

# 劳动力市场歧视与高校扩招的影响

## ——基于信号博弈模型的结构估计

徐 舒\*

**摘 要** 本文建立了一个信号博弈模型,在该模型框架下实证研究了大学教育的人力资本积累效应对大学生收入溢价的贡献以及高校扩招对大学生-非大学生收入差距的影响。基于模型的结构估计(structural estimation)证明该模型能很好地拟合大学生和非大学生的工资分布。模型的模拟结果表明:(1)大学教育的人力资本积累效应约能解释大学生工资溢价的72%;(2)与基于劳动力供给角度的分析不同,高校扩招可能通过劳动力市场歧视加剧大学生-非大学生的收入差距。

**关键词** 大学收入溢价,高校扩招,信号博弈

### 一、引 言

收入差距一直是政府、学术界和社会大众关注的焦点。近年来的经济学研究表明教育(尤其是高等教育)在其中扮演着重要的角色,大学教育的收入溢价对劳动者间的收入差距有重大影响(杜鹏,2005;张车伟,2006;刘泽云,2009)。就我国的实际情况来看,自1999年实行高校扩招后,随着大学生供给的大幅增加,大学生与非大学生间的工资差距不仅没有减少,反而呈现扩大的趋势。<sup>1</sup>这进一步加剧了劳动者间的收入不平等。从劳动力供给的角度分析,由于高校扩招在短期内大幅提高了大学生的相对供给,在需求相对稳定或平稳上升的情况下,应当观察到大学生相对工资的下降,而现实却恰恰相反。理论与实际现象的矛盾使得对大学收入溢价的研究有重要的理论和现实意义。具体而言,对该问题的讨论需要回答两个方面的问题:(1)大学收入溢价的根本原因是什么?(2)高校扩招与大学生-非大学生的收入差距扩

\* 西南财经大学经济与管理研究院。通信地址:四川省成都市光华村街55号西南财经大学经济与管理研究院,610074;电话:13982204113,E-mail:xsu@mail.swufe.edu.cn。本文受西南财经大学“211工程”三期建设项目资助。文章初稿完成于作者在香港岭南大学交流期间。特别感谢香港岭南大学魏向东教授在研究数据上的支持以及范承泽教授的帮助;与德州农工大学甘犁教授的讨论完善了文章的写作思路。感谢西南财经大学金融学院许志博士对文章初稿的宝贵修改意见。当然,文责自负。

<sup>1</sup> 附录A图A给出了1991—2006年间我国大学生与非大学生对数工资比,以及大学生在劳动者中所占的比重,两者同步上升的趋势非常明显。

大有怎样的联系?

回答上述两个问题的关键在于区分不同因素对大学收入溢价的影响。从经济学角度来看,大学教育的收入溢价有三个方面的原因:一是人力资本积累效应,指大学教育提高了劳动者的生产能力,从而提高了大学生的收入;二是能力信号效应(ability signaling),指大学教育是劳动者传递生产力信号的工具,而厂商在信息不对称的情况下将大学教育作为辨别劳动者生产力的标准,对非大学生进行工资歧视;三是劳动者的先天能力效应(innate ability),指进入大学的劳动者本身就具备较高的先天能力,因而在工作中往往表现出更高的生产能力,从而获得更高的工资,但这种能力与是否接受大学教育并没有关系。

正确区分这三种效应在大学生工资溢价中作用的意义在于,若大学生工资溢价全部是由大学教育的人力资本积累效应引起的,意味着可以通过让更多劳动者接受大学教育来降低收入不平等,同时高等教育带来的劳动者生产力提升能促进经济的发展。但是,能力信号效应的存在表明,简单地提高劳动者接受大学教育的机会(如高校扩招,虽然扩招并非完全出于此目的)不仅不能降低劳动者间的收入不平等,反而可能拉大大学生与非大学生间的收入差距。<sup>2</sup>最后,上述政策不会对劳动者的先天能力造成影响,从而不会改变由先天能力效应引起的收入不平等。

从实证研究的角度来看,回答上述两个问题最大的困难在于定量识别前文所述的三种不同效应对大学生工资溢价的影响,尤其是要区分人力资本积累效应与不可观测的劳动者先天能力这两种不同效应的作用。本文的主要贡献在于:建立了一个信号博弈模型,将劳动者先天能力、大学教育的人力资本积累和信号能力效应纳入一个统一的框架内,考察不同因素对劳动者收入的影响。基于模型的结构估计(structural estimation)得到模型参数,再通过反事实状态的政策模拟(counter-factual policy simulation)剥离劳动者先天能力的影响。上述步骤使我们能识别出大学教育的人力资本积累效应在大学生工资溢价中的作用,并判断高校扩招对大学生-非大学生收入差距的影响。研究结果表明,大学教育的人力资本积累效应能解释大学生-非大学生工资差距的71.86%;与基于劳动力供给视角的分析结论相反,高校扩招可能通过劳动力市场歧视扩大大学生-非大学生的收入差距。

文章其他部分的安排如下:第二部分是文献回顾;第三部分阐述信号博弈模型的设定并求解模型;第四部分介绍计量经济学的参数估计方法;第五部分给出数据描述和模型估计结果;第六部分是基于模型的政策模拟;第七部分是总结。

<sup>2</sup> 这主要是由于劳动者间的能力存在差异,其上大学的可能性也相应不同,扩招在一定条件下会加大厂商对非大学生的歧视程度。本文后面会有详细阐述。

## 二、文献回顾

经济学中对教育收益的研究由来已久，早在20世纪60年代，Becker (1964) 就从人力资本投资角度对教育收入溢价进行了解释，Mincer (1974) 则从实证角度给出了工资决定方程。Spence (1973) 在非对称信息的框架下提出了能力信号的概念，从博弈均衡的角度对收入差距提供了另一种解释。<sup>3</sup> 在此之后，许多文献试图从实证角度验证这两种不同的解释。Bedard (2001) 通过对比不同地区间的大学入学率和高中辍学率，认为经验证据主要支持能力信号模型的预测；Chevalier, Harmon, Walker and Zhu (2004) 的研究结果则表明英国的数据支持人力资本假说；Arcidiacono, Bayer and Hizmo (2008) 在雇主学习 (employer learning) 理论上对劳动者收入动态变化的研究表明，大学教育具有很强的信号效应，几乎能完美反映大学生的能力。但是，上述实证研究都不能定量区分出两种效应的相对重要性。Moro (2003) 的文章是一个重要的突破，他在统计歧视 (statistical discrimination) 的理论框架下建立了一个信号博弈模型，并相应发展出一套估计方法，能定量识别出能力信号在不同劳动力群体工资差距中的作用。Fang (2006) 在 Moro (2003) 的理论框架下，利用美国普查数据的估计表明，大学教育能提升劳动者约40%的生产力，同时能解释约2/3大学生-非大学生的工资差距。

将教育溢价与劳动者收入不平等相联系的研究始于20世纪90年代。实际数据显示，在劳动力市场上，随着高等教育劳动者相对供给的增加，其相对工资也在不断增长，造成了劳动力市场收入分配差距的扩大。对这类现象的研究产生了技能偏向型技术进步 (skill-biased technology change) 理论，成为解释这类现象的主要理论依据。国外学者基于技能偏向型技术进步理论进行的研究十分广泛，在20世纪90年代末已经得到丰富的成果。Acemoglu (2002) 对这一理论进行了详细的回顾和介绍；Card and DiNardo (2002) 则对实证研究中存在的与该理论不符之处进行了探讨。另一方面，Hendel, Shapiro and Willen (2005) 在技能偏向型技术进步理论框架外建立了一个信号博弈模型，表明高等教育机会的增加在一定条件下会加剧劳动力市场的收入不平等，对实际现象提供了新的解释。

对教育收益率上升与收入不平等这一问题，国内学者也进行了大量研究。李实和丁赛 (2003) 对抽样数据的分析表明，我国城镇劳动者的教育收益率在1990—1999年间呈现逐年上升趋势，并且教育的边际收益率也是递增的；杜鹃 (2005) 基于我国1995—2002年间的教育和收入分配数据的研究认为，

