

```

*****
/*      8 ноября 2018      Ратникова Т.А.      */
/*      Ошибки спецификации и гетероскедастичность      */
*****

use "housing.dta", clear
log using "house_new.smcl"

/* вывод описательных статистик */
sum

/* изучение распределений количественных показателей */
/* построение гистограммы и сглаженной плотности распределения
зависимой переменной */
twoway hist price || kdensity price
twoway hist lotsize || kdensity lotsize

/* формирование полного перечня регрессоров */
global HOUSE_1 "lotsize bedrooms bathrms stories driveway /// recroom
fullbase gashw airco garagepl prefarea"

/* оценивание полной линейной регрессионной модели */
reg price $HOUSE_1
est store model_1

/* создание ряда остатков оцененной линейной модели */
predict e_1, residual

/* тестирование остатков линейной модели на предмет отклонения от
предположений КЛРМ */

/* тестирование остатков на нормальность */

/* построение гистограммы и сглаженной плотности распределения
остатков линейной модели */
twoway hist e_1 || kdensity e_1
/* тест Харке-Бера на нормальность остатков */
Sktest e_1

/* тестирование ошибок спецификации линейной модели */

/* построение графика зависимости остатков от предсказанных регрессией
значений цены */
rvfplot
/* тест Рамсея на пропуск существенных переменных */
ovtest

/* тестирование гетероскедастичности остатков */

/* тест Бройша-Пагана на гетероскедастичность остатков */
hettest
/* тест Камеруна-Трайверди на гетероскедастичность и нормальность
остатков */
imtest
/* ранговый тест на гетероскедастичность остатков */
szroeter e_1

```

```

/* переход к другой функциональной форме */

/* генерирование логарифмов непрерывных переменных:
   цены и размера участка */
gen lprice=log(price)
gen llotsize=log(lotsize)
/* изучение распределений количественных показателей */
/* построение гистограммы и сглаженной плотности распределения*/
twoway hist lprice || kdensity lprice
twoway hist llotsize || kdensity llotsize

/* формирование полного перечня регрессоров */
global HOUSE_2 "llotsize bedrooms bathrms stories driveway recroom
fullbase gashw airco garagepl prefarea"

/* оценивание полной логарифмической регрессионной модели */
reg lprice $HOUSE_2
est store model_2

/* создание ряда остатков оцененной логарифмической модели */
predict e_2, residual
/* тестирование остатков логарифмической модели на предмет отклонения
от предположений КЛРМ */

/* тестирование остатков на нормальность */

/* построение гистограммы и сглаженной плотности распределения*/
twoway hist e_2 || kdensity e_2
/* тест Харке-Бера на нормальность остатков */
Sktest e_2

/* тестирование ошибок спецификации */

/* построение графика зависимости остатков от предсказанных регрессией
значений цены */
rvfplot
/* тест Рамсея на пропуск существенных переменных */
ovtest

/* тестирование гетероскедастичности остатков */

/* тест Бройша-Пагана на гетероскедастичность остатков */
hettest
/* тест Камеруна-Трайверди на гетероскедастичность и нормальность
остатков */
imtest
/* ранговый тест на гетероскедастичность остатков */
szroeter e_2

/* поиск источника гетероскедастичности */

/* построение графиков остатков от отдельных регрессоров */
rvpplot llotsize
/* построение графиков квадрата остатков от отдельных регрессоров */
gen esq_2=e_2*e_2
scatter esq_2 llotsize
scatter esq_2 bedrooms
scatter esq_2 bathrms

```

```

gen ae_2=abs(e_2)
/* проведение тестов Глейзера на выявление источников
гетероскедастичности */
foreach var of varlist $HOUSE_2 {
qui reg ae_2 `var'
est store ae2_`var'
}
est tab ae2_llotsize ae2_bedrooms ae2_bathrms ae2_stories ae2_driveway
ae2_recroom ae2_fullbase ae2_gashw ae2_airco ae2_prefarea, stats(r2)
b(%7.3f) star

/* выявление кластеризации в данных */

tab airco
tab prefarea

global HOUSE_3 "llotsize bedrooms bathrms stories driveway recroom
fullbase gashw airco garagepl"

/* выявление гетероскедастичности в остатках регрессий для разных
кластеров */

/* оценивание полной логарифмической регрессионной модели для
коттеджей в элитной зоне */
reg lprice $HOUSE_3 if prefarea==1
est store model_3_1
scalar rss1=e(rss)

/* создание ряда остатков модели для коттеджей в элитной зоне */
predict e_3, residual
/* тестирование остатков модели для коттеджей в элитной зоне на
гетероскедастичность и нормальность */
imtest

/* оценивание полной логарифмической регрессионной модели для
коттеджей в неэлитной зоне*/
reg lprice $HOUSE_3 if prefarea==0
est store model_3_0
scalar rss0=e(rss)

/* создание ряда остатков модели для коттеджей в неэлитной зоне */
predict e_4, residual
/* тестирование остатков модели для коттеджей в неэлитной зоне на
гетероскедастичность и нормальность */
imtest

/* создание перекрестных произведений регрессоров и дамми-переменной
для расположения в элитной зоне */
foreach var of varlist $HOUSE_3 {
gen pre_`var'=prefarea*`var'
}

/* оценивание полной логарифмической регрессионной модели с учетом
кластеризации по зонам */
global HOUSE_4 "$HOUSE_2 pre_llotsize pre_bedrooms pre_bathrms
pre_stories pre_driveway pre_recroom pre_fullbase pre_gashw pre_airco
pre_garagepl "

```

```

reg lprice $HOUSE_4
est store model_4

/* тестирование полной логарифмической регрессионной модели с учетом
кластеризации по зонам на гетероскедастичность и мультиколлинеарность
*/

imtest
vif
/* вывод матрицы парных корреляций переменных модели */
pwcorr price $HOUSE_4, star(.05)

/* проверка гипотез о группе лишних переменных */
test prefarea pre_llotsize pre_bedrooms pre_bathrms pre_stories
pre_fullbase pre_garagepl pre_airco pre_recroom

/* тест Чоу (нужна ли была кластеризация?) */
reg lprice lotsize bedrooms bathrms stories driveway recroom fullbase
gashw airco garagepl
scalar rss_01 =e(rss)
scalar ddf=546-2*11
scalar F=((rss_01-rss1-rss0)/11)/((rss1+rss0)/ddf)
scalar pvalF=Ftail(11,ddf,F)
scalar list F pvalF

log close

```