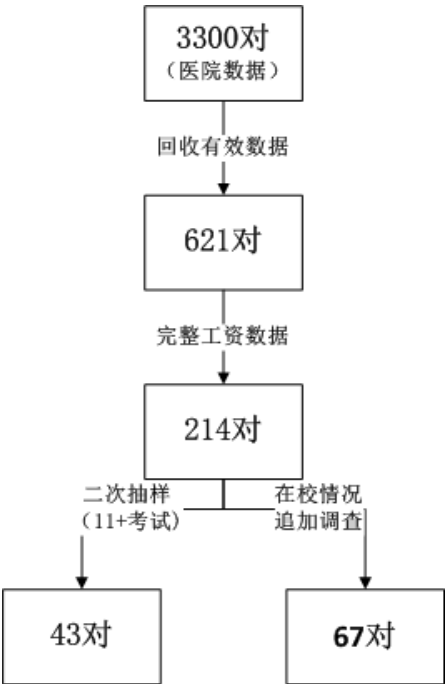


文献综述

一、主要研究内容

1.样本的选择



2.变量的选择

Variables	
Filter variables here	
<input checked="" type="checkbox"/> Variable	Label
<input checked="" type="checkbox"/> family	family number
<input checked="" type="checkbox"/> twinno	twin number within family
<input checked="" type="checkbox"/> earning	hourly wage
<input checked="" type="checkbox"/> schyear	reported years of schooling
<input checked="" type="checkbox"/> highqua	estimated years of schooling
<input checked="" type="checkbox"/> twihigh	twin's estimated years of schooling
<input checked="" type="checkbox"/> age	age
<input checked="" type="checkbox"/> LNandSE	London and South East
<input checked="" type="checkbox"/> part	part time employment
<input checked="" type="checkbox"/> full	full time employment
<input checked="" type="checkbox"/> self	self employment
<input checked="" type="checkbox"/> married	married
<input checked="" type="checkbox"/> own_exp	tenure in employment
<input checked="" type="checkbox"/> bweight	birth weight in ounces (oz)
<input checked="" type="checkbox"/> exp_par	partner's tenure in employment
<input checked="" type="checkbox"/> parted	partner's occupation
<input checked="" type="checkbox"/> sm16	smoking at 16
<input checked="" type="checkbox"/> sm18	smoking at 18

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

$$\log w_{if} = \beta S_{if} + A_{if} + \varepsilon_{if}$$

$$(1) \quad \log w_{1f} - \log w_{2f} = \beta_{WTP}(S_{1f} - S_{2f}) \\ + (a_{1f} - a_{2f}) + (\varepsilon_{1f} - \varepsilon_{2f})$$

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

- 1.→ 能力偏差将提高混合样本的教育收益（列（9）除外）；
- 2.→ 测量误差会将所有估计中的教育收益降低，在同胞间估计中尤为显著；
- 3.→ 在测量误差和能力偏差大致被抵消的情况下，OLS 估计得到的女性教育收益在 7.7%左右；

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，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）

1. 关于返回值的点估计，脚注 3 里的 7 篇研究中，对男性和女性的 IV 估计在 0.042-(Isacsson, 1999, for Swedish twins)到 0.167-(Ashenfelter and Krueger, 1994, for U.S. twins)之间。而我们的结果在估计的中点处：

