$R^2$ — не работает

=> Нецентрированный к-т. $R^2_{uc} = 1 - \frac{RSS}{\sum y_i^2} = \frac{\sum \hat{y}_i^2}{\sum y_i^2}$

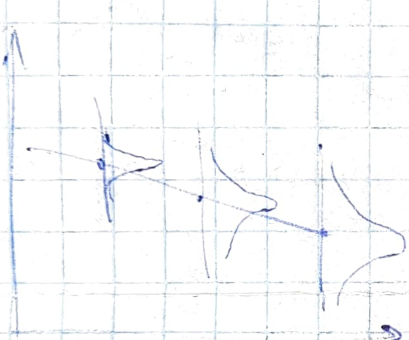
## 2.12 Гомоскедастичность и гетероскедастичность

Ошибка $u_i$ — гомоскедастична, если

$var(u_i | X_i = x)$ — const и, в частности,

не зависит от $x$ (однородность случ. ошибки).

Иначе ошибка гетероскедастична



e.g. разброс растет с увеличением $x$
(распр. станов. шире)

Следствие гомоскедастичности

① МНК-оценки остаются несмещ. и ас. нормальными
(нез. от гомо-гетеро-)

② при предп. МНК + гомоск. => $\hat\beta_0$, $\hat\beta_1$ — эффективны

③ $\hat\sigma^2_{\hat\beta_1} = \frac{s^2_{\hat u}}{\sum (X_i - \bar X)^2}$, где $s^2_{\hat u} = \frac{1}{n-2} \cdot \sum \hat u_i^2$

$\hat\sigma^2_{\hat\beta_0} = \frac{\left(\frac{1}{n} \sum X_i^2\right) s^2_{\hat u}}{\sum (X_i - \bar X)^2}$

▷ из $\sigma^2_{\hat\beta_1} = \frac{\sigma^2_u}{n \sigma^2_x}$ ; $\sigma^2_{\hat\beta_0} = \frac{E(X_i^2) \cdot \sigma^2_u}{n \sigma^2_x}$

эти формулы не действ. при опр. ур. знач.
при гетероск-ти, т.к. $t \not\to N$

=> $\hat\sigma^2_{\hat\beta_0}$ и $\hat\sigma^2_{\hat\beta_1}$ — heteroscedasticity consistent

2. (B) Теорема Гаусса – М[illegible]

THM: Если для $\forall i, i = \overline{1,n}$ вып. усл. Гаусса – Маркова

(1) $E(u_i \mid X_1, \ldots, X_n) = 0$

(2) $Var(u_i \mid X_1 \ldots X_n) = \sigma_u^2$, $\sigma_u^2 \in (0; \infty)$ (гомоск.)

(3) $E(u_i u_j \mid X_1, \ldots, X_n) = 0$, $i \neq j$ (незав.)

то МНК – оценка $\hat{\beta}_1$ – BLUE

ограничения: 1) при гетероск-ти $\hat{\beta}_1$ – не BLUE, но при НС – оценках гипотезы проверяются
2) могут $\exists$ не лин. или усл. несмещ. более эф. оценки

▷ $\hat{\beta}_1$ – линейна:

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X}) \cdot Y_i + 0}{\sum (X_i - \bar{X})^2} = \sum \hat{a}_i Y_i, \text{ где}$$

$$\hat{a}_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$$

$\Rightarrow$ веса нез. от $Y_i$ $\Rightarrow$ $\hat{\beta}_1$ – линейна

▷ $\hat{\beta}_1$ – усл. несмещена

$$E(\hat{\beta}_1 \mid X) = \beta_1 + \frac{1/n \sum (X_i - \bar{X})}{1/n \cdot \sum (X_i - \bar{X})^2} \cdot \underbrace{E(u_i \mid X)}_{=0} = \beta_1$$

$$Var(\hat{\beta}_1 \mid X) = \frac{\sigma_u^2}{\sum (X_i - \bar{X})^2}$$

▷ $\hat{\beta}_1$ – эффективна

$\tilde{\beta}_1$ – любая лин. усл. несмещ. оценка

$$\tilde{\beta}_1 = \sum a_i Y_i = \beta_0 \sum a_i + \beta_1 \sum a_i X_i + \sum a_i u_i$$

- $E(\sum a_i u_i \mid X) = E(\sum a_i u_i \mid X_i) = 0$

- $E(\tilde{\beta}_1 \mid X) = \beta_0 \sum a_i + \beta_1 \sum a_i X_i = 0 \iff \sum a_i = 0; \sum a_i X_i = 1$

$\Rightarrow \tilde{\beta}_1 - \beta_1 = \sum a_i u_i \Rightarrow Var(\tilde{\beta}_1 \mid X) = Var(\sum a_i u_i \mid X)$

$Var(\tilde{\beta} | X) = var(\sum a_i u_i | X) = \sum a_i a_j \cdot cov(u_i, u_j | X)$

$= \sigma_u^2 \cdot \sum a_i^2$ т.к. при $i \neq j$ $cov(u_i, u_j | X) = 0$

$\triangleright$ $\hat{\beta}_1$ : $Var(\tilde{\beta}_1 | X) > Var(\hat{\beta}_1 | X)$, $\forall \tilde{\beta}_1$

Пусть $a_i = \hat{a}_i + d_i$

причем $\sum a_i^2 = \sum \hat{a}_i^2 + \underbrace{2 \sum \hat{a}_i d_i}_{=0 \text{ по опр. } \hat{a}_i} + \sum d_i^2$

$$= \frac{\sum (X_i - \bar{X}) d_i}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i d_i - \bar{X} \cdot \sum d_i}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum a_i X_i - \sum \hat{a}_i \cdot \bar{X} - \bar{X}(\sum a_i - \sum \hat{a}_i)}{\sum (X_i - \bar{X})^2} = 0$$

$$= \frac{1 - 1 - \bar{X}(0 - 0)}{\sum (X_i - \bar{X})^2} = 0$$

$\Rightarrow \sigma_u^2 \sum a_i^2 = \sigma_u^2 \cdot \sum \hat{a}_i^2 + \sigma_u^2 \cdot \sum d_i^2 = var(\hat{\beta}_1 | X) + \sigma_u^2 \sum d_i^2$

$\Rightarrow Var(\tilde{\beta} | X) - Var(\hat{\beta} | X) = \sigma_u^2 \sum d_i^2$

$\Rightarrow \hat{\beta}$ - BLUE

## 2.14 Связь условий Гаусса - Маркова и предположений МНК

I) (2) $(X_i, Y_i)$ - i.i.d $\Rightarrow E(u_i | X_1, \dots, X_n) = E(u_i | X_i)$ $\Big\}$ $\Rightarrow E(u_i | X_1, \dots, X_n) = 0$

(1) $E(u_i | X_i) = 0$

II) (2) $(X_i, Y_i)$ - i.i.d. $\Rightarrow Var(u_i | X_1, \dots, X_n) = Var(u_i | X_i)$ $\Big\}$ $\Rightarrow Var(u_i | X) = \sigma_u^2$ - const

\+ гомоскедастичность

(3) $\mu^4 \in (0, \infty)$ $\Rightarrow$ $\sigma_u^2 \in (0; \infty)$

III) (2) $(X_i, Y_i)$ - i.i.d $\Rightarrow E(u_i u_j | X) = E(u_i u_j | X_i, X_j) = E(u_i | X_i) \cdot E(u_j | X_j)$

(1) $E(u_i | X_i) = 0$ $\Rightarrow$ $E(u_i u_j | X) = 0$

## ③ Проверка гипотез в модели п.л.р.

### 3.1. Тестирование двусторонних гипотез отн. $\beta_1$

$$\hat{Y}_i = \underset{(SE(\hat{\beta}_0))}{\hat{\beta}_0} + \underset{(SE(\hat{\beta}_1))}{\hat{\beta}_1} X_i$$

$H_0: \beta_1 = \beta_{1,0}$, где $\beta_{1,0}$ – значение $\beta_1$ при $H_0$

$H_1: \beta_1 \neq \beta_{1,0}$

① Вычисление станд. ошибки $\hat{\beta}_1$ – $(SE(\hat{\beta}_1))$

$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$ – оценка для $\sigma_{\hat{\beta}_1}$ (с.о. выб распр $\hat{\beta}_1$)

где: $$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2} \sum (X_i - \bar{X})^2 \cdot \hat{u}_i^2}{\left(\frac{1}{n} \sum (X_i - \bar{X})^2\right)^2} \left[ \xrightarrow{d} \frac{1}{n} \frac{Var[(X_i - \mu_X) u_i]}{[Var(X_i)]^2} = \sigma^2_{\hat{\beta}_1} \right.$$

② Расчет $t$-статистики

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

③ Отвержение / не отвержение нулевой гипотезы $H_0$.

1) при $\alpha$% ур. значимости

$H_0$ отв., если $|t^{act}| > |t^{\alpha}_{\alpha/2}| \left[= 1{,}96\right]$

(или $p\text{-value} < \alpha$)

2) расчет $p$-value

$$p\text{-value} = P_{H_0}\left[ |\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{act} - \beta_{1,0}| \right] =$$

$$= P_{H_0}\left[ \left| \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} \right| > \left| \frac{\hat{\beta}_1^{act} - \beta_{1,0}}{SE(\hat{\beta}_1)} \right| \right] = P_{H_0}\left[ |t| > |t^{act}| \right]$$

★ (при $n \to \infty$ $\hat{\beta}_1 \overset{a}{\sim} N(0;1) \Rightarrow p\text{-value} = P(|Z| > |t^{act}|) =$
$= 2\Phi(-|t^{act}|)$

3.2. **P-value** (вероятность значимости) – min вероятность отвержения $H_0$ (на основе имеющ. выборки), в предполож., что $H_0$ – верна (т.е. вер. соверш. ош. I рода)

↳ при $\forall \alpha$ : $\alpha > p$ $H_0$ – отвергается

3.3. **Тестирование односторонних гипотез отн. $\beta_1$**

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$H_0: \beta_1 = \beta_{1,0}$

$H_1: \beta_1 \lessgtr \beta_{1,0}$ (левосторонняя альтернатива)
право

1) $SE(\hat{\beta}_1)$

2) $t = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$

3) $H_0$ отв. если $t^{act} \lessgtr t^{cr}_{\alpha}$ $\left[ = \mp 1,645 \right]$

P-value $= P(Z \lessgtr t^{act}) = \Phi(t^{act})$

3.4. **Тестирование гипотез отн. $\beta_0$**

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$H_0: \beta_0 = \beta_{0,0}$

$H_1: \beta_0 \neq \beta_{0,0}$ (> или <)

1) $SE(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_0}}$,

где $\hat{\sigma}^2_{\hat{\beta}_0} = \dfrac{1}{n} \cdot \dfrac{\frac{1}{n-2} \sum H_i^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum H_i^2\right]^2}$, $\quad \to \sigma^2_{\hat{\beta}_0} = \dfrac{1}{n} \dfrac{Var[H_i u_i]}{(\mathbb{E}(H_i^2))^2}$,

где $H_i = 1 - \dfrac{\mu_X}{\mathbb{E}(X_i^2)} X_i$

$$H_i = 1 - \left[ \frac{\bar{X}}{\frac{1}{n} \sum X_i^2} \right] \cdot X_i$$

## 3.5. Построение доверительных интервалов

95%-ный д.и. для $\hat{\beta}_k$ - набор знач. $\beta_k$ равенство которых не может быть отв. на 5% зн. тесте

