cemmap

centre for microdata methods and practice

Estimating Endogenous Effects on Ordinal Outcomes

Andrew Chesher
Adam M. Rosen
Zahra Siddique

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP66/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Estimating Endogenous Effects on Ordinal Outcomes*

Andrew Chesher† UCL and CeMMAP
Adam M. Rosen‡ Duke University and CeMMAP
Zahra Siddique§ University of Bristol

November 27, 2019

Abstract

Recent research underscores the sensitivity of conclusions drawn from the application of econometric methods devised for quantitative outcome variables to data featuring ordinal outcomes. The issue is particularly acute in the analysis of happiness data, for which no natural cardinal scale exists, and which is thus routinely collected by ordinal response. With ordinal responses, comparisons of means across different populations and the signs of OLS regression coefficients have been shown to be sensitive to monotonic transformations of the cardinal scale onto which ordinal responses are mapped.

In many applications featuring ordered outcomes, including responses to happiness surveys, researchers may wish to study the impact of a ceteris paribus change in certain variables induced by a policy shift. Insofar as some of these variables may be manipulated by the individuals involved, they may be endogenous. This paper examines the use of instrumental variable (IV) methods to measure the effect of such changes. While linear IV estimators suffer from the same pitfalls as averages and OLS coefficient estimates when outcome variables are ordinal, nonlinear models that explicitly respect the ordered nature of the response variable can be used. This is demonstrated with an application to the study of the effect of neighborhood characteristics on subjective well-being among participants in the Moving to Opportunity housing voucher experiment. In this context, the application of nonlinear IV models can be used to estimate marginal effects and counterfactual probabilities of categorical responses induced by changes in neighborhood characteristics such as the level of neighborhood poverty.

*We have benefitted from comments from participants at research seminars given at Erasmus University Rotterdam, Duke, Simon Fraser, Leicester, and a 2018 joint CeMMAP/Northwestern conference on incomplete models. 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 usual disclaimer applies.

†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.

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 these individuals' well-being. Such characteristics might include for instance sex, age, education, quality of local public goods, and neighborhood characteristics, among others. A type of question often contemplated by economists is the following: If a policy were implemented that were to exogenously shift the value of one of these characteristics, all else equal, how would this affect the subsequent values of a measurable outcome of interest for the individuals in that community? The goal of this article is to consider such questions when the outcome of interest is ordinal in nature.

An essential qualifier in the preceding scenario is that the outcome on which the researcher focuses be *measurable*. When the outcome is the happiness of individuals, there is no obvious unit of measurement. A commonly used measure of happiness is self-reported subjective well-being (SWB), measured 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?"

How do individuals answer such questions? While the ranking of possible responses is clear, there is no unique measure of happiness to compare across individuals. This paper proposes a path to answering questions concerning the effect of exogenous changes in observable variables on ordinal outcomes despite this limitation on the interpretation of ordinal data collected from heterogeneous agents. In order to focus our analysis we consider specifically the use of ordinal SWB as the outcome variable of study, an outcome which has featured prominently in the recent literature on the determinants of happiness. However, the same issues arise with other ordinal outcomes, and our analysis is equally applicable to any other such ordinal outcomes that may lack a unique cardinal scale.

Two complications must be taken into account. The first is the aforementioned issue regarding the measurement of ordinal outcomes. Although there has been a great deal of interest in the determinants of happiness, comparisons of average happiness across different populations suffer from the fact that they are not invariant to monotone transformations of the measures of happiness that are employed. Consequently, such comparisons are sensitive to the cardinal scale imposed on the ordered responses, as previously pointed out by Schröder and Yitzhaki (2017) and Bond and Lang (2019). These papers show in several empirical examples that happiness data generally do not satisfy the properties required for these comparisions, 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. An alternative approach is put forward by Chen, Oparina, Powdthavee, and Srisuma (2019), who suggest the use of quantiles and quantile regression, since quantiles are equivariant under monotonic transformations. This is a sound alternative, but the methods used in that paper do not enable the measurement of ceteris paribus effects in the presence of potentially endogenous explanatory variables that is the goal of the analysis here. Another related but different analysis is provided by Kaplan and Zhou (2019), who 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 than that of this paper.

In this paper we address the rescaling issue by using nonlinear models that respect the ordinal nature of the outcomes 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, because of 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 questions 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 not require the construction of a happiness measure beyond what is reported by the households. They allow for the possibility that the households interpret the questions in different ways. For example households in Baltimore may answer differently than households in New York, but the counterfactual scenario is interpretable nonetheless, since we can measure the probability that households in Baltimore and New York provide each possible answer.

A further complication is the possibility of endogeneity. It is reasonable to expect individuals to make decisions that improve their well-being. Observable covariates that are themselves influenced by or chosen by households should therefore not be expected to be independent of unobservable determinants of well-being. For instance, in the context of the Moving to Opportunity housing experiment, this means that neighborhood poverty rates should be treated as endogenous. Households that believe they stand to benefit from moving to a lower poverty neighborhood will be more likely to do so when given the opportunity.

To deal with endogeneity we employ instrumental variables. While linear IV methods are easy to use, they are not appropriate for ordinal outcomes. This follows directly from the reasoning put forward by Schröder and Yitzhaki (2017) and Bond and Lang (2019). Linear IV estimates are weighted averages of observable data, and they are sensitive to monotonic transformations of the scale on which ordinal outcomes are placed, just as comparisons of means and OLS estimates are. This is made explicit in Section 2 below.

The most straightforward nonlinear models to use are ordered probit and logit models, easily implemented in STATA. However, these models restrict all covariates to be independent of unobserved heterogeneity, and are therefore inappropriate when some covariates are endogenous.

One way to allow for endogeneity while respecting the ordinal nature of the outcome is to employ a complete triangular model. These 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 of the endogenous covariates, each of which features an additive unobservable, all of which are normally distributed. This generalizes a triangular IV model for binary outcomes studied by e.g. Heckman (1978) and Rivers and Vuong (1988). The model provides a complete specification for the determination of all endogenous variables, such that all parameters are point identified under mild conditions, and can be consistently estimated by a two stage procedure or by maximum likelihood. Marginal effects and counterfactuals are smooth functions of these parameters, and as such standard approaches can be employed for asymptotic inference. The triangular IV model thus offers a coherent approach to perform estimation and inference on such quantities, respecting the ordinal nature of the outcome variable.

The triangular IV model is fully parametric, and, as is always the case in a structural analysis, the inferences drawn rely on the suitability of the restrictions imposed by the model. One may wonder how sensitive empirical findings are to relaxation of the restrictions employed. A sensitivity analysis is conducted in Section 4, which considers the use of a single equation instrumental variable model in which the auxiliary equations for the endogenous variables are dropped. The single equation IV model is incomplete because it does not uniquely pin down the value of all 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.

The results from Chesher and Rosen (2017) for generalized instrumental variable models can in fact be applied to characterize identified sets for a much broader range of IV models for ordered outcomes. Two further examples of bound characterizations in IV models that dispense with parametric distributional restrictions on unobservable variables are provided at the end of Section 4.1. Our goal in this paper is not to advocate for the use of any particular model(s), but rather to advocate for the use of models that are compatible with ordinal outcomes when that is what is called for, and for an explicit and careful consideration of the restrictions invoked and their identifying power.

Section 5 lays out the application in which the use of IV models for ordinal outcomes is demonstrated, namely the study of the effect of neighborhood poverty on subjective well-being using data from the Moving to Opportunity housing voucher experiment (MTO). The neighborhoods in which individuals live are chosen by the individuals themselves, and may therefore be in part determined by unobservable individual specific attributes that also effect SWB. This motivates the use of IVs to control for neighborhood characteristics. In this application, described in more detail in Section 5, random assignment of housing vouchers constitutes such an instrument. Empirical results are presented in Section 6, comparing the estimates obtained from employing different models. The triangular IV model point identifies the marginal effects and counterfactual probabilities of interest while respecting both the ordered nature of the outcome variable and the endogeneity of neighborhood characteristics. The partially identifying single equation IV models described in Section 4 provide a sensitivity analysis to the restrictions imposed on the determination of these neighborhood characteristics. Section 7 concludes. The remainder of this introduction describes the relation to the prior literature.

Related Literature

The methodology studied in this paper broadly pertains to applications with ordinal outcome variables, but the particular application in this paper is to data on SWB. The 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, Layard, and Sachs (2012)) commissioned for the United Nations Conference on Happiness in April 2012, 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). SWB is invariably measured by survey questions eliciting a discrete ordered response.

We focus attention on the effect of neighborhood characteristics on SWB, specifically among the economically disadvantaged subpopulation of 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' subjective well-being – 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 variables in the determination of SWB may be unjustified. The Moving to Opportunity data offers a unique opportunity to exploit random variation in housing possibilities – and neighborhood choice in particular – through random assignment of housing vouchers.

The MTO data has featured in previous studies, but these have predominantly focused on outcomes other than SWB. An important exception is Ludwig et al. (2012), who report OLS and linear IV estimates using MTO long-term outcome data to study neighborhood effects on a variety of outcomes, one of which is SWB. However, these estimators fall prey to the critiques of Bond and Lang (2019), and the study of SWB using MTO data is one of the nine applications Bond and Lang revisit. Moreover, these methods do not enable the construction of counterfactuals and marginal effects which are the focus of this paper.[1] We provide a fuller description of the related literature on MTO in Section 5.

From an econometric standpoint the sensitivity analyses of Section 6 employ bound characterizations obtained by application of general results in Chesher and Rosen (2017) to the models studied here. The single equation IV ordered probit type specification without any additional restrictions would fall in the class of models considered earlier by Chesher and Smolinski (2012), but for the fact that here the endogenous variable is continuously distributed. We implement an approach to inference developed by Belloni, Bugni, and Chernozhukov (2018) for the bound characterization delivered by application of the Chesher and Rosen (2017) analysis. The bound characterization features quite a large number of moment inequalities, for which the methods of Belloni, Bugni, and Chernozhukov (2018) are particularly well-suited for the purpose of conducting inference on subvectors and functions of partially identified parameter vectors.

2 The Drawbacks of Linear Model Estimators

In this section we briefly review some commonly used linear model estimators, and we highlight their drawbacks when employed with discrete ordered outcome variables, such as subjective well-being. In particular, these estimators all comprise weighted averages of the support points of the ordered outcomes. The cardinal value attributed to each ordered category is arbitrary, and such averages are prey to the criticisms leveled by Schröder and Yitzhaki (2017) and Bond and Lang (2019).

Consider the linear model

$$Y = W\beta + X\gamma + U, \tag{2.1}$$

where W is a k_w-vector of endogenous variables and X is k_x-vector of exogenous variables. Z are excluded instruments. Y is a discrete, ordered outcome variable taking values $y \in \{0, ..., J\}$.

Examination of the OLS estimand gives us

$$\begin{aligned}
\beta_{OLS} &\equiv E\left[(W,X)'(W,X)\right]^{-1} E\left[(W,X)'Y\right] \\
&= E\left[(W,X)'(W,X)\right]^{-1} E\left[(W,X)'\textstyle\sum_{y=1}^{J} yP\left[Y=y|W,X\right]\right].
\end{aligned}$$

The OLS estimator from a linear regression of Y on (X, W) is consistent for β_{OLS}. Each com-

[1] One might think to interpret a slope coefficient in a linear regression as a marginal effect. However, linear model coefficients are sensitive to the cardinal scale of the outcome variable. Schröder and Yitzhaki (2017) have shown that even the sign of a simple linear regression slope coefficient is generally sensitive to the scale on which ordered outcomes are mapped.

ponent of β_{OLS} is a weighted average of the population probabilities $P[Y=y|W,X]$ with weights dependent upon the distribution of (W,X) and proportional to y.

Another OLS estimator is obtained by a linear regression of Y on only the included and excluded exogenous variables X and Z. This estimator is consistent for

$$\begin{aligned}\beta_{ITT} &\equiv E\left[(Z,X)'(Z,X)\right]^{-1}E\left[(Z,X)'Y\right]\\ &= E\left[(Z,X)'(Z,X)\right]^{-1}E\left[(Z,X)'\textstyle\sum_{y=1}^{J}yP[Y=y|Z,X]\right].\end{aligned}$$

This is a different weighted average than β_{OLS}, comprising a weighted average of probabilities conditional on (Z,X) rather than on (W,X). It thus cannot reveal information about the effect of W on Y. The component of β_{ITT} comprising the slope coefficient on Z provides an intention to treat effect (ITT), capturing the effect of the instrument on the outcome.

