

Estimating Endogenous Effects on Ordinal Outcomes*

Andrew Chesher[†]

Adam M. Rosen[‡]

UCL and CeMMAP

Duke University and CeMMAP

Zahra Siddique[§]

University of Bristol

June 22, 2023

Abstract

We examine the use of instrumental variable (IV) methods to measure the effect of a ceteris paribus change in an endogenous variable on an ordered outcome. Specifically, we use these methods to investigate the effect of neighborhood characteristics on subjective well-being (SWB) among participants in the Moving to Opportunity (MTO)

*We are grateful for comments from the editor and two reviewers, Alex Belloni, Federico Bugni, and seminar and conference participants at Erasmus University Rotterdam, Duke, Simon Fraser, Leicester, Johns Hopkins, and Western Ontario, a 2018 joint CeMMAP/Northwestern conference on incomplete models and the 2020 EALE-SOLE-AASLE conference. Xinyue Bei and Muyang Ren provided excellent research assistance. 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. This is the final version of this paper produced for publication in *The Review of Economics and Statistics* where it is forthcoming.

[†]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.

housing voucher experiment. We find that the estimated positive effect of a decrease in neighborhood poverty on SWB is sensitive to the specification of the first stage auxiliary equation for endogenous neighborhood poverty. Our results highlight the influential role of control function restrictions in complete triangular models.

Keywords: Instrumental Variables, Ordered Choice, Incomplete Models, Partial Identification, Neighborhood Effects, Subjective Well-Being, Moving to Opportunity.

JEL classification: C25, C26, C35, I31, R2.

1 Introduction

Consider a community of individuals each possessing a set of observable characteristics, some of which are thought to bear on their well-being. Economists often aim to infer *ceteris paribus* effects that answer a question of the following type: If a policy were implemented that were to exogenously shift the value of one of these characteristics, all else equal, how would this affect subsequent values of a measurable outcome of interest?

An essential qualifier in the study of *ceteris paribus* effects on an outcome is that the outcome be *measurable*. There is no obvious unit of measurement, for instance, when the outcome is happiness of individuals. A commonly used measure of happiness is self-reported subjective well-being (SWB), elicited by individuals' responses to a query asking them to place themselves in one of an ordinal set of categories. For example, the General Social Survey asks, "Taken all together, how would you say things are these days – would you say that you are happy, pretty happy, or not too happy?" While the ranking of possible responses to such a question is clear, there is no unique measure of happiness to compare across individuals. We propose a path to addressing this issue in this paper; while our focus is on the use of ordinal SWB as the outcome, the same issues arise with other ordinal outcomes to which our analysis is also equally applicable.

Two complications need to be taken into account. The first is related to measurement of ordinal SWB. Comparisons of average happiness across different populations are not invariant

to monotonic transformations of the measures of happiness used; Schröder and Yitzhaki (2017) and Bond and Lang (2019) show that happiness data generally do not satisfy the properties required for these comparisons, or in the case of Schröder and Yitzhaki (2017), that linear regression coefficients do not satisfy the properties required to be robust to monotonic rescalings of the ordinal outcomes.¹ An alternative approach is given by Chen et al. (2019), who suggest the use of quantiles and quantile regression, since quantiles are equivariant under monotonic transformations. While a sound alternative, their methods do not enable measurement of ceteris paribus effects in the presence of potentially endogenous explanatory variables which is our goal. Kaplan and Zhou (2019) also study the problem of how ordinal data may be used to draw comparisons between the distributions of two latent continuous variables; this is an interesting endeavor, but once again an altogether different goal from that of this paper.

We circumvent these problems by using nonlinear models that respect the ordinal nature of the outcome to measure quantities that have a natural real-world interpretation. We do not aim to map responses onto a single scale of happiness on which to draw comparisons, due to the inherent subjectivity of how such a scale might be created. Instead, we consider questions about how individuals' responses to the questions actually asked would change in a counterfactual scenario. The first type of question we consider concerns marginal effects for the response variable itself. In the context of our application, the question is: "All else fixed, what is the marginal effect of a change in local neighborhood poverty on the probability that a head of household reports that they are happy (or not too happy)?" The second type of question is a counterfactual probability induced by a discrete shift in household circumstances, rather than a local shift: "Suppose that the poverty rate in the local community in which a household is situated were exogenously changed to a given level. What then would be the counterfactual probability that such a household, if asked, would respond that they were happy (or not too happy)?" These are coherent questions which do

¹Like ordinary least squares, linear IV estimates are weighted averages of the ordinal outcomes, and thus also sensitive to monotonic transformations of the scale on which these variables are placed.

not require the construction of a happiness measure beyond what is reported by households, and allow for the possibility that households interpret the questions in different ways.

A second complication is potential endogeneity. We expect individuals to make decisions that improve their well-being. Observable covariates chosen by households should therefore not be independent of unobservable determinants of well-being. For instance, when examining the effect of neighborhood poverty on SWB using the Moving to Opportunity (MTO) data, the neighborhood poverty rate should be treated as endogenous. Households that believe they stand to benefit from moving to a lower poverty neighborhood are more likely to do so when given the opportunity.

We review point-identifying, complete nonlinear models for ordered outcomes that may be used, such as ordered probit and logit, in Section 2. These models restrict covariates to be independent of unobserved heterogeneity. One way to allow for endogeneity is to employ a complete triangular model, hence referred to as the “CT” model. CT models specify the determination of endogenous covariates as a function of exogenous covariates, instruments, and unobservable heterogeneity. This can be done by maintaining the ordered probit equation for the ordinal variable, and augmenting it with additional linear equations for each endogenous covariate featuring additive unobservables, all of which are normally distributed. This generalizes a complete triangular model for binary outcomes studied by e.g. Heckman (1978) and Rivers and Vuong (1988). This model provides a complete specification for the determination of all endogenous variables. All parameters are point identified under mild conditions, and can be consistently estimated by a two stage procedure or maximum likelihood. Marginal effects and counterfactuals are smooth functions of these parameters, and standard approaches can be employed for asymptotic inference.²

The CT model is fully parametric, and, as is always the case in a structural analysis, the inferences drawn rely on the suitability of restrictions imposed by the model. One may

²Furthermore, although our focus here is on conditional marginal effects and counterfactual response probabilities, CT models can sometimes justify estimation of average structural functions and average partial effects using control functions, as considered e.g. by Blundell and Powell (2003) and Wooldridge (2014).

wonder how sensitive empirical findings are to relaxation of these restrictions. In Section 3 a nonlinear instrumental variable (IV) model is considered in which the auxiliary equations for the endogenous variables are dropped. The IV model is incomplete because it does not uniquely pin down the value of endogenous variables as a function of exogenous observed and unobserved variables. Such models are generally partially identifying, and results from Chesher and Rosen (2017) are applied to characterize the resulting identified sets.

Section 4 describes our application to the study of the effect of neighborhood poverty on SWB using MTO data. We compare estimates using different models in Section 4. The CT model point identifies marginal effects and counterfactual probabilities, while partially identifying IV models provide an analysis of the sensitivity of inference to restrictions imposed on the determination of endogenous neighborhood characteristics. Section 5 concludes.³

Related Literature

Our methodology broadly pertains to applications with ordinal outcome variables, although our focus in this paper is to data on SWB. Determinants of SWB, or happiness, have been of interest to both academics and policy makers in recent years as evidenced by the inception of the first *World Happiness Report* (Helliwell et al. (2012)) commissioned for the UN Conference on Happiness in April 2012, and which has been followed by world happiness reports each year since except 2014. Surveys that include references to the larger literature on happiness research in economics and the wide variety of topics studied within this literature include Stutzer and Frey (2010) and MacKerron (2012).

We focus attention on the effect of neighborhood characteristics on SWB among economically disadvantaged individuals eligible for MTO. The effect of neighborhood characteristics on SWB may be difficult to isolate due to the presence of unobservable factors that effect individuals' SWB – such as drug addiction or gang membership – that may also play a role in the individual's choice of neighborhood. Treating neighborhood characteristics as exogenous

³Several appendices are provided in an online supplement, covering details of bound derivations for the IV model presented in Section 3.1, linear model and first stage CT model estimates that do not appear in the main text, a discussion of the identifying power of a semi-monotone IV assumption, and computational details further to those provided in the main text.

variables in the determination of SWB may, therefore, be unjustified. MTO data offer a unique opportunity to exploit random variation in housing possibilities – and neighborhood choice in particular – through random assignment of housing vouchers.

MTO has been the focus of many previous studies; two studies to which our application is most closely related are those by Ludwig et al. (2012) and Pinto (2019). Ludwig et al. (2012) also focus on SWB, but instead using OLS and linear IV. Like Pinto (2019), our focus is on inferring neighborhood effects. Specifically, Pinto (2019) cleverly uses the incentives of the MTO experimental design in conjunction with revealed preference reasoning to identify average effects of a discrete classification of neighborhood poverty level on continuous and binary labor market outcomes. Unlike Pinto’s analysis – which did not analyze SWB – our approach employs structural models for ordinal outcomes. As previously discussed, when outcomes are ordered discrete such as SWB, average effects are not easily interpreted, lending motivation to our IV approach. Moreover, in contrast to Pinto (2019), our approach allows and indeed utilizes a continuously varying endogenous variable. In Pinto (2019) a monotonicity restriction is applied to a discretization of the endogenous neighborhood poverty variable into three categories, used to define a finite set of types. His model and our IV model are thus non-nested and offer complementary approaches to estimating the effects of neighborhood poverty on different types of outcomes.⁴

Although our pursuit of counterfactual probabilities and marginal effects differs from the ITT and 2SLS estimands that are the focus of Ludwig et al. (2012), our results using a CT model continue to suggest an increase in SWB from a reduction in neighborhood poverty. However, we find this result to be sensitive to the specification of the first stage auxiliary equation for the endogenous variable, in this case neighborhood poverty. When using an IV model, the empirical results are compatible with a wide range of effects and we can no longer rule out a *decrease* in SWB from a reduction in neighborhood poverty.

⁴The IV approach used here exploits instrument exclusion and independence restrictions with a structural model for an ordinal outcome. In Appendix E we illustrate how the addition of a monotonicity restriction similar to that considered by Pinto (2019) in combination with our IV model can potentially help to sign the effect of the endogenous variable.

The sensitivity analyses of Section 4 employ bound characterizations obtained by application of results in Chesher and Rosen (2017) to the models studied here. The IV ordered probit type specification without any additional restrictions falls in the class of models considered earlier by Chesher and Smolinski (2012), but for the fact that here the endogenous variable is continuously distributed. The bound characterization delivered by application of the Chesher and Rosen (2017) analysis features many moment inequalities. Methods employing many moment inequalities for inference on parameter vectors have been recently developed by Chernozhukov et al. (2019) (CCK19) and Bai et al. (2019). We employ methods from CCK19 as well as Belloni et al. (2018) (BBC18), which build on ideas in CCK19 to conduct inference on functions of partially identified parameter vectors.⁵

Summary of Contributions

Our main contributions are summarized as follows. First, we propose the use of structural modeling to study the effect of exogenous changes in endogenous variables when the outcome is an ordered variable lacking cardinal interpretation. We do this by focusing on changes induced in measurable quantities such as the probability of a given ordinal response, rather than on ad hoc cardinal scales onto which ordinal outcomes can be mapped. Second, we advocate for and illustrate the use of alternative modeling assumptions to investigate the sensitivity of empirical findings to modeling restrictions chosen by the researcher. Our third contribution is the detailed implementation of new inference methods from BBC18 to functions of parameters satisfying the bound characterization delivered by IV models for ordered outcomes using the Chesher and Rosen (2017) partial identification analysis. This apparently constitutes the first application of inference methods of BBC18 to data, as well as the first empirical application of IV models for ordered outcomes.⁶ A fourth contribution lies in the empirical contribution to MTO. Our analysis examines how neighborhood characteris-

⁵A working paper version of CCK19 appeared on arXiv in 2013.

⁶Chesher and Rosen (2020b) provide an application to binary outcome data on female employment using inference methods from Chernozhukov et al. (2013). Subsequent to this work, Chesher and Rosen (2020a) and Chesher et al. (2021) also use methods from CCK19 and BBC18 in applications to market structure featuring interdependent binary choices, and an application of IV methods for Tobit models to tobacco expenditure, respectively.

the interval

$$[c_y - w\beta - x\gamma, c_{y+1} - w\beta - x\gamma),$$

where $c_0 \equiv -\infty$, and $c_{J+1} \equiv \infty$. These intervals partition the real line according to each possible value of Y , and their probabilities correspond to likelihood contributions of the standard maximum likelihood estimator. Under the usual rank condition the expected value of $1/n$ times the log likelihood is uniquely maximized at the population parameter values. Estimation is easily carried out in modern software packages such as STATA, but the model does not allow for endogenous variables. In our application, choice of neighborhood and therefore neighborhood characteristics may be correlated with unobservable heterogeneity, so the required independence restriction may not be credible.

2.2 A CT Model

We now consider CT models that allow for potential endogeneity of W . We consider the same functional form for the ordered outcome given in (2.1) for the OP model, where again X is a vector of observed exogenous variables, and U is an unobserved exogenous variable. As in the OP model, U is restricted to be normal with mean zero and unit variance. It will not, however, be restricted to be independent of endogenous variables W .

With the components of random k_w -vector W allowed to be correlated with U , additional restrictions on the determination of W can play an important role for identification. We begin by considering a complete model that specifies how $W \equiv (W_1, \dots, W_{k_w})$ is determined as a function of X , instruments Z , and additional unobservable variables V :

$$W_k = X\delta_x^k + Z\delta_z^k + V_k, \quad k = 1, \dots, k_w, \quad (2.2)$$

where each $V_k \in \mathbb{R}$ is an unobserved random variable.

