Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight

by

Joseph V. Terza*
Department of Economics
Indiana University Purdue University Indianapolis
Indianapolis, IN 46202
Phone: 317-274-4747
Fax: 317-274-0097
Email: jvterza@iupui.edu

(April, 2014)

DO NOT QUOTE WITHOUT AUTHOR'S PERMISSION

*This work was supported by the Agency for Healthcare Research and Quality (R01 HS17434) The author is grateful for the excellent research assistance provided by F. Michael Kunz and Mujde Erten. This work has also benefitted greatly from seminar and workshop presentations at the following universities: Alabama, Emory, Florida, Georgia, Indiana, Maryland, Medical University of South Carolina, Michigan, Rochester, South Carolina, South Florida, UNC-Greensboro, UNC-Chapel Hill, and Southern California.

Abstract

Most empirical research in health economics is conducted with the goal of providing scientific evidence that will serve to inform current and future health policy. The use of parametric nonlinear regression (NR) methods for empirical analysis in health economics abounds. Studies that offer clear policy-relevant interpretations of NR results are, however, rare. We offer a comprehensive policy analytic framework within which the applied researcher can: 1) clearly define the policy-relevant estimation objective; 2) consistently estimate that objective using NR methods designed to account for the possible endogeneity of the policy variable of interest; 3) conduct correct asymptotic inference; and 4) offer policy-relevant interpretations of the empirical results. For binary policies, Rubin (1974, 1977) developed the *potential outcomes framework (POF)*. We propose a generally applicable extension of the POF (EPOF) which covers a broad range of policy analytic contexts. In particular, our EPOF accommodates: a) a non-binary policy variable of interest (X_p); b) policy-relevant counterfactual versions of X_p that are not fixed values; and c) a policy-defining increment to X_p that is not constant. Moreover, our EPOF facilitates the use of extant nonlinear regression (NR) methods that correct for potential bias due to the endogeneity of X_p. As a case in point, we consider the analysis of potential gains in infant birth weight that may result from a prenatal smoking prevention and cessation policy which, if fully effective, would maintain zero levels of smoking for non-smokers (prevention) and convince smokers to quit before becoming pregnant (cessation). In the context of our EPOF, using endogeneity-correcting NR methods, we re-analyze the data examined by Mullahy (1997) and estimate the potential effect of the smoking prevention/cessation policy described above. The EPOF should serve as a useful guide to applied health policy analysts.

1. Introduction

Most empirical research in health economics is conducted with the goal of providing scientific evidence that will serve to inform current and future health policy. Such policy analytic studies typically focus is on a particular variable (the *policy variable* -- X_p) that is at present, or will in the future be, under the control of a policy-making entity. Broadly stated, the key policy analytic objective is estimation of the effect that a change in X_p would have on a targeted policy relevant outcome of interest (Y) [henceforth the *policy effect (PE)*]. For contexts in which X_p is binary, Rubin (1974, 1977) developed the *potential outcomes framework (POF)* which facilitates clear definition and interpretation of various policy relevant treatment effects. The key concept in this framework is the *potential outcome* (Y_j) – the random variable representing the outcome as it would have manifested if the value of X_p were counterfactually fixed (i.e., exogenously mandated to be) at a specified value (j = 0 or 1) ceteris paribus. In the POF, the policy effect is measured as the difference between the distributions of Y_0 and Y_1 or some particular aspect (parameter) thereof. In many contexts relevant to health policy, however, two required features of the POF do not hold, viz.: 1) X_p is often non-binary -- i.e. it is a discrete (e.g. a count) or continuous variable; and 2) neither the policy relevant versions of X_p nor the policy relevant increment to X_p are to be fixed in value – instead they would vary across the population (i.e., they are random variables rather than fixed values). As a case in point, we consider the analysis of potential gains in infant birth weight (Y) that may result from effective prenatal smoking prevention and cessation policy. Here, X_p represents smoking during pregnancy and the policy of interest, if fully effective, would maintain zero levels of smoking for

non-smokers (prevention) and convince smokers to quit before becoming pregnant (cessation). It is clear in this example that the relevant pre-policy version of X_p is not fixed in value – it is the random variable representing the pre-policy distribution of smoking levels across the population of pregnant women. By the same token, the policy-driven increment required to bring individual prenatal smoking levels to zero must also vary across the population. We propose a generally applicable extension of the POF in which: a) X_p need not be binary; b) counterfactually imposed policy-relevant versions of X_p need not be fixed values; and c) the policy-defining increment to X_p need not be a constant.[1] Moreover, our extended POF is designed to facilitate empirical policy analysis using extant nonlinear regression (NR) methods that correct for possible bias due to the endogeneity of X_p.