$\Rightarrow$ для 95% выборок д.и. будет содержать ист. значение $\beta_k$

двуст. 95% д.и. для $\beta_k = [\hat{\beta}_k - 1{,}96\, SE(\hat{\beta}_k) ; \hat{\beta}_k + 1{,}96\, SE(\hat{\beta}_k)]$

одност. 95% д.и. для $\beta_k$: $\beta_i \le \hat{\beta}_k + 1{,}645\, SE(\hat{\beta}_k)$ (лево)

$\beta_i \ge \hat{\beta}_k - 1{,}645\, SE(\hat{\beta}_k)$ (право)

## 3.6 Доверительный интервал для $\Delta Y$ при $\Delta X$

Пусть $X$ изм. на $\Delta x \Rightarrow Y$ изм. на $\Delta Y = \beta_1 \Delta X$

95% д.и. для $\Delta Y$ =

$$\Delta Y = \beta_1 \Delta X \in [\{\hat{\beta}_1 - 1{,}96\, SE(\hat{\beta}_1)\} \Delta x ; \{\hat{\beta}_1 + 1{,}96\, SE(\hat{\beta}_1)\} \Delta x]$$

## 3.7. Регрессия с бинарной объясняющей переменной

$$D_i = \begin{cases} 0, & \text{если } \bar{A} \text{ происходит} \\ 1, & \text{если } A \text{ происходит} \end{cases}$$

$D_i$ - бинарная / индикаторная / дамми - переменная

$$Y_i = \beta_0 + \beta_1 \cdot D_i + u_i = \begin{cases} \beta_0 + u_i, & \text{если } D_i = 0 \\ \beta_0 + \beta_1 + u_i, & \text{если } D_i = 1 \end{cases}$$

$E(Y_i \mid D_i = 0) = \beta_0$ $\quad E(Y_i \mid D_i = 1) = \beta_0 + \beta_1$

$\Rightarrow \beta_1 = E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0)$ - разность между усл. средними

$\Rightarrow \hat{\beta}_0 = \bar{Y}_{i: D_i = 0}$ ; $\hat{\beta}_1 = \bar{Y}_{i: D_i = 1} - \bar{Y}_{i: D_i = 0}$

## 4 Нарушение предпосылок теоремы Гаусса–Маркова

### 4.1 Оценки дисперсий оценок коэф. при гомоск. ошибках

Для случая гомоскедастичных ошибок

$$\tilde{\sigma}^2_{\hat\beta_1} = \frac{S^2_{\hat u}}{\sum (X_i - \bar X)^2}, \text{ где } S^2_{\hat u} = \frac{1}{n-2}\sum \hat u_i^2 = \frac{RSS}{n-2}$$

$$\tilde{\sigma}^2_{\hat\beta_0} = \frac{\left(\frac{1}{n}\sum X_i^2\right)\cdot S^2_{\hat u}}{\sum (X_i - \bar X)^2}$$

▶ При гомоск-ти ошибок $Var(u_i | X_i) = \sigma^2_u$

$$\hat\beta_1 \overset{d}{\sim} N(\beta_1, \sigma^2_{\hat\beta_1})$$

$$\sigma^2_{\hat\beta_1} = \frac{1}{n}\cdot\frac{Var[(X_i-\mu_x)u_i]}{(Var(X_i))^2} = \frac{\sigma^2_u \cdot \sigma^2_x}{n\,\sigma^4_x} = \frac{\sigma^2_u}{n\,\sigma^2_x} \Rightarrow \tilde{\sigma}^2_{\hat\beta_1}$$

замена на выб X

где $Var[(X_i-\mu_x)u_i] = E\left[\left((X_i-\mu_x)u_i - E(X_i-\mu_x)u_i\right)^2\right]$

$$= E\left[\left((X_i-\mu_x)u_i\right)^2\right] + \underset{\text{т.к. } E(u_i|X_i)=0}{0} + \underset{\text{const}}{0} =$$

$$= E\left[(X_i-\mu_x)^2 \cdot u_i^2\right] = E\left[(X_i-\mu_x)^2 \cdot Var(u_i|X_i)\right]$$

$$= \sigma^2_u \cdot E\left[(X_i-\mu_x)^2\right] = \sigma^2_u \cdot \sigma^2_x$$

$$\hat\beta_0 \overset{d}{\sim} N(\beta_0, \sigma^2_{\hat\beta_0})$$

$$\sigma^2_{\hat\beta_0} = \frac{1}{n}\cdot\frac{Var(H_i u_i)}{[E(H_i^2)]^2}, \text{ где } H_i = 1 - \left[\frac{\mu_x}{E(X_i^2)}\right]\cdot X_i$$

аналогично упрощается до

$$\sigma^2_{\hat\beta_0} = \frac{E(X_i^2)}{n\cdot\sigma^2_x}\cdot\sigma^2_u \Rightarrow \tilde{\sigma}^2_{\hat\beta_0}$$

## 4.2 Следствия гетероскедастичности

1) МНК - оценки коэф. - несмещ. и сост.

и ас. нормальные и при гомо-, и гетероскед-ти

2) Эффективность только при гомоскед-ти
(при гетероскед-ти оценка дисперсии ст. ошибок смещена)

## 4.3. Недоступный ВМНК (при изв. форме гетероск-ти)

Взвешенный МНК дает оценки более эффективные

по сравнению с МНК - оценкой при гетероск-ти ошибок

Предположим, что $var(u_i|X_i)$ известна

с точностью до множителя

$$var(u_i|X_i) = \lambda h(X_i)$$

$$Y_i = \beta_0 + \beta_1 \cdot X_i + u_i \quad | : \sqrt{h(X_i)}$$

$$\frac{Y_i}{\sqrt{h(X_i)}} = \beta_0 \cdot \frac{1}{\sqrt{h(X_i)}} + \beta_1 \cdot \frac{X_i}{\sqrt{h(X_i)}} + \frac{u_i}{\sqrt{h(X_i)}}$$

$$\tilde{Y}_i = \beta_0 \cdot \tilde{X}_{0i} + \beta_1 \cdot \tilde{X}_{1i} + \tilde{u}_i$$

ВМНК - оценка - МНК - оценка для $\hat{\beta}_1$, получ.

из регрессии $\tilde{Y}_i$ на $\tilde{X}_{0i}$ и $\tilde{X}_{1i}$

$\hat{\beta}_1$ - BLUE - оценка к. наклона в $Y_i = \beta_0 + \beta_1 X_i + u_i$

т.к.

$$var(\tilde{u}_i|X_i) = var\left(\frac{u_i}{\sqrt{h(X_i)}} \Big| X_i\right) = \frac{var(u_i|X_i)}{h(X_i)} = \frac{\lambda h(X_i)}{h(X_i)} = \lambda$$

$\Rightarrow var(\tilde{u}_i|X_i) - const \Rightarrow BLUE \leftarrow \hat{\beta}_1$

## 4.4. Доступный ВМНК

Общий алгоритм:

1. Оценить регрессию $Y_i$ на $X_i$ при помощи МНК, вычислить остатки $\hat{u}_i$

2. Оценить модель ф-и усл. дисперсии $var(u_i | X_i)$ (оценить $\hat{u}_i^2$ на набор пер., от кот. она зависит)

   e.g.: $var(u_i | X_i) = \theta_0 + \theta_1 X_i^2$

   ↳ вычисл. регр. $\hat{u}_i^2$ на $X_i^2$, исп. МНК

   $\Rightarrow \hat{\theta}_0, \hat{\theta}_1 : \hat{\theta}_0 \xrightarrow{p} \theta_0, \hat{\theta}_1 \xrightarrow{p} \theta_1$

3. Исп. оцененную ф-ю для расч. предсказ. значений ф-и усл. дисперсии $\widehat{var}(u_i | X_i)$

4. Домножить зав. пер., регрессоры и св. член на $\frac{1}{\sqrt{\widehat{var}(u_i | X_i)}}$

5. Оценить получ. взвеш. регрессии при помощи МНК

$\Rightarrow$ полученные оценки - ВМНК-оценки

## 4.5. НС ~~[illegible]~~ Хьюбера-Уайта

(состоятельные при выполнении предп. МНК)

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\frac{1}{n-2} \sum (X_i - \bar{X})^2 \cdot \hat{u}_i^2}{\left(\frac{1}{n} \sum (X_i - \bar{X})^2\right)^2}$$

$$\hat{\sigma}^2_{\hat{\beta}_0} = \frac{1}{n} \frac{\frac{1}{n-2} \sum \hat{H}_i^2 \hat{u}_i^2}{\left(\frac{1}{n} \sum \hat{H}_i^2\right)^2}, \text{ где } H_i = 1 - \frac{\bar{X}}{\frac{1}{n} \sum X_i^2} \cdot X_i$$

## 4.6. ВМНК vs HC - станд. ошибки

ВМНК:

(+) более эффект. оценки, чем МНК - оценки получ. с исп. нескоррелир. регрессоров (по крайней мере асимптотически)

(−) необходимо знать ф-ю усл. дисперсии и оценки её параметров

↳ если функц. форма неверна, то ВМНК - ст. ошибки регрессии неверные ⇒ непр. стат. выводы (непр. размеры тестов)

HC - станд. ошибки

(+) асимпт. верные выводы (даже при неизв. $Var(u_i|x_i)$)

(+) легко вычисляются

(−) неэффективны по сравнению с ВМНК - оценкой (верной) (по крайней мере асимптотически)

## 4.7. Тестирование гетеро[...]

а) Графические методы

- Оценка модели при пом. МНК

- Построение графика ост. модели $\hat{u}_i$

(или стандарт. ост. модели $e_i = \frac{\hat{u}_i}{S_{\hat{u}}}$

1) От оценённых значений $\hat{y}_i$

↳ выявляются выбросы

возможная гетероскедастичность

неправильная спецификация модели

2) от отдельных объясняющих переменных

↳ выявляется возможная нелинейность

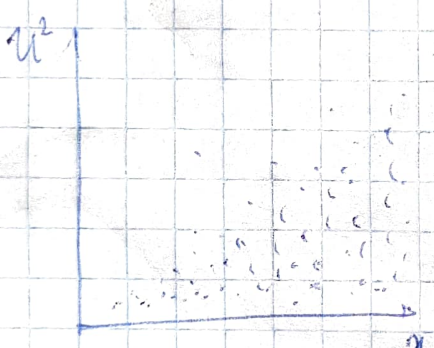
3) От номера наблюдения (в послед. $\Delta t$)

↳ выявляются изм. дисперсии ош. во времени

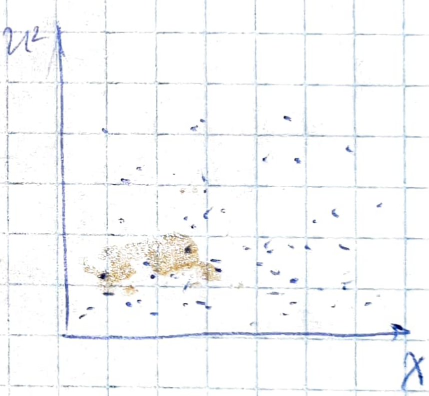
пропуск переменных

автокоррелированность ошибок

e.g.




