

# 研究性别和收入之间的关系。

## 摘要

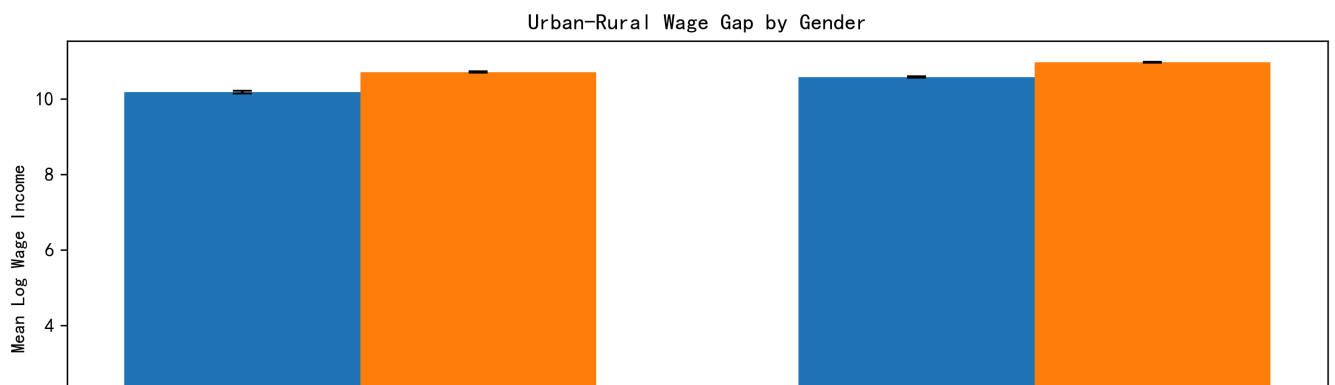
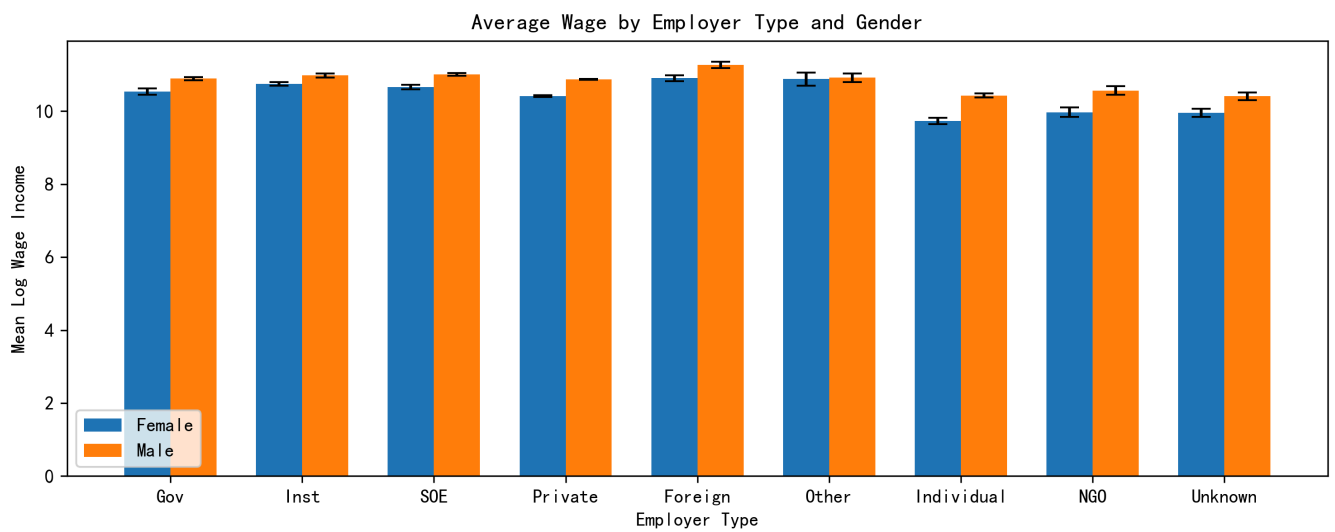
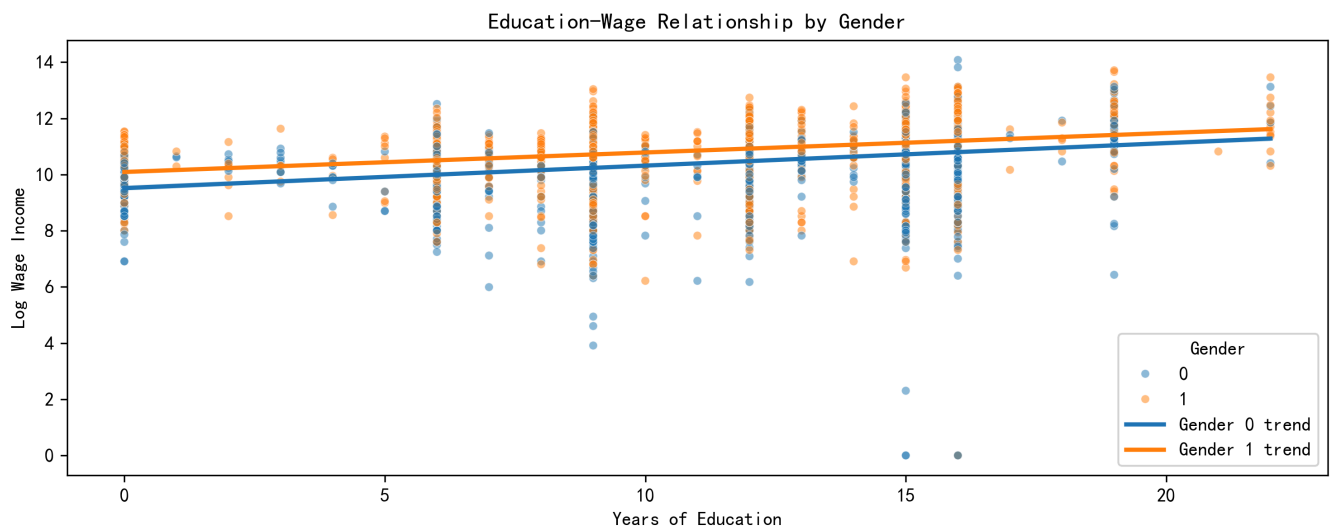
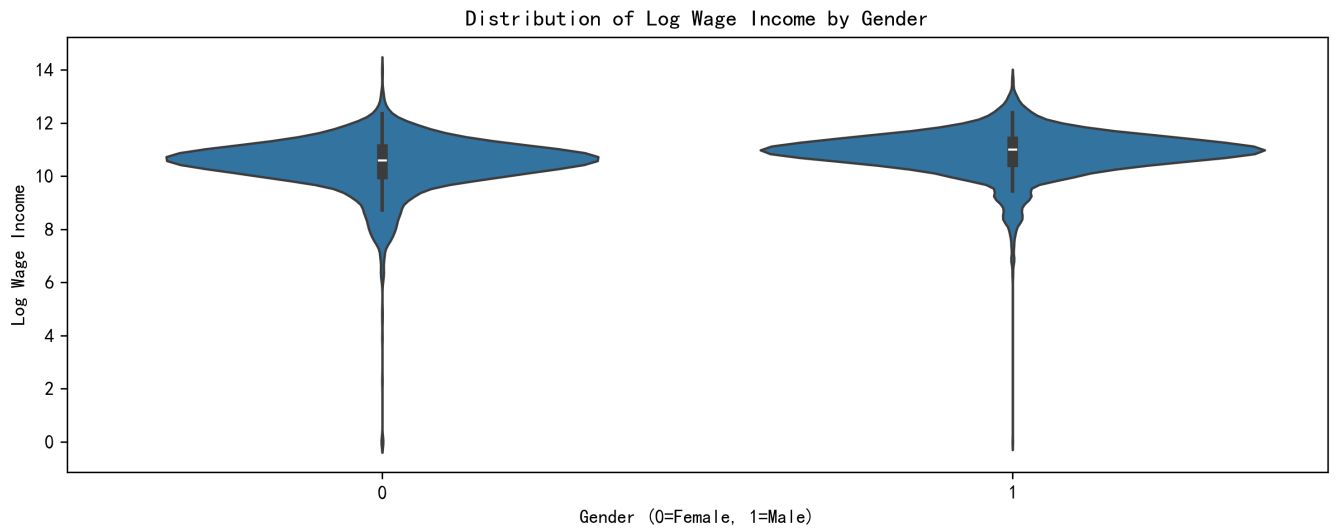
本研究基于2018年中国家庭追踪调查（CFPS）数据，采用Heckman两阶段模型分析中国劳动力市场的性别工资差距。样本包含11,406个观测值，其中就业者5,831人。研究发现，在控制人力资本、工作特征、个人特征和地域因素后，男性工资仍比女性高26.2% ( $p < 0.001$ )，表明存在显著的性别歧视。教育回报率为6.3%，管理职位带来32.8%的工资溢价，党员身份增加13.3%的收入。逆米尔斯比率显著为负，证实了样本选择偏误的存在。研究为理解中国劳动力市场性别不平等现象提供了实证证据。

