

读书何以改变命运？

——基于精英教育长期回报的考察

李逸飞 刘海琦 李静

附录 断点回归的有效性检验

断点回归的有效性检验包括以下几个方面：第一，驱动变量本身不受人为操纵，根据标准的检验方法，我们使用一本线附近考生（范围与基准回归中的带宽保持一致）的高考成绩进行了 McCrary 检验 (McCrary, 2008)。图 A1 是分数线的密度分布；可以看到，分数线的分布是比较平滑的正态分布，在一本线处并没有明显的断点且置信区间存在重叠。此外，McCrary 检验系数非常小且不显著，也从统计上证明了驱动变量（考生高考分数）的连续性。

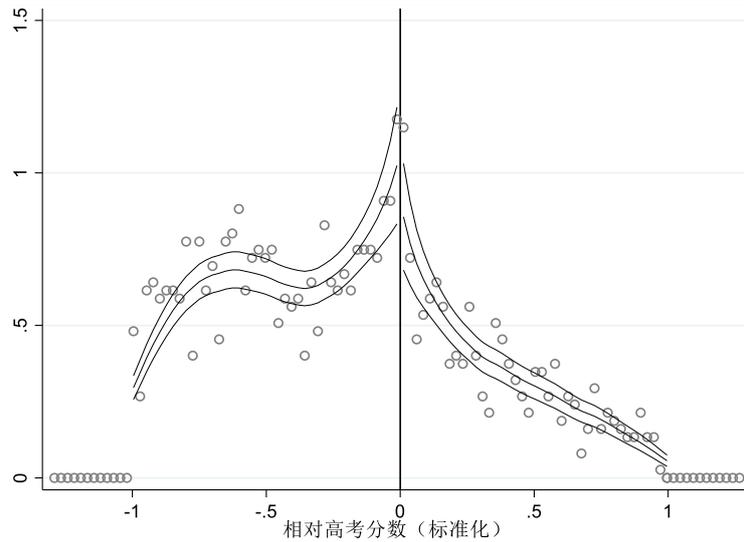


图 A1 断点回归检验：平滑性检验

第二，其他变量在断点两边应该平滑，没有明显的跳跃。如果断点的产生可以看做是接近随机的实验，断点两边的个体除了高考分数之外，则应该具有相似的特质。对此，我们将个体的性别、年龄、民族、高考前是否是农村户口、高考生源是否来自中西部省份等个人层面的特征变量作为被解释变量，使用与方程 (1) 一致的估计方法、估计参数，进行平衡性检验。由图 A2 可直观地观察到，断点两侧置信区间存在重叠，表 A1 结果进一步证明系数均在 10% 水平上不显著，表明了断点两边的个体差异不明显，符合断点回归的假设。

