

ESSAYS ON DEVELOPMENT AND HEALTH ECONOMICS

by

JOE D. MITCHELL-NELSON

A DISSERTATION

Presented to the Department of Economics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2022

DISSERTATION APPROVAL PAGE

Student: Joe D. Mitchell-Nelson

Title: Essays on Development and Health Economics

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Trudy Ann Cameron	Chair
Alfredo Burlando	Core Member
Shankha Chakraborty	Core Member
Richard York	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies
-------------------	-----------------------------------

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

© 2022 Joe D. Mitchell-Nelson
This work is licensed under a Creative Commons
Attribution-NonCommercial (United States) License.



DISSERTATION ABSTRACT

Joe D. Mitchell-Nelson

Doctor of Philosophy

Department of Economics

June 2022

Title: Essays on Development and Health Economics

This research examines the role of culture in two specific contexts—World Bank project management and preferences for pandemic mitigation strategies—and contributes a novel econometric method for sample selection correction for choice experiments.

Chapter 2 explores how the cultural background of World Bank project leaders affects the success of foreign aid projects, using a constructed measure of cultural proximity between project leaders and the countries where their projects take place. A principal-agent model of project leaders' incentives predicts that cultural proximity and a recipient country's institutional quality will interact to affect project quality. This prediction is borne out in data on project evaluations of 1,946 World Bank projects.

Chapter 3 examines individual preferences for local COVID-19 lockdown policies that force trade-offs between, on the one hand, deaths and illnesses averted, and, on the other hand, employment and income. We field a choice experiment to 993 respondents to determine individuals' willingness to make these trade-offs, and we specifically examine the effect of federal unemployment insurance on these decisions. We find that a stronger social safety net for the unemployed makes individuals, on average, *more willing* to accept county-level income losses but *less*

willing to accept increases in county-level unemployment rates in exchange for reduced COVID-19 deaths and illnesses. Split sample regressions reveal that this puzzling change in preferences is driven almost entirely by politically moderate and conservative respondents.

Finally, chapter 4 proposes a new method for sample selection correction for conditional logit models based on mixed logit estimation methods. Survey-based research methods can produce biased estimates if the responding sample is systematically different from the population of interest. A seminal paper by Heckman (1979) demonstrates how an explicit response/non-response model can be combined with a least-squares-based outcome model to correct for selection bias, but this approach is inappropriate for the conditional logit choice models typically used to analyze the data from choice experiments. Our new method, however, is appropriate for addressing sample selection in choice experiments, which are often used to value goods, services, and social policies that are not traded in markets.

This dissertation includes previously unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Joe D. Mitchell-Nelson

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Portland State University, Portland, OR
University of Maryland, College Park, MD
Arizona State University, Tempe, AZ

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2022, University of Oregon
Master of Science, Economics, 2017, University of Oregon
Bachelor of Science, Economics, 2015, Portland State University
Bachelor of Science, Philosophy, 2012, Portland State University

AREAS OF SPECIAL INTEREST:

Applied econometrics
Development economics

PROFESSIONAL EXPERIENCE:

Research Analyst, Oregon Health Authority, 2021-2022

GRANTS, AWARDS AND HONORS:

Graduate Teaching Award, University of Oregon, 2020
Kleinsorge Summer Fellowship, University of Oregon, 2018
Kleinsorge First-year Fellowship, University of Oregon, 2016
Harold Vatter Award, Portland State University, 2015

ACKNOWLEDGEMENTS

I thank professors Trudy Ann Cameron, Shankha Chakraborty, Alfredo Burlando, and Richard York for their insights and expertise as this research developed. Perhaps more importantly, I thank them for their mentorship, encouragement, professional advice, patience, and humanity. I am especially indebted to professor Cameron, who has been a tireless champion throughout the research process.

This research benefited from useful comments from participants in the University of Oregon Trade Group, the 2021 Econometric Society's Winter School, the 2021 International Choice Modelling Conference, and the 2022 Society for Benefit-Cost Analysis Annual Conference. I am deeply grateful to participants in the University of Oregon Development Economics Group, and the group's organizer, Alfredo Burlando, for many rounds of excellent feedback as this work evolved.

Finally, I wish to thank my fellow University of Oregon Economics graduate students, too numerous to name here, who provided insight, sanity checks, coffee, copy-editing, and moral support.

This work has been supported in part by the endowment accompanying the Raymond F. Mikesell Chair in Environmental and Resource Economics at the University of Oregon. Any remaining errors are my own.

To my wife Kati, who supports me in ways innumerable. To Matilda, who shines like the sun on my darkest days. And to my parents, who did more than raise me.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. CULTURAL INSIDERS AND TRANSNATIONAL PROJECT MANAGEMENT: EVIDENCE FROM THE WORLD BANK	4
2.1. Introduction	5
2.2. Related literature	10
2.3. Conceptual model	11
2.3.1. Setting up the model	12
2.3.2. Solving the model	13
2.4. Empirical strategy	17
2.4.1. Cultural proximity	19
2.4.2. Institutional quality	21
2.4.3. Outcome variable: project success	22
2.4.4. Other controls	24
2.4.5. Potential endogeneity of TTL cultural proximity	27
2.5. Results	31
2.5.1. IV results	31
2.6. The possibility of mismeasurement for female TTLs	36
2.7. Conclusion	39
III. WILLINGNESS TO PAY FOR PANDEMIC RISK REDUCTIONS: LOST LIVES VERSUS LOST LIVELIHOODS	41
3.1. Introduction	42
3.2. Data	48
3.2.1. Sample selection and response propensities	50

Chapter	Page
3.2.2. Estimating sample for policy choice models	52
3.2.3. Estimating specification	56
3.3. Results and Discussion	58
3.3.1. Homogeneous preferences	58
3.3.2. Role of federal unemployment insurance	61
3.3.3. Preference heterogeneity	65
3.4. Conclusions	69
IV. AMPLE CORRECTION FOR SAMPLE SELECTION IN THE ESTIMATION OF CHOICE MODELS USING ONLINE SURVEY PANELS	75
4.1. Introduction	76
4.2. Example: Willingness to pay for carbon emissions reductions	78
4.3. Data needs for modeling response propensities	79
4.4. Response/non-response and policy preferences as correlated choices	80
4.4.1. Algebra of the selection model	81
4.4.2. Algebra of the outcome model	82
4.4.3. Simplest correlation: Selection equation intercept and choice-model any-policy effect	85
4.4.3.1. Analog to probit selection model in conventional Heckman correction	91
4.4.3.2. Analog to a normal-error “outcome” model in conventional Heckman correction	92
4.4.3.3. Accommodating the truncation due to sample selection	93
4.4.3.4. FIML estimation by a generalization of mixed logit models	94

Chapter	Page
4.4.4. When sample selection also affects the marginal utility of policy attributes in the policy-choice submodel . . .	95
4.4.5. Joint estimation of the selection and policy choice submodels .	100
4.5. Simulations	102
4.5.1. Data generating process	102
4.5.2. One large-sample simulation	103
4.5.3. Many small simulations	107
4.6. Conclusion	108
V. CONCLUSION	110
APPENDICES	
A. CHAPTER 2 APPENDIX	112
A.1. Data appendix	112
A.1.1. Genetic distance	112
A.1.2. IEG evaluation	112
A.2. Structural estimation results (preliminary)	113
A.3. Instrument validation	118
A.4. Surnames for female TTLs	119
A.5. Alternative clustering and robustness checks	121
A.6. Dyadic regression results	124
B. CHAPTER 3 APPENDIX	130
B.1. Online Appendix: Other Pandemic Policy Choice-Experiment Surveys	130
B.2. Online Appendix: Survey development	135
B.3. Online Appendix: Selection model	138
B.3.1. Variables available for selection model	138

Chapter	Page
B.3.2. Coefficient estimates for selection model	140
B.4. Online Appendix: One example of a choice set, with pop-ups	144
B.4.1. One example of a policy-choice summary table	144
B.5. Online Appendix: Joint distribution, key choice-set design features	148
B.6. Online Appendix: Complete estimation results	149
B.7. Online Appendix: Other types of choice models	158
B.7.1. Mixed logit models	158
B.7.2. Latent class models	160
B.7.3. Sensitivity analysis: Preferences as a function of time spent on first-choice preamble about federal UI assumption	163
B.7.4. Sensitivity analysis: Preferences as a function of highest pandemic-month unemployment rate in respondent’s county relative to sample median	165
B.7.5. Sensitivity analysis: Preferences as a function of highest pandemic-month unemployment rate in respondent’s county relative to sample median	167
B.7.6. Sensitivity analysis: Preferences as a function of household income relative median household income in the respondents ZIP code	169
C. CHAPTER 4 APPENDIX	171
C.1. Online Appendix: Treatment of sample selection in choice models in the literature on choice modeling	171
C.1.0.1. Abstract claims a “representative sample”	171
C.1.0.2. Abstract mentions representativeness in terms of specific observable variables	172
C.1.0.3. Abstract mentions “representative household”	173
C.1.0.4. Abstract does not mention "representativeness"	173

Chapter	Page
C.2. Do we need to bother scaling the coefficients in the mixed-logit <i>policy-choice</i> equation?	174
REFERENCES CITED	176

LIST OF FIGURES

Figure	Page
1. Optimal TTL effort as a function of cultural proximity	18
2. Distribution of sample projects' CPIA ratings	23
3. Complementarity with Rees-Jones, D'Attoma, Piolatto, and Salvadori (2020) study	47
4. Timeline of survey responses relative to the American Rescue Plan of 2021	49
5. Fitted response propensities	53
6. Preference parameter distribution for simulated responders and non-responders	104
7. Distributions of marginal WTP for benefit in responder/non-responder sub-samples, full sample, and selection-corrected estimates using simulated data	105
8. Distributions of total WTP for a policy with <i>benefit</i> = 1.5 and <i>any policy</i> = 1 in responder/non-responder sub- samples, full sample, and selection-corrected estimates using simulated data	107
9. Distributions of WTP for benefit: 100 simulations of 2,000 invitees	108
A.1. Predicted effort	116
A.2. Fitting $h(I)$	117
A.3. Effect of selected variables on receiving specific IEG scores	123
D1. Immediate preamble to first summary table (one example)	144
D2. One instance of Policy A; contents of first three pop-ups	145
D3. Contents of fourth through seventh popups	145
D4. Contents of eighth through eleventh popups	146
D5. Contents of twelfth through 15th popups	146

Figure	Page
D6. Choice question that immediately follows the table	147
E1. Independent variation in average household costs and unemployment, by level of federal UI supplement	148

LIST OF TABLES

Table	Page
1. Summary statistics for high- and low-cultural proximity projects	25
2. OLS results, selected coefficients	32
3. IV: First stage regressions (selected coefficients)	33
4. IV results, selected coefficients	35
5. IV results, using WGI in place of CPIA	37
6. Correlation between measures of cultural proximity based on given name and surname	37
7. Other Covid-19 choice-experiment studies	46
8. Descriptive statistics for featured variables in our randomized design	55
9. Effects of Federal UI payments on preferences over pandemic policies; selected coefficients. (Complete models in Appendix B.6, Table F2)	60
10. Heterogeneity in preferences across socioeconomic groups; generalizations of Model 6 in Table 9; selected coefficients. (Complete models in Appendix B.6, Tables F3 and F4)	67
11. Parameter means for full sample and responder/non- responder sub-samples	103
12. Simulated marginal WTPs for full sample, responder sub- sample, and selection-corrected parametric bootstrap	106
13. Simulated total WTPs for full sample, responder sub-sample, and selection-corrected parametric bootstrap	106
B1. Structural parameter estimates	115
C1. Tests of the IV's exclusion restriction	118
D1. IV results with gender controls, selected coefficients	120

Table	Page
E1. IV results, standard errors clustered by country	124
E2. IV results, standard errors clustered by country-approval year	125
E3. IV results, standard errors unclustered	126
E4. OLS results with country fixed effects	127
E5. IV linear probability models	128
E6. IV Results for CPIA clusters	128
F1. Dyadic regressions	129
C1. Descriptive statistics: Basic variables for selection model, retained by LASSO model, either as individual variables or as part of a pairwise interaction term.	139
C2. Estimated coefficients for selection model, a binary logit specification employing all variables retained by the preliminary LASSO model (incompletely sorted)	140
F1. Descriptive statistics for restrictions on activities, across offered pandemic policies (included in this paper as incidental controls)	149
F2. Full set of parameter estimates for the models reported in Table 9 in the body of the paper: Effects of Federal UI (selected coefficients)	150
F3. Full sets of parameter estimates for the models reported in Panel A of Table 10 in the body of the paper: Heterogeneity in preferences	153
F4. Full sets of parameter estimates for the models reported in Panel B of Table 10 in the body of the paper: Heterogeneity in preferences across socioeconomic groups	155
F5. Standard errors clustered by respondent, with federal UI entering as in Models 3 and 6 in Table 9 (Panel A) and in “baseline plus interactions” form (Panel B).	157
G1. Two mixed logit specifications	159
G2. Specification with two latent preference classes, based on Model 3	162

Table	Page
G3. Reading time	164
G4. Worst month's unemployment rate relative to median worst pandemic unemployment rate prior to survey	166
G5. Heterogeneity according to correctness of answers on two comprehension questions during the survey's tutorial	168
G6. Zip code relative income. Household income below median zip code income (Poorer) versus above median zip code income (Richer)	170

CHAPTER I

INTRODUCTION

Many cultural attitudes and beliefs concern the meta-questions of economics: *How do we decide what is valuable? What kinds of trade-offs are ethically permissible? Whose well-being matters?* Economics, as a discipline, does not and should not attempt to answer these questions. But to the extent that answers to these questions vary across cultures, economics should concern itself with cultural differences. Chapters 2 and 3 investigate different dimensions of the role of culture in economic outcomes. Chapter 4 is an effort to improve stated preference methodology. Stated preference methods are not directly related to culture, but are an invaluable tool to any economist wishing to learn how individuals value goods and services that are not traded in markets (e.g. cultural goods or communal goods). Chapter 5 concludes.

Chapter 2 explores how the cultural background of project leaders affects the success of foreign aid projects. I use a new measure of cultural proximity between countries, based on the genetic distance measure compiled by Spolaore and Wacziarg (2018) and data from the World Bank, to quantify how much cultural overlap likely exists between project leaders and the countries where these projects take place. I then present a principal-agent model that illustrates theoretical reasons to infer that cultural background and a recipient country's institutional quality should interact to determine the project leader's level of effort. I find that this structural model describes the data better than a number of intuitive reduced-form specifications, and that both structural and reduced form models suggest that cultural background matters. To address possible endogeneity arising from assignment of managers to projects, I instrument for cultural proximity with the

average cultural proximity of other available project leaders. Where institutions are strong, culturally similar project managers outperform those who are more culturally distant, but this relationship is not present in countries with poor institutions.

Chapter 2 focuses on the cultural *match* between managers and recipients, without considering the *content* of any particular culture. Chapter 3 (co-authored with Trudy Ann Cameron), however, examines the relationship between cultural attributes and individual preferences for costly pandemic restrictions. In Chapter 3, we quantify the trade-offs that individuals are willing to make across the domains most relevant to pandemic mitigation policies—cases avoided, deaths prevented, and loss of access to goods and services and economic burdens—and examine how their preferences vary as a function of measurable cultural signifiers. Of particular interest is the relationship between an individual’s willingness to pay to prevent the illness and death of others and that individual’s political affiliation, which is currently perhaps the most salient cultural division within the US. Our online stated preference survey uses policy choice experiments for a regionally representative sample of respondents. In each policy choice scenario, respondents choose between costly sets of pandemic restrictions and a status quo option that allows individuals to make their own individual decisions about what measures to take, with resulting higher numbers of cases and deaths. Our estimating specification is a random utility model that permits estimation of overall and marginal willingness to bear the societal costs and willingness to accept restrictions of different levels of stringency imposed on ten different categories of businesses and activities. Preferences over these public health policies are allowed to vary

systematically as a function of individual characteristics such as age, expected vulnerability to COVID-19, and self-reported ideology.

Chapter 4 (co-authored with Trudy Ann Cameron) of this dissertation proposes a new econometric method for correcting selection bias in choice experiment surveys. Researchers in this domain often weight survey results so that their weighted sample approximately matches the population of interest in terms of observable attributes such as age or gender. Unfortunately, this technique and similar strategies cannot account for *unobservable* characteristics that may be correlated with both the propensity to respond to the survey and characteristics and attitudes the survey is designed to study. For example, a potential respondent's interest in the topic of the survey may make them more likely to respond and more likely place a high valuation on proposed policies in the survey. While sample selection correction methods exist for least-squares-based models, such as binary-outcome probit models, these methods are not appropriate for the many choice experiments that present respondents with more than two alternatives. The newly proposed method corrects for selection on unobservables (such as interest in the survey topic) in choice contexts with more than two alternatives.

CHAPTER II

CULTURAL INSIDERS AND TRANSNATIONAL PROJECT MANAGEMENT: EVIDENCE FROM THE WORLD BANK

This research explores how the cultural background of project leaders affects the success of foreign aid projects. I use a new measure of cultural proximity between countries, based on the genetic distance measure compiled by Spolaore and Wacziarg (2018) and data from the World Bank, to quantify how much cultural overlap likely exists between project leaders and the countries where these projects take place. I then present a principal-agent model that predicts the effort levels of project leaders as a function of cultural proximity and institutional quality. I find that this structural model describes the data better than a number of intuitive reduced-form specifications. To address possible endogeneity arising from assignment of managers to projects, I instrument for cultural proximity with the average cultural proximity of other available project leaders. Where institutions are strong, culturally similar project managers outperform those who are more culturally distant, but this relationship is not present in countries with poor institutions.

2.1 Introduction

Foreign aid projects can face serious and often unforeseeable challenges. Goods and services provided by foreign aid projects do not always reach recipients, often because of cost overruns and delays. The Soviet Union, for example, assisted Nigeria in building an \$8 billion steel mill starting in 1979. Despite intermittent efforts to complete the mill over the last 40 years, Ajaokuta Steel Mill has never produced any iron or steel (Mold, 2012). When goods and services do find their way into the right hands, the actual recipients may find little use for them. Point-of-use water sanitation products (e.g. chlorine tablets and ceramic filters) have notoriously low usage rates, even when they are subsidized or free and come with in-person education about the dangers of untreated drinking water (Luoto et al., 2011). Finally, much-needed aid can also have unintended consequences. A World Bank project in Kenya aimed at connecting poor residents to the public electrical grid has fueled the rise of electricity cartels in some districts of Nairobi. Cartels cut the metered connections to households and then offer to shunt those same households back into the grid for a fee (Langat, 2019). Cassen (1994) reports that about one-quarter to one-third of foreign aid projects, overall, are not successful. Given that net worldwide foreign aid has climbed to over \$160B annually, factors that can lower the failure rate of foreign aid projects have the potential to avert tremendous losses of scarce resources.¹

To this end, I identify the causal effect of one potentially important but as-yet unexplored factor in the success of foreign aid projects: the cultural

¹This measure of foreign aid includes only official development assistance (ODA). The OECD counts a transfer as ODA if (1) it comes from an official body (usually a government or a multilateral donor); (2) it “is administered with the promotion of the economic development and welfare of developing countries as its main objective”; and (3) it “is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 percent).”

background of project leaders. Specifically, I construct a measure of the likely cultural proximity between recipient countries and the individuals who are assigned to supervise the planning and execution of a project. As will be discussed in more detail in Section 2, several previous studies of foreign aid projects have concluded that the identities of project supervisors matter, but none of these studies has identified exactly which characteristics of these project supervisors are most important. A project supervisor's cultural proximity, or familiarity with the culture of the recipient country, may help that supervisor produce more effective outcomes. An understanding of the cultural complexities of a recipient population seems an obvious qualification for determining and supplying the needs of that population. Cultural proximity may lead project supervisors to ask the right questions when planning the project, to catch likely problems with implementation before they become apparent to others, and to rightly trust their gut.

I find that World Bank project supervisors, officially called task team leaders (TTLs), who are likely to be culturally similar to their recipient country have, on average, better project outcomes, provided that the recipient country has sufficiently strong institutions, as measured by the World Bank's Country Policy and Institutional Assessment (CPIA). When projects are located in countries with poor institutions, however, culturally close TTLs seem to have no advantage over their culturally more-distant counterparts. These results have clear implications for staffing and hiring decisions at multilateral aid organizations.

I do not observe each TTL's cultural backgrounds directly. Instead, I rely on *genetic distance*, an existing measure of cultural divergence between countries developed by Spolaore and Wacziarg (2009; 2016; 2018). Given that cultural traits are transmitted intergenerationally, as are genetic traits, Spolaore and Wacziarg

argue that the genetic distance between populations constitutes an appropriate proxy for the cultural distance between populations.² The most recent measure of genetic distance developed by Spolaore and Wacziarg—the data I use here—is based on neutral genetic features that geneticists believe are not affected by selection pressures. As a result, any variation in these features across different populations can be attributed to random drift. The longer any two populations have been spatially separated, the more drift will have occurred, and the more genetically distant these two populations will be. Like these genetic traits, we expect cultural traits also to diverge further as the duration of separation increases.

Though genetic distance is conceptually a measure that compares two populations, here I use it to compare individuals to populations—namely, the TTLs who oversee aid projects on behalf of the World Bank in relation to the countries that receive this aid. In doing so, I assume that TTLs are familiar through experience with the culture of their own ancestral country. Of course, I do not assume that TTLs personally adopt every aspect of their home country’s culture, which may in fact be a patchwork of regional subcultures. Throughout this paper, I refer to the *negative* of the measured genetic distance between a TTL’s ancestral country and the recipient country where the TTL supervises a foreign aid project as the *cultural proximity* between that TTL and the country where their assigned project is located.

A vast literature has examined the effectiveness of foreign aid. One branch of this literature has sought to measure the effect of foreign aid on GDP growth and other broad macroeconomic measures of development, but has arrived at

²Genetic distance is a measure based on the similarity of the distributions of genetic markers between two populations. Appendix A.1.1 provides a detailed description of how Spolaore and Wacziarg’s measure of genetic distance is calculated.

no consensus about the effect of foreign aid on GDP growth.³ In a thorough review of the literature to that date, Temple (2010) writes that the “cross-country evidence on the effects of aid must be regarded as a work-in-progress.” In a more-recent review of the literature, Qian (2015) similarly argues that the effect of foreign aid on a developing country’s economic growth “is perhaps among the most controversial [issues] in development and growth economics.”

The present research contributes to another branch of the literature on the effectiveness of foreign aid, which focuses on the outcomes of *individual* aid projects, rather than country-level aggregate measures of success. This body of research implicitly sets a much looser criterion for the success of aid. Aid is recorded as successful when projects demonstrably and efficiently meet their narrowly defined objectives. These projects may or may not contribute to improved growth rates. Furthermore, this literature is rarely able to show that aid does not simply substitute for government expenditures, which may then be diverted to less socially desirable objectives. Also, there may be complementarities among aid projects that are not captured by project-level analyses. Nevertheless, a foreign aid project can hardly be considered successful if it fails even on its own terms.

Much of the literature concerning the success of individual aid projects has focused on characteristics of recipient countries, and of the foreign aid projects themselves, as explanatory variables (see for instance Denizer, Kaufmann, and Kraay (2013); Duponchel, Chauvet, and Collier (2010)). For several reasons, many of these studies draw their samples from the same World Bank database of aid projects. First, the World Bank is a multilateral donor, and so is considered

³Some scholars find evidence that foreign aid can lift countries out of poverty (e.g. Arndt, Jones, and Tarp (2010); Burnside and Dollar (2000)). Others report that aid does not improve growth, even in countries with strong institutions (e.g. Doucouliagos and Paldam (2009, 2011); Easterly (2003); Rajan and Subramanian (2008)).

less likely to direct aid flows specifically to advance any single country's foreign policy objectives. Second, the World Bank is among the largest multilateral donors. In the last decade, the World Bank distributed more foreign aid than the United Nations, the International Monetary Fund and the World Health Organization combined. Finally, every World Bank project is uniformly assessed by the Independent Evaluation Group (IEG), a distinct arm of the World Bank, whose sole responsibility is objective evaluation of the development effectiveness of the World Bank group. For these reasons, I rely on the same World Bank data as (2013) and (2010) for the analysis of foreign aid projects undertaken in this study.

Section 2.2 discusses some related research concerning the determinants of successful foreign aid projects. In Section 2.3, I outline a model of TTL effort as a function of cultural proximity and institutional quality, and derive predictions for the joint effects on project success of cultural proximity and institutional quality. Section 2.4 provides details about the empirical strategy I use to (1) measure cultural proximity between the TTL and the recipient country and (2) estimate the effects on project success of cultural proximity and the recipient country's institutional quality. One contribution of this paper is to combine Spolaore and Wacziarg's data with other sources to develop a plausible measure of cultural proximity between individuals and countries that does not rely on demographic data about these individuals. Section 2.5 presents the main results and some robustness checks. Section 2.6 explores and comments upon one potential source of mismeasurement in my construction of the *cultural proximity* variable, and section 2.7 concludes.

2.2 Related literature

The development economics literature has produced a number of studies that examine factors that improve the outcomes of donor-funded development projects (Chauvet, Collier, & Fuster, 2017; Dollar & Svensson, 2000; Kilby, 2000), including many studies published by the World Bank itself (Duponchel et al., 2010; Guillaumont & Laajaj, 2006).⁴ While other papers in the economics literature have recognized that TTL characteristics matter, the present paper describes what seems to be the first effort to examine one specific characteristic of TTLs beyond just broad measures of their ability.

Studies published by the World Bank frequently identify recipient country institutions as one of the most important determinants of project success. Dollar and Levin (2005) use an instrumental variables strategy to determine the effect of institutional quality on development project outcomes, and find evidence of a causal, positive relationship between institutional quality and the proportion of projects in each country rated as “successful” by the World Bank’s Operations Evaluation Department.⁵ Geli, Kraay and Nobakht (2014) develop a relatively simple predictive model to determine World Bank project outcomes. They estimate a probit model to predict project success, using only a handful of variables: log of project cost, preparation time, initially planned project length, the outcome of TTLs’ other projects, and the recipient country’s institutional quality. They

⁴Scholars of project management have also given some attention to development projects (Diallo & Thuillier, 2005; Ika, Diallo, & Thuillier, 2012). The disciplines of economics and project management bring different assumptions and methodological procedures to the question of development project effectiveness. Where the economics literature emphasizes the role of recipient country and project characteristics, the project management literature focuses on the structure of working groups and characteristics of team members and their relationships with each other.

⁵The Operations Evaluation Department was the precursor to the IEG, whose ratings I use to determine project quality.

find that institutional quality and TTL track record are the strongest predictors of project success.

Denizer, Kaufmann and Kraay (2013) compare the effects on project outcomes of (a) country characteristics (e.g. institutions, GDP growth), and (b) project characteristics (e.g. duration, cost, sector), and find that both matter. Most relevant to the present analysis, these authors use the IEG scores that a given TTL has received on projects *other than* the current project to measure TTL quality. TTL quality is significantly and positively associated with project outcomes, but they do not address which observable TTL characteristics measure TTL quality. Furthermore, it is possible that the results in their paper may be driven by the TTL assignment process. I make two improvements: I examine a particular characteristic of each TTL—their shared cultural background with the recipient country—and I use a plausibly exogenous instrument for TTL assignment.

2.3 Conceptual model

TTL effort is likely to be an important factor in foreign aid project quality. Because I cannot observe effort directly, I introduce a principal-agent model to explain variation in TTL effort as a function of the TTL’s cultural proximity to the recipient country and the institutional quality of the recipient country. The World Bank provides a unique context for a principal-agent model, as TTLs are not directly remunerated on the basis of project quality, though performance may affect future wages. Instead I assume that TTLs receive professional and social benefits based on the level of effort inferred from their supervisors and peers, based on their project’s quality.⁶ The model begins with the following assumptions:

⁶These benefits may include access to more desirable projects or positions in the future, professional recognition from peers, etc.

1. TTLs are utility maximizers who dislike effort.
2. A TTL's effort is more effective when their cultural proximity is high.
3. A project's quality is more predictable when the recipient country's institutional quality is high.
4. TTL effort is not directly observable by the principal.
5. TTLs are rewarded in proportion to the probability, from the principal's perspective, that the TTL's effort exceeded some threshold.

2.3.1 Setting up the model. The agent chooses an effort level e to maximize the indirect utility function:

$$v(e) = R(e, \eta) - \phi(e) \tag{2.1}$$

where R is the reward the agent receives from the principal, η is an error term (to be described shortly), and $\phi(e)$ gives the agent's disutility of effort. R is increasing in e and η . Assume also that $\phi'(e) > 0$ and $\phi''(e) > 0$ so that marginal *disutility* of effort is *increasing*.

The principal observes project quality Q and forms a belief about the distribution of possible levels of the TTL's true effort e . Project quality Q is determined by the following production function:

$$Q = q(X) + g(C)e + \eta \tag{2.2}$$

where $q(X)$ is a function of variables X that affect project quality but are outside of the TTL's control, and $g(C)e$ is the total contribution of TTL effort to project quality. The function $g(C) > 0$ is increasing and gives the marginal contribution of effort to project quality, reflecting the assumption that TTL effort is more effective when their cultural proximity is high. The stochastic error term η is

distributed with mean $\mu_\eta = 0$ and variance $\sigma_\eta^2 = h(I)$. The variance $h(I)$ of η is a decreasing function of I , reflecting the assumption that project outcomes are more predictable in recipient countries with high institutional quality I . Let $F(x; h(I))$ and $f(x; h(I))$ be, respectively, the cumulative density function and the probability density function of η .

2.3.2 Solving the model. The principal knows the values of Q , and X (note that $C, I \in X$), along with the functional forms of Q , q , g and h , but is unable to observe η or e . The principal can then calculate an “observed” effort level $e_o = e + \frac{\eta}{g(C)}$ by rearranging the production function:

$$Q = q(X) + g(C)e + \eta$$

$$\frac{Q - q(X)}{g(C)} = e + \frac{\eta}{g(C)} = e_o$$

Note that $E(e_o) = e$. Recall that the principal rewards the TTL in proportion to the probability that TTL effort was above some threshold value m . How does the principal use e_o to determine the probability that the agent’s effort e was above the threshold m ? Intuitively, if the principal observes a low level of e_o , such that $e_o < m$, they must take into account the possibility that in fact $e > m$ but that η was negative and large in magnitude—the agent may have worked hard but been unlucky. Conversely, if the principal observes $e_o > m$, it remains possible that $e < m$ but the agent was unlucky and η was positive and large. For simplicity I suppose the principal has no prior belief about the agent’s likely level of effort.

Given the principal’s information set, the probability that $e > m$ is

$$P(e > m) = P[\eta < (e_o - m)g(C)] = F[(e_o - m)g(C); h(I)]$$

The TTL chooses their true effort level e before observing η . From the TTL's point of view, then, the reward R can be expressed as

$$\begin{aligned} R(e, \eta) &= F[(e_o - m)g(C); h(I)] \\ &= F\left[\left(e + \frac{\eta}{g(C)} - m\right)g(C); h(I)\right] \\ &= F[(e - m)g(C) + \eta; h(I)] \end{aligned}$$

The TTL then integrates over the set of rewards times their probability to calculate the *expected* reward $E(R(e))$ for any level of effort e :

$$E(R(e)) = \int_{-\infty}^{\infty} F[g(C)(e - m) + x; h(I)]f(x; h(I))dx$$

Then the agent's expected utility is given by

$$E(v(e)) = \int_{-\infty}^{\infty} F[g(C)(e - m) + x; h(I)]f(x; h(I))dx - \phi(e)$$

which yields the following necessary condition for a maximum:

$$\frac{\partial}{\partial e} \left(\int_{-\infty}^{\infty} F[g(C)(e - m) + x; h(I)]f(x; h(I))dx \right) \Big|_{e=e^*} - \phi'(e^*) = 0$$

Unfortunately, it is not possible to be more specific about the solution to the model without choosing an explicit distributional family for η . To proceed, I assume a logistic distribution for η . Given that η has mean 0 and variance $h(I)$, we have location parameter $\mu = 0$ and scale parameter $s = \frac{\sqrt{3h(I)}}{\pi}$. For notational parsimony, also let $a = g(C)(e - m)$. The TTL's first order condition is then:

$$\frac{\partial}{\partial e} \left(\int_{-\infty}^{\infty} \frac{1}{1 + \exp\left(\frac{-x-a}{s}\right)} \times \frac{\exp\left(\frac{-x}{s}\right)}{s \left(\exp\left(\frac{-x}{s}\right) + 1\right)^2} dx \right) \Big|_{e=e^*} - \phi'(e^*) = 0$$

Solving the integral above yields the antiderivative, denoted $B(x)$:

$$B(x) = -\frac{\exp\left(\frac{a}{s}\right) \left\{ \left[\exp\left(\frac{x}{s}\right) + 1\right] \left[\ln\left(\exp\left(\frac{x+a}{s}\right) + 1\right) - \ln\left(\exp\left(\frac{x}{s}\right) + 1\right)\right] + \exp\left(\frac{a}{s}\right) - 1 \right\}}{\left(\exp\left(\frac{a}{s}\right) - 1\right)^2 \left(\exp\left(\frac{x}{s}\right) + 1\right)}$$

Then we take the limits of $B(x)$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$ to evaluate the integral:

$$\lim_{x \rightarrow \infty} B(x) = \frac{-a \times \exp\left(\frac{a}{s}\right)}{s \left(\exp\left(\frac{a}{s}\right) - 1\right)^2}$$

$$\lim_{x \rightarrow -\infty} B(x) = \frac{-\exp\left(\frac{a}{s}\right)}{\exp\left(\frac{a}{s}\right) - 1}$$

Calculating the difference $\lim_{x \rightarrow \infty} B(x) - \lim_{x \rightarrow -\infty} B(x)$ produces the following definite integral:

$$\frac{\exp\left(\frac{a}{s}\right)}{\exp\left(\frac{a}{s}\right) - 1} - \frac{a \times \exp\left(\frac{a}{s}\right)}{s \left(\exp\left(\frac{a}{s}\right) - 1\right)^2}$$

We can then express the TTL's optimization problem as:

$$\max_e \frac{\exp\left(\frac{a}{s}\right)}{\exp\left(\frac{a}{s}\right) - 1} - \frac{a \times \exp\left(\frac{a}{s}\right)}{s \left(\exp\left(\frac{a}{s}\right) - 1\right)^2} - \phi(e)$$

Let e^* be the TTL's utility-maximizing level of effort, and let $a^* = g(C)(e^* - m)$.

Then the TTL's first order condition is:

$$g(C) \times \frac{\exp\left(\frac{a^*}{s}\right) \left((a^* - 2s) \exp\left(\frac{a^*}{s}\right) + a^* + 2s \right)}{s^2 \left(\exp\left(\frac{a^*}{s}\right) - 1\right)^3} - \phi'(e^*) = 0 \quad (2.3)$$

We are ultimately interested in how the TTL's chosen effort level e^* varies with cultural proximity C and institutional quality I . That is, we would like to say something about $\frac{\partial e^*}{\partial C}$ and $\frac{\partial e^*}{\partial I}$. To find and sign these partial derivatives, we turn to the implicit function theorem. Let $D(e^*, I, C)$ denote the left-hand side of equation 2.3.⁷ Then we have:

$$\frac{\partial e^*}{\partial C} = -\frac{D_C}{D_{e^*}} \quad (2.4)$$

$$\frac{\partial e^*}{\partial I} = -\frac{D_I}{D_{e^*}} \quad (2.5)$$

where the components take the form:

⁷It seems prudent to note again here that the logistic distribution's shape parameter $s = \frac{\sqrt{3h(I)}}{\pi}$, a parameterization chosen so that $\text{var}(\eta) = h(I)$.

$$\begin{aligned}
D_{e^*} &= -\frac{\exp\left(\frac{a^*}{s}\right) \left((a^* - 3s) \exp\left(\frac{2a^*}{s}\right) + 4a^* \exp\left(\frac{a^*}{s}\right) + a^* + 3s \right)}{s^3 \left(\exp\left(\frac{a^*}{s}\right) - 1 \right)^4} g(C) \frac{\partial a^*}{\partial e^*} - \phi''(e^*) \\
D_C &= g'(C) \left(\frac{\exp\left(\frac{a^*}{s}\right) \left((a^* - 2s) \exp\left(\frac{a^*}{s}\right) + a^* + 2s \right)}{s^2 \left(\exp\left(\frac{a^*}{s}\right) - 1 \right)^3} \right) \\
&\quad - g(C) \frac{\partial a^*}{\partial C} \left(\frac{\exp\left(\frac{a^*}{s}\right) \left((a^* - 3s) \exp\left(\frac{2a^*}{s}\right) + 4a^* \exp\left(\frac{a^*}{s}\right) + a^* + 3s \right)}{s^3 \left(\exp\left(\frac{a^*}{s}\right) - 1 \right)^4} \right) \\
D_I &= g(C) \frac{\partial s}{\partial I} \frac{\exp\left(\frac{a^*}{s}\right) \left((2s^2 - 4a^*s + a^{*2}) \exp\left(\frac{2a^*}{s}\right) + (4a^{*2} - 4s^2) \exp\left(\frac{a^*}{s}\right) + 2s^2 + 4a^*s + a^{*2} \right)}{s^4 \left(\exp\left(\frac{a^*}{s}\right) - 1 \right)^4}
\end{aligned}$$

Given that e^* is a maximum, we know that $D_{e^*} < 0$, since D_{e^*} is identical to the second derivative of the TTL's maximization problem. Given the sign of D_{e^*} and conditions 2.4 and 2.5, we have $\text{sign}\left(\frac{\partial e^*}{\partial C}\right) = \text{sign}(D_C)$ and $\text{sign}\left(\frac{\partial e^*}{\partial I}\right) = \text{sign}(D_I)$, and these functions turn out to have identical positivity conditions. We find that $D_C > 0$ and $D_I > 0$ if and only if:

$$\exp\left(\frac{a^*}{s}\right) > -\frac{(a^* \sqrt{4s^2 + 3a^{*2}} - 2s^2 + 2a^{*2})}{(2s^2 - 4a^*s + a^{*2})} \quad (2.6)$$

Usefully, the signs of D_C and D_I change precisely when the ratio $\left|\frac{a^*}{s}\right| = 2.3469413\dots \approx 2.347$, determined by plotting both functions over $a^* - s$ space. This property described by condition 2.6 suggests simpler rules for identifying the signs of $\frac{\partial e^*}{\partial C}$ and $\frac{\partial e^*}{\partial I}$. Substituting $a^* = g(C)(e^* - m)$ and $s = \frac{\sqrt{3h(I)}}{\pi}$ and solving for C and I , we find $\frac{\partial e^*}{\partial C} > 0$ and $\frac{\partial e^*}{\partial I} > 0$ if and only if:

$$C < g^{-1} \left(\frac{2.347 \sqrt{3h(I)}}{\pi |e^* - m|} \right) \quad (2.7)$$

$$I < h^{-1} \left(\frac{1}{3} \left(\frac{\pi g(C)(e^* - m)}{2.347} \right)^2 \right) \quad (2.8)$$

where inequality 2.8 simply re-expresses 2.7. These conditions demonstrate the following predictions of the model:

1. Increased cultural proximity or increased institutional quality makes TTL effort more observable to the principal.
2. Increased observability of effort motivates effort, but only up to a point.

When effort is highly observable, TTLs lose an incentive to work much harder than m , the threshold level of effort.

These predictions capture intuitive notions about how humans choose to exert effort. When our effort is not very effective relative to the degree of a task's unpredictability (i.e. C and I are both low), we will put in little effort. Failure is likely in any case, and exerting effort is tedious and unlikely to produce success. If our effort becomes more effective (but the task remains unpredictable), we may put in considerable effort to ensure success in case unpredictable factors conspire against us. But as our effort becomes still more effective, it is no longer necessary to work quite so hard. This story plays out in figure 1, which plots the TTL's optimal effort over various levels of cultural proximity, holding m and I constant.

2.4 Empirical strategy

With satisfactory measures of cultural proximity between the TTL and the recipient country, the institutional quality of the recipient country, and the success of the project outcome, the predictions of the model can be tested empirically. The basic estimating specification will be:

$$\begin{aligned}
 Outcome_i = & \beta_0 + \beta_1 Institutional\ quality_i + \beta_2 Cultural\ proximity_i + \\
 & \beta_3 (Institutional\ quality_i \times Cultural\ proximity_i) + X'_i \gamma + \varepsilon_i, \quad (2.9)
 \end{aligned}$$

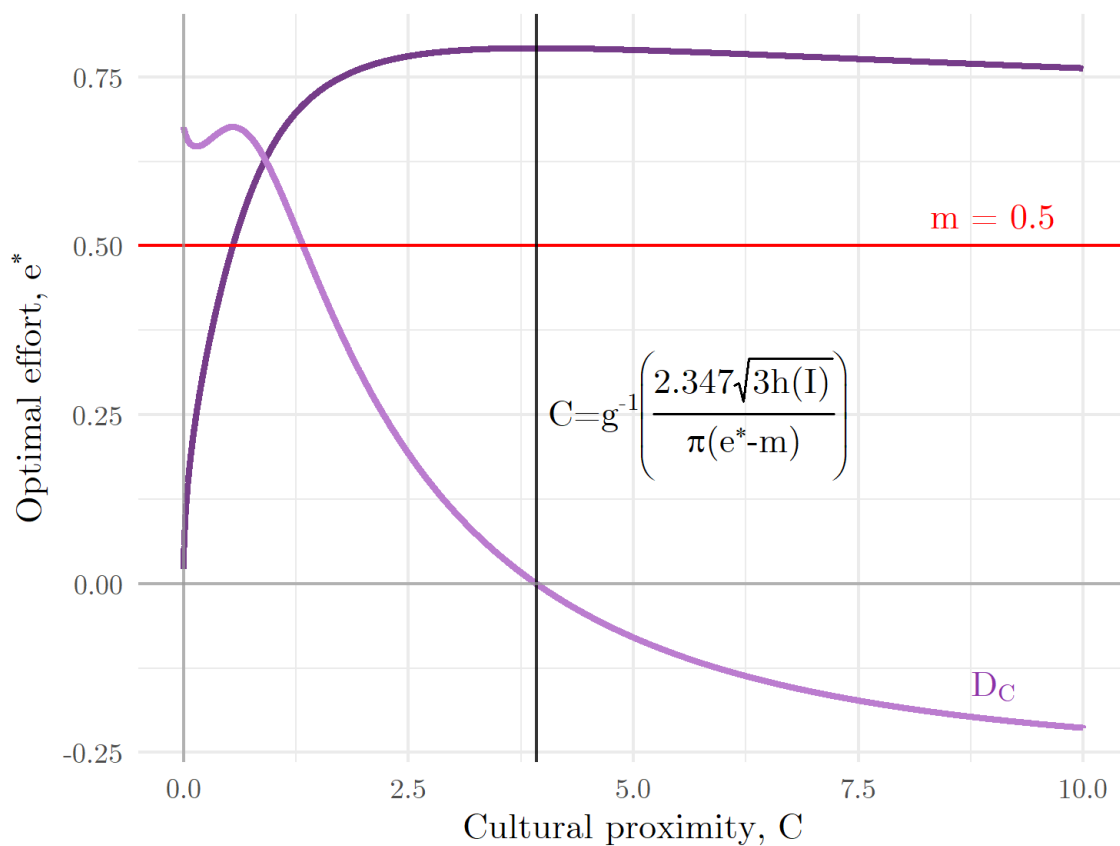


Figure 1. Optimal TTL effort as a function of cultural proximity
 The dark purple line plots e^* as a function of C . The lighter purple line plots D_C .
 The vertical line marks the point where $\frac{\partial e^*}{\partial C}$ becomes negative.
 $I = 5$, $m = .5$, $g(C) = \sqrt{C}$, $h(I) = I^{-1}$.

where $Outcome_i$ is an ordinal measure of success for project i , $Institutional\ quality_i$ is a measure of the institutional quality in project i 's recipient country in the year the project was approved, $Cultural\ proximity_i$ is the negative of genetic distance between the TTL assigned to project i and the recipient country for project i , rescaled such that standard deviation of $Cultural\ proximity_i = 1$. Finally, X_i is a vector of other controls. I choose a linear specification to allow for a rich set of fixed effects and an instrumental variables strategy. The construction of these variables of interest, and the variables included as additional controls, are described in more detail in the sections below.

2.4.1 Cultural proximity. I use a two-step process to estimate a TTL's cultural proximity relative to the recipient country. First, I use TTL surname to determine their likely country of ancestry. Then, I need a measure of genetic distance between the TTL's home country and the project's recipient country to use as a proxy for (negative) cultural proximity. The measure of genetic distance I use was developed by Pemberton et al. (2013) to describe differences in microsatellite variation between human populations, and then aggregated by Wacziarg and Spolaore (2018) to describe genetic distance between entire countries.⁸ The raw data assembled by Pemberton et al. (2013) provide pairwise genetic distances between 267 genetic populations. Wacziarg and Spolaore (2018) match those populations to ethnic groups in each country to aggregate these genetic distances to the country level, using country-level ethnic composition data from Alesina et al. (2003). The measure of genetic distance used by Pemberton et al. and adapted by Wacziarg and Spolaore is based on genetic differences that

⁸Microsatellites are sequences of repeating base pairs in DNA. Microsatellites have a higher mutation rate than other areas of DNA, making them ideal for measuring genetic diversity between populations.

geneticists believe to be neutral, in the sense that these differences are uncorrelated with genetic fitness and are not affected by genetic selection pressures.⁹

As further background, the Alesina et al. (2003) data set includes population proportions of 1120 distinct ethnic groups in each country. Wacziarg and Spolaore match each ethnic group to one of the 267 genetic populations in Pemberton et al. (2013)'s genetic distance data to calculate pairwise, country-level genetic distance. With minor adjustments, I rely on the Wacziarg and Spolaore data set as the basis for the subsequent calculations of my new cultural proximity measure in the context of World Bank TTLs and their assigned recipient countries.

To construct my new variable, I need to know each TTL's country of origin to determine their cultural proximity to the population of the relevant recipient country. Unfortunately, this information is not available in any public archive, although the name of each TTL is a matter of public record. I use TTL surnames to construct probability-weighted collections of likely ancestral countries. I first gather data on the global distribution of each TTL surname from *Forebears*, a globally representative surname database.¹⁰ Most surnames are found in more than one country. Where this is the case, I construct my cultural proximity measure as the negative of the weighted average of the genetic distances between recipient country and the 10 most likely ancestral countries based on TTL surname, where the weights are the proportion of all individuals with the TTL's surname who live

⁹In previous work Wacziarg and Spolaore (2009) use a measure of genetic distance aggregated to the country level from 42 genetic populations. Their data based on the work of Pemberton et al. (2013) represents a significant improvement in the granularity of their resulting data. Two countries, Togo and Tanzania, are not included in Wacziarg and Spolaore's updated data set for 2018. I regress Wacziarg and Spolaore's 2009 measure of genetic distance on their new measure and use interpolated values of the new measure for these countries.

¹⁰*Forebears* is a fairly new resource, having been launched in 2012, but already there is some precedent in academic research for using it, as I do, to determine likely country of origin for individuals based on surname. See Nguyen, Alexiou, and Singleton (2017); Pursiainen (2019).

in each country. I assume here that each TTL’s ancestral country represents a random draw from the global distribution of individuals with the same surname. If this assumption is approximately met, my constructed measure of cultural proximity introduces only attenuation bias in my OLS and IV coefficient estimates, relative to the information I could use if data were available concerning each TTL’s actual history of residence. Note that the maximum value of *cultural proximity* is 0 because it is constructed as the negative of *genetic distance*, an intrinsically non-negative measure. A cultural proximity of 0 would indicate that a TTL has a surname found exclusively in their project’s recipient country.¹¹

2.4.2 Institutional quality. As a measure of institutional quality within each country, I use the Country Policy and Institutional Assessment (CPIA) provided by the World Bank. For each project, I furthermore use the CPIA of the recipient country *in the year* the project was approved. The CPIA, measured on a six-point ordinal scale, is the simple average of four six-point “cluster” scores along the separate dimensions of economic policies, structural policies, quality of public administration, and social inclusion.¹²

The World Bank began measuring CPIA in the mid-1970s, and currently determines a CPIA value for nearly every country. Unfortunately, CPIA ratings before 2005 are not publicly available. Further, CPIA ratings are only provided for the 95 economies the World Bank classifies as being “low income.” Given that the CPIA measure is central to my analysis, I drop all projects for which no CPIA data

¹¹My use of surnames to infer cultural proximity may systematically over- or under-estimate cultural proximity for female TTLs because it is customary in many countries for married women to adopt their husband’s surname. I address this concern in Section 2.6.