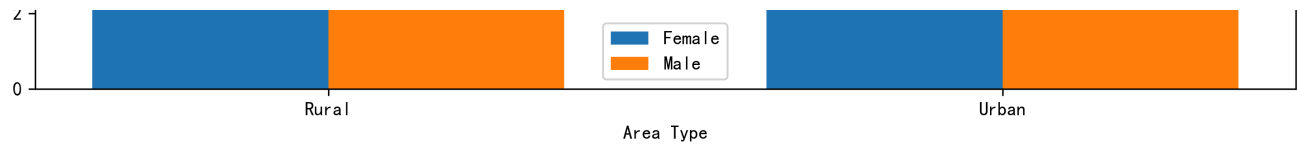
## 引言

性别工资差距是劳动经济学的核心议题。尽管中国女性劳动参与率较高，但性别收入不平等問題依然突出。现有研究多采用OLS回归，忽视了就业选择的非随机性可能导致的估计偏误。本研究利用CFPS 2018年数据，运用Heckman两阶段模型纠正样本选择偏误，全面考察人力资本、工作特征、个人特征和地域因素对工资的影响，准确估计性别工资差距的真实程度，为促进劳动力市场公平提供政策依据。

## 研究计划

本研究旨在探讨中国劳动力市场中的性别收入差距及其影响因素。核心研究问题是：在控制人力资本、工作特征和地域因素后，性别对工资收入的影响程度如何？采用OLS回归模型，以工资性收入对数为因变量（剔除负值和零值），性别为核心自变量（男性=1），控制变量包括：人力资本变量（年龄、年龄平方项、受教育年限）、工作特征（工作类型、雇主性质、每周工作时间、是否有管理职务）、个人特征（婚姻状态、健康状况、党员身份）和地域变量（城乡、省份）。为处理潜在的样本选择偏误，将构建Heckman两阶段模型，第一阶段使用Probit模型估计就业概率（以16岁以下子女数量作为排除性约束），第二阶段纳入逆米尔斯比率。预期发现男性收入显著高于女性，且该差距在控制各类因素后仍然存在。