The vector of unobservables (U, V_1, \dots, V_{k_w}) is restricted to be independent of (X, Z) and

distributed multivariate normal with mean zero and variance

$$\Sigma = \begin{pmatrix} 1 & R \\ R' & \Sigma_v \end{pmatrix},$$

where Σ_v denotes the variance of $V = (V_1, \dots, V_{k_w})$, assumed nonsingular. Component r_k of $R \equiv (r_1, \dots, r_{k_w})$ denotes the covariance of U with V_k . Both Σ_v and R comprise unknown parameters, while $\Sigma_{11} = 1$ imposes the same scale normalization as in the OP model.

The triangular model is complete because realization of the observed and unobserved exogenous variables X , Z , U , and V uniquely determines the realization of the endogenous variables Y and W . It thus remains that the conditional distribution of endogenous variables given observed exogenous variables is uniquely determined as a function of model parameters.

Taking (2.1) and (2.2) together with multivariate normality of (U, V) ,

$$\Pr[Y = y|x, z, w] = \Phi\left(\frac{c_{y+1} - w\beta - x\gamma - R\Sigma_v^{-1}v(w, x, z)}{\sigma(v)}\right) - \Phi\left(\frac{c_y - w\beta - x\gamma - R\Sigma_v^{-1}v(w, x, z)}{\sigma(v)}\right), \quad (2.3)$$

where $v(w, x, z) \equiv (w_1 - x\delta_x^1 - z\delta_z^1, \dots, w_K - x\delta_x^K - z\delta_z^K)'$, $\sigma(v) \equiv 1 - R\Sigma_v^{-1}R'$, and $\Phi(\cdot)$ denotes the standard normal CDF. Consequently, it can be shown that under standard rank conditions there is point identification of all model parameters $\beta, \gamma, R, \Sigma_v$ and δ_x^k and δ_z^k for each k . Estimation can proceed by way of a two stage procedure using estimated residuals from (2.2) obtained in a first stage as regressors in a second stage ordered probit regression that also includes observations of W and X , which generalizes the procedure developed by Rivers and Vuong (1988) for binary outcome models with endogenous variables. Algebraic manipulation of the second stage estimates can be used to consistently estimate all model parameters. Alternatively, (2.3) can be used as a basis for estimation by maximum likelihood.

Point estimators using either the two stage procedure or maximum likelihood are easy to compute. Marginal effects are also point identified and easily estimated. The model allows

such effects to be heterogeneous for individuals with different observable characteristics. The model explicitly accounts for the ordered nature of the outcome variable, but also relies on the specification for the determination of the endogenous variables W as a parametric function of exogenous variables with a normal unobservable as in (2.2). In Section 3 implications of the model are investigated in the absence of such a restriction.

2.3 Counterfactual Probabilities and Marginal Effects

Define the individual response function

$$\mathbf{y}(x, w, u, \theta) \equiv \sum_{j=1}^J j \times 1 [c_j < w\beta + x\gamma + u \leq c_{j+1}],$$

denoting for any individual the value of the ordered outcome y that would be chosen when faced with given values of (x, w, u) . The function $\mathbf{y}(\cdot, \cdot, u, \theta) : \text{Supp}(X, W) \rightarrow \{0, 1, \dots, J\}$ denotes the response function of an individual with unobservable u to values of (x, w) . For the sake of counterfactual analysis these are used to consider what would happen if a randomly selected individual in the population (or a randomly selected individual from the subpopulation with a given set of covariates x) were to have their value of w or x or both exogenously set to an alternative value, holding their value of u fixed.

Counterfactual probabilities and marginal effects can be expressed as functions of components of parameter vector θ through use of the ordered outcome equation (2.1) in conjunction with the restrictions $U \perp\!\!\!\perp (X, Z)$ and $U \sim \mathcal{N}(0, 1)$. The counterfactual probability that a person with observable characteristics $X = x$ randomly drawn from that subpopulation would achieve SWB $y \in \{0, \dots, J\}$ if their neighborhood characteristics were exogenously shifted to w is given by

$$p(\theta; y, x, w) \equiv P[\mathbf{y}(x, w, U, \theta) = y | X = x] = \Phi(c_{y+1} - w\beta - x\gamma) - \Phi(c_y - w\beta - x\gamma), \quad (2.4)$$

where $c_0 = -\infty$ and $c_J = \infty$. Marginal effects attributable to local changes in w are obtained as partial derivatives of these probabilities, or the corresponding finite differences for discrete components if W is discrete. For example, in a model with a single continuous endogenous variable W , the marginal effect with respect to w is given by

$$ME(\theta; y, x, w) \equiv \frac{\partial p(\theta; y, x, w)}{\partial w} = \beta(\phi(c_y - w\beta - x\gamma) - \phi(c_{y+1} - w\beta - x\gamma)). \quad (2.5)$$

In the OP and CT models laid out in this section parameters $\beta, \gamma, c_1, \dots, c_J$ are point identified under mild conditions. The counterfactual probabilities and marginal effects in (2.4) and (2.5) are known smooth functions of these parameters. Thus, plugging in consistent and asymptotically normal point estimators for $\beta, \gamma, c_1, \dots, c_J$ results in consistent and asymptotically normal estimators for counterfactual probabilities and marginal effects, with asymptotic variances obtainable by way of the delta method.

3 IV Models

In this section we maintain the specification (2.1) for the determination of the ordered outcome Y , but without assuming the “first stage” specification (2.2). We continue to assume that the unobservable U is a standard normal random variable independent of (X, Z) but – crucially – place no further restriction on its joint distribution with W . The analysis extends the nonparametric IV model for ordered outcomes studied by Chesher and Smolinski (2012). Here we consider a parametric version of their model, which generalizes the OP model commonly used in the absence of covariate endogeneity, and which allows for W to be either discrete or continuously distributed. We describe how the analysis of Chesher and Rosen (2017) can be applied to partially identify structural parameters, and we briefly describe how to obtain set estimates and confidence sets for marginal effects and counterfactual choice probabilities using these IV models.

3.1 Generalized Instrumental Variable Moment Inequalities

Despite the parametric specification in (2.1) and the normal distribution of U , this model is incomplete because it is silent as to the determination of endogenous W . As a result, the joint distribution of (Y, W) conditional on exogenous variables (X, Z) is no longer pinned down by knowledge of the distribution of U and model parameters. The model does however carry observable implications for the conditional distribution of (Y, W) in the form of conditional moment inequalities.

To see how such implications can be derived, recall that the ordered response specification (2.1) ensures that unobservable U lies in the interval $(c_Y - X\gamma - W\beta, c_{Y+1} - X\gamma - W\beta]$. Consider for the sake of argument any fixed interval $[s, t]$ on the real line. Then

$$\{ [c_Y - X\gamma - W\beta, c_{Y+1} - X\gamma - W\beta] \subseteq [s, t] \} \implies \{ U \in [s, t] \}. \quad (3.1)$$

That is, when the interval from $c_Y - X\gamma - W\beta$ to $c_{Y+1} - X\gamma - W\beta$ is contained in $[s, t]$, then U must be contained in $[s, t]$. The first event implies the second, so the conditional probability of the former event provides a lower bound on the conditional probability of the latter.

Application of Theorems 3 and 4 of Chesher and Rosen (2017) builds on this logic to characterize sharp bounds on model parameters that can be used for the construction of bound estimates. Specifically, the following inequality for all s, t pairs with $s \leq t$ characterizes sharp bounds on $\theta \equiv (\beta, c_1, c_2, \gamma, \sigma)$.⁷

$$\max_{x,z} \mathbb{P}[(s \leq c(Y, X, W; \theta) \wedge c(Y + 1, X, W; \theta) \leq t) | X = x, Z = z] \leq \Phi(t) - \Phi(s), \quad (3.2)$$

⁷Here s and t are elements of the extended real line, inclusive of $\pm\infty$ with the understanding that when $s = -\infty$ the inequality $s \leq c(y, x, w; \theta)$ always holds, as does the inequality $c(y + 1, x, w; \theta) \leq t$ when $t = +\infty$, even when $c(y, x, w; \theta)$ and $c(y + 1, x, w; \theta)$ are themselves $-\infty$ and $+\infty$, respectively, with $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$.

where “ \wedge ” denotes the logical “and” operator, and

$$\forall j \in \{1, \dots, J\} : c(j, x, w; \theta) \equiv c_j - x\gamma - w\beta, \quad c(0, x, w; \theta) \equiv -\infty, \quad c(J+1, x, w; \theta) \equiv \infty.$$

The inequality (3.2) is obtained by applying the conditional probability $\mathbb{P}[\cdot | X = x, Z = z]$ to events in (3.1). The right hand side of these inequalities employs the normal CDF because U is restricted to be standard normal, independent of (X, Z) .⁸

It is interesting to compare moment inequalities of the form (3.2) to the implications of the OP and CT models.⁹ The OP and CT models are complete, meaning that the value of all endogenous variables is uniquely determined as a function of exogenous observed and unobserved variables. Consequently these inequalities strengthen to equalities for specific values of s and t . In the OP model, the space of unobservable U can be partitioned into intervals that uniquely deliver each possible value of Y given realizations of X and W , and the probability that U lies in each of these intervals is known because U is standard normally distributed. For the collection of s and t pairs that correspond precisely to these intervals, we must have that the probability of the events on the left of (3.2) sum to one, and so must those on the right of (3.2), by virtue of the these $[s, t]$ intervals comprising a partition of \mathbb{R} . Therefore, these inequalities across this particular collection of s and t must hold with *equality*. The resulting equalities are in fact the conditional probabilities that comprise the OP likelihood function. In the CT model for any realization of (W, X, Z) , \mathbb{R} can also be partitioned into intervals for U that uniquely determine the outcome. The conditional probability that U is in each of these regions conditional on (W, X, Z) is a known function of parameters, and the resulting equalities are precisely those of (2.3) that can be used for maximum likelihood estimation in the triangular model.

⁸As long as $U \perp\!\!\!\perp (X, Z)$ is maintained, these inequalities hold with the CDF of U in place of the Gaussian CDF $\Phi(\cdot)$. For instance, if U were restricted to have the logit distribution, (3.2) would continue to hold, but with $\Phi(\cdot)$ replaced by the logit CDF $\Lambda(\cdot)$. In this case the same approach for estimation and inference employed in this paper could be used, simply with $\Lambda(\cdot)$ in place of $\Phi(\cdot)$.

⁹Inequalities on (conditional) probabilities are (conditional) moment inequalities because the probability of an event is the expectation of a binary indicator of that event.

In contrast, in the incomplete IV model there is no partition of \mathbb{R} into intervals of values for U that uniquely determine the endogenous variables, given exogenous variables (X, Z) . The observable implications take the form of inequalities in (3.2) that do not reduce to equalities, but whose components are nonetheless easily computed. The terms on the left hand side are probabilities involving the joint distribution of observable variables, while the probabilities on the right hand side of the inequalities are simply normal probabilities.

Corollary 3 and Theorem 6 of Chesher and Rosen (2017) provide further characterizations of parameter bounds when the assumption that U is normally distributed is relaxed. If U is assumed independent of exogenous variables (X, Z) but with an unknown distribution, Corollary 3 implies for example that parameters θ satisfy

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

for all pairs $s < t$ on the extended real line. When the stochastic independence restriction is weakened to only require that U is median independent of (X, Z) – with conditional median normalized to zero – then Theorem 6 implies that the identified set for θ are those that satisfy the inequalities

$$\max_{x,z} \mathbb{P}[c(Y + 1, X, W; \theta) \leq 0 | X = x, Z = z] \leq \frac{1}{2}, \quad (3.3)$$

$$\max_{x,z} \mathbb{P}[c(Y, X, W; \theta) \geq 0 | X = x, Z = z] \leq \frac{1}{2}. \quad (3.4)$$

These characterizations enable bound estimation for model parameters in the IV model without requiring a parametric specification for the unobservable variable U . The weakening of assumptions relative to the IV model with a parametric distributional restriction will however result in (weakly) wider bounds. Further details of the derivation of these bounds are provided in Appendix A.

3.2 Computing Sets for Counterfactuals and Marginal Effects

The expressions (2.4) and (2.5) are known functions of θ , and are both point identified in the OP and CT models. When instead the IV model is used, the expressions for counterfactual probabilities and marginal effects remain valid, but θ is in general only partially identified. Consequently, identified sets for counterfactual probabilities and marginal effects are characterized as the set of all possible values for (2.4) and (2.5), respectively, taking θ across all values in the identified set. In Section 4.3 we report set estimates and confidence sets for marginal effects using such a characterization when employing IV models, thus providing a sensitivity analysis for the CT model point estimates presented in Section 4.2. Before doing so, we now demonstrate the ability of the IV model to bound counterfactual probabilities and marginal effects with a numerical example. This illustrates the bounds obtained at the population level, purely in terms of identification analysis, in an ideal but infeasible setting in which the distribution of observable variables is known exactly. We then describe how computation of set estimates and confidence intervals is carried out with sample data before presenting our application in Section 4.

