Choosing Your Pond

Location Choices and Relative Income

Nicolas L. Bottan
Cornell University

Ricardo Perez-Truglia
UCLA-Anderson

Wharton
September 2019

Introduction

- Do individuals care about their absolute incomes or their relative incomes?
- Long-standing hypothesis dating back to Adam Smith.
- Large literature on relative concerns based on happiness data.
- However, revealed-preference evidence remains elusive.
  - Our contribution: propose revealed-preference test of this hypothesis.

Our Contribution

- Where you choose to live will determine your reference group (Frank, 1985).
  - Move to a rich pond: your relative income will be low.
  - Move to a poor pond: your relative income will be high.
- Key insight: preferences for relative income can be inferred from residential choices.
- **Research Question:** When choosing where to live, do individuals care about their prospective relative income?

Ideal Choice Data

In this Paper

- Use the National Resident Matching Program (NRMP) as a natural laboratory.
  - Participants must choose between programs located in different cities.
  - They get paid the same nominal income everywhere.
  - Thus, they face different relative incomes (and also cost of living) in different destinations.

Preview of Methods and Results

- Preview of Methods:
  - Field experiment with 1,100 NRMP participants.
  - Collect survey data to estimate preferences over relative income (and cost of living).
  - Use information-provision experiment to generate exogenous variation in beliefs.
- Preview of Results:
  - Average individual prefers higher relative income.
  - Substantial heterogeneity between single and non-single individuals.

Related Literature

- **Subjective well-being.** (e.g., Easterlin, 1974; Luttmer, 2005; Perez-Truglia, 2018).
- **Laboratory experiments.** (e.g., Kuziemko et al., 2014; Yamada and Sato, 2016).
- **Location preferences.** (e.g., Black, 1999; Bayer, Ferrerira and McMillan, 2007; Moretti, 2013; Diamond, 2016).

Outline

Institutional Context: NRMP

ARTHUR

1. CITY

SUNNY

1. CITY
2. MERCY

JOSEPH

1. CITY
2. GENERAL
3. MERCY

LATHA

1. MERCY
2. CITY
3. GENERAL

DARRIUS

1. CITY
2. MERCY
3. GENERAL

MERCY (2 POSITIONS)

1. DARRIUS
2. JOSEPH

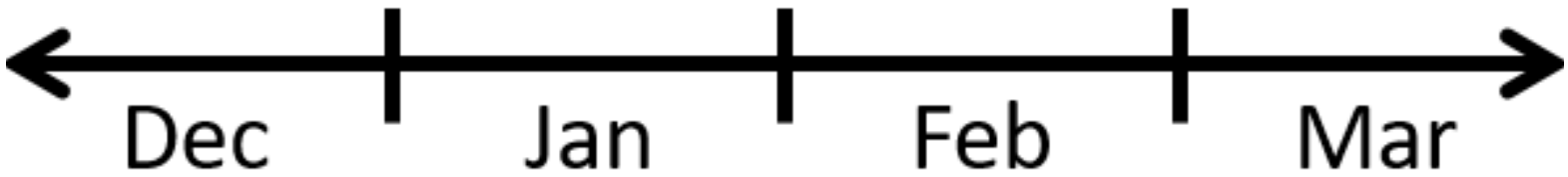
CITY (2 POSITIONS)

1. DARRIUS
2. ARTHUR
3. SUNNY
4. LATHA
5. JOSEPH

GENERAL (2 POSITIONS)

1. DARRIUS
2. ARTHUR
3. JOSEPH
4. LATHA

Institutional Context: Timeline



Institutional Context: NRMP Timeline

Dec

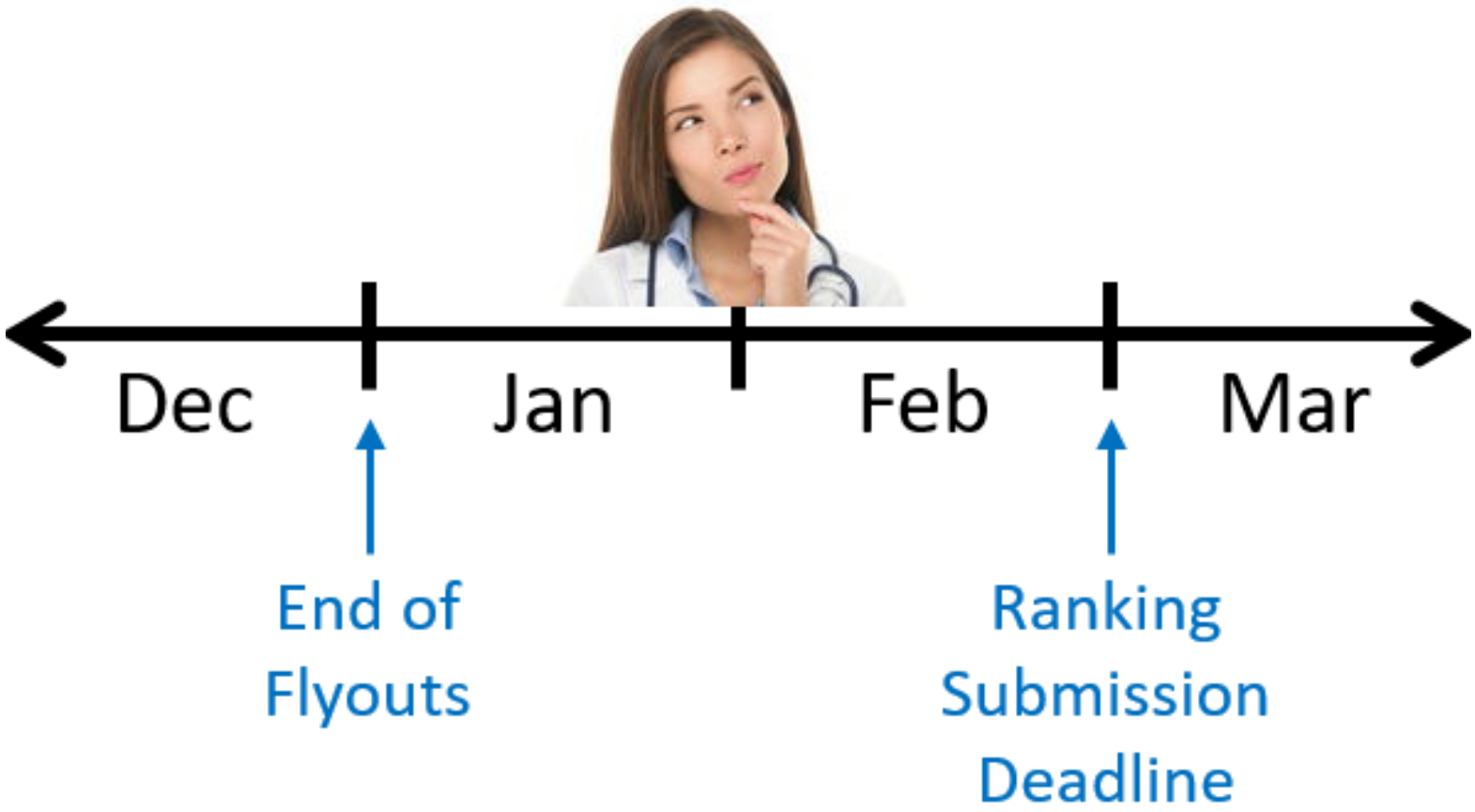
Jan

Feb

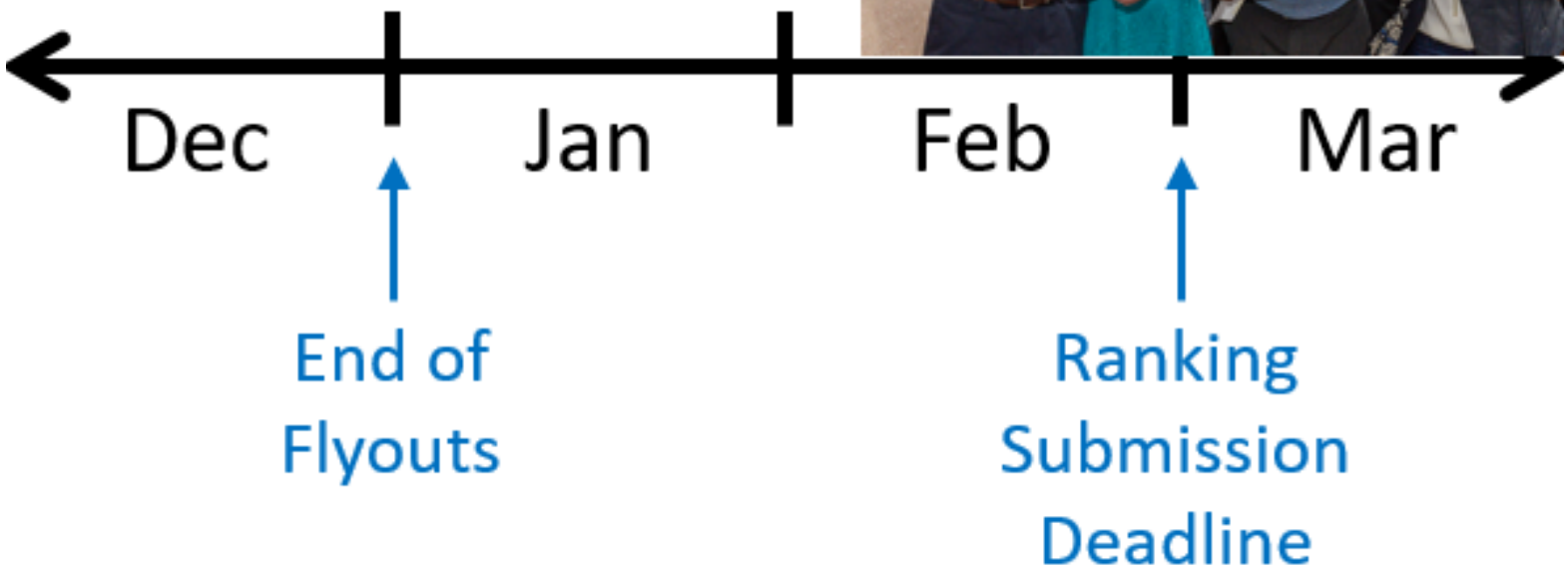
Mar

End of Flyouts

Institutional Context: NRMP Timeline



Institutional Context: NRMP Timeline



Convenient Features of NRMP

1. Due to incentive-compatible mechanism, observe preferences directly.
2. Identifiable choice set.
3. High-stakes choice to which students devote a lot of time and information.
4. Identifiable moment where decision is irreversibly committed.

Econometric Specification

- Define:
  - ER_j^i = Earnings Ranking in city j
  - COL_j^i = Cost of Living in city j
- Difference in attributes:
  - $ER_{1,2}^i = ER_1^i - ER_2^i$
  - $COL_{1,2}^i = COL_1^i - COL_2^i$

Econometric Specification

- Baseline Probit regression:

$$P(Prog_1 \succ_i Prog_2) = F(\beta^{ER} ER^i_{1,2} + \beta^{COL} COL^i_{1,2} + \theta Inc^i_{1,2})$$

- β^{ER}: preference for relative income.
  - Hypothesis: $\beta^{ER} \lesseqgtr 0$.
- β^{COL}: preference for absolute consumption.
  - Hypothesis: $\beta^{COL} < 0$.