гетероскедастичность




гомоскедастичность

b) Тест Голдф

предполагается, что 1) $\sigma_{u_i}$ пропорц. $X_i$

2) $u_i \sim N(\mu, \sigma^2)$

\+ др. усл. Гаусса-М

3) подходит для малых выборок

Алгоритм:

1. Все $n$ наблюдений в выборке упоряд. по предполагаемой переменной $X$ (по предпол. возр. дисперсии ошибок)

2. Отбросить $r$ ($\approx 20\%$) центральных наблюдений

3. Оценить исх. модель отдельно на первом и последнем интервале $\Rightarrow$ $F^{act} = \dfrac{RSS_2/n_2-k}{RSS_1/n_1-k}$

где $n_2$ – число набл. в верхней части выборки
$n_1$ – -//- нижней -//-
$k$ – число объясн. переменных в регр. уравнении

4. $F \sim F_{1-\alpha}\left(\frac{n-r}{2}-k;\ \frac{n-r}{2}-k\right)$

$H_0$: $var(u_i|X_i) = const$

$\Rightarrow$ $H_0$ отв., если $F^{act} > F^{cr}_{1-\alpha}(n_2-k; n_1-k)$

(!) если $RSS_2 < RSS_1$

$\Rightarrow$ $H_0$ не отв $\Rightarrow$ $F^{act}$ можно не вычислять

c) Тест Глейзера

- позволяет более тщательно рассмотреть характер гетероскедастичности. (т.к. снимается предп. о пропорц-зав-ти $\sigma_{u_i}$ от $X_i$)
- Предположим $var(u_i|X_i) = \theta_0 + \theta_1 X_i^{\gamma}$

Алгоритм:

1. Оценить регрессию $Y_i$ на $X_i$ при помощи МНК, вычислить остатки $\hat{u}_i$

2. Оценить регрессию $|\hat{u}_i|$ на константу и $X_i^{\gamma}$ при заданном знач. $\gamma$

3. $H_0: \theta_1 = 0$

$H_1: \theta_1 \neq 0$

$t = \frac{\hat{\theta}_1}{SE(\hat{\theta}_1)}$ => $H_0$ отв. на 5% ур. зн-ти, если $|t^{act}| > 1{,}96$

★ если $\theta_2$ значима при оценке более, чем одной ф-и (для разл. $\gamma$), то выбираете наилучшую

## 4.8. Автокоррелированность

Автокорреляция ошибок – частный случай нарушения III предпосылки Г-М: $E(u_i u_j | X_1, \ldots, X_T) \neq 0$

$\Rightarrow Cov(u_i u_j | X_1, \ldots, X_T) \neq 0$

$i \neq j$

<u>причины</u>: – "близость" наблюдений во вр. или пр-ве

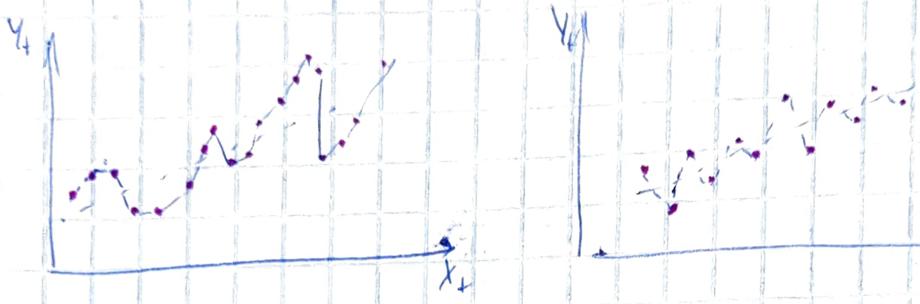
– наличие ненабл. факторов, действ. на соседние наблюдения

<u>последствия</u>: 1) ~~коэф.~~ коэф. оценки остаются <u>несмещ.</u> + сост., линейными, но становятся неэффективными.

2) может привести к <u>несост-ти</u> + смещенности оценок дисперсий ст. ошибок и дисп. оц коэф.

3) станд. ошибки часто занижены

(смещ. вниз) $\Rightarrow$ $\hat{\beta}_i$ несост. (требуется $se_{HAC}(\hat{\beta}_i)$)




полож. автокорреляция

(тенд. к сохр. полож. знач. после пол. и отриц. после отр.)

отрицательная автокорреляция

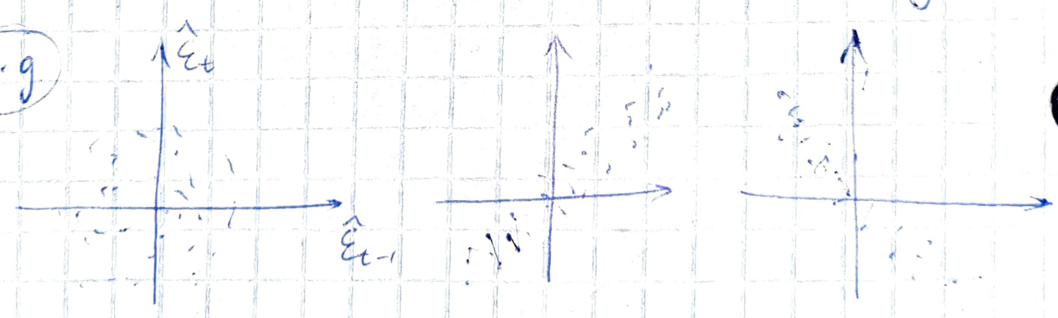
(скорее всего за полож. знач. в одном набл. идет отр. в след., и наоборот)

## 4.9. Тестирование автокоррелированности

### а) Графические методы

- оценка модели мет. МНК
- Построение графиков ост. или станд. ост.

e.g.




нет АС ($\rho=0$) АС > 0 АС < 0

- анализ коррелограммы

# 6) Тест Дарбина-Уотсона

Ограничения:

- применяется только при тест-и автокорреляции первого порядка

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t, \quad \rho \in (-1; 1)$$

- в регрессии должен присутствовать свободный член,
  не должно быть лаговых значений регрессора,
  не должно быть пропусков переменных
- требуется нормальность ошибок $u_t \sim N(\mu, \sigma^2)$
  строгая экзогенность $E(u_t | X) = 0$

Алгоритм:

$$Y_t = \beta_1 + \beta_2 \cdot X_t + u_t$$

где $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

① Оценить регрессию $Y_t$ на $X_t$ при пом. МНК, вычислить остатки $\hat{u}_t$

② Вычислить статистику $DW = \dfrac{\sum (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$

[для $H_1$: $\rho > 0$ : $0 < d^{cr}_{L\alpha} < d^{cr}_{\alpha} < d^{cr}_{U\alpha} < 2$]

(* $DW \xrightarrow{d} 2 - 2\rho$)

$$* \begin{cases} DW \to 0 \iff \rho = 1 \\ DW \to 2 \iff \rho = 0 \\ DW \to 4 \iff \rho = -1 \end{cases}$$

$\Rightarrow DW \in (0; 4)$

③ $H_0: \rho = 0 \qquad H_1: \rho > 0$

$H_0$ ⊖ отв., если $DW < d^{cr.L}_{\alpha}$; ⊖ неопр, если $DW \in (d_L; d_U)$; ⊖ не отв., если $DW > d^{cr.U}_{\alpha}$

## 4.10 Обобщенный МНК + процедура Кохрейна – Оркатта (GLS)

Пусть $Y_t = \beta_0 + \beta_1 X_t + u_t$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 \cdot X_{t-1} + \rho u_{t-1}$$

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + (\beta_1 X_t - \rho\beta_1 \cdot X_{t-1}) + \underbrace{u_t - \rho u_{t-1}}_{\varepsilon_t}$$

Алгоритм:

(1) Оценить регрессию $Y_t$ на $X_t$, вычислить $\hat{u}_t$

(2) Оценить $\hat{u}_t = \rho \cdot \hat{u}_{t-1} + \varepsilon_t$ => получить $\hat{\rho}$

(3) $\tilde{Y}_t = \beta_0' + \beta_1 \tilde{X}_t + \varepsilon_t$, оценить регрессию, получить $\hat{\beta}_0$ из $\beta_0'$, $\hat{\beta}_1$ из $\beta_1$

где $\tilde{Y}_t = Y_t - \hat{\rho} Y_{t-1}$

$\tilde{X}_t = X_t - \hat{\rho} \cdot X_{t-1}$

$\beta_0' = \beta_0(1-\hat{\rho})$

(4) Остатки пересчитываются, (возвращаться к п. 2, пока $|\hat{\beta}_{i,k} - \hat{\beta}_{i,k-1}| < \alpha$)

где $i = 1,2$; $k$ = число шагов

$\alpha$ – заданная точность

★ Поправка Прайса – Уинстена:

$\tilde{Y}_1 = \sqrt{1-\hat{\rho}^2} \cdot Y_1$, $\tilde{X}_1 = \sqrt{1-\hat{\rho}^2} \cdot X_1$

при исп. процедура Кохрейна – Оркатта

пропускает первое наблюдение
(что снижает хар-ки оценок на мал. выборках)

↳ если случ. сост. не коррелируют с 1ым (2, 3, ... ) остатком, то можно исп. 1ое набл. без преобразований

↳ однако при примен. МНК оно будет оказывать слишком большое влияние на оценки параметров

↳ вводится поправочный коэф. $K = \sqrt{1-\rho^2}$
$\Rightarrow$ 1ое уравнение становится соизмеримо остальным

## 5.3. Модель множественной линейной регрессии

$$Y_i = \beta_0 \cdot X_{0i} + \beta_1 X_{1i} + \beta_2 \cdot X_{2i} + \ldots + \beta_k X_{ki} + u_i, \quad \forall i = \overline{1,n}$$

где $Y_i$ – $i$-е наблюдение зав. переменной

$X_{1i}, \ldots, X_{ki}$ – $i$-е набл. каждого из $k$ регрессоров

$u_i$ – ошибка регрессии

$\beta_0$ – константа (ожид. знач. $Y_i$ при усл. $X_{ji} = 0$ $\forall j = \overline{1,k}$

$X_{0i} = 1$, $\forall i = \overline{1,n}$ – постоянный регрессор

$\beta_1, \ldots, \beta_k$ – коэф. наклона при пер. $X_1, \ldots, X_k$, соотв.

(коэф. $\beta_1$ отражает ожид. изм. $Y_i$

в результате изм. $\Delta X_1 = 1$

при фикс. $\overline{X}_2, \ldots, \overline{X}_k$) $\Rightarrow$ $\beta_1 = \frac{\Delta Y}{\Delta X_1}$

$$E(Y_i \mid X_{1i} = x_1, \ldots, X_{ki} = x_k) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

– линия теор. регрессии

(отнош., которое в среднем

выполняется между $X$ и $Y$ в ГС)

## 5.4 МНК - оценка коэффициентов м.л.р.

МНК - оценки коэф. регрессии $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_k$ -

это значения $\beta_0, \beta_1, \ldots \beta_k$:

$$\sum (Y_i - b_0 - b_1 X_{1i} - \ldots - b_k X_{ki})^2 \xrightarrow[b_0, b_1, \ldots, b_k]{} \min$$

где $\hat{\beta}_0 + \hat{\beta}_1 \cdot X_{1i} + \ldots + \hat{\beta}_k \cdot X_{ki}$ -

- МНК-оценка линии регрессии

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{1i} + \ldots + \hat{\beta}_k \cdot X_{ki}$ - предск. знач. $Y_i$

$\hat{u}_i = Y_i - \hat{Y}_i$ - остаток МНК-регрессии

## 5.5. Критерии кач-ва приближения данных моделью

SER - стандартная ошибка регрессии -

оценка стандартного откл. случ. ошибок регрессии $u_i$

(мера разброса $Y$ вокруг линии регрессии)

$SER = S_{\hat{u}}$

где $S_{\hat{u}}^2 = \frac{1}{n-k-1} \sum \hat{u}^2 = \frac{RSS}{n-k-1}$

число коэф. своб. гл.

