	13103
Июнь	9644	10248	11253	12282	12950	10588	14759	14933
Июль	10466	11030	11424	11637	12222	10794	13667	14147
Август	10748	10882	11391	11577	12246	12770	13731	14057
Сентябрь	9963	10333	10665	12417	13281	13812	15110	16234
Октябрь	8194	9109	9396	9637	10366	10857	12185	12389
Ноябрь	6848	7685	7775	8094	8730	9290	10645	11595
Декабрь	7027	7602	7933	9280	9614	10925	12161	12772



- ❖ Данные об авиалайнерах представляют собой образец сезонных колебаний, наслаивающихся на монотонно растущий тренд.
- ❖ Сезонный эффект легко расшифровывается. Например, в 1968 г. мы наблюдали три "всплеска" активности пассажирских авиaperевозок; все они объясняются перелетами в праздничный и отпускной периоды: один из них приходится на пасху, второй - на лето и третий - на рождественские праздники.
- ❖ Амплитуда колебаний меняется из года в год частично из-за возрастающего парка авиалайнеров, а частично из-за увеличения периода праздников.

**Пример.** В таблице представлены данные Financial Times (FT) о квартальной динамике среднего индекса курса акций ведущих компаний на лондонской бирже за 1960-1971 годы.

Индекс FT курса обыкновенных акций ведущих компаний: квартальные средние за 1960-1971 гг.								
Год	Квартал	Индекс	Год	Квартал	Индекс	Год	Квартал	Индекс
1960	1	323,8	1964	1	335,1	1968	1	409,1
	2	314,1		2	344,4		2	461,1
	3	321		3	360,9		3	491,4
	4	312,9		4	346,5		4	490,5
1961	1	323,7	1965	1	340,6	1969	1	491
	2	349,3		2	340,3		2	433
	3	310,4		3	323,3		3	378
	4	295,8		4	345,6		4	382,6
1962	1	301,2	1966	1	349,3	1970	1	403,4
	2	285,8		2	359,7		2	354,7
	3	271,7		3	320		3	343
	4	283,6		4	299,9		4	345,4
1963	1	295,7	1967	1	318,5	1971	1	330,4
	2	309,3		2	343,1		2	372,8
	3	295,7		3	360,8		3	409,2
	4	342		4	397,8		4	427,6



- Значения временного ряда формируются под воздействием факторов долговременных, циклических и случайных

**Пример.** В таблице и на рисунке представлены данные об урожае ячменя в Англии и Уэльсе за 56 лет (с 1884 по 1939 гг.). Временной интервал  $\Delta$  в данном случае равен одному году.

Год	Урожайность, ц/акр	Год	Урожайность, ц/акр	Год	Урожайность, ц/акр
1884	15,2	1903	15,1	1922	14
1885	16,9	1904	14,6	1923	14,5
1886	15,3	1905	16	1924	15,4
1887	14,9	1906	16,8	1925	15,3
1888	15,7	1907	16,8	1926	16
1889	15,1	1908	15,5	1927	16,4
1890	16,7	1909	17,3	1928	17,2
1891	16,3	1910	15,5	1929	17,8
1892	16,5	1911	15,5	1930	14,4
1893	13,3	1912	14,2	1931	15
1894	16,5	1913	15,8	1932	16
1895	15	1914	15,7	1933	16,8
1896	15,9	1915	14,1	1934	16,9
1897	15,5	1916	14,8	1935	16,6
1898	16,9	1917	14,4	1936	16,2
1899	16,4	1918	15,6	1937	14
1900	14,9	1919	13,9	1938	18,1
1901	14,5	1920	14,7	1939	17,5



- Анализ графика временного ряда приводит к выводу, что в генерировании этих данных не участвовали долговременные факторы (поскольку тренд, скорее всего, отсутствует).



#### 4.7. Аддитивная и мультипликативная модели временного ряда.

Модель, в которой временной ряд представлен в виде суммы тренда, сезонной (или циклической) и случайной компонент, называется аддитивной

$$y_t = f(t) + \varphi(t) + \varepsilon_t$$

Модель, в которой временной ряд представлен в виде произведения тренда, сезонной (или циклической) и случайной компонент, называется мультипликативной.

$$y_t = f(t) \cdot \varphi(t) \cdot \varepsilon_t$$

Мультипликативную модель часто можно свести к аддитивной логарифмированием.

$$\ln y_t = \ln f(t) + \ln \varphi(t) + \ln \varepsilon_t$$

## 4.8. Выделение основных компонентов временного ряда.

### 4.8.1. Порядок анализа модели.

В аддитивной форме ряд представляется в виде

$$y_t = f(t) + \varphi(t) + \varepsilon_t$$

В мультипликативной форме ряд представляется в виде

$$y_t = f(t) \cdot \varphi(t) \cdot \varepsilon_t$$

Построение модели включает следующие шаги

1. Сглаживание исходного ряда методом скользящей средней.
2. Расчёт значений периодической компоненты.
3. Устранение периодической компоненты из исходного ряда и получение аналитического выражения для тренда.
4. Анализ остатков.

### 4.8.2. Сглаживание временного ряда методом скользящей средней.

Чтобы уменьшить влияние случайных и циклических факторов, упростив нахождение тренда, используется сглаживание временного ряда методом скользящей средней.

Сглаживание представляет собой вычисление взвешенного среднего значений, наблюдаемых в окрестности рассматриваемой точки.

Оно определяется для каждого момента времени, за исключением нескольких первых и нескольких последних точек.

В простейшем случае вычисление производится следующим образом: берется  $k$  последовательных значений временного ряда ( $k$  нечётно), со средним, соответствующим текущему моменту времени  $t$ :

$$Y_t, Y_{t-1}, Y_{t+1}, Y_{t-2}, Y_{t+2},$$

и находится их среднее арифметическое:

$$\tilde{y}_t = \frac{y_t + y_{t-1} + y_{t+1} + \dots}{k}$$

Затем точка  $t$  сдвигается вправо на один шаг, опять производится усреднение значений временного ряда и т.д.

Число  $k$  (т.е. временной интервал, по которому производится усреднение), называется окном.

Замечание. Иногда требуется усреднять за период, равный чётному количеству шагов. Например, за год=12 месяцев. В этом случае применяют взвешенные средние. Например, в случае года:

$$\tilde{y}_t = \frac{y_{t-6} + 2y_{t-5} + 2y_{t-4} + \dots + 2y_t + 2y_{t+1} + \dots + 2y_{t+5} + y_{t+6}}{24}$$

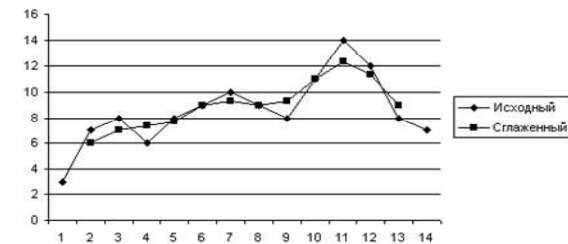
**Задача.** Сгладить временной ряд с окном 3.

$t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$y_t$	3	7	8	6	8	9	10	9	8	11	14	12	8	7

**Решение.** Вычислим поочерёдно

$$\tilde{y}_2 = \frac{y_1 + y_2 + y_3}{3} = \frac{3+7+8}{3} = 6, \quad \tilde{y}_3 = \frac{y_2 + y_3 + y_4}{3} = \frac{7+8+6}{3} = 7, \dots$$

$t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$y_t$	3	7	8	6	8	9	10	9	8	11	14	12	8	7
$\tilde{y}_t$	-	6	7	7,3	7,7	9	9,3	9	9,3	11	12,3	11,3	9	-



### 4.8.3. Экспоненциальное сглаживание.

Экспоненциально сглаженный ряд  $S_t$  определяется формулами

$$s_1 = y_1, \quad s_t = s_{t-1} + \alpha(y_t - s_{t-1})$$

где  $\alpha$  – параметр сглаживания,  $0 \leq \alpha \leq 1$ . Чем меньше  $\alpha$ , тем сильнее сглаживание.

Можно дать следующую интерпретацию экспоненциальной средней:

если  $S_{t-1}$  — прогноз значения ряда  $Y_t$ , то разность  $Y_t - S_{t-1}$  есть погрешность прогноза;

таким образом прогноз  $S_t$  для следующего момента времени  $t+1$  учитывает ставшую известной в момент  $t$  ошибку прогноза.

Метод экспоненциального сглаживания часто применяется для краткосрочного прогнозирования. При этом основной задачей является выбор параметра сглаживания.

**Задача.** Экспоненциально сгладить временной ряд с параметрами  $\alpha=0,5$  и  $\alpha=0,3$ .

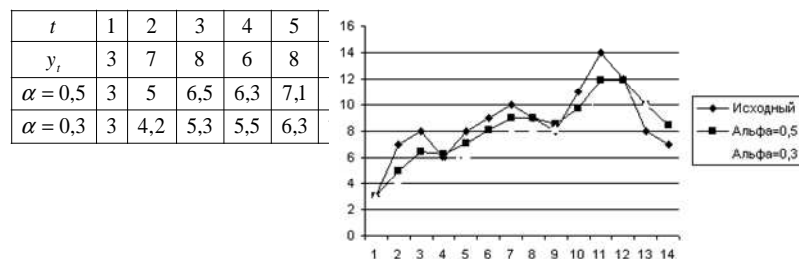
$t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$y_t$	3	7	8	6	8	9	10	9	8	11	14	12	8	7

**Решение.** Пусть  $\alpha=0,5$ .

$$s_1 = y_1 = 3, \quad s_2 = s_1 + 0,5 \cdot (y_2 - s_1) = 3 + 0,5 \cdot (7 - 3) = 5, \dots$$

Пусть теперь  $\alpha=0,3 \dots$

$$s_1 = y_1 = 3, \quad s_2 = s_1 + 0,3 \cdot (y_2 - s_1) = 3 + 0,3 \cdot (7 - 3) = 4,2, \dots$$



### 4.8.4. Выделение тренда временного ряда в аналитической форме.

Напомним, что трендом (или тенденцией) называют неслучайную медленно меняющуюся составляющую временного ряда, на которую могут накладываться случайные колебания или сезонные эффекты.

**Замечание.** Следует признать относительность слов «медленно меняющаяся составляющая» в определении тренда.  
 Например, при исследовании величины осадков в течении сотни лет, медленное увеличение в течение всей длительности изучаемого периода может быть понято как тренд, однако на самом деле этот рост осадков, характерный для данного столетия, может оказаться частью некоторого медленного колебательного процесса наблюдаемого в пределах нескольких тысячелетий.  
 Делать вывод на тысячелетие вперед на основе «тренда», выявленного по данным одного столетия, очевидно, неправильно.

Говоря о тренде, мы должны учитывать длину ряда, к которому относится формулируемое утверждение. При различении тренда и циклической компоненты колебания невозможно исключить из рассуждений элемент субъективности.

В качестве функции, моделирующей долговременную тенденцию, чаще всего берут

- ❖ Гиперболический тренд:  $f(t)=a+b/t$ ;
- ❖ Линейный тренд:  $f(t)=a+bt$ ;
- ❖ Экспоненциальный тренд:  $f(t)=a \cdot b^t$ ;
- ❖ Степенной тренд:  $f(t)=a \cdot t^b$ ;
- ❖ Полиномиальный тренд:  $f(t)=a+bt+ct^2+\dots$ ;
- ❖ Логистическая кривая:  $f(t)=1/(a+be^{-x})$

Выбор вида тренда обычно производится с учётом выводов экономической теории и визуального анализа графика ряда. При прочих равных можно использовать критерий наименьшего значения относительной ошибки аппроксимации.

Вычисление коэффициентов тренда производится обычным МНК (с предварительной линеаризацией при необходимости – см. . 3.6).

