Ошибки спецификации и выбор функциональной формы модели

План лекции

- Пропуск существенных переменных
  - Свойства оценок МНК
  - Примеры
  - Тест Рамсея на пропущенные переменные
- Включение лишних переменных
  - Свойства оценок МНК
  - Примеры
  - F-тест на лишние переменные
- Не вложенные регрессии
- Выбор функциональной формы
  - Преобразование Бокса-Кокса
  - Наиболее популярные в экономических приложениях формы моделей
  - Тестирование нормальности регрессионных остатков
  - Пример

Анализируемые модели

- Модель 1 (длинная регрессия)

$$Y = X\beta + Z\gamma + \varepsilon$$

- Модель 2 (короткая регрессия)

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

К чему ведет пропуск существенных переменных?

- Если пропущены существенные переменные, то
- (1) – $Y = X\beta + Z\gamma + \varepsilon$

  истинная модель
- (2) – $Y = X\beta + \varepsilon$

  оцениваемая модель,

и свойства ее оценок мы будем изучать

Свойства оценок

- Исследование свойств оценок коэффициентов

$$\widehat{\beta}^{(2)} = (X'X)^{-1}X'Y \overset{Y=X\beta+Z\gamma+\varepsilon}{=} \beta + (X'X)^{-1}X'Z\gamma + (X'X)^{-1}X'\varepsilon$$

Оценка оказывается смещенной:

$$E\left(\widehat{\beta}^{(2)}\right) = \beta + (X'X)^{-1}X'Z\gamma + (X'X)^{-1}X'E\left(\varepsilon\right) = \beta + \underbrace{(X'X)^{-1}X'Z\gamma}_{смещение} \neq \beta$$

Это смещение не исчезает при увеличении объема выборки, т.о. оценка становится несостоятельной

Интерпретация смещения: произведение коэффициентов регрессий Z на X и истинного значения γ.

Всегда ли будет смещение?

Если X и Z ортогональны, то есть $X'Z = 0$,

в этом случае смещения не будет, но чем значительнее корреляция между X и Z, тем серьезнее смещение

Свойства оценок

- Свойства ковариационной матрицы:

$$V\left(\widehat{\beta}^{(2)}\right)=\sigma_{\varepsilon}^{2}(X'X)^{-1}\neq V\left(\widehat{\beta}^{(1)}\right)=\sigma_{\varepsilon}^{2}\left[(X'X)^{-1}+LM_{Z}L'\right]=$$

$$=\sigma_{\varepsilon}^{2}\left[(X'X)^{-1}+\left(X'X\right)^{-1}X'Z\left(I-Z\left(Z'Z\right)^{-1}Z'\right)^{-1}Z'X\left(X'X\right)^{-1}\right]$$

Ковариационная матрица вычисляется неверно:

ее диагональные элементы меньше, чем должны быть теоретически

Всегда ли будет смещение?

Если $X'Z=0$, в этом случае смещения не будет, но чем значительнее корреляция между X и Z, тем серьезнее смещение

Свойства оценок

- Оценка дисперсии регрессии

$$\left(\hat{\sigma}_{\varepsilon}^{2}\right)^{(2)}=\frac{RSS^{(2)}}{n-k-1}=\frac{\hat{\varepsilon}_{(2)}'\hat{\varepsilon}_{(2)}}{n-k-1}=\frac{\left(Y-X\hat{\beta}^{(2)}\right)'\left(Y-X\hat{\beta}^{(2)}\right)}{n-k-1}=\frac{Y'M_xY}{n-k-1}$$

- Оценка оказывается смещенной вверх:

$$E\left\{\left(\hat{\sigma}_{\varepsilon}^{2}\right)^{(2)}\right\}=\frac{E\left\{\left(X\beta+Z\gamma+\varepsilon\right)'M_x\left(X\beta+Z\gamma+\varepsilon\right)\right\}}{n-k-1}\overset{M_xX\beta=0}{=}$$

$$=\frac{E\left\{\left(Z\gamma+\varepsilon\right)'M_x\left(Z\gamma+\varepsilon\right)\right\}}{n-k-1}\overset{E(M_x\varepsilon)=0}{=}\frac{E\left(\varepsilon'M_x\varepsilon\right)+\gamma'Z'M_xZ\gamma}{n-k-1}=$$

$$=\frac{\sigma_{\varepsilon}^{2}trM_x+\gamma'Z'M_xZ\gamma}{n-k-1}=\sigma_{\varepsilon}^{2}+\frac{\gamma'Z'M_xZ\gamma}{n-k-1}\neq\sigma_{\varepsilon}^{2}$$