<sup>3</sup> 见 Weiss(1995)和 Riley(2001)的理论综述。

教育拓展是“居民收入差距拉大的重要原因”；张车伟(2006)发现教育回报率在高收入劳动者中更高，认为人力资本投资必须向穷人倾斜以减小收入差距。与杜鹏(2005)相反，杨俊、黄潇和李晓羽(2008)基于内生增长理论的分析表明，教育不平等与收入不平等间存在非线性关系，但认为教育扩展有利于教育不平等和收入不平等的改善；刘泽云(2009)利用2005年1%人口抽样调查生成的省际数据进行的研究发现，教育水平的分布和教育收益率的分布都会对工资不平等产生显著影响。

虽然上述研究的结论及政策建议不尽相同，但都表明大学收入溢价(教育收益率上升)与收入不平等密切相关。本文力图在此基础上更进一步，在一个统一的信号博弈模型框架下区分引起大学收入溢价的三种不同效应，并研究高校扩招如何通过这三种不同效应对大学生-非大学生的收入差距产生影响。

### 三、信号博弈模型

本文的信号博弈模型建立在Moro(2003)的统计歧视理论框架下。文章的这个部分先给出模型的基本设定，然后详细讨论模型的求解过程和均衡条件。

#### (一) 模型设定

经济中有测度为1的劳动者，劳动者具有不同的以货币产出衡量的能力，记为 $a$ ， $a$ 服从 $[\underline{a}, \bar{a}]$ 上的均匀分布。劳动者在进入劳动力市场之前，根据自身能力及对未来收益的预期决定是否进行人力资本的投资。记投资决策函数为 $I(\cdot)$ ，若进行人力资本投资( $I=1$ )，则该劳动者成为高技能劳动者( $H$ )；若不进行人力资本投资( $I=0$ )，则为低技能劳动者( $L$ )。高技能劳动者以概率 $p_h^c \equiv p(c|I=1)$ 进入大学，低技能劳动者以概率 $p_l^c \equiv p(c|I=0)$ 进入大学，且 $p_h^c > p_l^c$ 。进入大学对劳动者的生产能力有促进作用(即大学教育的人力资本积累效应，也称为大学教育的能力溢价)。记 $\rho_h, \rho_l$ 分别为大学教育对高技能劳动者和低技能劳动者的生产力乘数，显然有 $\rho_h > 1$ 且 $\rho_l > 1$ 。劳动者进入劳动力市场时，被赋予生产能力信号 $\theta \in [0, 1]$ ，高技能劳动者的信号分布密度为 $f_h(\theta)$ ，低技能劳动者的信号分布密度为 $f_l(\theta)$ 。由于信号 $\theta$ 是劳动者生产力的有效信息，信号越强，意味着劳动者是高技能劳动力的可能性越大，根据Moro(2003)以及Fang(2006)，我们假定信号分布密度为 $f_h(\theta), f_l(\theta)$ 满足单调似然比性质(monotone likelihood ratio property, MLRP)。

**假设 1**  $f_h(\theta)/f_l(\theta)$ 是关于参数 $\theta$ 的单调增函数。

该假设表明相对于低技术劳动者而言，进行了人力资本投资的高技术劳

动者获得高生产力信号的可能性更大。

经济中的厂商是同质的，其生产需要资本与劳动两种投入，厂商根据劳动的边际产出来决定工人的工资。厂商能观测到劳动者是否为大学生，并对大学生（ $c$ ）与非大学生（ $nc$ ）两个劳动力群体中高技能劳动者所占比例形成先验预期，分别记为  $\delta^c$  与  $\delta^{nc}$ 。在进一步观测到劳动者的信号  $\theta$  后，厂商依据贝叶斯法则推断给定信号强度下该劳动者为高技能劳动者的后验概率，并按照概率平均确定工人的工资。

该信号博弈流程可以概括如下：

阶段1，劳动者根据自身能力及对未来收益的预期决定是否进行人力资本投资，并以概率  $p_h^c, p_l^c$  进入大学生与非大学生两个劳动力群体。接受大学教育的劳动者获得生产力乘数  $\rho_i (i=h, l)$ 。

阶段2，劳动者进入劳动力市场，被赋予生产力信号  $\theta \in [0, 1]$ 。若劳动者为高技能劳动者，则信号分布密度为  $f_h(\theta)$ ；若为低技能劳动者，则信号分布密度为  $f_l(\theta)$ 。

阶段3，厂商观测到劳动者是否为大学生，以及该劳动者的信号强度。基于该信息，厂商形成该劳动者是高技能劳动者的后验概率  $\Omega$ ，并确定劳动者的工资。

阶段4，劳动者观测到厂商的工资函数并决定为哪家厂商工作。

## （二）模型求解

该信号博弈模型的求解遵循逆向推导原则（backward induction）。本文先从厂商的最优化问题出发，给出工资方程，然后在给定工资方程的情况下求解劳动者的人力资本投资决策问题，最后定义该博弈的均衡。

### 1. 厂商问题

给定先验概率  $\delta^c$  与  $\delta^{nc}$ ，根据贝叶斯定理，在观测到劳动者的信号强度后，厂商推断该劳动者为高技能劳动者的后验概率为

$$\Omega(\theta, \delta^j) = \frac{\delta^j \cdot f_h^j(\theta)}{\delta^j \cdot f_h^j(\theta) + (1 - \delta^j) \cdot f_l^j(\theta)} \quad (j = c, nc), \quad (1)$$

这里，我们允许信号分布密度  $f_h(\theta)$  与  $f_l(\theta)$  的参数在大学生和非大学生两个劳动力群体中存在差异。同时，MLRP 性质保证了  $\Omega(\theta, \delta^j)$  也是关于  $\theta$  的增函数。大学生与非大学生的工资方程由下式给出<sup>4</sup>：

<sup>4</sup> Moro(2003)与 Fang(2006)均假设厂商的生产有复杂与简单两类劳动投入，复杂劳动的单位产出更高，但只有高技能劳动者能从事复杂劳动。厂商根据劳动者的信号强度得到该劳动者是高技能劳动者的后验概率，并且在两类劳动中对劳动者进行分配。在他们的设定中，存在一个信号的门槛值  $\vartheta$ ，当劳动者的信号强度高于  $\vartheta$  时从事复杂劳动，低于  $\vartheta$  时从事简单劳动， $\vartheta$  本身也是待估参数。为了保证模型简明，本文假设劳动投入是同质的。本文实证部分的结果表明，该简化不会对工资分布的拟合产生影响。

$$\begin{aligned}\omega^c(\theta, \delta^c) &= \Omega(\theta, \delta^c) \cdot \rho_h y_h + (1 - \Omega(\theta, \delta^c)) \cdot \rho_l y_l, \\ \omega^{nc}(\theta, \delta^{nc}) &= \Omega(\theta, \delta^{nc}) \cdot y_h + (1 - \Omega(\theta, \delta^{nc})) \cdot y_l,\end{aligned}\quad (2)$$

其中  $y_h$  是高技能劳动者的边际产出,  $y_l$  是低技能劳动者的边际产出。令  $G \equiv p(I=1)$  为劳动者进行人力资本投资的概率, 则  $y_h, y_l$  可以分别表示为

$$y_h = E[a | I = 1] = G^{-1} \cdot \int_{\{a, I=1\}} a \cdot f(a) da, \quad (3)$$

$$y_l = E[a | I = 0] = (1 - G)^{-1} \cdot \int_{\{a, I=0\}} a \cdot f(a) da. \quad (4)$$

## 2. 劳动者问题

劳动者通过比较人力资本投资的成本和收益做出投资决策。进行人力资本投资的期望收益可以表达为

$$\begin{aligned}B(\delta^c, \delta^{nc}) &= \int_{\theta} [\omega^c(\theta, \delta^c) \cdot p_h^c + \omega^{nc}(\theta, \delta^{nc}) \cdot (1 - p_h^c)] f_h(\theta) d\theta \\ &\quad - \int_{\theta} [\omega^c(\theta, \delta^c) \cdot p_l^c + \omega^{nc}(\theta, \delta^{nc}) \cdot (1 - p_l^c)] f_l(\theta) d\theta,\end{aligned}\quad (5)$$

