Online Supplement for "Estimating Endogenous Effects on Ordinal Outcomes"*

Andrew Chesher†
UCL and CeMMAP

Adam M. Rosen‡
Duke University and CeMMAP

Zahra Siddique§
University of Bristol

June 13, 2023

Abstract

This online supplement provides appendices for Chesher et al. (2022). Appendix A provides additional details for the derivation of bounds for the nonlinear IV model presented in Section 3.1. Appendix B provides mean values of baseline adult and household covariates across randomization sites. Appendix C presents linear model estimates using the ISCPR MTO data and Appendix D presents estimates for the first stage of the complete triangular (CT) model. Appendix E illustrates how a semi-monotone IV assumption can sign the effect of the endogenous variable on the outcome.

*We have benefitted from comments from Alex Belloni and participants at research seminars given at Erasmus University Rotterdam, Duke, Simon Fraser, Leicester, Johns Hopkins, a 2018 joint CeMMAP/Northwestern conference on incomplete models and the 2020 EALE-SOLE-AASLE conference. Adam Rosen and Andrew Chesher gratefully acknowledge financial support from the UK Economic and Social Research Council through a grant (RES-589-28-0001) to the ESRC Centre for Microdata Methods and Practice (CeMMAP). Adam Rosen gratefully acknowledges financial support from a British Academy mid-career Fellowship. The U.S. Department of Housing and Urban Development (HUD) provided the MTO data; we use the version made available by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. The views expressed in this paper are those of the authors and should not be interpreted as those of HUD or the U.S. Government.

†Address: Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Email: andrew.chesher@ucl.ac.uk.

‡Address: Department of Economics, Duke University, 213 Social Sciences, Box 90097 Durham, NC 27708, United States. Email: adam.rosen@duke.edu.

§Address: University of Bristol, Senate House, Tyndall Avenue, Bristol BS8 1TH, United Kingdom. Email: zahra.siddique@bristol.ac.uk.

Appendix F provides computational details for the implementation of IV set estimates and confidence sets, further to those provided in the main text.

Keywords: Instrumental Variables, Ordered Choice, Incomplete Models, Partial Identification, Neighborhood Effects, Subjective Well-Being, Moving to Opportunity.
JEL classification: C25, C26, C35, I31, R2.

A Details of Bound Derivations

This section provides further mathematical detail for the derivation of bounds for the IV model presented in Section 3.1. To proceed with set identification analysis for model parameters $\theta \equiv (\beta, \gamma, c_1, c_2)$, define the sets

$$\mathcal{U}(y, x, w; \theta) \equiv \left\{ \begin{array}{cc} (-\infty, c_1 - w\beta - x\gamma], & \text{if } y = 0, \\ (c_1 - w\beta - x\gamma, c_2 - w\beta - x\gamma], & \text{if } y = 1, \\ (c_2 - w\beta - x\gamma, \infty), & \text{if } y = 2. \end{array} \right\}. \tag{A.1}$$

From Chesher and Rosen (2017) we have for any set $\mathcal{S} \subseteq \mathbb{R}$ the conditional *containment* inequality

$$C_\theta(\mathcal{S}|x, z) \equiv \mathbb{P}[\mathcal{U}(Y, X, W; \theta) \subseteq \mathcal{S}|X = x, Z = z] \leq \mathbb{P}[U \in \mathcal{S}|X = x, Z = z],$$

as well as the conditional *capacity* inequality

$$\mathbb{P}[U \in \mathcal{S}|X = x, Z = z] \leq \mathbb{P}[\mathcal{U}(Y, X, W; \theta) \cap \mathcal{S} \neq \emptyset|X = x, Z = z],$$

where

$$\overline{C}_\theta(\mathcal{S}|x, z) \equiv 1 - C_\theta(\mathcal{S}^c|x, z) = \mathbb{P}[\mathcal{U}(Y, X, W; \theta) \cap \mathcal{S} \neq \emptyset|X = x, Z = z].$$

In the context of the ordered outcome IV model, the capacity and containment functional inequalities take a particular form, which is now derived. Define for $y \in \{0, 1, 2, 3\}$, $x \in \text{Supp}(X)$, and any $w \in \text{Supp}(W)$ the function $c(y, x, w; \theta)$ as follows.

$$\begin{aligned} c(0, x, w; \theta) &\equiv -\infty, \quad c(1, x, w; \theta) \equiv c_1 - x\gamma - w\beta, \\ c(2, x, w; \theta) &\equiv c_2 - x\gamma - w\beta, \quad c(3, x, w; \theta) \equiv \infty. \end{aligned}$$

Thus, we can express the set $\mathcal{U}(y, x, w; \theta)$ as

$$\mathcal{U}(y, x, w; \theta) = [c(Y, X, W; \theta), c(Y + 1, X, W; \theta)],$$

with the lower (upper) bound of the interval understood to be open in the event $c(Y, X, W; \theta) =$

$-\infty$ $(= +\infty)$.[1]

We can now re-express the containment and capacity functionals as

1. For all $t \in \mathbb{R}$:

$$\begin{aligned} C_\theta\left((-\infty, t]\,|x, z\right) &= \mathbb{P}\left[c\left(Y+1, X, W; \theta\right) \leq t | X = x, Z = z\right], \\ \overline{C}_\theta\left((-\infty, t]\,|x, z\right) &= \mathbb{P}\left[c\left(Y, X, W; \theta\right) \leq t | X = x, Z = z\right]. \end{aligned}$$

The difference $\overline{C}_\theta\left((-\infty, t]\,|x, z\right) - C_\theta\left((-\infty, t]\,|x, z\right)$ is equal to

$$\mathbb{P}\left[c\left(Y, X, W; \theta\right) \leq t < c\left(Y+1, X, W; \theta\right) | X = x, Z = z\right].$$

2. For all $s, t \in \mathbb{R}$, $s \leq t$,

$$\begin{aligned} C_\theta\left([t_1, t_2]\,|x, z\right) &= \mathbb{P}\left[t_1 \leq c\left(Y, X, W; \theta\right) \wedge c\left(Y+1, X, W; \theta\right) \leq t_2 | X = x, Z = z\right], \\ \overline{C}_\theta\left([t_1, t_2]\,|x, z\right) &= \mathbb{P}\left[c\left(Y, X, W; \theta\right) \leq t_2 \wedge c\left(Y+1, X, W; \theta\right) \geq t_1 | X = x, Z = z\right]. \end{aligned}$$

3. For all $t \in \mathbb{R}$:

$$\begin{aligned} C_\theta\left([t, \infty)\,|x, z\right) &= \mathbb{P}\left[c\left(Y, X, W; \theta\right) \geq t | X = x, Z = z\right], \\ \overline{C}_\theta\left([t, \infty)\,|x, z\right) &= \mathbb{P}\left[c\left(Y+1, X, W; \theta\right) \geq t | X = x, Z = z\right]. \end{aligned}$$