Numerical Illustration of IV Bounds

Table 1 illustrates bounds for counterfactual probabilities and marginal effects as described in (2.4) and (2.5) from a population in which the conditional distributions of (Y, W) given (X, Z) are generated using a CT structure as described in Section 2.2 with $X = 1$, $Z \in \{-1, 1\}$, and parameters:

$$c_1 = -0.5, \quad c_2 = 0.8, \quad \gamma = 0, \quad \beta = 1, \quad \delta_x^1 = 0, \quad \delta_z^1 = 0.5, \quad R = 0, \quad \Sigma_v \in \{1.00, 0.01\}.$$

The bounds shown were obtained by inequalities of the form (3.2) using intervals $[s, t]$ with endpoints in $\{-\infty, \Phi^{-1}(1/n), \Phi^{-1}(2/n), \dots, \Phi^{-1}((n-1)/n), \infty\}$.¹⁰ In all cases the bounds

¹⁰The interval $[-\infty, \infty]$ is omitted and n is set to 50 which results in $n(n+1)/2 - 1 = 1274$ inequalities at each of two values of Z and so 2548 inequalities in total. Doubling n leads to only slightly shorter identified

are informative. The binary instrumental variable Z is a much more accurate predictor of the value of W when $\Sigma_v = 0.01$ than when $\Sigma_v = 1$, resulting in much shorter identified sets. As Σ_v is driven further towards zero (not shown here), the IV model approaches a state in which it delivers point identification because when $\Sigma_v = 0$, exogenous Z is in fact a perfect predictor of W . The ability of the instrument to induce variation in W is helpful in achieving more informative bounds.

The covariance parameter R is zero in these illustrations, so in fact W is exogenous. Models incorporating this knowledge are point identifying, but that is not assumed in the IV model employed here, so this example is informative about the cost of not knowing an explanatory variable is exogenous when in fact it is.

Discrepancy Function, Set Estimates, and Critical Values

To put the moment inequalities (3.2) into a form convenient for inference using techniques from CCK19 and BBC18, note that they can be equivalently expressed as

$$\mathbb{P}[(s \leq c(Y, X, W; \theta) \wedge c(Y + 1, X, W; \theta) \leq t) \wedge X = x \wedge Z = z] - p(x, z) (\Phi(t) - \Phi(s)) \leq 0,$$

where as before “ \wedge ” denotes logical “and”, and $p(x, z)$ denotes $\mathbb{P}[X = x \wedge Z = z]$. The sample analog of the expression on the left is, for any $(s, t, x, z) = (s_j, t_j, x_j, z_j)$:

$$\hat{m}_j(\theta) \equiv \hat{E}_n[\omega_j(Y, W, X, Z; \theta)] = \frac{1}{n} \sum_{i=1}^n \omega_j(Y_i, W_i, X_i, Z_i; \theta),$$

$$\omega_j(Y, W, X, Z; \theta) \equiv (1[s_j - c_Y \leq -X\gamma - W\beta \leq t_j - c_{Y+1}] - (\Phi(t_j) - \Phi(s_j))) \cdot 1[X = x_j, Z = z_j].$$

To form moments indexed by $j = 1, \dots, J$ for estimation and inference we enumerate all possible combinations of x_j, z_j combined with pairs of values s_j, t_j as follows. The unit interval $[0, 1]$ is divided into K evenly spaced intervals. Pairs (s_j, t_j) are selected as the

sets. Details of the calculation are in the online supplementary material.

standard normal quantile function evaluated at the endpoints of all intervals that comprise unions of these $1/K$ length intervals, resulting in $K(K+1)/2$ pairs, less one for the removal of the uninformative pair $(-\infty, \infty)$. Crossed with 15 different points of support for (X, Z) , this yields $J = 15 \left(\frac{K(K+1)}{2} - 1 \right)$ moments. In the results reported in Section 4.3, $K = 24$ was used resulting in $J = 4,485$ moments.

We use the discrepancy function $\hat{Q}(\cdot)$ and profile discrepancy function $\hat{T}(\cdot)$ defined as:

$$\begin{aligned} \hat{Q}(\theta) &\equiv \max_{j \in [J]} \sqrt{n} \frac{\hat{m}_j(\theta)}{\hat{\sigma}_j(\theta)}, & \hat{T}(r) &\equiv \min_{\theta: g(\theta)=r} \hat{Q}(\theta), \\ \hat{\sigma}_j^2(\theta) &\equiv \hat{E}_n [(\omega_j(Y, W, X, Z) - \hat{m}_j(\theta))^2]. \end{aligned} \tag{3.5}$$

Here $[J] \equiv \{1, \dots, J\}$, and $g(\theta)$ denotes the function of interest, a counterfactual response probability or conditional marginal effect as defined in (2.4) and (2.5), respectively. Set estimates and confidence intervals are of the form

$$\left\{ r : \hat{T}(r) \leq c_n(r) \right\}, \tag{3.6}$$

for different choices of $c_n(r)$. For analog set estimates, the critical value is simply $c_n(r) = 0$ for all r . For inference on $g(\theta)$ we employ two different kinds of critical values.

The first type of critical value is a self-normalized (SN) critical value given by

$$c_n^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/J)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/J)^2/n}} \tag{3.7}$$

for nominal coverage probability $1 - \alpha$ provided by CCK19. The self-normalized critical values generally provide conservative asymptotic inference, but have the advantage that they do not depend on the value r under consideration and are easy to compute.

The second type of critical value we use for inference is a *Discard Resampling* (DR) bootstrap critical value. DR critical values for subvector inference using moment inequalities were introduced by Bugni et al. (2017) and adapted for use with many moment inequalities

by BBC18. We denote an asymptotic $1 - \alpha$ DR critical value as $c_n^{DR}(r, \alpha)$. Due to the use of the bootstrap and the need to recompute the critical values for each r , obtaining DR critical values takes substantially more computation time. However, because it accounts for the covariance structure of the sample moment inequalities that are close to binding, it generally results in less conservative inference.¹¹

Section 4.3 reports sets of the form (3.6) for critical values $c_n(r) \in \mathcal{C}_r$, where \mathcal{C}_r denotes the cutoff values for testing $g(\theta) = r$, i.e. $\mathcal{C}_r \equiv \{0, c_n^{SN}(0.50), c_n^{SN}(0.05), c_n^{DR}(r, 0.05)\}$. The steps used to compute these sets are outlined in Algorithm 1. The sample discrepancy function $\hat{Q}(\cdot)$ defined in (3.5) is not smooth, so the full parameter search step used 1000 random starting values sampled uniformly over the parameter space in addition to a small set of deterministic starting values.¹² For each target parameter $g(\theta)$ a separate search was conducted over regions of the parameter space in which β is nonpositive and nonnegative.¹³ Our implementation was executed in R, with computationally intensive components written in C++. Further computational details are provided in Appendix F.¹⁴

4 Application to MTO

MTO was a unique randomized housing voucher experiment implemented from 1994 to 1998. Through the program, 4,604 volunteer families living in “extreme-poverty” neighborhoods

¹¹BBC18 also consider an alternative *profile resampling* bootstrap critical value, similar in spirit to that of BCS, but also adapted to achieve asymptotic validity with many moment inequalities. Relative to self-normalized critical values, sufficient conditions for bootstrap critical values established by BBC18 to guarantee asymptotic size control are more stringent. We do not verify these conditions here. We note in particular that BBC18 employ convexity of the set $\{\theta \in \Theta : g(\theta) = r\}$ as a sufficient condition. This condition is satisfied for counterfactual probabilities for $Y = 0$ or $Y = 2$, but not for $Y = 1$ or for conditional marginal effects. Alternative sufficient conditions may be possible. We leave this to future investigation, noting that this condition is not needed to ensure validity of self-normalized critical values.

¹²Specifically, in addition to the randomly chosen starting values, nine deterministic starting values were used based on point estimates from the CT and OP model, as well as bound analysis using fewer (810) moment inequalities employed in an earlier draft.

¹³The sign of β dictates whether the thresholds $c_j - w\beta - x\gamma$ are increasing or decreasing in the value w of the endogenous variable. From results in for instance Chesher (2013) and Chesher and Rosen (2020b) for binary outcome models and Chesher and Smolinski (2012) for ordered outcome models, it is known that regions of the parameter space that correspond to different orderings of values of such thresholds across different values of the endogenous variable can be disconnected.

¹⁴Code for replication is available at <https://github.com/adammrosen/MTO-Replication>.

Algorithm 1 Computation of set estimates and confidence sets.

- 1: **Initialization:** Set $g(\cdot)$, K & parameter spaces Θ and $\mathcal{R} \equiv [\underline{g}, \bar{g}]$ for search.
 - 2: Load data. Compute $\Phi(t_j) - \Phi(s_j)$ and $1[X_i = x_j; Z_i = z_j]$, all $j = 1, \dots, J$.
 - 3: Set a collection of deterministic and random starting values $\tilde{\Theta} \subseteq \Theta$.
 - 4:
 - 5: **Full Parameter Search:** Call `MinDiscrepancyCpp`($\tilde{\Theta}$) to obtain $\hat{\Theta} \subseteq \{\text{argmin } \hat{Q}(\theta) : \theta \in \Theta\}$.
 - 6:
 - 7: **Target Parameter Search:** Find min/max values of $r \in \mathcal{R}$ such that $\hat{T}(r) \leq c$.
 - 8: **for** each $c \in \mathcal{C}_r$ **do**
 - 9: Set g_ℓ, g_u to min/max values of $g(\theta)$ from full parameter search such that $\hat{Q}(\theta) \leq c$.
 - 10: Set \mathcal{G} : a grid from g_ℓ to \underline{g} and g_u to \bar{g} with distance $\delta = 0.005$ between points.
 - 11: Call `ProfileDiscrepancyOnGridCpp`(\mathcal{G}) to compute $\hat{T}(r)$ for each $r \in \mathcal{G}$.
 - 12: **end for**
 - 13:
 - 14: **Endpoint Refinement:** Refine min/max values of r found to satisfy $\hat{T}(r) \leq c$.
 - 15: **for** each $c \in \mathcal{C}_r$ **do**
 - 16: Set r_ℓ, r_u to min/max values of $\{r = g(\theta) : \hat{T}(r) \leq c\}$ from target parameter search.
 - 17: Call `RefineBoundCpp`($r_\ell - \delta, r_\ell, \text{fromLower} = \text{T}, \Delta = 0.0005$).
 - 18: Call `RefineBoundCpp`($r_u + \delta, r_u, \text{fromLower} = \text{F}, \Delta = 0.0005$).
 - 19: **end for**
 - 20:
 - 21: **Report Results:** Report min/max r found to satisfy $\hat{T}(r) \leq c$, each $c \in \mathcal{C}_r$.
-

in Baltimore, Boston, Chicago, Los Angeles, and New York were randomly assigned to one of three treatments: receipt of a low poverty experimental housing voucher (which could be used only if the family moved to a low poverty area), a traditional section 8 voucher (which could be used without any location restriction), or no voucher (control group).¹⁵ Assignments are recorded as $Z = 2$, $Z = 1$, and $Z = 0$, respectively. Unlike traditional vouchers, low-poverty vouchers could only be applied toward housing in census tracts that had 1990 poverty rates below 10%. The vouchers made it more feasible for recipients to move, and in particular the low-poverty vouchers encouraged them to move to less distressed neighborhoods. Not all recipients chose to move, but many did.

Random assignment of housing vouchers in the experiment justifies the use of assignment Z as an IV for neighborhood characteristics. The SWB outcome we use is taken from long-

¹⁵“Extreme-poverty” neighborhoods are those in which at least 40% of residents’ income lies below the U.S. federal poverty threshold.

term data recorded 10-15 years after assignment. This outcome is categorical, with three allowable answers to the following question, taken from the General Social Survey (GSS): “Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?”¹⁶ The corresponding outcome is $Y \in \{0, 1, 2\}$, ranging from least to most happy.¹⁷ The responses are ordered, but do not have a cardinal interpretation. Only MTO adults were asked this question.

Families that were in the program were extremely economically disadvantaged; see Appendix B for details on the set of covariates which were elicited in a baseline survey before randomization took place in 1994-1998. The majority of household heads were minorities and less than 40% had completed high school. More than three quarters reported that one of the top two reasons for wanting to move was to get away from gangs and drugs.

Previous studies of MTO include Kling et al. (2007), Ludwig et al. (2012), Pinto (2019), and Chetty et al. (2016). Kling et al. (2007) used medium term (4-7 years after assignment) outcome data to measure the effect of the program on participants. They found mixed results of the effect of the program on traditional objective measures of well-being. No significant effects were found on adult economic self-sufficiency or physical health outcomes. Substantial mental health benefits were found for adults and female youths, but adverse effects were found for male youths. Chetty et al. (2016) subsequently studied the long-term impacts of treatment on the economic outcomes of those who were young children at the time of random assignment, finding significant effects that are decreasing in the child’s age at the time of assignment. As noted in the introduction, Pinto (2019) developed a model that combines revealed preference analysis with monotonicity assumptions motivated from the incentives of the MTO experiment to measure average treatment effects of a change in neighborhood on various labor market outcomes. This enabled estimation of effects of neighborhood transitions rather than of voucher assignment using a different approach than

¹⁶Further details about the GSS are available at <http://www3.norc.org/GSS+Website/>.

¹⁷Specifically, the values of Y correspond to 0 for ‘not too happy’, 1 for ‘pretty happy’ and 2 for ‘very happy’.

is taken here, for different (continuous and binary) outcomes than the one considered in our application. The measured effects aligned qualitatively with those of Kling et al. (2007), but in contrast produced statistically significant estimates for important labor market outcomes such as total income when the effect of the neighborhood was isolated.

