

NBER WORKING PAPER SERIES

CHOOSING YOUR POND:  
LOCATION CHOICES AND RELATIVE INCOME

Nicolas L. Bottan  
Ricardo Perez-Truglia

Working Paper 23615  
<http://www.nber.org/papers/w23615>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2017

We are thankful for comments from David Albouy, Dan Bernhardt, Jeff Brown, Leonardo Bursztyn, Mikhail Galashin, Craig Fox, Ori Heffetz, Erzo F.P. Luttmer, Ben Marx, Gautam Rao, Alex Rees-Jones, Chris Roth, Melanie Wasserman and seminar discussants at UIUC, Cornell, University of Warwick, Loyola Marymount, Tulane, University of Arkansas, Northwestern, Binghamton University, the Federal Reserve Bank of New York, the University of Pennsylvania, the 2017 SITE Workshop, the 2017 Advances with Field Experiments Conference, the 2017 NTA Annual Conference on Taxation, the 2017 SEA Annual Meeting, the 2017 GEMBPP Workshop and the 2018 NBER Summer Institute. We thank UCLA-Anderson and its Behavioral Lab for providing funding for the experiment; and thank the support from the Robert Ferber Dissertation Award. This project was reviewed and approved in advance by the Institutional Review Board at University of California Los Angeles (IRB #16-001968; #17-001449). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Nicolas L. Bottan and Ricardo Perez-Truglia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Choosing Your Pond: Location Choices and Relative Income  
Nicolas L. Bottan and Ricardo Perez-Truglia  
NBER Working Paper No. 23615  
July 2017, Revised August 2018  
JEL No. D62,D91,I31,Z13

### **ABSTRACT**

We propose that, when individuals are deciding where to live, they care about the position in the income distribution in their prospective location. To test this hypothesis, we develop a new methodology to estimate preferences over location characteristics that combines choice data, survey data and information-provision experiments. We implement this methodology with a sample of 1,080 senior medical students who participated in the National Resident Matching Program. These subjects were choosing the cities to which they would move and live in for the next five years. We provide suggestive evidence that individuals care about what their relative income would be in a prospective city, and that those preferences have substantial heterogeneity by relationship status. Using a separate subject pool, we replicate our findings and provide tests for the underlying mechanisms.

Nicolas L. Bottan  
Department of Economics  
University of Illinois at Urbana-Champaign  
1407 W Gregory Dr.  
214 David Kinley Hall, Mc-707  
Urbana, IL 61801  
bottan2@illinois.edu

Ricardo Perez-Truglia  
Anderson School of Management  
University of California, Los Angeles  
110 Westwood Plaza  
Los Angeles, CA 90095  
and NBER  
ricardo.truglia@anderson.ucla.edu

A data appendix is available at <http://www.nber.org/data-appendix/w23615>

# 1 Introduction

Location choices have been the subject of abundant economic research. A large body of work has explored location preferences in relation to a number of city amenities, such as cost of living, quality of schools, and climate. In this paper, we develop a new methodology to measure preferences for city amenities, and use it to measure preferences for a city amenity that researchers had not yet explored: an individual’s position in the income distribution.

There are a number of reasons why individuals may care about their relative income. On the one hand, the income of neighbors can impose a negative externality, giving individuals an incentive to locate into less affluent areas. This prediction is borne out by, for instance, models of social status and consumption aspirations. On the other hand, neighbors’ income can impose a positive externality, giving individuals incentives to locate into more affluent ponds. This prediction is borne out by models of local public goods and dating, for example.

The identification of preferences for any set of amenities, not only relative income, faces two fundamental challenges. The first challenge consists of observing not only the city and amenities that the individual chose, but the cities and amenities that the individual could have chosen instead. For instance, although some datasets identify individuals moving from one city to another,<sup>1</sup> they do not identify the other cities that the individual could have chosen or incomes they could have earned in those other locations. The second challenge is that we need sources of exogenous variation to identify the amenity preferences. In other words, preferences for a given amenity could be spuriously driven by preferences for another amenity that is not included in the model.

We develop a methodology for estimating location preferences that addresses these two identification challenges. Our methodology combines choice data, survey data, and information-provision experiments. While this methodology can be applied in a wider set of contexts, we implement it in a unique context: the National Resident Matching Program (NRMP) in the United States.

The NRMP uses an algorithm to pair graduating medical students with postgraduate hospital residency programs based on rankings submitted by both students and hospitals. When submitting their program rankings, the students choose the city where they will live for the duration of the residency (typically around five years). Several features of the NRMP make it desirable for this type of revealed-preference analysis (Benjamin, Heffetz, Kimball, and Rees-Jones 2014). First, it is possible to identify and specify the entire choice set faced by these individuals. Second, because most students are aware of the incentive-compatible matching algorithm used by the NRMP, it is possible to infer preferences directly, without the

---

<sup>1</sup>For instance, the National Survey of Families and Households (Luttmer 2005), or the United States Postal Service’s National Change of Address database (Perez-Truglia 2017).

need for estimating models that rely on additional assumptions. Third, this is a high-stakes choice, to which participants devote ample time and attention. A final, convenient aspect of our setting is that the graduating students choose between programs that offer almost identical nominal incomes, but in cities with largely different costs of living and income distributions. As a result, our subjects face substantial tradeoffs between cost of living and relative income.

We conducted the experiment with 1,080 senior medical students who participated in the 2017 Main Residency Match. The survey asked participants to consider their top two favorite residency programs. Even though students can rank several programs, the ranking of their top two is, in practice, the most important aspect of their decision. We elicited perceptions about two attributes of the cities where these two programs are located: their expected cost of living and their expected position in the city-wide distribution of individual earnings (i.e., their relative income). We then elicited each subject’s expected rank submission. Using these data on perceptions and choices, we can estimate how differences in the cities’ costs of living and relative incomes affect location choices.

These estimates of location preferences may be subject to omitted variable biases because these observable attributes can be correlated to some unobservable attributes excluded from the model. We use information-provision experiments to address this identification challenge. Immediately after eliciting perceptions about cost of living and earnings rankings, we provided all individuals with statistics about attributes of the cities where their programs are located. We randomized the value of this feedback in a non-deceptive way by randomizing the data source used to compute these statistics. For instance, students considering earning \$54,000 at a residency in Champaign-Urbana, IL, received one of two messages: their earnings rank would be 55.1% according to data from the Current Population Survey or 60.3% according to data from the American Community Survey. We then elicited their perceptions again after providing this feedback (i.e., the posterior beliefs). This source-randomization experiment creates exogenous variation in posterior beliefs. Our research design exploits that exogenous variation to estimate the causal effects of the perceived attributes on choice by means of an instrumental variables model.<sup>2</sup>

Our baseline estimates, which use experimental and non-experimental variations in perceptions, suggest that a 1-percentage-point decrease in cost of living increases the probability that a program will be chosen by 0.201 percentage points. This effect implies a behavioral elasticity of 0.201. Our baseline estimates suggest that, holding cost of living constant, individuals also care about relative income: on average, an increase of 1 percentage point in

---

<sup>2</sup>An alternative but equally valid approach consists of using a single source of feedback, but showing feedback to half of the subjects, chosen at random – see Cullen and Perez-Truglia (2018), for example.

earnings rank increases the probability that a program will be chosen by 0.186 percentage points – this effect implies a behavioral elasticity of 0.186.

The importance of financial aspects such as cost of living and relative income is consistent with survey data indicating that students chose medical school because of the financial rewards (Daniel and O’Brien 2008). Additionally, these findings are consistent with survey data indicating that, when choosing among programs, applicants care a lot about the geographic location of the programs (Sledge, Leaf, and Sacks 1989). Indeed, when asked to mention the most important factors they consider in their decision-making, 52% of surveyed applicants explicitly mention cost of living (NRMP 2015).<sup>3</sup>

These baseline estimates suggest that cost of living and relative income are both economically relevant for individuals in the sample and to a similar degree: the average individual would be willing to accept an increase of 0.925 percentage points in the cost of living in exchange for a 1 percentage point increase in earnings rank.

When it comes to preferences for cost of living, individuals have similar preferences regardless of whether they are single or non-single. On the other hand, preference for relative income has meaningful heterogeneity by relationship status: while non-singles prefer locating to less affluent ponds, single individuals prefer locating to more affluent ponds. This difference in preferences is large in magnitude and highly statistically significant ( $p$ -value  $< 0.001$ ). Moreover, this heterogeneity is consistent with prior evidence from the happiness literature: Luttmer (2005) finds that the positive effects of relative income on happiness are driven entirely by non-singles.

The baseline estimates of preferences for relative income satisfy several robustness checks. First, they are not sensitive to the inclusion of multiple residency and location characteristics as additional control variables. Second, the baseline estimates are consistent with the experimental estimates that focus on the variation generated by the information provision experiment. Third, we find that the information related to earnings rank, which is provided roughly one month before the submission deadline, has a lasting effect on the final rank order submitted to the NRMP.

We provide additional robustness checks based on an auxiliary experiment conducted with a sample of respondents recruited through Amazon Mechanical Turk. This sample has some disadvantages. For example, respondents are not planning to move anytime soon, so we must study their hypothetical location choices instead of their actual location choices. The auxiliary sample does have some advantages, however: it is more diverse than the primary

---

<sup>3</sup>Although these preferences for cost of living and relative income are statistically and economically significant, they do not imply that they are the main concern for medical students. Indeed, using perceptions about other characteristics of the programs, we find that doctors care substantially more about prestige and career prospects.

sample of medical students in many dimensions such as age and occupation; and it is possible to run further experiments on demand.

The results from this second survey replicate the main findings from the original experiment: we find that the average individual prefers a lower cost of living and a higher relative income; we find that the experimental estimates are similar to the non-experimental estimates; and we find large differences in preference for relative income by relationship status. Despite substantial differences in the distribution of observable characteristics between the two samples, the average marginal rate of substitution between relative income and cost of living are in the same order of magnitude, and their difference is statistically insignificant.

In the last part of the paper, we provide evidence about the potential interpretation for the estimated preferences for relative income. One potential explanation is that individuals do not care about their relative income *per se*, but that they use it as a signal for other unobservable attributes of a city. We think this possibility is unlikely in this high-stakes and high-information environment. Because they devote two months to making this decision, students can learn this information directly rather than relying on imperfect signals. We nonetheless provide two tests of this mechanism.

The first test explores whether individuals react to information about relative income because it teaches them about the expected cost of living. For example, if the students expect to compete with their neighbors for certain goods, like housing, then it may be natural for them to make such inferences. Contrary to this hypothesis, we show that feedback about earnings rank has a zero and precisely estimated effect on expected cost of living, both in the short term (the baseline survey) and in a longer horizon (the follow-up survey).

The second test explores whether individuals react to information about relative income because it teaches them about location characteristics such as public goods and crime rates. Although these inferences would not be unreasonable, this mechanism would probably run counter to our main finding: if more affluent ponds tend to have more desirable amenities, then the average individual should prefer to live in more affluent ponds, which is the opposite of what we find. To explore this channel, we extended the survey instrument of the auxiliary sample so that, after individuals received feedback about relative income, it elicits perceptions about several city amenities such as the quality of public goods and crime rates. We find that controlling for these other beliefs does not change the results significantly.

Our favorite explanation for the preference for less affluent ponds – though it is by no means the only explanation – is that individuals see their neighbor’s income as a negative externality. Moving to a more affluent pond, for instance, may result in a loss of social status (Luttmer 2005), a loss of self-esteem (Festinger 1954), higher consumption aspirations (Frank 1985b), and worse outcomes in social interactions (Doob and Gross 1968; Fennis

2008; Nelissen and Meijers 2011). Regarding the heterogeneity by relationship status, our preferred interpretation is that single individuals expect more positive externalities from affluent neighbors, such as better public goods and better dating prospects (Fisman et al. 2006; Gautier, Svarer, and Teulings 2010).

Our paper is related to several strands of literature. First, it relates to literature on the importance of relative income for subjective well-being. Since the seminal contribution by Easterlin (1974), several studies have argued that subjective well-being depends on relative, rather than absolute, income (Van de Stadt, Kapteyn, and Van de Geer 1985; Clark and Oswald 1996; Luttmer 2005; Ferrer-i Carbonell 2005; Perez-Truglia 2016).<sup>4</sup> Luttmer (2005), for instance, uses data from the United States to show that, holding own income constant, happiness increases with relative income in the area of residence. We contribute to this literature in two ways. Our main contribution is to develop a method to test the relative income hypothesis by using revealed-preference data instead of happiness data. Additionally, even if we take as granted that the income of neighbors impose a negative externality on happiness, it is unclear whether individuals anticipate these externalities when deciding where to live. Our results provide suggestive evidence that some individuals may anticipate these externalities, at least partially.

This paper is related to a series of survey and laboratory studies. Some studies use surveys that let subjects choose between pairs of hypothetical scenarios that encompass tradeoffs between income and status or between positional and non-positional goods. These studies find that individuals are sometimes willing to exchange absolute income for higher status (e.g., Solnick and Hemenway 1998; Johansson-Stenman, Carlsson, and Daruvala 2002; Yamada and Sato 2016; Clark, Senik, and Yamada 2017). Similarly, some laboratory experiments show that relative standing affects decision-making in lottery games and sharing games (e.g., Kuziemko, Buell, Reich, and Norton 2014). We contribute to this literature by estimating the tradeoffs between absolute and relative income in a real-world, high-stakes context.

This study is related to a strand of literature on conspicuous consumption, according to which individuals use consumption of highly visible goods as costly signals of their income. These papers show evidence consistent with a series of predictions of the signaling model. Charles, Hurst, and Roussanov (2009), for instance, show that visible consumption declines with the average income in the reference group. Heffetz (2011) shows that income elasticities can be predicted from the visibility of consumer expenditures. And Bursztyn, Ferman, Fiorin, Kanz, and Rao (2017) show that demand for platinum credit cards exceeds demand for a

---

<sup>4</sup>These studies often use a slightly different specification: holding own income constant, well-being decreases with the average income in the group of reference. It must be noted that some studies find the opposite effect (Senik 2004) or mixed evidence (Clark, Westergård-Nielsen, and Kristensen 2009). For an extensive review of the literature, see Tideman, Frijters, and Shields 2008.

nondescript control product with identical benefits, suggesting a demand for a signal of income.<sup>5</sup> All of these papers take the reference groups as exogenous and given. As noted by Frank (1985a), however, the prospective relative income should also affect the choice of reference group. We contribute to this literature by being the first to provide a test of this long-standing hypothesis.

