

Модель множественной регрессии всегда включает случайную составляющую

$$B = c + aYD + bP + \varepsilon$$

Наличие случайной составляющей связано в том, что помимо факторов, вошедших в модель на зависимую переменную могут влиять и другие неучтенные факторы.

Метод наименьших квадратов является наилучшим только если случайная составляющая обладает двумя свойствами

- 1) случайная составляющая является гомоскедастичной (имеет постоянную дисперсию)
- 2) в случайной составляющей отсутствует автокорреляция

Гетероскедастичность случайной составляющей

Пусть нарушено первое условие

$$D\varepsilon \neq const$$

Тогда говорят, что имеет место

гетероскедастичность, т.е. ошибки
регрессии имеют непостоянные дисперсии

В этом случае МНК не является лучшим методом.

Условия Гаусса-Маркова

Иллюстрация гомоскедастичности. МНК можно использовать. Разброс вокруг линии регрессии постоянный

Условия Гаусса-Маркова

Иллюстрация гетероскедастичности. МНК нельзя использовать, разброс вокруг линии регрессии увеличивается с ростом x .

ПРИМЕР

(зависимость инвестиций от ВРП в
2006г. по российским регионам)

Регион	VRP	INV
Белгородская область	119673,20	48422
Брянская область	61888,30	10973
Владимирская область	76328,10	20292
Воронежская область	70849,40	36265
Ивановская область	47949,80	14652

ПРИМЕР

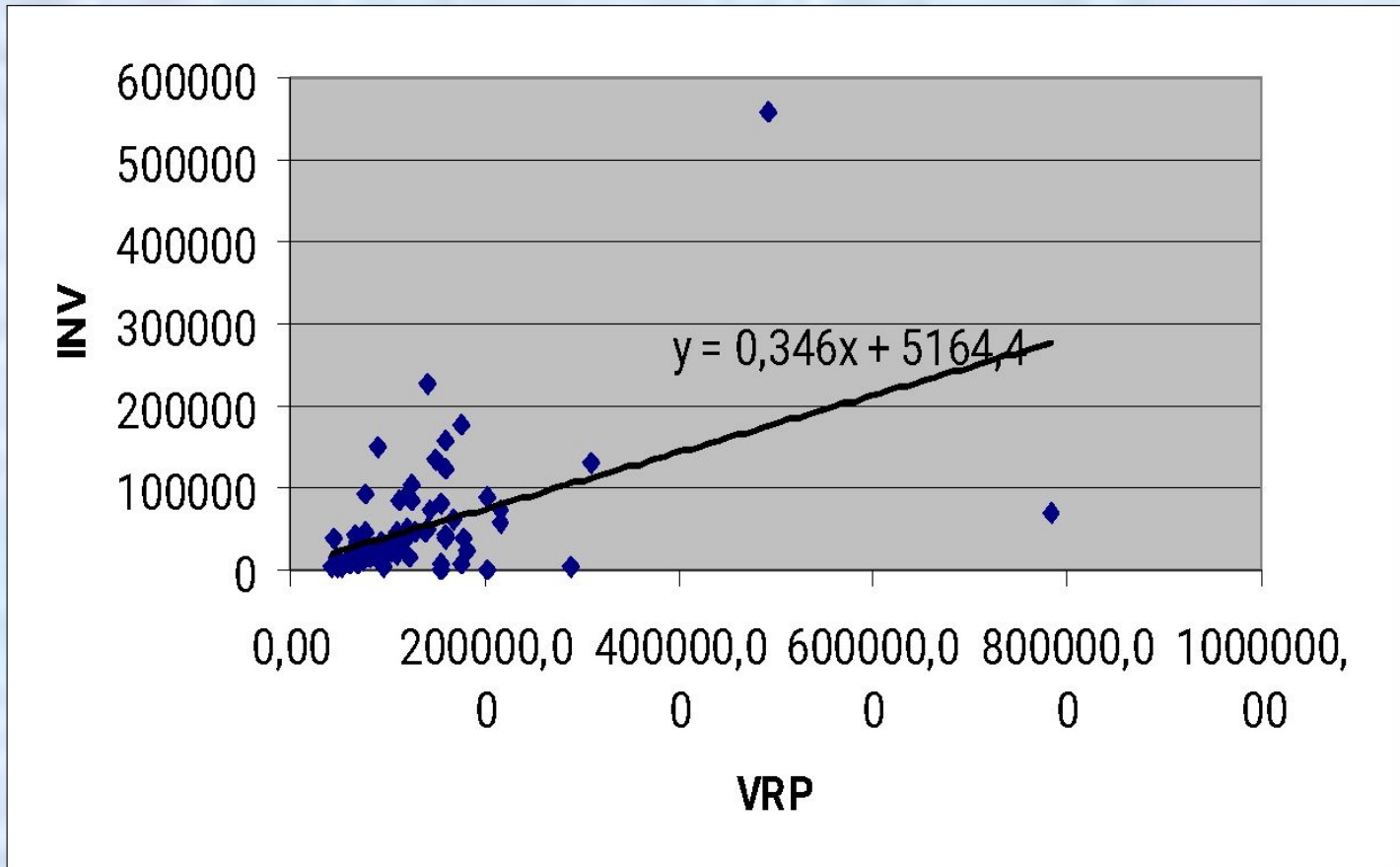
(зависимость инвестиций от ВРП в 2006г. по российским регионам)

