

# **REGRESSION ANALYSIS**

**OLS: GAUSS-MARKOV ASSUMPTIONS**

**ПРЕДПОСЫЛКИ МНК.  
ТЕОРЕМА ГАУССА-МАРКОВА**

## **OLS: GAUSS-MARKOV ASSUMPTIONS**

- Основные предпосылки МНК ассоциируются с теоремой Гаусса-Маркова и представляют собой **перечень условий** для случайных отклонений эконометрической модели, выполнение которых обеспечивает эффективную статистическую проверку значимости параметров регрессии.
- Часть предпосылок выполняется априори, невыполнение другой части не приводит к существенным нарушениям желаемых свойств оценок, получаемых с помощью МНК .

# OLS: GAUSS-MARKOV ASSUMPTIONS

- Нарушение некоторых предпосылок относительно метода наименьших квадратов может привести к тому, что полученные оценки не будут обладать необходимыми свойствами *несмещенности*, *состоятельности* и *эффективности*, т.е. не являться BLUE-оценками (BLUE: Best Linear Unbiased Estimators).
- Выполнение предпосылок обеспечивает несмещенность полученных оценок и, как минимум, их состоятельность.
- Предпосылки для случайных отклонений модели дополняются условиями относительно ошибок спецификации

# OLS ASSUMPTION №1

- *Математическое ожидание случайного отклонения равно нулю для всех наблюдений:*

$$M(\varepsilon_i) = 0$$

- другими словами, случайное отклонение в среднем не оказывает влияния на зависимую переменную
- если уравнение регрессии включает коэффициент свободного члена (постоянную), то обычно это условие выполняется априори

## OLS ASSUMPTION №2

- *Дисперсия случайных отклонений постоянна для любых наблюдений:*

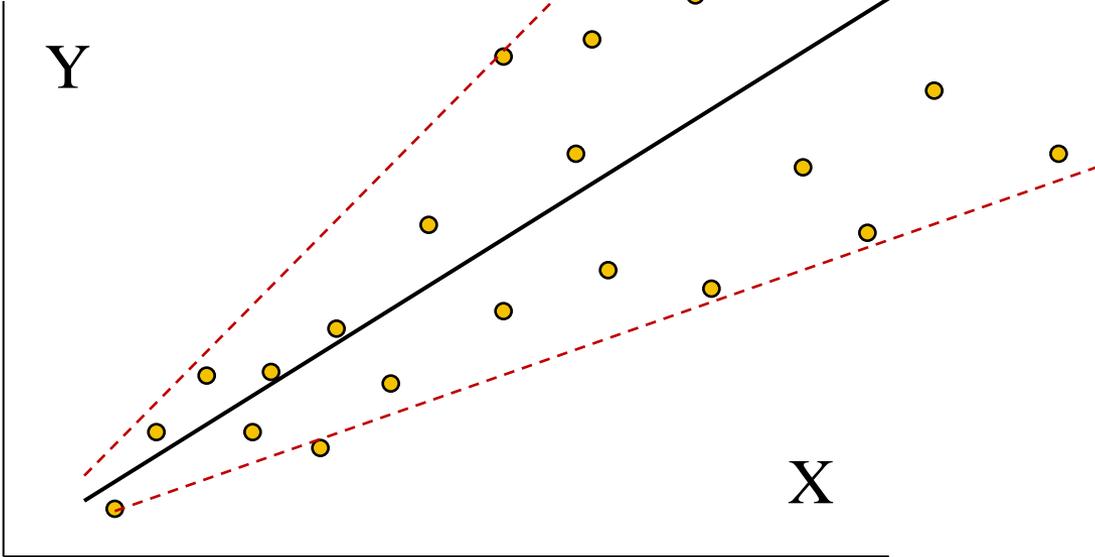
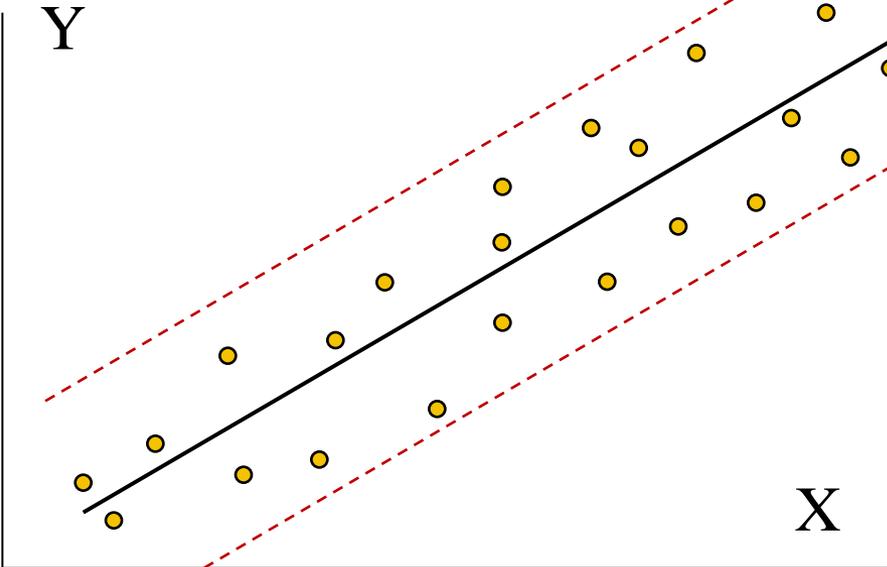
$$D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2 = const$$