The OLS estimands are the best linear predictors for Y given (W,X), and Y given (Z,X), respectively. That is, the quantities $(W,X)\beta_{OLS}$ and $(Z,X)\beta_{ITT}$ minimize the mean square error of $Y-(W,X)b$ and $Y-(Z,X)b$, respectively. They provide the best linear approximation to the conditional expectation of Y given (W,X), and Y given (Z,X), respectively. However, these are both sensitive to the cardinal scale of Y that is used.

Yet another weighted average of the probabilities $P[Y=y|Z,X]$ is given by the linear IV estimand for (2.1),

$$\beta_{IV} \equiv E\left[(Z,X)'(W,X)\right]^{-1}E\left[(Z,X)'Y\right] \tag{2.2}$$

$$= E\left[(Z,X)'(W,X)\right]^{-1}E\left[(Z,X)'\textstyle\sum_{y=1}^{J}yP[Y=y|Z,X]\right]. \tag{2.3}$$

The difference from β_{ITT} is due to the presence of $E\left[(Z,X)'(W,X)\right]$ in place of $E\left[(Z,X)'(Z,X)\right]$. Consequently, unlike β_{ITT}, the IV estimated β_{IV} does depend on W, and in particular on its joint distribution with both Z and X. Under easily interpreted conditions the IV estimand corresponds to a local average treatment effect (LATE), see Angrist, Imbens, and Rubin (1996). A key condition for the LATE interpretation is a monotonicity restriction, requiring that the direction of the effect that the instrument has on treatment is the same for all individuals. Monotonicity seems reasonable in the MTO application, as voucher assignment could reasonably be expected to increase the likelihood that any given family would move to a less distressed neighborhood. The LATE provided by β_{IV} corresponds to a weighted average of the average effect of W on Y among compliers. In the MTO context these are individuals who are induced to move by receiving a voucher, but who would not move without receiving a voucher. The β_{IV} parameter comprises a weighted average of these LATE's across different instruments as well as across different covariate values.[2]

[2]There is only one instrument in the MTO application, but it takes three different values. This instrument is equivalent to using two binary instruments, e.g. one for receiving a traditional voucher and another for receiving an experimental voucher. The parameter β_{IV} provides a weighted average of the LATEs for each of these binary

The weighted averages that all of these estimators measure are sensitive to the scale of y, so that different results are obtained if the support of y is for example $\{0, 1, 2\}$, or $\{0, 2, 4\}$. Assigning integers to the ordered outcomes gives them cardinal meaning when the data is only ordinal in nature. Ludwig et al. (2012) employed z-scores to encode ordinal Y, but this is only one of infinitely many possible cardinal transformations that would preserve the rank of the different categories. Because there is no natural scale on which to measure SWB, the resulting estimates correspond to weighted averages of inherently ordered outcomes whose magnitudes are not easily interpreted.

Aside from the issue of scale, these estimators have other notable drawbacks even if there were a single interpretable cardinal measure of SWB. β_{OLS} does nothing to deal with endogeneity. β_{ITT} measures the effect of the instrument Z on the outcome variable, but carries no information on the effect of the endogenous variable W on the outcome. β_{IV} deals with this by using linear IV to produce a LATE estimator. This provides information about the effect of manipulating the instrument. In the context of MTO, this targets the causal effect of neighborhood characteristics among those who would comply with receiving a housing voucher. However, the MTO housing voucher experiment seems unlikely to be widely replicated. Neither the ITT nor the LATE reveal the causal effect of a change in neighborhood characteristics induced by alternative means, such as policy changes or neighborhood evolution over time.

3 Complete Nonlinear Models

This section first considers the ordered probit model that does not allow for endogenous covariates, and then moves on to the triangular instrumental variable model that does allow for endogeneity. Expressions for counterfactual response probabilities and marginal effects applicable for both models as well as the single equation IV model of Section 4 are then derived.

3.1 The Ordered Probit Model

The ordered probit model specifies that ordered outcome $Y \in \{0, ..., J\}$ is determined by

$$Y = \left\{\begin{matrix} 0 & \text{if } W\beta + X\gamma + U \leq c_1, \\ 1 & \text{if } c_1 < W\beta + X\gamma + U \leq c_2, \\ \vdots & \vdots \\ J & c_J < W\beta + X\gamma + U. \end{matrix}\right\}, \tag{3.1}$$

where, in contrast to the treatment in the previous and subsequent section, *all* regressors are assumed exogenous, in the sense that $U \perp\!\!\!\perp (X, W)$. The values $0, ..., J$ for Y are labels used for

instruments.

ordered discrete outcomes and they play no role in the statistical analysis or in the policy use of the model other than as categorical labels. Random variable U is an unobserved exogenous variable, normally distributed with mean zero and unit variance. The thresholds $c_1, ..., c_J$ and vectors β, γ comprise model parameters.

This model is complete because for each realization of the exogenous observed variables (X, W) and the unobserved variables U, the endogenous variable Y is uniquely determined. For each $y \in \{0, ..., J\}$ and any realization (x, w) of the exogenous variables, the conditional probability that any $Y = y$ is given by the probability that normally distributed U lies within 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.

3.2 A Triangular IV Model

We now consider IV models that allow for potential endogeneity of W. We consider the same functional form for the ordered outcome given in (3.1) for the ordered probit model, where again X is a vector of observed exogenous variables, and U is an unobserved exogenous variable. As in the ordered probit 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 is determined as a function of X, instruments Z, and additional unobservable variables V. Each k-th component of W is determined by

$$W_k = X\delta_x^k + Z\delta_z^k + V_k, \quad k = 1, ..., k_w, \tag{3.2}$$

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

The vector of unobservables $(U, V_1, ..., V_k)$ 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 $V = (V_1, ..., V_K)$, assumed to be nonsingular. The $1 \times k_w$ vector

$$R \equiv (r_1, ..., r_{k_w})$$

is such that each r_k denotes the covariance of U with V_k. Both Σ_v and R comprise unknown parameters, while $\Sigma_{11} = 1$ incorporates the same scale normalization as in the ordered probit 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 (3.1) and (3.2) together with multivariate normality of (U, V),

$$\Pr[Y = y|x, z, w] = \\ \Phi\left(\frac{c_{y+1} - w\beta - x\gamma - R\Sigma_v^{-1}v(w, x, z)}{\sigma(v)}\right) - \Phi\left(\frac{c_y - w\beta - x\gamma - R\Sigma_v^{-1}v(w, x, z)}{\sigma(v)}\right), \quad (3.3)$$

where $\Phi(\cdot)$ denotes the standard normal CDF,

$$v(w, x, z) \equiv \left(w_1 - x\delta_x^1 - z\delta_z^1, ..., w_K - x\delta_x^K - z\delta_z^K\right)',$$

and

$$\sigma(v) \equiv 1 - R\Sigma_v^{-1}R'.$$

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 that uses estimated residuals from (3.2) 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 resulting second stage estimates can be used to consistently estimate all model parameters. Alternatively, (3.3) 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, as we shall see in the estimates reported in Section 6. The estimators derive from a model that 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 (3.2). Such knowledge of the process determining W may be questionable. In Section 4 implications of the model are investigated in the absence of such a restriction.

3.3 Counterfactual Probabilities and Marginal Effects

In order to define counterfactual probababilities and marginal effects, it is useful to first define individual response functions as

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

The value of $\mathbf{y}\left(x,w,u\right)$ denotes 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 $\mathbf{y}\left(\cdot,\cdot,u\right):\text{Supp}\left(X,W\right)\rightarrow\{0,1,...,J\}$ 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 shifted, 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 (3.1) in conjunction with the restrictions $U\perp\!\!\!\perp(X,Z)$ and $U\sim\mathcal{N}\left(0,1\right)$. The focus here is on such quantities conditional on $X=x$ in consideration of counterfactual shifts in the value of the endogenous variable W. Other counterfactual shifts can be considered in like manner.

The counterfactual probability that a person with observable characteristics $X=x$ randomly drawn from that subpopulation would achieve subjective well-being $y\in\{0,...,J\}$ if their neighborhood characteristics were exogenously shifted to w is given by

$$\begin{aligned}
p\left(0;x,w\right) &\equiv P\left[\mathbf{y}\left(x,w,U\right)=0|X=x\right]=\Phi\left(c_1-w\beta-x\gamma\right),\\
p\left(1;x,w\right) &\equiv P\left[\mathbf{y}\left(x,w,U\right)=1|X=x\right]=\Phi\left(c_2-w\beta-x\gamma\right)-\Phi\left(c_1-w\beta-x\gamma\right),\\
&\vdots\\
p\left(J;x,w\right) &\equiv P\left[\mathbf{y}\left(x,w,U\right)=J|X=x\right]=1-\Phi\left(c_{J-1}-w\beta-x\gamma\right),
\end{aligned}$$

equivalently

$$p\left(y;x,w\right)\equiv\Phi\left(c_{y+1}-w\beta-x\gamma\right)-\Phi\left(c_y-w\beta-x\gamma\right),\tag{3.4}$$

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

$$ME(\theta; y, x, w) \equiv \frac{\partial p(y; x, w)}{\partial w} = \beta \left(\phi\left(c_y - w\beta - x\gamma\right) - \phi\left(c_{y+1} - w\beta - x\gamma\right)\right). \tag{3.5}$$

Average partial effects are the averages of such quantities over the joint distribution of X and W.

In the ordered probit and triangular IV models laid out in this section parameters $\beta, \gamma, c_1, ..., c_{J-1}$ are point identified under mild conditions. The counterfactual probabilities and marginal effects in (3.4) and (3.5) are known smooth functions of these parameters. Thus, plugging in consistent and asymptotically normal point estimators for $\beta, \gamma, c_1, ..., c_{J-1}$ into these formulas results in consistent and asymptotically normal estimators for counterfactual probabilities and marginal effects, which are themselves asymptotically normal with variances obtained by way of the delta method.

4 Single Equation IV Models

In this section we continue to maintain the specification (3.1) for the determination of the ordered outcome Y as a function of W and X, but without assuming the "first stage" specification (3.2).

4.1 No Restrictions on the Influence of Instruments

To begin, 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 ordered probit model commonly used in the absence of covariate endogeneity, and which allows for W to be either discrete or continuously distributed.

Despite the parametric specification given by (3.1) and the normal distribution of U, this model is incomplete. That is, the realization of exogenous X and Z and unobservable U does not uniquely determine the realization of endogenous variables Y and W. This is precisely because the model is silent as to the determination of W. As a result, the joint distribution of Y and W conditional on exogenous variables (X, Z) is no longer pinned down by knowledge of the distribution of U together with 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, notice that the ordered response specification (3.1) ensures that the unobservable U lies in the interval running from $c_Y - X\gamma - W\beta$ to $c_{Y+1} - X\gamma - W\beta$.

Consider for the sake of argument any fixed interval $[s,t]$ on the real line. Then

$$\left\{ [c_Y - X\gamma - W\beta, c_{Y+1} - X\gamma - W\beta] \subseteq [s,t] \right\} \Longrightarrow \left\{ U \in [s,t] \right\}, \tag{4.1}$$

or in other words whenever 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 $[s,t]$. Since the first event implies the second, the probability of the former event conditional on exogenous variables provides a lower bound on the conditional probability of the latter event.

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 inequalities for all s,t pairs with $s \leq t$ characterize sharp bounds on the parameter vector $\theta \equiv (\beta, c_1, c_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \sigma)$.[3]

$$\begin{aligned}
\max_{x,z} \mathbb{P}\left[c\left(Y+1, X, W; \theta\right) \leq t | X = x, Z = z\right] &\leq \Phi\left(t\right), &(4.2)\\
\max_{x,z} \mathbb{P}\left[c\left(Y, X, W; \theta\right) \geq s | X = x, Z = z\right] &\leq 1 - \Phi\left(s\right), &(4.3)\\
\max_{x,z} \mathbb{P}\left[\left(s \leq c\left(Y, X, W; \theta\right); c\left(Y+1, X, W; \theta\right) \leq t\right) | X = x, Z = z\right] &\leq \Phi\left(t\right) - \Phi\left(s\right), &(4.4)
\end{aligned}$$

where

$$c\left(0, x, w; \theta\right) \equiv -\infty, \quad c\left(J+1, x, w; \theta\right) \equiv \infty,$$

and for $j = 1, ..., J$:

$$c\left(j, x, w; \theta\right) \equiv c_j - x\gamma - w\beta. \tag{4.5}$$

The inequalities (4.2)-(4.4) correspond to probability inequalities obtained by applying the conditional probability $\mathbb{P}\left[\cdot | X = x, Z = z\right]$ to the events in (4.1). The right hand side of these inequalities employ the normal CDF because U is restricted to be standard normal, and independent of (X, Z). The first two inequalities correspond to $s = -\infty$ and $t = +\infty$, respectively.[4]

Similar logic to that used for (4.1) yields the implication that

$$\left\{ U \in [s,t] \right\} \Longrightarrow \left\{ \left[c\left(Y, X, W; \theta\right), c\left(Y+1, X, W; \theta\right)\right] \cap [s,t] \neq \emptyset \right\}. \tag{4.6}$$

In words, whenever U is an element of the interval $[s,t]$, then $[s,t]$ must intersect with the interval $\left[c\left(Y, X, W; \theta\right), c\left(Y+1, X, W; \theta\right)\right]$. This is so because the model implies that

$$U \in \left[c\left(Y, X, W; \theta\right), c\left(Y+1, X, W; \theta\right)\right],$$

[3] Here s and t are elements of the extended real line, inclusive of $\pm\infty$.

