

# 择偶偏好与同型婚配

## ——来自择偶偏好调查实验的证据

石泽华 周 扬

### 目录

|                                     |   |
|-------------------------------------|---|
| 附录 I 检验受访者群体的择偶偏好应用于未受访群体的合理性 ..... | 1 |
| 附录 II 择偶偏好调查实验的随机性检验 .....          | 2 |
| 附录 III 稳健性检验 .....                  | 4 |
| 1. 限制样本为双收入家庭样本 .....               | 4 |
| 2. 对随机效用项的稳健性检验 .....               | 5 |
| 附录 IV 择偶偏好的婚龄以及年龄异质性检验 .....        | 6 |

# 附录 I 检验受访者群体的择偶偏好应用于未受访群体的合理性

本研究利用 CGSS 2021 构造的配对样本中，仅受访者群体参与了择偶偏好调查实验并由此分别估计出男性和女性的择偶偏好，受访者伴侣在 CGSS 2021 中属于未受访群体，并未参与调查。在构建未受访者（受访者伴侣）对潜在配偶的偏好排序时，需假定男性和女性未受访者的择偶偏好与其同性别的受访者的择偶偏好一致而进行外推。为此，我们检验了受访者群体的择偶偏好应用于未受访群体的合理性。具体而言，本文展示了分性别的受访者群体与未受访者群体在年龄、收入以及教育分布图（见图 I1）。可以看出，受访者和未受访者在三个维度上均基本一致。