Our paper is related to a large literature measuring preferences for city amenities (e.g., Albouy 2008, 2016). Starting with the neoclassical model pioneered by Rosen (1979) and Roback (1982), economists have been using structural models to estimate the value of amenities. While a few studies rely on quasi-experimental identification (Bayer, Ferreira, and McMillan 2007), most studies face challenges with omitted variable biases. Take preferences for cost of living, for example. Holding everything else constant, individuals should be more attracted to cities with a lower cost of living. However, this preference for cost of living is difficult to estimate from naturally-occurring data, because cost of living is correlated to other city characteristics such as labor productivity (Moretti 2013) and quality of life (Diamond 2016). We contribute to this literature by developing a new method to estimate preferences for city amenities.

Last, our study is also related to a strand of literature documenting that individuals substantially misperceive their relative incomes (e.g., Cruces, Perez-Truglia, and Tetaz 2013; Karadja, Mollerstrom, and Seim 2017). This literature shows that correcting these misperceptions has significant effects on stated preferences for redistribution. Yet there is no evidence that these misperceptions have a significant effect on behavior. We contribute to this literature by filling this gap.

The rest of the paper proceeds as follows. Section 2 describes the survey design. Section 3 presents the econometric model. Section 4 presents implementation details and descriptive statistics. Section 5 discusses the distribution of perceptions and learning. Section 6 presents results on location preferences. Section 7 presents the results from the auxiliary survey experiment. The last section concludes.

## 2 Survey Design

### 2.1 Timing of the Surveys

After graduating from medical school, students have to complete a residency to become a *Medicinae Doctor* (MD). A residency usually lasts from three to seven years, after which individuals may obtain their medical license. During the fall semester of 2016, fourth year

---

<sup>5</sup>For additional experimental evidence, see Roth (2015).

medical school students started their participation in the residency match by submitting applications to residency programs. Later in the semester, they were interviewed and flown out by some of the programs they applied to.<sup>6</sup> After all interviews were completed, the students spend almost two months deciding how to rank their favorite programs. Students have visited the cities during their interviews, and sometimes visit them again during this period. During this time, deciding on the rank order preference is the students' top priority: applicants claim to collect a lot of information to aid their decision, such as characteristics of the residency programs and characteristics of the cities where the programs are located.

We follow Benjamin et al. (2014) in using this context to study preferences. They conducted a survey of medical students after the students had submitted their rankings to the NRMP. The survey measured the submitted rankings as well as the perceived rank of many aspects of the programs, such as life satisfaction, happiness, and sense of control. In that study, Benjamin et al. (2014) measure and compare the preferences inferred from rank choices to those inferred from subjective well-being. We follow the survey collection method from Benjamin et al. (2014) closely, but we change the survey itself to test a different hypothesis, that is, whether individuals make a trade-off between relative income and cost of living. In doing so, we deviate from the survey design in an important aspect: we collect our baseline survey *before* subjects submit their rank choices to the NRMP. We also embed an information-provision experiment, which allows us to address concerns on causality.

In the 2017 Match, the submission window for rank order lists opened on January 15 and closed on February 22. We conducted a baseline survey weeks before the submission deadline, which we describe first, and a follow-up survey right after the submission window closed.

## 2.2 General Structure of Baseline Survey

The baseline survey starts and ends with some background questions, such as the subject's medical school and marital status (see Appendix B.1 for the full questionnaire of the baseline survey). The core of the survey comprises the following group of questions, in the order listed below:

1. Choice Set: Elicit the names of the two favorite programs that the individual was considering for his or her order rank submission.
2. Prior Beliefs: Elicit perceptions about the cost of living and the earnings rank in the cities where these two programs are located.

---

<sup>6</sup>In 2015, the median number of applications submitted was 30 and the median number of interviews 16 (NRMP 2015).

3. Feedback: Provide subjects with feedback related to their perceptions.
4. Posterior Beliefs: Re-elicite perceptions about the cost of living and the earnings rank.
5. Rank Choice: Elicite the individual’s expected rank submission (between the two programs).

The following sections provide details about each of these modules.

### 2.2.1 Choice Set

The survey asks individuals to list their top two preferred programs, in no particular order, from a user-friendly list of all the available programs organized by state and metro area. We limited the survey to two programs because otherwise it would have been too cognitively demanding. Most participants expect to be matched to one of their top-two choices: similar to previous years, 50.9% of the participants in the 2017 match were assigned to their first choice and 16.6% were assigned to their second choice. We concentrated on the participants’ top two programs rather than a random pair of programs because this happen to be the part of the decision with the highest stakes and to which individuals were paying the most attention. In any case, our focus on the top two choices does not challenge the validity of our estimates: the research design would be valid with any pair (or group) of options, not only the top-two.<sup>7</sup>

### 2.2.2 Perceptions about Cost of Living and Earnings Rank

One important feature of the residency match process is that salaries are relatively homogeneous across the different programs, even across specialties.<sup>8</sup> Indeed, each program offers the same salary to all its candidates (and that salary is often publicly available on the program’s website). Despite the homogeneity in nominal incomes, there is large heterogeneity in costs of living and earnings distributions in the cities where the programs are located. When designing the survey, we were constrained to using metropolitan areas rather than other geographical levels of aggregation (e.g., commuting zones) because the sources of data on cost of living are not collected at a finer level than the metro area.

---

<sup>7</sup>When individuals were listing the second program, we required respondents to make a selection from a different metro area because otherwise no differences would be present in relative income and cost of living across choices. Our survey data indicates that no more than 4% of individuals tried to select the same metro area. For those subjects, the comparison was between two of their top programs but not necessarily the top two.

<sup>8</sup>Even though there are no large income differences in residency salaries, there can be large differences in post-residency salaries, especially across specialties.

We asked two questions about cost of living (one for each metro area) and two questions about the earnings rank (one for each metro area), in that order. For the cost of living question, we provided the following brief introduction: “You probably noticed that the average prices of goods and services are different across different cities. As a result, with the same income, you would be able to buy more things in some cities and less in other cities.” After this introduction, we asked individual how much more or less expensive each metro area was, relative to the U.S. average. To make answering the question easier, we split it in two questions. The first question was: “Imagine that you chose to work in the [Metro Name] metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?” The respondents could choose either “cheaper” or “more expensive.” The second part of the question was: “How much [cheaper/more expensive] is the [Metro Name] metro area than the U.S. average?” Respondents could answer this second question with a drop-down menu ranging from 0% to 50%, in 1 percentage point increments.

We also provided an introduction for the question about earnings rank: “Now we want to ask you about your expected earnings rank. This rank is defined as the share of the working individuals of a city who earn less than you. You probably noticed that the distribution of earnings is different across different cities. As a result, with the same earnings, you may be relatively rich in some cities but relatively poor in other cities.” After this introduction, we asked the following question for each city: “Imagine that you chose to work in [Metro Name]. With your individual annual earnings of \$[Salary], you would be richer than what percentage of [Metro Name]’s individual earners?” Respondents could select their answer from a drop-down menu that ranges from “Richer than 1% of individual earners” to “Richer than 100% of individual earners,” in 1% increments.

We focus on this definition of reference group because it is the most widely used approach in the related literature: e.g., Luttmer (2005) studies how the happiness of an individual is affected by the income of her neighbors.<sup>9</sup> In practice, individuals may care about their ranking in finer reference groups: e.g., they may care disproportionately about their relative standing with respect to neighbors in the same age cohort, rather than caring about all neighbors equally. However, this source of measurement error is not a major source of concern, to the extent that it can only introduce attenuation bias.

---

<sup>9</sup>Moreover, this geographic definition of references group is used more generally in the literature of social interactions more generally: e.g., Perez-Truglia (2017) and Perez-Truglia and Cruces (2017) study how an individual’s political participation is affected by the participation of her neighbors.

### 2.2.3 Information-Provision Experiment

One limitation with using perceptions is the potential for omitted-variable bias. For instance, conditional on income and perceptions about cost of living, perceptions about relative income may happen to be correlated with perceptions about other characteristics of the area, such as the crime rate, amenities, public goods, and so forth. To address this concern, we generate exogenous variation in the perceptions about cost of living and earnings rank by embedding an information-provision experiment in the survey.

Immediately after respondents provided their prior beliefs on both measures, they were shown two messages: one page with statistics about the cost of living in the two cities being considered and a second page with statistics about the earnings rank in each of the two cities. The following message is a sample of the feedback page for cost of living: “Los Angeles-Long Beach-Anaheim, CA metro area is 17.0% more expensive than the U.S. average. The Champaign-Urbana, IL metro area is 6.6% cheaper than the U.S. average.” The following message is a sample of the feedback page for earnings rank: “With your individual annual earnings of \$54,000, you would be richer than 57.9% of Los Angeles-Long Beach-Anaheim, CA’s population. With your individual annual earnings of \$54,000, you would be richer than 60.3% of Champaign-Urbana, IL’s population.” In both of these feedback pages, individuals were asked to take a moment to review the information carefully and were alerted that the information was only going to be shown once. We did not allow respondents to continue to the next page until at least 10 seconds had elapsed.<sup>10</sup>

After individuals finished reviewing the feedback, we re-elicited their perceptions about cost of living and earnings rank, which we denote as the posterior beliefs. Given that our feedback entailed many figures for participants to remember and process, we wanted to make it easier for individuals to compare the options. Therefore, after eliciting respondent’s posterior beliefs, we gave subjects a third page of feedback based on their posterior beliefs. The following is a sample of that feedback page: “We understand this is a lot of information to process, so we will help you make the comparison simpler. According to your final answers about incomes, cost of living and earnings rank: If you chose to live in Los Angeles-Long Beach-Anaheim, CA, you would be able to afford 19.7% less than if you chose to live in Champaign-Urbana, IL. If you chose to live in Los Angeles-Long Beach-Anaheim, CA, your earnings rank would be 3.3% lower than if you chose to live in Champaign-Urbana, IL.”<sup>11</sup>

---

<sup>10</sup>The median time spent on the feedback page was 18.5 seconds.

<sup>11</sup>The difference in cost of living was calculated as  $100 \cdot \left( \frac{w_1}{w_2} \frac{COL_2}{COL_1} - 1 \right)$ , where  $w_i$  is the nominal wage for city  $i$  and  $COL_i$  is their posterior belief about cost of living (from 50 to 150). The difference in earnings rank was calculated as  $100 \cdot \left( \frac{ER_1}{ER_2} - 1 \right)$ , where  $ER_i$  is the posterior belief about earnings rank in city  $i$ . As with the other feedback pages, 10 seconds had to elapse before respondents could move to the next page. The median duration on the post feedback page was 19.5 seconds.

We computed the statistics shown to the subjects using two alternative data sources, and we cross-randomized which of the two sources were shown to each individual. The sources were randomized between individuals; that is, we used the same cost of living source for the two cities being considered by each individual, and the same earnings data source for the two cities. As a result, individuals were randomly assigned to one of four treatment groups. For cost of living estimates, the two sources used were the Regional Price Parity (RPP) data by the Bureau of Economic Analysis and the Cost of Living Index (COLI) data compiled by the Council for Community and Economic Research. For the earnings rank feedback, the two sources used were the American Community Survey (ACS) and the Current Population Survey (CPS), both conducted by the U.S. Census Bureau.<sup>12</sup>

This source randomization created a substantial amount of exogenous variation in signals. For instance, the correlation of the pairwise difference in cost of living shown to the respondents versus the corresponding pairwise difference from the alternative source is 0.656; the corresponding correlation for the earnings rank is 0.649. These differences across sources arise from a combination of several factors, most notably sampling variation and data definitions. For instance, the cost of living data is subject to sampling variation because it tracks the prices of a limited number of goods and services, and earnings rank data is subject to sampling variation because the estimates are based on a limited number of survey respondents. The variation in definitions arise because different cost of living indices give different weights to expenditure categories, and because the earnings rank measures are based on surveys with significant differences in the survey method and the phrasing of the questions used to elicit total annual earnings.<sup>13</sup>

For the sake of transparency and to ensure the validity of the information, the individuals were debriefed in the feedback messages on the name of the source of the information that they received. We would not expect the source name to have an effect in and of itself, given that the individuals did not have expertise on the data. Indeed, we find that the reaction of individuals to the information was orthogonal to the name of the information source.<sup>14</sup>

#### 2.2.4 Rank Submission Choices

The survey asked respondents to indicate which program they expected to rank higher when submitting to the NRMP: “As of this moment: of the two programs discussed so far, which

---

<sup>12</sup>For more details, see Appendix A.1.

<sup>13</sup>Our research design is not limited to having multiple information sources to randomize feedback between them. An alternative would be keeping the source constant and randomly choosing who receives information and who does not. This approach would require minor changes to the econometric framework, see Cullen and Perez-Truglia (2018) for details.

<sup>14</sup>Results reported in Appendix A.6.

one would you expect to rank higher for the NRMP?” Individuals could indicate their ranking on a 6-point scale ranging from “Very likely [Program 1] (in [Metro 1])” on one side to “Very likely [Program 2] (in [Metro 2])” on the other. In the baseline results we look at the binary choice of whether they expect to rank Program 1 over Program 2 because a comparison with the ex post submission choices is more straightforward. Nevertheless, results are similar when using the full likelihood scale.<sup>15</sup>

The algorithm used by NRMP was designed by Roth and Peranson (1999) to be 100% resistant to attempts of “strategic behavior,” meaning that it is a weakly dominant strategy for students to submit their true preferences (i.e., it is optimal regardless of the behavior of the other applicants). Students receive training from the NRMP that makes it explicit that it is in their best interest to submit truthful ranks. Indeed, survey data indicates that only 5% of participants attempt to misreport their true preferences with a strategic motive (Benjamin et al. 2014; Rees-Jones 2017).<sup>16</sup> Furthermore, almost all NRMP participants receive a match,<sup>17</sup> and backing out from a match entails serious sanctions.<sup>18</sup> As a result, the rank choices provide a direct proxy for the individuals’ true preferences.<sup>19</sup>

## 2.3 Follow-Up Survey

Shortly after the NRMP rank submission window closed, we conducted a follow-up survey with the subjects that responded to the baseline survey. Appendix B.2 shows the full questionnaire of the follow-up survey.

Most importantly, at the very beginning of the survey we collected data on the final rank order submitted to the NRMP. Additionally, we took the opportunity to ask individuals for some additional information. We elicited the perceptions about cost of living and earnings rank, which allows us to measure the persistence of the information learned in

---

<sup>15</sup>Results reported in Table A.12.