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	5164,439	11352,89	0,454901	0,650428
VRP	0,345983	0,070519	4,906268	4,88E-06

$$INV = 0,346 \cdot VRP + 5164$$

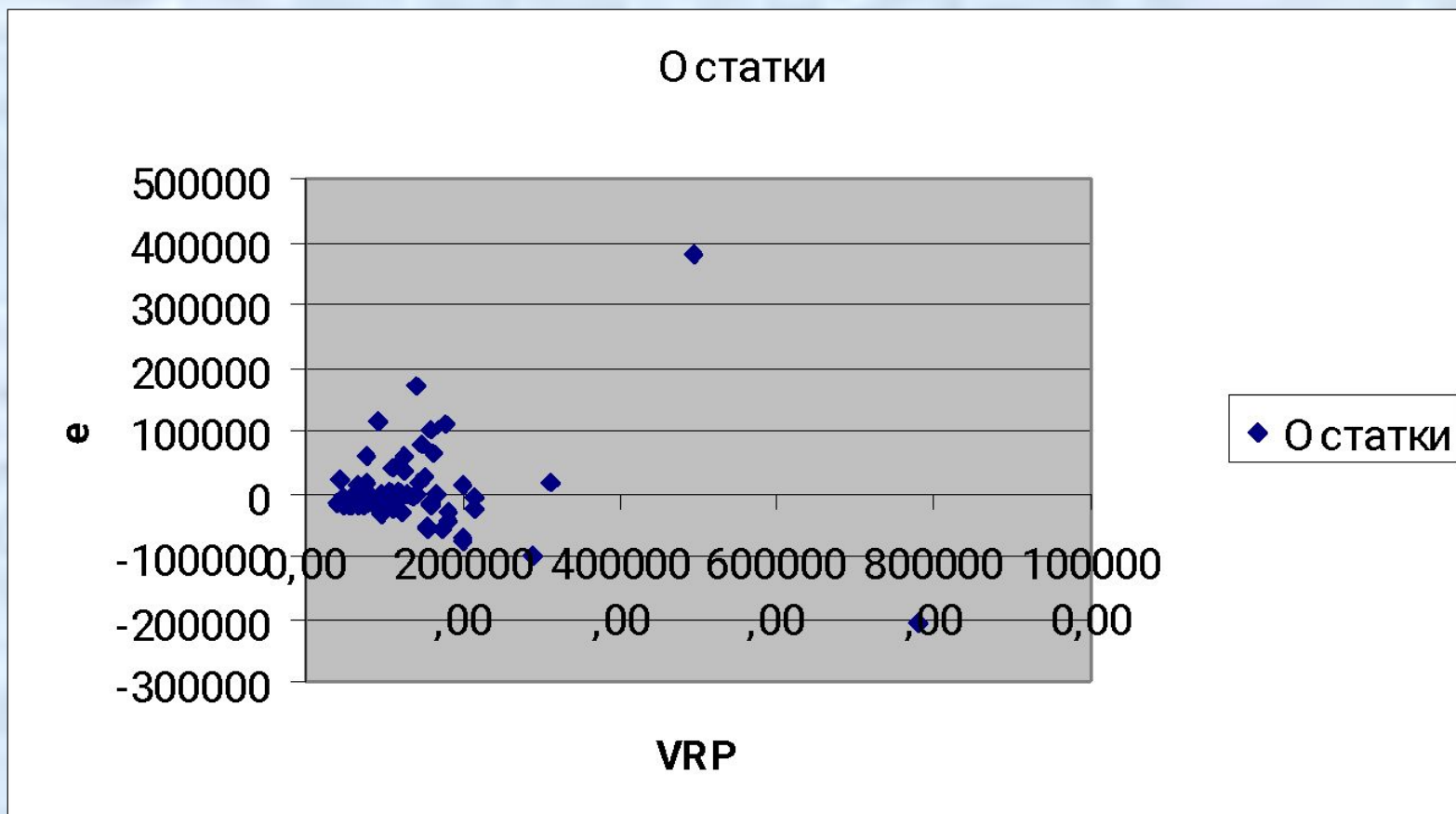
ПРИМЕР

(зависимость инвестиций от ВРП в 2006г. по российским регионам)



Разброс вокруг линии регрессии увеличивается с ростом ВРП. Есть непостоянство дисперсии (гетероскедастичность)

