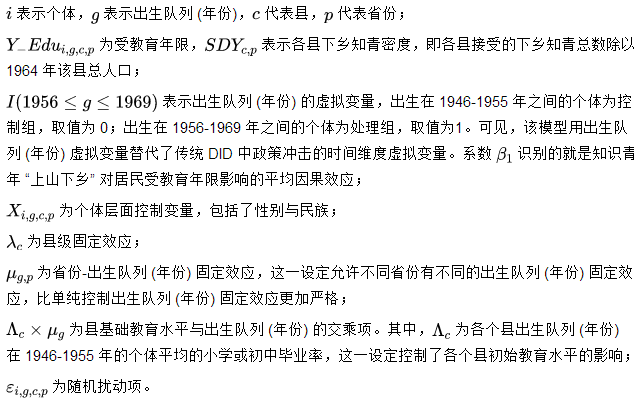
队列DID

Cohort DID（队列DID）是一种巧妙的计量识别策略，常用于评估特殊历史事件对个体和家庭的长期影响。传统 DID 一般有地区和时间两个维度。但是，截面数据由于缺少时间维度，只能由个体的出生队列 (年份) 代替，这也是 “队列 DID” 名称的由来。

2.1 标准模型





2.2 Stata 操作

Standard cohort DID估计

接下来，我们利用 Chen 等 (2020) 提供的数据和代码，来展示队列 DID 的 Stata 操作。

use"C:\Users\林豸\Desktop\03 DID（双重差分）\队列DID\知青下乡数据\census\_1990\_clean.dta"

global var\_abs\_cohort "region1990 prov#year\_birth c.primary\_base#year\_birth c.junior\_base#year\_birth"

//设置控制变量层次，就是同时控制这几个维度效应

region1990 prov表示县级固定效应

prov#year\_birth表示省份和出生控制效应

c.primary\_base#year\_birth表示基础教育和出生控制效应

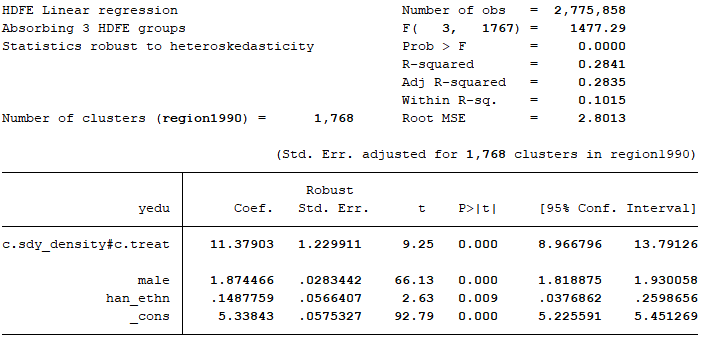
c.junior\_base#year\_birth表示c.junior和出生控制效应

gen treat = inrange(year\_birth,1956,1969) if inrange(year\_birth,1946,1969)

//生成处理组和控制组

reghdfe yedu c.sdy\_density#c.treat male han\_ethn if rural==1, absorb($var\_abs\_cohort) cluster(region1990)

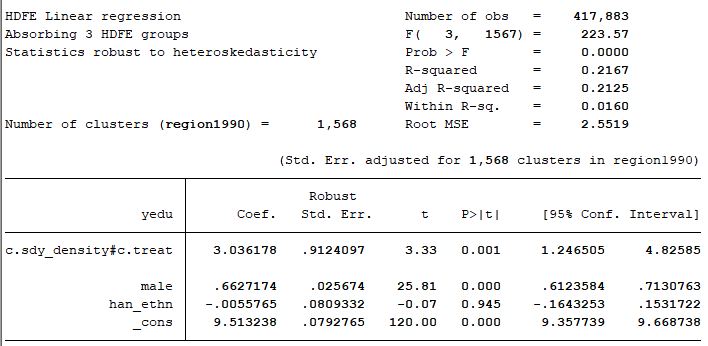
//c.sdy\_density#c.treat人口密度和控制住的交互项，absorb进行固定效应，同时采用region1990进行聚类。



est store m1

//保存回归结果

reghdfe yedu c.sdy\_density#c.treat male han\_ethn if rural==0, absorb($var\_abs\_cohort) cluster(region1990)