(5)式右边第一部分是进行人力资本投资获得的以平均工资水平衡量的期望收益, 它是进行人力资本投资后获得大学生工资(成为大学生)与非大学生工资的概率平均数, 并在信号分布上取期望。同理, (5)式右边的第二部分是不进行人力资本投资获得的平均工资水平, 两者相减得到人力资本投资的期望收益。

人力资本的投资成本只在进行投资时产生, 本文假设该成本为劳动者能力  $a$  的线性函数:  $c = \alpha + \beta a$ 。其中  $\beta < 0$ , 意味着劳动者的能力越高, 其人力资本投资所产生的成本就越小。由于  $a$  在  $[\underline{a}, \bar{a}]$  区间上服从均匀分布, 可以得到劳动者进行人力资本投资的概率为

$$\begin{aligned}G \equiv p(I = 1) &= p(c < B) = p(\alpha + \beta a < B) \\ &= \int_{\frac{B-\alpha}{\beta}}^{\bar{a}} f(a) da = \frac{\bar{a}\beta - B + \alpha}{(\bar{a} - \underline{a})\beta}.\end{aligned}\quad (6)$$

## 3. 均衡

该博弈的均衡条件可以由如下等式刻画:

$$\begin{aligned}p_c &= p(c | I = 1) \cdot p(I = 1) + p(c | I = 0) \cdot (1 - p(I = 1)) \\ &= p_h^c \cdot G + p_l^c \cdot (1 - G),\end{aligned}\quad (7)$$

$$\delta^c \equiv p(I = 1 | c) = \frac{p(c | I = 1) \cdot p(I = 1)}{p_c} = \frac{p_h^c \cdot G}{p_c}, \quad (8)$$

$$\delta^{nc} \equiv p(I = 1 | nc) = \frac{p(nc | I = 1) \cdot p(I = 1)}{1 - p_c} = \frac{(1 - p_h^c) \cdot G}{1 - p_c}, \quad (9)$$

其中,  $p_c$  是劳动力市场上大学生所占比重。(7)式表示的是在市场均衡情况下, 劳动力市场上大学生所占比重等于劳动者成为大学生的无条件概率。(8)式则表明在市场均衡时, 厂商对大学生群体中高技能劳动者所占比重的先验预期与实际大学生群体中高技能劳动者比重相一致。类似的, (9)式说明厂商对非大学生群体中高技能劳动者所占比重的先验预期也与实际非大学生群体中高技能劳动者的比重相一致。

#### 四、参数估计

根据 Moro(2003), 模型的估计可以分两步进行。首先, 给定信号分布密度  $f_h(\theta), f_l(\theta)$  的函数形式, 通过工资决定方程(2)可以估计出工资方程中的参数  $\psi \equiv \{\delta, \gamma_h, \gamma_l, \rho_h, \rho_l, \lambda_h, \lambda_l\}$ , 其中  $\{\lambda_h, \lambda_l\}$  分别是信号分布  $f_h(\theta)$  与  $f_l(\theta)$  中的参数。然后, 在给定工资方程参数的情况下, 根据均衡条件解出模型中剩余的参数  $\chi \equiv \{p_h^c, p_l^c, \underline{a}, \bar{a}, \alpha, \beta\}$ 。

##### (一) 信号分布及工资方程参数的估计

与 Antonovics(2002)类似, 假定  $f_h(\theta), f_l(\theta)$  分别服从参数为  $(\gamma_h, 1)$  及  $(1, \gamma_l)$  的 Beta 分布, 即密度函数分别为

$$\begin{aligned} f_h(\theta) &= \gamma_h \cdot \theta^{\gamma_h - 1}, \\ f_l(\theta) &= \gamma_l \cdot (1 - \theta)^{\gamma_l - 1}, \end{aligned} \quad (10)$$

其中  $\theta \in [0, 1]$ 。为了满足单调似然比性质 (MLRP), 必须限定  $\gamma_h \geq 1$  且  $\gamma_l \geq 1$ <sup>5</sup>。根据 Beta 分布的性质, 能力信号  $\theta$  的均值和方差可以表示为

$$\begin{aligned} E(\theta) &= 1/(1 + \gamma_i) & (i = h, l), \\ \text{var}(\theta) &= \gamma_i / [(1 + \gamma_i)^2 (2 + \gamma_i)] & (i = h, l), \end{aligned} \quad (11)$$

容易验证, 信号分布的均值和方差都是关于参数  $\gamma$  的减函数。这说明估计出的  $\gamma$  数值越大, 信号分布的方差就越小, 生产力信号的信息含量就越大, 意味着厂商能更加准确地从信号  $\theta$  判断劳动者类型 (高技能劳动者或低技能劳动者)。

从模型估计的角度出发, 生产力信号  $\theta$  在总体中的分布是上述两个 Beta 分布的混合, 其混合概率为厂商的先验概率  $\delta$ :

$$f(\theta) = \delta^j \cdot f_h^j(\theta) + (1 - \delta^j) \cdot f_l^j(\theta) \quad (j = c, nc). \quad (12)$$

<sup>5</sup> 在 Moro(2003)与 Fang(2006)的设定中,  $\gamma_h = \gamma_l$ 。在满足 MLRP 的前提下, 本文采用了更为一般的设定形式。对 Beta 分布的性质及其在回归建模中的应用, 可以参考 Cribari-Neto and Vasconcellos(2002)以及 Cribari-Neto and Ferrari(2004)的讨论。

如前文所述, MLRP 性质保证了后验概率  $\Omega(\theta, \delta^j)$  是关于  $\theta$  的增函数, 从方程 (2) 容易得到, 工资函数  $w(\theta, \delta^j)$  也是  $\theta$  的单调增函数。由工资函数的单调性可以将工资  $w$  的分布表达成信号  $\theta$  的分布形式。记  $\theta = \tau^j(w) = w^{-1}(\cdot)$ ;  $R_+ \rightarrow [0, 1]$  为工资函数的逆函数,  $J$  为相应的雅可比行列式, 显然有

$$J^j(w) = \frac{d\tau^j(w)}{dw} \equiv \left[ \frac{dw^j(\tau^j(w))}{d\theta} \right]^{-1}. \quad (13)$$

因此, 工资  $w$  的似然函数可以表示为

$$L^j = \prod_{n=1}^{N^j} \{ \delta^j \cdot f_h^j[\tau^j(w^j)] + (1 - \delta^j) \cdot f_l^j[\tau^j(w^j)] \} \cdot J^j(w^j), \quad (14)$$

给定模型的参数,  $\theta = \tau^j(w)$  可以通过数值方法从工资方程 (2) 解出, 而雅可比行列式  $J^j(w)$  则有显式解, 附录 B 给出了  $J^j(w)$  的具体形式。由此, 可以通过极大似然估计求得模型中的参数。令  $y_l^c \equiv \rho_l \cdot y_l$ ,  $y_h^c \equiv \rho_h \cdot y_h$ , 显然有  $L(w|\phi) = L(w^c|\phi^c) \cdot L(w^{nc}|\phi^{nc})$ , 因此可以分别对大学生和非大学生样本进行估计, 而不会造成信息损失。

在实际估计中, 可以对估计步骤进行进一步简化。注意到在工资方程 (2) 中, 对非大学生而言, 当  $\theta \rightarrow 0$  时,  $w \rightarrow y_l$ ; 当  $\theta \rightarrow 1$  时,  $w \rightarrow y_h$ 。又因为工资  $w$  是  $\theta$  的单调增函数, 这意味着  $y_l$  是非大学生样本中观测到的最低工资, 而  $y_h$  则是非大学生样本中观测到的最高工资。为了避免异常值的影响, 保证结果的稳健性, 同时尽可能增大样本量, 本文将非大学生样本中工资分布 1% 和 99% 的分位数分别作为  $y_l$  和  $y_h$  的估计; 相应的, 大学生样本中工资分布 1% 和 99% 的分位数则分别是  $\rho_l y_l$  和  $\rho_h y_h$  的估计。二者相除得到  $\rho_l$  和  $\rho_h$  的估计。<sup>6</sup> 这样, 在应用极大似然法估计 (14) 式时, 只需要估计  $\lambda = \{\delta, \gamma_h, \gamma_l\}$  三个参数。<sup>7</sup>