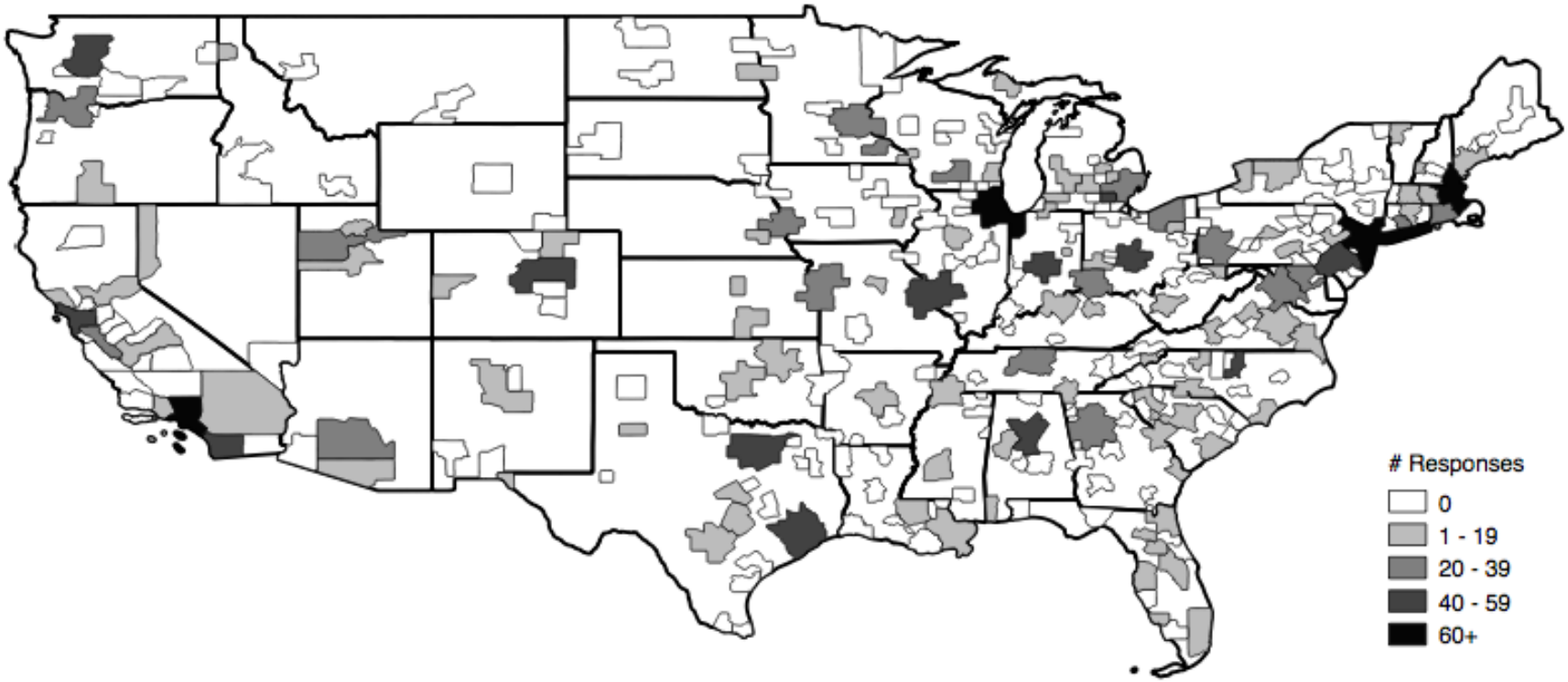
Survey Design

- Goal: collect data to estimate the Probit regression.
  - Preferences ($Prog_1 \succ_i Prog_2$).
  - Perceptions about relative income ($ER^i_{1,2}$) and cost of living ($COL^i_{1,2}$).
- Additionally: generate exogenous variation in $ER^i_{1,2}$, $COL^i_{1,2}$ through information-provision experiment.

Survey Design

1. Baseline Survey:
   - **Choice set**
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - Posterior Beliefs
   - Expected Ranking Choice
2. Follow-Up Survey:
   - Final Ranking Choice
   - Long-Term Beliefs

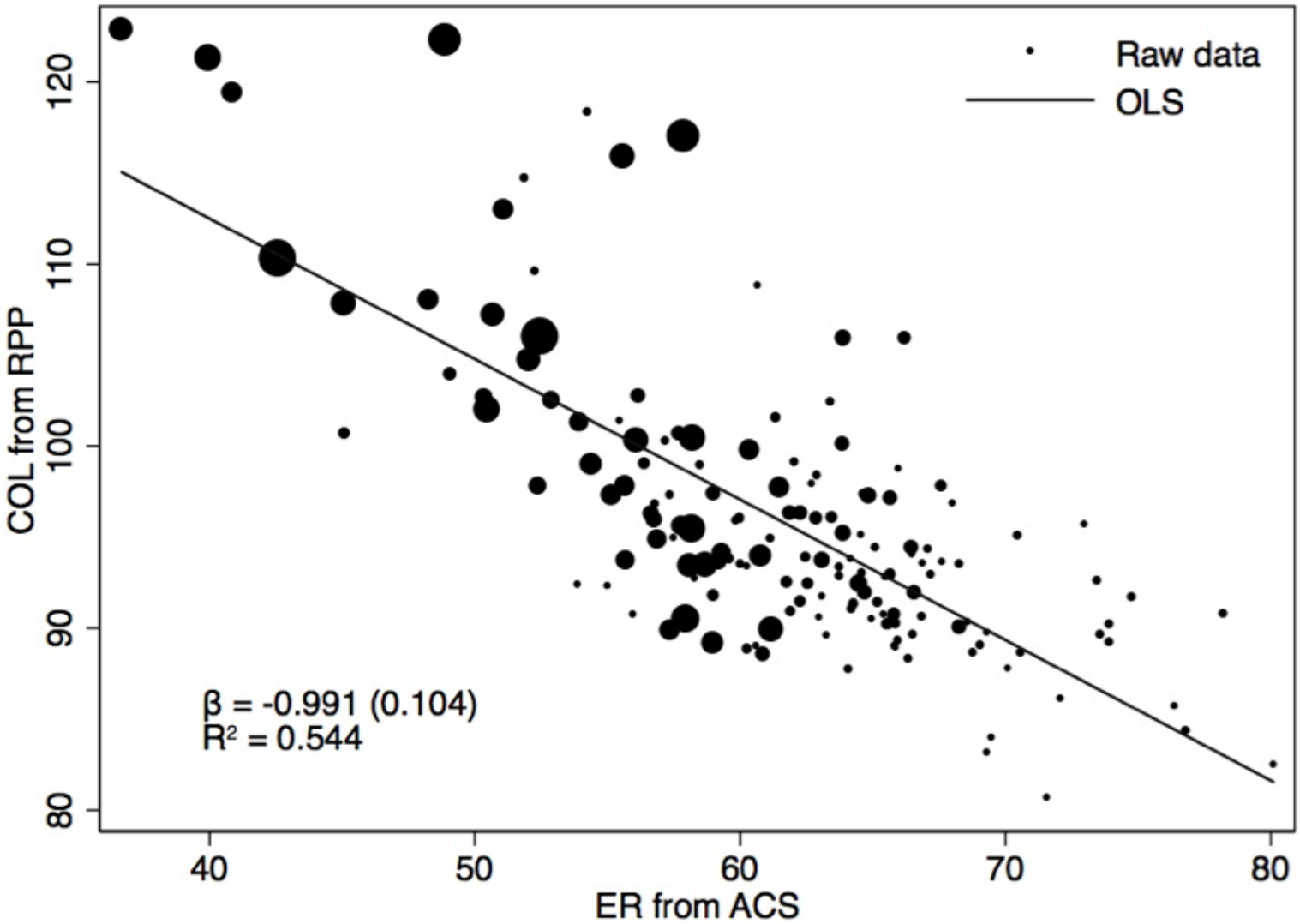
Geographic Distribution of Choices



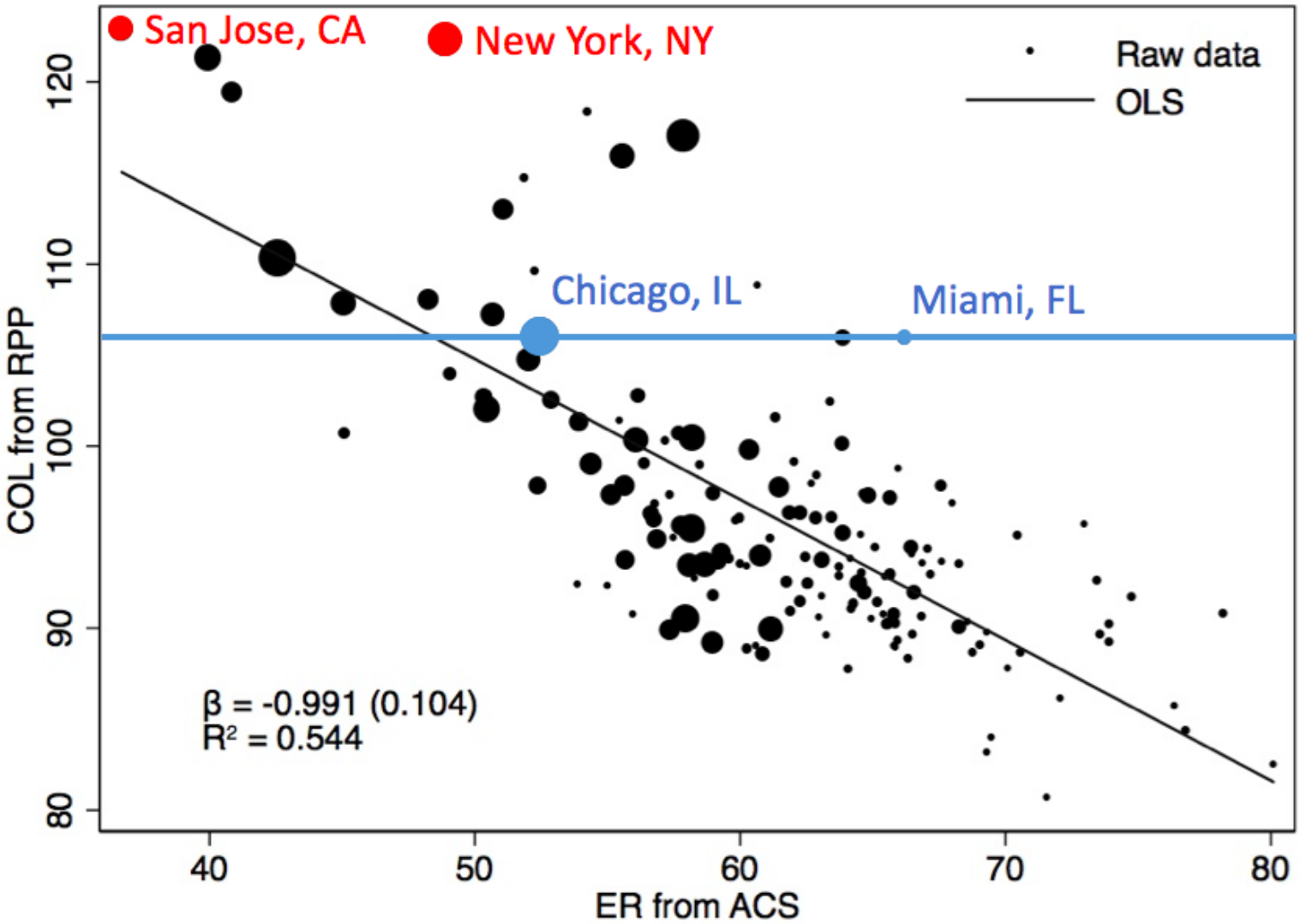
Survey Design

1. Baseline Survey:
   - Choice set
   - **Prior Beliefs (Earnings Rank & Cost of Living)**
   - Information-provision experiment
   - Posterior Beliefs
   - Expected Ranking Choice
2. Follow-Up Survey:
   - Final Ranking Choice
   - Long-Term Beliefs