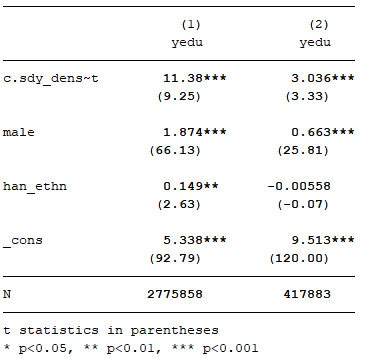


est store m2

//保存回归结果

esttab m1 m2

//合并展示展示回归结果



因为是高维固定效应，所以我们最好使用reghdfe命令，可以看出交互项的系数是3.036，这与原文报告的估计结果是一致的。其他列的估计方式都是类似的，只需更换一下数据或者是被解释变量即可。

Reduced-Form cohort DID动态估计

出生队列维度变异：个体i的出生年份，一个出生年份生成一个队列虚拟变量。从1946到1969共有24年，我们要生成24个队列虚拟变量（以1941-1945年为基准组）。地区维度变异：同上。

Reduced-Form cforvalues y = 1946/1969 {

gen I`y' = sdy\_density\*[year\_birth==`y']

}

cohort DID实际上就是一种动态DID模型，这个名称是follow了Duflo（2001，AER）那篇队列DID的“鼻祖”论文。我们首先需要生成队列虚拟变量和各县接受的下乡知青的人口比例的交互项（24个），将交互项作为解释变量进行回归。交互项的系数识别的就是知识青年上山下乡对特定年份出生的人群的受教育年限的平均因果效应。这些交互项的系数能够直观展示知青上山下乡对农村居民受教育年限的动态影响，这有助于我们检验DID模型的平行趋势假定。

reghdfe yedu **I1946-I1969** male han\_ethn if rural==1, absorb(region1990 prov#year\_birth c.primary\_base\_older#year\_birth c.junior\_base\_older#year\_birth) cluster(region1990)

**I1946-I1969表示从1946年开始到1969年。这个是stata表示方式。**



可以看出从1957年开始创，上山下乡对青年的平均受教育年限就产生了明显的影响。

接下来，我们可以进一步通过直观的图形，呈现交互项的系数的变化（知青上山下乡对农村居民受教育年限的动态影响）。作者使用的是一种较为复杂的绘图方法，有兴趣的朋友可以去看看，在这里我使用的是更加方便快捷的绘图命令coefplot。coefplot命令可以便捷地根据回归结果帮助我们绘制回归系数的取值和置信区间，常用于DID平行趋势检验制图。

coefplot, baselevels keep(I19\*) ///可以看出保存19开头的年份

vertical ///转置图形

coeflabels(I1946=1946 I1947=1947 I1948=1948 I1949=1949 I1950=1950 ///

I1951=1951 I1952=1952 I1953=1953 I1954=1954 I1955=1955 I1956=1956 ///

I1957=1957 I1958=1958 I1959=1959 I1960=1960 I1961=1961 I1962=1962 ///

I1963=1963 I1964=1964 I1965=1965 I1966=1966 I1967=1967 I1968=1968 ///

I1969=1969) ///

yline(0,lwidth(vthin) lpattern(solid) lcolor(teal)) ///设置x和y的属性

xline(10,lwidth(vthin) lpattern(solid) lcolor(teal)) ///

ylabel(-4(2)8,labsize(\*0.85) angle(0)) xlabel(,labsize(\*0.75) angle(45)) ///

ytitle("Coefficients") ///

xtitle("Birth cohort") ///

msymbol(O) msize(small) mcolor(gs1) ///plot样式

addplot(line @b @at,lcolor(gs1) lwidth(medthick)) ///增加点之间的连线

ciopts(recast(rline) lwidth(thin) lpattern(dash) lcolor(gs2)) ///置信区间样式

graphregion(color(white)) //白底