We denote the potential outcome as $Y_{X_p^*}$ and the counterfactually mandated version of the policy variable as X_p^* (possibly a random variable). In our extended POF, the *policy effect* of interest can be broadly stated as the difference between the distributions of $Y_{X_{p1}}$ and $Y_{X_{p2}}$ [or some particular aspect (parameter) thereof], where X_{p1} and X_{p2} represent well-defined and distinct counterfactually imposed pre- and post-intervention versions of the policy variable, respectively. Without loss of generality, we represent the policy increment and pre- and post-policy scenarios as Δ, $X_{p1} = X_p^*$ and $X_{p2} = X_p^* + \Delta$, respectively; all of which are, in general,

[1] Angrist and Pischke (2009) extend the POF to analyses of policies in which X_p is non-binary. Their extended POF does not, however, explicitly accommodate features (b) and (c).

random variables (possibly degenerate).[2,3] For the remainder of the discussion we focus on the following *average incremental effect* (AIE) as the policy effect of interest

$$\text{AIE}(\Delta) = \text{E}[\text{Y}_{\text{X}_\text{p}^*+\Delta}] - \text{E}[\text{Y}_{\text{X}_\text{p}^*}]. \quad (1)$$

Expression (1) is, in fact, quite general. For example, when the policy variable is binary, if we set $\text{X}_\text{p}^* = 0$ and $\Delta = 1$ then (1) measures the *average treatment effect* (ATE). When the policy variable is continuous and Δ approaches 0 then $\lim_{\Delta \to 0} (\text{AIE}(\Delta)/\Delta)$ represents the *average marginal effect* (AME) of an infinitesimal change in the policy variable. Placing (1) in the context of our smoking and birth weight policy analysis, we have that X_p^* is the random variable representing the pre-policy distribution of smoking levels for pregnant women and $\Delta = -\text{X}_\text{p}^*$ (also a random variable).

The remainder of the paper is organized as follows. In the next section, we discuss our general approach to the specification (estimation) of (1) [and all of its interesting variants] using extant NR models (methods) that are designed to account for the potential endogeneity of X_p. Section 3, details the implementation of two such endogeneity-correcting NR models (methods)

---

[2]Henceforth we will adhere to the following notational conventions: 1) uppercase letters for random variables (e.g., A); 2) lowercase letters for particular values in the support of the random variable in question (e.g., a); 3) lowercase with parenthetic omega (ω) for the realization of the random variable in question for a particular individual in the relevant population [e.g. a(ω)]; and 4) uppercase letters with an "i" subscript for the sampled version of the random variable in question (e.g. A_i).

[3] The thought experiment underlying the "policy" is counterfactual (or at least partially counterfactual) in the sense that, for a given individual in the relevant population (ω), it may be the case that: $\text{x}_\text{p}(\omega) \neq \text{x}_{\text{p}1}(\omega)$; $\text{x}_\text{p}(\omega) \neq \text{x}_{\text{p}2}(\omega)$; or $\text{x}_\text{p}(\omega) \neq \text{x}_{\text{p}1}(\omega)$ and $\text{x}_\text{p}(\omega) \neq \text{x}_{\text{p}2}(\omega)$.

for the analysis of the impact on birth weight of effective smoking prevention/cessation policy for pregnant women: the generalized method of moments (GMM) estimator of Mullahy (1997) for exponential models; and the two-stage residual inclusion (2SRI) estimator of Terza et al. (2008) for the general NR case. In section 4, we apply these methods to the analysis of prenatal smoking policy using the analysis sample examined by Mullahy (1997).[4] For the purpose of comparison, we also conduct the analysis under a linearity assumption using the conventional linear instrumental variables (IV) method. In addition, to set a baseline for comparison, we ignore the potential endogeneity of smoking during pregnancy and estimate (1) based on OLS (in the linear case) and NLS (in the exponential case) regression results. The results are discussed in section 5. The final section summarizes and concludes.