¹²I also consider specifications that employ each of the six-point cluster scores individually in place of the CPIA average. These results do not differ qualitatively and are reported in Appendix A.5.

are available, essentially excluding all middle-income countries, despite the fact that several of these countries (e.g. Argentina, Brazil, China, Poland, South Africa) qualified for World Bank assistance at some point after 2005.

For projects that begin before 2005, but were evaluated for their success after 2005, I use the average CPIA rating for the country in question over all years for which data are available. Imputing CPIA data in this way is not ideal, but country-level CPIA ratings are relatively stable over time, so the average of the CPIA in available years is likely to be a reasonable approximation when data are missing.¹³

2.4.3 Outcome variable: project success. After a World Bank foreign aid project is completed or abandoned, the IEG evaluates the project on several dimensions and determines an overall outcome score, which is, like CPIA, measured on a six-point scale. I use this overall ordinal outcome score as my main dependent variable.

Though IEG evaluators strive to be objective in their assessments, some subjective judgments are unavoidable in project evaluation. Each project is evaluated against its own stated objectives, and the objectives are themselves evaluated for their relevance to the needs of the recipient country. While a universal and objective measure of project success would be preferable, no such

¹³For the 75 countries represented in my sample, CPIA data are on average available for 11.95 years (among 13 possible years, 2005-2017). The average standard deviation of CPIA within each country is 0.12 for the available years. Zimbabwe's CPIA fluctuates the most, with a standard deviation of 0.47. Excluding Zimbabwe from my sample and repeating my analysis yields qualitatively similar results.

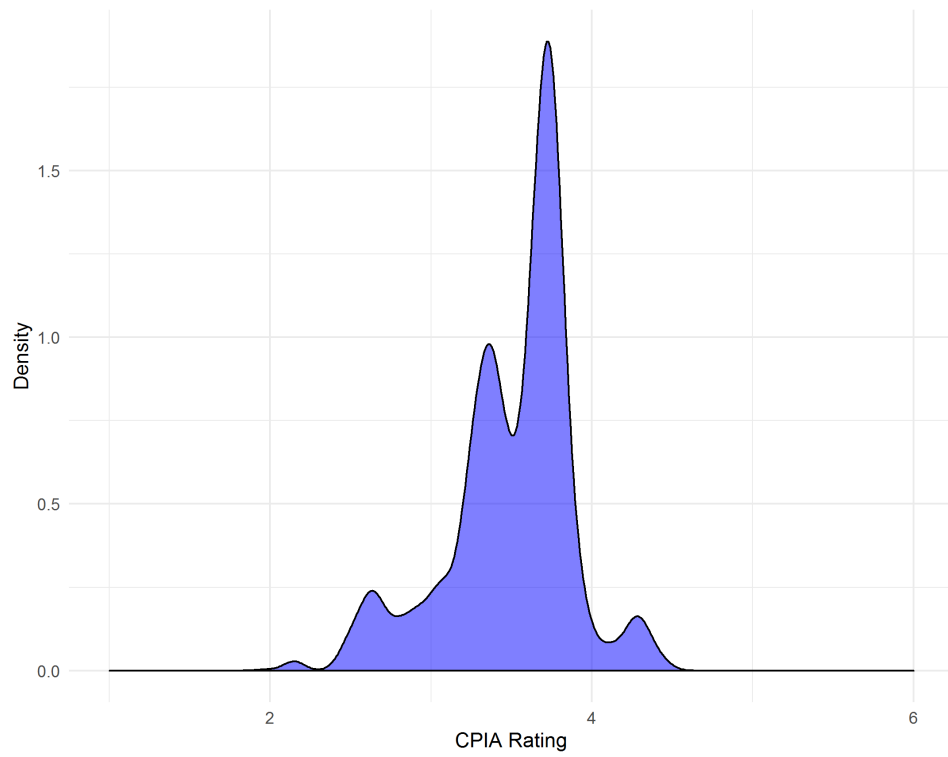


Figure 2. Distribution of sample projects' CPIA ratings

measure exists, so IEG evaluations have become a widely used proxy for project success (see Geli et al. (2014), for a concise review).¹⁴

The IEG provides data on every project it has ever evaluated, including outcome score, location, cost, year and project ID. This last variable allows me to link projects uniquely with other data provided by the World Bank, including the names of the TTLs.

The IEG updated its evaluation methodology in 2005, so I limit my sample of projects to those evaluated in 2005 or later. I also drop projects that were rated “Not Applicable” or “Not Rated,” or for which the necessary data were otherwise incomplete.

The IEG rates each project at one of six levels, which I convert to a numerical score: *highly unsatisfactory* (1), *unsatisfactory* (2), *moderately unsatisfactory* (3), *moderately satisfactory* (4), *satisfactory* (5), and *highly satisfactory* (6). Nearly half of all projects in my sample are rated as *moderately satisfactory*. The name of the evaluator is given in the IEG’s published assessment, called the Implementation Completion and Results Report Review (ICRR). This information allows me to control for evaluator fixed effects.¹⁵

2.4.4 Other controls. I also include a number of control variables in my estimation specification. For example, it is possible that some global regions simply produce better TTLs than others, or that some regions are more likely to host successful aid projects due to factors not accounted for by other variables. To control for this, I employ fixed effects for the project region and the TTL home

¹⁴In defending their decision to use IEG evaluation data, Chauvet, Collier and Duponchel (2010) write, “the evaluation procedure is independent, staff are experienced, the process has been on-going for more than three decades, and a lot of resources are put into it.”

¹⁵See Appendix A.1.2 for details about how the IEG assigns ratings.

Table 1. Summary statistics for high- and low-cultural proximity projects

	High proximity		Low proximity		Full sample	
	<i>Mean</i>	<i>St Dev</i>	<i>Mean</i>	<i>St Dev</i>	<i>Mean</i>	<i>St Dev</i>
Cultural proximity	-0.57	0.42	-2.33	0.51	-1.45	1
CPIA	3.54	0.39	3.47	0.38	3.50	0.39
IEG outcome	4.00	0.96	3.79	0.99	3.89	0.98
Approval year	2004.57	4.18	2004.85	4.38	2004.71	4.28
Evaluation year	2012.16	3.63	2011.94	3.67	2012.05	3.65
N	973		973		1946	

Notes: Summary statistics for projects divided into subsamples at median cultural proximity.

region. Project region is designated in each IEG evaluation, and TTL home region is, like my measure of cultural proximity, a weighted measure based on the global distribution of individuals who share a TTL's surname.

TTL experience may also be relevant. To capture the effect of TTL experience on project success, I include an indicator variable equal to 1 if that TTL has completed at least one other project prior to the approval date of the current project, even if that project does not appear in my sample.

I control for the cost of each project, a variable that is included in the IEG's data set. Very expensive projects may be more complicated and therefore prone to failure. Alternatively, large projects, when they succeed, may have positive effects that are especially salient to IEG reviewers, leading to higher outcome scores. In either case, cost is likely to be an important control. The marginal effect of project cost on project outcome is likely to decline as cost increases, so I include the logarithm of project cost in each regression.

Similarly, I control for the recipient country's population in logs. Countries with larger populations may pose larger challenges for foreign aid projects, but the marginal effect of population size is likely to decrease as population increases.

I calculate growth in each country’s GDP per capita (in terms of purchasing power parity, PPP) over the lifetime of each project. Rather than use the simple average of the yearly growth rates, I use the GDP per capita (PPP) in the first and last year of the project—as given by the World Bank—and the duration of the project in years to calculate the counterfactual growth rate as if this rate had been constant over the life of the project. This procedure yields growth estimates that are very similar to the simple average, but “penalizes” countries that experience uneven growth.¹⁶ I include this control variable because heterogeneous growth rates across countries are a potential source of variation in project outcomes that is partly unpredictable by the TTL or the World Bank.

Following Kaufmann, Kraay and Denizer (2013), I include in nearly every regression a set of indicator variables that captures the time interval when the project began, the time interval when it was evaluated, the project’s sector, and interactions between sector and the two timing variables. Specifically, I control for the project’s approval date, in five-year bins, to capture changes over time in the mix of projects the World Bank approves. I also control for project evaluation date, likewise in five-year bins, as standards for evaluation may have shifted over time. Additionally, because both kinds of changes may vary by sector, I interact both groups of year bins with a set of indicator variables designating project sector. A project’s sectors are specified in its IEG evaluation. I classify each project as belonging to the sector listed first in the IEG evaluation.¹⁷

¹⁶For instance, if a country’s GDP is halved one year, and then doubled the next year, the simple average of its GDP growth rates over the two years is 25% (the mean of -50% in the first year and 100% in the second). The counterfactual-constant growth rate over these two years is 0%, as the starting and ending annual GDPs are identical.

¹⁷The project sectors that appear in my data set are Agriculture and Rural Development; Economic Policy; Education; Energy and Mining; Environment; Finance; Global Information/Communications Technology; Health, Nutrition and Population; Public Sector

I also control for the length of time between a project’s conclusion and its evaluation. Most projects are evaluated between two and three years after completion, but some are evaluated much later, which may limit the information available to evaluators and add noise to the assigned outcome score. Alternatively, a longer delay between project conclusion and evaluation may bring the project’s longer-term effects into clearer focus. In either case, it may be important to control for any evaluation lag.

Finally, for the IEG rating of project success that serves as my outcome variable, two kinds of evaluations appear in my sample: ICRRs and PARs.¹⁸ Both evaluation types use the same criteria and are intended to be broadly comparable. However, PARs are based on a more thorough review of evidence related to project success, and are issued by a group rather than by a single IEG reviewer. PARs also tend to assign slightly lower outcome scores, so I include an indicator variable that distinguishes between these two evaluation types.

2.4.5 Potential endogeneity of TTL cultural proximity.

Estimation results from ordinary least squares regression cannot be interpreted as causal effects. The biggest threat to causal identification is that assignment of TTLs to projects may be endogenous. For example, if culturally close TTLs are preferentially assigned to more difficult projects, OLS will bias toward 0 the coefficients on *Cultural proximity_i* and its interaction with *Institutional quality_i*. Alternatively, culturally proximate TTLs may be assigned to particularly difficult projects in countries with low institutional quality, but they may be assigned to the easiest projects in countries with high institutional quality. Such an assignment

Governance; Social Protection; Transport; and Urban Development. Additionally, 29 project sectors are listed as “Unknown.”

¹⁸See Appendix A.1.2 for descriptions of these two evaluation types.

pattern could conceivably account for a positive coefficient the interaction between *Cultural proximity_i* and *Institutional quality_i*, even if the true causal effect of this interaction term on project quality were 0.

To address this concern, I use an instrumental variables strategy to isolate the variation in cultural proximity explained by conditions exogenous to the characteristics of the recipient country: the availability of culturally close TTLs. As an instrument, I use the average cultural proximity to that project’s recipient country of all TTLs who led other projects that are both (a) located in low-income countries and (b) evaluated *before* 2005. That is, I use the average cultural proximity of TTLs who led projects that are excluded from my estimating sample only because they were evaluated before the starting date for my sample. I refer to this set of projects as “pre-sample” projects. By construction, this instrument varies between countries but not within countries. The instrumental variable’s value is identical for every project in a given country, so the results from these instrumental variables regressions cannot be an artifact of a TTL assignment policy that simply gives the most difficult projects to culturally proximate TTLs in countries with low institutional quality and easier projects to culturally proximate TTLs in countries with high institutional quality.

Conceptually, the average cultural proximity of the earlier pre-sample TTLs to the recipient country for a given aid project should provide a good summary statistic for the availability of culturally close TTLs for projects in that country. Where culturally close TTLs are scarce, such that average cultural proximity of the earlier TTLs is low, we should expect, on average, to see a more culturally *distant* TTL assigned. The first-stage results (presented in Table 3) support this claim. To overcome endogeneity, however, instrumental variables (IVs) need to meet the usual

exclusion restriction—they need to be uncorrelated with the main regression’s error term. In practice, this means the IV should be related to the outcome variable only through the endogenous variable (or through other included exogenous regressors). My IV varies only between countries (every project in a given recipient country has the same average cultural proximity to the fixed set of pre-sample TTLs). Thus my instrument is guaranteed to eliminate any source of endogeneity arising from within-country variation, and we need be concerned only with between-country sources of endogeneity in my IV models.

There are two reasons why the instrument may fail the exclusion restriction because of between-country sources of endogeneity. I examine the plausibility of each concern.

First, an abundant supply of culturally close TTLs (relative to a given country) may result from the World Bank’s unequal involvement with various countries or cultures. The World Bank may seek out or cultivate talented individuals from favored countries or cultures. If projects in those preferred countries or cultures also receive other support that promotes project success, the IV fails the exclusion restriction. However, we have two good measures of the World Bank’s investment in individual countries—the number of projects they complete in each country and the amount they spend in each country. To determine whether the supply of TTLs that are culturally close to a country is correlated with observable measures of World Bank investment in that country, I regress the instrument on the number of projects in each recipient country since 1990, controlling for the logarithms of population and GDP per capita, both of which I include as controls in every regression. I find that this cumulative number of projects does not predict average TTL cultural distance (p-value=.89). A similar

reduced-form regression, including the total cost of all projects in each recipient country since 1990, also yields an insignificant effect of total project cost on the instrument (p-value=.71). The results for both of these regressions are reported in Appendix A.3, Table C1. These results support the contention that if the World Bank favors some countries over others, it does not express this favoritism through differential recruitment of people to serve as TTLs, so the exclusion restriction appears not to be violated in this manner.

A second concern is that cultures that are most conducive to project success may also produce a disproportionate number of qualified World Bank applicants. A culture's strong emphasis on the value of education, for example, may lead to a large pool of qualified, motivated World Bank applicants from that culture. If those same cultural traits also produce conditions which are conducive to project success, the proposed IV may fail the exclusion restriction. Fortunately, this concern also appears to be unfounded. The best available measure for conditions that may be conducive to project success is the CPIA measure, which is intended to capture explicitly the suitability of a country's institutions for development projects and poverty-reduction projects. Institutions are not the same as culture, but where culture facilitates World Bank project success, this likely occurs at least partly through the culture's influence on institutions. Again, regressing the proposed instrument on each recipient country's average CPIA, controlling for the logarithms of population and GDP per capita, I find no statistically significant relationship between CPIA and this instrument (p-value=.63). These results are also included in Table C1 of Appendix A.3.

2.5 Results

I begin by estimating the effects of cultural proximity and institutional quality on project success using OLS methods. After presenting the basic OLS results, I estimate an instrumental variables specification.

According to the conceptual model described in Section 2.3, cultural proximity should be beneficial for project outcomes unconditional on institutional quality, implying $\beta_2 > 0$ in the empirical model in equation (11). If we force $\beta_3 = 0$ by excluding the interaction term from the equation (11), as in model 1 in Table 2, we indeed find that $\beta_2 > 0$, i.e. that cultural proximity is associated with improved project outcomes, independent of institutional quality. In Table 2, model 2, we allow the effect of cultural proximity on project success to vary with institutional quality by including the appropriate interaction term in the regression. We find a significant and positive coefficient on the interaction term ($\beta_3 = 0.062$, $p < .01$). This coefficient implies that cultural proximity is *more beneficial* in countries with strong institutions than in those with weak institutions.

The effect sizes of cultural proximity and its interaction with institutional quality are moderate. At the mean institutional quality, a one-standard-deviation increase in cultural proximity implies a 0.067 increase in IEG project score on a six-point scale. This is about two-thirds as large as the effect size of institutional quality. This seems economically important, given that institutional quality has been repeatedly identified in the literature as a strong predictor of the success of foreign aid projects.

2.5.1 IV results. Table 3 presents first-stage results for the main IV specification. My analysis relies on cultural proximity and its interaction with institutional quality, so I need two instruments and two first-stage equations, one to

Table 2. OLS results, selected coefficients

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.069** (0.033)	0.068** (0.033)	0.063* (0.036)
Institutional quality	0.096*** (0.030)	0.192*** (0.049)	0.198*** (0.050)
Cultural proximity \times Institutional quality		0.062** (0.026)	0.055** (0.027)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes:

*** < .01, ** < .05, * < .1

Project success is IEG's evaluation of the project on a 1-6 scale. Institutional quality is CPIA standardized to mean = 0, sd = 1. Cultural proximity is re-scaled such that sd = 1. Standard errors are clustered by country-evaluation year. Each model includes controls for project cost, TTL experience, project region, probability-weighted average of TTL's likely ancestral region, recipient country GDP per capita (log) and GDP growth, recipient country population (log), project sector (as defined by the World Bank), project approval year, project evaluation year, and the length of time between the end of the project and its evaluation.

explain *cultural proximity* and one to explain *cultural proximity* \times *institutional quality*. My second instrument, then, is just my first instrument (*Pre-sample average TTL cultural proximity*) interacted with *Institutional quality*. Other than the two instruments, all first stages include the same exogenous controls as their second stage counterparts.

Table 3. IV: First stage regressions (selected coefficients)

	Cultural proximity	Cultural proximity \times Institutional quality
	(1)	(2)
Pre-sample TTL cultural proximity	0.455*** (0.034)	-0.028 (0.045)
Institutional quality	-0.083 (0.053)	0.267*** (0.094)
Pre-sample TTL cultural proximity \times Institutional quality	-0.041*** (0.015)	0.550*** (0.030)
Other controls	Yes	Yes
Observations	1,946	1,946
R ²	0.597	0.838
Joint <i>F</i> -stat for IVs	93.9***	221.8***

Notes:

*** $< .01$, ** $< .05$, * $< .1$

For a summary of additional control variables included in all models, see notes to Table 2.

The *F* statistics for the joint significance of the pair of instruments in the first-stage regressions are 93.9 ($p < .001$) and 221.8 ($p < .001$) in models (1) and (2), respectively, indicating that the instruments are highly relevant. Furthermore, the instruments function in the manner expected: *pre-sample TTL cultural proximity* is an excellent proxy for the actual *cultural proximity* between

TTL and recipient country and *pre-sample TTL cultural proximity* \times *institutional quality* is highly predictive of *cultural proximity* \times *institutional quality*, and both of the key coefficients are positive as expected.

Table 4 presents results for the second-stage instrumental variable regressions. Model 2 gives the IV results for my main specification. The estimated effect of the interaction term is almost twice the magnitude estimated by OLS (model 3 in Table 2). The estimated coefficient on *cultural proximity*, representing the marginal effect of cultural proximity at the mean institutional quality (since CPIA is standardized), is also much larger, but is significant only at the 10% level. Why might the IV regression produce a larger estimate for the effect of cultural proximity? If the World Bank wants to ensure that projects meet a minimum level of success and operates under the impression that culturally close TTLs have an advantage in producing high-quality projects, then assigning culturally close leaders to difficult projects, no matter the recipient country's institutional quality, would be a sensible policy. This assignment strategy would bias downwards—toward zero—the OLS coefficient on *cultural proximity*, since culturally proximate TTLs would be assigned to more difficult projects. If some components of project difficulty are unobservable to the econometrician (but observable to World Bank staff), and therefore cannot be controlled for, then the positive effect of *cultural proximity* will be partially offset by endogenous TTL assignment. As a result, the OLS estimate for *cultural proximity* \times *institutional quality* will also be biased toward zero. The increased magnitudes of the IV coefficient estimates, relative to the OLS estimates, are not surprising because the IV estimates should better reflect the true causal effects of *cultural proximity* and *cultural proximity* \times *institutional quality*.

quality in the absence of an endogenous assignment strategy that would be expected to attenuate OLS estimates.

Model 3 includes fixed effects for the 230 IEG reviewers who rated the success of projects in my sample. As noted above, projects are rated according to a standardized set of criteria, but there remains an element of subjectivity in IEG outcome scores. Some reviewers may be “tougher” than others, and if toughness is correlated with institutional quality or cultural proximity, estimates will be biased. Model 3 provides evidence which suggests that this concern is unfounded.¹⁹

Table 4. IV results, selected coefficients

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185 (0.122)	0.205* (0.121)	0.185 (0.123)
Institutional quality	0.095*** (0.030)	0.261*** (0.074)	0.313*** (0.078)
Cultural proximity × Institutional quality		0.107** (0.046)	0.130** (0.048)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

For a summary of additional control variables included in all models, see notes to Table 2.

¹⁹If a project’s rating comes from a more rigorous Project Performance Assessment Report (PAR), which is conducted by a group rather than an individual, I categorize the reviewer as $PAR = 1$ so that all such projects are given the same fixed effect. Fifteen observations had no reviewer listed, so I assign these observations the same fixed effect. Additionally, 72 observations were evaluated by reviewers who reviewed no other projects in my sample. These 72 observations are assigned a single fixed effect.

Table 5 demonstrates that the results do not change qualitatively if I use an alternative measures of institutional quality, the World Governance Indicators (WGI) (Kaufmann, Kraay, & Mastruzzi, 2011). The WGI is a set of six separate measures of institutional quality: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. Crucially, the WGI is not specifically designed to measure the suitability of a country for aid projects, and the World Bank does not use the WGI to allocate resources. Table 5 is analogous to Table 4, but uses the average of the six WGI indicators (standardized to $\mu = 0$, $\sigma = 1$) in place of the CPIA as the measure of institutional quality. Using WGI in place of CPIA yields comparable results for the interaction term *Cultural proximity* \times *Institutional quality*, but the effect of *Cultural proximity* at the mean WGI is not statistically significant, even at the 10% level.

2.6 The possibility of mismeasurement for female TTLs

Given that my measure of cultural proximity is based on TTL surnames, cultural proximity for female TTLs may be systematically mismeasured. Specifically, female TTLs may take a new surname at marriage, and their new surname may not be representative of their cultural background. In this section I provide evidence, using a second measure of cultural proximity based on TTLs' given names, that this possibility is unlikely to influence my findings significantly and report the key coefficients of my main IV specification with controls for gender.

To test empirically the possibility that female TTLs' surnames are less informative than those of male TTLs, I reconstruct my cultural proximity measure using *given names* instead of surnames and compare this new variable to my original measure of cultural proximity. This new measure of cultural proximity

Table 5. IV results, using WGI in place of CPIA

	IEG Outcome: 1-6 scale		
	(1)	(2)	(3)
Cultural proximity	0.183 (0.124)	0.146 (0.123)	0.117 (0.125)
Inst. quality (WGI)	0.095*** (0.026)	0.259*** (0.064)	0.301*** (0.067)
Cultural proximity \times Inst. quality (WGI)		0.108*** (0.040)	0.130*** (0.042)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** $< .01$, ** $< .05$, * $< .1$

WGI is standardized to $\mu = 0$, $\sigma = 1$. For a summary of additional control variables included in all models, see notes to Table 2.

Table 6. Correlation between measures of cultural proximity based on given name and surname

	Full sample	Men only	Women only
Correlation	0.2630	0.2647	0.2606
95% CI	(0.2165, 0.3084)	(0.20756, 0.3201)	(0.1792, 0.3385)
N	1578	1051	527

Notes: Correlation is Pearson's product-moment correlation. The total number of observations does not equal 1946 because instances in which a TTL oversees multiple aid projects in the same country yield duplicate observations, which are removed. Additionally, observations are removed for TTLs who use only a first initial because a first initial is likely to be wholly uninformative about cultural background.

is constructed in precisely the same way as the original measure, but I draw the global distribution of each TTL's given name from Forebears.io to determine a set of probable home countries for each TTL.²⁰ We should expect these two measures of cultural proximity to be correlated for both male and female TTLs. If the correlation between these two measures of cultural proximity is significantly higher for male TTLs than for female TTLs, then we have evidence that female TTLs' surnames may yield a mismeasurement of cultural proximity. Specifically, this analysis relies on the following assumptions:

1. Given names are somewhat informative about cultural background.
2. Given names are equally as informative in this respect for women as for men.
3. Individuals—men or women—do not change their given names upon marriage.

TTLs' genders are unknown, so I use their given names to infer gender. I processed each given name using Genderize.io, a database of the gender rates for more than 200,000 given names, which predicted gender for most TTLs.²¹

Some given names could not be matched by the Genderize.io database. For these remaining 293 TTLs who have unusual given names, gender-neutral names, or only a first initial, I use publicly available online sources to determine the gender of that particular World Bank employee. For each TTL gender determined in this way, I recorded the URL of the source.

As reported in Table 6, the correlation between cultural proximity calculated based on each TTL's given name and cultural proximity calculated based on each

²⁰See section 2.4.1 for details on the construction of my main cultural proximity measure.

²¹See Topaz and Sen (2016), Greenberg and Mollick (2015), and Mohammadi and Shafi (2018) for other instances of research using Genderize.io to infer subject gender.

TTL's surname is positive and significantly different from 0 for both the male and female subsamples. Furthermore, these correlations are not significantly different from each other, suggesting that cultural proximity is not mismeasured for female TTLs relative to male TTLs.²² These results are perhaps unsurprising given the pervasiveness of assortative mating, or the tendency of individuals to seek partners similar to themselves. Individuals tend to marry others with similar education levels (Mare, 1991), religion (McClendon, 2016), and personality traits (Glicksohn & Golan, 2001). Kalmijn (1994), in a study of recently married husbands and wives in the 1970 and 1980 US censuses, finds that “[t]he strong degree of assortative mating by cultural status, as measured by occupational schooling, suggests that the tendency to seek cultural similarity plays a central role in the selection process.”

2.7 Conclusion

This study examines the relationships between the cultural proximity between World Bank TTLs and foreign aid recipient countries, the institutional quality of recipient countries, and the success of foreign aid projects. Examining these relationships poses two problems. First, there is no existing measure of cultural proximity between TTLs and recipient countries. Second, the assignment of TTLs to recipient countries is likely to be endogenous. I overcome the first of these challenges by constructing a plausible measure of cultural proximity from an existing measure of genetic distance between countries and the global distribution of surnames. I overcome the second of these challenges by instrumenting for cultural proximity with a measure of the availability of culturally close TTLs.

I find evidence of a positive causal relationship between the cultural proximity of foreign aid project leaders and their projects' outcomes. Furthermore,

²²Appendix A.4 reports coefficient estimates from my main IV specification with an indicator variable for female TTLs interacted with the coefficients of interest.

this relationship appears to be significantly stronger in countries with strong institutions. These findings imply that multilateral aid organizations should give special attention to the cultural background of project leaders they assign in countries with high institutional capacities.

CHAPTER III

WILLINGNESS TO PAY FOR PANDEMIC RISK REDUCTIONS: LOST LIVES VERSUS LOST LIVELIHOODS

The survey described in this chapter was developed primarily by Trudy Ann Cameron and me, with assistance from Garrett Stanford and Shan Zhang. Dr. Cameron handled the dynamic aspects of the survey, and we collaborated to determine the correct econometric specification. I conducted the analysis and wrote the paper with editorial assistance from Dr. Cameron.

In the U.S., the generosity of supplementary federal unemployment insurance (UI) was a controversial issue throughout the early part of the COVID-19 pandemic. The debate focused mostly on worries about economic disincentives for workers. However, federal UI may also have undermined support for local-level pandemic mitigation strategies. We quantify the effect of federal UI on the trade-offs that individuals are willing to make with respect to county-level pandemic policies. We use choice experiments from an online survey, and both model, and correct for, systematic response/non-response propensities. When respondents are asked to assume that federal UI will be zero, they tend to be averse to losses in average household income but favorably disposed toward increased unemployment. With positive federal UI payments, however, respondents become more willing to accept losses in average household income but view increased unemployment less favorably. The reversal with respect to losses in average household income is driven by younger, white, non-college and lower-income respondents. The reversal with respect to unemployment is driven by middle-aged and conservative respondents. Our findings demonstrate policy-relevant heterogeneity in support for county-level health policies as a function of national-level social safety net policy.

3.1 Introduction

The U.S. suffered major public health disruptions during the COVID-19 pandemic. Future episodes of pandemic disease are likely to be a continuing threat, especially as humans continue to encroach on wildlife habitats (Grange et al., 2021; Wilkinson, Marshall, French, & Hayman, 2018). While the specific risks are unknown, future pandemics will again force societies to choose the extent to which they are willing to sacrifice economic and social activity to protect or improve public health.

During the COVID-19 pandemic, the U.S. government delegated most policy decisions about non-pharmaceutical interventions (NPIs)—such as mask-wearing and social-distancing—to state, county and local-level authorities. The federal government, however, made important national-level decisions to enhance the overall social safety net, primarily by supplementing state-level unemployment insurance benefits with additional federal unemployment insurance (UI) benefits. When pandemic policies are instituted at different levels of government, however, it is important to consider the potential interactions between these policies.

Using a survey-based choice experiment with nearly 1,000 U.S. respondents in California, Oregon, and Washington State, we focus on how the generosity of the federal-level social safety net—namely, the availability of supplementary federal UI—shapes preferences for pandemic mitigation policies implemented at the county level.¹ In the absence of federal UI, respondents are significantly *averse to* net losses in county-level average household incomes, but they are *in favor of* increases in unemployment in their county caused by county-level pandemic mitigation

¹Best practices for survey-based stated-preference research are reviewed in Johnston et al. (2017). The federal UI program examined in this study is more formally known as Federal Pandemic Unemployment Compensation.

policies. When federal UI is provided, however, respondents are *in favor of* net losses in county-level average household incomes, on average. They also become *indifferent to* increases in county-level unemployment, or even *opposed to* increases in county-level unemployment if we focus on the actual levels of federal UI that had most recently been experienced at the time of our survey. Thus, paradoxically, the presence of generous federal UI may have undermined support for local pandemic mitigation policies that could reasonably be expected to increase unemployment. Our results are consistent with the more-general notion that the federal policy environment can potentially affect public support for county-level policies more broadly.

One plausible explanation for this federal UI-induced change in county-level policy preferences might be a relatively widespread fear that unemployment benefits create perverse economic incentives, especially if federal UI, in numerous cases, exceeds the lost wages it is meant to replace. This view has had vocal adherents in the U.S. Congress. Senator Ron Johnson (R-WI) called the original \$600-per-week unemployment benefits a “perverse incentive to keep people out of the economy,” citing the proportion of recipients who earned more from unemployment insurance than they lost in wages (UPFRONT, 2020). House Minority Leader Kevin McCarthy (R-CA) made similar remarks in July 2020, saying, “We made a mistake when we overpaid on unemployment insurance where now it’s hard for people to come back to work because they’re making more on unemployment than they can working” (Stein & Werner, 2020). Indeed, concerns about the disincentive effects of the continuation of \$300-per-week federal UI benefits became even more apparent in May of 2021, as reported by Sainato (2021) and Zeleny and Luhby (2021). Republican-led states began seeking to end

these benefits earlier than the Biden administration had planned, citing workforce shortages.

Early evidence, such as that reported in Bartik, Bertrand, Lin, Rothstein, and Unrath (2020), Marinescu, Skandalis, and Zhao (2021) and Dube (2021), contrasts with the claims of legislators Johnson and McCarthy by suggesting that high levels of federal UI during the first wave of the pandemic did not contribute to increased unemployment. Political commentator Sean Hannity raised a different objection in an on-air interview with then-Treasury Secretary Steve Mnuchin in March 2020, citing a widespread feeling of aggrievement, rather than the threat of distortionary incentives. Apparently speaking for his 3+ million *Hannity* viewers, the host explained, “This idea that you’re going to make more money unemployed, that angers my audience. That angers me too. Why couldn’t somebody just have to show a pay stub, and that’s the money you’re going to get?” (Concha, 2020). Hannity argues, *without referencing economic disincentives*, that a segment of the US population prefers that those who experience unemployment not be compensated beyond their previous earnings.

We find evidence consistent with Hannity’s claim, especially for respondents with lower incomes and those who are politically conservative or moderate, and we show that these preferences have a significant effect on support for pandemic mitigation policies. We hypothesize that respondents infer that increased unemployment reduces social contact, limiting the spread of COVID-19 and making increased unemployment a desirable feature of a pandemic mitigation policy. In the presence of federal UI, however, this effect is offset by concerns about economic incentives or preferences over distributional fairness, and increased unemployment is, on net, an insignificant or even negative factor in respondents’

decision-making. Conversely, larger decreases in county average household incomes due to unemployment—expected to be an unambiguous economic “bad”—may signal to concerned respondents that unemployment is appropriately burdensome, even with federal UI payments.²

A growing body of survey-based choice experiments has already revealed some of the tradeoffs people are willing to make with respect to COVID-19 pandemic policies. Some of the key features of these studies are summarized in Table 7. Four studies were fielded during the so-called “first wave” of the pandemic—a Dutch study (Chorus, Sandorf, & Mouter, 2020), a French study (Blayac et al., 2021), a study for the entire U.S. (Reed, Gonzalez, & Johnson, 2020), and a study just in the state of Missouri (Wilson et al., 2020).³ These surveys may have been fielded too early to capture the effects of federal-level policies on local preferences. The actual dates for a fifth study (described in Genie et al. (2020) and introduced as a protocol for an upcoming survey) seem not yet to have been published. Our survey was fielded between January 13 and February 16, 2021, in the latter part of the so-called “third wave” of the pandemic. During this period, future levels for U.S. Federal Pandemic Unemployment Compensation were still uncertain, allowing us to vary the levels of federal UI in our choice scenarios without straining credulity.⁴

²When federal UI is not present, we find that respondents are on average strongly averse to policies that reduce average household income in their counties, as expected.

³For the very early Reed et al. (2020) survey, fielded between May 9 and May 20 of 2020, the authors warn that “If we had taken the time to follow standard, good-practice procedures, we almost certainly would have ended up with a different instrument.”

⁴More-detailed summaries of these closely related studies are included in Appendix B.1, where we also describe a pre-COVID-19 study by Cook, Zhao, Chen, and Finkelstein (2018) in Singapore in the wake of the previous 2003 SARS-CoV and 2009 H1N1 influenza outbreaks.

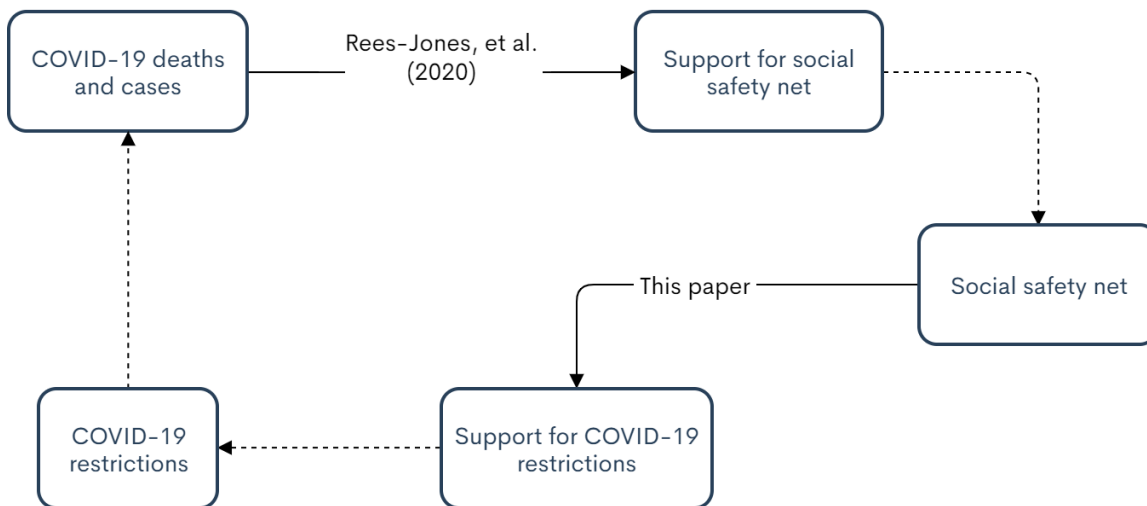
Table 7. Other Covid-19 choice-experiment studies

	Chorus et al. (2020)	Blayac et al. (2021)	Reed et al. (2020)	Wilson et al. (2020)	Genie et al. (2021)
Health attributes	Number of deaths, physical injuries, mental injuries	None	Cases	Risk of infection	Excess deaths, infections, postponed non-pandemic medical care
Economic attributes	Lost income, taxes	Financial compensation	Percent below poverty line, time until economy recovers	Lost income	Job losses, ability to buy things
Other attributes	Educational disadvantages, healthcare worker stress	Duration, masks, restrictions on transportation, vacations, and bars/restaurants	Restrictions on non-essential business	Restrictions on gatherings, social venues, and schools	Policy duration, severity of lockdown (four tiers)
Method	Latent class (3 classes)	Heterogeneity by age, gender vulnerability	Latent class (4 classes)	Latent class (4 classes)	Mixed logit; heterogeneity by 5 core “moral foundations”
Region	Netherlands	France	whole U.S.	Missouri only	United Kingdom
N	1009	1154	5953	2428	4021

Our work also complements Rees-Jones et al. (2020), who document increased support for federal safety-net programs, including federal UI, in response to the severity of COVID-19’s negative effects on local (county-level) health and employment. We identify a related link in the causal chain: a strengthened federal social safety net reduces support for county-level NPI pandemic policies that (a) lead to unemployment and thereby (b) increase take-up rates for the federal UI that provides this social safety net. Figure 3 illustrates the different emphases of Rees-Jones et al. (2020) and the present paper.

Finally, our work contributes to the literature examining the relation between political partisanship and attitudes toward NPI policies and associated behaviors. Reed et al. (2020) note that on simple questions of concern or support, respondents seem to answer ideologically. In the context of their choice experiments, however, ideology “plays a more complicated role.” They find that

Figure 3. Complementarity with Rees-Jones et al. (2020) study



the preferences of self-identified Republicans and Democrats are more similar to each other than to the preferences of Independents. In related work, Allcott et al. (2020) use human mobility data (SafeGraph cell-phone GPS information) to show that people in areas with more Republicans are less likely to practice social-distancing, and Kahane (2021) uses survey data at the county level to detect lower mask-wearing in counties where the Republican candidate was strongly supported in the 2016 Presidential election.

In contrast to other pandemic policy choice experiments in the literature, we invited respondents to consider different baseline pandemic conditions with varied levels of expected cases and deaths, both without and with each pandemic policy in place. Furthermore, since state-level Health Authorities have tended to publish daily statistics on cases and deaths for each county, we express the numbers of cases and deaths in our choice scenarios in absolute numbers, in comparison to the population of the respondent’s own county.⁵ We also express the economic costs

⁵This required regular updating of actual cases and deaths for each county, based on daily county-level cases and deaths at <https://usafacts.org/visualizations/coronavirus-covid-19-spread->

of each pandemic policy in terms of both the expected unemployment rate in the respondent’s county and the average number of dollars lost per household. For this paper, one key policy attribute is the presence of federal UI, which drives a wedge between these two types of pandemic cost measures. Finally, we break out ten specific categories of restrictions to permit our respondents the opportunity to differentiate among policies that are relatively more or less restrictive for different types of activities.⁶

Our study also differs from earlier pandemic policy choice experiments in its attention to the possibility of systematic selection of potential survey respondents into the estimating sample. In the absence of sufficient lead time for researchers to develop, submit, and hear back about proposals for significant research funding from the usual sources, most existing studies have used modestly priced sampling strategies. Convenience samples can be inexpensive, but they present a significant risk of being systematically selected. Earlier survey-based studies of COVID-19 policy preferences have commented on the representativeness of their estimating samples, but none has proposed a strategy to correct for systematic selection into the estimating sample. In contrast, we model and correct for selection bias, in a systematic fashion.

3.2 Data

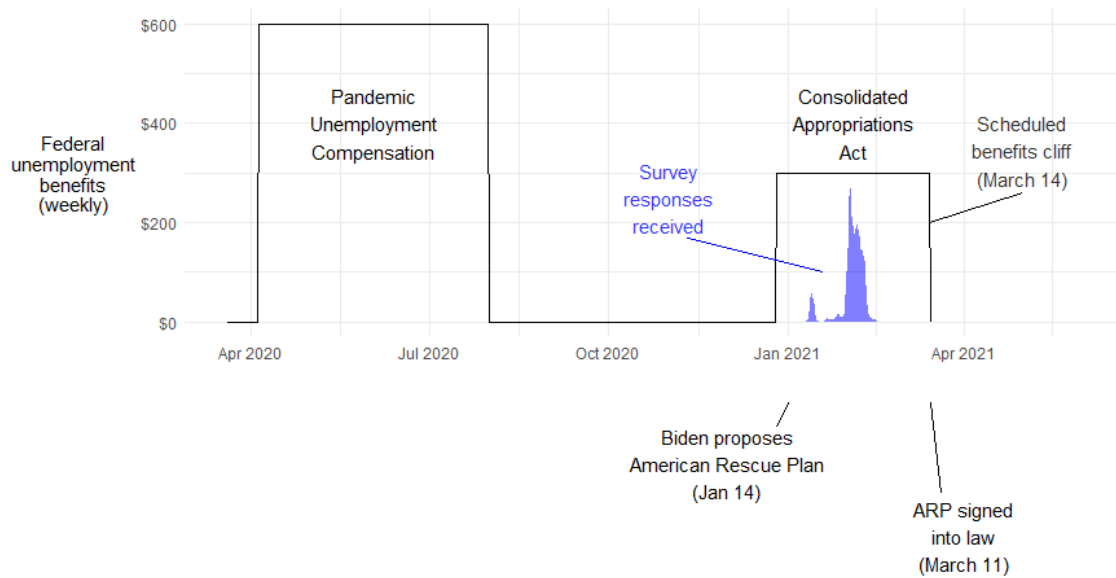
We use a survey instrument distributed via Qualtrics to residents of the west-coast U.S. states of California, Oregon and Washington during January 13–February 16, 2021. This timing is important for the present study because it

map/, and javascript code within the survey software to take randomized mixes of cases and deaths in our externally survey design, normalized for a population of 50,000, and scale these to the population of the county selected by each anonymous respondent.

⁶Preferences over these ten categories of restrictions will be the focus of a future paper. Here, these differences in restrictions are included merely as controls.

coincides with a period of uncertainty about the future level of supplementary federal UI benefits. Figure 4 describes the federal policy context under which respondents participated in our choice experiments.⁷

Figure 4. Timeline of survey responses relative to the American Rescue Plan of 2021



We over-sampled the states of Oregon and Washington, relative to California, because of California’s much larger population (and compensate by weighting the resulting subsamples during estimation). We requested quotas that matched the marginal distributions within each state of gender, age (18 to 34 years, 35 to 54 years, and 55 years and older), race (White, Black, Asian and other), and income (less than \$50,000, \$50,000 to \$99,000 and \$100,000 or more). We also asked about the respondent’s zip code, with the explanation that we would need

⁷The process of survey development is outlined in Appendix B.2. Our survey was approved under University of Oregon IRB Protocol Number 07022020.002. A codebook for the survey (including the wording of each question in generic format) may be viewed at pages.uoregon.edu/cameron/UE_COVID_survey.pdf (172 pp.)

to assess whether eligible participants were broadly geographically representative. We asked these screening questions at the very beginning of the survey to prevent wasted effort by respondents who turn out to be ineligible (due to quotas for their categories having already been met). Respondents who were unwilling, for any reason, to provide data on their age or race bracket were automatically excused from the study.⁸

3.2.1 Sample selection and response propensities. If a potential respondent was confirmed to be eligible to participate in the survey, they were then introduced to the topic of the survey during the preamble to the formal consent-to-participate question. While our 70 percent response rate is respectable, a critically important feature of our study is our ability to formally model the individual decisions of 993 of our 1,412 eligible respondents to complete our particular survey. In addition to the sociodemographic screening questions described above, the survey platform passively collects information about the operating system upon which the respondent began taking the survey, and the date and time when they initiated their session.

The zip code data are especially valuable because they allow us to merge our sample with external data at the zip code level for all eligible panelists, and to associate with each eligible non-respondent some county-level variables. Non-respondents who opted out of the survey after the screening process provided no information concerning the pandemic policy preferences that are the topic of our survey study. However, the screening questions for eligibility allow us to build a wide array of variables that explain systematic differences in propensities to

⁸A reluctance to divulge such screening information could be systematically related to concerns about privacy in the general population, but regular participants in consumer surveys are presumably accustomed to providing this basic information.

respond to a survey about COVID-19 pandemic policies. Especially useful is our ability to link every eligible respondent to prevailing rates of COVID-19 cases and deaths in their *own* county over the duration of the pandemic leading up to our survey.

Candidate explanatory variables for our selection model include the sets of category indicators for individual screening sociodemographics as well as the device type and timing of the session. We also assemble a large inventory of potential explanatory variables at the zip code or county level for our sample-selection model. These candidate explanatory variables include zip code or county proportions for age groups, income groups, racial and ethnic groups, industries, rural/urban mix, ownership of computers and type of internet access. Relating to the local area's history with the current pandemic, we include county-level COVID-19 cases and deaths, by month, since the beginning of the pandemic. These variables include days since the first COVID-19 death in the potential respondent's county, and cases and deaths in the potential respondent's own county in the four weeks leading up to the date when they started the survey, which were duly quoted to actual respondents in the body of the survey. The variables that are retained in the final binary-outcome specification for selection are identified by LASSO methods. Appendix B.3 reproduces the descriptive statistics for these retained variables and the parameter estimates for our selection model.

Here, we note only that there is considerable heterogeneity in the predicted response propensities from our selection model. In Figure 5 the distribution shown with a black outline describes the de-meaned fitted response propensities for the entire set of 1,412 eligible survey subjects. Figure 5 also shows the separate distributions of this variable for non-respondents and respondents. As expected,

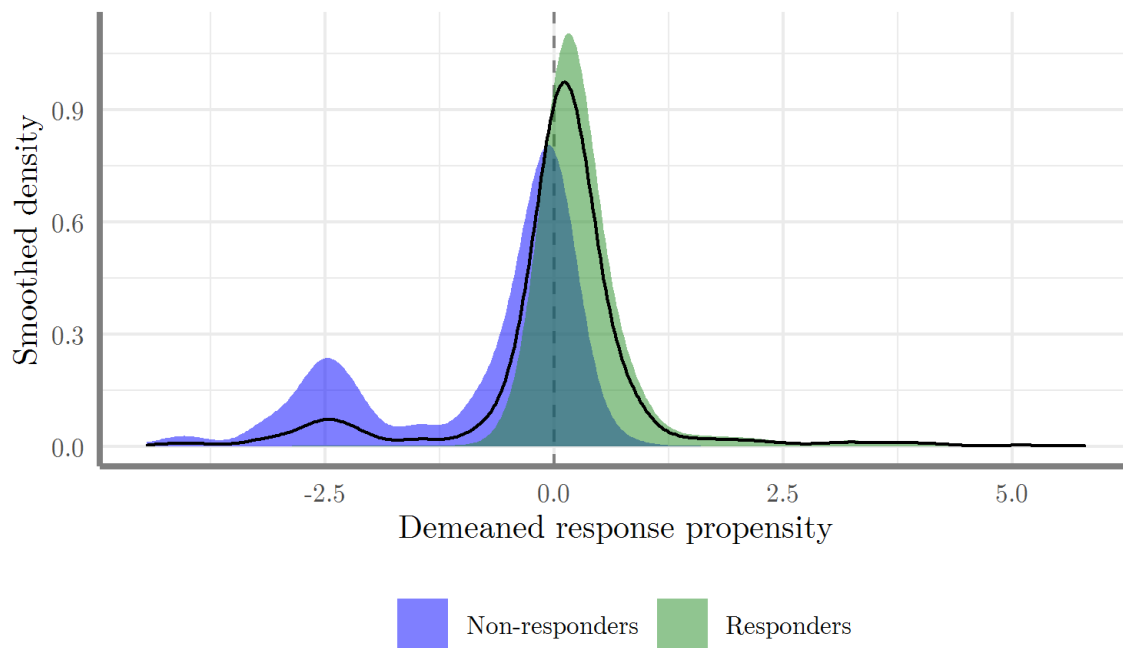
response propensities are lower for people who do not complete the survey and higher for those who do, but there is considerable overlap in the two distributions. We allow all of the estimated parameters in our subsequent policy-preference models to vary systematically with the de-meaned fitted response propensities for the estimating sample (consisting of the 993 actual respondents). Then we can simulate the parameter estimates that would obtain had everyone’s de-meaned response propensity been exactly zero—that is, if everyone in the estimating sample shared the mean response propensity in the eligible population. Selection-correction, in general, seeks to identify preferences under counterfactual conditions where everyone from the population of interest is equally likely to show up in the estimating sample.⁹

3.2.2 Estimating sample for policy choice models. Our estimating sample for policy choice modeling in this analysis consists of a total of 1,986 choices by 993 respondents. Each policy scenario describes the baseline numbers of cases and deaths in the absence of any policy, the reductions in cases and deaths that would occur under each offered policy, the economic consequences of each policy for the respondent’s county in terms of unemployment rates and the resulting average cost per household, and a set of restrictions (described in terms of levels 0 through 3) on each of ten different categories of activities or businesses.¹⁰ Each county-level policy-choice scenario was presented in the context of a specific

⁹We acknowledge, of course, that this two-step correction process utilizes estimated quantities as interaction terms in the second step without correcting the parameter standard errors for the extra noise that the first stage introduces into the model.