## (二) 剩余模型参数的估计

在得到参数  $\phi = \{\delta, \gamma_h, \gamma_l, \rho_h, \rho_l, \lambda\}$  的估计后, 可以利用均衡条件, 通过解非线性联立方程组的方法得到模型的剩余参数。

首先, 注意到投资收益  $B$  (见 (5) 式) 可以写成未知参数  $p_h^c, p_l^c$  的线性形式:

<sup>6</sup> 根据此方法得到的  $\rho_l, \rho_h$  的估计分别为 1.29 和 1.43。若采用 2% 与 98%, 3% 与 97% 的分位数作为  $y_l$  和  $y_h$  的估计, 得到的  $\rho_l$  与  $\rho_h$  估计值则分别为 {1.31, 1.32} 和 {1.42, 1.45}, 表明该估计值具有稳健性, 不受所选择分位数的影响。

<sup>7</sup> 上述处理方法也意味着在利用极大似然法估计参数  $\{\delta, \gamma_h, \gamma_l\}$  时, 小于 1% 和大于 99% 分位数的样本没有被使用到。

$$\begin{aligned}
 B(\delta^c, \delta^m) &= p_h^c \cdot \int_{\theta} [\omega^c(\theta, \delta^c) - \omega^m(\theta, \delta^m)] f_h(\theta) d\theta \\
 &\quad - p_l^c \cdot \int_{\theta} [\omega^c(\theta, \delta^c) - \omega^m(\theta, \delta^m)] f_l(\theta) d\theta \\
 &\quad + \int_{\theta} \omega(\theta, \delta^c) (f_h - f_l) d\theta.
 \end{aligned} \tag{15}$$

给定参数集  $\psi$ , (15) 式中的积分可以通过数值方法计算得到。这样, 根据 (6) 式可以将进行人力资本投资的概率  $G$  表达成未知参数  $\chi \equiv \{p_h^c, p_l^c, a, \bar{a}, \alpha, \beta\}$  的函数。

其次, 从数据中可以得到大学生在全体劳动者中所占比重, 这是劳动者成为大学生的无条件概率  $p_c$  的一致估计。给定  $G$  的表达式, (7) 式构成非线性方程组的第一个方程。同时, 从第一步估计中已经得到  $\delta^c$  与  $\delta^m$  的估计值, 这样, (8) 式和 (9) 式就构成非线性方程组的第二和第三个方程。

再次, 在  $a$  的均匀分布假设下, 第一步估计出的  $y_l, y_h$  有如下表达式:

$$y_h = E[a | I = 1] = G^{-1} \cdot \int_{\frac{B-a}{\beta}}^{\bar{a}} a \cdot f(a) da = \frac{1}{2G(\bar{a} - \underline{a})} \left[ \bar{a}^2 - \frac{(B-a)^2}{\beta^2} \right], \tag{16}$$

$$\begin{aligned}
 y_l &= E[a | I = 0] = (1-G)^{-1} \cdot \int_{\underline{a}}^{\frac{B-a}{\beta}} a \cdot f(a) da \\
 &= \frac{1}{2(1-G)(\bar{a} - \underline{a})} \left[ \frac{(B-a)^2}{\beta^2} - \underline{a}^2 \right].
 \end{aligned} \tag{17}$$

(16) 式和 (17) 式构成非线性方程组的第四与第五个方程。

最后, 由于参数集  $\chi$  包含六个未知参数, 而根据模型只能得到五个方程, 为了求解方程组, 必须对参数施加一个额外约束。在实际估计中, 我们令参数  $\alpha = (\bar{a} - \underline{a})/2$ 。这样, 在该约束条件下联立求解方程 (7)–(9) 以及 (16) 和 (17), 就得到参数  $\chi = \{p_h^c, p_l^c, a, \bar{a}, \alpha, \beta\}$  的估计。

上述参数估计的标准差可以由模拟得到。具体做法是, 在第一步估计中得到参数  $\psi$  及其方差-协方差矩阵  $V_{\psi}$  后, 可以从多元正态分布  $N(\psi, V_{\psi})$  随机产生  $R$  个  $\psi^r (r=1, 2, \dots, R)$  序列。给定  $\psi^r$ , 重复求解上述非线性方程组, 得到  $R$  个  $\chi$  的估计值, 然后计算其标准差。在实际计算中, 由于在第一步中是分别估计参数  $\{y_h, y_l, \rho_h, \rho_l\}$  与  $\{\delta_h, \delta_l, \lambda_h, \lambda_l\}$ , 我们无法得到参数集  $\psi$  整体的方差-协方差矩阵 (主要是两部分参数的协方差), 因此在模拟中采用了固定参数  $\{y_h, y_l, \rho_h, \rho_l\}$  的处理方法, 只对  $\{\delta_h, \delta_l, \lambda_h, \lambda_l\}$  进行随机抽样。这会在一定程度上低估参数  $\psi$  的标准差。

## 五、数据及参数估计结果

本文的实证研究使用了来自“中国城镇家庭调查”(China's Urban Household Survey, CUHS) 1998—2003年的部分数据。该调查由国家统计局自1988年开始实施,详细记录了受访家庭及家庭成员的人口特征、收入以及消费支出等方面的信息。从1998—2001年,调查覆盖的家庭总数约为17000户;自2002年起,调查样本实行了扩充,达到56000户家庭,覆盖全国418个城市。本文使用的数据是该调查的一个子样本,包括北京、辽宁、浙江、广东、四川和陕西六个省(市)的样本数据。一方面,这六个省(市)在地理位置和经济发展水平上存在较大差异,使样本具有代表性;另一方面,在样本期内,高校扩招的效果还没有显现出来,可以认为劳动力市场处于相对稳定的状态<sup>8</sup>,与模型的假设相契合。

### (一) 数据处理与样本选择

本文研究的收入指的是劳动者的工资性收入,包括了工资、奖金及津贴,因此属于个体或私营企业主的劳动者没有包括在样本中。CUHS记录的是劳动者的名义年收入,我们先用不同省(市)的CPI将劳动者的名义工资收入调整为以1997年为基期的实际收入,再除以52换算成周平均收入的概念。同时,本文剔除了受教育程度在初中及以下的劳动者。这样,样本中劳动者的受教育程度为:高中、中专和大学及以上学历。高中和中专学历的劳动者构成非大学生劳动力群体,大学及以上学历的劳动者构成大学生劳动力群体。上述处理方式主要是与现有国外文献保持一致。此外,由于本文的模型是一个静态模型,为了尽可能剔除工作经验的影响<sup>9</sup>,同时保证样本数量,我们将劳动者的年龄限制在20—35岁之间。最后,劳动者间性别工资差异的存在也会对模型估计造成影响,为了剔除性别工资差异的影响,先拟合如下线性回归模型:

$$\text{wage} = \beta_0 + \beta_1 \cdot \text{female} + \varepsilon, \quad (18)$$

其中wage是计算出的周平均工资, female是虚拟变量,当劳动者是女性时取1,否则为0。在得到回归系数 $\hat{\beta}_0, \hat{\beta}_1$ 后,计算 $\hat{w} = \hat{\beta}_0 + \hat{\varepsilon}$ ,作为本文实证研究的工资测度。<sup>10</sup>表1给出了主要变量的描述统计结果。从表1可以看出,大学

<sup>8</sup> 高校扩招始于1999年,到2004年第一批扩招的大学生才进入劳动力市场。当然,这样的说法并不严格成立,可以认为自1999年起,劳动者进入大学的概率已经发生了改变。但从本文使用的样本来看,在20—60岁的劳动者中,大学生所占比例十分稳定,从1998年到2003年分别为11.4%,12.3%,12.1%,12.2%,11.5%和11.6%。

<sup>9</sup> 随着劳动者工作经验的增加,厂商对劳动者的劳动能力的认识是一个动态的贝叶斯学习(Bayesian learning)过程,需要建立一个动态模型来刻画这一情形,这超出了本文的研究范围。