$R^2$ регрессии - коэффициент детерминации

$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ (при вкл. новых регрессоров: TSS-const; RSS↓; $R^2$↑)

доля выб. дисперсии $Y_i$, предсказ. (объясн.) $X_{ji}$

где $TSS = \sum (Y_i - \bar{Y})^2$

$ESS = \sum (\hat{Y}_i - \bar{Y})^2$

$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$

$\overline{R}^2$ - скорректированный коэф. детерминации регрессии

( рост $R^2$ при вкл. новой переменной м. дать завышенную оценку качества подгонки модели)

$$\overline{R}^2 = 1 - \underbrace{\frac{n-1}{n-k-1}}_{\text{всегда} > 1} \cdot \frac{RSS}{TSS} = 1 - \frac{S_u^2}{S_y^2}$$

$\Rightarrow \overline{R}^2 < R^2$

Эффекты вкл. переменных:

$\downarrow RSS \Rightarrow \overline{R}^2 \uparrow$

$\uparrow \frac{n-1}{n-k-1} \Rightarrow \overline{R}^2 \downarrow$

## 5.6. Предположение МНК для м.л.р.

$$Y_i = \beta_0 + \beta_1 \cdot X_{1i} + \ldots \beta_k \cdot X_{ki} + u_i$$

1. Усл. распр. $u_i$ отн. $X_i$ имеет нул. среднее

$$E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$$

2. $(X_{1i}, \ldots, X_{ki}, Y_i)$ - i.i.d.

3. Большие выбросы маловероятны:

$X_{1i}, \ldots, X_{ki}$ и $Y_i$ имеют ненул. конечные четвертые моменты

4. Отсутствует совершенная мультиколлинеарность

▷ 1. гарантия несмещенности МНК-оценок

2. выполняется, если данные получены с помощью простого случ. отбора

3. $0 < E(X_{1i}^4) < \infty, \ldots, 0 < E(X_{ki}^4) < \infty, \quad 0 < E(Y_i^4) < \infty$

$\Rightarrow \quad \chi_i = \dfrac{E(Y_i - \mu_Y)^4}{\sigma_x^4} \in (0; \infty)$ – конечные коэф. эксцесса для $X_i$, $Y_i$

большие выбросы могут привести к смещению МНК-оценок

## 5.7. Совершенная (строгая) мультиколлинеарность регрессоров

– наличие регрессора, который является точной линейной комбинацией др. регрессоров.

Последствия:

– МНК-оценки получить невозможно (из-за деления на 0)

▷ $Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + u_i$

где $X_{1i} = \lambda X_{0i}$

$\Rightarrow \quad \hat{\beta}_1 = \dfrac{\sum (X_{1i} - \bar{X}_1)(Y_i - \bar{Y})}{\sum (X_{1i} - \bar{X}_1)^2} = \dfrac{\sum (\lambda X_{0i} - \lambda) Y_i}{\sum (\lambda - \lambda)^2}$

$\Rightarrow$ обычно вызвана ошибкой спецификации (требуется изменить набор объясн. переменных)

e.g. ловушка фиктивных переменных

## 5.8. Несовершенная мультиколлинеарность регрессоров

– две или более объясн. переменные (их лин. комб.) сильно коррелированы между соб.

* в м.л.р. присутствует практически всегда

Причины:

– ошибочное включение (≥2) мн. зависимых переменных

– (≥2) объясн. переменные слабо коррелированы в ГС, но сильно корр-ны в конкретной выборке

– в модель включается объясняющая переменная, сильно корр-ая с зависимой ("доминирующая")

Последствия:

– с мат. т.з. оценки м.б. расчитаны, но оценки МНК при коллинеарных переменных не явл. эффективными (хотя и не смещены)

– станд. ошибки оценок коэф. завышены

=> сост. t-статистики занижены

– неустойчивость оценок (вкл. нового регрессора сильно меняет оценки коэф.)

– трудности с интерпретацией коэф. при коррел-ых перем-ых

– общее качество уравнения, а так-же оценки коэф. при некор-ых пер. остаются незатронуты

! Чем сильнее мультиколлинеарность, тем "больше" последствия

☆ Часто коэф. при кор. пер. получаются незначим/с "неверн." знаком

("1. $\beta_i > 0$; $\beta_j > 0$". 2. $\rho_{x_j y_i} > 0$, $\forall j = \overline{1,k}$; 3. $\rho_{x_i x_j} > \rho_{x_j y_i}$ $\forall j$ => $\beta_i$ или $\beta_j$)

Признаки:

- "Частные" коэф. корреляции $|\rho_{x_i x_j}| > 0,7$
- Определитель м-цы коэф. парной корреляции близок к нулю !
- Оценки коэф. имеют большие ст. ошибки (незнач. значимость), но в целом модель значима ($R^2$ высокий, F-статистика)
- Введение в модель новой объясн. переменной (коллинеарной) приводит к существ. изм. оценок коэф. (даже до смены знака) и небольшому изм. $R^2$

Тестирование мультиколлинеарности

VIF (variance inflation factor) —

фактор обесценения (возрастания) дисперсии

$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots \beta_k X_{ki} + u_i$

1 Оценить регрессию $X_j$ на $X_1, \ldots, X_k$ ($i = \overline{1,k} / j$), определить $R_j^2$ — к. дет. в этой регрессии

2 Считаем статистику:

$$VIF(X_j) = \frac{1}{1 - R_j^2}$$, если $VIF(X_j) > 10$

$\Rightarrow$ мультиколлин-ть

Методы "борьбы":

прямые:

1) попытаться уменьшить $\sigma_u^2$

(если $\exists$ важная переменная,

не вкл. в модель, то её

добавление снизит теор. $\sigma_u^2$)

2) увеличить число наблюдений

(для ср. рядов уменьшить вр. интервалы)

↳ м. б. усилен. автокор-ия

↳ м. б. внесен смещение, т.к. данные будут менее точными

3) увеличить дисп. объясняющ. перем.

(группировка)

4) на стадии планирования опроса

выбрать наименее зависимые перем.

косвенные:

1) если кор. переменные имеют концептуально,

то можно объед. их в совок. индекс

2) исключить нек. из кор. перем-ых,

если их коэф. незначимы (опасно)

3) исп. внешнюю инф. отн. коэф. одной из пер.

$$Y = \beta_1 + \beta_2 X + \beta_3 P + u$$

$$\hat{Y}' = \beta_1' + \beta_2' X' \Rightarrow b_2' \Rightarrow \underbrace{Y - b_2' X}_{Z} = \beta_1 + \beta_3 P + u$$

? $\hat{\beta}_3$ зависит от точности $b_2'$

4) ограничения (e.g. $\beta_3 = \beta_4 \Rightarrow \beta_3 (X_i + Y_i)$)

## 5.9. Распределение МНК-оценок коэф. м.л.р.

Если вып. все 4 предположения МНК для м.л.р., то МНК-оценки коэф. модели совместно норм. распр-ны, являются уст. несмещ. и состоят-ми, и оценка любого коэф. $\hat{\beta}_i$:

$$\hat{\beta}_i \overset{d}{\sim} N(\beta_i, \sigma^2_{\hat{\beta}_i})$$

▷ прим. ЦПТ, т.к. $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ — явл. ср. знач. оценок, получ. по разл. случ. выборкам ($\Rightarrow \overset{d}{\sim} N(\mu, \sigma^2)$)

## 5.10. Теорема Фриша-Во

МНК-оценки м.л.р. можно получить, последовательно оценивая опр. набор регрессий

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k \cdot X_{ki} + u_i$$

Алгоритм вычисления МНК-оценки:

1. Оценить регрессию $X_1$ на $X_2, \ldots, X_k$, вычислить остатки этой регрессии $\tilde{X}_1$
2. Оценить регрессию $Y$ на $X_2, \ldots, X_k$, вычислить остатки этой регрессии $\tilde{Y}$
3. Оценить регрессию $\tilde{Y}$ на $\tilde{X}_1$,

где все регрессии был. своб. член

## Теорема Фриша - Во:

МНК - оценка коэф. в регр. $\tilde{Y}$ на $\tilde{X}_1$

равна МНК - оценке при $X_1$ в регр. $Y_i$ на $X_i$

Интерпретация:

$\hat{\beta}_1$ - оценка влияния переменной $X_1$ на $Y$,

при контрол. $X_2, \dots, X_k$

- Оценив (1) и (2) регрессии, мы очистили $Y$ и $X_1$ от влияния других регрессоров
- (3) регрессия дает влияние очищ. переменной $X_1$ на очищ. переменную $Y$

## 6 Проверка гипотез в модели м. л. р.

### 6.1. Проверка гипотез для одного коэффициента

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$$
$$(SE(\hat{\beta}_0)) \quad (-\text{"}-) \qquad\qquad (-\text{"}-)$$

, причем $SE(\hat{\beta}_j)$ – HC (Уайта)

$H_0: \beta_j = \beta_{j,0}$ , $j = \overline{0,k}$

$H_2: \beta_j \neq \beta_{j,0}$ , $j = \overline{0,k}$

1. Вычислить станд. ошибку $(SE(\hat{\beta}_j))$

2. Вычислить $t$-статистику:

$$t^{act} = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}$$

3. Вычислить p-value:

p-value $= 2\Phi(-|t^{act}|)$

$H_0$ отв. на 5% ур. зн., если p-value $< \alpha$

$\Leftrightarrow |t^{act}| > |t_{\frac{\alpha}{2}}| = 1{,}96$

### 6.2. Построение дов. инт. для одного коэф-та.

95%-дов. инт. для $\beta_j$: $\{\hat{\beta}_j \pm 1{,}96 \cdot SE(\hat{\beta}_j)\}$

### 6.3. Проверка совместных гипотез

$$Y_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \beta_3 \cdot X_{3i} + u_i$$

$H_0: \beta_1 = 0$ и $b_2 = 0$

$H_2: \beta_1 \neq 0$ или $\beta_2 \neq 0$ (или оба)

где $H_0$ – совместная гипотеза (накл. ограничение на два (или более) коэф. регр

## Использование $t$-статистик не оправдано

т.к. $H_0$ отвергается слишком часто (чаще 5%)

пусть $t_1 = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)}$ ; $t_2 = \frac{\hat{\beta}_2 - 0}{se(\hat{\beta}_2)}$

– независимо распр. (для простоты)

тогда $H_0$ отв. если

$$|t_1| > 1{,}96 \quad \text{и/или} \quad |t_2| > 1{,}96$$

$$\text{p-value} = P_{H_0}[\,|t_1| > 1{,}96 \text{ и/или } |t_2| > 1{,}96\,]$$

$$= 1 - P_{H_0}[\,|t_1| \leq 1{,}96 \text{ и } |t_2| \leq 1{,}96\,] = (\text{из незав-ти})$$

$$= 1 - P_{H_0}[\,|t_1| \leq 1{,}96\,] \cdot P_{H_0}[\,|t_2| \leq 1{,}96\,] =$$

$$= 1 - 0{,}95 \cdot 0{,}95 = 0{,}0975$$

$$9{,}75\% \neq 5\%$$

Вывод:

- предложенная процедура неверна:
  размер критерия неправильный ($\neq 5\%$),
  зависит от корреляции между $t_1$ и $t_2$
  $\Rightarrow$ и между $\hat{\beta}_1$ и $\hat{\beta}_2$

Решение:

- Метод Бонферрони: использовать правильное крит. значение (а не $t^{cr}_{\alpha/2}$)
- исп. другие процедуры тестирования

## Метод Бонферрони:

Критерий Бонферрони для проверки совм. гипотезы

$\begin{cases}\beta_1 = \beta_{1,0}, & \beta_2 = \beta_{2,0}\end{cases}$, исп. крит. значение $C > 0$:

$H_0$ не отв., если $|t_1| \le C$ и $|t_2| \le C$,

иначе отвергается

↳ $C$ выбирается так, что p-value $\le \alpha$

Неравенство Бонферрони:

$$P(A \cup B) \le P(A) + P(B)$$

$$P(\{|t_1| > c\} \cup \{|t_2| > c\}) \le P(\{|t_1| > c\}) + P(\{|t_2| > c\})$$

Критерий Бонферрони:

Если $H_0$ - верна, то

$$P_{H_0}(\text{вер. отвергнуть инд. гипотезы}) \le \underbrace{2}_{q=2} P(|z| > c)$$

$\Rightarrow$ при $\alpha = 5\%$; $c = 2{,}241$

## 6.4 F-статистика для гомоскедастичных ошибок

F-статистика позволяет проверять одновременно все ограничения совм. гипотезы

в частности:

$H_0$: $\beta_1 = \beta_{1,0}$ и $\beta_2 = \beta_{2,0}$

$$F^{HC} = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1 t_2} t_1 \cdot t_2}{1 - \hat{\rho}^2_{t_1 t_2}}\right),$$

где $\hat{\rho}_{t_1 t_2}$ - оценка коэф. корреляции сет. $t_1$ и $t_2$

→ F-статистика большая, если $t_1$ и/или $t_2$ большие

→ F-статистика учитывает корреляцию $t_1$ и $t_2$

→ $F \overset{d}{\sim} F_{\underset{q=2}{2}, \infty}$ вне зависимости от $\rho_{t_1 t_2}$

$F \overset{d}{\sim} \chi^2_q / q \Rightarrow$ p-value $= \Pr [F_{q,\infty} > F^{act}]$

## При гомоскедастичности:

① Оценить регрессию "с ограничениями" и "без"

(т.е. "в предположениях $H_0$" и "в предполож. $H_1$")

② Вычислить $R^2_R$ и $R^2_{UR}$

③ $$F = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/n - k_{UR}(-1)} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/n - k_{UR}(-1)}$$

где $q$ – число ограничений в $H_0$

$k_{UR}$ – число объясн. переменных (без конст.) в модели без ограничений

↳ чем больше разность $R^2_{UR}$ и $R^2_R$,

тем больше улучшение качества подгонки при включении новых переменных

тем больше $F^{act} \Rightarrow H_0$ – отв.

## Устойчивая к гетероскедастичности F-статистика

$$F^{HC} = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1 t_2} t_1 t_2}{1 - \hat{\rho}^2_{t_1 t_2}}\right) \overset{d}{\sim} F_{q,\infty} \left[= \chi^2_q / q\right]$$

при использовании $HC - se(\hat{\beta}_i)$ → вне зав-ти от гомо/гетероск-ти

(или HC - оценки "ковар. м-цы")

## 6.5. Распределение F-статистики

a) для случая гомоскедастичности ош. регрессии (при гетеро-) больших и небольших $n$

если ошибки регрессии (помимо IV предп. МНК)

$u_i$ – гомоскедастичны, т.е. $var(u_i | X_i) = const$

$u_i \sim N(\mu_i, \sigma_i^2)$, $\forall i = \overline{1,k}$

то F-статистика (вычисл. для случая гомоск. ош.)

$F \sim F_{q, n-k-1}$, где $q$ – число огр. в $H_0$

$k$ – число регрессоров без конст. в LR-регр.

b) $F \overset{d}{\sim} F_{q,\infty} = \chi^2_q / q$ ($n \geq 100$)

c) при гетероскедастичности $F \not\sim \chi^2_q / q$

=> приходится исп. HC-F статистику

=> $F^{HC} \overset{d}{\sim} \chi^2_q / q$

## 6.6. Св-ва F-статистики (для случая гомоск-ти ошибок)

- можно пользоваться только при выполнении всех VI предпосылок (4 + гомоск. + N)
- лучше исп. HCF-статистику ($\sim F_{q,\infty}$)
- $F \overset{a}{\sim} F_{q,\infty} = \chi^2_q / q$ (при $n>100$ верно)
- при малых $n$ => $F \sim F_{q, n-k-1}$, т.к. оно имеет большие крит. знач
- Но отв., если включение в модель регрессоров "достаточно" улучш. кач-во модели

## 6.7 - Использование F-статистики для тестирования гипотезы о незначимости регр.

Пусть $Y_i = \beta_1 + \beta_2 X_i + u_i$

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

* при проверке t-тестом в 5% случаев мы отвергнем верную $H_0$

(!) F-тест на основе анализа дисперсии

Пусть $Var(Y) = Var(\hat{Y}) + Var(u_i)$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum \hat{u}_i^2$$

$$TSS = ESS + RSS$$

F-стат. для проверки общ. кач-ва регр. $\rightarrow$ $$F = \frac{ESS/k-1}{RSS/n-k} = \frac{R^2/k-1}{(1-R^2)/n-k}$$

где k - число оц. пар-ров (м. чл. + (k-1))

В частности, при k=2

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

$H_0$ отв. при $F > F^{cr}$

Если $$F^{cr} = \frac{R^2_{cr}/(k-1)}{(1-R^2_{cr})/(n-k)} \Rightarrow R^2_{cr} = \frac{(k-1)F^{cr}}{(k-1)F^{cr} + (n-k)}$$

$$R^2 > R^2_{cr}$$

$\Rightarrow H_0$ - отв.

## 6.8. Тестирование одного ограничения для нескольких коэффициентов

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 \cdot X_{2i} + u_i$$

$H_0: \beta_1 = \beta_2$ — одно ограничение $q = 1$ на неск. коэф. (не совм. гип.)

$H_1: \beta_1 \neq \beta_2$

<u>Способ I</u> Прямой тест

$F^{act} \overset{?}{\lessgtr} F^{кр,\alpha}_{1,\infty} \Rightarrow (1,96)^2 = 3,84$ (при $\alpha = 5\%$)

<u>Способ II</u> Преобразование регрессии

↳ так, чтобы ограничение стало на один коэф.

$$Y_i = \beta_0 + (\beta_1 - \beta_2) X_{1i} + \beta_2 \cdot (X_{1i} + X_{2i}) + u_i$$

$$Y_i = \beta_0 + \gamma_1 X_{1i} + \beta_2 \cdot W_i + u_i$$

где $\gamma_1 = \beta_1 - \beta_2$

$W_i = X_{1i} + X_{2i}$

$H_0: \gamma_1 = 0$

$H_1: \gamma_1 \neq 0$ $\Rightarrow$ $t = \dfrac{\hat{\gamma}_1}{SE(\hat{\gamma}_1)}$

☆ Регрессии "R" и "UR" имеют одинак. $R^2$, ESS, RSS

☆☆ ∃ расширение для $q > 1$ с исп. F-стат.

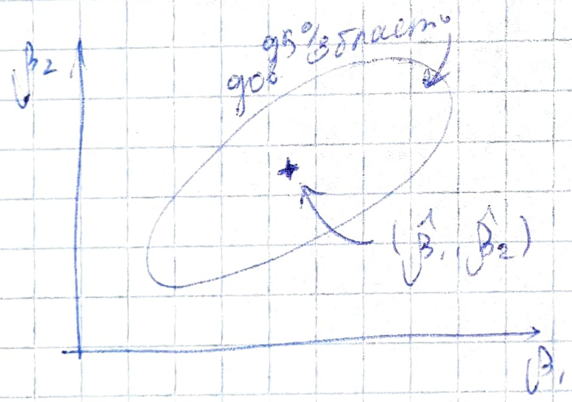
## 6.9. Доверительные области для неск. коэф.

95% довер. мн-во — это

- мн-во знач. коэф., которое содержит ист. коэф. в 95% повтор. выборок
- мн-во знач. коэф., равенство которым не м.б. отвергнуто на 5% ур. зн.

95% дов. обл. для неск. коэф. строится при помощи F-теста.

в частности при $q=2$, (т.е. $\beta_1$ и $\beta_2$) дов области явл. эллипсами




Пусть $H_0: \beta_1 = \beta_{1,0}, \ \beta_2 = \beta_{2,0}$

$F(\beta_{1,0}, \beta_{2,0})$ — HC-F-статистика для проверки совм. гипотезы

95% дов. область $= \{\beta_{1,0}, \beta_{2,0} : F(\beta_{1,0}, \beta_{2,0}) < 3{,}00\}$

где $F^{\alpha}_{\alpha=5\%, 2, \infty} = 3{,}00$

$$F(\beta_{1,0}, \beta_{2,0}) = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat\rho_{t_1 t_2} t_1 t_2}{1 - \hat\rho^2_{t_1 t_2}}\right) =$$

$$= \frac{1}{2(1-\hat\rho^2_{t_1 t_2})} \times \left[\left(\frac{\hat\beta_1 - \beta_{1,0}}{SE(\hat\beta_1)}\right)^2 + \left(\frac{\hat\beta_2 - \beta_{2,0}}{SE(\hat\beta_2)}\right)^2 - 2\hat\rho_{t_1 t_2}\left(\frac{\hat\beta_1 - \beta_{1,0}}{SE(\hat\beta_1)}\right)\left(\frac{\hat\beta_2 - \beta_{2,0}}{SE(\hat\beta_2)}\right)\right]$$

квадр. форма по $\beta_{1,0}, \beta_{2,0}$ $\Rightarrow$ граница — эллипс

## 7 Выбор спецификации модели м.л.р.

### 7.1. Контрольные переменные W

– это переменная, которая коррелирована с пропущенным фактором в регрессии Y на X, но сама не обязательно влияет на Y

Св-ва:

1. Эффективная контрольная переменная при включении в регрессию делает интересующую нас переменную (не коррелированной со случ. ошибкой)
2. В предположении постоянства контрольной (ых) переменной (ых) интересующая нас переменная "как бы" случайно задана
3. Среди объектов с одинаковыми значениями контрольной (ых) переменной (ых), интересующая нас переменная не коррелирована с пропущ. факторами, влияющими на Y

★ т.к. контр. переменная не всегда оказывает влияние на зав. переменную

⇒ коэф. при ней не имеет станд. интерпрет.

## 7.2. Независимость условного среднего:

(I) предположение МНК:

Предположение о нулевом усл. среднем

$\Rightarrow$ предположение о независимости усл. среднего

(☆ коэф. при контр. перем. м.б. смещен

$\Rightarrow E(u_i | X_i) \neq 0 \Rightarrow$ новое предполож.)

<u>Независимость условного среднего.</u>

↳ при наличии <u>эффект.</u> контрольной переменной

усл. среднее $u_i$ <u>не зависит</u> от интересующей нас переменной

Пусть $X_i$ – интересующая нас переменная,

$W_i$ – контр. переменные

$W$ – эффект. контрольные переменные, если

выполняется усл. нез-ти усл. среднего:

$$E(u_i | X_i, W_i) = E(u_i | W_i)$$

☆ (это усл заменяет (I) предп. МНК

для случая наличия контр. переменной)

## 7.3. Предположение МНК для м.л.р. с контр. переменной

1. $E(u_i | X_i, W_i) = E(u_i | W_i)$

2. $(X_i, Y_i)$ – i.i.d.

3. большие выбросы маловероятны

4. отсутствует соверш. мультиколлинеарность

## 7.4. Св-ва МНК-оценок м.л.р. с контр. переменными

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot W_i + u_i$$

при выполнении предполож. для МНК с контр пер

(1) $\beta_1$ имеет причинно-следственную интерпретацию

(2) $\hat{\beta}_1$ – несмещенная оценка (при интерес. иск пер.)

(3) $\hat{\beta}_2$ – (в общ. случае) смещена (при контр. пер.)

