
经济数据的特点与类型。

- 1、横截面数据：多个经济个体的变量在同一时间点上的取值，如 2012 年中国各省的 GDP
- 2、时间数列数据：指的是某个经济个体的变量在不同时点上的取值，如 1978-2012 年山东省每年的 GDP
- 3、面板数据：多个经济个体的变量在不同时点上的取值，如 1978-2012 年中国各省的 GDP

小样本 OLS（最小二乘法）：单一方程线性回归最常见方法

条件：解释变量与扰动项正交、扰动项无自相关、同方差。

拟合优度：衡量线性回归模型对样本数据的拟合程度 (R^2)，越高说明模型拟合程度越好。

单系数 T 检验：对回归方程扰动项的具体概率进行假设

显著性水平进行检验

F 检验：整个回归方程是否显著

STATA 操作简介：

如果数据中包含 1949-10-01 或 1949/10/01 的时间变量，导入 stata 后可能会被视为字符串，因此对于日度数据，可以使用命令 `gen newvar=date(varname,YMD)`，将其转换为整数日期变量，其中 YMD 说明原始数据的格式为年月日，如果原始数据的格式为月日年则使用 MDY；对于月度数据则 `gen newvar=monthly(varname,YM)`。

`.describe`：数据的概貌

`.drop keep`：删除和保留

`.su`：统计特征

`Pwcorr`：变量之间相关系数

`Star (.05)`：5% 显著性水平

`gen`：产生

`g intc=log(tc)`：取自然对数

`reg`：OLS 回归

`.Vce`：协方差矩阵

`reg... , noc` 表示在进行回归时不要常数项

大样本 OLS：只要求解释变量与同期的扰动项正交即可

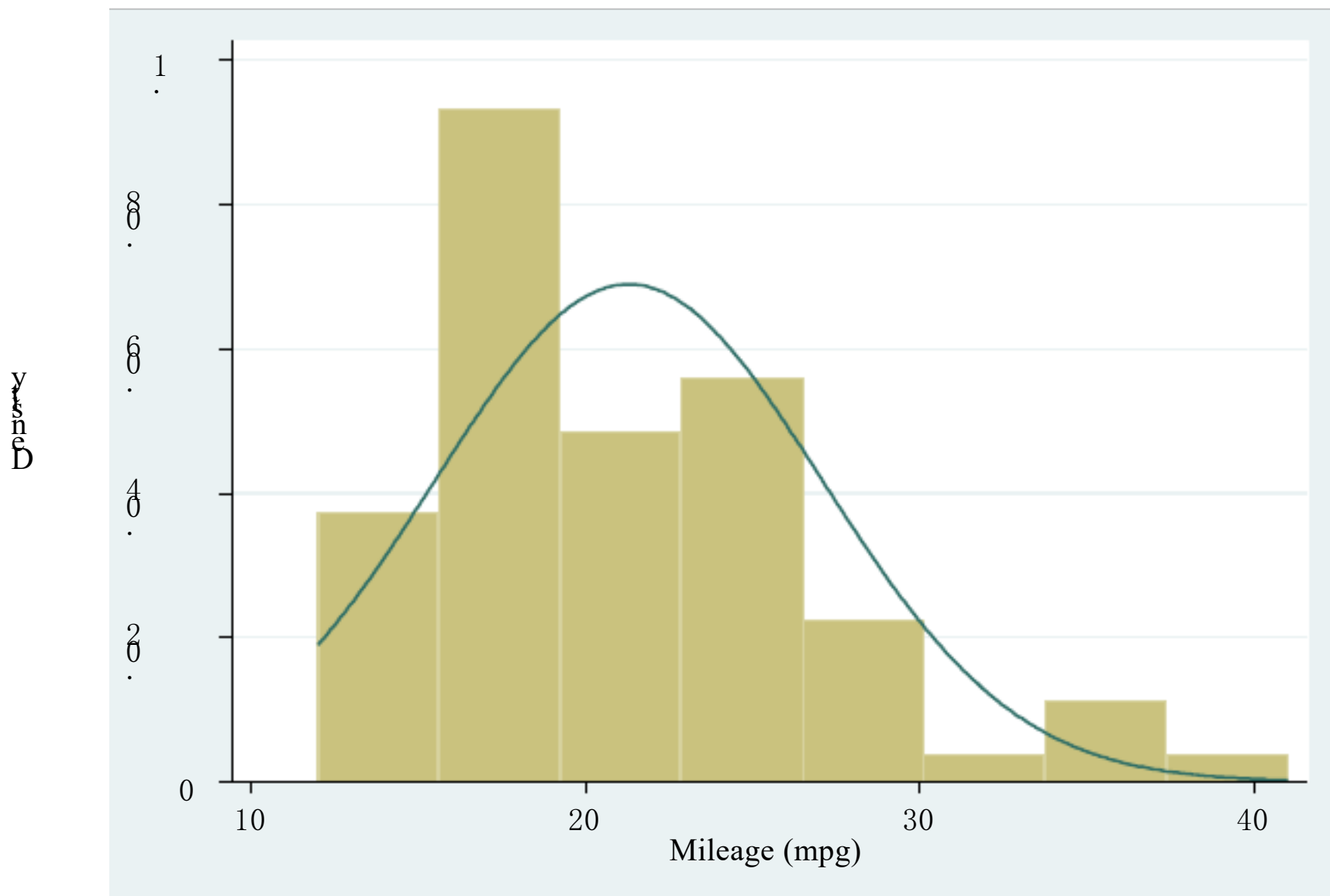
Robust：稳健标准误，如果存在异方差，则应使用稳健标准误

最大似然估计法：如果回归方程存在非线性，则使用最大似然估计法（MLE）或非线性最小二乘法（NLS）

三类在大样本下渐进等价的统计检验：Wald test LR（似然比检验） LM

操作步骤如下：`sysuse auto`（调用数据集）

`Hist mpg, normal`（画变量 `mpg` 的直方图，并与正态密度比较）



直方图显示，变量 `mpg` 的分布于正态分布有一定差距。
变量可以取对数解决非正态分布的问题。

异方差与 GLS（广义最小二乘法）

异方差的检验：看残差图、怀特检验（white test）、BP 检验（Breusch and Pagan）

异方差的处理：1、OLS 稳健标准误（最好的）

2、广义最小二乘法（GLS）

3、加权最小二乘法（WLS）

实例操作：

1、使用数据：`use nerlove.dta,clear`

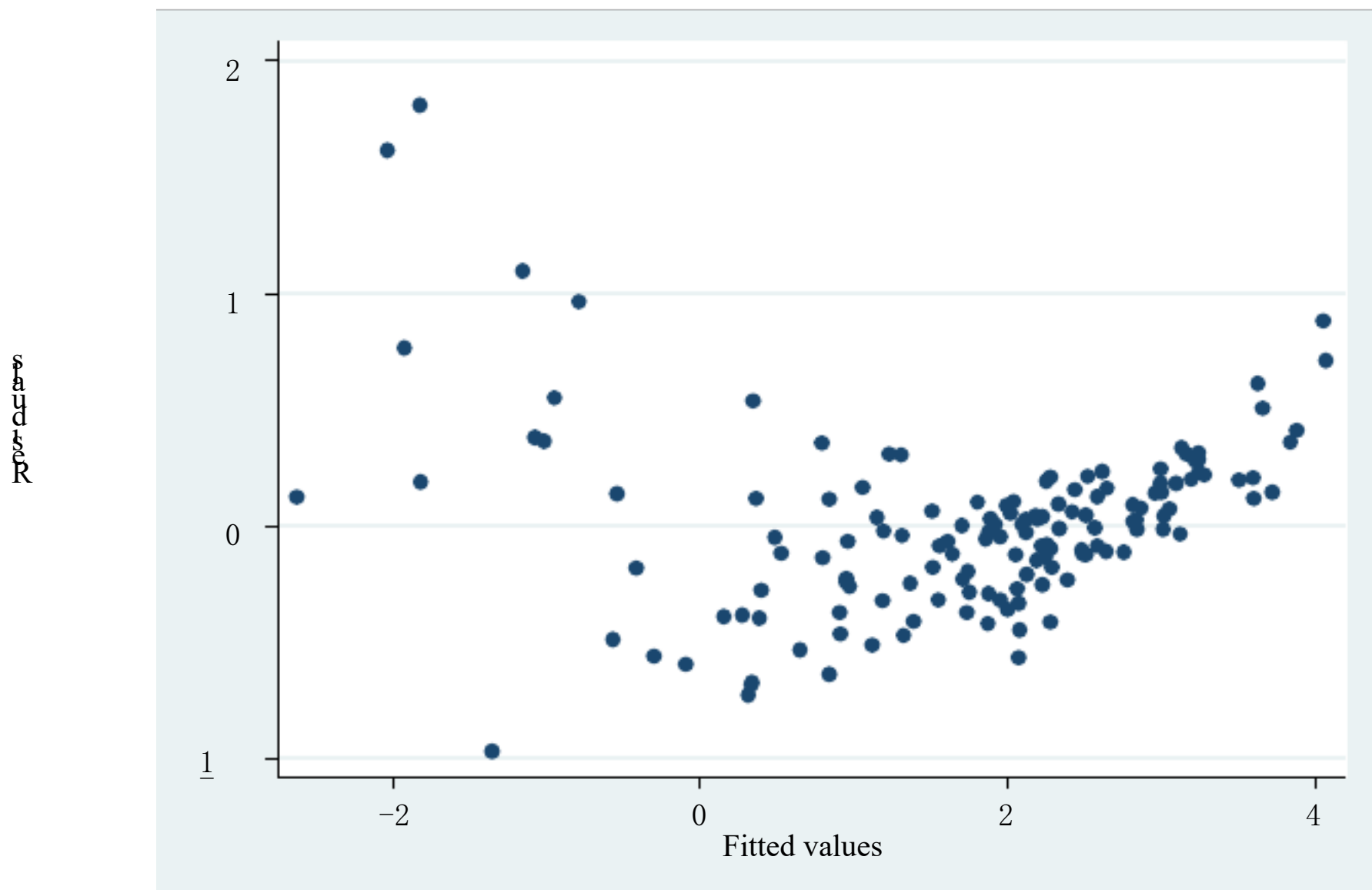