If $U \sim \mathcal{N}(0,1)$ and $U \perp\!\!\!\perp (X, Z)$, then using results from Chesher and Rosen (2017) Theorem 4 we have that the identified set for $\theta \equiv (\beta, c_1, c_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are those parameters such that for all $s, t \in \mathbb{R}$, $s < t$:

$$\begin{aligned} \max_{x,z} \mathbb{P}\left[c\left(Y+1, X, W; \theta\right) \leq t | X = x, Z = z\right] &\leq \Phi(t), \\ \max_{x,z} \mathbb{P}\left[c\left(Y, X, W; \theta\right) \geq t | X = x, Z = z\right] &\leq 1 - \Phi(t), \\ \max_{x,z} \mathbb{P}\left[\left(s \leq c\left(Y, X, W; \theta\right) \wedge c\left(Y+1, X, W; \theta\right) \leq t\right) | X = x, Z = z\right] &\leq \Phi(t) - \Phi(s). \end{aligned}$$

If we continue to assume that $U \perp\!\!\!\perp (X, Z)$ but without imposing $U \sim \mathcal{N}(0, \sigma^2)$, we have from Chesher and Rosen (2017) Corollary 3 that bounds on θ are given by the following

[1] When the endpoints of the intervals in (A.1) are finite it is convenient to define these intervals as closed intervals which include their endpoints, although this is of no substantive consequence with continuously distributed U.

inequalities for all $t_1, t_2 \in \mathbb{R}_{\pm\infty}$ with $t_1 < t_2$, where $\mathbb{R}_{\pm\infty}$ denotes the extended real line (i.e. inclusive of $\pm\infty$):

$$\max_{x,z} C_\theta\left([t_1, t_2]\,|x, z\right) \leq \min_{x,z} \overline{C}_\theta\left([t_1, t_2]\,|x, z\right).$$

Substitution for C_θ and $\overline{C}_\theta$ then delivers the inequalities displayed in the main text. With this assumption in place we require a location normalization, for which we can use the restriction that $\text{Median}\,(U|X, Z) = 0$, giving the inequalities

$$\max_{x,z} C_\theta\left((-\infty, 0]\,|x, z\right) \leq \frac{1}{2} \leq \min_{x,z} \overline{C}_\theta\left((-\infty, 0]\,|x, z\right). \tag{A.2}$$

If we then drop the independence restriction $U \perp\!\!\!\perp (X, Z)$ and replace it with only the weaker restriction that $\text{Median}\,(U|X, Z) = 0$, we obtain the inequalities given in (3.3) and (3.4).

$$\begin{aligned} \max_{x,z} \mathbb{P}\left[c\left(Y+1, X, W; \theta\right) \leq 0 | X = x, Z = z\right] &\leq \frac{1}{2}, \\ \max_{x,z} \mathbb{P}\left[c\left(Y, X, W; \theta\right) \geq 0 | X = x, Z = z\right] &\leq \frac{1}{2}. \end{aligned}$$

B Data Description

Table B.1 shows the set of covariates which were elicited in a baseline survey before randomization took place in 1994-1998. These covariates include randomization site, gender, age, race and ethnicity, marital status, work and education, whether on welfare, household income, household size, and covariates on the kind of neighborhood the individual was living in and reasons why they wanted to move. As may be seen in Table B.1 the baseline covariates are quite balanced across different treatment arms.

Prior to our access, some observations for baseline covariates in the ICPSR data were replaced with imputed values (or group averages), either when the observation on the covariate was missing or to maintain data confidentiality; further details are provided in the codebook documentation of the MTO Restricted Access Dataset (ICPSR 34860) for the *Science* article Ludwig et al. (2012). We thus report estimates in all point-identifying models with and without controlling for these baseline covariates.

In empirical analysis that produces point estimates we use weights in our empirical analysis following Ludwig et al. (2012) to account for differences in random assignment proportions across sites and time as well as various aspects of survey administration. Further details regarding these weights can be found in the supplementary material to Ludwig et al. (2012). In unreported results, we also carried out all computations without incorporating sampling weights and obtained only small numerical differences, resulting in qualitatively similar conclusions. These weights were not used in empirical analysis using moment inequalities, as weighting different covariate subpopulations has no effect on identification analysis at the population level, and its effect on inference with partial identification using many moment inequality has not been studied.

Ludwig et al. (2012) construct residential poverty using the z-score of duration weighted share poor in an individual's neighborhood while share minority is constructed using the z-score of duration weighted share minority. Share poor is the fraction of census tract residents living below the poverty threshold while share minority is the fraction of census tract minority residents. These variables are constructed using interpolated data from the 1990 and 2000 decennial census as well as the 2005-2009 American Community Survey for all neighborhoods MTO adults lived in between random assignment and the start of the long term survey fielding period. Duration weighted share poor and share minority are the 'average measures weighted by the amount of time respondents lived at each of their addresses between random assignment and May 2008 (just prior to the start of the long term survey fielding period)'.

Z-scores of these variables are standardized values of the duration-weighted neighborhood characteristic, using the control group weighted average and standard deviation.[2]

Figures B.1 and B.2 show the distributions of these across different treatment groups. Adults belonging to the experimental voucher group lived in less poor neighborhoods than either the MTO traditional section 8 voucher group or the control group (Figure B.1). Adults belonging to the experimental voucher group also lived in neighborhoods that had fewer minority residents, but the difference from the MTO traditional section 8 voucher group or the control group is less striking than for neighborhood poverty (Figure B.2).

[2]Ludwig et al. (2012) use duration weighted measures rather than current measures of neighborhood environment in their main analysis since an individual's life outcomes may depend on cumulative exposure to the neighborhood environment. Nevertheless, they find that their main conclusions remain robust to the use of current measures of neighborhood environment, or neighborhood poverty and share minority measured at the start of the MTO long-term fieldwork (May 2008).

Table B.1: Baseline characteristics of MTO adults or covariates X across randomization groups

	Experimental	Section 8	Control
Site:			
Baltimore	0.14	0.14	0.14
Boston	0.19	0.19	0.22
Chicago	0.27	0.16	0.16
Los Angeles	0.19	0.22	0.27
New York	0.22	0.29	0.20
Demographic characteristics:			
African American (non-hispanic)	0.67	0.59	0.63
Hispanic ethnicity (any race)	0.28	0.36	0.32
Female	0.99	0.98	0.98
<= 35 years old	0.14	0.14	0.15
36-40 years old	0.21	0.23	0.22
41-45 years old	0.25	0.22	0.23
46-50 years old	0.19	0.19	0.18
Never married	0.64	0.61	0.64
Parent while younger than 18 years old	0.26	0.26	0.25
Working	0.27	0.28	0.24
Enrolled in school	0.16	0.18	0.16
High school diploma	0.40	0.35	0.37
General Education Development (GED) certificate	0.16	0.18	0.19
Receiving Aid to Families with Dependent Children (AFDC)	0.77	0.74	0.78
Household characteristics:			
Household income (dollars)	12,659	12,799	12,655

continued on next page

Table B.1: Baseline characteristics of MTO adults or covariates X across randomization groups

