



HHS Public Access

Author manuscript

J Econom Method. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

J Econom Method. 2020 January ; 9(1): . doi:10.1515/jem-2018-0030.

Regression-Based Causal Analysis from the Potential Outcomes Perspective

Joseph V. Terza¹

¹Department of Economics, Indiana University - Purdue University Indianapolis, Indianapolis, IN 46202, USA

Abstract

Most empirical economic research is conducted with the goal of providing scientific evidence that will be informative in assessing causal relationships of interest based on relevant counterfactuals. The implementation of regression methods in this context is ubiquitous. With this as motivation, we detail a comprehensive regression-based potential outcomes framework for causal modeling, estimation and inference. This framework facilitates rigorous specification of the effect parameter of interest and makes clear the sense in which it is causally interpretable, when appropriately defined in a potential outcomes setting. It also serves to crystallize the conditions under which the effect parameter and the underlying regression parameters are identified. The consistent sample analog estimator of the effect parameter is discussed. Juxtaposing this framework with a stylized version of a commonly implemented and routinely applied modeling and estimation protocol reveals how the latter is deficient in recognizing, and fully accounting for, conditions required for identification of the relevant effect parameter and the causal interpretability of estimation results. In the context of an example, we demonstrate the conceptual advantages of this general potential outcomes framework for regression modeling by showing how it resolves fundamental shortcomings in the conventional approach to characterizing and remedying omitted variable bias.

Keywords

causal effect parameter estimation; causal interpretability; conditional independence; conditional mean independence; mean independence

1 Introduction

The main motivation for nearly all empirical economic research is to provide scientific evidence that can be used to assess causal relationships of interest based on relevant counterfactuals. Such assessments usually focus on the rigorous specification and accurate estimation of parameters that characterize the relationship between a *presumed causal variable of interest* (\mathbf{X}), whose value is to be exogenously set and altered in the context of the relevant counterfactual, and a designated outcome of interest (\mathbf{Y}).¹ For example,

Joseph V. Terza is the corresponding author. jvterza@iupui.edu.

¹Henceforth, \mathbf{X} and \mathbf{Y} are to be taken as global replacements for the phrases “presumed causal variable of interest” and “outcome of interest,” respectively.

consider the analysis of the effect on infant birth weight (an example of a \mathbf{Y}) in the context of a counterfactual in which a fully effective smoking ban is imposed on pregnant women (an example of a \mathbf{Y}). Relationships of this type are typically characterized by an effect parameter (EP), and estimation of the EP is the usual objective of the empirical analysis. Rubin (1974, 1977) developed a framework for analyzing such counterfactual effects for contexts in which the \mathbf{X} is binary. In his framework, the key concept is the *potential outcome* (Y_j) – the random variable representing the \mathbf{Y} as it would have manifested if a specified value of the \mathbf{X} ($j = 0$ or 1) were counterfactually imposed on the relevant population. Here the EP is informative regarding the difference in the probability distributions of Y_0 and Y_1 (or specified features of their respective distributions) that can be exclusively attributed to the relevant counterfactually imposed change in the \mathbf{X} . The main difficulty to be confronted in the identification and estimation of the EP stems from the essential counterfactuality of the potential outcomes. Rubin (1974, 1977) and many other authors give conditions for identification of the EP and propose methods for its consistent estimation using observable data in this binary \mathbf{X} context.

The objective of the present paper is two-fold. First, we offer students of econometrics a comprehensive and coherent review of a general version of Rubin’s framework in which the \mathbf{X} can be any type of variable (e.g. continuous, count-valued, qualitative, semi-continuous, etc.) – not necessarily binary. Secondly, implicit in our discussion of this general potential outcomes framework (GPOF) is our advocacy for this approach to empirical modeling and identification which subtly but substantively differs from the more common paradigm in which modeling focuses on the relevant data generating process (DGP). As we hope the present discussion makes clear to students and practitioners alike, the GPOF more directly aligns with the causal inferential objectives of the typical empirical econometric analysis.

Although, the basic concepts discussed in this paper are, in one form or another, given coverage in some econometrics texts [e.g. Cameron and Trivedi (2005), Angrist and Pischke (2009), Wooldridge (2010), and Stock and Watson (2003, 2007, and 2015)], there appears to be no comprehensive and detailed textbook development of the GPOF cast in a conventional nonlinear parametric regression context.² The present paper is aimed at filling this void. Because parametric regression methods still constitute a substantial fraction of the coverage in these texts and in the typical econometrics course syllabus, and because parametric methods remain in wide use among practitioners, we confine the present discussion to cases in which the relevant potential outcome can be characterized using a conventional parametric model. In fact, we focus here on *minimally parametric potential outcomes models* that are based on a conditional mean regression assumption (possibly nonlinear); and *fully parametric potential outcomes models* based on a conditional probability mass/density function assumption. There are other less parametric contexts in which to cast the potential outcomes approach to causal modeling, estimation and inference; but we view the class of minimally parametric and fully parametric regression models as the logical starting point for introducing students and practitioners to the GPOF. In this context, we explicate conditions under which the relevant EP is identified and, as a by-product of this discussion, we show

²I direct the reader to Angrist and Pischke (2009) [in particular, pp. 52–59] for an overview of some of the concepts discussed in detail in the present paper.

how conventional estimates of the model parameters can be used to estimate the targeted EP. Inference based on such two-stage EP estimates is detailed elsewhere, so it is not included in the present discussion (see Terza 2016a, 2016b, 2016c, 2017, and 2018).

The remainder of the paper is organized as follows. In the next section, in order to frame the discussion, we detail the GPOF. To fix ideas, in Section 3 we then review Rubin's potential outcomes framework, in which the \mathbf{X} is binary, in the context of a specific example. In Rubin's framework, using the example, we also outline how regression models and methods can be implemented for identification and estimation of the targeted EP. In Section 4, we move on to a general discussion of the use of regression models and methods for specifying, identifying and estimating the relevant EP in the context of the GPOF. Therein we define the *conditional potential outcomes model (CPOM)* – a key concept in the implementation of regression modeling in the GPOF. We articulate conditions under which the CPOM is identified and discuss consistent estimation of its parameters and the relevant causally interpretable EP. In Section 5, we juxtapose the GPOF with a stylized version of a commonly applied protocol for empirical analysis that focuses on modeling the DGP. We illustrate all aforementioned issues and concepts with an example in Section 6. In the context of this example, we also demonstrate the conceptual advantages of the GPOF (CPOM) vs. the DGP-based approach, by showing how the former resolves fundamental conceptual shortcomings of the latter in characterizing and defining omitted variable bias. The final section summarizes the paper.

