

Эконометрика, 2017-2018, 2 модуль

Семинар 5.2

27.11.17 для

Группы Э_Б2015_Э_3

Семинарист О.А.Демидова

Выбор функциональной формы модели

Материалы из учебника О.Демидовой и Д.Малахова «Эконометрика. Учебник и практикум»

Задача 8.1.

По месячным данным с 01.2001 по 06.2003 были оценены три регрессии:

$$Y = 18 - 0.427P + 0.000007I \quad (1)$$

(0.23) (0.006) (0.00001)

$$\ln Y = 4.7 - 0.096P + 0.000001I \quad (2)$$

(0.15) (0.004) (0.000007)

$$\ln Y = 12 - 3.11 \ln P + 0.0317 \ln I \quad (3)$$

(0.73) (0.15) (0.029)

где Y – агрегированные расходы на медицинские услуги (в млрд. руб.), P – индекс цен на медицинские услуги, I – средний среднемесячный доход россиян (руб.), в скобках указаны стандартные отклонения.

Дайте экономическую интерпретацию полученным результатам.

Задача 8.2.

С помощью теста Бокса-Кокса оценили зависимость веса индивида от его роста:

$$W^{(\lambda)} = \beta_1 + \beta_2 H^{(\theta)} + \varepsilon$$

Исходя из результатов оценки, какую спецификацию модели (линейную, линейную в логарифмах, полулогарифмическую) Вы предпочтете и почему.

	Number of obs	=	540
	LR chi2(2)	=	230.68
Log likelihood = -2659.5656	Prob > chi2	=	0.000

-----+-----						
WEIGHT02	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
/lambda	1.055498	1.892654	0.56	0.577	-2.654035	4.76503
/theta	-.0263371	.1471576	-0.18	0.858	-.3147607	.2620865

Estimates of scale-variant parameters

	Coef.
Notrans	
_cons	2.936809
Trans	
HEIGHT	.0237224
/sigma	.1660251

Test	Restricted		
H0:	log likelihood	chi2	Prob > chi2
theta=lambda = -1	-2680.8693	42.61	0.000
theta=lambda = 0	-2659.7618	0.39	0.531
theta=lambda = 1	-2685.5201	51.91	0.000

Упражнение 8.1.

На основе встроенной в статистическом пакете STATA базы данных nlsw88.dta, оцените линейную в логарифмах модель для зарплаты (описание всех переменных приведено в приложении 1):

$$\ln wage_i = \beta_0 + \beta_1 \ln_ttl_exp_i + \beta_2 \ln_tenure_i + \beta_3 \ln_grade_i + \beta_4 race_i + \beta_5 union_i + \beta_6 south_i + u_i$$

и полулогарифмическую модель:

$$\ln wage_i = \beta_0' + \beta_1' \ln_ttl_exp_i + \beta_2' tenure_i + \beta_3' grade_i + \beta_4' race_i + \beta_5' union_i + \beta_6' south_i + \varepsilon_i$$

Проинтерпретируйте полученные результаты.

Решение:

Поскольку переменные tenure и grade принимают в том числе значения, равные 0, а переменные wage и ttl_exp всегда положительны (в этом легко убедиться, используя команду:

```
sum wage ttl_exp tenure grade,
```

в результате которой получим:

Variable		Obs	Mean	Std. Dev.	Min	Max
----------	--	-----	------	-----------	-----	-----

```

-----+-----
      wage |      2246      7.766949      5.755523      1.004952      40.74659
    ttl_exp |      2246      12.53498      4.610208      .1153846      28.88461
      tenure |      2231      5.97785      5.510331           0      25.91667
      grade |      2244      13.09893      2.521246           0           18,

```

то, чтобы не потерять наблюдения, в которых переменные `tenure` и `grade` принимают значения, равные нулю, сначала элементарно их преобразуем (добавим 1, эти новые переменные достаточно легко интерпретировать):

```

gen tenure1 = tenure+1,
gen grade1 = grade+1.

```

Создадим логарифмы непрерывных переменных, используя команду `gen`:

```

gen ln_wage=ln(wage)

gen ln_ttl_exp=ln(ttl_exp)

gen ln_tenure1=ln(tenure1)

gen ln_grade1=ln(grade1)

```

Фиктивные переменные `race`, `union`, `south` по очевидным причинам не преобразуются логарифмически.

Теперь оценим первую модель с помощью команды:

```

reg ln_wage ln_ttl_exp ln_tenure1 ln_grade1 race union south,

```

получим:

```

      Source |      SS      df      MS              Number of obs =      1866
-----+-----
      Model | 176.741105      6 29.4568509          F( 6, 1859) = 171.87
    Residual | 318.620761    1859  .171393631          Prob > F      = 0.0000
-----+-----
      Total | 495.361866    1865  .26560958          R-squared      = 0.3568
                                          Adj R-squared = 0.3547
                                          Root MSE      =  .414

```

```

-----+-----
      ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
    ln_ttl_exp |   .2950968   .0230577     12.80  0.000   .2498752   .3403184
    ln_tenure1 |   .0851056   .0136712      6.23  0.000   .058293   .1119183
    ln_grade1 |   .8264256   .0491267     16.82  0.000   .7300764   .9227749
      race |  -.0539839   .020811     -2.59  0.010  -.0947993  -.0131684
      union |   .1349628   .0229037      5.89  0.000   .0900431   .1798825
      south |  -.1324289   .0202091     -6.55  0.000  -.1720638  -.092794

```

```

      _cons | -1.061651      .136037      -7.80      0.000      -1.328452      -.7948496
-----+-----

```

Как следует из результатов оценки, эта модель адекватна, $R_{adj}^2 = 0.35$, все коэффициенты значимы на 5% уровне значимости. Исходя из полученных коэффициентов модели, если общий стаж возрастет на 1%, то зарплата возрастет на 0.295% при прочих равных (интерпретацию остальных коэффициентов оставляем в качестве упражнения читателю).

Теперь оценим полулогарифмическую модель с помощью команды:

```
reg ln_wage ttl_exp tenure1 grade1 race union south,
```

получим:

Source	SS	df	MS	Number of obs =	1866
-----+-----				F(6, 1859) =	186.40
Model	186.075738	6	31.0126231	Prob > F	= 0.0000
Residual	309.286128	1859	.166372312	R-squared	= 0.3756
-----+-----				Adj R-squared =	0.3736
Total	495.361866	1865	.26560958	Root MSE	= .40789

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
ttl_exp	.0335061	.0025626	13.07	0.000	.0284802 .0385321
tenure1	.0083856	.0020851	4.02	0.000	.0042962 .012475
grade1	.0733228	.0038303	19.14	0.000	.0658106 .0808349
race	-.0578975	.0204558	-2.83	0.005	-.0980163 -.0177787
union	.1320118	.0225986	5.84	0.000	.0876904 .1763331
south	-.1328175	.019902	-6.67	0.000	-.1718501 -.0937849
_cons	.4598454	.0652003	7.05	0.000	.3319718 .5877189
-----+-----					

Аналогично, эта модель адекватна, $R_{adj}^2 = 0.37$, все коэффициенты значимы на любом адекватном уровне значимости. Исходя из оценок коэффициентов данной модели, если общий стаж работы увеличиться на 1 год, то зарплата возрастет на 3.35% при прочих равных. Поскольку R_{adj}^2 в обеих моделях очень близки, то на основе этих показателей трудно сказать, какая модель лучше.

Упражнение 8.2.

На основе встроенной в статистическом пакете STATA базы данных nlsw88.dta (описание всех переменных приведено в приложении 1), сравните линейную модель для зарплат:

$$wage_i = \beta_0 + \beta_1' ttl_exp_i + \beta_2' tenure_i + \beta_3' grade_i + \beta_4' race_i + \beta_5' union_i + \beta_6' south_i + \varepsilon_i$$

и линейную в логарифмах модель:

$\ln_wage_i = \beta_0 + \beta_1 \ln_ttl_exp_i + \beta_2 \ln_tenure_i + \beta_3 \ln_grade_i + \beta_4 race_i + \beta_5 union_i + \beta_6 south_i + u_i$,
используя тест РЕ тест Дэвидсона, Уайта и МакКиннона.

Решение.

Сначала оценим линейную модель с помощью команды:

```
reg wage ttl_exp tenure1 grade1 race union south,
```

получим:

Source		SS	df	MS	Number of obs = 1866	
-----+-----					F(6, 1859) =	132.10
Model		9693.82304	6	1615.63717	Prob > F	= 0.0000
Residual		22736.2036	1859	12.2303408	R-squared	= 0.2989
-----+-----					Adj R-squared	= 0.2967
Total		32430.0267	1865	17.3887542	Root MSE	= 3.4972

wage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
ttl_exp		.2358167	.0219718	10.73	0.000	.1927246 .2789088
tenure1		.0418487	.0178776	2.34	0.019	.0067864 .076911
grade1		.5840443	.0328406	17.78	0.000	.5196359 .6484526
race		-.2507745	.1753863	-1.43	0.153	-.5947493 .0932003
union		.800398	.1937585	4.13	0.000	.4203909 1.180405
south		-.9433636	.1706377	-5.53	0.000	-1.278025 -.6087019
_cons		-3.503466	.5590217	-6.27	0.000	-4.599842 -2.40709

Сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict y_hat
```

Аналогично оценим линейную в логарифмах модель с помощью команды:

```
reg ln_wage ln_ttl_exp ln_tenure1 ln_grade1 race union south,
```

получим:

Source	SS	df	MS	Number of obs = 1866		
-----+-----				F(6, 1859) =	171.87	
Model	176.741105	6	29.4568509	Prob > F =	0.0000	
Residual	318.620761	1859	.171393631	R-squared =	0.3568	

-----+-----					Adj R-squared = 0.3547	
Total		495.361866	1865	.26560958	Root MSE	= .414
-----+-----						
ln_wage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
ln_ttl_exp		.2950968	.0230577	12.80	0.000	.2498752 .3403184
ln_tenure1		.0851056	.0136712	6.23	0.000	.058293 .1119183
ln_grade1		.8264256	.0491267	16.82	0.000	.7300764 .9227749
race		-.0539839	.020811	-2.59	0.010	-.0947993 -.0131684
union		.1349628	.0229037	5.89	0.000	.0900431 .1798825
south		-.1324289	.0202091	-6.55	0.000	-.1720638 -.092794
_cons		-1.061651	.136037	-7.80	0.000	-1.328452 -.7948496
-----+-----						

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 линейная модели:

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

и оценим эту модель:

```
reg wage ttl_exp tenure1 grade1 race union south lin_add
```

получим:

Source		SS	df	MS	Number of obs = 1863	
-----+-----					F(7, 1855) = 116.17	
Model		9864.22938	7	1409.17563	Prob > F = 0.0000	
Residual		22501.5701	1855	12.1302265	R-squared = 0.3048	
-----+-----					Adj R-squared = 0.3021	
Total		32365.7995	1862	17.3822768	Root MSE = 3.4828	
-----+-----						
wage		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
ttl_exp		.264288	.023213	11.39	0.000	.2187616 .3098144
tenure1		.0204572	.0188192	1.09	0.277	-.0164518 .0573663
grade1		.6791962	.0403594	16.83	0.000	.6000416 .7583507
race		-.243289	.175403	-1.39	0.166	-.587297 .1007191

union	.6942744	.1955297	3.55	0.000	.310793	1.077756
south	-.9924902	.1704138	-5.82	0.000	-1.326713	-.6582672
ln_add	3.135945	.8953066	3.50	0.000	1.380031	4.891859
_cons	-4.783241	.6414437	-7.46	0.000	-6.041268	-3.525213

Создадим также дополнительную разность для линейной в логарифмах модели:

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

оценим эту модель:

```
reg ln_wage ln_ttl_exp ln_tenure1 ln_gradel race union south log_add
```

и получим:

Source	SS	df	MS	Number of obs =	1866
-----+-----				F(7, 1858) =	153.46
Model	181.478253	7	25.9254648	Prob > F	= 0.0000
Residual	313.883612	1858	.168936282	R-squared	= 0.3664
-----+-----				Adj R-squared =	0.3640
Total	495.361866	1865	.26560958	Root MSE	= .41102

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
ln_ttl_exp	.2281519	.0261507	8.72	0.000	.1768641 .2794396
ln_tenure1	.128159	.0158217	8.10	0.000	.0971289 .1591892
ln_gradel	.4784654	.0818331	5.85	0.000	.317971 .6389598
race	-.0737616	.0209962	-3.51	0.000	-.1149402 -.032583
union	.1666843	.0235148	7.09	0.000	.1205662 .2128024
south	-.1260247	.0201001	-6.27	0.000	-.1654459 -.0866036
log_add	.1486221	.0280664	5.30	0.000	.0935771 .203667
_cons	-.1370307	.2207465	-0.62	0.535	-.5699679 .2959064

Поскольку оба коэффициента при дополнительных разностях значимы, то выбрать посредством теста Дэвидсона, Уайта и МакКиннона невозможно.

База данных woman (файл nlws88.xls, nlws88.csv, nlws88.dta)

Данная база данных содержит информацию о зарплате и социально-демографических характеристиках 2246 женщин, проживавших в 1988 году в США.

idcode – id респондентов, участвующих в опросе,

age- возраст респондента (в годах),

race – переменная, равная 2, если женщина чернокожая и 1, если белая, 0 – иначе,
 black – дамми-переменная, равная 1, если женщина чернокожая, 0 – иначе,
 other – дамми-переменная, равная 1, если переменная race=0 (то есть женщина – латинос или азиатка и т.д.), 0 – иначе,
 married – дамми-переменная, равная 1, если на момент опроса женщина замужем и 0 иначе,
 never_married- дамми-переменная равная 1, если женщина никогда не была замужем и 0 – иначе,
 grade – количество лет образования женщины,
 collgrad- переменная, равная 1, если женщина является выпускницей колледжа и 0 иначе,
 south- дамми-переменная, равная 1, если женщина проживает на юге США и 0 иначе,
 smsa- дамми-переменная, равная 1, если женщина проживает в standard metropolitan statistical area (стандартная учётная территориальная единица расселения в США (с наличием не менее одного города с 50 000 жителей)),
 c_city - дамми-переменная, равная 1, если женщина проживает в центральном городе и 0 иначе,
 industry - название отрасли, в которой работает женщина (полный список можно получить, набрав в STATA команду tab industry),
 occupation - название позиции (должности), на которой работает респондент (полный список можно получить, набрав в STATA команду tab occupation),
 union - дамми-переменная, равная 1, если женщина состоит в профсоюзе и 0 иначе,
 wage - уровень почасовой зарплаты женщины (долл.),
 hours – продолжительность средней рабочей недели респондента (в часах),
 ttl_exp - общий рабочий стаж респондента (в годах),
 tenure – время, которое женщина занимает текущую должность (в годах).