График остатков в зависимости от ВРП также позволяет увидеть гетероскедастичность, чем больше ВРП, тем больше по модулю остатки

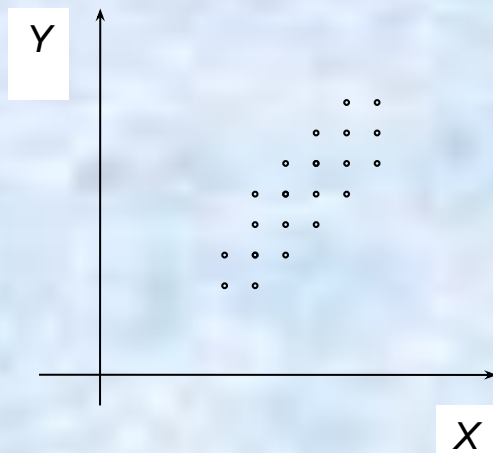


ПОСЛЕДСТВИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ

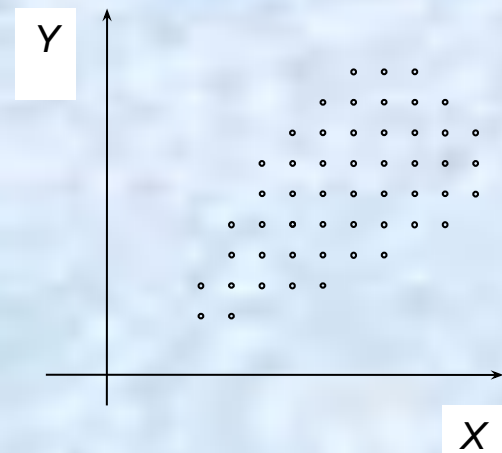
1. Обычная МНК оценка становится не самой точной (есть более точные оценки).
2. Проверка значимости коэффициентов становится некорректной

ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

1. Визуальный метод.
Диаграмма рассеяния



Гомоскедастичность

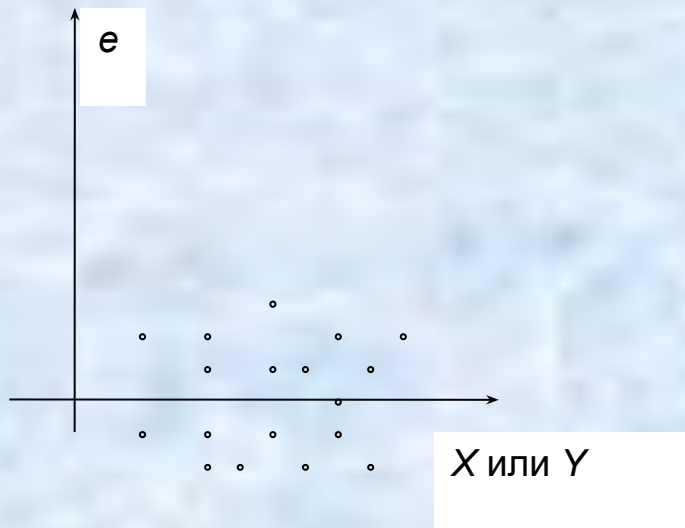


Гетероскедастичность

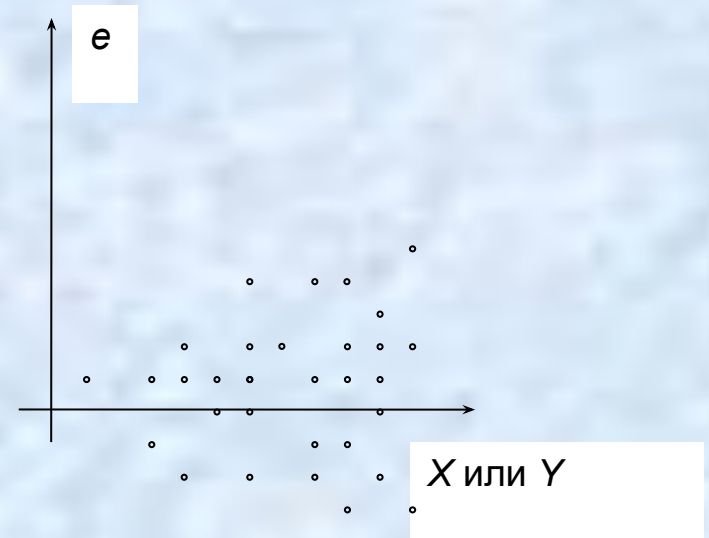
ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

1. Визуальный метод.

Графики остатков после построения оценок по методу МНК



Гомоскедастичность



Гетероскедастичность

ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

Кроме визуального метода используются более точные методы

Тесты:

1. Тест ранговой корреляции Спирмена.
2. Тест Глейзера.
3. Тест Голдфелда-Квандта.
4. Тест Уайта.

В специализированных эконометрических пакетах эти тесты есть

Устранение ГЕТЕРОСКЕДАСТИЧНОСТИ

1) Использовать обобщенный метод наименьших квадратов

В этом методе предполагается, что стандартное отклонение остатков пропорционально одной из объясняющих переменных

Например, $INV = aVRP + b + \varepsilon \quad \sigma_{\varepsilon} = k \cdot VRP$

Устранение гетероскедастичности пример

	1 Region	2 VRP	3 INV
1	Белгородская область	119673,20	48422
2	Брянская область	61888,30	10973
3	Владимирская облас	76328,10	20292
4	Воронежская область	70849,40	36265

$$INV = aVRP + b + \varepsilon \quad \sigma_{\varepsilon} = k \cdot VRP$$

Делим уравнение на эту объясняющую переменную (в примере на ВРП)

$$\frac{INV}{VRP} = a + b \frac{1}{VRP} + \frac{\varepsilon}{VRP}$$

Тогда

$$D \frac{\varepsilon}{VRP} = \frac{1}{VRP^2} D\varepsilon = \frac{1}{VRP^2} k^2 VRP^2 = k^2 = const$$

Устранение гетероскедастичности

пример

$$\frac{INV}{VRP} = a + b \frac{1}{VRP} + \frac{\varepsilon}{VRP}$$

Создаем новые переменные

$$\frac{INV}{VRP}, \frac{1}{VRP}$$

Регион	VRP	INV	INV/VRP	1/VRP
Белгородская область	119673,20	48422	0,40	8,36E-06
Брянская область	61888,30	10973	0,18	1,62E-05
Владимирская область	76328,10	20292	0,27	1,31E-05
Воронежская область	70849,40	36265	0,51	1,41E-05
Ивановская область	47949,80	14652	0,31	2,09E-05
Калужская область	83817,40	16268	0,19	1,19E-05