2、`reg intc inq inpl inpk inpf`（进行回归）

3、

Source	SS	df	MS	Number of obs = 145	
F(4, 140) = 437.90					
Model	269.524728	4	67.3811819	Prob > F	= 0.0000
Residual	21.5420958	140	.153872113	R-squared	= 0.9260
Adj R-squared = 0.9239					
Total	291.066823	144	2.02129738	Root MSE	= .39227

intc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
inq	.7209135	.0174337	41.35	0.000	.6864462	.7553808
inpl	.4559645	.299802	1.52	0.131	-.1367602	1.048689
inpk	-.2151476	.3398295	-0.63	0.528	-.8870089	.4567136
inpf	.4258137	.1003218	4.24	0.000	.2274721	.6241554
_cons	-3.566513	1.779383	-2.00	0.047	-7.084448	-.0485779

4、画残差图：rvfplot



上图可以发现当拟合值较小时，扰动项方差较大，继续考察残差与解释变量 inq 的散点图：rvpplot inq，结果与上图几乎一致，可能存在异方差，即扰动项的方差随着观测值而变。

5、完成回归后，进行怀特检验：estat imtest, white

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(14) = 73.88
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	73.88	14	0.0000
Skewness	22.79	4	0.0001
Kurtosis	2.62	1	0.1055
Total	99.29	19	0.0000

P 值显著，认为存在异方差

6、完成回归后，进行 BP 检验：`estat hettest,iid` `estat hottest,rhs iid` `estat hottest inq,iid`

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance
Variables: inq inpl inpk inpf

chi2(4) = 36.16
Prob > chi2 = 0.0000

三种形式的检验都强烈拒绝同方差的原假设，存在异方差（这里只放一个形式的检验结果）

7、处理异方差

自相关：扰动项之间自相关