▶ $E(Y \mid X = \Delta X + X, W = w) - E(Y \mid X = x, W = w)$

$= [\beta_0 + \beta_1(X + \Delta X) + \beta_2 W + E(u \mid X = X + \Delta X, W = w)]$

$- [\beta_0 + \beta_1 X + \beta_2 W + E(u \mid X = x, W = w)]$

$= \beta_1 \cdot \Delta X + [E(u \mid X = x + \Delta x, W = w) - E(u \mid X = x, W = w)]$

$= \beta_1 \cdot \Delta X + 0 = \beta_1 \cdot \Delta X$

ожидаемое $\Delta Y$ вследствие изм. $\Delta X$ при фикс. W

где $E(u \mid X = x + \Delta X, W = w) = E(u \mid X = x, W = w) =$
$= E(u \mid W = w)$

в силу (I) предпосылки

▶ Пусть $E(u_i \mid W_i) = \gamma_0 + \gamma_2 W_i$

и пусть $v_i = u_i - E(u_i \mid X_i, W_i)$

$\Rightarrow E(v_i \mid X_i, W_i) = [E(u_i \mid X_i, W_i) -$

$E(u_i \mid X_i, W_i)] = 0$

$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \gamma_0 + \gamma_2 \cdot W_i + v_i$

$\Rightarrow$ $(\beta_0+\gamma_0)+\beta_1 x_i+(\beta_2+\gamma_2)w_i+v_i=$

$=\delta_0+\beta_1 x_i+\delta_2 \cdot w_i+v_i$

где $E(v_i \mid x_i, w_i)=0$

$\Rightarrow$ (I) предп. МНК удовл.

$\Rightarrow$ $\delta_0, \beta_1, \delta_2$ – не смещены

т.к. регрессоры в моделях одинаковы

$E(\hat{\beta}_1)=\beta_1$

$E(\hat{\beta}_2)=\delta_2=\beta_2+\gamma_2 \neq \beta_2$

$\hookrightarrow$ где $\hat{\beta}_1$ при инт. иск. переменной – несмещ.

$\hat{\beta}_2$ смещена из-за кор-ти контрольной переменной и пропущенной переменной,

где $\gamma_2$ – смещ. из-за проп. пер.

## 7.5. Спецификация модели:

– Если мы хотим измерить переменную, мы включаем её в регрессию $\Rightarrow$ избежать смещения из-за пропущенной перем.

– Смещение из-за пропущенной переменной приводит к нарушению (I) предпосылки МНК

$\Rightarrow$ формулы для ст. ошибок коэф. и тест. статистики становятся неприменимы

$$\hat{\beta}_2 = \beta_2 + \underbrace{\beta_3 \underbrace{\frac{Cov(X_2, X_3)}{Var(X_2)}}_{= \rho_{X_2, X_3} \frac{\sigma_{X_3}}{\sigma_{X_2}}}}_{\text{смещение из-за проп. пер.}} + \underbrace{\frac{Cov(X_2, u)}{Var(X_2)}}_{= 0 \text{, т.к. экзогенност.}}$$

где $X_3$ – интерес. нас переменная

! Если мы не можем измерить проп. переменную, мы можем контролировать её влияние, включив контрольную перем.

· Алгоритм:

① Определить целевую переменную

② Определить источник смещения из-за пропущ. перем.

③ Если возможно, включить пер. в регрессию, иначе – коррелированные с ними контрольные переменные

↳ Контр. переменные эффективны, если выполняется усл. нез-ти усл. среднего

↳ полученная модель – "базовая"

④ Определить мн-во возможных альтернативных моделей, вкл. доп. переменные

⑤ Оценить "базовую" модель и предполагаемые альтернативы

↳ "проверка чувствительности"

? изм. ли коэф. при интерес. ($\beta_1$) переменной при вкл. доп. регрессоров

? коэф. при доп. регрессоре стат. значим.

$R^2$ и $\bar{R}^2$ ~~→~~ max, лучше $E(\hat{\beta}) = \beta$

добиться несмещ. цел. пок-лей

1) Высокий $R^2$ означает, что регрессоры хорошо объясн. вариацию Y

2) Высокий $R^2$ не означает, что

- нет смещения из-за проп. переменной
- оц. ценного пар-ра $\beta_1$ не смещ.
- включ. переменные статистически значимы

Подход: специфицировать базовую модель, и затем проверять чувствительность оценок ключ. показателей в альт. специф.

## 7.6. Информационные критерии Акаике и Шварца

$Y_i = \beta_0 + \beta_1 \cdot X_{1i} + \ldots + \beta_k X_{ki}$

Информационный критерий – мера отн. кач-ва модели ("степени подгонки") под данные с корректировкой (штрафом) на используемое кол-во параметров

↳ суть: найти компромисс между точностью и сложностью модели

Инф. к-р критерии связаны с концепцией информац. энтропии и расст. Кульбака – Лейблера (на этой основе был разработан AIC)

$$*\left(D_{KL}(p\|q) = E(-\ln q(x)) - \underbrace{E(-\ln p(x))}_{\text{энтропия}}\right)$$

И.к. используются для сравнения моделей между собой (без содержательной интерпретации)

Критерий Акаике

$$AIC = \left(2(k+1) - 2\ln(L)\right)\frac{1}{n} = \frac{1}{n}\left[2(k+1) + n\left[\ln\left(2\pi\frac{RSS}{n}\right)+1\right]\right]$$

$$= \ln\frac{RSS_k}{n} + \frac{2(k+1)}{n} + 1 + \ln 2\pi$$

Критерий Шварца (байесовский)

$$BIC = \frac{1}{n}\left((k+1)\ln n + n\left[\ln\left(2\pi\frac{RSS}{n}\right)+1\right]\right)$$

$$= \ln\frac{RSS_k}{n} + (k+1)\frac{\ln n}{n} + 1 + \ln 2\pi$$

## 7.7. Тест Харке – Бера

– стат. тест, проверяющий ошибки наблюдений на нормальность посредством сверки их третьего момента (ассиметрия) и четвертого (эксцесс) с моментами нормального распр., у которого $S=0$, $K=3$

skewness (к. ассим-и) Kurtosis (к. эксцесса)

$H_0: S=0, K=3$

$H_1: S\neq 0$ или $K\neq 3$

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right) \sim \chi^2_2$$

где $S = \frac{m_3}{\sigma^3}$; $K = \frac{m_4}{\sigma^4} - 3$; $m_k = \frac{1}{N}\sum(\hat{u}_i - \bar{u})^k$

# ⑧ Нелинейные регрессионные модели

## 8.1. Нелинейность по переменным –

$\partial \hat{Y} / \partial X_j \neq g_j(\beta)$ – ф-е не содерж. регрессоров

Нелинейность по параметрам –

$\partial \hat{Y} / \partial b_j \neq g_j(X)$ – ф-е не сод. неизв. пар-ров

e.g. модель лин. по параметрам и нелин. по переменным

$$Y = \beta_1 + \beta_2 \cdot X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + u$$

$$\Downarrow$$

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 \cdot Z_3 + \beta_4 \cdot Z_4 + u$$

e.g. нелин. по перем. и параметру $a_1$

$$Y = a_0 K^{a_1} \cdot L^{(1-a_1)}$$

## 8.2 Моделирование нелинейности по переменным

1. Идентифицировать возможное нелинейное соотношение
2. Специфицируйте нелинейную функцию и оцените её параметры при помощи МНК
3. Определить, улучшает ли нелинейная модель линейную
4. Изобразить графически оцененную ф-ю нелинейной регрессии
5. Оценить изменение $Y$ в зависимости от изменения $X_i$

$$\Delta Y = f(\Delta X_1 + X_1, X_2, \ldots, X_k) - f(X_1, \ldots, X_k)$$

$$\Delta \hat{Y} = \hat{f}(\Delta X_1 + X_1, X_2, \ldots, X_k) - \hat{f}(X_1, \ldots, X_k)$$

## 8.3. Полиномиальная регрессия

Модель полином. регрессии степени $r$ –

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot X_i^2 + \ldots + \beta_r \cdot X_i^r + u_i$$

Выбор спецификации:

↳ определить компромисс между гибкостью модели и точностью

$r \uparrow$ $\uparrow$ гибкость

$r \uparrow$ $\downarrow$ точность оцененных коэф.

Процедура последовательного тестирования гипотез:

(1) Выбрать max значение $r$,

Оценить полиномиальную регрессию ст. $r$

(2) $H_0: \beta_r = 0 \Rightarrow t = \frac{\hat{\beta}_r}{SE(\hat{\beta}_r)}$

если $H_0$ отв., то $X^r$ ост. в регрессии

если $H_0$ не отв., то $X^r$ искл. из регрессии

(3) Оценить полин. регрессию ст. $r-1$

(4) –//– продолжать шаги 2, 3, пока коэф. при наиб. степени в полиноме не будет ст. значим

Тестирование (гипотеза о линейности)

$$\left.\begin{array}{l} H_0: \beta_2 = \ldots = \beta_r = 0 \\ H_1: \exists j : \beta_j \neq 0, \ j = \overline{2, r} \end{array}\right\} \Rightarrow F^{act} > F^{\alpha}_{r-1, \infty}$$

Интерпретация коэффициентов

↳ построить граф. оц. ф-и регрессии;

вычислить оцененное $\Delta\hat{Y}$ связ. с изм. $\Delta X$;

в частности $r=2$

$$\Delta\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 (X+\Delta X) + \hat{\beta}_2 (X+\Delta X)^2 - \hat{\beta}_0 - \hat{\beta}_1 X - \hat{\beta}_2 X^2 =$$

$$= \hat{\beta}_1 \cdot \Delta X + \hat{\beta}_2 \cdot (2X \cdot \Delta X + \Delta X^2)$$

$$SE(\Delta\hat{Y}) = \frac{|\Delta\hat{Y}|}{\sqrt{F}}$$

где $F$ : стат. для тест $H_0 : \Delta\hat{Y} = 0$

## 8.4. Логарифмические регрессии

a) Линейно-логарифмическая ($Y$ от $\ln X$)

$$Y_i = \beta_0 + \beta_1 \cdot \ln(X_i) + u_i$$

$$\Delta\hat{Y} = \beta_0 + \beta_1 \cdot \ln(X+\Delta X) - \beta_0 - \beta_1 \cdot \ln(X) =$$

$$= \beta_1 \ln\left(\frac{X+\Delta X}{X}\right) \cong \beta_1 \cdot \frac{\Delta X}{X}$$

b) Логарифмически-линейная ($\ln Y$ от $X$)

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

$$\ln(\Delta Y + \hat{Y}) - \ln(\hat{Y}) = \beta_0 + \beta_1 \cdot (X + \Delta X) - \beta_0 - \beta_1 \cdot X = \beta_1 \Delta X$$

$$\frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

c) Линейная в логарифмах ($\ln Y$ от $\ln X$)

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

$$\ln(\Delta Y + \hat{Y}) - \ln(\hat{Y}) = \beta_1 (\ln(\Delta X + X) - \ln(X)) \Rightarrow \frac{\Delta Y}{Y} \cong \frac{\Delta X}{X}$$

## 8.5. Интерпретация $\beta_1$

a) $dY \approx \beta_1 \frac{dX}{X}$ $X \uparrow 1\%$ $Y \uparrow \beta_1$

b) $\frac{dY}{Y} \approx \beta_1 \cdot dX$ $X \uparrow 1$ $Y \uparrow \beta_1 \%$

$\Big\}$ => $\beta$ - полуэласт.

c) $\frac{dY}{Y} \approx \beta_1 \frac{dX}{X}$ $X \uparrow 1\%$ $Y \uparrow \beta_1 \%$ => $\beta$ - эластичность

## 8.6. Эластичность

$$E_{Y/X} = \frac{\partial \ln y}{\partial \ln x} = \frac{dY/Y}{dX/X}$$

## 8.7. Эластичность в моделях парной и лог. регр.

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad E_{Y/X} = \frac{\beta_1 \cdot x_i}{\beta_0 + \beta_1 x_i}$$

$$y_i = \beta_0 + \beta_1 \cdot \ln x_i + u_i \qquad E_{Y/X} = \frac{\beta_1}{\beta_0 + \beta_1 x_i}$$

$$\ln y_i = \beta_0 + \beta_1 \cdot x_i + u_i \qquad E_{Y/X} = \beta_1 x_i$$

$$\ln y_i = \beta_0 + \beta_1 \ln x_i + u_i \qquad E_{Y/X} = \beta_1$$

## 8.8 Выбор между нелинейными моделями:

Используя критерии качества

подгонки можно сравнить с один. регрессантом

Тест Бокса - Кокса

- метод ~~преобразования~~ выбора функц.

