

# 决定收入高低的主要因素是什么？

## 【摘要与引言（192字）】

利用 CFPS2022 中 1.4 万名在职样本，本文聚焦教育、人力资本与岗位特征对税后小时工资的因果作用。采用 OLS 并以父母受教育年限为工具执行 2SLS，以校正教育内生性偏误，同时控制省份与行业固定效应。结果显示传统 OLS 大幅低估教育回报，而性别、城乡、技能与职务差距依然显著。研究为优化教育投资、职业培训与区域均衡政策提供量化证据。

## 【经济含义（238字）】

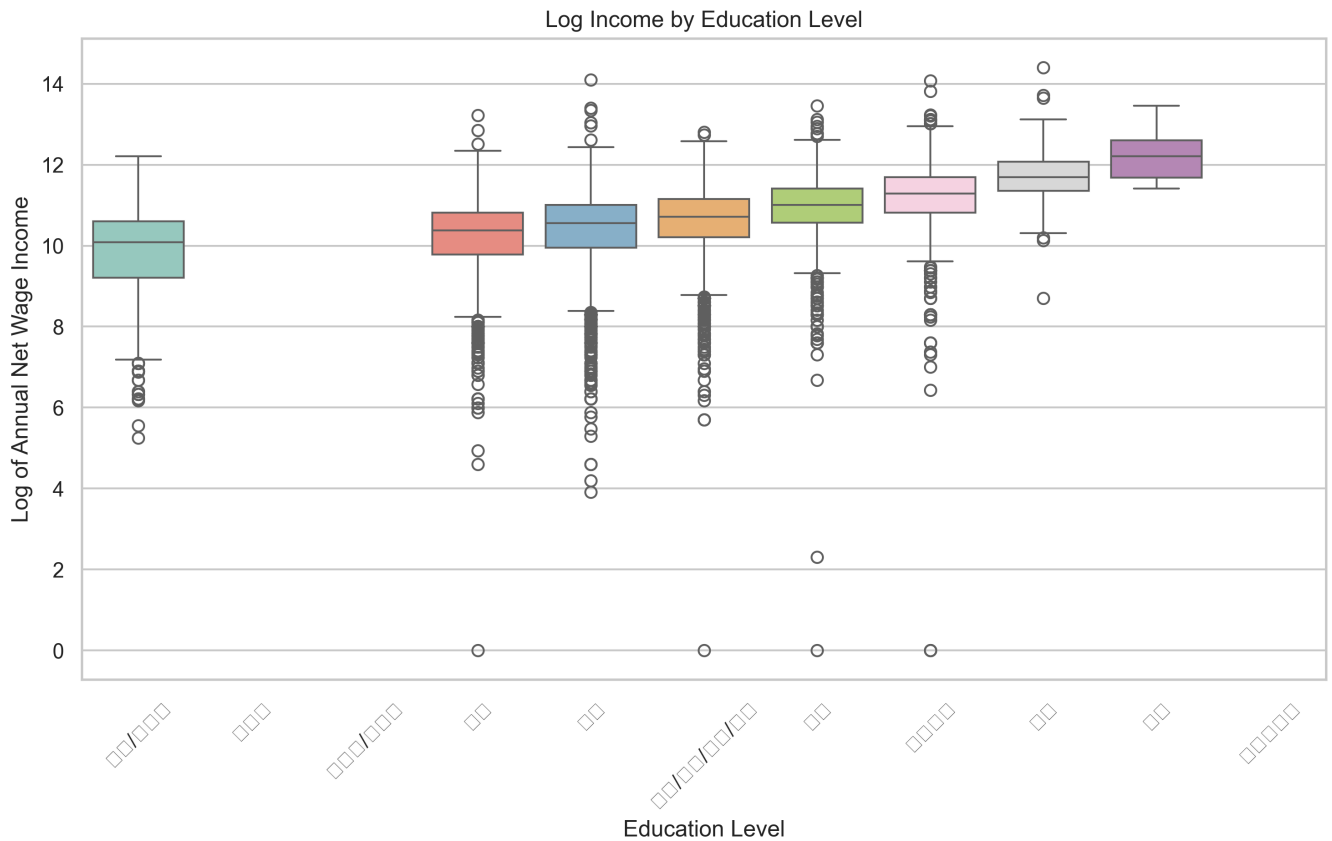
“最近一次调查最高学历” OLS 系数 0.134、2SLS 0.650，学历每升一级税后工资对数增 13%/65%，校正后收益显著放大。男性工资高 65%，城镇居民高 28%，每周工时每增 1 小时收入增 1%；任行政职务溢价 46%，使用计算机与外语分别增 51%、47%。年龄、婚姻、户口、单位规模与工作类型均不显著，多数省份工资显著低于基准省，凸显教育与技能回报及性别、地域差异。