<sup>10</sup> 在文章的附录D中,我们在回归式(18)中加入了地区虚拟变量,生成新的工资序列并重新估计模型参数,以检验实证结果的稳健性。

生的平均工资比非大学生高约43%。而从年龄结构上看,样本中的大学生比非大学生年长1岁左右。同时,大学生中女性比例比非大学生低8个百分点。

表1 数据基本统计描述

变量名	均值	中位数	标准差	最小值	最大值	样本数
大学生						
周工资	329.98	261.29	239.85	20.37	1655.82	1772
年龄	29.57	30.00	3.93	20.00	35.00	1772
女性哑变量	0.45	0	0.50	0	1	1772
非大学生						
周工资	231.03	190.93	168.24	19.01	1604.02	10919
年龄	28.84	29.00	4.42	20.00	35.00	10919
女性哑变量	0.53	1	0.50	0	1	10919

注:样本包括了下列估计中没有使用到的小于1%分位数以及大于99%分位数的观测值,因此均值和标准差相对较大,但中位数非常接近。

## (二) 模型估计结果

### 1. 工资决定方程的估计结果

根据本文第四部分介绍的估计方法,采用极大似然估计分别对大学生样本和非大学生样本进行估计,得到的结果如表2所示。

表2 工资决定方程中的参数

参数	大学生(标准差)		非大学生(标准差)	
$\delta$	0.255	(0.005)	0.219	(0.002)
$\gamma_h$	1.581	(0.026)	1.584	(0.010)
$\gamma_l$	1.814	(0.040)	1.695	(0.016)
$\rho_l \times y_l$	56.862	(3.424)	—	—
$\rho_h \times y_h$	1074.32	(54.75)	—	—
$y_l$	—	—	43.968	(1.137)
$y_h$	—	—	751.29	(16.83)
$\rho_l$	1.293	(0.085)	—	—
$\rho_h$	1.430	(0.081)	—	—
对数似然函数值	-11 209.27		-64 945.2	
样本量	1 736		10 712	

注:括号中为稳健标准差。其中 $y_l, y_h, \rho_l$ 和 $\rho_h$ 的标准差通过对全样本自举抽样(bootstrap)得到,自举抽样次数为1000。

表2给出了工资决定方程的参数估计结果,所有估计值均在1%的显著性水平上显著。上述结果在下列两个方面值得注意:

第一,有25.5%的大学生进行了人力资本投资,成为高技术劳动者;这一比例在非大学生中为21.9%。两个劳动力群体中高技术劳动者的比例较为接近,只相差3.6个百分点,这个结果与直观预期存在一定差异。另一方面,大学教育的生产力乘数在低技术劳动者中约为1.29,而在高技术劳动者中为1.43,表明大学教育能带来29%—43%的工资溢价,这是一个很高的数值。

同时,高技术劳动者在大学教育中获益更大,比低技术劳动者高14%。综合这两个方面的结果,可以认为大学生与非大学生的收入差异主要是由大学教育的溢价引起的<sup>11</sup>,本文第七部分基于模型的政策模拟验证了这一结论。

第二,从生产力信号在大学生和非大学生两个劳动力群体的分布来看,高技术劳动者的信号分布在两个群体中非常接近, $\gamma_h$ 的估计值分别为1.581和1.584;但低技术劳动者的信号分布在两个劳动力群体中的差别较大, $\gamma_l$ 的估计值分别为1.814和1.695。如前文所述,信号分布的方差是 $\gamma$ 的减函数, $\gamma$ 的估计值越大,表明信号分布的方差越小,其信息含量越大,厂商越能准确地根据信号识别出劳动者的技能水平。上述估计结果意味着在两个劳动力群体中,厂商都能更精确地辨别出低技能劳动者(在两个劳动力群体中都有 $\gamma_h < \gamma_l$ ),但这种识别能力在大学生中更强。

由于求解模型剩余参数和下文根据研究目的进行的政策模拟都建立在工资方程的估计上,因此有必要先对工资方程的拟合情况进行评估,以保证模型剩余参数求解以及政策模拟结果的准确性。评价工资方程的拟合优度可以从两个方面入手:一是根据极大似然原理,比较模型拟合的密度函数与非参数核密度估计得到的经验密度函数之间的差异程度;二是给定参数,通过模型中的工资决定方程生成工资分布,比较通过模型生成的工资分布和数据中实际工资分布的差异。

图1给出了根据极大似然估计原理得到的工资分布拟合情况,实线是模型极大似然估计得到的密度函数拟合结果,虚线是核密度估计的结果。可以

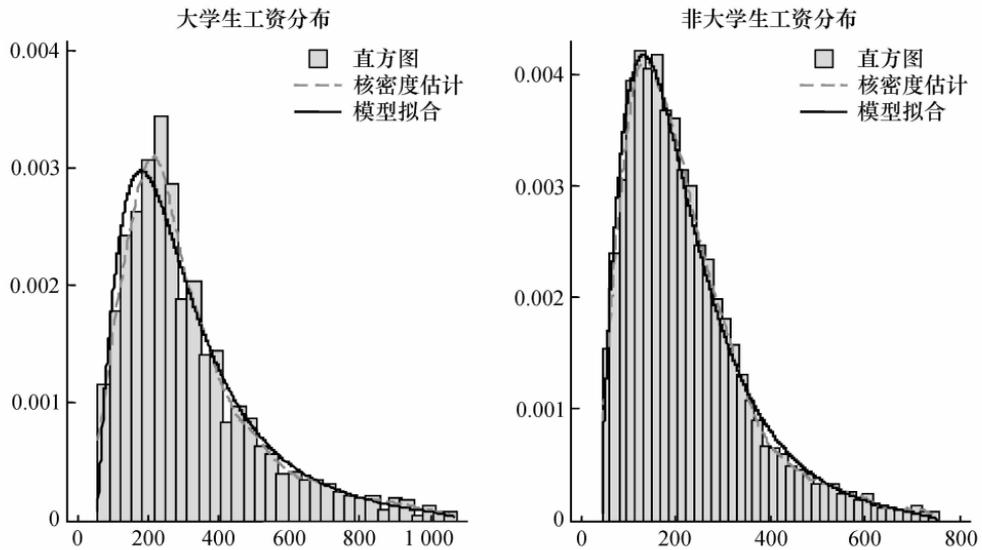


图1 大学生与非大学生工资分布拟合

<sup>11</sup> 这里需要阐明的是,在模型中,人力资本的投资决策是内生的,是考虑到大学工资溢价后的均衡决策结果。因此,这个结论并不能说明人力资本的投资对工资差异没有影响。

看出, 对非大学生工资分布的拟合非常理想, 模型拟合的曲线与核密度估计的曲线几乎重合, 说明模型能很好地反映非大学生的实际工资分布。相对而言, 模型对大学生工资分布的拟合精度略低, 主要表现为拟合分布在峰度上略低于经验分布, 这与大学生群体的样本量相对较小有一定关系。但是, 无论是对大学生样本还是非大学生样本, 模型都能精确地捕捉到工资分布显著右偏的特征, 并且在分布的尾部有良好的拟合效果, 因此, 可以认为模型的总体拟合效果仍然是非常出色的。

根据工资决定方程 (2), 给定估计出的模型参数可以进一步对大学生和非大学生的工资分布进行模拟。由大学生和非大学生的样本量, 模拟生成的大学生和非大学生观测数分别为 10 000 和 3 000。画出其直方图如图 2 和图 3 所示。

