

Семинар 11.

Ошибки спецификации модели.

1. (Исключение существенных переменных) Дана стандартная модель парной регрессии

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

- (а) Чему равна МНК-оценка коэффициента β_2 при ограничении $\beta_1 = 0$.
- (б) Чему равна дисперсия оценки в пункте (а)? Покажите, что она меньше, чем $\sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$ — дисперсия МНК-оценки β_2 в регрессии без ограничения. Противоречит ли это теореме Гаусса–Маркова?
2. (Включение лишних переменных) Пусть процесс, порождающий данные, имеет вид:

$$y = X\beta + \varepsilon.$$

Модель, которую мы оцениваем:

$$y = X\beta + Z\gamma + \varepsilon.$$

Здесь X — $n \times k$ матрица, Z — $n \times l$ матрица, y — $n \times 1$ вектор, β — $k \times 1$ вектор, γ — $l \times 1$ вектор, ε — $n \times 1$ вектор.

- (а) Будет ли МНК-оценка вектора параметров β несмещённой?
- (б) Что произойдёт с оценкой ковариационной матрицы $\widehat{\text{Var}}(\hat{\beta})$?
- (в) Будет ли несмещённой МНК-оценка дисперсии случайной ошибки σ^2 ?
3. Для 400 голландских магазинов модной одежды с помощью трёх моделей оценили зависимость продаж в расчете на квадратный метр в гульденах, $Sales$, от:

- общей площади магазина, $Size$, в м^2 ;
- количества сотрудников, работающих целый день, $Nfull$;
- количества временных рабочих, $Ntemp$;
- дамми-переменной $Owner$, равной единице, если собственник один, и нулю иначе.

$$\widehat{Sales}_i = \underset{(718)}{6083} - \underset{(1.59)}{15.25}Size_i + \underset{(171)}{1452.8}Nfull_i + \underset{(423)}{420.15}Ntemp_i - \underset{(361)}{1464.1}Owner_i$$

$$\ln \widehat{Sales}_i = \underset{(0.11)}{8.59} - \underset{(0.00024)}{0.0024}Size_i + \underset{(0.026)}{0.183}Nfull_i + \underset{(0.066)}{0.102}Ntemp_i - \underset{(0.056)}{0.209}Owner_i$$

$$\ln \widehat{Sales}_i = \underset{(0.21)}{10.08} - \underset{(0.043)}{0.31} \ln Size_i + \underset{(0.061)}{0.22} \ln Nfull_i + \underset{(0.118)}{0.066} \ln Ntemp_i - \underset{(0.059)}{0.19} \ln Owner_i$$

В скобках приведены стандартные ошибки.

- (а) Дайте интерпретацию коэффициента при переменной *Size* в каждой из трёх моделей;
- (б) Подробно опишите, как выбрать наилучшую из этих моделей.
4. По данным для 23 демократических стран оценили зависимость индекса Джини от ВВП на душу населения с учетом ППС (паритета покупательной способности). Затем провели тест Рамсея.

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

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

- (а) Сформулируйте нулевую и альтернативную гипотезу теста Рамсея.
- (б) Опишите пошагово, как проводится тест Рамсея.
- (в) Прокомментируйте результаты теста Рамсея.