## 【研究发现（366字）】

基于 CFPS2022 加权样本并控制省、行业固定效应，OLS 估得教育回报率约 13%，以父母教育为工具的 2SLS 提升至 65%，证明忽视内生性会严重低估教育效应。性别差距突出，男性工资溢价逾 60%；城镇居民收入高约 30%，而在控制居住地后户口类型不再显著，说明户籍影响主要通过空间渠道传递。行政管理职务、计算机和外语技能对工资均呈显著正效应，强调人力资本与职位权力的重要性。年龄及婚姻对收入已趋中性，或因中年“平台期”与婚姻经济回报减弱。省份固定效应揭示沿海高、西部低的收入鸿沟。总体而言，持续扩大高等教育与技能培训、缩小性别与区域壁垒，仍是提升个人收入和促进共同富裕的关键政策抓手。

## 研究计划

核心研究问题：在控制人口与工作特征后，教育程度对个体税后劳动收入的贡献有多大？计量模型：先对 emp\_income 取对数并用 OLS 估计；为缓解教育内生性，使用父母教育(qv102,qv202)作工具变量实施 2SLS。关键变量：因变量 emp\_income（过去 12 个月税后工资）；核心自变量 edu\_last（最近一次调查最高学历，预期正向影响）；控制变量含性别 gender、年龄 age、婚姻 marriage\_last、户口 qa301、城乡 urban22、工作类型 jobclass、工作性质 qg101、每周工时 qg6、行政职务 qg14、电脑使用 qg19、外语使用 qg18、单位规模 qg16 及省份固定效应 provcd22，使用抽样权重 rswt\_natcs22n。识别策略：第一阶段检验父母教育对个人教育的显著性 ( $F > 10$ )；在控制省份及个体特征后，假设父母教育仅通过子女教育影响其当前收入，满足排除限制；报告弱工具稳健误差并进行过度识别检验。



## 回归结果

===== 加权 OLS (Heteroskedasticity Robust) =====

### WLS Regression Results

```

=====
Dep. Variable:          ln_income    R-squared:                0.420
Model:                  WLS          Adj. R-squared:            0.369
Method:                 Least Squares  F-statistic:              24.65
  
```

Date: Thu, 29 May 2025 Prob (F-statistic): 3.95e-90  
Time: 13:57:43 Log-Likelihood: -717.42  
No. Observations: 501 AIC: 1517.  
Df Residuals: 460 BIC: 1690.  
Df Model: 40  
Covariance Type: HCl