2. Specification and Estimation of the AIE: Potential Outcomes and Nonlinear Models

In the following we define the potential outcome concept, as discussed above, in a way that makes it amenable to the specification [estimation] of (1) and all of its interesting variants (e.g. ATE and AME) via nonlinear regression (NR) models [methods] that are designed to accommodate the possible endogeneity of the observed value of the policy variable. We begin with some definitions. Let Y and X_p denote the observable versions of the outcome and policy variables, respectively. We define a *confounder for a specified variate* U to be any other variate that is correlated with *both* Y *and* U. The conditional mean regression model

$$E[Y \mid X_p, V] = \mu(X_p, V, \tau) \qquad (2)$$

[4] Mullahy (1997) applies his GMM method, but does not go on to use the results to analyze potential policy effects.

is said to be *causal* if V is *comprehensive* in that it comprises all possible confounders for X_p and its own elements. An implication of this definition is that, conditional on V, differences in the mean of the observed value of Y can be exclusively attributed to differences in the observed value of X_p.[5] We define the *causal regression model* as the following version of (2)

$$Y = \mu(X_p, X_o, X_u, \tau) + e \qquad (3)$$

where $V = [X_o \quad X_u]$, X_p is the observable value of the policy variable, X_o denotes a vector of observable confounders that is *exogenous* in that there are no unobservable confounders for its elements, X_u represents a scalar comprising all of the unobservable (unobserved or omitted) confounders for X_p, τ is a vector of parameters to be estimated; and $e \equiv Y - \mu(X_p, X_o, X_u, \tau) \equiv Y - E[Y \mid X_p, X_o, X_u]$ is the error term.[6] Note that due to the

---

[5] Strictly speaking, for any vector of conditioning variates (C), we can write a regression model of the form

$$Y = E[Y \mid X_p, C] + \upsilon \qquad (*)$$

where $\upsilon = Y - E[Y \mid X_p, C]$ so that $E[\upsilon \mid X_p, C] = 0$. This implies that $[X_p \quad C]$ is not correlated with υ (i.e., $\text{cov}([X_p \quad C], \upsilon) = 0$). Note that $E[\upsilon \mid X_p, C] = 0$ [or $\text{cov}([X_p \quad C], \upsilon) = 0$] is not enough to qualify (*) as "causal." The key here is that, although the regressors $[X_p \quad C]$ are not correlated with the error term υ (which implies that consistent estimates of the parameters of $E[Y \mid X_p, C]$ can be obtained, if it is parametric), if C is not comprehensive in the sense that it does not include all of the confounders for X_p and its own elements, then $E[Y \mid X_p, C]$ cannot be used for causal analysis.

[6] Note that (3) is general in all relevant respects. In particular, note that we could have begun the discussion with the following more primitive form

$$Y = \psi(X_p, X_o, X_u, \varepsilon, \kappa) \qquad (**)$$

where ε is a non-additive unobservable component that is mean independent of X_p, X_o and X_u, and κ is a vector of parameters. The function $\mu(\ )$ in (3) could, therefore, be defined as the

presence of the unobservable confounder X_u in (3), direct application of conventional NR methods for the estimation of τ is precluded. Up to this point in the discussion, we have been using the term *endogeneity* without formal definition. Henceforth, we define a conditioning variate for Y to be *endogenous* if it has an unobservable confounder. Including X_u in (3) serves two important and coincident purposes. First, it accounts for the possible endogeneity of X_p. Secondly, it guarantees that (3) is causal because, by design, the vector of conditioning variates $[X_o \quad X_u]$ comprises all possible confounders for X_p and its own elements (i.e., $[X_o \quad X_u]$ is comprehensive).

We use (3) to develop an operational definition of the potential outcome, $Y_{X_p^*}$. We first draw the distinction between the observable value of the policy variable (X_p) and the version of that variate *as if it were exogenously mandated* (say X_p^{exog}). Next we assume that the observable value of the outcome for any individual in the relevant population (Y) is the same as it would have been if the observable value of the policy variable were exogenously imposed rather than the product of individual choice.[7] In other words, for individual ω

$$y(\omega) = y_{X_p^{exog}}(\omega). \tag{4}$$

---

result of integrating (**) with respect to ε conditional on X_p, X_o and X_u. For example, suppose Y is binary and

$$\psi(X_p, X_o, X_u, \varepsilon, \kappa) = I(X_p\kappa_p + X_o\kappa_o + X_u\kappa_u + \varepsilon > 0)$$

and $(\varepsilon \mid X_p, X_o, X_u)$ is standard normal distributed, then

$$\mu(X_p, X_o, X_u, \tau) = \Phi(X_p\kappa_p + X_o\kappa_o + X_u\kappa_u)$$

where $\tau = [\kappa_p \quad \kappa_o' \quad \kappa_u]$.

[7] This is a standard assumption in the POF.

