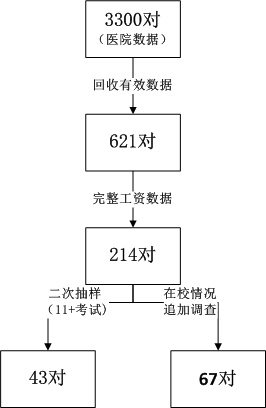
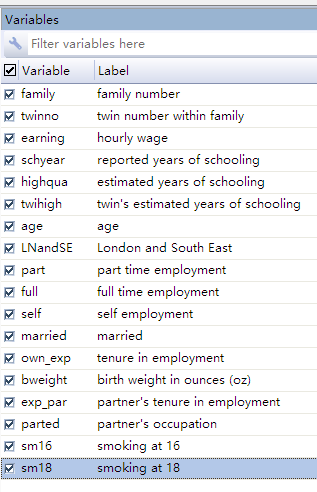
# 文献综述

## 一、主要研究内容

#### 1.样本的选择



#### 2.变量的选择



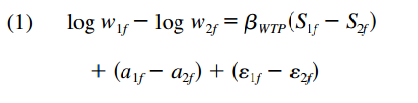
3.主要数据的处理

（1）线性回归

（2）工具变量

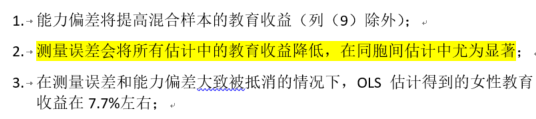
（3）多变量相关性矩阵

微信截图_20161006182048



## 二、验证的已有观点

1.通过同胞间估计修正测量误差、反映能力偏差



2.使用相关性分析，如是否在职、配偶的职位等。我们发现除初次结婚的年龄外，同胞间特征差异与同胞间教育差异无关。

这一模型也应用于出生体重因素的探究上（Ashenfelter and Rouse (1998)没有使用）。由此，我们发现没有迹象表明能力偏差会对同胞间估计结果造成偏差而对混合估计影响更大。因此我们认为能力偏差在同胞间估计中比未控制同胞变量情况下要小。所以，我们找到了至少使得教育收益上限更小的条件正能力偏差。

## 三、创新点

1.对于我们的全部样本，我们拥有早期吸烟行为数据。我们的结果表明吸烟情况更多反映的是家庭背景而非个人折现率。因此，吸烟情况可以被当做可能加剧能力偏差的教育收益相关解释变量

2. 对于更小的样本，我们拥有双胞胎测试和阅读能力成绩。正如其他特征一般，我们发现它们与同胞间估计教育差异相关性不大。在测量能力差异的程度上，这也表明同胞间估计中的能力差异小于家庭间估计中的能力差异；

3.我们拥有双胞胎在读学校及在读班级的信息。这表明了我们的结果不受不同学校和班级质量的影响。

## 四、疑问

#### 1.数据的问题

(1)三个教育年限的选择 schyear，highqua，twihigh

Schyear 定义reported years of schooling

Highqua 定义estimated years of schooling

twihigh 定义twin's estimated years of schooling

三个数据都可以用来表示教育年限，文章中一般用highqua数据，然后用twihigh数据来做修正，往往得出的结果会有差距。所以比较大的疑惑点是这个数据选取的标准是怎么样的，twihigh与highqua具体的差距表现在哪，为何选用的不同数据得出结果差距较大。

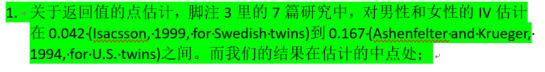
(2)工作状态的选择 part，full，self

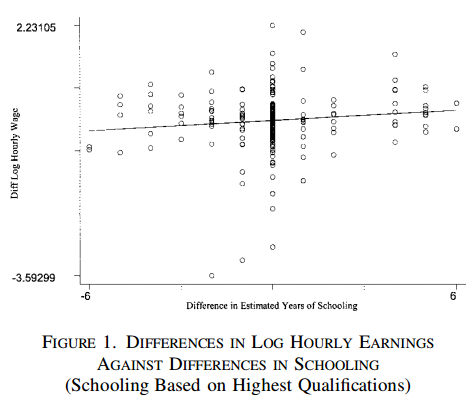
三个变量分别定义工作状态为兼职、全职、个人劳动者

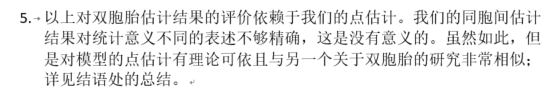
然后有发现把每行三个变量加一起发现有0的结果，说明有不是以上三者工作状态的样本出现，但是调查的样本应该都是已工作的英国同卵双胞胎女性，所以有点疑问

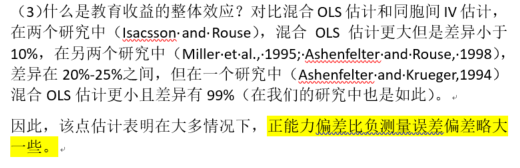
#### 2.数据处理的问题

**（1）点估计（具体如何应用不了解）**

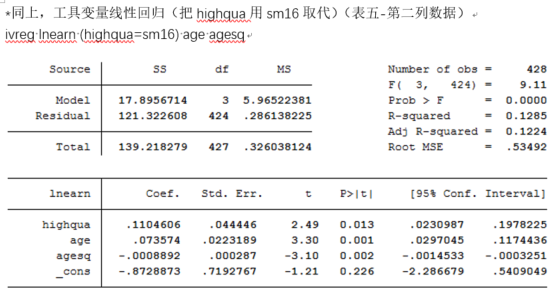




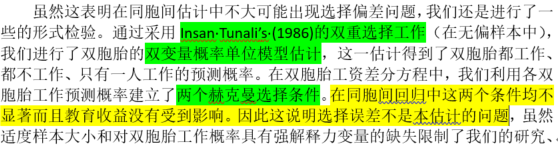




**（2）工具变量（具体instrumented与instruments的变量如何处理不了解）**



**（3）样本选择误差的影响检验**





**（4）为什么会去做一个关于年龄age的二次项回归**