自相关的例子：1、时间序列数据中通常具有某种连续性和持久性，如相邻两年的 GDP 增长率；2、截面数据中相邻的观测单位之间可能存在溢出效应，如相邻地区的农业产量收到类似天气变化的影响；3、对数据的人为处理如数据中包含移动平均数等；4、如果模型设定中遗漏了某个自相关的解释变量并被纳入到扰动项中，则会引起扰动项的自相关。

自相关的检验：1、画图（不推荐）2、BG 检验 `estat bgodfrey` 3、BOX-Pierce Q 检验 4、DW 检验 `estat dwatson`。 检验都要在 OLS 做完后才能做。

自相关的处理：1、使用 OLS 异方差自相关稳健的标准误；2、OLS 聚类稳健的标准误；3、使用可行广义最小二乘法（FGLS）；4、修改模型设定

自相关处理实例：1、使用数据 `icecream` 然后进行回归

. reg consumption temp price income

Source	SS	df	MS	Number of obs = 30	
				F(3, 26) = 22.17	
Model	.090250523	3	.030083508	Prob > F	= 0.0000
Residual	.035272835	26	.001356647	R-squared	= 0.7190
				Adj R-squared = 0.6866	
Total	.125523358	29	.004328392	Root MSE	= .03683

consumption	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
temp	.0034584	.0004455	7.76	0.000	.0025426	.0043743
price	-1.044413	.834357	-1.25	0.222	-2.759458	.6706322
income	.0033078	.0011714	2.82	0.009	.0008999	.0057156
_cons	.1973149	.2702161	0.73	0.472	-.3581223	.752752

BG 检验

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	4.237	1	0.0396

H0: no serial correlation

显著 拒绝了原假设无自相关, 则认为存在自相关

Q 检验 (略)、DW 检验如下

. estat dwatson

Durbin-Watson d-statistic(4, 30) = 1.021169

DW=1.02 距离 2 很远 可以认为存在自相关。

由以上的检验可以看出扰动项之间存在自相关, 因此 OLS提供的标准误是不准确的, 应使用异方差自相关稳健标准误, 由于样本为 30 个, $n \text{ 四分之一} = 2.34$, 故取 NEWey-West估计量的滞后值为 P=3, 结果如下:

. newey consumption temp price income,lag(3)

Regression with Newey-West standard errors Number of obs = 30
maximum lag: 3 F(3, 26) = 27.63
 Prob > F = 0.0000

consumption	Newey-West					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
temp	.0034584	.0004002	8.64	0.000	.0026357	.0042811
price	-1.044413	.9772494	-1.07	0.295	-3.053178	.9643518
income	.0033078	.0013278	2.49	0.019	.0005783	.0060372
_cons	.1973149	.3378109	0.58	0.564	-.4970655	.8916952

上图显示标准误与 OLS标准误无多大区别，因此将滞后阶数增加为 6，
.newey consumption temp price income,lag(6)

Regression with Newey-West standard errors Number of obs = 30
maximum lag: 6 F(3, 26) = 52.97
 Prob > F = 0.0000

consumption	Newey-West					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
temp	.0034584	.0003504	9.87	0.000	.0027382	.0041787
price	-1.044413	.9821798	-1.06	0.297	-3.063313	.9744864
income	.0033078	.00132	2.51	0.019	.0005945	.006021
_cons	.1973149	.3299533	0.60	0.555	-.4809139	.8755437

从上图可以看到无论截断参数是 3 还是 6，标准误都变化不大，比较稳健。

此外，前面提到自相关存在可能是因为模型设定不正确，因此考虑在解释变量中加入 temp 的滞后值，然后再进行 OLS 回归：

. reg consumption temp L.temp price income

Source	SS	df	MS	Number of obs = 29		
				F(4, 24) = 28.98		
Model	.103387183	4	.025846796	Prob > F = 0.0000		
Residual	.021406049	24	.000891919	R-squared = 0.8285		
				Adj R-squared = 0.7999		
Total	.124793232	28	.004456901	Root MSE = .02987		

consumption	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
temp						
--	.0053321	.0006704	7.95	0.000	.0039484	.0067158
L1.	-.0022039	.0007307	-3.02	0.006	-.0037119	-.0006959
price	-.8383021	.6880205	-1.22	0.235	-2.258307	.5817025
income	.0028673	.0010533	2.72	0.012	.0006934	.0050413
_cons	.1894822	.2323169	0.82	0.423	-.2899963	.6689607

然后使用 BG 检验是否存在自相关：

. estat bgo

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.120	1	0.7292

H0: no serial correlation

结果显示无自相关,而后 DW 值也改进为 1.58,因此修改模型后扰动项基本不再存在自相关。

模型设定与数据问题

遗漏变量: 被解释变量可能被加入到扰动项中

解决方式: 加入尽可能多的控制变量; 使用代理变量; 工具变量法; 使用面板数据; 随机试验或自然实验。

其中代理变量应满足两个条件: 多余性, 仅通过影响遗漏变量而作用于被解释变量; 剩余独立性, 遗漏变量中不受代理变量影响的剩余部分与所有解释变量均不相关。

多重共线性: 某一解释变量可以由其他解释变量线性表出, 即存在多重共线性。

检测: 先回归, 然后 estat vif VIF 低于 10 即不存在多重共线性。

工具变量, 2SLS 与 GMM

工具变量的适用条件: OLS 成立的最重要条件是解释变量与扰动项不相关, 如出现相关可以使用工具变量法来解决。其中在计量经济学中, 将所有与扰动项相关的解释变量成为内生变量, 一个有效的工具变量应满足一下两个条件: 1、工具变量与内生解释变量相关; 2、工具变量与扰动项不相关。

过程: 传统的工具变量法一般通过二阶段最小二乘法 (2SLS 或 TSLS 来实现: 1、用内生解释变量对工具变量进行回归, 得到拟合值 Pt (Pt 实际上是内生变量中的外生部分, 而另一部分是与扰动项相关); 2、用被解释变量对第一阶段的拟合值 Pt 进行回归。

工具变量的检测: 1、不可识别检验; 2、弱工具变量检验; 3、过度识别检验 (estat overid)。

豪斯曼检验: 原假设为所有解释变量均为外生变量, 若拒绝假设的话应该使用工具变量法, 若接受的话使用 OLS

豪斯曼过程: reg X1 x2

Estimates store ols(存储 OLS 的结果)

Ivregress 2sls y x1(x2=z1 z2)(假设怀疑 X2 为内生变量)

Estimates store iv(存储 2SLS 结果)

Hausman iv ols,constant sigmamore(根据存储的结果进行豪斯曼检验)

若存在异方差的问题, 则可以使用杜宾-吴-豪斯曼检验。它在异方差的情况下也适用。

在球形扰动项的假定下, 2SLS 最有效, 但如果扰动项存在异方差或自相关, 则存在 GMM 广义矩估计这一更有效的方法: 首先使用 2SLS 得到残差, 然后 GMM, 但是在实际操作中使