¹⁰D-efficient designs were employed for the ten sets of restrictions on activities and businesses, albeit in two batches of five because a single design involving ten attributes with four levels each proved too time-consuming to generate at the scale required for our project. Initial attempts permitted the D-efficient design algorithm to process for several days without completion.

Figure 5. Fitted response propensities



LASSO model; n responders = 993; n non-responders = 419. Black line gives density for full sample (n = 1412).

level of federal UI payments that would be available whether or not the county-level policy was implemented.¹¹

Summary statistics for the key attributes of policy scenarios for this paper are presented in Table 8. The distributions of cases and premature deaths avoided were loosely conditioned on the severity of these restrictions, as were the average household costs and unemployment, to preserve plausibility of the policy scenarios.¹²

In this paper we present estimates based on just the first two dichotomous choices made by each respondent, since binary-choice formats are more incentive-compatible (Carson & Groves, 2007). Future work on choice experiment methodology will take advantage of the subsequent three-way and conditional two-way choices.¹³

¹¹Policies were also described as remaining in effect for a specified duration. These durations were randomly assigned to be one or two months (40% each) or three months (20%). After standardizing total cases and total deaths with and without each policy to their per-month equivalents for our analysis, policy duration has no statistically discernible effect on respondents' propensities to choose any policy over the status quo, so we do not include policy duration in any of the models reported in this paper.

¹²Randomizations were generated outside the survey and one instance from the inventory was dynamically selected at random, without replacement, for each respondent. The externally randomized attributes were scaled to a standardized county population of 50,000 and then scaled "on the fly" during the survey to match the population of the respondent's own county. This level of aggregation seemed appropriate because much public data on actual cases and deaths has been quoted at the county level. Extensive details about the randomized design of our policy choice scenarios are contained in the Online Supplementary Materials, available at http://pages.uoregon.edu/cameron/UO_COVID_description_of_randomizations.pdf.

¹³For each respondent, six different NPI pandemic policies (A–F) are described, individually or in pairs, with "No restrictions" (N) always included as the status quo alternative, where everyone merely takes whatever precautions they think are appropriate. Respondents' policy choice scenarios were presented first as two consecutive dichotomous choices: AN, BN. Only these two choices are employed in this paper. Policies C–F were combined into three-alternative choices with conditional pairwise follow-up choices, so we save their more-complex analysis for future research. One image of a choice set is provided in Appendix B.4, along with an example of the pop-up reminders about how to interpret the indicated level of restrictions on each activity or business.

Table 8. Descriptive statistics for featured variables in our randomized design

	Mean	SD	Min	Max
Federal policy context (does not vary by county-level policy):				
Federal UI (\$/week)	193.15	138.9	0	400
Federal UI = 0	0.20			
Federal UI = 100	0.21			
Federal UI = 200	0.20			
Federal UI = 300	0.21			
Federal UI = 400	0.17			
Without policy (status quo):				
Unempl rate, federal UI = 0	4.34	1.65	2.3	10.4
Unempl rate, federal UI > 0	4.26	1.87	2.1	19.8
Avg. hhld cost/mo	0	0	0	0
Absolute '00s cases/mo/50,000	9.74	3.31	3.08	22.2
Absolute deaths/mo/50,000	12.75	5.30	0	37.25
With policy:				
Unempl rate, federal UI = 0	14.96	4.13	6.00	30.10
Unempl rate, federal UI > 0	14.81	4.03	5.60	32.20
Avg. hhld cost/mo, federal UI = 0	355.	144.	60.0	930.
Avg. hhld cost/mo, federal UI > 0	238.	137.	5.00	895.
Absolute '00s cases/mo/50,000	4.85	2.42	0.003	17.6
Absolute deaths/mo/50,000	6.25	3.56	0	27.42
Respondents	993			
Policies (choices)	1986			

Notes: We rely upon the designed-in independent variation in federal UI payments, unemployment rates, and average household costs per county. The joint distribution for the federal policy context variable and the two key attributes of the county-level policies is summarized in Appendix B.5, Figure E1. Cases and deaths are reported in terms of absolute levels of cases under each alternative. These rates are summarized per 50,000 people because 50,000 is roughly the average population of counties in our three-state region. In the survey itself, for each respondent, cases and deaths were scaled to the corresponding numbers for their own county.

3.2.3 Estimating specification. Conditional logit choice models based on random utility models (RUM) are by now familiar and algorithms for estimation are readily available in most standard econometric software packages. Thus we will not repeat here the basics of the RUM approach to preference estimation. The key part of the specification is the particular functional form used for indirect utility function under each alternative. The respondent is assumed to choose the alternative that yields the highest level of utility, up to an error term that is unobservable by the researcher.

Our model is based on the indirect utility of respondent i under policy A . In contrast to the usual specification in a random utility model, however, the cost of the policy is expressed in terms of its effect on average household incomes in the respondent's county—namely, the *social* cost, as opposed to the *private* cost specifically to the respondent's own household. Our specification is additively separable in the respondent's income and in average household costs for the policy in the respondent's county:

$$\begin{aligned}
V_i^A &= \beta_0 Y_i + \beta_1 avcost_i^A + \beta_2 unempl_i^A + \beta_3 cases_i^A + \beta_4 deaths_i^A & (3.1) \\
&+ \sum_{k=5}^{14} (\beta_k restr_{ki}^A) + \beta_{15} SQ^A + \gamma_0 Y_i \times \hat{RP}_i + \gamma_1 avcost_i^A \times \hat{RP}_i \\
&+ \gamma_2 unempl_i^A \times \hat{RP}_i + \gamma_3 cases_i^A \times \hat{RP}_i + \gamma_4 deaths_i^A \times \hat{RP}_i \\
&+ \sum_{k=5}^{14} (\gamma_k restr_{ki}^A \times \hat{RP}_i) + \gamma_{15} SQ^A \times \hat{RP}_i + \varepsilon_i^A,
\end{aligned}$$

where the status quo indicator, SQ^A is zero for policy A and \hat{RP}_i is the individual's de-meanded fitted response propensity from our preliminary selection model.

Indirect utility of respondent i under the status quo, V_i^N , involves no policy and therefore no decrease in average county-level incomes (so that

$avcost_i^N = 0$), only baseline unemployment in the respondent's county and no extra unemployment related to Policy A (so that $unempl_i^N \neq 0$), and no pandemic restrictions (thus $restr_{ki}^N = 0$). This utility is given by:

$$\begin{aligned} V_i^N &= \beta_0 Y_i + \beta_2 unempl_i^N + \beta_3 cases_i^N + \beta_4 deaths_i^N + \beta_{15} SQ^N \\ &+ \gamma_0 Y_i \times \hat{RP}_i + \gamma_2 unempl_i^N \times \hat{RP}_i + \gamma_3 cases_i^N \times \hat{RP}_i \\ &+ \gamma_4 deaths_i^N \times \hat{RP}_i + \gamma_{15} SQ^N \times \hat{RP}_i + \varepsilon_i^N, \end{aligned} \quad (3.2)$$

where $SQ^N = 1$ for the status quo alternative.

Each respondent's choice between policy A and the status quo is determined by whether policy A yields greater utility. Let $\Delta V^A = V^A - V^N$ be the difference in indirect utilities from policy A and the status quo option N, so that policy A is chosen if and only if $V^A \geq V^N$ or $\Delta V^A > 0$. The individual's baseline level of income drops out, so our basic econometric specification is as follows.¹⁴

$$\begin{aligned} \Delta V^A &= \beta_1 cost_i^A + \beta_2 \Delta unempl_i^A + \beta_3 \Delta cases_i^A + \beta_4 \Delta deaths_i^A \\ &+ \sum_{k=5}^{14} (\beta_k restr_{ki}^A) + \beta_{15}(-1) + \gamma_1 cost_i^A \times \hat{RP}_i + \gamma_2 \Delta unempl_i^A \times \hat{RP}_i \\ &+ \gamma_3 \Delta cases_i^A \times \hat{RP}_i + \gamma_4 \Delta deaths_i^A \times \hat{RP}_i \\ &+ \sum_{k=5}^{14} (\gamma_k restr_{ki}^A \times \hat{RP}_i) + \gamma_{15}(-1) \times \hat{RP}_i + \varepsilon_i, \end{aligned} \quad (3.3)$$

where $\Delta cases_i^A$ and $\Delta deaths_i^A$ are the differences in COVID-19 cases and deaths under policy A relative to N, $avcost_i^A$ is the average net loss in average household income in respondent i 's county of residence under policy A, and $\Delta unempl_i^A$ is the *extra* unemployment in county A created by policy A. The specific levels of restrictions on the $k = 1, \dots, 10$ different activities and businesses, are rendered

¹⁴We assume that the individual's disutility from county-level average household costs includes the probabilistic implication of this change for their own household's income. With a larger sample, it might have been possible to allow the marginal (dis)utility of average household costs to be greater for respondents who have greater exposure to job losses.

as $restr_{ki}^A \in \{0, 1, 2, 3\}$ (and are treated in our different specifications as either continuous cardinal measures or sets of indicators). These restrictions are present under policy A but absent under the status quo. Given that \hat{RP}_i is respondent i 's fitted de-measured response propensity, the γ terms serve to correct for sample selection bias, and \hat{RP}_i will be counterfactually set to zero when we interpret the β coefficients as representative preferences.

Our approach departs from conventional choice models in one very important way. Our model is expressed in terms of *social costs*, rather than *private costs*, so this specification does not lend itself to straightforward calculations of private willingness-to-pay for private risk reductions. Our policy choices explicitly involve the respondent's willingness to impose *social costs* on their community, in the form of lost jobs and lost incomes, as well as restrictions on activities and businesses, to reduce the risk of cases and deaths in that same community.¹⁵

3.3 Results and Discussion

3.3.1 Homogeneous preferences. Selected parameter estimates for the model in equation (3.3), namely the estimates for $(\beta_1, \beta_2, \beta_3, \beta_4$ and $-\beta_{15})$ are shown as Model 1 in Table 9.¹⁶ Model 2 in Table 9 replaces $avcost_i^A$ and $\Delta unempl_i^A$ in equation 3.3 with interactions of $avcost_i^A$ and $\Delta unempl_i^A$ with each of five indicator variables for the full set of five possible randomized levels of federal

¹⁵This study is expressly NOT designed to yield an estimate of the value of a statistical life or the value of a statistical illness in the conventional sense. Our choice scenarios do not elicit the respondent's willingness to give up their own (private) household income to gain a reduction in their own (private) risk of illness or death. This study also differs from Bosworth, Cameron, and DeShazo (2009) and Bosworth, Cameron, and DeShazo (2015), where choice experiments elicited the respondent's willingness to give up their own (private) household income to pay for public health prevention programs and for public health treatment programs that would reduce illnesses and deaths in their communities.

¹⁶Coefficients on restrictions and response propensity interactions are suppressed in the body of this paper, since these features of the model are not our current focus. The full set of parameter estimates for each model in the body of the paper is provided in Appendix B.6.

UI we asked respondents to assume. A simpler specification, Model 3 in Table 9, replaces $avcost_i^A$ and $\Delta unempl_i^A$ with interactions of $avcost_i^A$ and $\Delta unempl_i^A$ with just two indicators, in this case for the presence and absence of federal UI in the choice scenario. Models 4-6 in Table 9 are analogous to Models 1-3 but among the suppressed estimates, substitute sets of *indicator* variables to capture the effects of the policy’s restrictions on activities or businesses, rather than treating these restrictions as continuous variables. In each model, all variables are also interacted with the respondent’s fitted response propensity. Estimates for the corresponding γ parameters are provided in Appendix B.6. Note that Models 2, 3, 5 and 6 *omit the baseline levels* of for average household income lost and unemployment rate, and use instead the full set of interactions with the federal UI variables in each model. Thus the estimated coefficients are the *levels* of the effects in each case, rather than a base effect and a differential relative to that base effect.¹⁷

The basic specification of Model 1 in Table 9 suggests that respondents prefer policies that reduce the expected number of COVID-19 deaths in their county, as expected, though this result is significant only at the 10% level. Respondents are also highly averse to the status quo alternative, in which no restrictions are placed on activities or businesses and everyone is allowed to decide for themselves what precautions to take, if any. We do not find statistically significant evidence that subjects, on average, respond to expected reductions in COVID-19 cases. We do find, however, that the presence of federal unemployment insurance shapes the effects of lost income and increased unemployment on respondents’ decisions (Table 9, Models 2-3 and 5-6).

¹⁷The estimates in Table 9 do not employ clustered standard errors. Appendix B.6, Table F5 shows the consequences for the key parameter estimates in Models 3 and 6 if we cluster at the respondent level. The qualitative results are the same (i.e., same-signed and statistically significant *differences* between contexts with and without federal UI supplements), so we focus on the simpler models in the body of the paper.

Table 9. Effects of Federal UI payments on preferences over pandemic policies; selected coefficients. (Complete models in Appendix B.6, Table F2)

Model:	<i>Dependent variable:</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
	1=Preferred policy					
	(1)	(2)	(3)	(4)	(5)	(6)
NOTE: Coefficients for each specified condition for federal UI, rather than base coefficients and coefficient differentials						
(β_1) Avg. hhld cost for cty (\$100)	0.004 (0.051)			-0.018 (0.050)		
Avg. hhld cost for cty (fed UI = 0)		-0.284*** (0.093)	-0.291*** (0.093)		-0.301*** (0.098)	-0.310*** (0.098)
Avg. hhld cost for cty (fed UI > 0)			0.174*** (0.066)			0.153*** (0.065)
Avg. hhld cost for cty (fed UI = 100)		0.125 (0.102)			0.088 (0.103)	
Avg. hhld cost for cty (fed UI = 200)		0.131 (0.173)			0.105 (0.164)	
Avg. hhld cost for cty (fed UI = 300) ^a		0.665*** (0.124)			0.659*** (0.124)	
Avg. hhld cost for cty (fed UI = 400)		0.166 (0.144)			0.194 (0.138)	
(β_2) Unempl rate for cty	-0.0003 (0.020)			0.002 (0.021)		
Unempl rate for cty (fed UI = 0)		0.097** (0.039)	0.093** (0.039)		0.098** (0.040)	0.092** (0.040)
Unempl rate for cty (fed UI > 0)			-0.024 (0.020)			-0.023 (0.020)
Unempl rate for cty (fed UI = 100)		-0.028 (0.034)			-0.017 (0.035)	
Unempl rate for cty (fed UI = 200)		-0.001 (0.045)			0.001 (0.043)	
Unempl rate for cty (fed UI = 300) ^a		-0.079*** (0.026)			-0.083*** (0.026)	
Unempl rate for cty (fed UI = 400)		-0.015 (0.025)			-0.014 (0.026)	
(β_3) Absolute '00s cases/mo/50,000	-0.041 (0.032)	-0.036 (0.031)	-0.035 (0.031)	-0.042 (0.032)	-0.037 (0.032)	-0.036 (0.031)
(β_4) Absolute deaths/mo/50,000	-0.034* (0.020)	-0.037* (0.020)	-0.038* (0.020)	-0.027 (0.020)	-0.026 (0.020)	-0.029 (0.020)
(β_{15}) 1=Status quo alternative	-1.982*** (0.296)	-2.008*** (0.284)	-1.994*** (0.291)	-2.433*** (0.360)	-2.552*** (0.370)	-2.472*** (0.368)
Activity restrictions (continuous variables)	✓	✓	✓			
Activity restrictions (sets of indicators)				✓	✓	✓
All response propensity interactions	✓	✓	✓	✓	✓	✓
Total estimated coefs in model	30	46	34	68	84	72
Respondents	993	993	993	993	993	993
Choices	1986	1986	1986	1986	1986	1986
Log likelihood	-1205.77	-1184.09	-1194.52	-1180.75	-1158.56	-1169.80
AIC	2471.55	2460.17	2457.05	2497.5	2485.12	2483.61
BIC	2639.36	2717.49	2647.24	2877.88	2955	2886.37

Notes: See equation 3.3 for the econometric specification for Model 1. To reduce leading 0s in coefficient estimates, average household costs are measured in 100s of dollars per month, cases are measured per 50 county residents, and deaths are measured per 50,000 county residents. ^aFederal UI payments of \$300 were the most-recently experienced level of generosity.

3.3.2 Role of federal unemployment insurance. Respondents were not given the opportunity to vote for or against any level of federal unemployment insurance. They were simply instructed to assume that federal unemployment benefits would take a given level (\$0, \$100, \$200, \$300 or \$400 in additional benefits per week) no matter which policy they selected. Respondents were told, “**Assume that any Federal unemployment benefits, as described, will be in place regardless of any pandemic rules that apply in [respondent’s county]**” (bold in original). This design allows us to determine how the level of federal unemployment insurance affects preferences over our two measures of a policy’s economic costs: increases in county-level unemployment and corresponding reductions in county-level average household income. We emphasized to respondents that average household incomes in their county would fall mostly as a result of the increased unemployment. We also instructed respondents that the economic costs would be unevenly distributed across residents of their county, with some households losing a lot of income and others losing almost no income, so they should consider their own household’s chances of losing income under each policy.¹⁸

Our survey was launched January 13, 2021, one day before the American Rescue Plan (ARP) was first proposed. We concluded data collection February 16, 2021, while the inclusion of federal UI in the ARP (and its level) were still being debated, and more than three weeks before the ARP was signed into law. The ARP ultimately extended the existing \$300/week in additional unemployment insurance provided by the federal government, which would otherwise have expired March 14, 2021. We are fortunate to have fielded our survey during a period when future

¹⁸We included a comprehension check to ensure respondents understood the relationship between unemployment and reduced household income. If they failed the comprehension check, they were reminded, “These ‘Average \$/month lost’ because of a policy are mostly a RESULT of unemployment and lost business earnings. They are not an extra cost on top of that.”

federal contributions to unemployment insurance were still uncertain, permitting us to vary the levels of federal aid in our choice scenarios without straining credulity.¹⁹

Casual intuition and observation suggest competing hypotheses about the likely effects of increased unemployment insurance on preferences. On the one hand, it seems reasonable to surmise that, because unemployment insurance serves to smooth consumption over negative income shocks, increased assistance to the unemployed should make a policy’s “unemployment costs” less painful and therefore less salient. Furthermore, those who lose their jobs because of a coordinated response to a global pandemic—and through no fault of their own—may be considered especially deserving of aid. On the other hand, lawmakers and public figures, especially on the political right, have voiced concerns that overly generous unemployment insurance packages create perverse incentives for workers.²⁰ If these concerns resonate with with a broad swath of Americans, we may find that increases in unemployment are less tolerable in the presence of federal unemployment insurance. Our results are consistent with the latter of these two hypotheses, though we find a great deal of heterogeneity across groups in the effect of federal UI on preferences over the effects of these policies on unemployment and average household costs.

¹⁹For context, respondents were told: “During the first part of the current pandemic, there was an extra unemployment benefit of \$600 per week from the Federal government under the CARES Act. These benefits made the pandemic’s ‘Average \$/month lost’ much lower than they would normally be, for any given level of unemployment. The \$600/week extra benefit ended July 31. A \$300/week extra benefit was then provided in December. The incoming Administration is proposing \$400/week. It is not yet clear whether extra unemployment benefits will continue to be available, at what level, or for how long, as the pandemic drags on.”

²⁰In addition to examples given in Section 1, South Carolina Sen. Lindsey Graham said of 2020’s stimulus bill, “You’re literally incentivizing taking people out of the workforce at a time when we need critical infrastructure supplied with workers.” <https://www.cnn.com/2020/03/25/politics/senate-stimulus-unemployment-benefits-coronavirus/index.html>

In the absence of additional federal UI, we find that respondents have a negative and statistically significant response to a policy’s monetary cost, expressed as the average lost income over the households in their county (Table 9, Models 3 and 6). Again in the absence of supplemental federal UI, we find that respondents have a *positive* and statistically significant response to increases in unemployment caused by a given policy. In contrast, when respondents are instructed to assume a *non-zero* level of federal UI (i.e., \$100, \$200, \$300 or \$400 in additional benefits per week), the effects of these economic costs change considerably. In the presence of federal UI, respondents are significantly *in favor of* lost household income and indifferent to increases in unemployment. Federal UI, then, shapes preferences for local pandemic policies. We discuss possible mechanisms for this apparent preference reversal in Section 3.4.

The results in Table 9 also raise another question: Why would respondents ever respond positively to losses in average household incomes (as they do when federal UI is present) or to increased unemployment (as they tend to do when federal UI is absent)? Attitudes toward the possibly perverse incentives of large federal UI payments or a sense of aggrievement over benefits perceived to be unfairly generous could plausibly explain the first result. If generous federal UI payments seem to be unfair or distortionary, then respondents may view lost income due to unemployment as an important mechanism to maintain desirable economic incentives, and therefore a beneficial feature of a pandemic mitigation policy. A relatively large increase in lost income may specifically signal to respondents that the level of federal UI is sufficiently low that most folks cannot earn more money on unemployment benefits than they would make in wages. However, subjective scenario adjustment on the part of respondents could account for the second result. Respondents may suppose that high unemployment

is associated with reduced social contact, slowing the spread of COVID-19 and resulting in greater reductions in cases and deaths than are specified in the choice scenario.²¹

The point estimates for the effects of \$100, \$200, \$300 and \$400/week federal UI policies on preferences over increased unemployment and average household costs do not display monotonicity, as can be seen in Models 2 and 5 in Table 9. For both average household costs and unemployment rates, we see the largest (and the only individually statistically significant) effects of federal UI at the \$300/week level. Respondents perhaps found this level to be more plausible than other positive levels, given that \$300/week was the true level of federal UI at the time the survey was fielded. A likelihood ratio test suggests that Model 5, with indicators for each level of federal UI, dominates Model 6, with indicators for only the presence or absence of federal UI. However, we select the Model 6 as our preferred (conservative) specification because the Akaike Information Criterion yields the opposite conclusion.²² Overall, subjects appear to respond mostly to the presence, rather than the specific level, of federal unemployment insurance, while anchoring on the very salient \$300/week level when it is presented. Given this pattern, in subsequent specifications we interact policies' economic costs (both average household costs and unemployment) with indicators for the presence or absence of federal UI described for their the choice scenario.

Among the other coefficients featured for the models with homogeneous preferences in Table 9, the number of COVID-19 cases does not appear to affect support for pandemic policies. COVID-19 deaths affect policy preferences only

²¹Scenario adjustment in stated preference research is specifically addressed in Cameron, DeShazo, and Johnson (2011).

²²Model selection based on AIC yields $AIC(\text{Model 5}) = 2485.1$; $AIC(\text{Model 6}) = 2483.6$.

in Models 1–3 that capture the extent of our ten categories of restrictions on activities and business as continuous variables. In the fully saturated model with each of these ten types of restrictions included as sets of indicators, the marginal (dis)utility of COVID-19 deaths becomes statistically insignificant. However, for all specifications, the coefficient on the status quo indicator is negative and strongly statistically significant, indicating that on average, respondents prefer *some policy* to *no policy*, regardless of the policy’s attributes.

3.3.3 Preference heterogeneity. Three of the other choice-experiment studies of pandemic policies summarized in Table 1 employ latent-class analysis when they consider heterogeneity in policy preferences. However, in this paper, we emphasize *systematic* sources of heterogeneity in policy preferences, because it seems likely that preferences are more complicated than just a finite mixture of some small number of preference classes.²³ In Table 10, we split our sample according to different sociodemographic characteristics to illustrate numerous dimensions of preference heterogeneity with respect to the economic impacts of pandemic policy according to the presence or absence of federal UI.

Panels A and B in Table 10 present split-sample estimates for separate demographic groups. Each set of columns (e.g. “Age Groups,” comprising columns (1)–(3) in Table 10) shows the results from a *single* specification. Each such specification includes indicator variables for membership in the full set of the relevant groups (e.g. *18–34*, *35–64*, or *65+*) interacted with the attributes of each alternative. The baseline non-interacted attribute levels are again excluded from

²³We report on our explorations of alternative models, including latent class and mixed-logit specifications, in Appendix B.7

the models. Significance levels for coefficients for each group are determined with respect to zero, rather than with respect to an arbitrarily chosen reference group.²⁴

In Table 10, panels A and B reveal significant preference reversals with respect to average household costs and unemployment rates in the presence of federal UI for both men and women, and for both white and non-white racial groups. When partitioning the sample by age, political ideology, and income, however, we find mixed effects of federal UI.

The young (18–34) and middle (35–64) age groups are significantly averse to higher average household costs when federal UI is absent and significantly in favor of higher average household costs when federal UI is present. The preference reversal with respect to unemployment for these two age groups is less stark but is still apparent. For respondents 65 years and older, however, the coefficients for cost and unemployment are statistically indistinguishable from zero, whether or not federal UI is present. This may reflect the availability of social security and other retirement income that insulates retirees from the risk of unemployment.

Political moderates and conservatives also show strong evidence of a preference reversal in the presence of federal UI, although liberals do not. Liberals have a larger point estimate for the coefficient for unemployment when federal UI is present than when it is absent, although this difference is not statistically significant. Nevertheless, across the dimensions of heterogeneity presented here, this is the only instance that stands apart from the prevailing pattern where the availability of federal UI is associated with greater aversion to unemployment.

We also see strong evidence for preference reversals among respondents whose

²⁴An alternative to this mode of estimation is to estimate completely separate specifications for each group. However, the estimated parameters from a conditional logit model are identified only up to an unknown scale parameter, which may differ across groups. By estimating the parameters for all groups with a single specification, we constrain this scale parameter to be the same across groups, permitting legitimate comparisons of same-scale utility-function parameters across groups.

Table 10. Heterogeneity in preferences across socioeconomic groups; generalizations of Model 6 in Table 9; selected coefficients. (Complete models in Appendix B.6, Tables F3 and F4)

Panel A							
Heterogeneity by:	(1) Age			(2) Race		(3) Gender	
	18 to 34	35 to 64	65 +	Non-white	White	Women	Men
<i>Dep. var:</i> 1=Chosen policy	(a)	(b)	(c)	(a)	(b)	(a)	(b)
Avg. hhld cost (fed UI = 0)	-0.597*** (0.224)	-0.382*** (0.139)	-0.307 (0.376)	-0.842*** (0.294)	-0.228** (0.112)	-0.497*** (0.165)	-0.220* (0.114)
Avg. hhld cost (fed UI > 0)	0.251** (0.124)	0.212* (0.119)	0.181 (0.140)	0.052 (0.126)	0.216*** (0.079)	0.155 (0.106)	0.134 (0.082)
Unempl rate (fed UI = 0)	0.251*** (0.093)	0.080 (0.055)	0.219 (0.161)	0.268** (0.118)	0.090* (0.046)	0.141** (0.064)	0.093 (0.057)
Unempl rate (fed UI > 0)	0.023 (0.044)	-0.066** (0.031)	-0.012 (0.069)	0.011 (0.050)	-0.029 (0.024)	-0.049 (0.036)	0.020 (0.029)
Absolute '00s cases/mo/50k	-0.047 (0.055)	0.068 (0.049)	-0.045 (0.099)	-0.055 (0.071)	-0.048 (0.040)	-0.086* (0.044)	0.008 (0.049)
Absolute deaths/mo/50k	-0.074** (0.033)	-0.011 (0.035)	-0.059 (0.056)	-0.070* (0.039)	-0.001 (0.027)	0.001 (0.032)	-0.061** (0.028)
1=Status quo alternative	-4.533*** (0.729)	-1.943*** (0.559)	0.966 (1.211)	-2.603*** (0.761)	-2.584*** (0.443)	-2.925*** (0.557)	-2.039*** (0.508)
Number of respondents	317	453	223	295	698	507	480
Number of choices	634	906	446	590	1396	1014	960
Total estimated coefficients	102			68		68	

Panel B							
Heterogeneity by:	(4) Ideology			(5) Education		(6) Income	
	Liberal	Moderate	Conservative	Non-college	College	< \$75k/yr	> \$75k/yr
<i>Dep. var:</i> 1=Chosen policy	(a)	(b)	(c)	(a)	(b)	(a)	(b)
Avg. hhld cost (fed UI = 0)	0.094 (0.212)	-0.891*** (0.229)	-0.411** (0.173)	-0.347* (0.210)	-0.239* (0.123)	-0.452*** (0.157)	-0.192 (0.143)
Avg. hhld cost (fed UI > 0)	-0.081 (0.207)	0.123 (0.118)	0.123 (0.108)	0.150* (0.082)	0.119 (0.109)	0.367*** (0.091)	0.028 (0.087)
Unempl rate (fed UI = 0)	-0.116 (0.087)	0.316*** (0.104)	0.089 (0.068)	0.140 (0.086)	0.015 (0.054)	0.250*** (0.073)	-0.015 (0.055)
Unempl rate (fed UI > 0)	0.063 (0.078)	-0.047 (0.040)	-0.082*** (0.031)	-0.042 (0.034)	-0.016 (0.031)	-0.017 (0.033)	-0.022 (0.028)
Absolute '00s cases/mo/50k	-0.036 (0.105)	-0.146** (0.057)	-0.081 (0.061)	-0.164*** (0.061)	0.081** (0.037)	-0.092* (0.055)	-0.012 (0.045)
Absolute deaths/mo/50k	-0.035 (0.068)	0.028 (0.044)	-0.083*** (0.030)	-0.029 (0.034)	-0.029 (0.026)	-0.028 (0.030)	-0.014 (0.029)
1=Status quo alternative	-5.006*** (1.017)	-3.647*** (0.832)	-2.854*** (0.650)	-2.497*** (0.514)	-2.869*** (0.559)	-2.705*** (0.504)	-2.321*** (0.578)
Number of respondents	338	309	303	398	589	489	475
Number of choices	676	618	606	796	1178	978	950
Total estimated coefficients	102			68		68	

Notes: Panels A and B each contain three models, split into multiple columns for easier comparisons of these marginal utilities across groups. Each model includes interactions between all variables and indicators for membership in each group (e.g. “18–34”, “35–64,” and “65+”). To avoid the dummy variable trap, the non-interacted base levels of these variables are not included in the models. Similarly, average household cost and unemployment rate are interacted with indicators for the presence or absence of federal UI, but the uninteracted cost and unemployment variables are excluded from the models to avoid the dummy variable trap. All models are corrected for sample selection (see section 2.1), and include indicators for the levels of restrictions on each of the 10 categories of activities or businesses in our choice experiment. For five of the six models, we find that a likelihood-ratio test of the restrictions implied by the homogenous-preferences model (model 6 in Table 9) rejects these restrictions ($p < 0.05$). The model that differentiates by race (Panel A, Model 2, columns (a) and (b)) does not offer a statistically significant improvement over Model 6 in Table 9. Average household costs are measured in \$100.

household income is less than \$75k per year, but not among those whose household income is greater than \$75k per year.

Several other differences in preferences across groups warrant comment. The main benefits of these policies—reduced cases and deaths from COVID-19—are more important to some groups than to others. Table 10 reveals that COVID-19 *cases* are of concern to women, political moderates, and lower-income individuals. Somewhat perplexing is the result that college-educated respondents appear to prefer more COVID-19 cases to fewer. Positive point estimates for the marginal (dis)utility of COVID-19 cases are also obtained for the 35–64 age group and for men, but these estimates are not statistically different from zero. COVID-19 *deaths* are of concern to the 18–34 age group, to Whites, to men, and to conservatives.

Three subgroups in Table 10 display no systematic responsiveness to average household costs, unemployment rates, cases, or deaths. These subgroups are: people aged 65 and above (who may anticipate their imminent eligibility for the vaccine), liberals, and those with annual household incomes greater than \$75,000. People with predictable retirement income and high-income households are likely to find issues of unemployment and its costs to be less salient. While the liberal group certainly includes young people with lower incomes, the most notable feature of their preferences is the greatest aversion to the status quo across all groups in Table 10. This group simply prefers *any* policy to no policy, regardless of the economic impact on their county or the morbidity/mortality consequences.

Political moderates wish to reduce *cases* (but not necessarily deaths), while conservatives are significantly motivated to reduce *deaths*. Every demographic group we examine, with the exception of those 65 or older, exhibits a statistically significant aversion to the status quo alternative. While there is considerable disagreement about the ideal features of a pandemic mitigation policy, we find a

near consensus across groups that, in the face of a pandemic, something must be done to limit its spread.

Appendix B.7 provides results for a number of alternative specifications and well as some sensitivity analyses that explore heterogeneity across split samples according to (a) time spent reading the instructions about assumptions concerning federal UI for the first choice task; (b) the salience of unemployment as measured by whether the worst monthly actual unemployment rate during the pandemic in the respondent’s own county was better or worse than the median worst unemployment rate across all respondents in the sample; (c) whether the respondent who made no mistakes, versus any mistakes, on two key comprehension questions in the survey; and (d) whether the respondent’s household income was more or less than the median income in the same zip code. The point estimates and levels of significance for the coefficients on our four key interaction terms vary from case to case, but the typical pattern of signs persists.

3.4 Conclusions

Harmonization of local-level policies or programs with federal-level counterparts can be challenging. We specifically explore the relationship between (a) the federal social safety net and (b) preferences over county-level pandemic mitigation strategies. However, we note that adjustments to federal policies may change people’s preferences for related policies across jurisdictional levels and policy domains. Sometimes a straight line can be drawn from federal policies to state or local policy preferences. For example, a number of localities passed “sanctuary laws” in 2017 in response to the Trump administration’s deportation policies. Similarly, state-level efforts to pass “heartbeat bills” that restrict abortion access are almost certainly a response to the legal status of abortion at the federal level. Less directly, changes in federal immigration policy may affect preferences for state-

level voter ID requirements. Or, increased trade liberalization could plausibly be expected to change preferences for legal protections for labor unions that preserve domestic manufacturing jobs.

In this study, we employ a new survey, fielded between mid-January and mid-February of 2021, to elicit preferences over alternative county-level pandemic policies under different assumptions about the generosity of supplementary federal unemployment insurance. Each county-level policy is described in terms of its individually randomized effects on pandemic cases and deaths in the respondent's own county, the economic consequences (namely, the average cost per household and unemployment rates) and the levels of restrictions on a set of ten activities or businesses that would be imposed at the county level to achieve the specified expected reductions in cases and deaths.

We focus in this paper on the role played by supplementary federal unemployment insurance. More specifically, we highlight the effect of federal UI supplements on the level of disutility that respondents associate with (a) the county-level average costs of these pandemic policies, and (b) the unemployment that these policies generate. Federal unemployment insurance is intended to alleviate the economic hardship caused by pandemic restrictions.

What might one expect, intuitively? *Without federal UI*, unemployment rates will be highest for people who cannot work from home, so that unemployment rates may also proxy for avoidance of workplace exposure to infection for essential workers. This socially desirable aspect of higher unemployment may dominate people's policy preferences. But unemployment also imposes costs on households, and these costs can be expected to make people worse off. *With federal UI*, if these payments succeed in completely removing people's concerns about the economic impact of a county-level pandemic policy, we might

expect that neither average household costs, nor the unemployment rate from a policy, would then be viewed as a bad thing. Greater unemployment might even be desirable if its role in social distancing is perceived to be sufficiently great.

In the following discussion, we summarize our findings. Some of the results from our analysis may seem counter-intuitive at first. In some cases, we speculate about possible mechanisms behind these effects, but we emphasize that additional data and further research will likely be necessary to confirm or refute these conjectures.

What do we find, in models with homogeneous preferences?

Without federal UI, it appears that people dislike policies that result in higher average household costs, although they are in favor of policies that cause more county-level unemployment (again, perhaps because unemployment is associated with greater social distancing). *With federal UI*, policies that result in higher average household costs become acceptable, but greater unemployment becomes undesirable (perhaps because respondents object to people being paid by the federal government *not* to work). The difference is especially discernible when federal UI is described as matching the \$300-per-week regime that had most recently been experienced by respondents.

What do we find, in models with heterogeneous preferences, for average household costs? *Without federal UI*, people tend to derive disutility from county-level policies with higher average household costs. This marginal utility is statistically significantly negative for all of the subgroups we considered except those 65 years and older, liberals, and those with incomes in excess of \$75,000 per year. A negative marginal utility from higher average household costs can be interpreted as concern about one's own livelihood or the livelihoods of others in the community. *With federal UI*, however, people derive statistically

significant *positive* utility from higher average household costs in their county. This marginal utility is statistically significantly positive for people under 65 years of age, for Whites, for those without college educations and for people earning less than \$75,000 per year. Positive marginal utility from higher policy costs, ordinarily, would be unexpected. With generous federal UI benefits, however, a subset of respondents may be concerned that some unemployed workers will earn more money from unemployment benefits than those workers could have earned at their jobs. For these respondents, a larger average household cost may make a given policy more attractive because it signals that fewer workers find unemployment to be more lucrative than working.

What do we find, in models with heterogeneous preferences, for county-level unemployment rates? *Without federal UI*, people derive positive marginal utility from higher county-level unemployment. These effects are statistically significant for the 18–34 age group, for both racial groups we consider, for women, for moderates and for people with incomes less than \$75,000 per year. These positive marginal utilities may stem from the perception that unemployment corresponds to reduced exposure to COVID-19. *With federal UI*, people may suffer a loss in utility from higher county-level unemployment rates, although this effect is not always statistically significant. This disutility from policies that result in more unemployment is statistically significant, however, for the 35–64 age group and for conservatives. The loudest warnings about undesirable economic incentives of federal UI have come from the political right, which is consistent with a sentiment that higher unemployment is a bad thing.

What do we find for COVID-19 cases and deaths? In our models with *homogeneous* preferences, all specifications yield negative point estimates for the disutility associated with COVID-19 cases and deaths in one’s county,

but only the disutility from additional deaths is statistically significant, and only in some model variants. In our models with *heterogeneous* preferences, there is greater variety. All statistically significant groupwise marginal disutilities from cases and deaths are negative, although there is one puzzling exception where college-educated respondents appear to prefer *more* COVID-19 cases. If college-educated respondents are themselves less likely to contract COVID-19, one could speculate that they might favor higher infection rates because this would hasten “herd immunity,” but we have no evidence to support this conjecture.

What do we find for status quo effects? Except for the 65+ age group in our split-sample models with heterogeneous preferences, respondents are (in all cases) strongly statistically significantly more likely to choose some policy over the status quo, regardless of the attributes of that policy.

What is our most notable empirical insight? When respondents are asked to assume that federal UI will be zero, they tend to be averse to losses in average household income in their county but favorably disposed toward increased unemployment. With positive federal UI payments, however, respondents are more willing to accept losses in average household income but view increased unemployment less favorably. The first reversal, with respect to losses in average household income, is driven primarily by younger, white, non-college and lower-income respondents. The second reversal, with respect to unemployment, is driven primarily by middle-aged and conservative respondents.

Overall? Our results suggest that the generosity of the social safety net at the federal level can alter support for county-level public health policies when these county-level policies will have substantial economic impacts. This insight suggests that policymakers should carefully consider the potential effects of federal-level economic stimulus and unemployment insurance policies on support for

other policies where authority devolves to lower-level jurisdictions. If too many people are concerned that federal UI is distortionary or unfair, then “stimulus checks” that are independent of employment status may represent a better solution. Alternatively, if federal UI is capped at some fraction of previous income (as are unemployment payments in many states), this may circumvent the objection that paying folks more to stay home than to work creates an especially perverse incentive. Unconditional stimulus checks, or federal UI that is indexed to previous income, might give state and local governments more flexibility to enact pandemic mitigation policies—or other emergency policies that foreseeably increase reliance on public assistance—while retaining popular support.

CHAPTER IV

AMPLE CORRECTION FOR SAMPLE SELECTION IN THE ESTIMATION OF CHOICE MODELS USING ONLINE SURVEY PANELS

Trudy Ann Cameron and I wrote this chapter together. Dr. Cameron contributed the econometric theory in section 4 of this chapter. I developed and wrote the code for the full information maximum likelihood estimation, executed and analyzed the simulations, and wrote much of the exposition.

Online survey panels with quota-based sampling are often used for choice experiments designed for non-market valuation of public goods. Quotas can ensure a sample of respondents that is representative in terms of its marginal distributions for a limited number of observable sociodemographic characteristics, such as age, race, gender, income brackets or geography. But economists are well aware of the additional potential for systematic selection on the basis of unobservable traits or attitudes. Systematic selection can yield an estimating sample of respondents who have different preferences than the general population. A seminal paper by Heckman (1979) demonstrates how an explicit response/non-response model can be combined with a least-squares-based outcome model based on the respondent sample, under a maintained hypothesis that the errors in the selection equation and the outcome equation are bivariate normal and potentially correlated. However, a Heckman-type approach is inappropriate for the conditional logit choice models typically used to analyze the data from choice experiments. This is because these outcome models are based on fundamentally uncorrelated Type I Extreme Value distributions which thwart reliance on the assumption of potentially correlated bivariate normal errors. We propose and demonstrate a novel method of sample selection correction for conditional logit models based on mixed logit estimation methods.

4.1 Introduction

Researchers across many different fields have used internet panel data for choice experiments designed to reveal the demand for non-market public goods. Across most of these studies, there has been relatively little attention to systematic sample selection, and there are only a few examples of attempts to correct for sample selection bias.

Quota sampling from an online panel can readily yield a sample with marginal distributions for a set of observable sociodemographic variables—say age, gender, race, income brackets or geographies—that match the marginal distributions for these variables in the general population. However, once respondents have been informed about the subject matter of the survey (usually during the Consent to Participate section), some will find the topic to be more interesting or salient than will others. Alternatively, attrition may occur during the course of the survey, perhaps because of the complexity of the choice experiments, leaving fewer choices by respondents who are less engaged with the topic, have less patience with difficult choice questions, or have less time to work carefully through a long survey. Of course, if response propensity is independent of the outcome variables of interest, even a low response rate can produce unbiased estimates if the sample remains large enough to permit statistical identification of key marginal utilities. But we should always be concerned that completed surveys may be a non-random sample from the representative set of invited participants.

How have researchers using internet panels for choice experiments addressed the problem of potentially systematic sample selection? We conducted a cursory search of Web of Science using the search string (survey and “choice experiment” and representative) and chose a number of examples from each of the main research areas where choice experiments are common (i.e., environmental economics, health

economics, marketing research, and transportation economics). We describe the treatment (or non-treatment) of sample selection issues in these papers in greater detail in Appendix C.1, and merely summarize here.

Some choice-modeling studies claim a “representative sample” in their abstracts, but most turn out to be representative only in terms of the marginal distributions of a handful of basic sociodemographic categories or income brackets. Only a few acknowledge that this representativeness is only in terms of observable characteristics, rather than unobservables such as ideologies or attitudes. Many cite their relatively higher response rates as insurance against sample selection bias. However, it is far more rare for researchers to address the representativeness of the pool of potential respondents from which their quota-based samples are drawn.¹

The abstracts of some papers describe the willingness-to-pay (WTP) estimates from their choice experiments as indicative of a “representative household,” while other papers do not advertise the representativeness of their sample in the paper’s abstract but address the topic in the body of the paper, or at least acknowledge in passing the potential for non-response bias in their results. Only a very few researchers mention that a sample which is representative on observables may still be systematically selected on unobservables.

The unobservables that affect survey response propensities may include the salience of the survey topic, or the way that the survey’s topic relates to the individual’s own political ideology or attitudes. As an example, political polling with “representative” samples of likely voters prior to the 2020 Presidential election in the U.S. often suggested a strong win for the Democratic Party candidate, but the actual election produced a much narrower win. This result could have been

¹The new method for selection correction we propose here remains conditioned on the assumption that the sampling frame is representative. We focus on the problem that a representative sample of survey invitees may yield a non-random sample of respondents.

expected if voters in favor of the Republican Party candidate were systematically less likely to cooperate with poll-takers. This systematic bias appeared even when pollsters weighted respondents using observable characteristics.

4.2 Example: Willingness to pay for carbon emissions reductions

A concrete “running example” may help clarify the problem of selection on unobservables. Suppose we are conducting a stated choice experiment to determine average WTP across a population for a proposed environmental policy. The benefits of the policy are described to respondents in terms of greenhouse gas reductions, and the costs are described in terms of dollars (e.g. higher prices). A large random sample from the population of interest is invited to participate in the choice experiment, and a sizable fraction of invitees complete the survey. Participants appear to be representative—the sample and the target population have similar proportions of men and women, age groups, ethnicities, etc. That is, the sample is representative *on observables*. However, suppose some invitees have a higher degree of trust in science and that this attitude is unobservable. Let the scalar m_i^* represent i 's trust in science. Invitees with high values of m_i^* perceive more value in scientific research, including survey-based social science research, and are more likely to respond. Additionally, these same respondents are more likely to take seriously the warnings of climate scientists about the effects of greenhouse gas emissions and to derive utility from prospective emissions reductions. As a result, those who are most concerned about greenhouse gas reductions are also most likely to respond to the survey, biasing upwards the estimated WTP for the proposed policy.²

²To complete the argument that estimated WTP will be biased upward in this example, we must also make an assumption about the relation between trust in science the marginal utility of income. For illustrative purposes, we may simply assume that trust in science has no relationship to the marginal utility of income.

4.3 Data needs for modeling response propensities

Selection correction methods, such as the Heckman correction, typically employ two submodels. The first (or *selection*) submodel predicts each invitee's propensity to respond to the survey (or to be part of the effective sample for the second model, in the typical application outside of the choice-experiment survey context). The second (or *outcome*) submodel predicts the outcome of interest and uses information from the first submodel to correct the values of parameters in the second submodel. In the Heckman correction, the selection submodel employs a probit specification, and the outcome submodel includes as a regressor the conditional expected value of the error term from the selection submodel, calculated as the inverse Mills ratio associated with the fitted "index" for the selection submodel.³ The outcome submodel in a Heckman correction may be either ordinary least squares or probit, but its error term must be distributed normally for this correction method to be appropriate.

To correct for selection bias as described above, the selection submodel requires conformable explanatory variables for every individual who is *invited* to participate in a choice experiment study, regardless of whether they respond to the invitation and complete a survey. The key requirement, however, is that the specification for the latent response propensity in the selection model must include exogenous variables that can be excluded, a priori, from the outcome equation wherein preferences for the public good in question are estimated. In the context of our running example, the model requires determinants of survey response propensities that are expected to have no effect on respondents' absolute level of utility derived from any of the carbon-reduction programs offered in the survey, or

³In practice, both submodels are typically estimated jointly, but it is also possible to estimate a Heckman correction in two stages, provided one corrects the standard errors in the second stage.

on the marginal utility that respondents derive from either their net income or the extent of the carbon reduction in any proposed program.

Collecting information about non-respondents can be challenging, and the problem is compounded if we limit our scope to variables that are related to response propensity and unrelated to the preferences under study (e.g. WTP for carbon reductions). We propose a straightforward solution. Survey researchers can “treat” invited participants in systematically different ways by randomizing the administration of the survey to different groups. Extant research has explored the efficacy of different survey administration practices to increase response rates. At the cost of somewhat lower overall response rates, therefore, it is possible to create exogenous differences in survey invitations that systematically affect response rates. There are several options. Survey invitations (assuming online administration) can be issued in smaller batches, on different days of the week and at different times of day. The sizes of monetary survey incentives, now often consisting of redeemable gift certificates to one or more online vendors, can be randomized across survey instances. Or, the type of gift certificate could be randomized. Different waves of survey invitations can be subjected to different numbers of follow-up reminders, at different time intervals after the original invitation. If multiple such treatments are planned, there must be enough different waves of the survey so that there is independent variation in each type of treatment. Randomly assigned (and therefore exogenous) treatments in survey administration can then be included as regressors in the selection model.