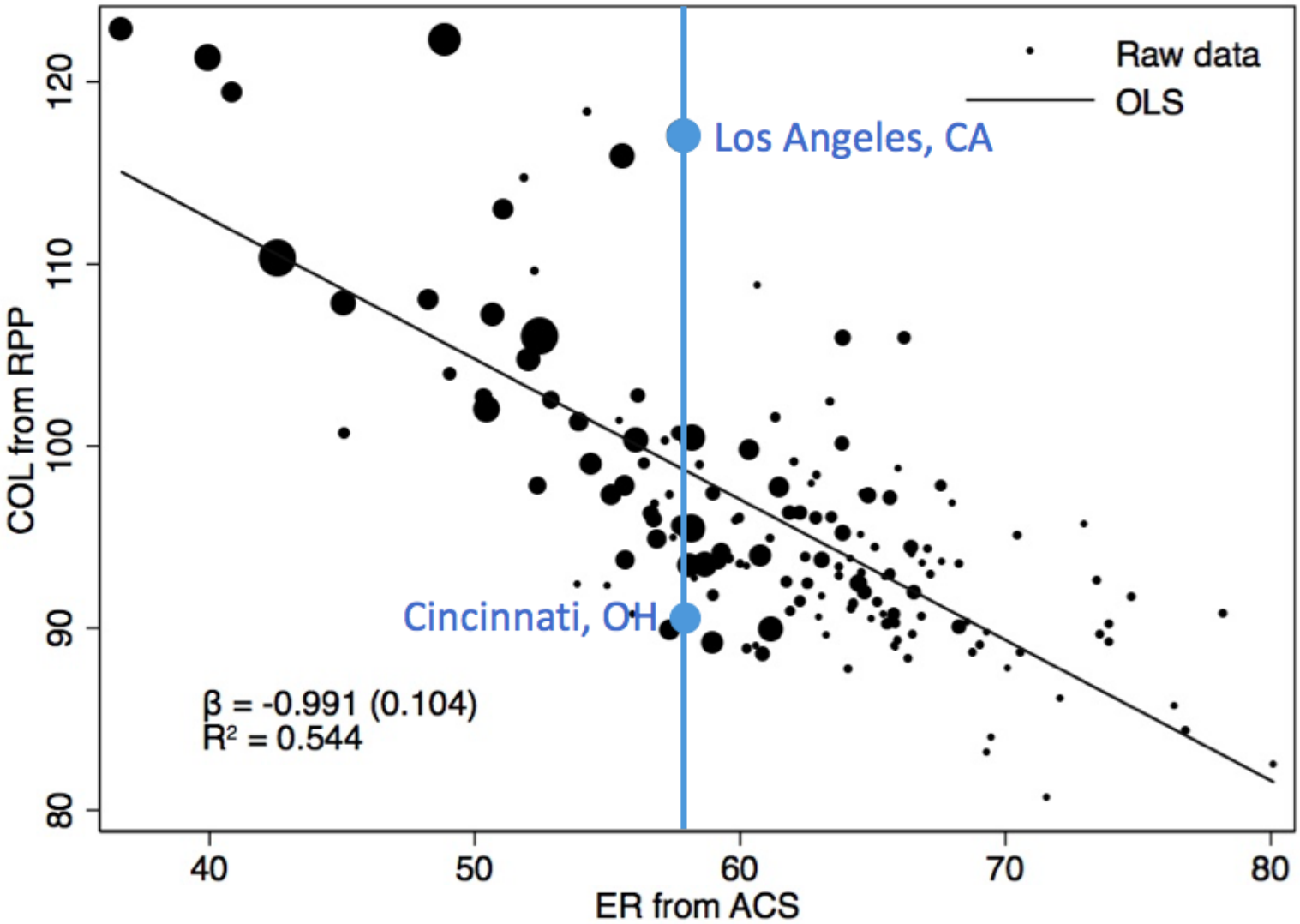
Substantial variation between ER and COL



Substantial variation between ER and COL



Substantial variation between ER and COL



Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - **Information-provision experiment**
   - Posterior Beliefs
   - Expected Ranking Choice
2. Follow-Up Survey:
   - Final Ranking Choice
   - Long-Term Beliefs

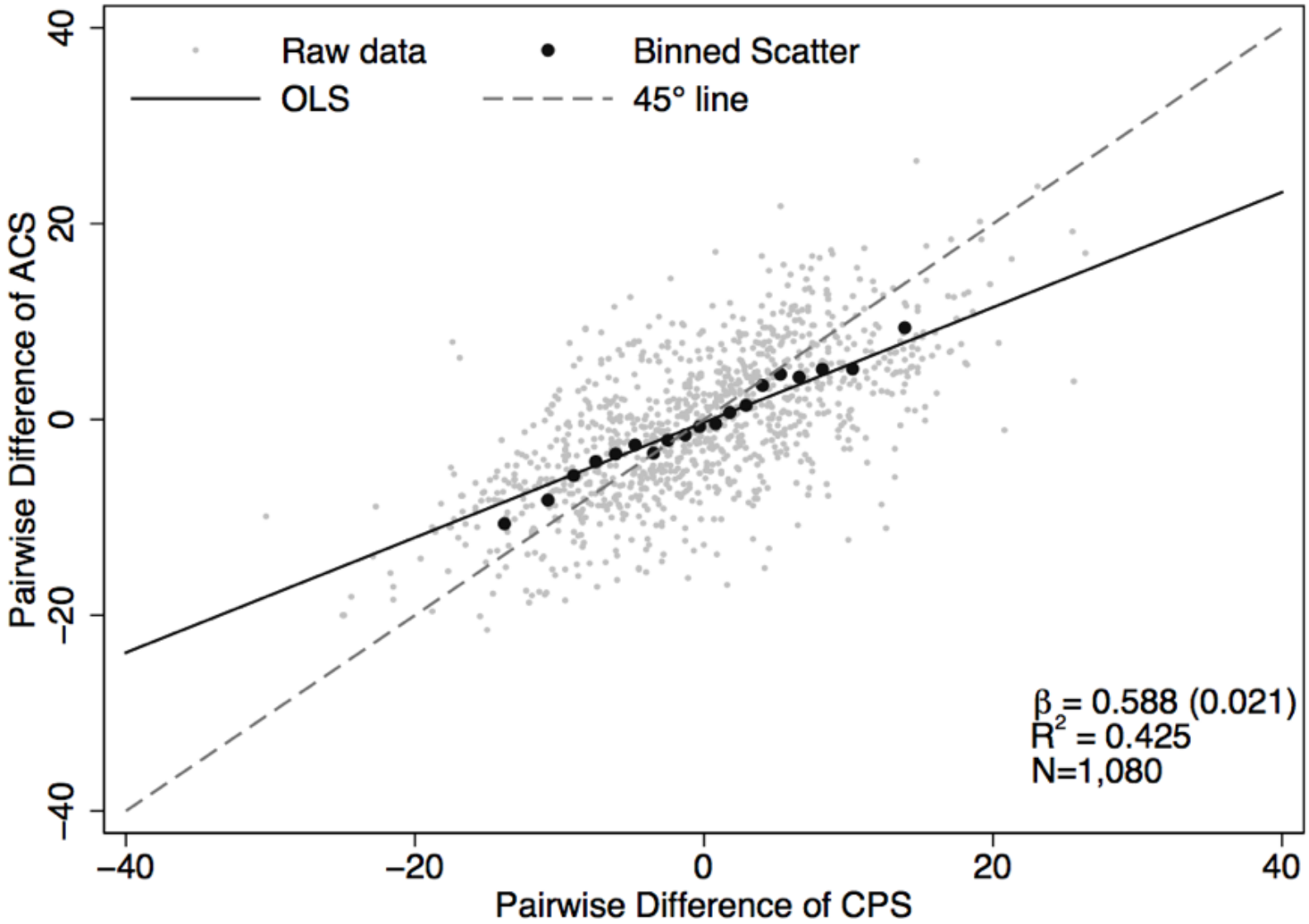
Survey Design: Information Experiment

- Goal: generate exogenous variation in perceptions in a non-deceptive way.
- All respondents shown feedback for ER and COL in both cities.
- By a flip of a coin, use different data source to generate feedback:
  - Earnings Ranking: ACS vs. CPS.
  - Cost of Living: RPP vs. COLI.
- Source-randomization generates exogenous variation in feedback.

Survey Design: Information Experiment

- Example: $54,000 earnings in Champaign-Urbana, IL.
- Randomly assigned to one of two messages:
  - In this city, you would be richer than **55.1%** of income earners (according to data from the Current Population Survey).
  - In this city, you would be richer than **60.3%** of income earners (according to data from the American Community Survey).

Example: Variation in ER by Source



Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - **Posterior Beliefs**
   - Expected Ranking Choice
2. Follow-Up Survey:
   - Final Ranking Choice
   - Long-Term Beliefs

Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - Posterior Beliefs
   - **Expected Ranking Choice** ($Prog_1 \succ_i Prog_2$)
2. Follow-Up Survey:
   - Final Ranking Choice
   - Long-Term Beliefs

Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - Posterior Beliefs
   - Expected Ranking Choice
2. **Follow-Up Survey:**
   - Final Ranking Choice
   - Long-Term Beliefs

Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - Posterior Beliefs
   - Expected Ranking Choice
2. Follow-Up Survey:
   - **Final Ranking Choice**
   - Long-Term Beliefs

Survey Design

1. Baseline Survey:
   - Choice set
   - Prior Beliefs (Earnings Rank & Cost of Living)
   - Information-provision experiment
   - Posterior Beliefs
   - Expected Ranking Choice
2. Follow-Up Survey:
   - Final Ranking Choice
   - **Long-Term Beliefs**

Outline

Survey Deployment

- Reached out to all accredited U.S. medical schools.
- Schools forwarded email invitation to graduating students participating in NRMP.
- $10 Amazon gift card incentive for baseline, $5 for follow-up.
- Verified the identity of participants with schools using email addresses.

Survey Implementation

- 27 out of 135 accredited U.S. medical schools accepted to participate.
  - Sample is representative in observable program characteristics.

Survey Implementation

- 27 out of 135 accredited U.S. medical schools accepted to participate.
  - Sample is representative in observable program characteristics.
- 1,087 (29.57%) of 3,600 eligible students completed the baseline survey.
  - Observable characteristics balanced across treatment groups.

Survey Implementation

- 27 out of 135 accredited U.S. medical schools accepted to participate.
  - Sample is representative in observable program characteristics.
- 1,087 (29.57%) of 3,600 eligible students completed the baseline survey.
  - Observable characteristics balanced across treatment groups.
- 985 (90.6%) completed the follow-up survey.
  - No evidence of selective attrition.

Outline

Results

- Baseline Specification:
  - Outcome Variable: Expected Choice ($Prog_1 \succ_i Prog_2$) from Baseline Survey.
  - All (experimental and non-experimental) variation in posterior beliefs ($ER^i_{1,2}$ and $COL^i_{1,2}$).
- For later:
  - Sensitivity to control variables.
  - Focus on experimental variation.
  - Effects on final submission.
  - Auxiliary sample.

Baseline Results

	Probit (1)	MFX
β^{ER}	0.995*	0.186*
	(0.539)	(0.100)
β^{COL}	-1.073**	-0.201**
	(0.485)	(0.090)
Observations	1,080	

Baseline Results

	Probit (1)	MFX
β^{ER}	0.995*	0.186*
	(0.539)	(0.100)
β^{COL}	-1.073**	-0.201**
	(0.485)	(0.090)
Observations	1,080	

- Prestige and career prospects 2x-3x as important.