[4] These inequalities may be alternatively represented by (4.2) alone, with the understanding that when $s = -\infty$ the inequality $s \leq c\left(y, x, w; \theta\right)$ always holds, as does the inequality $c\left(y+1, x, w; \theta\right) \leq t$ when $t = +\infty$, even when $c\left(y, x, w; \theta\right)$ and $c\left(y+1, x, w; \theta\right)$ are themselves $-\infty$ and $+\infty$, respectively.

and so U being an element of a set $[s,t]$ which does not intersect with $[c(Y,X,W;\theta), c(Y+1,X,W;\theta)]$ would be a contradiction. Probability inequalities can be derived from this implication as well, but they are implied by particular inequalities of the form (4.2)-(4.4).

It is interesting to compare the moment inequalities (4.2)-(4.4) to the implications of the ordered probit model and the complete triangular IV model. In the probit and triangular IV models, the strengthening of these inequalities to equalities for specific values of s and t stems from the fact that both the probit and triangular IV models are complete, i.e. that the value of all endogenous variables is uniquely determined as a function of exogenous observed and unobserved variables. In the probit 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 regions 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 (4.4) sum to one, and so must those on the right of (4.4), 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 ordered probit likelihood function. In the triangular IV model for any realization of (W,X,Z), $\mathbb{R}$ can also partitioned into intervals for U that uniquely determine the outcome, resulting in the strengthening of inequalities to equalities. 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 (3.3) that can be used for maximum likelihood estimation in the triangular model.

In contrast, in the incomplete single equation IV model there is no partition of $\mathbb{R}$ into intervals that uniquely determine the endogenous variables, and whose conditional probability can be written as a known function of model parameters. This is because values of U that produce a given realization of Y for a particular value of W can result in a different value of Y if W takes on a different value. W is not exogeneous, and is not uniquely determined as a function of unobservable U even after fixing exogenous variable (X,Z). Consequently the support of the unobservable variables of the model cannot be partitioned into unique regions for each realization of the endogenous variables as it can in the probit and triangular IV models. The observable implications thus take the form of inequalities (4.2)-(4.4) 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 variate 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 that parameters θ satisfy

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

for all pairs $s < t$ on the extended real line. When the stochastic independence restriction is further 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

$$\max_{x,z} \mathbb{P}\left[c\left(Y+1,X,W;\theta\right) \leq 0 | X=x, Z=z\right] \leq \frac{1}{2}, \tag{4.7}$$
$$\max_{x,z} \mathbb{P}\left[c\left(Y,X,W;\theta\right) \geq 0 | X=x, Z=z\right] \leq \frac{1}{2}. \tag{4.8}$$

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

4.2 Counterfactual Probabilities and Marginal Effects

The expressions (3.4) and (3.5) are known functions of θ, and are both point identified in the ordered probit and triangular IV models. When instead the single equation IV model is used, with or without either treatment choice restriction, then these expressions for counterfactual probabilities and marginal effects remain valid, but θ is in general only partially identified. Consequently, identified sets for counterfactual probabilities or marginal effects are characterized as the set of all possible values for (3.4) and (3.5), respectively, taking θ across all values in the identified set. In Section 6.2.3 we report set estimates and confidence sets for marginal effects using such a characterization when employing the partially identifying single equation IV model, thus providing a sensitivity analysis for the triangular IV model point estimates presented in Section 6.2.2.

5 Moving to Opportunity

Moving to Opportunity (MTO) was a unique randomized housing voucher experiment that was implemented from 1994-1998. Through the program, 4,604 volunteer families living in "extreme-poverty" neighborhoods in Baltimore, Boston, Chicago, Los Angeles, and New York were randomly assigned to one of three treatments.[5] The treatment assignments were receipt of a low poverty experimental voucher (which could be used only if the family moved to a low poverty area), the MTO traditional section 8 voucher (which could be used without any location restriction), or no voucher (control group). 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 instrumental variable for neighborhood characteristics. That is, the experimental design guarantees that Z is independent of unobservable variables that might effect neighborhood choice. The outcome data we use is taken from long-term data on outcomes recorded 10-15 years after assignment. The subjective well-being 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 happy, pretty happy, or not too happy?"[6] The corresponding outcome is $Y \in \{0, 1, 2\}$, ranging from least to most happy.[7] 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. 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, Liebman, and Katz (2007), Ludwig et al. (2012), Pinto (2015), and Chetty, Hendren, and Katz (2016). Kling, Liebman, and Katz (2007) used medium term (4 to 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, Hendren, and Katz (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. Pinto (2015) developed a model that combines revealed preference analysis with tools from the literature on Causal Bayesian Networks 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

[5] "Extreme-poverty" neighborhoods are those in which at least 40% of residents' income lies below the U.S. federal poverty threshold.

[6] Further details about the GSS are available at http://www3.norc.org/GSS+Website/.

[7] Specifically, the values of Y correspond to 0 for 'not too happy', 1 for 'pretty happy' and 2 for 'happy'.

than is taken here. The measured effects align qualitatively with those of Kling, Liebman, and Katz (2007), but produced statistically significant estimates for some labor market outcomes such as total income when the effect of the neighborhood is isolated. As previously discussed, when outcomes are ordered discrete variables such as SWB, average effects are not easily interpreted, lending motivation to the IV approach taken here.

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.[8]

Data Description

The neighborhood characteristics that we examine are residential poverty and share minority. Residential poverty is estimated using the z-score of duration weighted share poor in an individual's neighborhood while share minority is estimated 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 a duration weighted neighborhood characteristic, using the control group weighted average and standard deviation. Figures 1 and 2 show the distributions of these variables across different treatment groups. As is clear from Figure 1, adults belonging to the low poverty experimental voucher group lived in less poor neighborhoods

[8] In 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.

than either the MTO traditional section 8 voucher group or the control group. From Figure 2 these adults also live in neighborhoods that had fewer minority residents, but the difference from the MTO traditional section 8 voucher group or the control group is smaller than for neighborhood poverty.

Table 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 from the Table, these covariates are quite balanced across different treatment arms.

We also use weights in our empirical analysis to account for differences in random assignment proportions across sites and time as well as various aspects of survey administration. These are the same weights as Ludwig et al. (2012).[9]

6 Empirical Results

6.1 Linear model estimates

As a first step linear model parameters are estimated using the ICPSR MTO data. These linear models are the same as the ones used by 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.

Results using the linear model (2.1) are presented in Panel A of Table 2. Results are given using the poverty rate and minority rate 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, which may not be a credible assumption.

In columns (1)-(3) of Panel A in Table 2, 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 neighborhood minority on SWB. When both neighborhood poverty and minority are included, the negative effect of neighborhood 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 1) is included, and the results remain qualitatively unchanged.

Interactions of MTO assignment and randomization site are used as instrumental variables Z in

[9]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.

the results reported in Panel B of Table 2. 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 2 report results without the inclusion of additional 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 2 are qualitatively similar, with minor differences likely caused by the aforementioned differences in the two estimation samples. The coefficient on neighborhood poverty on SWB in Table S5 is -0.141 while here it has been estimated as -0.096 (both with 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 3 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 of Ludwig et. al (2012). Specifically, the coefficient on MTO voucher assignment is the ITT effect. Columns (1)-(3) of Table 3 report ITT estimates without including a complete set of covariates while columns (4)-(6) report ITT estimates with inclusion of such 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 3 is slightly larger. Estimates still indicate a positive and statistically significant effect of being offered an MTO voucher on SWB.

6.2 Nonlinear Model Estimates

This section reports results using non-linear models to estimate the effect of neighborhood characteristics on SWB. We start with the ordered probit model and then report estimates using the triangular IV model of Section 3 that allows for endogeneity of neighborhood characteristics.

6.2.1 Ordered Probit Estimates

Columns (1)-(3) of Table 4 report parameter estimates for the ordered probit model with only dummy variables for randomization site used for 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.

Figures 3 and 4 give the predicted probabilities that a person reports 'not too happy' (0) or 'very happy' (2) across different values of neighborhood characteristics. These are conditional on the adult being in NY. Moreover, other exogenous explanatory variables X are held fixed at the NY sample median values. So for instance, 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). As expected, Figure 3 shows that the probability of being 'not too happy' for such a person increases with neighborhood poverty and the probability of being 'very happy' decreases with neighborhood poverty. Figure 4 illustrates that this also holds true when the level of neighborhood poverty is changed while holding neighborhood minority constant at the NY sample median. Figure 3 also shows that the probability of being 'not too happy' is increasing in neighborhood minority and the probability of being 'very happy' is declining in neighborhood minority. However, Figure 4 illustrates that the effect of changes in neighborhood minority on SWB is relatively small once neighborhood poverty is held constant at the NY median value, so that the predicted probability is almost flat across changes in neighborhood minority.

Figures 5 and 6 present the corresponding marginal effects (3.5). Figures 5 and 6 indicate that neighborhood poverty has a negative effect on SWB, with and without holding neighborhood minority constant. Also from these Figures the effect of neighborhood minority on SWB is less clear; there is some negative effect without controlling for neighborhood poverty in Figure 5 but these effects are close to zero once neighborhood poverty is held constant at the NY sample median as may be seen in Figure 6.

These estimates are broadly in line with the signs and statistical significance of effects measured using linear models, despite the difference in the 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 neighborhood poverty; conversely a positive statistically significant effect on the probability of being 'not too happy' with a unit increase in neighborhood poverty. Relative to linear model estimates, the nonlinear model estimates enable construction of heterogeneous and intrepretable counterfactual probabilities and marginal effects.

6.2.2 Triangular IV Model Estimates

The restriction requiring that neighborhood characteristics are exogenous is now relaxed, and estimates are reported for the triangular IV model presented in Section 3. The endogenous variables W are duration-weighted z-scores for neighborhood poverty and neighborhood minority. In all cases estimation is carried out by maximum likelihood.

Parameter estimates for the outcome equation (3.1) are given in Table 5.[10] Neighborhood poverty has a negative and statistically significant effect across all specifications. Neighborhood minority has a negative but statistically insignificant effect when it is included in the absence of the neighborhood minority variable. When both variables are included, neighborhood minority has a positive effect on SWB, although it 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 neighborhood minority is not statistically significantly different from zero when neighborhood poverty is not included accords with the linear IV estimates reported in Table 2. Likewise, so do the findings that neighborhood poverty has a negative impact on SWB, and the finding that when both neighborhood characteristics are included neighborhood minority has a (borderline significant) positive impact. The ordered probit model estimates reported in the previous section, which did not allow for endogeneity of neighborhood characteristics, did not produce a positive effect for neighborhood minority.

Like the ordered probit model, however, the triangular 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. Figures 7 and 8 give predicted probabilities 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. Figure 7 shows that the probability of being 'not too happy' for such a person is increasing in neighborhood poverty and the probability of being 'very happy' is declining in neighborhood poverty. Figure 7 also indicates that the probability of being 'not too happy' is increasing in neighborhood minority and the probability of being 'very happy' decreasing with neighborhood minority. Note that the specifications used in the top panels of Figure 7 include

[10] Estimates from the first stage are reported in Table 10 in Appendix C.

only neighborhood poverty and those in the bottom panels of Figure 7 include *only* neighborhood minority as endogenous variables.

Figure 8 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 8 shows that the direction of change in the probability of SWB taking on different values across values of neighborhood poverty remains the same as in Figure 7. Compared to Figure 7, Figure 8 provides a different picture for the effect of neighborhood minority on SWB. With neighborhood poverty held fixed, the probability of being 'not too happy' is decreasing in neighborhood minority and the probability of being 'very happy' is flat at very low levels and then increasing in neighborhood minority.