- в случае выполнения предпосылки, т.е. постоянства дисперсии отклонений, говорят о *гомоскедастичности*, в противном случае, т.е. когда предпосылка не выполняется и дисперсия отклонения непостоянная – о *гетероскедастичности*

## OLS ASSUMPTION №2

- другими словами, несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, не должно быть некой априорной причины, вызывающей большее или меньшее отклонение, т.е. нельзя проследить закономерность в величинах отклонений
- *гетероскедастичность* случайных отклонений приводит к неэффективности оценок, полученных с помощью МНК (речь о коэффициентах, свойство несмещенности и состоятельности сохраняется, что позволяет, не смотря ни на что использовать МНК), а также к смещенной и несостоятельной оценке дисперсионно-ковариационной матрицы МНК (речь о значениях дисперсий коэффициентов, т.е. стандартных ошибках)

# HOMOSCEDASTICITY VS. HETEROSCEDASTICITY



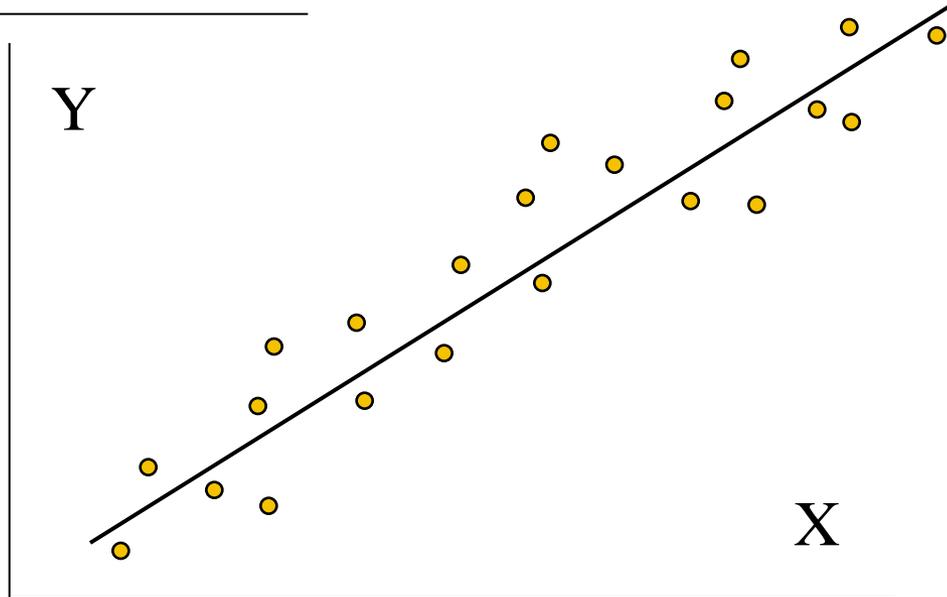
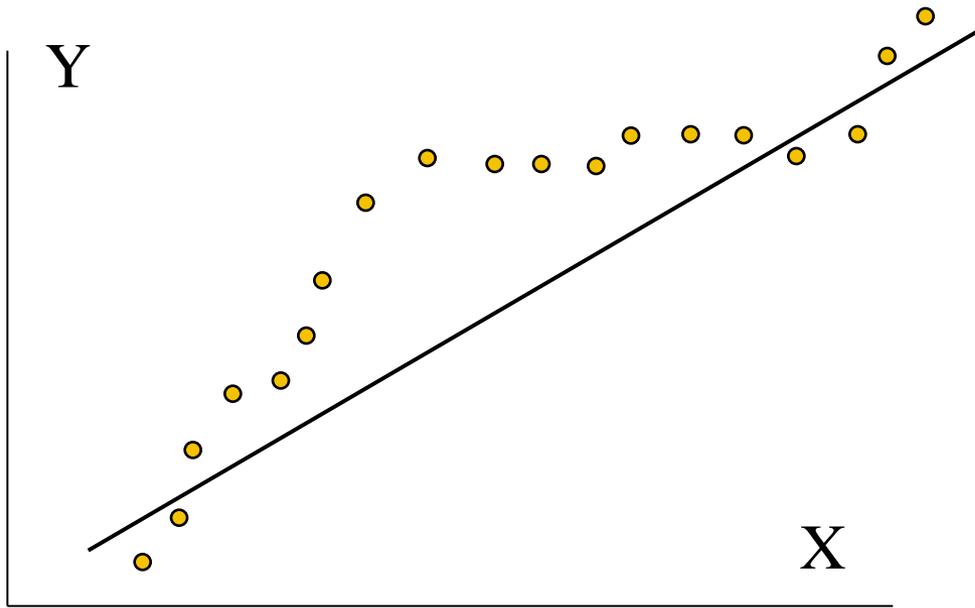
## OLS ASSUMPTION №3