2 The General Potential Outcomes Framework

We now extend Rubin's potential outcomes framework to cases in which the \mathbf{X} can be any type of variable. We begin the discussion by drawing the distinction between two versions of the \mathbf{X} :

\mathbf{X}
 \equiv the random variable representing the observable (*factual*) version of the distribution of the \mathbf{X} (the sampled values of the \mathbf{X} are drawn from the distribution of \mathbf{X})

and

\mathbf{X}^* \equiv the random variable representing a hypothetical (*counterfactual*) exogenously mandated version of the \mathbf{X} .

Likewise, we distinguish two versions of the \mathbf{Y} :

\mathbf{Y}
 \equiv the random variable representing the factual version of the distribution of the \mathbf{Y} (the sampled values of the \mathbf{Y} are drawn from the distribution of \mathbf{Y})

and

$\mathbf{Y}_{\mathbf{X}^*}$
 \equiv the random variable representing the distribution of *potential outcomes*, defined as the distribution of values of the \mathbf{Y} that would have manifested for a particular \mathbf{X}^*

Remark 1

Although X^* is a random variable, it differs in character from X , Y and Y_{X^*} . X^* is a random variable in the sense that its value differs for each elementary unit (e.g. person) in the relevant population but unlike X , Y and Y_{X^*} , the value of X^* (per elementary unit) is determinate and knowable in the context of the relevant counterfactual. As such, X^* is not a component of the relevant DGP as are X and Y . Moreover, it is independent of all other variates germane to the present discussion.

Henceforth, we characterize the relevant counterfactual as an exogenously imposed change in the \mathbf{X} from X^{pre} to X^{post} - pre- and post- counterfactual versions of the \mathbf{X} , respectively. Without loss of generality we write $X^{post} = X^{pre} + \Delta$. Because X^{pre} and X^{post} are specific versions of X^* , and $\Delta = X^{post} - X^{pre}$, Remark 1 applies thereto. With this terminology and notation, we give the following definition.

Definition 1

A parameter characterizing a relationship between the \mathbf{X} and the \mathbf{Y} is *causally interpretable* if it can be written as a function of the moments of $Y_{X^{pre}}$ and $Y_{X^{pre} + \Delta}$.

The most commonly encountered causally interpretable EPs in the literature are³ *average treatment effect*

$$ATE = E[Y_1 - Y_0] \tag{1}$$

average treatment effect on the treated

$$ATET = E[Y_1 - Y_0 | X = 1] \tag{2}$$

average incremental effect

$$AIE(\Delta) = E[Y_{X^{pre} + \Delta} - Y_{X^{pre}}] \tag{3}$$

marginal AIE

$$MAIE = \lim_{\Delta \rightarrow 0} \frac{AIE(\Delta)}{\Delta} \tag{4}$$

In defining the MAIE, without loss of generality we assume that Δ is a constant.

The primary objectives in most empirical studies are consistent estimation of the relevant causally interpretable EP [e.g. (1) through (4)] and related causal inference to be drawn therefrom. There exists a substantial literature covering the latter and, therefore, it will not be explicitly discussed here. In the present paper we concentrate on the former and its antecedent – identification of the relevant EP.

³A parameter characterizing a relationship between the \mathbf{X} and the \mathbf{Y} is *first-order causally interpretable* if it is exclusively formulated in terms of $E[Y_{X^{pre}}]$ and $E[Y_{X^{pre} + \Delta}]$. For simplicity of exposition, without loss of generality, we henceforth consider first-order causally interpretable parameters only.

3 Identification and Estimation of the ATE in Rubin’s Framework

To get a handle on the above concepts and notation, let’s consider an example of Rubin’s potential outcomes framework in which the \mathbf{X} and the \mathbf{Y} are binary. Suppose the \mathbf{X} is whether or not an expectant mother smokes during pregnancy and the \mathbf{Y} is whether or not the infant she delivers is of low birth weight (LBW).⁴ Suppose Table 1 represents a population of 10 mothers. Individuals in cell A are those who, in fact, smoked during pregnancy.

Those in cell B, in fact, did not smoke during pregnancy. The bold value cells C and F contain the observable data on the \mathbf{Y} . The values of \mathbf{Y} in C correspond to birth weight status ($Y = 1$ if LBW, 0 if not) for mothers who smoked during pregnancy. The values of \mathbf{Y} in F correspond to birth weight status for the infants of the non-smokers. The bold value cells G and J contain the observable data on the \mathbf{X} . The values of \mathbf{X} in G correspond to mothers who smoked during pregnancy ($X = 1$). The values of \mathbf{X} in J correspond to those who did not smoke during pregnancy ($X = 0$). Italic value cells D and E contain the unobservable (counterfactual) data on the potential outcomes. The values of Y_1 in cell D are the LBW outcomes that would have manifested for non-smoking mothers had they instead been smokers. The values of Y_0 in cell E are the LBW outcomes that would have manifested for smokers had they been forced to quit during pregnancy. Cells H and I contain the counterfactually imposed smoking status values for the observed non-smokers and the observed smokers, respectively. The last column of the table (cells K and L) contains observable values of the indicator of whether or not the infant was the mother’s first-born ($FIRST = 1$ if so, 0 otherwise). In summary, the bold value cells contain factual (observable) data; the italic value cells contain counterfactual data.

When specifying the relevant EP in this context, the focus is on the difference between the distributions of the potential outcomes Y_0 and Y_1 that can be exclusively attributed to a counterfactual change in the \mathbf{X} from $X^* = 0$ to $X^* = 1$. In this example we take the relevant EP to be the ATE given in (1). The specific version of (1) that applies in the present example is

$$ATE = p_{Y_1}(1) - p_{Y_0}(1) \tag{5}$$

where $P_{Y_j}(y_j)$ denotes the probability mass function of Y_j (for $j = 0$ or 1) evaluated at y_j (for $y_j = 0$ or 1). Clearly (5) is first-order causally interpretable (by Definition 1 and footnote 3) because it is a function of the first-order moments of $Y_{X^{pre}}$ and $Y_{X^{pre} + \Delta}$ (in this case $X^{pre} = 0$ and $X^{pre} + \Delta = 1$). From the second and third columns of Table 1 we get

$$E[Y_1] = p_{Y_1}(1) = 0.7$$

and

⁴We assume all mothers have live births and thus abstract from the “bad control” issue arising from the fact that smoking may affect infant mortality [see Angrist and Pischke (2009) section 3.2.3 for a discussion of the bad control issue].

$$E[Y_0] = p_{Y_0}(1) = 0.4$$

so

$$ATE = 0.7 - 0.4 = 0.3. \quad (6)$$

In summary, (6) is the relevant causally interpretable EP in the present example. In other words, the ATE of smoking during pregnancy on the likelihood of a LBW infant is a 0.3 increase in the probability of that event.