Figures 9 and 10 show the estimated conditional marginal effects. As before the figures indicate a clear negative effect of higher neighborhood poverty on subjective well-being. The effect of neighborhood minority on SWB, as also indicated in Table 5 is found to be slightly negative but statistically insignificant without holding neighborhood poverty fixed as in Figure 9. Yet in Figure 10 when both variables are included with neighborhood poverty fixed at the sample median, the marginal effects indicate that SWB is generally increasing in neighborhood minority with varying levels of statistical significance.

What then can be concluded from these results further to what can been learned using linear IV? To answer this question, consider what can be inferred from Figures 7 - 10 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 degree of the effect of a change in the neighborhood poverty will in general differ with the neighborhood poverty level at which the effect is measured. That is, the model allows for estimation of heterogeneous effects across values of observable variables. 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 triangular IV model relative to the linear IV model rely on the suitability of the restrictions embodied in the structural equations (3.1) and (3.2) and the joint normality of unobserved variables. In order to investigate, information matrix (IM) tests of the triangular IV 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.[11]

[11]See 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

For specifications both including and excluding the neighborhood minority variable, 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 neighborhood poverty, on 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 neighborhood poverty and all city dummy variables for the specification in which neighborhood minority is omitted, with a resulting p-value of 0.0554.

The results in hand indicate that the IM test statistic for the triangular IV 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 neighborhood minority is included. Accordingly some exploration of the sensitivity of the estimates to a relaxation of the triangular IV 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 equation and the joint normality restriction.

6.2.3 Sensitivity Analysis with Single Equation IV Models

Table 6 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 single equation IV (SEIV) model. These are compared to point estimates and confidence intervals obtained using the more restrictive triangular IV (TIV) specification, the results of which were presented in Section 6.2.2. These conditional marginal effects, defined in (3.5), depend 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 observed median in New York, namely -0.2014303 for neighborhood poverty and 0.4906762 for neighborhood minority.[12] The TIV estimates and confidence intervals are reported both with and without the inclusion of additional exogenous variables. When included, the marginal effects are also measured conditional on their median values, as in Section 6.2.2. Results reported for the SEIV model use only city dummies as included exogenous variables for computational tractability.

To compute the SEIV estimates and confidence intervals, moment inequalities were used as described in Section 4 at specified values of s, t, x, and z for the SEIV inequalities. For values of s, t in (4.2)-(4.4) we used all pairs $s < t$ with $s, t \in \left\{-\infty, \Phi^{-1}(0.1), \Phi^{-1}(0.2), ..., \Phi^{-1}(0.9), \infty\right\}$, 54

of the test in terms of uncontrolled parameter heterogeneity.

[12]These are same median values used for these variables in results reported in application of the ordered probit and triangular IV models for estimation of marginal effects 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.

pairs in total.[13] Since the support of (X, Z) comprises 15 distinct values, there were consequently 810 inequalities (4.2)-(4.4) in the SEIV specification. We thus employed an inference method set out in Belloni, Bugni, and Chernozhukov (2018) which is specifically designed for inference on functions of parameters that are bounded by a large number of moment inequalities.

Specifically, we report analog set estimates, median corrected set estimates, and 95% confidence intervals on these conditional marginal effects, which we here denote by $g(\theta)$ for convenience, where $g(\theta)$ corresponds to the conditional marginal effect defined in (3.5). Each of these sets are of the form

$$\left\{ r : \inf_{\{\theta : g(\theta) = r\}} \hat{Q}(\theta) \leq c_n \right\}, \tag{6.1}$$

for an appropriately defined value of c_n described below, where

$$\hat{Q}(\theta) \equiv \max_j \sqrt{n} \frac{\hat{m}_j(\theta)}{\hat{\sigma}_j(\theta)}, \tag{6.2}$$

and each $j \in \{1, ..., 810\}$ corresponds to distinct quadruplets (s, t, x, z), denoted (s_j, t_j, x_j, z_j). The quantities $\hat{m}_j(\theta)$ and $\hat{\sigma}_j(\theta)$ are given by

$$\hat{m}_j(\theta) \equiv \hat{E}_n[\omega_j(Y, W, X, Z)], \qquad \hat{\sigma}_j(\theta) \equiv \hat{E}_n\left[(\omega_j(Y, W, X, Z) - \hat{m}_j(\theta))^2\right],$$

where

$$\begin{aligned} &\omega_j(Y, W, X, Z) \\ &\quad \equiv (1[s_j \leq c_Y - X\gamma - W\beta \wedge c_{Y+1} - X\gamma - W\beta \leq t_j] - (\Phi(t) - \Phi(s))) \cdot 1[X = x_j, Z = z_j], \end{aligned}$$

and $\hat{E}_n[\cdot]$ denotes the sample mean, such that for example

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

Each of the $J = 810$ terms over which the maximum is taken in (6.2) is the sample counterpart of one of the inequalities of the form (4.2)-(4.4) for the corresponding value of (s, t, x, z), multiplied by $\mathbb{P}[(X, Z) = (x, z)]$.[14] Since each $\mathbb{P}[(X, Z) = (x, z)]$ is positive, scaling the inequalities in this way does not alter the region defined by the inequalities, but it avoids the use of ratios of sample estimates as would be required without scaling. It also results in a test statistic of the form used by Belloni, Bugni, and Chernozhukov (2018), which we use for inference.

Analog set estimates are obtained simply by setting $c_n = 0$. The resulting set is comprised of values of the conditional marginal effect for which there is some θ such that all sample inequalities

[13] The pair $(s, t) = (-\infty, \infty)$ is trivially uninformative and was therefore excluded.

[14] The counterpart of (4.2) and (4.3) are obtained when $s_j = -\infty$ and when $t_j = +\infty$, respectively.

$\hat{m}_j(\theta) \leq 0$ hold, equivalently such that $\max_j \hat{m}_j(\theta) \leq 0$. In finite samples this analog set will be inward biased, because 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 set will be inward biased. The corrected set estimates in Table 6 provide a correction for this by setting c_n in (6.1) 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, Lee, and Rosen (2013). The set (6.1) with this value of c_n provides an asymptotic 50% confidence set for the conditional marginal effect. The third set estimate provided is a 95% confidence set. For the choice of c_n for the corrected set estimates and 95% confidence sets, we used the self-normalized critical value

$$c_n(J,\alpha) = \frac{\Phi^{-1}(1-\alpha/J)}{\sqrt{1-\Phi^{-1}(1-\alpha/J)^2/n}} \tag{6.3}$$

proposed by Belloni, Bugni, and Chernozhukov (2018) for $\alpha \in \{0.5, 0.05\}$, respectively, with $J = 810$. The self-normalized critical values generally provide conservative asymptotic inference, but have the advantage that they do not depend on θ and are easy to compute.

Computation of analog set estimates and confidence intervals for marginal effects and counterfactual response probabilities $g(\theta)$ based on (6.1) was conducted as follows.[15] For each such target parameter $g(\theta)$ a separate search was conducted over regions of the parameter space in which β is nonpositive and nonnegative.[16] The first step of each search was a search for the minimum of $\hat{Q}(\theta)$, defined in (6.2), among all values of θ with the corresponding sign for β, with all parameters restricted to the hyperrectangle with each component bounded by ± 2.5. Starting values for this search included point estimates from the triangular IV and ordered probit specifications when compatible with the sign restriction on β, as well as 10 randomly drawn starting values. The value of target parameter $\hat{g}^* \equiv g(\hat{\theta}^*)$ was then computed at the minimizing value of θ, $\hat{\theta}^*$. A grid of starting values was constructed in increments of 0.05 spanning the relevant parameter space for

[15]Computations were done in R (R Core Team (2019)) using the nloptr package (Johnson (2007–2019), Ypma (2018)) for optimization. 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 in conjunction with a C++ implementation of the discrepancy function $\hat{Q}(\theta)$ defined below in order to speed up computation of the 810 sample moments and variances for each θ evaluated.

[16]The 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 (forthcoming) 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. Thus we searched separately over regions of the parameter space in which β is restricted to be nonpositive and nonnegative.

$g(\theta)$ above and below this minimizing value $\hat{g}^*$.[17] At each point r on this grid, the value of $\hat{Q}(\theta)$ was minimized subject to the constraint that $g(\theta) = r$. Regions of the form (6.1) were then constructed by then evaluating points r starting from the maximum (minimum) grid point outside (6.1) less than (greater than) $\hat{g}^*$, incrementally increasing (decreasing) r by 0.0005 until a point was found to be in the set (6.1). The tables below report minimal and maximal values of $g(\theta)$ afforded by this search over each such region, rounded to the nearest three decimal places.[18,19]

Consider first the results from the SEIV specification reported in Table 6 in comparison to those obtained from the triangular IV specification (TIV). For the marginal effect of neighborhood poverty on respondents answering they are either "not too happy" or "happy" we see that the SEIV interval estimates are fairly wide in terms of magnitude, but agree in sign with the TIV estimates. For the marginal effect on respondents answering "not too happy", the TIV point estimates are near the lower bound of the SEIV point estimates. Similarly, the marginal effect on respondents answering they are "happy" is measured as slightly negative using the TIV model, but can range over a much wider domain under the SEIV specification, all the way to -0.290 in consideration according to the analog estimate. Thus, if one doubts the veracity of the second equation of the triangular IV model, one must consider the possibility that the effect of neighborhood poverty on happiness could be substantially stronger than what the triangular IV model implies. Moreover, the effect of sampling variation on both the TIV and SEIV estimates is not negligible, and the corrected SEIV interval estimates and 95% confidence intervals are substantially wider than both the SEIV analog estimates and the TIV 95% confidence intervals. Nonetheless, they agree in sign with the SEIV estimates, ruling out the possibility that a marginal increase in neighborhood poverty results in an increase in self-reported happiness.

The results reported in Table 7 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 6. 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 SEIV analog interval estimates always include the TIV point estimates, and never overturn their sign, although now both intervals include marginal effects

[17] When $g(\theta)$ was a counterfactual response probability the parameter space was set to $[0, 1]$. When $g(\theta)$ was a marginal effect it was set to either $[-0.8, 0]$ or $[0, 0.8]$, depending on the sign of β in the region subject to search. The upper bound of 0.8 on the magnitude of marginal effects did not appear to bind.

[18] The reported confidence regions comprise intervals and in some cases unions of intervals. The algorithm implies that the extreme points of such intervals were found to belong the corresponding interval, but there may be values inside these intervals that would not pass the criterion (6.1) for inclusion. Furthermore the profiled discrepancy function $\hat{D}(r) = \min_{\{\theta: g(\theta)=r\}} \hat{Q}(\theta)$ need not be monotonic or even continuous in r. Thus, while it cannot be guaranteed with complete certainty that there are no points outside the reported regions that would result in a sufficiently low discrepancy to be included in the reported sets, the algorithm employed to construct these regions attempted to explore these regions as thorougly as possible given the irregular nature of the problem and computational constraints.

[19] Note that when zero is included in a region this indicates that a value within 0.0005 of zero was found to be in the region over which the search was conducted.

of zero in addition to effects much larger in magnitude than those obtained from the TIV model.[20] Furthermore, once sampling variation is taken into account, we see that the 95% confidence sets no longer sign either marginal effect. This is just barely the case for the marginal effect on respondents answering they are "not too happy", with only a small range in the negative region included in the confidence set, but the confidence set for the marginal effect on answering "happy" includes a wide range of both negative and positive values. The ability to measure this marginal effect to any reasonable accuracy thus seems especially sensitive to the inclusion of the second equation in the TIV specification.

Tables 8 and 9 report point and set estimates and 95% confidence regions on counterfactual response probabilities using the TIV and SEIV specifications. The probabilities in Tables 8 and 9 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 and one standard error below the median level in New York, respectively. Both tables present results for the subpopulation of individuals with exogenous covariates X at the same values as used for Table 6, using specifications in which the neighborhood minority rate was not included. In each case we see that analog estimates from the SEIV model contain point estimates from the TIV model. As was also the case when estimating marginal effects, the sensitivity analysis afforded by the SEIV specification generally accords with the TIV results, but indicates that without the additional restrictions used in the TIV specification, the ranges of possible values of counterfactual probabilities are much wider in magnitude. For these probabilities, the SEIV 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 TIV estimates for this probability are close to the lower end of what is indicated by the SEIV specification. Thus, if the additional restrictions of the TIV specification are incorrect, it could be that the TIV specification substantially under estimates the fraction of individuals that would report they were happy at these neigborhood poverty levels, in particular for the counterfactual in which neighborhood poverty is lowered by one standard error.