- *Случайные отклонения являются независимыми друг от друга:*

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & \text{if } i \neq j \\ \sigma^2, & \text{if } i = j \end{cases}$$

- в случае невыполнения предпосылки, говорят об *автокорреляции* случайных отклонений
- автокорреляция – «корреляция внутри себя», корреляционная связь между значениями одного и того же случайного процесса  $x(t)$  в моменты времени  $t_1$  и  $t_2$

# AUTOCORRELATION



## OLS ASSUMPTION №3

- Другими словами, отсутствует систематическая связь между любыми случайными отклонениями, т.е. ни величина, ни определенный знак любого случайного отклонения не должны быть причинами величины и знака любого другого отклонения
- *автокорреляция* случайных отклонений приводит к неэффективности оценок, полученных с помощью МНК (речь о коэффициентах, свойство несмещенности и состоятельности сохраняется, что позволяет, не смотря ни на что использовать МНК), а также к смещенной и несостоятельной оценке дисперсионно-ковариационной матрицы МНК (речь о значениях дисперсий коэффициентов, т.е. стандартных ошибках)

## **OLS ASSUMPTIONS №4-5**

- **Случайное отклонение должно быть независимо от объясняющих переменных:**

$$\text{cov}(x_i, \varepsilon_j) = 0$$

- другими словами, случайное отклонение (или зависимая переменная) есть величина случайная, а объясняющая переменная - величина неслучайная
- **Модель является линейной относительно параметров.**

## **OLS ASSUMPTION №6-7**

- **Случайное отклонение должно быть распределено согласно нормальному закону распределения:**

$$\varepsilon_i \sim N(\mu, \sigma)$$

- **Должна отсутствовать высокая взаимная коррелированность экзогенных переменных:**
- **в случае невыполнения предпосылки, говорят о мультиколлинеарности**

# GAUSS-MARKOV THEOREM

**Теорема Гаусса-Маркова.** Если предпосылки 1-7 выполнены, то оценки, полученные по МНК, обладают следующими свойствами:

**(1) Оценки являются несмещенными**, т.е.  $M(b_0) = \beta_0$ ,  $M(b_1) = \beta_1$ , где  $b_0, b_1$  – коэффициенты эмпирического уравнения регрессии, полученные с помощью МНК, т.е. точечные оценки для теоретических коэффициентов  $\beta_0, \beta_1$ . Это обеспечивается выполнением первой предпосылки и говорит об отсутствии систематической ошибки в определении положения линии регрессии.