We now turn to the identification and estimation of the EP in (6). The true value of the ATE is 0.3; measured as the difference in the relative frequencies of the value “1” in the 2nd and 3rd columns of Table 1. Note, however, that values from these columns are only partially observable. Data on the Y can only be sampled from the cells bold value in gray (C and F). In general (and in this example in particular), the partial observability (counterfactuality) of the ATE impedes identification and consistent estimation. To see this, note that we cannot hope to obtain an accurate estimate of the EP by simply sampling from the observable data in cells C and F and differencing their respective sample relative frequencies, because even if we could census the observable data in those cells such an estimator would yield

$$p_{(Y|X)}(1, 1) - p_{(Y|X)}(1, 0) \quad (7)$$

where $p_{(Y|X)}(y, x)$ denotes the conditional probability mass function of Y evaluated at y , given $X = x$. From the table it is easy to see that $p_{(Y|X)}(1, 1) = 0.8$ and $p_{(Y|X)}(1, 0) = 0.2$, so (7) is equal to 0.6. This makes consistent estimation of the relevant EP (the ATE) challenging. Consider the following intuitive simple difference of means estimator

$$\hat{p}_{(Y|X)}(1, 1) - \hat{p}_{(Y|X)}(1, 0) \quad (8)$$

where $\hat{p}_{(Y|X)}(y, x)$ represents the sample relative frequency of $Y = y$ among those for whom $X = x$. The estimator in (8) is consistent for (7), i.e.

$$p \lim [\hat{p}_{(Y|X)}(1, 1) - \hat{p}_{(Y|X)}(1, 0)] = p_{(Y|X)}(1, 1) - p_{(Y|X)}(1, 0) = 0.6 \quad (9)$$

but, by the same token, it is inconsistent for the EP in (6).

The problem is that an estimate of the ATE based on the differences in Y associated with variation in X , will not truly represent the targeted counterfactual EP. For example, such observable differences in the outcome may, in part, be attributable to the FIRST variable which is likely to be correlated with X . Can we somehow control for the influence of FIRST as a means of solving the identification and estimation problem? With this in mind we note that, using the law of iterated expectations (LIE) we can rewrite the ATE (5) as

$$\begin{aligned} ATE &= E[Y_1 - Y_0] \\ &= E[E[Y_1|FIRST] - E[Y_0|FIRST]]. \end{aligned} \quad (10)$$

In the present case, we note that

$$E[Y_{X^*}|FIRST] \equiv \text{probability of a LBW potential outcome } (Y_{X^*} = 1) \text{ for counterfactual smoking status } X^* \text{ conditional on FIRST} \\ = P(Y_{X^*}=1|FIRST, X^*) \tag{11}$$

where $P(Y_{X^*}|FIRST)(Y_{X^*}, FIRST, X^*)$ denotes the conditional probability mass function of Y_{X^*} given $FIRST$ evaluated at $Y_{X^*}=y_{X^*}$, X^* and $FIRST$. A reasonable assumption for the form of (11) might in the present context be

$$P(Y_{X^*}|FIRST)(1, FIRST, X^*) = \Phi(FIRST\pi_F + X^*\pi_X) \tag{12}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function and $\pi = (\pi_F, \pi_X)$ is the vector of parameters. Note that (12) has a sort of “hybrid” form in that it includes both counterfactual (X^* and Y_{X^*}) and observable ($FIRST$) components [in addition, of course, to the parametric component (π)]. Related to this is a kind of notational awkwardness that follows from the distinctive nature of X^* (see Remark 1 above). Although X^* is a random variable that appears as an argument of $P(Y_{X^*}|FIRST)(1, FIRST, X^*)$, it is not one of the variates upon which the probability of a LBW potential outcome is conditioned (indeed, it is independent of all other variates in the analysis – see Remark 1). Instead, it is a known (in the context of the relevant counterfactual) population unit-specific shifter of the probability of a LBW potential outcome. The expression in (12) is an example of an important general concept (the *conditional potential outcomes model*) which we define and discuss in detail in the next section of the paper. Note that (12) serves as a bridge between the inherently counterfactual targeted EP (the ATE) and the factual (observable) data at our disposal for estimating it. To wit, combining (12) and (10) yields

$$ATE = E[\Phi(FIRST\pi_F + \pi_X) - \Phi(FIRST\pi_F)] . \tag{13}$$

Assuming, for the moment, that we have an estimate of π in hand [say $\hat{\pi} = (\hat{\pi}_F, \hat{\pi}_X)$], we can consider the following sample analog estimator of (13)

$$\widehat{ATE} = \frac{1}{n} \sum_{i=1}^n \{ \Phi(FIRST_i\hat{\pi}_F + \hat{\pi}_X) - \Phi(FIRST_i\hat{\pi}_F) \} \tag{14}$$

where $FIRST_i$ is the value of $FIRST$ for the i th observation in a sample of size n ($i = 1, \dots, n$). In the next section of the paper, we also discuss the conditions on Y_{X^*} , Y , X and $FIRST$ (more generally, the vector of relevant controls) under which (12), π and (13) are identified and $\hat{\pi}$. and (14) are consistent.

4 EP Specification, Identification and Estimation in the GPOF

Here we return to the general case in which the \mathbf{X} , the \mathbf{Y} and the random variables X , Y , X^{pre} , Y_X^{pre} and $Y_{X^{pre} + \Delta}$ can be of any type. Moreover, in addition to (1) through (4), there are innumerable other possible relevant causally interpretable EPs. For ease of exposition and to fix ideas, without any substantive loss of generality, we focus the

discussion on (1) through (4). We seek to specify conditions under which the targeted EP is identified and can be consistently estimated. We saw in the example depicted in Table 1, identification and the consistency of the simple difference-of-means estimator for the ATE was thwarted by the presence of a variable that obfuscated the causal relationship between the \mathbf{X} and the \mathbf{Y} . The following definitions serve to formalize this discussion.