Ludwig et al. (2012) revisited MTO using long-term data on outcomes recorded 10-15 years after assignment. They used a linear model to measure ITT effects on various outcomes Y . This provides a measure of the effect of offering a voucher on the outcome. Ludwig et al. (2012, p.1508) concluded that, "...the opportunity to move through MTO had mixed (null to positive) long-term effects on objective measures of well-being of the type that have been the traditional focus of the neighborhood effects literature." In previous work Ludwig and coauthors showed that MTO had significant long-term effects on some important long-term health outcomes, specifically extreme obesity and diabetes. Relating specifically to SWB, Ludwig et al. (2012) wrote that their paper includes, "the first time the effect of neighborhoods on SWB has been assessed in an experimental analysis." They found significant effects of neighborhood characteristics on SWB. These conclusions however rely on linear model estimates generally sensitive to the scale on which SWB is measured, and which do not enable computation of counterfactual probabilities or their marginal sensitivity to endogenous variables.¹⁸ Following Ludwig et al. (2012), the neighborhood characteristics we examine are also measures of neighborhood poverty and residential share minority, denoted W_{pov} and W_{min} , respectively; Appendix B discusses how these characteristics are constructed and their distribution across different treatment groups.

The following subsections report empirical results using a sequence of progressively less restrictive models to estimate the effect of neighborhood characteristics on SWB. Subsection 4.1 provides results using the ordered probit (OP) model, making no allowance for endo-

¹⁸In the supplementary materials of Ludwig et al. (2012) it is noted that "As a sensitivity analysis we also relax this assumption and re-estimate equations (S2) and (S3) using instrumental variables probit following the control function approach from Rivers and Vuong and obtain qualitatively similar results." Unfortunately no further details on these estimates are provided in the main paper or in the supplementary materials.

generity. Subsection 4.2 reports estimates using the CT model of Section 2 that allows for endogeneity of neighborhood characteristics. Subsection 4.3 reports results obtained using the IV model. Appendix C reports results using linear models as previously used in Ludwig et al. (2012).

4.1 OP Model Estimates

Columns (1)-(3) of Table 2 report parameter estimates for the OP model with only dummy variables for randomization site used as exogenous variables X . The base site is New York, and the positive coefficient estimates indicate that households in other cities tended to report higher SWB. Neighborhood characteristics W are also restricted to be exogenous. Columns (4)-(6) report parameter estimates after inclusion of a complete set of covariates; we find that the parameter estimates remain qualitatively unchanged.

The solid curves in Figures 1 and 2 depict OP estimates of the predicted probabilities that a person reports being ‘not too happy’ ($y = 0$) or ‘very happy’ ($y = 2$) across different values of the neighborhood characteristics. Dashed lines present the corresponding estimates for the CT model, discussed in Section 4.2. These are all conditional on the adult being in NY. Other exogenous explanatory variables X are held fixed at the NY sample median values. Thus the predicted probabilities are conditional on the adult being a white, Hispanic female with age between 41 and 45 who is not married (among other characteristics).¹⁹ Focusing now on the OP estimates, Figure 1 shows that the probability of being ‘not too happy’ for such a person increases with neighborhood poverty W_{pov} and the probability of being ‘very happy’ decreases with W_{pov} . Figure 2 illustrates that this also holds true when W_{pov} is changed while holding neighborhood minority W_{min} constant at the NY sample median. Figure 1 also shows that the probability of being ‘not too happy’ is increasing in W_{min} and the probability of being ‘very happy’ is declining in W_{min} . However, Figure 2 illustrates that the effect of changes in W_{min} on SWB is relatively small once W_{pov} is held constant at the

¹⁹In unreported results we also examine the predicted probabilities and marginal effects which do not condition on baseline covariates and find that the pattern of estimates is very similar.

NY sample median, so that the predicted probability is almost flat across changes in W_{min} .

Figures 3 and 4 present the corresponding marginal effects (2.5). Again focusing on the OP estimates shown by the solid curves, these figures indicate that W_{pov} has a negative effect on SWB, with and without holding W_{min} constant. Also from these figures the effect of W_{min} on SWB is less clear; there is some negative effect without controlling for W_{pov} in Figure 3 but these effects are close to zero once W_{pov} is held constant at the NY sample median as may be seen in Figure 4.²⁰

The OP estimates are broadly in line with the signs and statistical significance of effects measured using linear models (see Appendix C), despite the difference in interpretation afforded by the linear model. There is a negative statistically significant effect on the probability of being ‘very happy’ with a unit increase in W_{pov} ; conversely a positive statistically significant effect on the probability of being ‘not too happy’ with a unit increase in W_{pov} . Relative to linear model estimates, the nonlinear model estimates enable construction of heterogeneous and interpretable counterfactual probabilities and marginal effects.

It is important to keep in mind that these OP estimates continue to be sensitive to the cardinal scale used for categorical SWB outcomes. Bond and Lang (2019) show that simply transforming the underlying MTO long term happiness data using a transformation that applies a slightly more right-skewed distribution than a standard log-normal reverses the result that recipients of MTO experimental vouchers were happier than those in the control group. This is because individuals in the MTO control group have lower mean SWB, but also a higher variance in SWB. Since variances across experimental and control groups are unequal, a necessary condition for rank order identification under normality is not satisfied.

²⁰In unreported results we also estimate ordered logit and multinomial logit models on the data to construct predicted probabilities and marginal effects. We find that these models give very similar patterns of predicted probabilities and marginal effects as the ordered probit.

4.2 CT Model Estimates

The restriction requiring that neighborhood characteristics are exogenous is now relaxed, and estimates are reported for the CT model presented in Section 2. Endogenous variables W are the same as before; in all cases estimation is carried out by maximum likelihood.²¹

Parameter estimates for the outcome equation (2.1) are given in Table 3.²² Neighborhood poverty W_{pov} has a negative and statistically significant effect across all specifications. Neighborhood minority W_{min} has a negative but statistically insignificant effect when it is included in the absence of W_{pov} . When both variables are included, W_{min} has a positive effect on SWB, although it is just statistically significant at the 10% level when both neighborhood characteristics are included and only site indicators are used as exogenous explanatory variables.

The finding that the effect of W_{min} is not statistically significantly different from zero when W_{pov} is not included accords with the linear IV estimates reported in Appendix Table C.1. Likewise, so do the findings that W_{pov} has a negative impact on SWB, and that when both neighborhood characteristics are included W_{min} has a (borderline significant) positive impact. The OP model estimates reported in the previous section, which did not allow for endogeneity of neighborhood characteristics, did not produce a positive effect for W_{min} .

Like the OP model, however, the CT model enables further investigation of the effect of endogenous variables on SWB through the consequent formulas for counterfactual probabilities and marginal effects, while also accounting for endogeneity. The dashed curves in Figures 1 and 2 give predicted probabilities using the CT model for different values of neighborhood characteristics, again conditional on the adult being in NY, and having values of exogenous explanatory variables X that correspond to the NY sample median values.²³

²¹In unreported results we also estimate model parameters using an alternative control function approach. The results are close to those obtained using maximum likelihood, but estimators obtained by maximum likelihood are more asymptotically efficient.

²²Estimates from the first stage are reported in Table D.1 in Appendix D.

²³As in the case of the OP, in unreported results we also examine the predicted probabilities and marginal effects which do not condition on baseline covariates and find that the pattern of estimates is very similar.

Figure 1 shows that, as expected, the probability of being ‘not too happy’ for such a person is increasing in W_{pov} and the probability of being ‘very happy’ is decreasing in W_{pov} . Figure 1 also indicates that the probability of being ‘not too happy’ is increasing in W_{min} and the probability of being ‘very happy’ decreasing with W_{min} . Note that the specifications used in the top panels of Figure 1 include *only* W_{pov} and those in the bottom panels of Figure 1 include *only* W_{min} as endogenous variables.

Figure 2 further investigates the effect of these variables when both are included as endogenous variables, where the value of the variable whose effect is not being illustrated is held fixed at the NY median value in the sample. Figure 2 shows that the direction of change in the probability of SWB taking on different values across values of W_{pov} remains the same as in Figure 1. Compared to Figure 1, Figure 2 provides a different picture for the effect of W_{min} on SWB. With W_{pov} held fixed, the probability of being ‘not too happy’ is decreasing in W_{min} and the probability of being ‘very happy’ is flat at very low levels and then increasing in W_{min} . The CT estimates indicate a greater degree of heterogeneity in the effect of these variables on counterfactual response probabilities than do the OP estimates.

Figures 3 and 4 show the estimated conditional marginal effects, again with the CT model estimates indicated by the dashed curves. As before the Figures indicate a clear negative effect of higher neighborhood poverty W_{pov} on SWB. The effect of W_{min} on SWB, as also indicated in Table 3 is found to be slightly negative but statistically insignificant without holding W_{pov} fixed as in Figure 3. Yet in Figure 4 when both variables are included with W_{pov} fixed at the sample median, the marginal effects indicate that SWB is generally increasing in W_{min} with varying levels of statistical significance.

What then can be concluded from these results further to what can be learned using linear IV? To answer this question, consider what can be inferred from Figures 1 - 4 about the impact of shifts in exogenous neighborhood characteristics, both in sign and *magnitude*. The figures are in agreement in indicating that shifts in policy that serve to *lower* the poverty rate can have a positive effect on the well-being of households in affected neighborhoods.

This neighborhood effect is separate from any direct effect that particular households in that neighborhood may experience from their own increase in earnings. Not only are the results clear as to the direction of the effect but they also allow us to measure its magnitude, and to observe that the size of the effect of a change in neighborhood poverty will in general differ with the level at which the effect is measured. That is, the model allows for estimation of different effects at different poverty levels. Such information is useful to policy makers who must decide where to distribute public resources for the best possible impact.

In any application, the virtues of the CT model relative to the linear IV model rely on the suitability of the restrictions embodied in the structural equations (2.1) and (2.2) and the joint normality of unobserved variables. In order to investigate, information matrix (IM) tests of the CT specifications were conducted. These IM procedures test for equality of components of the expected outer product of score and expected negative Hessian forms of the information matrix whose inverse is the asymptotic variance of ML estimates of model parameters.²⁴

For specifications both including and excluding the neighborhood minority variable W_{min} , without additional exogenous regressors, the test was carried out by comparing the two forms of the IM matrix diagonal terms associated with the coefficients on W_{pov} , the coefficient on the dummy variable for residence in Baltimore, and the off diagonal cross-derivative term. The p-values for these tests were 0.0576 and 0.0083. The test was also computed using all derivative and cross-derivative terms associated with the coefficients on W_{pov} and all city dummy variables for the specification in which W_{min} is omitted, with a resulting p-value of 0.0554.

The results in hand indicate that the IM test statistic for the CT specification is at best near the margin of the rejection region at conventional testing levels, and possibly well inside the rejection region, e.g. for the case where the neighborhood minority variable is included.

²⁴See White (1982) for details of the information matrix test. Implementation as described in Chesher (1983) was used, in which it was shown that the test statistic can be computed as n times the R^2 of a least squares regression of a vector of ones on first and second order derivatives of the log density, see also Chesher (1984) for an interpretation of the test in terms of uncontrolled parameter heterogeneity.

Accordingly some exploration of the sensitivity of the estimates to relaxation of the CT specification is warranted. The next section sets out a brief investigation of the sensitivity of some of the empirical findings to the removal of the first stage specification and the joint normality restriction.

4.3 Sensitivity Analysis with IV Models

Table 4 reports set estimates and confidence intervals for the conditional marginal effect of the neighborhood poverty index on the conditional probability that a head of household in New York responds that they are “not too happy” ($y = 0$) and that they are happy ($y = 2$) using the IV model described in Section 3. The model maintains the specification for Y in equation (2.1) with $U \sim N(0, 1)$, but without any specification for the determination of endogenous W as in equation (2.2) in the CT model. The set estimates and confidence intervals are compared to point estimates and confidence intervals obtained using the more restrictive CT specification, the results of which were presented in Section 4.2. These quantities are functions of parameter vector θ defined in (2.5), comprising the threshold parameters c_1 and c_2 , the four dimensional parameter γ multiplying randomization site indicators (with base site taken to be New York), and the coefficients on neighborhood characteristics W . It is seven dimensional in specifications in which only neighborhood poverty is included in W , and eight dimensional when share neighborhood minority is also included.

The conditional marginal effects depend additionally on the particular value of the endogenous right hand side variable(s) at which they are evaluated. The values reported here are obtained with the endogenous variables fixed at their sample median in New York.²⁵ The CT estimates and confidence intervals are reported both with and without the inclusion of additional exogenous variables. When included, the marginal effects are also measured

²⁵The observed median values are approximately -0.20 for neighborhood poverty and 0.49 for neighborhood minority. The same values are used for results reported in application of the OP and CT models for estimation of marginal effects and counterfactual response probabilities. The median poverty rate is slightly sensitive to the treatment of observations in which SWB is missing, of which there were two such observations in New York. These were kept in the sample for the sake of computing the median.

conditional on their median values, as in Section 4.2. Results reported for the IV model use only city dummies as included exogenous variables for computational tractability.

Computations were carried out as described in Section 3.2. Recall that all reported sets are of the form $\{r : \hat{T}(r) \leq c_n(r)\}$ for a given $c_n(\cdot)$. Analog set estimates are obtained by setting $c_n(\cdot) = 0$, corresponding to those values of the conditional marginal effect for which there is some θ such that all sample inequalities $\hat{m}_j(\theta) \leq 0$ hold, equivalently such that $\max_j \hat{m}_j(\theta) \leq 0$. In finite samples application of the max operator to a noisy sample estimate will tend to result in an overestimate of $\hat{m}_j(\theta)$. Consequently, the set of θ such that $\max_j \hat{m}_j(\theta) \leq 0$ will tend to be smaller than the set of θ such that $\max_j m_j(\theta) \leq 0$, where each $m_j(\theta)$ is a corresponding population moment, i.e. the analog set estimate will be inward biased. The corrected set estimates in Table 4 adjust for this by setting $c_n(\cdot) = c_n^{SN}(0.50)$ such that the asymptotic probability that the maximum (minimum) value in the set estimate for $g(\theta)$ is less (more) than its population value is no greater than one half. That is, with probability at least one half asymptotically, each endpoint estimate is no tighter than its population target. Such a correction is a half-median-unbiased set estimate, as proposed in Chernozhukov et al. (2013). The set (3.6) with this value of $c_n(\cdot)$ provides an asymptotic 50% confidence set for the conditional marginal effect. The third set estimate provided is a 95% confidence set obtained using the self-normalized critical value $c_n(\cdot) = c_n^{SN}(0.05)$ from (3.7). The fourth set estimate provided uses the discard resampling (DR) bootstrap critical value, $c_n(\cdot) = c_n^{DR}(\cdot, 0.05)$ discussed in Section 3.2.