# GAUSS-MARKOV THEOREM

**Теорема Гаусса-Маркова.** Если предпосылки 1-7 выполнены, то оценки, полученные по МНК, обладают следующими свойствами:

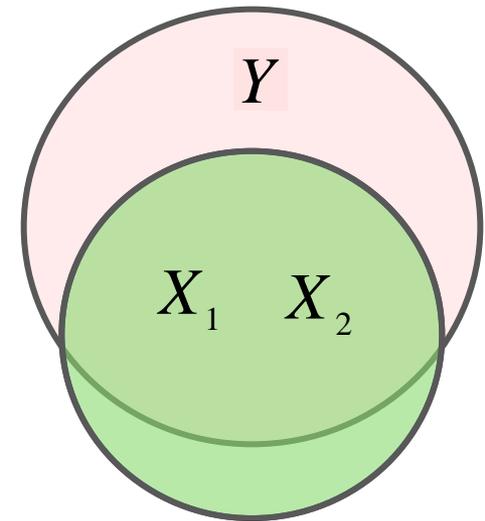
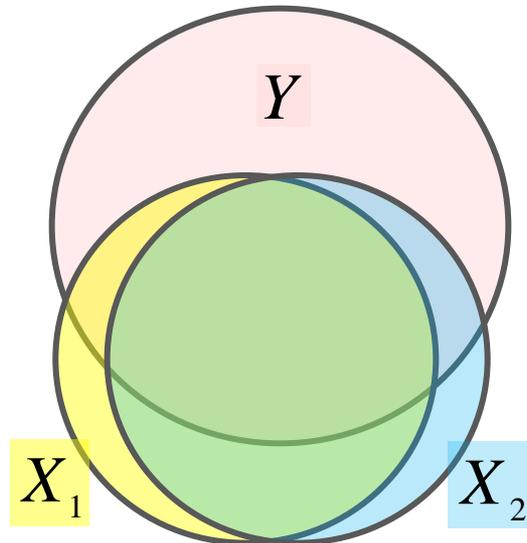
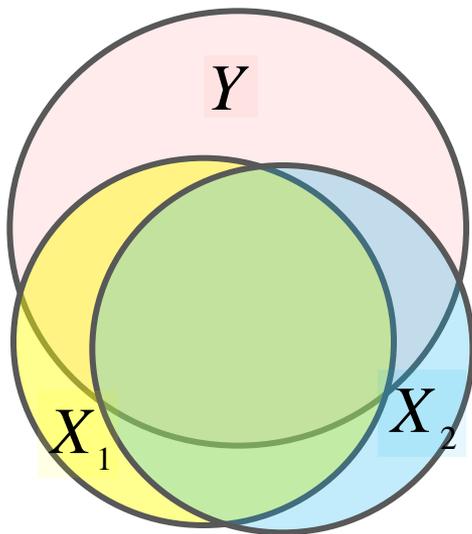
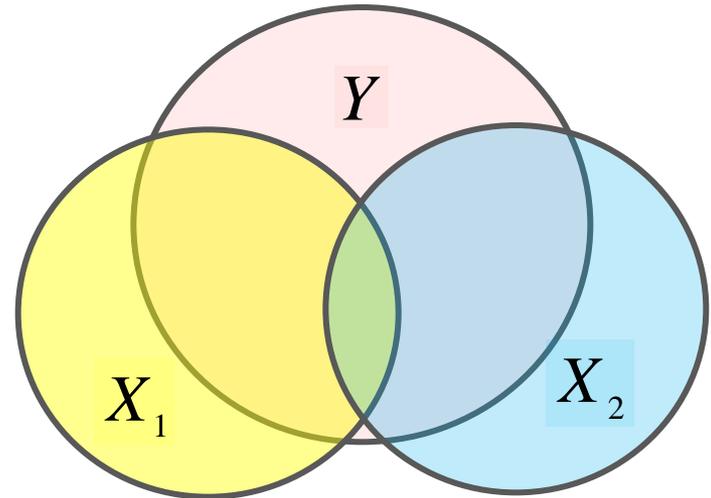
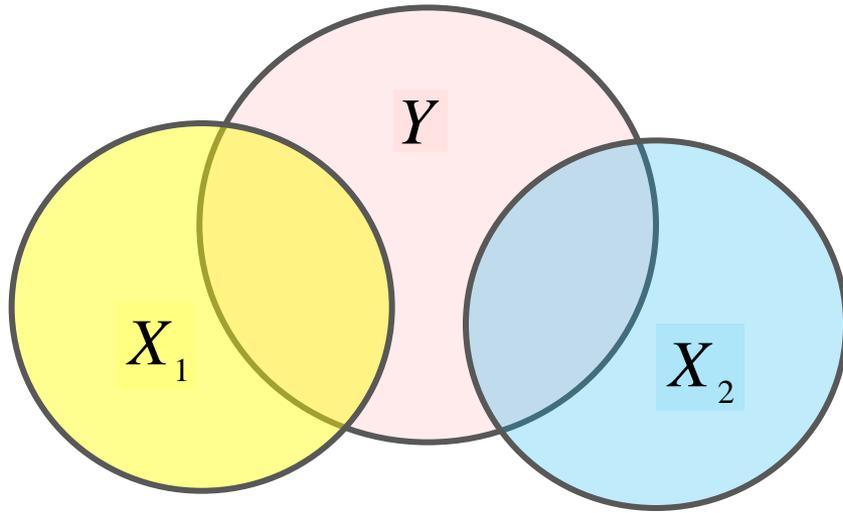
- (2) Оценки состоятельны,** так как дисперсия оценок параметров при возрастании числа  $n$  наблюдений стремится к нулю. Другими словами, при увеличении объема выборки надежность оценок увеличивается (коэффициенты теоретического и эмпирического уравнений регрессии практически совпадают).
- (3) Оценки эффективны,** то есть они имеют наименьшую дисперсию по сравнению с любыми оценками данных параметров, линейными относительно величин  $y_i$ .