Definition 2

Let A , B and C be vector or scalar variates. B induces *conditional mean independence* between A and C if

$$E[A|B, C] = E[A|B]. \quad (15)$$

Definition 3

Suppose B induces conditional mean independence between A and C . We say that an element of B is a *confounder* for A and C if deleting it from B invalidates (15).

According to Definition 3, the variable `FIRST` in the example characterized by Table 1 is indeed a confounder for Y_{X^*} and X . To see this, note that for Table 1

$$E[Y_{X^*}|\text{FIRST}, X] = E[Y_{X^*}|\text{FIRST}]. \quad (16)$$

In the presence of uncontrolled confounders, there is typically insufficient structure to identify the targeted EP or devise a consistent estimator of it. Recall that this was true for the simple example depicted in Table 1 of the previous section. Without additional modeling structure, the presence of the confounder `FIRST` precluded identification and consistent estimation of the ATE. For this reason, we posited (12) and claimed that under certain conditions the ATE in (13) is identified and can be consistently estimated using its intuitive sample analog estimator (14). In the following, we prove this claim in the context of the GPOF. For the remainder of the discussion, we assume that \mathbf{V} is a vector of variates that subsumes all of the confounders for Y_{X^*} and X . Note that under this assumption \mathbf{V} induces conditional mean independence between Y_{X^*} and X . We begin with the following definition.

Definition 4

The *conditional potential outcomes model (CPOM)* specifies all moments of the distribution of $(Y_{X^*} | \mathbf{V})$ up to a given order.

The following defines the specific classes of regression models on which we focus the discussion.

Definition 5

The class of CPOMs that we call *minimally parametric (MP)* [fully parametric (FP)] comprises those for which it is assumed that

$$E[Y_{X^*}|\mathbf{V}] = m(\mathbf{V}, X^*; \boldsymbol{\pi}) \quad (17)$$

$$[\text{pmf/pdf}(Y_{X^*}|V) = f_{(Y_{X^*}|V)}(Y_{X^*}, V, X^*; \pi)] \quad (18)$$

where $m(\cdot) [f_{(Y_{X^*}|V)}(\cdot)]$ is a known function, π is the vector of parameters and $f_{(A|B)}(A, B; \xi)$ denotes the pmf/pdf of A conditional on B, written as a function of those variates and a parameter vector ξ . The FP CPOM in (18), of course implies a known form for the MP CPOM in (17). Clearly (12) for the example depicted in Table 1 of the previous section is a special case of (17) and (18). In this example, the LBW potential outcome (Y_{X^*}) is binary, as is the counterfactual version of the smoking variable (X^*) so the relevant version of (18) is

$$\begin{aligned} \text{pmf}(Y_{X^*}|V) &= f_{(Y_{X^*}|V)}(Y_{X^*}, V, X^*; \pi) \\ &= \Phi(\text{FIRST}\pi_F + X^*\pi_X)^{Y_{X^*}} [1 - \Phi(\text{FIRST}\pi_F + X^*\pi_X)]^{(1 - Y_{X^*})} \end{aligned} \quad (19)$$

which implies (12) as the relevant version of the CPOM in (17).

Returning now to the general model, in the MP case [or (17) as implied by (18) in the FP case], using the LIE, we can rewrite ATE, ATET, AIE and MAIE [(1) through (4), respectively] as

$$\begin{aligned} \text{ATE} &= E[E[Y_1|V] - E[Y_0|V]] \\ &= E[m(V, 1; \pi) - m(V, 0; \pi)] \end{aligned} \quad (20)$$

$$\begin{aligned} \text{ATET} &= E[E[Y_1|V, X = 1]] - E[E[Y_0|V, X = 1]] \\ &= E[m(V, 1; \pi)|X = 1] - E[m(V, 0; \pi)|X = 1] \end{aligned} \quad (21)$$

$$\begin{aligned} \text{AIE}(\Delta) &= E[E[Y_{X^{\text{pre}}} + \Delta|V] - E[Y_{X^{\text{pre}}}|V]] \\ &= E[m(V, X^{\text{pre}} + \Delta; \pi) - m(V, X^{\text{pre}}; \pi)] \end{aligned} \quad (22)$$

and

$$\begin{aligned} \text{MAIE} &= E \left[\left. \frac{\partial \mathcal{E}[Y_b|a]}{\partial b} \right|_{a=V, b=X^{\text{pre}}} \right] \\ &= E \left[\left. \frac{\partial m(a, b; \pi)}{\partial a} \right|_{a=V, b=X^{\text{pre}}} \right]. \end{aligned} \quad (23)$$

where $\mathcal{E}[Y_b|a]$ is the nonstochastic representation of $\mathcal{E}[Y_b|a]$ - a deterministic function whose arguments are a and b . Assuming that the relevant CPOM [e.g. (17)] is identified and that we have a consistent estimator of π (say $\hat{\pi}$), under very general conditions, we can consistently estimate (20) through (23) using the following intuitive sample analog estimators, respectively⁵

⁵The sample analog statistics in (20) through (23) can be cast as two-stage optimization estimators (see Terza 2016c). The general conditions under which such estimators are consistent are well established (see Newey and McFadden 1994; White 1994; or Woodridge 2010).

$$\widehat{ATE} = \frac{1}{n} \sum_{i=1}^n \{m(V_i, 1; \hat{\pi}) - m(V_i, 0; \hat{\pi})\} \quad (24)$$

$$\widehat{ATET} = \frac{1}{n_1} \sum_{i_1=1}^{n_1} \{m(V_{i_1}, 1; \hat{\pi}) - m(V_{i_1}, 0; \hat{\pi})\} \quad (25)$$

$$\widehat{AIE(\Delta)} = \frac{1}{n} \sum_{i=1}^n \{m(V_i, X_i^{\text{pre}} + \Delta_i; \hat{\pi}) - m(V_i, X_i^{\text{pre}}; \hat{\pi})\} \quad (26)$$

$$\widehat{MAIE} = \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial m(a, b; \pi)}{\partial a} \right|_{a=V_i, b=X_i^{\text{pre}}; \pi=\hat{\pi}} \quad (27)$$

where X_i^{pre} and i_1 are the exogenously determined counterfactual values of X^{pre} and i_1 for the i th observation; V_i is the sampled value of V ; and i_1 denotes the i th observation in the treated subsample of size n_1 . The two-stage sample analog statistics in (24) through (27) for the MP and FP cases considered here, can typically be implemented in Stata® using the relevant packaged command followed by an appropriately specified version of the “margins” command. This is described in detail in Terza (2017) along with explicit instructions on how to code such statistics in Mata® (the matrix language in Stata®) in cases for which the “margins” command is unavailable.^{6,7} Numerous real data examples are given in Terza (2017).