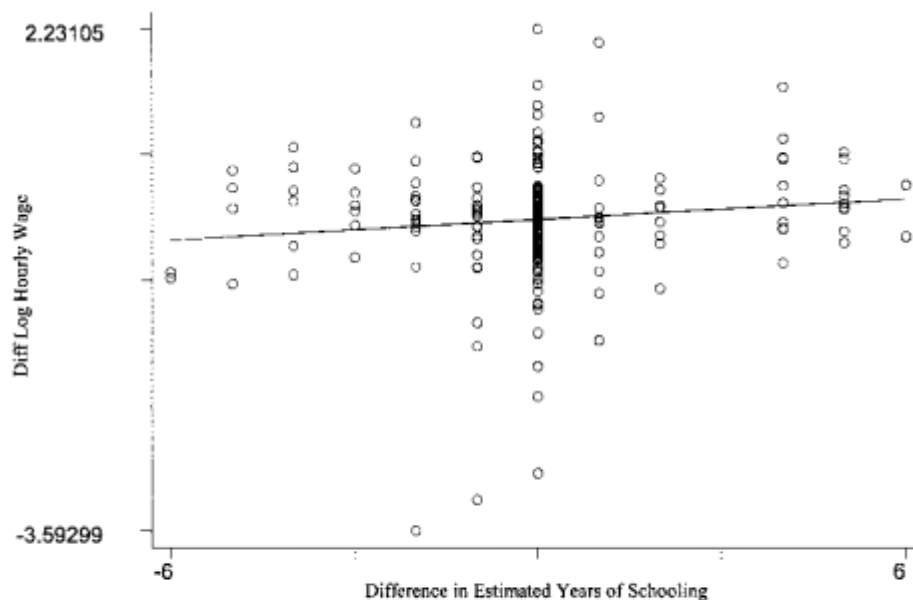


FIGURE 1. DIFFERENCES IN LOG HOURLY EARNINGS
AGAINST DIFFERENCES IN SCHOOLING
(Schooling Based on Highest Qualifications)

5. 以上对双胞胎估计结果的评价依赖于我们的点估计。我们的同胞间估计结果对统计意义不同的表述不够精确，这是没有意义的。虽然如此，但是对模型的点估计有理论可依且与另一个关于双胞胎的研究非常相似；详见结语处的总结。

(3) 什么是教育收益的整体效应？对比混合 OLS 估计和同胞间 IV 估计，在两个研究中 (Isacsson and Rouse)，混合 OLS 估计更大但是差异小于 10%，在另两个研究中 (Miller et al., 1995; Ashenfelter and Rouse, 1998)，差异在 20%-25% 之间，但在一个研究中 (Ashenfelter and Krueger, 1994) 混合 OLS 估计更小且差异有 99% (在我们的研究中也是如此)。

因此，该点估计表明在大多情况下，正能力偏差比负测量误差偏差略大一些。

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

*同上，工具变量线性回归 (把 highqua 用 sm16 取代) (表五-第二列数据)。

`ivreg lnearn (highqua=sm16) age agesq`

Source	SS	df	MS	Number of obs = 428		
Model	17.8956714	3	5.96522381	F(3, 424) = 9.11		
Residual	121.322608	424	.286138225	Prob > F = 0.0000		
Total	139.218279	427	.326038124	R-squared = 0.1285		
				Adj R-squared = 0.1224		
				Root MSE = .53492		

lnearn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
highqua	.1104606	.044446	2.49	0.013	.0230987	.1978225
age	.073574	.0223189	3.30	0.001	.0297045	.1174436
agesq	-.0008892	.000287	-3.10	0.002	-.0014533	-.0003251
_cons	-.8728873	.7192767	-1.21	0.226	-2.286679	.5409049

(3) 样本选择误差的影响检验

虽然这表明在同胞间估计中不大可能出现选择偏差问题，我们还是进行了一些的形式检验。通过采用 Insan Tunali's (1986) 的双重选择工作 (在无偏样本中)，我们进行了双胞胎的双变量概率单位模型估计，这一估计得到了双胞胎都工作、都不工作、只有一人工作的预测概率。在双胞胎工资差分方程中，我们利用各双胞胎工作预测概率建立了两个赫克曼选择条件。在同胞间回归中这两个条件均不显著而且教育收益没有受到影响。因此这说明选择误差不是本估计的问题，虽然适度样本大小和对双胞胎工作概率具有强解释力变量的缺失限制了我们的研究、方面的选择偏差可能也会影响我们的混合统计。所以我们将我们的混合统计回归与传统的赫克曼误差修正模型 (使用相关方程中孩子及丈夫职业的数据) 进行比较，发现没有选择误差显著影响本估计的任何迹象。

(4) 为什么会去做一个关于年龄 age 的二次项回归

*定义每小时工资对数和年龄平方、

gen lnearn=ln(earning)、

、

gen agesq=age*age、

、

*做线性回归、工资对数关于教育年限、年龄的二元二次项回归（表 2-第二列数据）、

reg lnearn highqua age agesq、

Source	SS	df	MS	Number of obs = 428		
Model	20.7258534	3	6.9086178	F(3, 424) = 24.72		
Residual	118.492426	424	.279463268	Prob > F = 0.0000		
Total	139.218279	427	.326038124	R-squared = 0.1489		
				Adj R-squared = 0.1429		
				Root MSE = .52864		

lnearn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
highqua	.0767543	.0105917	7.25	0.000	.0559355	.0975731
age	.0778154	.0213949	3.64	0.000	.0357622	.1198687
agesq	-.0009675	.0002658	-3.64	0.000	-.0014899	-.0004451
_cons	-.4282208	.4347756	-0.98	0.325	-1.282805	.4263631