	Experimental	Section 8	Control
Household owns a car	0.17	0.18	0.17
Household member had a disability	0.15	0.17	0.15
No teens in household	0.61	0.62	0.64
Household size is <= 2	0.21	0.22	0.20
Household size is 3	0.30	0.30	0.32
Household size is 4	0.24	0.24	0.22
Neighborhood characteristics:			
Household member was a crime victim in past 6 months	0.43	0.42	0.41
Neighborhood streets very unsafe at night	0.49	0.54	0.51
Very dissatisfied with neighborhood	0.47	0.48	0.45
Household living in neighborhood > 5 years	0.60	0.63	0.60
Household moved more > $3x$ in last 5 yrs	0.09	0.08	0.11
Household has no family living in neighborhood	0.63	0.64	0.63
Household has no friends living in neighborhood	0.40	0.41	0.41
Household head chatted with neighbor >= $1x$ per week	0.53	0.50	0.54
Household head very likely to to tell on neighborhood kid	0.55	0.52	0.56
Household head very sure of finding apartment	0.47	0.50	0.46
Housheold head applied for Section 8 before	0.39	0.40	0.43
Primary or secondary reason for wanting to move:			
Want to move to get away from gangs and drugs	0.78	0.75	0.78
Want to move for better schools for children	0.49	0.54	0.47
Want to move to get a bigger/better apartment	0.45	0.44	0.46

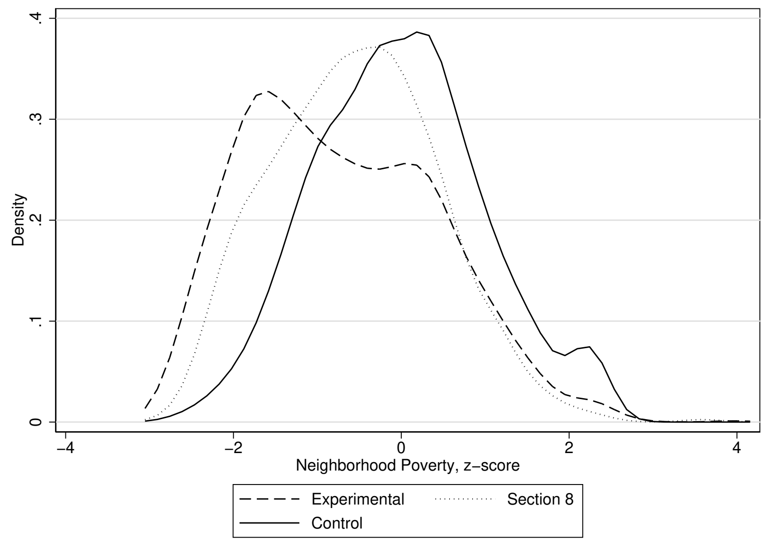
continued on next page

Table B.1: Baseline characteristics of MTO adults or covariates X across randomization groups

	Experimental	Section 8	Control
Want to move to get a job	0.07	0.05	0.06
N	1422	655	1098

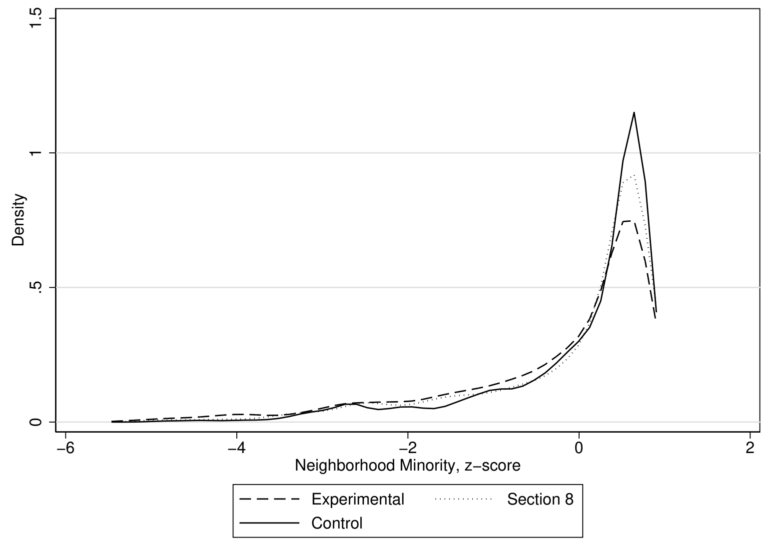
Notes: Each cell gives the average value of a variable in the sub-sample. Only observations with non-missing values for Subjective Well Being (SWB), neighbourhood characteristics and x covariates are used. There are seven observations with missing SWB, three observations with missing neighborhood characteristics and 89 out of 3,273 observations with missing household income. Some observations of covariates include imputed values.
Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Figure B.1: Distribution of neighborhood poverty by randomization group



Notes: Only observations with non-missing values for neighborhood poverty are used (neighborhood poverty is missing for 3 out of 3,273 adults). These include 1,453 adults in the Experimental group, 678 adults in the Section 8 group and 1,139 adults in the Control group. *Source:* Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Figure B.2: Distribution of neighborhood minority by randomization group



Notes: Only observations with non-missing values for neighborhood minority are used (neighborhood minority is missing for 3 out of 3,273 adults). These include 1,453 adults in the Experimental group, 678 adults in the Section 8 group and 1,139 adults in the Control group. *Source:* Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

C Linear Model Estimates

We estimate linear model parameters using the ICPSR MTO data using similar methods as Ludwig et al. (2012). Our results are very similar to theirs, with minor differences seemingly down to small discrepancies between their data and that available through ICPSR. The estimation sample in the original analysis has $3,273$ adults while the estimation sample using data from ICPSR has $3,175$ adults; this is due to the missing observations on SWB, neighborhood characteristics and household income in the ICPSR data. Additionally, some observations for baseline covariates in the ICPSR data are replaced with imputed values (or group averages) either when the observation is missing or to maintain data confidentiality.

The results obtained using least squares linear regression are presented in Panel A of Table C.1; results are given using neighborhood poverty and share minority separately and together as neighborhood characteristics W. The coefficients on W give the effect of neighborhood characteristics on SWB under the assumption that W is uncorrelated with U.

In columns (1)-(3) of Panel A in Table C.1, dummy variables for randomization site are used as the only covariates X. The results show a statistically significant and negative effect of neighborhood poverty and share minority on SWB. When both neighborhood poverty and share minority are included, the negative effect of share minority on SWB becomes statistically indistinguishable from zero. In columns (4)-(6) of the table a complete set of baseline covariates (as given in Table B.1) is included, and the results remain qualitatively unchanged.

Interactions of MTO assignment and randomization site are used as instrumental variables Z in the results reported in Panel B of Table C.1.[3] Unlike the results in Panel A these estimators allow for the possibility that neighborhood characteristics are endogenous. Under an instrument monotonicity assumption the estimated coefficients are consistent for weighted averages of LATE parameters; see Chapter 4.5 of Angrist and Pischke (2009) for details regarding the mixture of LATE parameters estimated when there are multiple endogenous variables and additional covariates. These are, however, sensitive to the cardinal scale used for the categorical SWB outcomes.

Columns (1)-(3) in Panel B of Table C.1 report results without the inclusion of additional

[3]That is, instrumental variables Z here refer to interactions of both the included exogenous variable randomization site as well as excluded exogenous treatment assignment; these are identical to the instruments used by Ludwig et al. (2012). For tests on the validity of these instruments and alternative estimates (including limited information maximum likelihood (LIML) and Fuller-modified LIML) designed to adjust for weak instruments we refer the reader to Tables S5 and S9 of the supplementary materials to Ludwig et al. (2012).

covariates. As before the coefficient on neighborhood poverty is negative and statistically significantly different from zero. The coefficient on the neighborhood minority variable is closer to zero. In column (2) it is statistically insignificant and in column (3) it is positive and larger in magnitude, but remains statistically insignificant.