Устранение гетероскедастичности пример

$$\frac{INV}{VRP} = a + b \frac{1}{VRP} + \frac{\varepsilon}{VRP}$$

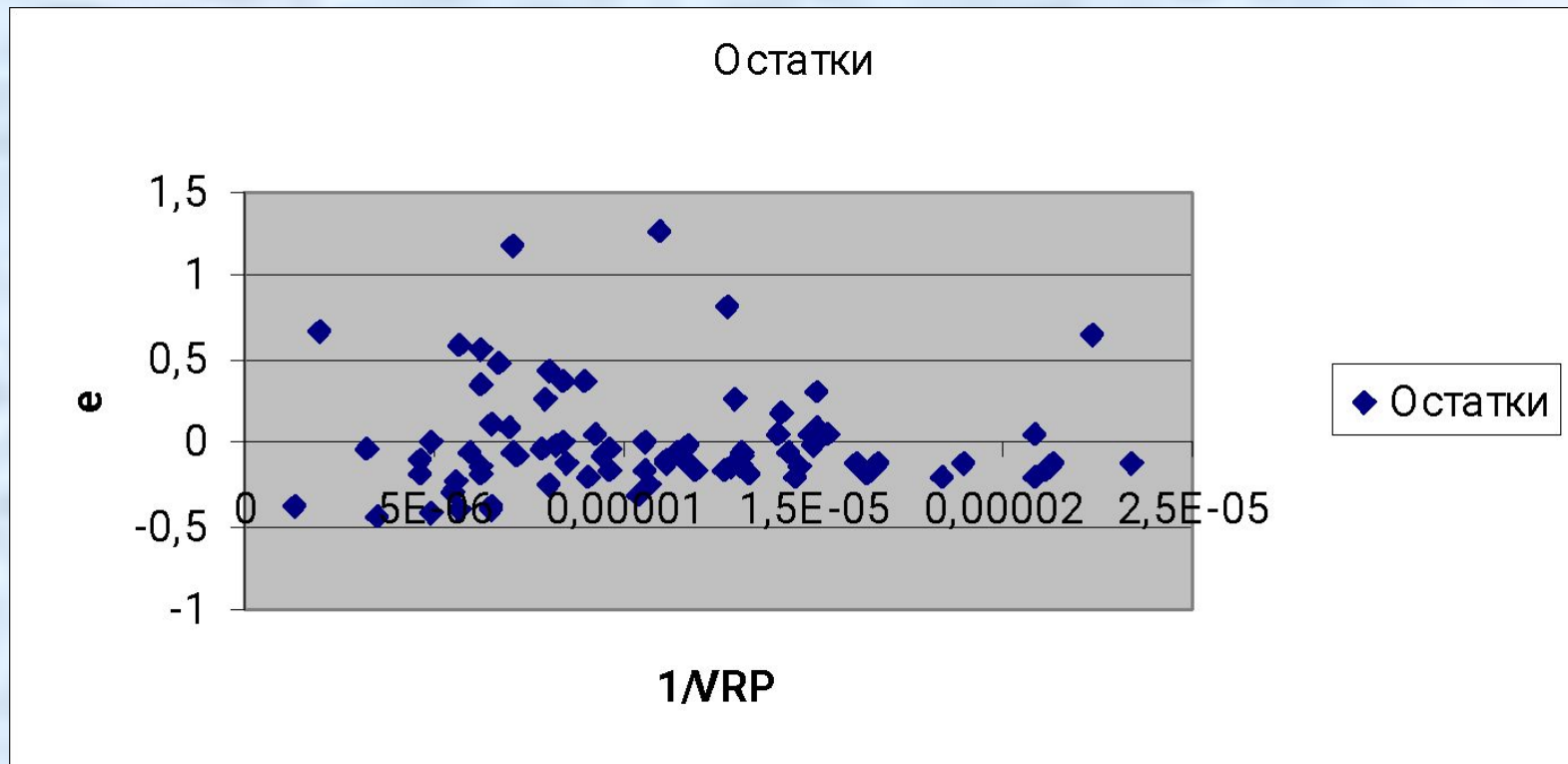
	Коэффициент ы	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	0,491025	0,0864	5,683151	2,13E-07
1/VRP	-11337,2	7354,391	-1,54156	0,127178

$$\frac{INV}{VRP} = 0,5 - 11,337 \frac{1}{VRP},$$

$$INV = 0,5 \cdot VRP - 11,337$$

Устранение гетероскедастичности пример

$$\frac{INV}{VRP} = a + b \frac{1}{VRP} + \frac{\varepsilon}{VRP}$$



Визуально гетероскедастичность отсутствует

2) Изменить функциональную форму модели

С ростом x растет разброс вокруг линии регрессии. Есть гетероскедастичность

2) Изменить функциональную форму модели

Прологарифмируем данные и построим модель в логарифмах

Разброс постоянный вокруг линии регрессии. Нет гетероскедастичности. Таким образом, нужно строить степенную модель вместо линейной