用迭代法。

GMM的命令:

```
ivregress gmm y x1(x2=z1 z2)          两步最优 GMM
ivregress gmm y x1 (x2=z1 z2) igmm     迭代 GMM
estat overid                            过度识别检验
```

工具变量法的 STATA命令和实例:

```
Use grilic.dta,clear
```

```
Sum
```

然后考察智商和受教育年限的相关关系 (本文研究的是工资与受教育年限的关系)

```
. pwcorr iq s,sig
```

	iq	s
iq	1.0000	
s	0.5131	1.0000 0.0000

具有较强的正相关关系。

然后作为一个参照系, 进行 OLS回归并使用稳健标准误。

```
. reg lw s expr tenure rns smsa,r
```

```
Linear regression          Number of obs = 758
                          F( 5, 752) = 84.05
                          Prob > F   = 0.0000
                          R-squared   = 0.3521
                          Root MSE  = .34641
```

lw	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
s	.102643	.0062099	16.53	0.000	.0904523	.1148338
expr	.0381189	.0066144	5.76	0.000	.025134	.0511038
tenure	.0356146	.0079988	4.45	0.000	.0199118	.0513173
rns	-.0840797	.029533	-2.85	0.005	-.1420566	-.0261029
smsa	.1396666	.028056	4.98	0.000	.0845893	.194744
_cons	4.103675	.0876665	46.81	0.000	3.931575	4.275775

教育投资率 10.26%, 显然过高, 可能是遗漏了变量能力, 使得能力对工资的贡献也被纳入教育的贡献。

因此使用 iq 作为能力的代理变量, 再进行 OLS回归, 可以发现加入 iq 作为能力的代理变量后, 教育投资回报率降低了一些, 但还是过高。(如下图)


```
. reg lw s iq expr tenure rns smsa,r
```

```
Linear regression                Number of obs =   758
                                F( 6, 751) = 71.89
                                Prob > F   = 0.0000
                                R-squared   = 0.3600
                                Root MSE  = .34454
```

lw	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
s	.0927874	.0069763	13.30	0.000	.0790921	.1064826
iq	.0032792	.0011321	2.90	0.004	.0010567	.0055016
expr	.0393443	.0066603	5.91	0.000	.0262692	.0524193
tenure	.034209	.0078957	4.33	0.000	.0187088	.0497092
rns	-.0745325	.0299772	-2.49	0.013	-.1333815	-.0156834
smsa	.1367369	.0277712	4.92	0.000	.0822186	.1912553
_cons	3.895172	.1159286	33.60	0.000	3.667589	4.122754

使用 iq 来度量能力存在测量误差，因此 iq 为内生变量，考虑使用 med kww mrt age 作为 iq 的工具变量，进行 2SLS 回归，并使用稳健标准误。

```
. ivregress 2sls lw s expr tenure rns smsa (iq=med kww mrt age),r
```

```
Instrumental variables (2SLS) regression    Number of obs =   758
                                             Wald chi2(6) = 355.73
                                             Prob > chi2  = 0.0000
                                             R-squared   = 0.2002
                                             Root MSE   = .38336
```

lw	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	-.0115468	.0056376	-2.05	0.041	-.0225962	-.0004974
s	.1373477	.0174989	7.85	0.000	.1030506	.1716449
expr	.0338041	.0074844	4.52	0.000	.019135	.0484732
tenure	.040564	.0095848	4.23	0.000	.0217781	.05935
rns	-.1176984	.0359582	-3.27	0.001	-.1881751	-.0472216
smsa	.149983	.0322276	4.65	0.000	.0868182	.2131479
_cons	4.837875	.3799432	12.73	0.000	4.0932	5.58255

Instrumented: iq

Instruments: s expr tenure rns smsa med kww mrt age

受教育年限回报上升，而 iq 竟然是负相关，因此不可信，使用工具变量法需要验证其工具变量的有效性因此进行过度识别来检验所有工具变量是否外生。

```
. estat overid
```

Test of overidentifying restrictions:

Score chi2(3) = 51.5449 (p = 0.0000)

上图显示有些工具变量不合格，与扰动项相关。怀疑 mrt 和 age 不满足外生性，因此仅适用 med 和 kww 作为 iq 的工具变量，再次进行 2SLS 回归，同时显示第一阶段的回归结果。

```
. ivregress 2sls lw s expr tenure rns smsa (iq=med kww),r first
```