**Задача.** Вычислить коэффициенты линейного тренда

$f(t) = a + bt$  для временного ряда

$t$	1	2	3	4	5	6	7
$y_t$	82	87	99	104	107	121	118

**Решение.** Применим МНК

$$Y = \begin{pmatrix} 82 \\ 87 \\ 99 \\ \vdots \end{pmatrix}, X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \end{pmatrix}, X^T X = \dots, \dots \quad \begin{pmatrix} a \\ b \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 76,29 \\ 6,57 \end{pmatrix}$$

$$f(t) = 76,29 + 6,57 \cdot t$$

**Задача.** Для того же временного ряда вычислить коэффициенты степенного тренда  $f(t) = a \cdot t^b$

**Решение.** Линеаризуем, взяв логарифм от обеих частей:  $\ln(f(t)) = \ln a + b \ln t$

$$z = \begin{pmatrix} 4,4 \\ 4,5 \\ \dots \end{pmatrix}, X = \begin{pmatrix} 1 & 0 \\ 1 & 0,7 \\ \dots & \dots \end{pmatrix}, X^T X = \dots, \dots$$

$\ln t$	0	0,7	1,1	1,4	1,6	1,8	1,9
$\ln y_t$	4,4	4,5	4,6	4,6	4,7	4,8	4,8

$$\begin{pmatrix} \ln a \\ b \end{pmatrix} = \begin{pmatrix} 4,37 \\ 0,21 \end{pmatrix} \quad a = e^{4,37} = 79,13 \quad f(t) = 79,13 \cdot t^{0,21}$$

#### 4.8.5. Использование MS Excel для выделения тренда.

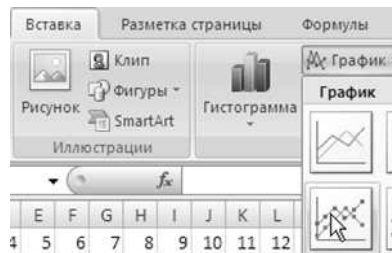
**Пример.** Найдём аналитическое выражение тренда в различных формах для временного ряда

$t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$y_t$	3	5	4	4	5	4	7	6	11	9	11	14	18	21	22	29

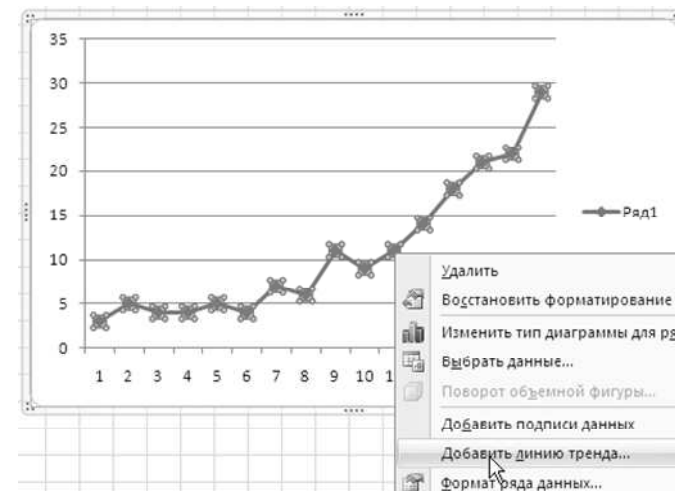
**Решение.** Занесём данные в таблицу MS Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	3	5	4	4	5	4	7	6	11	9	11	14	18	21	22	29

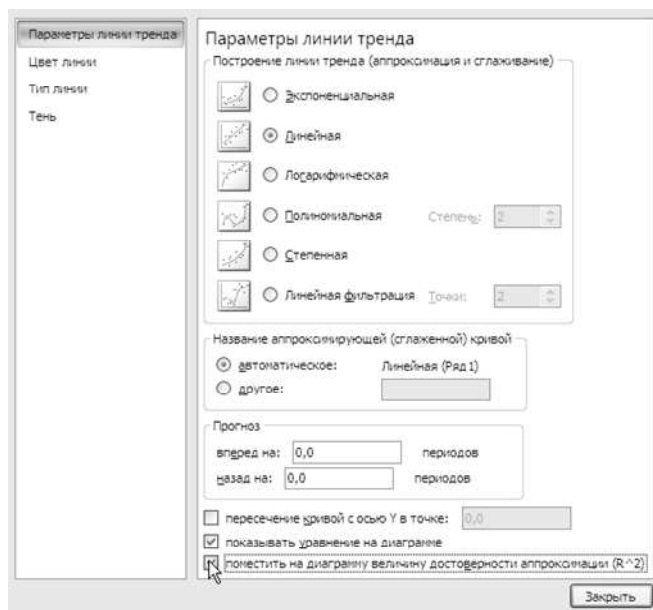
Меню  
Вставка  
-> Диаграмма  
-> График:



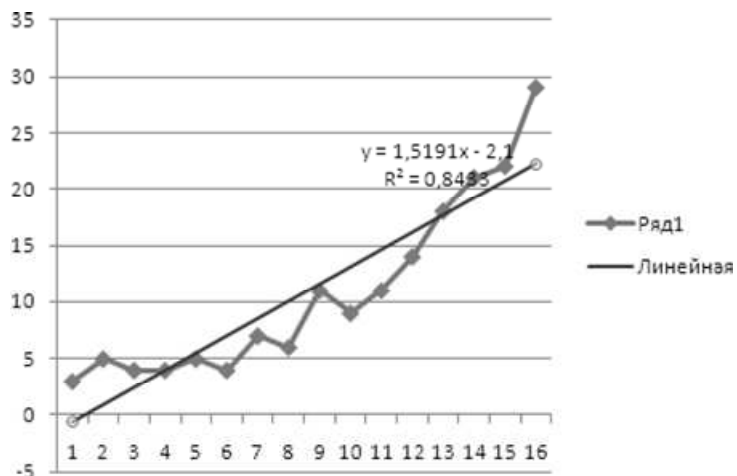
Выделим мышкой график и выберем пункт меню **Добавить линию тренда:**



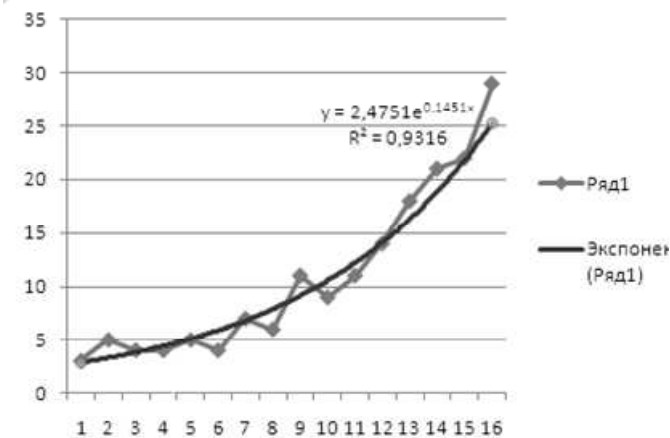
Появится диалог, в котором можно выбрать тип тренда:



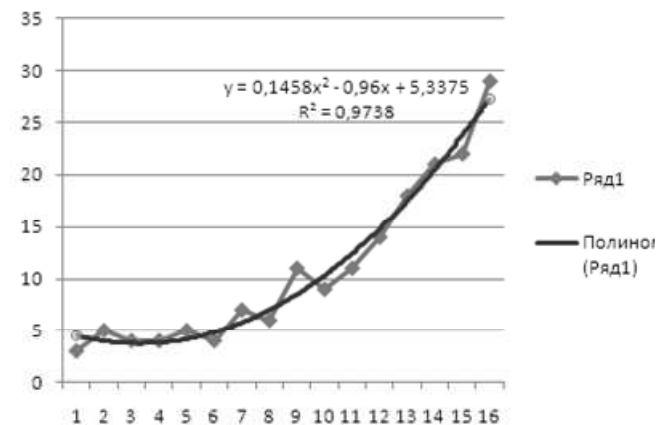
Выберем, например, линейный:



Экспоненциальный:



Полиномиальный второй степени:



По критерию R<sup>2</sup> наилучшее значение у последнего вида тренда.

### 4.8.6. Выделение компонентов временного ряда в случае аддитивной модели.

**Пример.** Выделить основные компоненты временного ряда, считая, что циклическая имеет период 4

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y <sub>t</sub>	6	4,4	5	9	7,2	4,8	6	10	8	5,6	6,4	11	9	6,6	7	10,8

**Шаг 1. Сглаживание.** Произведём сглаживание с периодом 4. Так как период чётный, будем использовать формулу

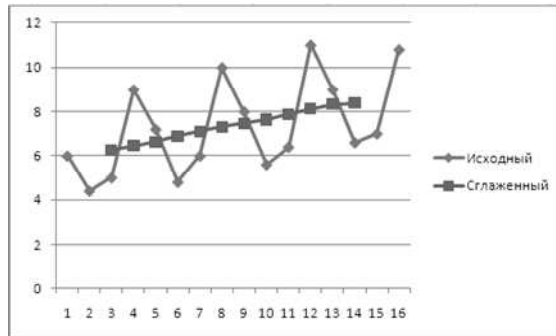
$$\tilde{y}_t = \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}$$

Например,

$$\tilde{y}_3 = \frac{6 + 2 \cdot 4,4 + 2 \cdot 5 + 2 \cdot 9 + 7,2}{8} = 6,25$$

t	1	2	3	4	5	6	7	8
y <sub>t</sub>	6	4,4	5	9	7,2	4,8	6	10
$\tilde{y}_t$	-	-	6,25	6,45	6,625	6,875	7,1	7,3

t	9	10	11	12	13	14	15	16
y <sub>t</sub>	8	5,6	6,4	11	9	6,6	7	10,8
$\tilde{y}_t$	7,45	7,625	7,875	8,125	8,325	8,325	-	-



Получим сглаженный ряд



**Шаг 2. Выделение периодической составляющей.** Вычислим разности между исходным и сглаженным рядом

t	1	2	3	4	5	6	7	8
y <sub>t</sub> - $\tilde{y}_t$	-	-	-1,25	2,55	0,575	-2,075	-1,1	2,7

t	9	10	11	12	13	14	15	16
y <sub>t</sub> - $\tilde{y}_t$	0,55	-2,025	-1,475	2,875	0,675	-1,775	-	-

Расположим в таблице из четырёх колонок (так как период равен 4)

1	2	3	4
-	-	1,25	2,55
0,575	-2,075	-1,1	2,7
0,55	-2,025	-1,475	2,875
0,675	-1,775	-	-
0,6	-1,958	-1,275	2,708

В последней строке записаны средние арифметические чисел в столбцах:  $0,6 = (0,575 + 0,55 + 0,675) / 3$  и т.д.

**Периодическая компонента** находится вычитанием из каждого числа последней строки среднего арифметического  $\Delta$ .

$$\Delta = \frac{0,6 - 1,958 - 1,275 + 2,708}{4} = 0,01875$$

$\varphi(1)$	$\varphi(2)$	$\varphi(3)$	$\varphi(4)$
$0,6 - \Delta =$	$-1,958 - \Delta =$	$-1,275 - \Delta =$	$2,708 - \Delta =$
0,581	-1,977	-1,294	2,69

**Шаг 3. Выделение тренда.** Удалим из исходного ряда периодическую составляющую.

Из  $y_1, y_5, y_9, \dots$  вычитается  $\varphi(1)$ , из  $y_2, y_6, y_{10}, \dots$  вычитается  $\varphi(2)$ , из  $y_3, y_7, y_{11}, \dots$  вычитается  $\varphi(3)$  и т.д. Получаем ряд  $f(t) + \varepsilon_t$ .

t	1	2	3	4	5	6	...	16
$f(t) + \varepsilon_t =$	5,419	6,337	6,694	6,31	6,619	6,777	...	8,11
$y_t - \varphi(t)$								

Определим коэффициенты линейного тренда  $f(t) = a + bt$  по последней таблице.

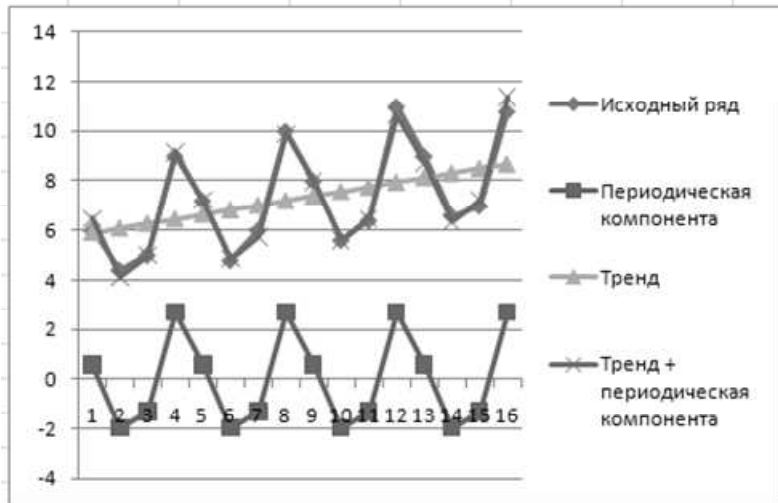
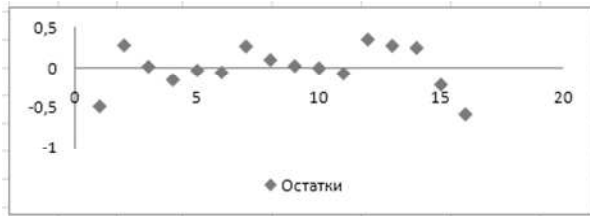
$t$	1	2	3	4	5	6	...	16
$f(t) + \varepsilon_t = y_t - \varphi(t)$	5,419	6,337	6,694	6,31	6,619	6,777	...	8,11

$$Y = \begin{pmatrix} 5,419 \\ 6,337 \\ 6,694 \\ \vdots \end{pmatrix}, X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \end{pmatrix}, X^T X = \dots, \dots \quad \begin{pmatrix} a \\ b \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 5,715 \\ 0,186 \end{pmatrix}$$

$$f(t) = 5,715 + 0,186 \cdot t$$

**Шаг 4. Остаток ряда.**

Вычитаем из исходного ряда тренд и периодическую составляющую.