4.4 Response/non-response and policy preferences as correlated choices

If the decision about whether to respond to the survey is related to the policy preferences the survey is designed to reveal, the estimating specification must accommodate these correlated choices. Furthermore, each respondent may be

asked to make more than one policy choice. Ultimately, we need a joint model for all of the choices made by each respondent that allows for unobserved factors that may make the individual simultaneously (1) more likely than otherwise expected to provide a set of responses, and (2) more likely than otherwise expected to express a preference for, or against, any policies that are offered by the survey. These unobserved factors are captured by the error terms in the selection model and the policy choice model. First we discuss the theory underlying each submodel, then explain how to “pool” the data for each submodel to be estimated jointly.

4.4.1 Algebra of the selection model. In our two-alternative choice for the selection equation, only the “respond” indicator variable is alternative-specific. All of the other explanatory variables are individual-specific, rather than individual-and-alternative-specific. Thus a two-alternative conditional logit model for the selection equation (where r = response and nr = non-response) could be written as follows:

$$V_i^r = \alpha_{1i}(1) + \alpha_2^r z_i + \omega_i^r \quad (4.1)$$

$$V_i^{nr} = \alpha_{1i}(0) + \alpha_2^{nr} z_i + \omega_i^{nr}$$

With nr as the numeraire alternative, the utility-difference that dictates the individual’s response/non-response decision is given by $\Delta V_i^r = V_i^r - V_i^{nr}$, so that the differenced equation in the case of the numeraire alternative is then degenerate:

$$\begin{aligned} \Delta V_i^r &= \alpha_{1i} & + \alpha_2 z_i & + \omega_i & \text{[Utility-difference for response alternative]} & (4.2) \\ 0 &= 0 & + 0 & + 0 & \text{[Utility-difference for non-response alternative]} \end{aligned}$$

where only the parameter difference(s) $\alpha_2 = (\alpha_2^r - \alpha_2^{nr})$ can be estimated, rather than α_2^r and α_2^{nr} separately, because the individual-specific characteristics z_i do not differ across the two alternatives. The new error term $\omega_i = (\omega_i^r - \omega_i^{nr})$ is the difference in the original two error terms (which will be assumed to be distributed

Type I Extreme Value so that ω_i is logistic). Returning to our running example, the characteristic *trust in science* (m_i^*) does not appear directly in 4.2 because it is unobservable. Instead m_i^* affects the utility difference indirectly, either through the individual-specific constant α_{1i} or through the error term ω_i (or both).⁴

4.4.2 Algebra of the outcome model. Suppose that each policy choice scenario asks the respondent to consider J alternatives, where the J^{th} alternative is the status quo, considered to be “no policy.” In the indirect-utility equations that make up the policy-choice portion of the pooled model, there is no intercept, per se. However, the Δx_i^j variables may include alternative-specific indicators. In particular, it is considered best practice to include a status-quo (or an any-policy) alternative in policy choice tasks. In terms of the levels of utility for each policy alternative, the policy-choice submodel can be written as:

$$\begin{aligned}
 V_i^A &= \beta_1 1(\text{any policy}_i^A) + \beta_2 x_i^A + \epsilon_i^A && \text{[policy A]} && (4.3) \\
 V_i^B &= \beta_1 1(\text{any policy}_i^B) + \beta_2 x_i^B + \epsilon_i^B && \text{[policy B]} \\
 &\vdots \\
 V_i^{J-1} &= \beta_1 1(\text{any policy}_i^{J-1}) + \beta_2 x_i^{J-1} + \epsilon_i^{J-1} && \text{[policy J-1]} \\
 V_i^J &= \beta_1(0) + \beta_2 x_i^J + \epsilon_i^J && \text{[status quo policy J]}
 \end{aligned}$$

where x_i^j is row-vector of the levels of alternatives for policy j , and β_2 is a conformable column-vector of parameters. Error terms ϵ_i^j are assumed to be distributed i.i.d. Type I Extreme Value.

The policy choice portion of the model must also be differenced, relative to a numeraire alternative, to produce unique parameter estimates. We can difference all of the policy attributes for each choice set relative to their levels for alternative J ,

⁴We will show momentarily that this is a distinction without a difference.

so the utility-difference for the J^{th} alternative is again degenerate:⁵

$$\begin{aligned}
\Delta V_i^A &= \beta_1 1(\text{any policy}_i^A) + \beta_2 \Delta x_i^A + \eta_i^A && \text{[policy A]} \\
\Delta V_i^B &= \beta_1 1(\text{any policy}_i^B) + \beta_2 \Delta x_i^B + \eta_i^B && \text{[policy B]} \\
&\vdots \\
\Delta V_i^{J-1} &= \beta_1 1(\text{any policy}_i^{J-1}) + \beta_2 \Delta x_i^{J-1} + \eta_i^{J-1} && \text{[policy J-1]}
\end{aligned} \tag{4.4}$$

where $\eta_i^j = \epsilon_i^j - \epsilon_i^J$ and $\Delta x_i^j = x_i^j - x_i^J$ for $j = A, \dots, J - 1$.⁶

The stacked utility-differences for the selection (response/non-response) choices and outcome (policy) choices then take the following form:

$$\begin{aligned}
\Delta V_i^r &= \alpha_1(1) + \alpha_2 z_i + \beta_1(0) + \beta_2(0) + \omega_i && \text{[selection]} \\
\Delta V_i^A &= \alpha_1(0) + \alpha_2(0) + \beta_1(1) + \beta_2 \Delta x_i^A + \eta_i^A && \text{[policy A]} \\
\Delta V_i^B &= \alpha_1(0) + \alpha_2(0) + \beta_1(1) + \beta_2 \Delta x_i^B + \eta_i^B && \text{[policy B]} \\
&\vdots \\
\Delta V_i^{J-1} &= \alpha_1(0) + \alpha_2(0) + \beta_1(1) + \beta_2 \Delta x_i^{J-1} + \eta_i^{J-1} && \text{[policy } J - 1\text{]}
\end{aligned} \tag{4.5}$$

Note that the variables in the selection submodel do not overlap with the variables in each equation for the policy submodel, in the sense that no coefficients are constrained to be equal across the two submodels. Without cross-model parameter constraints, all the different error terms ($\omega_i, \eta_i^A, \eta_i^B \dots \eta_i^{J-1}$) are implicitly normalized relative to the dispersions of the relevant error terms and are thus i.i.d.

⁵Most conditional logit software packages permit the user to employ data in the form of the absolute levels of the attributes that enter into the utility function in for each alternative, with the differencing relative to an arbitrarily specified alternative (often the respondent's chosen alternative) happening invisibly in the course of estimation. But the attributes of alternative in policy choices are often explained to respondents in terms of differences relative to the status quo, to save the researcher from having to specify realistic respondent-specific status quo levels. If the numeraire alternative is the status quo, selection of that alternative would lead to "no change" in conditions, i.e., zeros for all of these changes.

⁶As we've described the algebra here, x_i^J is a vector of 0s, so $\Delta x_i^j = x_i^j$.

logistic(0,1) (and therefore automatically uncorrelated due to the properties of the logit model).

While the typical structure of the data for conditional logit models requires one row of data for each alternative in a choice task, we instead structure the data for the Apollo package in R (Hess & Palma, 2019), version 0.2.4, which we use for joint mixed logit estimation of both submodels. Apollo requires the data to be structured such that each choice occupies one row, so that responders have $n + 1$ rows and non-responders have 1 row, where n is the number of policy choices each responder makes. This structure is in contrast to the typical data structure for conditional logit, as in the Stata software package, for example, where the stacked data for each respondent would have $2 + nJ$ rows and the data for each non-respondent would have 2 rows.

One other context in which choice data are pooled across different types of choices is where data are combined across revealed-preference (RP) choices (i.e., choices that people have made in the real world) with analogous stated-preference (SP) choices (i.e. choices that people have made in hypothetical choice contexts in surveys). In that case, marginal utility parameters that *should* reflect the same preferences appear in the submodels for both type of data. However, SP and RP data tend to have different amounts of “noise,” implying that when an identical specification is used for the two types of data, the error variances for each submodel should be allowed to differ.⁷ For our purposes, however, we constrain the error

⁷In the ordinary fixed-parameter conditional logit framework, the marginal utility parameters are identified only up to a scale parameter, call it κ , which implicitly divides all the coefficients and the error term, such that the effective error dispersion is normalized to one. Any conditional-logit error dispersion parameter must be non-negative, so this dispersion parameter is often conceptualized as $\kappa^* = \exp(\kappa)$ where κ is unconstrained. In a homoskedastic model, the usual normalization implies $\kappa^* = \exp(\kappa) = \exp(0) = 1$. However, if we wish to allow the error dispersion to differ across two groups in the data, we can select a baseline group for which the error dispersion is 1, and allow the error dispersion for the alternate group to be some positive multiple of that dispersion, either greater than or less than one. This can be achieved by using an

variances to be the same across the two kinds of choices participants make. See Appendix C.2 for discussion on this point.

4.4.3 Simplest correlation: Selection equation intercept and choice-model any-policy effect. Given an exogenous predictor of response propensity z_i and appropriately structured data, we can begin estimation. We will first consider a scenario where unobserved heterogeneity that makes an invited participant relatively more likely to respond to the survey also makes the individual either more or less likely to vote in favor of the status quo alternative, and then consider the more general case in which the decision to respond is correlated with specific preference parameters in the outcome submodel. For our simplest selection-corrected choice model, we allow an individual’s latent propensity to respond to the survey to be correlated, potentially, with a discrete increment (or decrement) to utility that they associated with any policy, as opposed to the status quo. This can be accomplished by employing a modified mixed logit model where α_1 in the selection submodel and the β_1 on the “any policy” indicator in the policy choice submodel are permitted to be random parameters with a non-zero correlation.⁸

By pooling the selection submodel and policy-choice submodel and estimating them simultaneously, we can allow the intercept coefficient in the

indicator variable, say $D_i = 1$ for the alternate group, where $D_i = 0$ for the baseline group. The relevant dispersion parameter for an observation is then given by $\kappa_i^* = \exp(0 + \delta D_i)$, where δ is an unrestricted parameter. Then the effective dispersion parameter for baseline group is 1 and for the alternate group, it is $\kappa_A = \exp(\delta)$. The error dispersion for the alternate group will be non-negative, and will be larger or smaller than the error dispersion for the baseline group according to whether the estimated δ parameter is greater or less than one. Many researchers who work with pooled samples in choice models choose to reparameterize the model, switching from the dispersion parameter to a “scale” parameter, $\lambda = 1/\kappa_A = \exp(-\delta)$. In this case, the appropriate transformation of the data for the alternative group is to multiply all of the variables (including the intercept term) by λ , rather than dividing by κ_A .

⁸While it is typical to use a status-quo indicator variable (equal to 1 for the status quo alternative and 0 otherwise), we instead use an “any policy” indicator, equal to 1 minus the status quo indicator. This reparameterization merely aids the intuition behind the resulting willingness to pay function.

selection equation to be random such that $\alpha_1 = \bar{\alpha}_1 + \epsilon_{1i}^\alpha$, and the coefficient on the any-policy indicator in the policy choice equations also to be random so that $\beta_{1i} = (\bar{\beta}_1 + \epsilon_{1i}^\beta)$. Thirdly, we can allow these two random coefficients to be correlated so that $\rho(\epsilon_{1i}^\alpha, \epsilon_{1i}^\beta) \neq 0$.

When the intercept coefficient in the selection equation becomes random because α_1 is replaced by $(\bar{\alpha}_1 + \epsilon_{1i}^\alpha)$, the ϵ_{1i}^α portion effectively becomes a component of the error term in the selection equation, which is now $(\epsilon_{1i}^\alpha + \omega_i)$. Likewise, if the coefficient β_1 on $1(\text{any policy}_i^j)$ is permitted to be random, it is replaced by $(\bar{\beta}_1 + \epsilon_{1i}^\beta)$. The random portion of the parameter can be collected with the original equation error to yield a new compound error term, $(\epsilon_{1i}^\beta + \eta_i)$, for the policy choice utility-differences. If ϵ_{1i}^α and ϵ_{1i}^β are correlated, so are $(\epsilon_{1i}^\alpha + \omega_i)$ and $(\epsilon_{1i}^\beta + \eta_i)$, even if the ordinary conditional logit error terms ω_i and η_i are uncorrelated.

The crux of our approach is recognition that a standard logistic distribution can be closely approximated by a normal distribution with mean zero and $\sigma = \pi/\sqrt{3}$. Thus the sum of a normal and a standard logistic error term in each of the two compound errors in our joint model for selection and policy choices can be approximated by an appropriate normal distribution. This approximation allows our specification to mimic the simple correlation in the error terms in the selection and outcome equations that motivates the conventional Heckman-type selection-correction model in the least-squares context.⁹

Calculation of the covariance between the approximately normal compound error terms $(\epsilon_{1i}^\alpha + \omega_i)$ and $(\epsilon_{1i}^\beta + \eta_i)$ is simplified because all four random terms are mean-zero and the two standard logistic errors, ω_i and η_i , are distributed

⁹This mixed logit framework has also been used to mimic nested logit models. Researchers introduce alternative-specific constants with mean-zero normally distributed coefficients if they wish to introduce error correlations across alternatives. Here, we employ a similar strategy to allow for error correlations between the selection equation and the utility-difference equations in the policy-choice portion of the model.

independently from the two ϵ terms.

$$\begin{aligned}
Cov[(\epsilon_{1i}^\alpha + \omega_i), (\epsilon_{1i}^\beta + \eta_i)] &= E[(\epsilon_{1i}^\alpha + \omega_i)(\epsilon_{1i}^\beta + \eta_i)] - E[(\epsilon_{1i}^\alpha + \omega_i)]E[(\epsilon_{1i}^\beta + \eta_i)] \quad (4.6) \\
&= \left(E[\epsilon_{1i}^\alpha \epsilon_{1i}^\beta] + E[\epsilon_{1i}^\beta \omega_i] + E[\epsilon_{1i}^\alpha \eta_i] + E[\omega_i \eta_i] \right) \\
&\quad - \left(E[\epsilon_{1i}^\alpha] + E[\omega_i] \right) \left(E[\epsilon_{1i}^\beta] + E[\eta_i] \right) \\
&= \left(E[\epsilon_{1i}^\alpha \epsilon_{1i}^\beta] + 0 + 0 + 0 \right) - \left((0 + 0)(0 + 0) \right) \\
&= E[\epsilon_{1i}^\alpha \epsilon_{1i}^\beta]
\end{aligned}$$

Converting this covariance into a correlation involves dividing by the standard deviations of the two compound error terms, where the variance of each compound error is the variance of the corresponding ϵ plus the variance of a normal approximation to each of the independent standard logistic errors:

$$\rho\left((\epsilon_{1i}^\alpha + \omega_i), (\epsilon_{1i}^\beta + \eta_i)\right) = \frac{E[\epsilon_{1i}^\alpha \epsilon_{1i}^\beta]}{\left((\sigma_\epsilon^\alpha)^2 + \left(\frac{\pi^2}{3}\right)\right)^{.5} \left((\sigma_\epsilon^\beta)^2 + \left(\frac{\pi^2}{3}\right)\right)^{.5}} \quad (4.7)$$

where this correlation will be smaller than the correlation between just the two random parameters in the model:

$$\rho\left(\epsilon_{1i}^\alpha, \epsilon_{1i}^\beta\right) = \frac{E[\epsilon_{1i}^\alpha \epsilon_{1i}^\beta]}{\sigma_\epsilon^\alpha \sigma_\epsilon^\beta} \quad (4.8)$$

Spelling out the structure of our stacked model with strategically selected random parameters can cement the intuition behind our approach to selection correction in models for choice experiments. If the intercept in the selection portion of the model (α_1) and the coefficient on the “status quo” effect in the policy choice portion of the model (β_1) are permitted to be random and correlated, the model in equation (4.5)

becomes:

$$\begin{aligned}
\Delta V_i^r &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(1) + \alpha_2 z_i + (\bar{\beta}_1 + \epsilon_{1i}^\beta)(0) + \beta_2(0) + \omega_i & (4.9) \\
\Delta V_i^A &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)(1) + \beta_2 \Delta x_i^A + \eta_i^A \\
\Delta V_i^B &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)(1) + \beta_2 \Delta x_i^B + \eta_i^B \\
&\vdots \\
\Delta V_i^{J-1} &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)(1) + \beta_2 \Delta x_i^{J-1} + \eta_i^{J-1}
\end{aligned}$$

Rearranging, dropping any resulting error components that will be zero, and assuming mean-zero and symmetric distributions for $\epsilon_{1i}^\alpha, \epsilon_{1i}^\beta$, this model is equivalent to:

$$\begin{aligned}
\Delta V_i^r &= \bar{\alpha}_1(1) + \alpha_2 z_i + \bar{\beta}_1(0) + \beta_2(0) + (\epsilon_{1i}^\alpha + \omega_i) & (4.10) \\
\Delta V_i^A &= \bar{\alpha}_1(0) + \alpha_2(0) + \bar{\beta}_1(1) + \beta_2 \Delta x_i^A + (\epsilon_{1i}^\beta + \eta_i^A) \\
\Delta V_i^B &= \bar{\alpha}_1(0) + \alpha_2(0) + \bar{\beta}_1(1) + \beta_2 \Delta x_i^B + (\epsilon_{1i}^\beta + \eta_i^B) \\
&\vdots \\
\Delta V_i^{J-1} &= \bar{\alpha}_1(0) + \alpha_2(0) + \bar{\beta}_1(1) + \beta_2 \Delta x_i^{J-1} + (\epsilon_{1i}^\beta + \eta_i^{J-1})
\end{aligned}$$

If ϵ_{1i}^α and ϵ_{1i}^β are correlated, there will now be a correlation between the compound error term $(\epsilon_{1i}^\alpha + \omega_i)$ for the selection equation and the compound error terms $(\epsilon_{1i}^\beta + \eta_i^j)$ for $j = 1, \dots, J-1$ for the choice model, even though there is no correlation among the usual i.i.d. η_i^j conditional logit error terms for each alternative or between the η_i^j errors and the ω_i error term in the ordinary selection equation.

The conventional Heckman correction for sample selection assumes a normally distributed outcome variable. In the case of systematic sample selection in conditional logit choice models, the error terms in the expressions for ΔV_i^A (the η_i^A), have traditionally been assumed to be i.i.d. standard logistic, which has precluded correlations with any other error terms. If we were to estimate

our selection model using a probit estimator, allowing the noise in the intercept parameter to be combined with the regular normally distributed probit error term, the model for the selection data would be:

$$\begin{aligned}\Delta V_i^r &= \bar{\alpha}_1(1) + \alpha_2 z_i + (\epsilon_{1i}^\alpha + \omega_i) \\ &= \alpha z_i + (\epsilon_{1i}^\alpha + \omega_i)\end{aligned}\tag{4.11}$$

where $\alpha z_i = (\bar{\alpha}_1 \ \alpha_2)(1 \ z_i)'$ the normally distributed probit error term would be a normal approximation to the sum of the independent random error terms ϵ_{1i}^α and ω_i .

In the selection submodel of our pooled mixed-logit model, the sum of the independent normal and standard logistic errors ($\epsilon_{1i}^\alpha + \omega_i$) can be well-approximated with a normal distribution. The coefficient on the intercept in the selection submodel is assumed by our mixed-logit specification to be distributed normally with mean $\bar{\alpha}_1$ and standard deviation σ_ϵ^α). Using the normal approximation to a standard logistic, the composite error in the selection submodel, $(\epsilon_{1i}^\alpha + \omega_i)$, is distributed approximately normal with mean 0 and standard deviation $\left((\sigma_\epsilon^\alpha)^2 + \left(\frac{\pi^2}{3}\right)\right)^{.5}$.

If we stack the selection model and the policy-choice model in a pooled mixed logit specification, we get the following (where we make explicit the fact that to yield standard logistic distributions for the ω_i and the η_i^j error terms, each equation must be normalized by dividing through by the corresponding scale parameter for the implicit non-standardized logistic error, κ_ω and κ_η , respectively):

$$\begin{bmatrix} V_i^{*r}/\kappa_\omega \\ \Delta V_i^{*A}/\kappa_\eta \\ \vdots \\ \Delta V_i^{*J-1}/\kappa_\eta \end{bmatrix} = \begin{bmatrix} z_i & 0 \\ 0 & \Delta x_i^{*A} \\ \vdots & \vdots \\ 0 & \Delta x_i^{*J-1} \end{bmatrix} \begin{bmatrix} \alpha/\kappa_\omega \\ \beta/\kappa_\eta \\ \vdots \\ \beta/\kappa_\eta \end{bmatrix} + \begin{bmatrix} (\epsilon_{1i}^\alpha + \omega_i) \\ (\epsilon_{1i}^\beta + \eta_i^A) \\ \vdots \\ (\epsilon_{1i}^\beta + \eta_i^{J-1}) \end{bmatrix} \quad (4.12)$$

In equation (4.12), data vector z_i for each respondent includes an intercept term with a random coefficient that has an estimated mean value $\bar{\alpha}_1$ and a standard deviation of σ_ϵ^α . The fixed portion of the intercept parameter, $\bar{\alpha}_1$, is then subsumed in the parameter vector α . Likewise, each of the policy alternatives shares a common increment/decrement to utility relative to the status quo alternative, so each utility difference equation for non-status-quo alternatives includes among its Δx_i^j variables an indicator variable with a normally distributed random coefficient that has mean $\bar{\beta}_1$ and standard deviation σ_ϵ^β . The $\bar{\beta}_1$ coefficient is subsumed within the β coefficient vector, and the ϵ_{1i}^β random component is combined with the η_i^j errors in equation (4.12). To simplify the notation in what follows, the parameter vectors α/κ_ω and β/κ_η will be denoted simply as α and β , since these parameter vectors can only ever be identified up to its respective scale factor.

For the selection submodel, the mixed-logit pooled estimation gives us an estimate of the standard deviation of ϵ_{1i}^α , namely $\hat{\sigma}_\epsilon^\alpha$. If we approximate the standard logistic error ω in the selection submodel with a mean-zero normal distribution, the standard deviation of that independent normally distributed component is about $\pi/\sqrt{3}$. Thus the *effective* compound error in the selection submodel, $(\epsilon_{1i}^\alpha + \omega_i)$, has an approximately normal distribution with mean zero and standard deviation of approximately $= \left((\hat{\sigma}_\epsilon^\alpha)^2 + \left(\frac{\pi^2}{3}\right) \right)^{.5}$. Call this standard deviation $\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}$.

Analogously, for the pooled-data mixed logit model with correlated random coefficients on the intercept in the selection equation and the any-policy indicator in the outcome equation, the effective compound error in the outcome submodel is $(\epsilon_{1i}^\beta + \eta_i)$. This is also the sum of the normally distributed ϵ_{1i}^β and the independent standard logistic (0,1) error term η_i , whereas in the pooled model, both ω_i and η_i are constrained to have the same implicit dispersion, $\kappa_{\omega\eta}$ that scales the equation so that the resulting error is standard logistic. Thus the compound error, $(\epsilon_{1i}^\beta + \eta_i)$ has an approximately normal distribution with mean zero and standard deviation $((\hat{\sigma}_\epsilon^\beta)^2 + (\frac{\pi^2}{3}))^{.5}$. Call this standard deviation $\hat{\sigma}_{(\epsilon_{1i}^\beta + \eta_i)}$.

4.4.3.1 *Analog to probit selection model in conventional*

Heckman correction. If we were to estimate the selection submodel

independently, using a conventional binary probit specification, we would get:

$$\Delta V_i^{*r} = \left(\frac{\alpha}{\sigma_u} \right) z_i + u_i, \quad \text{where } u_i \sim N(0, 1) \quad (4.13)$$

where σ_u normalizes the scale of the equation so that the error variance is one.

In contrast, in our pooled-data model with random coefficients on the selection-equation intercept and the status quo indicator in the outcome submodel, the selection submodel is estimated as:

$$\Delta V_i^{*r} = \alpha z_i + (\epsilon_{1i}^\alpha + \omega_i), \quad (4.14)$$

where $\epsilon_{1i}^\alpha \sim N(0, \sigma_{\epsilon^\alpha}^2)$ and $\omega_i \sim \text{logistic}(0, 1)$

so that $(\epsilon_{1i}^\alpha + \omega_i) \sim N(0, \sigma_{(\epsilon_{1i}^\alpha + \omega_i)}^2)$ (approximately)

$$\text{where } \sigma_{(\epsilon_{1i}^\alpha + \omega_i)}^2 = \hat{\sigma}_{\epsilon^\alpha}^2 + \left(\frac{\pi^2}{3} \right)$$

keeping in mind, of course, that the conditional logit parameters α and the error term ω_i have already been normalized on κ_ω , the dispersion of ω_i , the logistic component of the error before its standardization.

To convert the equation in (4.14) so that it approximately matches the equation in (4.13) with its standard-normal error, we need to divide equation (4.14) by its error standard deviation, which yields:

$$\frac{\Delta V_i^{*r}}{\sigma_{(\epsilon_{1i}^\alpha + \omega_i)}} = \left(\frac{\alpha}{\sigma_{(\epsilon_{1i}^\alpha + \omega_i)}} \right) z_i + \left(\frac{\epsilon_{1i}^\alpha + \omega_i}{\sigma_{(\epsilon_{1i}^\alpha + \omega_i)}} \right) \quad (4.15)$$

The selection equation estimated via mixed logit using the pooled selection and policy-choice data, with correlated coefficients on the intercept of the selection submodel and the status quo indicator in the policy-choice equation, yields estimates of the mean coefficient on the intercept of the selection submodel and all the slopes in the selection submodel. Thus, to approximate what would be the systematic portion of the equivalent probit selection model, one need only divide all the mixed logit coefficients for the selection submodel by $\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}$ which can be calculated using the first diagonal element of the variance-covariance matrix for the random parameters in the joint model.

4.4.3.2 Analog to a normal-error “outcome” model in conventional Heckman correction. The same strategy can be used to infer the analogous probit-like parameters for the policy-choice model, if the standard deviation is calculated for the compound error from the pooled mixed logit model when the coefficient on the any-policy indicator is allowed to be random. If there were no systematic selection, the indirect utility-differences for the policy-choice submodel would all take the form:

$$\Delta V_i^{*j} = \beta \Delta x_i + (\epsilon_{1i}^\beta + \eta_i), \quad (4.16)$$

where $\epsilon_{1i}^\beta \sim N(0, \sigma_{\epsilon^\beta}^2)$ and $\eta_i \sim \text{logistic}(0, 1)$

and $(\epsilon_{1i}^\beta + \eta_i) \sim N\left(0, \sigma_{(\epsilon_{1i}^\beta + \eta_i)}^2\right)$

where $\sigma_{(\epsilon_{1i}^\beta + \eta_i)}^2 = \hat{\sigma}_{\epsilon^\beta}^2 + \left(\frac{\pi^2}{3}\right)$

keeping in mind, again, that the conditional logit parameters β and the error term η have already been normalized on κ_η , the dispersion of the logistic component of the error, η_i , before its standardization.

The the probit analog to the policy-choice submodel of the pooled mixed logit specification is:

$$\frac{\Delta V_i^{*j}}{\sigma_{(\epsilon_{1i}^\beta + \omega_i)}} = \left(\frac{\beta}{\sigma_{(\epsilon_{1i}^\beta + \omega_i)}} \right) \Delta x_i + \left(\frac{\epsilon_{1i}^\alpha + \omega_i}{\sigma_{(\epsilon_{1i}^\beta + \omega_i)}} \right) \quad (4.17)$$

4.4.3.3 *Accommodating the truncation due to sample selection.*

With sample selection, the policy choices are observed only if the the unobserved selection propensity in the selection submodel, ΔV_i^{*r} , is large enough so that a response is provided by that person. The approximate bivariate normality of the compound effective error terms in the selection submodel and the policy-choice submodel makes it possible for us to employ an analog to Heckman's two-step method, where the correction is based on the expected value of a singly truncated bivariate normal joint density. A correction term must be added to the vector of policy attributes in the policy-choice submodel. This correction term is the nonselection hazard (also known as the inverse of the Mills ratio, IMR). The nonselection hazard is computed from the estimated coefficients for the selection submodel after these coefficients have been scaled so that the corresponding error terms for that portion of our pooled mixed-logit model are distributed approximately standard normal.

For the selection submodel with a random coefficient on the intercept term, mixed-logit estimates for the pooled model will provide estimates of the coefficient vector $\hat{\alpha}$ and the standard error of the random intercept, $\hat{\sigma}_{\epsilon^\alpha}$:

$$\Delta V_i^{*r} = \alpha z_i + (\epsilon_{1i}^\alpha + \omega_i), \text{ where } \hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)} \approx \hat{\sigma}_{\epsilon^\alpha} + \left(\frac{\pi}{\sqrt{3}} \right) \quad (4.18)$$

Given the approximate normality of the error term and its standard deviation (calculated from the standard deviation of the random intercept and the standard deviation of the logistic equation error), that standard deviation can be used to scale the equation so that the resulting error is approximately standard normal. The scaled value of the selection-equation “index” can be used to construct the relevant inverse Mills ratio needed to accommodate the truncation of the approximately bivariate normal distribution due to selection:

$$\widehat{IMR}_i = \frac{\phi\left(\frac{-\hat{\alpha}z_i}{\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}}\right)}{1 - \Phi\left(\frac{-\hat{\alpha}z_i}{\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}}\right)} = \frac{\phi\left(\frac{\hat{\alpha}z_i}{\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}}\right)}{\Phi\left(\frac{\hat{\alpha}z_i}{\hat{\sigma}_{(\epsilon_{1i}^\alpha + \omega_i)}}\right)} \quad (4.19)$$

where ϕ and Φ are the probability density and cumulative density functions of the standard normal distribution.

The formula for the \widehat{IMR}_i can then be included in each of utility-difference equations in the outcome submodel to remove the selection bias in the estimated β coefficients that could otherwise distort the inferences drawn from the outcome submodel:¹⁰

$$\Delta V_i^{*j} = \beta \Delta x_i + \left(\rho \sigma_{(\epsilon_{1i}^\beta + \eta_i)}\right) \times \widehat{IMR}_i + (\epsilon_{1i}^\beta + \eta_i), \quad (4.20)$$

$$\text{where } (\epsilon_{1i}^\beta + \eta_i) \sim N(0, \sigma_{(\epsilon_{1i}^\beta + \eta_i)}^2) \text{ and } \sigma_{(\epsilon_{1i}^\beta + \eta_i)} \approx \hat{\sigma}_{\epsilon^\beta} + \left(\frac{\pi}{\sqrt{3}}\right)$$

If needed, an approximate estimate of the error correlation ρ between the selection submodel and the policy-choice submodel can then be recovered using the estimate of the approximate value of the standard deviation of the compound error.

4.4.3.4 *FIML estimation by a generalization of mixed logit*

models. Full information maximum likelihood estimation is difficult to accomplish with basic packaged mixed-logit models, but we provide sample code demonstrating

¹⁰The \widehat{IMR}_i variable is not alternative-specific. It is most convenient to interact this variable with the $1(\text{any policy}^j)$ or $1(\text{status quo}_j)$ variable or it will drop out in the differencing process.

how to accomplish this task using the Apollo package for R. The full stacked model with all parameters estimated simultaneously substitutes the formula for \widehat{IMR}_i directly into the policy-choice submodel and all of the unique parameters (α , $\sigma_{(\epsilon^\alpha)}$, β , $\sigma_{(\eta^\beta)}$ and ρ) feature in a generalization of the likelihood to be estimated analogously to an ordinary mixed logit model, maximized simultaneously with respect to all of these parameters. Note that the ρ parameter that enters into in the policy-choice equations as part of the coefficient on the IMR term, with its formula given in equation (4.8), needs to be consistent via parameter constraints with the estimated correlation between the pair of random parameters on the intercept in the selection submodel and on the $1(\text{any policy}^j)$ variable in the policy-choice submodel, with the formula given in equation (4.7)

4.4.4 When sample selection also affects the marginal utility of policy attributes in the policy-choice submodel. Including only one random parameter in the policy-choice submodel, solely on an any-policy indicator, preserves homoskedasticity within the policy-choice portion of the pooled-data model. However, we may also wish to allow the *marginal utilities* of specific program attributes, measured up to a scale factor by the coefficients β_k on the attribute differences Δx_{ki}^j , to covary with response propensities. If respondents have systematically different marginal utilities than non-respondents, the effective error structure becomes more complex. To streamline notation, consider a random-parameters specification for the policy-choice model that allows (potentially) all of the marginal utilities of the policy attributes to be correlated with response

propensities via the selection-equation intercept. The pooled model becomes:

$$\begin{aligned}
\Delta V_i^r &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(1) + \alpha_2 z_i + (\bar{\beta}_1 + \epsilon_{1i}^\beta)(0) + \dots + (\bar{\beta}_k + \epsilon_{ki}^\beta)(0) + \omega_i \quad (4.21) \\
\Delta V_i^A &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)\Delta x_{1i}^A + \dots + (\bar{\beta}_k + \epsilon_{ki}^\beta)\Delta x_{ki}^A + \eta_i^A \\
\Delta V_i^B &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)\Delta x_{1i}^B + \dots + (\bar{\beta}_k + \epsilon_{ki}^\beta)\Delta x_{ki}^B + \eta_i^B \\
&\vdots \\
\Delta V_i^{J-1} &= (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(0) + \alpha_2(0) + (\bar{\beta}_1 + \epsilon_{1i}^\beta)\Delta x_{1i}^{J-1} + \dots + (\bar{\beta}_k + \epsilon_{ki}^\beta)\Delta x_{ki}^{J-1} + \eta_i^{J-1}
\end{aligned}$$

Or, we can rearrange these equations to collect the error terms, dropping the error components that will be zero. We can subsume the means of the random parameters, $\bar{\alpha}_1$ in the selection model and $\bar{\beta}_1, \dots, \bar{\beta}_k$ in the policy-choice submodel, within the coefficient vectors α on the selection-equation variables z_i and β on the policy-choice variables Δx_i^j in the systematic portion of the two submodels, but retain the detail in the compound error terms.

$$\begin{aligned}
\Delta V_i^r &= \alpha z_i + \beta(0) + (\epsilon_{1i}^\alpha + \omega_i) \quad (4.22) \\
\Delta V_i^A &= \alpha(0) + \beta \Delta x_i^A + (\epsilon_{1i}^\beta \Delta x_{1i}^A + \dots + \epsilon_{ki}^\beta \Delta x_{ki}^A + \eta_i^A) \\
\Delta V_i^B &= \alpha(0) + \beta \Delta x_i^B + (\epsilon_{1i}^\beta \Delta x_{1i}^B + \dots + \epsilon_{ki}^\beta \Delta x_{ki}^B + \eta_i^B) \\
&\vdots \\
\Delta V_i^{J-1} &= \alpha(0) + \beta \Delta x_i^{J-1} + (\epsilon_{1i}^\beta \Delta x_{1i}^{J-1} + \dots + \epsilon_{ki}^\beta \Delta x_{ki}^{J-1} + \eta_i^{J-1})
\end{aligned}$$

Of course, not all of the marginal utility (β) coefficients need to be random, and the Δx_i^j variables may contain an any-policy indicator, or even a full set of alternative-specific constants, so this model subsumes our earlier simple model.

For a Heckman-like IMR term to be appropriate for selection correction, it is again the case that the error terms in the selection submodel and the policy-choice submodel would need to be jointly normal. But now the pooled mixed-logit model has a random parameter for the intercept in the selection submodel and

(potentially) random parameters for all k of the β parameters in the policy-choice submodel. The mixed-logit algorithm will produce estimates for the covariance matrix of this set of $(1 + k)$ random parameters. The mean-zero random parameter error terms that are components of the compound errors for pooled mixed-logit model will share this covariance matrix.

Within the selection submodel, the compound error term $(\epsilon_{1i}^\alpha + \omega_i)$ remains homoskedastic and the selection model can again be scaled by the standard deviation of the approximately normal compound error term to permit calculation of an approximate *IMR* term to be appended to the list of regressors in the policy-choice submodel. However, the compound error terms in the policy-choice submodel are no longer homoskedastic due to the presence of the Δx_{ki}^j variables multiplying the random components, ϵ_{ki}^β , of the marginal utility parameters. The scale of the error term for a given individual i and alternative j thus depends upon the levels of these differenced policy attributes, Δx_i^j , for that ij^{th} observation. Converting these heteroskedastic errors to an approximately standard normal error for each ij is a little more complicated.

A pooled mixed-logit model will produce point estimates for the fixed parameters and the means of the random parameters, as well as estimates of the $(1 + k) \times (1 + k)$ covariance matrix for the random parameters on the intercept in the selection submodel and the differenced policy attributes in the policy-choice submodel. Call this $(1 + k) \times (1 + k)$ covariance matrix $\Sigma_{\alpha\beta}$. We can again sum the parameter variance for the intercept of the selection submodel (in the $[1, 1]$ element of $\Sigma_{\alpha\beta}$) and the $\left(\frac{\pi^2}{3}\right)$ variance of the normal approximation to the independent standard logistic error to get the variance of the effective approximately normal compound error in the selection submodel. The corresponding standard deviation can again be used to scale the “index” for the selection equation before it is used in

the IMR formula to yield the appropriate selection-correction term for the policy-choice model.

However, we also need to take steps to make the policy-choice submodel homoskedastic with approximately standard-normal error terms. For this, we need to extract the $k \times k$ sub-matrix of the full random-parameters covariance matrix that corresponds to the random coefficients on the attributes in the policy-choice model. Call this covariance sub-matrix Σ_β . For alternative j in choice task i , the error term will be:

$$e_i^j = (\epsilon_{1i}^\beta \Delta x_{1i}^j + \dots + \epsilon_{ki}^\beta \Delta x_{ki}^j + \eta_i^j) \quad (4.23)$$

where the random components of the marginal utility coefficients, $\epsilon_{1i}^\beta, \dots, \epsilon_{ki}^\beta$ are correlated, but the standard logistic equation error, η_i^j , is distributed independently from the marginal utility coefficients. For each alternative within each choice, we can treat the $\Delta x_{1i}^j, \dots, \Delta x_{ki}^j$ data as fixed values, so that the distribution of the error term, for each alternative j within each choice i , is a linear combination of k correlated mean-zero normal random parameter error components and one independent standard logistic random equation error term, which can be approximated by a normally distributed error with $\sigma = \pi/\sqrt{3}$.

We can calculate the (scalar) variance of each compound error term in the policy-choice submodel, e_{ij} , as follows:

$$Var(e_{ij}) = \begin{bmatrix} \Delta x_{1i}^j & \dots & \Delta x_{ki}^j \end{bmatrix} \begin{bmatrix} \Sigma_\beta \end{bmatrix} \begin{bmatrix} \Delta x_{1i}^j \\ \dots \\ \Delta x_{ki}^j \end{bmatrix} + \left(\frac{\pi^2}{3}\right) \quad (4.24)$$

If we divide all of the Δx_{1i}^j through Δx_{ki}^j variables for each ij^{th} observation in the policy-choice submodel by the square root of the corresponding $Var(e_{ij})$, the resulting equation error will be rendered homoskedastic and i.i.d., with errors that

are distributed approximately $N(0,1)$. However, we cannot rescale the Δx_{1i}^j through Δx_{ki}^j variables separately for each alternative within a choice occasion, or we risk re-ordering i 's preferences.¹¹ To avoid this, we instead calculate a single rescaling factor for all policies within a choice set:

$$Var(e_i) = \begin{bmatrix} \overline{\Delta x_{1i}} & \dots & \overline{\Delta x_{ki}} \end{bmatrix} \begin{bmatrix} \Sigma_\beta \end{bmatrix} \begin{bmatrix} \overline{\Delta x_{1i}} \\ \dots \\ \overline{\Delta x_{ki}} \end{bmatrix} + \left(\frac{\pi^2}{3}\right) \quad (4.25)$$

where $\overline{\Delta x_{ki}}$ is the mean across alternatives of the $J - 1$ different Δx_{ki}^j terms in a given choice occasion for individual i .

Along with the transformation of the selection submodel, this transformation of the policy-choice submodel creates a suitable context for the intuition of Heckman's two-step approach to apply. Note, however, that the transformation of the data in the policy-choice submodel is not optional, as it was for the simpler specification we considered first, where the effective error term in the policy-choice submodel remained homoskedastic when only the any-policy effect was modeled as random across respondents.

The ratios of the estimated *means* of the random coefficients on the attributes in the policy-choice submodel can then be used to compute marginal willingness-to-pay measures. The precision of the estimates of these means should be factored in, as always, by an appropriate strategy for calculating the mean of a ratio of random variables (e.g. the delta method, Feiller's method, or the Krinsky-Robb approach). Given the computation time required to estimate one

¹¹To see how this can occur, consider an extreme case. Suppose an individual has indirect utility function $V_i = -x_1 + 2x_2$, and suppose we have estimated $\Sigma_\beta = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. The individual chooses between policy A ($x_1 = 2, x_2 = 1.1$), policy B ($x_1 = 200, x_2 = 101$), and a status quo policy N ($x_1 = 0, x_2 = 0$). Then $0 < \Delta V^A = 0.2 < \Delta V^B = 2$, and the individual chooses policy B . But $\sqrt{Var(e_{iA})} = 3.6$ and $\sqrt{Var(e_{iB})} = 265.3$, so our (separately) rescaled indirect utilities are reversed: $\frac{V^A}{\sqrt{Var(e_{iA})}} > \frac{V^B}{\sqrt{Var(e_{iB})}}$.

set of parameters for this model, it is likely impractical at typical processing speeds to attempt conventional bootstrap resampling-with-replacement methods for calculating interval estimates for the desired marginal willingness-to-pay estimates.¹²

4.4.5 Joint estimation of the selection and policy choice

submodels. When only the coefficient on the any-policy indicator in the choice data is allowed to be random and correlated with the intercept in the selection model, there are just two random parameters in the joint model, α_1 and β_k , which we assume have a correlated joint distribution characterized by five parameters $(\mu_1^\alpha, \mu_k^\beta, \sigma_\alpha, \sigma_\beta, \rho)$. The numeraire alternative in the selection choices is the non-response alternative, and the numeraire alternative in the policy choices is the status-quo alternative. For the simplest pooled model, assume fixed marginal utilities, α_2 , for each of the z_i explanatory variables in the selection model, and fixed marginal utilities of net income, β_1 and the other policy attributes, β_2 , in the policy-choice model.

Now let $S_{it} = 1$ indicate that choice task t for individual i is a response/non-response (i.e., selection model) choice, so that $S_{it} = 0$ for each of individual i 's policy choice(s). Then the stacked model can be written as follows, using S_{it} as a switch to activate different types of variables:

$$\begin{aligned} \Delta V_{it}^j &= \alpha_1 S_{it} + \alpha_2 z_{it} S_{it} + \beta_1 (-cost_{it}^A)(1 - S_{it}) \\ &+ \beta_2 \Delta x_{it}^A (1 - S_{it}) + \beta_k (1)(1 - S_{it}) + \eta_{it}^j \end{aligned} \tag{4.26}$$

¹²The first simulation described in section 4.5 took 21 hours 47 minutes total computing time, using 15 of 16 available threads running in parallel on a Ryzen 7 processor. Conditional logit estimation without selection correction on the same simulated data takes 0.29 seconds.

for alternative j (relative to the numeraire alternative) in choice task t by individual i . The coefficient β_k in the utility-differences for the policy-choice submodel is the any-policy effect in the policy choices.

The conditional probability that individual i in the sample of respondents will choose alternative j in choice task t , for given values of α_{10} and β_{20} is:

$$P_{it}^j | (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_k) = \frac{\exp(\alpha_1 S_{it} + \alpha_2 z_{it} S_{it} + \beta_1 (-\text{cost}_{it}^j)(1 - S_{it}) + \beta_2 \Delta x_{it}^j (1 - S_{it}) + \beta_k (1)(1 - S_{it}))}{\sum_{m=1}^{J_t^i} [\exp(\alpha_1 S_{it} + \alpha_2 z_{it} S_{it} + \beta_1 (-\text{cost}_{it}^m)(1 - S_{it}) + \beta_2 \Delta x_{it}^m (1 - S_{it}) + \beta_k (1)(1 - S_{it}))]} \quad (4.27)$$

To get the unconditional probability of choosing alternative j , we need to integrate out the four random parameters, i.e. α_1 , β_1 , β_2 , and β_k (but not α_2 , which we model as a fixed parameter).

$$P_{it}^j = E(P_{it}^j | (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_k)) = \int_{\alpha_1} \int_{\beta_1} \int_{\beta_2} \int_{\beta_k} (P_{it}^j | (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_k)) f(\bar{\mu}, \Sigma) d\beta_k d\beta_2 d\beta_1 d\alpha_1 \quad (4.28)$$

where $f(\bar{\mu}, \Sigma)$ is the joint density of the random parameters $(\alpha_1, \beta_1, \beta_2, \beta_k)$ that accommodate selection on unobservables in our joint model of response/non-response decisions and subsequent policy choice by respondents.

If we let $y_{it}^j = 1$ if alternative j is chosen by individual i for choice task t out of T_i choices observed for that individual, the log-likelihood function for the choice model will be:

$$\text{Log}\mathcal{L} = \sum_{i=1}^n \left(\sum_{t=1}^{T_i} \left(\sum_{m=1}^{J_t^i} [y_{it}^m \log(P_{it}^m)] \right) \right) \quad (4.29)$$

where T_i is the number of choice tasks observed for individual i and J_i^t is the number of alternatives in each choice task for that individual. For *non-respondents*, $T_i = 1$ and $J_i^1 = 2$ because we can estimate only the selection equation for this

group. For respondents, T_i will be the number of policy choices they are observed to make, plus one (for their response/non-response choice).

4.5 Simulations

To determine the reliability of our estimation strategy, we ran several simulations. Specifically, we take the pooled model described in 4.21 as the data generating process, drawing correlated ϵ 's such that simulated individuals with higher marginal WTP for the policy's beneficial attribute (which we call, generically, *benefit*) are more likely to respond, thereby biasing upward an uncorrected estimate of marginal WTP. Given the policy choices of a non-representative sub-sample of responders and minimal data on non-responders, we can estimate a selection-corrected marginal WTP for the policy's benefit attribute and compare it to the true marginal WTP in the simulated population. For simplicity, we set $J = 3$ as the number of alternatives and $k = 3$ as the number of attributes, which are an *any policy* effect, *cost*, and *benefit*.

4.5.1 Data generating process. The resulting data generating process is as follows:

$$\Delta V_i^r = (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)(1) + \alpha_2 z_i + \omega_i \tag{4.30}$$

$$\Delta V_i^A = (\bar{\beta}_1 + \epsilon_{1i}^\beta)(1) + (\bar{\beta}_2 + \epsilon_{2i}^\beta)(cost_i^A) + (\bar{\beta}_3 + \epsilon_{3i}^\beta)(benefit_i^A) + \eta_i^A$$

$$\Delta V_i^B = (\bar{\beta}_1 + \epsilon_{1i}^\beta)(1) + (\bar{\beta}_2 + \epsilon_{2i}^\beta)(cost_i^B) + (\bar{\beta}_3 + \epsilon_{3i}^\beta)(benefit_i^B) + \eta_i^B$$

where $\bar{\alpha}_1 = 0$, $\alpha_2 = 1$, $\bar{\beta}_1 = 1$ (utility from *any policy*), $\bar{\beta}_2 = -5$ (marginal disutility of *cost*), and $\bar{\beta}_3 = 5$ (marginal utility of *benefit*). The exogenous (scalar) variable z_i is distributed $N(0, 1)$. Error terms ω_i , η_i^A , and η_i^B are i.i.d. $Logistic(0, 1)$. The individual-specific random components for the preference parameters— ϵ_{1i}^α , ϵ_{1i}^β , ϵ_{2i}^β ,

and ϵ_{3i}^β —are each distributed $N(0, 1)$ with covariance matrix

$$\Sigma_{\alpha\beta} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 1 \end{bmatrix} \quad (4.31)$$

so that individual-specific marginal utilities (i.e. the individual-specific β s) are positively correlated with propensity to respond but uncorrelated with each other. For parsimony, let $\alpha_i = (\bar{\alpha}_1 + \epsilon_{1i}^\alpha)$, $\beta_i^{\text{any policy}} = (\bar{\beta}_1 + \epsilon_{1i}^\beta)$, $\beta_i^{\text{cost}} = (\bar{\beta}_2 + \epsilon_{2i}^\beta)$, and $\beta_i^{\text{benefit}} = (\bar{\beta}_3 + \epsilon_{3i}^\beta)$.

