

面板数据的线性回归-

固定效应模型、随机效应模型与 hausman 检验

王裕 15320171151908

构建一个线性回归模型，检验制造业发展水平、信息化程度、人力资本水平、政府规模和城市规模这五个因素对安徽省生产性服务业集聚的影响，设定线性回归模型如下：

$$lq_{it} = \alpha + \beta_1 manu_{it} + \beta_2 internet_{it} + \beta_3 gov_{it} + \beta_4 edu_{it} + \beta_5 popu_{it} + \varepsilon_{it}$$

其中 i 代表安徽 16 个地级市，t 代表年份， α 代表常数项， $manu_{it}$ 、 $internet_{it}$ 、 gov_{it} 、 edu_{it} 、 $popu_{it}$ 分别是反应 i 地级市第 t 年的制造业发展水平、信息化程度、政府规模、人力资本水平和城市规模的变量， α 代表常数项， $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ 代表待估系数， ε_{it} 代表误差项。

选用区位商作为被解释变量的测度指标，测度安徽省 2006 至 2015 年各地级市的生产性服务业集聚水平。区位商的公式为， $LQ_{ij} = \frac{x_{ij}/\sum_j x_{ij}}{\sum_i x_{ij}/\sum_i \sum_j x_{ij}}$ ，其中 i 表第 i 地区，j 代表第 j 行业， x_{ij} 代表第 i 地区第 j 行业的从业人数。

被解释变量选取如下表：

解释变量	英文缩写	度量指标	检验假说
制造业发展水平	manu	i 城市第 t 年工业增加值占 i 城市第 t 年的地区生产总值的比重	假说 1
信息化程度	internet	i 城市第 t 年每百人国际互联网用户数	假说 2
人力资本水平	edu	i 城市第 t 年每万人高等学校在校学生数	假说 3
政府规模	gov	i 城市第 t 年地方财政支出占 i 城市第 t 年的地区生产总值的比重	假说 4
城市规模	popu	i 城市第 t 年人口数量占安徽省第 t 年全省城市人口数量均值	假说 5

以上数据来自主要来源于 2007-2016 年的《中国城市统计年鉴》与《安徽统计年鉴》

使用 STATA 软件进行模型回归分析

1、导入数据，声明截面变量和时间变量

```
. tsset city year
      panel variable:  city (strongly balanced)
      time variable:  year, 2006 to 2015
      delta:  1 unit
```

2、面板数据模型回归分析

首先采用固定效应模型

. xtreg lq edu internet gov popu manu , fe

Fixed-effects (within) regression	Number of obs	=	160
Group variable: city	Number of groups	=	16
R-sq: within = 0.2307	Obs per group: min =		10
between = 0.0014	avg =		10.0
overall = 0.0046	max =		10
	F(5,139)	=	8.34
corr(u_i, Xb) = -0.9069	Prob > F	=	0.0000

	lq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
+							
edu		-.0000771	.0002544	-0.30	0.762	-.0005801	.0004259
internet		-.0050749	.0027319	-1.86	0.065	-.0104763	.0003264
gov		.9862992	.403545	2.44	0.016	.188419	1.784179
popu		-.5697519	.1212959	-4.70	0.000	-.8095754	-.3299283
manu		-.4805528	.2832698	-1.70	0.092	-1.040628	.079522
_cons		1.48934	.1512436	9.85	0.000	1.190305	1.788376
+							
sigma_u		.45560161					
sigma_e		.11216163					
rho		.94285689	(fraction of variance due to u_i)				

F test that all u_i=0: F(15, 139) = 18.48 Prob > F = 0.0000

根据这里给出的 F 统计量和 P 值，这里固定效应十分显著

再采用随机效应模型，进行对比

. xtreg lq edu internet gov popu manu , re

Random-effects GLS regression	Number of obs	=	160
Group variable: city	Number of groups	=	16
R-sq: within = 0.1282	Obs per group: min =		10
between = 0.1580	avg =		10.0
overall = 0.1399	max =		10
Random effects u_i ~ Gaussian	Wald chi2(5)	=	21.26
corr(u_i, X) = 0 (assumed)	Prob > chi2	=	0.0007

	lq	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
+						

edu	.000461	.0002033	2.27	0.023	.0000625	.0008594
internet	-.0074294	.0027641	-2.69	0.007	-.0128468	-.0020119
gov	1.161239	.3424933	3.39	0.001	.4899646	1.832514
popu	-.1092189	.0651641	-1.68	0.094	-.2369382	.0185004
manu	-.5332193	.2318327	-2.30	0.021	-.987603	-.0788357
_cons	.9442453	.1084649	8.71	0.000	.7316581	1.156833

sigma_u	.15114579
sigma_e	.11216163
rho	.64487986 (fraction of variance due to u_i)

3、模型的筛选与检验

1) 检验个体效应的显著性

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

$$lq[city,t] = Xb + u[city] + e[city,t]$$

Estimated results:

	Var	sd = sqrt(Var)
lq	.0459837	.2144382
e	.0125802	.1121616
u	.0228451	.1511458

Test: $Var(u) = 0$

chi2(1) =	194.21
Prob > chi2 =	0.0000

P 值为 0.0000，表明随机效应十分显著

2) Hausman 检验，具体步骤为：

Step1 估计固定效应模型，存储记忆结果

Step2 估计随机效应模型，存储记忆结果

Step3 进行 hausman 检验

. qui xtreg lq edu internet gov popu manu , fe

. est store fe

. qui xtreg lq edu internet gov popu manu , re

. est store re

. hausman fe

---- Coefficients ----				
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	fe	re	Difference	S.E.
-----+-----				
edu	-.0000771	.000461	-.000538	.0001529
internet	-.0050749	-.0074294	.0023544	.
gov	.9862992	1.161239	-.17494	.2134172
popu	-.5697519	-.1092189	-.4605329	.1023051
manu	-.4805528	-.5332193	.0526666	.1627741

b = consistent under Ho and Ha; obtained from xtreg

B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 25.84
 Prob>chi2 = 0.0001
 (V_b-V_B is not positive definite)

固定效应模型和随机效应模型的参数估计方差的差是一个非正定矩阵，此时随机效应模型的基本假设得不到满足，此时应采用固定效应模型。

参考文献: <https://wenku.baidu.com/view/c089a9bbe518964bce847c67.html>