Общие выводы

При пропуске существенных переменных

- оценки $\widehat{\beta}^{(2)} = (X'X)^{-1}X'Y$
  - смещены (кроме случая $X'Z = 0$)
  - несостоятельны (кроме случая $X'Z = 0$)
- ковариационная матрица оценок коэффициентов $V\left(\widehat{\beta}^{(2)}\right) = \sigma_{\varepsilon}^{2}(X'X)^{-1}$
  - вычисляется неверно (кроме случая $X'Z = 0$)
- оценка дисперсии регрессии $\hat{\sigma}_{\varepsilon}^{2}$
  - смещена

Пример 1

- Какие переменные часто пропускаются?
  - Сезонность, степени, перекрестные произведения, тренд, лаги.
- Рассмотрим произвольную модель распределенных лагов, записанную в отклонениях
- Истинное уравнение модели: $y_t = \beta x_t + \gamma x_{t-1} + \varepsilon_t$
- Оцениваемое уравнение: $y_t = \beta x_t + u_t$

$$E\left(\hat{\beta}\right) = \beta + \underbrace{\frac{\sum x_t x_{t-1}}{\sum x_t^2}}_{смещение}\gamma = \beta + \rho\gamma \qquad \rho = corr\left(x_t, x_{t-1}\right)$$

- Очевидно, если ρ=0, смещения не будет

Пример 2: гедонистическая регрессия цен на коттеджи

Гедонистическая цена приписывается неявной цене определенного характерного признака оцениваемого объекта и влияет на его продажную цену (для коттеджа, например, таким характерным признаком может быть число спален).

Типичными товарами, для которых оцениваются гедонистические ценовые функции, являются компьютеры, автомобили и недвижимость.

Гедонистическая ценовая функция описывает саму ожидаемую цену дома (или ее логарифма) как функцию множества характеристик:

- размера участка земли в собственности,
- числа спален,
- числа полностью оборудованных ванных комнат,
- числа мест в гараже,
- числа этажей,
- наличия подъездной дороги,
- наличия комнаты отдыха,
- наличия обустроенного подвального помещения,
- наличия центрального кондиционирования воздуха,
- наличия водяного отопления на газе и
- расположения в привилегированном районе.

Пример 2

Регрессионная зависимость цен на коттеджи от размера участка, квадрата размера и участка и дамми для расположения в привилегированной зоне

Variable	m1	m2
lotsize	13.692***	5.977***
lotsize2	-0.001***	
prefarea	12845.248***	13547.941***
_cons	12820.184**	34163.993***
r2	0.362	0.331
r2_a	0.359	0.328
rss	2.48e+11	2.60e+11
rmse	21384.452	21884.052

legend: * p<0.05; ** p<0.01; *** p<0.001

Вопрос: почему при пропуске существенной переменной `lotsize2` оценка при `lotsize` упала, а оценка при `prefarea` выросла?

Пример 2

Корреляционная матрица регрессоров

	lotsize	lotsize2	prefarea
lotsize	1.0000		
lotsize2	**0.9618***	1.0000	
prefarea	0.2348*	**0.2096***	1.0000

Пример 2

Сопоставление оценок стандартных ошибок регрессий

```
--------------------------------------
    Variable |         m1           m2
-------------+------------------------
     lotsize |     13.6923       5.9767
             |    (1.5560)     (0.4448)
    lotsize2 |     -0.0006
             |    (0.0001)
    prefarea |    1.28e+04     1.35e+04
             |  (2226.5177)  (2274.2763)
       _cons |    1.28e+04     3.42e+04
             |  (4759.6651)  (2415.6753)
--------------------------------------
                          legend: b/se
```