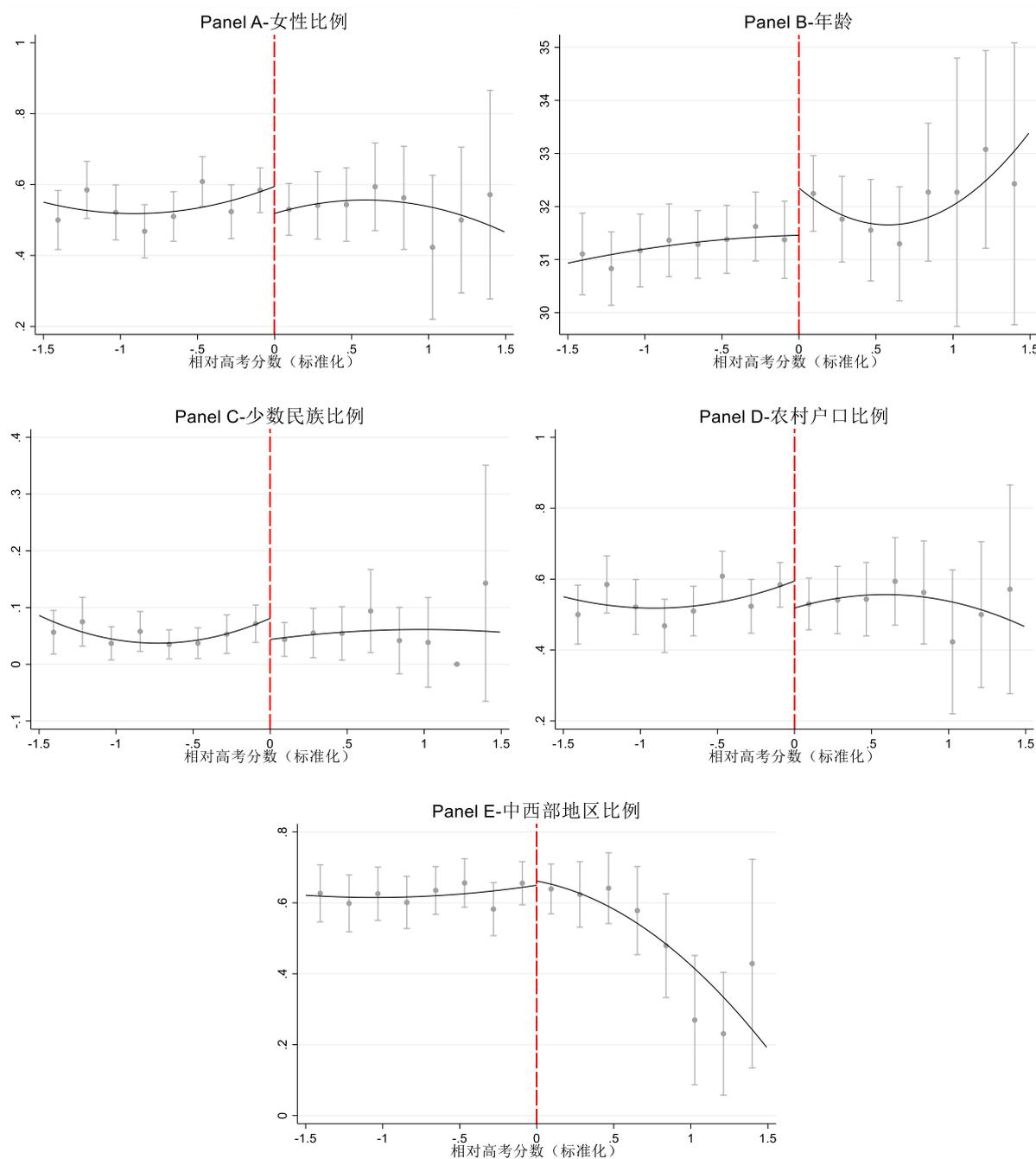


图 A2 个人特征变量的分数断点检验

注：Panel A-E 分别对应是否是女性、年龄、是否是少数民族、是否是农村户口、是否来自中西部地区作为被解释变量时的情形。

表 A1 个人特征变量的平衡性检验系数结果

被解释变量：	女性	年龄	少数民族	农村户口	中西部地区
	(1)	(2)	(3)	(4)	(5)
就读于重点大学=1	-1.34e-14 (5.24e-16)	0.385 (2.424)	-0.022 (0.314)	0.129 (0.187)	2.183 (1.457)

二次项	Yes	Yes	Yes	Yes	Yes
省级固定效应	Yes	Yes	Yes	Yes	Yes
带宽	1.395	1.393	1.559	1.395	1.396
样本数	1,334; 551	1,330; 551	1,703; 576	1,334; 551	1,334; 551

注：平衡性检验使用与方程（1）一致的估计方法、估计参数。最优带宽有 CCT 法确定，所选择的带宽范围与主回归加入所有控制变量与固定效应后（正文表 4-列 4）的带宽范围基本一致。其余同正文表 2。

第三，断点回归可以视为局部自然实验，带宽的范围对于自然实验的效度与系数估计的准确性均有很大影响。在正文中，全部最优带宽均由 CCT 法确定。在稳健性检验中，我们使用 IK 法，并分别调整带宽为 0.5 倍与 2 倍进行回归。回归结果如表 A2 所示，结果均保持显著。

表 A2 稳健性检验：带宽的调整

	(1)	(2)	(3)	(4)	(5)	(6)
	IK 法	0.5 倍	2 倍	IK 法	0.5 倍	2 倍
	最优带宽	最优带宽	最优带宽	最优带宽	最优带宽	最优带宽
	2018 年全年总收入（对数化）			2018 年主要工作时薪（对数化）		
就读于重点大学=1	1.989***	2.532***	1.676**	1.889***	2.234***	1.796**
	(0.640)	(0.354)	(0.691)	(0.525)	(0.400)	(0.696)
二次项	Yes	Yes	Yes	Yes	Yes	Yes
个人控制变量	Yes	Yes	Yes	Yes	Yes	Yes
省级固定效应	Yes	Yes	Yes	Yes	Yes	Yes
带宽	1.282	0.641	2.564	1.241	0.621	2.482
样本量	1,258; 532	688; 404	2,079; 578	1,227; 526	661; 395	2,043; 577

注：使用 IK 法确定最优带宽，并调整至 0.5 倍与 2 倍进行稳健性检验。其余同正文表 2。

参考文献

- [1] McCrary, J., "Manipulation of the Running Variable in The Regression Discontinuity Design: A Density Test", *Journal of Econometrics*, 2008,142(2), 698-714.

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。