This assumption, combined with (3) implies that

$$Y_{X_p^{exog}} = \mu(X_p^{exog}, X_o, X_u, \tau) + e. \quad (5)$$

We extend (5) to any counterfactually mandated version of the policy variable (X_p^*) and thereby obtain the following general and practical definition of the *potential outcome*

$$Y_{X_p^*} = \mu(X_p^*, X_o, X_u, \tau) + e. \quad (6)$$

In the context of our birth weight and smoking policy model discussed in the previous section, we can use Terza's (2006) recast of Mullahy's (1997) model and posit the following version of (3)[8]

$$Y = \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u) + e \quad (7)$$

where $\beta = [\beta_p \quad \beta_o' \quad \beta_u]$ is a vector of parameters to be estimated, and $e \equiv Y - \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)$ is the error term. The corresponding potential outcome at X_p^* is

$$Y_{X_p^*} = \exp(X_p^*\beta_p + X_o\beta_o + X_u\beta_u) + e. \quad (8)$$

[8] Here and for the remainder of the discussion, wherever possible, we use the symbol "β" to denote regression coefficient parameters. This is an abuse of notation because distinct model specifications warrant distinct parametric notation. Given the multiplicity of models that we consider, however, we are compelled to adhere to this convention in order to conserve notation.

Returning now to the general case, using (6) and the law of iterated expectations it can be shown that

$$E[Y_{X_p^*}] = E_{X_p^*, X_o, X_u}\left[\mu(X_p^*, X_o, X_u, \tau)\right]. \quad (9)$$

Using (9) we can rewrite (1) as

$$AIE(\Delta) = E_{X_p^*+\Delta, X_o, X_u}\left[\mu(X_p^* + \Delta, X_o, X_u, \tau)\right] - E_{X_p^*, X_o, X_u}\left[\mu(X_p^*, X_o, X_u, \tau)\right]. \quad (10)$$

If for the moment, we assume the existence of a consistent estimator of τ (say $\hat{\tau}$) and an appropriate observable proxy value for the unobservable X_u [say $\hat{X}_u(W, \hat{\delta})$ -- which, as we discuss later, may require auxiliary estimation of a parameter vector δ using additional variables W] then (10) could be consistently estimated as[9]

$$\widehat{AIE(\Delta)} = \sum_{i=1}^{n} \frac{1}{n}\left\{\mu(X_{pi}^* + \Delta_i, X_{oi}, \hat{X}_u(W, \hat{\delta}); \hat{\tau}) - \mu(X_{pi}^*, X_{oi}, \hat{X}_u(W, \hat{\delta}); \hat{\tau})\right\} \quad (11)$$

where the i subscript denotes the ith individual in a sample of size n (i = 1, …, n). In the context of our birth weight and smoking policy analysis, using (7) we can write (10) and (11) as[10]

$$AIE(\Delta) = E_{X_p^*+\Delta, X_o, X_u}\left[\exp(X_o\beta_o + X_u\beta_u)\right] - E_{X_p^*, X_o, X_u}\left[\exp(X_p^*\beta_p + X_o\beta_o + X_u\beta_u))\right] \quad (12)$$

and

[9] In an appendix that will be supplied upon request, we show that specifications and estimators similar to (10) and (11), respectively, can be devised for ATE and AME.

[10] Expressions (12) and (13) make use of the fact that, for the smoking prevention/cessation policy that we are considering, $X_{pi}^* + \Delta_i = 0$.

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \Big\{ \exp(X_{oi}\hat{\beta}_o + \hat{X}_{ui}(W_i, \hat{\delta})\hat{\beta}_u) - \exp(X^*_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}(W_i, \hat{\delta})\hat{\beta}_u) \Big\} . \qquad (13)$$

Later, we will discuss methods for consistent estimation of τ and the construction of an appropriate proxy for X_u [if needed for the estimation of (10)]. For now, we turn to the asymptotic properties of (11) [in particular, the formulation of its correct asymptotic standard error]. To simplify the notation, we rewrite (11) as

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \widehat{aie}_i \qquad (14)$$

where

$$\widehat{aie}_i = aie(X^*_{pi}, X_{oi}, W_i, \hat{\tau}, \hat{\delta})$$

and $aie(X^*_p, X_o, W, \tau, \delta) = \mu(X^*_p + \Delta, X_o, X_u(W,\delta), \tau) - \mu(X^*_p, X_o, X_u(W,\delta), \tau)$. Terza (2012) shows that (14) can be cast as a two-stage optimization (2SOPT) estimator, and using standard asymptotic theory for 2SOPT he demonstrates that[11]

$$\sqrt{\frac{n}{\widehat{a\,var}\left(\widehat{\mathrm{AIE}(\Delta)}\right)}} \left(\widehat{\mathrm{AIE}(\Delta)} - \mathrm{AIE}(\Delta)\right) \xrightarrow{d} n(0,1) . \qquad (15)$$

where

[11] For comprehensive discussions of 2SOPT estimators and their asymptotic properties see White (1994, Chapter 6); and Newey and McFadden (1994). These authors extend the results of Murphy and Topel (1985) for two-stage maximum likelihood estimators to the more general class of 2SOPT estimators.