Columns (4)-(6) report results when a complete set of baseline covariates is included. These results can be directly compared with those in Tables S5 and S9 in the supplementary material to Ludwig et al. (2012), where estimates from IV regressions that included baseline covariates were also reported. The results reported in Panel B of Table C.1 are qualitatively similar, with minor differences likely caused by the aforementioned differences in the two estimation samples and small modifications to the data available through ICPSR. The coefficient on neighborhood poverty on SWB in Table S5 is -0.141 while here it has been estimated as -0.096, both having p-values less than 0.05. The coefficient on neighborhood minority on SWB in Table S5 is -0.069 while here it has been estimated as -0.063, both with p-values higher than 0.1. The coefficient on neighborhood poverty, while controlling for neighborhood minority, in Table S9 is -0.261 while here it has been estimated as -0.186, both with p-values less than 0.01. The coefficient on neighborhood minority, while controlling for neighborhood poverty, in Table S9 is 0.279 with a p-value between 0.05 and 0.1 while here it has been estimated as 0.202 with a p-value of 0.105.

Table C.2 reports ITT effects obtained by linear regression of SWB on X and Z using the ICPSR MTO data, which correspond roughly to those of Table S4 of the supplementary material to Ludwig et. al (2012). Specifically, the coefficient on MTO voucher assignment is the ITT effect. Columns (1)-(3) of Table C.2 report ITT estimates without including a complete set of covariates while columns (4)-(6) report ITT estimates with inclusion of these covariates. Column (1) pools both kinds of vouchers (experimental and section 8) together. Column (2) excludes adults who were randomly assigned the section 8 voucher, so gives the ITT effect of the experimental voucher on SWB. Column (3) excludes adults who were randomly assigned the experimental voucher, so gives the ITT effect of the section 8 voucher on SWB. In all three cases the ITT effect of an MTO voucher is positive with a p-value between 0.01 and 0.10, consistent with a positive effect of being offered an MTO voucher on SWB.

Compared to the case without covariates, the coefficient on the MTO voucher reported in columns (4)-(6) of Table C.2 is slightly larger. Estimates still indicate a positive and statistically significant effect of being offered an MTO voucher on SWB.

Table C.1: Linear model estimation (OLS and IV) of neighborhood effects on SWB

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS estimation						
$\beta_{Poverty}$	-0.0551***		-0.0491***	-0.0546***		-0.0534***
	(0.0130)		(0.0148)	(0.0131)		(0.0151)
$\beta_{Minority}$		-0.0367***	-0.0135		-0.0287**	-0.0029
		(0.0129)	(0.0147)		(0.0136)	(0.0157)
N	3263	3263	3263	3175	3175	3175
Panel B: IV estimation						
$\beta_{Poverty}$	-0.0916**		-0.1803***	-0.0962**		-0.1859***
	(0.0382)		(0.0675)	(0.0376)		(0.0687)
$\beta_{Minority}$		-0.0383	0.2048		-0.0632	0.2019
		(0.0694)	(0.1245)		(0.0688)	(0.1247)
N	3263	3263	3263	3175	3175	3175

Notes: The dependent variable is Subjective Well Being (SWB) which takes the value zero for not too happy, one for pretty happy and two for very happy; columns (1)-(3) use a set of dummy variables for randomization site as covariates X while columns (4)-(6) use a complete set of baseline characteristics (as given in Table B.1), and whether a sample adult was included in the first release of the long-term evaluation survey fielding period, as covariates X; all regressions are weighted; * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.
Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Table C.2: Linear model estimation (ITT) of neighborhood effects on SWB

	(1)	(2)	(3)	(4)	(5)	(6)
Z = Any voucher	0.0630**			0.0660**		
	(0.0283)			(0.0284)		
Z = Low pov voucher		0.0521*			0.0546*	
		(0.0298)			(0.0298)	
Z = Sec 8 voucher			0.0793**			0.0875**
			(0.0385)			(0.0440)
N	3266	2593	1811	3178	2523	1753

Notes: The dependent variable is Subjective Well Being (SWB) which takes the value zero for not too happy, one for pretty happy and two for very happy; columns (1)-(3) use a set of dummy variables for randomization site as covariates X while columns (4)-(6) use a complete set of baseline characteristics (as given in Table B.1), and whether a sample adult was included in the first release of the long-term evaluation survey fielding period, as covariates X; all regressions are weighted; * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.
Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

D Complete Triangular Model First Stage Estimates

First stage estimates for the CT model estimated in Section 4.2 are presented below in Table D.1.

Table D.1: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\delta_1^{exp,Balt}$	-1.0912***	-0.8235***	-1.0910***	-1.1591***	-0.9015***	-1.1589***
	(0.1017)	(0.1143)	(0.1021)	(0.1038)	(0.1190)	(0.1042)
$\delta_1^{exp,Bos}$	-1.2798***	-1.7007***	-1.2819***	-1.2154***	-1.4961***	-1.2168***
	(0.0877)	(0.1264)	(0.0876)	(0.0896)	(0.1189)	(0.0898)
$\delta_1^{exp,Chi}$	-0.3068***	0.0993	-0.3055***	-0.3470***	0.0334	-0.3450***
	(0.0852)	(0.0736)	(0.0853)	(0.0915)	(0.0792)	(0.0917)
$\delta_1^{exp,LA}$	-0.8787***	-0.3421***	-0.8801***	-0.8137***	-0.3436***	-0.8142***
	(0.1007)	(0.0822)	(0.1003)	(0.1044)	(0.0916)	(0.1042)
$\delta_1^{exp,NY}$	-0.8052***	-0.1401	-0.8021***	-0.7993***	-0.1326	-0.7945***
	(0.0875)	(0.0854)	(0.0877)	(0.0891)	(0.0873)	(0.0895)
$\delta_1^{sec8,Balt}$	-1.0427***	-0.6651***	-1.0412***	-1.1065***	-0.7093***	-1.1032***
	(0.1184)	(0.1992)	(0.1194)	(0.1254)	(0.1986)	(0.1264)
$\delta_1^{sec8,Bos}$	-1.0880***	-1.2662***	-1.0838***	-1.0376***	-1.1362***	-1.0328***
	(0.1055)	(0.1428)	(0.1058)	(0.1130)	(0.1455)	(0.1134)
$\delta_1^{sec8,Chi}$	-0.1905*	0.2901***	-0.1863*	-0.2696**	0.1970**	-0.2643**
	(0.1092)	(0.0802)	(0.1092)	(0.1171)	(0.0894)	(0.1168)
$\delta_1^{sec8,LA}$	-0.8139***	0.0257	-0.8072***	-0.7508***	-0.0071	-0.7429***
	(0.0960)	(0.0988)	(0.0960)	(0.1041)	(0.1066)	(0.1038)
$\delta_1^{sec8,NY}$	-0.3742***	-0.0448	-0.3728***	-0.3945***	-0.0548	-0.3920***
	(0.0913)	(0.0838)	(0.0916)	(0.0923)	(0.0867)	(0.0926)
$\delta_1^{cont,Balt}$	-0.5220***	-0.3311***	-0.5184***	-0.5641***	-0.3941***	-0.5590***
	(0.0899)	(0.0931)	(0.0902)	(0.0949)	(0.0998)	(0.0951)