Consider first the results from the IV specification reported in Table 4 in comparison to those obtained from the complete triangular (CT) specification. When computing set estimates and confidence sets for marginal effects, separate searches were conducted over the region in which $\beta \leq 0$ and that in which $\beta \geq 0$ with the results reported in Columns (3) and (4), respectively. The region in which $\beta \leq 0$ is that in which neighborhood poverty has a zero or negative effect on SWB, and that in which $\beta \geq 0$ is that in which neighborhood poverty has a zero or positive effect on SWB. In principle the restrictions of the IV model

can be sufficient to sign β , and hence the conditional marginal effect, but with the MTO data at hand this is not the case. Using the inequalities implied by the IV model, the set estimates and confidence sets obtained comprise the union of those in columns (3) and (4). A researcher who finds it reasonable to assume that the effect of neighborhood poverty on counterfactual SWB is negative ($\beta \leq 0$) can simply take the IV model set estimates and confidence regions to be those reported in column (3).²⁶

For the marginal effect of neighborhood poverty on respondents answering they are either “not too happy” or “happy” we see that the IV interval estimates are fairly wide. For the marginal effect on respondents answering “not too happy”, the CT point estimates are near the lower bound of the IV point estimates in column (3). Similarly, the marginal effect on respondents answering they are “happy” is measured as slightly negative using the CT model, but can range over a much wider domain under the IV specification, all the way to -0.258 according to the analog estimate. Thus, if one doubts the veracity of the second equation of the CT model, one must consider the possibility that the effect of neighborhood poverty on happiness could be substantially stronger than what the CT model implies. Moreover, the effect of sampling variation on both the CT and IV estimates is not negligible, and the corrected IV interval estimates and 95% confidence intervals are substantially wider than both the IV analog estimates and the CT 95% confidence intervals.

The results reported in Table 5 are for specifications in which the neighborhood minority index is included as an additional exogenous variable. By construction these result in larger intervals than those reported in Table 4. This is because the specification that excludes this variable can be viewed as the special case of the more general specification in which the coefficient on neighborhood minority is fixed at zero. Once again the IV analog interval esti-

²⁶There are alternative routes to signing the effect of β . Note that, by contrast, the CT model produces point estimates and confidence sets that imply that the effect of neighborhood poverty on SWB is negative. The “first-stage” specification of the CT model implies that the direction of the effect of voucher receipt on the poverty level of a household’s neighborhood in a given city is the same for all households. That is, the CT model embeds an instrument monotonicity restriction as in Imbens and Angrist (1994). In Appendix E we discuss how the addition of such a restriction to the IV model can also serve to sign the effect of neighborhood poverty on SWB.

mates always include the CT point estimates, although now both intervals include marginal effects of zero in addition to effects much larger in magnitude than those obtained from the CT model.²⁷ The ability to measure this marginal effect to any reasonable accuracy thus seems especially sensitive to the inclusion of the second equation in the CT specification.

Table 6 reports point and set estimates and 95% confidence regions on counterfactual response probabilities using the CT and IV specifications, the latter imposing the restriction $\beta \leq 0$ throughout. The probabilities in Table 6 correspond to response probabilities for individuals in New York that would be obtained by exogenously shifting the endogenous neighborhood poverty index to the median level in New York (columns (1)-(3)) and to one standard error below the median level in New York (columns (4)-(6)), respectively. This presents results for the subpopulation of individuals with exogenous covariates X at the same values as used for Table 4, using specifications in which the neighborhood minority rate was not included. In all but one case the analog estimates from the IV model contain the CT model point estimates, with the sole exception that the point estimate of 0.166 in column (5) lies barely outside the lower bound of the IV set estimate in column (6). The sensitivity analysis afforded by the IV model generally accords with the CT results, but indicates that without the additional restrictions used in the CT specification, the ranges of possible values of counterfactual probabilities are much wider in magnitude. For these probabilities, the IV specification is not as informative. However, bounds and confidence intervals on the probability that individuals answer they are in the highest happiness category indicate that the CT estimates for this probability are close to the lower end of what is indicated by the IV specification. Thus, if the additional restrictions of the CT specification are incorrect, it could be that the CT specification substantially *under estimates* the fraction of individuals that would report they were happy at these neighborhood poverty levels, in particular for the counterfactual in which neighborhood poverty is lowered by one standard error.

²⁷Note that while the IV bound estimates would generally be expected to contain the CT point estimates, this need not occur if the CT model is misspecified.

5 Conclusion

In this paper we considered the use of nonlinear instrumental variable models for ordered outcomes to measure marginal effects and counterfactual probabilities. In the context of research on happiness data, it has recently been shown that comparisons of means across populations and the use of OLS estimates are problematic with ordinal outcomes. The inherent problem is that the methods are sensitive to the cardinal scale on which the ordered outcome is measured. There is no natural cardinal measure for happiness. The use of linear model IV estimation methods intended to deal with endogeneity are similarly problematic.

Instead, estimators employing nonlinear IV models may be used that respect the ordinal nature of the outcome data. This enables the measurement of *ceteris paribus* effects, which are often of interest to economists, and useful for studying the impact of exogenous changes. With these methods, researchers need not impose a cardinal scale for the ordered outcome.

We demonstrated the use of nonlinear IV models in an application to data from the MTO housing voucher experiment. Point estimates of structural parameters as well as marginal effects and counterfactual probabilities for reported household happiness induced by changes in neighborhood poverty were provided using a CT specification. As is the case with any structural model, the results rely on the restrictions employed by the model used. Thus, we turned to consideration of partially identifying IV models that nested the CT specification, and allowed some degree of investigation of the sensitivity of the structure of the triangular model to relaxation of the auxiliary equation for the endogenous variable. In the absence of the complete specification provided by the CT model the data have substantially less to say about the magnitudes of marginal effects and counterfactual probabilities. This analysis highlights the under-appreciated power of the often-used control function restrictions that are embodied in the CT model.

Recently, there have been many studies of happiness, and there are other contexts in which outcomes are measured on an inherently ordinal scale. Often, there may be endogenous variables, and IV methods are called for. This paper has presented some methods that can

be used in such contexts, which are compatible with ordinal outcome data. There is ample scope for application and further development of such methods.

References

- BAI, Y., A. SANTOS, AND A. M. SHAIKH (2019): “A Practical Methods for Testing Many Moment Inequalities,” BFI working paper 2019-116.
- BELLONI, A., F. BUGNI, AND V. CHERNOZHUKOV (2018): “Subvector Inference in PI Models with Many Moment Inequalities,” ArXiv:1806.11466.
- BLUNDELL, R. AND J. L. POWELL (2003): “Endogeneity in Nonparametric and Semiparametric Regression Models,” in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Cambridge University Press, vol. 2, 312–357.
- BOND, T. N. AND K. LANG (2019): “The Sad Truth About Happiness Scales,” *Journal of Political Economy*, 127, 1629–1640.
- 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.
- CHEN, L.-Y., E. OPARINA, N. POWDTHAVEE, AND S. SRISUMA (2019): “Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales,” IZA DP No. 12152.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2019): “Inference on Causal Parameters Using Many Moment Inequalities,” *Review of Economic Studies*, 86, 1867–1900.

- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81, 667–737.
- CHESHER, A. (1983): “The Information Matrix Test: Simplified Calculation via a Score Test Interpretation,” *Economics Letters*, 13, 45–48.
- (1984): “Testing for Neglected Heterogeneity,” *Econometrica*, 52, 865–872.
- (2013): “Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification,” *Econometric Theory*, 29, 231–266.
- CHESHER, A., D. KIM, AND A. M. ROSEN (2021): “IV Methods for Tobit Models,” CeMMAP working paper CWP26/21.
- CHESHER, A. AND A. M. ROSEN (2017): “Generalized Instrumental Variable Models,” *Econometrica*, 85, 959–989.
- (2020a): “Econometric Modeling of Interdependent Discrete Choice with Applications to Market Structure,” CeMMAP working paper CWP25/20.
- (2020b): “Generalized Instrumental Variable Models, Methods, and Applications,” in *The Handbook of Econometrics*, ed. by S. Durlauf, L. P. Hansen, J. J. Heckman, and R. Matzkin, Elsevier, vol. 7a, 1–110.
- CHESHER, A. AND K. SMOLINSKI (2012): “IV Models of Ordered Choice,” *Journal of Econometrics*, 166, 33–48.
- CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiments,” *American Economic Review*, 106, 855–902.
- HECKMAN, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System,” *Econometrica*, 46, 931–959.

- HELLIWELL, J., R. LAYARD, AND J. SACHS, eds. (2012): *World Happiness Report*, Earth Institute, Columbia University.
- IMBENS, G. W. AND J. D. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62, 467–475.
- KAPLAN, D. AND L. ZHOU (2019): “Comparing Latent Inequality with Ordinal Data,” Working Paper, University of Missouri.
- KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75, 83–119.
- 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.
- MACKERRON, G. (2012): “Happiness Economics from 35,000 Feet,” *Journal of Economic Surveys*, 26, 705–735.
- PINTO, R. (2019): “Noncompliance as a Rational Choice: A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects,” Working paper, UCLA.
- RIVERS, D. AND Q. H. VUONG (1988): “Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models,” *Journal of Econometrics*, 39, 347–366.
- SCHRÖDER, C. AND S. YITZHAKI (2017): “Revisiting the Evidence for Cardinal Treatment of Ordinal Variables,” *European Economic Review*, 92, 337–358.
- STUTZER, A. AND B. S. FREY (2010): “Recent Advances in the Economics of Individual Subjective Well-Being,” *Social Research*, 77, 679–714.
- WHITE, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–25.

WOOLDRIDGE, J. M. (2014): “Quasi-maximum Likelihood Estimation and Testing for Non-linear Models with Endogenous Explanatory Variables,” *Journal of Econometrics*, 186, 226–234.

Tables and Figures

Table 1: Conditional counterfactual probabilities and marginal effects in numerical examples

		Bounds on Counterfactual Probabilities			Bounds on Marginal Effects		
y_1	y_2	$p(\theta_0, y, x, w)$	$s_{22} = 1.00$	$s_{22} = 0.01$	$\frac{\partial p(\theta_0, y, x, w)}{\partial w}$	$s_{22} = 1.00$	$s_{22} = 0.01$
	0.5	0.16	[0.03, 0.64]	[0.10, 0.23]	-0.24	[-0.49, -0.06]	[-0.31, -0.19]
0	0	0.31	[0.07, 0.76]	[0.24, 0.39]	-0.35	[-0.63, -0.05]	[-0.57, -0.23]
	-0.5	0.50	[0.14, 0.88]	[0.41, 0.62]	-0.40	[-0.66, -0.06]	[-0.61, -0.26]
	0.5	0.46	[0.17, 0.67]	[0.31, 0.59]	-0.14	[-0.45, 0.16]	[-0.36, -0.05]
1	0	0.48	[0.18, 0.69]	[0.32, 0.61]	0.06	[-0.23, 0.31]	[-0.03, 0.15]
	-0.5	0.40	[0.17, 0.61]	[0.27, 0.52]	0.23	[-0.09, 0.56]	[0.12, 0.42]
	0.5	0.38	[0.09, 0.83]	[0.30, 0.51]	0.38	[0.05, 0.67]	[0.23, 0.64]
2	0	0.21	[0.05, 0.69]	[0.15, 0.29]	0.29	[0.07, 0.58]	[0.19, 0.47]
	-0.5	0.10	[0.02, 0.57]	[0.06, 0.15]	0.17	[0.04, 0.45]	[0.13, 0.23]

Table 2: Ordered probit estimation of neighborhood effects on SWB

	(1)	(2)	(3)	(4)	(5)	(6)
$\beta_{Poverty}$	-0.0886*** (0.0209)		-0.0791*** (0.0238)	-0.0904*** (0.0217)		-0.0885*** (0.0248)
$\beta_{Minority}$		-0.0585*** (0.0206)	-0.0213 (0.0235)		-0.0471** (0.0223)	-0.0044 (0.0256)
$\gamma_{Baltimore}$	0.3068*** (0.0766)	0.3150*** (0.0772)	0.2996*** (0.0772)	0.2197** (0.0856)	0.2375*** (0.0859)	0.2180** (0.0862)
γ_{Boston}	0.1812*** (0.0669)	0.1577** (0.0716)	0.1589** (0.0717)	0.0917 (0.0755)	0.0866 (0.0788)	0.0877 (0.0791)
$\gamma_{Chicago}$	0.2800*** (0.0667)	0.2666*** (0.0665)	0.2828*** (0.0667)	0.1609** (0.0785)	0.1452* (0.0784)	0.1613** (0.0786)
γ_{LA}	0.0957 (0.0645)	0.1045 (0.0647)	0.0967 (0.0645)	0.0676 (0.0728)	0.0691 (0.0729)	0.0677 (0.0728)
c_1	-0.4935*** (0.0472)	-0.5243*** (0.0467)	-0.4986*** (0.0476)	-0.7239*** (0.2600)	-0.6989*** (0.2613)	-0.7225*** (0.2609)
c_2	0.8962*** (0.0485)	0.8620*** (0.0475)	0.8914*** (0.0489)	0.7033*** (0.2603)	0.7240*** (0.2617)	0.7048*** (0.2611)
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 3: Complete triangular model estimation of neighborhood effects on SWB