4.5.2 One large-sample simulation. Using the data generating process above, we draw individual characteristics (i.e., α_i , $\beta_i^{\text{status quo}}$, β_i^{cost} , β_i^{benefit} , z_i , ω_i , η_i^A , η_i^B) for a panel of 20,000 simulated invitees. Means for each of these parameters are described in Table 11 both for the full sample of invited participants and for the responder and non-responder sub-samples. Figure 6 presents the distributions of two key parameters— β_i^{benefit} and β_i^{cost} —in more detail.

Parameter	Population	Full sample	Responders	Non-responders
α_i	0	-0.013	0.351	-0.373
$\beta_i^{\text{any policy}}$	1	1.003	0.827	1.176
β_i^{cost}	-5	-5.015	-4.825	-5.203
β_i^{benefit}	5	5.004	5.181	4.829
N	-	20,000	9,947	10,053

Table 11. Parameter means for full sample and responder/non-responder sub-samples

As expected, given the design, the simulation’s response rate is approximately 50%.¹³ Note that responders receive more utility from the program’s benefit and are less cost-averse, on average, than are non-responders. These two

¹³Individuals respond if $V^r > 0$; V^r is distributed symmetrically and has an expected value of 0 in this simulation.

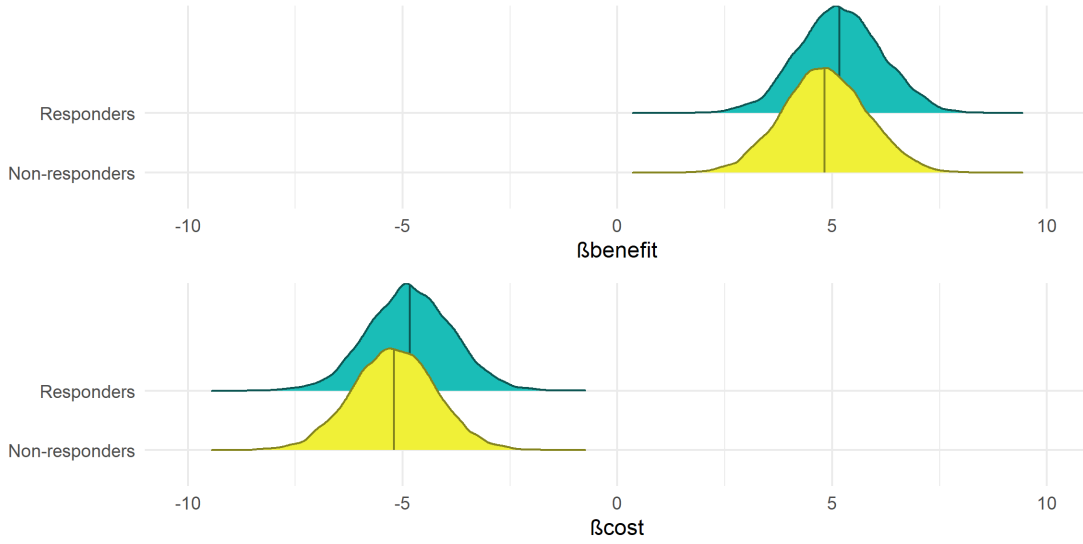


Figure 6. Preference parameter distribution for simulated responders and non-responders

considerations entail that estimates of WTP (total or marginal) for the program’s benefit based only on responders, and without correcting for selection effects, will be systematically higher than those in the population.

Individuals in the responder sub-sample each make one policy choice between alternatives A , B , and the status quo. Cost and benefit for alternatives A and B are i.i.d. $\text{Uniform}(0,3)$, and the any-policy indicators are set to 1. For the status quo policy, cost and benefit are 0 and the any-policy indicator is set to 0. Respondents choose the alternative that maximizes their utility. Individuals in the non-responder sub-sample (for whom $\Delta V_i^r \leq 0$) do not make a policy choice.

The distributions of marginal WTP for each sub-sample, which are not observable to the econometrician, are plotted in Figure 7, where marginal WTP for *benefit* for the i^{th} individual is calculated as $-\beta_i^{\text{benefit}}/\beta_i^{\text{cost}}$. Because the numerator and denominator of WTP are each normally distributed, the resulting distribution of this ratio is Cauchy. Cauchy distributions have no defined mean, so we use the

median to characterize the center of WTP distributions. In Figure 7, the median, rather than the mean, of each distribution is marked by a vertical line (selected means are reported in Table 13).

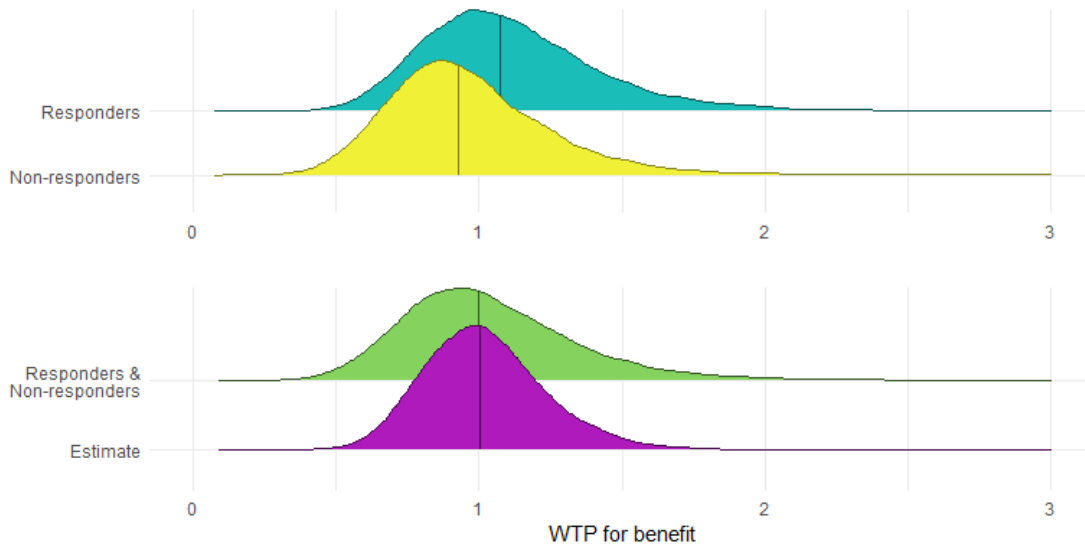


Figure 7. Distributions of marginal WTP for benefit in responder/non-responder sub-samples, full sample, and selection-corrected estimates using simulated data

The selection-corrected estimated distribution of WTP (Figure 7, in purple) comes from a Krinsky-Robb parametric bootstrap (Krinsky & Robb, 1990) of 20,000 draws of β_i^{benefit} and β_i^{cost} from the estimated *joint* distribution of α_i , $\beta_i^{\text{anypolicy}}$, β_i^{cost} , and β_i^{benefit} . The median of this distribution produces an unbiased estimate of the population median. However, the distribution of marginal WTP estimates for *benefit* produced by the Krinsky-Robb method is narrower than the actual distribution in the simulated population. The medians for the full sample and corrected estimate of marginal WTP for *benefit* are not statistically distinguishable at any conventional significance level.¹⁴ The median estimates for

¹⁴The *means* are statistically distinguishable at the 10% level, but given that the mean of a Cauchy distribution is undefined, this is not a significant concern.

marginal WTP for *any policy* are, however, significantly different between the full sample and the corrected estimate.

marginal WTP for...	Full Sample		Responders		Corrected estimate	
	Median	Mean	Median	Mean	Median	Mean
...Benefit	0.997	1.043	1.072	1.122	1.003	1.018
...Any policy	0.201	0.209	0.172	0.181	0.174	0.161
N	20,000		9,947		9,947	

Table 12. Simulated marginal WTPs for full sample, responder sub-sample, and selection-corrected parametric bootstrap

In addition to the relevant *marginal* WTP estimates, we may also wish to calculate *total* WTP for a given policy. Figure 8 presents the distributions of total WTP for a policy with *benefit* = 1.5 and *any policy* = 1 for the simulated responder and non-responder sub-samples (cyan and yellow, respectively), the full sample (green), and the Krinsky-Robb parametric bootstrap of 20,000 draws from the selection-corrected estimate (purple). For median total WTP, the corrected estimate represents a substantial improvement relative to the median total WTP in the responders-only subsample but is still distinguishable from the median total WTP in the full sample at any conventional significance level.

	Full Sample		Responders		Corrected estimate	
	Median	Mean	Median	Mean	Median	Mean
Total WTP	1.692	1.774	1.785	1.864	1.675	1.688
	(0.0041)	(0.0040)	(0.0066)	(0.0059)	(0.0022)	(0.0019)
N	20,000		9,947		9,947	

Table 13. Simulated total WTPs for full sample, responder sub-sample, and selection-corrected parametric bootstrap

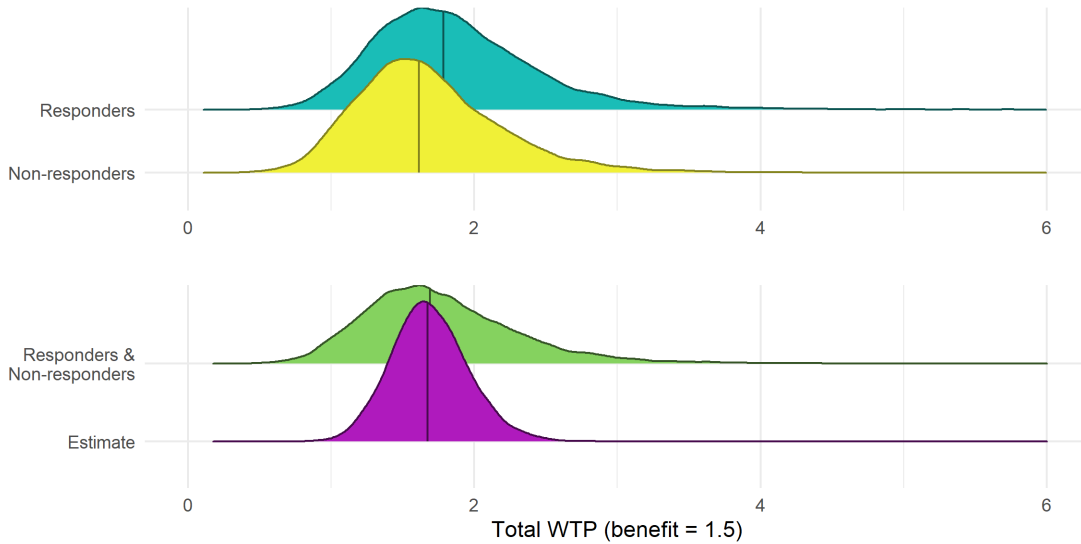


Figure 8. Distributions of total WTP for a policy with $benefit = 1.5$ and $any\ policy = 1$ in responder/non-responder sub-samples, full sample, and selection-corrected estimates using simulated data

4.5.3 Many small simulations. This simulation demonstrates that, given a large enough sample and an exogenous response-predictor z_i , our correction algorithm is able to recover an unbiased estimate of marginal WTP for a policy’s continuous attributes, even in the presence of non-random selection effects.

In practice, however, few researchers are able to achieve a sample size for respondents as large as 9,947. To approximate more realistic conditions, we repeat this simulation 100 times, each with 2,000 invitees, for an average responder sample of 1,000. For each simulation, we estimate marginal WTPs separately by conditional logit using both the uncorrected responder sample only, and the full sample (i.e. including both respondents and non-respondents), as well as using our selection correction algorithm with the respondent-only sample.

Figure 9 plots the distributions of marginal WTP estimates from these three methods. As before, the true median marginal WTP for $benefit$ in the population is 1. Note that the responders-only distribution again demonstrates

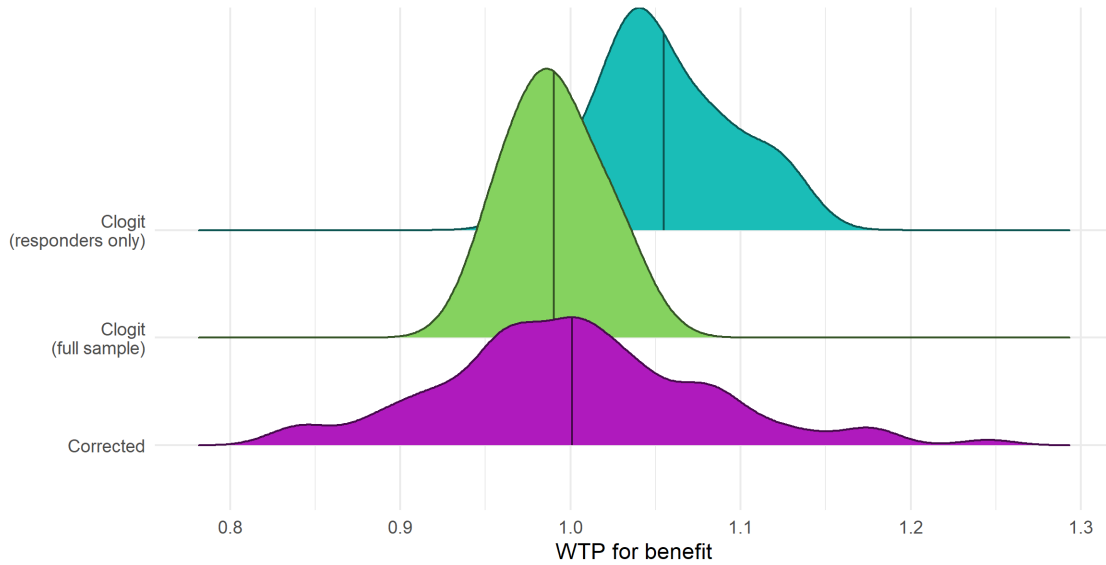


Figure 9. Distributions of WTP for benefit: 100 simulations of 2,000 invitees

significant upward bias in its central tendency. The corrected WTPs are more widely distributed than those for either the full samples or the selected samples estimated by the conditional logit. However, the mean marginal WTP for *benefit* across the 100 corrected samples (mean = 1.001) is not statistically distinguishable from the true marginal WTP of 1.

4.6 Conclusion

A viable method of correcting for sample selection bias in multiple discrete-choice specifications (traditionally estimated by conditional logit models) has long eluded choice modelers. We propose a correction method analogous to Heckman’s (1979) approach in the conventional least-squares context. Our method capitalizes on the close approximation of a normal distribution with $\pi^2/3$ variance to a standard logistic distribution and leverages software designed for mixed-logit models to achieve full-information maximum likelihood estimates for a conditional logit model allowing for potential systematic selection. Based on a large-scale

simulation and the distribution of results across 100 small-scale simulations, we illustrate the extent to which our proposed correction method recovers the central tendency for key marginal willingness-to-pay estimates that would be obtained using a true random sample from the population. The corrected estimates, however, may be noisier than those from a true random sample, suggesting that our method for selection-correction may involve a bias/efficiency trade-off.

CHAPTER V

CONCLUSION

This dissertation applies multiple empirical approaches to examine the role of culture in economic phenomena, and establishes that cultural traits play a role both in production (of foreign aid projects) and the formation of preferences (for pandemic lockdown policies). This dissertation also makes a methodological contribution to stated preference research, most commonly used to value goods and services that are not traded in markets, such as natural and cultural goods.

Chapter 2 identifies a complex relationship between (1) the cultural match between foreign aid project leaders and recipient countries and (2) the success of foreign aid projects. This relationship is predicted by a principal-agent model, and is identified as a *causal* relationship using an instrumental variables strategy. Additionally, this chapter offers a novel strategy for estimating the cultural match between individuals and countries using only the individual's name, leveraging information about the global distribution of surnames and existing measures of cultural proximity between countries. These findings imply that multilateral aid organizations should give special attention to the cultural background of project leaders they assign in countries with high institutional capacities. Future research is required to determine if a culturally knowledgeable and diverse workforce can improve the productivity of profit-driven international firms, in addition to aid organizations.

Chapter 3 uses a choice experiment to identify a preference reversal caused by the presence of generous federal unemployment benefits in the context of pandemic lockdown policies. In scenarios where we asked respondents to assume that there would be large federal unemployment payments, self-identified political moderates and conservatives became more likely to vote in favor of lockdown

policies that reduced average income in their counties. At the same time, generous unemployment payments caused these same groups to become more averse to policies that would increase in unemployment. These results suggest that policymakers should carefully consider the potential effects of federal-level economic stimulus and unemployment insurance policies on support for other policies where authority devolves to lower-level jurisdictions.

Chapter 4 presents the first viable method of correcting for sample selection bias in multiple discrete-choice specifications. We develop and implement a FIML estimation strategy for this correction method, and demonstrate its properties in simulations.

APPENDIX A

CHAPTER 2 APPENDIX

A.1 Data appendix

A.1.1 Genetic distance. Wacziarg and Spolaore calculate genetic distance between countries according to the following formula:

$$F_{st}^{12} = \sum_{i=1}^I \sum_{j=1}^J (s_{1i} \times s_{2j} \times d_{ij}),$$

where $i = 1, \dots, I$ indexes the genetic groups in country 1, $j = 1, \dots, J$ indexes those of country 2, s_{1i} and s_{2j} are the population shares of genetic groups i and j in countries 1 and 2 as given by Alesina et al. (2003), and d_{ij} is the genetic distance between i and j as given by Pemberton et al. (2013). F_{st}^{12} then is the expected genetic distance between two randomly selected citizens, one from each country.

A.1.2 IEG evaluation. Every World Bank aid project receives at least two ratings. First, the TTL evaluates their own project according to the same criteria the IEG uses. This preliminary evaluation, called an Implementation, Completion and Results Report (ICR), and the documents on which it is based, are then reviewed by an IEG evaluator, who issues a second report—an ICR Review (ICRR). The ICRR may revise the outcome score given in the initial ICR. Where insufficient supporting documentation is provided, IEG evaluators are instructed to penalize the project's assessment.

Some projects receive a third evaluation—a Project Performance Assessment Report (PAR). These evaluations follow the same rating criteria as ICRRs, but are based on a broad range of independently gathered evidence, including site visits and interviews with stakeholders. I use the PAR score if it is available, and otherwise use the ICRR score. ICR scores are determined by the TTL in charge

of the project, so I do not use ICR scores in my analysis. Because PARs are far more rigorous than ICRRs and represent the research and judgments of a group of evaluators, I assign every PAR the same fixed effect.

A.2 Structural estimation results (preliminary)

To test the validity of my principal-agent model, outlined in section 2.3, I use the model's production function as the estimating equation:

$$Q = q(X) + g(C)e + \eta$$

where $q(X)$ is a linear combination of control variables, and e is predicted utility maximizing level of effort for each TTL, given their cultural proximity to their project's recipient country and the institutional quality of the recipient country. Additionally, a number of other parameters that govern the TTL's choice of effort require estimation. I choose the following simple functional forms for $g(C)$ and $h(I)$:

$$g(C) = \gamma_1 C^{\gamma_2}$$

$$h(I) = \lambda_1 + \lambda_2 I$$

where γ_1 , γ_2 , λ_1 , and λ_2 are parameters to be estimated, with the restrictions that $\gamma_1 > 0$, $0 < \gamma_2 < 1$, $\lambda_1 > 0$, $\lambda_2 < 0$, and $\lambda_1 > \lambda_2 \bar{I}$, where \bar{I} is the largest observed value of I . The threshold effort level m is also a parameter to be estimated.

To limit the parameter-space to a manageable dimensionality, I impose the following function for the TTL's disutility of effort:

$$\phi(e) = \frac{1}{2}e^2$$

I re-scale C and I by subtracting each variable's (negative) minimum observed value and then adding 0.1, so that $C > 0$ and $I > 0$. This re-scaling

preserves the feature that a 1-unit increase in C or I is associated with a one-standard deviation increase.

Structural estimation proceeds iteratively as follows.

Step 1

Begin with guesses for values of the parameters in $h(I)$: λ_1 and λ_2 .

Step 2

Using these values of λ_1 and λ_2 , run a numeric optimization routine to find values of γ_1 , γ_2 , and m that minimize the residual sum of squares from OLS estimation of the production function.¹ Values of e for each project are determined within the optimization routine. For the OLS specification, $q(X)$ includes precisely the same set of variables described in section 4 and used in estimating the reduced form results, with the exception of the interaction between C and I . The quantity $g(C)e$ is included as a regressor with its coefficient restricted to 1.

Step 3

Save the residuals from OLS in the last step of OLS before convergence. Group these residuals into 50 bins by their associated value of I and calculate the variance within each bin. Then regress these 50 variances on each bin's mean value of I . The resulting OLS estimates provide new values for λ_1 and λ_2 , the parameters in $h(I)$, so that the variance in the residuals (as a function of I) provides an estimate of the variance in η (as a function of I). Return to step 2.

This process terminates when two successive iterations produce estimates for λ_1 , λ_2 , γ_1 , γ_2 , and m that are sufficiently close. The stopping criterion I've been

¹I recover similar parameter estimates using Nelder-Mead, Broyden-Fletcher-Goldfarb-Shanno, or Conjugate Gradient methods. I present results from Nelder-Mead, which converges most quickly.

using is:

$$\sqrt{(\lambda_1^t - \lambda_1^{t-1})^2 + (\lambda_2^t - \lambda_2^{t-1})^2 + (\gamma_1^t - \gamma_1^{t-1})^2 + (\gamma_2^t - \gamma_2^{t-1})^2 + (m^t - m^{t-1})^2} < 5 \times 10^{-5}$$

where t indexes iterations.

The parameter estimates resulting from this estimation process are given in Table B1. Figure A.1 plots the estimated effort levels for each project at the last iteration. Figure A.2 plots the variances of the residuals within each bin at the last iteration. Comparing Akaike information criteria for the structural model (5320.09) and the reduced form model estimated in Table 2, column 3 (5325.757) shows the structural model to be an improvement.

Parameter	Estimate
λ_1	1.278
λ_2	-0.128
γ_1	1.460
γ_2	0.417
m	0.934

Table B1. Structural parameter estimates

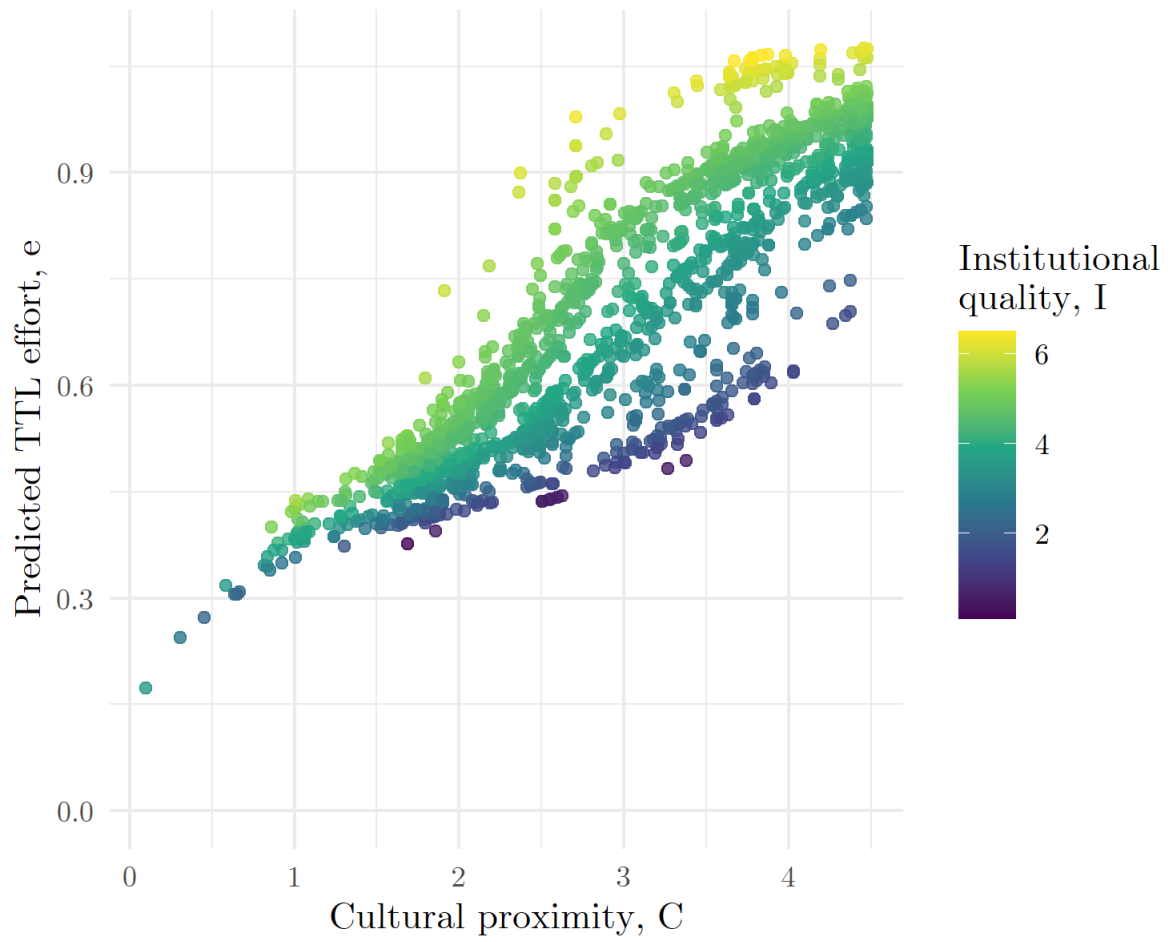


Figure A.1. Predicted effort

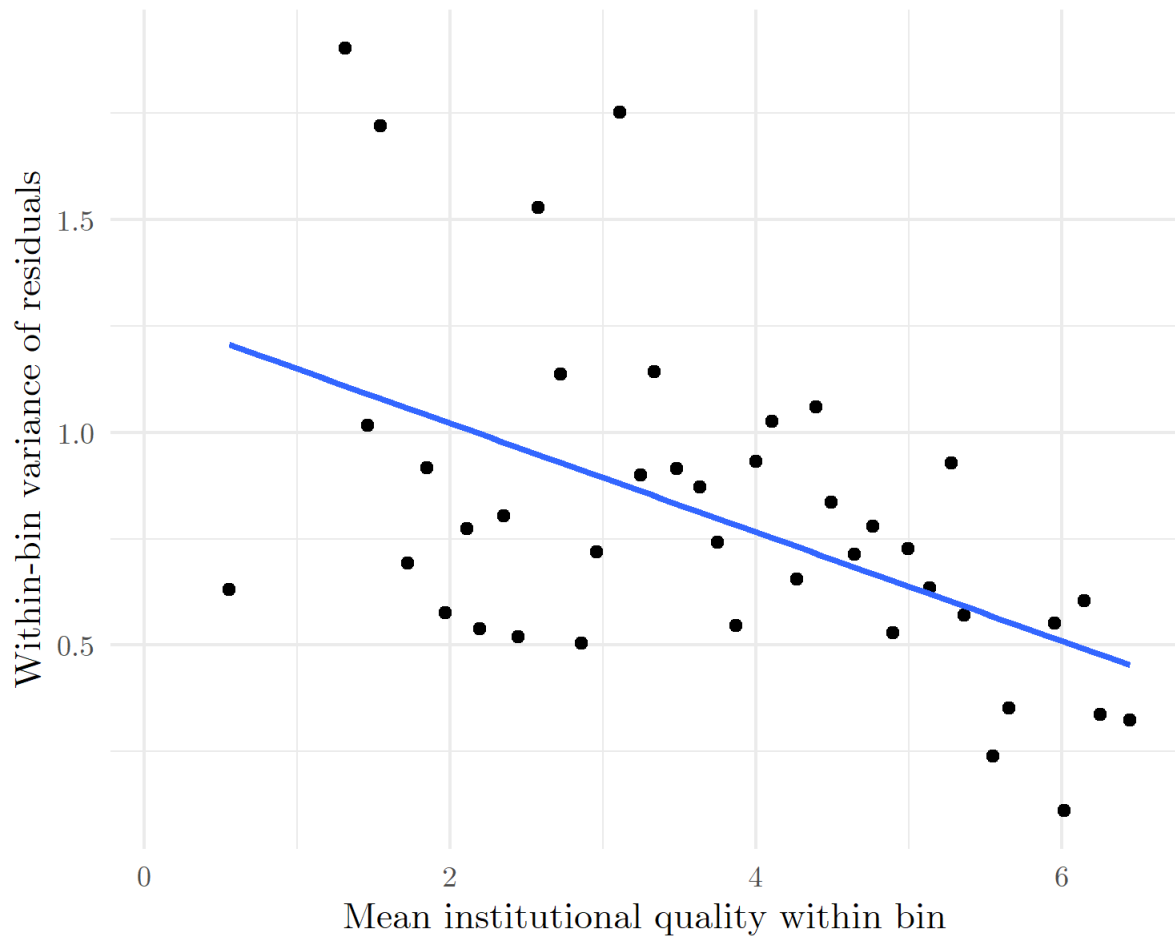


Figure A.2. Fitting $h(I)$

A.3 Instrument validation

Table C1 presents estimation results for tests of the IV's exclusion restriction as discussed in Section 2.5. The instrument, the average cultural proximity to each recipient country of TTLs who led any World Bank project that was evaluated before 2005, is intended to capture the availability of culturally close TTLs to each recipient country.

Table C1. Tests of the IV's exclusion restriction

	<i>Dependent variable is the instrument</i>		
	(1)	(2)	(3)
Number of projects (in hundreds)	-0.048 (0.406)		
Total project costs (in 10s of billions)		-0.047 (0.115)	
CPIA			0.118 (0.235)
log(population)	0.284*** (0.097)	0.297*** (0.084)	0.269*** (0.066)
log(GDP)	0.393** (0.160)	0.406** (0.160)	0.359** (0.164)
Observations	75	75	75
R ²	0.199	0.200	0.201

Note: *p<0.1; **p<0.05; ***p<0.01

Instrument is average distance to each recipient country of TTLs who led pre-sample projects.

A.4 Surnames for female TTLs

Section 2.6 evaluates the possibility that cultural proximity is systematically mismeasured for female TTLs by comparing my primary measure of cultural proximity with an alternative measure of cultural proximity that should not vary in accuracy between male and female TTLs. Another way to evaluate the possibility that cultural proximity is systematically mismeasured for female TTLs is to allow the key coefficients in my model to vary by gender. Table D1 reports results from two IV specifications with key coefficients interacted with an indicator variable equal to 1 if the TTL is female. Model 1 in Table D1 reports coefficients estimated by a model including the full set of interaction terms between *Cultural proximity*, *Institutional quality*, and *Female*. The coefficient on *Cultural proximity*, which is now identified only by variation among male TTLs, is qualitatively similar to the estimate from my preferred specification (model 2 in Table 4), but the key interaction term *Cultural proximity* \times *Institutional quality* is smaller and not statistically distinguishable from 0.

However, none of *Cultural proximity* \times *Female*, *Institutional quality* \times *Female*, or *Cultural proximity* \times *Institutional quality* \times *Female* has a statistically significant effect on project success. Furthermore, these three variables are not jointly statistically significant ($F = 1.134$, $p = 0.334$). In short, we have no evidence that the gender of a project's TTL is an important determinant for the success of the project, either on its own or in conjunction with *Cultural proximity* or *Institutional quality*.

Table D1. IV results with gender controls, selected coefficients

	IEG Outcome: 1-6 scale		
	(1)	(2)	(3)
Cultural proximity	0.215* (0.127)	0.216* (0.126)	0.233* (0.125)
Institutional quality	0.204** (0.097)	0.265*** (0.079)	0.267*** (0.075)
Female	0.085 (0.140)	0.119 (0.142)	-0.021 (0.053)
Cultural proximity × Institutional quality	0.074 (0.061)	0.114** (0.047)	0.111** (0.046)
Cultural proximity × Female	0.091 (0.097)	0.101 (0.099)	
Institutional quality × Female	0.179 (0.127)	0.017 (0.054)	
Cultural proximity × Institutional quality × Female	0.109 (0.079)		

Notes: For a summary of additional control variables included in all models, see notes to Table 2. Twenty-one observations for which TTL gender could not be determined have been excluded.

A.5 Alternative clustering and robustness checks

Tables E1, E2 and E3 are analogous to Table 4, but allow standard errors to cluster at alternative levels of aggregation. Table E1 presents estimates with standard errors clustered at the country level. Table E2 presents estimates with standard errors clustered at the level of country-project approval year. Finally, Table E3 presents estimates with unclustered standard errors. In all cases, the coefficient estimates for *Cultural proximity* \times *Institutional quality* are statistically significant from 0 at the 5% level. I do not cluster standard errors by only approval year or only project evaluation year. These levels of aggregation yield only 20 and 14 clusters, respectively. As a rule of thumb, fewer than 40 clusters is not sufficient to estimate clustered standard errors.

Table E4 presents OLS results with controls for country fixed effects. The inclusion of country fixed effects precludes use of the instrument, which varies only between countries. Institutional quality has an insignificant effect on project outcomes after including country fixed effects because there is little variation in institutional quality within a country over time. The direct effect of cultural proximity remains significant at the 5% level in specification (1), but the interaction of cultural proximity and institutional quality is insignificant.

Table E5 presents IV results for a series of linear probability models, where the cut off for “success” is assigned to every possible value. That is, in column (1), the outcome variable is 1 if the project received an IEG score greater than 1. In column (2), the outcome variable is 1 if the project received an IEG score greater than 2, and so on. Together, these specifications demonstrate that cultural proximity is especially important for producing projects that receive the highest IEG outcome score of 6. Additionally, cultural proximity and its interaction with institutional quality are especially important for achieving at least an IEG outcome

of 4. On the other hand, cultural proximity plays little role in saving projects from the worst scores of 1 and 2.

Figure A.3 plots the coefficient estimates from six linear probability models, where each dependent variable is coded as 1 if a project received score i , $i \in 1, 6$. A $1-\sigma$ increase in cultural proximity (above the mean), for instance, is associated with a significantly lower probability of receiving an IEG score of 3, but is associated with a significantly higher probability of receiving a score of 6. None of the variables—institutional quality, cultural proximity, or the interaction of the two—significantly changes a project’s probability of receiving a score of 1 or 2. In general terms, better institutions and more cultural proximity are associated with fewer marginal failures (scores of 3) but not with fewer abject failures (scores of 1 or 2).

Table E6 presents IV results using only a subset of the CPIA composite score as a measure of institutional quality. The World Bank constructs its CPIA score as a simple average of four cluster subscores for economic management, structural policies, social inclusion/equity, and public sector management. Using the simple average of these as a proxy for overall institutional quality may overstate or understate the relative importance of each component.

Table E6 reproduces the results from the main IV specification (model 2 in Table 4), using each CPIA cluster subscore in place of the single CPIA aggregate. The results indicate that social and economic institutions have the strongest interaction with cultural proximity.

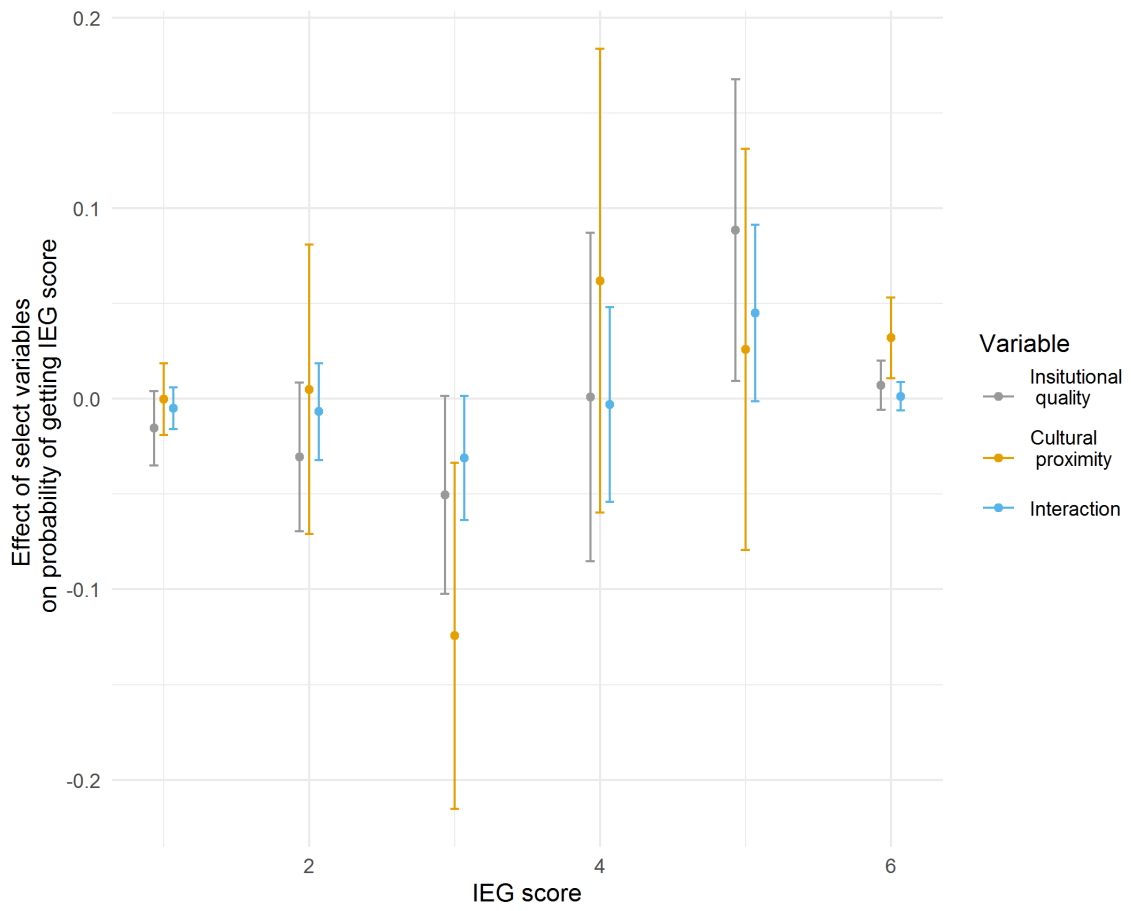


Figure A.3. Effect of selected variables on receiving specific IEG scores

Table E1. IV results, standard errors clustered by country

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185 (0.128)	0.205* (0.123)	0.185 (0.123)
Institutional quality	0.095*** (0.036)	0.261*** (0.080)	0.313*** (0.101)
Cultural proximity \times Institutional quality		0.107** (0.045)	0.130** (0.055)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes:

*** $< .01$, ** $< .05$, * $< .1$

For a summary of additional control variables included in all models, see notes to Table 2.

A.6 Dyadic regression results

Dyadic regressions are used to model network connectivity. Often, dyadic regressions model the formation of links between nodes, where any two nodes may share a link. Every pair of nodes is an observation, and the outcome variable is equal to 1 for pairs that share a link, 0 otherwise. Explanatory variables can be features of one or both nodes, or of the relation between them. Here I use dyadic regressions to model the probability that TTLs are assigned to projects. In this case, an observation is a TTL-project pair, and all possible TTL-project pairs are observations. Note that this differs from the typical case in that TTL-TTL and project-project pairs are not possible. The outcome variable is equal to 1 if that TTL was the primary TTL for the project. Regressing the outcome variable on the cultural distance between the TTL and the project's recipient country measures

Table E2. IV results, standard errors clustered by country-approval year

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185 (0.118)	0.205* (0.118)	0.185 (0.121)
Institutional quality	0.095*** (0.030)	0.261*** (0.075)	0.313*** (0.079)
Cultural proximity × Institutional quality		0.107** (0.044)	0.130*** (0.047)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

For a summary of additional control variables included in all models, see notes to Table 2.

Table E3. IV results, standard errors unclustered

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185* (0.111)	0.205* (0.113)	0.185* (0.112)
Institutional quality	0.095*** (0.026)	0.261*** (0.071)	0.313*** (0.073)
Cultural proximity × Institutional quality		0.107** (0.043)	0.130*** (0.044)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

For a summary of additional control variables included in all models, see notes to Table 2.

the role of cultural proximity in the assignment process of TTLs to projects. The results in Table F1 indicate that TTLs who are culturally close to a recipient country are more likely to be assigned to a project in that country. Additionally, this tendency is somewhat stronger for recipient countries with strong institutions, though this relationship is significant only at the 10% level. Taken together, these results suggest that the World Bank’s TTL assignment policies already conform with the recommendation that culturally close TTLs should be preferentially assigned to projects, but that World Bank projects may benefit further if this assignment policy is emphasized in countries with strong institutions.

Table E4. OLS results with country fixed effects

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.067** (0.034)	0.063* (0.034)	0.063 (0.038)
Institutional quality	-0.067 (0.125)	-0.009 (0.132)	0.068 (0.150)
Cultural proximity × Institutional quality		0.044 (0.030)	0.028 (0.034)
Country FEs	Yes	Yes	Yes
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

All models include controls listed in the notes to Table 2, with the exception of the geographic region of recipient countries, which are subsumed by country fixed effects.

Table E5. IV linear probability models

	Project success = 1 if..				
	IEG score > 1 (1)	IEG score > 2 (2)	IEG score > 3 (3)	IEG score > 4 (4)	IEG score > 5 (5)
Cultural proximity	0.000 (0.010)	-0.005 (0.039)	0.120** (0.057)	0.058 (0.055)	0.032*** (0.011)
Institutional quality	0.015 (0.010)	0.046** (0.019)	0.096*** (0.033)	0.096** (0.042)	0.007 (0.007)
Cultural proximity × Institutional quality	0.005 (0.006)	0.012 (0.013)	0.043** (0.021)	0.046* (0.024)	0.001 (0.004)

Notes:

*** < .01, ** < .05, * < .1

For a summary of additional control variables included in all models, see notes to Table 2.

Table E6. IV Results for CPIA clusters

CPIA cluster:	IEG Outcome: 1-6 scale			
	Social inclusion (1)	Economic management (2)	Public sector (3)	Structural policies (4)
Cultural proximity	0.186 (0.135)	0.193 (0.124)	0.194 (0.121)	0.140 (0.123)
CPIA cluster	0.256*** (0.087)	0.236*** (0.070)	0.234*** (0.064)	0.157* (0.085)
Cultural proximity × CPIA cluster	0.114** (0.055)	0.108** (0.046)	0.082* (0.047)	0.067 (0.059)
N	1841	1946	1946	1946

Notes:

*** < .01, ** < .05, * < .1

CPIA cluster scores are standardized $\mu = 0, \sigma = 1$. Social inclusion cluster scores are missing for projects that took place in Indonesia (105 observations). For a summary of additional control variables included in all models, see notes to Table 2.

Table F1. Dyadic regressions

	<i>Dependent variable:</i>		
	= 1 if TTL _{<i>i</i>} is assigned to project _{<i>j</i>}		
	(1)	(2)	(3)
Cultural proximity	0.550*** (0.022)	0.555*** (0.022)	0.552*** (0.022)
Institutional quality		-0.052*** (0.009)	0.021 (0.041)
Cultural proximity × Institutional quality			0.033* (0.018)
Observations	2,247,023	2,247,023	2,247,023

Note: *p<0.1; **p<0.05; ***p<0.01

All models include TTL fixed effects. Standard errors are clustered at the project level. Coefficients and standard errors are multiplied by 1000 for readability.

APPENDIX B

CHAPTER 3 APPENDIX

B.1 Online Appendix: Other Pandemic Policy Choice-Experiment Surveys

Chorus et al. (2020) collected their entire sample of 1009 completed surveys on April 22, 2020 using an online survey panel (Kantar Public) intended to be representative of the Dutch population in terms of gender, age, and education level. Rather than using specific types of pandemic restrictions as their policy attributes, they elect to use the generic consequences (impacts) of differing policy measures, making theirs a so-called “unlabelled” or “unbranded” choice experiment. Specifically, they described the increase in the number of deaths (8k, 11.5k, 15k and 18.5k), the increase in people with lasting injuries (30k, 80k, 130k and 180k), the increase in people with lasting mental injuries (30k, 80k, 140k and 200k), the increase in the number of children with educational disadvantages (10k, 90k, 170k, and 250k), the increase in households with a net income loss of more than 15% for more than three years (400k, 700k, 1M, 1.3M), a one-time tax per household in 2023 (1k, 2.5k, 4k and 5.5k Euros), and four levels of work pressure in the health-care sector. Their latent-class models differentiate preferences by gender, education level, and age. Their estimated preference classes suggest deontological preferences for some people, and consequentialist preferences for others. In separate specifications, they also explore the effect of having had a relative infected with COVID-19, and the effect of choices that involve “the taboo tradeoff” that involves simultaneously both a lower tax and a higher number of fatalities.

Blayac et al. (2021) report on survey-based choice experiments fielded in France during May 4 to May 16, 2020, yielding 1154 respondents with three pair-wise policy choices each. Both options, in each case, were active options (i.e. there

was no status quo, or no-intervention, policy). Their choice scenarios featured seven policy attributes: (1) weeks of additional lockdown (no extension, one week or three weeks, treated as continuous), (2) mask requirements (indicators for mandatory in public places or mandatory everywhere), (3) bar and restaurant closure (indicator for closed all the summer season), (4) daily public transportation, urban and regional (indicator for limited to working hours), (5) vacation and leisure travels (indicators for limited to France or limited to within 100km), (6) digital tracking (an indicator for a free app), (7) financial compensation (in Euros, 0, 500, 1500 or 2200, treated as continuous). They differentiate their estimated preferences according to clinical vulnerability, age, and gender.

Reed et al. (2020) surveyed 5953 adults across all 50 U.S. states between May 8 and May 20 of 2020, during the first wave of the pandemic. They collaborated with a health care market-research firm, SurveyHealthcareGlobus, which sent email invitations to adults throughout the U.S., with oversampling in New York, California, Texas and North Carolina. The choice scenarios that they presented to respondents distinguished between just four attributes: The month when all non-essential businesses would be re-opened, the total number of COVID-19 cases per 100 people by January 2021, the percent of families below the poverty line, and when the economy would recover. Outside of the choice tasks, their survey did include a ranking exercise, where respondents were asked to rank their subjective importance of relaxing six categories of restrictions. In order of importance their respondents prioritized the relaxation of restrictions on non-essential businesses, schools and colleges, restaurants, parks and museums, religious ceremonies, and then sporting events.

These authors focus on latent class analysis for their main results. About 40% of the people in their sample are classed as “risk-minimizers” who are not

willing to accept any level of risk brought about by the re-opening of non-essential businesses or reducing the impact of the pandemic on the economy. There is also a class of “waiters” who are less concerned about the health risks, but also do not want to re-open the economy in any hurry. About 25% are “recovery-supporters,” and about the remainder are labelled as “openers” who are unconcerned about health risks or any impacts upon the poor. The individual characteristics that have systematic effects on membership in these latent preference classes include Democrat/Republican/Independent political affiliations, Employment status, income in each of three brackets, and whether the respondent is non-white.

During May and June of 2020, Wilson et al. (2020) recruited respondents through “randomly allocated social media advertising on Facebook and Instagram.” They also sought greater diversity in their sample by recruiting by email through a community health organization. They argue that Missouri broadly represents the U.S. Midwest and South in terms of socioeconomic characteristics. Like Reed et al. (2020), Wilson et al. (2020) identify four latent preference classes in this population, dubbing them “risk eliminators,” “risk balancers,” “altruistic,” and “risk takers.” They detect that younger respondents (18-24 years) have a stronger relative preference for keeping schools and social venues open and to prevent income loss, compared to older age groups. Men, in contrast to women, prefer to keep outdoor recreational facilities and social/lifestyle venues open, to allow large gatherings, to minimizing lost incomes, and to have shorter policy durations. Whites, similar to men, but in contrast to other racial groups, want to keep services open, to minimize lost incomes, and to have shorter policy durations.

Genie et al. (2020) published their protocol for a choice experiment study of pandemic policy preferences, not including results, in November of 2020. Their choice experiments feature seven attributes: (1) restrictiveness of the

lockdown (summarized as one of four levels, each associated with a fixed pattern of restrictions), (2) lockdown duration (3, 6, 10 or 16 weeks), (3) postponement of non-urgent medical care (for no, some, or all non-urgent care), (4) excess deaths (1, 4, 9, or 13 additional people per 10,000), (5) number of infections (100, 600, 1300 or 2000 people infected per 10,000), (6) ability to buy things (illustrated as 100%, 90%, 80% or 70% of current purchases) and (7) the proportion of people who will lose their job (0, 4, 15 or 25 out of 100).

One limitation of the Genie et al. (2020) choice set design is that their four overall levels of activity restrictions are aggregated into just “green,” “yellow,” “amber,” and “red.” The key to interpretation of their four levels of “lockdown” references the extent to which level requires (or limits) staying at home, group sizes for social gatherings, non-essential activities other than groceries and work-related trips, schools and youth activities, businesses and shops, and outdoor activities, but there is no variation within each of these four overall levels in the relative severity of the different constituent categories of restrictions.