**4.8.7. Выделение компонентов временного ряда в случае мультипликативной модели.**

**Пример.** Выделить основные компоненты временного ряда, считая, что циклическая имеет период 4

$t$	1	2	3	4	5	6	7	8	9	...	16
$y_t$	72	100	90	64	70	92	80	58	62	...	30

**Шаг 1. Сглаживание.** Произведём сглаживание с периодом 4. Этот этап не отличается от случая аддитивной модели. Так как период чётный, будем использовать формулу

$$\tilde{y}_t = \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}$$

Например,

$$\tilde{y}_3 = \frac{72 + 2 \cdot 100 + 2 \cdot 90 + 2 \cdot 64 + 70}{8} = 81,5$$

Получим сглаженный ряд.

$t$	1	2	3	4	5	6	7	8	9	...	16
$y_t$	72	100	90	64	70	92	80	58	62	...	30
$\tilde{y}_t$	-	-	81,5	81	79	76,5	75	73	70	...	-

**Шаг 2. Выделение периодической компоненты.** Вычислим частное от деления исходного ряда на сглаженный

$t$	1	2	3	4	5	6	7	8	...	14	15	16
$y_t / \tilde{y}_t$	-	-	1,08	0,8	0,9	1,215	1,081	0,811	...	1,194	-	-

Расположим в таблице из четырёх колонок (так как период равен 4)

1	2	3	4
-	-	1,08	0,8
0,9	1,215	1,081	0,811
...	...	...	...
...	1,194	-	-
0,918	1,209	1,088	0,808

В последней строке записаны средние арифметические чисел в столбцах



**Периодическая компонента** находится делением каждого числа последней строки на среднее арифметическое  $\Delta$ .

$$\Delta = \frac{0,918 + 1,209 + 1,088 + 0,808}{4} = 1,00575$$

$\varphi(1)$	$\varphi(2)$	$\varphi(3)$	$\varphi(4)$
$0,918 / \Delta =$	$1,209 / \Delta =$	$1,088 / \Delta =$	$0,808 / \Delta =$
0,913	1,202	1,082	0,803

**Шаг 3. Выделение тренда.** Удалим из исходного ряда периодическую составляющую.

Значения  $Y_1, Y_5, Y_9, \dots$  делятся на  $\varphi(1)$ , значения  $Y_2, Y_6, Y_{10}, \dots$  делятся на  $\varphi(2)$ , значения  $Y_3, Y_7, Y_{11}, \dots$  делятся на  $\varphi(3)$  и т.д. Получаем ряд  $f(t) \cdot \varepsilon_t$ .

$t$	1	2	3	4	5	6	...	16
$f(t) \cdot \varepsilon_t =$	78,86	83,19	83,18	79,7	76,67	76,54	...	37,36
$y_t / \varphi(t)$								

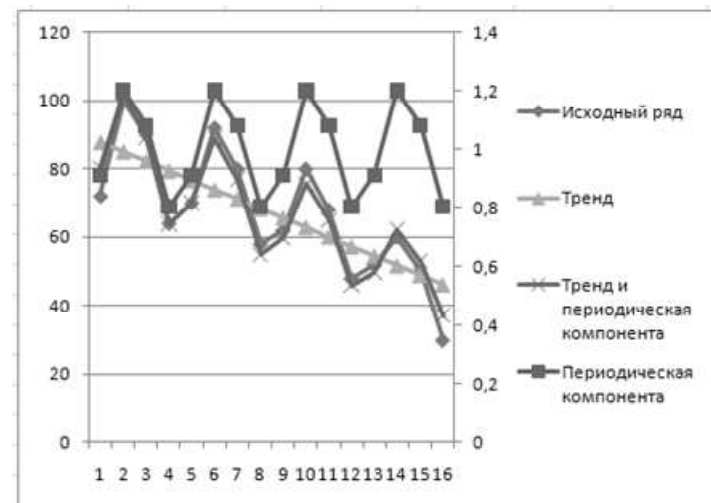
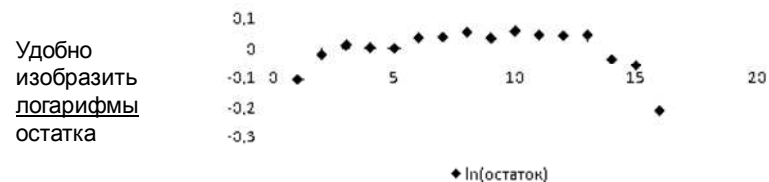
Определим коэффициенты линейного тренда  $f(t) = a + bt$  по последней таблице.

$t$	1	2	3	4	5	6	...	16
$f(t) \cdot \varepsilon_t =$	78,86	83,19	83,18	79,7	76,67	76,54	...	37,36
$y_t / \varphi(t)$								

$$Y = \begin{pmatrix} 78,86 \\ 83,19 \\ 83,18 \\ \vdots \end{pmatrix}, X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \end{pmatrix}, X^T X = \dots, \dots \begin{pmatrix} a \\ b \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 90,59 \\ -2,773 \end{pmatrix}$$

$$f(t) = 90,59 - 2,773 \cdot t$$

**Шаг 4. Остатки.** Разделим исходный ряда на тренд и периодическую составляющую.



### 7. Автокорреляционная функция. Коррелограмма.

Если временной ряд  $y(t)$  содержит тренд или периодическую составляющую, то, очевидно, имеется зависимость между предыдущими и последующими значениями ряда.

Коэффициент корреляции между уровнями исходного ряда  $y(t)$  и уровнями этого ряда, сдвинутыми на  $s$  шагов  $y(t-s)$  называется **коэффициентом автокорреляции** временного ряда **порядка  $s$**  и обозначается  $r(s)$ . Число  $s$  ещё называют **лагом**.

Как известно, коэффициент корреляции показывает тесноту линейной связи между случайными величинами. Поэтому  $r(s)$  показывает наличие (или отсутствие) линейной связи между значениями временного ряда, отстоящими друг от друга на время  $s$

#### 4.9. Автокорреляционная функция. Коррелограмма.

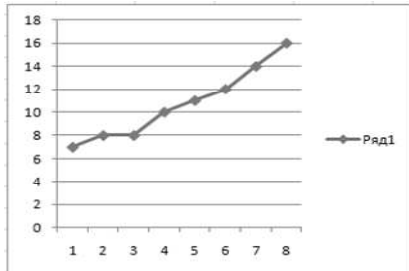
Если временной ряд  $y(t)$  содержит тренд или периодическую составляющую, то, очевидно, имеется зависимость между предыдущими и последующими значениями ряда.

Коэффициент корреляции между уровнями исходного ряда  $y(t)$  и уровнями этого ряда, сдвинутыми на  $s$  шагов  $y(t-s)$  называется коэффициентом автокорреляции временного ряда порядка  $s$  и обозначается  $r(s)$ . Число  $s$  ещё называют лагом.

Как известно, коэффициент корреляции показывает тесноту линейной связи между случайными величинами. Поэтому  $r(s)$  показывает наличие (или отсутствие) линейной связи между значениями временного ряда, отстоящими друг от друга на время  $s$

**Задача.** Вычислить коэффициенты автокорреляции порядка 1 и 2 для временного ряда

$t$	1	2	3	4	5	6	7	8
$y_t$	7	8	8	10	11	12	14	16



**Решение.**

Формула для коэффициента корреляции случайных величин  $x$  и  $y$  имеет вид

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}$$

Найдём сначала

$$r(1) = r_{y_t, y_{t-1}}$$

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}$$

$$\bar{y}_t = \frac{8 + \dots + 16}{7}, \quad \bar{y}_{t-1} = \frac{7 + \dots + 14}{7},$$

$$\overline{y_t^2} = \frac{8^2 + \dots + 16^2}{7},$$

$$\overline{y_t y_{t-1}} = \frac{8 \cdot 7 + \dots + 16 \cdot 14}{7}$$

$t$	1	2	3	4	5	6	7	8
$y_t$	7	8	8	10	11	12	14	16
$y_{t-1}$	-	7	8	8	10	11	12	14

$$r(1) = r_{y_t, y_{t-1}} = 0,976$$

Теперь найдём

$$r(2) = r_{y_t, y_{t-2}}$$

$t$	1	2	3	4	5	6	7	8
$y_t$	7	8	8	10	11	12	14	16
$y_{t-2}$	-	-	7	8	8	10	11	12

$$\bar{y}_t = \frac{8 + \dots + 16}{6}, \quad \bar{y}_{t-2} = \frac{7 + \dots + 12}{6}, \dots$$

$$r(2) = r_{y_t, y_{t-2}} = 0,973$$

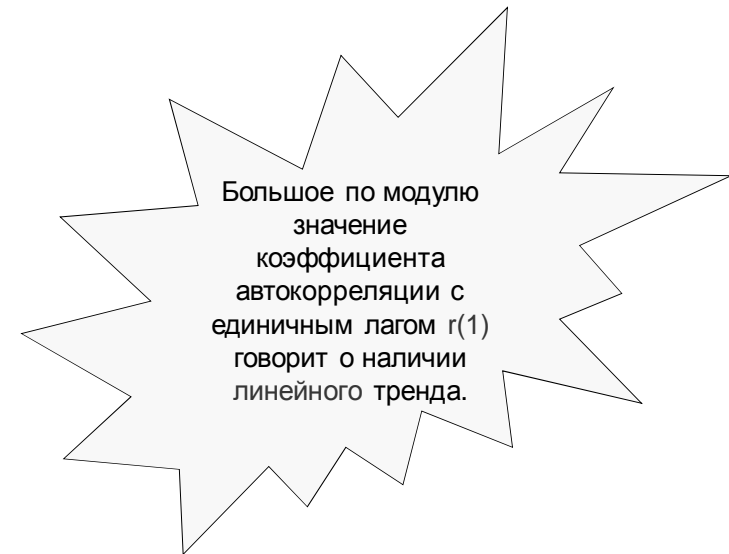
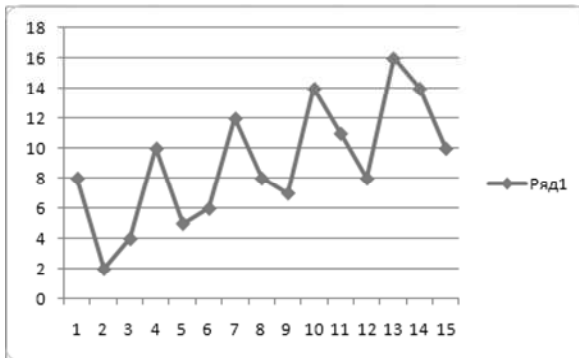


График зависимости  $r(s)$  называется **коррелограммой**

**Задача.** Рассчитать коррелограмму для временного ряда до порядка 8 включительно



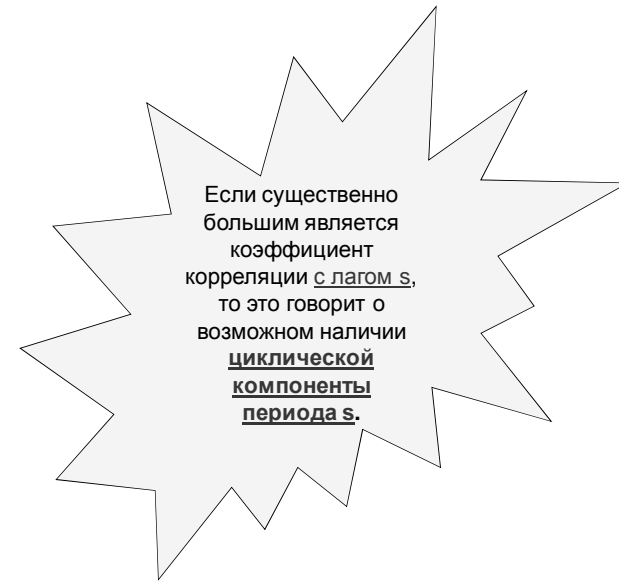
t	y <sub>t</sub>
1	8
2	2
3	4
4	10
5	5
6	6
7	12
8	8
9	7
10	14
11	11
12	8
13	16
14	14
15	10

Аналогично предыдущему находим автокорреляции с лагом 1,2,3,...,8. Получаем

Автокорреляционная функция для y

Лаг	ACF	PACF
1	0,3000	0,3000
2	0,0091	-0,0889
3	0,5182 ***	0,6002 ***
4	0,1000	-0,4751 *
5	-0,2545	0,0720
6	0,1273	-0,1892
7	-0,0500	-0,0258
8	-0,3818	-0,1456

Столбец ACF содержит искомые коэффициенты автокорреляции. Видно, что максимальная по модулю автокорреляция (помечена звёздочками) имеет лаг 3, что на диаграмме примерно соответствует периоду циклической компоненты.



**4.10. Стационарные и нестационарные временные ряды. Стационарность в узком и широком смысле.**

Остаток временного ряда, освобождённый от периодической составляющей и тренда, должен быть в каком-то смысле стационарным.

На интуитивном уровне **стационарность временного ряда** мы связываем с требованием, чтобы он имел **постоянное среднее значение** и колебался вокруг этого среднего с **постоянной дисперсией**.

Ряд  $y(t)$  называется **строго стационарным (или стационарным в узком смысле)**, если для любого числа  $T$  совместное распределение вероятностей  $m$  наблюдений  $y(t_1), y(t_2), \dots, y(t_m)$  такое же, как и для  $m$  наблюдений  $y(t_1 + T), y(t_2 + T), \dots, y(t_m + T)$ .