Тест Рамсея на наличие пропущенной переменной

Суть теста состоит в исследовании объясняющей силы вспомогательной регрессии остатков исходной модели от полинома 3-ей степени предсказанных исходной моделью значений зависимой переменной

Ramsey RESET test using powers of the fitted values of price

Ho: model has no omitted variables

Результат для модели m1:

F = 0.16

Prob > F = 0.9244

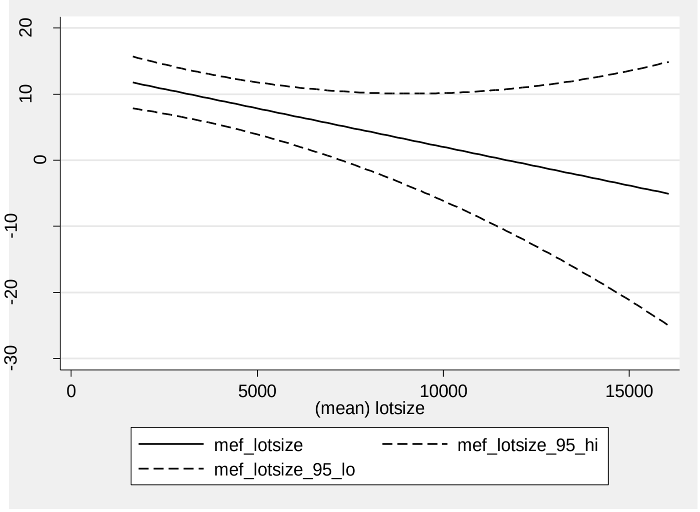
Результат для модели m2:

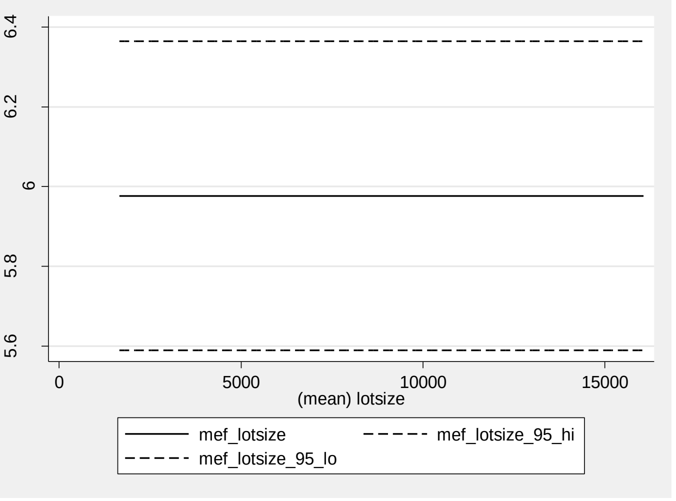
F = 8.98

Prob > F = 0.0000

Доверительный интервал для предельного эффекта размера участка на прогноз цены дома

$$\frac{\partial price}{\partial lotsize} = \hat{\beta}_1 + \hat{\gamma}_1 lotsize = 13.6923 - 0.0006\, lotsize \qquad \frac{\partial price}{\partial lotsize} = \hat{\beta}_2 = 5.9767$$





Интерпретация результатов

- Очевидно, что в модели 2 предполагается равномерный рост цены на 5.98 канадских долларов при увеличении размера участка на 1 квадратный фут.
- В то время как в более реалистичной и правильной с точки зрения проделанных тестов модели 1 цена дома растет, но с ростом размера участка рост цены замедляется.
- Например, для участка размером в 5000 квадратных футов цена вырастет, согласно модели, на 13.69-2*.0006*5000 = 7.69 долларов при увеличении размера участка на 1кв.фут, а для участка размером в 10000 квадратных футов цена изменится при увеличении размера на 1 кв.фут только на 13.69-2*.0006*10000=1.69 долларов.

К чему ведет включение лишних переменных?