continued on next page

Table D.1: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\delta_1^{cont,Bos}$	-0.7145***	-1.1184***	-0.7106***	-0.6409***	-0.9028***	-0.6350***
	(0.0722)	(0.1018)	(0.0721)	(0.0823)	(0.1047)	(0.0825)
$\delta_1^{cont,Chi}$	0.2299**	0.2621***	0.2297**	0.1848*	0.2124***	0.1855*
	(0.0999)	(0.0718)	(0.0999)	(0.1078)	(0.0809)	(0.1077)
$\delta_1^{cont,LA}$	0.1584*	0.2110***	0.1597*	0.2360**	0.2192***	0.2381**
	(0.0907)	(0.0664)	(0.0906)	(0.0959)	(0.0729)	(0.0959)
$\delta_1^{cont,NY}$	0.1371**	0.1735***	0.1354**	0.5305**	-0.2895	0.5258**
	(0.0547)	(0.0567)	(0.0548)	(0.2234)	(0.2631)	(0.2235)
$\delta_2^{exp,Balt}$			-0.8173***			-0.8886***
			(0.1118)			(0.1171)
$\delta_2^{exp,Bos}$			-1.7055***			-1.4949***
			(0.1226)			(0.1166)
$\delta_2^{exp,Chi}$			0.0880			0.0216
			(0.0733)			(0.0798)
$\delta_2^{exp,LA}$			-0.3677***			-0.3710***
			(0.0828)			(0.0929)
$\delta_2^{exp,NY}$			-0.1588*			-0.1447*
			(0.0846)			(0.0877)
$\delta_2^{sec8,Balt}$			-0.6895***			-0.7349***
			(0.1946)			(0.1968)
$\delta_2^{sec8,Bos}$			-1.2842***			-1.1477***
			(0.1401)			(0.1415)
$\delta_2^{sec8,Chi}$			0.2812***			0.1982**
			(0.0799)			(0.0877)
$\delta_2^{sec8,LA}$			0.0347			0.0240
			(0.0883)			(0.0941)

continued on next page

Table D.1: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\delta_2^{sec8,NY}$			-0.0617			-0.0641
			(0.0843)			(0.0863)
$\delta_2^{cont,Balt}$			-0.3553***			-0.4127***
			(0.0982)			(0.1033)
$\delta_2^{cont,Bos}$			-1.1349***			-0.9192***
			(0.1018)			(0.1077)
$\delta_2^{cont,Chi}$			0.2454***			0.1999**
			(0.0743)			(0.0824)
$\delta_2^{cont,LA}$			0.1980***			0.2055***
			(0.0667)			(0.0731)
$\delta_2^{cont,NY}$			0.1858***			-0.2870
			(0.0565)			(0.2625)
N	3263	3263	3263	3175	3175	3175

Notes: Each column reports first stage estimates of a triangular model for specifications reported in corresponding columns of Table 3. The dependent variables in the first stage are neighborhood poverty and neighborhood minority. Columns (1)-(3) exclude while columns (4)-(6) include a complete set of baseline characteristics (as given in Table B.1), as well as whether a sample adult was included in the first release of the long-term evaluation survey fielding period, as covariates X; all regressions are weighted; * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.
Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

E Semi-Monotone Instrument

In this section we briefly consider the ability of a type of instrument monotonicity restriction to sign the effect of neighborhood poverty on SWB, when used in conjunction with the IV modeling restrictions. This constitutes a blend of the sort of monotonicity restriction used by Pinto (2019) with the IV model, in a setting in which the endogenous variable is continuous and latent type space is infinite.

Receipt of either type of voucher by a household expands the menu of housing options available. This may be credibly thought to induce low income participants to choose a housing location in a neighborhood of no higher poverty level than would have been chosen had they not received the voucher. Further, if an experimental voucher were received that stipulates it can only be used in a neighborhood with poverty rate below some threshold, one might reason that this would induce individuals to live in an even (weakly) lower poverty neighborhood. On the other hand, it could be that some participants who would move if awarded a traditional voucher might choose not to move at all if given an experimental voucher, because of the additional restrictions imposed on those neighborhoods to which they could move. Nonetheless, receipt of a traditional voucher could induce those families to move to a lower poverty neighborhood than the one they would otherwise be in, even if it were not of low enough poverty level to use the experimental voucher.

This can be used to motivate an instrument semi-monotonicity restriction, in which it is assumed that counterfactual neighborhood poverty is weakly lower under receipt of either kind of voucher than it would be if no voucher were received. This does not impose any restriction on the relationship between neighborhood choice under the two different kinds of vouchers.

To formalize this restriction, suppose that neighborhood minority is the sole endogenous neighborhood characteristic W, and let the random vector (W_0, W_1, W_2) denote a given individual's *potential* value of W from treatment assignment, or equivalently instrument value Z. Motivated by random assignment of the treatment, we assume that $(W_0, W_1, W_2) \perp\!\!\!\perp Z|X$. The observed value of the endogenous variable is $W = W_Z$. The instrument semi-monotonicity restriction can then be written as follows.

Restriction SMI: $W_0 \geq W_1$ and $W_0 \geq W_2$ almost surely.

As previously noted, the rationale for this restriction follows similar reasoning to the instrument monotonicity restriction used in Pinto (2019) in conjunction with random assignment of treatment. Here we adapt his logic to the present analysis, which differs in that

(i) the outcome of interest is ordinal rather than continuous, and (ii) the endogenous variable is continuous rather than discrete. The reasoning extends the well-known instrument monotonicity assumption used with a binary instrument in Imbens and Angrist (1994) and Angrist et al. (1996) to more general cases, in conjunction with revealed preference arguments.

Depending on the data, the inequalities delivered by the IV model presented in Section 3 may or may not be sufficient to sign the effect of the endogenous variable W on subjective well-being, whereas additionally imposing instrument semi-monotonicity can do this in a transparent fashion. To understand how, consider the probability of response $Y = 0$ conditional on X and Z corresponding to no voucher. Suppose one compares this to the same conditional probability holding the value of X fixed but now conditioning on Z corresponding to receipt of a voucher. If the second conditional probability is higher (lower), then, because Z is excluded from the outcome equation, the increase must be due to the effect of the change in voucher receipt on W. Since voucher receipt weakly lowers W for all households under Restriction SMI, this means that lower W, all else equal, leads to a higher (lower) conditional probability of $Y = 0$. Formally, if $\beta > 0$, then we have for $\tilde{z} \in \{1, 2\}$:

$$\begin{aligned}\mathbb{P}\left[Y = 0|X = x, Z = 0\right] &= \mathbb{P}\left[U \leq c_1 - W_0\beta - X\gamma|X = x, Z = 0\right] \\ &= \mathbb{P}\left[U \leq c_1 - W_0\beta - X\gamma|X = x, Z = \tilde{z}\right] \\ &\leq \mathbb{P}\left[U \leq c_1 - W_{\tilde{z}}\beta - X\gamma|X = x, Z = \tilde{z}\right] = \mathbb{P}\left[Y = 0|X = x, Z = \tilde{z}\right],\end{aligned}$$

where the second equality follows by random assignment conditional on X and the inequality follows since $W_0 \geq W_{\tilde{z}}$ under Restriction SMI. Therefore if we observe that

$$\mathbb{P}\left[Y = 0|X = x, Z = 0\right] > \mathbb{P}\left[Y = 0|X = x, Z = \tilde{z}\right], \tag{E.1}$$

we can conclude that $\beta \leq 0$. Similar reasoning applies to changes in the conditional probability of $Y = 2$, from which it follows that

$$\mathbb{P}\left[Y = 2|X = x, Z = 0\right] < \mathbb{P}\left[Y = 2|X = x, Z = \tilde{z}\right]. \tag{E.2}$$

also implies that $\beta \leq 0$. Thus, a researcher who imposes Restriction SMI together with conditional independence from random assignment, in addition to the IV assumptions of Section 3.1, can test whether (E.1) and (E.2) hold for all $x \in \text{Supp}(X)$ and all $\tilde{z}$.