Heterogeneity by Relationship Status

- Related findings point to differences in relationship status:
  - Effects of relative income in Luttmer (2005) entirely driven by individuals who are married/cohabiting.
  - Evidence on different location preferences by relationship status in urban literature (e.g., Couture and Handbury 2015, Gautier et al. 2010)
- We elicited relationship status a-la-Luttmer:
  - Single (35% of respondents)
  - Married (24%)
  - Long-term relationship (41%)

Heterogeneity by Relationship Status

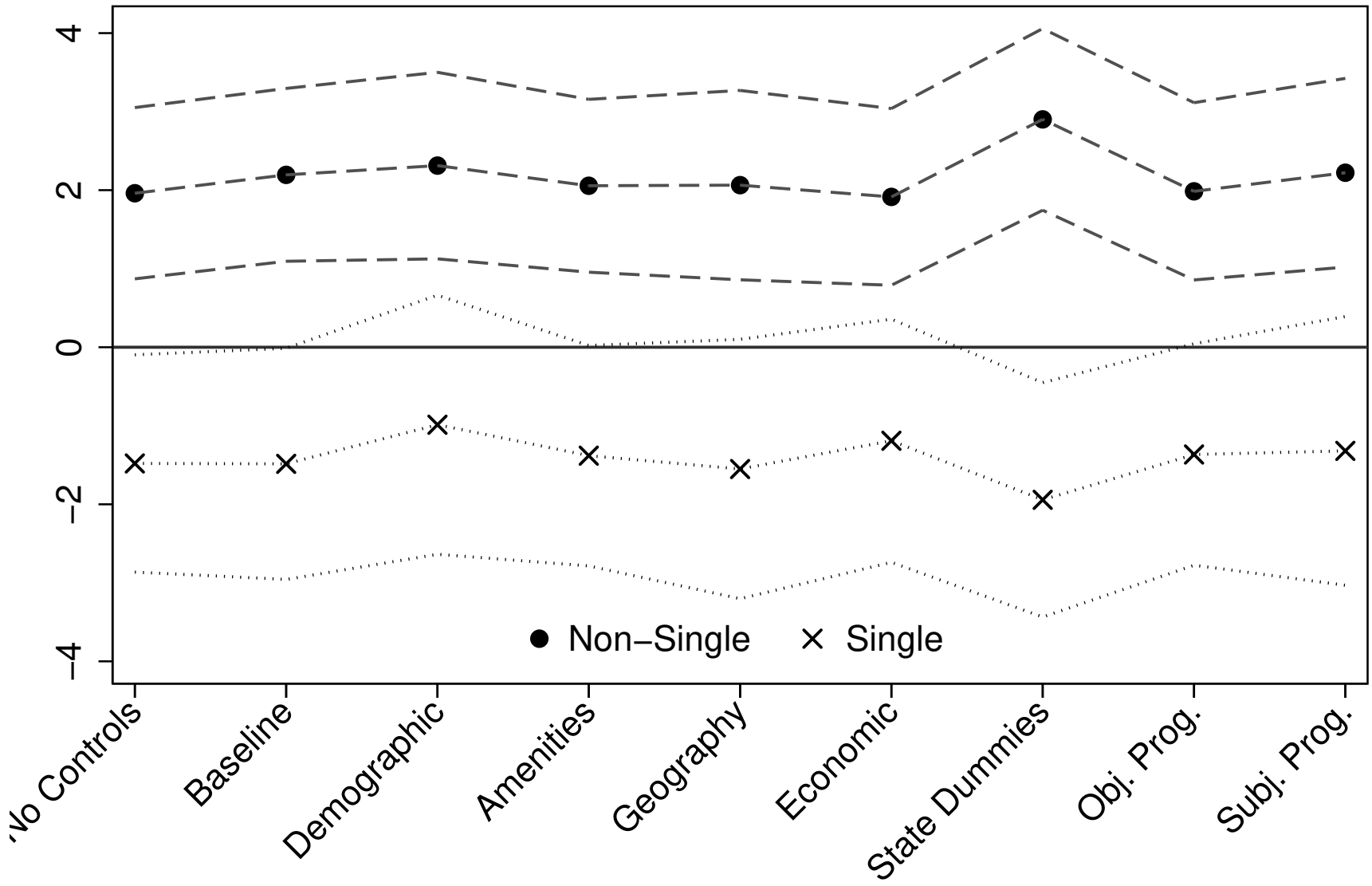
	All (1)	Non-Single (2)	Single (3)
β^{ER}	0.995*	2.236***	-1.538*
	(0.539)	(0.669)	(0.880)
β^{COL}	-1.073**	-1.087	-1.058
	(0.485)	(0.663)	(0.749)
Diff. P-value:			
ER		0.001 *[0.030]*	
COL		0.977 *[0.977]*	
Observations	1,080	698	382

Note: Multiple-testing q-values based on Benjamini and Yekutieli (2001) in brackets.

Heterogeneity by Relationship Status

- Potential explanations for singles:
  - Live in more affluent cities to find richer partners (Fisman et al. 2006, Hitsch et al. 2010).
  - Single individuals may care about other positive spillovers from the rich.
- No other sources of heterogeneity seem to be important.

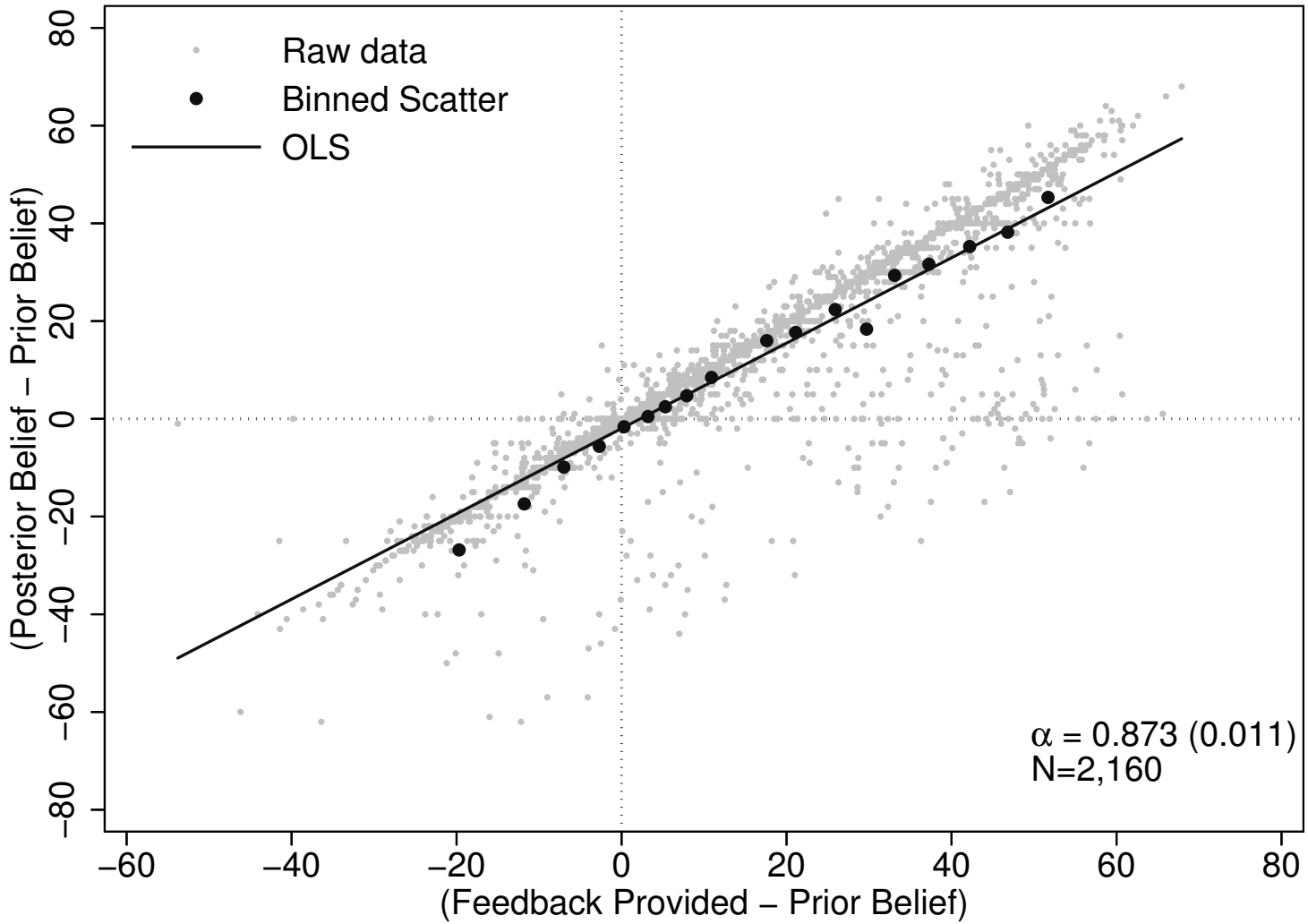
Sensitivity to Controls: β^{ER}



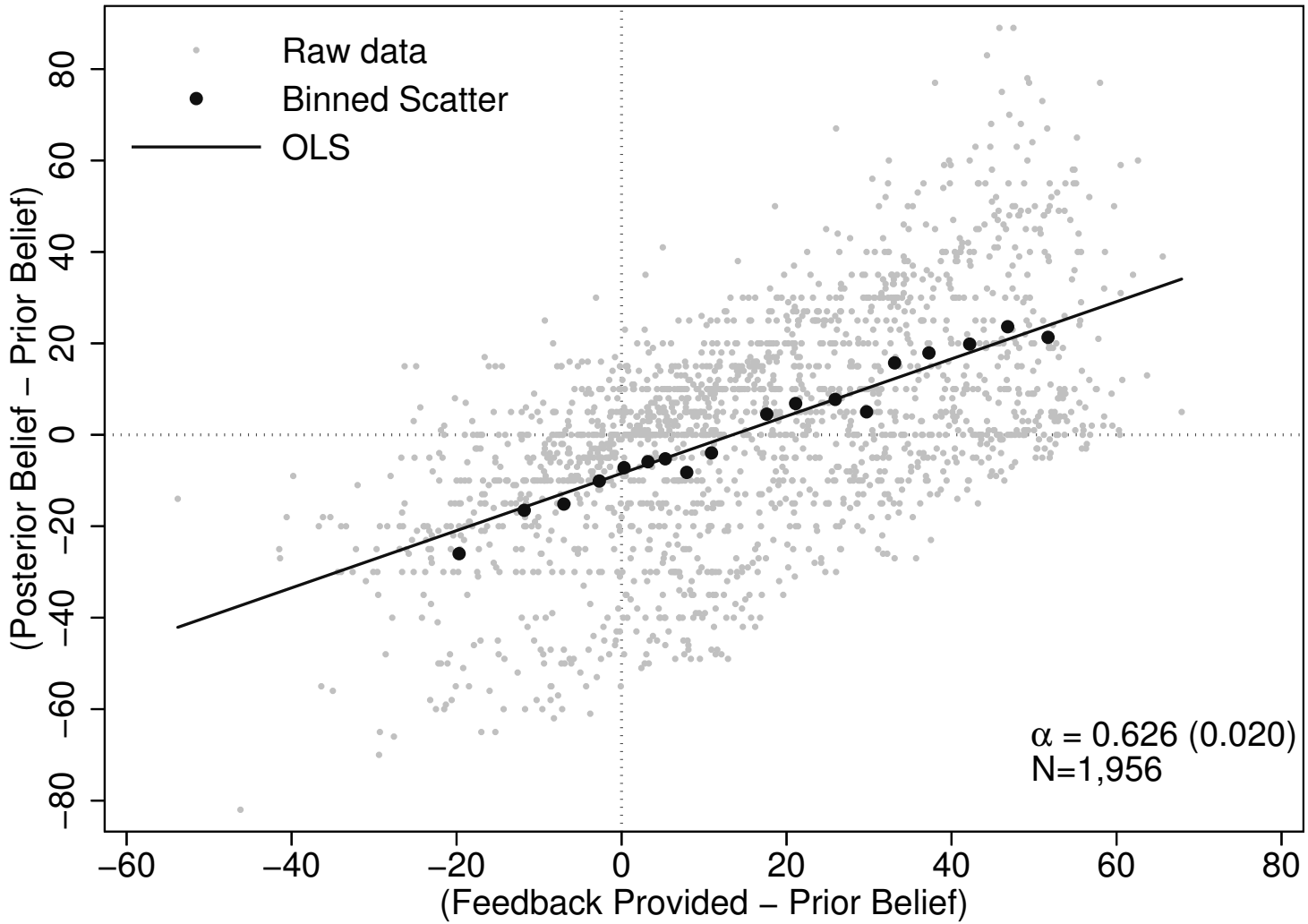
Robustness: Experimental Estimates

- Exploit exogenous variation from source-randomization.
- Define "random" part of the information shown:
  - $\Delta ER^{i}_{1,2} = ER^{i,shown}_{1,2} - ER^{i,alt}_{1,2}$
  - $\Delta COL^{i}_{1,2} = COL^{i,shown}_{1,2} - COL^{i,alt}_{1,2}$
- Estimate IV-Probit model: $\Delta ER^{i}_{1,2}$ and $\Delta COL^{i}_{1,2}$ as instruments for $ER^{i}_{1,2}$ and $COL^{i}_{1,2}$.