$$\widehat{\text{a var}}\left(\widehat{\text{AIE}(\Delta)}\right) = \left(\frac{\sum_{i=1}^{n} \nabla_{[\tau\ \delta]}\widehat{\text{aie}}_i}{n}\right) \widehat{\text{AVAR}}([\hat{\tau}\quad \hat{\delta}]) \left(\frac{\sum_{i=1}^{n} \nabla_{[\tau\ \delta]}\widehat{\text{aie}}_i}{n}\right)' + \left(\frac{\sum_{i=1}^{n} (\widehat{\text{aie}}_i - \widehat{\text{AIE}(\Delta)})^2}{n}\right) \tag{16}$$

where $\nabla_{[\tau\ \delta]}\widehat{\text{aie}}_i$ denotes the gradient of $\text{aie}(X_p, X, W, \tau, \delta)$ [a row vector] evaluated at X^*_{pi}, X_{oi}, W_i and $[\hat{\tau}\quad \hat{\delta}]$; and $\widehat{\text{AVAR}}([\hat{\tau}\ \hat{\delta}])$ is the estimated asymptotic covariance matrix of $[\hat{\tau}\quad \hat{\delta}]$.[12] For the pseudo causal regression specification in (7) we have

$$\widehat{\text{aie}}_i = \exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u) - \exp(X^*_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)$$

where $\hat{\tau} = [\hat{\beta}_p \quad \hat{\beta}'_o \quad \hat{\beta}_u]$.

3. Parameter Estimation and Accommodating the Unobservable Confounder

Two issues related to the implementation of the estimator in (11) have yet to be resolved: 1) consistent estimation of τ; and 2) an appropriate accommodation for the unobservable confounder X_u. We begin by noting that there are at least two special cases in which X_u can, as a practical matter, be ignored in the formulations of (10) and (11). First, if the pseudo regression model in (3) is linear, i.e.

$$\mu(X_p, X_o, X_u, \tau) = X_p\beta_p + X_o\beta_o + X_u\beta_u \tag{17}$$

then (10) and (11) can be written

[12] In an appendix that will be supplied upon request, we show that asymptotic properties similar to (15) and (16) can be derived for the ATE and AME variants of (11).

$$\mathrm{AIE}(\Delta) = \mathrm{E}[\Delta\beta_p] \tag{18}$$

and

$$\widehat{\mathrm{AIE}(\Delta)} = \hat{\beta}_p \sum_{i=1}^{n} \frac{1}{n} \Delta_i \tag{19}$$

respectively. Terza (2006) shows that, in this case, the conventional linear instrumental variables estimator can be used to obtain a consistent estimate of β_p (say $\hat{\beta}_p$) if we have at least one identifying instrumental variable (IV); i.e. a variable (or vector of variables) W^+ that:[13,14]

i) is not included in the vector of observable confounders,

ii) is sufficiently correlated with X_p (i.e., the IVs are not weak),

iii) satisfies $\mathrm{E}[X_u \mid W] = \rho$ (a constant), where $W = [X_o \quad W^+]$,

and

iv) does not systematically affect Y; i.e., $\mathrm{E}[Y \mid X_p, W, X_u] = X_p\beta_p + X_o\beta_o + X_u\beta_u$. (20)

It is easy to see that, using (16), the true asymptotic standard error of (19) can be obtained as

$$\widehat{\mathrm{a\,var}}\left(\widehat{\mathrm{AIE}(\Delta)}\right) = \left(\frac{\sum_{i=1}^{n} \Delta_i}{n}\right)^2 \mathrm{a\,var}(\hat{\beta}_p) + \left(\frac{\sum_{i=1}^{n} (\hat{\beta}_p\Delta_i - \widehat{\mathrm{AIE}(\Delta)})^2}{n}\right) \tag{21}$$

where $\mathrm{a\,var}(\hat{\beta}_p)$ denotes the asymptotic standard error of the conventional linear IV estimator of β_p.

---

[13] See Wooldridge (2010), Chapter 4, for a discussion of the conventional linear instrumental variables estimator.

[14] Note that although X_u can be ignored in the formulations of (18) and (19), it must be accounted for in the estimation of β_p via the conventional linear IV method (see Terza, 2006).