- Если включены лишние переменные, то
- (2) – $Y = X\beta + \varepsilon$

  истинная модель
- (1) – $Y = X\beta + Z\gamma + \varepsilon$

  оцениваемая модель,

и свойства ее оценок мы будем изучать

Свойства оценок

- Исследование свойств оценок коэффициентов

$$\hat{\beta}^{(1)} = (X'X)^{-1}X'Y - LM_Z^{-1}Z'M_X Y \overset{Y=X\beta+\varepsilon}{=} \beta + (X'X)^{-1}X'\varepsilon - LM_Z^{-1}Z'M_X\varepsilon$$

$$\hat{\gamma}^{(1)} = M_Z^{-1}Z'M_X Y \overset{Y=X\beta+\varepsilon}{=} M_Z^{-1}Z'M_X\varepsilon$$

Оценка оказывается несмещенной:

$$E\left(\hat{\beta}^{(1)}\right) = \beta + (X'X)^{-1}X'E\left(\varepsilon\right) - LM_Z^{-1}Z'M_X E\left(\varepsilon\right) = \beta$$

$$\hat{\gamma}^{(1)} = M_Z^{-1}Z'M_X E\left(\varepsilon\right) = 0$$

Свойства оценок

- Свойства ковариационной матрицы:

$$V\left(\hat{\beta}^{(1)}\right) = \sigma_\varepsilon^2 \left[(X'X)^{-1} + LM_Z L' \right] =$$

$$= \sigma_\varepsilon^2 \left[(X'X)^{-1} + (X'X)^{-1} X'Z \left(I - Z(Z'Z)^{-1} Z' \right)^{-1} Z'X (X'X)^{-1} \right] \neq$$

$$\neq V\left(\hat{\beta}^{(2)}\right) = \sigma_\varepsilon^2 (X'X)^{-1}$$

Ковариационная матрица вычисляется неверно:
ее диагональные элементы больше, чем должны быть теоретически

Всегда ли будет смещение?
Если $X'Z = 0$, в этом случае смещения не будет, но чем значительнее корреляция между X и Z, тем серьезнее смещение

Свойства оценок

- Оценка дисперсии регрессии

$$\left(\hat{\sigma}_\varepsilon^2\right)^{(1)} = \frac{RSS^{(1)}}{n-k-1-r} = \frac{\hat{\varepsilon}'_{(1)}\hat{\varepsilon}_{(1)}}{n-k-r-1} = \frac{\left(Y - X\hat{\beta}^{(1)}\right)'\left(Y - X\hat{\beta}^{(1)}\right)}{n-k-r-1} =$$

$$= \frac{\varepsilon' M_X \left(I - Z'M_Z^{-1}ZP_X\right)\left(I - P_X ZM_Z^{-1}Z'\right)M_X\varepsilon}{n-k-r-1}$$

- Оценка оказывается несмещенной:

$$E\left\{\left(\hat{\sigma}_\varepsilon^2\right)^{(1)}\right\} = \frac{\sigma_\varepsilon^2 tr\left[\left(I - Z'M_Z^{-1}ZP_X\right)\left(I - P_X ZM_Z^{-1}Z'\right)M_X\right]}{n-k-r-1} \overset{?}{=} \sigma_\varepsilon^2$$

Общие выводы

При включении лишних переменных

- оценки
  - несмещены
  - состоятельны

$$\widehat{\beta}^{(1)} = (X'X)^{-1}X'Y - LM_Z^{-1}Z'M_XY$$

$$\hat{\gamma}^{(1)} = M_Z^{-1}Z'M_XY$$

- ковариационная матрица оценок коэффициентов
  - дает завышенные значения дисперсий (кроме случая $X'Z = 0$) (неэффективность)
- оценка дисперсии регрессии $\hat{\sigma}_{\varepsilon}^2$
  - несмещена

Пример 2

Регрессионная зависимость цен на коттеджи от размера участка, квадрата размера и участка и дамми для расположения в привилегированной зоне

Variable	m1	m3
lotsize	13.692***	13.617***
lotsize2	-0.001***	-0.001***
prefarea	12845.248***	11525.605
lotpref		0.231
_cons	12820.184**	13171.228**
r2	0.362	0.362
r2_a	0.359	0.358
rss	2.48e+11	2.48e+11
rmse	21384.452	21403.043