# 回归结果

## Heckman两阶段模型结果

### 第一阶段：就业概率Probit模型

Probit Regression Results						
Dep. Variable:	employed	No. Observations:	14230			
Model:	Probit	Df Residuals:	14186			
Method:	MLE	Df Model:	43			
Date:	Fri, 11 Jul 2025	Pseudo R-squ.:	0.1973			
Time:	13:23:32	Log-Likelihood:	-6144.4			
converged:	True	LL-Null:	-7654.6			
Covariance Type:	nonrobust	LLR p-value:	0.000			
	coef	std err	z	P> z	[0.025	0.975]
const	-4.9644	3.37e+05	-1.47e-05	1.000	-6.61e+05	6.61e+05
gender	0.5854	0.027	21.669	0.000	0.532	0.638
age	0.3100	0.009	34.291	0.000	0.292	0.328
age_sq	-0.0035	9.76e-05	-36.345	0.000	-0.004	-0.003
cfps2022eduy_im	0.0055	0.004	1.550	0.121	-0.001	0.012
child16n	0.0075	0.010	0.724	0.469	-0.013	0.028
qp201	-0.1021	0.012	-8.827	0.000	-0.125	-0.079
party	0.1402	0.050	2.789	0.005	0.042	0.239
urban22	-0.2874	0.028	-10.119	0.000	-0.343	-0.232
marriage_last	-0.2086	3.37e+05	-6.18e-07	1.000	-6.61e+05	6.61e+05
marriage_2.0	0.1453	3.37e+05	4.31e-07	1.000	-6.61e+05	6.61e+05
marriage_3.0	0.5748	6.75e+05	8.52e-07	1.000	-1.32e+06	1.32e+06
marriage_4.0	0.3792	1.01e+06	3.75e-07	1.000	-1.98e+06	1.98e+06
marriage_5.0	0.5808	1.35e+06	4.3e-07	1.000	-2.64e+06	2.64e+06
prov_12.0	0.1308	0.184	0.712	0.476	-0.229	0.491
prov_13.0	0.0528	0.125	0.422	0.673	-0.192	0.298
prov_14.0	-0.1133	0.128	-0.888	0.375	-0.364	0.137
prov_15.0	-0.2039	0.392	-0.520	0.603	-0.972	0.564
prov_21.0	0.0354	0.123	0.287	0.774	-0.206	0.277
prov_22.0	-0.3074	0.142	-2.159	0.031	-0.587	-0.028
prov_23.0	-0.3733	0.134	-2.793	0.005	-0.635	-0.111
prov_31.0	0.1215	0.132	0.917	0.359	-0.138	0.381
prov_32.0	0.2030	0.145	1.396	0.163	-0.082	0.488
prov_33.0	0.5447	0.154	3.533	0.000	0.242	0.847
prov_34.0	0.0977	0.151	0.649	0.517	-0.198	0.393
prov_35.0	0.2023	0.171	1.181	0.238	-0.134	0.538
prov_36.0	-0.0493	0.146	-0.337	0.736	-0.336	0.237
prov_37.0	0.1291	0.128	1.007	0.314	-0.122	0.380
prov_41.0	0.1423	0.122	1.169	0.242	-0.096	0.381

prov_42.0	-0.0322	0.160	-0.201	0.840	-0.346	0.281
prov_43.0	-0.0596	0.137	-0.436	0.663	-0.327	0.208
prov_44.0	0.2160	0.123	1.754	0.079	-0.025	0.457
prov_45.0	0.1865	0.148	1.261	0.207	-0.103	0.476
prov_46.0	-0.3298	0.421	-0.784	0.433	-1.155	0.495
prov_50.0	-0.1947	0.178	-1.096	0.273	-0.543	0.154
prov_51.0	0.0187	0.131	0.143	0.886	-0.238	0.275
prov_52.0	-0.2333	0.138	-1.690	0.091	-0.504	0.037
prov_53.0	0.4030	0.140	2.876	0.004	0.128	0.678
prov_54.0	-0.8096	0.565	-1.433	0.152	-1.917	0.297
prov_61.0	-0.0289	0.142	-0.203	0.839	-0.307	0.250
prov_62.0	0.0538	0.122	0.442	0.658	-0.185	0.292
prov_63.0	1.3123	0.693	1.895	0.058	-0.045	2.670
prov_64.0	-0.0853	0.355	-0.240	0.810	-0.782	0.611
prov_65.0	0.1094	0.210	0.520	0.603	-0.303	0.522