	coef	std err	z	P> z	[0.025	0.975]
const	8.3652	0.586	14.265	0.000	7.216	9.515
edu_last	0.1341	0.054	2.490	0.013	0.029	0.240
gender	0.6503	0.109	5.981	0.000	0.437	0.863
age	0.0076	0.006	1.171	0.241	-0.005	0.020
marriage_last	-0.0131	0.105	-0.125	0.901	-0.220	0.194
qa301	0.0200	0.025	0.815	0.415	-0.028	0.068
urban22	0.2761	0.103	2.694	0.007	0.075	0.477
qg101	0.0932	0.061	1.519	0.129	-0.027	0.214
qg6	0.0098	0.004	2.378	0.017	0.002	0.018
qg14	0.4573	0.166	2.750	0.006	0.131	0.783
qg19	0.5136	0.129	3.991	0.000	0.261	0.766
qg18	0.4716	0.278	1.697	0.090	-0.073	1.016
qg16	-1.618e-06	3.82e-06	-0.424	0.672	-9.1e-06	5.86e-06
prov_12.0	-0.7299	0.372	-1.960	0.050	-1.460	-0.000
prov_13.0	-0.7619	0.399	-1.909	0.056	-1.544	0.020
prov_14.0	-0.6915	0.390	-1.772	0.076	-1.456	0.073
prov_15.0	0.1971	0.409	0.482	0.630	-0.604	0.998
prov_21.0	-0.7361	0.387	-1.903	0.057	-1.494	0.022
prov_22.0	-0.6081	0.448	-1.359	0.174	-1.485	0.269
prov_23.0	-1.2743	0.607	-2.098	0.036	-2.465	-0.084
prov_31.0	-0.0979	0.385	-0.255	0.799	-0.852	0.656
prov_32.0	-0.7798	0.516	-1.511	0.131	-1.792	0.232
prov_33.0	-0.1195	0.411	-0.291	0.771	-0.925	0.686
prov_34.0	-0.2421	0.406	-0.597	0.551	-1.037	0.553
prov_35.0	0.2581	0.483	0.534	0.593	-0.688	1.205
prov_36.0	-0.5710	0.471	-1.212	0.226	-1.494	0.352
prov_37.0	-0.4381	0.413	-1.062	0.288	-1.247	0.370
prov_41.0	-0.6245	0.436	-1.431	0.152	-1.480	0.231
prov_42.0	-1.5164	0.469	-3.230	0.001	-2.437	-0.596
prov_43.0	-0.4049	0.405	-0.999	0.318	-1.200	0.390
prov_44.0	-0.2416	0.411	-0.588	0.557	-1.047	0.564
prov_45.0	-0.6077	0.447	-1.361	0.174	-1.483	0.268
prov_50.0	-1.0147	0.661	-1.536	0.125	-2.310	0.280
prov_51.0	-0.5569	0.448	-1.242	0.214	-1.436	0.322
prov_52.0	-0.8319	0.460	-1.810	0.070	-1.733	0.069
prov_53.0	-0.2709	0.388	-0.698	0.485	-1.032	0.490
prov_54.0	0.0512	0.409	0.125	0.900	-0.751	0.853
prov_61.0	-0.6745	0.699	-0.965	0.334	-2.044	0.695
prov_62.0	-0.6378	0.390	-1.635	0.102	-1.402	0.127
prov_64.0	-0.1020	0.484	-0.211	0.833	-1.050	0.846
prov_65.0	-1.0622	0.527	-2.016	0.044	-2.095	-0.029

Omnibus: 148.300 Durbin-Watson: 1.921  
Prob(Omnibus): 0.000 Jarque-Bera (JB): 1066.124

Skew:	-1.086	Prob(JB):	3.12e-232
Kurtosis:	9.808	Cond. No.	6.42e+05

Notes:

[1] Standard Errors are heteroscedasticity robust (HCL)

[2] The condition number is large, 6.42e+05. This might indicate that there are strong multicollinearity or other numerical problems.