7 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 both problematic when used 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

[20]Note that while the SEIV bound estimates would generally be expected to contain the TIV point estimates, this need not occur if the TIV model is misspecified.

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. The use of such models additionally enables the measurement of ceteris paribus effects, 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 Moving to Opportunity 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 triangular instrumental variable model specification. As is the case with any structural model, the results rely on the restrictions employed by the model that is used. Thus, we turned to consideration of partially identifying single equation IV models to compute set estimates for marginal effects. These set-identifying models nested the triangular IV 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 triangular IV 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 triangular IV 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

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(434), 444–455.

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

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

Bond, T. N., and K. Lang (2019): "The Sad Truth About Happiness Scales," *Journal of Political Economy*, 127(4), 1629–1640.

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., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chesher, A. (1983): "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation," *Economics Letters*, 13(1), 45–48.

——— (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 52(4), 865–872.

——— (2013): "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *Econometric Theory*, 29(2), 231–266.

Chesher, A., and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85(3), 959–989.

——— (forthcoming): "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, vol. 7a. Elsevier.

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(4), 855–902.

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

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

Eddelbuettel, D., and R. François (2011): "Rcpp: Seamless R and C++ Integration," *Journal of Statistical Software*, 40(8), 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.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46(4), 931–959.

Helliwell, J., R. Layard, and J. Sachs (eds.) (2012): *World Happiness Report.* Earth Institute, Columbia University.

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

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(1), 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(4), 705–735.

Pinto, R. (2015): "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity," working paper, UCLA.

R Core Team (2019): *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rivers, D., and Q. H. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39(3), 347–366.

Sanderson, C., and R. Curtin (2016): "Armadillo: a template-based C++ library for linear algebra.," *Journal of Open Source Software*, 1(26).

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(2), 679–714.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.

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

Appendices

A Details of Bound Derivations

This section provides further mathematical detail for the derivation of bounds for the SEIV model presented in Section 4.1. To proceed with set identification analysis for model parameters $\theta \equiv (\beta, \gamma, c_1, c_2)$, define the sets

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

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

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

as well as the conditional *capacity* inequality

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

where

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

In the context of the single equation IV model, the capacity and containment functional inequalities take a particular form, which is now derived. Define for $y \in \{0, 1, 2, 3\}$, $x \in \{0, 1, 2, 3, 4\}$ and any $w \in \text{Supp}(W)$, the function $c(y, x, w; \theta)$ as follows.

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

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

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

with the lower (upper) bound of the interval understood to be open in the event $c(Y, X, W; \theta) = -\infty$ $(= +\infty)$.[21]

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

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

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

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

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

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

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

$$C_\theta\left(\left[t_1,t_2\right]|x,z\right) = \mathbb{P}\left[t_1 \leq c\left(Y,X,W;\theta\right) \wedge c\left(Y+1,X,W;\theta\right) \leq t_2 | X=x, Z=z\right], \tag{A.2}$$

$$\overline{C}_\theta\left(\left[t_1,t_2\right]|x,z\right) = \mathbb{P}\left[c\left(Y,X,W;\theta\right) \leq t_2 \wedge c\left(Y+1,X,W;\theta\right) \geq t_1 | X=x, Z=z\right]. \tag{A.3}$$

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

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

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

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

If we continue to assume that $U \perp\!\!\!\perp (X,Z)$ but without imposing $U \sim \mathcal{N}\left(0,\sigma^2\right)$, we have from Chesher and Rosen (2017) Corollary 3 that bounds on θ are given by the following inequalities for all $t_1, t_2 \in \mathbb{R}_{\pm\infty}$ with $t_1 < t_2$, where $\mathbb{R}_{\pm\infty}$ denotes the extended real line (i.e. inclusive of $\pm\infty$):

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

Substitution for C_θ and $\overline{C}_\theta$ then delivers the inequalities displayed in the main text. With this assumption in place we require a location normalization, for which we can use the restriction that

Median $(U|X,Z)=0$, giving the inequalities

$$\max_{x,z} C_\theta\left(\left(-\infty,0\right]|x,z\right) \le \frac{1}{2} \le \min_{x,z} \overline{C}_\theta\left(\left(-\infty,0\right]|x,z\right). \tag{A.4}$$

If we then drop the independence restriction $U \perp\!\!\!\perp (X,Z)$ and replace it with only the weaker restriction that Median $(U|X,Z)=0$, we obtain the inequalities given in (4.7) and (4.8).

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

B Tables and Figures

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

	Experimental	Section 8	Control
Site:			
Baltimore	0.1378	0.1435	0.1430
Boston	0.1920	0.1893	0.2195
Chicago	0.2651	0.1634	0.1621
Los Angeles	0.1892	0.2183	0.2723
New York	0.2159	0.2855	0.2031
Demographic characteristics:			
African American (non-hispanic)	0.6749	0.5865	0.6274
Hispanic ethnicity (any race)	0.2801	0.3624	0.3167
Female	0.9886	0.9767	0.9780
<= 35 years old	0.1375	0.1408	0.1458
36-40 years old	0.2125	0.2343	0.2218
41-45 years old	0.2470	0.2228	0.2348
46-50 years old	0.1904	0.1891	0.1808
Never married	0.6356	0.6126	0.6418
Parent while younger than 18 years old	0.2578	0.2565	0.2521
Working	0.2676	0.2765	0.2449
Enrolled in school	0.1583	0.1757	0.1614
High school diploma	0.3980	0.3511	0.3679
General Education Development (GED) certificate	0.1632	0.1847	0.1876
Receiving Aid to Families with Dependent Children (AFDC)	0.7681	0.7359	0.7764
Household characteristics:			
Household income (dollars)	12,659	12,799	12,655
Household owns a car	0.1734	0.1802	0.1734
Household member had a disability	0.1536	0.1743	0.1462
No teens in household	0.6097	0.6244	0.6375

continued on next page

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

	Experimental	Section 8	Control
Household size is <= 2	0.2103	0.2198	0.1976
Household size is 3	0.3017	0.3023	0.3233
Household size is 4	0.2405	0.2351	0.2195
Neighborhood characteristics:			
Household member was a crime victim in past 6 months	0.4293	0.4186	0.4110
Neighborhood streets very unsafe at night	0.4898	0.5357	0.5122
Very dissatisfied with neighborhood	0.4710	0.4831	0.4490
Household living in neighborhood > 5 years	0.6001	0.6264	0.5984
Household moved more > $3x$ in last 5 yrs	0.0913	0.0782	0.1089
Household has no family living in neighborhood	0.6252	0.6351	0.6346
Household has no friends living in neighborhood	0.4000	0.4052	0.4087
Household head chatted with neighbor >= $1x$ per week	0.5286	0.5025	0.5447
Household head very likely to to tell on neighborhood kid	0.5459	0.5246	0.5648
Household head very sure of finding apartment	0.4718	0.5046	0.4582
Housheold head applied for Section 8 before	0.3892	0.3974	0.4349
Primary or secondary reason for wanting to move:			
Want to move to get away from gangs and drugs	0.7827	0.7497	0.7816
Want to move for better schools for children	0.4876	0.5430	0.4710
Want to move to get a bigger/better apartment	0.4469	0.4370	0.4649
Want to move to get a job	0.0655	0.0477	0.0614
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 7/3,273 observations with missing SWB, 3/3,273 observations with missing neighborhood characteristics and 89/3,273 observations with missing household income.

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

Table 2: Linear model estimation (OLS and IV) of neighborhood effects on SWB

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

Notes: The dependent variable is Subjective Well Being (SWB) which takes the value zero for not too happy, one for pretty happy and two for very happy; columns (1)-(3) use a set of dummy variables for randomization site as covariates X while columns (4)-(6) use a complete set of baseline characteristics (as given in Table 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: Linear model estimation (ITT) of neighborhood effects on SWB

	(1)	(2)	(3)	(4)	(5)	(6)
Z = Any MTO voucher	0.0636**			0.0660**		
	(0.0285)			(0.0284)		
Z = MTO low poverty voucher		0.0517*			0.0546*	
		(0.0301)			(0.0298)	
Z = MTO section 8 voucher			0.0803**			0.0875**
			(0.0387)			(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 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: Ordered probit estimation of neighborhood effects on SWB

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

	(1)	(2)	(3)	(4)	(5)	(6)
$\beta_{Poverty}$	-0.1487**		-0.2876***	-0.1609**		-0.3141***
	(0.0630)		(0.0994)	(0.0642)		(0.1119)
$\beta_{Minority}$		-0.0619	0.3417*		-0.1147	0.3656
		(0.1359)	(0.2042)		(0.1399)	(0.2322)
$\gamma_{Baltimore}$	0.2778***	0.3131***	0.3823***	0.1814**	0.1949	0.3127***
	(0.0809)	(0.1052)	(0.0965)	(0.0906)	(0.1200)	(0.1153)
γ_{Boston}	0.1435*	0.1531	0.4967**	0.0526	0.0093	0.3779*
	(0.0771)	(0.1954)	(0.2243)	(0.0822)	(0.1801)	(0.2253)
$\gamma_{Chicago}$	0.2989***	0.2676***	0.2363***	0.1803**	0.1591*	0.1394*
	(0.0694)	(0.0742)	(0.0788)	(0.0808)	(0.0831)	(0.0845)
γ_{LA}	0.0893	0.1045	0.0688	0.0663	0.0690	0.0600
	(0.0640)	(0.0647)	(0.0628)	(0.0726)	(0.0729)	(0.0703)
c_1	-0.4750***	-0.5247***	-0.3704***	-0.7291***	-0.6788**	-0.8070***
	(0.0508)	(0.0488)	(0.0923)	(0.2599)	(0.2694)	(0.2528)
c_2	0.9119***	0.8616***	0.9403***	0.6944***	0.7409***	0.5407*
	(0.0502)	(0.0494)	(0.0508)	(0.2602)	(0.2675)	(0.2803)
ρ	0.0668	0.0036		0.0769	0.0676	
	(0.0681)	(0.1400)		(0.0675)	(0.1383)	
ρ_1			0.0563			0.0732
			(0.0662)			(0.0653)
ρ_2			-0.2718			-0.2579
			(0.1702)			(0.1844)
$cov(v_1, v_2)$			0.4502***			0.4200***
			(0.0242)			(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 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.

	TIV (1)	TIV (2)	SEIV
marginal effect on "not too happy"			
analog estimate	0.052	0.062	$[0.044, 0.293]$
corrected estimate	–	–	$[0.000, 0.445]$
95% interval	$[0.008, 0.096]$	$[0.013, 0.110]$	$[0.000, 0.483]$
marginal effect on "happy"			
analog estimate	-0.040	-0.040	$[-0.290, 0.000]$
corrected estimate	–	–	$[-0.398, 0.000]$
95% interval	$[-0.074, -0.007]$	$[-0.074, -0.007]$	$[-0.417, 0.000]$

Table 6: Triangular IV and Single Equation 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) taken with respect to neighborhood poverty with neighborhood minority excluded, at the New York median level. Columns TIV (1) and TIV (2) correspond to the triangular IV model results that exclude and include additional exogenous covariates, respectively.

	TIV (1)	TIV (2)	SEIV
marginal effect on "not too happy"			
analog estimate	0.096	0.115	$[0.000, 0.348]$
corrected estimate	–	–	$[0.000, 0.498]$
95% interval	$[0.034, 0.159]$	$[0.038, 0.192]$	$[-0.041, 0.000] \cup [0.000, 0.498]$
marginal effect on "happy"			
analog estimate	-0.089	-0.081	$[-0.341, 0.000]$
corrected estimate	–	–	$[-0.490, 0.000] \cup [0.000, 0.373]$
95% interval	$[-0.162, -0.016]$	$[-0.160, -0.002]$	$[-0.540, 0.000] \cup [0.000, 0.376]$

Table 7: Triangular IV and Single Equation 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) taken with respect to neighborhood poverty with neighborhood minority included, at the New York median level. Columns TIV (1) and TIV (2) correspond to the triangular IV model results that exclude and include additional exogenous covariates, respectively.