First-stage regressions

```
Number of obs = 758
F( 7, 750) = 47.74
Prob > F = 0.0000
R-squared = 0.3066
Adj R-squared = 0.3001
Root MSE = 11.3931
```

iq	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
s	2.467021	.2327755	10.60	0.000	2.010052	2.92399
expr	-.4501353	.2391647	-1.88	0.060	-.9196471	.0193766
tenure	.2059531	.269562	0.76	0.445	-.3232327	.7351388
rns	-2.689831	.8921335	-3.02	0.003	-4.441207	-.938455
smsa	.2627416	.9465309	0.28	0.781	-1.595424	2.120907
med	.3470133	.1681356	2.06	0.039	.0169409	.6770857
kww	.3081811	.0646794	4.76	0.000	.1812068	.4351553
_cons	56.67122	3.076955	18.42	0.000	50.63075	62.71169

```
Instrumental variables (2SLS) regression      Number of obs = 758
Wald chi2(6) = 370.04
Prob > chi2 = 0.0000
R-squared = 0.2775
Root MSE = .36436
```

lw	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	.0139284	.0060393	2.31	0.021	.0020916	.0257653
s	.0607803	.0189505	3.21	0.001	.023638	.0979227
expr	.0433237	.0074118	5.85	0.000	.0287968	.0578505
tenure	.0296442	.008317	3.56	0.000	.0133432	.0459452
rns	-.0435271	.0344779	-1.26	0.207	-.1111026	.0240483
smsa	.1272224	.0297414	4.28	0.000	.0689303	.1855146
_cons	3.218043	.3983683	8.08	0.000	2.437256	3.998831

Instrumented: iq

量对第一阶段回归的拟合值进行回归。

上图中教育回报率较为合理，而且 iq 系数也为整数，再次进行过度识别检验。

```
. estat overid
```

Test of overidentifying restrictions:

```
Score chi2(1) = .151451 (p = 0.6972)
```

结果没有拒绝外生的原假设。

阶段的回归看出，med 和 kww 对 iq 有较好的解释力，但为稳健起见，还是使用对弱工具变量更不敏感的有限信息最大似然法（LIML）。

```
. ivregress liml lw s expr tenure rns smsa (iq=med kww),r
```

```
Instrumental variables (LIML) regression      Number of obs =   758
      Wald chi2(6) = 369.62
      Prob > chi2  = 0.0000
      R-squared    = 0.2768
      Root MSE    = .36454
```

lw	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	.0139764	.0060681	2.30	0.021	.0020831	.0258697
s	.0606362	.019034	3.19	0.001	.0233303	.0979421
expr	.0433416	.0074185	5.84	0.000	.0288016	.0578816
tenure	.0296237	.008323	3.56	0.000	.0133109	.0459364
rns	-.0433875	.034529	-1.26	0.209	-.1110631	.0242881
smsa	.1271796	.0297599	4.27	0.000	.0688512	.185508
_cons	3.214994	.4001492	8.03	0.000	2.430716	3.999272

Instrumented: iq

以上结果与 2SLS 非常接近，侧面验证了不存在弱工具变量。还有，使用工具变量法的前提是存在内生解释变量，因此进行豪斯曼检验。

```
. qui reg lw iq s expr tenure rns smsa
. estimates store ols
. qui ivregress 2sls lw s expr tenure rns smsa (iq=med kww)
. estimates store iv
. hausman iv ols,constant sigmamore
```

Note: the rank of the differenced variance matrix (1) does not equal the number of coefficients being tested (7); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients			
	(b) iv	(B) ols	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
iq	.0139284	.0032792	.0106493	.0054318
s	.0607803	.0927874	-.032007	.0163254
expr	.0433237	.0393443	.0039794	.0020297
tenure	.0296442	.034209	-.0045648	.0023283
rns	-.0435271	-.0745325	.0310054	.0158145
smsa	.1272224	.1367369	-.0095145	.0048529
_cons	3.218043	3.895172	-.6771285	.3453751

b = consistent under Ho and Ha; obtained from ivregress
B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

$$\chi^2(1) = (b-B)'[(V_b-V_B)^{-1}](b-B) = 3.84$$

Prob>chi2 = 0.0499

(V_b-V_B is not positive definite)

iq 为内生变量，又因为传统的豪斯曼检验在异方差的情况下不成立，下面进行异方差稳健的 DWH 检验：

Tests of endogeneity

Ho: variables are exogenous

Durbin (score) chi2(1) = 3.87962 (p = 0.0489)

DWH 的 P 值小于 0.05，故可以认为 iq 为内生解释变量。

另外如果存在异方差，则 GMM 比 2SLS 更有效，因此进行最优 GMM 估计：

```
. ivregress gmm lw s expr tenure rns smsa (iq=med kww)
```

```
Instrumental variables (GMM) regression      Number of obs =   758
              Wald chi2(6) = 372.75
              Prob > chi2 = 0.0000
              R-squared   = 0.2750
GMM weight matrix: Robust                  Root MSE   = .36499
```

lw	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	.0140888	.0060357	2.33	0.020	.0022591	.0259185
s	.0603672	.0189545	3.18	0.001	.0232171	.0975174
expr	.0431117	.0074112	5.82	0.000	.0285861	.0576373
tenure	.0299764	.0082728	3.62	0.000	.013762	.0461908
rns	-.044516	.0344404	-1.29	0.196	-.1120179	.0229859
smsa	.1267368	.0297633	4.26	0.000	.0684018	.1850718
_cons	3.207298	.398083	8.06	0.000	2.427069	3.987526