	(1)	(2)	(3)	(4)	(5)	(6)
$\beta_{Poverty}$	-0.1487** (0.0630)		-0.2876*** (0.0994)	-0.1609** (0.0642)		-0.3141*** (0.1119)
$\beta_{Minority}$		-0.0619 (0.1359)	0.3417* (0.2042)		-0.1147 (0.1399)	0.3656 (0.2322)
$\gamma_{Baltimore}$	0.2778*** (0.0809)	0.3131*** (0.1052)	0.3823*** (0.0965)	0.1814** (0.0906)	0.1949 (0.1200)	0.3127*** (0.1153)
γ_{Boston}	0.1435* (0.0771)	0.1531 (0.1954)	0.4967** (0.2243)	0.0526 (0.0822)	0.0093 (0.1801)	0.3779* (0.2253)
$\gamma_{Chicago}$	0.2989*** (0.0694)	0.2676*** (0.0742)	0.2363*** (0.0788)	0.1803** (0.0808)	0.1591* (0.0831)	0.1394* (0.0845)
γ_{LA}	0.0893 (0.0640)	0.1045 (0.0647)	0.0688 (0.0628)	0.0663 (0.0726)	0.0690 (0.0729)	0.0600 (0.0703)
c_1	-0.4750*** (0.0508)	-0.5247*** (0.0488)	-0.3704*** (0.0923)	-0.7291*** (0.2599)	-0.6788** (0.2694)	-0.8070*** (0.2528)
$\ln(c_2 - c_1)$	0.3271*** (0.0233)	0.3267*** (0.0229)	0.2705*** (0.0721)	0.3531*** (0.0237)	0.3505*** (0.0246)	0.2984*** (0.0769)
ρ	0.0668 (0.0681)	0.0036 (0.1400)		0.0769 (0.0675)	0.0676 (0.1383)	
ρ_1			0.0563 (0.0662)			0.0732 (0.0653)
ρ_2			-0.2718 (0.1702)			-0.2579 (0.1844)
$cov(v_1, v_2)$			0.4502*** (0.0242)			0.4200*** (0.0224)
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 4: Complete triangular (CT) and nonlinear IV estimates and confidence sets for marginal effects for the probability of reporting SWB “not too happy” (0, the lowest category) and “happy” (2, the highest category) with respect to neighborhood poverty, with neighborhood minority excluded, at the New York median level. Columns CT (1) and CT (2) report results for the CT model excluding and including additional exogenous covariates, respectively. Columns IV (1) and IV (2) report results from the IV model with parameter restrictions $\beta \leq 0$ and $\beta \geq 0$, respectively.

	(1) <u>CT (1)</u>	(2) <u>CT (2)</u>	(3) <u>IV (1)</u> $\beta \leq 0$	(4) <u>IV (2)</u> $\beta \geq 0$
<i>marginal effect on “not too happy”</i>				
analog estimate	0.052	0.062	[0.019, 0.261]	[-0.211, -0.036]
corrected estimate	–	–	[0.000, 0.410]	[-0.340, 0.000]
95% interval	[0.008, 0.096]	[0.013, 0.110]	[0.000, 0.472]	[-0.351, 0.000]
95% bootstrap interval	–	–	[0.000, 0.391]	[-0.340, 0.000]
<i>marginal effect on “happy”</i>				
analog estimate	-0.040	-0.034	[-0.258, -0.032]	[0.035, 0.199]
corrected estimate	–	–	[-0.375, 0.000]	[0.000, 0.342]
95% interval	[-0.074, -0.007]	[-0.061, -0.006]	[-0.398, 0.000]	[0.000, 0.372]
95% bootstrap interval	–	–	[-0.375, 0.000]	[0.000, 0.342]

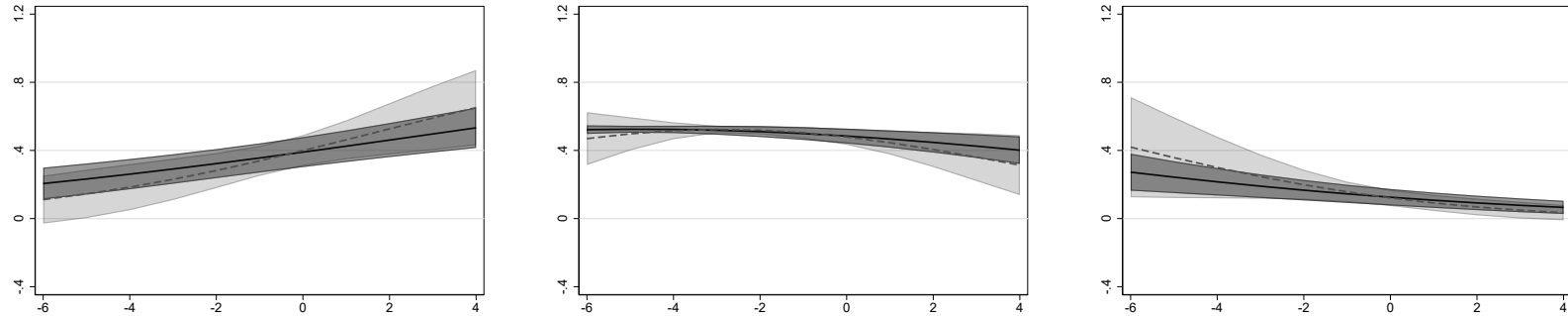
Table 5: Complete triangular (CT) and nonlinear IV estimates and confidence sets for marginal effects for the probability of reporting SWB “not too happy” (0, the lowest category) and “happy” (2, the highest category) with respect to neighborhood poverty, with neighborhood minority included, at the New York median level. Columns CT (1) and CT (2) report results for the CT model excluding and including additional exogenous covariates, respectively. Columns IV (1) and IV (2) report results from the IV model with parameter restrictions $\beta \leq 0$ and $\beta \geq 0$, respectively.

	(1) <u>CT (1)</u>	(2) <u>CT (2)</u>	(3) <u>IV (1)</u> $\beta \leq 0$	(4) <u>IV (2)</u> $\beta \geq 0$
<i>marginal effect on “not too happy”</i>				
analog estimate	0.096	0.115	[0.001, 0.289]	[-0.221, 0.000]
corrected estimate	–	–	[0.000, 0.503]	[-0.456, 0.000]
95% interval	[0.034, 0.159]	[0.038, 0.192]	[0.000, 0.550]	[-0.465, 0.000]
95% bootstrap interval	–	–	[0.000, 0.496]	[-0.435, 0.000]
<i>marginal effect on “happy”</i>				
analog estimate	-0.089	-0.081	[-0.300, 0.000]	[0.000, 0.198]
corrected estimate	–	–	[-0.482, 0.000]	[0.000, 0.393]
95% interval	[-0.162, -0.016]	[-0.160, -0.002]	[-0.571, 0.000]	[0.000, 0.428]
95% bootstrap interval	–	–	[-0.480, 0.000]	[0.000, 0.392]

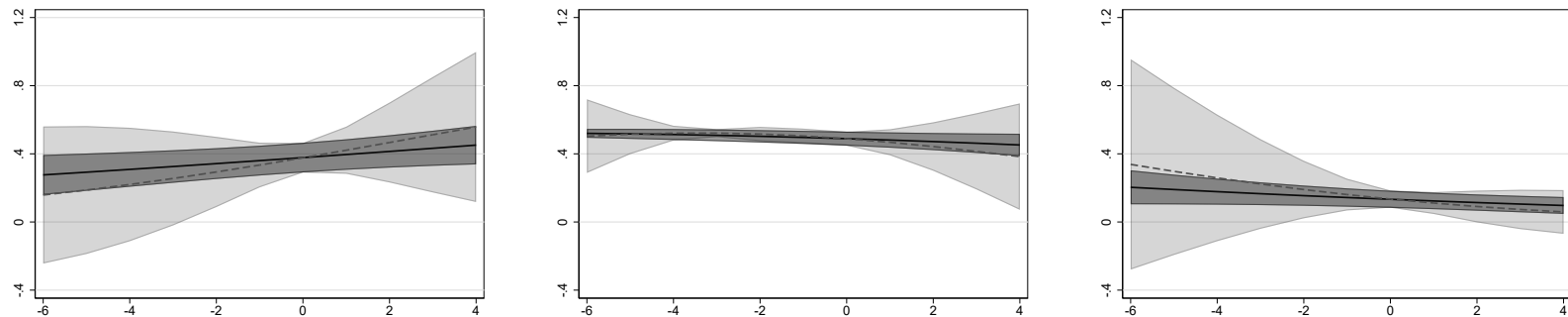
Table 6: Complete triangular (CT) and nonlinear IV estimates and confidence sets for counterfactual response probabilities of reporting SWB “not too happy” (0, the lowest category), “pretty happy” (1, the middle category), and “happy” (2, the highest category) with respect to neighborhood poverty, with neighborhood minority excluded, at the New York median and one standard deviation below the NY median level. Columns CT (1) and CT (2) report results for the CT model excluding and including additional exogenous covariates, respectively. IV results were obtained under the restriction $\beta \leq 0$.

	(1)	(2)	(3)	(4)	(5)	(6)
	<u>CT (1)</u>	<u>CT (2)</u>	<u>IV</u>	<u>CT (1)</u>	<u>CT (2)</u>	<u>IV</u>
	<i>NY median poverty</i>			<i>NY median poverty minus 1 s.e.</i>		
<i>counterfactual response probability for “not too happy”</i>						
analog estimate	0.307	0.387	[0.167, 0.532]	0.256	0.326	[0.119, 0.427]
corrected estimate	–	–	[0.055, 0.714]	–	–	[0.027, 0.605]
95% interval	[0.274, 0.339]	[0.299, 0.475]	[0.041, 0.750]	[0.210, 0.302]	[0.236, 0.416]	[0.024, 0.626]
95% bootstrap	–	–	[0.067, 0.727]	–	–	[0.027, 0.621]
<i>counterfactual response probability for “pretty happy”</i>						
analog estimate	0.504	0.485	[0.136, 0.662]	0.512	0.509	[0.058, 0.651]
corrected estimate	–	–	[0.000, 0.844]	–	–	[0.000, 0.832]
95% interval	[0.484, 0.525]	[0.442, 0.528]	[0.000, 0.874]	[0.492, 0.531]	[0.476, 0.541]	[0.000, 0.883]
95% bootstrap	–	–	[0.000, 0.807]	–	–	[0.000, 0.828]
<i>counterfactual response probability for “happy”</i>						
analog estimate	0.189	0.128	[0.121, 0.383]	0.233	0.166	[0.171, 0.529]
corrected estimate	–	–	[0.036, 0.603]	–	–	[0.071, 0.717]
95% interval	[0.164, 0.214]	[0.079, 0.177]	[0.017, 0.633]	[0.187, 0.279]	[0.101, 0.230]	[0.044, 0.744]
95% bootstrap	–	–	[0.037, 0.613]	–	–	[0.084, 0.733]

Figure 1: $p(y; x, w)$ across w for NY respondents



(a) $p(0; x, w)$ across neighborhood poverty (b) $p(1; x, w)$ across neighborhood poverty (c) $p(2; x, w)$ across neighborhood poverty

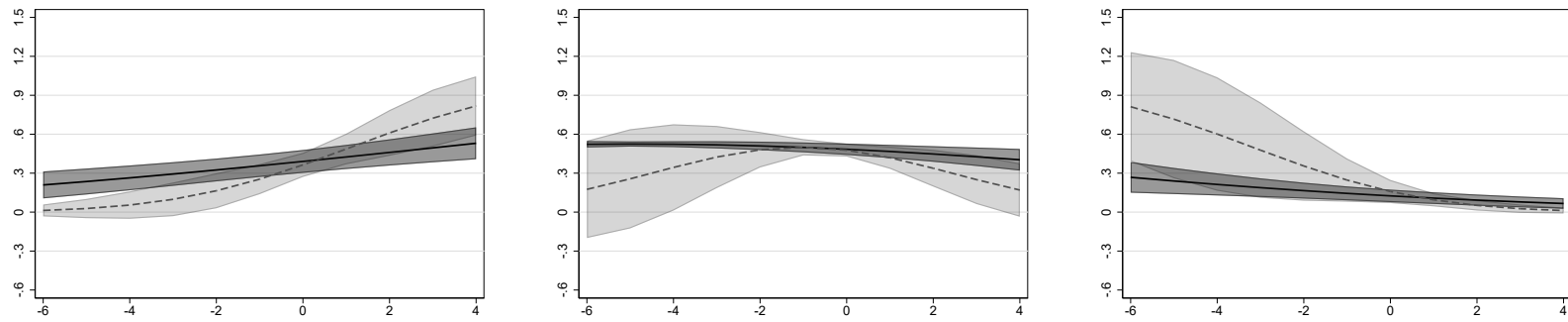


(d) $p(0; x, w)$ across neighborhood minority (e) $p(1; x, w)$ across neighborhood minority (f) $p(2; x, w)$ across neighborhood minority