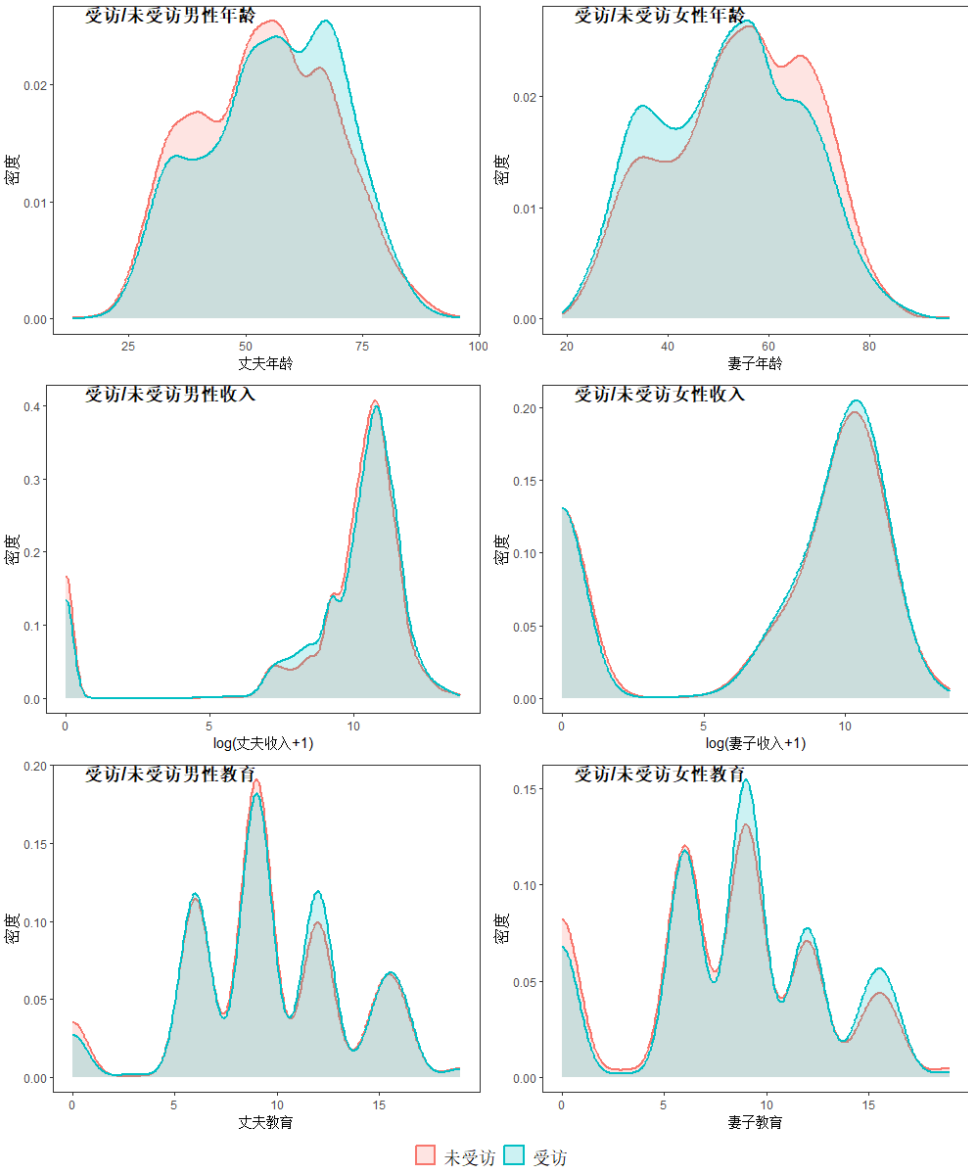


图 I1 受访/未受访女性与受访/未受访男性在年龄、收入和教育上的分布

## 附录 II 择偶偏好调查实验的随机性检验

对择偶偏好调查实验的随机性检验包括虚拟配偶特征生成的随机性和分配的随机性两个方面。图 II1 展示的是基于受访者性别而生成的虚拟配偶在年龄、收入、教育三项择偶维度上的特征分布以及三项维度特征组合的联合分布情况，均基本满足均匀分布，从而证实了虚拟配偶在各（组合）维度生成上具有随机性。

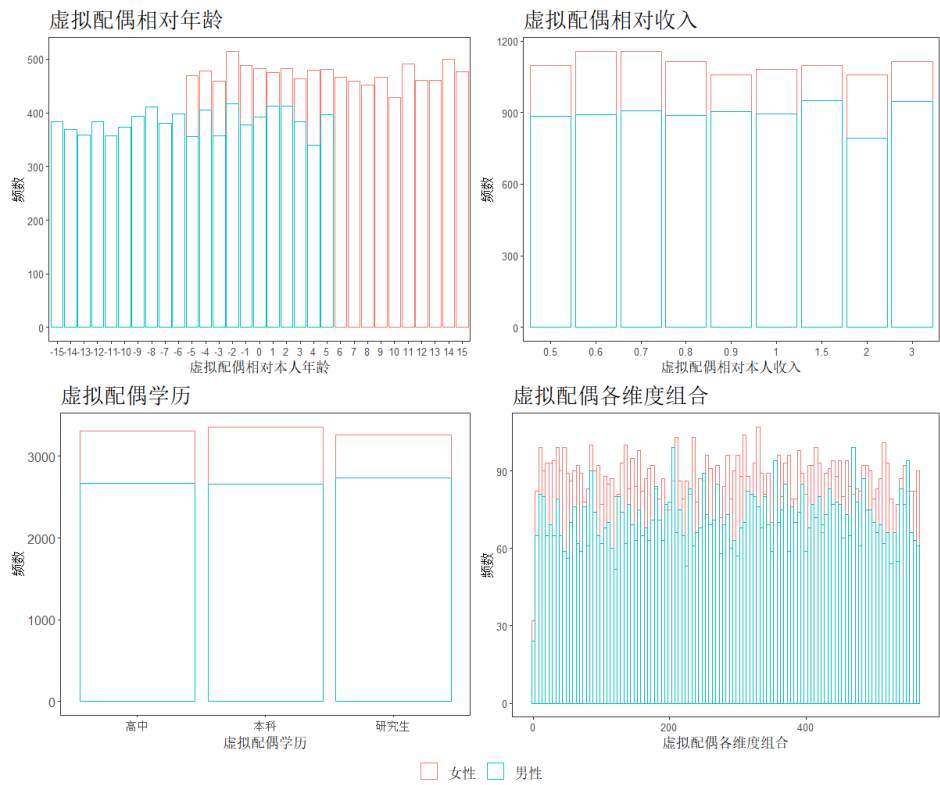


图 II1 分受访者性别的虚拟配偶在年龄、收入、教育及三维度组合上的分布

进而需要检验分配给各受访者的虚拟配偶（相对）特征是否与受访者本人特征无显著性关联，结果见表 II1。对于女性受访者来说，被分配的虚拟配偶无论是相对年龄（“与本人的年龄差几岁”）、相对收入（“是本人收入的几倍”）还是教育维度上的特征值对受访者本人年龄、收入、教育和户口特征回归的联合统计检验 F 统计值均不显著。对于男性受访者来说，虚拟配偶的相对收入和教育维度均与受访者本人无关，仅虚拟配偶的相对年龄与受访者教育存在一定程度的正向关系。总体上看，可以认为调查实验中虚拟配偶是被随机分配给每一位受访者的。

表 111 分受访者性别的虚拟配偶的分配随机性检验结果

| 变量             | 女性受访者的虚拟配偶        |                   |                   | 男性受访者的虚拟配偶        |                   |                   |
|----------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                | 相对本人的<br>年龄差异     | 相对本人的<br>收入倍数     | 教育                | 相对本人的<br>年龄差异     | 相对本人的<br>收入倍数     | 教育                |
|                | (1)               | (2)               | (3)               | (4)               | (5)               | (6)               |
| 受访者年龄          | 0.002<br>(0.005)  | -0.000<br>(0.001) | 0.003<br>(0.002)  | 0.007<br>(0.006)  | 0.000<br>(0.001)  | -0.002<br>(0.003) |
| 受访者收入          | 0.020<br>(0.013)  | -0.002<br>(0.002) | 0.001<br>(0.006)  | 0.006<br>(0.012)  | -0.001<br>(0.002) | -0.005<br>(0.006) |
| 受访者教育          | 0.013<br>(0.019)  | 0.004+<br>(0.002) | 0.017*<br>(0.009) | 0.047*<br>(0.022) | 0.002<br>(0.003)  | 0.017<br>(0.011)  |
| 受访者户口 (城市=1)   | -0.162<br>(0.156) | -0.018<br>(0.020) | -0.081<br>(0.074) | 0.133<br>(0.160)  | 0.013<br>(0.021)  | 0.013<br>(0.076)  |
| 联合显著性检验        |                   |                   |                   |                   |                   |                   |
| F 统计值          | 0.942             | 1.188             | 1.080             | 2.598             | 0.477             | 1.119             |
| P 值            | 0.439             | 0.314             | 0.365             | 0.034*            | 0.753             | 0.346             |
| N              | 9957              | 9957              | 9957              | 8064              | 8064              | 8064              |
| R <sup>2</sup> | 0.000             | 0.001             | 0.000             | 0.001             | 0.000             | 0.001             |

注：括号内为标准误，\*\*\*p<0.001，\*\*p<0.01，\*p<0.05，+p<0.1。

## 附录 III 稳健性检验

### 1. 限制样本为双收入家庭样本

考虑到比较 GS 算法模拟的稳定匹配与真实匹配时（正文表 3），相较于年龄和教育两项择偶维度，无论是配对样本中的真实匹配还是 GS 模拟得到的稳定匹配中，配偶双方的收入相关性均较低。这可能是由于配偶双方存在一方无收入的情况，附录 I 图 I1 也印证了样本中确实存在着大量无收入的个体（收入加 1 的对数在 0 处的密度较高），从而导致收入匹配上的低相关性。基于此，我们进一步观察全样本 GS 模拟匹配中的双收入配对以比较真实匹配中的双收入家庭。根据表 III1（1）和（2）列的结果，真实匹配中的双收入家庭，年龄和教育维度上配偶双方特征的相关性基本与全样本无差异，收入维度上的相关性确实有所上升，由 0.464 上升至 0.503。而在 GS 模拟匹配中，稳定匹配的双收入配偶在收入和教育维度上的相关性较全样本时有一定的提高，且教育相关性的提升幅度更高。

表 III1 双收入配偶的真实匹配与 GS 模拟匹配中的特征相关性

| 配偶双方特征相关性 | 真实匹配  |       | GS 模拟匹配          |                  |                  |                  |
|-----------|-------|-------|------------------|------------------|------------------|------------------|
|           |       |       | 男性最优             |                  | 女性最优             |                  |
|           | 全样本   | 双收入家庭 | 全样本              | 双收入家庭            | 全样本              | 双收入家庭            |
|           | (1)   | (2)   | (3)              | (4)              | (5)              | (6)              |
| 年龄        | 0.964 | 0.956 | 0.922<br>(0.006) | 0.912<br>(0.006) | 0.984<br>(0.001) | 0.981<br>(0.003) |
| 收入        | 0.464 | 0.503 | 0.406<br>(0.031) | 0.411<br>(0.036) | 0.841<br>(0.016) | 0.704<br>(0.020) |
| 教育        | 0.657 | 0.655 | 0.404<br>(0.026) | 0.493<br>(0.023) | 0.641<br>(0.025) | 0.702<br>(0.021) |

注：相关性汇报的是 Pearson 相关系数。其中，GS 算法预测匹配双方各维度上的相关系数是 100 次模拟匹配后求取的均值，括号内是这 100 次模拟匹配后所获相关系数的标准差。

## 2. 对随机效用项的稳健性检验

由于正文公式 (1) 中的误差项 ( $\epsilon_{mw}$ ) 是择偶效用函数中的随机效用项, 无法被观测。因此, 我们在构建偏好排序时, 为每个个体从正态分布中随机抽取了随机效用值。但换一种思考角度, 误差项也可能纯粹是受访者在择偶偏好调查实验中出现的噪声, 即当受访者给每位虚拟配偶评分时, 他们会在评分时偶尔出错。而这些可能出现的随机性错误会在真实匹配时得到纠正。因此, 在利用由实验获得的择偶偏好来进行 GS 模拟匹配以对比真实婚姻匹配时, 应去除掉这些噪声, 也即将误差项设定为零。结果见表 III2, 随机效用项设定为零时获得的稳定匹配结果与主结果基本一致。

表 III2 对随机效用项的稳健性检验

| 配偶双方<br>特征<br>相关性 | GS 模拟匹配          |                  |                  |                  |
|-------------------|------------------|------------------|------------------|------------------|
|                   | 男性最优             |                  | 女性最优             |                  |
|                   | 主结果              | 随机效用项为零          | 主结果              | 随机效用项为零          |
|                   | (1)              | (2)              | (3)              | (4)              |
| 年龄                | 0.922<br>(0.006) | 0.906<br>(0.006) | 0.984<br>(0.001) | 0.968<br>(0.004) |
| 收入                | 0.406<br>(0.030) | 0.403<br>(0.035) | 0.841<br>(0.016) | 0.848<br>(0.012) |
| 教育                | 0.404<br>(0.026) | 0.391<br>(0.029) | 0.641<br>(0.025) | 0.627<br>(0.027) |

## 附录IV 择偶偏好的婚龄以及年龄异质性检验

首先，考虑到已婚样本实际上是婚后参与的调查实验，其择偶偏好存在被其现实配偶和婚姻生活所改变的可能性，由此才得到了本文中择偶偏好形成的GS稳定匹配与现实匹配较为一致的结果。针对这一可能性解释，在一定程度上可以通过考察样本中新婚群体和久婚群体在择偶偏好上是否存在异质性而进行检验，其检验逻辑在于，择偶偏好如果会被其现实配偶和生活境遇所长期塑造，则需相对更长的婚龄来展现，而那些新近结婚群体的择偶偏好可能尚未显著地受其现实配偶或婚姻生活境遇所塑造，因此，若新婚群体和久婚群体在择偶偏好上没有显著的差异，则在一定程度上可以认为，虽然择偶偏好随机实验是已婚被试在其婚后参与的调查实验，但仍能在平均意义上反映其择偶偏好。

为此，我们将样本依据婚龄分为新婚群体（婚龄在三年及以内）和久婚群体（婚龄大于三年），在正文公式（1）的个人固定效应模型基础上，加入各项择偶维度特征变量（包括水平偏好和垂直偏好）与受访者婚龄变量（“婚龄 $\leq 3$ ”=1，“婚龄 $> 3$ ”=0）的交互项，用以检验择偶偏好在婚龄上的异质性。为了方便阅读，这里仅展示各个交互项的估计结果，详见表IV1：对于男性受访者，新婚男性和久婚男性对各个择偶维度特征的择偶偏好在95%的置信度下并无显著性差异，仅在负向年龄差异项具有边际显著性（ $p < 0.1$ ）。

表 IV1 择偶偏好的婚龄异质性检验

|                                   | 女性受访者             | 男性受访者              |
|-----------------------------------|-------------------|--------------------|
|                                   | (1)               | (2)                |
| 虚拟配偶年龄*受访者婚龄 $\leq 3$             | 0.023<br>(0.061)  | -0.038<br>(0.065)  |
| 正向年龄差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | -0.004<br>(0.005) | -0.011<br>(0.029)  |
| 负向年龄差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | -0.004<br>(0.024) | -0.008+<br>(0.005) |
| 虚拟配偶收入*受访者婚龄 $\leq 3$             | 0.014<br>(0.026)  | 0.000<br>(0.012)   |
| 正向收入差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | 0.018<br>(0.101)  | 0.165<br>(0.113)   |
| 负向收入差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | -0.423<br>(1.425) | 0.825<br>(1.228)   |
| 虚拟配偶教育*受访者婚龄 $\leq 3$             | -0.016<br>(0.083) | -0.110<br>(0.081)  |
| 正向教育差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | 0.000<br>(0.008)  | 0.016<br>(0.008)   |
| 负向教育差异 $\wedge 2$ *受访者婚龄 $\leq 3$ | -0.019<br>(0.025) | -0.030<br>(0.031)  |
| R <sup>2</sup>                    | 0.944             | 0.955              |
| N                                 | 7488              | 6009               |

注：仅展示交互项系数估计结果。括号内为标准误，\*\*\* $p < 0.001$ ，\*\* $p < 0.01$ ，\* $p < 0.05$ ，+ $p < 0.1$ 。

考虑到已婚样本平均年龄偏大，我们也检验了择偶偏好的年龄异质性。具体来说，在正文公式（1）的个人固定效应模型基础上，加入各项择偶维度特征变量（包括水平偏好和垂直偏好）与受访者年龄变量的交互项，用以检验择偶偏好在年龄上的异质性。为了方便阅读，这里仅展示各个交互项的估计结果，详见表IV2：不同年龄男性对各个择偶维度特征上的择偶偏好在95%的置信度下并无显著性差异；对于女性受访者，仅在虚拟配偶教育

项具有边际显著性 ( $p<0.1$ )。

表 IV2 择偶偏好的年龄异质性分析结果

|                | 女性受访者             | 男性受访者             |
|----------------|-------------------|-------------------|
|                | (1)               | (2)               |
| 虚拟配偶年龄*受访者年龄   | -0.001<br>(0.001) | 0.000<br>(0.001)  |
| 正向年龄差异^2*受访者年龄 | 0.000<br>(0.001)  | 0.001<br>(0.000)  |
| 负向年龄差异^2*受访者年龄 | 0.000<br>(0.000)  | 0.000<br>(0.000)  |
| 虚拟配偶收入*受访者年龄   | -0.001<br>(0.001) | -0.000<br>(0.000) |
| 正向收入差异^2*受访者年龄 | 0.001<br>(0.001)  | -0.001<br>(0.001) |
| 负向收入差异^2*受访者年龄 | 0.007<br>(0.016)  | 0.010<br>(0.016)  |
| 虚拟配偶教育*受访者年龄   | 0.002+<br>(0.001) | 0.001<br>(0.000)  |
| 正向教育差异^2*受访者年龄 | -0.000<br>(0.000) | -0.000<br>(0.000) |
| 负向教育差异^2*受访者年龄 | 0.000<br>(0.001)  | 0.000<br>(0.001)  |
| R <sup>2</sup> | 0.945             | 0.955             |
| N              | 9957              | 8064              |

注：仅展示交互项系数估计结果。括号内为标准误，\*\*\* $p<0.001$ ，\*\* $p<0.01$ ，\* $p<0.05$ ，+ $p<0.1$ 。

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。