

Задачи по Эконометрике-2: Качество подгонки и сравнение моделей

Н.В. Артамонов (МГИМО МИД России)

Содержание

1	Качество подгонки	1
1.1	labour force equation #1 (probit)	1
1.2	approve equation #1 (logit)	2
2	Сравнение моделей	3
2.1	labour force equation #1 (probit)	3
2.2	approve equation #1 (logit)	5

1 Качество подгонки

1.1 labour force equation #1 (probit)

Для датасета TableF5-1.csv рассмотрим несколько probit-регрессий. Результаты оценивания

Зависимая переменная				
	LFP			
	(1)	(2)	(3)	(4)
WA	0.0076 (0.0701)		-0.0184 (0.0685)	-0.0056 (0.0681)
I (WA2)	-0.0005 (0.0008)		-0.0002 (0.0008)	-0.0003 (0.0008)
WE	0.1088*** (0.0241)	0.1238*** (0.0237)	0.1088*** (0.0238)	0.1230*** (0.0221)
KL6	-0.8513*** (0.1154)	-0.6209*** (0.0977)	-0.8473*** (0.1149)	-0.8554*** (0.1151)
K618	-0.0632 (0.0417)	0.0306 (0.0363)		
CIT	-0.1277 (0.1070)	-0.1654 (0.1053)		
UN	-0.0106	-0.0169		

	(0.0157)	(0.0155)		
log(FAMINC)	0.1996*	0.1704*	0.1626	
	(0.1049)	(0.1028)	(0.1016)	

Constant	-2.0046	-2.6731***	-1.4351	-0.2815
	(1.7039)	(0.9574)	(1.6682)	(1.4961)

Observations	753	753	753	753
Log Likelihood	-462.3402	-475.6991	-464.6961	-465.9906
Akaike Inf. Crit.	942.6804	965.3982	941.3922	941.9812

Note: *p<0.1; **p<0.05; ***p<0.01

Для каждой регрессии вычислите следующие показатели качества подгонки модели: R^2_{pseudo} , $adj.R^2_{pseudo}$, Cox & Snell, Nagelkerke/Cragg & Uhler, Efron, McKelvey & Zavoina. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

	Model	pseudo.R2	adj.pseudo.R2	CoxSnell	Nagelkerke	Efron	McKelvey	Zavoina
1		0.102	0.085	0.130	0.175	0.133	0.207	
2		0.076	0.062	0.099	0.133	0.100	0.158	
3		0.097	0.086	0.125	0.167	0.127	0.199	
4		0.095	0.085	0.122	0.163	0.123	0.195	

1.2 approve equation #1 (logit)

Для датасета loanapp рассмотрим несколько logit-регрессий. Результаты оценивания

	Зависимая переменная			
	approve			
	(1)	(2)	(3)	(4)
appinc	0.0042*		0.0033	0.0042*
	(0.0024)		(0.0023)	(0.0024)
I(appinc2)	-0.00001**		-0.00001**	-0.00001**
	(0.000003)		(0.000003)	(0.000003)
mortno	0.7072***	0.7700***	0.6941***	0.7363***
	(0.1750)	(0.1717)	(0.1748)	(0.1737)
unem	-0.0498*	-0.0521*	-0.0623**	
	(0.0296)	(0.0294)	(0.0286)	
dep	-0.1545**	-0.1683***	-0.1565**	-0.1032*
	(0.0652)	(0.0645)	(0.0649)	(0.0606)

male	-0.0207 (0.1876)	0.0216 (0.1857)		
married	0.3977** (0.1632)	0.4173** (0.1624)	0.3962** (0.1554)	
yjob	-0.0101 (0.0647)	-0.0066 (0.0651)		
self	-0.3593* (0.1995)	-0.3144 (0.1951)		
Constant	1.6806*** (0.2282)	1.8602*** (0.1928)	1.7311*** (0.2067)	1.6158*** (0.1632)

Observations	1971	1971	1986	1986
Log Likelihood	-713.7313	-717.4311	-717.3869	-722.7259
Akaike Inf. Crit.	1447.4630	1450.8620	1448.7740	1455.4520

Note: *p<0.1; **p<0.05; ***p<0.01

Для каждой регрессии вычислите следующие показатели качества подгонки модели: R^2_{pseudo} , $adj.R^2_{pseudo}$, Cox & Snell, Nagelkerke/Cragg & Uhler, Efron, McKelvey & Zavoina. Ответ округлите до 3-х десятичных знаков.

Ответ:

	Model pseudo.R2	adj.pseudo.R2	CoxSnell	Nagelkerke	Efron	McKelveyZavoina
1	0.033	0.019	0.024	0.046	0.027	0.066
2	0.028	0.017	0.021	0.039	0.022	0.060
3	0.030	0.021	0.022	0.043	0.026	0.062
4	0.023	0.017	0.017	0.033	0.019	0.049

2 Сравнение моделей

2.1 labour force equation #1 (probit)

Для датасета TableF5-1.csv рассмотрим несколько probit-регрессий. Результаты оценивания

	Зависимая переменная			
	LFP			
	(1)	(2)	(3)	(4)
WA	0.0076 (0.0701)		-0.0184 (0.0685)	-0.0056 (0.0681)
I(WA2)	-0.0005 (0.0008)		-0.0002 (0.0008)	-0.0003 (0.0008)

WE	0.1088*** (0.0241)	0.1238*** (0.0237)	0.1088*** (0.0238)	0.1230*** (0.0221)
KL6	-0.8513*** (0.1154)	-0.6209*** (0.0977)	-0.8473*** (0.1149)	-0.8554*** (0.1151)
K618	-0.0632 (0.0417)	0.0306 (0.0363)		
CIT	-0.1277 (0.1070)	-0.1654 (0.1053)		
UN	-0.0106 (0.0157)	-0.0169 (0.0155)		
log(FAMINC)	0.1996* (0.1049)	0.1704* (0.1028)	0.1626 (0.1016)	
Constant	-2.0046 (1.7039)	-2.6731*** (0.9574)	-1.4351 (1.6682)	-0.2815 (1.4961)

Observations 753 753 753 753
Log Likelihood -462.3402 -475.6991 -464.6961 -465.9906
=====

Note: *p<0.1; **p<0.05; ***p<0.01

Для каждой модели вычислите показатели информационных критерии AIC & BIC и adj.R²_{pseudo}. Ответ округлите до 3-х десятичных знаков.

Ответ

Модель	AIC	BIC	adj.pseudo.R2
1	942.680	984.297	0.085
2	965.398	997.767	0.062
3	941.392	969.137	0.086
4	941.981	965.102	0.085

Какая модель предпочтительней по информационным критериям и adj.R²_{pseudo}?

Ответ:

Критерий	Регрессия
AIC	3
BIC	4
adj.pseudo.R2	3

2.2 approve equation #1 (logit)

Для датасета loanapp рассмотрим несколько probit-регрессий. Результаты оценивания

Зависимая переменная				
	approve			
	(1)	(2)	(3)	(4)
appinc	0.0022*		0.0018	0.0022*
	(0.0013)		(0.0012)	(0.0012)
I(appinc2)	-0.000004**		-0.000003**	-0.000004**
	(0.000002)		(0.000002)	(0.000002)
mortno	0.3650***	0.3983***	0.3569***	0.3758***
	(0.0885)	(0.0866)	(0.0882)	(0.0875)
unem	-0.0280*	-0.0293*	-0.0341**	
	(0.0163)	(0.0162)	(0.0158)	
dep	-0.0784**	-0.0861**	-0.0798**	-0.0510
	(0.0353)	(0.0351)	(0.0351)	(0.0329)
male	-0.0076	0.0143		
	(0.1007)	(0.0997)		
married	0.2065**	0.2172**	0.2060**	
	(0.0873)	(0.0869)	(0.0830)	
yjob	-0.0039	-0.0010		
	(0.0342)	(0.0345)		
self	-0.1961*	-0.1750		
	(0.1088)	(0.1064)		
Constant	1.0102***	1.1051***	1.0366***	0.9715***
	(0.1231)	(0.1048)	(0.1115)	(0.0861)
<hr/>				
Observations	1971	1971	1986	1986
Log Likelihood	-713.9286	-717.5349	-717.6389	-722.9362
<hr/>				

Note:

*p<0.1; **p<0.05; ***p<0.01

Для каждой модели вычислите показатели информационных критериев AIC & BIC и adj.R²_{pseudo}. Ответ округлите до 3-х десятичных знаков.

Ответ

Модель	AIC	BIC	adj.pseudo.R2

1	1447.857	1503.720	0.019
2	1451.070	1495.760	0.017
3	1449.278	1488.435	0.018
4	1455.872	1483.842	0.014

Какая модель предпочтительней по информационных критериям и $adj.R^2_{pseudo}$?

Ответ:

=====
Критерий Регрессия

AIC 1
BIC 4
adj.pseudo.R2 1