Lotpref= lotsize * prefarea

Пример 2

Сопоставление оценок стандартных ошибок регрессий

Variable	m1	m3
lotsize	13.6923	13.6174
	(1.5560)	(1.5877)
lotsize2	-0.0006	-0.0006
	(0.0001)	(0.0001)
prefarea	1.28e+04	1.15e+04
	(2226.5177)	**(5880.3421)**
lotpref		0.2311
		(0.9528)
_cons	1.28e+04	1.32e+04
	(4759.6651)	(4978.8855)

legend: b/se

Пример 2

Серьезные потери в эффективности понес регрессор prefarea, очень сильно коррелированный с дополнительной переменной lotpref, почти не пострадали регрессоры слабо коррелированные с lotpref

Анализ корреляционной матрицы

	lotsize	lotsize2	prefarea	lotpref
lotsize	1.0000			
lotsize2	0.9618*	1.0000		
prefarea	0.2348*	0.2096*	1.0000	
lotpref	0.4299*	0.3978*	**0.9134***	1.0000

Тест на лишние переменные

В модели 3 имеет смысл провести тест на мультиколлинеарность, посмотрев VIFы

```
 Variable |       VIF       1/VIF
-------------+----------------------
     lotsize |     14.10    0.070934
    lotsize2 |     13.41    0.074573
     lotpref |      8.58    0.116611
    prefarea |      7.40    0.135192
```

Можно попробовать провести F-тест на возможность исключения незначимых регрессоров из модели

```
 ( 1)  prefarea = 0
 ( 2)  lotpref = 0
       F(  2,   541) =   16.64
            Prob > F =    0.0000
```

Тест Рамсея

Оказывается, что гипотеза о совместной незначимости двух регрессоров отвергается, следовательно, в них заложена ценная информация, влияние которой на цену мешает выявить квазимультиколлинеарность.

О том, что эта группа важна, говорит и тест Рамсея, проведенный для результатов регрессии без этих переменных и диагностирующий ошибку спецификации.

```
qui reg price lotsize lotsize2
ovtest
```

Результат:

```
Ramsey RESET test using powers of the fitted values
of price
       Ho:  model has no omitted variables
                  F(3, 540) =      6.93
                   Prob > F =      0.0001
```

Использование теоремы о корне из r

price	Coef.	Std. Err.	t	P>|t|
lotsize	13.61742	1.587664	8.58	0.000
lotsize2	-.0005834	.0001132	-5.15	0.000
prefarea	11525.6	5880.342	1.96	0.051
lotpref	.2310658	.9528311	**0.24**	0.808
_cons	13171.23	4978.886	2.65	0.008

Если мы намереваемся исключить два регрессора, t-статистики при каждом из них должны быть меньше =1.414, но такой пары регрессоров, судя по результатам, у нас не наблюдается. Тогда нам остается удалить один, t-статистика при котором меньше единицы, а это и есть лишняя переменная `lotpref`.

Тестирование вложенности регрессий

Если увидеть влияние важных для нас переменных мешает квазимультиколлинеарность, имеет смысл строить не вложенные регрессии, которые будут дополнять друг друга.

Пусть имеется 2 альтернативные модели с пересекающимися наборами регрессоров:

(А) $Y = X\beta + \varepsilon$

(Б) $Y = Z\gamma + \varepsilon$

$$X \cap Z = Q$$

Существуют специальные методы диагностики, которые позволяют выявить скрытую вложенность моделей и различить три ситуации:

(А) вложена в (Б)

(Б) вложена в (А)

(А) и (Б) не вложенные

Тест Дэвидсона и Мак-Киннона

- Оценивается модель А
- Вычисляется прогнозное значение зависимой переменной по модели А
- Оценивается модель Б, куда включается вспомогательный регрессор – значение зависимой переменной, предсказанной по модели А, и откуда исключаются общие для двух моделей регрессоры
- Изучается значимость вспомогательного регрессора, если он незначим, вся информация о нем уже была вложена неявно в модель Б, следовательно, А вложена в Б.
- И таким же образом тест воспроизводится после того, как модели меняются ролями для большей достоверности.

Результат:

Variable	test_1_4	test_4_1
lotpref	.04948441	
price1_hat	.99414433***	
prefarea		2102.5158
price4_hat		.97159665***
_cons	328.48914	1441.9841

legend: * p<0.05; ** p<0.01; *** p<0.001

Все коэффициенты в тестовых регрессиях, кроме коэффициентов при вспомогательных регрессоров, незначимы.

Таким образом, и в этой модификации теста мы наблюдаем практически полную вложенность моделей друг в друга.

Вывод

- Модель 1 все же чуть лучше, поскольку для нее не показывает ошибки спецификации тест Рамсея.
- Таким образом, в нашем примере наилучшей оказывается модель1, в которой предполагается квадратичная зависимость цены от размере участка и различие в уровне цен для объектов, расположенных внутри и вне привилегированной зоны.

Выбор функциональной формы

- Выбор формы и экономическая теория
- Сравнение функциональных форм
  - Преобразование Бокса-Кокса
  - Популярные функциональные формы
- Тест Харке-Бера на нормальность остатков

Выбор формы и экономическая теория

- Производственные функции
  - ПФ с совершенными заменителями

$$Y_t = \alpha + \beta L_t + \gamma K_t + \varepsilon_t$$

  - ПФ Кобба-Дугласа

$$\ln Y_t = \alpha + \beta \ln L_t + \gamma \ln K_t + \varepsilon_t$$

  - ПФ CES

$$\ln Y_t = \ln \gamma - \frac{\nu}{\rho} \ln \left[\delta K_t^{-\rho} + (1-\delta) L_t^{-\rho} \right] + \varepsilon_t$$

- Функция спроса

$$\ln Q_i = \alpha + \beta \ln P_i + \gamma \ln Y_i + \varepsilon_i$$

Сравнение функциональных форм

- Сравнение по коэффициенту детерминации больше невозможно, поскольку зависимые переменные становятся несопоставимы и TSS несопоставимы
- Используются специальные тесты, в которых производится преобразование зависимых переменных к сопоставимому виду

  (тесты Бокса-Кокса, PE-тест Дэвидсона и Мак Киннона)
- Тестирование нормальности остатков – еще одна возможность убедиться в адекватности функциональной формы регрессии

Преобразование Бокса-Кокса

- Это преобразование при различных значениях параметров охватывает практически все используемые в экономических приложениях функциональные формы
- Вместо исходных переменных вводятся их аналоги:

$$Y^{(\lambda_1)} = \begin{cases} \dfrac{Y^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0 \\ \ln Y, & \lambda_1 = 0 \end{cases} \qquad X^{(\lambda_2)} = \begin{cases} \dfrac{X^{\lambda_2} - 1}{\lambda_2}, & \lambda_2 \neq 0 \\ \ln X, & \lambda_2 = 0 \end{cases}$$

- Существует непрерывный предельный переход

$$\lim_{\lambda_1 \to 0} Y^{(\lambda_1)} = \lim_{\lambda_1 \to 0} \frac{Y^{\lambda_1} - 1}{\lambda_1} \overset{\textit{по правилу Лопиталя}}{=} \lim_{\lambda_1 \to 0} \frac{\frac{d}{d\lambda_1}\left(Y^{\lambda_1} - 1\right)}{\frac{d}{d\lambda_1}\left(\lambda_1\right)} = \lim_{\lambda_1 \to 0} Y^{\lambda_1} \ln Y = \ln Y$$

Наиболее популярные в экономических приложениях функциональные формы

- Линейная модель
- Логарифмически-линейная модель
- Полулогарифмическая модель
- Обратная зависимость
- Логарифмическая обратная зависимость

Линейная модель

$$\lambda_1 = \lambda_2 = 1$$

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

- Смысл коэффициентов – предельные эффекты X на Y

$$\frac{\partial Y}{\partial X} = \beta$$

- Интерпретация коэффициентов – при изменении X на единицу измерения Y меняется на β

Логарифмически-линейная модель

$$\lambda_1 = \lambda_2 = 0$$

$$\ln Y = \alpha + \beta \ln X + \varepsilon$$

- Смысл коэффициентов – эластичности

$$\frac{\partial \ln Y}{\partial \ln X} = \frac{\partial Y / Y}{\partial X / X} = \beta$$

- Интерпретация коэффициентов – при изменении X на 1% измерения Y меняется на β %-ов

Полулогарифмическая модель

$$\lambda_1 = 0 \quad \lambda_2 = 1$$

$$\ln Y = \alpha + \beta X + \varepsilon$$

- Смысл коэффициентов – темп роста

$$\frac{\partial \ln Y}{\partial X} = \frac{\partial Y / Y}{\partial X} = \beta$$

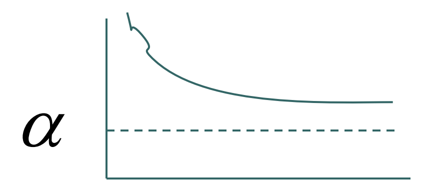
- Интерпретация коэффициентов – при изменении X на единицу измерения Y меняется на β *100%-ов

Обратная зависимость

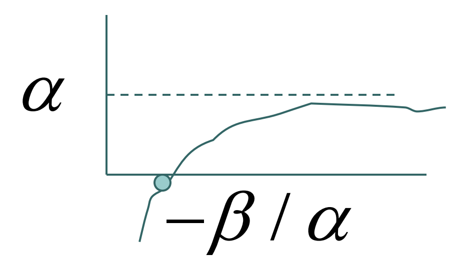
$$\lambda_1 = 1 \quad \lambda_2 = -1$$

$$Y = \alpha + \beta\left(1/X\right) + \varepsilon$$

- Прозрачной интерпретации коэффициентов больше нет
- Примеры:
  - Кривая Филлипса $\beta > 0$



  - Зависимость расходов на некое благо от общих расходов $\beta < 0$



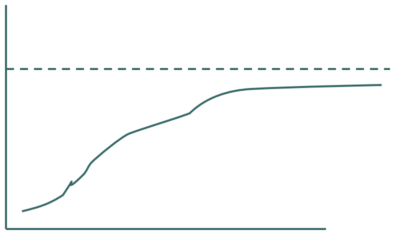
Логарифмическая обратная зависимость

$$\lambda_1 = 0 \quad \lambda_2 = -1$$

$$\ln Y = \alpha + \beta\left(1/X\right) + \varepsilon$$

- Прозрачной интерпретации коэффициентов больше нет
- Примеры:
  - Кривая роста с выходом на насыщение
  - Заменяет логистическую функцию $Y = \dfrac{\alpha}{1+\beta\exp\left(-\gamma X\right)}$

$$\beta < 0$$



Тест Харке-Бера на нормальность остатков

- Тест связан с вычислением выборочных асимметрии S и эксцесса (K - 3) для остатков регрессии и выяснении близости последних к соответствующим моментам нормального распределения.
- Осуществляется на основании статистики

  Харке-Бера (Jarque-Bera, J.-B.)
- Проверяется близость выборочной асимметрии и выборочного эксцесса к 0 (выборочный куртозис K должен быть близок к 3)
- Может косвенно свидетельствовать об ошибках выбора функциональной формы уравнения регрессии

Тест Харке-Бера на нормальность остатков

$$H_0: \quad \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$$

$$J.-B. = \frac{n-k-1}{6}\left[S^2 + \frac{1}{4}(K-3)^2\right] \overset{H_0}{\sim} \chi_2^2$$

$$S = \frac{\sum_{i=1}^{n} (\hat{\varepsilon}_i)^3 / n}{\hat{\sigma}_\varepsilon^{\;3}}$$