Instrumented: iq

Instruments: s expr tenure rns smsa med kww

上图显示两步最优 GMM 与 2SLS 很接近，再进行过度识别检验

```
. estat overid
```

Test of overidentifying restriction:

Hansen's J chi2(1) = .151451 (p = 0.6972)

结果接受原假设，说明所有工具变量外生。

然后再做迭代 GMM：下图显示与两步 GMM 系数估计值相差不大。

```
. ivregress gmm lw s expr tenure rns smsa (iq=med kww),igmm
Iteration 1: change in beta = 1.753e-05 change in W = 1.100e-02
Iteration 2: change in beta = 4.872e-08 change in W = 7.880e-05
Iteration 3: change in beta = 2.507e-10 change in W = 2.303e-07
```

```
Instrumental variables (GMM) regression      Number of obs = 758
Wald chi2(6) = 372.73
Prob > chi2 = 0.0000
R-squared = 0.2750
GMM weight matrix: Robust                Root MSE = .36499
```

lw	Robust					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	.0140901	.0060357	2.33	0.020	.0022603	.02592
s	.0603629	.0189548	3.18	0.001	.0232122	.0975135
expr	.0431101	.0074113	5.82	0.000	.0285841	.057636
tenure	.0299752	.0082729	3.62	0.000	.0137606	.0461898
rns	-.0445114	.0344408	-1.29	0.196	-.1120142	.0229913
smsa	.1267399	.0297637	4.26	0.000	.0684041	.1850757
_cons	3.207224	.3980878	8.06	0.000	2.426986	3.987462

Instrumented: iq

```
. est sto ols_no_iq
. qui reg lw iq s expr tenure rns smsa,r
. est sto ols_with_iq
. qui ivregress 2sls lw s expr tenure rns smsa (iq=med kww),r
. est sto tsls
. qui ivregress liml lw s expr tenure rns smsa (iq=med kww),r
. est sto liml
. qui ivregress gmm lw s expr tenure rns smsa (iq=med kww)
. qui ivregress gmm lw s expr tenure rns smsa (iq=med kww)
. est sto gmm
. qui ivregress gmm lw s expr tenure rns smsa (iq=med kww),igmm
. est sto igmm
. estimates table ols_no_iq ols_with_iq tsls liml gmm igmm,b se
```

其中，选项 b 表示显示回归系数，se 表示显示标准误差

Variable	ols_no_iq	ols_with~q	tsls	liml	gmm	igmm
s	.10264304	.09278735	.06078035	.06063623	.06036723	.06036285
expr	.0381189	.03934425	.04332367	.04334159	.04311171	.04311006
tenure	.03561456	.03420896	.02964421	.02962365	.02997643	.02997521
rns	-.08407974	-.07453249	-.04352713	-.04338751	-.04451599	-.04451145
smsa	.13966664	.13673691	.12722244	.1271796	.12673682	.12673991
iq	.00327916	.01392844	.01397639	.01408883	.01409011	
_cons	4.103675	3.8951718	3.2180433	3.2149943	3.2072978	3.2072239

10% 显著性水平等等：

. estimates table ols_no_iq ols_with_iq tsls liml gmm igmm,star(0.1 0.05 0.01)

Variable	ols_no_iq	ols_with_iq	tsls	liml	gmm
s	.10264304***	.09278735***	.06078035***	.06063623***	.06036723***
expr	.0381189***	.03934425***	.04332367***	.04334159***	.04311171***
tenure	.03561456***	.03420896***	.02964421***	.02962365***	.02997643***
rns	-.08407974***	-.07453249**	-.04352713	-.04338751	-.04451599
smsa	.13966664***	.13673691***	.12722244***	.1271796***	.12673682***
iq	.00327916***	.01392844**	.01397639**	.01408883**	
_cons	4.103675***	3.8951718***	3.2180433***	3.2149943***	3.2072978***

legend: * p<.1; ** p<.05; *** p<.01

Variable	igmm
s	.06036285***
expr	.04311006***
tenure	.02997521***
rns	-.04451145
smsa	.12673991***
iq	.01409011**
_cons	3.2072239***

legend: * p<.1; ** p<.05; *** p<.01

如果想像论文一样显示，则如下表：se 表示在括弧中显示标准误差，p 表示显示 P 值，r2 表示显示 R 的平方，mtitle 显示使用模型名字，

```
. esttab ols_no_iq ols_with_iq tsls liml gmm igmm,se r2 mtitle star
```