# **REGRESSION ANALYSIS**

**OLS ASSUMPTIONS: MULTICOLLINEARITY**

**ПРЕДПОСЫЛКИ МНК:  
МУЛЬТИКОЛЛИНЕАРНОСТЬ**

# MULTICOLLINEARITY: VIENNA DIAGRAM



# MULTICOLLINEARITY

## *Методы выявления:*

- анализ матрицы коэффициентов парной корреляции (в случае  $m=2$ ) и матрицы коэффициентов частной корреляции (при  $m>2$ )

явление мультиколлинеарности в исходных данных считается установленным, если коэффициент парной корреляции между двумя переменными больше 0,8

# MULTICOLLINEARITY

## Методы выявления:

Нахождение коэффициентов частной корреляции (при  $m=3$ )

$r_{ij}$  — коэффициент парной корреляции между переменными  $x_i$  и  $x_j$

$r_{ij.k}$  — коэффициент частной корреляции между переменными  $x_i$  и  $x_j$ , исключая их корреляцию с  $x_k$

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

# MULTICOLLINEARITY

## Методы выявления:

Нахождение матрицы коэффициентов частной корреляции (при  $m > 3$ , например, при  $m = 4$ )

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{bmatrix} \quad R^{-1} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

$$r_{ij.kl} = \frac{-c_{ij}}{\sqrt{c_{ii}} \cdot \sqrt{c_{jj}}}$$

$$r_{13.24} = \frac{-c_{13}}{\sqrt{c_{11}} \cdot \sqrt{c_{33}}}$$

# MULTICOLLINEARITY

## Методы выявления:

- исследование матрицы перекрестных произведений значений экзогенных переменных

$$(X^T \cdot X)$$

Если определитель матрицы близок к нулю, т.е. матрица вырожденная, то это свидетельствует о наличии мультиколлинеарности

Аналогичным образом, можно рассмотреть вариант анализа близости к нулю определителя матрицы коэффициентов парной корреляции

# MULTICOLLINEARITY

## Методы выявления:

- метод инфляционных факторов: для каждой экзогенной переменной модели строится вспомогательная модель регрессии на остальные экзогенные переменные, выписывается коэффициент детерминации такой модели и рассчитывается дисперсионно-инфляционный фактор VIF (Variance Inflationary Factor)

$$VIF_{X_i} = \frac{1}{1 - R_{X_i}^2}$$

Если значение  $VIF > 5$ , говорят о присутствии мультиколлинеарности (по некоторым источникам  $VIF > 10$ )