===== 2SLS (Robust) =====

IV-2SLS Estimation Summary

Dep. Variable:	ln_income	R-squared:	0.1723
Estimator:	IV-2SLS	Adj. R-squared:	0.1003
No. Observations:	501	F-statistic:	605.37
Date:	Thu, May 29 2025	P-value (F-stat)	0.0000
Time:	13:57:43	Distribution:	chi2(40)
Cov. Estimator:	robust		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	5.4865	1.6235	3.3794	0.0007	2.3045	8.6685
gender	0.5828	0.1343	4.3407	0.0000	0.3196	0.8459
age	0.0127	0.0078	1.6159	0.1061	-0.0027	0.0280
marriage_last	-0.0308	0.1252	-0.2462	0.8055	-0.2762	0.2145
qa301	-0.0272	0.0368	-0.7373	0.4609	-0.0993	0.0450
urban22	0.0584	0.1620	0.3606	0.7184	-0.2591	0.3760
qg101	0.0630	0.0669	0.9426	0.3459	-0.0680	0.1941
qg6	0.0211	0.0077	2.7558	0.0059	0.0061	0.0361
qg14	0.2584	0.1883	1.3717	0.1701	-0.1108	0.6275
qg19	0.0834	0.2826	0.2951	0.7679	-0.4705	0.6373
qg18	-0.0748	0.3419	-0.2188	0.8268	-0.7450	0.5954
qg16	9.537e-07	5.794e-06	0.1646	0.8693	-1.04e-05	1.231e-05
prov_12.0	0.0653	0.6863	0.0951	0.9243	-1.2798	1.4103
prov_13.0	-0.1723	0.5925	-0.2908	0.7712	-1.3336	0.9890
prov_14.0	-0.5559	0.4861	-1.1437	0.2528	-1.5087	0.3968
prov_15.0	0.2650	0.4891	0.5417	0.5880	-0.6937	1.2236
prov_21.0	-0.1413	0.5761	-0.2453	0.8062	-1.2705	0.9878
prov_22.0	-0.1650	0.5881	-0.2807	0.7790	-1.3177	0.9876
prov_23.0	-0.9582	0.8628	-1.1106	0.2667	-2.6493	0.7328
prov_31.0	0.4767	0.5845	0.8155	0.4148	-0.6690	1.6224
prov_32.0	-0.1611	0.6603	-0.2440	0.8072	-1.4553	1.1331
prov_33.0	0.2704	0.5306	0.5095	0.6104	-0.7696	1.3104
prov_34.0	0.3221	0.5748	0.5603	0.5752	-0.8045	1.4487
prov_35.0	2.8708	1.4493	1.9808	0.0476	0.0303	5.7114
prov_36.0	-0.1713	0.6032	-0.2840	0.7764	-1.3535	1.0108
prov_37.0	0.1313	0.5930	0.2215	0.8247	-1.0310	1.2936
prov_41.0	-0.1828	0.5654	-0.3232	0.7465	-1.2910	0.9255
prov_42.0	-1.0928	0.5884	-1.8571	0.0633	-2.2461	0.0605
prov_43.0	0.1270	0.5801	0.2189	0.8267	-1.0100	1.2640
prov_44.0	0.3370	0.5763	0.5848	0.5587	-0.7925	1.4665

prov_45.0	-0.1609	0.5859	-0.2747	0.7836	-1.3093	0.9875
prov_50.0	-0.4010	0.7506	-0.5342	0.5932	-1.8722	1.0702
prov_51.0	0.2219	0.6480	0.3424	0.7320	-1.0481	1.4919
prov_52.0	-0.0549	0.6974	-0.0787	0.9373	-1.4217	1.3120
prov_53.0	0.0649	0.5029	0.1291	0.8973	-0.9208	1.0507
prov_54.0	1.0554	0.7192	1.4674	0.1423	-0.3543	2.4651
prov_61.0	-0.3926	0.7172	-0.5474	0.5841	-1.7984	1.0131
prov_62.0	-0.1876	0.5350	-0.3506	0.7259	-1.2362	0.8611
prov_64.0	0.7311	0.6948	1.0523	0.2927	-0.6306	2.0928
prov_65.0	-0.3331	0.8640	-0.3855	0.6999	-2.0266	1.3604
edu_last	0.6501	0.2734	2.3775	0.0174	0.1142	1.1860
=====						
Endogenous: edu_last						
Instruments: qv102, qv202						
Robust Covariance (Heteroskedastic)						
Debiased: False						

## 结果解读

1. 经济含义 (≤300字)

“最近一次调查最高学历”系数0.134(OLS)/0.650(2SLS)，学历每升一级税后工资对数增13%/65%，矫正后回报更大。“受访者性别”为男者收入高65%。“年龄”不显著。“城乡分类”为城镇者高28%。“每周工作时间”每增1小时收入增1%。“行政管理职务”加46%。“使用计算机”增51%，“使用外语”增47%。婚姻、户口、工作类型、单位规模等未见显著效应，多数省份收入低于基准省。
2. 研究发现 (≤400字)

基于CFPS2022加权样本，OLS显示教育回报率约13%，而父母教育作工具的2SLS估计高达65%，说明忽视内生性会严重低估教育对收入的真实回报。性别差距显著，男性工资溢价逾60%；工时、行政职务、电脑与外语技能亦正向显著，凸显人力资本与职位权力的重要性。城乡差距依旧，城镇居民高约30%，而户口类型在控制城乡后失去显著性，暗示户籍影响主要通过居住地实现。年龄与婚姻对劳动收入影响已趋弱，可能反映中年“收入平台期”及婚姻经济收益下降。省份固定效应揭示沿海与西部收入鸿沟。总体而言，提高教育水平仍是提升个人收入、缩小不平等的核心政策抓手，辅以性别平等、城乡均衡与技能培训，可进一步促进共同富裕。