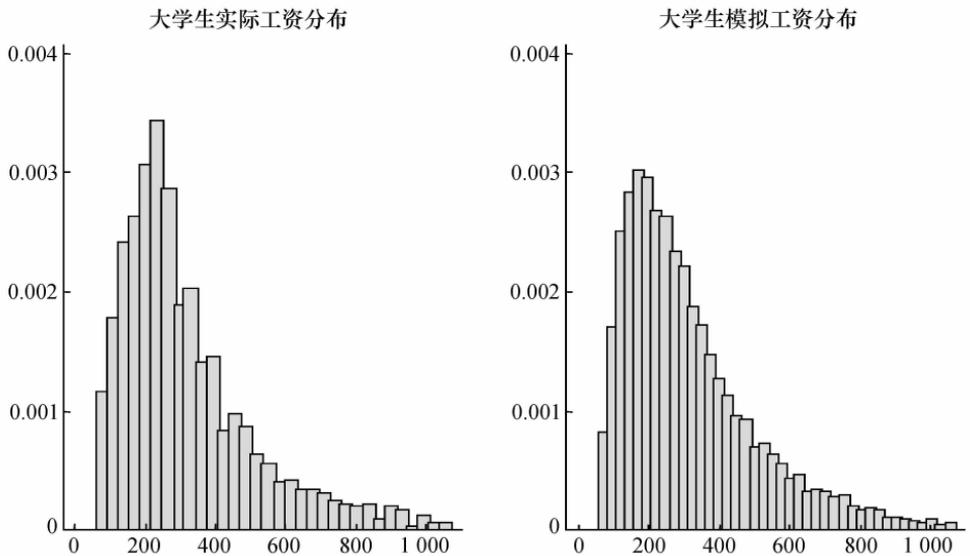


图2 大学生的实际工资分布与模拟工资分布

图2和图3左侧分别给出了大学生和非大学生样本实际工资分布的直方图, 右侧则是根据工资决定方程(2)得到的模拟工资分布直方图。可以发现, 模拟工资分布和实际工资分布非常接近, 为了定量描述模拟分布与实际分布的特征, 表3进一步给出了实际工资分布和模拟工资分布的描述统计量。从表3可以看出, 模拟工资分布在均值、中位数以及方差上都与实际分布非常接近, 尤其在均值和标准差上, 两者的差异非常小。而在偏度和峰度上, 模型工资分布则略低于实际分布。上述结果再次表明, 模拟分布对实际分布的近似是十分理想的。

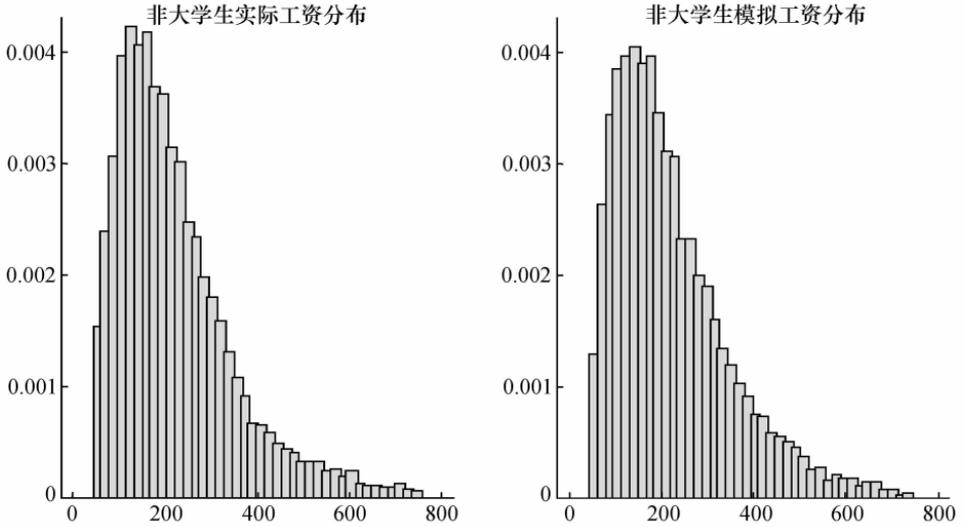


图3 非大学生的实际工资分布与模拟工资分布

表3 实际工资分布与模拟工资分布

	均值	中位数	标准差	偏度	峰度
大学生					
实际分布	315.53	261.29	193.11	1.40	4.86
模拟分布	315.47	266.39	189.61	1.27	4.46
非大学生					
实际分布	204.30	190.87	128.06	1.34	4.94
模拟分布	206.44	191.50	126.48	1.20	4.40

注:对大学生工资分布的模拟样本数为3000,对非大学生工资分布的模拟样本数为10000。

## 2. 模型剩余参数的估计结果

上述模型评估结果表明工资决定方程的拟合有很高的精确度,由此在得到工资决定方程的参数估计后,给定约束条件 $\alpha = (\bar{a} - a)/2$ ,通过联立求解方程(7)、(8)、(9)、(16)以及(17)可以得到模型剩余部分的参数估计。估计结果如表4所示。

表4 剩余模型参数估计结果

参数	估计值(标准差)	参数	估计值(标准差)
$p_h^c$	0.159 (0.002)	$\alpha$	172.61 (2.267)
$p_i^c$	0.134 (0.001)	$\beta$	-0.607 (0.019)
$\bar{a}$	350.56 (3.614)	$B$	122.50 (2.106)
$a$	5.345 (0.920)	$G$	0.224 (0.002)

注:参数估计的标准差通过对工资决定方程的参数估计值及其方差-协方差矩阵抽样产生,抽样次数为1000。

在得到模型剩余参数估计的同时,表4还列出了均衡情况下人力资本投资收益 $B$ 以及人力资本投资概率 $G$ 的估计。从表4可以看出,劳动者能力分布的区间为 $[5.345, 350.56]$ 。在进行人力资本投资后,工人进入大学的概率约为0.16,较不进行人力资本投资的情况高2.5%,两者的差别很小,这和工资

方程估计中大学生与非大学生群体中高技术工人比重十分接近的结论是一致的。另外， $\beta$ 的估计值约为-0.61，与模型的约束一致，表明劳动者的能力越高，其进行人力资本投资产生的成本就越小。最后，人力资本投资的平均收益为122.5，略高于大学生与非大学生间111.23的平均工资差距；而均衡的人力资本投资概率（即全体劳动者中高技术工人所占比例）为22.4%。

## 六、政策模拟

在得到模型的参数估计后，可以根据模型进行反事实状态的政策模拟(counter-factual policy simulation)，分解出模型中各个因素对大学生和非大学生间收入差距的影响。本文主要关心两个方面的问题：一是大学教育的能力溢价（即生产力乘数 $\rho_l, \rho_h$ ）能在多大程度上解释大学生和非大学生间的收入差距；二是高校扩招对大学生和非大学生工资差距的影响。

对于大学教育的工资溢价问题，限定 $\rho_l=1, \rho_h=1$ ，在其他参数不变的情况下重新求解模型<sup>12</sup>，得到新的均衡状态下大学生和非大学生两类劳动者的工资。通过与实际工资的比较，可以得到大学教育的工资溢价对大学生及非大学生收入差距的影响。高校扩招对大学生及非大学生收入差距的影响在模型中主要是通过提高进入大学的概率 $p_h^c$ 及 $p_l^c$ 来体现的。大学扩招对进入大学的概率有三种不同的影响：(1)扩招同等程度地提高了高技能劳动者和低技能劳动者进入大学的可能性，但两类劳动者进入大学的概率差保持不变；(2)扩招对进行了人力资本投资的高技能劳动者更加有利，高技能劳动者进入大学的可能性提升得更多；(3)扩招对低技能劳动者更有利，低技能劳动者进入大学的可能性提升得更多。根据上述三种情况，分别设定新的 $p_h^c, p_l^c$ 的数值，重新求解模型，同样可以得到新的均衡状态下大学生与非大学生的工资。通过比较实际工资与新均衡状态下的工资，能够识别高校扩招对大学生和非大学生收入差距的影响。政策模拟的结果如表5所示。

表5 政策模拟

实际分布	反事实状态分布				
	$\rho_l=1$ $\rho_h=1$	$p_h^c=0.21$ $p_l^c=0.185$	$p_h^c=0.21$ $p_l^c=0.145$	$p_h^c=0.21$ $p_l^c=0.205$	
均值					
大学生	315.53	263.4	285.86	309.76	277.17
非大学生	204.30	232.1	183.85	167.62	194.16
工资差距	111.23	<b>31.3</b>	<b>102.01</b>	<b>142.14</b>	<b>83.01</b>

注：模拟样本数为10000。

表5第二列显示的是数据中大学生与非大学生的实际工资差距，从均值上看，工资差距约为111元，大学生的平均工资较非大学生高出约54.4%。

<sup>12</sup> 在给定参数下重新求解模型的步骤见附录C。

而在消除大学教育的能力溢价后(第三列),大学生的平均工资只较非大学生高31.3元,说明大学教育的能力溢价能解释大学生与非大学生间71.86%的收入差距。<sup>13</sup>这个结果与Fang(2006)利用美国数据得到的2/3的结论较为接近。