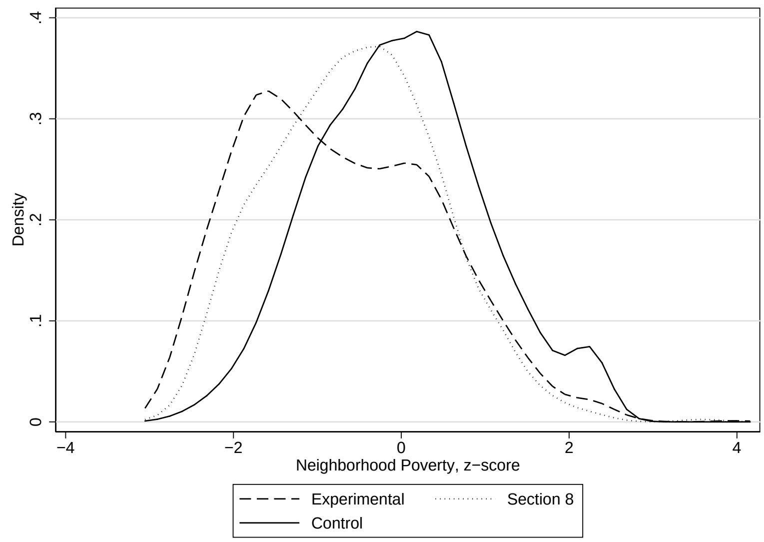
	TIV (1)	TIV (2)	SEIV
counterfactual response probability for "not too happy", NY median poverty.			
analog estimate	0.307	0.387	$[0.140, 0.539]$
corrected estimate	–	–	$[0.053, 0.589]$
95% interval	$[0.274, 0.339]$	$[0.299, 0.475]$	$[0.044, 0.610]$
counterfactual response probability for "pretty happy", NY median poverty.			
analog estimate	0.504	0.485	$[0.169, 0.718]$
corrected estimate	–	–	$[0.000, 0.866]$
95% interval	$[0.484, 0.525]$	$[0.442, 0.528]$	$[0.000, 0.917]$
counterfactual response probability for "happy", NY median poverty.			
analog estimate	0.189	0.128	$[0.093, 0.392]$
corrected estimate	–	–	$[0.011, 0.557]$
95% interval	$[0.164, 0.214]$	$[0.079, 0.177]$	$[0.009, 0.607]$

Table 8: Triangular IV and Single Equation 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) taken with respect to neighborhood poverty with neighborhood minority excluded, at the New York median level. Columns TIV (1) and TIV (2) correspond to the triangular IV model results that exclude and include additional exogenous covariates, respectively.

	TIV (1)	TIV (2)	SEIV
counterfactual response probability for "not too happy", NY median poverty - 1 s.e.			
analog estimate	0.256	0.326	$[0.041, 0.554]$
corrected estimate	–	–	$[0.004, 0.754]$
95% interval	$[0.210, 0.302]$	$[0.236, 0.416]$	$[0.002, 0.804]$
counterfactual response probability for "pretty happy", NY median poverty - 1 s.e.			
analog estimate	0.512	0.509	$[0.175, 0.674]$
corrected estimate	–	–	$[0.000, 0.858]$
95% interval	$[0.492, 0.531]$	$[0.476, 0.541]$	$[0.000, 0.870]$
counterfactual response probability for "happy", NY median poverty - 1 s.e.			
analog estimate	0.233	0.166	$[0.163, 0.612]$
corrected estimate	–	–	$[0.008, 0.838]$
95% interval	$[0.187, 0.279]$	$[0.101, 0.230]$	$[0.004, 0.853]$

Table 9: Triangular IV and Single Equation 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) at one standard deviation below the New York median, with neighborhood minority excluded. Columns TIV (1) and TIV (2) correspond to the triangular IV model results that exclude and include additional exogenous covariates, respectively.

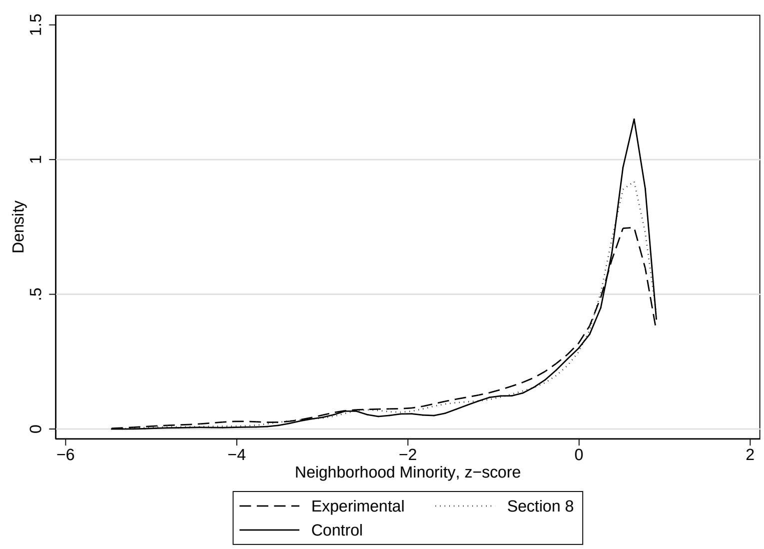
Figure 1: Distribution of neighborhood poverty by randomization group



Notes: Only observations with non-missing values for neighborhood poverty are used (neighborhood poverty is missing for 3/3,273 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.

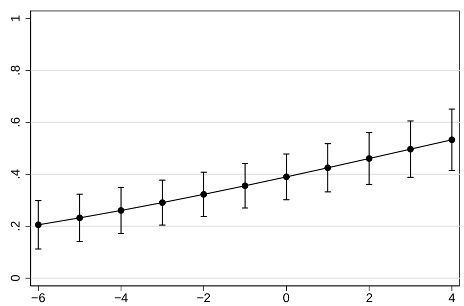
Figure 2: Distribution of neighborhood minority by randomization group



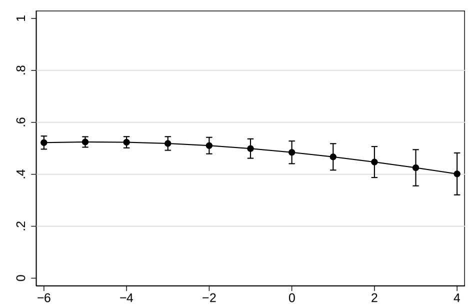
Notes: Only observations with non-missing values for neighborhood minority are used (neighborhood minority is missing for 3/3,273 adults). These include 1,453 adults in the Experimental group, 678 adults in the Section 8 group and 1,139 adults in the Control group.

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

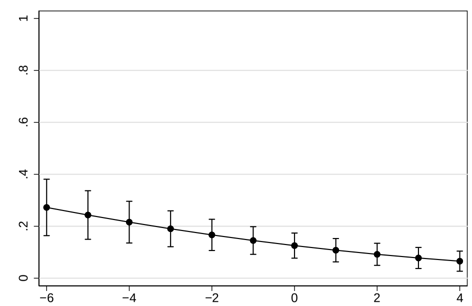
Figure 3: Ordered probit estimation of $p(y; x, w)$ across w for NY respondents



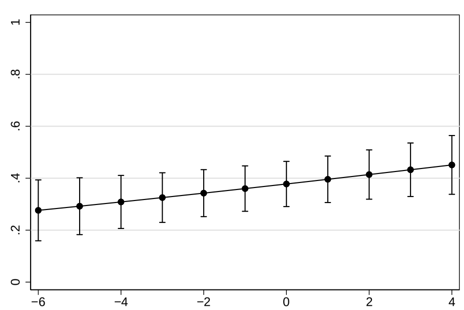
(a) $p(0; x, w)$ across neighborhood poverty



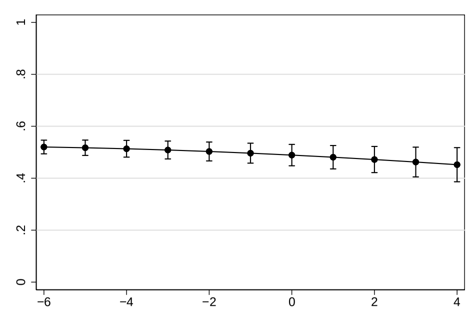
(b) $p(1; x, w)$ across neighborhood poverty



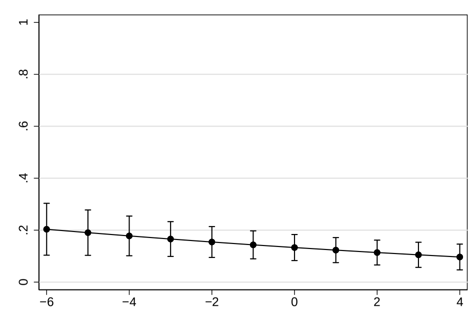
(c) $p(2; x, w)$ across neighborhood poverty



(d) $p(0; x, w)$ across neighborhood minority



(e) $p(1; x, w)$ across neighborhood minority

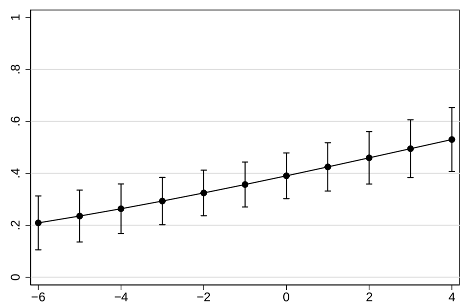


(f) $p(2; x, w)$ across neighborhood minority

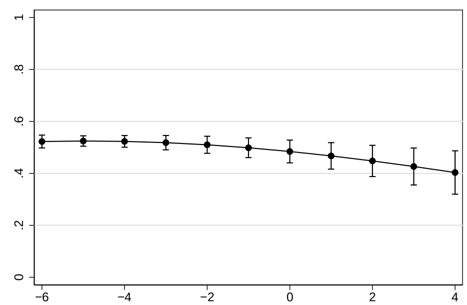
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using results reported in columns (4)-(5) of Table 4. Values of X variables are held fixed at the NY sample median, and randomization assignment is to the experimental voucher group. 95% confidence intervals are estimated using the delta method.

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

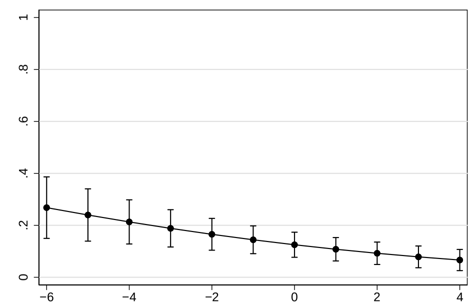
Figure 4: Ordered probit estimation of $p(y; x, w)$ across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



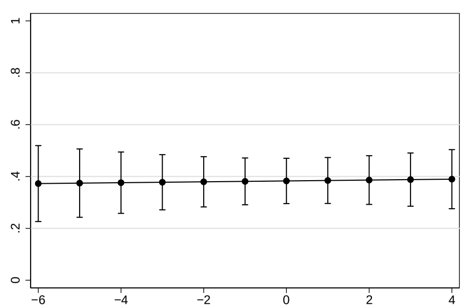
(a) $p(0; x, w)$ across neighborhood poverty



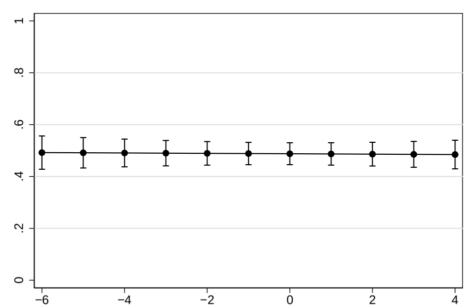
(b) $p(1; x, w)$ across neighborhood poverty



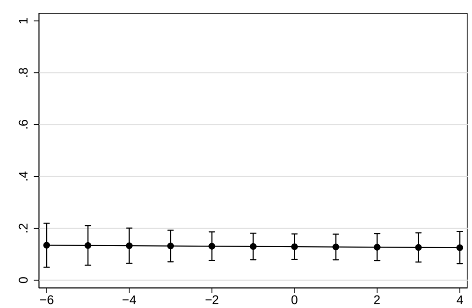
(c) $p(2; x, w)$ across neighborhood poverty



(d) $p(0; x, w)$ across neighborhood minority



(e) $p(1; x, w)$ across neighborhood minority

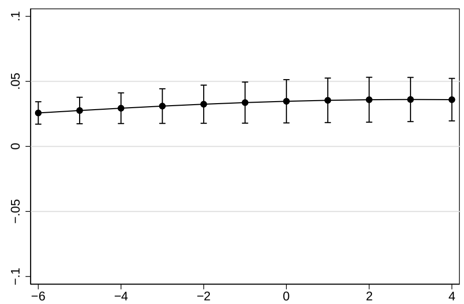


(f) $p(2; x, w)$ across neighborhood minority

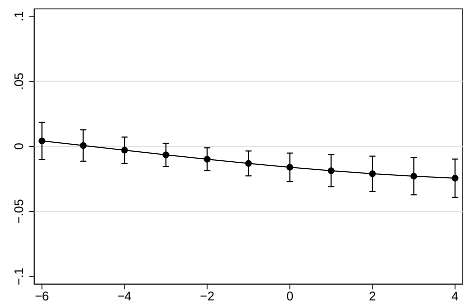
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using results reported in column (6) of Table 4. Values of X variables are held fixed at the NY sample median, and randomisation assignment is to the experimental voucher group. 95% confidence intervals are estimated using the delta method.