First Stage Results: ER



Persistence of Learning



Robustness Checks: β^{ER}

	Non-Single (1)	Single (2)	All (3)
Baseline	2.380***	-1.656*	1.141**
	(0.702)	(0.991)	(0.577)
Experimental			
Experimental, Long-Term			

Robustness Checks: β^{ER}

	Non-Single (1)	Single (2)	All (3)
Baseline	2.380***	-1.656*	1.141**
	(0.702)	(0.991)	(0.577)
Experimental	2.977**	-4.964**	0.867
	(1.331)	(1.974)	(1.151)
Experimental, Long-Term			

Robustness Checks: β^{ER}

	Non-Single (1)	Single (2)	All (3)
Baseline	2.380***	-1.656*	1.141**
	(0.702)	(0.991)	(0.577)
Experimental	2.977**	-4.964**	0.867
	(1.331)	(1.974)	(1.151)
Experimental, Long-Term	1.993*	-5.285***	-0.029
	(1.188)	(1.984)	(1.071)

Robustness Check: Auxiliary Experiment

- Recruited over 1,200 U.S. respondents on Amazon Mechanical Turk.
- Repeated the resident survey with two cities that they would consider moving to.
- Disadvantages:
  - Intended choice instead of actual choice.
  - Low information, low attention environment.
- Advantages:
  - Can run new experiments on demand.
  - A more diverse subject pool.

Subject Pools: Main vs Auxiliary

	Main Experiment *Med. Students*	Auxiliary Experiment *Online Sample*	Difference
Age	27.091	37.476	-10.385***
	(2.725)	(11.980)	(0.350)
% Male	0.481	0.391	0.090***
	(0.500)	(0.488)	(0.021)
% Married	0.240	0.461	-0.221***
	(0.427)	(0.499)	(0.019)
% Has children	0.089	0.527	-0.438***
	(0.285)	(0.499)	(0.017)
Observations	1,080	1,245	

Results: Auxiliary Experiment

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline	1.408***	1.095**	1.293***	-2.203***	-1.618***	-1.962***
	(0.376)	(0.478)	(0.292)	(0.463)	(0.566)	(0.364)
Experimental	2.578**	0.664	1.706**	-2.385***	-2.956***	-2.528***
	(1.019)	(1.272)	(0.816)	(0.666)	(0.917)	(0.531)

Results: Auxiliary Experiment

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline	1.408***	1.095**	1.293***	-2.203***	-1.618***	-1.962***
	(0.376)	(0.478)	(0.292)	(0.463)	(0.566)	(0.364)
Experimental	2.578**	0.664	1.706**	-2.385***	-2.956***	-2.528***
	(1.019)	(1.272)	(0.816)	(0.666)	(0.917)	(0.531)

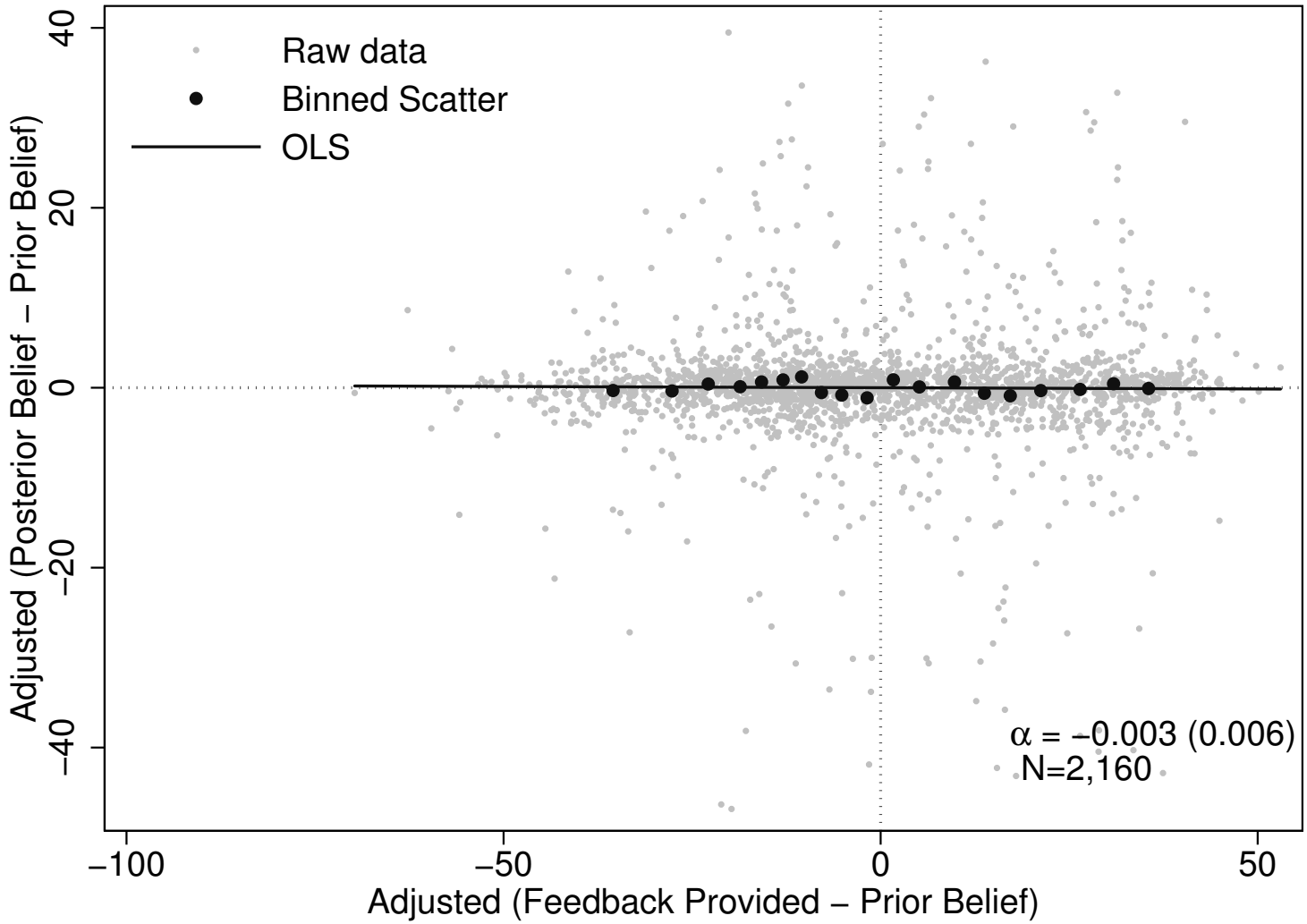
- MRS: 0.90 (s.e. 0.64) in main vs. 0.66 (0.20) in auxiliary.

Outline

Confounding Factors

- Could individuals be using ER to make inferences about COL?
  - ER feedback does not affect beliefs of COL.

COL unaffected by ER feedback



Confounding Factors

- Could individuals be using ER to make inferences about other city attributes besides COL? (e.g., public goods).
  - Could explain preferences for singles.
  - But would go against preferences for non-singles.
  - We address this directly in auxiliary experiment.

Outline

Conclusions

- Evidence that individuals care about prospective relative income when making location choices.
- Key insight: we can infer relative income concerns from location choices.
  - Break the "dependency" on happiness data.

Conclusions

- Broader methodological contribution: measure preferences over amenities.
- Literature is largely structural (e.g., Albouy, 2016, Moretti, 2013, Diamond, 2016).
  - Although a few cases of quasi-experimental identification (e.g., Black, 1999; Bayer, Ferrerira and McMillan, 2007).
- Natural experiments are hard to come by.
  - Our methodology creates the exogenous variation on demand, through information-provision experiments.

Thank you!

Survey Design: Information Experiment

Now, we want to share some information with you, related to the characteristics of the two cities that you are considering living in. Please take a moment to review the information carefully.

***Note**: this information is only shown once and you will not be able to come back to it.*

First, find below some estimates of the cost of living:

The **Champaign-Urbana, IL** metro area is **6.6% cheaper** than the U.S. average.

The **Los Angeles-Long Beach-Anaheim, CA** metro area is **17.0% more expensive** than the U.S. average.

Source: based on most recent data from the Bureau of Economic Analysis.

Survey Design: Information Experiment

Second, find below some estimates of the earnings ranking:

With your individual annual earnings of **$ 54000**, you would be richer than **60.3%** of **Champaign-Urbana, IL**'s population.

With your individual annual earnings of **$ 54000**, you would be richer than **57.9%** of **Los Angeles-Long Beach-Anaheim, CA**'s population.

Source: based on most recent data from the American Community Survey.

Auxiliary Experiment

- Replicate medical student survey experiment:
  - Re-framed as a hypothetical choice
  - At very end: added questions measuring perceptions regarding other city attributes (e.g., crime, quality of public services, etc.)
  - Recruited over 1,200 U.S. respondents on Amazon Mechanical Turk
- Important caveats:
  - Hypothetical choice, not moving any time soon.
  - Low information, low stakes.
  - Very different in observable (and likely unobservable) characteristics

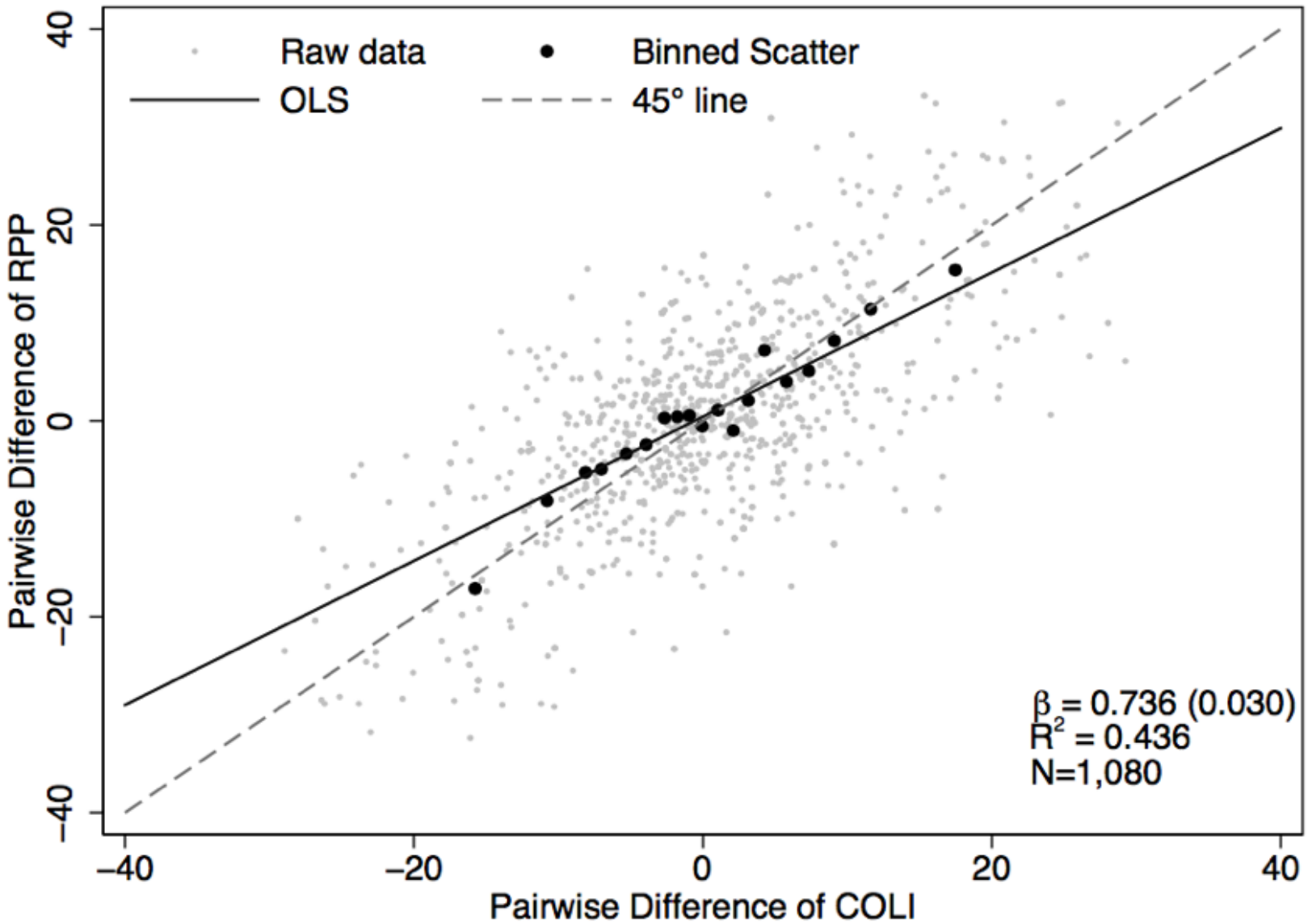
Samples very different in observable characteristics