In some of our empirical analysis W denotes both a measure of neighborhood poverty and the proportion of minority households in a neighborhood. Extensions of Restriction

SMI for multivariate W could also be considered. Without placing some restrictions on counterfactual values of the additional component(s) of W the inequalities derived above need not follow, and further care would need to be taken regarding assumptions on the impact of instrument Z on multivariate potential outcomes.

F Additional Computational Details

This section provides computational details further to those of Section 3.2.

F.1 Numerical Illustrations

Computations for numerical illustration of IV bounds discussed in Section 3.2 and reported in Table 1 were done by executing an R script using the nloptr package (Ypma (2018)) that wraps functionality in the nonlinear optimization package nlopt (Johnson (2007–2019)). The lower and upper bounds reported in Table 1 were computed by minimizing $p(\theta; y, x, w)$ and $ME(\theta; y, x, w)$, and $-p(\theta; y, x, w)$ and $-ME(\theta; y, x, w)$, respectively, subject to inequalities of the form (3.2) using values $s < t$ both in $\{-\infty, \Phi^{-1}(1/n), \Phi^{-1}(2/n), ..., \Phi^{-1}((n-1)/n), \infty\}$, as described in Section 3.2.

Minimization was done using the COBYLA algorithm of `nloptr`. At termination of the COBYLA algorihm there are typically a few inequalities that are violated by small amounts, of the order of $1e-9$. In the calculations reported in the paper the inequalities were adjusted by subtracting an amount $\varepsilon < 1e-5$ from the Gaussian probabilities on the right hand side of the inequalities (3.2) - (3.4). The amount subtracted varies with y_2 and y_1 and in most cases is $1e-6$. With this adjustment at the termination of COBYLA there are no violations of the original unadjusted inequalities. The adjustment has no effect on the bounds to the accuracy reported here.

For the sake of illustrating identified sets delivered by the data generating structures employed, the probabilities on the left hand side of (3.2) were calculated using the probability distribution of (Y, W) given $Z = z$ delivered by the structure employed in the numerical example, in which $X = 1$ is a constant. This expression, here denoted

$$\wp(s, t, x, z; \theta) \equiv \mathbb{P}\left[(s \leq c(Y, X, W; \theta) \wedge c(Y+1, X, W; \theta) \leq t) \,|X = x, Z = z\right],$$

where "$\wedge$" denotes "and" simplifies as follows depending on the values of s and t employed.

Case 1: $s = -\infty$, $t < \infty$.

$$\wp(-\infty, t, x, z; \theta) = \\ \mathbb{P}\left[(Y = 0 \wedge W\beta \geq c_1 - X\gamma - t) \vee (Y = 1 \wedge W\beta \geq c_2 - X\gamma - t) \,|X = x, Z = z\right]. \quad \text{(F.1)}$$

Case 2: $s > -\infty$, $t = \infty$.

$$\begin{aligned}&\wp(s,\infty,x,z;\theta) = \\ &\quad \mathbb{P}\left[(Y=1 \wedge W\beta \leq c_1 - X\gamma - s) \vee (Y = 2 \wedge W\beta \leq c_2 - X\gamma - s)\,|X=x, Z=z\right]. \quad \text{(F.2)}\end{aligned}$$

Case 3: $-\infty < s < t < \infty$.

$$\wp(s,t,x,z;\theta) = \mathbb{P}\left[Y = 1 \wedge c_2 - X\gamma - t \leq W\beta \leq c_1 - X\gamma - s|X=x, Z=z\right]. \quad \text{(F.3)}$$

These probabilities can be computed for any (s,t,x,z,θ) given the values of population parameters, making use of the CT structure used for these illustrations, in which

$$Y = \sum_{j=1}^{J} j \times 1\left[c_{0,j} < W\beta_0 + X\gamma_0 + U \leq c_{0,j+1}\right], \quad W = \delta_x + Z\delta_z + V,$$

$$(U,V) \sim \text{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & R \\ R & \Sigma_v \end{pmatrix}\right),$$

with population parameters $\theta_0 \equiv (c_{0,1}, c_{0,2}, \beta_0, \gamma_0, \delta_x, \delta_z, R, \Sigma_v)'$ taking values as specified on page 16 under "Numerical Illustration of IV Bounds".[4] Substituting for W in equation 2.1 for the determination of Y there is

$$Y = \left\{\begin{array}{ccc} 0 & , & V\beta_0 + U \leq c_{0,1} - X\gamma_0 - (\delta_x + Z\delta_z)\,\beta_0 \\ 1 & , & c_{0,1} - X\gamma_0 - (\delta_x + Z\delta_z)\,\beta_0 < V\beta_0 + U \leq c_{0,2} - X\gamma_0 - (\delta_x + Z\delta_z)\,\beta_0 \\ 2 & , & c_{0,2} - X\gamma_0 - (\delta_x + Z\delta_z)\,\beta_0 < V\beta_0 + U \end{array}\right\}.$$

Defining $(V_1, V_2) \equiv (V\beta_0 + U, V\beta)$ such that

$$(V_1, V_2) \sim \text{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \beta_0^2\Sigma_v + 2\beta_0 R + 1 & \beta\beta_0\Sigma_v + \beta R \\ \beta\beta_0\Sigma_v + \beta R & \beta_0^2\Sigma_v \end{pmatrix}\right),$$

and making use of $V \perp\!\!\!\perp (X,Z)$, the probabilities (F.1), (F.2) and (F.3) are as follows.

[4] Here the parameter vector $\theta = (c_1, c_2, \gamma\beta)$ whose identifed set is of interest comprises fewer elements than θ_0. This is because the true data generating process in these illustrations is from a CT structure specified by more parameters than the single equation IV model.

$$\wp(-\infty, t, x, z; \theta) =$$
$$\mathbb{P}\left[(V_1 \le c_{0,1} - x\gamma_0 - \delta_x\beta_0 - z\delta_z\beta_0 \wedge V_2 \ge c_1 - x\gamma - \delta_x\beta - z\delta_z\beta - t)\right]$$
$$+ \mathbb{P}\left[c_{0,1} \le V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \le c_{0,2} \wedge V_2 \ge c_2 - x\gamma - \delta_x\beta - z\delta_z\beta - t\right],$$

$$\wp(s, \infty, x, z; \theta) =$$
$$\mathbb{P}\left[c_{0,1} \le V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \le c_{0,2} \wedge V_2 \le c_1 - x\gamma - \delta_x\beta - z\delta_z\beta - s)\right]$$
$$+ \mathbb{P}\left[V_1 \ge c_{0,2} - x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \wedge V_2 \le c_2 - x\gamma - \delta_x\beta - z\delta_z\beta - s\right],$$

$$\wp(s, t, x, z; \theta) =$$
$$\mathbb{P}\left[c_{0,1} \le V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \le c_{0,2} \wedge c_2 - t \le V_2 + x\gamma + \delta_x\beta + z\delta_z\beta \le c_1 - s\right].$$

Since (V_1, V_2) are bivariate normal with parameters given in (F.1) these probabilities can be computed using standard software. In our numerical examples, such probabilities were calculated using the `pmvnorm` function in the R package `mvtnorm`, Genz et al. (2021), which additionally refers to Genz and Bretz (2009).

When $\beta = 0$ there are further simplifications, as follows:

$$\begin{aligned}
\wp(-\infty, t, x, z; \theta) &= 1\left[c_1 - x\gamma \le t\right] \cdot \mathbb{P}[Y = 0|x, z] + 1\left[c_2 - x\gamma \le t\right] \cdot \mathbb{P}[Y = 1|x, z], \\
\wp(s, \infty, x, z; \theta) &= 1\left[c_1 - x\gamma \ge s\right] \cdot \mathbb{P}[Y = 1|x, z] + 1\left[c_2 - x\gamma \ge s\right] \cdot \mathbb{P}[Y = 2|x, z], \\
\wp(s, t, x, z; \theta) &= 1\left[c_2 - t \le x\gamma \le c_1 - s\right] \cdot \mathbb{P}[Y = 1|x, z],
\end{aligned}$$

where

$$\mathbb{P}[Y = 0|x, z] = \Phi\left(\frac{c_{0,1} - x\gamma_0 - (\delta_x + z\delta_z)\,\beta_0}{\left(\beta_0^2\Sigma_v + 2\beta_0 R + 1\right)^{1/2}}\right)$$

$$\mathbb{P}[Y = 1|x, z] = \Phi\left(\frac{c_{0,2} - x\gamma_0 - (\delta_x + z\delta_z)\,\beta_0}{\left(\beta_0^2\Sigma_v + 2\beta_0 R + 1\right)^{1/2}}\right) - \Phi\left(\frac{c_{0,1} - x\gamma_0 - (\delta_x + z\delta_z)\,\beta_0}{\left(\beta_0^2\Sigma_v + 2\beta_0 R + 1\right)^{1/2}}\right)$$

$$\mathbb{P}[Y = 2|x, z] = 1 - \Phi\left(\frac{c_{0,2} - x\gamma_0 - (\delta_x + z\delta_z)\,\beta_0}{\left(\beta_0^2\Sigma_v + 2\beta_0 R + 1\right)^{1/2}}\right)$$

The above expressions for the case $\beta = 0$ are employed for all β such that $|\beta| < 0.00001$.

In the reported calculations for the numerical example there are no included exogenous variables X and no parameter γ. In this case the terms $X\gamma_0$ and $X\gamma$ are zero and may be removed.

F.2 Application to MTO

Computations using the MTO data were implemented by executing an R script (R Core Team (2022)) using the nonlinear optimization package nlopt linked through the nloptr package (Johnson (2007–2019), Ypma (2018)) for optimization and the rcpp and rcpparmadillo packages (Eddelbuettel and François (2011), Eddelbuettel (2013), Eddelbuettel and Sanderson (2014), Sanderson and Curtin (2016), Eddelbuettel and Balamuta (2017)) were used to employ C++ implementations of the most computationally intensive aspects. The full parameter search, target parameter search, and endpoint refinement steps of Algorithm 1, as well as the DR Bootstrap computations, each involve solving a large number of successive minimization problems inside a loop. These loops were executed in C++ rather than R for computational efficiency.

Simply computing the discrepancy function $\hat{Q}(\theta)$ defined in (3.5) at just a single value of θ requires first computing and then taking the maximum of the ratio of $4,485$ sample moments and standard errors. The steps described in Algorithm 1 entail computing $\hat{Q}(\theta)$ for a large number of different values of θ in order to solve the constituent constrained optimization problems. As is typical for set estimates and confidence sets using moment inequalities, and especially when there is such a large number of moment inequalities, implementation is computationally intensive. All computations reported here were carried out on a Dell Precision 3620 i7-6700 desktop with a 3.4 gigahertz processor and 16 gigabytes of memory. Total computation time for all eight sequences of results for conditional marginal effects reported in Tables 4 and 5 executed in parallel took just over 20 hours, with the first one having finished in just under 19 hours.[5] For counterfactual response probabilities, twelve sequences of computations were conducted in parallel, corresponding to the six sets of results reported in Table 6 along with six sets of results with neighborhood minority included as an

[5] A sequence of results is obtained by executing Algorithm 1 for a given choice of (i) whether β is restricted nonpositive or nonnegative, (ii) whether the marginal effect under consideration is for the response $y = 0$ or $y = 2$, and (iii) whether the neighborhood minority variable is included or excluded from the specification. The eight sequences for which computation time is reported refers to all such combinations. For counterfactual response probabilities, the twelve sequences referenced correspond to all possible configurations of (i) counterfactual response $y = 0$, $y = 1$, or $y = 2$, (ii) whether neighborhood poverty is fixed at the NYC median or one standard error below, and (iii) whether the neighborhood minority variable is included or excluded.

additional endogenous variable.[6] All six sequences of computations corresponding to results reported in Table 6 were completed in a little over two days, and all twelve sequences were finished in a little more than three days.

To give a rough reflection of the computational complexity involved, note that each sequence of computations reported comprised roughly between 129,000 and 153,000 evaluations of $\hat{Q}(\theta)$ at different values of θ, not including bootstrap computations. For the eight sequences of computations for the conditional marginal effects a total of $1,161,946$ evaluations of $\hat{Q}(\cdot)$ were executed, each one comprising a maximum $4,485$ studentized moment functions. Implementation of the DR bootstrap additionally requires repeated computation of the the maximization problem (F.7) and the bootstrap statistic in (F.9) below. Computations for counterfactual response probabilities required substantially more computation of bootstrap critical values, which is apparently what resulted in the longer execution time reported above.

Details of the DR Bootstrap procedure are provided below, followed by a table that provides a rough outline of key functions employed for computations described in Algorithm 1. The code used to carry out these computations can be found at `https://github.com/adammrosen/MTO-Replication`.

Discard Resampling Bootstrap

Computation of the DR Bootstrap critical value is implemented in the C++ function BootstrapCV, which takes r, the hypothesized value of $g(\theta)$ as an argument. We follow the steps described by Belloni et al. (2018) on pages 12–13 and define the bootstrap process

$$\hat{v}^*_{\theta,j} \equiv n^{-1/2} \sum_{i=1}^{n} \xi_i \{\omega_j(Y, W, X, Z; \theta) - \hat{m}_j(\theta)\}, \tag{F.4}$$

where $\{\xi_i : i = 1, ..., n\}$ denote i.i.d. standard normal bootstrap draws independent of the data. We further define

$$\hat{\Theta}(r) \subseteq \{\theta \in \Theta(r) : \hat{Q}(\theta) = \hat{T}(r)\}, \tag{F.5}$$

which is a set of values of θ such that $g(\theta) = r$ and for which the discrepancy function $\hat{Q}(\theta)$ attains the value of the profile discrepancy at r, $\hat{T}(r)$. For the results reported in Tables

[6] To save on space IV estimates and confidence sets for counterfactual response probabilities with the neighborhood minority variable included are not reported. By construction, they produce larger sets than those obtained with the neighborhood minority variable excluded as was the case with conditional marginal effects.

4 – 6 we specify $\hat{\Theta}(r)$ as the singleton set $\{\hat{\theta}\}$, where $\hat{\theta}$ is the value of θ achieved in the constrained minimization defining $\hat{\Theta}(r)$ in (3.5). From a computation standpoint this is the easiest choice of $\hat{\Theta}(r)$, but employing a larger collection of values of θ will produce (weakly) smaller critical values. Finally,

$$\widehat{\Psi}_{\theta} \equiv \{j \in [J] : \sqrt{n}\hat{m}_j(\theta)/\hat{\sigma}_j(\theta) \geq \max_{\tilde{j} \in [J]} \sqrt{n}\hat{m}_{\tilde{j}}(\theta)/\hat{\sigma}_{\tilde{j}}(\theta) - M_n\} \tag{F.6}$$

denotes the indices of the studentized moments that come within distance M_n of achieving the maximum value $\hat{Q}(\theta)$. Here M_n is an appropriately chosen sequence that diverges to ∞ as $n \to \infty$. Such a sequence is also used in Bugni et al. (2017), but as BBC18 explain, in a many moment inequality setting it is required additionally that $M_n/\bar{w}_n \to \infty$, with $\bar{w}_n$ as specified in BBC equation (4.2). We follow their recommendation for approximating $\bar{w}_n$ by approximating with $\bar{w}_n^*$ the $1 - \gamma_n$ quantile of

$$\sup_{\theta \in \Theta(r), j \in [J]} |\hat{v}_{\theta,j}^*|, \tag{F.7}$$

and in order to ensure $M_n/\bar{\omega}_n \to \infty$ we set

$$M_n = \log(n) \cdot \bar{w}_n^*. \tag{F.8}$$

Finally, using (F.4) – (F.7), the DR bootstrap test statistic is defined as

$$R_n^{DR*} \equiv \inf_{\theta \in \widehat{\Theta}(r)} \max_{j \in \widehat{\Psi}_{\theta}(M_n)} \hat{v}_{\theta,j}^*. \tag{F.9}$$

For a given r the DR bootstrap test statistic critical value $c_n^{DR}(r, \alpha)$ is then computed by first taking B bootstrap samples of independent standard normal variates $\xi_1, ..., \xi_n$ and computing the bootstrap process (F.4). For each bootstrap sample the following steps are then conducted:

1. Compute $\bar{\omega}_n$ by solving the maximization problem in (F.7) and set $M_n = \log(n) \cdot \bar{\omega}_n$.

2. Compute R_n^{DR*} defined in (F.4).

Once these steps are finished the DR critical value $c_n^{DR}(r, \alpha)$ is set to the $1 - \alpha$ quantile of R_n^{DR*} in the B bootstrap iterations. For results presented here, $B = 99$ was used.

Functions Referenced in Algorithm 1

Function MinDiscrepancyCpp($\tilde{\Theta}$)
$\forall \theta_s \in \tilde{\Theta}$ minimize $\hat{Q}(\theta)$ using θ_s as starting value.
return for each $\theta_s \in \tilde{\Theta}$
– Vector θ^* at which minimization terminated.
– Discrepancy value $\hat{Q}(\theta^*)$ achieved.
– Target parameter value $g(\theta^*)$.
End Function

Function ProfileDiscrepancyOnGridCpp($\mathcal{G}$, runDR)
for each $r \in \mathcal{G}$ **do** Compute $\hat{T}(r) \leftarrow \min_\theta\{\hat{Q}(\theta) : g(\theta) = r\}$
    if (runDR & $\hat{T}(r) > 0$) **then**
        Compute DR critical value $c_n^{DR}(r, \alpha)$ by calling BootstrapCV(r)
    end if
end for
return for each $r \in \mathcal{G}$
– Profile discrepancy value $\hat{T}(r)$
– Vector θ such that $\hat{Q}(\theta) = \hat{T}(r)$
– DR critical value $c_n^{DR}(r, \alpha)$
End Function

Function RefineBoundCpp($r_{\text{out}}, r_{\text{in}}$, fromLower, Δ)
Construct a grid $\mathcal{G}$ from r_{out} to r_{in} in increments of Δ.
if (fromLower) **then**
    return $\min\{r \in \mathcal{G} : \hat{T}(r) \leq c\}$
else
    return $\max\{r \in \mathcal{G} : \hat{T}(r) \leq c\}$
end if
End Function

Function BootstrapCV(r)
for each $b = 1, ..., B$ **do**
    Compute $\bar{\omega}_n$ and set $M_n = \log(n) \cdot \bar{\omega}_n$
    Compute $R_n^{DR*} = \inf_{\theta \in \widehat{\Theta}(r)} \max_{j \in \widehat{\Psi}_\theta(M_n)} \hat{v}^*_{\theta,j}$
end for
return $1 - \alpha$ quantile of R_n^{DR*}
End Function

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455.

Angrist, J. D. and J.-S. Pischke, eds. (2009): *Mostly Harmless Econometrics*, Princeton, NJ: Princeton University Press.

Belloni, A., F. Bugni, and V. Chernozhukov (2018): "Subvector Inference in PI Models with Many Moment Inequalities," ArXiv:1806.11466.

Bugni, F., I. Canay, and X. Shi (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 8, 1–38.

Chesher, A., A. Rosen, and Z. Siddique (2022): "Estimating Endogenous Effects on Ordinal Outcomes," Working paper, Duke, University of Bristol, and UCL.

Chesher, A. and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989.

Eddelbuettel, D. (2013): *Seamless R and C++ Integration with Rcpp*, New York: Springer, iSBN 978-1-4614-6867-7.

Eddelbuettel, D. and J. J. Balamuta (2017): "Extending extitR with extitC++: A Brief Introduction to extitRcpp," *PeerJ Preprints*, 5, e3188v1.

Eddelbuettel, D. and R. François (2011): "Rcpp: Seamless R and C++ Integration," *Journal of Statistical Software*, 40, 1–18.

Eddelbuettel, D. and C. Sanderson (2014): "RcppArmadillo: Accelerating R with high-performance C++ linear algebra," *Computational Statistics and Data Analysis*, 71, 1054–1063.

Genz, A. and F. Bretz (2009): *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, Heidelberg: Springer-Verlag.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn (2021): *mvtnorm: Multivariate Normal and t Distributions.*, R package version 1.1-3.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Johnson, S. G. (2007–2019): "The NLopt nonlinear-optimization package," `http://github.com/stevengj/nlopt`.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2012): "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults," *Science*, 337, 1505–1510.

Pinto, R. (2019): "Noncompliance as a Rational Choice: A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects," Working paper, UCLA.

R Core Team (2022): *R: A Language and Environment for Statistical Computing.*, R Foundation for Statistical Computing, Vienna, Austria.

Sanderson, C. and R. Curtin (2016): "Armadillo: a template-based C++ library for linear algebra." *Journal of Open Source Software*, 1.

Ypma, J. (2018): "Introduction to nloptr: an R interface to NLopt," `https://cran.r-project.org/web/packages/nloptr/vignettes/nloptr.pdf`.