зав-ти между переменными

$$Z = \frac{Y^\lambda - 1}{\lambda} = \begin{cases} Y - 1, & \lambda = 1 \\ \to \log Y, & \lambda = 0 \end{cases}$$

Пусть $Y^* = \frac{Y^{\lambda_1} - 1}{\lambda_1}$ $X^* = \frac{X^{\lambda_2} - 1}{\lambda_2}$

$y^* = \beta_0 + \beta_1 x^* + u$

Оценить ~~при~~ помощи МНК $\Rightarrow \hat{\beta}_0, \hat{\beta}_1$

$\lambda_1, \lambda_2$ оцениваются НМНК

при $\lambda_2 = 1, \lambda_1 = 1$ – лин. мод.

при $\lambda_2 = 0 \;\; \lambda_1 = 1$ – лин.-логар. мод.

при $\lambda_2 = 1 \;\; \lambda_1 = 0$ – лог.-лин. мод.

при $\lambda_2 = 0 \;\; \lambda_1 = 0$ – лин. в лог. мод.

Тест сравнивает кач-ва подгонки моделей $\lambda_1, \lambda_2$

Тест Зарембки

- один из вариантов теста Бокса – Кокса специально для случая выбора между лин. и логарифм. мод. регрессии

$$\bar{\bar{y}} = \sqrt[n]{\Pi y_i} = \left(e^{\log \Pi y_i}\right)^{\frac{1}{n}} = e^{\frac{1}{n} \log \Pi y_i} = e^{\frac{1}{n} \Sigma \log y_i}$$

$$\sqrt[n]{\Pi y_i} = \exp(\overline{\log y_i})$$

① Вычислить $\sqrt[n]{\Pi y_i} = \bar{\bar{y}}$

② Рассчитать $y_i^* = \dfrac{y_i}{\bar{\bar{y}}}$

③ Оценить $y_i^*$ и $\log y_i^*$ на $x_i$

$\Rightarrow$ $RSS_1 > RSS_2 \Rightarrow 2$ модель лучше

④ $\hookrightarrow \frac{n}{2} \left| \log \frac{RSS_1}{RSS_2} \right| \sim \chi^2_1$

## 8.9. Регрессионные модели с колич. взаимод-е

Взаимодействие между бин. и непр. переменными:

пусть $X_i \times D_i$ - компонента взаимодействия, теор. линия регрессии может иметь коэф. наклона, зависящий от $D_i$ - бин. пер.

① Различные конст., один угол наклона

$$Y_i = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot D_i + u_i$$

② Различны и конст., и угол накл.

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot D_i + \beta_3 (X_i D_i) + u_i$$

③ Одинаковые конст., разл. угл. коэф.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_3 (X_i D_i) + u_i$$

Взаимодействие между двумя бин. перем.

$$Y_i = \beta_0 + \beta_1 \cdot D_{1i} + \beta_2 D_{2i} + u_i$$

$$+ \beta_3 D_{1i} \times D_{2i}$$

- компонента взаимодействие

Взаимодействие между двумя непр. пер.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \cdot X_{2i}) + u_i$$

$$\hookrightarrow \Delta Y = (\beta_1 + \beta_3 X_2) \Delta X_1$$

$$\Delta Y = (\beta_2 + \beta_3 X_1) \Delta X_2$$

$\triangle$ Включение комп. взаимод-е $X_1 \times X_2$ позв. моделировать эф. влияние изм. $\Delta X_1$ и $\Delta X_2$

## 8.10 Примеры моделей, нелинейных по пар-рам

a) Модель Михаэлиса – Ментона

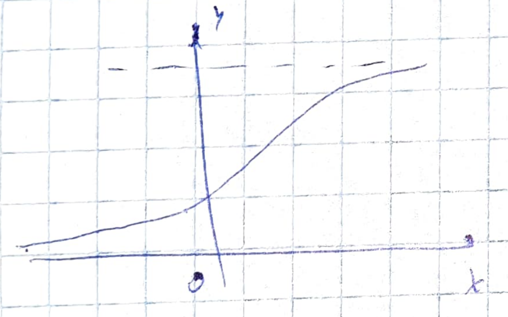
$$Inf = \frac{\theta_1 \cdot Unjob}{\theta_2 + Unjob}$$

$$Y = \frac{1}{Inf} = \frac{\theta_2/\theta_1}{Unjob} + \frac{1}{\theta_1} = \alpha + \beta \cdot X$$

$$Y = \frac{1}{Inf}, \quad X = \frac{1}{Unjob}, \quad \alpha = \frac{1}{\theta_1}, \quad \beta = \frac{\theta_2}{\theta_1}$$

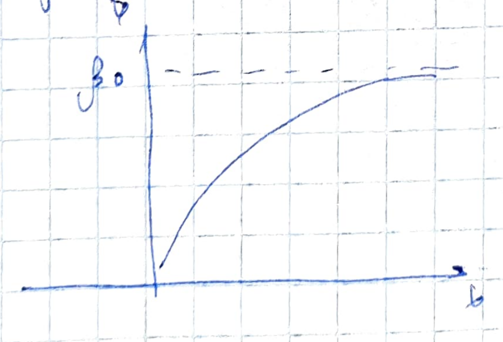
b) Логистическая функция

$$Y_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i) + u_i}}$$




c) Ф-я отрицательного эксп. роста

$$Y_i = \beta_0 \left[ 1 - e^{-\beta_1 (X_i - \beta_2)} \right] + u_i$$



8.11. Нелинейный МНК

$y_i = f(x_{1i}, \dots, x_{ki}; \beta_0, \dots, \beta_m) + u_i$

НМНК:

$$\sum (y_i - f(x_{1i}, \dots, x_{ki}; \beta_0, \dots, \beta_m))^2 \to \min_{\beta_0, \dots, \beta_m}$$

## ⑨ Внутренняя и внешняя обоснованность исследования

### 9.1 Внутренняя обоснованность — анализ

(стат. проверка гипотез о причинных эффектах) обоснован для изучаемой ГС

Внешняя обоснованность — анализ

(стат. проверка гипотез и стат. выводы) м.б. обобщены с исслед. ГС и зад. условий на другие ГС и желаемые условия

Изучаемая ГС — ГС объектов, из кот взяты выборки

Целевая ГС — ГС объектов, на кот. распр. результаты стат. исслед. (исходя из прич-след. связей

### 9.2. Условия внутр. обосн-ти

Внутренняя обоснованность (условие)

1) несмещ. и сост. оценок
2) тест. гипотез должно иметь желаемый ур. зн. (д.и. — ур. доверия)

### 9.3 ~~Внешняя~~ внутр обоснованность (угрозы)

1) пропущенные переменные
2) неправильная спецификация функц. формы регр
3) неточное измерение независим. переменных
4) одновременная причинность отбор наблюдений

## 9.4. Смещение из-за пропущенных переменных:

↳ пропущенная переменная должна

(1) влиять на Y

(2) быть коррелированной с 1 и более перем, вкл. в регр.

Смещение сохраняется (даже асимпт.)

=> МНК-оценка несостоятельна

Если переменная <u>измерима</u>

(∃ адекватная (контрольная) переменная)

① определить интересующие коэф-ты

② определить возм. ист. смещ. из-за проп. пер.

③ проверка дополн. "сомнительных" переменных на нецелевые коэф-ты

④ оценка изменения оценок при включении контр. переменных

Если переменная <u>неизмерима</u>

① использование панельных данных (т.к. они учитывают ненаблюдаемые пропущенные переменные, не мен. во времени)

② регрессия с инструмент. переменными

③ строгий управляемый эксперимент

## 9.5. Смещение из-за неправильной спецификации

Неправильная спецификация ф. ф. — отличие оцененной функции регрессии от функциональной формы теор. ф-и регрессии

=> Оценка частного эффекта от изм. одной или более пер. будет в общ. случае смещенной

(A) м.б. обнаружена, если изобразить граф-и данные и оц. ф-ю регрессии

(A) м.б. скорректирована, используя др. ф.-форму

Решение:

1. Непрерывная зависимая переменная => использование "подходящей" нелин. спецификации по X
2. Дискр. (бинарная, e.g.) зав. переменная => использование спец. методов оценки регрессий ("пробит" и "логит" модели для бин. з. п.)

# 9.6. Смещение из-за ошибок измерения

Смещ. из-за ошибок в экзогенных переменных возникает при оценке МНК, если незав. перем. измерена неточно

Причины:

- ошибки ввода / сбора данных
- вопросы, на кот. респонденты не хотят отвечать
- намеренная ложь

Пусть $Y_i = \beta_0 + \beta_1 X_i + u_i$

$E(u_i | X_i) = 0$

где $X_i$ – истинное знач. $X$

$\tilde{X}_i$ – набл. (с ошибкой) значение $X$

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + \beta_1 \cdot X_i - \beta_1 \tilde{X}_i + u_i$$

$$= \beta_0 + \beta_1 \tilde{X}_i + [\underbrace{\beta_1 (X_i - \tilde{X}_i) + u_i}_{v_i}]$$

=

=

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 - \beta_1 \cdot \frac{\sigma_w^2}{\sigma_X^2 + \sigma_w^2} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2} \beta_1$$

Если $(w_i, u_i)$ : $\rho_{wu} \neq 0$ => Смещение

Методы борьбы:

① Найти точную меру $X$

② регрессия с инструментальными переменными

(в основе: использование "инстр." переменной, которая коррелирована с факт. знач., но некоррелирована с ошибками измерения)

③ Разработать мат модель с ошибками измерения &, если возможно, в использ. полученные ф-л для коррекции оценок

## 9.7 Классическая модель с ошибками измерения

Пусть $\tilde{X}_i = X_i + w_i$

$$\left.\begin{array}{l} \operatorname{corr}(w_i, X_i) = 0 \\ \operatorname{corr}(w_i, u_i) = 0 \end{array}\right\} \Rightarrow \hat{\beta}_1 \xrightarrow{p} \frac{\sigma_X^2}{\sigma_X^2 + \sigma_w^2} \beta_1$$

▷ $v_i = \beta_1 (X_i - \tilde{X}_i) + u_i = -\beta_1 w_i + u_i$

$E(u_i | X_i) = 0 \Rightarrow \operatorname{cov}(u_i, X_i) = 0$

$\Rightarrow \operatorname{cov}(w_i, \tilde{X}_i) =$ ~~...~~

$\operatorname{cov}(w_i ; X_i + w_i) = \sigma_w^2$

$\Rightarrow \operatorname{cov}(v_i | \tilde{X}_i) = \operatorname{cov}(-\beta_1 w_i + u_i ; X_i + w_i) =$

$= -\beta_1 \sigma_w^2$

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum (x_i - \bar{x}) \cdot u_i \to cov(v_i, \tilde{x}) = -\beta_1 \sigma_w^2}{\frac{1}{n}\sum (x_i - \bar{x})^2 \to \sigma_{\tilde{x}}^2 = \sigma_x^2 + \sigma_w^2}$$

$$\Rightarrow \quad \hat{\beta}_1 \xrightarrow{p} \beta_1 \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2} = \beta_1 + \underbrace{\frac{-\sigma_w^2}{\sigma_x^2 + \sigma_w^2}\beta_1}_{\text{смещ. вниз}}$$

## 9.8. Ошибки измерения: следствия

$\hat{\beta}_1$ не смещена, но не эффективна

$\tilde{Y} - Y_i - W_i \qquad E(W_i) = 0 \qquad (W_i, X_i)$ iid.

$$E(W_i | X_i) = 0$$

$\tilde{Y} = \beta_0 + \beta_1 X_i + V_i \quad , V_i = u_i + W_i$

$E(W_i | X_i) = 0 \Rightarrow E(V_i | X_i) = 0 \Rightarrow \hat{\beta}_1$ – несмещ.