The other instance in which the unobservable confounder X_u does not directly appear in (10) and (11) is the case in which the pseudo causal regression is specified as in (7). This is, in essence, the model considered by Mullahy (1997).[15] Suppose, IV conditions (i) and (ii) in (20) hold with (iii) and (iv) replaced by

iii*) $E[\exp(X_u\beta_u) \mid W] = \kappa$ (a constant)

and

iv*) $E[Y \mid X_p, W, X_u] = \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)$.

Under these conditions Terza (2006) shows that (10) [(12)] can be rewritten as

$$AIE(\Delta) = E\left[\exp([X_p^*+\Delta]\beta_p + X_o\beta_o^+) - \exp(X_p^*\beta_p + X_o\beta_o^+)\right]$$

where β_o^+ is the same as β_o except for a shift of its intercept element by +ln(κ). In the context of our birth weight and smoking policy analysis we have[16]

$$AIE(\Delta) = E\left[\exp(X_o\beta_o^+) - \exp(X_p^*\beta_p + X_o\beta_o^+)\right]. \qquad (22)$$

Therefore, a proxy value for the unobservable X_u is not required in this case. Under the above assumptions, Mullahy (1997) derives a very clever generalized method of moments (GMM) estimator that is consistent for β_p and β_o^+. The relevant version of (10) in this case is[17]

[15] Terza (2006) draws the connection between the approach we take in the present discussion and that of Mullahy (1997).

[16] Here again we use the fact that for the smoking prevention/cessation policy that we are considering $X_{pi}^* + \Delta_i = 0$.

[17] Mullahy (1997) does not extend his analysis to the estimation of policy effects like (12).

$$\widehat{\mathrm{AIE}}(\Delta) = \sum_{i=1}^{n} \frac{1}{n} \left\{ \exp(X_{oi}\hat{\beta}_o^+) - \exp(X_{pi}^*\hat{\beta}_p + X_{oi}\hat{\beta}_o^+) \right\} \tag{23}$$

where $\hat{\beta}_p$ and $\hat{\beta}_o^+$ are the GMM estimates. The correct asymptotic standard error for (23) is obtained using (16) with

$$\widehat{\mathrm{aie}}_i = \exp(X_{oi}\hat{\beta}_o^+) - \exp(X_{pi}^*\hat{\beta}_p + X_{oi}\hat{\beta}_o^+)$$

$$\nabla_{[\tau\ \delta]}\widehat{\mathrm{aie}}_i = \nabla_{\beta}\widehat{\mathrm{aie}}_i = [\nabla_{\beta_p}\widehat{\mathrm{aie}}_i \quad \nabla_{\beta_o^+}\widehat{\mathrm{aie}}_i]$$

$$\nabla_{\beta_p}\widehat{\mathrm{aie}}_i = -\exp(X_{pi}^*\hat{\beta}_p + X_{oi}\hat{\beta}_o^+)X_{pi}^*$$

$$\nabla_{\beta_o^+}\widehat{\mathrm{aie}}_i = \left[\exp(X_{oi}\hat{\beta}_o^+) - \exp(X_{pi}^*\hat{\beta}_p + X_{oi}\hat{\beta}_o^+) \right] X_{oi}.$$

Similar methods for ATE (binary X_p) and AME (continuous X_p with infinitesimal and constant Δ) estimation can be derived.[18]

For the general case characterized by the pseudo causal regression (3) in which direct inclusion of X_u in (10) and (11) cannot be avoided, Terza et al. (2008) suggest a two-stage residual inclusion (2SRI) estimator based on assumptions (i), (ii) and (iii) in (20) and the following

iv**) $E[Y \mid X_p, W, X_u] = \mu(X_p, X_o, X_u, \tau)$

and

v) $X_p = r(W, \delta) + X_u$ (24)

[18] An appendix detailing these derivations will be supplied upon request.

where $r(W,\delta)$ is a known function of W and a vector of parameters to be estimated, δ. Under these conditions (10) can be rewritten as

$$\begin{aligned} AIE(\Delta) &= E\left[\mu(X_p^* + \Delta,\, X_o,\, X_u, \tau) - \mu(X_p^*,\, X_o,\, X_u, \tau)\right] \\ &= E\left[\mu(X_p^* + \Delta,\, X_o,\, X_p - r(W,\delta), \tau) - \mu(X_p^*,\, X_o,\, X_p - r(W,\delta), \tau)\right]. \end{aligned} \tag{25}$$

The AIE in (25) can be consistently estimated using (11), with

$$\hat{X}_{ui}(W_i, \hat{\delta}) = X_{pi} - r(W_i, \hat{\delta}) \tag{26}$$

where $\hat{\delta}$ is a consistent estimate of δ. Under the above assumptions, Terza et al. (2008) show that the following 2SRI estimator is consistent for $[\tau \quad \delta]$

First Stage

Obtain a consistent estimate of the vector δ ($\hat{\delta}$) by applying the nonlinear least squares (NLS) method to (v) in (24).[19] Next, compute the residual using (26).