As we have seen, the identification of the relevant EP [e.g. (20) through (23)] is tantamount to the identification of the relevant version of the CPOM. In the parametric context this identification argument has two levels. First, we must establish conditions under which the following holds

$$E[Y|V, X] = m(V, X; \pi) \quad (28)$$

where the $m(\cdot)$ function is defined as in (17). Next, full identification of the CPOM is established by maintaining (28) and validating conditions under which the parameter vector π is identified. This second (parametric) level of the identification argument is covered elsewhere in a number of alternative sources and, as such, need not be repeated here.⁸ Instead, we focus here on reviewing sufficient conditions under which (28) is valid. For simplicity of exposition, we consider the case in which the CPOM is MP as specified in (17).

⁶Formulation and calculation of the correct asymptotic standard errors for two-stage estimators like (24) through (27) are discussed in Terza (2016a, 2016b, 2016c, 2017, and 2018).

⁷Note that (14) is a special case of (24).

⁸See, for example Woodridge (2010).

Before we formally establish conditions under which (28) holds, we need the following definitions.

Definition 6

Conditional outcome mean invariance holds if

$$E[Y_a | X = a, V] = E[Y | X = a, V] \quad (29)$$

Where “a” is a particular value in the support of X.

We know that because the vector V subsumes all confounders for Y_{X^*} and X, it induces conditional mean independence between them, i.e.

$$E[Y_{X^*} | V, X] = E[Y_{X^*} | V] \quad (30)$$

from which, by Definition 6, it follows that

$$E[Y_a | V] = E[Y | V, X = a]. \quad (31)$$

This implies that, regardless of whether the value of the \mathbf{X} (“a”) is counterfactually mandated or is the product of the DGP, the random variable representing the potential outcomes version of the \mathbf{Y} defined at “a” and conditional on V [$(Y_a|V)$] and the random variable representing the observable version of the \mathbf{Y} conditional on V and X (at a) [$(Y|V, X = a)$] will have the same mean.⁹ We need an additional condition for establishing (28).

Definition 7

Overlap holds if¹⁰

$$0 < p_{(X|V)}(x, v) < 1 \quad (32)$$

where $p_{(X|V)}(x, v)$ denotes the conditional probability mass/density function of X given $V = v$ evaluated at $X = x$.

Overlap ensures that each unit in the relevant population for which the observable value of the vector of controls is v, has some nontrivial but uncertain chance of having an observable value of X equal to x.¹¹ The following theorem establishes conditions under which (28) holds.

Theorem 1

In the context of the MP CPOM in (17) (Definition 5), in which it is implicit that

⁹The *stable unit treatment value assumption* (SUTVA) is usually imposed when the \mathbf{X} is binary (Imbens and Wooldridge 2009, p. 13). The SUTVA posits that only the level of treatment for a specific individual affects the outcomes for that individual. In the GPOF, conditional outcome mean invariance can be viewed as a weaker version of the SUTVA. Therefore, strictly speaking, the SUTVA is not required to establish (28). It should be noted however, that in the context of estimating the deep parameter vector π and the relevant EP, random sampling is typically assumed. As Woodridge (2010) notes, random sampling implies the SUTVA.

¹⁰Imbens and Wooldridge (2009) discuss this condition for the case in which the \mathbf{X} is binary.

¹¹Note that the population represented by Table 1 satisfies this condition.

- a. V induces conditional mean independence between Y_{X^*} and X ,
if
- b. conditional outcome mean invariance holds (Definition 6)
and
- c. overlap holds (Definition 7)
- then

$$E[Y|V, X] = m(V, X; \pi).$$

Proof

See Appendix A

Thus, the first level of identification is established if condition (a) and premises (b) and (c) of Theorem 1 hold. The identification argument is completed by maintaining (28) and invoking the conditions for establishing the parametric identification of π therefrom. If the CPOM (17) is identified and the parametric identification of π is established, then under general conditions, π can be consistently estimated by applying a conventional regression method [e.g. the nonlinear least squares (NLS) method] to (28) using a sample of data on Y , V and X . Moreover, as we have already discussed, this consistent estimator of π facilitates consistent estimation of the relevant EP [e.g. as in (24) through (27)].

Although the targeted EP may be first-order, the underlying CPOM may involve higher order conditional moments of the potential outcome. In the extreme, the CPOM may be FP as defined in (18). Recall that knowledge of (18) implies knowledge of (17) so that all of the discussion surrounding equations (20) through (27) also holds if the assumed CPOM is FP. Here the two levels of the identification argument comprise: 1) validation of conditions under which the FP CPOM in (18) implies

$$\text{pmf/pdf}(Y|V, X) = f_{(Y_{X^*}|V)}(Y, V, X; \pi) \quad (33)$$

and 2) proof of the identification of π with (33) maintained. The conditions for establishing the validity of (33) are given as the premises of Theorem 1' in Appendix B. Second level identification of π in the context of FP DGP specifications like (33) has been given extensive textbook treatment and will not be discussed here.¹² Given that the CPOM is identified, π can be consistently estimated by applying the relevant maximum likelihood estimator (MLE) based on (33) to a sample of data on Y , V and X .

¹²See Woodridge (2010).

5 Juxtaposing the CPOM and a Stylized Version of the DGP-Based Regression Protocol

It is common in the empirical literature for modeling to commence with direct specification of the DGP without mentioning the relevant potential outcomes framework and, in particular, without specifying the relevant CPOM. Typically, such a DGP specification is based on either the conditional mean or conditional pmf/pdf of Y given X and a vector of observable controls (say C). Formalizing such minimally and fully parametric DGP specifications we write

$$\text{MP: } E[Y|C, X] = \ell(C, X; \gamma) \quad (34)$$

and

$$\text{FP: } \text{pmf/pdf}(Y|C, X) = g_{(Y|C, X)}(Y, C, X; \gamma) \quad (35)$$

where ℓ and g are known functions and γ is a vector of unknown parameters. The empirical analysis is then conducted according to the following stylized protocol:

Step 1 Estimate γ using the appropriate M-estimator (e.g. NLS in the MP model, or MLE in the FP model).