表5第四到第六列描述的是高校扩招对大学生和非大学生收入差距的影响。可以看出,若高校扩招同等程度地提高了两类劳动者(高技能劳动者和低技能劳动者)上大学的可能性,大学生与非大学生间的工资差距基本保持不变,从111元下降到102元。但差距减少的部分主要是由于低技能劳动者上大学的概率基数较低,因而概率上升的百分比更大引起的。当扩招对高技能劳动者更有利时,大学生与非大学生间的收入差距明显增加。在高技术劳动者上大学的概率增量比低技术劳动者高5个百分点的情况下,大学生与非大学生间的工资差距增加到142.14元,增长了约30%。与此相反,当低技能劳动者上大学的概率增量更大时,大学生与非大学生的工资差距明显缩小,从111元下降到83元,下降幅度约为25%。

表5第五列描述的情形最值得关注。在高技能劳动者上大学的可能性相对于低技能劳动者大得多时,意味着大学生中高技能劳动者的比例增加,厂商能更好地根据劳动者是否是大学生来判断该劳动者是否属于高技能劳动者。极端的情况是,只有进行了人力资本投资的高技能劳动者能上大学,这样厂商能完美地判断劳动者类型而不受生产力信号的影响,劳动者完全凭借其边际产出获得工资,此时大学生和非大学生的收入差距是最大的。

基于劳动力供给视角的分析认为,大学扩招后随着大学生供给的大幅增加,大学生与非大学生间的相对工资差距应该减小。但是,这样的理论预测与实际数据并不相符。正如附录中的图A所示,基于CHNS数据计算出的1991—2006年间大学生与非大学生的对数工资比以及大学生在劳动者群体中所占的比重,两者呈现明显的同步上升趋势。<sup>14</sup>虽然技能偏向型技术进步理论能较好地解释上述现象,但在我国特殊的国情下,本文的模型提供了另外一种解释:在信息不对称情况下,大学扩招也可能会加大劳动力市场上对非大学生劳动者的歧视,从而在一定程度上扩大了大学生和非大学生间的收入差距,本文称之为大学扩招的**统计歧视效应**。

## 七、结论与扩展

本文通过一个信号博弈模型讨论了大学教育的人力资本积累效应、劳动

<sup>13</sup>  $71.86\% = (111.23 - 31.3) \div 111.23 \times 100\%$ 。

<sup>14</sup> 由于本文使用的CUHS数据长度有限,这里采用CHNS数据说明大学生工资溢价的变化趋势。由于CHNS数据已经被国内学者广泛使用,这里不对数据进行详细介绍,感兴趣的读者可以通过CHNS网站(<http://www.cpc.unc.edu/projects/china>)了解更多信息。

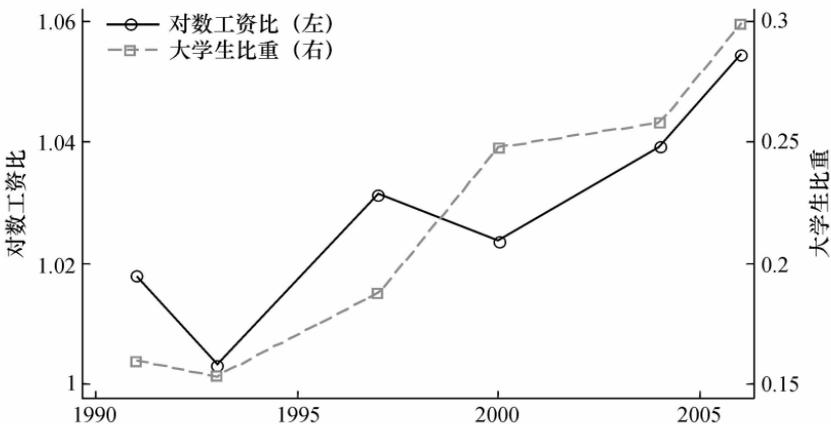
者的先天能力,以及劳动力市场上的能力信号效应在大学生工资溢价中的作用。基于模型的结构估计表明该模型能很好地拟合现实工资数据的分布;反事实状态的模拟表明,在我国劳动力市场上,大学生的工资溢价主要是由大学教育的人力资本积累效应引起的,约能解释大学工资溢价的72%。另一方面,模型证实了能力信号效应的存在,大学生与非大学生在人力资本投资以及能力信号分布上的不同也会对其工资分布造成影响。劳动力市场上的信息不对称与高校扩招相互作用,当高校扩招在更大程度上提高了高技术劳动者进入大学的概率时,会使劳动力市场上对非大学生的歧视程度上升,在其他条件不变的情况下导致大学生与非大学生间收入差距扩大。

针对20世纪90年代开始出现的大学生供给增加的同时相对收入也同步上升的现象,本文建立的模型在技能偏向型技术进步的理论框架外,结合我国特殊的政策背景,给出了新的可能的解释,即高校扩招的统计歧视效应也可能引起大学生-非大学生收入差距的扩大。

需要注意的是,本文的模型是一个静态模型。从动态角度来看,随着厂商对劳动者能力信号的不断学习,其对劳动者真实生产能力的判断也会越来越准确。直观上判断,此时能力信号在劳动者工资中所起的作用会逐渐减弱,而大学教育的人力资本积累效应在大学生-非大学生收入差距中的作用会更大。当然,实际工资差距还取决于大学生和非大学生中高技能劳动者的比例。另一方面,如何从数据中区分技能偏向型技术进步和大学扩招的统计歧视效应这两种不同因素对大学生-非大学生收入差距的影响,并计算它们间的相对重要性是一个值得探索的方向,本文将其留作后继研究。

## 附录

### A. 大学生和非大学生对数工资比及大学生所占比重



图A 1991—2006年大学生比重与大学生-非大学生对数工资比

注:计算基于CHNS数据,与文章中的定义保持一致,非大学生指的是最高学历为高中或中专的劳动者。

## B. 雅可比行列式

在给定  $f_h(\theta)$  与  $f_l(\theta)$  分布形式的情况下, 有

$$\Omega(\theta, \delta) = \frac{\delta \cdot \gamma_h \cdot \theta^{\gamma_h - 1}}{\delta \cdot \gamma_h \cdot \theta^{\gamma_h - 1} + (1 - \delta) \cdot \gamma_l \cdot (1 - \theta)^{\gamma_l - 1}},$$

$$\omega(\theta, \delta) = \frac{\delta \cdot \gamma_h \cdot \theta^{\gamma_h - 1} \cdot y_h + (1 - \delta) \cdot \gamma_l \cdot (1 - \theta)^{\gamma_l - 1} \cdot y_l}{\delta \cdot \gamma_h \cdot \theta^{\gamma_h - 1} + (1 - \delta) \cdot \gamma_l \cdot (1 - \theta)^{\gamma_l - 1}},$$

所以,

$$J(\omega) = \frac{d\tau(\omega)}{d\omega} \equiv \left[ \frac{d\omega}{d\theta} \right]^{-1}$$

$$= \frac{\theta(1 - \theta)(\delta \cdot \gamma_h \theta^{\gamma_h - 1} + \gamma_l(1 - \theta)^{\gamma_l - 1}(1 - \delta))^2}{\delta \cdot \gamma_h \theta^{\gamma_h - 1} \cdot \gamma_l(1 - \theta)^{\gamma_l - 1} \cdot (y_h - y_l)(1 - \gamma_h + \gamma_h \theta - \gamma_l \theta)(\delta - 1)}.$$

## C. 反事实状态模拟

反事实状态的模拟可以分 4 步进行, 这里只给出  $\rho_l = 1, \rho_h = 1$  时的求解模型的步骤, 不同  $\rho_h^c, \rho_l^c$  数值下的模型求解可通过类似步骤得到。