Second Stage

Consistently estimate τ by applying NLS to the following version of (3)

$$Y = \mu(X_p,\, X_o, \hat{X}_{ui}(W_i, \hat{\delta}); \tau) + e^{2SRI} \tag{27}$$

[19]Any consistent estimator of δ can be used here. The choice of estimator depends upon the available non-sample information. For example, if $(X_p \mid W)$ is known to be a member of a particular family of parametric distributions, then the full information maximum likelihood (FIML) method can be used to estimate δ.

where e^{2SRI} is the regression error term.[20,21]

The 2SRI estimator is consistent and asymptotically normal. The asymptotic properties of this estimator can be derived as a special case of the generic 2SOPT estimator.[22,23,24]

4. Smoking During Pregnancy and Infant Birth Weight

As a running illustration for the concepts and methods discussed above, we analyze the potential impact of effective smoking prevention/cessation policy on birth weight. We apply the following three estimators to the same data and variable specification analyzed by Mullahy (1997) in this context: (19) – using conventional linear IV estimates of the parameters of (17); (23) – using Mullahy's (1997) GMM estimates of the parameters of (7); and (11) – using 2SRI estimates (Terza et al., 2008) of the parameters of a flexible-form version of (3). All of these estimators account for the possible endogeneity of X_p. The analysis sample was taken from the

[20]See footnote 17. A similar comment is true for second-stage estimation of τ.
[21] If (17) holds and the r( ) function in condition (v) of (24) is also linear, 2SRI is identical to the conventional linear IV estimator.
[22] An appendix detailing these derivations will be supplied upon request.
[23] The 2SRI approach can similarly be used to estimate the parameters underlying an ATE or AME. An appendix detailing these methods will be supplied upon request.
[24] There exists some controversy regarding the use of the 2SRI approach for the analysis of ATE (the case in which X_p is binary); despite the clear consistency of 2SRI in this case under assumptions (i), (ii) and (iii) in (20); and (iv**) and (v) in (24). The main criticism appears to be that these assumptions are "nonstandard" vis-a-vis, and difficult to reconcile with, the assumptions underlying conventional models involving endogenous binary variables [e.g., Heckman (1978) for linear models, Terza (1998) for count models, and the bivariate probit model for binary outcome models]. To avoid controversy when X_p is binary, practitioners can implement the general framework proposed by Terza (2009). This framework conforms to conventional behavioral assumptions for binary X_p and subsumes Heckman (1978), Terza (1998), and bivariate probit as special cases. In fact, the modeling framework and estimation method of Terza (2009) accommodates any nonlinear regression model with an endogenous binary regressor.

Child Health Supplement to the 1988 National Health Interview Survey and has 1,388 observations. The definitions of the variables included in the regressions are given in Table 1. In estimating the AIE we take X_p^{exog} [defined in assumption (4)] as the value of X_p^*. In other words, we take the pre-policy distribution of the policy variable to be as it was at the time of sampling – i.e., the same as the observable policy random variable X_p. The linear IV and GMM based estimation results for AIE [(19) and (23), respectively] are displayed in the first and second rows of Table 2.[25]

To investigate the robustness of the GMM estimates, we consider the following flexible functional form for the pseudo causal regression model (3)

$$E[Y \mid X_p, X_o, X_u] \equiv \mu(X_p, X_o, X_u, \tau) = \begin{cases} \left(\left(\frac{\gamma}{2}\left(X_p\beta_p + X_o\beta_o + X_u\beta_u\right) + 1 \right)^2 \right)^{\frac{1}{\gamma}} & \text{if} \quad \gamma \neq 0 \\ \exp\left(X_p\beta_p + X_o\beta_o + X_u\beta_u\right) & \text{if} \quad \gamma \to 0 \end{cases} \tag{28}$$

where $\tau = [\beta_p \quad \beta_o' \quad \beta_u \quad \gamma]$, and $\gamma \in (0,\ 2]$ is a scalar parameter. This is a variant of the inverse of the flexible functional form suggested by Box and Cox (1964) which was first considered and implemented by Wooldridge (1992) and has since been widely applied.[26] The

[25] Details of the conventional IV and the GMM parameter (β) estimates are shown in the first and second columns of Table 3. In order to maintain comparability with the 2SRI estimates for a more general version of the model (discussed later), particularly with regard to the sum of squared residuals goodness-of-fit measure, we implemented the 2SRI version of the IV estimator (see Hausman, 1978; and Terza et al., 2008).