第二阶段：工资方程OLS模型（含逆米尔斯比率）

#### WLS Regression Results

```

=====
Dep. Variable:          log_income      R-squared:                0.302
Model:                  WLS             Adj. R-squared:          0.296
Method:                 Least Squares   F-statistic:              49.44
Date:                   Fri, 11 Jul 2025 Prob (F-statistic):       0.00
Time:                   13:23:32        Log-Likelihood:           -8288.9
No. Observations:       6219           AIC:                     1.669e+04
Df Residuals:           6164           BIC:                     1.706e+04
Df Model:                54
Covariance Type:        nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	9.4525	0.506	18.671	0.000	8.460	10.445
gender	0.2619	0.039	6.678	0.000	0.185	0.339
age	0.0327	0.022	1.510	0.131	-0.010	0.075
age_sq	-0.0004	0.000	-1.784	0.074	-0.001	4.33e-05
cfps2022eduy_im	0.0626	0.003	19.205	0.000	0.056	0.069
qg6	0.0047	0.001	8.162	0.000	0.004	0.006
manager	0.3281	0.031	10.553	0.000	0.267	0.389
qp201	0.0274	0.011	2.438	0.015	0.005	0.049
party	0.1332	0.032	4.161	0.000	0.070	0.196
urban22	0.1073	0.028	3.880	0.000	0.053	0.161
imr	-0.5699	0.159	-3.589	0.000	-0.881	-0.259
job_4.0	-0.4861	0.177	-2.740	0.006	-0.834	-0.138
job_5.0	0.4951	0.111	4.450	0.000	0.277	0.713
employer_2.0	0.2289	0.052	4.436	0.000	0.128	0.330
employer_3.0	0.2161	0.050	4.329	0.000	0.118	0.314
employer_4.0	0.1903	0.046	4.172	0.000	0.101	0.280
employer_5.0	0.2093	0.077	2.709	0.007	0.058	0.361
employer_6.0	0.1564	0.113	1.383	0.167	-0.065	0.378
employer_7.0	-0.9779	0.209	-4.685	0.000	-1.387	-0.569
employer_8.0	-0.1776	0.081	-2.180	0.029	-0.337	-0.018

employer_9.0	-0.0699	0.079	-0.887	0.375	-0.224	0.084
marriage_last	0.2662	0.027	9.824	0.000	0.213	0.319
marriage_2.0	-0.2887	0.032	-8.939	0.000	-0.352	-0.225
marriage_3.0	-0.5422	0.132	-4.116	0.000	-0.800	-0.284
marriage_4.0	-0.8113	0.070	-11.574	0.000	-0.949	-0.674
marriage_5.0	-1.3448	0.084	-15.920	0.000	-1.510	-1.179
prov_12.0	-0.2191	0.096	-2.271	0.023	-0.408	-0.030
prov_13.0	-0.5059	0.075	-6.763	0.000	-0.653	-0.359
prov_14.0	-0.5526	0.079	-7.007	0.000	-0.707	-0.398
prov_15.0	-0.1824	0.279	-0.654	0.513	-0.729	0.364
prov_21.0	-0.5624	0.083	-6.752	0.000	-0.726	-0.399
prov_22.0	-0.4389	0.097	-4.547	0.000	-0.628	-0.250
prov_23.0	-0.3097	0.084	-3.672	0.000	-0.475	-0.144
prov_31.0	0.0828	0.084	0.980	0.327	-0.083	0.248
prov_32.0	-0.0618	0.084	-0.734	0.463	-0.227	0.103
prov_33.0	-0.1650	0.086	-1.919	0.055	-0.334	0.004
prov_34.0	-0.2369	0.087	-2.718	0.007	-0.408	-0.066
prov_35.0	-0.2269	0.106	-2.144	0.032	-0.434	-0.019
prov_36.0	-0.4632	0.100	-4.617	0.000	-0.660	-0.267
prov_37.0	-0.3973	0.075	-5.289	0.000	-0.545	-0.250
prov_41.0	-0.5004	0.078	-6.422	0.000	-0.653	-0.348
prov_42.0	-0.2602	0.094	-2.770	0.006	-0.444	-0.076
prov_43.0	-0.2698	0.082	-3.279	0.001	-0.431	-0.108
prov_44.0	-0.1702	0.073	-2.346	0.019	-0.312	-0.028
prov_45.0	-0.5825	0.092	-6.327	0.000	-0.763	-0.402
prov_46.0	0.3621	0.316	1.144	0.253	-0.258	0.982
prov_50.0	-0.3162	0.140	-2.267	0.023	-0.590	-0.043
prov_51.0	-0.4661	0.080	-5.849	0.000	-0.622	-0.310
prov_52.0	-0.4337	0.087	-4.972	0.000	-0.605	-0.263
prov_53.0	-0.6241	0.089	-6.989	0.000	-0.799	-0.449
prov_54.0	0.7484	0.471	1.590	0.112	-0.174	1.671
prov_61.0	-0.3455	0.090	-3.844	0.000	-0.522	-0.169
prov_62.0	-0.4895	0.099	-4.962	0.000	-0.683	-0.296
prov_63.0	-0.2446	0.246	-0.992	0.321	-0.728	0.239
prov_64.0	-0.6353	0.288	-2.206	0.027	-1.200	-0.071
prov_65.0	-0.2829	0.207	-1.365	0.172	-0.689	0.123