Другими словами, свойства строго стационарного временного ряда **не меняются при изменении начала отсчета времени**.

В частности, из предположения о строгой стационарности временного ряда  $y(t)$  следует, что **закон распределения вероятностей случайной величины  $y(t)$  не зависит от  $t$ ,**

а значит, **не зависят от  $t$  и все его основные числовые характеристики**, в том числе:

**Математическое ожидание  $My(t)=m$**

**Дисперсия  $Dy(t)=\sigma^2$**

Поскольку закон распределения вероятностей случайной величины  $y(t)$  одинаков при всех  $t$ , то он сам и его основные числовые характеристики могут быть оценены по наблюдениям  $y(1), y(2), \dots, y(N)$ .

В частности:

$$\hat{m} = \bar{y} = \frac{\sum y_i}{n}, \quad \hat{\sigma}^2 = \frac{n}{n-1} (\overline{y^2} - \bar{y}^2)$$

Можно расширить класс моделей стационарных временных рядов, используемых при анализе конкретных рядов экономической динамики.

С этой целью введем понятие **стационарного в широком смысле** временного ряда.

Ряд  $y(t)$  называется **слабо стационарным (или стационарным в широком смысле)**, если его среднее значение и дисперсия не зависят от  $t$ .

#### 4.11. Проверка гипотез о структуре временного ряда

##### 4.11.1. Проверка гипотез о наличии тренда

###### 4.11.1.1. Критерий рядов (критерий Вальда-Волфовитца).

Одним из способов проверки гипотезы о наличии тренда является рассмотренный ранее в пункте 3.3.2.2 **критерий рядов**. Опишем его ещё раз в приложении ко временным рядам

Напомним, что медианой называется **средний элемент** упорядоченного по возрастанию числового ряда, если число элементов нечётно, и полусумма двух средних, если число элементов чётно.

В критерии рядов рассматриваются **знаки отклонений  $y_t$  от медианы**. Если  $y_t \geq y_{med}$ , то ставится знак плюс, иначе минус.

“Ряд” это **непрерывная последовательность одинаковых знаков**.

При достаточно большом числе наблюдений временного ряда:

1) Рассчитываются статистики.

$$M = \frac{2n_+n_-}{n} + 1, \quad D = \frac{2n_+n_-(2n_+n_- - n)}{n^2(n-1)}$$

где  $n_+$  - количество плюсов,  $n_-$  - количество минусов,  $n$  - общее количество.

2) Если

$$M - u_{1-\alpha/2} \cdot D < k < M + u_{1-\alpha/2} \cdot D$$

где  $k$  - количество рядов, то **не отвергается гипотеза об отсутствии тренда**.

Для **небольших объёмов выборки** есть таблицы критических значений  $k$  (пример для  $\alpha=0,01$ ):

$n_+ \setminus n_-$	5	6	7	8	9	10	...
3	1-8	2-8	2-8	2-8	2-8	2-8	...
4	2-9	2-9	2-10	3-10	3-10	3-10	...
5	2-10	3-10	3-11	3-11	3-12	3-12	...
6		3-11	3-12	3-12	4-13	4-13	...
7			3-13	4-13	4-14	4-14	...
8				4-14	4-15	4-15	...
9					5-15	5-16	...
10						6-16	...
...	...	...	...	...	...	...	...

**Задача.** Проверить гипотезу об отсутствии тренда по критерию рядов при  $\alpha=0,01$

**Решение.** Медиана равна 8

Знаки отклонений имеют вид:

+ - - + - - + + - + + + + + +.

Здесь  $k=7$ ,  $n_+=10$ ,  $n_-=5$ . По таблице находим интервал: 3-12.

Число  $k$  попадает в эти пределы, так что на уровне  $\alpha=0,01$  гипотеза об отсутствии тренда принимается.

| t  | $y_t$ |
|----|-------|
| 1  | 8     |
| 2  | 2     |
| 3  | 4     |
| 4  | 10    |
| 5  | 5     |
| 6  | 6     |
| 7  | 12    |
| 8  | 8     |
| 9  | 7     |
| 10 | 14    |
| 11 | 11    |
| 12 | 8     |
| 13 | 16    |
| 14 | 14    |
| 15 | 10    |

#### 4.11.1.2. Критерий Рамачандрана-Ранганатана

1) Рассчитывается статистика.  $R = \sum_j j^2 n_j$

где  $n_j$  - количество рядов длины  $j$ .

2) Если  $R < R(n, \alpha)$

то **не отвергается гипотеза об отсутствии тренда**.

Таблица значений  $R(n, \alpha)$ .

| $n \setminus \alpha$ | 0,1 | 0,05 | 0,01 | $n$ | 0,1 | 0,05 | 0,01 |
|----------------------|-----|------|------|-----|-----|------|------|
| 6                    | 18  | 18   | 18   | 18  | 68  | 78   | 90   |
| 8                    | 26  | 32   | 32   | 20  | 68  | 78   | 100  |
| 10                   | 34  | 38   | 50   | 22  | 76  | 86   | 108  |
| 12                   | 38  | 44   | 58   | 24  | 84  | 94   | 118  |
| 14                   | 46  | 52   | 68   | 26  | 92  | 102  | 128  |
| 16                   | 54  | 60   | 80   | 28  | 98  | 110  | 136  |

**Задача.** Проверить гипотезу об отсутствии тренда методом Рамачандрана-Ранганатана при  $\alpha=0,01$

**Решение.** Знаки отклонений имеют вид:  
+ - - + - - + + - + + + + +.

Здесь  $n=15, n_1=3, n_2=3, n_6=1$ .

$$R = \sum_j j^2 n_j = 1^2 \cdot 3 + 2^2 \cdot 3 + 6^2 \cdot 1 = 51$$

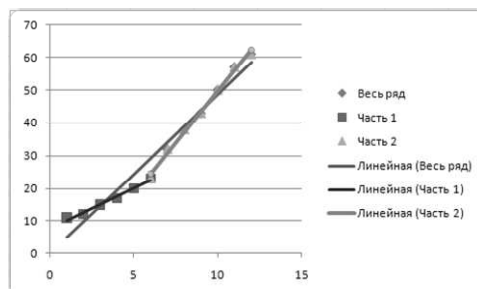
По таблице находим  $R(15;0,01) \approx 74$ .  
Так как  $51 < 74$ , то на уровне  $\alpha=0,01$  гипотеза об отсутствии тренда не отвергается.

| t  | y <sub>t</sub> |
|----|----------------|
| 1  | 8              |
| 2  | 2              |
| 3  | 4              |
| 4  | 10             |
| 5  | 5              |
| 6  | 6              |
| 7  | 12             |
| 8  | 8              |
| 9  | 7              |
| 10 | 14             |
| 11 | 11             |
| 12 | 8              |
| 13 | 16             |
| 14 | 14             |
| 15 | 10             |

#### 4.11.2. Проверка гипотез о структурном изменении

##### 4.11.2.1. Тест Чоу (Chow).

В некоторых случаях можно наблюдать изменения характера тренда, вызванные структурными изменениями в экономике или иными факторами. В таком случае уравнения тренда на разных участках будут различаться



В ситуации на рисунке возникает альтернатива – описывать тренд одной синей линией или совокупностью из двух линий

#### 4.11.2.2. Использование фиктивных переменных.

**Замечание.** Другой способ проверки состоит в использовании фиктивных переменных.

Добавим в модель фиктивную переменную  $z$ , равную 0 для первой выборки и 1 для второй, а также переменные  $zx_1, zx_2, \dots$ . Оценим соответствующее уравнение регрессии. Если эти переменные окажутся незначимыми, то объединение выборок возможно.

**Пример.** Проверить возможность объединения выборок при  $\alpha=0,1$ .

| t | y <sub>t</sub> | t  | y <sub>t</sub> |
|---|----------------|----|----------------|
| 1 | 11             | 7  | 32             |
| 2 | 12             | 8  | 38             |
| 3 | 15             | 9  | 43             |
| 4 | 17             | 10 | 50             |
| 5 | 20             | 11 | 57             |
| 6 | 23             | 12 | 61             |

**Решение. 1 способ (критерий Чоу).**

Для первой выборки:

$$y_t = 7,73 + 2,46 \cdot t, ESS = 1,67$$

Для второй выборки:

$$y_t = -9,9 + 5,67 \cdot t, ESS = 2,81$$

Для объединённой выборки:  $y_t = -0,08 + 4,87 \cdot t, ESS = 132,52$

$$F = \frac{(132,52 - 1,67 - 2,81)/(1+1)}{(1,67 + 2,81)/(12 - 2 \cdot 1 - 2)} \approx 114,32$$

$$F_{0,9}(1+1; 12 - 2 \cdot 1 - 2) = F_{0,9}(2; 8) = 3,11$$

Так как  $114,32 > 3,11$ , то наличие структурного изменения не отвергается.

**2 способ (фиктивная переменная).**  
Добавляем новые переменные  $z, zt$ .

Проверяем их значимость

$$y = 7,73 + 2,46 \cdot t - 17,63 \cdot z + 3,51 \cdot zt$$

(0,7)      (0,18)      (1,87)  
(0,25)

Все коэффициенты перед добавленными переменными значимы

| y  | t  | z | zt |
|----|----|---|----|
| 11 | 1  | 0 | 0  |
| 12 | 2  | 0 | 0  |
| 15 | 3  | 0 | 0  |
| 17 | 4  | 0 | 0  |
| 20 | 5  | 0 | 0  |
| 23 | 6  | 0 | 0  |
| 32 | 7  | 1 | 7  |
| 38 | 8  | 1 | 8  |
| 43 | 9  | 1 | 9  |
| 50 | 10 | 1 | 10 |
| 57 | 11 | 1 | 11 |
| 61 | 12 | 1 | 12 |

$\sum z = 12$

### 4.11.3. Проверка гипотез о наличии единичного корня.

#### 4.11.3.1. Модель авторегрессии AR(p).

**Моделью авторегрессии AR(p)** называется временной ряд, удовлетворяющий соотношению

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

где случайная составляющая  $\varepsilon_t$  – нормально распределённый стационарный процесс с нулевым средним и нулевой автокорреляцией (“гауссов белый шум”).

В частности, **AR(1)** – это временной ряд, удовлетворяющий соотношению

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

Интерес представляют условия стационарности таких рядов.

Чтобы ряд AR(p) был стационарен, необходимо, чтобы все корни уравнения

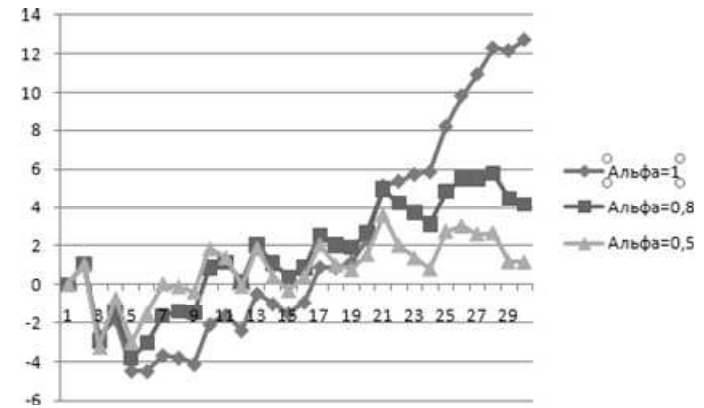
$$Z^p = c + \alpha_1 Z^{p-1} + \dots + \alpha_p Z$$

по модулю были меньше 1

В частности,

**Ряд AR(1) стационарен, если и только если  $|\alpha| < 1$**

Примеры реализации временных рядов AR(1) при различных  $\alpha$ :



#### 4.11.3.2. Интегрируемые временные ряды

Если в модели A(1) коэффициент  $\alpha$  равен 1:

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

то ряд не стационарен. В этом случае стационарным оказывается ряд для первых разностей  $\Delta y_t = y_t - y_{t-1}$ :

$$\Delta y_t = c + \varepsilon_t,$$

1. Ряд называется **интегрируемым порядка 1** (обозначение **I(1)**), если **ряд из его первых разностей стационарен**. Фактически это означает наличие **линейного тренда** и отсутствие периодических компонент.
2. Ряд называется **интегрируемым порядка k** (обозначение **I(k)**), если **ряд из разностей порядка k интегрируем**.

### 4.11.3.3. Тест Дики-Фуллера (Dickey-Fuller)

Перепишем уравнение AR(1) в виде

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

Если  $\rho=0$ , что соответствует  $\alpha=1$ , то исходный ряд не стационарен. Если  $\rho<0$ , то  $\alpha<1$  и исходный ряд стационарен.

**Тест Дики-Фуллера** проверяет гипотезу  $H_0: \rho=0$  при альтернативе  $H_1: \rho<0$ .

Если  $\tau = \frac{\hat{\rho}}{s_{\rho}} < \tau_{крит}$ ,

где  $\tau_{крит}$  – табличное значение, то гипотеза  $H_0$  принимается (ряд нестационарный), в противном случае отвергается (ряд считается стационарным).

Для использования теста Дики-Фуллера требуются соответствующие табличные значения. Они были получены эмпирически, с использованием метода Монте-Карло. За основу был взят процесс AR(1).

| $\alpha \setminus n$ | 25    | 50    | 100   | $\infty$ |
|----------------------|-------|-------|-------|----------|
| 0,01                 | -3,75 | -3,58 | -3,51 | -3,43    |
| 0,05                 | -3,33 | -3,22 | -3,17 | -3,12    |
| 0,1                  | -3    | -2,93 | -2,89 | -2,86    |

Модифицированный тест Дики-Фуллера рассматривает авторегрессионные процессы более высокого порядка A(p).

Основное уравнение вместо

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

принимает следующий вид:

$$\Delta y_t = c + \rho y_{t-1} + \Delta y_{t-1} + \dots + \Delta y_{t-p} + \varepsilon_t,$$

Дополнительные слагаемые позволяют убрать автокорреляцию остатков.

При проверке гипотезы используются та же таблица.

## 5. Взаимосвязь временных рядов

### 5.1. Специфика оценки взаимосвязи временных рядов

В анализе временных рядов важное место отводится изучению причинно-следственной взаимосвязи двух или большего числа временных рядов.

Примером может служить зависимость общих расходов населения  $Y_t$  от его наблюдаемых доходов  $X_t$  и множество других зависимостей.

Применение в этих целях традиционных методов корреляционно-регрессионного анализа может приводить к неправильным результатам вследствие наличия тренда и периодической составляющей в каждом из анализируемых рядов..

Предварительный этап анализа включает выявление структуры каждого ряда. Если в рассматриваемых временных рядах выявляются периодические составляющие, то перед проведением дальнейших исследований их необходимо устранить из числовых данных.



Наличие тренда в изучаемых временных рядах также вносит изменения в причинно-следственную связь. Большой коэффициент корреляции двух временных рядов может быть результатом того, что оба содержат ярко выраженный временной тренд.

Для того, чтобы получить модели, характеризующие истинную причинно-следственную связь между изучаемыми рядами, следует избавиться от так называемой ложной корреляции, вызванной наличием тенденции в каждом ряде.

Для этой цели применяются следующие методы:

- ❖ Метод отклонений от трендов;
- ❖ Метод последовательных разностей;
- ❖ Включение фактора времени в модель регрессии.

## 5.2. Метод отклонений от трендов

Для исходных временных рядов  $y_t$  и  $x_t$ , каждый из которых содержит трендовую компоненту, определяются расчетные по тренду уровни. Затем из фактических уровней каждого ряда вычитают расчетные значения:

$$y'_t = y_t - \hat{y}_t, \quad x'_t = x_t - \hat{x}_t$$

Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а отклонений от трендов.

## 5.3. Метод последовательных разностей

Если временной ряд содержит ярко выраженную линейную тенденцию, ее можно устранить путем замены исходных уровней ряда первыми разностями:

$$\Delta y_t = y_t - y_{t-1}, \quad \Delta x_t = x_t - x_{t-1}$$

Если, например,  $y_t = a + bt + \varepsilon_t$ ,

то первые разности не содержат линейного тренда:

$$\Delta y_t = a + bt + \varepsilon_t - (a + b(t-1) + \varepsilon_{t-1}) = b + \varepsilon_t - \varepsilon_{t-1}$$

Для **параболы** (многочлен второй степени) применяют **вторые разности**.

Если имеет место **экспоненциальный** или **степенной** тренд, метод последовательных разностей применяют к **логарифмам** исходных уровней ряда.

## 5.4. Включение фактора времени

Еще один путь устранения воздействия фактора времени на причинно-следственную связь анализируемых временных рядов – включение в модель фактора времени в качестве независимой переменной:

$$y_t = a + bx_t + ct + \varepsilon_t,$$

Модель строится по всей совокупности данных за рассматриваемый период в отличие от метода последовательных разностей, который приводит к потере числа наблюдений.

Параметры модели с включением фактора времени определяются обычным МНК.

## 5.5. Коинтеграция временных рядов

Общий недостаток методов исключения тенденции заключается в том, что эти методы предполагают некоторую модификацию исходной модели  $y_t = a + b \cdot x_t + e_t$  при помощи либо замены переменных, либо добавления в модель фактора времени.

Однако большая часть соотношений, постулируемых экономической теорией, сформулирована для исходной модели. А в экономике большая часть временных рядов содержит тренды какого-либо типа.

Стремление получить модель, которая учитывала бы краткосрочные и долгосрочные особенности процессов и в то же время поддерживала бы стационарность всех переменных, подтолкнуло к пересмотру принципов построения регрессии

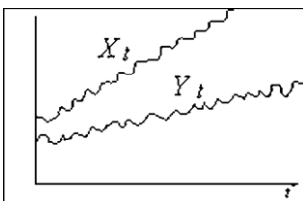
Новый подход основывается на таких экономических данных, которые, будучи нестационарными, могут быть скомбинированы в один ряд, который будет уже стационарным. Ряды, обладающие такой особенностью, называются коинтегрированными рядами

Напомним, что

Временной ряд  $Y_t$  называется интегрируемым порядка 1 (обозначение  $I(1)$ ), если его первая разность  $\Delta Y_t$  стационарна.

Два временных ряда, интегрируемые порядка 1, называются коинтегрируемыми, если некоторая линейная комбинация этих временных рядов есть стационарный временной ряд.

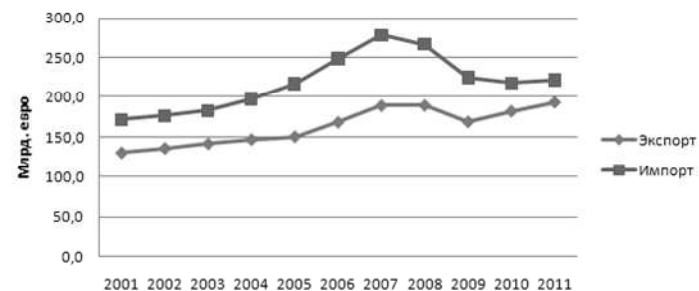
Пример.



Здесь ряд  $Y_t - 0,5X_t$  стационарен.

## Пример предположительно коинтегрированных рядов

Объёмы экспорта и импорта Испании



## Проверка гипотезы о коинтеграции при помощи критерия Энгеля-Гранжера (Engel-Granger)

Предварительно проверяется гипотезы об интегрируемости  $X_t$  и  $Y_t$ .

1. Выдвигается гипотеза  $H_0$  об отсутствии коинтеграции между рядами  $X_t$  и  $Y_t$ .
2. Строится уравнение регрессии  $Y_t$  на  $X_t$  и вычисляются остатки  $e_t$ .

3. Рассчитывают параметры уравнения регрессии вида

$$\Delta e_t = a + \rho \cdot e_{t-1},$$

где  $\Delta e_t$  - первые разности остатков.

4. Нулевой гипотезе  $H_0$  соответствует случай  $\rho=0$ .
5. Фактическое значение t-критерия для коэффициента регрессии  $\rho$  в полученном уравнении сравнивают с критическим значением, которое берётся из таблицы.

| $\alpha$ | $t_{крит}$ |
|----------|------------|
| 0,01     | -3,9       |
| 0,05     | -3,34      |
| 0,1      | -3,04      |

## 5.6. Модель распределённых лагов DL(q).

**Моделью распределённых лагов DL(q)** называется связь двух стационарных временных рядов вида

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t,$$

где случайная составляющая  $\varepsilon_t$  – нормально распределённый стационарный процесс с нулевым средним и нулевой автокорреляцией (“гауссов белый шум”).

В частности, **DL(1)** – это связь между временными рядами вида

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t,$$

Оценивание коэффициентов  $c, \beta_j$  модели распределённых лагов (в отличие от модели авторегрессии AR(p)), как правило, может быть произведено обычным методом наименьших квадратов, используя в качестве объясняющих переменных  $x_t, x_{t-1}, \dots$

Однако на практике может оказаться, что количество неизвестных параметров слишком велико.

Основная идея, поэтому, состоит в том, чтобы, отталкиваясь от содержательной сущности коэффициентов  $\beta_j$ , определить их структуру путём введения небольшого количества параметров, через которое они выражаются.

### 5.6.1. Модель полиномиальных лагов Алмон.

В этой модели предполагают, что зависимость  $\beta_j$  от  $j$  является полиномом некоторой степени  $r < q$ .

$$\beta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2 + \dots + \gamma_r j^r$$

После подстановки в основное уравнение

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t,$$

получается модель, содержащая  $r+2$  неизвестные  $c, \gamma_j$  вместо  $q+2$  неизвестных  $c, \beta_j$ .

$$y_t = c + \gamma_0 z_{0t} + \gamma_1 z_{1t} + \dots + \gamma_r z_{rt} + \varepsilon_t,$$

где

$$z_{0t} = x_t + x_{t-1} + \dots + x_{t-q},$$

$$z_{1t} = x_{t-1} + 2x_{t-2} + 3x_{t-3} + \dots + q \cdot x_{t-q},$$

$$z_{2t} = x_{t-1} + 2^2 x_{t-2} + 3^2 x_{t-3} + \dots + q^2 \cdot x_{t-q},$$

...

Коэффициенты последнего уравнения находятся обычным методом наименьших квадратов, а исходные выражаются через них по формулам

$$\beta_0 = \gamma_0,$$

$$\beta_1 = \gamma_0 + \gamma_1 + \dots + \gamma_q,$$

$$\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 + \dots + 2^q \gamma_q,$$

$$\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2 + \dots + 3^q \gamma_q,$$

...

**Пример.** Построить модель с распределёнными лагами

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t,$$

используя полиномиальные лаги степени 2.

**Решение.**  $\beta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2$

$$z_{0t} = x_t + x_{t-1} + x_{t-2} + x_{t-3},$$

$$z_{1t} = x_{t-1} + 2x_{t-2} + 3x_{t-3},$$

$$z_{2t} = x_{t-1} + 2^2 x_{t-2} + 3^2 x_{t-3}$$

$$z_{04} = 3 + 2 + 2 + 3 = 10,$$

$$z_{14} = 2 + 2 \cdot 2 + 3 \cdot 3 = 15,$$

$$z_{24} = 2 + 2^2 \cdot 2 + 3^2 \cdot 3 = 37, \dots$$

| t  | y <sub>t</sub> | z <sub>0t</sub> | z <sub>1t</sub> | z <sub>2t</sub> |
|----|----------------|-----------------|-----------------|-----------------|
| 1  | 3,0            | -               | -               | -               |
| 2  | 5,0            | -               | -               | -               |
| 3  | 2,0            | -               | -               | -               |
| 4  | 3,0            | 10              | 15              | 37              |
| 5  | 2,7            | 11              | 13              | 29              |
| 6  | 3,6            | 14              | 16              | 34              |
| 7  | 4,6            | 16              | 22              | 48              |
| 8  | 4,7            | 15              | 26              | 60              |
| 9  | 3,3            | 16              | 25              | 63              |
| 10 | 3,8            | 17              | 21              | 49              |
| 11 | 5,8            | 17              | 22              | 44              |
| 12 | 5,1            | 18              | 31              | 73              |

| t  | y <sub>t</sub> | x <sub>t</sub> |
|----|----------------|----------------|
| 1  | 3,0            | 3,0            |
| 2  | 5,0            | 2,0            |
| 3  | 2,0            | 2,0            |
| 4  | 3,0            | 3,0            |
| 5  | 2,7            | 4,0            |
| 6  | 3,6            | 5,0            |
| 7  | 4,6            | 4,0            |
| 8  | 4,7            | 2,0            |
| 9  | 3,3            | 5,0            |
| 10 | 3,8            | 6,0            |
| 11 | 5,8            | 4,0            |
| 12 | 5,1            | 3,0            |

Построив регрессию, получаем

$$y_t = 1,17 - 0,06z_{0t} + 0,71z_{1t} - 0,23z_{2t} + \delta_t,$$

откуда

$$\beta_0 = \gamma_0 = -0,06,$$

$$\beta_1 = \gamma_0 + \gamma_1 + \gamma_2 = -0,06 + 0,71 - 0,23 = 0,42,$$

$$\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 = -0,06 + 2 \cdot 0,71 - 4 \cdot 0,23 = 0,44,$$

$$\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2 = -0,06 + 3 \cdot 0,71 - 9 \cdot 0,23 = 0$$

Итак,

$$y_t = 1,17 - 0,06x_t + 0,42_1 x_{t-1} - 0,44x_{t-2} + \varepsilon_t,$$

### 5.6.2. Модель геометрических лагов Койка.

В этой модели предполагают, что зависимость  $\beta_j$  от  $j$  имеет вид

$$\beta_j = \beta_0 \cdot \lambda^j, \quad j = 0, 1, 2, \dots, \quad 0 < \lambda < 1$$

Здесь всего **два неизвестных параметра**,  $\beta_0$  и  $\lambda$ , но оценить их нелегко, так как зависимость от них нелинейная.

Один из вариантов – перебор различных  $\lambda$  из интервала  $0 < \lambda < 1$  с некоторым шагом, аналогично процедуре Хилдрета-Лу (3.3.3.2).

### 5.7. Модель ADL(p,q).

Эта модель содержит как частные случаи модели AR(p) и DL(q)

**Моделью ADL(p,q)** называется связь двух стационарных временных рядов вида

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t,$$

где случайная составляющая  $\varepsilon_t$  – гауссов белый шум.

### 5.8. Оценка параметров модели ADL(p,q).

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t,$$

В случае  $p=0$  получается модель MA(q), которую, как говорилось ранее, можно оценивать методом наименьших квадратов

Если  $p>0$ , то прежде всего следует проверить стационарность (например, используя тест Дики-Фуллера).

Далее, в правой части равенства есть переменная  $y_{t-1}$ , которая выражается через  $\varepsilon_{t-1}$ . Поэтому имеется корреляция между матрицами

$$\begin{pmatrix} 1 & y_{t-1} & \dots & x_{t-q} \\ 1 & y_{t-2} & \dots & x_{t-q-1} \\ \dots & \dots & \dots & \dots \end{pmatrix} \text{ и } \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \dots \end{pmatrix}$$

Поэтому оценки по методу наименьших квадратов будут смещёнными

Один из методов оценивания (метод инструментальных переменных), состоит в замене переменных  $y$  в правой части на их оценки  $\hat{y}$  в регрессии на переменные  $X$ :

$$\hat{Y} = d + e \cdot X$$

**Пример.** Построить модель ADL(1,0)

$$y_t = c + \alpha_1 y_{t-1} + \beta_0 x_t + \varepsilon_t,$$

**Решение.** Решим при помощи инструментальной переменной

1 шаг. Построим регрессию  $y_t = d + e x_t$

Обычным образом находим уравнение  $y_t = 11,9029 - 1,2621x_t$   
(0,99) (0,2)

Из него вычисляем значения инструментальной переменной Z, которая будет заменять Y в нужном нам уравнении

$$z_1 = 11,9029 - 1,2621 \cdot 3 = 8,12$$

$$z_2 = 11,9029 - 1,2621 \cdot 4 = 6,85, \dots$$

2 шаг. Построим регрессию

$$y_t = c + \alpha_1 z_{t-1} + \beta_0 x_t + \varepsilon_t,$$

Обычным образом находим уравнение

$$y_t = 8,67 + 0,28z_{t-1} - 0,99x_t$$

(1,19) (0,13) (0,15)

Отсюда находим уравнение модели ADL(1,0):

$$y_t = 8,67 + 0,28y_{t-1} - 0,99x_t + \varepsilon_t$$

(1,19) (0,13) (0,15)

| t | y <sub>t</sub> | x <sub>t</sub> |
|---|----------------|----------------|
| 1 | 9              | 3              |
| 2 | 7              | 4              |
| 3 | 6              | 5              |
| 4 | 6              | 4              |
| 5 | 5              | 6              |
| 6 | 6              | 4              |
| 7 | 4              | 6              |
| 8 | 3              | 7              |

| t | y <sub>t</sub> | x <sub>t</sub> | z <sub>t</sub> |
|---|----------------|----------------|----------------|
| 1 | 9              | 3              | 8,12           |
| 2 | 7              | 4              | 6,85           |
| 3 | 6              | 5              | 5,59           |
| 4 | 6              | 4              | 6,85           |
| 5 | 5              | 6              | 4,33           |
| 6 | 6              | 4              | 6,85           |
| 7 | 4              | 6              | 4,33           |
| 8 | 3              | 7              | 3,07           |

| t | y <sub>t</sub> | x <sub>t</sub> | z <sub>t-1</sub> |
|---|----------------|----------------|------------------|
| 1 | 9              | 3              | -                |
| 2 | 7              | 4              | 8,12             |
| 3 | 6              | 5              | 6,85             |
| 4 | 6              | 4              | 5,59             |
| 5 | 5              | 6              | 6,85             |
| 6 | 6              | 4              | 4,33             |
| 7 | 4              | 6              | 6,85             |
| 8 | 3              | 7              | 4,33             |

### 5.9. Тест Гранжера (Granger) на причинно-следственную зависимость.

В экономике нередко возникает вопрос о причинно-следственной связи между факторами. Например, верно ли, что увеличение цен на нефть влечёт за собой инфляцию?

Идея теста такова: если X является причиной Y, то изменения X предшествуют изменениям Y, но не наоборот.

Другими словами, регрессия с лагами Y на X должна быть значимой, а регрессия с лагами X на Y – незначимой. Если же обе регрессии значимы, то, скорее всего, есть другой фактор, влияющий и на X и на Y.

Чтобы тестировать гипотезу  $H_0$ : “X не влияет на Y”, рассматривается ARMA модель вида

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t,$$

Гипотеза  $H_0$  записывается в виде  $\beta_1 = \beta_2 = \dots = \beta_q = 0$ . Проверяется она обычным F-тестом.

Чтобы тестировать гипотезу  $H_0$ : “Y не влияет на X”, рассматривается аналогичная модель

$$x_t = c + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \dots + \beta_q y_{t-q} + \varepsilon_t,$$

Если первая гипотеза  $H_0$  отклоняется, а вторая принимается, то, возможно, X является причиной для Y

### 6. Системы эконометрических уравнений.

#### 6.1. Системы одновременных уравнений.

При моделировании часто приходится вводить не одно, а несколько связанных между собой эконометрических уравнений, т.е. описывать модель **системой уравнений**.

*Система одновременных уравнений* — это система эконометрических уравнений, содержащая взаимозависимые переменные

Примером системы одновременных уравнений может служить следующая гипотетическая модель:

$$\begin{cases} Y_1 = a_0 + a_1 Y_2 + a_2 X_1 + \varepsilon_1 \\ Y_2 = b_0 + b_1 Y_1 + b_2 X_2 + \varepsilon_2 \end{cases}$$

Классическим примером является одновременное формирование спроса  $Q^D$  и предложения  $Q^S$  товара в зависимости от его цены P и доходов I

$$\begin{cases} Q^S_t = a_0 + a_1 P_t + \varepsilon_1 \\ Q^D_t = b_0 + b_1 P_t + b_2 I_t + \varepsilon_2 \end{cases}$$

В ситуации равновесия на рынке  $Q^D = Q^S$  и получаем систему уравнений

$$\begin{cases} Q_t^D = a_0 + a_1 P_t + \varepsilon_1 \\ Q_t^S = b_0 + b_1 P_t + b_2 I_t + \varepsilon_2 \\ Q_t^D = Q_t^S \end{cases}$$

Уравнения, не содержащие неизвестных коэффициентов (как третье в примере), называются тождествами

## 6.2. Структурная и приведённая форма записи системы уравнений.

Роль переменных в системе уравнений различна.

**Эндогенные** переменные – это **зависимые** переменные, значения которых определяются **внутри** модели (системы). Как правило, их количество равно числу уравнений в системе.

**Экзогенные** переменные – это **независимые** переменные, значения которых формируются вне модели.  
С математической точки зрения их особенность в том, что они не коррелируют с ошибками измерений  $\varepsilon$ .

Разделение переменных на эндогенные и экзогенные зависит от теоретической концепции.

Внеэкономические переменные (например, климатические условия) являются экзогенными.

Близкую к экзогенным роль играют значения эндогенных переменных в предыдущие моменты времени (так как они уже сформировались), т.е. лаговые переменные.

Экзогенные и лаговые эндогенные переменные называются предопределёнными

Так, потребление текущего года  $Y_t$  зависит не только от одновременных экономических факторов, но и от потребления предыдущего года  $Y_{t-1}$  (и это уже предопределённая переменная).

Исходная форма записи системы, когда в правой присутствуют и предопределённые и эндогенные переменные, называется структурной.

Как правило, каждое уравнение в левой части содержит свою эндогенную переменную, а в правой – какие-то другие переменные, эндогенные и предопределённые

$$\begin{cases} y_1 = a_{12}y_2 + \dots + b_{11}x_1 + b_{12}x_2 + \dots, \\ y_2 = a_{21}y_1 + \dots + b_{21}x_1 + b_{22}x_2 + \dots, \\ \dots, \\ y_m = a_{m1}y_1 + \dots + b_{m1}x_1 + b_{m2}x_2 + \dots \end{cases}$$

Из-за наличия в правой части эндогенных переменных для нахождения коэффициентов нельзя применять МНК непосредственно к структурной форме.

С этой целью её преобразуют в приведённую форму.

В приведённой форме записи эндогенные переменные выражены через предопределённые, в каждом уравнении в левой части стоит своя эндогенная переменная.

$$\begin{cases} y_1 = \pi_{11}x_1 + \pi_{12}x_2 + \dots, \\ y_2 = \pi_{21}x_1 + \pi_{22}x_2 + \dots, \\ \dots, \\ y_m = \pi_{m1}x_1 + \pi_{m2}x_2 + \dots \end{cases}$$

**Пример.** Рассмотрим модель 
$$\begin{cases} Q_t = b_0 + a_1P_t + \varepsilon_1 \\ Q_t = b_1 + a_2P_t + b_2I_t + \varepsilon_2 \end{cases}$$

Здесь  $Q$  – равновесный спрос-предложение (эндогенная переменная),  $P$  – цена (эндогенная переменная),  $I$  – доход (экзогенная переменная).

Это структурная форма записи, числа  $a_j, b_j$  – структурные коэффициенты.

Чтобы получить приведённую форму записи, выразим  $P, Q$  через  $I$ .

Перенесём все переменные в левую часть

$$\begin{cases} Q_t - a_1P_t - b_0 = \varepsilon_1 \\ Q_t - a_2P_t - b_1 - b_2I_t = \varepsilon_2 \end{cases}$$

и запишем в матричной форме

$$AY + BX = \varepsilon,$$

$$A = \begin{pmatrix} 1 & -a_1 \\ 1 & -a_2 \end{pmatrix}, Y = \begin{pmatrix} Q_t \\ P_t \end{pmatrix}, B = \begin{pmatrix} -b_0 & 0 \\ -b_1 & -b_2 \end{pmatrix}, X = \begin{pmatrix} 1 \\ I_t \end{pmatrix},$$

Если существует обратная матрица

$$A^{-1} = \begin{pmatrix} 1 & -a_1 \\ 1 & -a_2 \end{pmatrix}^{-1}$$

то, домножив на неё равенство  $AX+BY=\varepsilon$ , получим

$$Y + A^{-1}BX = A^{-1}\varepsilon \Rightarrow$$

$$Y = PX + \xi, \quad \xi = B^{-1}\varepsilon,$$

где

$$P = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_3 & \pi_4 \end{pmatrix} = -A^{-1}B$$

т.е. получаем **приведённую форму** модели

$$\begin{cases} Q_t = \pi_1 + \pi_2 I_t + \xi_{1t} \\ P_t = \pi_3 + \pi_4 I_t + \xi_{2t} \end{cases}$$

В правой части только предопределённые переменные, так что оценку коэффициентов здесь уже можно проводить методом наименьших квадратов

### 6.3. Косвенный метод наименьших квадратов.

#### 6.3.1. Идея метода

Попробуем применить подход из предыдущего примера в общем случае.

1) Пусть исходная система дана в структурной форме

$$AY + BX = \varepsilon,$$

2) Перепишем её (если возможно) в приведённой форме

$$Y = PX + \xi, \quad P = -A^{-1}B$$

3) Рассчитаем оценки коэффициентов матрицы  $P$  методом наименьших квадратов после чего,

4) Выразим через них **(если это окажется возможным!)** исходные структурные коэффициенты (т.е. матрицы  $A$  и  $B$ ).

Такой способ оценивания структурных коэффициентов называется косвенным методом наименьших квадратов.

### 6.3.2. Проблема идентификации.

При использовании косвенного метода наименьших квадратов возникает проблема идентификации: удастся ли выразить коэффициенты исходной модели через коэффициенты приведённой модели?

**Пример.** Рассмотрим модель 
$$\begin{cases} Q_t = a_0 + a_1 P_t + \varepsilon_1 \\ Q_t = b_0 + b_1 P_t + b_2 I_t + \varepsilon_2 \end{cases}$$

В приведённой форме она имеет вид

$$\begin{cases} Q_t = c_0 + c_1 I_t + \xi_{1t} \\ P_t = c_2 + c_3 I_t + \xi_{2t} \end{cases}$$

Ясно, что **пять** неизвестных параметров исходной модели  $a_0, a_1, b_0, b_1, b_2$  невозможно выразить через **четыре** коэффициента приведённой модели  $c_0, c_1, c_2, c_3$ .

- ❖ Структурный параметр называется **идентифицируемым**, если он может быть **однозначно выражен** через коэффициенты приведённой формы
- ❖ Уравнение называется **идентифицируемым**, если идентифицируемы все входящие в него коэффициенты
- ❖ Структурный параметр называется **неидентифицируемым**, если его **невозможно выразить** через коэффициенты приведённой формы.
- ❖ Структурный параметр называется **сверхидентифицируемым**, если он **может быть выражен** через коэффициенты приведённой формы **несколькими различными способами**.

В случае неидентифицируемости косвенный метод наименьших квадратов, очевидно, неприменим.



### 6.3.3. Признаки идентифицируемости

Пусть исходная система дана в структурной форме

$$AY + BX = \varepsilon,$$

всего  $m$  уравнений,  $m_1$  эндогенных переменных  $Y$  и  $p$  предопределённых переменных  $X$ .

Все тождества можно исключить из системы уравнений.

#### Первое необходимое условие идентифицируемости.

Число уравнений системы должно быть равно количеству эндогенных переменных:

$$m = m_1$$

и матрица  $A$  обратима.

В этом случае приведённая форма системы имеет вид:

$$Y = \Pi X + \xi, \quad \Pi = -A^{-1}B$$

#### Второе необходимое условие идентифицируемости.

Матрица наблюдений предопределённых переменных  $X$  должна иметь ранг  $p$  (т.е. ранг, равный количеству этих переменных).

#### Третье необходимое условие идентифицируемости.

Наборы переменных в различных уравнениях системы должны быть разными.

#### Четвёртое необходимое условие идентифицируемости.

Для каждого уравнения системы количество отсутствующих в нём предопределённых переменных должно быть не меньше числа включённых в него эндогенных переменных минус 1.

Если равно, то это уравнение идентифицируемо, если больше, то сверхидентифицируемо.

Четвёртое необходимое условие называется «правило порядка». Сформулируем его ещё раз.

Общий вид  $j$ -го уравнения модели в структурной форме можно записать как ( $Y$  – эндогенные переменные,  $X$  – предопределённые):

$$a_{j1} \cdot y_{1t} + \dots + a_{jm} \cdot y_{mt} + b_{j1} \cdot x_{1t} + \dots + b_{jp} \cdot x_{pt} = \varepsilon_{jt}$$

Коэффициент при  $y_{jt}$  обычно равен 1 («правило нормировки»).

Часть коэффициентов может быть равна нулю.

Обозначим через  $cnt_j$  количество на самом деле присутствующих в  $j$ -м уравнении переменных.

**Теорема (правило порядка).** Пусть  $j$ -ое уравнение системы идентифицируемо или сверхидентифицируемо,  $p$  – количество предопределённых переменных в системе. Тогда справедливо неравенство

$$p \geq cnt_j - 1$$

Сверхидентифицируемости соответствует строгое неравенство, идентифицируемости – равенство.

**Пример.** Рассмотрим следующую эконометрическую модель:

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 S_t + a_3 t + \varepsilon_{1t}, \\ I_t = b_0 + b_1 Y_{t-1} + \varepsilon_{2t}, \\ S_t = c_0 + c_1 Y_t + c_2 Y_{t-1} + \varepsilon_{3t}, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где  $C_t$  – расходы на конечное потребление в период  $t$ ;

$Y_t, Y_{t-1}$  – совокупный доход в периоды  $t$  и  $t-1$  соответственно;

$I_t$  – валовые инвестиции периода  $t$ ;

$S_t$  – расходы на зарплату в период  $t$ ;

$G_t$  – государственные расходы период  $t$ ;

$\varepsilon_1, \varepsilon_2, \varepsilon_3$  – случайные факторы.

В модели четыре эндогенных переменных:  $C_t, I_t, S_t, Y_t$ .

Остальные три переменные модели:  $1, t$  и  $G_t$  – экзогенные.

Кроме того, модель содержит лаговую эндогенную переменную  $Y_{t-1}$ . Таким образом, общее количество эндогенных переменных  $m = 4$ , количество предопределённых переменных  $p = 4$ .

Для первого уравнения общее количество включенных в него переменных  $\text{cnt}_1=5$  (в него входят переменные  $C_t, 1, Y_t, S_t, t$ ). Имеем:  $p=4, \text{cnt}_1-1=5-1=4$ .  $4=4$ , следовательно, первое уравнение идентифицировано.

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 S_t + a_3 t + \varepsilon_{1t}, \\ I_t = b_0 + b_1 Y_{t-1} + \varepsilon_{2t}, \\ S_t = c_0 + c_1 Y_t + c_2 Y_{t-1} + \varepsilon_{3t}, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

Для второго уравнения:  $\text{cnt}_2 = 3$  ( $I_t, 1, Y_{t-1}$ ). Имеем:  $\text{cnt}_2-1=3-1=2$ .  $p=4 > 2$ , следовательно, второе уравнение сверхидентифицировано.

Для третьего уравнения:  $\text{cnt}_3 = 4$  ( $S_t, 1, Y_t, Y_{t-1}$ ). Имеем:  $\text{cnt}_3 - 1 = 4 - 1 = 3$ .  $p=4 > 3$ , следовательно, третье уравнение сверхидентифицировано.

Последнее уравнение модели представляет собой тождество, его не надо проверять на идентификацию.

Приведенная форма модели будет иметь вид:

$$\begin{cases} C_t = \pi_0 + \pi_1 Y_{t-1} + \pi_2 G_t + \pi_3 t + \xi_{1t}, \\ I_t = \pi_4 + \pi_5 Y_{t-1} + \pi_6 G_t + \pi_7 t + \xi_{2t}, \\ S_t = \pi_8 + \pi_9 Y_{t-1} + \pi_{10} G_t + \pi_{11} t + \xi_{3t}, \\ Y_t = \pi_{12} + \pi_{13} Y_{t-1} + \pi_{14} G_t + \pi_{15} t + \xi_{4t}. \end{cases}$$

Здесь 12 неизвестных коэффициентов.

Когда они будут найдены, через них потребуется выразить 9 коэффициентов исходной модели ( $a_0, a_1, a_2, a_3, b_0, b_1, c_0, c_1, c_2$ ).

Эта процедура неоднозначна, что лишний раз подтверждает сверхидентифицируемость модели

Правило порядка – это необходимое условие идентифицируемости, но оно не является достаточным. Это означает, что, когда неравенство несправедливо, то  $j$ -ое уравнение заведомо неидентифицируемо. Однако при выполнении неравенства ещё нельзя сделать вывод о идентифицируемости данного уравнения. Достаточное условие выражается в терминах рангов матриц.

### **Необходимое и достаточное условие идентифицируемости.**

Уравнение идентифицируемо, если матрица, составленная по остальным уравнениям из коэффициентов отсутствующих в нём переменных имеет **ранг не меньше, чем общее количество эндогенных переменных минус 1**.

**Пример.** Проверим достаточное условие идентифицируемости.

Перенесём все в левую часть и выпишем матрицу коэффициентов при переменных

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 S_t + a_3 t + \varepsilon_{1t}, \\ I_t = b_0 + b_1 Y_{t-1} + \varepsilon_{2t}, \\ S_t = c_0 + c_1 Y_t + c_2 Y_{t-1} + \varepsilon_{3t}, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

Матрица коэффициентов при переменных, не входящих в первое уравнение:

| $C_t$ | 1      | $Y_t$  | $S_t$  | $t$    | $I_t$ | $Y_{t-1}$ | $G_t$ |
|-------|--------|--------|--------|--------|-------|-----------|-------|
| 1     | $-a_0$ | $-a_1$ | $-a_2$ | $-a_3$ | 0     | 0         | 0     |
| 0     | $-b_0$ | 0      | 0      | 0      | 1     | $-b_1$    | 0     |
| 0     | $-c_0$ | $-c_1$ | 1      | 0      | 0     | $-c_2$    | 0     |
| -1    | 0      | 1      | 0      | 0      | -1    | 0         | -1    |

$$\begin{pmatrix} 1 & -b_1 & 0 \\ 0 & -c_2 & 0 \\ -1 & 0 & -1 \end{pmatrix}$$

Её определитель не равен 0, так что ранг равен 3. Общее количество эндогенных переменных равно  $m=4$ . Имеем:  $4-1=3$ , так что первое уравнение идентифицируемо.

Матрица коэффициентов при переменных, не входящих в второе уравнение:

$$\begin{pmatrix} -1 & a_1 & a_2 & a_3 & 0 \\ 0 & c_1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}$$

|       |        |        |        |        |       |           |       |
|-------|--------|--------|--------|--------|-------|-----------|-------|
| $C_t$ | 1      | $Y_t$  | $S_t$  | $t$    | $I_t$ | $Y_{t-1}$ | $G_t$ |
| 1     | $-a_0$ | $-a_1$ | $-a_2$ | $-a_3$ | 0     | 0         | 0     |
| 0     | $-b_0$ | 0      | 0      | 0      | 1     | $-b_1$    | 0     |
| 0     | $-c_0$ | $-c_1$ | 1      | 0      | 0     | $-c_2$    | 0     |
| -1    | 0      | 1      | 0      | 0      | -1    | 0         | -1    |

Ранг матрицы равен 3, так как определитель следующей подматрицы 3x3 не равен нулю. Второе уравнение идентифицируемо.

$$\begin{pmatrix} -1 & a_2 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Матрица коэффициентов при переменных, не входящих в третье уравнение:

$$\begin{pmatrix} -1 & a_3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Матрица коэффициентов при переменных, не входящих в четвёртое уравнение:

$$\begin{pmatrix} -a_0 & -a_2 & -a_3 & 0 \\ -b_0 & 0 & 0 & -b_1 \\ -c_0 & -1 & 0 & -c_2 \end{pmatrix} \dots$$

Ранг матрицы аналогичным образом равен 3. Третье уравнение идентифицируемо.

### 6.3.4. Косвенный метод наименьших квадратов

Подведём итоги.



Косвенный метод наименьших квадратов (КМНК) применяется в случае точной идентифицируемости уравнений модели

Алгоритм применения КМНК:

1. От структурной формы модели переходят к приведенной
2. Определяются МНК-оценки параметров приведенной формы модели
3. По МНК-оценкам приведенной формы вычисляются оценки параметров структурной формы модели.

Исходная система в структурной форме:  $AY + BX = \varepsilon$ ,

В приведённой форме:  $Y = \Pi X + \xi$ ,  $\Pi = -A^{-1}B$

Следовательно,

$$A\Pi = A \cdot (-A^{-1}B) = -B$$

Это – система  $m \times p$  уравнений.

Из неё, в случае идентифицируемости, по матрице  $\Pi$ , можно вычислить значения структурных коэффициентов  $A$  и  $B$ .

**Задача.** Дана модель в структурной форме. Найти оценки структурных коэффициентов по исходным данным косвенным методом наименьших квадратов

$$\begin{cases} y_{1t} = c_{10} + a_{12} y_{2t} + b_{11} x_{1t} + \varepsilon_{1t}, \\ y_{2t} = c_{20} + a_{21} y_{1t} + b_{22} x_{2t} + \varepsilon_{2t}, \end{cases}$$

| $t$ | $y_1$ | $y_2$ | $x_1$ | $x_2$ |
|-----|-------|-------|-------|-------|
| 1   | 22    | 24    | 2     | 7     |
| 2   | 24    | 15    | 3     | 5     |
| 3   | 17    | 23    | 2     | 5     |
| 4   | 19    | 21    | 3     | 4     |
| 5   | 22    | 21    | 4     | 5     |
| 6   | 25    | 13    | 5     | 3     |
| 7   | 25    | 15    | 4     | 4     |
| 8   | 31    | 12    | 6     | 3     |
| 9   | 33    | 8     | 7     | 2     |
| 10  | 31    | 5     | 8     | 1     |

**Решение.** 1) Проверим на идентифицируемость.

Перенесём все в левую часть и выпишем матрицу коэффициентов при переменных

|           |           |           |           |           |
|-----------|-----------|-----------|-----------|-----------|
| $y_1$     | $y_2$     | 1         | $x_1$     | $x_2$     |
| 1         | $-a_{12}$ | $-c_{10}$ | $-b_{11}$ | 0         |
| $-a_{21}$ | 1         | $-c_{20}$ | 0         | $-b_{22}$ |

Количество эндогенных переменных  $m=2$  ( $y_1$  и  $y_2$ ).

Матрица коэффициентов при переменных, не входящих в первое уравнение состоит из одного числа  $-b_{21}$ , ранг = 1 =  $m-1$ , уравнение точно идентифицируемо. Аналогично со вторым уравнением.