Genie et al. (2020) plan to explore observed heterogeneity according to the usual array of sociodemographic variables, but include a novel measure of moral attitudes. They plan to use mixed logit models to test for unobserved preference heterogeneity.

There have also been a few stated-preference surveys concerning people’s preferences concerning COVID-19 vaccines. Cerda and Garcia (2021) using a sample of 531 individuals from middle- and high-income sociodemographic groups in Chile. They use a double-bounded dichotomous choice elicitation format. Their survey was in the field between July 10 and August 10 of 2020. Muqattash, Niankara, and Traoret (2020) describe data from a Google Forms survey about vaccine preferences, fielded between July 4 and August 4 of 2020, administered to

respondents in the UAE and garnering 1109 responses, but no models or statistical results are reported in the paper. Kreps et al. (2020) report a choice-experiment survey fielded on July 9, 2020, using the Lucid survey platform, which contacted 3708 U.S. adults and yielded 2000 responses. This study also concerns tradeoffs between COVID-19 vaccine attributes.

We should not omit to mention a pre-COVID-19 study by Cook et al. (2018) using a survey conducted with 500 respondents in Singapore between November 2012 and February 2013 in the wake of the previous 2003 SARS-CoV and 2009 H1N1 influenza outbreaks. Their policy options concerned seven attributes, including quarantines (mandatory or voluntary), isolation of actual cases (voluntary or mandatory), cancellation of mass gatherings (none, just schools, or also other gatherings of 30 or more), temperature screenings (none, at border checkpoints, at border checkpoints and internally), a fee to fund public health measures (S\$15, S\$20, S\$40 or S\$50), the expected number of infections (200, 500, 1k, 10k or 1M), and the expected number of infection-related deaths (0, 30, 80, 120, or 180).

B.2 Online Appendix: Survey development

Survey development took place over the summer and early fall of 2020. Several generations of choice scenarios were assessed using sequential one-on-one think-aloud protocols with test subjects who helped us to understand the most efficient and accessible ways to present the tutorial information in the survey.

Each paragraph of the survey was processed through MS Word's reading-level utility, where we pursued revisions until passages fell into the 60-70 range on the Flesch Reading Ease test, and achieved a Flesch-Kincaid Grade Level score between 7.0 and 8.0.

Using javascript in the background of the survey, modal popups are used extensively to make access to optional explanations , or the review of key concepts, as easy as possible for respondents.

Information about the population of the respondent's own county, as well as actual COVID-19 cases and deaths in the respondent's county, was supplied dynamically. Values were keyed to the respondent's selection of their county of residence and updated with the most-recent four weeks of COVID-19 data every few days over the course of the survey.

Information about unemployment insurance that underlies our scenarios about unemployment rates and average household income lost in the respondent's county was keyed to their state's formula for state-level unemployment insurance and the baseline level of unemployment in their county, according to the Bureau of Labor Statistics, just prior to the start of the pandemic.

The survey's scenarios about potential future baseline pandemic conditions were based on draws from the distributions of county-level cases and deaths per 50,000 people during a recent four-week period across all counties in the U.S. These randomized draws of actual rates of cases and deaths were converted into

the numbers of cases and deaths that would be implied for the respondent's own county.

Respondents were reminded of actual cases and deaths in their own county during a recent four-week period. They were also informed about what would have been the cases and deaths over the last four weeks in their own county if their county's rates had corresponded to the 90th percentiles, across all counties, during that recent four-week period. These actual data about potential COVID-19 risks were important for conveying the scope of the pandemic elsewhere in the country. Before we provided this contextual information, several test subjects had no real idea just how bad the pandemic could be if their county had been among the worst 10% across all U.S. counties (i.e. approximately the worst 300 counties) in terms of COVID-19 cases and deaths. These benchmarks were important for preventing rejection of choice scenarios as being implausible.

All of our randomized choice scenarios were generated outside the survey in rates per 50,000 people (for baseline cases and deaths as well as the reductions in these baselines that would be achieved by each proposed pandemic policy). All of the fields for the standardized choice scenarios were converted to json format and uploaded to Qualtrics via the survey platform's Web Service utility, making use of a block of php code designed to draw instances of these standardized choice scenarios, without replacement, from the universe of randomized scenarios.

One instance of the information for the choice scenarios was selected when a respondent accessed the survey, and javascript was used to convert the baseline rates for cases and deaths, and for reductions in cases and deaths in each policy choice scenario from rates per 50,000 to absolute counts of cases and deaths for their own county's population, as given by the 2018 5-year American Community Survey. The quoted costs of each policy were systematically related to the stringency of the

restrictions that the policy would place on each of ten activities or businesses, for plausibility, and these costs were associated with rates of unemployment that would be consistent with these costs for a single-earner household with the median annual income in that county (after state-level unemployment insurance). The quoted weekly amounts of federal UI were incorporated dynamically into the specific choice scenario, reducing these costs but having no effect on unemployment levels.

Each level of restrictions on each of our ten types of activities or businesses was associated with a verbal description appropriate to that activity or business, and these descriptions were available via popup associated with the terse name for that type of restriction. Each activity/business could have restrictions at level 0, 1, 2 or 3, except for “Grocery, essential retail,” which was never subjected to level 3 restrictions.

The survey was designed to start with two binary choices between a specified policy and a No Policy alternative, described as being a policy where everyone is simply allowed to take whatever pandemic precautions that they deem to be appropriate. Two further choice scenarios were employed, where each involved first a three-way choice between either of two policies and a status quo alternative. If the status quo was not chosen, the respondent was then given a choice between the other, non-chosen policy versus the status quo. These three-way choices, with a two-way conditional follow-up, are more complex and require more complex statistical analysis, so we reserve their analysis for subsequent research and focus on the two binary (and most incentive-compatible) choice sets for this analysis.

The survey contains a substantial number of follow-up questions about respondent’s pandemic perceptions and attitudes, for which detailed analysis will also be addressed in later research.

B.3 Online Appendix: Selection model

B.3.1 Variables available for selection model. We designed our survey to permit extensive modeling of the decisions by eligible respondents about whether or not to complete the survey after they learned its topic (or at any subsequent point). In a separate paper, Mitchell-Nelson and Cameron (2021), we explain the survey design considerations and the variety of sources from which additional explanatory variables can be recruited. We used LASSO methods to identify the most important regressors to use in a model to yield fitted response propensities. In our models reported in the main paper, we employ a full set of interaction terms between every explanatory variable and these response propensities, which have first been de-measured relative to the average response propensity across all eligible respondents, regardless of whether they completed the survey and ended up in the estimating sample.

Briefly, we built our selection model using the following process. First, we assembled a set of all candidate variables and their interactions, dropping variables that are perfectly collinear with other variables, e.g. an indicator and that same indicator squared. We used 10-fold cross-validation to select the value of lambda for LASSO that minimizes prediction error for a binary logit selection model, and retained the list of variables selected by LASSO at this value of lambda. Finally, we estimate a conventional logit selection model using these variables, supplemented by all of the base variables employed in any retained interaction terms, even if these base variables were not themselves retained by our LASSO model.

Table C1. Descriptive statistics: Basic variables for selection model, retained by LASSO model, either as individual variables or as part of a pairwise interaction term.

	mean	sd	min	max
Days since Jan 13, 2021 launch	20.307	6.658	0	33
CDC SVI cnty-Minority, language	0.855	0.166	0.256	0.998
CDC SVI cnty-Hsg type, transp.	0.759	0.183	0.063	0.999
1=Own gender female	0.511	0.5	0	1
1=Own race black	0.046	0.21	0	1
1=Own race white	0.703	0.457	0	1
1=Own age is 25 to 34	0.193	0.395	0	1
1=Own age is 35 to 44	0.229	0.42	0	1
1=Own age is 45 to 54	0.115	0.319	0	1
1=Own age is 55 and up	0.056	0.231	0	1
1=Own hhld inc less than 20K	0.115	0.319	0	1
1=Own hhld inc 20K to 25K	0.055	0.229	0	1
1=Own hhld inc 30K to 50K	0.117	0.321	0	1
1=Own hhld inc 100K to 125K	0.109	0.311	0	1
1=Own hhld inc 125K to 150K	0.09	0.286	0	1
1=Started survey on Monday	0.298	0.458	0	1
1=Started survey on Tuesday	0.206	0.405	0	1
1=Started survey on Wednesday	0.142	0.349	0	1
1=Started survey on Friday	0.109	0.311	0	1
1=Started survey on Saturday	0.103	0.304	0	1
1=Start: hour ending at 7:00	0.017	0.13	0	1
1=Start: hour ending at 9:00	0.047	0.212	0	1
1=Start: hour ending at 14:00	0.069	0.254	0	1
1=Start: hour ending at 15:00	0.082	0.274	0	1
1=Start: hour ending at 17:00	0.077	0.266	0	1
1=Start: hour ending at 19:00	0.045	0.208	0	1
1=Start: hour ending at 21:00	0.039	0.194	0	1
1=Start: hour ending at 23:00	0.031	0.174	0	1
Zip pr: vote republican	0.353	0.143	0.093	0.769
Days since first case, in 100s	3.438	0.24	2.43	3.89
1=At least one county death	0.995	0.071	0	1
Days since first death, in 100s	3.045	0.543	0	3.68
Cases/50K last 4 weeks	786.802	505.472	60.705	2835.951
Zip pr: age 18-24	0.084	0.04	0	0.359
Zip pr: age 55-64	0.117	0.039	0	0.385
Zip pr: race black	0.047	0.06	0	0.662
Zip pr: race asian	0.108	0.124	0	0.746
Zip pr: indus = agric.	0.019	0.042	0	0.54
Zip pr: indus = wholes.	0.027	0.015	0	0.122
Zip pr: indus = retail.	0.108	0.036	0	0.396
Zip pr: indus = transp.	0.049	0.029	0	0.28
Zip pr: indus = edu. serv.	0.205	0.066	0	0.66
Zip pr: indus = arts, ent.	0.098	0.042	0	0.316
County unemp rate Mar'20	5.411	2.055	2.8	23.5

Continued on next page

Table C1 – continued from previous page

County unemp rate Nov'20	6.904	2.141	3.7	16.2
County cases/50k Jun 2020	111.564	115.721	0	1278.466
County cases/50k Jul 2020	239.921	180.576	0	943.458
County cases/50k Aug 2020	188.713	132.845	0	1042.168
County cases/50k Oct 2020	141.259	63.762	0	537.281
County cases/50k Jan 2021	851.739	566.377	60.705	1808.252
County deaths/50k Jul 2020	2.872	3.009	0	32.461
County deaths/50k Aug 2020	3.489	3.038	0	23.028
County deaths/50k Sep 2020	2.533	2.171	0	18.074
Zip pr. urban (if havzip==1)	0.872	0.264	0	1
Zip pr. rural (if havzip==1)	0.089	0.198	0	1
County pr. other intnt	0.007	0.007	0	0.045
County pr. intnt not subscr	0.02	0.01	0	0.054
County pr. dialup cmpr	0.003	0.002	0	0.034
County pr. broadband cmpr	0.86	0.041	0.641	0.914
County pr. no intnt cmpr	0.066	0.019	0.038	0.192
County 2018 population	2652.225	3578.59	1.146	10098.052
1=Used a mobile device	0.534	0.499	0	1
1=Have ZIP code data	0.957	0.204	0	1

B.3.2 Coefficient estimates for selection model. Table C2 reports parameter estimates for our sample selection model. All available variables or sets of indicators, along with their pairwise interactions, were employed in a LASSO model for binary logit specifications. This algorithm employed ten-fold cross-validation to winnow down this extensive set of regressors to the set that makes the most accurate cross-validation predictions of which respondents who passed the screening phase will go on to complete the survey. This model yields the de-meaned fitted response propensities that serve as the variable $\hat{R}P_i$ in equation (3.3) in the paper. The coefficients on these interaction terms are not reported in the body of the paper because the relevant counterfactual simulation involves setting all of these de-meaned response propensities to zero, which is equivalent to dropping these terms from the model.

Table C2. Estimated coefficients for selection model, a binary logit specification employing all variables retained by the preliminary LASSO model (incompletely sorted)

Variable or interaction term	Coef. estimate	Std. error
Days since Jan 13, 2021 launch	-0.023	(0.060)
Days since Jan 13, 2021 launch \times Zip pr: age 18-24	0.200	(0.335)
Days since Jan 13, 2021 launch \times Zip pr: indus = edu. serv.	-0.004	(0.250)
Days since Jan 13, 2021 launch \times Zip pr: indus = arts, ent.	0.260	(0.327)
CDC SVI cnty-Minority, language	4.263	(4.557)
CDC SVI cnty-Minority, language \times Zip pr: age 55-64	-35.881	(32.166)

Continued on next page

Table C2 – continued from previous page

CDC SVI cnty-Hsg type, transp.	6.012	(4.449)
CDC SVI cnty-Hsg type, transp. × 1=Have ZIP code data	-6.224	(4.437)
1=Own gender female	1.213*	(0.621)
1=Own gender female × 1=Start: hour ending at 23:00	-1.126	(0.940)
1=Own gender female × 1=Own hhld inc 30K to 50K	-0.336	(0.541)
1=Own gender female × County unemp rate Nov'20	-0.086	(0.097)
1=Own gender female × County cases/50k Aug 2020	-0.004*	(0.002)
Zip pr: race black	3.374*	(1.863)
Zip pr: race black × 1=Start: hour ending at 17:00	-12.186**	(5.011)
1=Own race black	0.257	(0.508)
1=Own race black × 1=Own age is 55 and up	-18.577	(1,130.179)
1=Own race black × Zip pr: indus = agric.	-33.493	(48.676)
County pr. rural × 1=Own race black	-6.035*	(3.664)
1=Own race white	-0.218	(0.231)
Zip pr: race asian	-0.113	(0.739)
1=Started survey on Friday × Zip pr: race asian	13.397**	(5.317)
Zip pr: race asian × 1=Own hhld inc 20K to 25K	-7.906**	(3.292)
1=Own age is 25 to 34	0.007	(0.265)
1=Own age is 25 to 34 × 1=Start: hour ending at 9:00	5.385	(6.881)
1=Own age is 25 to 34 × 1=Start: hour ending at 17:00	-1.244*	(0.723)
1=Own hhld inc 30K to 50K × 1=Own age is 25 to 34	-0.355	(0.576)
1=Own age is 25 to 34 × 1=Own hhld inc 100K to 125K	2.198*	(1.163)
1=Own age is 35 to 44	0.216	(0.247)
1=Own age is 35 to 44 × 1=Start: hour ending at 19:00	-3.231*	(1.770)
1=Own age is 35 to 44 × 1=Start: hour ending at 21:00	-2.837***	(0.972)
1=Own age is 45 to 54	0.205	(0.290)
1=Own age is 55 and up	0.153	(0.385)
1=Own hhld inc less than 20K	0.404	(0.562)
1=Own hhld inc less than 20K × 1=Used a mobile device	-0.493	(0.540)
1=Own hhld inc less than 20K × County deaths/50k Jul 2020	-0.182**	(0.093)
1=Own hhld inc 20K to 25K	1.963*	(1.029)
1=Own hhld inc 20K to 25K × 1=Start: hour ending at 15:00	-1.521	(1.571)
1=Own hhld inc 20K to 25K × 1=Own race white	-1.576*	(0.864)
1=Own hhld inc 20K to 25K × County pr. other intnt	-52.940	(57.599)
1=Own hhld inc 30K to 50K	0.921	(0.674)
1=Started survey on Saturday × 1=Own hhld inc 30K to 50K	-1.369**	(0.651)
1=Own hhld inc 30K to 50K × County cases/50k Jul 2020	-0.005**	(0.002)
1=Own hhld inc 30K to 50K × County cases/50k Jan 2021	0.0003	(0.001)
1=Own hhld inc 100K to 125K	-0.784**	(0.374)
1=Used a mobile device × 1=Own hhld inc 100K to 125K	1.287**	(0.643)
1=Own hhld inc 125K to 150K	-5.479**	(2.343)
1=Start: hour ending at 7:00 × 1=Own hhld inc 125K to 150K	-8.139***	(3.003)
1=Own hhld inc 125K to 150K × County unemp rate Mar'20	0.298	(0.381)
1=Own hhld inc 125K to 150K × County pr. rural	18.081*	(10.176)
1=Own hhld inc 125K to 150K × County pr. dialup cmptr	222.663	(558.748)
1=Own hhld inc 125K to 150K × County pr. no intnt cmptr	61.696	(42.617)
1=Started survey on Monday	0.287	(0.352)
1=Started survey on Monday × County 2018 population	-0.0001*	(0.00005)
1=Started survey on Tuesday	-0.653	(0.683)

Continued on next page

Table C2 – continued from previous page

1=Started survey on Tuesday × 1=Start: hour ending at 7:00	-18.173	(1,702.151)
1=Started survey on Tuesday × Zip pr: age 18-24	4.800	(6.288)
1=Started survey on Tuesday × County deaths/50k Sep 2020	0.205	(0.128)
1=Started survey on Wednesday	-0.264	(0.394)
1=Started survey on Wednesday × 1=Start: hour ending at 14:00	-17.882	(1,715.373)
1=Started survey on Wednesday × 1=Start: hour ending at 17:00	-1.002	(0.638)
1=Started survey on Friday	-0.986*	(0.507)
1=Started survey on Saturday	0.507	(0.418)
1=Start: hour ending at 21:00	-0.262	(0.457)
1=Start: hour ending at 23:00	0.156	(0.815)
1=Start: hour ending at 23:00 × County pr. rural	-2.050	(2.416)
1=Start: hour ending at 7:00	0.268	(0.771)
1=Start: hour ending at 9:00	-0.541	(0.438)
1=Start: hour ending at 14:00	-0.334	(0.292)
1=Start: hour ending at 15:00	-0.238	(0.307)
1=Start: hour ending at 15:00 × 1=Own age is 45 to 54	-1.754***	(0.678)
1=Start: hour ending at 17:00	1.394**	(0.639)
1=Start: hour ending at 17:00 × 1=Own race white	-1.546***	(0.598)
1=Start: hour ending at 19:00	-0.412	(0.391)
Zip pr: vote republican	-3.379**	(1.722)
1=Used a mobile device × Zip pr: vote republican	0.022	(1.425)
Days since first case, in 100s	1.771	(3.932)
Days since first case, in 100s × 1=Have ZIP code data	-0.243	(3.734)
1=At least one county death	7.406	(8.193)
Zip pr: age 55-64 × 1=At least one county death	-16.974	(45.388)
Days since first death, in 100s	-4.556**	(2.058)
1=Have ZIP code data × Days since first death, in 100s	4.711**	(2.086)
Cases/50K last 4 weeks	0.003	(0.002)
Zip pr: age 55-64 × Cases/50K last 4 weeks	0.007	(0.008)
Cases/50K last 4 weeks × Zip pr: indus = transp.	-0.002	(0.010)
Zip pr: age 18-24	-1.292	(6.471)
Zip pr: age 55-64	50.383	(44.900)
Zip pr. urban (if havzip==1) × Zip pr: age 55-64	6.561	(16.008)
Zip pr: indus = agric.	-0.705	(2.785)
Zip pr: indus = wholes.	-16.712	(20.162)
Zip pr: indus = wholes. × County cases/50k Jun 2020	0.074	(0.080)
Zip pr: indus = wholes. × County pr. intnt not subscr	1,010.169	(872.524)
Zip pr: indus = retail.	-14.775*	(8.659)
Zip pr: indus = retail. × County cases/50k Oct 2020	0.129**	(0.059)
Zip pr: indus = transp.	-17.816**	(7.992)
Zip pr: indus = transp. × County deaths/50k Aug 2020	3.813**	(1.934)
Zip pr: indus = edu. serv.	1.974	(5.301)
Zip pr: indus = arts, ent.	-1.380	(6.897)
County unemp rate Nov'20	-0.027	(0.183)
County unemp rate Mar'20	0.053	(0.085)
County cases/50k Jun 2020	0.002	(0.003)
County cases/50k Jul 2020	0.0004	(0.002)
County cases/50k Aug 2020	0.0002	(0.003)
County cases/50k Oct 2020	-0.014**	(0.007)

Continued on next page

Table C2 – continued from previous page

County cases/50k Jan 2021	-0.002**	(0.001)
County deaths/50k Jul 2020	0.031	(0.107)
County deaths/50k Aug 2020	-0.179	(0.110)
County deaths/50k Sep 2020	0.014	(0.087)
Zip pr. urban (if havzip==1)	-24.499	(717.006)
1=Used a mobile device × Zip pr. urban (if havzip==1)	15.607	(716.903)
Zip pr. urban (if havzip==1) × County pr. broadband cmpr	-1.740	(13.642)
County pr. other intnt	20.184	(19.224)
County pr. intnt not subscr	-11.215	(24.551)
County pr. dialup cmpr	-100.434	(106.977)
County pr. broadband cmpr	-6.598	(16.954)
County pr. no intnt cmpr	2.704	(20.853)
County 2018 population	-0.0002	(0.0001)
County pr. rural	2.437	(1.723)
1=Used a mobile device	-16.092	(716.903)
Zip pr. rural (if havzip==1)	-22.074	(716.910)
1=Have ZIP code data	-2.147	(9.896)
1=Used a mobile device × Zip pr. rural (if havzip==1)	13.761	(716.905)
Constant	19.985	(717.258)
Observations	1,412	
Log Likelihood	-504.736	
Akaike Inf. Crit.	1,255.472	

B.4 Online Appendix: One example of a choice set, with pop-ups

B.4.1 One example of a policy-choice summary table. This is a two-month policy, with total cases and deaths over that two-month period, without and with Policy A. We convert these to per-month cases and deaths for analysis, commensurate with county unemployment rate and average cost per household stemming from this unemployment. This instance has very high county-level unemployment, but average \$/month lost for county households is limited by unemployment insurance and federal unemployment insurance supplements, in the amount of \$200/week, in this instance.

The sets of images below show the first choice task in the survey, for this one instance of the randomizations, as displayed on the screen of a mobile device. These tasks are preceded by an extensive tutorial about how to interpret these compact policy scenarios.

Figure D1. Immediate preamble to first summary table (one example)

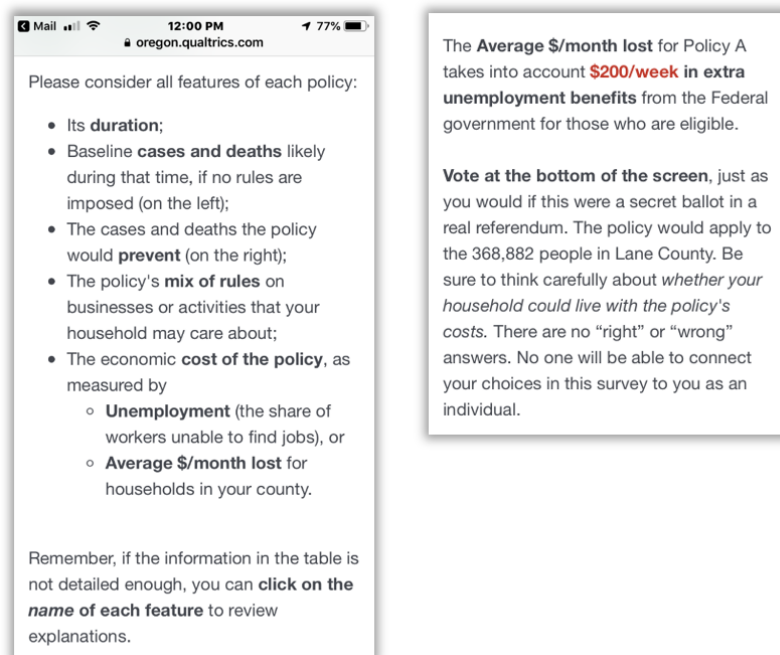
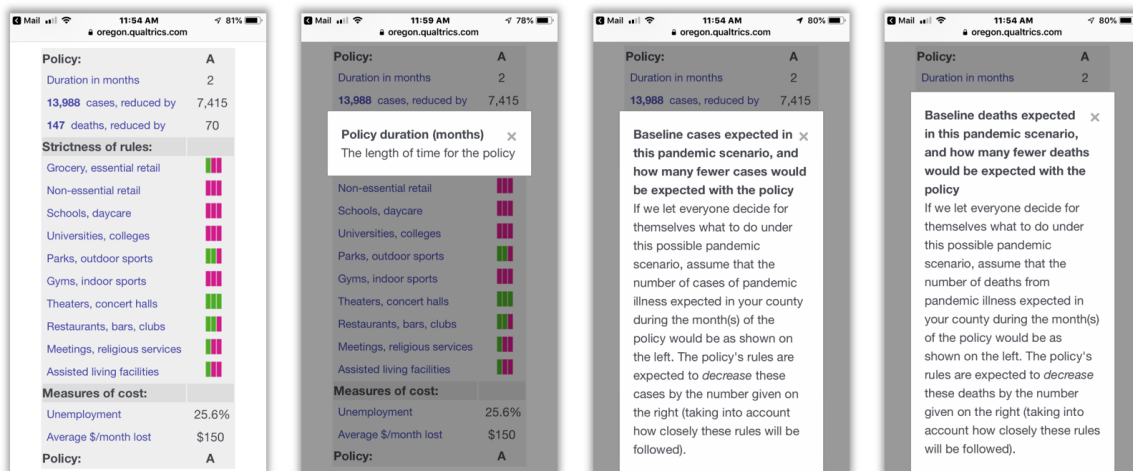


Figure D2. One instance of Policy A; contents of first three pop-ups



145

Figure D3. Contents of fourth through seventh popups

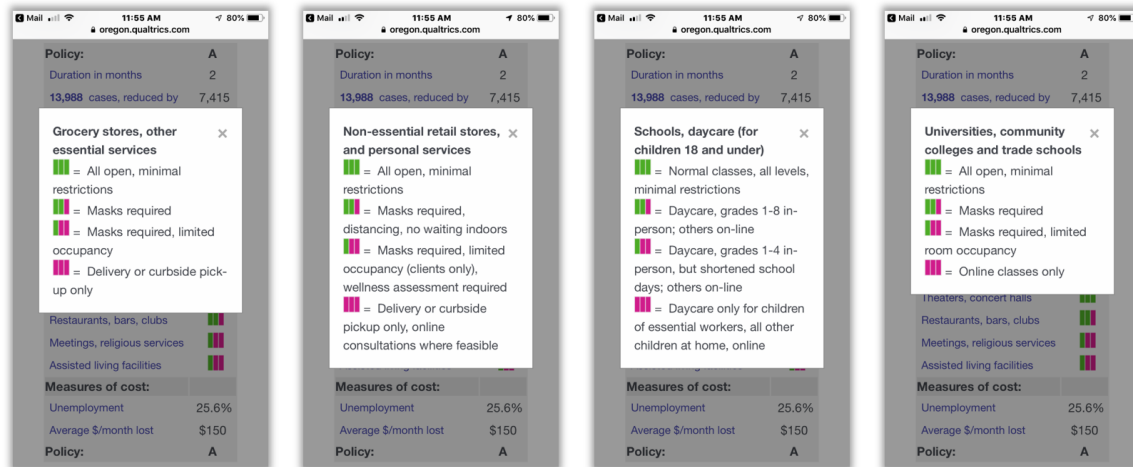
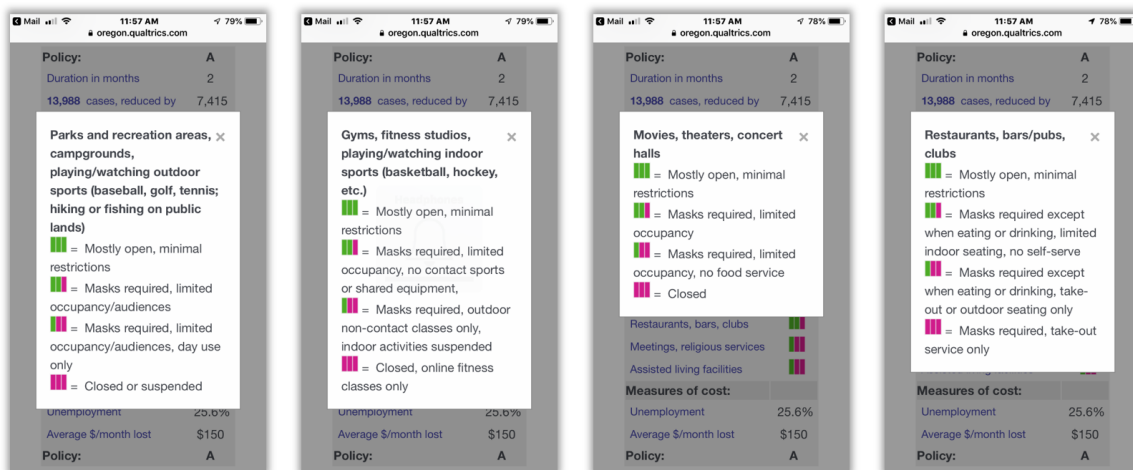


Figure D4. Contents of eighth through eleventh popups



146

Figure D5. Contents of twelfth through 15th popups

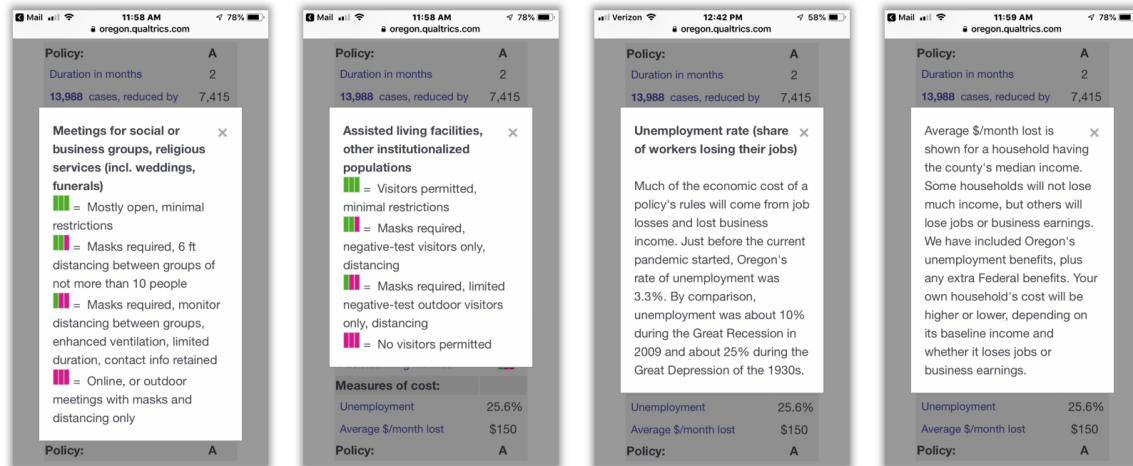


Figure D6. Choice question that immediately follows the table

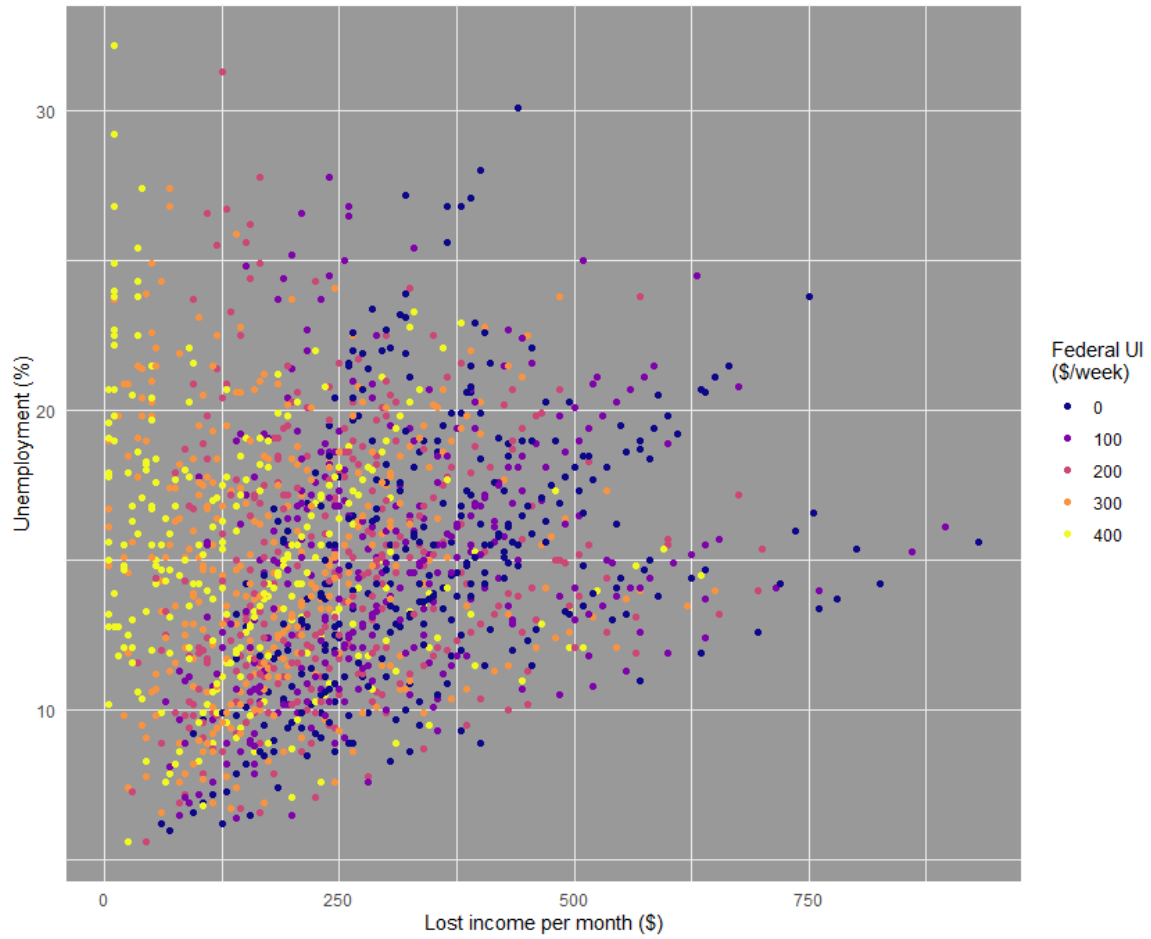
In a vote about whether to adopt Policy A, or to set no pandemic rules, I would choose:

Policy A with its cost and its mix of rules
(reduces cases by 7,415 and deaths by 70)

No pandemic rules, and just accept the
13,988 cases and 147 deaths expected if
we let everyone decide for themselves what
to do

B.5 Online Appendix: Joint distribution, key choice-set design features

Figure E1. Independent variation in average household costs and unemployment, by level of federal UI supplement



B.6 Online Appendix: Complete estimation results

Table F1 provides descriptive statistics, across all offered county-level pandemic policies, for the levels of restrictions placed on ten different categories of activities and businesses. The statistics in this table assume that the levels of restrictions (level 0, 1, 2, or 3) are a cardinal variable. For each policy, these restriction levels are randomized, although some implausible combinations are excluded. For details, see the online supplementary materials that describe the attribute randomization in detail.¹

Controls for these restrictions are included in all specifications in the body of the paper, either as cardinal variables (for example, Model 3 in Table 9), or as sets of indicators for each level (for example, Model 6 in Table 9). The coefficients on these variables (and interactions with these variables) are contained amongst the full sets of parameter estimates in Tables F2, F3 and F4 in this appendix.

Table F1. Descriptive statistics for restrictions on activities, across offered pandemic policies (included in this paper as incidental controls)

	Mean	SD	Min	Max
Grocery, essential retail	1.18	0.85	0	2
Non-essential retail	1.65	1.11	0	3
Schools, daycare	1.32	1.09	0	3
Parks, outdoor sports	1.34	1.09	0	3
Gyms, indoor sports	1.61	1.11	0	3
Theaters, concert halls	1.48	1.12	0	3
Restaurants, bars, clubs	1.53	1.11	0	3
Meetings, religious services	1.51	1.1	0	3
Assisted living facilities	1.62	1.11	0	3
Universities, colleges	1.49	1.12	0	3

¹The randomization process for the choice sets is described at http://pages.uoregon.edu/cameron/UO_COVID_description_of_randomizations.pdf. The joint distributions of the randomized design variables are shown in the document at http://pages.uoregon.edu/cameron/UO_COVID_survey_orthogonality.pdf

Table F2. Full set of parameter estimates for the models reported in Table 9 in the body of the paper: Effects of Federal UI (selected coefficients)

	<i>Dependent variable:</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Selected coefficients, as reported in Table 9 in the body of the paper:						
Avg. hhld cost for cty	0.004 (0.051)			-0.018 (0.050)		
Avg. hhld cost for cty (fed UI = 0)		-0.284*** (0.093)	-0.291*** (0.093)		-0.301*** (0.098)	-0.310*** (0.098)
Avg. hhld cost for cty (fed UI > 0)			0.174*** (0.066)			0.153** (0.065)
Avg. hhld cost for cty (fed UI = 100)		0.125 (0.102)			0.088 (0.103)	
Avg. hhld cost for cty (fed UI = 200)		0.131 (0.173)			0.105 (0.164)	
Avg. hhld cost for cty (fed UI = 300)		0.665*** (0.124)			0.659*** (0.124)	
Avg. hhld cost for cty (fed UI = 400)		0.166 (0.144)			0.194 (0.138)	
Unempl rate for cty	-0.0003 (0.020)			0.002 (0.021)		
Unempl rate for cty (fed UI = 0)		0.097** (0.039)	0.093** (0.039)		0.098** (0.040)	0.092** (0.040)
Unempl rate for cty (fed UI > 0)			-0.024 (0.020)			-0.023 (0.020)
Unempl rate for cty (fed UI = 100)		-0.028 (0.034)			-0.017 (0.035)	
Unempl rate for cty (fed UI = 200)		-0.001 (0.045)			0.001 (0.043)	
Unempl rate for cty (fed UI = 300)		-0.079*** (0.026)			-0.083*** (0.026)	
Unempl rate for cty (fed UI = 400)		-0.015 (0.025)			-0.014 (0.026)	
Absolute '00s cases/mo/50,000	-0.041 (0.032)	-0.036 (0.031)	-0.035 (0.031)	-0.042 (0.032)	-0.037 (0.032)	-0.036 (0.031)
Absolute deaths/mo/50,000	-0.034* (0.020)	-0.037* (0.020)	-0.038* (0.020)	-0.027 (0.020)	-0.026 (0.020)	-0.029 (0.020)
1=Status quo alternative	-1.982*** (0.296)	-2.008*** (0.284)	-1.994*** (0.291)	-2.433*** (0.360)	-2.552*** (0.370)	-2.472*** (0.368)
Other coefficients, suppressed in Table 9 in the body of the paper:						
Grocery, essential retail	-0.045 (0.078)	-0.072 (0.079)	-0.056 (0.078)			
Non-essential retail	-0.059 (0.061)	-0.078 (0.062)	-0.069 (0.062)			
Schools, daycare	-0.138** (0.067)	-0.153** (0.069)	-0.144** (0.068)			
Universities, colleges	-0.023 (0.059)	-0.025 (0.060)	-0.025 (0.059)			
Parks, outdoor sports	-0.058 (0.061)	-0.088 (0.062)	-0.080 (0.061)			
Gyms, indoor sports	-0.067 (0.059)	-0.078 (0.060)	-0.082 (0.059)			
Theaters, concert halls	-0.121** (0.060)	-0.128** (0.060)	-0.127** (0.059)			
Restaurants, bars, clubs	-0.070 (0.062)	-0.085 (0.063)	-0.069 (0.063)			
Meetings, religious services	-0.021 (0.062)	-0.033 (0.061)	-0.028 (0.061)			
Assisted living facilities	0.059 (0.061)	0.052 (0.062)	0.060 (0.062)			
1(Grocery, essential retail=1)				-0.103 (0.179)	-0.165 (0.185)	-0.143 (0.181)
1(Grocery, essential retail=2)				-0.155 (0.162)	-0.234 (0.167)	-0.185 (0.164)
1(Non-essential retail=1)				-0.430** (0.198)	-0.496** (0.201)	-0.453** (0.200)
1(Non-essential retail=2)				0.023 (0.212)	-0.012 (0.219)	0.017 (0.217)
1(Non-essential retail=3)				-0.378* (0.197)	-0.440** (0.202)	-0.394** (0.200)
1(Schools, daycare=1)				-0.114 (0.180)	-0.124 (0.182)	-0.111 (0.178)
1(Schools, daycare=2)				-0.077 (0.206)	-0.078 (0.211)	-0.070 (0.209)
1(Schools, daycare=3)				-0.466** (0.207)	-0.473** (0.214)	-0.457** (0.209)

Continued on next page

Table F2 – continued from previous page

	(1)	(2)	(3)	(4)	(5)	(6)
1(Universities, colleges=1)				-0.190 (0.176)	-0.221 (0.180)	-0.219 (0.179)
1(Universities, colleges=2)				-0.124 (0.202)	-0.117 (0.204)	-0.117 (0.204)
1(Universities, colleges=3)				-0.134 (0.195)	-0.152 (0.199)	-0.155 (0.197)
1(Parks, outdoor sports=1)				-0.117 (0.180)	-0.181 (0.184)	-0.154 (0.181)
1(Parks, outdoor sports=2)				-0.206 (0.185)	-0.248 (0.190)	-0.223 (0.187)
1(Parks, outdoor sports=3)				-0.112 (0.209)	-0.217 (0.210)	-0.178 (0.207)
1(Gyms, indoor sports=1)				-0.028 (0.211)	-0.076 (0.215)	-0.072 (0.211)
1(Gyms, indoor sports=2)				-0.334* (0.201)	-0.360* (0.205)	-0.361* (0.204)
1(Gyms, indoor sports=3)				-0.181 (0.201)	-0.203 (0.208)	-0.220 (0.203)
1(Theaters, concert halls=1)				-0.038 (0.191)	-0.048 (0.194)	-0.030 (0.190)
1(Theaters, concert halls=2)				-0.287 (0.188)	-0.249 (0.193)	-0.257 (0.192)
1(Theaters, concert halls=3)				-0.281 (0.191)	-0.327* (0.196)	-0.307 (0.191)
1(Restaurants, bars, clubs=1)				-0.104 (0.196)	-0.128 (0.202)	-0.108 (0.197)
1(Restaurants, bars, clubs=2)				-0.076 (0.190)	-0.152 (0.194)	-0.124 (0.192)
1(Restaurants, bars, clubs=3)				-0.196 (0.202)	-0.260 (0.210)	-0.197 (0.207)
1(Meetings, religious services=1)				-0.221 (0.186)	-0.247 (0.189)	-0.226 (0.186)
1(Meetings, religious services=2)				-0.156 (0.194)	-0.162 (0.196)	-0.158 (0.192)
1(Meetings, religious services=3)				-0.072 (0.204)	-0.095 (0.200)	-0.077 (0.200)
1(Assisted living facilities=1)				-0.124 (0.197)	-0.166 (0.202)	-0.122 (0.199)
1(Assisted living facilities=2)				0.047 (0.206)	0.010 (0.208)	0.044 (0.207)
1(Assisted living facilities=3)				0.129 (0.201)	0.083 (0.210)	0.127 (0.206)
$\hat{R}P \times \text{Avg. hhld costs}$	0.085*** (0.031)			0.087*** (0.032)		
$\hat{R}P \times \text{Avg. hhld costs (fed UI = 0)}$		0.396* (0.216)	0.377* (0.216)		0.497* (0.273)	0.476* (0.271)
$\hat{R}P \times \text{Avg. hhld costs (fed UI = 100)}$		-0.004 (0.165)			-0.015 (0.188)	
$\hat{R}P \times \text{Avg. hhld costs (fed UI = 200)}$		-0.241 (0.193)			-0.150 (0.183)	
$\hat{R}P \times \text{Avg. hhld costs (fed UI = 300)}$		-0.167 (0.209)			0.0005 (0.214)	
$\hat{R}P \times \text{Avg. hhld costs (fed UI = 400)}$		0.183 (0.214)			0.105 (0.243)	
$\hat{R}P \times \text{Avg. hhld costs (fed UI > 0)}$			-0.067 (0.092)			0.002 (0.097)
$\hat{R}P \times \text{Unempl rate}$	0.104 (0.085)			0.193*** (0.074)		
$\hat{R}P \times \text{Unempl rate (fed UI = 0)}$		0.024 (0.071)	0.023 (0.071)		0.023 (0.092)	0.025 (0.090)
$\hat{R}P \times \text{Unempl rate (fed UI = 100)}$		0.108* (0.058)			0.103* (0.059)	
$\hat{R}P \times \text{Unempl rate (fed UI = 200)}$		0.152** (0.062)			0.147*** (0.049)	
$\hat{R}P \times \text{Unempl rate (fed UI = 300)}$		0.108** (0.047)			0.106* (0.058)	
$\hat{R}P \times \text{Unempl rate (fed UI = 400)}$		0.079** (0.031)			0.083** (0.037)	
$\hat{R}P \times \text{Unempl rate (fed UI > 0)}$			0.101*** (0.032)			0.104*** (0.033)
$\hat{R}P \times \text{'00s cases/mo/50,000}$	-0.319 (0.485)	-0.271 (0.353)	-0.343 (0.369)	-0.553 (0.388)	-0.507 (0.421)	-0.575 (0.367)
$\hat{R}P \times \text{deaths/mo/50,000}$	-0.014 (0.024)	-0.023 (0.024)	-0.022 (0.026)	-0.040 (0.027)	-0.066** (0.033)	-0.056* (0.030)
$\hat{R}P \times 1=\text{Status quo alternative}$	0.616 (0.436)	0.655** (0.326)	0.573 (0.372)	1.165** (0.504)	1.360** (0.543)	1.111** (0.522)
$\hat{R}P \times \text{Grocery, essential retail}$	-0.147 (0.097)	-0.135 (0.100)	-0.138 (0.104)			
$\hat{R}P \times \text{Non-essential retail}$	0.114 (0.082)	0.105 (0.085)	0.099 (0.080)			
$\hat{R}P \times \text{Schools, daycare}$	-0.067 (0.097)	-0.070 (0.098)	-0.071 (0.097)			

Continued on next page

Table F2 – continued from previous page

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{R}P \times$ Universities, colleges	-0.070 (0.083)	-0.089 (0.085)	-0.083 (0.079)			
$\hat{R}P \times$ Parks, outdoor sports	0.007 (0.096)	0.037 (0.090)	0.025 (0.087)			
$\hat{R}P \times$ Gyms, indoor sports	-0.012 (0.078)	0.038 (0.087)	0.034 (0.083)			
$\hat{R}P \times$ Theaters, concert halls	-0.032 (0.083)	-0.014 (0.076)	-0.014 (0.077)			
$\hat{R}P \times$ Restaurants, bars, clubs	-0.068 (0.084)	-0.075 (0.080)	-0.095 (0.086)			
$\hat{R}P \times$ Meetings, religious services	-0.084 (0.084)	-0.057 (0.084)	-0.062 (0.083)			
$\hat{R}P \times$ Assisted living facilities	-0.199** (0.081)	-0.239*** (0.079)	-0.228*** (0.080)			
$\hat{R}P \times$ 1(Grocery, essential retail=1)				-0.269 (0.242)	-0.210 (0.270)	-0.256 (0.241)
$\hat{R}P \times$ 1(Grocery, essential retail=2)				-0.234 (0.245)	-0.130 (0.254)	-0.241 (0.249)
$\hat{R}P \times$ 1(Non-essential retail=1)				0.936*** (0.310)	0.988*** (0.318)	0.951*** (0.305)
$\hat{R}P \times$ 1(Non-essential retail=2)				0.876*** (0.327)	0.828** (0.327)	0.821*** (0.316)
$\hat{R}P \times$ 1(Non-essential retail=3)				0.676** (0.313)	0.635** (0.319)	0.590* (0.309)
$\hat{R}P \times$ 1(Schools, daycare=1)				-0.024 (0.266)	-0.056 (0.283)	-0.018 (0.263)
$\hat{R}P \times$ 1(Schools, daycare=2)				-0.268 (0.298)	-0.349 (0.312)	-0.278 (0.292)
$\hat{R}P \times$ 1(Schools, daycare=3)				-0.229 (0.311)	-0.419 (0.340)	-0.339 (0.312)
$\hat{R}P \times$ 1(Universities, colleges=1)				-0.055 (0.261)	0.063 (0.263)	0.022 (0.258)
$\hat{R}P \times$ 1(Universities, colleges=2)				0.117 (0.354)	0.067 (0.324)	0.056 (0.326)
$\hat{R}P \times$ 1(Universities, colleges=3)				-0.361 (0.299)	-0.391 (0.294)	-0.332 (0.275)
$\hat{R}P \times$ 1(Parks, outdoor sports=1)				0.509** (0.254)	0.581** (0.292)	0.504** (0.249)
$\hat{R}P \times$ 1(Parks, outdoor sports=2)				0.594** (0.292)	0.724** (0.316)	0.599** (0.291)
$\hat{R}P \times$ 1(Parks, outdoor sports=3)				-0.166 (0.321)	-0.062 (0.327)	-0.109 (0.316)
$\hat{R}P \times$ 1(Gyms, indoor sports=1)				-0.229 (0.299)	-0.023 (0.309)	-0.073 (0.283)
$\hat{R}P \times$ 1(Gyms, indoor sports=2)				0.642** (0.294)	0.746** (0.311)	0.765** (0.301)
$\hat{R}P \times$ 1(Gyms, indoor sports=3)				-0.162 (0.305)	-0.042 (0.319)	-0.050 (0.309)
$\hat{R}P \times$ 1(Theaters, concert halls=1)				-0.647** (0.268)	-0.710** (0.285)	-0.692*** (0.259)
$\hat{R}P \times$ 1(Theaters, concert halls=2)				-0.308 (0.270)	-0.483* (0.270)	-0.405 (0.262)
$\hat{R}P \times$ 1(Theaters, concert halls=3)				-0.559* (0.297)	-0.436 (0.307)	-0.462 (0.291)
$\hat{R}P \times$ 1(Restaurants, bars, clubs=1)				-0.107 (0.316)	-0.094 (0.317)	-0.111 (0.308)
$\hat{R}P \times$ 1(Restaurants, bars, clubs=2)				-0.567** (0.271)	-0.455 (0.280)	-0.526* (0.277)
$\hat{R}P \times$ 1(Restaurants, bars, clubs=3)				-0.421 (0.307)	-0.357 (0.310)	-0.433 (0.308)
$\hat{R}P \times$ 1(Meetings, religious services=1)				0.137 (0.311)	0.092 (0.344)	0.054 (0.310)
$\hat{R}P \times$ 1(Meetings, religious services=2)				-0.265 (0.283)	-0.280 (0.303)	-0.296 (0.278)
$\hat{R}P \times$ 1(Meetings, religious services=3)				-0.262 (0.310)	-0.236 (0.317)	-0.291 (0.308)
$\hat{R}P \times$ 1(Assisted living facilities=1)				-0.468 (0.349)	-0.430 (0.352)	-0.509 (0.342)
$\hat{R}P \times$ 1(Assisted living facilities=2)				-0.292 (0.333)	-0.297 (0.341)	-0.366 (0.328)
$\hat{R}P \times$ 1(Assisted living facilities=3)				-0.988*** (0.302)	-1.033*** (0.315)	-1.063*** (0.306)
Restrictions on activities (continuous vars)	✓	✓	✓			
Restrictions on activities (indicators)				✓	✓	✓
Respondents	993	993	993	993	993	993
Choices	1986	1986	1986	1986	1986	1986
Log likelihood	-1205.77	-1184.09	-1194.52	-1180.75	-1158.56	-1169.8

Notes: *p<0.1; **p<0.05; ***p<0.01

Table F3. Full sets of parameter estimates for the models reported in Panel A of Table 10 in the body of the paper: Heterogeneity in preferences

<i>Dep. var:</i> 1=Preferred policy	18 to 34 (1)	35 to 64 (2)	65 + (3)	Non-white (4)	White (5)	Women (6)	Men (7)
Avg. hhld cost (fed UI = 0)	-0.597*** (0.224)	-0.382*** (0.139)	-0.307 (0.181)	-0.842*** (0.376)	-0.228** (0.112)	-0.497*** (0.165)	-0.220* (0.114)
Avg. hhld cost (fed UI > 0)	0.251** (0.124)	0.212* (0.119)	0.181 (0.140)	0.052 (0.126)	0.216*** (0.079)	0.155 (0.106)	0.134 (0.082)
Unempl rate (fed UI = 0)	0.251*** (0.093)	0.080 (0.055)	0.219 (0.161)	0.268** (0.118)	0.090* (0.046)	0.141** (0.064)	0.093 (0.057)
Unempl rate (fed UI > 0)	0.023 (0.044)	-0.066** (0.031)	-0.012 (0.069)	0.011 (0.050)	-0.029 (0.024)	-0.049 (0.036)	0.020 (0.029)
Absolute '00s cases/mo/50k	-0.047 (0.055)	0.068 (0.049)	-0.045 (0.099)	-0.055 (0.071)	-0.048 (0.040)	-0.086* (0.044)	0.008 (0.049)
Absolute deaths/mo/50k	-0.074** (0.033)	-0.011 (0.035)	-0.059 (0.056)	-0.070* (0.039)	-0.001 (0.027)	0.001 (0.032)	-0.061** (0.028)
1=Status quo alternative	-4.533*** (0.729)	-1.943*** (0.559)	0.966 (1.211)	-2.603*** (0.761)	-2.584*** (0.443)	-2.925*** (0.557)	-2.039*** (0.508)
Other coefficients, suppressed in Table 10 Panel A in the body of the paper:							
Grocery, essential retail.1	-0.556 (0.342)	0.375 (0.295)	-0.835 (0.537)	0.110 (0.334)	-0.313 (0.229)	-0.058 (0.258)	-0.055 (0.276)
Grocery, essential retail.2	-1.117*** (0.373)	0.251 (0.335)	2.156*** (0.624)	0.149 (0.410)	-0.239 (0.265)	-0.192 (0.311)	-0.217 (0.291)
Non-essential retail.1	-0.656** (0.300)	-0.022 (0.254)	-0.825 (0.540)	0.374 (0.317)	-0.476** (0.202)	-0.011 (0.234)	-0.382 (0.240)
Non-essential retail.2	-0.454 (0.410)	-0.019 (0.293)	0.762 (0.539)	-0.625 (0.497)	0.281 (0.250)	-0.522 (0.338)	0.336 (0.291)
Non-essential retail.3	-0.954** (0.371)	0.230 (0.277)	-0.360 (0.514)	-0.276 (0.366)	-0.462** (0.226)	-0.543** (0.269)	0.112 (0.272)
Schools, daycare.1	-0.127 (0.374)	-0.677** (0.282)	-0.700 (0.631)	-0.165 (0.378)	-0.777*** (0.258)	-0.325 (0.308)	-0.424 (0.287)
Schools, daycare.2	-0.558 (0.398)	-0.131 (0.301)	0.077 (0.659)	-0.762 (0.465)	-0.163 (0.237)	-0.668** (0.327)	0.005 (0.277)
Schools, daycare.3	-1.117*** (0.351)	0.297 (0.295)	0.962 (0.686)	-0.059 (0.358)	-0.058 (0.250)	-0.228 (0.299)	-0.039 (0.285)
Universities, colleges.1	-1.016*** (0.357)	0.119 (0.277)	1.374** (0.563)	-0.005 (0.383)	-0.291 (0.234)	-0.331 (0.283)	-0.0003 (0.249)
Universities, colleges.2	-0.254 (0.342)	-0.301 (0.344)	1.165* (0.648)	-0.030 (0.421)	-0.114 (0.245)	-0.074 (0.323)	-0.356 (0.294)
Universities, colleges.3	-1.201*** (0.413)	0.104 (0.317)	0.164 (0.606)	-0.134 (0.422)	-0.197 (0.237)	0.068 (0.315)	-0.602** (0.292)
Parks, outdoor sports.1	-0.153 (0.384)	0.072 (0.376)	0.751 (0.667)	0.503 (0.445)	-0.289 (0.266)	0.157 (0.325)	-0.005 (0.300)
Parks, outdoor sports.2	-0.837** (0.381)	0.094 (0.300)	-0.260 (0.698)	-0.391 (0.458)	-0.093 (0.231)	-0.612* (0.313)	0.026 (0.284)
Parks, outdoor sports.3	-0.403 (0.389)	0.208 (0.277)	-0.277 (0.434)	0.153 (0.432)	-0.240 (0.242)	0.152 (0.306)	-0.396 (0.277)
Gyms, indoor sports.1	-0.876** (0.399)	-0.163 (0.311)	-0.806 (0.508)	0.405 (0.397)	-0.768*** (0.251)	-0.181 (0.317)	-0.506* (0.284)
Gyms, indoor sports.2	-0.583* (0.325)	0.329 (0.289)	0.114 (0.457)	-0.780* (0.429)	0.157 (0.229)	-0.065 (0.269)	-0.111 (0.304)
Gyms, indoor sports.3	0.809** (0.343)	-0.645** (0.305)	0.426 (0.595)	-0.236 (0.384)	-0.227 (0.249)	0.259 (0.283)	-0.546* (0.295)
Theaters, concert halls.1	-0.861*** (0.315)	0.316 (0.285)	-0.315 (0.567)	-0.428 (0.376)	-0.096 (0.225)	-0.433* (0.251)	0.326 (0.258)
Theaters, concert halls.2	-0.888*** (0.325)	-0.072 (0.312)	1.426** (0.667)	-0.864** (0.420)	-0.014 (0.239)	-0.093 (0.276)	-0.598** (0.289)
Theaters, concert halls.3	0.742** (0.361)	-0.242 (0.319)	0.674 (0.627)	0.134 (0.446)	0.042 (0.254)	0.500 (0.310)	-0.473 (0.305)
Restaurants, bars, clubs.1	-0.561 (0.373)	-0.016 (0.320)	0.892 (0.678)	-0.378 (0.420)	-0.074 (0.255)	-0.089 (0.332)	0.061 (0.307)
Restaurants, bars, clubs.2	-1.023*** (0.361)	0.093 (0.303)	-1.016* (0.596)	-1.098*** (0.374)	-0.067 (0.248)	-0.134 (0.270)	-0.735*** (0.284)
Restaurants, bars, clubs.3	0.734** (0.358)	-0.077 (0.327)	1.056 (0.672)	0.028 (0.412)	0.072 (0.246)	0.387 (0.302)	-0.305 (0.294)
Meetings, relig. services.1	-0.919** (0.397)	-0.071 (0.333)	-0.379 (0.732)	-0.823** (0.407)	-0.332 (0.261)	-0.579* (0.311)	-0.449 (0.296)
Meetings, relig. services.2	-0.224 (0.340)	0.614 (0.384)	0.290 (0.531)	-0.219 (0.399)	-0.020 (0.242)	-0.376 (0.267)	0.209 (0.315)
Meetings, relig. services.3	-1.345*** (0.403)	0.098 (0.286)	0.049 (0.510)	-0.533 (0.371)	-0.172 (0.223)	-0.491* (0.270)	-0.127 (0.274)
Assisted living facilities.1	-0.860** (0.360)	0.305 (0.288)	1.611*** (0.493)	0.139 (0.351)	-0.198 (0.226)	-0.248 (0.277)	0.117 (0.268)
Assisted living facilities.2	0.222 (0.371)	-0.204 (0.349)	1.291* (0.696)	-0.018 (0.384)	-0.118 (0.230)	-0.234 (0.284)	0.132 (0.289)
Assisted living facilities.3	-1.445*** (0.403)	0.338 (0.324)	2.134*** (0.613)	-0.725* (0.405)	0.177 (0.260)	-0.444 (0.311)	0.175 (0.317)
Avg. hhld cost (fed UI > 0) × \hat{RP}	0.221 (0.186)	-0.094 (0.208)	-1.737*** (0.456)	-0.428 (0.401)	0.017 (0.112)	0.717*** (0.250)	-0.229* (0.130)

Continued on next page

Table F3 – continued from previous page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Avg. hhld cost (fed UI = 0) $\times \hat{R}P$	0.573 (0.725)	0.304 (0.370)	0.017 (0.979)	1.201 (0.778)	0.301 (0.349)	1.106*** (0.406)	0.087 (0.426)
Unempl rate (fed UI > 0) $\times \hat{R}P$	-0.016 (0.080)	0.291*** (0.069)	0.307 (0.234)	0.103 (0.103)	0.144*** (0.043)	0.161*** (0.053)	0.029 (0.069)
Unempl rate (fed UI = 0) $\times \hat{R}P$	-0.182 (0.259)	0.288* (0.152)	-0.364 (0.279)	-0.389 (0.291)	0.143 (0.125)	-0.108 (0.144)	0.081 (0.161)
Absolute '00s cases/mo/50k $\times \hat{R}P$	-0.882 (1.126)	-2.055*** (0.663)	-4.835* (2.804)	0.166 (1.792)	-0.795* (0.420)	-0.158 (0.861)	-1.185* (0.609)
Absolute deaths/mo/50k $\times \hat{R}P$	0.030 (0.068)	-0.160** (0.066)	-0.152 (0.135)	-0.085 (0.105)	-0.081** (0.032)	-0.168*** (0.059)	-0.042 (0.044)
Grocery, essential retail.1 $\times \hat{R}P$	-0.329 (0.616)	-1.112** (0.538)	4.400*** (1.667)	-0.112 (0.966)	-0.082 (0.299)	-0.904* (0.499)	0.081 (0.457)
Grocery, essential retail.2 $\times \hat{R}P$	2.067*** (0.629)	-1.724** (0.681)	-5.040*** (1.508)	-2.513** (1.211)	-0.005 (0.372)	-0.906 (0.643)	-0.331 (0.455)
Non-essential retail.1 $\times \hat{R}P$	0.018 (0.519)	-0.422 (0.507)	5.049*** (1.539)	-1.585* (0.936)	0.222 (0.312)	-0.432 (0.487)	-0.217 (0.375)
Non-essential retail.2 $\times \hat{R}P$	-0.799 (0.693)	1.160** (0.507)	-2.743* (1.405)	-0.063 (1.165)	-0.393 (0.330)	-0.413 (0.545)	-0.583 (0.431)
Non-essential retail.3 $\times \hat{R}P$	1.068* (0.623)	-1.011* (0.529)	-0.305 (1.331)	1.204 (1.108)	0.097 (0.353)	0.275 (0.558)	-0.487 (0.475)
Schools, daycare.1 $\times \hat{R}P$	1.621** (0.717)	1.389*** (0.462)	1.950 (1.637)	1.222 (0.987)	1.149*** (0.393)	-0.265 (0.634)	1.166*** (0.436)
Schools, daycare.2 $\times \hat{R}P$	0.828 (0.860)	-0.059 (0.542)	-0.337 (1.407)	1.900** (0.966)	0.425 (0.399)	1.665*** (0.600)	-0.750 (0.530)
Schools, daycare.3 $\times \hat{R}P$	1.412** (0.590)	-1.460*** (0.552)	-3.425** (1.588)	-0.927 (1.038)	-0.519 (0.327)	-0.848 (0.577)	-0.640 (0.506)
Universities, colleges.1 $\times \hat{R}P$	0.956 (0.654)	0.077 (0.451)	-0.130 (1.502)	-0.086 (0.982)	0.372 (0.382)	0.436 (0.517)	0.749* (0.431)
Universities, colleges.2 $\times \hat{R}P$	0.126 (0.636)	-1.597*** (0.536)	-1.851 (1.433)	-2.014* (1.051)	-0.475 (0.337)	-1.136* (0.551)	-0.133 (0.490)
Universities, colleges.3 $\times \hat{R}P$	0.760 (0.534)	0.248 (0.559)	-4.176*** (1.473)	0.813 (0.973)	-0.109 (0.345)	0.542 (0.529)	-0.018 (0.469)
Parks, outdoor sports.1 $\times \hat{R}P$	2.042*** (0.694)	0.811 (0.611)	0.029 (1.771)	0.177 (1.075)	1.102*** (0.369)	-0.307 (0.627)	1.444*** (0.506)
Parks, outdoor sports.2 $\times \hat{R}P$	-0.223 (0.717)	-1.389** (0.554)	3.066* (1.789)	-0.936 (0.980)	-0.216 (0.348)	-0.674 (0.534)	0.123 (0.513)
Parks, outdoor sports.3 $\times \hat{R}P$	0.429 (0.621)	-1.967*** (0.556)	-0.050 (1.818)	-0.620 (1.050)	-0.171 (0.361)	-0.785 (0.551)	0.305 (0.482)
Gyms, indoor sports.1 $\times \hat{R}P$	1.468** (0.677)	0.315 (0.606)	1.910 (1.273)	-1.472 (1.077)	1.164*** (0.350)	-1.293** (0.658)	1.414*** (0.500)
Gyms, indoor sports.2 $\times \hat{R}P$	0.483 (0.636)	-1.798*** (0.478)	-1.117 (1.217)	0.649 (0.974)	-1.113*** (0.314)	-0.678 (0.506)	-0.289 (0.511)
Gyms, indoor sports.3 $\times \hat{R}P$	-2.048*** (0.626)	0.022 (0.546)	-4.689*** (1.538)	0.448 (0.964)	-0.447 (0.405)	-0.712 (0.531)	-0.177 (0.577)
Theaters, concert halls.1 $\times \hat{R}P$	1.064* (0.567)	0.161 (0.440)	-0.057 (1.461)	1.218 (0.906)	0.044 (0.361)	0.810 (0.505)	-0.842 (0.532)
Theaters, concert halls.2 $\times \hat{R}P$	0.365 (0.519)	-0.402 (0.563)	-7.383*** (1.711)	0.636 (1.035)	-0.931*** (0.303)	-1.101** (0.520)	0.633 (0.493)
Theaters, concert halls.3 $\times \hat{R}P$	-1.235* (0.711)	-0.284 (0.505)	-5.552*** (1.892)	0.041 (1.036)	-0.478 (0.342)	0.082 (0.525)	-0.757* (0.454)
Restaurants, bars, clubs.1 $\times \hat{R}P$	-0.086 (0.600)	0.137 (0.712)	-2.260 (1.725)	2.002* (1.149)	-0.543 (0.342)	0.162 (0.652)	-1.314** (0.543)
Restaurants, bars, clubs.2 $\times \hat{R}P$	1.004* (0.602)	-0.775 (0.563)	-2.085 (1.533)	1.024 (0.955)	-0.789** (0.358)	-0.302 (0.589)	0.659 (0.481)
Restaurants, bars, clubs.3 $\times \hat{R}P$	-2.326*** (0.712)	-1.385** (0.599)	-3.882** (1.771)	-0.467 (1.005)	-0.968*** (0.335)	-1.059** (0.530)	-0.900** (0.422)
Meetings, relig. services.1 $\times \hat{R}P$	0.022 (0.619)	-1.406*** (0.535)	-2.001 (2.164)	1.530 (1.044)	-0.763** (0.357)	0.023 (0.530)	-0.462 (0.528)
Meetings, relig. services.2 $\times \hat{R}P$	-0.354 (0.653)	-1.449** (0.584)	0.172 (1.302)	-1.399 (1.090)	0.008 (0.372)	-0.189 (0.518)	-0.182 (0.570)
Meetings, relig. services.3 $\times \hat{R}P$	1.414** (0.630)	1.074** (0.494)	-0.111 (1.114)	0.945 (0.877)	-0.279 (0.325)	1.734*** (0.537)	-0.403 (0.400)
Assisted living facilities.1 $\times \hat{R}P$	2.711*** (0.656)	-1.940*** (0.502)	0.293 (1.255)	-1.515 (0.969)	0.407 (0.282)	0.233 (0.521)	0.037 (0.393)
Assisted living facilities.2 $\times \hat{R}P$	-0.525 (0.554)	-0.347 (0.603)	-5.309*** (1.721)	-0.998 (0.887)	-0.557* (0.313)	-0.651 (0.502)	-0.589 (0.574)
Assisted living facilities.3 $\times \hat{R}P$	0.760 (0.534)	0.248 (0.559)	-4.176*** (1.473)	0.813 (0.973)	-0.109 (0.345)	0.542 (0.529)	-0.018 (0.469)
1=Status quo alternative $\times \hat{R}P$	0.756 (0.657)	0.468 (0.499)	-1.521 (1.434)	-0.378 (0.966)	-0.298 (0.339)	-0.294 (0.650)	0.189 (0.416)
Respondents	317	453	223	295	698	507	480
Choices	634	906	446	590	1396	1014	960

Notes: *p<0.1; **p<0.05; ***p<0.01

Table F4. Full sets of parameter estimates for the models reported in Panel B of Table 10 in the body of the paper: Heterogeneity in preferences across socioeconomic groups

<i>Dep. var:</i> 1=Preferred policy	Lib (1)	Mod (2)	Cons (3)	Non-coll (4)	College (5)	< \$75k/yr (6)	> \$75k/yr (7)
Avg. hhld cost (fed UI = 0)	0.094 (0.212)	-0.891*** (0.229)	-0.411** (0.173)	-0.347* (0.210)	-0.239* (0.123)	-0.452*** (0.157)	-0.192 (0.143)
Avg. hhld cost (fed UI > 0)	-0.081 (0.207)	0.123 (0.118)	0.123 (0.108)	0.150* (0.082)	0.119 (0.109)	0.367*** (0.091)	0.028 (0.087)
Unempl rate (fed UI = 0)	-0.116 (0.087)	0.316*** (0.104)	0.089 (0.068)	0.140 (0.086)	0.015 (0.054)	0.250*** (0.073)	-0.015 (0.055)
Unempl rate (fed UI > 0)	0.063 (0.078)	-0.047 (0.040)	-0.082*** (0.031)	-0.042 (0.034)	-0.016 (0.031)	-0.017 (0.033)	-0.022 (0.028)
Absolute '00s cases/mo/50k	-0.036 (0.105)	-0.146** (0.057)	-0.081 (0.061)	-0.164*** (0.061)	0.081** (0.037)	-0.092* (0.055)	-0.012 (0.045)
Absolute deaths/mo/50k	-0.035 (0.068)	0.028 (0.044)	-0.083*** (0.030)	-0.029 (0.034)	-0.029 (0.026)	-0.028 (0.030)	-0.014 (0.029)
1=Status quo alternative	-5.006*** (1.017)	-3.647*** (0.832)	-2.854*** (0.650)	-2.497*** (0.514)	-2.869*** (0.559)	-2.705*** (0.504)	-2.321*** (0.578)
Other coefficients, suppressed in Table 10 Panel B in the body of the paper:							
Grocery, essential retail.1	0.077 (0.527)	0.441 (0.329)	-0.361 (0.293)	0.093 (0.289)	-0.398 (0.252)	-0.493* (0.264)	0.096 (0.276)
Non-essential retail.1	0.230 (0.406)	-0.381 (0.322)	-0.214 (0.279)	-0.352 (0.236)	-0.066 (0.256)	-0.413 (0.266)	-0.103 (0.248)
Schools, daycare.1	-0.546 (0.530)	-1.181*** (0.384)	-0.152 (0.357)	-0.103 (0.319)	-0.554** (0.270)	-0.449 (0.302)	-0.703** (0.296)
Parks, outdoor sports.1	0.583 (0.649)	0.563 (0.441)	-0.760** (0.358)	0.157 (0.324)	0.080 (0.313)	-0.119 (0.333)	0.009 (0.342)
Gyms, indoor sports.1	-0.250 (0.555)	-0.092 (0.423)	-0.715** (0.335)	-0.269 (0.318)	-0.268 (0.280)	-0.503 (0.316)	-0.274 (0.305)
Theaters, concert halls.1	-0.929 (0.569)	-0.437 (0.336)	0.841*** (0.296)	-0.062 (0.307)	-0.078 (0.249)	-0.416 (0.286)	0.053 (0.264)
Restaurants, bars, clubs.1	-0.574 (0.574)	0.588 (0.444)	-0.048 (0.333)	0.104 (0.309)	-0.253 (0.294)	-0.315 (0.311)	-0.091 (0.298)
Meetings, relig services.1	-0.603 (0.554)	-0.251 (0.450)	-0.572 (0.350)	-0.679** (0.325)	-0.455 (0.294)	-0.987*** (0.321)	-0.242 (0.307)
Assisted living facilities.1	-1.370* (0.774)	0.066 (0.385)	-0.075 (0.301)	-0.501* (0.296)	0.121 (0.254)	0.182 (0.286)	-0.372 (0.251)
Universities, colleges.1	-2.786*** (0.715)	-0.140 (0.349)	0.475 (0.289)	-0.641** (0.293)	0.137 (0.260)	-0.355 (0.276)	-0.063 (0.276)
Grocery, essential retail.2	-2.384*** (0.724)	-0.325 (0.415)	-0.181 (0.301)	-0.395 (0.319)	-0.042 (0.293)	-0.202 (0.306)	-0.189 (0.307)
Non-essential retail.2	-0.823 (0.562)	-0.054 (0.404)	-0.445 (0.357)	-0.424 (0.379)	0.105 (0.288)	-0.503 (0.387)	0.270 (0.279)
Schools, daycare.2	-0.916 (0.595)	-1.020** (0.397)	-0.184 (0.336)	-0.928*** (0.340)	-0.059 (0.290)	-0.983*** (0.365)	0.126 (0.313)
Parks, outdoor sports.2	-0.019 (0.519)	-0.526 (0.363)	-0.002 (0.332)	-0.805** (0.335)	0.146 (0.288)	-0.905*** (0.346)	0.510* (0.287)
Gyms, indoor sports.2	0.786 (0.625)	-0.308 (0.377)	-0.239 (0.328)	0.403 (0.294)	-0.269 (0.274)	0.314 (0.290)	-0.487 (0.310)
Theaters, concert halls.2	-0.212 (0.548)	-0.747* (0.390)	-0.885*** (0.343)	-0.160 (0.292)	-0.210 (0.277)	-0.142 (0.293)	-0.567* (0.318)
Restaurants, bars, clubs.2	0.152 (0.587)	-0.823** (0.389)	-0.358 (0.322)	-0.401 (0.308)	-0.266 (0.287)	-0.508* (0.282)	-0.312 (0.324)
Meetings, relig services.2	-1.072* (0.576)	-0.266 (0.401)	0.253 (0.328)	-0.758** (0.310)	0.174 (0.283)	-0.652** (0.289)	0.787** (0.322)
Assisted living facilities.2	-0.361 (0.494)	-0.312 (0.387)	-0.067 (0.336)	-0.398 (0.303)	0.045 (0.276)	-0.485* (0.290)	0.436 (0.301)
Universities, colleges.2	-0.521 (0.594)	-0.052 (0.403)	-0.404 (0.359)	-0.756** (0.335)	0.161 (0.301)	-0.314 (0.330)	0.273 (0.301)
Non-essential retail.3	-0.408 (0.528)	-0.298 (0.360)	-0.058 (0.329)	-0.011 (0.326)	-0.486* (0.266)	0.117 (0.285)	-0.671** (0.285)
Schools, daycare.3	0.210 (0.571)	-0.403 (0.348)	-0.410 (0.321)	-0.171 (0.310)	-0.184 (0.290)	-0.294 (0.269)	-0.021 (0.319)
Parks, outdoor sports.3	0.904 (0.623)	0.131 (0.384)	-0.666* (0.344)	-0.120 (0.340)	-0.074 (0.276)	0.093 (0.326)	-0.256 (0.291)
Gyms, indoor sports.3	-0.191 (0.545)	-0.540 (0.410)	-0.798** (0.348)	0.252 (0.328)	-0.455 (0.284)	0.268 (0.319)	-0.557* (0.297)
Theaters, concert halls.3	1.044* (0.623)	-0.508 (0.436)	-0.956*** (0.336)	0.278 (0.325)	-0.027 (0.283)	0.080 (0.326)	-0.044 (0.307)
Restaurants, bars, clubs.3	0.494 (0.570)	-0.020 (0.438)	-0.422 (0.353)	0.252 (0.294)	0.231 (0.305)	0.136 (0.312)	0.263 (0.312)
Meetings, relig services.3	-0.185 (0.560)	-0.522 (0.389)	-0.139 (0.328)	0.060 (0.290)	-0.388 (0.246)	-0.039 (0.290)	-0.067 (0.279)
Assisted living facilities.3	0.166 (0.476)	-0.359 (0.398)	-0.042 (0.333)	-0.332 (0.318)	0.027 (0.284)	-0.406 (0.305)	0.100 (0.322)
Universities, colleges.3	0.143 (0.665)	-0.557 (0.404)	-0.604* (0.341)	-0.199 (0.314)	-0.193 (0.276)	-0.085 (0.315)	-0.356 (0.302)

Continued on next page

Table F4 – continued from previous page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Absolute deaths/mo/50k $\times \hat{RP}$	-0.203** (0.094)	-0.327*** (0.101)	0.068 (0.050)	0.036 (0.077)	-0.102** (0.048)	-0.265*** (0.079)	-0.010 (0.032)
Absolute '00s cases/mo/50k $\times \hat{RP}$	-0.034 (1.626)	0.616 (1.049)	0.537 (1.137)	-0.532 (1.304)	-1.170*** (0.447)	1.114 (0.982)	-1.156** (0.500)
Avg. hhld cost (fed UI > 0) $\times \hat{RP}$	0.100 (0.309)	-0.133 (0.237)	0.054 (0.237)	0.211 (0.263)	-0.091 (0.143)	-0.498** (0.250)	0.205 (0.126)
Avg. hhld cost (fed UI = 0) $\times \hat{RP}$	0.196 (0.604)	2.234*** (0.678)	-0.199 (0.417)	0.745 (0.646)	0.257 (0.338)	1.007** (0.405)	0.002 (0.505)
Unempl rate (fed UI > 0) $\times \hat{RP}$	0.508*** (0.126)	0.305*** (0.089)	-0.011 (0.068)	0.184*** (0.065)	0.100* (0.053)	0.223*** (0.065)	0.104** (0.046)
Unempl rate (fed UI = 0) $\times \hat{RP}$	0.756*** (0.195)	-0.601*** (0.227)	0.110 (0.157)	0.007 (0.242)	0.120 (0.125)	-0.321** (0.140)	0.291 (0.186)
Grocery, essential retail.1 $\times \hat{RP}$	0.346 (0.990)	-2.302*** (0.580)	0.542 (0.514)	1.116 (0.745)	-0.547* (0.313)	-0.473 (0.705)	-0.586** (0.298)
Non-essential retail.1 $\times \hat{RP}$	-1.729** (0.833)	0.358 (0.629)	-0.060 (0.521)	-0.016 (0.651)	-0.270 (0.397)	-1.172* (0.689)	-0.058 (0.322)
Schools, daycare.1 $\times \hat{RP}$	2.755** (1.203)	1.585*** (0.596)	0.471 (0.734)	-0.444 (0.930)	1.069*** (0.353)	2.382*** (0.733)	1.293*** (0.403)
Parks, outdoor sports.1 $\times \hat{RP}$	2.382** (1.096)	0.642 (0.697)	3.326*** (0.747)	-0.490 (0.962)	1.058** (0.461)	1.740** (0.708)	0.828* (0.442)
Gyms, indoor sports.1 $\times \hat{RP}$	-0.142 (1.093)	1.333 (0.882)	1.363** (0.564)	-0.555 (1.008)	0.380 (0.415)	1.148* (0.683)	0.814* (0.426)
Theaters, concert halls.1 $\times \hat{RP}$	-1.024 (0.861)	-0.567 (0.668)	0.257 (0.683)	1.521** (0.763)	-0.559 (0.378)	0.968 (0.699)	-0.028 (0.402)
Restaurants, bars, clubs.1 $\times \hat{RP}$	-2.070* (1.123)	-2.446** (1.013)	1.171 (0.783)	-0.741 (0.772)	-0.092 (0.401)	0.061 (0.610)	-0.169 (0.431)
Meetings, relig services.1 $\times \hat{RP}$	-2.574** (1.206)	-1.814** (0.804)	2.490*** (0.804)	0.691 (0.803)	-0.615 (0.404)	1.569** (0.706)	-0.983** (0.446)
Assisted living facilities.1 $\times \hat{RP}$	3.990*** (1.132)	-0.254 (0.683)	0.822 (0.616)	0.058 (0.701)	0.372 (0.327)	0.668 (0.620)	0.156 (0.348)
Universities, colleges.1 $\times \hat{RP}$	4.064*** (1.261)	0.905 (0.633)	0.894 (0.671)	0.523 (0.793)	0.316 (0.410)	1.341* (0.758)	0.070 (0.371)
Grocery, essential retail.2 $\times \hat{RP}$	2.065 (1.262)	0.598 (0.794)	0.256 (0.750)	-0.534 (0.871)	-0.410 (0.473)	-0.320 (0.724)	0.158 (0.507)
Non-essential retail.2 $\times \hat{RP}$	-3.299*** (1.048)	-0.207 (0.621)	1.411** (0.622)	1.527* (0.861)	-0.386 (0.369)	-0.059 (0.829)	0.149 (0.377)
Schools, daycare.2 $\times \hat{RP}$	-0.136 (1.120)	2.793*** (0.888)	-0.013 (0.659)	2.057*** (0.774)	0.466 (0.430)	0.844 (0.784)	0.847 (0.516)
Parks, outdoor sports.2 $\times \hat{RP}$	-4.274*** (1.423)	-0.179 (0.700)	-0.114 (0.584)	1.608** (0.780)	-0.450 (0.410)	-0.553 (0.744)	-0.120 (0.385)
Gyms, indoor sports.2 $\times \hat{RP}$	-3.966*** (0.952)	-1.551** (0.778)	-0.599 (0.574)	-0.731 (0.751)	-0.440 (0.379)	-0.088 (0.688)	-0.591* (0.346)
Theaters, concert halls.2 $\times \hat{RP}$	-1.635* (0.972)	-0.338 (0.588)	0.382 (0.612)	-1.810** (0.787)	-0.112 (0.343)	-0.669 (0.749)	0.073 (0.349)
Restaurants, bars, clubs.2 $\times \hat{RP}$	-4.357*** (1.260)	-0.070 (0.860)	-0.221 (0.604)	-0.124 (0.771)	-0.135 (0.423)	-0.127 (0.648)	-0.345 (0.465)
Meetings, relig services.2 $\times \hat{RP}$	3.316*** (1.123)	-0.481 (0.733)	-0.059 (0.564)	0.341 (0.770)	0.342 (0.414)	-0.494 (0.695)	-0.930** (0.391)
Assisted living facilities.2 $\times \hat{RP}$	-1.318 (1.071)	-0.424 (0.691)	-0.968* (0.587)	-0.599 (0.751)	-0.474 (0.403)	-0.889 (0.611)	-0.999** (0.393)
Universities, colleges.2 $\times \hat{RP}$	-2.083* (1.099)	-0.992 (0.741)	-0.180 (0.545)	-1.108 (0.784)	-0.326 (0.462)	-1.305* (0.679)	-0.869* (0.465)
Non-essential retail.3 $\times \hat{RP}$	-1.130 (0.956)	1.867*** (0.629)	-1.249* (0.648)	-0.178 (0.998)	0.478 (0.395)	-0.302 (0.638)	0.532 (0.397)
Schools, daycare.3 $\times \hat{RP}$	-1.820* (0.992)	1.969** (0.842)	0.083 (0.690)	-0.605 (0.944)	-0.290 (0.372)	-1.137* (0.622)	-0.475 (0.366)
Parks, outdoor sports.3 $\times \hat{RP}$	-1.325 (1.036)	0.228 (0.791)	0.668 (0.738)	-0.499 (0.893)	0.072 (0.428)	-0.960 (0.662)	-0.052 (0.429)
Gyms, indoor sports.3 $\times \hat{RP}$	-0.283 (1.284)	0.584 (0.817)	1.047 (0.686)	-1.571* (0.919)	0.130 (0.473)	-1.736** (0.728)	-0.341 (0.433)
Theaters, concert halls.3 $\times \hat{RP}$	-2.990** (1.316)	1.352* (0.747)	2.048*** (0.633)	-0.534 (0.909)	-0.417 (0.392)	-0.686 (0.724)	-0.751* (0.442)
Restaurants, bars, clubs.3 $\times \hat{RP}$	-1.687 (1.048)	-1.629** (0.698)	-0.090 (0.707)	-0.591 (0.801)	-1.322*** (0.377)	-1.007 (0.693)	-1.658*** (0.468)
Meetings, relig services.3 $\times \hat{RP}$	0.695 (1.172)	0.272 (0.645)	0.026 (0.635)	0.521 (0.631)	0.365 (0.363)	0.442 (0.704)	-0.031 (0.361)
Assisted living facilities.3 $\times \hat{RP}$	0.293 (0.980)	-1.658** (0.694)	0.011 (0.688)	0.285 (0.809)	0.171 (0.450)	0.754 (0.810)	-0.146 (0.433)
Universities, colleges.3 $\times \hat{RP}$	0.293 (0.980)	-1.658** (0.694)	0.011 (0.688)	0.285 (0.809)	0.171 (0.450)	0.754 (0.810)	-0.146 (0.433)
1=Status quo alternative $\times \hat{RP}$	-0.947 (1.127)	0.407 (0.633)	-0.176 (0.759)	-0.750 (0.781)	-0.082 (0.361)	-0.002 (0.752)	0.131 (0.383)
Respondents	338	309	303	398	589	489	475
Choices	676	618	606	796	1178	978	950

Notes: *p<0.1; **p<0.05; ***p<0.01

Table F5. Standard errors clustered by respondent, with federal UI entering as in Models 3 and 6 in Table 9 (Panel A) and in “baseline plus interactions” form (Panel B).

		<i>Dependent variable:</i>	
		1=Preferred policy	
Panel A	Model:	(3)	(6)
Avg. hhld cost for county (federal UI = 0)		-0.291** (0.142)	-0.310** (0.147)
Avg. hhld cost for county (federal UI > 0)		0.174* (0.092)	0.153* (0.091)
Unempl rate for county (federal UI = 0)		0.093 (0.057)	0.092 (0.058)
Unempl rate for county (federal UI > 0)		-0.024 (0.028)	-0.023 (0.028)
Absolute '00s cases/mo/50,000		-0.035 (0.038)	-0.036 (0.039)
Absolute deaths/mo/50,000		-0.038 (0.025)	-0.029 (0.025)
1=Status quo alternative		-1.994*** (0.324)	-2.472*** (0.380)
Panel B	Model:	(3')	(6')
Avg. hhld cost for county (baseline)		-0.291** (0.142)	-0.310** (0.147)
× 1=Non-zero federal UI (differential)		0.465*** (0.163)	0.463*** (0.166)
Unempl rate for county (baseline)		0.093 (0.057)	0.092 (0.058)
× 1=Non-zero federal UI (differential)		-0.116** (0.055)	-0.115** (0.057)
Absolute '00s cases/mo/50,000		-0.035 (0.382)	-0.036 (0.389)
Absolute deaths/mo/50,000		-0.038 (0.025)	-0.029 (0.025)
1=Status quo alternative		-1.994*** (0.324)	-2.472*** (0.380)
Respondents		993	993
Choices		1,986	1,986

Notes: *p<0.1; **p<0.05; ***p<0.01. All models are corrected for sample selection and include relevant variables to control for the 10 categories of activities or businesses. Panel A reports estimates for key coefficients with standard errors clustered at the level of the respondent. Models 3 and 6 here are comparable to Models 3 and 6 in Table 9. The results in Panel B are equivalent to those of Panel A, except that results for cost and unemployment are reported in terms of base levels and shifters in Panel B. Panel B makes it clear that, while the two coefficients for unemployment are not significantly different from zero (in Panel A), they are significantly different from each other.

B.7 Online Appendix: Other types of choice models

B.7.1 Mixed logit models. Mixed logit models allow some marginal utility coefficients in a choice model to be distributed randomly across the population, while others are held fixed. Each random parameter must be assigned a specific type of distribution. In Table G1, Model 1 assumes independent normal distributions for each of the seven key parameters in our basic model, while Model 2 assumes a log-normal distribution for the coefficients on cases and deaths and normal distributions for the other five key parameters. Cases and deaths are transformed to be negative for Model 2 so that their positive coefficients (after exponentiation) are sensible.

For each parameter, the mixed-logit algorithm produces both an estimate of the mean of that distribution (with its standard error) and the standard deviation of that distribution (likewise with its standard error). When the standard deviation of a random parameter is statistically significantly different from zero, there is unobserved heterogeneity in that parameter across the population. Table G1 shows that a zero value is well within two standard deviations of all except the effect of average household cost when federal UI is not present (*Avg. hhld cost for county (federal UI = 0)*), though the standard deviation is only significantly different from 0 for this variable and the status quo effect. Model 2 differs from Model 1 in that the log-normal distributions for (negative) cases and deaths do not permit indirect utility to increase with additional cases and deaths. Model 2 suggests more *heterogeneity* in the marginal disutility associated with unemployment rates, household costs when federal UI is present, and cases than does Model 1.

Both models in Table G1 concur with the results in Model 6 in Table 9 for the *signs* of the means for the first four coefficients (for the terms in average household costs and unemployment).

We note that yet another alternative for a mixed logit specification employs triangular distributions for random parameters, often used to prevent inappropriate signs. We estimated the means and standard deviations of a set of triangular random parameters for our seven featured marginal disutilities. The means of the triangular random parameters all bear the expected signs, consistent with Models 3 and 6 in Table 9. All point estimates of these means are statistically significantly different from zero and there is statistically significant unobserved heterogeneity in each of these parameters as well. However, the magnitudes of all of the resulting parameters are dramatically larger than in all other specifications considered in this paper, a result we are not yet able to explain. Thus we do not pursue this specification in the current paper.

Table G1. Two mixed logit specifications

	<i>Dependent variable:</i>			
	1=Preferred policy			
	(1)		(2)	
	Coefficients all normal		Cases/deaths lognormal	
	Mean	SD	Mean	SD
Avg. hhld cost for county (federal UI = 0)	-0.604** (0.289)	0.096 (0.273)	-0.583* (0.316)	0.194 (0.266)
Avg. hhld cost for county (federal UI > 0)	0.388** (0.181)	0.230 (0.185)	0.679*** (0.262)	0.617*** (0.201)
Unempl rate for county (federal UI = 0)	0.230* (0.128)	0.122 (0.075)	0.249* (0.136)	0.150* (0.086)
Unempl rate for county (federal UI > 0)	-0.001 (0.063)	0.166*** (0.054)	-0.029 (0.067)	0.136* (0.073)
Absolute '00s cases/mo/50,000	-0.066 (0.069)	0.074 (0.078)		
Absolute deaths/mo/50,000	-0.044 (0.045)	0.065 (0.056)		
Absolute '00s cases/mo/50,000 × (-1)			-2.986** (1.454)	2.557** (1.016)
Absolute deaths/mo/50,000 × (-1)			-3.569* (1.962)	0.829 (0.960)
1=Status quo alternative	-3.250*** (0.600)	2.511*** (0.348)	-2.455*** (0.764)	1.930*** (0.616)
height				
Respondents	993	993	993	993
Choices	1,986	1,986	1,986	1,986

Coefficients on reported variables are normally distributed in Model 1. Coefficients on *Absolute '00s cases/mo/50,000 × (-1)* and *Absolute deaths/mo/50,000 × (-1)* are distributed log-normal in Model 2, while the rest of the coefficients in Model 2 are normally distributed. All other variables in Model 3 of Table 9 (i.e. continuous controls for business restrictions and response propensity interactions) are included in estimation of both models but are not permitted to vary.

B.7.2 Latent class models. Given the emphasis on latent class models in the previous literature on pandemic policy choice experiments, we have estimated some 2-class and 3-class models (4-class models would not converge), but only for specifications analogous to Model 3, where the ten types of restrictions enter as continuous variables. For Model 6, where the ten types of restrictions enter non-parametrically as sets of indicators, the specification has 29, rather than just ten, parameters associated with the restrictions on activities and businesses (implying 19 more coefficient in 2-class models, or 28 more in 3-class models). Latent class generalization of Model 6 would not converge.

It appears not to be possible to allow every parameter in our homogeneous specifications to differ across latent classes. It proved necessary to constrain the nuisance parameters on the interaction terms between the policy attributes and the fitted response propensity to be the same for both/all preference classes. In our latent class models, the separating equation for the different classes is permitted to depend upon the same set of characteristics that we employ to split the sample in our analysis of heterogeneity in the body of the paper.

The estimation algorithms are somewhat finicky and require very high-quality starting values. We resort to Stata’s `llogit2` with its EM algorithm, followed up by `llogitml2` by maximum likelihood once a simplified model can be induced to converge, because unlike the latent class algorithm in R, these latent-class estimators in Stata permit some of the preference coefficients to remain fixed while others are allowed to differ across some specified number of latent classes). We have managed to achieve convergence for latent class models with either two or three classes, but for a sample slightly larger than our final estimating sample.

Table G2 describes our 2-class model. As other researchers have tended to find, preferences over pandemic policies differ across political ideologies. For this two-class model, the only statistically significant coefficients in the class-separating equation are the indicators for “liberal” or “conservative” ideologies (relative to “moderate” ideology).

These estimates are similar in spirit to those reported in Online Appendix Table F4. Conservatives are more likely to belong to preference class 1, and liberals to class 2. Liberals are again more likely to vote for any policy, regardless of its characteristics. Conservatives pay more attention to death rates. The preferences of moderates seem to be subsumed with those of liberals in preference class 2 (reflecting their disutility from average household costs and positive utility from unemployment rates in the absence of federal UI payments). Conservatives object to restrictions on non-essential retail, parks and outdoor sports, and meetings and religious services. Liberals object to restrictions on schools and daycare, etc. Finally, for either group, people who are more likely to complete the survey tend to derive less disutility from restrictions on gyms and indoor sports, and greater disutility from restrictions on assisted living facilities.

However, Table 10 in the body of the paper reveals that there are likely to be more than just two or three classes of preferences, given the heterogeneity in preference parameters across the different splits of the sample. Latent class models with two or three classes appear to be too restrictive to reveal all of the interesting heterogeneity in preferences in these data. Latent class models with three latent preference classes (but without corrections for response propensities) do bring age, gender, and college graduation status into one or both of the class-separating equations. However, for the third class of preferences, only two of the 17 coefficients are statistically significantly different from zero, and only at the 10 percent level. Three-class models analogous to our two-class model (namely, including the interaction terms involving response propensities) have been reluctant to converge, so we do not pursue them in this paper.

Table G2. Specification with two latent preference classes, based on Model 3

<i>Dependent variable: 1=Preferred policy</i>				
<i>Class separating equation: Propensity to exhibit Class 1 preferences = f(respondent characteristics)</i>				
	coef.	(s.e.)		
Age less than 35 years	0.379	(0.252)		
Age greater than 64 years	-0.377	(0.288)		
Non-white	-0.419	(0.261)		
Remale	0.31	(0.227)		
Liberal ideology	-1.567	(0.393)***		
Conservative ideology	0.889	(0.237)***		
College graduate	-0.401	(0.267)		
Income \$75,000 or more	0.105	(0.239)		
Constant	-1.471	(0.372)***		
<i>Main preference parameters</i>				
	Class 1		Class 2	
	coef.	(s.e.)	coef.	(s.e.)
Avg. hhld cost for county (federal UI = 0)	0.00292	(0.00503)	-0.00455	(0.00188)***
Avg. hhld cost for county (federal UI > 0)	-0.00074	(0.00215)	0.00021	(0.0019)
Unempl rate for county (federal UI = 0)	-0.222	(0.178)	0.207	(0.081)**
Unempl rate for county (federal UI > 0)	-0.043	(0.058)	0.087	(0.056)
Absolute '00s cases/mo/50,000	-0.039	(0.096)	-0.045	(0.065)
Absolute deaths/mo/50,000	-0.167	(0.065)***	-0.012	(0.045)
Grocery, essential retail	0.21	(0.348)	-0.193	(0.186)
Non-essential retail	-0.6	(0.225)***	0.017	(0.159)
Schools, daycare	-0.107	(0.188)	-0.39	(0.144)***
Universities, colleges	-0.00141	(0.262)	-0.122	(0.156)
Parks, outdoor sports	-0.514	(0.227)**	-0.043	(0.169)
Gyms, indoor sports	-0.407	(0.278)	-0.077	(0.143)
Theaters, concert halls	0.003	(0.183)	-0.096	(0.129)
Restaurants, bars, clubs	-0.04	(0.219)	-0.041	(0.134)
Meetings, religious services	-0.619	(0.252)***	0.245	(0.154)
Assisted living facilities	0.256	(0.305)	0.15	(0.135)
1=Status quo alternative	-0.738	(1.269)	-2.366	(0.755)
<i>Interactions with fitted response propensities, $\hat{R}P$</i>				
	coef.	(s.e.)		
$\hat{R}P \times$ Avg. hhld cost for county (federal UI = 0)	0.00298	(0.00574)		
$\hat{R}P \times$ Avg. hhld cost for county (federal UI > 0)	0.00021	(0.00187)		
$\hat{R}P \times$ Unempl rate for county (federal UI = 0)	0.105	(0.183)		
$\hat{R}P \times$ Unempl rate for county (federal UI > 0)	0.026	(0.058)		
$\hat{R}P \times$ Absolute '00s cases/mo/50,000	-0.047	(0.079)		
$\hat{R}P \times$ Absolute deaths/mo/50,000	-0.052	(0.04)		
$\hat{R}P \times$ Grocery, essential retail	-0.128	(0.184)		
$\hat{R}P \times$ Non-essential retail	0.078	(0.164)		
$\hat{R}P \times$ Schools, daycare	-0.031	(0.156)		
$\hat{R}P \times$ Universities, colleges	-0.047	(0.141)		
$\hat{R}P \times$ Parks, outdoor sports	0.163	(0.174)		
$\hat{R}P \times$ Gyms, indoor sports	0.266	(0.155)*		
$\hat{R}P \times$ Theaters, concert halls	0.023	(0.135)		
$\hat{R}P \times$ Restaurants, bars, clubs	-0.131	(0.133)		
$\hat{R}P \times$ Meetings, religious services	-0.006	(0.152)		
$\hat{R}P \times$ Assisted living facilities	-0.363	(0.138)***		
1=Status quo alternative	0.442	(0.57)		
Log likelihood	-825.60747			

Notes: *p<0.1; **p<0.05; ***p<0.01. Even by starting with a minimal specification, it was extremely difficult to nurse these models towards convergence due to the large number of preference parameters to be estimated. These estimates employ a sample that is slightly larger than the 993 respondents used for our main results because we had not yet excluded 34 respondents who rejected pandemic policy(ies) because they rejected the choice scenario, rather than for reasons related to preferences or economic considerations. For analogous models employing only the 993 respondents, it was even more difficult to achieve convergence. Given that our models with heterogeneous preferences reveal that there are many different types of preferences, we gave up on latent-class models (despite their popularity in other studies of pandemic-policy preferences).

B.7.3 Sensitivity analysis: Preferences as a function of time spent on first-choice preamble about federal UI assumption. Some respondents read survey screens at roughly the average reading speed in the population. Others spend less time than this on each page, reinforcing the importance of providing them with succinct information. When respondents are offered several choices in a similar format, it can be expected that they will develop choice heuristics that will allow them to process choice tasks more quickly. In our survey, the tutorial section used that individual’s (unique) first choice set for training purposes, highlighting and explaining each section of the choice task summary table in turn, so respondents will have had an opportunity, in advance of their first choice, to think about which policy attributes matter the most to them and where in the choice table they will be able to see those attributes. Our use of “meters” for the severity of restrictions on activities and businesses allows respondents to see these restrictions without having to read any numbers. The pop-up descriptions explain how to interpret each meter reading in the context of that particular type of restriction.

It is generally prudent to determine the extent to which there may be systematically different apparent preferences between (a) people who read survey pages quickly and (b) people who read more deliberately. In this study, we are particularly concerned about whether respondents took the time to review the preamble to the first choice set, where the the assumptions to be made concerning federal UI benefits were described. These benefits were also described as a fixed characteristic of the policy choice context, invariant to the respondent’s choice of whether or not to “vote” for a given policy.

Table G3 shows that while there is some movement in the point estimates for our four coefficients on the average household cost and unemployment variables and their interactions with federal UI payments, all of these coefficients retain their signs and significance levels when they are accompanied by their interactions with each of two measures of the respondent’s time on that page. None of the four interaction terms is individually statistically significant, and neither are the four extra terms jointly statistically significant from zero.

Table G3. Reading time

	<i>Dependent variable:</i>		
	1=Preferred policy		
	(1)	(2)	(3)
Avg. hhld cost for county (federal UI = 0)	-0.310*** (0.121)	-0.292** (0.164)	-0.336** (0.171)
× Time on page		-0.0002 (0.006)	
× Time on page (censored)			0.002 (0.006)
Avg. hhld cost for county (federal UI > 0)	0.153** (0.068)	0.137* (0.081)	0.137* (0.086)
× Time on page		0.001 (0.002)	
× Time on page (censored)			0.001 (0.003)
Unempl rate for county (federal UI = 0)	0.092** (0.052)	0.085 (0.068)	0.096* (0.071)
× Time on page		0.0004 (0.002)	
× Time on page (censored)			-0.0001 (0.003)
Unempl rate for county (federal UI > 0)	-0.023 (0.025)	-0.026 (0.027)	-0.033 (0.029)
× Time on page		0.0001 (0.001)	
× Time on page (censored)			0.0005 (0.001)
Absolute '00s cases/mo/50,000	-0.036 (0.033)	-0.029 (0.033)	-0.026 (0.033)
Absolute deaths/mo/50,000	-0.029 (0.022)	-0.030 (0.023)	-0.030 (0.023)
1=Status quo alternative	-2.472*** (0.399)	-2.509*** (0.402)	-2.501*** (0.401)
Respondents	993	993	993
Choices	1986	1986	1986
Log Likelihood	-1,169.803	-1,165.983	-1,165.560

Notes: *p<0.1; **p<0.05; ***p<0.01. Column 1 reproduces our preferred specification (Model 6 in Table 3) for easy comparison. Column 2 includes interactions of our four variables of interest with the amount of time (in seconds) respondents spent on the survey page that instructed them to “**Assume that any Federal unemployment benefits, as described, will be in place regardless of any pandemic rules that apply in [respondent’s county].**” A few respondents spent long periods of time (e.g. more than 10 minutes) on that page. To ensure that these outliers are not driving our results, we also interact each variable of interest with a right-censored measure of time. This measure replaces all values above 92 seconds (i.e. double the time it would take a typical reader to read the page) with 92.

B.7.4 Sensitivity analysis: Preferences as a function of highest pandemic-month unemployment rate in respondent’s county relative to sample median. The extent to which a respondent responds to a policy that changes the unemployment rate may depend on unemployment rates in that respondent’s own county relative to the unemployment rates experienced by other respondents. For each respondent, we calculate the highest monthly county-level unemployment rate experienced over March 2020 through December 2020. To split the sample into two roughly equal groups, we calculate the median of these highest unemployment rates across all respondents. We then split the sample according to whether the respondent’s highest pandemic-era monthly county-level unemployment rate is higher or lower than the sample median.

The estimates in Table G4 are all from the same model, as in our models for heterogeneous preferences across partitions of the sample by sociodemographic groups. Column 1 shows our seven featured parameter estimates for respondents who have experienced lower-than-median “worst” pandemic-era unemployment rates. Column 2 shows the corresponding parameter estimates for respondents who have experienced higher-than-median “worst” pandemic-era unemployment rates. For both groups, the first four coefficients retain the same signs that they exhibit in our other specifications.

Without federal UI, the disutility from average household costs is about three times greater for the group that has experienced higher-than-median “worst” unemployment levels. Decreasing marginal utilities of income may explain some of this disparity. Additionally, recent experience with high local unemployment may make people more cost averse. With federal UI, the marginal utility from average household costs is positive (as usual), but is statistically significant only for the group with lower-than-median “worst” unemployment levels during the pandemic.

Without federal UI, higher unemployment rates confer statistically significant positive utility only for the group with higher-than-median “worst” unemployment levels during the pandemic. Counties that have already experienced very high unemployment are likely to have higher shares of workers in sectors that cannot transition to remote work and require in-person interaction, so residents of those counties may perceive a larger public health benefit associated with unemployment. With federal UI, both groups dislike higher unemployment rates, but neither point estimate is statistically significantly different from zero.

The disutility from greater numbers of deaths is marginally statistically significant only for the group with higher-than-median “worst” unemployment rates during the pandemic. The worst unemployment rates are likely experienced in regions where the pandemic has been especially severe, which may increase the salience of pandemic deaths.

Table G4. Worst month's unemployment rate relative to median worst pandemic unemployment rate prior to survey

	<i>Dependent variable:</i>	
	1=Preferred policy	
	Better Unemp	Worse Unemp
	(1)	(2)
Avg. hhld cost for county (federal UI = 0)	-0.205*	-0.641**
	(0.109)	(0.253)
Avg. hhld cost for county (federal UI > 0)	0.190**	0.122
	(0.092)	(0.120)
Unempl rate for county (federal UI = 0)	0.035	0.212**
	(0.049)	(0.086)
Unempl rate for county (federal UI > 0)	-0.038	-0.031
	(0.033)	(0.029)
Absolute '00s cases/mo/50,000	-0.040	-0.063
	(0.042)	(0.056)
Absolute deaths/mo/50,000	0.004	-0.052*
	(0.031)	(0.029)
1=Status quo alternative	-2.258***	-3.000***
	(0.500)	(0.582)
Respondents	503	490
Choices	1006	980

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are corrected for sample selection and include relevant variables to control for the 10 categories of activities or businesses.

B.7.5 Sensitivity analysis: Preferences as a function of highest pandemic-month unemployment rate in respondent’s county relative to sample median. Our survey employs occasional comprehension questions to assess whether the respondent is paying attention to the tutorial portion of the survey. One such question was: “We need to be sure that everyone interprets their choice tasks the same way. Thus there are several points at which we ‘check your understanding.’ For example, can we be sure that exactly X people will get sick and Y will die, if no rules are in place?” In the tutorial, the wildcard amounts X and Y are set to their values in the respondent’s first choice task.²

A second comprehension question was: “For your Policy A, the ‘Average \$/month lost’ across all households will be \$X. Does that mean your household, and every other household in [your county], will end up losing \$X of income each month during the policy? In addition to any unemployment?”³

For this sensitivity analysis, we split the sample according to whether the respondent was one of the 370 respondents who answered both of these questions correctly, or whether they were one of the 623 respondents who answered at least one question incorrectly. Note that the subsequent screen in the survey either confirms the respondent’s correct answer, or goes into more detail to explain the correct answer if the respondent’s answer is either incorrect or “Don’t know/not sure.” Thus an incorrect answer does not imply that the respondent makes their policy choices based on incorrect information, only that they needed more help to fully understand important points that were being made in the tutorial portion of the survey.

Again, the first four coefficients in both columns of Table G5 display the familiar pattern in their signs. Respondents who were more attentive to the information in the tutorial have less precise estimates for the first three coefficients, but with federal UI, their marginal disutility from unemployment is strongly statistically significantly negative, whereas this disutility for respondents who made at least one mistake in these two comprehension questions shows no discernible response to unemployment rates when federal UI is non-zero. Respondents who are more engaged with the survey may be more concerned about county-level pandemic policies in general, including their interactions with federal policies.

²The correct answer here is *no*. We instructed respondents that X and Y are to be treated as the best available estimates of cases and deaths.

³Again, the correct answer is *no*. We instructed respondents that X is an average over households that will lose much more than X in income and households that lose less—or no— income.

Table G5. Heterogeneity according to correctness of answers on two comprehension questions during the survey's tutorial

	<i>Dependent variable:</i>	
	1=Preferred policy Both answers right (1)	At least one answer wrong (2)
Avg. hhld for county (federal UI = 0)	-0.447* (0.258)	-0.333*** (0.109)
Avg. hhld for county (federal UI > 0)	0.123 (0.133)	0.195** (0.077)
Unempl rate for county (federal UI = 0)	0.066 (0.102)	0.137*** (0.047)
Unempl rate for county (federal UI > 0)	-0.095** (0.044)	-0.005 (0.025)
Absolute '00s cases/mo/50,000	-0.105 (0.070)	-0.011 (0.039)
Absolute deaths/mo/50,000	-0.019 (0.039)	-0.033 (0.025)
1=Status quo alternative	-3.205*** (0.811)	-2.284*** (0.426)
Respondents	370	623
Choices	740	1246

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are corrected for sample selection and include relevant variables to control for the 10 categories of activities or businesses.

B.7.6 Sensitivity analysis: Preferences as a function of household income relative median household income in the respondents ZIP code. It is possible that preferences for pandemic policy attribute may vary systematically according to whether the respondent’s household income is higher or lower than the median household income in their ZIP code. We merge ZIP code level sociodemographic data, including median household income, with our survey response data. In Table G6, column 1 gives our seven featured parameter estimates for respondents with household incomes below their ZIP code median (“Poorer”) and column 2 gives the corresponding estimates for respondents with household incomes above their ZIP code median (“Richer”).

This model reveals that most of the action on the average household cost and unemployment rate variables is driven by the preferences of relatively poorer household. These estimates are complementary to those in Table 10 where we separate respondents according to the absolute level of their household income, using \$75,000/year as the dividing line. We estimate the model in Table G6 because median household incomes differ substantially across ZIP codes. It is sometimes argued that “relative income” compared to one’s neighbors may be a stronger determinant of preferences than the absolute level of income, it seemed prudent to check the effects of relative household income within ZIP codes.

The main difference between the estimates in Table G6 and those in Table 10 in the body of the paper, is that in Table G6, the marginal disutility from average household costs in the absence of federal UI is marginally statistically significant, whereas in Table 10, none but the coefficient on the indicator for the status quo alternative is statistically significantly different from zero for the group of households distinguished by having an absolute income greater than \$75,000.

Table G6. Zip code relative income. Household income below median zip code income (Poorer) versus above median zip code income (Richer)

	<i>Dependent variable:</i>	
	1=Preferred policy Poorer	Richer
	(1)	(2)
Avg. hhld for county (federal UI = 0)	-0.514*** (0.162)	-0.236* (0.138)
Avg. hhld for county (federal UI > 0)	0.319*** (0.105)	0.026 (0.084)
Unempl rate for county (federal UI = 0)	0.268*** (0.077)	0.009 (0.052)
Unempl rate for county (federal UI > 0)	0.010 (0.039)	-0.041 (0.025)
Absolute '00s cases/mo/50,000	-0.094* (0.057)	-0.016 (0.045)
Absolute deaths/mo/50,000	0.006 (0.030)	-0.048* (0.028)
1=Status quo alternative	-2.830*** (0.577)	-2.444*** (0.542)
Respondents	440	524
Choices	880	1048

Notes: *p<0.1; **p<0.05; ***p<0.01. Models are corrected for sample selection and include relevant variables to control for the 10 categories of activities or businesses.

APPENDIX C
CHAPTER 4 APPENDIX

C.1 Online Appendix: Treatment of sample selection in choice models in the literature on choice modeling

C.1.0.1 Abstract claims a “representative sample”. Some papers have abstracts that mention representative samples but the authors acknowledge in the body of the paper that the sample may not actually be representative. Hauber, Gonzalez, Schenkel, Loffland, and Martin (2011) mention that the relevant population characteristics are unknown. Sandorf, Aanesen, and Navrud (2016) have concerns about possible sample selection by topic salience in a workshop-related survey that complements their internet survey, but they rely on quota sampling in their internet survey to yield observable sociodemographics that represent the national population.

Three papers claim a representative sample in their abstracts, but do not seem to acknowledge, specifically, that their samples are only representative in terms of a handful of observable demographics. These papers include Kilambi, Johnson, Gonzalez, and Mohamed (2014), Cantillo, Martin, and Roman (2021), and Cao, Cranfield, Chen, and Widowski (2021).

Some authors do elaborate, in the body of the paper, on the type of representativeness they are claiming for their samples, and these authors either assert or assess their sample’s representativeness in terms of specific demographics. In this category, Shigeoka and Yamada (2019) also use weights to improve the representativeness of their estimates.

Some researchers do go so far as to acknowledge, explicitly, that there is some potential for their samples to be subject to selection on unobservables. Grimsrud, Nielsen, Navrud, and Olesen (2013) must deal with a response rate of 33%, and they do mention representativeness among their non-response cases. Carattini, Baranzini, Thalmann, Varone, and Vohringer (2017) use a market research firm with aggressive non-response conversion efforts. While theirs is not an internet panel survey, they argue that the resulting response rate of 75% mitigates potential self-selection issues. Craig and Rand (2018) find that people with incomplete surveys differ from those who completed the surveys, and they compare the marginal distributions for basic sociodemographics in their sample and the U.S. population, but they do not pursue any corrections.

Baji et al. (2020) use quota sampling with an internet panel but note that the use of an internet panel means that study samples might be somewhat selective, compared to other samples. Urrutia et al. (2021) use quota samples and acknowledge the possibility of selection bias in online sampling, such as the likelihood that people with long-term health conditions may be less able or willing to participate in online surveys. However, they argue that quota-based sampling will "ensure a closer representativeness of the sample to the target population."

Numerous papers, however, do not seem to acknowledge possible systematic selection in samples of completed responses from an internet panel, especially when there is a lower response rate. For the most part, researchers using internet panels seem less likely to report response rates, perhaps because the internet survey firm contracts to deliver a specific number of "completes" and does not share information about the universe of people from which they draw their samples or the response rate to their invitations.

Carson, Louviere, and Wei (2010) work with choice data from a survey having a response rate of 76.5%. Botzen, de Boer, and Terpstra (2013) (internet?), have a 69% response rate. Farsky, Schnittka, Sattler, Hofer, and Lorth (2017) do not seem to report a response rate.

Hoek, Pearson, James, Lawrence, and Friel (2017) make an effort to compare observables for sample and population; Rogers and Burton (2017) note that their sample is stratified and apparently compare the sample and population statistics, but the appendix containing this table seems to be inaccessible. Rahmani and Loureiro (2019) (internet?) report marginal distributions for the sample only. Plum, Olschewski, Jobin, and van Vliet (2019) use quota sampling but achieve a response rate of only 19.7%, which would seem to point to a need to assess reasons for non-response.

Kormos, Axsen, Long, and Goldberg (2019) have a narrower target population, limited to people who intend to purchase or lease a new vehicle within 12 months, and they give marginal distributions for their sample and the general population (what comment about these?), but do not mention their response rate. Armstrong, Aanesen, van Rensburg, and Sandorf (2019) use stratified random samples and different scale factors to accommodate different amounts of noise in their different samples but make no mention of response rates. Goranitis, Best, Christodoulou, Stark, and Boughtwood (2020) use quota sampling, compare their samples to Census data, and report a "cooperation rate" of 73%. Huber, Wicki, and Bernauer (2020) sample by three "interlocked quotas" by gender, age, and region (by which they may imply the joint distribution of these observable characteristics in both the sample and the population), plus quotas by urban/rural areas and degree of agglomeration.

The online survey used by Long, Kormos, Sussman, and Axsen (2021) is representative by gender, age, education, income, race and region, but they acknowledge that demographic information on the target population (intended vehicle purchasers) is not available. Nevertheless, they still compare their study sample to the Census. Finally, Manipis, Street, Cronin, Viney, and Goodall (2021) compare their sample demographics to population demographics, but they report an overall usable response rate of only 52.7% out of all the people who started the survey. They do not explore reasons why people did not complete the survey.

C.1.0.2 Abstract mentions representativeness in terms of specific observable variables. Rowen et al. (2016) report marginal relative

frequencies for a set of observable sociodemographic characteristics for their sample and the population, but do not mention their response rate or the issue of potential sample selection. Comans, Nguyen, Ratcliffe, Rowen, and Mulhern (2020) acknowledge that their sample is different from the population on a couple of variables, even though they use quota sampling. They mention that "Nonresponse or invalid survey were very low in this sample, increasing the confidence in the representativeness," but they make no mention of how they calculate their response rates. Kantor (2021) uses an internet-based sample that is stratified on age, sex and race and is claimed to be representative of the U.S. population, and also reports a completion rate of 97% for this survey!

C.1.0.3 Abstract mentions "representative household". The goal of many choice experiments is to estimate willingness to pay for specified alternatives. Some papers refer in their abstracts to WTP for a representative household. However, they go on to mention in the body of the paper that the sample is different from the population on some demographic dimensions. Despite these differences, they do not specifically mention the potential for systematic selection into the estimating sample. Rosston, Savage, and Waldman (2010), for example, use Knowledge Networks and do mention that cases are dropped for incomplete data, although they do not compare completed responses with incomplete responses. Bosworth, Cameron, and DeShazo (2010) also use Knowledge Networks and claim a representative sample. They do acknowledge the sample selection issue, but do not undertake corrections in this paper.

C.1.0.4 Abstract does not mention "representativeness". A handful of papers that do not mention representativeness in their abstracts nevertheless model the selection process and undertake at least a crude correction for sample selection, similar to the ad hoc method explained in this paper. Cameron, DeShazo, and Johnson (2010) and ? allow estimated preference parameters to vary systematically with fitted response propensities. Cameron, DeShazo, and Stiffler (2010) undertake selection corrections for one of the two samples in this study, but do not have sufficient data for sample selection corrections for the other sample.

Other authors also make no claim for representativeness in their abstracts but nevertheless acknowledge the potential for nonresponse bias or other selection issues. Mulhern et al. (2014) use quota sampling to produce three internet samples. Of those respondents who passed the quota screening, the three samples had response rates of 60%, 33.6% and 40.8%. For two of these samples, they do compare the characteristics of responders and non-responders.

A final category is one where representativeness is not advertised in the paper's abstract, but the body of the paper claims a nationally representative sample. Bosworth et al. (2009) do not, however, go into any detail about sample representativeness.

C.2 Do we need to bother scaling the coefficients in the mixed-logit policy-choice equation?

When the necessary IMR selection-correction term is included among the regressors for the policy-choice submodel, and the coefficients of both submodels are scaled by the standard deviation of their approximately normal compound errors, the coefficient on the IMR term is an estimate of ρ (given that the error standard deviation for the scaled policy-choice submodel is one). Simulating the systematic portion of the policy-choice submodel in the absence of selection is equivalent to setting ρ to zero. The corrected coefficients in the policy-choice equation are thus merely the coefficients on the original Δx_i variables in the utility-difference equations for the policy alternatives, netting out the effect of the IMR term. The IMR term can be viewed as a source of omitted-variables bias in the naive model that omits this correction term.

In the case of a selection and outcome equation with correlated *standard* normal error terms, there is a risk that the empirical estimate of the coefficient on the IMR term will lie outside the $(-1, +1)$ admissible range for a correlation. In FIML estimation, this result could be precluded by instead estimating an unconstrained inverse hyperbolic tangent transformation of the ρ parameter, $\rho^* = 0.5 * \ln[(1 + \rho)/(1 - \rho)]$. If we wish to rely upon packaged mixed logit estimation algorithms, however, the coefficient on the IMR term will not approximate ρ until *after* all the coefficients of the policy-choice equation are normalized on the approximate error standard deviation of the mixed-logit policy-choice submodel, $\sigma_{\epsilon^{\beta+\eta}}$. Thus it appears to be difficult to impose this constraint on the underlying ρ parameter. Note that this parameter is not simply the estimated correlation between the random intercept coefficient, ϵ_{1i}^{α} , in the selection submodel and the random any-policy effect, (ϵ_{1i}^{β}) , in the policy-choice model. Instead, It is the correlation between $(\epsilon_{1i}^{\alpha} + \omega_i)$ and $(\epsilon_{1i}^{\beta} + \eta_i)$, where these two compound error terms are only approximately joint normal. Their correlation, however, should be *less* than the estimated correlation between ϵ_{1i}^{α} and ϵ_{1i}^{β} .

There seems to be no reason to bother scaling the coefficients on the policy attributes in the policy-choice submodel, since standard normal errors are not required in the estimation of the coefficients on the policy attributes in these “outcome” equations. Utility functions, fortunately, are invariant to scale, and the quantities we are interested in measuring via the policy-choice submodel, such as marginal willingnesses to pay for specific policy attributes, are all calculated from *ratios* of the estimated coefficients in the policy-choice submodel. If the $1(\text{Any policy}^j)$ indicator is the only indirect utility-function variable modeled as having a random coefficient, the error standard deviation $\sigma_{(\epsilon_{1i}^{\beta} + \omega_i)}$ is merely a scalar that can be ignored in the calculations of coefficient ratios. Let β_1 be the marginal utility of net income, and let β_k be the marginal utility of attribute x_k in the policy-choice model. Then the marginal willingness to pay for one unit of attribute x_k will be given by the ratio of these two parameters under any scalar

normalization:

$$\frac{\left(\frac{\beta_k/\kappa_{\omega\eta}}{\sigma_{(\epsilon_{1i}^\beta+\omega_i)}}\right)}{\left(\frac{\beta_1/\kappa_{\omega\eta}}{\sigma_{(\epsilon_{1i}^\beta+\omega_i)}}\right)} = \frac{\beta_k/\kappa_{\omega\eta}}{\beta_1/\kappa_{\omega\eta}} = \frac{\beta_k}{\beta_1} \quad (\text{C.1})$$

REFERENCES CITED

- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8(2), 155–194.
- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic [Journal Article]. *Journal of Public Economics*, 191, 11. Retrieved from <GotoISI>://WOS:000579860600008 doi: 10.1016/j.jpubeco.2020.104254
- Armstrong, C. W., Aanesen, M., van Rensburg, T. M., & Sandorf, E. D. (2019). Willingness to pay to protect cold water corals [Journal Article]. *Conservation Biology*, 33(6), 1329-1337. Retrieved from <GotoISI>://WOS:000481505600001 doi: 10.1111/cobi.13380
- Arndt, C., Jones, S., & Tarp, F. (2010). Aid, growth, and development: have we come full circle? *Journal of Globalization and Development*, 1(2).
- Baji, P., Farkas, M., Golicki, D., Rupel, V. P., Hoefman, R., Brouwer, W. B. F., ... Pentek, M. (2020). Development of population tariffs for the carerqol instrument for hungary, poland and slovenia: A discrete choice experiment study to measure the burden of informal caregiving [Journal Article]. *Pharmacoeconomics*, 38(6), 633-643. Retrieved from <GotoISI>://WOS:000521143600002 doi: 10.1007/s40273-020-00899-2
- Bartik, A. W., Bertrand, M., Lin, F., Rothstein, J., & Unrath, M. (2020). Measuring the labor market at the onset of the covid-19 crisis [Journal Article]. *National Bureau of Economic Research Working Paper Series, No. 27613*. Retrieved from <http://www.nber.org/papers/w27613> doi: 10.3386/w27613
- Blayac, T., Dubois, D., Duchene, S., Nguyen-Van, P., Ventelou, B., & Willinger, M. (2021). Population preferences for inclusive covid-19 policy responses [Journal Article]. *Lancet Public Health*, 6(1), E9-E9. Retrieved from <GotoISI>://WOS:000608270600006 doi: 10.1016/s2468-2667(20)30285-1
- Bosworth, R., Cameron, T. A., & DeShazo, J. R. (2009). Demand for environmental policies to improve health: Evaluating community-level policy scenarios [Journal Article]. *Journal of Environmental Economics and Management*, 57(3), 293-308. Retrieved from <GotoISI>://WOS:000266115100005 doi: 10.1016/j.jeem.2008.07.009

- Bosworth, R., Cameron, T. A., & DeShazo, J. R. (2010). Is an ounce of prevention worth a pound of cure? comparing demand for public prevention and treatment policies [Journal Article]. *Medical Decision Making*, 30(4), E40-E56. Retrieved from <GotoISI>://WOS:000279934800016 doi: 10.1177/0272989x10371681
- Bosworth, R., Cameron, T. A., & DeShazo, J. R. (2015). Willingness to pay for public health policies to treat illnesses [Journal Article]. *Journal of Health Economics*, 39, 74-88. Retrieved from <GotoISI>://WOS:000349266800006 doi: 10.1016/j.jhealeco.2014.10.004
- Botzen, W. J. W., de Boer, J., & Terpstra, T. (2013). Framing of risk and preferences for annual and multi-year flood insurance [Journal Article]. *Journal of Economic Psychology*, 39, 357-375. Retrieved from <GotoISI>://WOS:000328720500029 doi: 10.1016/j.joep.2013.05.007
- Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American economic review*, 90(4), 847-868.
- Cameron, T. A., DeShazo, J. R., & Johnson, E. H. (2010). The effect of children on adult demands for health-risk reductions [Journal Article]. *Journal of Health Economics*, 29(3), 364-376. Retrieved from <GotoISI>://WOS:000278297200004 doi: 10.1016/j.jhealeco.2010.02.005
- Cameron, T. A., DeShazo, J. R., & Johnson, E. H. (2011). Scenario adjustment in stated preference research [Journal Article]. *Journal of Choice Modelling*, 4(1), 9-43. Retrieved from <GotoISI>://WOS:000414832500002
- Cameron, T. A., DeShazo, J. R., & Stiffler, P. (2010). Demand for health risk reductions: A cross-national comparison between the u.s. and canada [Journal Article]. *Journal of Risk and Uncertainty*, 41(3), 245-273. Retrieved from <GotoISI>://WOS:000284238200004 doi: 10.1007/s11166-010-9106-9
- Cantillo, J., Martin, J. C., & Roman, C. (2021). Determinants of fishery and aquaculture products consumption at home in the eu28 [Journal Article]. *Food Quality and Preference*, 88, 14. Retrieved from <GotoISI>://WOS:000594542500006 doi: 10.1016/j.foodqual.2020.104085
- Cao, Y., Cranfield, J., Chen, C., & Widowski, T. (2021). Heterogeneous informational and attitudinal impacts on consumer preferences for eggs from welfare enhanced cage systems [Journal Article]. *Food Policy*, 99, 11. Retrieved from <GotoISI>://WOS:000633368500005 doi: 10.1016/j.foodpol.2020.101979

- Carattini, S., Baranzini, A., Thalmann, P., Varone, F., & Vohringer, F. (2017). Green taxes in a post-paris world: Are millions of nays inevitable? [Journal Article]. *Environmental Resource Economics*, 68(1), 97-128. Retrieved from <GotoISI>://WOS:000410761400005 doi: 10.1007/s10640-017-0133-8
- Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions [Journal Article]. *Environmental Resource Economics*, 37(1), 181-210. Retrieved from <GotoISI>://WOS:000246592000011 doi: 10.1007/s10640-007-9124-5
- Carson, R. T., Louviere, J. J., & Wei, E. (2010). Alternative australian climate change plans: The public's views [Journal Article]. *Energy Policy*, 38(2), 902-911. Retrieved from <GotoISI>://WOS:000273985700025 doi: 10.1016/j.enpol.2009.10.041
- Cassen, R., et al. (1994). Does aid work?: report to an intergovernmental task force. *OUP Catalogue*.
- Cerda, A. A., & Garcia, L. Y. (2021). Willingness to pay for a covid-19 vaccine [Journal Article]. *Applied Health Economics and Health Policy*, 9. Retrieved from <GotoISI>://WOS:000620413600001 doi: 10.1007/s40258-021-00644-6
- Chauvet, L., Collier, P., & Fuster, A. (2017). Supervision and project performance: A principal-agent approach.
- Chorus, C., Sandorf, E. D., & Mouter, N. (2020). Diabolical dilemmas of covid-19: An empirical study into dutch society's trade-offs between health impacts and other effects of the lockdown [Journal Article]. *Plos One*, 15(9), 19. Retrieved from <GotoISI>://WOS:000573658000036 doi: 10.1371/journal.pone.0238683
- Comans, T. A., Nguyen, K. H., Ratcliffe, J., Rowen, D., & Mulhern, B. (2020). Valuing the ad-5d dementia utility instrument: An estimation of a general population tariff [Journal Article]. *Pharmacoeconomics*, 38(8), 871-881. Retrieved from <GotoISI>://WOS:000527522300001 doi: 10.1007/s40273-020-00913-7
- Concha, J. (2020, Mar). *Hannity challenges mnuchin on coronavirus bill: Expanding unemployment benefits "angers my audience"*. Retrieved from <https://thehill.com/homenews/media/489665-hannity-challenges-mnuchin-on-coronavirus-bill-expanding-unemployment-benefits>

- Cook, A. R., Zhao, X. H., Chen, M. I. C., & Finkelstein, E. A. (2018). Public preferences for interventions to prevent emerging infectious disease threats: a discrete choice experiment [Journal Article]. *Bmj Open*, 8(2), 12. Retrieved from <GotoISI>://WOS:000433129800084 doi: 10.1136/bmjopen-2017-017355
- Craig, B. M., & Rand, K. (2018). Choice defines qalys a us valuation of the eq-5d-5l [Journal Article]. *Medical Care*, 56(6), 529-536. Retrieved from <GotoISI>://WOS:000440798800012 doi: 10.1097/mlr.0000000000000912
- Denizer, C., Kaufmann, D., & Kraay, A. (2013). Good countries or good projects? macro and micro correlates of world bank project performance. *Journal of Development Economics*, 105, 288–302.
- Diallo, A., & Thuillier, D. (2005). The success of international development projects, trust and communication: an african perspective. *International journal of project management*, 23(3), 237–252.
- Dollar, D., & Levin, V. (2005). *Sowing and reaping: institutional quality and project outcomes in developing countries*. The World Bank.
- Dollar, D., & Svensson, J. (2000). What explains the success or failure of structural adjustment programmes? *The Economic Journal*, 110(466), 894–917.
- Doucouliaqos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys*, 23(3), 433–461.
- Doucouliaqos, H., & Paldam, M. (2011). The ineffectiveness of development aid on growth: An update. *European journal of political economy*, 27(2), 399–404.
- Dube, A. (2021). Aggregate employment effects of unemployment benefits during deep downturns: Evidence from the expiration of the federal pandemic unemployment compensation [Journal Article]. *National Bureau of Economic Research Working Paper Series, No. 28470*. Retrieved from <http://www.nber.org/papers/w28470> doi: 10.3386/w28470
- Duponchel, M., Chauvet, L., & Collier, P. (2010). *What explains aid project success in post-conflict situations?* The World Bank.
- Easterly, W. (2003). Can foreign aid buy growth? *Journal of economic Perspectives*, 17(3), 23–48.

- Farsky, M., Schnittka, O., Sattler, H., Hofer, B., & Lorth, C. (2017). Brand-anchored discrete choice experiment (bdce) vs. direct attribute rating (dar): An empirical comparison of predictive validity [Journal Article]. *Marketing Letters*, *28*(2), 231-240. Retrieved from <GotoISI>://WOS:000400767300005 doi: 10.1007/s11002-016-9402-5
- Geli, P., Kraay, A., & Nobakht, H. (2014). *Predicting world bank project outcome ratings*. The World Bank.
- Genie, M. G., Loria-Rebolledo, L. E., Paranjothy, S., Powell, D., Ryan, M., Sakowsky, R. A., & Watson, V. (2020). Understanding public preferences and trade-offs for government responses during a pandemic: a protocol for a discrete choice experiment in the uk [Journal Article]. *Bmj Open*, *10*(11), 9. Retrieved from <GotoISI>://WOS:000595876500015 doi: 10.1136/bmjopen-2020-043477
- Glicksohn, J., & Golan, H. (2001). Personality, cognitive style and assortative mating. *Personality and Individual Differences*, *30*(7), 1199–1209.
- Goranitis, I., Best, S., Christodoulou, J., Stark, Z., & Boughtwood, T. (2020). The personal utility and uptake of genomic sequencing in pediatric and adult conditions: eliciting societal preferences with three discrete choice experiments [Journal Article]. *Genetics in Medicine*, *22*(8), 1311-1319. Retrieved from <GotoISI>://WOS:000530613100003 doi: 10.1038/s41436-020-0809-2
- Grange, Z. L., Goldstein, T., Johnson, C. K., Anthony, S., Gilardi, K., Daszak, P., ... others (2021). Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proceedings of the National Academy of Sciences*, *118*(15).
- Grimsrud, K. M., Nielsen, H. M., Navrud, S., & Olesen, I. (2013). Households' willingness-to-pay for improved fish welfare in breeding programs for farmed atlantic salmon [Journal Article]. *Aquaculture*, *372*, 19-27. Retrieved from <GotoISI>://WOS:000312390700003 doi: 10.1016/j.aquaculture.2012.10.009
- Guillaumont, P., & Laajaj, R. (2006). *When instability increases the effectiveness of aid projects*. The World Bank.
- Hauber, A. B., Gonzalez, J. M., Schenkel, B., Lofland, J. H., & Martin, S. (2011). The value to patients of reducing lesion severity in plaque psoriasis [Journal Article]. *Journal of Dermatological Treatment*, *22*(5), 266-275. Retrieved from <GotoISI>://WOS:000295976100004 doi: 10.3109/09546634.2011.588193

- Heckman, J. J. (1979). Sample selection bias as a specification error [Journal Article]. *Econometrica*, 47(1), 153-161. Retrieved from <GotoISI>://WOS:A1979GH66400010 doi: 10.2307/1912352
- Hess, S., & Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application [Journal Article]. *Journal of Choice Modelling*, 32.
- Hoek, A. C., Pearson, D., James, S. W., Lawrence, M. A., & Friel, S. (2017). Healthy and environmentally sustainable food choices: Consumer responses to point-of-purchase actions [Journal Article]. *Food Quality and Preference*, 58, 94-106. Retrieved from <GotoISI>://WOS:000395845500011 doi: 10.1016/j.foodqual.2016.12.008
- Huber, R. A., Wicki, M. L., & Bernauer, T. (2020). Public support for environmental policy depends on beliefs concerning effectiveness, intrusiveness, and fairness [Journal Article]. *Environmental Politics*, 29(4), 649-673. Retrieved from <GotoISI>://WOS:000475124700001 doi: 10.1080/09644016.2019.1629171
- Ika, L. A., Diallo, A., & Thuillier, D. (2012). Critical success factors for world bank projects: An empirical investigation. *International journal of project management*, 30(1), 105–116.
- Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., ... Vossler, C. A. (2017). Contemporary guidance for stated preference studies [Journal Article]. *Journal of the Association of Environmental and Resource Economists*, 4(2), 319-405. Retrieved from <GotoISI>://WOS:000398822900001 doi: 10.1086/691697
- Kahane, L. H. (2021). Politicizing the mask: Political, economic and demographic factors affecting mask wearing behavior in the usa [Journal Article]. *Eastern Economic Journal*, 21. Retrieved from <GotoISI>://WOS:000605093900001 doi: 10.1057/s41302-020-00186-0
- Kalmijn, M. (1994). Assortative mating by cultural and economic occupational status. *American journal of Sociology*, 100(2), 422–452.
- Kantor, J. (2021). Willingness to pay for surgical treatments for basal cell carcinoma: A population-based cross-sectional study [Journal Article]. *Dermatologic Surgery*, 47(4), 467-472. Retrieved from <GotoISI>://WOS:000658825800013 doi: 10.1097/dss.0000000000002874
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2), 220–246.

- Kilambi, V., Johnson, F. R., Gonzalez, J. M., & Mohamed, A. F. (2014). Valuations of genetic test information for treatable conditions: The case of colorectal cancer screening [Journal Article]. *Value in Health*, 17(8), 838-845. Retrieved from <GotoISI>://WOS:000346918100010 doi: 10.1016/j.jval.2014.09.001
- Kilby, C. (2000). Supervision and performance: the case of world bank projects. *Journal of Development Economics*, 62(1), 233-259.
- Kormos, C., Axsen, J., Long, Z., & Goldberg, S. (2019). Latent demand for zero-emissions vehicles in canada (part 2): Insights from a stated choice experiment [Journal Article]. *Transportation Research Part D-Transport and Environment*, 67, 685-702. Retrieved from <GotoISI>://WOS:000464890900048 doi: 10.1016/j.trd.2018.10.010
- Kreps, S., Prasad, S., Brownstein, J. S., Hswen, Y., Garibaldi, B. T., Zhang, B. B., & Kriner, D. L. (2020). Factors associated with us adults' likelihood of accepting covid-19 vaccination [Journal Article]. *Jama Network Open*, 3(10), 13. Retrieved from <GotoISI>://WOS:000586431300007 doi: 10.1001/jamanetworkopen.2020.25594
- Krinsky, I., & Robb, A. (1990, FEB). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, 72(1), 189-190. doi: 10.2307/2109761
- Long, Z., Kormos, C., Sussman, R., & Axsen, J. (2021). Mpg, fuel costs, or savings? exploring the role of information framing in consumer valuation of fuel economy using a choice experiment [Journal Article]. *Transportation Research Part a-Policy and Practice*, 146, 109-127. Retrieved from <GotoISI>://WOS:000631021700008 doi: 10.1016/j.tra.2021.02.004
- Luoto, J., Najnin, N., Mahmud, M., Albert, J., Islam, M. S., Luby, S., ... Levine, D. I. (2011). What point-of-use water treatment products do consumers use? evidence from a randomized controlled trial among the urban poor in bangladesh. *PloS one*, 6(10), e26132.
- Manipis, K., Street, D., Cronin, P., Viney, R., & Goodall, S. (2021). Exploring the trade-off between economic and health outcomes during a pandemic: A discrete choice experiment of lockdown policies in australia [Journal Article]. *Patient-Patient Centered Outcomes Research*, 14(3), 359-371. Retrieved from <GotoISI>://WOS:000627190900001 doi: 10.1007/s40271-021-00503-5
- Mare, R. D. (1991). Five decades of educational assortative mating. *American sociological review*, 15-32.

- Marinescu, I., Skandalis, D., & Zhao, D. (2021). The impact of the federal pandemic unemployment compensation on job search and vacancy creation [Journal Article]. *National Bureau of Economic Research Working Paper Series, No. 28567*. Retrieved from <http://www.nber.org/papers/w28567> doi: 10.3386/w28567
- McClendon, D. (2016). Religion, marriage markets, and assortative mating in the united states. *Journal of Marriage and Family, 78*(5), 1399–1421.
- Mitchell-Nelson, J., & Cameron, T. A. (2021). *Addressing sample selection in the estimation of choice models using online survey panels* [Manuscript].
- Mold, A. (2012). Will it all end in tears? infrastructure spending and african development in historical perspective. *Journal of International Development, 24*(2), 237–254.
- Mulhern, B., Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., . . . Tsuchiya, A. (2014). Preparatory study for the revaluation of the eq-5d tariff: methodology report [Journal Article]. *Health Technology Assessment, 18*(12), 1-+. Retrieved from <GotoISI>://WOS:000331905700001 doi: 10.3310/hta18120
- Muqattash, R., Niankara, I., & Traoret, R. I. (2020). Survey data for covid-19 vaccine preference analysis in the united arab emirates [Journal Article]. *Data in Brief, 33*, 9. Retrieved from <GotoISI>://WOS:000600652300112 doi: 10.1016/j.dib.2020.106446
- Nguyen, H. H., Alexiou, A., & Singleton, A. (2017). Personal name classification using collective data.
- Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3: Genes, Genomes, Genetics, 3*(5), 891–907.
- Plum, C., Olschewski, R., Jobin, M., & van Vliet, O. (2019). Public preferences for the swiss electricity system after the nuclear phase-out: A choice experiment [Journal Article]. *Energy Policy, 130*, 181-196. Retrieved from <GotoISI>://WOS:000471083700015 doi: 10.1016/j.enpol.2019.03.054
- Pursiainen, V. (2019). Cultural biases in equity analysis. *Available at SSRN 3153900*.
- Qian, N. (2015). Making progress on foreign aid. *Annu. Rev. Econ., 7*(1), 277–308.

- Rahmani, D., & Loureiro, M. L. (2019). Assessing drivers' preferences for hybrid electric vehicles (hev) in Spain [Journal Article]. *Research in Transportation Economics*, *73*, 89-97. Retrieved from <GotoISI>://WOS:000472704400010 doi: 10.1016/j.retrec.2018.10.006
- Rajan, R. G., & Subramanian, A. (2008). Aid and growth: What does the cross-country evidence really show? *The Review of Economics and Statistics*, *90*(4), 643-665.
- Reed, S., Gonzalez, J. M., & Johnson, F. R. (2020). Willingness to accept trade-offs among COVID-19 cases, social-distancing restrictions, and economic impact: A nationwide US study [Journal Article]. *Value in Health*, *23*(11), 1438-1443. Retrieved from <GotoISI>://WOS:000582722700007 doi: 10.1016/j.jval.2020.07.003
- Rees-Jones, A., D'Attoma, J., Piolatto, A., & Salvadori, L. (2020). COVID-19 changed tastes for safety-net programs [Journal Article]. *National Bureau of Economic Research Working Paper Series, No. 27865*. Retrieved from <http://www.nber.org/papers/w27865> doi: 10.3386/w27865
- Rogers, A. A., & Burton, M. P. (2017). Social preferences for the design of biodiversity offsets for shorebirds in Australia [Journal Article]. *Conservation Biology*, *31*(4), 828-836. Retrieved from <GotoISI>://WOS:000405457000010 doi: 10.1111/cobi.12874
- Rosston, G. L., Savage, S. J., & Waldman, D. M. (2010). Household demand for broadband internet in 2010 [Journal Article]. *BE Journal of Economic Analysis Policy*, *10*(1), 44. Retrieved from <GotoISI>://WOS:000283285700011 doi: 10.2202/1935-1682.2541
- Rowen, D., Brazier, J., Mukuria, C., Keetharuth, A., Hole, A. R., Tsuchiya, A., . . . Shackley, P. (2016). Eliciting societal preferences for weighting QALYs for burden of illness and end of life [Journal Article]. *Medical Decision Making*, *36*(2), 210-222. Retrieved from <GotoISI>://WOS:000367472900006 doi: 10.1177/0272989x15619389
- Sainato, M. (2021, May 21). *Millions of unemployed in US face hardship under republican benefit cuts*. Retrieved from https://www.theguardian.com/us-news/2021/may/21/us-unemployment-benefits-pandemic-cut-republican-states?CMP=oth_b-aplnews_d-1
- Sandorf, E. D., Aanesen, M., & Navrud, S. (2016). Valuing unfamiliar and complex environmental goods: A comparison of valuation workshops and internet panel surveys with videos [Journal Article]. *Ecological Economics*, *129*, 50-61. Retrieved from <GotoISI>://WOS:000382350400006 doi: 10.1016/j.ecolecon.2016.06.008

- Shigeoka, H., & Yamada, K. (2019). Income-comparison attitudes in the united states and the united kingdom: Evidence from discrete-choice experiments [Journal Article]. *Journal of Economic Behavior Organization*, 164, 414-438. Retrieved from <GotoISI>://WOS:000480667400021 doi: 10.1016/j.jebo.2019.06.012
- Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *The Quarterly journal of economics*, 124(2), 469–529.
- Spolaore, E., & Wacziarg, R. (2016). Ancestry, language and culture. In *The palgrave handbook of economics and language* (pp. 174–211). Springer.
- Spolaore, E., & Wacziarg, R. (2018). Ancestry and development: New evidence. *Journal of Applied Econometrics*, 33(5), 748–762.
- Stein, J., & Werner, E. (2020). Trump demands payroll tax cut while gop eyes benefit cuts for unemployed. *Washington Post*. Retrieved from <https://www.washingtonpost.com/business/2020/07/19/republican-stimulus-unemployment-coronavirus/>
- Temple, J. R. (2010). Aid and conditionality. In *Handbook of development economics* (Vol. 5, pp. 4415–4523). Elsevier.
- UPFRONT, W.-T. (2020, Aug). *Upfront recap: Sen. ron johnson: Democrats want too much in second stimulus*. Retrieved from <https://www.wisn.com/article/upfront-recap-sen-ron-johnson-democrats-want-too-much-in-second-stimulus/33481405>
- Urrutia, C. P. I., Erdem, S., Birks, Y. F., Taylor, S. J. C., Richardson, G., Bower, P., ... Manca, A. (2021). People’s preferences for self-management support [Journal Article]. *Health Services Research*, 11. Retrieved from <GotoISI>://WOS:000621640800001 doi: 10.1111/1475-6773.13635
- Wilkinson, D. A., Marshall, J. C., French, N. P., & Hayman, D. T. S. (2018). Habitat fragmentation, biodiversity loss and the risk of novel infectious disease emergence [Journal Article]. *Journal of the Royal Society Interface*, 15(149), 10. Retrieved from <GotoISI>://WOS:000456783800003 doi: 10.1098/rsif.2018.0403
- Wilson, I. E., Mody, A., McKay, G., Hlatshwayo, M., Bradley, C., Thompson, V., ... Geng, E. H. (2020). Public preferences for social distancing behaviors to mitigate the spread of covid-19: A discrete choice experiment [Journal Article]. *medRxiv*, 2020.12.12.20248103. Retrieved from <https://www.medrxiv.org/content/medrxiv/early/2020/12/14/2020.12.12.20248103.full.pdf> doi: 10.1101/2020.12.12.20248103

Zeleny, J., & Luhby, T. (2021, May 21). *Biden administration unable to continue \$300 weekly pandemic unemployment benefits that gop governors are slashing*. Retrieved from <https://www.cnn.com/2021/05/20/politics/pandemic-unemployment-benefits-300-dollars/index.html>