

The empirical method of economic research and stata code

在做任何分析之前都要做协变量平衡分析，防止由于对照组和控制组变量分布造成的误差。

- 1.Random Experiment
- 2.OLS
- 4.Instrument Variable
 - 弱工具变量检验
 - 外生性（排除性）检验
 - 过度识别检验
- Limit dependent variable
- Matching
- 实用小代码stata
- 一些方法

1.Random Experiment

1. 在进行因果估计之前为了避免存在样本分布问题，或者选择性问题，通常会对对照组和样本组进行随机化分析，即计算对照组和实验组具有近似的样本分布。这样可以表示条件独立性。

```
// 随机实验验证 对于分组进行验证 检查子组内的平衡
gen subgroup = group(变量) // 生成分组变量 这个公式会生成一个新的变量，这个变量
// 是根据原来的变量进行取分组值的
bysort subgroup: summarize(变量) // 按照分组变量进行分组，然后对变量进行描述性统计
// 因为产生的太快了，需要一个变量一个变量跑，然后j子组内对照组和实验组进行对比
```

分组求回归等公式

```
// 分组求回归等公式
bys subgroup: logit/reg y x
```

2.OLS

1. OLS回归

```
reg y x1 x2 x3//robust 异方差情况
```

2. 加权回归 利用加权对于其残差平方和进行加权，进行最小化.其权重可以选择倾向得分

```
```stata
reg y x1 x2 x3 [aweight = weight] //加权回归
```
```

4. Instrument Variable

我们在使用工具变量时，需要进行检验，最常见的就是排除性和相关性。

进行IV时我们需要讲故事，并且数据检验其合理性：同时其最基础的工具变量回归的代码如下

```
ivregress 2sls y (x1 = z1 z2) x2 x3, robust
```

弱工具变量检验

1. F检验

```
reg y x ,robust // OLS回归估计
ivregress 2sls y (x=z1,z2),robust // 2SLS回归估计
reg x z1 z2,robust // 第一阶段回归估计
test z1 z2 //查看是否有弱工具变量问题·F检验 大于10即可 F估计与弱IV的关系来自于
causal inference
```

可以通过以上的第一阶段回归查看第一阶段的参数从而判断工具变量的相关性

也可以比较OLS和2SLS的结果，看看是否有差异

2. Cragg-Donald检验

一般条件是同方差，无自相关

```
ivreg2 y (x1 x2 = z1 z2), robust //Cragg-Donald检验,要大于 10
```

3. Kleibergen-Paap检验 无iid假设

```
ivreg2 y (x1 x2 = z1 z2), robust //Kleibergen-Paap检验,要大于 10
```

外生性（排除性）检验

1. Hausman检验

```
//豪斯曼检验 这是在同方差条件下的检验
reg y x1 x2
estimates store ols
ivregress 2sls y (x1 = z) x2
estimates store iv
hausman iv ols, constant sigmamore
//chi - squared和p - value.p 小于0.05·拒原·认为变量是内生变量,p最好大一点
```

2. DWH检验

用上一个检验的结果就行，也会输出DWH检验的结果。这是在异方差条件下的检验

3. GMM估计

```
ivregress gmm y (x1 = z1 z2), twostep robust
estat overid //原假设：工具变量是有外生的
```

过度识别检验

1. Sargan检验 用于线性模型中的工具变量过度识别检验

```
ivregress 2sls y (x1 = z1 z2)
```

2. Anderson - Rubin 检验 用于非线性模型或联立方程模型中的工具变量过度识别检验 以联立方程模型为例

```
sysreg (eq1: y1 = x1 x2 (y2 = z1 z2)) (eq2: y2 = x3 x4 (y1 = z3 z4))
test [eq1_y2] [eq2_y1] // 原假设是不存在过度识别问题
```

3. Hansen J统计量 非iid时用Hansen J统计量 和Sargan检验类似 非iid时用Hassen统计量

Limit dependent variable

当相关变量是虚拟变量或选择变量时，我们必须使用其他模型，例如 logit 或probit模型来估计模型

```
logit y x1 x2 x3 //默认使用最大似然估计
//关于logit的迭代以及公式可以看崔学彬的ppt，就是MLE和回归的替换
```

Matching

1. 精确匹配

2. 模糊匹配

3. 倾向得分匹配PSM

其具有降维的力量，同时避免了因协变量较多带来的维度诅咒问题。由于倾向得分匹配是被处理的概率，因此可以通过被处理概率来进行匹配。即可以用Logit或Probit模型来估计倾向得分 这是由于倾向得分定理表示得分值也满足条件独立性，因此可以消除选择偏误。

- 倾向得分回归

```
logit treat x1 x2 x3 //使用treat作为因变量，其他协变量进行估计得分，这估计的是协变量相同时被处理的概率
```

```
predict pscore, pr
```

实用小代码stata

- ```
count if contact == 1 //统计contact为1的个数
```
- ```
drop if var==. //删除变量的缺失值
```

一些方法

- 证伪实验：证伪实验的目的不是证明某个假设是正确的，而是尝试找到证据来反驳它，证伪实验中，研究者会设计一个实验来检验假设的预测结果。如果实验结果与假设的预测不一致，那么就可以认为该假设被证伪了。例如：如果认为打电话对于02年的选举有影响，那证伪实验就是在98年进行打电话对于选举的影响，如果没有影响，那么就认为打电话对选举有影响（之前得出结论有影响）。