Step 2 Estimate a “marginal effect” based on the first-step estimate (say $\hat{\gamma}$). For example,

$$\frac{1}{n} \sum_{i=1}^n \{ \ell(C_i, 1; \hat{\gamma}) - \ell(C_i, 0; \hat{\gamma}) \} \quad (36)$$

$$\frac{1}{n_1} \sum_{i_1=1}^{n_1} \{ \ell(C_{i_1}, 1; \hat{\gamma}) - \ell(C_{i_1}, 0; \hat{\gamma}) \} \quad (37)$$

$$\frac{1}{n} \sum_{i=1}^n \{ \ell(C_i, X_i^{\text{pre}} + \Delta_i; \hat{\gamma}) - \ell(C_i, X_i^{\text{pre}}; \hat{\gamma}) \} \quad (38)$$

$$\frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \ell(a, b; \gamma)}{\partial a} \right|_{a=C_i, b=X_i^{\text{pre}}; \gamma=\hat{\gamma}} \quad (39)$$

This step can be, and often is, automated [e.g. via the Stata[®] “margins” command (StataCorp 2017)].

Step 3 Report the calculated value of the relevant “marginal effect,” and explicitly or implicitly use it to conduct causal inference.

This protocol is clearly incomplete and invalid from the perspective of causal inference. Even if (17) holds and the premises of Theorem 1 are satisfied for Y_{X^*} , Y , X^* , X and V (not

C); there is no reason to believe that the relevant EP [e.g. (1) through (4)] will be identified. Likewise, there is no basis for interpretation of results obtained from (36) through (39) as causal. In short, the minimally or fully parametric DGP specifications in (34) and (35), respectively, have no necessary connection whatsoever with the true CPOM specifications in (17) and (18), respectively, and are therefore devoid of any causally interpretable content. Moreover, the fact that (36) through (39) are automated (e.g. in Stata[®]) is often taken as license by the researcher to forego explanation of the specifics of the “marginal effect” that is being calculated and reported. This lack of specificity facilitates the use of veiled and obfuscatory language in describing the empirical results. Phrases like “the association effect of X on Y” or “the predictive margin of Y with respect to X” are often used to connote causality in describing results from statistics such as (36) through (39). Such concepts are, however, entirely uninformative from the perspective of causal inference and should be avoided in that context.

6 Smoking and Birth Weight: Making Sense of Omitted Variables Bias Using the CPOM

To illustrate the concepts that we have developed and discussed to this point, we re-visit the DGP-based model of smoking during pregnancy and birth weight considered by Mullahy (1997). He considers an exponential conditional mean regression version of (34) that has the following essential form

$$E[Y|C, X] = \exp(C\kappa_c + X\kappa_x) \quad (40)$$

where

$Y \equiv$ observed infant birth weight in lbs. (now a continuous variable)

$X \equiv$ observed number of cigarettes smoked per day during pregnancy (now a count variable)

$C \equiv$ [PARITY WHITE MALE 1]

PARITY \equiv birth order

WHITE \equiv 1 if white, 0 otherwise

MALE \equiv 1 if male, 0 otherwise

and $\kappa' = (\kappa'_c, \kappa'_x)$ is the vector of parameters. Mullahy posits an alternative version of the model of the form

$$E[Y|C, U, X] = \exp(C\beta_c + U + X\beta_x) \quad (41)$$

where $\beta = (\beta'_c, \beta'_x)$ is the vector of parameters and U is an arbitrary unobserved control variable that is possibly correlated with X . It is clear that in the DGP-based specifications in (40) and (41) the values of κ and β are likely to differ. For example, suppose the correlation between U and X can be formalized as

$$E[\exp(U)|C, X] = \exp(\alpha_0 + X\alpha_x) \tag{42}$$

where α_0 and α_x are parameters.¹³ Combining (41) and (42) we can rewrite the parameters of (40) as

$$\kappa'_C = \beta'_C + (0, \alpha_0)$$

$$\kappa_x = \beta_x + \alpha_x$$

where 0 denotes a row vector of zeros of dimension one less than that of β_C . In Mullahy’s DGP-based framework, such parameter differences would be characterized as “omitted variables” bias. To that end, he devises a very clever consistent generalized method of moments (GMM) estimator for estimation of β that does not require explicit specification of the mean of $(\exp(U)|C, X)$ as in (42).¹⁴ He does not discuss EP estimation or causal inference. To empirically validate the difference between (40) and (41), we considered the corresponding versions of (38) relevant to a conventional DGP-based analysis of the “marginal effect” of a counterfactual in which all nonsmokers are absolutely prevented from smoking during pregnancy and all smokers are forced to quit.¹⁵ The specific versions of (38) that we calculated corresponded to

$$\begin{aligned} & \exp(C_i \hat{\kappa}_C + (X_i + \Delta_i) \hat{\kappa}_x) \text{ for (40)} \\ & \ell(C_i, X_i^{\text{pre}} + \Delta_i; \hat{\gamma}) \\ & = \exp(C_i \hat{\beta}_C + (X_i + \Delta_i) \hat{\beta}_x) \text{ for (41)} \end{aligned} \tag{43}$$

where, in calculating (38), Δ_i is set equal to either $-X_i$ or 0 and $\hat{\kappa}' = (\hat{\kappa}'_C, \hat{\kappa}'_x)$ ($\hat{\beta}' = (\hat{\beta}'_C, \hat{\beta}'_x)$) denotes the vector of Poisson quasi-maximum likelihood (GMM) estimates given in the second (fourth) column of Mullahy’s Table 2.¹⁶ The “marginal effect” estimates are displayed in Table 2 below. As can be seen therein, the Poisson-based version of (38) is less than half the size of the version obtained using the GMM estimates. We also conducted a Hausman-Wu test (Wu 1973; Hausman 1978) of the difference between the Poisson and GMM versions of (38) and found it to be statistically significant at nearly a 5% level.¹⁷

From the conventional DGP-based modeling perspective, one might argue that this exercise reveals the omitted variable bias inherent in the conditional mean regression specification in (40), and that (41) is the correct (or more correct) model for causal estimation and inference.

¹³This would be the case, for instance, if $(U|X)$ were normally distributed with mean $X\alpha_x$ and variance σ^2 . In this case we would get $\alpha_0 = \sigma^2/2$.

¹⁴His estimator does, however, require that $E[\exp(U)|W] = 1$, where W is a vector of instrumental variables.

¹⁵Mullahy did not calculate “marginal effects” or discuss causal interpretability/inference.

¹⁶It is interesting that the results in the second and fourth columns of Mullahy’s Table 2 align with an underlying model in which the correlation between U and X is formalized as in (42). In comparing the parameter estimates for the two models [(40) in the second column vs. (41) in the fourth column] we see that there is very little difference in the estimated coefficients of *PARITY*, *WHITE*, *MALE* and the constant. Whereas, the GMM estimate of the coefficient of X is more than double the size of the Poisson estimate.