# MULTICOLLINEARITY

## *Методы выявления:*

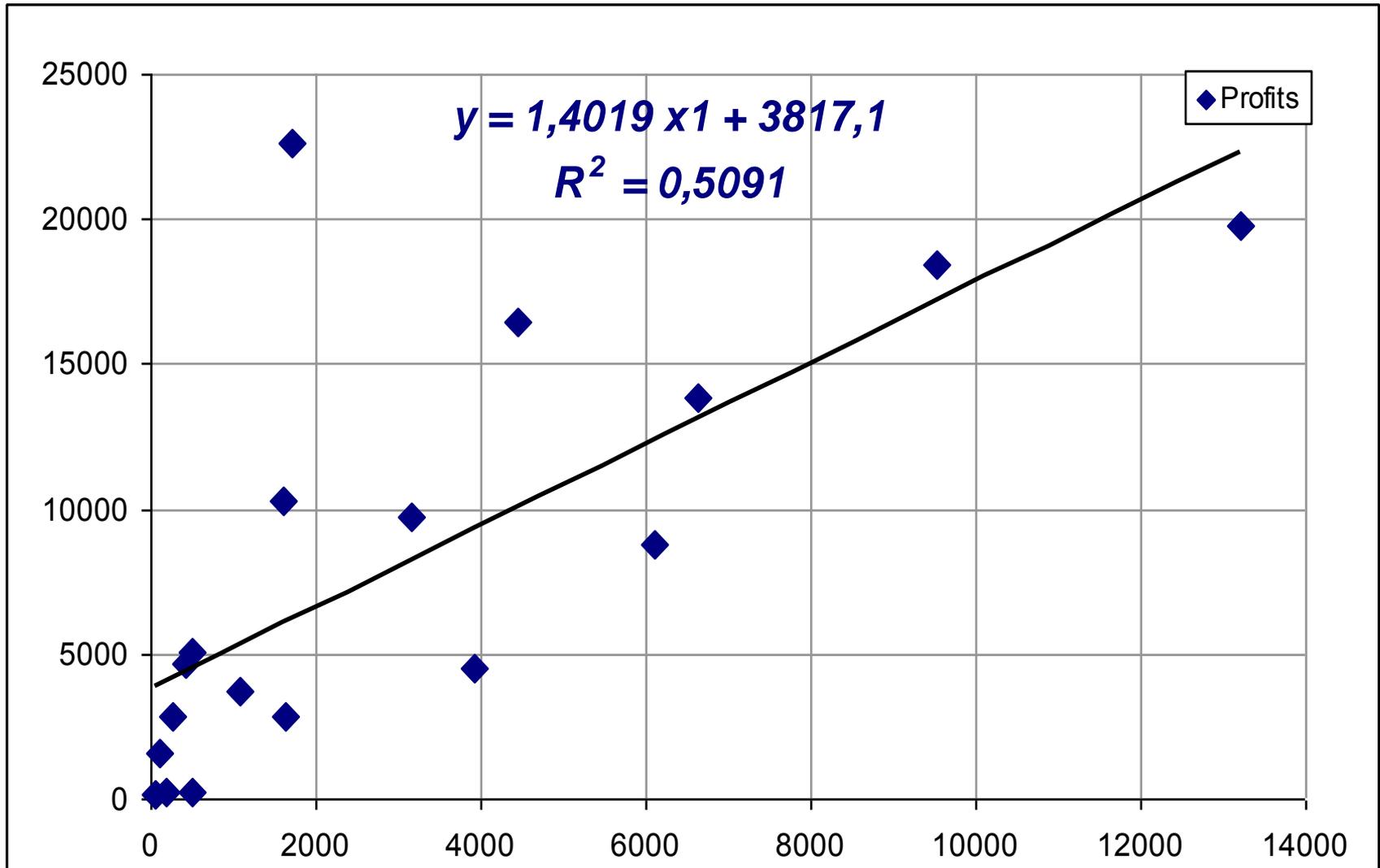
Значения дисперсионно-инфляционного фактора VIF можно получить и другим способом: при вычислении элементов главной диагонали обратной матрицы к матрице коэффициентов парной корреляции

$$VIF_{x_i} = \frac{1}{1 - R_{x_i}^2}$$

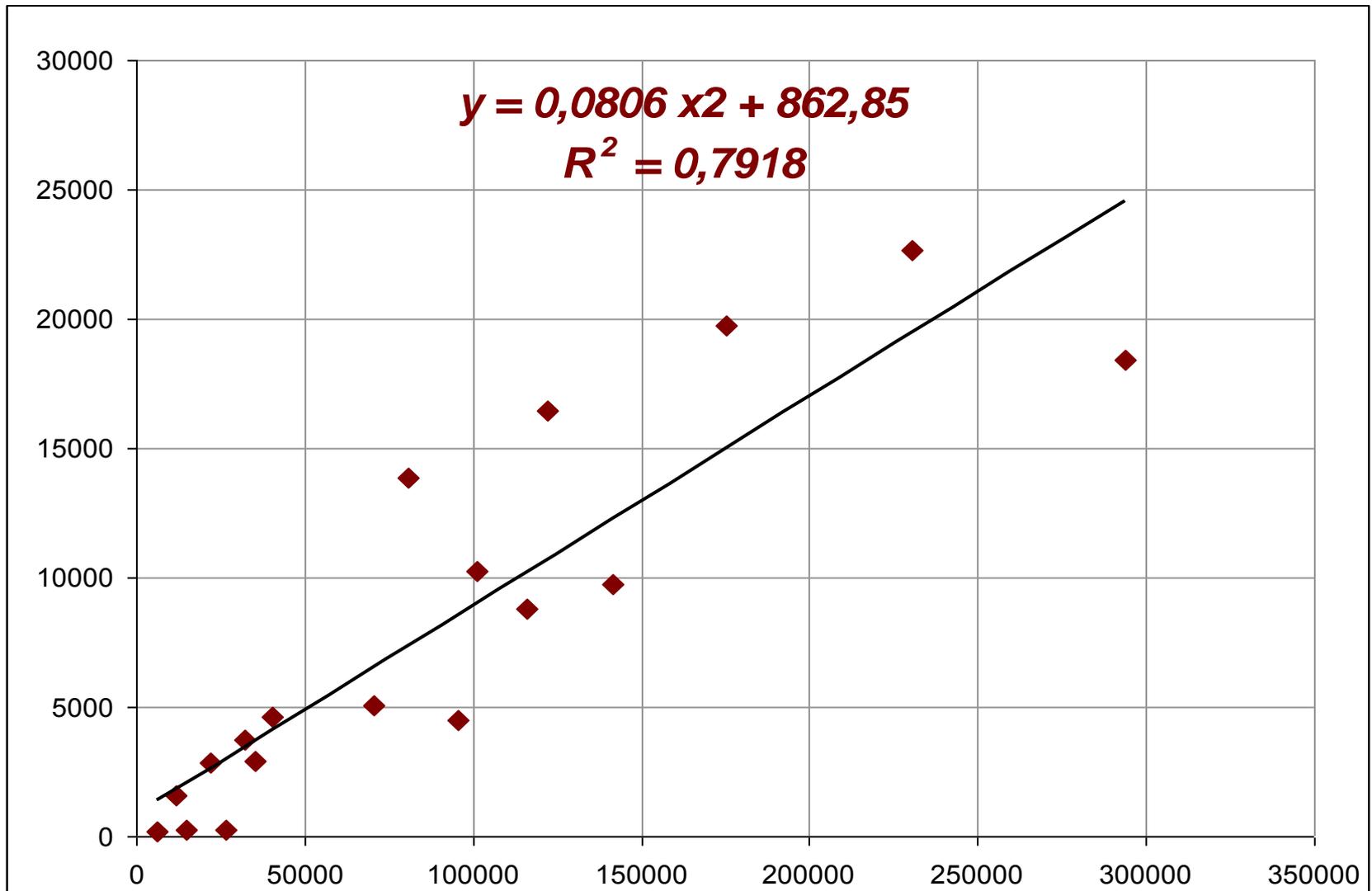
## EXAMPLE: MULTICOLLINEARITY, M=2

i	Y=Profits	X1=R&D Expenditure	X2=Sales
1	185,1	62,5	6375,3
2	1569,5	92,9	11626,4
3	276,8	178,3	14655,1
4	2828,1	258,4	21869,2
5	225,9	494,7	26408,3
6	3751,9	1083	32405,6
7	2884,1	1620,6	35107,7
8	4645,7	421,7	40295,4
9	5036,4	509,2	70761,6
10	13869,9	6620,1	80552,8
11	4487,8	3918,6	95294
12	10278,9	1595,3	101314,1
13	8787,3	6107,5	116141,3
14	16438,8	4454,1	122315,7
15	9761,4	3163,8	141649,9
16	19774,5	13210,7	175025,8
17	22626,6	1703,8	230614,5
18	18415,4	9528,2	293543

# EXAMPLE: MULTICOLLINEARITY



# EXAMPLE: MULTICOLLINEARITY



# EXAMPLE: MULTICOLLINEARITY

$$Y = 791,536 + 0,3695 \cdot X_1 + 0,0689 \cdot X_2 + \varepsilon \quad R^2 = 0,8103$$

$$(t) \quad (0,652) \quad (1,208) \quad (4,879)$$

$$(P) \quad (0,524) \quad (0,246) \quad (0,0002)$$

$$r_{X_1 X_2} = 0,6916$$