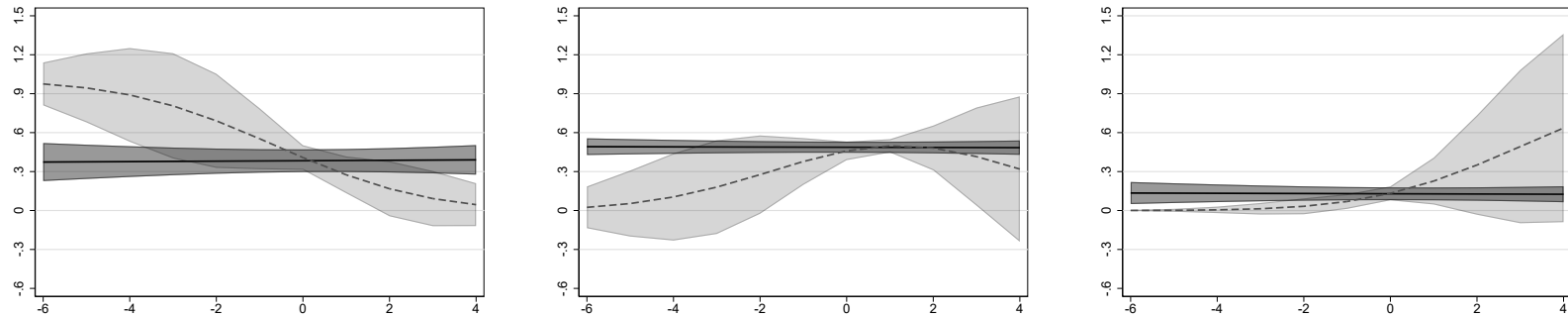
Notes: The dependent variable Y or SWB takes the value zero for not too happy, one for pretty happy and two for very happy. Solid lines connect eleven predicted conditional probabilities for $w \in \{-6, -5, \dots, +3, +4\}$ using an ordered probit model as reported in columns (4)-(5) of Table 2 while dashed lines connect eleven predicted conditional probabilities for $w \in \{-6, -5, \dots, +3, +4\}$ using a complete triangular model as reported in columns (4)-(5) of Table 3. Values of X variables are held fixed at the NY sample median, and randomization assignment is to the experimental voucher group. The dark shaded area gives 95% confidence intervals for ordered probit predictions using the delta method while the light shaded area gives 95% confidence intervals for complete triangular model predictions using the `nlcom` command in STATA.

Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Figure 2: $p(y; x, w)$ across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



(a) $p(0; x, w)$ across neighborhood poverty (b) $p(1; x, w)$ across neighborhood poverty (c) $p(2; x, w)$ across neighborhood poverty

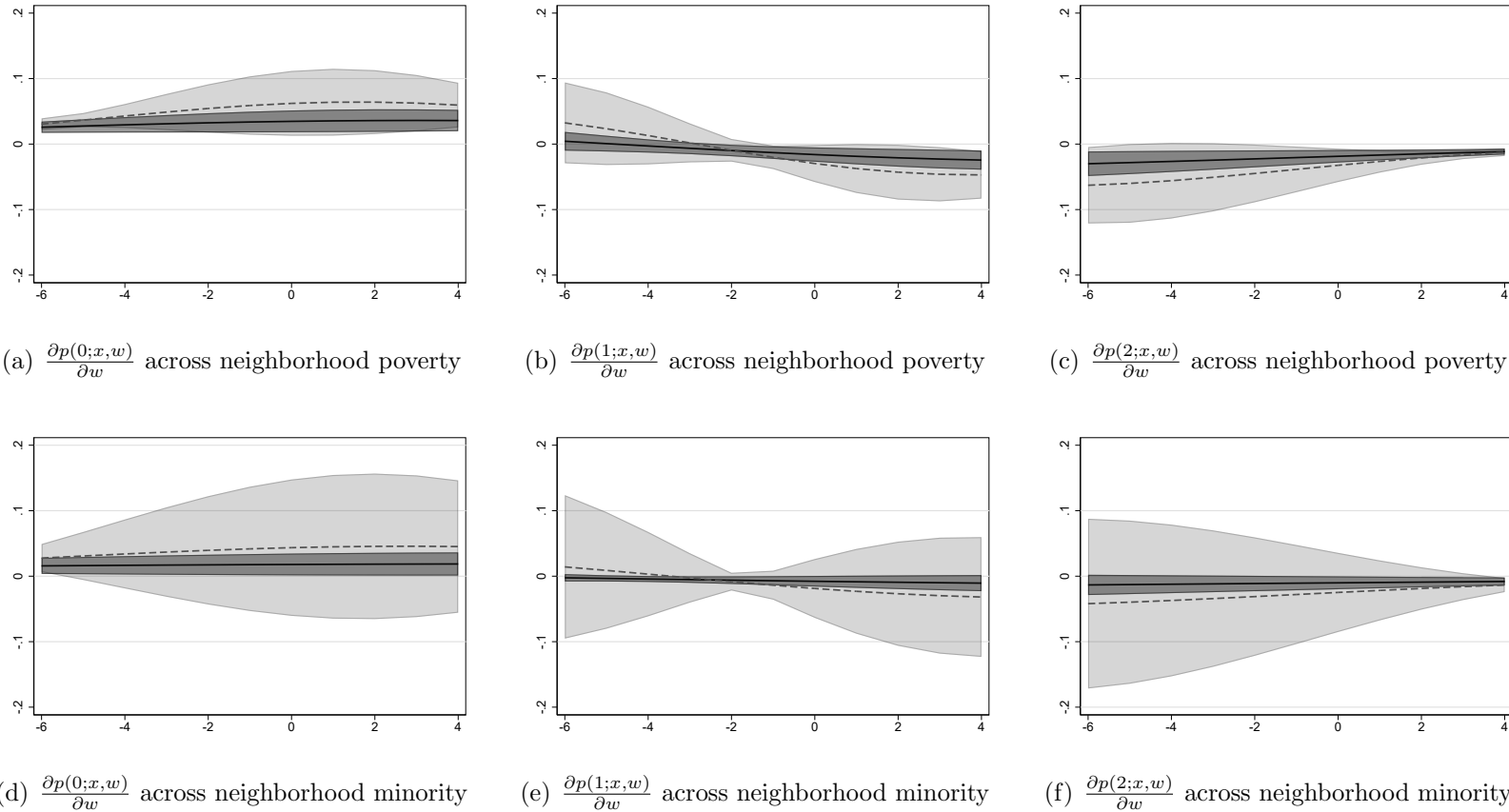


(d) $p(0; x, w)$ across neighborhood minority (e) $p(1; x, w)$ across neighborhood minority (f) $p(2; x, w)$ across neighborhood minority

Notes: The dependent variable Y or SWB takes the value zero for not too happy, one for pretty happy and two for very happy. Solid lines connect eleven predicted conditional probabilities for $w \in \{-6, -5, \dots, +3, +4\}$ using an ordered probit model as reported in column (6) of Table 2 while dashed lines connect eleven predicted conditional probabilities for $w \in \{-6, -5, \dots, +3, +4\}$ using a complete triangular model as reported in column (6) of Table 3. Values of X variables are held fixed at the NY sample median, and randomization assignment is to the experimental voucher group. The dark shaded area gives 95% confidence intervals for ordered probit predictions using the delta method while the light shaded area gives 95% confidence intervals for complete triangular model predictions using the `nlncom` command in STATA.

Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

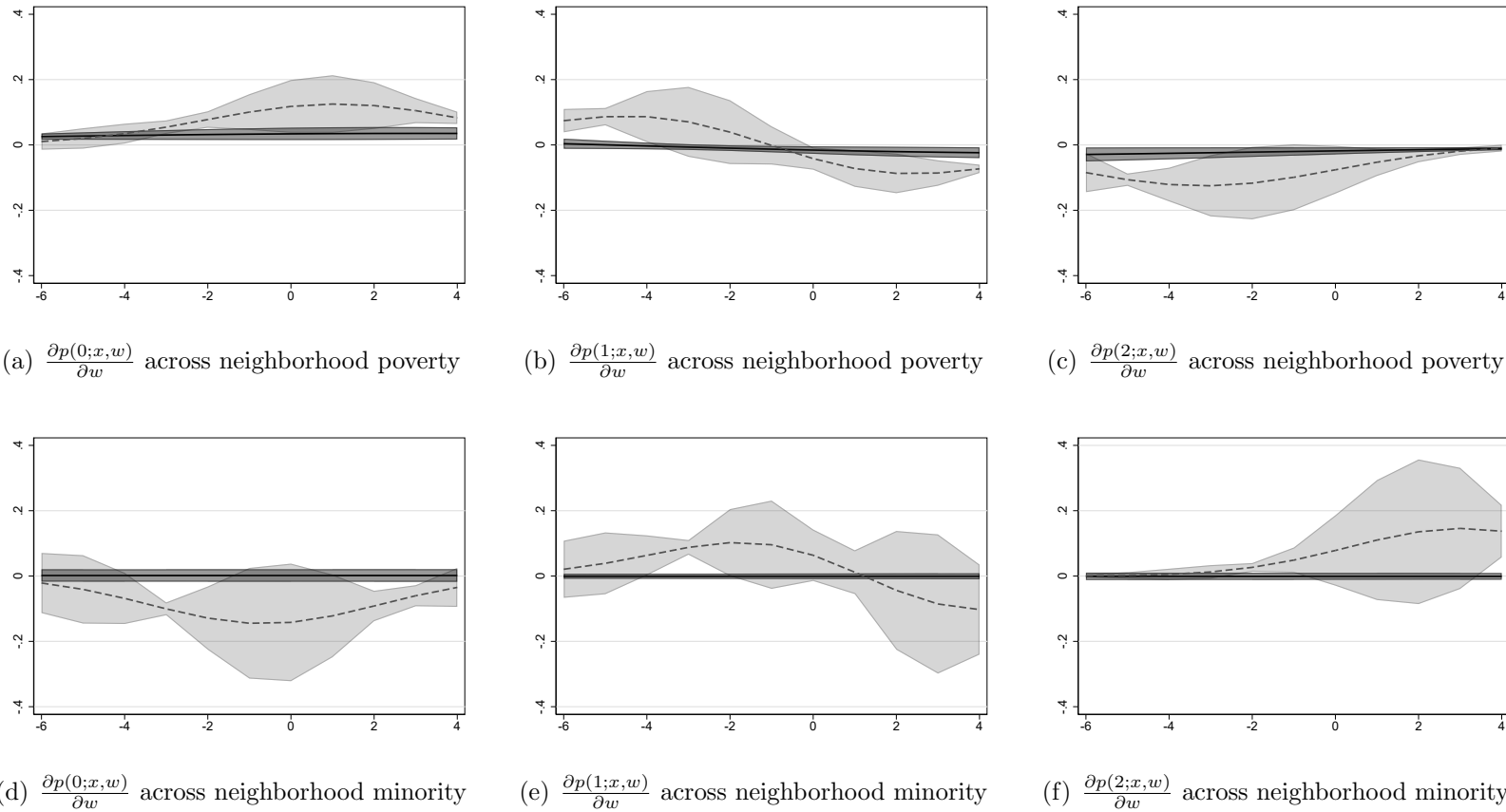
Figure 3: Marginal effects across w for NY respondents



Notes: The dependent variable Y or SWB takes the value zero for not too happy, one for pretty happy and two for very happy. Solid lines connect eleven predicted marginal effects for $w \in \{-6, -5, \dots, +3, +4\}$ using an ordered probit model as reported in columns (4)-(5) of Table 2 while dashed lines connect eleven predicted marginal effects for $w \in \{-6, -5, \dots, +3, +4\}$ using a complete triangular model as reported in columns (4)-(5) of Table 3. Values of X variables are held fixed at the NY sample median, and randomization assignment is to the experimental voucher group. The dark shaded area gives 95% confidence intervals for ordered probit predictions using the delta method while the light shaded area gives 95% confidence intervals for complete triangular model predictions using the `nlncom` command in STATA.

Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Figure 4: Marginal effects across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



Notes: The dependent variable Y or SWB takes the value zero for not too happy, one for pretty happy and two for very happy. Solid lines connect eleven predicted marginal effects for $w \in \{-6, -5, \dots, +3, +4\}$ using an ordered probit model as reported in column (6) of Table 2 while dashed lines connect eleven predicted marginal effects for $w \in \{-6, -5, \dots, +3, +4\}$ using a complete triangular model as reported in column (6) of Table 3. Values of X variables are held fixed at the NY sample median, and randomization assignment is to the experimental voucher group. The dark shaded area gives 95% confidence intervals for ordered probit predictions using the delta method while the light shaded area gives 95% confidence intervals for complete triangular model predictions using the `nlncom` command in STATA.

Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

Online Supplement for “Estimating Endogenous Effects on Ordinal Outcomes”^{*}

Andrew Chesher[†] Adam M. Rosen[‡]
UCL and CeMMAP Duke University and CeMMAP

Zahra Siddique[§]
University of Bristol

June 20, 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 ISCPM 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 are grateful for comments from the editor and two reviewers, Alex Belloni, Federico Bugni, and seminar and conference participants at Erasmus University Rotterdam, Duke, Simon Fraser, Leicester, Johns Hopkins, and Western Ontario, 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

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

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

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

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.

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

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 $t \in \mathbb{R}$:

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

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

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

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

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

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

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

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 $s, t \in \mathbb{R}$, $s < t$:

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

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

¹When 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 $\pm\infty$):

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

Substitution for C_θ and \overline{C}_θ 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

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

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).

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

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).