Но $var(V_i) = Var(u_i) + Var(W_i) > Var(u_i)$

$\Rightarrow \hat{\beta}_1$ – не эф. оценка

## 9.9. Смещение из-за отбора наблюдений

Причина: когда процесс отбора влияет на доступность данных и этот процесс связан с объясняемой переменной вне зав-ти от регрессоров

$\Rightarrow$ наличие корреляции между $\geq 1$ регрессорами и компонентой ошибки

$\Rightarrow$ (смещенность и несост. МНК-оценок)

## [illegible] одновр. причинности

X влияет на Y (+) Y влияет на X

↳ одновременная причинность

↳ корреляция между регр. и ошибкой

↳ смещенность и несост. оценок

$$\begin{cases} Y_i = \beta_0 + \beta_1 \cdot X_i + u_i \\ X_i = \gamma_0 + \gamma_1 \cdot Y_i + v_i \end{cases}$$

$$Cov(X_i, u_i) = \frac{\gamma_1}{1-\gamma_1 \beta_1} \cdot \sigma_u^2$$

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{\gamma_1}{1-\gamma_1\beta_1} \cdot \frac{\sigma_u^2}{\sigma_X^2}$$

если $\beta_1 < 1/\gamma_1$, то $\hat{\beta}$ – смещена вверх вниз

Решение:

1. регрессия с инстр. переменными
2. стр. управляемый эксперимент
3. сист. одновреи уравнений

## [illegible] Угрозы для внутр. обоснованности

1. пропущенные переменные
2. неправильная спецификация
3. ошибки изм. переменных
4. отбор наблюдений
5. одновр. причинность

$\Rightarrow$ нарушение (I) предп. МНК

9.12. Нарушения предпосылок:

$E(u_i \mid X_i) \neq 0$

=> МНК - оценка смещена; несостоятельна

+ гетероскедастичность,
зависимость перем. между набл.

=> нарушение $(X_i; Y_i) = i.i.d.$

↳ смещение вычисления
станд. ошибок коэф.
и оценок дисперсии

9.13. Внешняя обоснованность

Угрозы:

- различия в ГС
- различия в условиях

=> прогнозирование

=> важна внешн. обосн. (больше, чем несмещ. коэф.)

=> модель должна быть стабильной
и повсеместно применимой

# ⑩ Линейная модель м.р. в матр. форме

## 10.1. Терминология:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix} = \underset{n \times k+1}{\begin{pmatrix} 1 & X_{11} & \dots & X_{k1} \\ \vdots & & & \vdots \\ 1 & & & X_{kn} \end{pmatrix}} \quad U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$$

вектор наблюдений набл. переменной; м-ца набл. регрессоров $(X_i' = (1 X_{1i}; X_{1i} \dots X_{ki}))$; вект. ош.; вект. неизв. к-тов регрессии

$$Y_i = X_i' \cdot \beta + u_i$$

## 10.2. МНК - оценки

$$X'(Y - X\hat{\beta}) = 0$$

$$X'Y = X'X\hat{\beta} \Rightarrow \hat{\beta} = (X'X)^{-1} X'Y$$

## 10.3. Предпосылки МНК

$$Y_i = X_i'\beta + u_i$$

1. $E(u_i | X_i) = 0$
2. $(X_i, Y_i)$ – i.i.d. из совм. распр.
3. $X_i$ и $Y_i$ имеют ненулевые конечные четв. моменты
4. $X$ – полного ранга (отсут. сов. мультиколл-ти)
5. $var(u_i | X_i) = \sigma_u^2$ (гомоскедастичность)
6. $(u_i | X_i) \sim N(\mu, \sigma)$

## 10.4. Следствия выполнения предпосылок МНК

$(1)+(2) \Rightarrow E(u_i|X_i) = E(u_i|X) = 0 \Rightarrow E(U|X) = 0$

$cov(u_i, u_j|X) = E(u_i, u_j|X)$
$= E(u_i u_j | X_i, X_j) = E(u_i|X_i) \cdot E(u_j|X_j) = 0, \quad i \neq j$

$(1)+(2)+(5) \Rightarrow E(u_i^2|X) = E(u_i^2|X_i) = \sigma_u^2 \Rightarrow E(UU'|X) = \sigma_u^2 \cdot I_n$

$(1)+(2)+(5)+(6) \Rightarrow (U|X) \sim N(0_n, \sigma_u^2 I_n)$

## 10.5. Св-ва МНК-оценок (асимпт.)

и.м. ас. норм.

$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0_{k+1}, \Sigma_{\sqrt{n}(\hat{\beta}-\beta)})$

где $\Sigma_{\sqrt{n}(\hat{\beta}-\beta)} = Q_x^{-1} \Sigma_v Q_x^{-1}$

где $Q_x^{-1} = (E(X_i X_i'))^{-1}$

м-ца 2ых моментов

где $\Sigma_v = E(V_i V_i')$

ков. м-ца $V_i = X_i u_i$

$\hat{\beta} \xrightarrow{d} N(\beta, \Sigma_{\hat{\beta}})$

где $\Sigma_{\hat{\beta}} = \frac{\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}}{n} = Q_x^{-1} \Sigma_v Q_x^{-1} / n$

к.м.
прибл. норм.
распр.

## 10.6. Проверка гипотез.

$H_0: \ R\beta = r$

$$\left| \begin{array}{l} R = [\, I_q \ \ 0_{q\times(k+1-q)} \,] \\ r = 0_q \end{array} \right.$$

$$F^{ht} = (R\hat{\beta} - r)' [\, R \hat{\Sigma}_{\hat{\beta}} R' \,]^{-1} \cdot (R\hat{\beta} - r) / q$$

$$F \xrightarrow{d} F_{q,\infty}$$

$$\Rightarrow \{ \delta : (\hat{\beta} - \delta_0)' [\, R \hat{\Sigma}_{\hat{\beta}} R' \,]^{-1} (\hat{\beta} - \delta) / q \le c \}$$

## 10.7 Теорема Гаусса - Маркова

1. $E(U \mid X) = 0_n$
2. $E(UU' \mid X) = \sigma_u^2 I_n$
3. $X$ имеет полный ранг столбцов

$\Rightarrow$ $\hat{\beta}$ - BLUE $\left( var(c'\hat{\beta} \mid X) \le var(c'\tilde{\beta} \mid X) \right)$

# ⑪ Метод инструментальных переменных

## 11.1. Экзогенные переменные – переменные, которые не коррелируют со случ. ошибкой (т.е. опр вне модели)

**Эндогенная переменная** – переменные, кот. кор-ют со случ. ошибкой (опр. внутри модели)

## 11.2 Допустимый инструмент

– такая инстр переменная:

(1) Условие релевантности $cov(Z_i, X_i) \neq 0$

(2) Условие экзогенности $cov(Z_i, u_i) = 0$

## 11.3. Обобщ. регрессия с инстр. переменными

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \ldots + \beta_{k+r} W_{ri} + u_i, \; i = \overline{1,n}$$

$X_{ji}$ – k эндогенных регрессоров

$W_{ji}$ – r экзогенных регрессоров (не кор. с $u_i$ или мы. контр. пер.)

Предположения:

(1). $E(u_i \mid W_i') = 0$

(2) $(X_i', W_i', Z_i', Y_i)$ i.i.d.

(3) $X, Y, Z, W$ имеют конеч. пол. 4 мом.

(4) $\hat{X}_i', W_i'$ и $\beta_0$ не должны быть сов. мультикол. (Relevance)

(5) $cov(u_i, Z_{ji}) = 0$ (Exogen.)

$\Rightarrow \hat{\beta}_i$ сост. и ас. норм распр.

## 11.4. Двухшаговый МНК пар. лин. м. р.

$Y_i = \beta_0 + \beta_1 X_i + u_i$

$\hat{\beta}_1$ – смещ. и несост.

=> ввод. инструмент $Z$ "изолирующий" ту часть $X_i$, кот. кор-на с $u_i$

Шаг 1 $X_i = \pi_0 + \pi_1 Z_i + v_i$

$\hat{X} = \hat{\pi}_0 + \hat{\pi}_1 Z_i$

$corr(u_i, Z_i) = 0 \Rightarrow corr(u_i, \hat{X}_i) = 0$

Шаг 2 $Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$

$\Rightarrow \hat{\beta}_0^{TSLS}$ и $\hat{\beta}_1^{TSLS}$

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)} = \beta_1$$

$\hat{\beta}_1^{TSLS} \xrightarrow{d} N(\beta_1, \sigma^2_{\hat{\beta}_1^{TSLS}}) \Rightarrow$ t-стат. ...

$$\sigma^2_{\hat{\beta}_1^{TSLS}} = \frac{1}{n} \frac{Var[(Z_i - \mu_Z) u_i]}{[cov(Z_i, X_i)]^2}$$

## 11.5. TSLS мн. л. р.

$m \overset{>}{\underset{<}{=}} k$ коэф. $\begin{matrix}\text{пере}\\ \text{точно}\\ \text{недо}\end{matrix}$ опр. $\begin{matrix}+\\ +\\ \end{matrix}$

Шаг 1: "очистим" эндогенных переменных

оценить $X_{ji}$ на $\beta_0, Z_{ji}, W_{ji}$, $\forall j = \overline{1,k}$

получаем $\hat{X}_{ji}$, $\forall j = \overline{1,k}$
"очищенные"

Шаг 2 по МНК оцениваем

$Y_i = \beta_0 + \beta_1 \hat{X}_{1i} + \ldots + \beta_k \hat{X}_{ki} + \beta_{k+1} W_{1i} + \ldots \beta_{k+r} W_{ri} + u_i$

$\Rightarrow \hat{\beta}_j^{TSLS}$

11.6. Слабый инстр. — инстр., кот. объясняет малую часть изм. X

$\Rightarrow$ TSLS оценки смещ. и не ас. норм.

↳ F – тест $H_0: \pi_{1i} = \ldots = \pi_{m} = 0$

$F < 10 \Rightarrow$ набор инстр. слабый

Сверхидентифицирующие ограничения

$\hat{u}_i^{TSLS} = \delta_0 + \delta_1 Z_{1i} + \ldots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \ldots + \delta_{m+r} W_{ri} + e_i$

F – тест $H_0: \delta_1 = \ldots = \delta_m = 0 \Rightarrow F$

$J = mF \sim \chi^2_{m-k}$

– стат. для J тест. сверхиден. ограничений

m – k – "ст. сверх."