¹⁷The method detailed in Terza (2016a, 2016b, and 2016c) was implemented for calculating the correct asymptotic standard errors of the Poisson and GMM versions of (38). Corresponding Stata/Mata[®] code will be supplied upon request.

However, without additional modeling structure (presumably in the relevant GPOF), neither (40) nor (41) serves to identify a causally interpretable EP. As we saw in the previous section, this is because they are cast entirely in the DGP context which provides no foundation upon which to base causal interpretation and inference. By the same token, this DGP-based modeling framework provides no foundation for distinguishing (40) from (41) with regard to causal correctness or causal interpretability. The same can be said of the probability limits of the corresponding versions of (38) detailed in (43).

If, however, modeling were cast in the GPOF such ambiguity is resolved. In the GPOF one could, for instance, specify the relevant MP CPOM as

$$E[Y_{X^*} | C, U] = \exp(C\beta_c + U + X^*\beta_x) \tag{44}$$

where

1. $Y_{X^*} \equiv$ infant birth weight potential outcome
2. and
3. $X^* \equiv$ counterfactually imposed number of cigarettes smoked per day during pregnancy.

The CPOM in (44) resolves the ambiguity with respect to what one is seeking to identify. Moreover, if for $V = (C, U)$ and the CPOM in (44), premises (a), (b) and (c) of Theorem 1 hold, then the latter is identified and causally interpretable so that (41) is a legitimate DGP representation.¹⁸ Under these conditions, the relevant causally interpretable EP can be derived as¹⁹

$$AIE(\Delta) = E[\exp(C\beta_c + [X^{pre} + \Delta]\beta_x) - \exp(C\beta_c + X^{pre}\beta_x)] \tag{45}$$

where $\Delta = -X^{pre}$. Because (44) is identified so is (45). Therefore, if premises (a), (b) and (c) of Theorem 1 hold for V and (44) then under general conditions

$$\widehat{AIE(\Delta)} = \frac{1}{n} \sum_{i=1}^n \left\{ \exp(C_i \hat{\beta}_c + [X_i + \Delta_i] \hat{\beta}_x) - \exp(C_i \hat{\beta}_c + X_i \hat{\beta}_x) \right\} \tag{46}$$

is consistent for (45) where $\hat{\beta}_c$ and $\hat{\beta}_x$ are Mullahy's GMM estimates. Because (46) is consistent, the EP estimate it produces is causally interpretable.

This example demonstrates how the ambiguity with regard to identification and causal interpretability inherent in conventional DGP-based regression modeling can be resolved by casting model development in the GPOF. The foregoing discussion reveals that the two key steps in GPOF-based causal regression modeling and inference are: A) specifying the relevant CPOM; and B) arguing that the premises of Theorem 1 hold. The latter establishes the identification of the CPOM and the relevant EP. It also ensures the consistency of both

¹⁸It seems reasonable to assume that U comprises all unobserved confounders for X and Y_{X^*} (see Definition 3) so that V satisfies condition (a) in the statement of Theorem 1.

¹⁹The derivation of (45) makes use of Mullahy's assumption that $E[\exp(U)|W] = 1$ (see Terza 2006).

the regression-based estimates of the parameters of the CPOM and the corresponding estimate of the EP. Finally, it guarantees the causal interpretability of the CPOM parameters, the EP and the results produced by their consistent estimators.

7 Discussion

We have here highlighted the importance of the potential outcome concept for characterizing causally interpretable effect parameters and their estimators in a general nonlinear parametric regression framework in which the \mathbf{X} and the \mathbf{Y} can be of any type. The foregoing discussion also makes clear the need for the GPOF (and in particular the CPOM) in establishing the requisite conditions for effect specification, identification and consistent estimation. The commonly implemented DGP-based protocol makes no mention of potential outcomes and, therefore, does not afford a clear and rigorous definition of the causal parameter of interest [e.g. the parameters presumably targeted by estimators like (36) through (39)]. This lack of specificity regarding estimation objectives renders discussions of truly causally interpretable effect parameters, their identification and consistent estimation, and related asymptotic causal inference, meaningless in the DGP protocol context. Because the GPOF involves modeling at a more primitive level via the CPOM, it: 1) allows rigorous formulation of the causally interpretable estimation objective (the EP) and its intuitive sample analog estimator; 2) facilitates articulation of the conditions under which the true causally interpretable version of the DGP is implied by the fundamental potential outcome specification (the CPOM); and 3) makes clear how knowledge of the true DGP can be used to establish identification and consistent sample analog estimation of the EP. Using an empirical example involving an omitted variable, we demonstrate the conceptual advantages held by the GPOF/CPOM approach over the DGP-based protocol.

Acknowledgement

This work was supported by the Agency for Healthcare Research and Quality (R01 HS17434) and by the Alcohol Research Group through a center grant from the National Institute of Alcohol Abuse and Alcoholism (P50 AA05595-36). The author would like to thank an anonymous reviewer, the anonymous associate editor, and the co-editor, Tong Li, for comments and suggestions that served to greatly improve the substance and presentation of the paper.

Appendix A

Proof of Theorem 1

Proof: We first note that premise (c) is necessary for the existence of the left-hand side of the theorem's conclusion (viz. $E[Y|V, X]$). To show that $E[Y|V, X] = m(V, x; \pi)$ we can equivalently establish that for any value in the support of X (say “ x ”) $E[Y|V, X = x] = m(V, x; \pi)$. The MP CPOM in (17) implies that

$$E[Y_x|V] = m(V, x; \pi) \tag{47}$$

and, as we have seen, condition (a) implies (31). Now let $X^* = X^{\text{exog}}$ where X^{exog} denotes the exogenously imposed version of the \mathbf{X} whose distribution is the same as the marginal distribution of X . For any x in the support of X (i.e. the support of X^{exog}), from (31) we get

$$E[Y_x|V] = E[Y_x|V, X = x] \quad (48)$$

[(31) holds for *any* X^* ; therefore it must hold for $X^* = x$]. Moreover, by premise (b)

$$E[Y_x|V, X = x] = E[Y|V, X = x]. \quad (49)$$

Combining (47), (48) and (49) yields

$$E[Y|V, X = x] = m(V, x; \pi).$$

This completes the proof. ■

Appendix B

Identification of Fully Parametric CPOM Specifications

Here, in the FP case in which the CPOM is specified as in (18), we establish conditions under which (33) is legitimate. We begin with the following FP (stronger) versions of Definition 2, Definition 3 and Definition 6.