²Ludwig 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
<i>Site:</i>			
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
<i>Demographic characteristics:</i>			
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
<i>Household characteristics:</i>			
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
<i>Neighborhood characteristics:</i>			
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 $\geq 1x$ per week	0.53	0.50	0.54
Household head very likely to tell on neighborhood kid	0.55	0.52	0.56
Household head very sure of finding apartment	0.47	0.50	0.46
Household head applied for Section 8 before	0.39	0.40	0.43
<i>Primary or secondary reason for wanting to move:</i>			
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

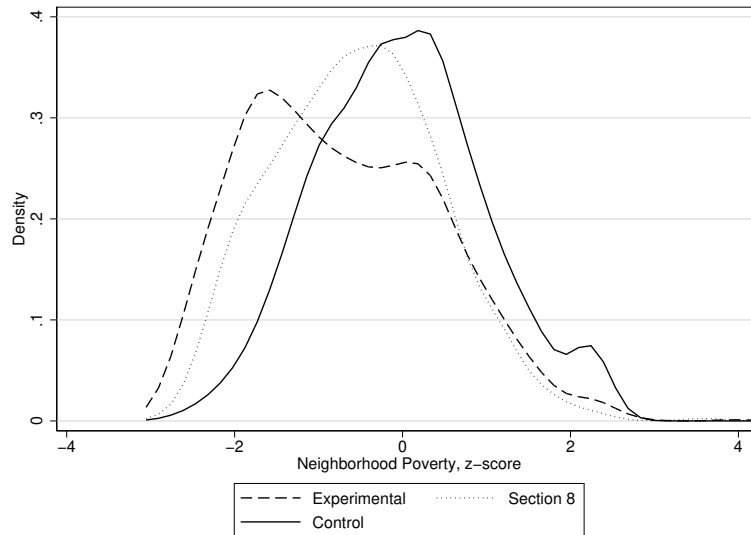
Table B.1: Baseline characteristics of MTO adults or covariates X across randomization groups

	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.

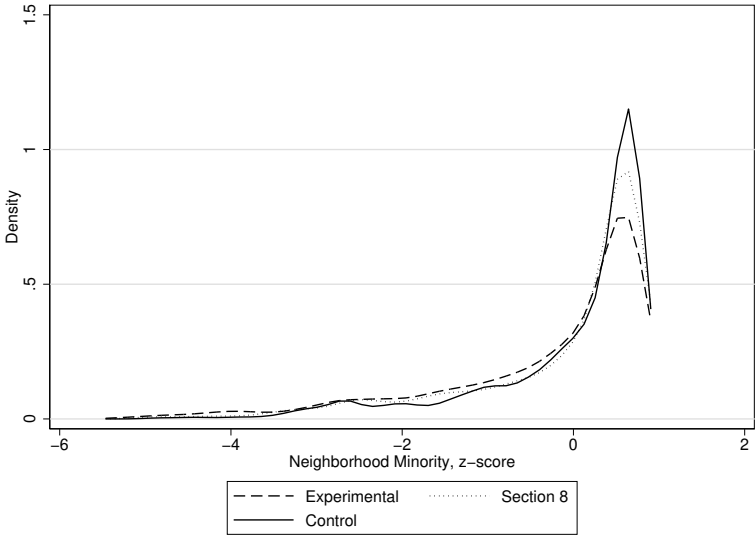
Source: Data from ICPSR Study 34860: Moving to Opportunity: Final Impacts Evaluation Science Article Data, 2008-2010.

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

³That 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*** (0.0130)		-0.0491*** (0.0148)	-0.0546*** (0.0131)		-0.0534*** (0.0151)
$\beta_{Minority}$		-0.0367*** (0.0129)	-0.0135 (0.0147)		-0.0287** (0.0136)	-0.0029 (0.0157)
N	3263	3263	3263	3175	3175	3175
Panel B: IV estimation						
$\beta_{Poverty}$	-0.0916** (0.0382)		-0.1803*** (0.0675)	-0.0962** (0.0376)		-0.1859*** (0.0687)
$\beta_{Minority}$		-0.0383 (0.0694)	0.2048 (0.1245)		-0.0632 (0.0688)	0.2019 (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)
$Z = \text{Any voucher}$	0.0630** (0.0283)			0.0660** (0.0284)		
$Z = \text{Low pov voucher}$		0.0521* (0.0298)			0.0546* (0.0298)	
$Z = \text{Sec 8 voucher}$			0.0793** (0.0385)			0.0875** (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*** (0.1017)	-0.8235*** (0.1143)	-1.0910*** (0.1021)	-1.1591*** (0.1038)	-0.9015*** (0.1190)	-1.1589*** (0.1042)
$\delta_1^{exp,Bos}$	-1.2798*** (0.0877)	-1.7007*** (0.1264)	-1.2819*** (0.0876)	-1.2154*** (0.0896)	-1.4961*** (0.1189)	-1.2168*** (0.0898)
$\delta_1^{exp,Chi}$	-0.3068*** (0.0852)	0.0993 (0.0736)	-0.3055*** (0.0853)	-0.3470*** (0.0915)	0.0334 (0.0792)	-0.3450*** (0.0917)
$\delta_1^{exp,LA}$	-0.8787*** (0.1007)	-0.3421*** (0.0822)	-0.8801*** (0.1003)	-0.8137*** (0.1044)	-0.3436*** (0.0916)	-0.8142*** (0.1042)
$\delta_1^{exp,NY}$	-0.8052*** (0.0875)	-0.1401 (0.0854)	-0.8021*** (0.0877)	-0.7993*** (0.0891)	-0.1326 (0.0873)	-0.7945*** (0.0895)
$\delta_1^{sec8,Balt}$	-1.0427*** (0.1184)	-0.6651*** (0.1992)	-1.0412*** (0.1194)	-1.1065*** (0.1254)	-0.7093*** (0.1986)	-1.1032*** (0.1264)
$\delta_1^{sec8,Bos}$	-1.0880*** (0.1055)	-1.2662*** (0.1428)	-1.0838*** (0.1058)	-1.0376*** (0.1130)	-1.1362*** (0.1455)	-1.0328*** (0.1134)
$\delta_1^{sec8,Chi}$	-0.1905* (0.1092)	0.2901*** (0.0802)	-0.1863* (0.1092)	-0.2696** (0.1171)	0.1970** (0.0894)	-0.2643** (0.1168)
$\delta_1^{sec8,LA}$	-0.8139*** (0.0960)	0.0257 (0.0988)	-0.8072*** (0.0960)	-0.7508*** (0.1041)	-0.0071 (0.1066)	-0.7429*** (0.1038)
$\delta_1^{sec8,NY}$	-0.3742*** (0.0913)	-0.0448 (0.0838)	-0.3728*** (0.0916)	-0.3945*** (0.0923)	-0.0548 (0.0867)	-0.3920*** (0.0926)
$\delta_1^{cont,Balt}$	-0.5220*** (0.0899)	-0.3311*** (0.0931)	-0.5184*** (0.0902)	-0.5641*** (0.0949)	-0.3941*** (0.0998)	-0.5590*** (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*** (0.0722)	-1.1184*** (0.1018)	-0.7106*** (0.0721)	-0.6409*** (0.0823)	-0.9028*** (0.1047)	-0.6350*** (0.0825)
$\delta_1^{cont,Chi}$	0.2299** (0.0999)	0.2621*** (0.0718)	0.2297** (0.0999)	0.1848* (0.1078)	0.2124*** (0.0809)	0.1855* (0.1077)
$\delta_1^{cont,LA}$	0.1584* (0.0907)	0.2110*** (0.0664)	0.1597* (0.0906)	0.2360** (0.0959)	0.2192*** (0.0729)	0.2381** (0.0959)
$\delta_1^{cont,NY}$	0.1371** (0.0547)	0.1735*** (0.0567)	0.1354** (0.0548)	0.5305** (0.2234)	-0.2895 (0.2631)	0.5258** (0.2235)
$\delta_2^{exp,Balt}$			-0.8173*** (0.1118)			-0.8886*** (0.1171)
$\delta_2^{exp,Bos}$			-1.7055*** (0.1226)			-1.4949*** (0.1166)
$\delta_2^{exp,Chi}$			0.0880 (0.0733)			0.0216 (0.0798)
$\delta_2^{exp,LA}$			-0.3677*** (0.0828)			-0.3710*** (0.0929)
$\delta_2^{exp,NY}$			-0.1588* (0.0846)			-0.1447* (0.0877)
$\delta_2^{sec8,Balt}$			-0.6895*** (0.1946)			-0.7349*** (0.1968)
$\delta_2^{sec8,Bos}$			-1.2842*** (0.1401)			-1.1477*** (0.1415)
$\delta_2^{sec8,Chi}$			0.2812*** (0.0799)			0.1982** (0.0877)
$\delta_2^{sec8,LA}$			0.0347 (0.0883)			0.0240 (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 (0.0843)			-0.0641 (0.0863)
$\delta_2^{cont,Balt}$			-0.3553*** (0.0982)			-0.4127*** (0.1033)
$\delta_2^{cont,Bos}$			-1.1349*** (0.1018)			-0.9192*** (0.1077)
$\delta_2^{cont,Chi}$			0.2454*** (0.0743)			0.1999** (0.0824)
$\delta_2^{cont,LA}$			0.1980*** (0.0667)			0.2055*** (0.0731)
$\delta_2^{cont,NY}$			0.1858*** (0.0565)			-0.2870 (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 $\tilde{z} \in \{1, 2\}$:

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

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

$$\mathbb{P}[Y = 0|X = x, Z = 0] > \mathbb{P}[Y = 0|X = x, Z = \tilde{z}], \quad (\text{E.1})$$

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

$$\mathbb{P}[Y = 2|X = x, Z = 0] < \mathbb{P}[Y = 2|X = x, Z = \tilde{z}]. \quad (\text{E.2})$$

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 $\{-\infty, \Phi^{-1}(1/n), \Phi^{-1}(2/n), \dots, \Phi^{-1}((n-1)/n), \infty\}$, as described in Section 3.2.

Minimization was done using the COBYLA algorithm of `nloptr`. At termination of the COBYLA algorithm 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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} Y = \sum_{j=1}^J j \times 1 [c_{0,j} < W\beta_0 + X\gamma_0 + U \leq c_{0,j+1}], \quad W = \delta_x + Z\delta_z + V, \\ (U, V) \sim \text{BVN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & R \\ R & \Sigma_v \end{pmatrix} \right), \end{aligned}$$

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

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

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.

⁴Here the parameter vector $\theta = (c_1, c_2, \gamma\beta)$ whose identified 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.

$$\begin{aligned}\varphi(-\infty, t, x, z; \theta) = & \\ & \mathbb{P}[(V_1 \leq c_{0,1} - x\gamma_0 - \delta_x\beta_0 - z\delta_z\beta_0 \wedge V_2 \geq c_1 - x\gamma - \delta_x\beta - z\delta_z\beta - t)] \\ & + \mathbb{P}[c_{0,1} \leq V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \leq c_{0,2} \wedge V_2 \geq c_2 - x\gamma - \delta_x\beta - z\delta_z\beta - t],\end{aligned}$$

$$\begin{aligned}\varphi(s, \infty, x, z; \theta) = & \\ & \mathbb{P}[c_{0,1} \leq V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \leq c_{0,2} \wedge V_2 \leq c_1 - x\gamma - \delta_x\beta - z\delta_z\beta - s] \\ & + \mathbb{P}[V_1 \geq c_{0,2} - x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \wedge V_2 \leq c_2 - x\gamma - \delta_x\beta - z\delta_z\beta - s],\end{aligned}$$

$$\begin{aligned}\varphi(s, t, x, z; \theta) = & \\ & \mathbb{P}[c_{0,1} \leq V_1 + x\gamma_0 + \delta_x\beta_0 + z\delta_z\beta_0 \leq c_{0,2} \wedge c_2 - t \leq V_2 + x\gamma + \delta_x\beta + z\delta_z\beta \leq c_1 - s].\end{aligned}$$

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:

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

where

$$\begin{aligned}\mathbb{P}[Y = 0|x, z] &= \Phi\left(\frac{c_{0,1} - x\gamma_0 - (\delta_x + z\delta_z)\beta_0}{(\beta_0^2\Sigma_v + 2\beta_0R + 1)^{1/2}}\right) \\ \mathbb{P}[Y = 1|x, z] &= \Phi\left(\frac{c_{0,2} - x\gamma_0 - (\delta_x + z\delta_z)\beta_0}{(\beta_0^2\Sigma_v + 2\beta_0R + 1)^{1/2}}\right) - \Phi\left(\frac{c_{0,1} - x\gamma_0 - (\delta_x + z\delta_z)\beta_0}{(\beta_0^2\Sigma_v + 2\beta_0R + 1)^{1/2}}\right) \\ \mathbb{P}[Y = 2|x, z] &= 1 - \Phi\left(\frac{c_{0,2} - x\gamma_0 - (\delta_x + z\delta_z)\beta_0}{(\beta_0^2\Sigma_v + 2\beta_0R + 1)^{1/2}}\right)\end{aligned}$$

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

⁵A 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/MT0-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

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

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

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

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

⁶To 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,

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

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 $n \rightarrow \infty$. 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 $M_n/\bar{w}_n \rightarrow \infty$, 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

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

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

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

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

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

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 ξ_1, \dots, ξ_n and computing the bootstrap process (F.4). For each bootstrap sample the following steps are then conducted:

1. Compute \bar{w}_n by solving the maximization problem in (F.7) and set $M_n = \log(n) \cdot \bar{w}_n$.
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($r_{\text{out}}, r_{\text{in}}, \text{fromLower}, \Delta$)

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 $b = 1, \dots, B$ **do**

 Compute $\bar{\omega}_n$ and set $M_n = \log(n) \cdot \bar{\omega}_n$

 Compute $R_n^{DR*} = \inf_{\theta \in \hat{\Theta}(r)} \max_{j \in \hat{\Psi}_{\theta}(M_n)} \hat{v}_{\theta, j}^*$

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, iISBN 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>.