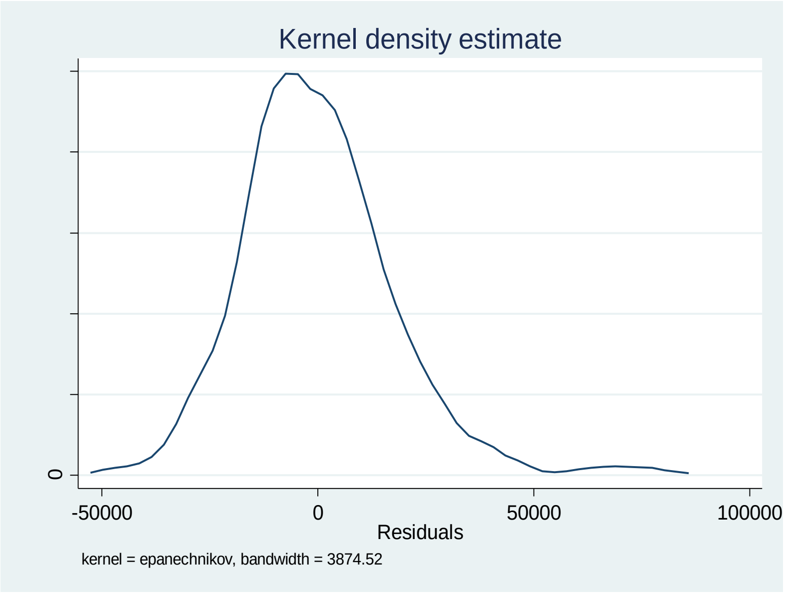
- выборочная асимметрия (должна быть близка к 0)

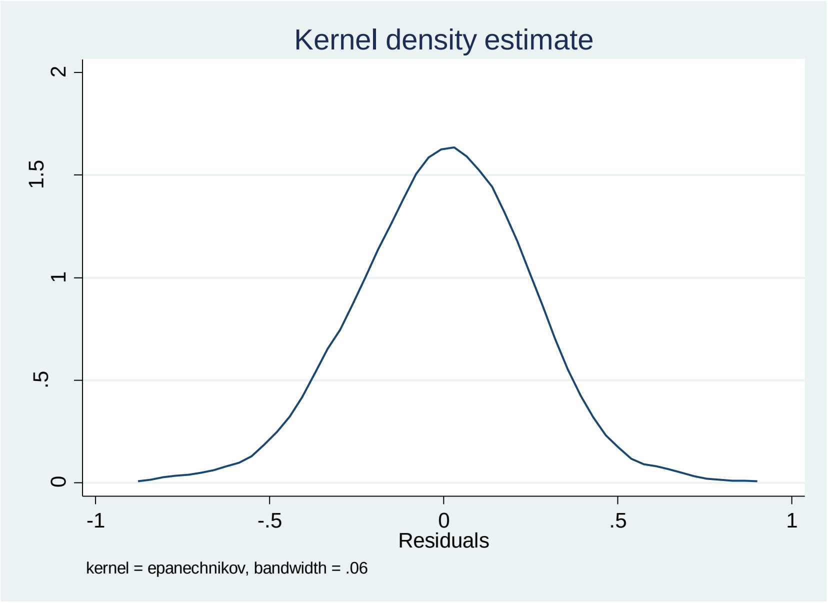
$$K = \frac{\sum_{i=1}^{n} (\hat{\varepsilon}_i)^4 / n}{\hat{\sigma}_\varepsilon^{\;4}}$$

- выборочный куртозис (должен быть близок к 3)

Пример: цены на коттеджи

- Полезно сравнить сначала визуально графики функций плотностей для остатков
- Остатки линейной модели   Остатки логарифмической модели





Пример: цены на коттеджи

- Skewness/Kurtosis tests for Normality

```
                                                  ------- joint ------
Variable | Pr(Skewness)  Pr(Kurtosis)  adj chi2(2)     Prob>chi2
  ------+----------------------------------------------------------
       e |       0.000        0.000           69.58         0.0000
      el |       0.383        0.151            2.82         0.2436
```

- Результаты теста указывают p-значения для асимметрии Pr(Skewness) и куртозиса Pr(Kurtosis), а также статистику Харке-Бера adj chi2(2) и ее p-значение Prob>chi2
- В верхней строке указаны результаты теста для остатков линейной модели, а в нижней – для остатков модели логарифмической.
- Вывод: для линейной модели гипотеза нормальности ошибок отвергается, **для логарифмической – нет оснований отвергнуть гипотезу нормальности.**

Спасибо за внимание!