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

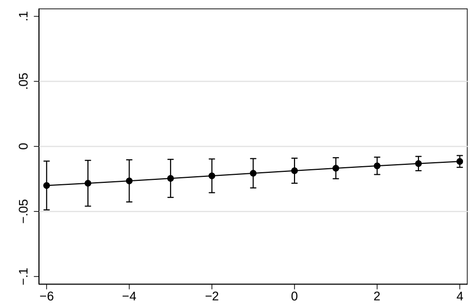
Figure 5: Ordered probit estimation of marginal effects across w for NY respondents



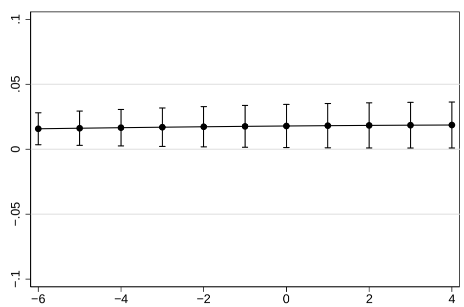
(a) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood poverty



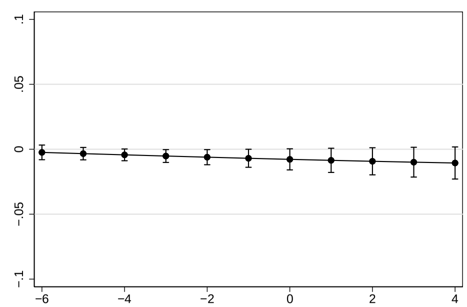
(b) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood poverty



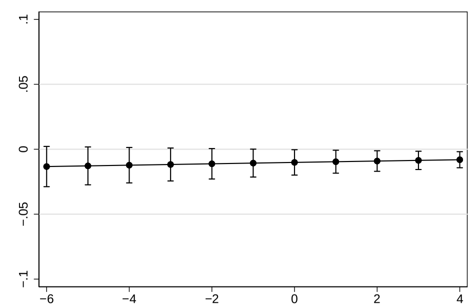
(c) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood poverty



(d) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood minority



(e) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood minority

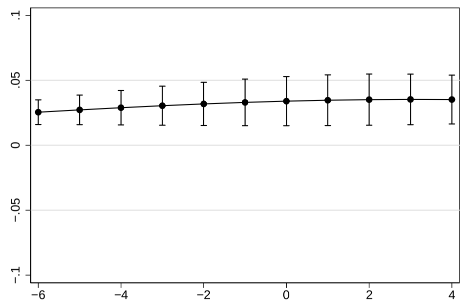


(f) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood minority

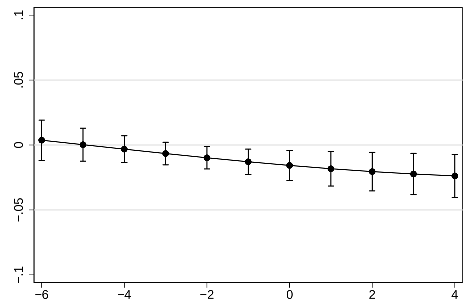
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using results reported in columns (4)-(5) of Table 4. Values of X variables are held fixed at the NY sample median, and randomisation assignment is to the experimental voucher group. 95% confidence intervals are estimated using the delta method.

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

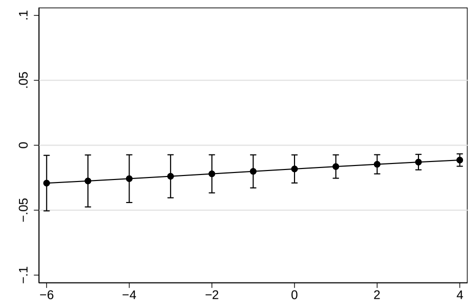
Figure 6: Ordered probit estimation of marginal effects across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



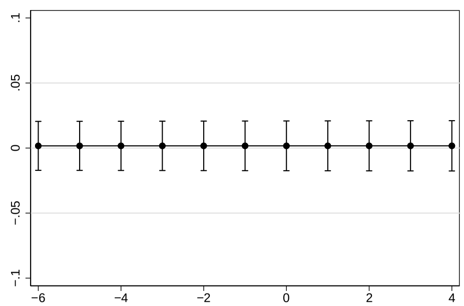
(a) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood poverty



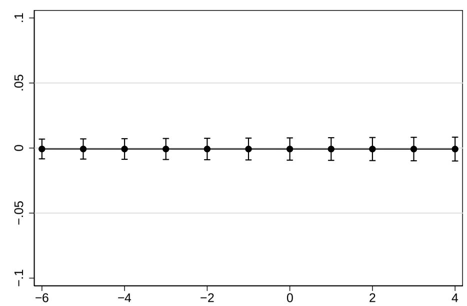
(b) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood poverty



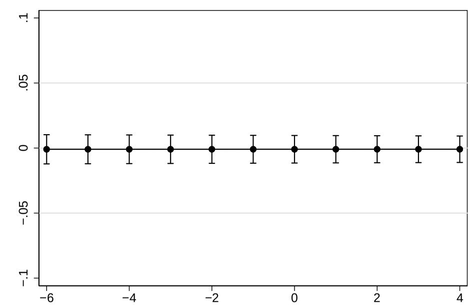
(c) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood poverty



(d) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood minority



(e) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood minority

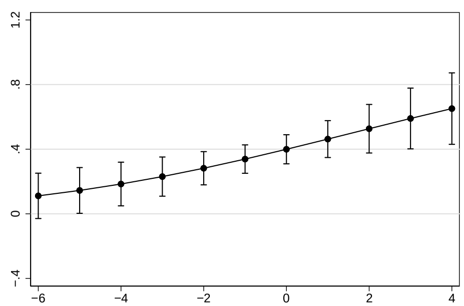


(f) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood minority

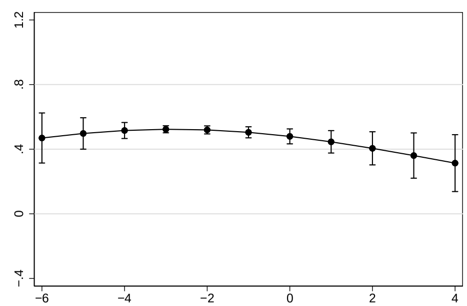
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using results reported in columns (6) of Table 4. Values of X variables are held fixed at the NY sample median, and randomisation assignment is to the experimental voucher group. 95% confidence intervals are estimated using the delta method.

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

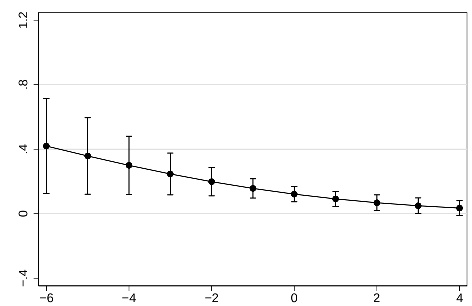
Figure 7: Triangular IV estimation of response probabilities $p(y; x, w)$ across w for NY respondents



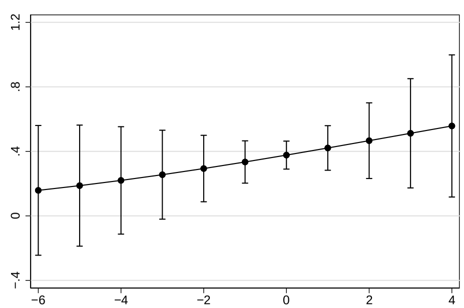
(a) $p(0; x, w)$ across neighborhood poverty



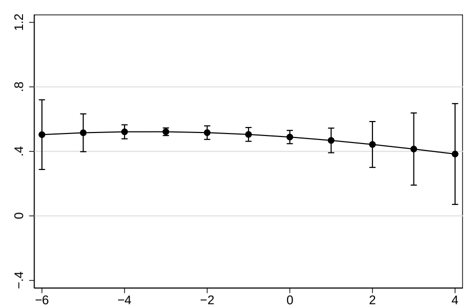
(b) $p(1; x, w)$ across neighborhood poverty



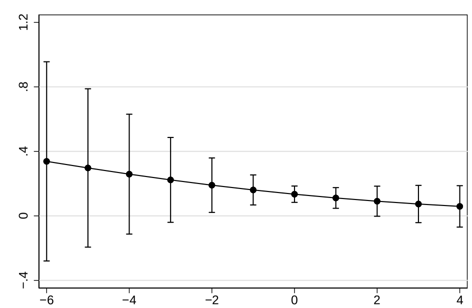
(c) $p(2; x, w)$ across neighborhood poverty



(d) $p(0; x, w)$ across neighborhood minority



(e) $p(1; x, w)$ across neighborhood minority

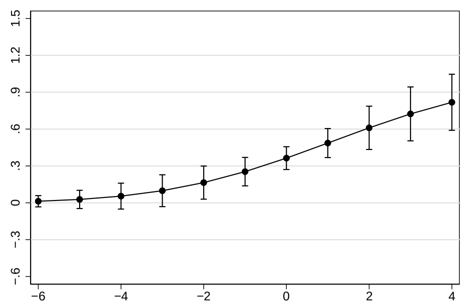


(f) $p(2; x, w)$ across neighborhood minority

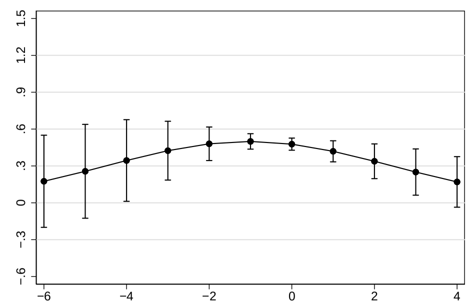
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using the results reported in columns (4)-(5) of Table 5. Values of all X variables are held fixed at the NY sample median. Standard errors are estimated using the `nlcom` command in STATA.

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

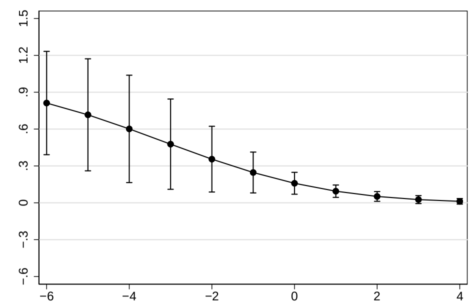
Figure 8: Triangular IV estimation of response probabilities $p(y; x, w)$ across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



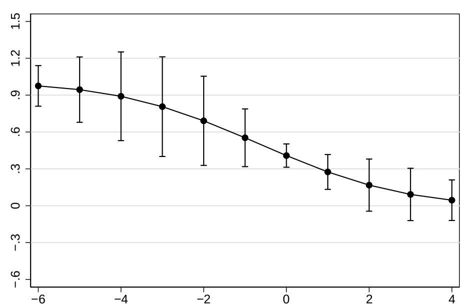
(a) $p(0; x, w)$ across neighborhood poverty



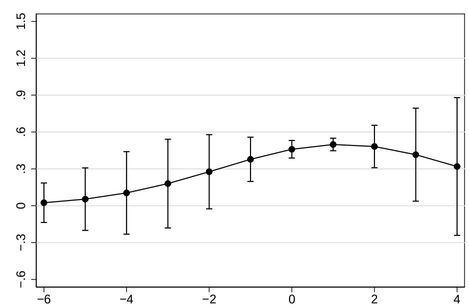
(b) $p(1; x, w)$ across neighborhood poverty



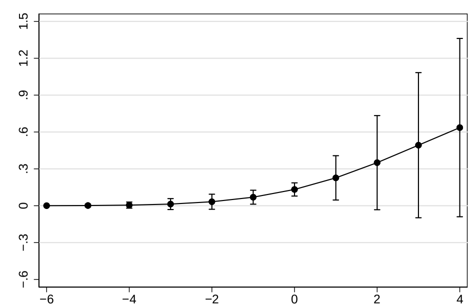
(c) $p(2; x, w)$ across neighborhood poverty



(d) $p(0; x, w)$ across neighborhood minority



(e) $p(1; x, w)$ across neighborhood minority

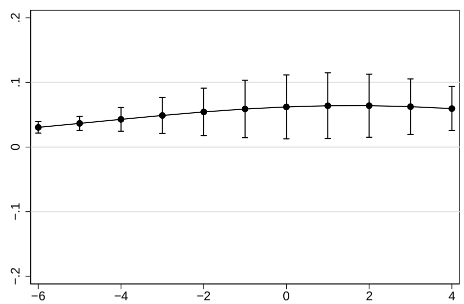


(f) $p(2; x, w)$ across neighborhood minority

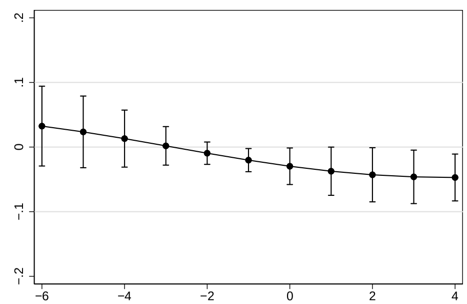
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using the results reported in column (6) of Table 5. Values of all X variables are held fixed at the NY sample median. Standard errors are estimated using the `nlcom` command in STATA.