<sup>16</sup>These results are consistent with other surveys (NRMP 2015). Given the small share of individuals attempting to manipulate rankings, we decided not to include questions about this. Relatedly, Rees-Jones and Skowronek (2017) provide complementary behavioral evidence that NRMP participants may fail to fully optimize, including a discussion of the source of those frictions.

<sup>17</sup>For instance, 95% of the 27,048 U.S. graduating medical students received a successful match in 2017.

<sup>18</sup>For example, applicants with confirmed violations of NRMP policies are subject to a one year bar from accepting or starting a position in any program sponsored by a Match-participating institution, from one year to a lifetime bar from participation in future NRMP Matches, and from one year to a lifetime identification in the matching system as a match violator (Source: <http://www.nrmp.org/policies/the-match-commitment/>). Additionally, the NRMP has established rules prohibiting programs from contacting candidates to ask or coordinate their rank orders.

<sup>19</sup>Since most of the evidence on relative concerns is based on the happiness literature (e.g., Luttmer 2005), we wanted to be able to compare preferences inferred from choice data with respect to the preferences inferred from happiness data, in the spirit of Benjamin et al. (2012; 2014). For this purpose, we included the following question about happiness rank: “If assigned to it, in which of the two programs would you expect to live a happier life?” Responses used the same likelihood scale as for rank.

the information-provision experiment. Also, we measured additional characteristics of the subjects that we were not able to measure in the baseline survey due to space and time constraints, such as the places where the individuals grew up.

### 3 Econometric Model

#### 3.1 Baseline Model

In this baseline model, we exploit all the variation in perceptions of earnings rank and cost of living, which includes the experimental variation induced by our information-provision as well as the remaining non-experimental variation.

Let  $i$  index subjects and  $j \in \{1, 2\}$  denote the two programs being considered by the subject. We define  $ER_j^{i,posterior}$  and  $COL_j^{i,posterior}$  as the posterior beliefs for earnings rank and cost of living for program  $j$  in the baseline survey. Let  $ER_{1,2}^{i,posterior} = ER_1^{i,posterior} - ER_2^{i,posterior}$  be the perceived difference in earnings rank between the two programs. Similarly, let  $COL_{1,2}^{i,posterior} = COL_1^{i,posterior} - COL_2^{i,posterior}$  be the perceived difference in cost of living between the two programs. Let  $Program_1 \succ_i Program_2$  denote that individual ranks program 1 over program 2, and let  $I(\cdot)$  be an indicator function. The regression specification is:

$$I(Program_1 \succ_i Program_2) = I\left(\beta^{ER} \cdot ER_{1,2}^{i,posterior} + \beta^{COL} \cdot COL_{1,2}^{i,posterior} + \theta X^i + \varepsilon_i \geq 0\right), \quad (1)$$

where  $X^i$  is a vector of control variables and  $\theta$  is the corresponding vector of coefficients. We always include a constant and the log-difference of nominal residency wages as control variables. In the baseline specification, we include an additional set of controls consisting of pairwise differences in some residency and location characteristics: residency program rank (from Doximity), quality of life inferred from compensating differentials (Albouy 2016), population size, population density, share of African-American residents, share of Democrat residents, and share of urban population.<sup>20</sup> In any case, we present results with alternative sets of control variables.

In the baseline specification, we estimate a Probit model, which implies that the error term ( $\varepsilon_i$ ) is normally distributed. As is typical in discrete-choice models, using a Probit model is convenient in the sense that the ratio between parameters can be readily interpreted as

---

<sup>20</sup>The source for the demographic characteristics is the 2011-2014 American Community Survey. For the share of Democrat residents, we use the share of Obama voters between all voters in the 2008 Presidential Elections.

marginal rates of substitution. However, this specification choice is irrelevant in practice: the results are virtually identical if we use alternatives such as Logit or Linear Probability models.

The two key parameters of interest are  $\beta^{ER}$  and  $\beta^{COL}$ . The parameter  $\beta^{COL}$  measures preferences for purchasing power during the duration of the residency. The hypothesis is that  $\beta^{COL} < 0$ : i.e., individuals prefer to live in places with lower cost of living. The parameter  $\beta^{ER}$  measures preferences for relative income during the residency. Depending on the mechanism at play, we may expect  $\beta^{ER}$  to be positive or negative. On the one hand, if individuals see their neighbor's income as a negative externality, as in the models of social status, then we would expect  $\beta^{ER} > 0$  (i.e., individuals want to choose less affluent ponds). On the other hand, if individuals see the neighbor's income as a positive externality, then we would expect  $\beta^{ER} < 0$  (i.e., individuals want to choose more affluent ponds).

The duration of a residency depends on the specialty: it lasts for a minimum of three years, it typically takes five years, and in some cases it may require a minimum of seven years.<sup>21</sup> Note that earnings rank and cost of living *after* the end of the residency would be part of the error term.

### 3.2 Instrumental Variables Model

The second model exploits the variation in beliefs induced by the source-randomization experiment to estimate the causal effects of perceptions on choice. Let  $ER_{1,2}^{i,shown}$  be the information randomly chosen to be shown to the individual, and  $ER_{1,2}^{i,alt}$  be the alternative information that could have been shown to the individual, but was not shown. Let  $\Delta ER_{1,2}^i = ER_{1,2}^{i,shown} - ER_{1,2}^{i,alt}$  be the difference between the information shown and the alternative information that could have been shown. We estimate an IV-Probit model that uses  $\Delta ER_{1,2}^i$  and  $\Delta COL_{1,2}^i$  as instrumental variables. In other words, this model uses the variation introduced by the random assignment of sources to estimate the effect of perceptions on choice:

$$I(\text{Program}_1 \succ_i \text{Program}_2) = I(\beta^{ER} \cdot ER_{1,2}^{i,prior} + \beta^{COL} \cdot COL_{1,2}^{i,prior} + \lambda_1 \cdot ER_{1,2}^{i,alt} + \lambda_2 \cdot COL_{1,2}^{i,alt} + \theta X^i + \varepsilon_i \geq 0)$$

$$ER_{1,2}^{i,prior} = \gamma_1^{ER} \cdot \Delta ER_{1,2}^i + \gamma_2^{ER} \cdot \Delta COL_{1,2}^i + \gamma_3^{ER} \cdot ER_{1,2}^{i,alt} + \gamma_4^{ER} \cdot COL_{2,1}^{i,alt} + \gamma_5^{ER} X^i + \epsilon_{1,i}$$

$$COL_{1,2}^{i,prior} = \gamma_1^{COL} \cdot \Delta ER_{1,2}^i + \gamma_2^{COL} \cdot \Delta COL_{1,2}^i + \gamma_3^{COL} \cdot ER_{1,2}^{i,alt} + \gamma_4^{COL} \cdot COL_{2,1}^{i,alt} + \gamma_5^{COL} X^i + \epsilon_{2,i}$$

---

<sup>21</sup>A small minority of subjects may expect to continue living in the same city after the residency, in which case the cost of living and the distribution of earnings may also be relevant for the post-residency period.

There is a simple way to understand the intuition behind this instrumental variables approach. In a deceptive design, subjects would be shown the statistic from a certain source, but with random noise added to this statistic. Then we would only exploit the variation in beliefs generated by the random noise. In our context,  $\Delta ER_{1,2}^i$  and  $\Delta COL_{1,2}^i$  play the role of the random noise added to the feedback, only that they are generated in a non-deceptive manner.

## 4 Implementation Details and Summary Statistics

Our recruitment strategy is similar to that of Benjamin et al. (2014). During December 2016 we contacted the Associate Dean of Student Affairs at all 135 accredited medical schools in the United States by email to ask for permission to invite fourth year students participating in the 2017 Main Residency Match to take part in our study (a sample of the invitation email is shown in Appendix B.3). Our goal was to recruit as many respondents as possible, so we followed up, by email and phone, with all the deans who showed interest. Of the 79 schools that answered our invitation, 27 agreed to participate. The main reason given by the schools that declined to participate was school policy restricting external surveys, in place to avoid survey fatigue. Our sample of participating schools includes 22 of the 50 U.S. states, and it is quite representative of the whole sample of 135 accredited medical schools – we do not find statistically significant differences in observable characteristics such as total enrollment, average MCAT scores, undergraduate GPA at admission, acceptance rate, and U.S. News rank.<sup>22</sup>

For confidentiality reasons, we were not given email lists to directly invite students to participate in our study. Instead, the deans agreed to forward our invitation email containing the link to the survey to eligible students (i.e., senior medical students participating in the NRMP). This email invitation, a sample of which is shown in Appendix B.4, asked students to participate in a confidential survey about the Main Residency Match for a study on how medical students select residency programs. The message mentioned that the survey would take less than 10 minutes to complete and respondents would be sent a \$10 Amazon gift card by email as a token of appreciation. Finally, the email stressed the eligibility criteria for participating in the survey: being a graduating medical student participating in the Main Residency Match who has not yet submitted his or her rank to the NRMP.<sup>23</sup>

The only reason why we excluded individuals who had previously submitted their ranks

---

<sup>22</sup>For details, see Appendix A.2.

<sup>23</sup>There are a number of alternative matches for some specialties that have different deadlines than the Main Residency Match.

was because we wanted individuals who were still deciding and thus prone to using the signals from the information-provision experiment. However, this concern is not important in the sense that submissions can be modified anytime before February 22. Even if some students had already submitted their rank at the time of responding to the survey, they would still be able to modify their rank. In any case, the vast majority of our subjects responded to the baseline survey quite early in the submission period.

We took several measures to minimize the chance that non eligible students would participate in the survey. First, deans were asked to carefully forward the invitation to senior students participating in the Main Residency Match. This request was not an issue since such a mailing list already existed; targeted announcements were already being sent to this group during the semester regarding the Match. Second, individuals were reminded of these restrictions in the invitation email and on the consent page of the survey. Third, the first questions of the surveys acted as filters; we asked what match the respondent was participating in and whether they had already submitted their ranks. If they responded with a match other than the Main Residency or “yes” to already submitting their rank, the survey ended there, and they were excluded from taking the survey again.<sup>24</sup>

Last, at the end of the survey, respondents were required to submit their university email address to “sign” a statement claiming that they were medical students participating in the NRMP and they understood that we reserved the right to verify their status before making a payment. We were able to confirm the validity of 100% of respondents for a subset of schools. Given all the measures taken and the evidence obtained, we are confident that the survey data are of high quality.

The invitation emails were sent to students in a staggered way, with the first round of invitations sent on January 6, 2017, and the last round of invitations and reminders sent on February 7, 2017. We estimated that the student invitations were forwarded to around 3,676 students in total, with 1,080 finishing the baseline survey, implying an overall response rate of 29.38%. The median survey completion time was almost 9 minutes. At the end of the baseline survey we included an attention check question that was passed by 96.4% of respondents. For the sake of transparency, we do not drop the group that did not pass the attention check – indeed, we do not drop any other group from the baseline sample.<sup>25</sup>

On February 23, 2017, the day after the NRMP rank submission deadline, we sent respondents who participated in our baseline survey an invitation to participate in the follow-up

---

<sup>24</sup>The survey platform blocks users from taking the survey again by using their I.P. address and cookies, although students could circumvent this restriction by opening the survey link from a different device.

<sup>25</sup>Upon inspection of the data, the 3.6% of respondents who failed the attention check seem to have answered the survey as consistently as everyone else. And, as reported in Appendix A.8, the results are virtually the same if we drop this 3.6% of the sample.

survey. We offered participants an additional \$5 Amazon gift card for participating in this shorter follow-up survey. We closed the follow-up survey on March 12, one day before Match Week started (i.e., the time when the students find out where they are matched). The response rate to the follow-up survey was 90.62%. Moreover, the characteristics of the individuals who responded to the follow-up survey are similar to the characteristics of individuals who did not respond to the follow-up.<sup>26</sup>

Figure 1 presents the distribution of dates when subjects responded to the baseline survey, when they responded to the follow-up survey, and when they submitted their ranks to the NRMP (for those who provided this information in the follow-up survey). On average, students responded to the baseline survey 24.5 days (s.d. 12.9) before submitting their ranks, and responded to the follow-up survey 13.9 days (s.d. 11.8) after submitting their ranks.

Table 1 provides descriptive statistics for the key variables used in the analysis. Column (1) corresponds to all respondents to the baseline survey. Around 48% of respondents were male, the average age was 27 years, 35.4% of respondents were single, 23.9% were married, and 40.7% were in a long-term relationship. On average, students were offered a salary of \$54,000 for the first year of their residency – this salary would make them richer than 56% of earners in the average metro area. Of course, this sample is not representative of the general U.S. population of adults: most notably, our subject pool is younger and more educated. Nevertheless, our subject pool is close to the U.S. average in terms of nominal wages and gender composition.<sup>27</sup>

To verify that the randomization was successful, Table 1 breaks down the descriptive statistics by each treatment group. This table also reports the p-value for the test of the alternative hypothesis that at least one mean is different across the four treatment groups. First, this table shows that the number of respondents was almost identical number of respondents across all groups. Second, this table shows that the differences in individual characteristics are economically small and statistically insignificant across the treatment groups, thus confirming that the random assignment was successful.

Last, Figure 2 shows the geographic distribution of the metropolitan areas in which students' top-two programs are located. This figure shows that there is a broad geographical coverage of the U.S. territory. This broad coverage results in significant trade-offs between cost of living and relative income. To demonstrate this, Figure 3 shows a scatterplot of the pairwise differences in cost of living (according to the RPP measure) vs. the differences in earnings rank (according to the ACS measure).<sup>28</sup> The substantial dispersion along the

---

<sup>26</sup>Results presented in Appendix A.2.

<sup>27</sup>For more details, see Appendix A.3.

<sup>28</sup>Using the alternative data sources yield similar results.

y-axis suggests that there are large differences in cost of living across the pairs of cities that the individuals must choose from. The substantial dispersion along the x-axis suggests that there are large differences in earnings rank across the pairs of cities.<sup>29</sup> Most important, the  $R^2 = 0.22$  indicates that, even though the two are correlated,<sup>30</sup> substantial orthogonal variation exists between cost of living and relative income.

## 5 Results: Prior Beliefs and Learning

### 5.1 Distribution of Prior Beliefs

To the best of our knowledge, ours is the first paper to measure perceptions about cost of living and earnings ranks across different cities. To get a sense of how informed individuals are about these aspects of their decision-making, we start by comparing their prior beliefs (that is, perceptions prior to the feedback) to the statistics from the baseline sources.

Respondents seem to have a relatively good idea of the cost of living in the cities they are considering. Figure 4.a shows respondents' prior beliefs about cost of living along with the corresponding RPP estimates. The RPP is meant to reflect all sources of expenditures, and for that they employ data on prices in apparel, education, food, housing, medical, recreation, rents, transportation and other goods and services. If prior beliefs were completely accurate, we would expect to see all responses on the 45 degree line. On average, prior beliefs overestimate the baseline estimate by just 4 percentage points; and the prior belief and RPP estimates are highly correlated, with an  $R^2$  of 0.550.

However, individuals are substantially less well informed about their earnings ranks. Figure 4.b plots prior beliefs about earnings rank against the ACS estimates. On average, individuals underestimate earnings ranks by almost 16 percentage points; and the prior belief and ACS estimates are positively correlated, but with an  $R^2$  of just 0.029. Because we are ultimately interested in relative differences for their decision making, we repeat this exercise using pairwise differences instead of levels. It seems that respondents have a better understanding of relative differences in earnings rank, though they still remain far less accurate than perceptions over cost of living.<sup>31</sup> This finding suggests that, while prior evidence suggests that individuals have significant biases when assessing their position in the national income distribution (Cruces et al. 2013; Karadja et al. 2017), these biases are even more sub-

---

<sup>29</sup>Furthermore, the vast majority of these differences in cost of living and earnings rank are orthogonal to differences in nominal income – see Appendix A.4 for details.

<sup>30</sup>The slope of  $-0.664$  suggests that, on average, relatively more expensive cities tend to have a higher distribution of nominal earnings.

<sup>31</sup>Detailed results reported in Appendix A.5.

stantial when individuals try to predict their position in places where they are not currently living.

## 5.2 Learning from Feedback

We next examine whether respondents learned from the information we provided. If respondents learn from the information provided, we would expect a positive relation between their perception gaps (i.e., the signal received minus the prior belief) and their revisions (i.e., the posterior belief minus the prior belief); that is, respondents who originally overestimated would revise their beliefs downwards, while those who underestimated would revise in the opposite direction. Moreover, the slope of this relationship can be used to quantify the degree of learning from the feedback: a parameter that takes the value from 0 to 1, indicating the weight assigned to the signal relative to the weight assigned to the prior belief.<sup>32</sup>

Respondents strongly updated their beliefs after being provided with feedback. Figure 5 presents the reduced-form effects of information for cost of living and earnings rank, respectively. Figures 5.a and 5.d present the short-term effect, that is, the revision made by respondents directly after being given the information. The short-term learning rates, given by the slopes reported in these figures, are 0.879 (s.e. 0.010) for the cost of living and 0.873 (s.e. 0.011) for the earnings rank. These two learning rates are statistically significant, precisely estimated, and we cannot reject the null hypothesis that they are equal to each other (p-value=0.754). These learning rates are remarkably close to 1, meaning that respondents almost fully reacted to the signals.

One limitation with this evidence is that individuals may have revised their beliefs towards the truth regardless of the feedback we provided. For instance, they may have taken extra time to think about the question, leading to a more accurate response. The source experiment was designed to test this specific hypothesis. We construct two variables: the information actually shown and the “alternative” information that could have been shown. If the alternative information had any effect beyond the information shown, that would be evidence that part of the revisions were due to reversion to the truth rather than reversion to the information provided. Figures 5.b and 5.e show the relation between the alternative information and the revision adjusted for the information actually shown. The alternative information indeed has no effect: the coefficients are close to zero ( $-0.034$  for cost of living and  $0.060$  for earnings ranking) and precisely estimated. Furthermore, in Section 7.3 we show that there was no cross-learning (i.e., feedback on cost of living did not affect beliefs about earnings ranking and vice versa).

---

<sup>32</sup>See Appendix A.6 for more details.

In survey experiments, one main concern is that instead of inducing genuine learning, the information provided in the experiment may elicit spurious reactions. For instance, if an individual is told that the cost of living in a city is “10% more expensive than the U.S. average” and then later asked about the cost of living in the same city, he or she may report a cost of living that is closer to “10% more expensive than the U.S. average” for spurious reasons, such as unconscious numerical anchoring (Kahneman and Tversky 1972). Under the assumption that these effects are temporary, we can disentangle genuine from spurious learning by looking at the reaction to the information provided in the experiment that persisted over time (Cavallo et al. 2017).

We look at the persistence of the effect of feedback between the time participants responded to the baseline and follow-up surveys, which was 38.4 days on average. Figures 5.c and 5.f show the relation between the initial perception gap and the long-term revision based on beliefs reported in the follow-up survey (i.e.,  $b_k^{posterior,LT} - b_k^{prior}$ ). There is substantial persistence the effects of the feedback: the estimated slope for the initial perception gap and the long-term revision (i.e., the difference between long-term belief and the initial prior belief) for cost of living is 0.752 (s.e. 0.016), while for earnings rank it is 0.626 (s.e. 0.020). The persistent effect of our feedback suggests that individuals indeed care about these city features. These longer-term revisions are slightly weaker than the short-term revisions, but that result is expected given that individuals must have gathered some additional information in the time between the two surveys.

## 6 Results: Location Preferences

### 6.1 Average Preferences

We start with the baseline specification, which uses the Probit model from Section 3 with the expected rank submission as dependent variable. This specification exploits all the variation in perceptions, which includes the experimental variation induced by our information-provision as well as the remaining non-experimental variation. We introduce variations of this specification later.

The Probit coefficients are presented in Table 2. Column (1) presents the results for the full sample, while columns (2) through (7) present results by demographic subgroups. The estimated  $\beta^{COL}$  from column (1) is negative and statistically significant (p-value=0.027), suggesting that the average individual prefers programs with lower costs of living. To better understand the magnitude of these Probit coefficients, we can transform them into the corresponding marginal effects, where increasing the cost of living by 1 percentage point at a

program’s location decreases the probability of choosing that program by 0.201 percentage points (which can be interpreted as a behavioral elasticity of -0.201).<sup>33</sup>

The average subject also prefers a higher earnings rank: the estimated  $\beta^{ER}$  from column (1) is positive and statistically significant (p-value=0.065). This coefficient suggests that the average individual prefers to live in a city where, holding her cost of living constant, she earns more than her neighbors. The corresponding marginal effect indicates that increasing the earnings rank at a program’s location by 1 percentage point increases the probability of choosing that program by 0.186 percentage points (for a behavioral elasticity of 0.186). The elasticity for cost of living (-0.201) is similar in absolute value to the elasticity for earnings rank (0.186) – indeed, their difference is statistically insignificant. This finding suggests that individuals care about relative income nearly about as much as they care about cost of living: the average individual would be willing to accept an increase of 0.925 percentage points in the cost of living in exchange for a 1 percentage point increase in earnings rank.

The fact that medical students care about cost of living and relative income is consistent with the view that money is a primary motivation for doctors. For instance, according to a 2008 survey, 49% of pre-med students self-reported being primarily motivated by money in their career choice (Daniel and O’Brien 2008).<sup>34</sup> Our findings are also consistent with prior survey evidence indicating that NRMP applicants care about the geographic location of the programs. For example, Sledge et al. (1989) surveyed a sample of graduating medical students to elicit the most important features in determining the program they ranked first in their NRMP submission. They allowed subjects to choose from 15 features that included characteristics of the program (e.g., program curriculum) and other characteristics (e.g., geographic location). They found that “geographic location” was rated as the most important feature (selected 63.8% of the time) as well as the most desirable feature (selected 37.6% of the time). Using more recent data, NRMP (2015) also find that geographic location is the most popular feature; additionally, they find that, when asked to mention the most important factors taken into account in their decision-making, 52% of surveyed medical students explicitly mention the cost of living.

Even though  $\beta^{COL}$  and  $\beta^{ER}$  are statistically and economically significant, they do not imply that cost of living and relative income during the residency are the main features that medical students pay attention to. Indeed, if doctors are motivated by financial rewards, we would expect them to seek a higher post-residency income through more prestigious programs. Indeed, in complementary analysis, we find that, when deciding their NRMP submissions, doctors care more about the prestige of the program and the career prospects than about

---

<sup>33</sup>These marginal effects are reported in Appendix A.7.

<sup>34</sup>These survey results are based on responses from 461 takers of the Kaplan MCAT test in February 2008 and 453 takers of the Kaplan LSAT test in February 2008.

the cost of living and relative income during their residency.<sup>35</sup> This finding is also consistent with survey data indicating that the program’s prestige is of higher importance than cost of living (NRMP 2015).

Under the assumption that our revealed-preference estimates reflect the same type of neighbor externalities from Luttmer (2005), we can compare the magnitude of relative concerns in this paper to the magnitude reported in Luttmer (2005). This comparison, presented in Appendix A.11, suggests that our estimates imply a somewhat smaller role for relative concerns.<sup>36</sup>

## 6.2 Heterogeneity by Relationship Status

The average preferences could potentially mask substantial heterogeneity. For instance, Luttmer (2005) finds that the effect of relative income on happiness is driven entirely by non-single individuals. Also, evidence from the urban economics literature indicates that single and non-single individuals have different location preferences (e.g., Couture and Handbury 2016; Gautier et al. 2010). With the goal of exploring this form of heterogeneity, we elicited the relationship status using the same categories as in Luttmer (2005).

Columns (2) and (3) of Table 2 present the results on this form of heterogeneity, by estimating the model separately for singles and non-singles. Column (2) of Table 2 shows the estimates for non-single individuals (i.e., the 65% of the sample who are married or in a long-term relationship) and column (3) for the sample of single individuals (35% of the sample).<sup>37</sup> It is important to note that by non-single we only refer to their relationship status, not to whether the respondent participates as a dual match, which is a special regime used by roughly 7% of subjects—indeed, the results are similar if we drop the small minority of subjects participating in dual matches.<sup>38</sup>

The coefficients from columns (2) and (3) suggest a large heterogeneity in  $\beta^{ER}$  by relationship status. For non-single individuals, the estimated  $\beta^{ER}$  (2.236) is positive and statistically significant at the 1% level. For the sample of single individuals,  $\beta^{ER}$  (−1.538) is negative and statistically significant at the 10% level. The direction of the difference in relative concerns between non-singles and singles is consistent with the evidence from Luttmer (2005). The difference in  $\beta^{ER}$  between non-singles and singles is also highly statistically significant

---

<sup>35</sup>Results presented in Appendix A.9.

<sup>36</sup>For a more direct comparison between happiness and choice data, we can also exploit the survey responses on expected happiness. We find that the marginal rates of substitution inferred by happiness are statistically indistinguishable from the marginal rates of substitution inferred by choice; however, due the lack of precision of the happiness estimates, we cannot rule out large discrepancies. Results presented in Appendix A.12.

<sup>37</sup>Appendix A.10 shows results breaking down the non-single individuals into married and in a long-term relationship. The relative concerns are similar between these two groups.

<sup>38</sup>See Appendix A.8 for more details.

(p-value=0.001). And contrary to the case of preferences for earnings rank, the relationship status does not seem to affect the preferences for cost of living. According to columns (2) and (3) of Table 2, the estimated  $\beta^{COL}$  is -1.087 for non-singles and -1.058 for singles, with the difference being statistically insignificant (p-value=0.977).

These estimates suggest that while non-single individuals prefer to live in less affluent ponds, single individuals would rather live in more affluent ponds. In other words, single individuals tend to see more affluent neighbors as a positive rather than a negative externality. One potential explanation is that they expect to benefit from some public goods that are provided by the more affluent neighbors. Another potential explanation lies in the dating market: since more affluent individuals prefer to date more affluent partners (Fisman et al. 2006, Hitsch et al. 2010), this can naturally create a preference for locating in more affluent ponds.<sup>39</sup> Indeed, this result is consistent with Gautier et al. (2010), who show that singles are willing to pay higher housing prices to benefit from a denser dating market in cities. Consistent with our findings for non-singles, Gautier et al. (2010) find that after getting married, the dating-market benefits no longer matter for them, and couples move out of the city.

It is important to note that the singles in our subject pool are a special group of individuals, and as a result the dating motive in this sample may be stronger than for the general population of singles. First, these subjects are at their prime dating age, and thus may expect to find a long-term partner during their residency. Second, even though their wages during the residency put them near the middle of the national earnings distribution, their expected *post-residency* earnings will place them at the very top of the national earnings distribution. As a result, these subject would probably struggle to meet a partner that can match their permanent income unless they move into one of the most affluent cities.

Columns (4) through (7) of Table 2 explore other forms of heterogeneity based on characteristics measured in the baseline survey. Columns (4) and (5) explore potential differences in preferences by gender:  $\beta^{ER}$  is similar for females (1.041) and males (0.896), and  $\beta^{COL}$  is also similar for females (-0.972) and males (-1.443). Moreover, neither of these two differences are statistically significant (p-values of 0.894 and 0.642, respectively). Columns (6) and (7) explores heterogeneity by future income. Even though all these subjects receive a similar income during the residency, they have different expected incomes after they finish their residencies. Then, it is possible that individuals who selected high-earning specialties may be more concerned about cost of living or relative income. To test this hypothesis, columns (6)

---

<sup>39</sup>Prior evidence suggests that, relative to single men, single women may have a stronger preference for finding more affluent partners (Bertrand et al. 2015; Bursztyn et al. 2017). Consistent with this view, we find that the preference for more affluent ponds among singles is driven primarily by single women, although this result is imprecisely estimated. Results reported in Appendix A.8.

and (7) split the sample in specialties with above and below median post-residency average income. The  $\beta^{ER}$  is 1.433 for below-median specialties and 0.777 for above-median specialties, and  $\beta^{COL}$  is -0.690 for below-median specialties and -1.238 for above-median specialties, with neither of those differences being statistically significant (p-values of 0.544 and 0.580, respectively).

To address the possibility of false positives due to multiple hypothesis testing, for each p-value reported in Table 2, we report the corresponding q-value in brackets (Benjamini and Yekutieli 2001). The q-value indicates the minimum false discovery rate (i.e., the expected proportion of rejected null hypotheses that are actually true) at which the null hypothesis would be rejected for that test given all tests reported in the same table. The adjustment for multiple hypothesis testing does not change any of the results: most importantly, the difference in  $\beta^{ER}$  between singles and non-singles has a q-value of 0.030, which indicates that this heterogeneity is unlikely to be spurious.

We also computed heterogeneity by other characteristics measured in the follow-up survey. None of these dimensions are nearly as important as relationship status for predicting heterogeneity in preferences for relative income, both in terms of magnitude and statistical significance.<sup>40</sup> Because of the magnitude of the heterogeneity by relationship status, in the remainder of the paper we report estimates for the entire sample as well as broken down by relationship status.

### 6.3 Robustness Checks: Additional Controls

One potential concern with the baseline specification is that of omitted-variable biases. For instance, if places where an individual expects higher earnings rank (i.e., less affluent metro areas) are systematically worse in terms of quality of life, then failing to account for quality of life would introduce a negative bias in  $\beta^{ER}$ . This would make relative concerns look weaker than they actually are.

As a first approximation to addressing these omitted variable biases, we assess the robustness of our estimates to the inclusion of different sets of control variables. We present the baseline results with alternative sets of control variables in Table 3. Each row corresponds to a different regression, with a different set of control variables. The first row presents results for our baseline specification, but without including any control variables for the characteristics of the program or the metro area. The second row corresponds to the baseline specification from Table 2, which includes the six baseline controls listed in Section 3.1. The results in the first two rows of Table 3 indicate that  $\beta^{ER}$  and  $\beta^{COL}$  are qualitatively and quantitatively

---

<sup>40</sup>Results reported in Appendix A.16.

similar between the baseline specification and the specification without controls.

The third through last rows of Table 3 include different sets of additional controls. These sets of controls were selected based on attributes that could potentially be relevant for the options of the subjects and at the same time may be correlated to the earnings rank. For instance, we may want to control for crime rates: living in a less affluent city may be desirable for medical students interested in certain specialties where they must learn to treat injuries that are more common in high-crime areas, such as gunshot wounds. Also, we may want to account for place of origin: medical students, who tend to grow up in affluent areas, may want to remain in the the same areas where they grew up (Agarwal 2015).

We examine the following groups of attributes: demographic characteristics (population, population density, percentage urban population, percentage same gender, percentage age 25 to 34, share of college graduates, share foreign, share Hispanic, and share black); amenities (quality of life from Albouy 2016, per capita spending on local public goods, per capita spending on education and health, overall crime rate and violent crime rate, share of registered Democrat voters in the 2012 election); geography (distance of program to city where they grew up, distance of program to current medical school); economic factors (estimated income taxes, federal and state income taxes, local sales tax, rent prices, and the Gini coefficient); a set of state dummies; objective program characteristics (residency program rank from Dximity, dummies for university hospitals and for community hospitals), and subjective program characteristics (the subjective rank in prestige, purpose, and prospect, as reported in the follow-up survey).

Comparing the results across rows of  $\beta^{ER}$  and  $\beta^{COL}$  of Table 3 suggests that these estimates are robust to the choice of control variables, both in terms of statistical significance and economic significance. Of course, small differences occur in the point estimates across specifications. For instance, relative to the baseline  $\beta^{ER}$  of 0.995 for the entire sample in column (3),  $\beta^{ER}$  ranges from a minimum of 0.703 with all controls to a maximum of 1.199 with subjective program characteristics.<sup>41</sup> However, all of these differences are statistically insignificant.

## 6.4 Robustness Checks: Experimental Estimates and Persistence

In this section, we present results from two robustness checks. The first check addresses concerns about omitted-variable bias by exploiting the exogenous variation in beliefs generated

---

<sup>41</sup>According to the pseudo- $R^2$ , including these variables increases the explanatory power of our model to some degree. For the full sample, the pseudo- $R^2$  increases from 0.015 in the specification with no additional controls to a minimum of 0.018 with controls for objective program characteristics or amenities, and a maximum of 0.123 with controls for subjective programs characteristics (or 0.218 with all controls).

by the source-randomization experiment. The second check is intended to address potential concerns about spurious effects of the information-provision experiment, by comparing the short-term effects to the long-term effects of the information-provision. To make these estimates directly comparable to the long-term effects, in this section we restrict the sample to the 90.6% of subjects who responded to the follow-up survey.

Panel A of Table 4 presents the results for  $\beta^{ER}$ . The first row presents the baseline specification, while the second row presents the experimental estimates. The experimental estimates are less precisely estimated than the baseline estimates because they only use a portion of the variation in beliefs. For each of the subgroups of single and non-single respondents, shown in columns (1) and (2), the estimated  $\beta^{ER}$  is qualitatively consistent across the baseline and experimental specifications. For non-singles, the coefficient is 2.380 (p-value=0.001) in the baseline specification vs. 2.977 (p-value=0.025) in the experimental specification. And for singles, the coefficients are -1.656 (p-value=0.095) in the baseline specification vs. -4.964 (p-value=0.012) in the experimental specification.

Column (3) shows that, for the entire sample,  $\beta^{ER}$  is slightly lower in the second row (0.867) than in the first row (1.141) and, due to the lower precision, becomes statistically insignificant in the second row. However, we must take this finding with a grain of salt. First, due to the precision of the experimental coefficient, this difference between the first and second rows is statistically insignificant. Second, the reduction in the average  $\beta^{ER}$  is driven primarily by the fact that the coefficient becomes more negative (and thus more consequential) among singles.

Panel B of Table 4 presents the results for  $\beta^{COL}$ . The results from the baseline specification (first row) are qualitatively different from the results in the experimental specification (second row). All the coefficients (for the entire sample, singles and non-singles) become positive, are imprecisely estimated, and are statistically insignificant. One potential interpretation could be that individuals anticipate that unobservable city amenities translate into higher cost of living, thus reducing the desire to move to areas with lower cost of living (Roback 1982; Albouy 2008). However, this may simply be a spurious finding: since the experimental estimates are not precisely estimated, we cannot rule out large negative values for  $\beta^{COL}$ , and in most cases we cannot reject that the experimental coefficients are equal to those from the baseline specification.<sup>42</sup> Indeed, consistent with the interpretation of spurious finding, we find a negative and significant  $\beta^{COL}$  when we replicate the experiment with the

---

<sup>42</sup>Also, the coefficients from the first and second row are not be expected to be equal, to the extent that the experimental coefficients identify local average preferences instead of average preferences. For instance, it is plausible that the information-provision experiment disproportionately affected individuals who were the most unsure about their priors beliefs about cost of living, who likely are those who care the least about cost of living.

secondary sample (results reported in section 7 below).

As discussed above, the treatment groups were balanced in observable characteristics, suggesting that the randomization was indeed successful. As an additional robustness check, we re-estimate the instrumental variables model but, instead of the rank order, we use the list order as dependent variable (i.e., the order in which the individual listed the residency programs at the beginning of the survey). Because it takes place before the provision of feedback, the feedback should not have any effect on the list order. We present results for this falsification in the fourth row of Table 4. As expected, the estimated values of  $\beta^{ER}$  and  $\beta^{COL}$  are close to zero and statistically insignificant, in the full sample as well as in the sub-samples of non-singles and singles.

The Appendix presents some additional results. In all the instrumental variable specifications, we strongly reject the null hypothesis of weak instruments. Also, the learning rates implied by the first-stage coefficients are always close to 1, and for that reason the instrumental variables estimates are similar to the reduced form estimates.<sup>43</sup>

The second robustness test is intended to address potential concerns about spurious effects of the information-provision experiment. By asking individuals questions about the cost of living and earnings rank, the baseline survey makes those aspects more salient, which may make individuals overweight them in their expected choice. However, it must be noted that this salience effect may not necessarily exaggerate the importance of relative income, because it would be expected to inflate both  $\beta^{ER}$  and  $\beta^{COL}$ . Another potential concern is that of experimenter-demand effects: by providing information about cost of living and earnings rank, the experimenter may be putting pressure on the subjects to use this information in their expected choice. Again, this source of bias would not necessarily exaggerate the importance of relative income: since most individuals do not want to reveal to others that they care about status (Shigeoka and Yamada 2016), the experimenter-demand bias would probably shrink  $\beta^{ER}$  towards zero.<sup>44</sup>

To address these remaining concerns, we estimate the effects of the information provision on the final rank submission, which takes place an average of 38.4 days after the information provision. This can be achieved by using the same instrumental variable model, but using the final submission rank (elicited in the follow-up survey) instead of the expected submission rank (elicited in the baseline survey) as the dependent variable. If the effects were purely due to salience or experimenter-demand effect, we would expect that the information provided in

---

<sup>43</sup>Reduced-form and first-stage estimates are presented in Appendix A.16.

<sup>44</sup>Also, our survey was conducted confidentially and online, which reduces the scope for experimenter-demand effects (Van Gelder et al. 2010). Additionally, it would be difficult to reconcile the experimenter-demand channel with the finding that the earnings rank had a positive effect on non-singles and a negative effect on singles.

the experiment would not have any effect on the final submission choice.

The third row of Table 4 presents the experimental estimates based on the long-term effects of the experiment. By comparing the coefficients in the third row to those from the second row, we can compare the short-term and long-term effects of the information. The long-term experimental coefficients are somewhat different from the short-term experimental coefficients, but those differences are mostly statistically insignificant. Most important, the coefficient on  $\beta^{ER}$  is still positive (1.993) and statistically significant for non-singles, and negative (-5.285) and statistically significant for singles.

## 7 Auxiliary Experiment

### 7.1 Survey Design

In this section, we present complementary evidence from an auxiliary experiment. We designed a variation of the survey instrument, attached in Appendix B.7, that can be used in other contexts besides the medical residency match. At the beginning of this survey, we ask respondents to list two cities that they know well to which they would consider moving to. The following questions are identical to the baseline survey instrument from our main experiment with medical students: we elicit prior and posterior beliefs about cost of living and earnings rank, we conduct the information-provision experiment, and we elicit preferences for the two cities under consideration.

We conducted this auxiliary experiment using a sample of 1,245 U.S. respondents recruited through Amazon Mechanical Turk.<sup>45</sup> Compared to the main residency match experiment, the primary difference is that the subjects in the auxiliary experiment are not moving anytime soon, so we cannot followup with them to measure the effects of the information provision experiment on their actual location choices. Instead, we measure the effects on their expected location choices. This is a limitation of the auxiliary experiment.<sup>46</sup> On the other hand, this auxiliary sample has some advantages. Due to a wide availability of subjects, it is possible to run additional experiments on demand to address alternative mechanisms and to disentangle the mechanisms at play. Also, this secondary sample is more diverse than the main sample of medical students in many observable characteristics.<sup>47</sup> As a result, this secondary sample

---

<sup>45</sup>Details about the recruitment are presented in Appendix A.17.

<sup>46</sup>The results from the previous section suggests that using hypothetical choices may not be as problematic as generally thought, to the extent that preferences inferred from expected choices are consistent with preferences inferred from actual choices.

<sup>47</sup>For example, on average, participants in the auxiliary sample are older and less educated. Descriptive statistics comparing sample characteristics between the main and auxiliary experiment are presented in Appendix A.18.

can help to assess the external validity of the results – for example, due to their high incomes or the competitive nature of their profession, it is possible that doctors have stronger relative concerns compared to the general population.

Also, we extended the survey instrument used in this auxiliary sample to test a specific explanation for the preference for relative income: individuals may use information about relative income to learn about other city characteristics such as school quality and crime rates. Towards the end of the auxiliary survey, after individuals received feedback about the cost of living and the relative income, we included a set of additional questions eliciting beliefs about other attributes of the two cities under consideration. We picked eight attributes that individuals could arguably perceive to be correlated to the affluence of a city: quality of schools, crime rates, quality of health services, quality of public spaces, quality of the environment, quality of entertainment, share of college graduates, and share of supporters of Donald Trump. If individuals learn about these attributes from the earnings rank, then controlling for those attributes should mute the association between perceived earnings rank and choice.

## 7.2 Results: Replication of the Main Experiment

Table 5 replicates the preference estimation from Table 4, but using data from the auxiliary experiment instead of the main experiment. The comparison between Tables 5 and 4 can shed light on the robustness and external validity of the results from the main experiment.

The first row of Table 5 corresponds to the baseline estimates, which uses the experimental and non-experimental variations in beliefs. The coefficients  $\beta^{ER}$  and  $\beta^{COL}$  are similar between the main experiment and the auxiliary experiment. Focusing on the entire sample, the estimated  $\beta^{ER}$  and  $\beta^{COL}$  are 1.141 and -1.262 in the main experiment (p-value=0.048 and p-value=0.017), and 1.293 and -1.962 in the auxiliary experiment (p-value<0.001 for both). That is, the coefficients have the same signs and similar magnitudes. The coefficients are more precisely estimated in the auxiliary sample, in part due to the larger sample size.

To compare the magnitude of relative income concerns, we can compare the marginal rate of substitution between relative income and cost of living: i.e.,  $\frac{\beta^{ER}}{-\beta^{COL}}$ . In the main experiment,  $\frac{\beta^{ER}}{-\beta^{COL}}$  is 0.90 (s.e. 0.64; from the first row, columns (3) and (6) of Table 4). In the auxiliary experiment, we find a corresponding ratio of 0.66 (s.e. 0.20; from the first row, columns (3) and (6) of Table 5). That is, the auxiliary experiment suggests slightly weaker preferences for relative income than the main experiment, but that difference is statistically insignificant. This constitutes suggestive evidence that, despite large observable differences in observable characteristics, medical students are not special in terms of their preferences for relative income.

The second row presents the experimental estimates. The results from the auxiliary experiment are even more robust than the results from the main experiment. In the main experiment, the baseline estimates for  $\beta^{ER}$  are similar to the experimental estimates, and this is true again in the auxiliary experiment. In the main experiment, the experimental estimates for  $\beta^{COL}$  are statistically insignificant and smaller in magnitude than the baseline estimates. In the auxiliary sample, the experimental estimates for  $\beta^{COL}$  are negative, precisely estimated, and statistically significant at the 1% level.

Another important finding from the main experiment is the heterogeneity in  $\beta^{ER}$  by relationship status. In the auxiliary experiment, we find evidence consistent with this heterogeneity, although it is less extreme. Table 5 shows that, when we break down  $\beta^{ER}$  by relationship status, the coefficient of  $\beta^{ER}$  is smaller among singles than among non-singles. This difference, however, becomes stronger in the experimental estimates:  $\beta^{ER}$  is 2.578 for singles and 0.664 for non-singles. However, there are two notable differences: the difference is statistically insignificant in the auxiliary experiment (p-value=0.244), and less pronounced in magnitude than the corresponding heterogeneity in the main experiment. One potential interpretation for this finding is that the results from the main sample exaggerate the degree of heterogeneity by relationship status. Another potential interpretation is that the difference in results are due to differences in the characteristics across the two samples. For example, compared to singles in the auxiliary sample, singles in the medical student sample are much more likely to be in prime dating age, less likely to have children from previous relationships, and they expect a much higher permanent income.

### 7.3 Results: Disentangling Mechanisms

One possible interpretation of the coefficient  $\beta^{ER}$  is that individuals use their prospective relative income as a signal for other city attributes. Given that medical students are highly informed and are making a high-stakes decision, this possibility seems unlikely. These students devote their entire fourth year of medical school to the Match. After the hospital visits, they have about two months to finalize their rankings. During this time, they continue to gather information to aid their decision. Because they have been to these locations and can easily obtain additional information directly, it is unlikely that they would rely on earnings rank to learn about other features of the locations. In any case, we present two tests for this mechanism.

One version of this mechanism is that individuals use information about relative income to make inferences about the expected cost of living. If participants in the NRMP believe that their earnings rank reflects the degree of competition with their neighbors for some goods, such as housing, it would be natural for them to learn about cost of living from information

about their relative income. We can test this directly by examining how the information provision experiment affected posterior beliefs about the expected cost of living. In Figure 6 we use similar learning regressions from Section 5.2, only that we measure the effect of relative income feedback on beliefs about cost of living (and vice versa).

Figure 6.a shows the effect of earnings rank feedback on posterior beliefs about the cost of living from the baseline survey (i.e., the short-term effect). The slope is close to zero (-0.003), precisely estimated (s.e. 0.006), and statistically insignificant. This coefficient suggests that increasing the observed earnings rank by 1 percentage point reduces posterior beliefs about cost of living by 0.003 percentage points. To put this magnitude in context, Figure 5.d suggests that the effect of earnings rank on posterior beliefs about earnings rank is 0.873 (s.e. 0.011). The difference between this 0.873 and the -0.003 effect is economically large and statistically highly significant. Figure 6.b shows the effect of earnings rank feedback on beliefs about the cost of living from the follow-up survey (i.e., the long-term effect). Again, the effect is close to zero (-0.011), precisely estimated (s.e. 0.011), and statistically insignificant.<sup>48</sup>

A second version of this mechanism is that individuals use information about relative income to learn about other city characteristics, such as school quality and crime rates. Although these inferences would not be unreasonable, this mechanism would probably work against our main finding: if more affluent ponds tend to have desirable amenities, then individuals should prefer to live in more affluent ponds, which is the opposite of what we find.

We can test this hypothesis using the perceptions about additional characteristics elicited in the auxiliary experiment. One important difference in context is that, whereas subjects in the main experiment made a high-stakes decision for which they obtained substantial information, subjects in the auxiliary experiment had no immediate incentives to be informed about the attributes of these cities and then may make more sense for them to use information about earnings rank to make inferences about other unobserved city attributes. In other words, this mechanism would arguably play a bigger role in the auxiliary experiment than in the main experiment.

The third row of Table 5 estimates the experimental model, but adding these eight perceptions as additional control variables. If individuals care about relative income because they learn about the other characteristics, the coefficient on  $\beta^{ER}$  should be muted after controlling for these additional perceptions.<sup>49</sup> On the contrary, the comparison between the second and

---

<sup>48</sup>As additional evidence that subjects see cost of living and relative income as two distinct features of the city, Figure 6.c and 6.d show that the converse also is true: feedback about cost of living does not affect short-term or long-term beliefs about relative income.

<sup>49</sup>Consistent with the less informed nature of this subject pool, we find that the feedback about cost of living and earnings rank did affect a few of these additional beliefs. Results presented in Appendix A.15.

third rows of Table 5 suggests that controlling for the additional characteristics, if anything, increases the value of  $\beta^{ER}$ . For instance, among non-singles, the experimental estimate for  $\beta^{ER}$  is 2.578 (p-value=0.011) without these additional controls and 3.048 (p-value=0.004) with these additional controls. The difference between these two coefficients is statistically insignificant. Also, the increase in  $\beta^{ER}$  caused by adding the extra controls is consistent with the previous argument that, if anything, this mechanism leads to an underestimation of preferences for relative income.

## 8 Conclusions

In this paper, we develop a new methodology to measure preferences for city amenities, and apply it to an amenity that had not yet explored: relative income. We implemented this methodology with a sample of 1,080 medical students who participated in the National Resident Matching Program. We find unique revealed-preference evidence that, when choosing where to live, individuals care about their prospective relative income. Furthermore, we find that individuals can differ substantially in their preferences for relative income: non-single individuals want to live in less affluent ponds, whereas single individuals prefer to live in more affluent ponds.

One interpretation for the average preferences, although not necessarily the only one, is that the average individual sees more affluent neighbors as a negative externality. For example, more affluent ponds may result in lower social status (Luttmer 2005), lower self-esteem (Festinger 1954), unfavorable treatment in social interactions (Doob and Gross 1968; Fennis 2008; Nelissen and Meijers 2011) and higher consumption aspirations (Frank 1985b). On the other hand, the heterogeneity by relationship status suggests that, relative to non-singles, singles are less pessimistic about the externalities from affluent neighbors. For example, singles may expect to benefit from public goods financed by affluent neighbors (Gautier et al. 2010), and from mingling and dating with affluent neighbors (Fisman et al. 2006; Gautier et al. 2010).

An avenue for future research is to find other contexts in which this revealed-preference method could be used to estimate preferences for relative income as well as other city amenities. For instance, although the settings may not be as clear as they are for the medical residency, multiple job markets require job seekers to choose between offers in different cities. Using a broader subject pool will help generalize the findings from this study and provide more data to study heterogeneity in preferences.

Future research should also investigate the mechanisms underlying individuals' concerns about relative income. There is, for instance, little evidence as to whether relative concerns

respond to instrumental motives (e.g., dating prospects) or non-instrumental motives (e.g., envy). Similarly, there is little evidence about whether individuals care about their self-perceptions of relative income (i.e., their self-image) or about how others perceive their relative income (i.e., their social image). These additional hypotheses can be explored by using the same empirical framework proposed in this paper, but with additional treatment arms designed to test specific mechanisms.

## References

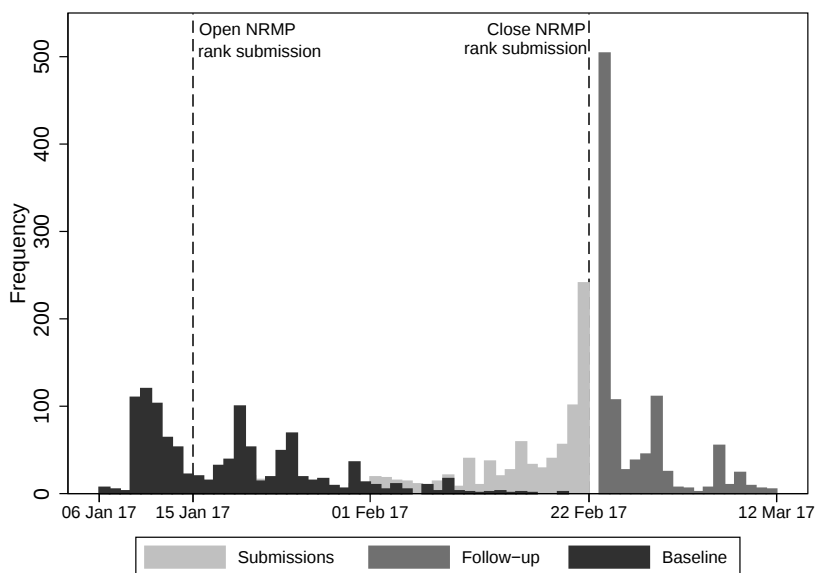
- Agarwal, N. (2015). An empirical model of the medical match. *The American Economic Review* 105(7), 1939–1978.
- Albouy, D. (2008). Are big cities bad places to live? estimating quality of life across metropolitan areas. *NBER Working Paper No. 14472*.
- Albouy, D. (2016). What are cities worth? land rents, local productivity, and the total value of amenities. *Review of Economics and Statistics* 98(3), 477–487.
- Armantier, O., S. Nelson, G. Topa, W. van der Klaauw, and B. Zafar (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics* 98(3), 503–523.
- Bayer, P., F. Ferreira, and R. McMillan (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* 115(4), 588–638.
- Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2012). What do you think would make you happier? what do you think you would choose? *The American Economic Review* 102(5), 2083–2110.
- Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2014). Can marginal rates of substitution be inferred from happiness data? evidence from residency choices. *The American Economic Review* 104(11), 3498–3528.
- Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4), 1165–1188.
- Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics* 130(2), 571–614.
- Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2017). Status goods: experimental evidence from platinum credit cards. *NBER Working Paper No. 23414*.
- Bursztyn, L., T. Fujiwara, and A. Pallais (2017). Acting wife: Marriage market incentives and labor market investments. *The American Economic Review* 107(11), 3288–3319.
- Cavallo, A., G. Cruces, and R. Perez-Truglia (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics* 9(3), 1–35.
- Charles, K. K., E. Hurst, and N. Roussanov (2009). Conspicuous consumption and race. *The Quarterly Journal of Economics* 124(2), 425–467.

- Clark, A., C. Senik, and K. Yamada (2017). When experienced and decision utility concur: The case of income comparisons. *Journal of Behavioral and Experimental Economics* 70, 1 – 9.
- Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics* 61(3), 359–381.
- Clark, A. E., N. Westergård-Nielsen, and N. Kristensen (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association* 7(2-3), 519–527.
- Couture, V. and J. Handbury (2016). Urban revival in america, 2000 to 2010. *University of Pennsylvania. Mimeo.*
- Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics* 98, 100–112.
- Cullen, Z. and R. Perez-Truglia (2018). How much does your boss make? the effects of salary comparisons. *NBER Working Paper No. 24841.*
- Daniel, C. and M. O’Brien (2008). Why study medicine? *The Student Doctor Network April 24.*
- Diamond, R. (2016). The determinants and welfare implications of us workers’ diverging location choices by skill: 1980-2000. *American Economic Review* 106(3), 479–524.
- Doob, A. N. and A. E. Gross (1968). Status of frustrator as an inhibitor of horn-honking responses. *The Journal of Social Psychology* 76(2), 213–218.
- Easterlin, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. *Nations and Households in Economic Growth* 89, 89–125.
- Fennis, B. M. (2008). Branded into submission: Brand attributes and hierarchization behavior in same-sex and mixed-sex dyads. *Journal of Applied Social Psychology* 38(8), 1993–2009.
- Ferrer-i Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics* 89(5), 997–1019.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations* 7(2), 117–140.
- Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics* 121(2), 673–697.
- Frank, R. H. (1985a). *Choosing the right pond: Human behavior and the quest for status.* Oxford University Press.

- Frank, R. H. (1985b). The demand for unobservable and other nonpositional goods. *The American Economic Review* 75(1), 101–116.
- Gautier, P. A., M. Svarer, and C. N. Teulings (2010). Marriage and the city: Search frictions and sorting of singles. *Journal of Urban Economics* 67(2), 206 – 218.
- Heffetz, O. (2011). A test of conspicuous consumption: Visibility and income elasticities. *Review of Economics and Statistics* 93(4), 1101–1117.
- Hitsch, G., A. Hortacsu, and D. Ariely (2010). What makes you click: An empirical analysis of online dating. *Quantitative Marketing and Economics* 8(4), 393–427.
- Johansson-Stenman, O., F. Carlsson, and D. Daruvala (2002). Measuring future grandparents’ preferences for equality and relative standing. *The Economic Journal* 112(479), 362–383.
- Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3), 430–454.
- Karadja, M., J. Mollerstrom, and D. Seim (2017). Richer (and holier) than thou? the effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics* 99(2), 201–212.
- Kuziemko, I., R. W. Buell, T. Reich, and M. I. Norton (2014). Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics* 129(1), 105–149.
- Luttmer, E. F. (2005). Neighbors as negatives: relative earnings and well-being. *The Quarterly Journal of Economics* 120(3), 963–1002.
- Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics* 5(1), 65–103.
- Nelissen, R. M. and M. H. Meijers (2011). Social benefits of luxury brands as costly signals of wealth and status. *Evolution and Human Behavior* 32(5), 343–355.
- NRMP (2015). Results of the 2015 nrmp applicant survey by preferred specialty and applicant type. *National Resident Matching Program, Washington, DC. 2015.*
- Perez-Truglia, R. (2016). The effects of income transparency on well-being: evidence from a natural experiment. *Mimeo.*
- Perez-Truglia, R. (2017). Political conformity: Event-study evidence from the united states. *Review of Economics and Statistics forthcoming.*
- Perez-Truglia, R. and G. Cruces (2017). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy* 125(4), 1208–1243.

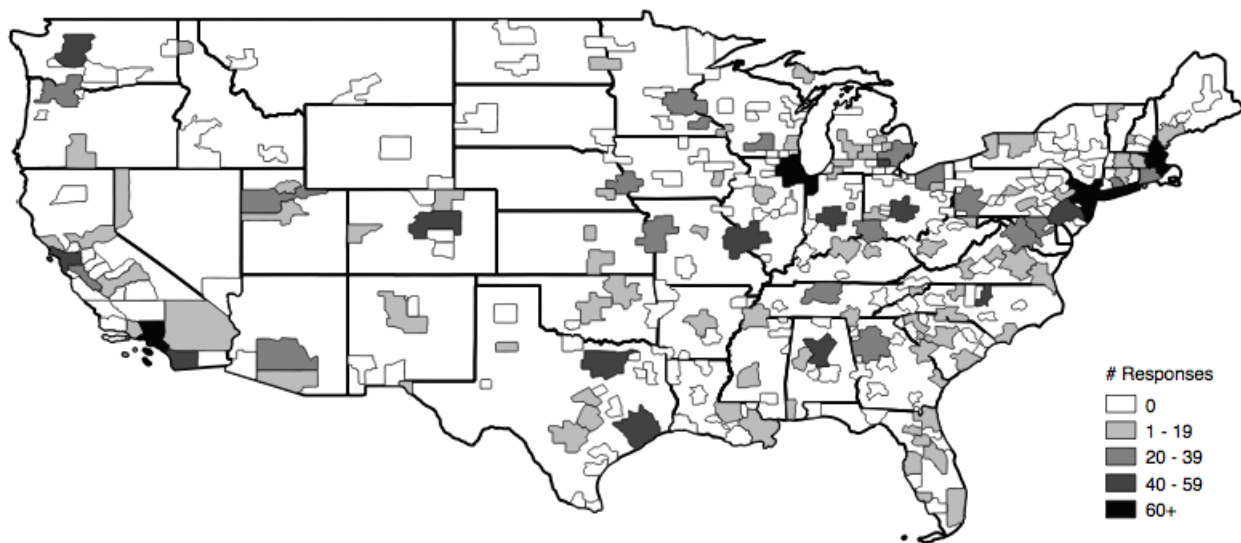
- Rees-Jones, A. (2017). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior In Press*.
- Rees-Jones, A. and S. Skowronek (2017). Why do we lie in incentive-compatible mechanisms? evidence from the residency match. *Mimeo*.
- Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy* 90(6), 1257–1278.
- Rosen, S. (1979). Wage-based indexes of urban quality of life. *Current Issues in Urban Economics*, 74–104.
- Roth, A. E. and E. Peranson (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *The American Economic Review* 89(4), 748.
- Roth, C. (2015). Conspicuous consumption and peer effects: Evidence from a randomized field experiment. *Mimeo*.
- Senik, C. (2004). When information dominates comparison: Learning from russian subjective panel data. *Journal of Public Economics* 88(9), 2099–2123.
- Shigeoka, H. and K. Yamada (2016). Income-comparison attitudes in the us and the uk: Evidence from discrete-choice experiments. *NBER Working Paper No. 21998*.
- Sledge, W. H., P. Leaf, and M. Sacks (1989). Applicants perceptions of psychiatric residency training programs. *Academic Psychiatry* 13(1), 24–30.
- Solnick, S. J. and D. Hemenway (1998). Is more always better? a survey on positional concerns. *Journal of Economic Behavior & Organization* 37(3), 373–383.
- Tideman, S., P. Frijters, and M. A. Shields (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic Literature* 46(1), 95–144.
- Van de Stadt, H., A. Kapteyn, and S. Van de Geer (1985). The relativity of utility: Evidence from panel data. *The Review of Economics and Statistics* 67(2), 179–187.
- Van Gelder, M. M., R. W. Bretveld, and N. Roeleveld (2010). Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology* 172(11), 1292–1298.
- Wiswall, M. and B. Zafar (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies* 82(2), 791–824.
- Yamada, K. and M. Sato (2016). *Another avenue for anatomy of income comparisons: Evidence from hypothetical choice experiments*. Behavioral Economics of Preferences, Choices, and Happiness; Springer.

Figure 1: Distribution Over Time of Survey Responses and NRMP Rank Submissions



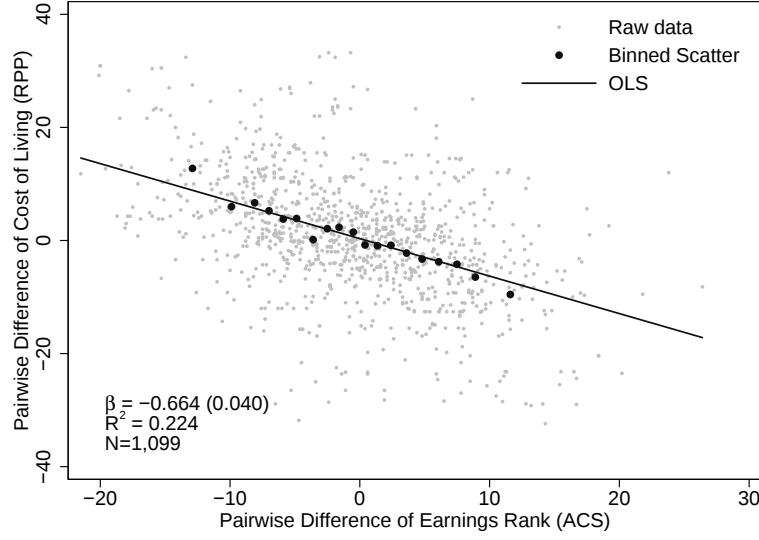
Notes: Distribution of timing of responses to Baseline and Follow-up Surveys, and NRMP rank submission dates (as reported by respondents in the follow-up survey).

Figure 2: Geographic Distribution of Choice Set



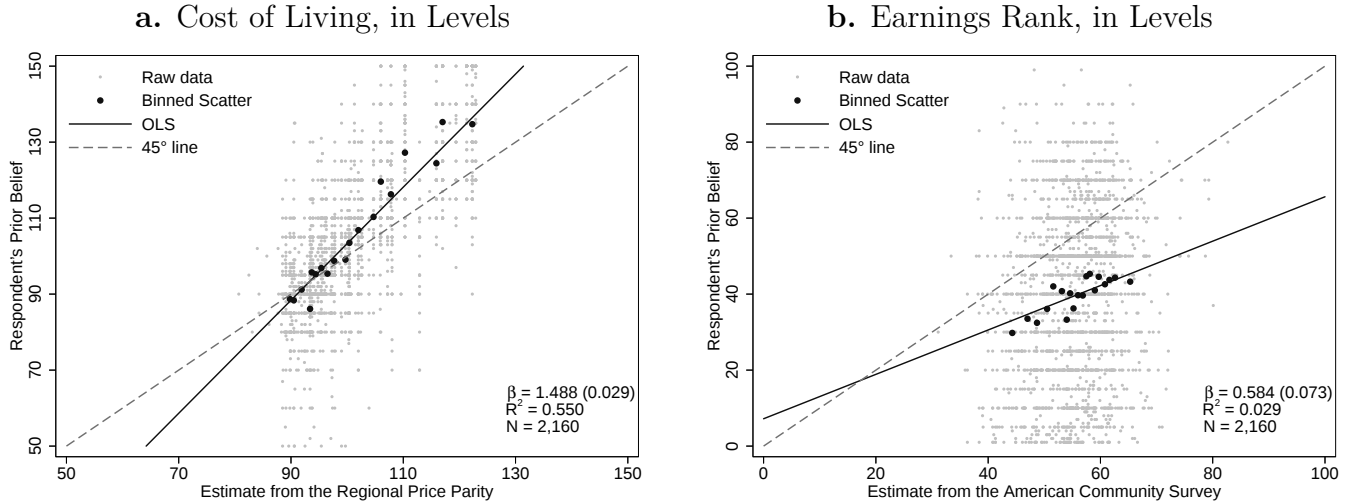
Notes: Geographical distribution of metropolitan areas where top-2 residency programs of respondents are located, for the continental United States. No responses were located in Hawaii, while Alaska only has 2 responses. Only metropolitan areas with a residency program participating in the 2017 NRMP are displayed (279 in total).

Figure 3: Variation in Cost of Living and Earnings Rank



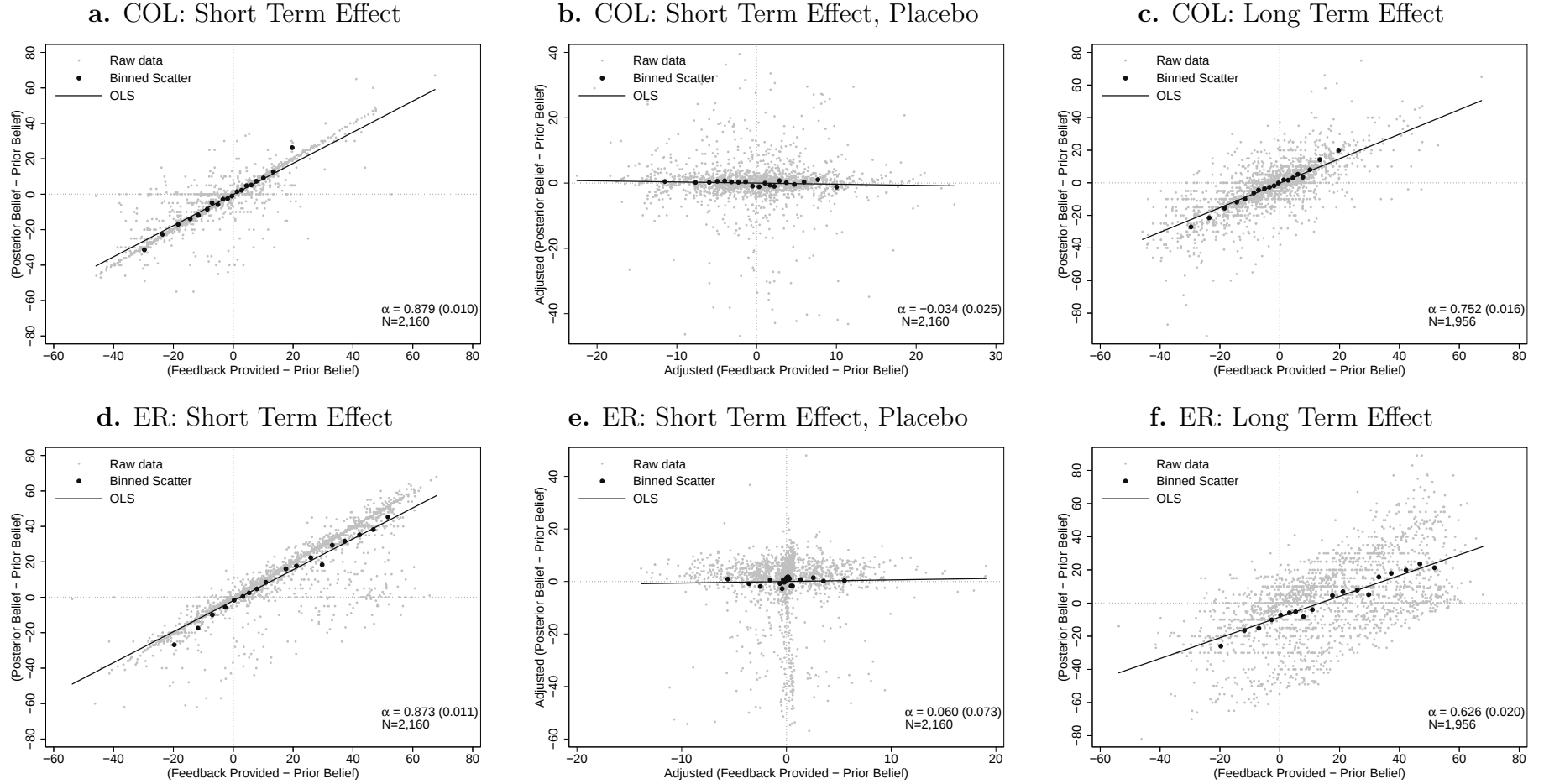
Notes: The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Slopes ( $\beta$ , with robust standard errors in parentheses) and  $R^2$  are based on a linear regression. All variables for x-axis and y-axis correspond to pairwise differences across the two cities that the subject is considering submitting to the algorithm. Data from survey responses, the Regional Price Parity Index (for cost of living) and the American Community Survey (for earnings rank).

Figure 4: Comparison Between Prior Beliefs and Statistics



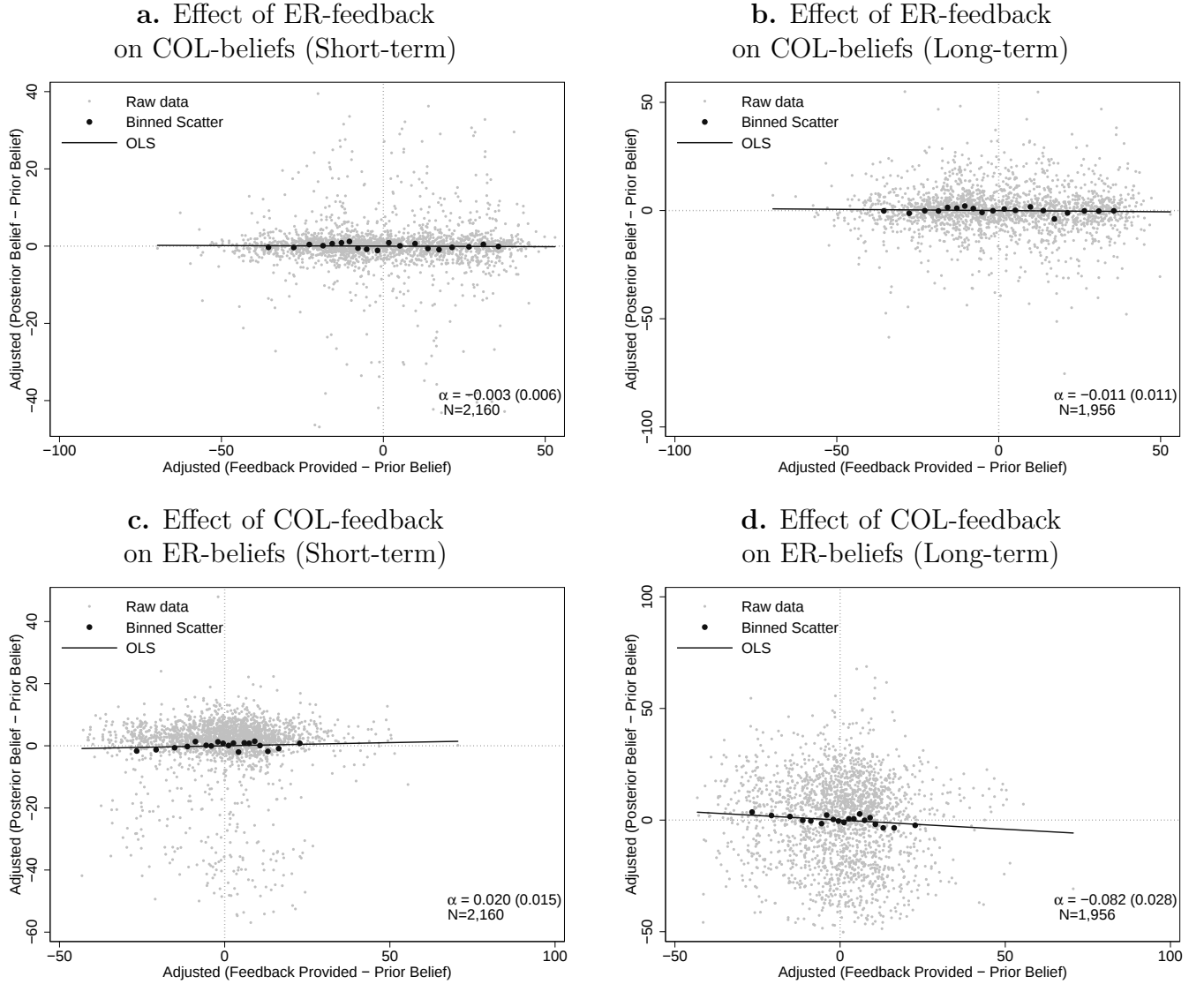
Notes: Comparison between respondent's perceptions before the information provision (i.e., prior beliefs) and statistics. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels a and b present data in levels (i.e., two observations per individual, one for each of their options). The slope ( $\beta$ , with robust standard errors in parentheses) and  $R^2$  are based on a linear regression.

Figure 5: Learning from the Experimental Feedback



Notes: Comparison between the difference in statistics and respondent's perceptions before the information provision (i.e., prior beliefs), and difference in respondent's perceptions after the information provision (i.e., posterior beliefs) and prior beliefs. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panel b and e shows a placebo test where we compare the difference between the alternative feedback and prior belief to the difference between the posterior and prior beliefs, adjusting for the shown statistic. Panel c and f uses respondent's perceptions measured in the follow-up survey as posterior belief. The slope ( $\alpha$ , with robust standard errors in parentheses) is based on a linear regression.

Figure 6: Effect of Earning Rank Feedback on Posterior Belief on Cost of Living (and vice-versa)



Notes: Comparison between the difference in statistics and respondent's perceptions before the information provision (i.e., prior beliefs), and difference in respondent's perceptions after the information provision (i.e., posterior beliefs) and prior beliefs. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels shows the extent to which respondents adjust their perceptions on earnings rank (cost of living) as a result in their perception gap in cost of living (earnings rank) adjusted for the perceptions gap in ER (COL). The slope ( $\alpha$ , with robust standard errors in parentheses) is based on a linear regression.

Table 1: Descriptive Statistics and Randomization Balance

|                      | All               | RPP; ACS          | RPP; CPS          | COLI; ACS         | COLI; CPS         | F-test<br>P-value |
|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Male (=1)            | 0.481<br>(0.015)  | 0.452<br>(0.030)  | 0.491<br>(0.031)  | 0.481<br>(0.031)  | 0.502<br>(0.030)  | 0.688             |
| Age                  | 27.091<br>(0.083) | 27.092<br>(0.164) | 27.104<br>(0.165) | 26.985<br>(0.145) | 27.181<br>(0.187) | 0.863             |
| Nr Kids              | 0.132<br>(0.014)  | 0.125<br>(0.027)  | 0.164<br>(0.033)  | 0.104<br>(0.026)  | 0.137<br>(0.029)  | 0.553             |
| Single (=1)          | 0.354<br>(0.015)  | 0.401<br>(0.030)  | 0.312<br>(0.028)  | 0.343<br>(0.029)  | 0.358<br>(0.029)  | 0.189             |
| Dual Match (=1)      | 0.074<br>(0.008)  | 0.077<br>(0.016)  | 0.059<br>(0.014)  | 0.104<br>(0.019)  | 0.055<br>(0.014)  | 0.157             |
| US News Rank         | 58.81<br>(0.787)  | 58.849<br>(1.612) | 59.104<br>(1.560) | 58.604<br>(1.568) | 58.683<br>(1.565) | 0.996             |
| Prior: $COL_{1,2}^i$ | 0.409<br>(0.640)  | 0.445<br>(1.364)  | -0.238<br>(1.134) | -0.567<br>(1.308) | 1.982<br>(1.303)  | 0.506             |
| Prior: $ER_{1,2}^i$  | 0.394<br>(0.467)  | 0.162<br>(0.903)  | 0.71<br>(0.925)   | -0.526<br>(0.906) | 1.221<br>(0.999)  | 0.595             |
| Observations         | 1,080             | 272               | 269               | 268               | 271               |                   |

Notes: Individual characteristics obtained from baseline survey. Column (1) corresponds to all respondents, and columns (2) through (4) correspond to each of the four treatment groups given by all the possible combinations from the source-randomization experiment. RPP and COLI are the two sources used to compute the cost of living feedback (corresponding to the Regional Price Parity Index and the Cost of Living Index, respectively). ACS and CPS are the two sources used to compute the earnings ranking feedback (corresponding to the American Community Survey and the Current Population Survey, respectively). The final column presents p-value for test of the null hypothesis that the mean characteristic is equal across all four treatment groups. All variables constructed from the survey data, except for the U.S. News Rank which was taken from the U.S. News rank of medical schools for 2016.

Table 2: Location Preferences: Baseline Estimates

|                                   |                     | By Relationship Status |                    | By Gender         |                    | By Specialty Salary     |                      |
|-----------------------------------|---------------------|------------------------|--------------------|-------------------|--------------------|-------------------------|----------------------|
|                                   | All<br>(1)          | Non-Single<br>(2)      | Single<br>(3)      | Female<br>(4)     | Male<br>(5)        | $\leq$ \$229,000<br>(6) | $>$ \$229,000<br>(7) |
| $\beta^{ER}$                      | 0.995*<br>(0.539)   | 2.236***<br>(0.669)    | -1.538*<br>(0.880) | 1.041<br>(0.755)  | 0.896<br>(0.781)   | 1.433*<br>(0.732)       | 0.777<br>(0.797)     |
| $\beta^{COL}$                     | -1.073**<br>(0.485) | -1.087<br>(0.663)      | -1.058<br>(0.749)  | -0.972<br>(0.679) | -1.443*<br>(0.753) | -0.690<br>(0.713)       | -1.238*<br>(0.690)   |
| Diff. P-value [ <i>q-value</i> ]: |                     |                        |                    |                   |                    |                         |                      |
| ER                                |                     | 0.001 [0.030]          |                    | 0.894 [0.974]     |                    | 0.544 [0.954]           |                      |
| COL                               |                     | 0.977 [0.977]          |                    | 0.642 [0.954]     |                    | 0.580 [0.954]           |                      |
| Observations                      | 1,080               | 698                    | 382                | 560               | 520                | 549                     | 531                  |

Notes: Heteroskedasticity-robust standard errors in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Raw Probit coefficients. Each column corresponds to a different Probit regression of expected rank order submission on posterior beliefs about cost of living and earnings rank, from the baseline survey, including the baseline controls listed in section 3. Columns (2) through (7) show estimates when restricting sample to different subgroups: columns 2 and 3 by Non-Single (i.e., married or in a long-term relationship) or Single, columns 4 and 5 by gender, columns 5 and 7 by expected salary of specialty after residency (over and below the median value of \$229,000). P-values corresponds to the test of the null hypothesis that the coefficients are equal between the two sub-groups, multiple-testing q-values based on Benjamini and Yekutieli (2001) presented in brackets.

Table 3: Location Preferences: Robustness to Alternative Control Variables

|                      | Panel A: $\beta^{ER}$ |                     |                    | Panel B: $\beta^{COL}$ |                      |                      |
|----------------------|-----------------------|---------------------|--------------------|------------------------|----------------------|----------------------|
|                      | Non-Single<br>(1)     | Single<br>(2)       | All<br>(3)         | Non-Single<br>(4)      | Single<br>(5)        | All<br>(6)           |
| No Controls          | 1.961***<br>(0.663)   | -1.480*<br>(0.841)  | 0.873*<br>(0.531)  | -0.812<br>(0.523)      | -1.131*<br>(0.589)   | -0.894**<br>(0.382)  |
| Baseline             | 2.236***<br>(0.669)   | -1.538*<br>(0.880)  | 0.995*<br>(0.539)  | -1.087<br>(0.663)      | -1.058<br>(0.749)    | -1.073**<br>(0.485)  |
| Demographic          | 2.288***<br>(0.715)   | -0.871<br>(0.977)   | 1.176**<br>(0.578) | -1.219*<br>(0.628)     | -1.712**<br>(0.713)  | -1.342***<br>(0.468) |
| Amenities            | 2.056***<br>(0.669)   | -1.381<br>(0.853)   | 0.958*<br>(0.538)  | -0.718<br>(0.630)      | -1.265<br>(0.816)    | -0.898*<br>(0.481)   |
| Geography            | 2.064***<br>(0.733)   | -1.551<br>(1.004)   | 1.001*<br>(0.593)  | -1.626**<br>(0.652)    | -1.783***<br>(0.646) | -1.572***<br>(0.461) |
| Economic             | 1.914***<br>(0.684)   | -1.191<br>(0.941)   | 0.946*<br>(0.566)  | -0.467<br>(0.670)      | -1.647**<br>(0.812)  | -0.868*<br>(0.498)   |
| State Dummies        | 2.901***<br>(0.703)   | -1.943**<br>(0.907) | 1.084*<br>(0.555)  | -1.090<br>(0.671)      | -1.219<br>(0.939)    | -0.968*<br>(0.502)   |
| Obj. Program Chars.  | 1.987***<br>(0.686)   | -1.451*<br>(0.842)  | 0.919*<br>(0.540)  | -0.868*<br>(0.527)     | -1.141*<br>(0.597)   | -0.912**<br>(0.386)  |
| Subj. Program Chars. | 2.222***<br>(0.730)   | -1.320<br>(1.040)   | 1.199**<br>(0.605) | -1.210**<br>(0.587)    | -1.678***<br>(0.619) | -1.277***<br>(0.425) |
| All Controls         | 2.221**<br>(0.941)    | -3.177*<br>(1.854)  | 0.703<br>(0.700)   | -0.295<br>(1.083)      | -5.067***<br>(1.712) | -0.862<br>(0.704)    |

Notes: Heteroskedasticity-robust standard errors in parenthesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Raw Probit coefficients. Each row corresponds to a separate regression of expected rank on posterior beliefs about cost of living and earnings rank, from the baseline survey. All regressions include as controls the log difference in nominal income and a constant. The first row does not include any additional controls. The second row includes the baseline controls listed in section 3. The third to last rows use different sets of additional controls, listed in section 6.3. Results are based on 1,080 individual responses (698 from non-singles and 382 from singles), except for the last row, which is restricted to the follow-up sample (978 responses, 595 from non-singles and 311 from singles).

Table 4: Location Preferences: Experimental Estimates

|                             | Panel A: $\beta^{ER}$ |                      |                    | Panel B: $\beta^{COL}$ |                    |                     |
|-----------------------------|-----------------------|----------------------|--------------------|------------------------|--------------------|---------------------|
|                             | Non-Single<br>(1)     | Single<br>(2)        | All<br>(3)         | Non-Single<br>(4)      | Single<br>(5)      | All<br>(6)          |
| Baseline                    | 2.380***<br>(0.702)   | -1.656*<br>(0.991)   | 1.141**<br>(0.577) | -1.234*<br>(0.743)     | -1.379*<br>(0.772) | -1.262**<br>(0.531) |
| Experimental                | 2.977**<br>(1.331)    | -4.964**<br>(1.974)  | 0.867<br>(1.151)   | 0.353<br>(1.160)       | 1.663<br>(1.286)   | 0.662<br>(0.881)    |
| Experimental, Long Term     | 1.993*<br>(1.188)     | -5.285***<br>(1.984) | -0.029<br>(1.071)  | 1.662*<br>(1.005)      | 0.251<br>(1.359)   | 1.012<br>(0.821)    |
| Experimental, Falsification | -0.007<br>(0.998)     | 0.040<br>(1.732)     | 0.004<br>(0.837)   | 0.037<br>(0.855)       | 0.021<br>(1.123)   | 0.031<br>(0.651)    |

Notes: Heteroskedasticity-robust standard errors in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Raw Probit (or IV-Probit) coefficients restricting sample to respondents who completed the follow-up survey. All regressions include the baseline controls listed in section 3. The independent variables are the posterior beliefs about cost of living and earnings rank, from the baseline specification. The first row corresponds to the baseline Probit specification. The second through third row correspond to IV-Probit regressions, using the variation in perceptions generated by the source-randomization experiment as instrumental variables. The first and second rows use the expected rank order submission (from the baseline survey) as dependent variable. The third row uses the final rank order submission (from the follow-up survey) as dependent variable. The fourth row provides corresponds to a falsification test that uses the same IV-Probit specification from the second row, but using the list order (i.e., the order in which programs are listed at the beginning of the survey) as dependent variable instead of the rank order. To estimate this IV-Probit model, we randomly assign programs to be program 1 and program 2, and then we use as dependent variable a dummy that takes the value 1 if program 1 was listed first at the beginning of the survey. We repeat this procedure 1,000 times and report the average and standard error from the distribution of coefficients. All results based on the sample of individuals who completed the follow-up survey (978 responses, 647 from non-singles and 311 from singles).

Table 5: Location Preferences: Auxiliary Experiment

|                                   | Panel A: $\beta^{ER}$ |                    |                     | Panel B: $\beta^{COL}$ |                      |                      |
|-----------------------------------|-----------------------|--------------------|---------------------|------------------------|----------------------|----------------------|
|                                   | Non-Single<br>(1)     | Single<br>(2)      | All<br>(3)          | Non-Single<br>(4)      | Single<br>(5)        | All<br>(6)           |
| Baseline                          | 1.408***<br>(0.376)   | 1.095**<br>(0.478) | 1.293***<br>(0.292) | -2.203***<br>(0.463)   | -1.618***<br>(0.566) | -1.962***<br>(0.364) |
| Experimental                      | 2.578**<br>(1.019)    | 0.664<br>(1.272)   | 1.706**<br>(0.816)  | -2.385***<br>(0.666)   | -2.956***<br>(0.917) | -2.528***<br>(0.531) |
| Experimental, Additional Controls | 3.048***<br>(1.064)   | 0.452<br>(1.430)   | 1.902**<br>(0.872)  | -2.329***<br>(0.691)   | -3.753***<br>(0.906) | -2.688***<br>(0.563) |

Notes: Heteroskedasticity-robust standard errors in parenthesis. All regressions include the baseline controls listed in section 3 with the exception of program characteristics. The independent variables are the posterior beliefs about cost of living and earnings rank, from the baseline specification. The first row corresponds to the baseline Probit specification. The second through third row correspond to IV-Probit regressions, using the variation in perceptions generated by the source-randomization experiment as instrumental variables. The third row includes additional controls for differences in city perceptions: quality of schools, crime rates, quality of health, quality of public spaces, quality of the environment, quality of entertainment, quality of colleges, fraction voting Trump in presidential election. All results based on the sample of respondents in the United States on Amazon Mechanical Turk (1,245 responses, 829 from non-singles and 416 from singles).