1. 给定  $\rho_l = 1, \rho_h = 1$ , 根据公式 (5) 计算新的投资收益  $B$ 。
2. 给定  $B$  及估计出的参数  $\alpha, \beta, \underline{a}$  和  $\bar{a}$ , 根据公式 (6) 计算人力资本投资概率  $G$ 。
3. 给定  $G$  及参数  $\rho_h^c, \rho_l^c$ , 根据公式 (8) 和 (9) 计算出新的大学生和非大学生中高技术劳动者的比例  $\delta, \delta^*$ 。在新的均衡下  $\delta, \delta^*$  与厂商的先验预期一致。
4. 给定  $\delta, \delta^*$  和参数  $\gamma = \{\gamma_l^c, \gamma_h^c, \gamma_l^*, \gamma_h^*\}$ , 根据工资决定方程模拟新的工资分布。

## D. 稳健性检验

为了对文章结论进行稳健性检验, 我们进一步控制了地域因素(省份虚拟变量)对工资的影响, 拟合如下回归方程:

$$\text{wage} = \beta_0 + \beta_1 \cdot \text{female} + \text{prov-dummies} \cdot \beta_3 + \epsilon, \quad (19)$$

同样取  $\hat{\omega} = \hat{\beta}_0 + \epsilon$  作为结构估计中的工资测度。表 D-1 给出了  $\hat{\omega}$  的描述统计, 需要注意的是, 此时经过处理的周工资序列存在负值。

表 D-1 工资变量的描述统计

周工资	均值	中位数	标准差	最小值	最大值	样本数
大学生	364.32	212.94	318.83	-34.91	1623.63	1772
非大学生	271.50	149.69	246.76	-64.16	1589.36	10919

注: 样本包括了下列估计中没有使用到的小于 1% 分位数以及大于 99% 分位数的观测值。

表 D-2 和 D-3 分别给出了工资决定方程和剩余模型参数的估计结果。可以看出, 由于工资序列的变化, 模型的参数的绝对数值与文章中差别很大, 但相对大小都比较接近。在结构估计使用到的 1%—99% 工资分位数的样本中, 大学生的平均工资为 353.15 元, 非大学生的工资为 263.92 元, 两者相差 89.23 元。令  $\rho_l = 1, \rho_h = 1$  进行反事实状态的模拟, 得到大学生的平均工资为 266.18 元, 非大学生的工资为 243.40 元, 两者相差 22.78 元。由此得到大学教育的能力溢价所能解释的大学生与非大学生收入差距百分比为:  $(89.23 - 22.78) \div 89.23 \times 100\% = 74.46\%$ , 这个数值也与文章中得到的 71.86% 的数值非常接近。综上所述, 可以认为文章的计量方法与估计结果是具有统计稳健性的。

表 D-2 工资决定方程中的参数估计

参数	大学生(标准差)		非大学生(标准差)	
$\delta$	0.343	(0.004)	0.307	(0.002)
$\gamma_h$	1.441	(0.027)	1.492	(0.010)
$\gamma_l$	1.632	(0.035)	1.540	(0.011)
$\rho_l \times y_l$	54.06	(3.360)	—	—
$\rho_h \times y_h$	1029.11	(50.82)	—	—
$y_l$	—	—	38.110	(1.137)
$y_h$	—	—	696.44	(11.52)
$\rho_l$	1.418	(0.291)	—	—
$\rho_h$	1.478	(0.076)	—	—
对数似然函数值	-11 143.64		-64 570.01	
样本量	1 736		10 712	

注:括号中为稳健标准差。其中  $y_l, y_h, \rho_l$  和  $\rho_h$  的标准差通过对全样本自举抽样得到,自举抽样次数为 1000。

表 D-3 模型剩余参数估计结果

参数	估计值(标准差)		参数	估计值(标准差)	
$\rho_h^c$	0.146	(0.002)	$\alpha$	317.76	(2.651)
$\rho_l^c$	0.127	(0.001)	$\beta$	-1.573	(0.009)
$\bar{a}$	566.17	(3.875)	$B$	88.68	(1.541)
$\underline{a}$	-69.34	(1.426)	$G$	0.338	(0.002)

注:参数估计的标准差通过对工资决定方程的参数估计值及其方差-协方差矩阵抽样产生,抽样次数为 1000。虽然这里  $\underline{a}$  的估计值为负,但不能说明模型拟合出现了异常。这主要是与  $a$  服从均匀分布的假设以及  $\alpha = (\bar{a} - \underline{a})/2$  的限制有关。Fang(2006)也得到了类似的结果。

## 参考文献

- [1] Acemoglu, D., "Technical Change, Inequality and the Labor Market", *Journal of Economic Literature*, 2002, 40(1), 7—72.
- [2] Antonovics, K., "Persistent Racial Wage Inequality", Working Paper, University of California-San Diego, 2002.
- [3] Arcidiacono, P., P. Bayer, and A. Hizmo, "Beyond Signaling and Human Capital: Education and the Revelation of Ability", Working Paper, Duke University, 2008.
- [4] Becker, G., *Human Capital*. New York: NBER, 1964.
- [5] Bedard, K., "Human Capital versus Signaling Models: University Access and High School Dropout", *Journal of Political Economics*, 2001, 109(4), 749—775.
- [6] Card, D., and J. Dinardo, "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles", *Journal of Labor Economics*, 2002, 20(4), 733—783.
- [7] Chevalier, A., C. Harmon, I. Walker, and J. Zhu, "Does Education Raise Productivity or Just Reflect It?" *Economic Journal*, 2004, 114(4), 499—517.
- [8] Cribari-Neto, F., and K. Vasconcellos, "Nearly Unbiased Maximum Likelihood Estimation for The Beta Distribution", *Journal of Statistical Computation and Simulation*, 2002, 72(2), 107—118.
- [9] Cribari-Neto, F., and S. Ferrari, "Beta Regression for Modeling Rates and Proportions", *Journal of Applied Statistics*, 2004, 31(7), 799—815.
- [10] 杜鹏, "我国教育发展对收入差距影响的实证分析",《南开经济研究》,2005年第4期,第47—52页。
- [11] Fang, H., "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices", *International Economic Review*, 2006, 47(4), 1151—1185.

- [12] Hendel, I., J. Shapiro, and P. Willen, "Educational Opportunity and Income Inequality", *Journal of Public Economics*, 89(5-6), 841-870.
- [13] 李实、丁赛, "中国城镇教育收益率的长期变动趋势", 《中国社会科学》, 2003年第6期, 第58-72页。
- [14] 刘泽云, "教育与工资不平等——中国城镇地区的经验研究", 《统计研究》, 2009年第4期, 第53-59页。
- [15] Mincer, J., *Schooling, Experience and Earnings*. New York: Columbia University Press, 1974.
- [16] Moro, A., "The Effect of Statistical Discrimination on Black-White Wage Inequality: Estimating a Model with Multiple Equilibria", *International Economic Review*, 2003, 44(2), 467-500.
- [17] Riley, J., "Silver Signals: Twenty-Five Years of Screening and Signaling", *Journal of Economic Literature*, 2001, 39(2), 432-478.
- [18] Spence, M., 1973, "Job Market Signaling", *Quarterly Journal of Economics*, 87, 355-374
- [19] Weiss, A., "Human Capital vs. Signalling Explanations of Wages", *Journal of Economic Perspectives*, 1995, 9(4), 133-154.
- [20] 杨俊、黄潇、李晓羽, "教育不平等与收入分配差距——中国的实证分析", 《管理世界》, 2008年第8期, 第38-47页。
- [21] 张车伟, "人力资本回报率变化与收入差距: '马太效应'及其政策含义", 《经济研究》, 2006年第12期, 第59-70页。

## Labor Market Discrimination and Consequences of the University Enrollment Expansion: A Structural Estimation Based on A Signaling Game

SHU XU

(*Southwestern University of Finance and Economics*)

**Abstract** We establish a model of signaling and explore the contribution of human capital accumulation to college wage premiums, as well as the effect of university enrollment expansion to the college-non-college wage differential. Results from the structural estimation indicate that the model performs very well in fitting the wage distribution. Counterfactual simulation suggests that (1) human capital accumulation accounts for nearly 72% of the college wage premium in China; (2) unlike the analysis from the perspective of labor supply, under very weak conditions the model shows that the university enrollment expansion enlarges college-non-college wage differentials by way of statistical discrimination.

**JEL Classification** C14, I21, J31