	Main Experiment *Med. Students*	Auxiliary Experiment *Online Sample*	Difference
Age	27.091	37.476	-10.385***
	(2.725)	(11.980)	(0.350)
% Male	0.481	0.391	0.090***
	(0.500)	(0.488)	(0.021)
% Married	0.240	0.461	-0.221***
	(0.427)	(0.499)	(0.019)
% Has children	0.089	0.527	-0.438***
	(0.285)	(0.499)	(0.017)
Observations	1,080	1,245	

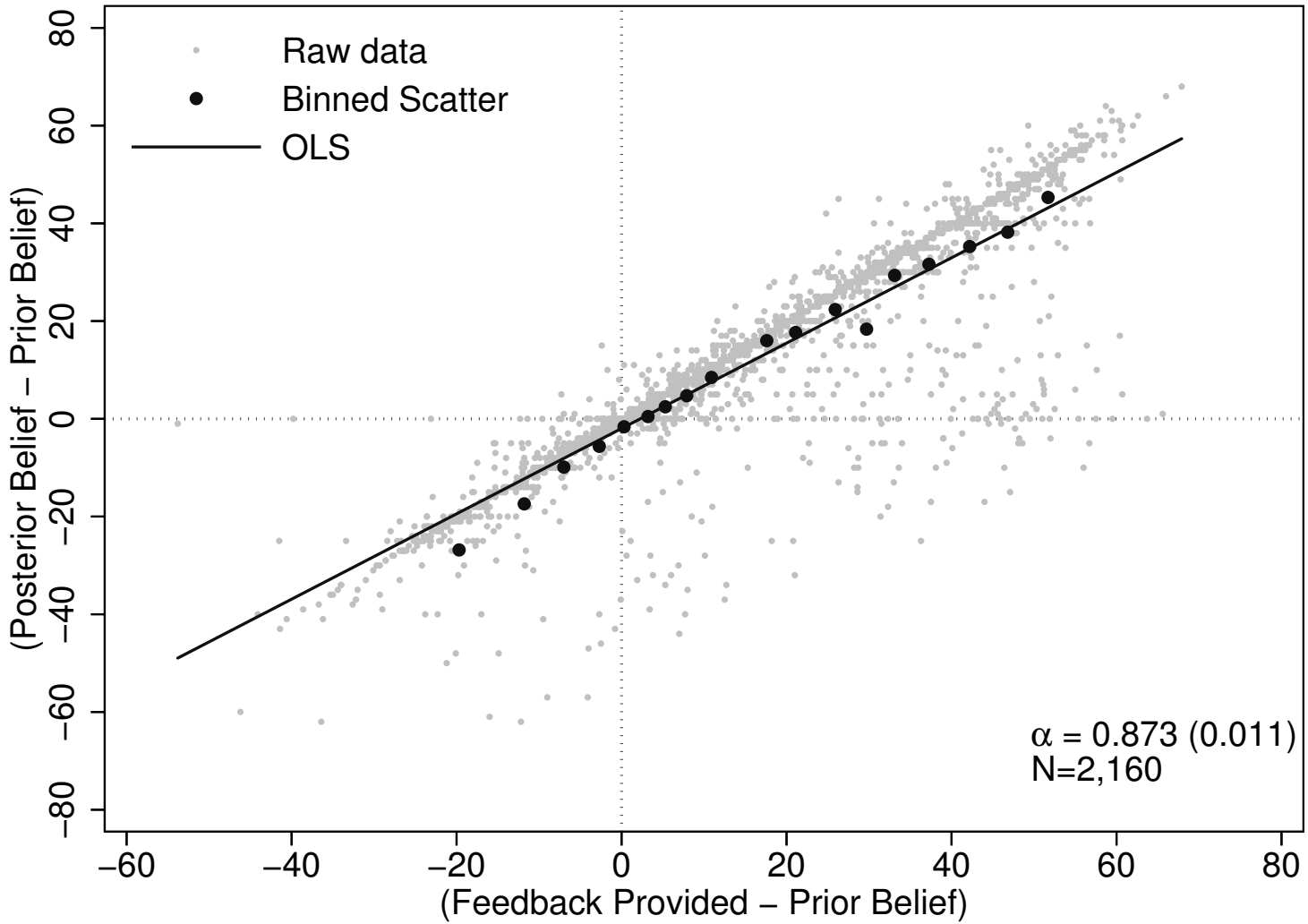
Results: Auxiliary Experiment

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

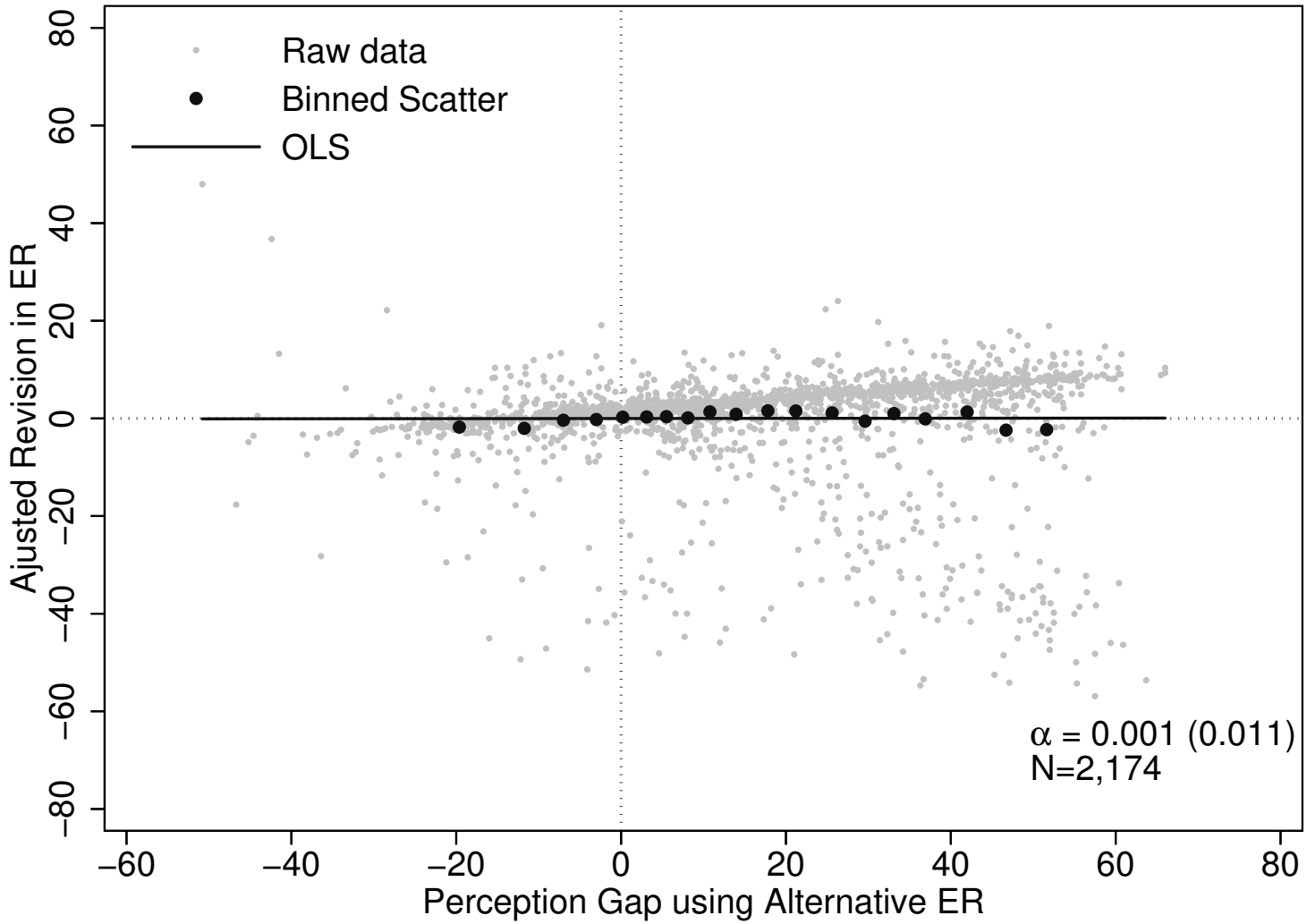
Substantial Variation in COL by Source



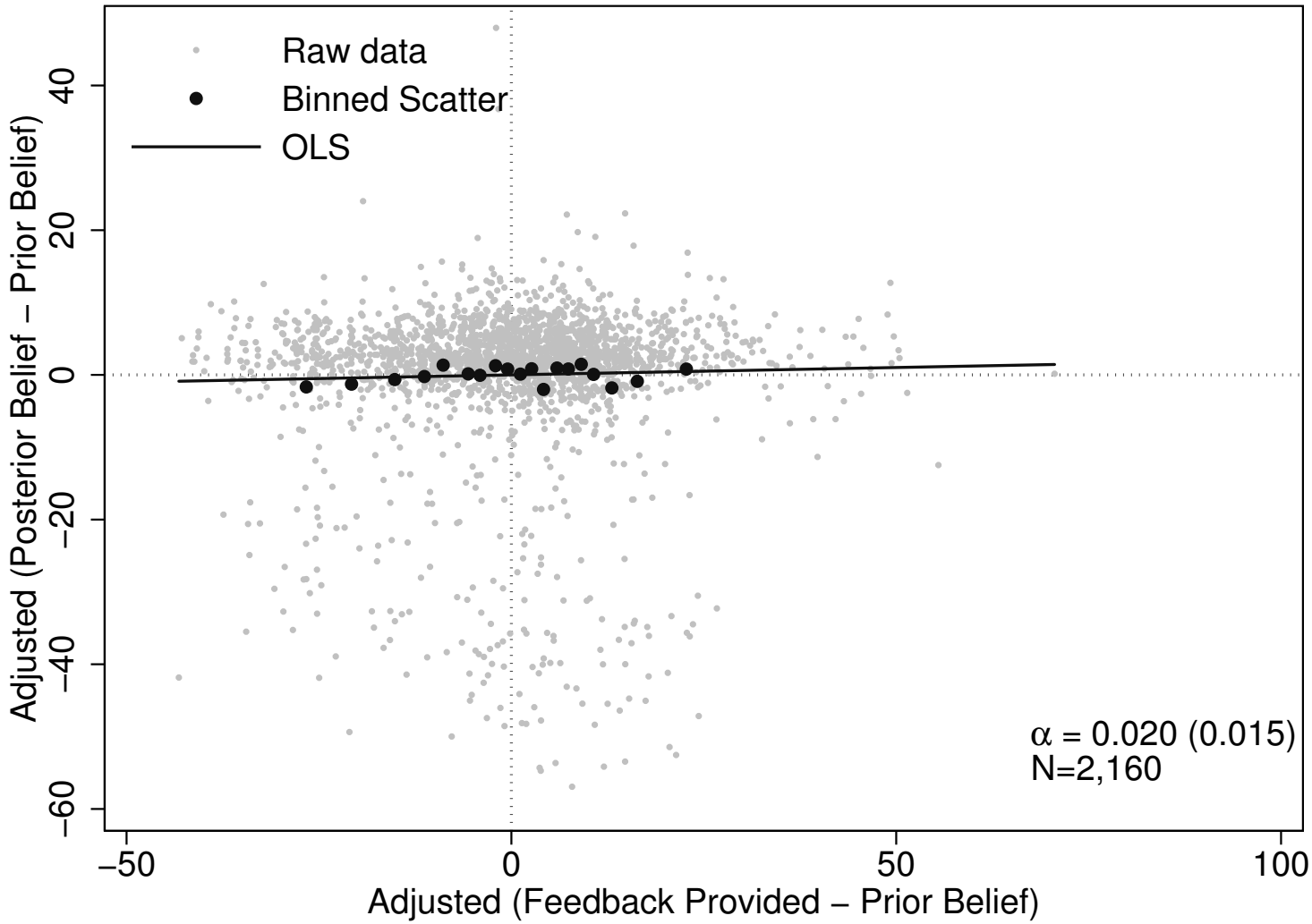
ER^{post}: Strong update in beliefs



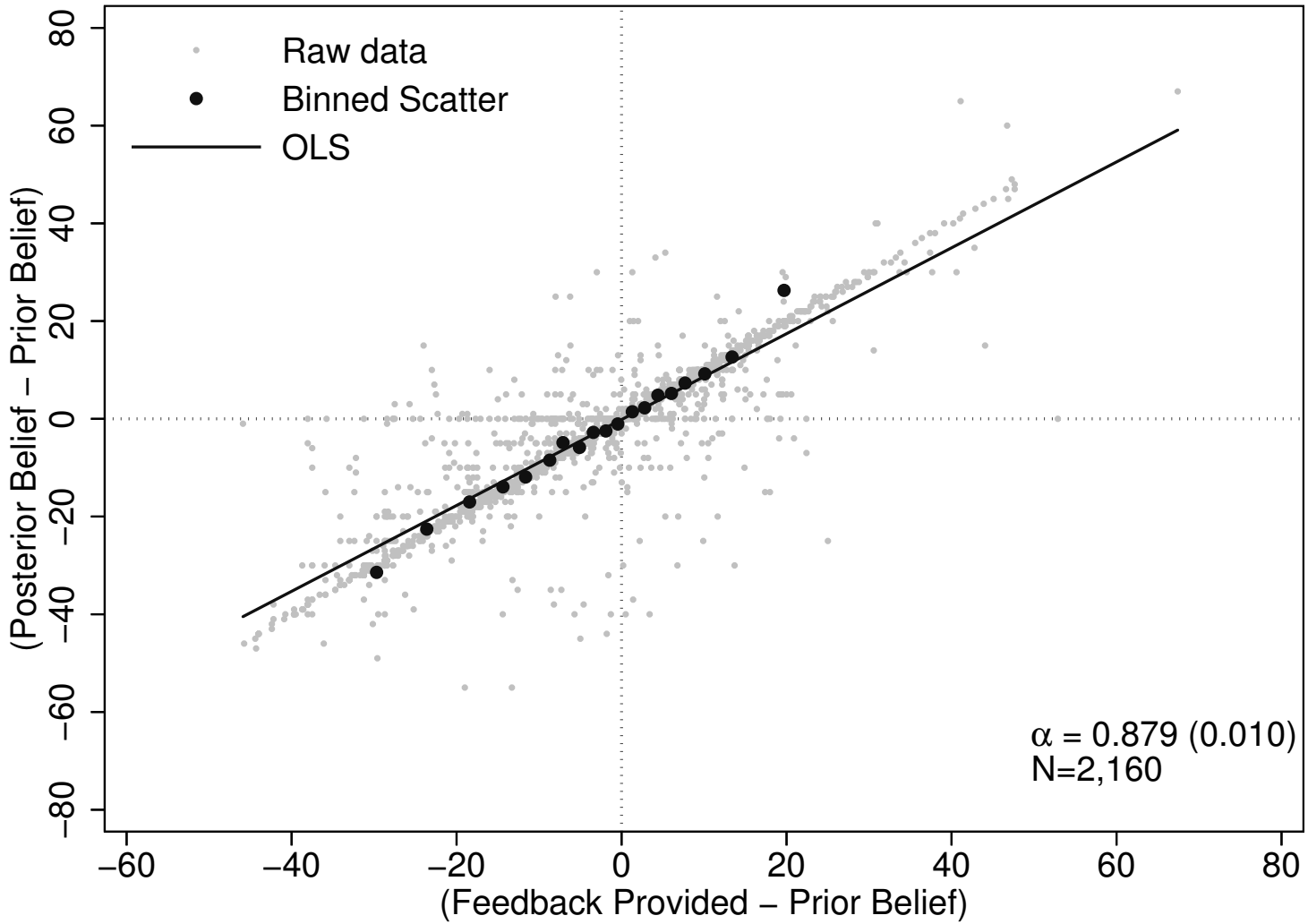
ER^{post} not affected by alternative source



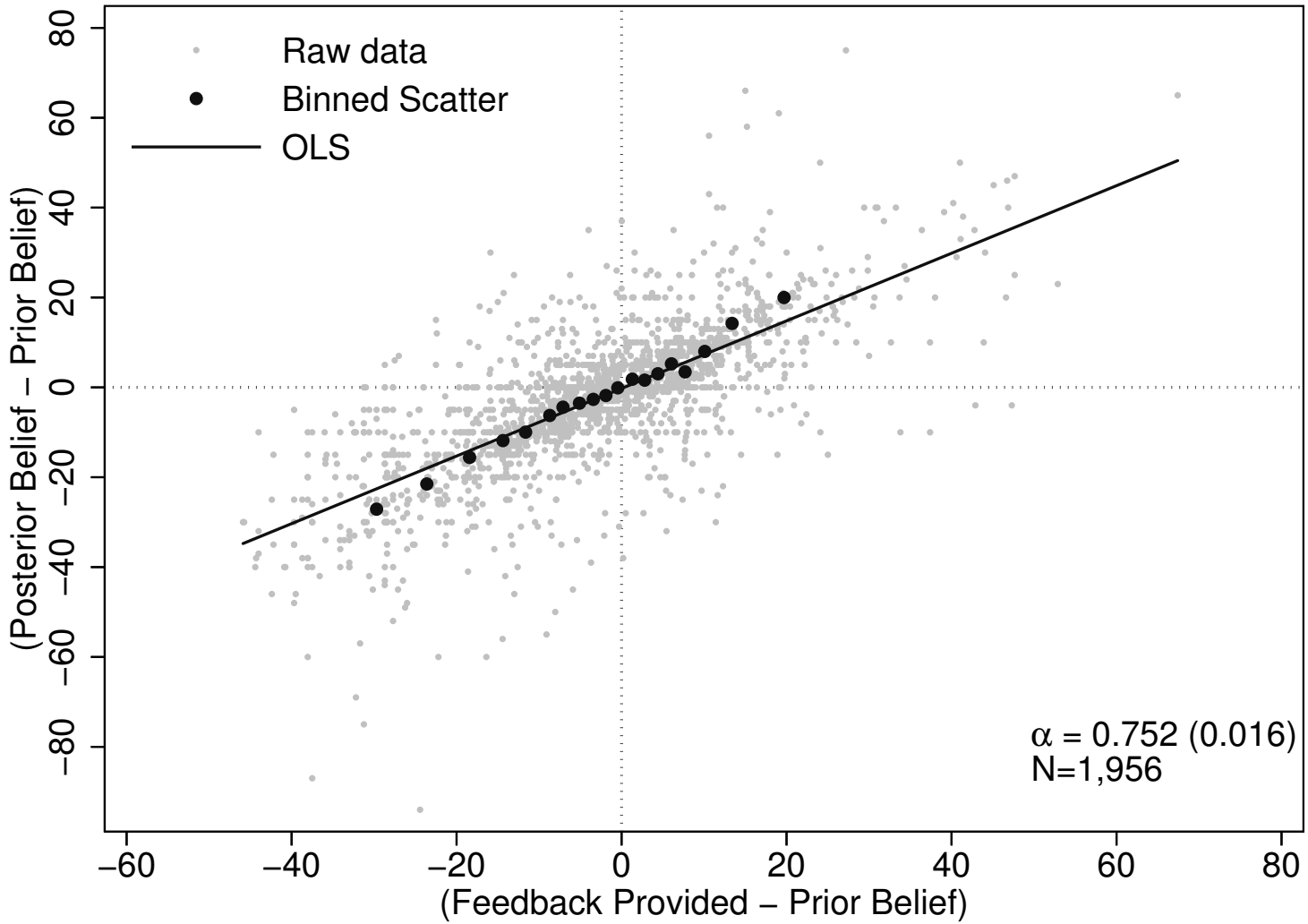
ER unaffected by COL feedback



Learning about COL



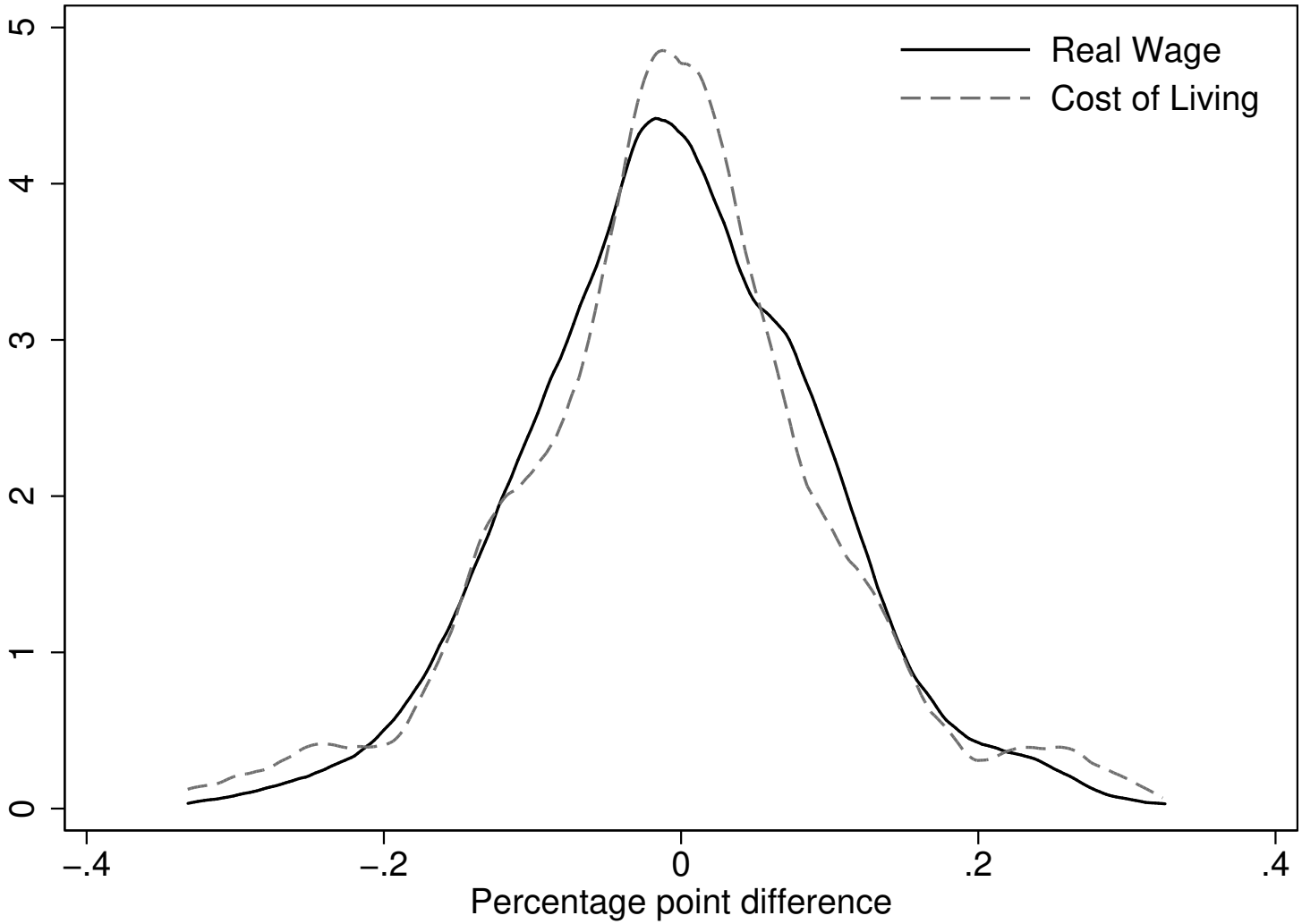
Learning about COL: Persistence



Response Rates

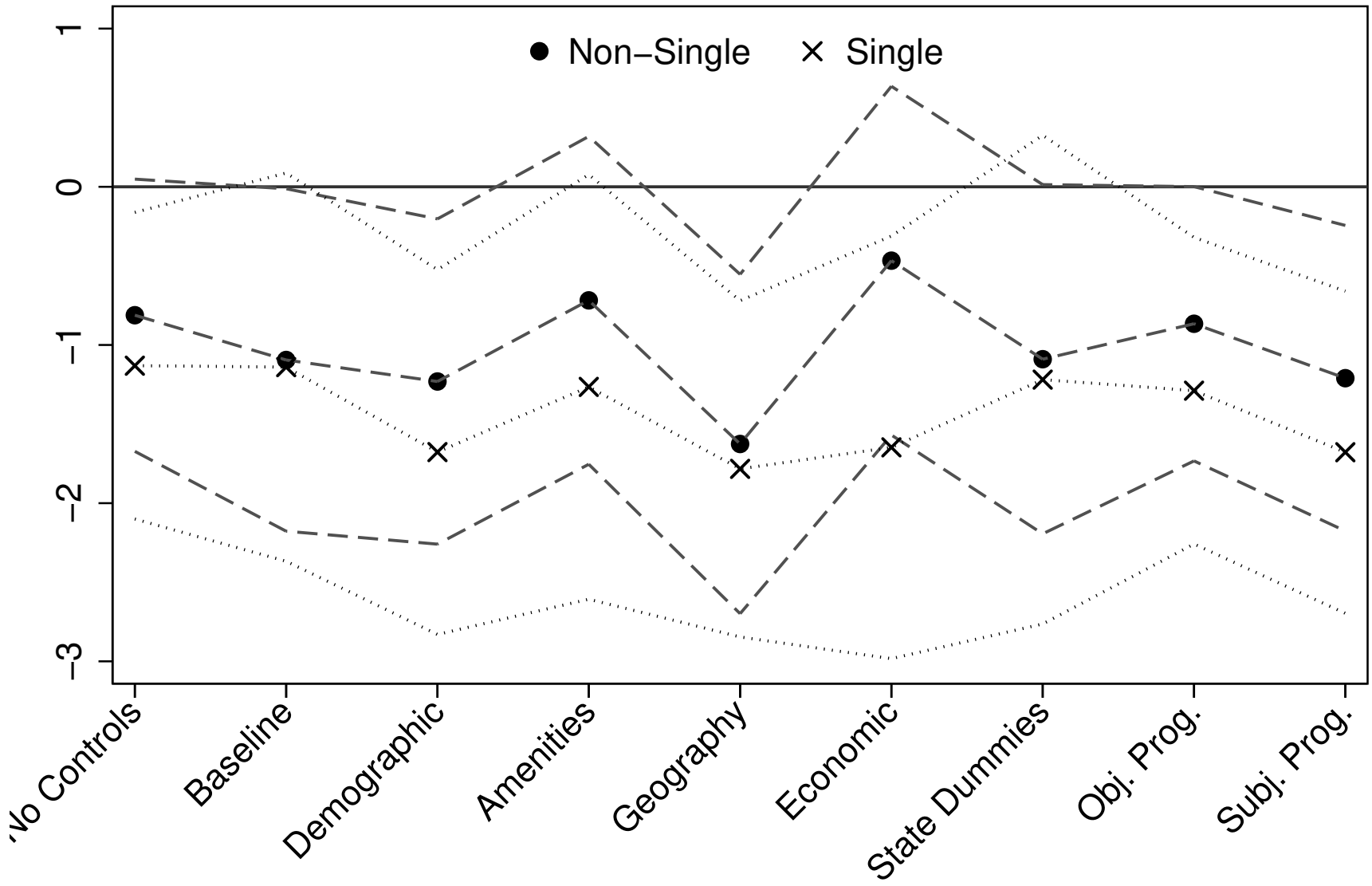
State	University	Est. Senior Cohort	Nr Finished Surveys	Est. Response Rate (%)
Alabama	University of Alabama	174	47	27.0
Alabama	University of South Alabama	73	21	28.8
Arizona	University of Arizona	72	18	25.0
California	UC San Diego	124	39	31.5
Connecticut	Yale University	121	25	20.7
Florida	University of Florida	135	52	38.5
Illinois	Loyola University	145	67	46.2
Illinois	University of Illinois	20	8	40.0
Indiana	Indiana University	345	89	25.8
Massachusetts	Tufts University	194	43	22.2
Michigan	Michigan State University	183	76	41.5
Missouri	Saint Louis University	165	70	42.4
Missouri	University of Missouri (Kansas City)	101	34	33.7
Nebraska	University of Nebraska	125	46	36.8
New Mexico	University of New Mexico	97	27	27.8
New York	Stony Brook University	126	17	13.5
New York	University of Rochester	103	38	36.9
Ohio	Ohio State University	172	61	35.5
Oklahoma	University of Oklahoma	147	47	32.0
Pennsylvania	Pennsylvania State University	139	4	2.9
Rhode Island	Brown University	126	34	27.0
South Carolina	University of South Carolina	90	21	23.3
Texas	Baylor	180	44	24.4
Texas	Paul L. Foster School of Medicine (TTU)	89	30	33.7
Vermont	University of Vermont	105	39	37.1
Virginia	Virginia Commonwealth University	215	67	31.2
West Virginia	West Virginia University	110	23	20.9
	Total	3,676	1,087	29.57

Variation in COL

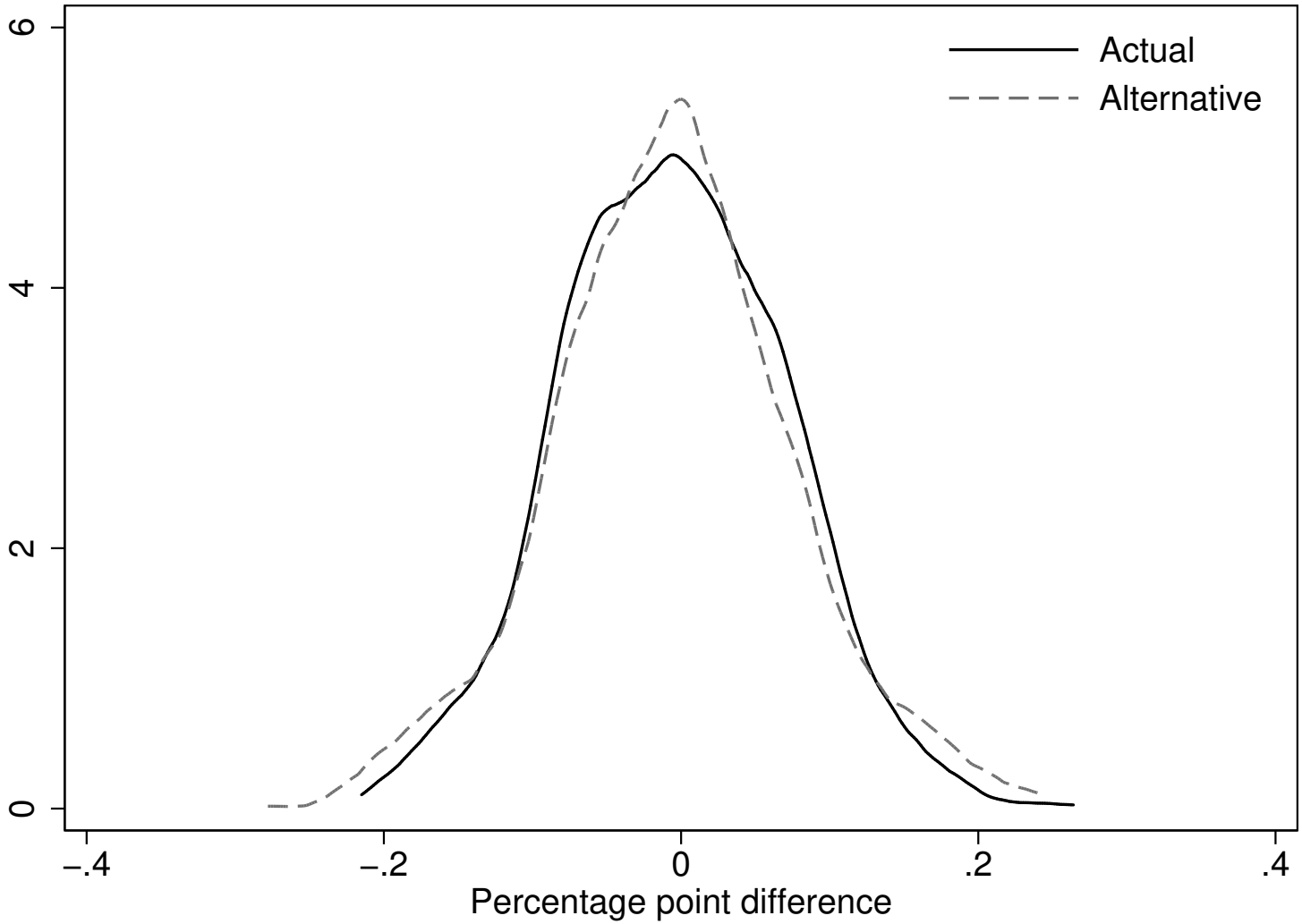


Back

Sensitivity Analysis: β^{COL}



Variation in ER



Back

First Stage

Back

	All (1)	Non-Single (2)	Single (3)
Panel A: First Stage			
Dep. Var.: $ER^i_{1,2}$			
$\Delta ER^i_{1,2}$	0.796***	0.854***	0.687***
	(0.045)	(0.055)	(0.081)
$\Delta COL^i_{2,1}$	-0.013	-0.021	-0.007
	(0.039)	(0.049)	(0.064)
Dep. Var.: $COL^i_{2,1}$			
$\Delta ER^i_{1,2}$	0.058	0.101***	-0.036
	(0.037)	(0.036)	(0.087)
$\Delta COL^i_{2,1}$	0.928***	0.893***	0.985***
	(0.048)	(0.064)	(0.070)

Robustness Checks: β^{COL}

	Non-Single (4)	Single (5)	All (6)
Baseline	-1.234*	-1.379*	-1.262**
	(0.743)	(0.772)	(0.531)
Experimental	0.353	1.663	0.662
	(1.160)	(1.286)	(0.881)
Experimental, Long-Term	1.662*	0.251	1.012
	(1.005)	(1.359)	(0.821)