	(1)	(2)	(3)	(4)	(5)	(6)
	ols_no_iq	ols_with_iq	tsls	liml	gmm	igmm
s	0.103***	0.0928***	0.0608**	0.0606**	0.0604**	0.0604*
	(0.00621)	(0.00698)	(0.0190)	(0.0190)	(0.0190)	(0.0190)
expr	0.0381***	0.0393***	0.0433***	0.0433***	0.0431***	0.0431*
	(0.00661)	(0.00666)	(0.00741)	(0.00741)	(0.00741)	(0.00741)
tenure	0.0356***	0.0342***	0.0296***	0.0296***	0.0300***	0.0300*
	(0.00800)	(0.00790)	(0.00832)	(0.00832)	(0.00832)	(0.00827)
rns	-0.0841**	-0.0745*	-0.0435	-0.0434	-0.0445	-0.0445
	(0.0295)	(0.0300)	(0.0345)	(0.0345)	(0.0345)	(0.0344)
smsa	0.140***	0.137***	0.127***	0.127***	0.127***	0.127*
	(0.0281)	(0.0278)	(0.0297)	(0.0298)	(0.0298)	(0.0298)
iq	0.00328**	0.0139*	0.0140*	0.0141*	0.0141*	
		(0.00113)	(0.00604)	(0.00607)	(0.00604)	(0.00604)
_cons	4.104***	3.895***	3.218***	3.215***	3.207***	3.207*
	(0.0877)	(0.116)	(0.398)	(0.398)	(0.400)	(0.398)
N	758	758	758	758	758	758
R-sq	0.352	0.360	0.278	0.277	0.275	0.275

Standard errors in parentheses

离散选择模型、定性反应模型或被解释变量取非负整数时，都不适宜使用 OLS 回归。

1、二值选择模型：只有两种选择，是否。

Probit $y \ x_1 \ x_2 \ x_3, r$ (probit 模型)

Logit $y \ x_1 \ x_2 \ x_3, or \ vce(\text{cluster } \text{clustvar})$ (logit 模型)

其中， r 代表使用稳健标准误， or 显示几率比而不是系数， vce 表示使用以 $clustvar$ 为聚类变量的聚类稳健标准误。

Stata 举例：美国妇女就业与否的二值选择模型。

```
. use "C:\Users\Administrator\Desktop\stata\y\womenwk.dta"
. reg work age married children education, r
```

Linear regression

Number of obs = 2000
F(4, 1995) = 192.58
Prob > F = 0.0000
R-squared = 0.2026
Root MSE = .41992

work	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
age	.0102552	.0012236	8.38	0.000	.0078556 .0126548
married	.1111116	.0226719	4.90	0.000	.0666485 .1555748
children	.1153084	.0056978	20.24	0.000	.1041342 .1264827
education	.0186011	.0033006	5.64	0.000	.0121282 .025074
_cons	-.2073227	.0534581	-3.88	0.000	-.3121622 -.1024832

然后使用 logit 进行估计：

```
. logit work age married children education, nolog
```

Logistic regression

Number of obs = 2000
LR chi2(4) = 476.62
Prob > chi2 = 0.0000
Pseudo R2 = 0.1882

Log likelihood = -1027.9144

work	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.0579303	.007221	8.02	0.000	.0437773 .0720833
married	.7417775	.1264705	5.87	0.000	.4938998 .9896552
children	.7644882	.0515289	14.84	0.000	.6634935 .865483
education	.0982513	.0186522	5.27	0.000	.0616936 .134809
_cons	-4.159247	.3320401	-12.53	0.000	-4.810034 -3.508461

结果显示所有系数的联合显著性很高，继续使用稳健标准误进行 logit 回归：

```
. logit work age married children education, r nolog
```

Logistic regression

Number of obs = 2000
Wald chi2(4) = 344.54
Prob > chi2 = 0.0000
Pseudo R2 = 0.1882

Log pseudolikelihood = -1027.9144

work	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
age	.0579303	.0072054	8.04	0.000	.0438079 .0720527
married	.7417775	.1272191	5.83	0.000	.4924326 .9911224
children	.7644882	.0497584	15.36	0.000	.6669635 .8620129
education	.0982513	.019011	5.17	0.000	.0609904 .1355121
_cons	-4.159247	.327398	-12.70	0.000	-4.800936 -3.517559

对比以上两个表格显示标准误相差不大，因此不用担心模型设定问题。

二值选择模型中的异方差问题：`hetprob y x1 x2 x3,het(varlist)` 如果接受原假设则为同方差。此外，二值选择模型中一般都没有扰动项的存在。

二值选择模型中的异方差问题可以进行似然比检验 (LR)：`hetprob y x1 x2 x3,het(varlist)`(这是在异方差情况下进行 Probit 估计的 stata 命令，`het (varlist)` 制定对扰动项方差有影响的所

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/975132123330011232>