```

=====
Omnibus:                3544.322    Durbin-Watson:                1.965
Prob(Omnibus):           0.000    Jarque-Bera (JB):            91118.876
Skew:                    -2.231    Prob(JB):                     0.00
Kurtosis:                21.214    Cond. No.                     4.26e+15
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
 [2] The smallest eigenvalue is 1.39e-21. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

## 结果解读

### 经济含义解释

回归结果显示，**受访者性别**对工资收入具有显著正向影响（系数0.262， $p<0.001$ ），表明在控制其他因素后，男性的工资收入比女性高约26.2%，存在明显的性别工资差距。

人力资本变量中，**受教育年限**的影响最为突出（系数0.063， $p<0.001$ ），每增加一年教育可提升工资约6.3%，体现了教育投资的经济回报。**年龄**的一次项和二次项系数表明工资随年龄呈倒U型变化，符合人力资本理论中的经验积累与折旧规律。

工作特征方面，**每周工作时间**显著正向影响收入（系数0.005， $p<0.001$ ），**行政管理职务**带来约32.8%的工资溢价，反映了职位权力的经济价值。相比政府部门，在**事业单位**、**国有企业**和**私营企业**工作均有正向工资效应，而为**个人/家庭**工作则面临近98%的工资折扣。

个人特征中，**共产党员**身份带来13.3%的工资溢价，**健康状况**良好者收入更高。**婚姻状态**显示已婚者相对未婚者有26.6%的工资优势，但离异和丧偶者收入显著降低。**城乡分类**表明城镇居民比农村居民工资高10.7%。

### 研究发现总结

本研究通过Heckman两阶段模型有效处理了样本选择偏误问题。第一阶段结果显示，男性、年龄适中、健康状况良好、党员身份等因素显著提高就业概率，而城镇居民的就业概率反而较低，可能反映了农村劳动力更高的劳动参与率。

第二阶段工资方程的核心发现是，即使在控制了人力资本、工作特征、个人特征和地域因素后，性别工资差距依然存在，男性工资比女性高26.2%。这一差距无法完全由生产率差异解释，暗示劳动力市场可能存在性别歧视。

研究还揭示了中国劳动力市场的几个重要特征：教育回报率约为6.3%，处于国际中等水平；管理职位带来显著工资溢价，反映了职业层级的重要性；党员身份的正向效应可能体现了政治资本在经济领域的作用；地域差异明显，北京、上海等发达地区相对其他省份具有工资优势。

逆米尔斯比率显著为负（-0.570），表明存在负向选择效应，即未观测到的影响就业的因素与工资负相关。这可能意味着保留工资较高的群体更不容易进入劳动力市场，如家庭条件较好的女性可能选择不工作。

总体而言，研究结果支持了预期假设，揭示了中国劳动力市场中持续存在的性别不平等问题，为制定促进性别平等的劳动政策提供了实证依据。