Definition 2'

Let A, B and C be vector or scalar variates. B induces *conditional independence* between A and C if

$$\text{pmf/pdf}(A|B, C) = \text{pmf/pdf}(A|B). \quad (50)$$

Definition 3'

Suppose B induces conditional independence between A and C. We say that an element of B is a *full confounder* for A and C if deleting it from B invalidates (50).

Definition 6'

Conditional outcome invariance holds if

$$(Y|V, X = a) \text{ and } (Y|V, X = a) \text{ are identical} \quad (51)$$

Where “a” is a particular value in the support of X.

For the remainder of the discussion in this appendix, we assume that V is a vector of variates that subsumes all of the full confounders for Y_{X^*} and X. Note that under this assumption V induces conditional independence between Y_{X^*} and X, i.e.

$$\text{pmf/pdf}(Y_{X^*}|V, X) = \text{pmf/pdf}(Y_{X^*}|V). \quad (52)$$

If, in addition, conditional outcome invariance holds, then

$$\text{pmf/pdf}(Y_a|V) = \text{pmf/pdf}(Y|V, X = a) \quad (53)$$

which says that, regardless of whether the value of the \mathbf{X} (“a”) is counterfactually mandated or is the product of the data generating process (i.e. randomly sampled and then treated as a conditioning variate) the random variable representing the potential outcomes version of the \mathbf{Y} defined at “a” and conditioned on V [$(Y_a | V)$] and the random variable representing the observable version of the \mathbf{Y} conditioned on V [$(Y | V)$] will have the same pmf/pdf value. The following theorem establishes conditions under which (33) holds.

Theorem 1'

In the context of the FP CPOM in (18) (Definition 5), in which it is implicit that

a. V induces conditional independence between Y_{X^*} and X ,

if

b. conditional outcome invariance holds (Definition 2')

and

c. overlap holds (Definition 7)

then

$$\text{pmf/pdf}(Y|V, X) = f_{(Y_{X^*}|V)}(Y, V, X; \pi).$$

Proof: Analogous to the proof of Theorem 1.

References

- Angrist JD, and Pischke JS. 2009 *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.
- Cameron AC, and Trivedi PK. 2005 *Microeconometrics: Methods and Applications*. New York, NY: Cambridge University Press.
- Hausman J 1978 “Specification Tests in Econometrics.” *Econometrica* 46: 1251–1271.
- Imbens GW, and Wooldridge JM. 2009 “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47: 5–86.
- Mullahy J 1997 “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior.” *Review of Economics and Statistics* 79: 586–593.
- Newey WK, and McFadden DL. 1994 “Large Sample Estimation and Hypothesis Testing, Chapter 36” In *Handbook of Econometrics*, edited by Engle RF and McFadden DL, 2111–2245. Amsterdam: Elsevier Science B. V.
- Rubin DB 1974 “Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies.” *Journal of Educational Psychology* 66: 688–701.
- Rubin DB 1977 “Assignment to a Treatment Group on the Basis of a Covariate.” *Journal of Educational Statistics* 2: 1–26.
- StataCorp. 2017 *Stata: Release 15. Statistical Software*. College Station, TX: StataCorp LLC.
- Stock JH, and Watson MW. 2003 *Introduction to Econometrics*. Boston: Addison-Wesley.

- Stock JH, and Watson MW. 2007 Introduction to Econometrics. 2nd ed Boston: Addison-Wesley.
- Stock JH, and Watson MW. 2015 Introduction to Econometrics. 3rd ed Boston: Addison-Wesley.
- Terza JV 2006 “Estimation of Policy Effects Using Parametric Nonlinear Models: A Contextual Critique of the Generalized Method of Moments.” Health Services and Outcomes Research Methodology 6: 177–198.
- Terza JV 2016a “Simpler Standard Errors for Two-Stage Optimization Estimators.” Stata Journal 16: 368–385.
- Terza JV 2016b “Inference Using Sample Means of Parametric Nonlinear Data Transformations.” Health Services Research 51: 1109–1113. [PubMed: 27091770]
- Terza JV 2016c “Supplementary Appendix to ‘Inference Using Sample Means of Parametric Nonlinear Data Transformations.’” Health Services Research. doi: 10.1111/1475-6773.12494.
- Terza JV 2017 “Causal Effect Estimation and Inference Using Stata.” Stata Journal 17:939–961.
- Terza JV. Even Simpler Standard Errors for Two-Stage Optimization Estimators: Mata Implementation via the DERIV Command; Presented at the Stata Conference; Columbus, OH: 2018. July. https://www.stata.com/meeting/columbus18/slides/columbus18_Terza.pdf.
- White H 1994 Estimation, Inference and Specification Analysis. New York: Cambridge University Press.
- Woodridge JM 2010 Econometric Analysis of Cross Section and Panel Data. 2nd Ed Cambridge, MA: MIT Press.
- Wu D 1973 “Alternative Tests of Independence between Stochastic Regressors and Disturbances.” Econometrica 41: 733–750.

Table 1:

Low Birth Weight and Any Smoking.

Person	Y_1	Y_0	$X^* = 1$	$X^* = 0$	FIRST
A	C (Y)	E	G (X)	I	K
1	1	1	1	0	0
2	1	1	1	0	0
3	1	1	1	0	0
4	1	0	1	0	1
5	0	0	1	0	1
B	D	F (Y)	H	J (X)	L
6	1	1	1	0	0
7	1	0	1	0	1
8	1	0	1	0	1
9	0	0	1	0	1
10	0	0	1	0	1

Y_1 \equiv low birth weight potential outcome for counterfactually imposed smoking ($Y_1 = 1$ if LBW).

Y_0 \equiv low birth weight potential outcome for counterfactually imposed nonsmoking ($Y_0 = 1$ if LBW).

Y \equiv observable low birth weight outcome ($Y = 1$ if LBW).

X^* \equiv counterfactually imposed smoking status ($X^* = 1$ if smoker).

X \equiv observable smoking status ($X = 1$ if smoker).

FIRST \equiv infant is mother's first-born (FIRST = 1 if first-born).

Bold value cells contain observable population values.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

DGP-Based “Marginal Effect” Estimates.

Method	“Marginal Effect” (38)	t-stat	p-value
Poisson	0.069	5.30	<0.001
GMM	0.143	3.57	<0.001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript