

## Эконометрика, 2017-2018, 2 модуль

### Семинар 6

4.12.17 для

Группы Э\_Б2015\_Э\_3

Семинарист О.А.Демидова

### Выбор включаемых в модель факторов

**Задача 9.1.** (К.Доугерти, Введение в эконометрику, изд.3, стр. 216, задача № 6.7).

Исследователь считает, что уровень активности в теневой экономике  $Y$  зависит либо положительно от налогового бремени  $X$ , либо отрицательно от уровня государственных расходов на предотвращение теневой экономической деятельности  $Z$ . Переменная  $Y$  может также зависеть от обеих переменных  $X$  и  $Z$ . Получены международные данные двух перекрестных выборок по  $Y$ ,  $X$  и  $Z$  (в млн долл. США): для группы из 30 индустриально развитых и для группы из 30 развивающихся стран. Исследователь оценивает регрессионные зависимости: 1)  $\log Y$  от  $\log Z$ ; 2)  $\log Y$  только от  $\log X$ ; 3)  $\log Y$  только от  $\log Z$  одновременно для каждой выборки, получая следующие результаты (в скобках приведены стандартные отклонения):

	Индустриально развитые страны			Развивающиеся страны		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log X$	0.699	0.201	-	0.806	0.727	-
s.e.	(0.154)	(0.112)		(0.137)	(0.090)	
$\log Z$	-0.646	-	-0.053	-0.091	-	0.427
s.e.	(0.162)		(0.124)	(0.117)		(0.116)
Константа	-1.137	-1.065	1.230	-1.122	-1.024	2.824
s.e.	(0.863)	(1.069)	(0.896)	(0.873)	(0.858)	(0.835)

$R^2$	0.44	0.10	0.01	0.71	0.70	0.33

Переменная  $X$  положительно коррелирована с  $Z$  в обеих выборках. Выполните соответствующие статистические тесты, напишите краткий обзор, дайте рекомендации исследователю относительно интерпретации полученных результатов. Выберите, какая из моделей лучше

А) Для индустриально развитых стран, Б) Для развивающихся стран. Объясните изменения в оценках коэффициентов и их стандартных отклонений в других моделях.

### Задача 9.2.

- 1) По 150 наблюдениям оценили зависимость почасовой заработной платы от пола (переменная  $MALE$  равно 1 для мужчин и 0 для женщин), длительности обучения  $S$  и возраста  $AGE$ .

$$\hat{Y} = 3.6 + 3.5 MALE + 3.24 S + 0.44 AGE, RSS = 7632$$

(3.09) (1.21) (0.53) (0.057)

Используя результаты двух вспомогательных регрессий, приведенных ниже, проведите RESET – тест и ответьте, правильная ли спецификация модели выбрана.

$$\hat{Y} = 12.37 - 0.29 MALE - 0.49 S - 0.08 AGE + 0.0064 \hat{Y}^2, RSS = 7154$$

(4.09) (1.7) (1.3) (0.17) (0.002)

$$\hat{Y} = -18.1 + 9.2 MALE + 7.93 S + 1.1 AGE - 0.012 \hat{Y}^2 + 1.75 \cdot 10^{-10} \hat{Y}^3, RSS = 6069$$

(4.42) (2.44) (2.05) (0.28) (0.004) (3.45 \cdot 10^{-11})

### Задача 9.3.

При применении к модели, результаты оценки которой приведены ниже,

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
AGE	-10.70493	9.211662	-1.16	0.246	-28.80062	7.390769
agesq	.1300605	.1125515	1.16	0.248	-.0910395	.3511605
EXP	.4429137	.1442633	3.07	0.002	.159518	.7263094
S	2.578227	.2288185	11.27	0.000	2.128729	3.027726
MALE	6.364055	1.111968	5.72	0.000	4.179668	8.548442
ETHBLACK	-4.014172	2.152185	-1.87	0.063	-8.241996	.2136528
ETHHISP	-1.078255	2.268688	-0.48	0.635	-5.534941	3.378432

```

      _cons |    193.7202    187.6859        1.03    0.302    -174.9761    562.4165
-----+-----
. vif

      Variable |          VIF        1/VIF
-----+-----
      AGE |    1411.96    0.000708
    agesq |    1411.13    0.000709
      EXP |         1.29    0.778114
        S |         1.14    0.875122
  ETHBLACK |         1.04    0.962602
      MALE |         1.03    0.966488
  ETHHISP |         1.02    0.983851
-----+-----
    Mean VIF |        404.09

```

метода последовательного исключения, какая переменная будет удалена из уравнения регрессии на ближайшем шаге?

2) Оценено уравнение регрессии (в скобках указаны значения  $t$  – статистик)

$$\hat{Y} = 1 + \underset{(0.5)}{2.1} X_1 + \underset{(1.5)}{5.7} X_2 - \underset{(-1.7)}{7.5} X_3 + \underset{(2.1)}{3} X_4 - \underset{(-3.5)}{6.2} X_5$$

При удалении каких переменных качество подгонки регрессии может увеличиться?

#### Задание 9.4.

Оцененная с помощью МНК зависимость заработной платы индивида EARNINGS от его возраста AGE, опыта EXP, пола MALE, длительности обучения S, длительности обучения матери SM имеет вид (в скобках стандартные отклонения коэффициентов):

$$\widehat{EARN} = -24 - \underset{(10.6)}{0.099} AGE + \underset{(0.25)}{2.49} S + \underset{(0.249)}{0.26} SM + \underset{(0.24)}{0.46} EXP + \underset{(0.14)}{6.23} MALE, R^2 = 0.247$$

Были оценены также вспомогательные регрессии:

$$\widehat{AGE} = -0.007 + 0.53S - 0.6SM + 0.23EXP + 1.23MALE, R^2 = 0.2,$$

$$\widehat{S} = 8.47 + 0.095AGE + 0.4SM - 0.2EXP + 0.12MALE, R^2 = 0.25,$$

$$\hat{SM} = 6.16 - 0.045AGE + 0.42S + 0.08EXP + 0.42MALE, R^2 = 0.18,$$

$$\hat{EXP} = -.07 + 0.53AGE - 0.6S + 0.23SM + 1.23MALE, R^2 = 0.2,$$

Найдите VIF для переменных AGE, S, SM, EXP.

### Задание 9.5.

По данным для 23 демократических стран оценили зависимость индекса Джини (меры неравенства, 0 – полное равенство, по мере роста этого показателя степень неравенства увеличивается) от ВНР на душу населения с учетом ППС (паритета покупательной способности) и провели тест Рамсея. Результаты оценивания указаны в таблице. Прокомментируйте результаты теста Рамсея.

```
reg gini gdp if democ==1
```

Source	SS	df	MS	Number of obs =	23
-----+-----				F( 1, 21) =	13.05
Model	506.853501	1	506.853501	Prob > F =	0.0016
Residual	815.572523	21	38.8367868	R-squared =	0.3833
-----+-----				Adj R-squared =	0.3539
Total	1322.42602	22	60.1102738	Root MSE =	6.2319

  

gini	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
gdp	-.0006307	.0001746	-3.61	0.002	-.0009937 -.0002676
_cons	44.30983	3.572733	12.40	0.000	36.87993 51.73974
-----+-----					

```
ovtest
```

Ramsey RESET test using powers of the fitted values of gini

```
Ho: model has no omitted variables
      F(3, 18) =      5.16
      Prob > F =      0.0095
```

**9.3** По 30 наблюдениям при помощи метода наименьших квадратов оценена модель  $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , для которой  $RSS = 150$ . При помощи вспомогательной регрессии  $\hat{y} = \hat{\gamma}_1 + \hat{\gamma}_2x + \hat{\gamma}_3z + \hat{\gamma}_4\hat{y}^2 + \hat{\gamma}_5\hat{y}^3$ , для которой  $RSS = 120$ , выполните тест Рамсея на уровне значимости 5%.

**9.4** По 35 наблюдениям при помощи метода наименьших квадратов оценена модель  $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , для которой  $R^2 = 0.7$ . При помощи вспомогательной регрессии  $\hat{y} = \hat{\gamma}_1 + \hat{\gamma}_2x + \hat{\gamma}_3z + \hat{\gamma}_4\hat{y}^2 + \hat{\gamma}_5\hat{y}^3$ , для которой  $R^2 = 0.8$ , выполните тест Рамсея на уровне значимости 5%.

**9.5** Используя 80 наблюдений, исследователь оценил две конкурирующие модели:  $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , в которой  $RSS_1 = 36875$  и  $\widehat{\ln y} = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , в которой  $RSS_2 = 122$ .

Выполнив преобразование  $y_i^* = y_i / \sqrt[n]{\prod y_i}$ , исследователь также оценил две вспомогательные регрессии:  $\hat{y}^* = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , в которой  $RSS_1^* = 239$  и  $\widehat{\ln y^*} = \hat{\beta}_1 + \hat{\beta}_2x + \hat{\beta}_3z$ , в которой  $RSS_2^* = 121$ .

Завершите тест Бокса-Кокса на уровне значимости 5%.