

Эконометрика, 2020-2021, 2 модуль
Семинар 4
16.11.20

для
Группы Э_Б2018_Э_3
Семинарист О.А.Демидова

Проверка нормальности распределения остатков регрессии.

Робастные методы оценивания

По данным файла *clothing.dta*, содержащем данные о продажах одежды в 400 голландских магазинах мужской одежды

1) Оцените коэффициенты уравнения регрессии

$$sales = \beta_0 + \beta_1 hoursw + \beta_2 size + u.$$

2) Постройте гистограмму распределения остатков регрессии и *qqplot*.

С помощью тестов Колмогорова-Смирнова, Харке-Бера, Шапиро-Уилка проверьте гипотезу о нормальности распределения случайных ошибок.

3) Рассчитайте Стьюдентизированные остатки регрессии, постройте их график и с его помощью определите возможные наблюдения, которые являются выбросами. Если такие наблюдения будут выявлены, то оцените регрессию

а) без наблюдений, являющихся выбросами,

б) R-регрессию Хубера.

4) Сравните результаты оценки регрессий, оцененных с помощью МНК (с выбросами и без), медианной регрессии, R-регрессии Хуберта.

Методические рекомендации

1) Открыв файл *clothing.dta* в статистическом пакете STATA, оцените необходимую регрессию с помощью команды:

```
reg sales hoursw ssize
```

2) Сохраните остатки регрессии с помощью команды

```
predict res, resid
```

3) Постройте их гистограмму с помощью команды *hist res*

Qq plot с помощью команды

```
qnorm res
```

4) Проведите тест Колмогорова-Смирнова с помощью команды

```
sum res
```

```
ksmirnov res = normal((res-r(mean))/r(sd))
```

5) Проведите тест Харке-Бера с помощью команды *sktest res, noadjust*

6) Проведите тест Шапиро-Уилка с помощью команды *swilk res*

7) Постройте график стьюдентизированных остатков регрессии с помощью команды

```
predict residst, rstudent
```

```
twoway(scatter residst number, sort)
```

```
list number if abs(residst) > 2
```

Оценить регрессию без наблюдений-выбросов можно с помощью команды

```
reg sales hoursw ssize if abs(residst) < 2
```

8) Оценить медианную регрессию можно с помощью команды

```
qreg sales hoursw ssize
```

9) Оценить R-регрессию Хубера можно с помощью команды

```
rreg sales hoursw ssize
```

10) Сохранить результаты оценки регрессии можно с помощью команды

```
est store reg1 (и т.д.)
```

Для сравнения результатов удобно сформировать общую таблицу с помощью команды

```
est tab reg1 (и т.д.), star (0.1 0.05 0.01) b(%7.3f)
```

Выбор функциональной формы модели

I. По данным файла `dougherty.dta`, используя тест Бокса-Кокса, с помощью статистического пакета STATA, оцените параметры модели

$$EARNINGS^{(\theta)} = \beta_0 + \beta_1 S^{(\lambda)} + \beta_2 ASVABC^{(\lambda)} + \varepsilon$$

1) Когда переменные левой и правой части преобразуются не одинаково:

Это можно сделать с помощью команды:

```
boxcox EARNINGS S ASVABC, model (theta)
```

2) Когда переменные в обеих частях модели преобразуются одинаково.

Это можно сделать с помощью команды:

```
boxcox EARNINGS S ASVABC, model (lambda),
```

3) Когда преобразуется только зависимая переменная

Это можно сделать с помощью команды

```
boxcox EARNINGS S ASVABC, model (lhsonly),
```

4) Когда преобразуются только независимые переменные

Это можно сделать с помощью команды

```
boxcox EARNINGS S ASVABC, model (rhsonly),
```

5) поэкспериментируйте с проведением теста Бокса-Кокса в модели с другим набором переменных (если Вы не хотите преобразовывать какие-то переменные, например, `dummy`, то это можно сделать с помощью команды, аналогичной следующей:

```
boxcox EARNINGS S ASVABC, notrans(MALE) model (theta)
```

```
boxcox EARNINGS S ASVABC, notrans(MALE) model (lambda)
```

II. По данным файла `dougherty.dta` выберите между линейной, полулогарифмической и линейной в логарифмах моделями с помощью теста

А) РЕ теста Дэвидсона, Уайта и МакКиннона

б) Бера и МакАлера.

а) РЕ теста Дэвидсона, Уайта и МакКиннона

Необходимые команды

Выбор между линейной и линейной в логарифмах моделью

Сначала оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC,
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Предварительно создав необходимые дополнительные переменные,

аналогично оценим линейную в логарифмах модель с помощью команды,

```
gen lnEARNINGS = ln(EARNINGS)
```

```
gen lnS = ln(S)
```

```
gen lnASVABC = ln(ASVABC)
```

```
reg lnEARNINGS lnS lnASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict ln_y_hat.
```

Теперь переходим к шагу 2, оценим дополнительные модели.

Сначала создадим дополнительную разность для линейной модели:

```
gen lin_add= ln(y_hat)- ln_y_hat
```

и оценим эту модель:

```
reg EARNINGS S ASVABC lin_add
```

Создадим также дополнительную разность для линейной в логарифмах модели:

```
gen log_add=y_hat-exp(ln_y_hat)
```

оценим эту модель:

```
reg lnEARNINGS lnS lnASVABC log_add
```

Если

- 1) оба коэффициента при дополнительных разностях значимы или оба незначимы, то выбрать посредством теста Дэвидсона, Уайта и МакКиннона невозможно,
- 2) если незначим только коэффициент при дополнительной разности в линейной модели, то лучше линейная модель,
- 3) если незначим только коэффициент при дополнительной разности в линейной в логарифмах модели, то лучше линейная в логарифмах модель.

Выбор между линейной и полулогарифмической моделями

Сначала оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Аналогично оценим полулогарифмическую модель с помощью команды:

```
reg lnEARNINGS S ASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict semiln_y_hat.
```

Теперь переходим к шагу 2, оценим дополнительные модели.

Сначала создадим дополнительную разность для линейной модели:

```
gen lin_adds= ln(y_hat)- semiln_y_hat
```

и оценим эту модель:

```
reg EARNINGS S ASVABC lin_adds
```

Создадим также дополнительную разность для полулогарифмической модели:

```
gen semilog_add=y_hat-exp(ln_y_hat)
```

оценим эту модель:

```
reg lnEARNINGS S ASVABC semilog_add
```

Если

- 1) оба коэффициента при дополнительных разностях значимы или оба незначимы, то выбрать посредством теста Дэвидсона, Уайта и МакКиннона невозможно,
- 2) если незначим только коэффициент при дополнительной разности в линейной модели, то лучше линейная модель,
- 3) если незначим только коэффициент при дополнительной разности в полулогарифмической модели, то лучше полулогарифмическая модель.

б) Тест Бера и МакАлера

Необходимые команды

Выбор между линейной и полулогарифмической моделями

Сначала оценим полулогарифмическую модель с помощью команды:

```
reg lnEARNINGS S ASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict semiln_y_hat.
```

Аналогично оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Теперь переходим к шагу 2, оценим дополнительные модели.

```
1) gen y1= exp(semiln_y_hat)
```

```
reg y1 S ASVABC
```

Сохранить остатки регрессии можно с помощью команды

```
predict res1, resid
```

```
2) gen y2= ln(y_hat)
```

```
reg y2 S ASVABC
```

Сохранить остатки регрессии можно с помощью команды

```
predict res2, resid
```

Теперь переходим к шагу 3, оценив еще 2 дополнительные модели.

```
reg lnEARNINGS S ASVABC res1
```

```
reg EARNINGS S ASVABC res2
```

Если

- 1) оба коэффициента при res1 и res2 незначимы или оба значимы, то выбрать посредством теста Бера и МакАлера невозможно,
- 2) если незначим только коэффициент при res1, то лучше полулогарифмическая модель,
- 3) если незначим только коэффициент при res2, то лучше линейная модель.

III. По данным файла `dougherty.dta` выберите между линейной и полулогарифмической моделями с помощью теста Зарембки.

Необходимые команды

Выбор между линейной и полулогарифмической моделями

```
reg EARNINGS S ASVABC
```

```
gen lnEARNINGS = ln(EARNINGS)
```

```
reg lnEARNINGS S ASVABC
```

```
means EARNINGS
```

Variable	Type	Obs	Mean	[95% Conf. Interval]	
EARNINGS	Arithmetic	540	19.71924	18.48493	20.95355
	Geometric	540	16.3442	15.54379	17.18584
	Harmonic	540	13.77391	13.05586	14.57555

```
gen EARNINGSstar= EARNINGS/16.3442
```

```
gen lnEARNINGSstar = ln(EARNINGSstar)
```

```
reg EARNINGSstar S ASVABC
```

```
сохраните RSS с помощью команды scalar rss3=e(rss)
```

```
reg lnEARNINGSstar S ASVABC
```

```
сохраните RSS с помощью команды scalar rss4=e(rss)
```

Используя RSS из оцененных регрессий, следует рассчитать тестовую F – статистику

$$\text{scalar } xi2 = 0.5*540*abs(\ln(rss4/rss3))$$

```
display xi2
```

и p-value для этой статистики:

```
display chi2tail(1, xi2)
```

Если рассчитанное p-value не превышает выбранного уровня значимости, то основная гипотеза отвергается, есть разница между исходными линейной и полулогарифмическими моделями.