2) В приведенной форме модель примет вид:

$$\begin{cases} y_{1t} = \pi_{10} + \pi_{11}x_{1t} + \pi_{12}x_{2t} + \xi_{2t}, \\ y_{2t} = \pi_{20} + \pi_{21}x_{1t} + \pi_{22}x_{2t} + \xi_{2t}, \end{cases}$$

Оценивая уравнения по отдельности обычным МНК, получаем

$$\begin{cases} y_{1t} = 0,71 + 3,8x_{1t} + 1,91x_{2t}, & R^2 = 0,88 \\ y_{2t} = 20,44 - 2,06x_{1t} + 1,11x_{2t}, & R^2 = 0,88 \end{cases}$$

| t  | y <sub>1</sub> | y <sub>2</sub> | x <sub>1</sub> | x <sub>2</sub> |
|----|----------------|----------------|----------------|----------------|
| 1  | 22             | 24             | 2              | 7              |
| 2  | 24             | 15             | 3              | 5              |
| 3  | 17             | 23             | 2              | 5              |
| 4  | 19             | 21             | 3              | 4              |
| 5  | 22             | 21             | 4              | 5              |
| 6  | 25             | 13             | 5              | 3              |
| 7  | 25             | 15             | 4              | 4              |
| 8  | 31             | 12             | 6              | 3              |
| 9  | 33             | 8              | 7              | 2              |
| 10 | 31             | 5              | 8              | 1              |

3) Записываем равенство для нахождения структурных коэффициентов

$$A\Pi = -B$$

$$\begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \begin{pmatrix} 0,71 & 3,8 & 1,91 \\ 20,44 & -2,06 & 1,11 \end{pmatrix} = - \begin{pmatrix} -c_{10} & -b_{11} & 0 \\ -c_{20} & 0 & -b_{22} \end{pmatrix}$$

Из шести уравнений есть два с нулевой правой частью, откуда находятся коэффициенты матрицы А.

$$\begin{cases} 1,91 - 1,11a_{12} = 0 \\ 3,8a_{21} + 2,06 = 0 \end{cases} \Rightarrow \begin{cases} a_{12} \approx 1,72 \\ a_{21} \approx -0,54 \end{cases} \Rightarrow A = \begin{pmatrix} 1 & -1,72 \\ 0,54 & 1 \end{pmatrix}$$

Далее находим матрицу В

$$\begin{pmatrix} c_{10} & b_{11} & 0 \\ c_{20} & 0 & b_{22} \end{pmatrix} = \begin{pmatrix} 1 & -1,72 \\ 0,54 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0,71 & 3,8 & 1,91 \\ 20,44 & -2,06 & 1,11 \end{pmatrix} = \begin{pmatrix} -34,44 & 7,34 & 0 \\ 20,82 & 0 & 2,14 \end{pmatrix}$$

Теперь выписываем модель в структурной форме.

$$\begin{cases} y_{1t} = c_{10} + a_{12}y_{2t} + b_{11}x_{1t} + \varepsilon_{1t}, \\ y_{2t} = c_{20} + a_{21}y_{1t} + b_{22}x_{2t} + \varepsilon_{2t}, \end{cases}$$

$$\begin{cases} y_{1t} = -34,44 + 1,72y_{2t} + 7,34x_{1t} + \varepsilon_{1t}, \\ y_{2t} = 20,82 - 0,54y_{1t} + 2,14x_{2t} + \varepsilon_{2t}, \end{cases}$$

**Задача.** Дана модель в структурной форме

$$\begin{cases} Y_1 = a_0 + a_1X_1 + a_2X_2 + a_3Y_3 + \varepsilon_{1t}, \\ Y_2 = b_0 + b_1X_2 + b_2X_3 + b_3Y_1 + \varepsilon_{2t}, \\ Y_3 = c_0 + c_1X_1 + c_2X_3 + \varepsilon_{3t} \end{cases}$$

В приведённой форме получены коэффициенты

$$\begin{cases} Y_1 = 4 + 7X_1 + 9X_2 + 4X_3 + \varepsilon_{1t}, \\ Y_2 = 6 - 12X_1 - 7X_2 + 8Y_1 + \varepsilon_{2t}, \\ Y_3 = 10 - 5X_1 - 2X_2 + 5X_3 + \varepsilon_{3t} \end{cases}$$

Определить значения всех структурных коэффициентов, которые можно найти.

**Решение.**

Общий вид структурной формы  $AY + BX = \varepsilon$ ,

Общий вид приведённой формы  $Y = \Pi X + \xi$ ,

Выпишем матрицы А, В

$$A = \begin{pmatrix} 1 & 0 & -a_3 \\ -b_3 & 1 & 0 \\ -c_3 & 0 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} -a_0 & -a_1 & -a_2 & 0 \\ -b_0 & 0 & -b_1 & -b_2 \\ -c_0 & -c_1 & 0 & -c_3 \end{pmatrix}$$

$$\begin{cases} Y_1 = a_0 + a_1X_1 + a_2X_2 + a_3Y_3 + \varepsilon_{1t}, \\ Y_2 = b_0 + b_1X_2 + b_2X_3 + b_3Y_1 + \varepsilon_{2t}, \\ Y_3 = c_0 + c_1X_1 + c_2X_3 + c_3Y_1 + \varepsilon_{3t} \end{cases}$$

Выпишем матрицу П

$$\Pi = \begin{pmatrix} 4 & 6 & 1 & 4 \\ 6 & -12 & -7 & 8 \\ 10 & -5 & -2 & 5 \end{pmatrix}$$

$$\begin{cases} Y_1 = 4 + 6X_1 + X_2 + 4X_3 + \varepsilon_{1t}, \\ Y_2 = 6 - 12X_1 - 7X_2 + 8Y_1 + \varepsilon_{2t}, \\ Y_3 = 10 - 5X_1 - 2X_2 + 5X_3 + \varepsilon_{3t} \end{cases}$$

3) Запишем равенство для нахождения структурных коэффициентов

$$A\Pi = -B$$

$$\begin{pmatrix} 1 & 0 & -a_3 \\ -b_3 & 1 & 0 \\ -c_3 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 6 & 1 & 4 \\ 6 & -12 & -7 & 8 \\ 10 & -5 & -2 & 5 \end{pmatrix} = \begin{pmatrix} a_0 & a_1 & a_2 & 0 \\ b_0 & 0 & b_1 & b_2 \\ c_0 & c_1 & 0 & c_3 \end{pmatrix}$$

Из 12 уравнений три имеют правую часть равную нулю.

$$\begin{cases} 4 - 5a_3 = 0, \\ -6b_3 - 12 = 0, \\ -c_3 - 2 = 0 \end{cases}$$

Отсюда находим все коэффициенты матрицы А.

$$\begin{cases} a_3 = 0,8, \\ b_3 = -2, \\ c_3 = -2 \end{cases}$$

Перемножая матрицы, находим остальные коэффициенты:

$$\begin{pmatrix} a_0 & a_1 & a_2 & 0 \\ b_0 & 0 & b_1 & b_2 \\ c_0 & c_1 & 0 & c_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -0,8 \\ 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 6 & 1 & 4 \\ 6 & -12 & -7 & 8 \\ 10 & -5 & -2 & 5 \end{pmatrix}$$

### 6.4. Двухшаговый метод наименьших квадратов

Как уже отмечалось, если система сверхидентифицируема, то косвенный МНК применить не удаётся, так как структурные коэффициенты не выражаются однозначным способом через коэффициенты приведённой системы

**Пример.** Рассмотрим следующую модель в структурной форме.

$$\begin{cases} y_{1t} = a_1 y_{2t} + b_1 x_{1t} + b_2 x_{2t} + \varepsilon_{1t}, \\ y_{2t} = a_2 y_{1t} + b_3 x_{3t} + \varepsilon_{2t}, \end{cases}$$

В приведённой форме она будет иметь вид

$$\begin{cases} y_{1t} = \pi_1 x_{1t} + \pi_2 x_{2t} + \pi_3 x_{3t} + \xi_{1t}, \\ y_{2t} = \pi_4 x_{1t} + \pi_5 x_{2t} + \pi_6 x_{3t} + \xi_{2t}, \end{cases}$$

Пусть получены следующие оценки параметров приведённой формы

$$\begin{cases} y_{1t} = 2 x_{1t} + 3 x_{2t} + 3 x_{3t}, \\ y_{2t} = 3 x_{1t} + 4 x_{2t} + 2 x_{3t}, \end{cases}$$

Попробуем найти оценки структурной формы:

$$AP = -B$$

$$\begin{pmatrix} 1 & -a_1 \\ -a_2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 3 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} -b_1 & -b_2 & 0 \\ 0 & 0 & -b_{23} \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & 0 \\ 0 & 0 & b_{23} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -a_1 \\ -a_2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 3 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & 0 \\ 0 & 0 & b_{23} \end{pmatrix}$$

Из шести уравнений есть три с нулевой правой частью, откуда получается три уравнения для нахождения  $a_1, a_2$ .

$$\begin{cases} -2a_2 + 3 = 0 \\ -3a_2 + 4 = 0 \\ 3 - 2a_1 = 0 \end{cases}$$

Видно, что  $a_2$  выражается двумя способами, противоречащими друг другу – это и есть сверхидентифицируемость.

### Алгоритм оценки коэффициентов структурной формы двухшаговым МНК

1. Оцениваются параметры приведенной формы модели при помощи МНК
2. Оцениваются параметры структурной формы модели, в правую часть которой вместо значений эндогенных переменных подставляются их оценки, рассчитанные по приведенной форме

**Пример (продолжение).** Пусть в предыдущем примере исходные данные имели вид:

| t  | y <sub>1</sub> | y <sub>2</sub> | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> |
|----|----------------|----------------|----------------|----------------|----------------|
| 1  | 33,4           | 40,0           | 2              | 7              | 3              |
| 2  | 24,0           | 30,3           | 3              | 5              | 1              |
| 3  | 21,3           | 28,6           | 2              | 5              | 1              |
| 4  | 23,5           | 29,8           | 3              | 4              | 2              |
| 5  | 24,9           | 34,7           | 4              | 5              | 1              |
| 6  | 21,8           | 28,9           | 5              | 3              | 1              |
| 7  | 26,1           | 31,9           | 4              | 4              | 2              |
| 8  | 24,5           | 32             | 6              | 3              | 1              |
| 9  | 32,7           | 36,8           | 7              | 2              | 4              |
| 10 | 33,9           | 37,9           | 8              | 1              | 5              |

Подставляя в найденные формулы, мы можем найти расчётные значения:

$$\begin{cases} \hat{y}_{1t} = 2 x_{1t} + 3 x_{2t} + 3 x_{3t}, \\ \hat{y}_{2t} = 3 x_{1t} + 4 x_{2t} + 2 x_{3t}, \end{cases}$$

| t  | y <sub>1</sub> | y <sub>2</sub> | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> | $\hat{y}_1$ | $\hat{y}_2$ |
|----|----------------|----------------|----------------|----------------|----------------|-------------|-------------|
| 1  | 33,4           | 40,0           | 2              | 7              | 3              | 34          | 40          |
| 2  | 24,0           | 30,3           | 3              | 5              | 1              | 24          | 31          |
| 3  | 21,3           | 28,6           | 2              | 5              | 1              | 22          | 28          |
| 4  | 23,5           | 29,8           | 3              | 4              | 2              | 24          | 29          |
| 5  | 24,9           | 34,7           | 4              | 5              | 1              | 26          | 34          |
| 6  | 21,8           | 28,9           | 5              | 3              | 1              | 22          | 29          |
| 7  | 26,1           | 31,9           | 4              | 4              | 2              | 26          | 32          |
| 8  | 24,5           | 32             | 6              | 3              | 1              | 24          | 32          |
| 9  | 32,7           | 36,8           | 7              | 2              | 4              | 32          | 37          |
| 10 | 33,9           | 37,9           | 8              | 1              | 5              | 34          | 38          |

Эти рассчитанные значения используем в качестве новых переменных в правой части структурной формы

$$\begin{cases} y_{1t} = a_1 \hat{y}_{2t} + b_1 x_{1t} + b_2 x_{2t} + \varepsilon_{1t}, \\ y_{2t} = a_2 \hat{y}_{1t} + b_3 x_{3t} + \varepsilon_{2t}, \end{cases}$$

Опять обычным методом получаем коэффициенты структурной формы

$$\begin{cases} y_{1t} = 1,49 \hat{y}_{2t} - 2,4 x_{1t} - 3,09 x_{2t} + \varepsilon_{1t}, & R^2 = 0,999 \\ y_{2t} = 1,38 \hat{y}_{1t} - 1,9 x_{3t} + \varepsilon_{2t}, & R^2 = 0,999 \end{cases}$$

Таким образом, двухшаговый МНК даёт следующий ответ:

$$\begin{cases} y_{1t} = 1,49 y_{2t} - 2,4 x_{1t} - 3,09 x_{2t} + \varepsilon_{1t}, & R^2 = 0,999 \\ y_{2t} = 1,38 y_{1t} - 1,9 x_{3t} + \varepsilon_{2t}, & R^2 = 0,999 \end{cases}$$

#### **Замечания.**

- 1) В случае точной идентифицируемости двухшаговый МНК даёт те же результаты, что и косвенный МНК.
- 2) Двухшаговый МНК является частным случаем метода инструментальных переменных, когда “неудобные” эндогенные переменные в правой части заменяются на новые переменные.

## **Список литературы**

1. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. – М.: Дело, 2004.
2. Эконометрика: Учебник./Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2005.
3. Практикум по эконометрике: Учеб. пособие. /И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2005.
4. Кремер Н.Ш., Путко Б.А. Эконометрика: учебник для вузов. /Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ–ДАНА, 2002.
5. Бородич С.А. Эконометрика: Учеб. пособие. – Мн.: Новое знание, 2001.
6. Берндт Э. Р. Практика эконометрики: классика и современность: Учебник. – М.: ЮНИТИ–ДАНА, 2005.
7. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. – М.: ЮНИТИ, 1998.

## Оглавление

|   |     |
|---|-----|
| 1. Введение   | 3   |
| 2. Классическая модель множественной линейной регрессии | 9   |
| 3. Нарушения допущений классической линейной модели     | 33  |
| 4. Временные ряды                                       | 57  |
| 5. Взаимосвязь временных рядов                          | 94  |
| 6. Системы эконометрических уравнений                   | 106 |
| <br>  |     |
| Список литературы                                       | 122 |