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

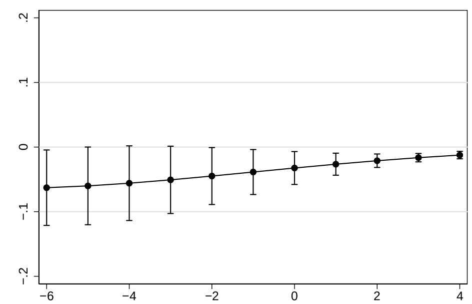
Figure 9: Triangular IV estimation of marginal effects across w for NY respondents



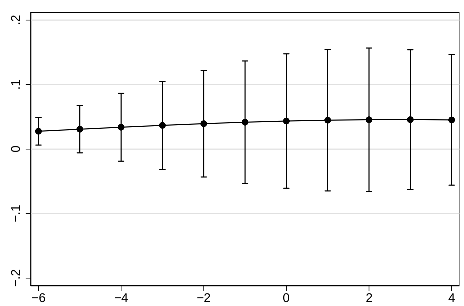
(a) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood poverty



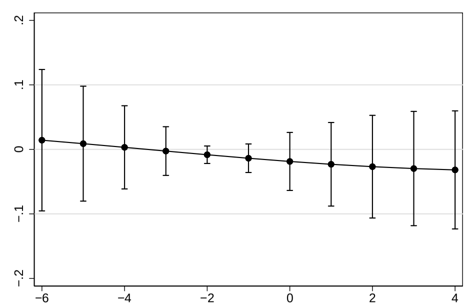
(b) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood poverty



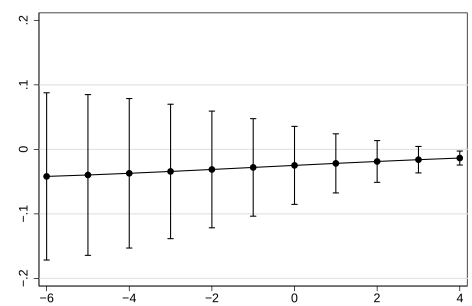
(c) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood poverty



(d) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood minority



(e) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood minority

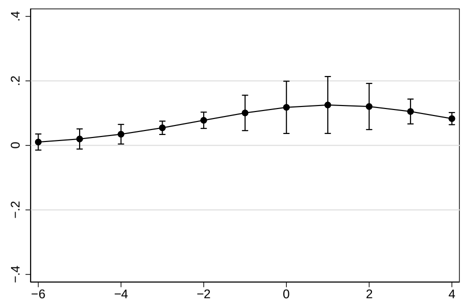


(f) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood minority

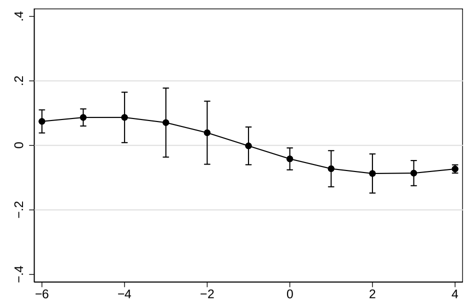
Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using the results reported in columns (4)-(5) of Table 5. Values of all X variables are held fixed at the NY sample median. Standard errors are estimated using the `nlcom` command in STATA.

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

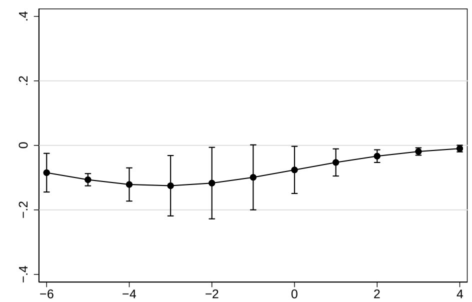
Figure 10: Triangular IV estimation of marginal effects across hood poverty (minority) while holding the value of hood minority (poverty) constant at it's median value for NY respondents



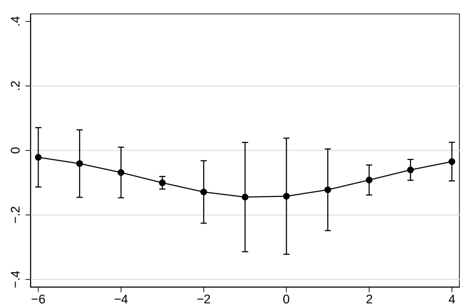
(a) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood poverty



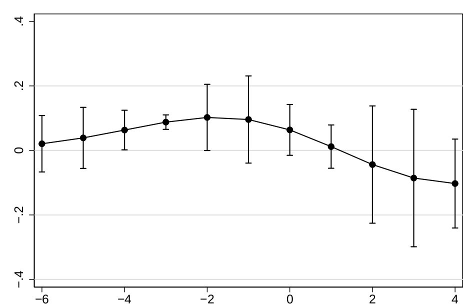
(b) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood poverty



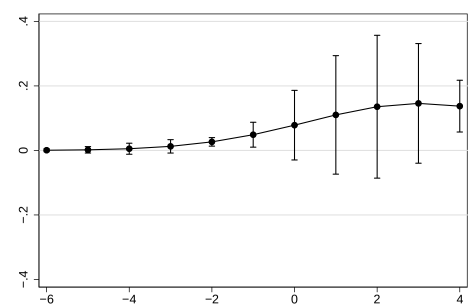
(c) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood poverty



(d) $\frac{\partial p(0;x,w)}{\partial w}$ across neighborhood minority



(e) $\frac{\partial p(1;x,w)}{\partial w}$ across neighborhood minority



(f) $\frac{\partial p(2;x,w)}{\partial w}$ across neighborhood minority

Notes: The dependent variable Y or Subjective Well Being (SWB) takes the value zero for not too happy, one for pretty happy and two for very happy. Conditional probabilities are estimated using the results reported in column (6) of Table 5. Values of all X variables are held fixed at the NY sample median. Standard errors are estimated using the `nlcom` command in STATA.

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

C Supplement: Triangular IV Model First Stage Estimates

First stage estimates for the triangular IV model estimated in Section 6.2.2 are presented here in Table 10.

Table 10: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\delta^{exp,Balt,1}$	-1.0912***	-0.8235***	-1.0910***	-1.1591***	-0.9015***	-1.1589***
	(0.1017)	(0.1143)	(0.1021)	(0.1038)	(0.1190)	(0.1042)
$\delta^{exp,Bos,1}$	-1.2798***	-1.7007***	-1.2819***	-1.2154***	-1.4961***	-1.2168***
	(0.0877)	(0.1264)	(0.0876)	(0.0896)	(0.1189)	(0.0898)
$\delta^{exp,Chi,1}$	-0.3068***	0.0993	-0.3055***	-0.3470***	0.0334	-0.3450***
	(0.0852)	(0.0736)	(0.0853)	(0.0915)	(0.0792)	(0.0917)
$\delta_{exp,LA,1}$	-0.8787***	-0.3421***	-0.8801***	-0.8137***	-0.3436***	-0.8142***
	(0.1007)	(0.0822)	(0.1003)	(0.1044)	(0.0916)	(0.1042)
$\delta_{exp,NY,1}$	-0.8052***	-0.1401	-0.8021***	-0.7993***	-0.1326	-0.7945***
	(0.0875)	(0.0854)	(0.0877)	(0.0891)	(0.0873)	(0.0895)
$\delta^{sec8,Balt,1}$	-1.0427***	-0.6651***	-1.0412***	-1.1065***	-0.7093***	-1.1032***
	(0.1184)	(0.1992)	(0.1194)	(0.1254)	(0.1986)	(0.1264)
$\delta^{sec8,Bos,1}$	-1.0880***	-1.2662***	-1.0838***	-1.0376***	-1.1362***	-1.0328***
	(0.1055)	(0.1428)	(0.1058)	(0.1130)	(0.1455)	(0.1134)
$\delta^{sec8,Chi,1}$	-0.1905*	0.2901***	-0.1863*	-0.2696**	0.1970**	-0.2643**
	(0.1092)	(0.0802)	(0.1092)	(0.1171)	(0.0894)	(0.1168)
$\delta^{sec8,LA,1}$	-0.8139***	0.0257	-0.8072***	-0.7508***	-0.0071	-0.7429***
	(0.0960)	(0.0988)	(0.0960)	(0.1041)	(0.1066)	(0.1038)
$\delta^{sec8,NY,1}$	-0.3742***	-0.0448	-0.3728***	-0.3945***	-0.0548	-0.3920***
	(0.0913)	(0.0838)	(0.0916)	(0.0923)	(0.0867)	(0.0926)
$\delta^{cont,Balt,1}$	-0.5220***	-0.3311***	-0.5184***	-0.5641***	-0.3941***	-0.5590***
	(0.0899)	(0.0931)	(0.0902)	(0.0949)	(0.0998)	(0.0951)
$\delta^{cont,Bos,1}$	-0.7145***	-1.1184***	-0.7106***	-0.6409***	-0.9028***	-0.6350***
	(0.0722)	(0.1018)	(0.0721)	(0.0823)	(0.1047)	(0.0825)
$\delta^{cont,Chi,1}$	0.2299**	0.2621***	0.2297**	0.1848*	0.2124***	0.1855*
	(0.0999)	(0.0718)	(0.0999)	(0.1078)	(0.0809)	(0.1077)
$\delta^{cont,LA,1}$	0.1584*	0.2110***	0.1597*	0.2360**	0.2192***	0.2381**

continued on next page

Table 10: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.0907)	(0.0664)	(0.0906)	(0.0959)	(0.0729)	(0.0959)
$\delta^{cont,NY,1}$	0.1371**	0.1735***	0.1354**	0.5305**	-0.2895	0.5258**
	(0.0547)	(0.0567)	(0.0548)	(0.2234)	(0.2631)	(0.2235)
$\delta^{exp,Balt,2}$			-0.8173***			-0.8886***
			(0.1118)			(0.1171)
$\delta^{exp,Bos,2}$			-1.7055***			-1.4949***
			(0.1226)			(0.1166)
$\delta^{exp,Chi,2}$			0.0880			0.0216
			(0.0733)			(0.0798)
$\delta_{exp,LA,2}$			-0.3677***			-0.3710***
			(0.0828)			(0.0929)
$\delta_{exp,NY,2}$			-0.1588*			-0.1447*
			(0.0846)			(0.0877)
$\delta^{sec8,Balt,2}$			-0.6895***			-0.7349***
			(0.1946)			(0.1968)
$\delta^{sec8,Bos,2}$			-1.2842***			-1.1477***
			(0.1401)			(0.1415)
$\delta^{sec8,Chi,2}$			0.2812***			0.1982**
			(0.0799)			(0.0877)
$\delta^{sec8,LA,2}$			0.0347			0.0240
			(0.0883)			(0.0941)
$\delta^{sec8,NY,2}$			-0.0617			-0.0641
			(0.0843)			(0.0863)
$\delta^{cont,Balt,2}$			-0.3553***			-0.4127***
			(0.0982)			(0.1033)
$\delta^{cont,Bos,2}$			-1.1349***			-0.9192***
			(0.1018)			(0.1077)
$\delta^{cont,Chi,2}$			0.2454***			0.1999**
			(0.0743)			(0.0824)
$\delta^{cont,LA,2}$			0.1980***			0.2055***
			(0.0667)			(0.0731)

continued on next page

Table 10: Triangular IV estimation of neighborhood effects on SWB, first stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\delta^{cont,NY,2}$			0.1858***			-0.2870
			(0.0565)			(0.2625)
N	3263	3263	3263	3175	3175	3175

Notes: Each column reports first stage estimates of a triangular model for specifications reported in corresponding columns of Table 5. 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 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.