[26] See Abrevaya (2002), Abrevaya and Hausman (2004), Arvin-Rad (1997), Basu and Rathouz, (2005), Basu (2005), Basu et al. (2006), Berndt et al. (1990), Blackburn (2007), Gencay and

inverse Box-Cox (IBC) functional form in (28) approaches the exponential model as $\gamma \to 0$; and when $\gamma = 2$ and $X_p\beta_p + X_o\beta_o + X_u\beta_u > -1$, it reduces to a simple linear regression model.[27] With a view toward estimation of τ via the 2SRI method detailed in the previous section, we also posit the following IBC version of the auxiliary regression defined in condition (v) of (24)

$$r(W,\delta) = \begin{cases} \left(\left(\frac{\eta}{2}(W\alpha)+1\right)^2\right)^{\frac{1}{\eta}} & \text{if} \quad \eta \neq 0 \\ \exp(W\alpha) & \text{if} \quad \eta \to 0 \end{cases} \tag{29}$$

where $\delta = [\alpha \quad \eta]$ and $\omega \in (0,\ 2]$ is a scalar parameter. The model defined by (28) and (29) subsumes the conventional linear IV model as the special case in which $\gamma = \eta = 2$. We applied 2SRI and found that the NLS optimizing value of the second-stage 2SRI-IBC estimate of γ is arbitrarily close to zero. Therefore based on (28), we take (7) to be the appropriate specification for the pseudo causal regression model and apply NLS to the following version of (27) in the second stage of the 2SRI estimator

$$Y = \exp(X_p\beta_p + X_o\beta_o + \hat{X}_u(W,\hat{\delta})\beta_u) + e^{2SRI} \tag{30}$$

Yang (1996, a and b), Kenkel and Terza (2001), Machadoa and Mata (2000), Showalter (1994), Taylor (2008), Terza-Basu-Rathouz (2008), Terza-Bradford-Dismuke (2008), Wooldridge (1994).

[27] When $\gamma = 2$, the pseudo regression (3) becomes $E[Y \mid X_p, X_o, X_u] = g(Z) = |Z+1|$, where $Z = X_p\beta_p + X_o\beta_o + X_u\beta_u$. In general, g(Z) is V-shaped with vertex (-1,0), but if $Z > -1$ then only the positively sloped linear portion of the function is relevant. In this case (28) corresponds to the linear version of (3) given in (17).

where $\hat{X}_u(W,\hat{\delta})$ is defined as in (26) with $r(W,\delta) = \left(\left(\frac{\eta}{2}(W\alpha)+1\right)^2\right)^{\frac{1}{\eta}}$ and $\hat{\delta}$ is the first stage estimate of δ obtained by applying NLS to (24). The results of this 2SRI re-estimation are given in the third and fourth columns of Table 3. Under our assumption that $X_p^* = X_p^{exog}$, the relevant version of (25) in this 2SRI setup is

$$AIE(\Delta) = E\left[\exp(X_o\beta_o + [X_p - r(W,\delta)]\beta_u) - \exp(X_p^{exog}\beta_p + X_o\beta_o + [X_p - r(W,\delta)]\beta_u)\right] \tag{31}$$

which we consistently estimated using the following variant of (11)

$$\widehat{AIE(\Delta)} = \sum_{i=1}^{n}\frac{1}{n}\left\{\exp(X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u)\right\} \tag{32}$$

where $\hat{X}_{ui} = X_{pi} - r(W_i,\hat{\delta})$. The asymptotic properties of (32) follow (15) and (16) with

$$\widehat{aie}_i = \exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)$$

$$\nabla_{[\tau\ \delta]}\widehat{aie}_i = [\nabla_{\beta_p}\widehat{aie}_i \quad \nabla_{\beta_o}\widehat{aie}_i \quad \nabla_{\beta_u}\widehat{aie}_i \quad \nabla_{\alpha}\widehat{aie}_i \quad \nabla_{\eta}\widehat{aie}_i]$$

$$\nabla_{\beta_p}\widehat{aie}_i = -\exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u)X_{pi}$$

$$\nabla_{\beta_o}\widehat{aie}_i = \left[\exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)\right]X_{oi}$$

$$\nabla_{\beta_u}\widehat{aie}_i = \left[\exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)\right]\hat{X}_{ui}$$

$$\nabla_{\alpha}\widehat{aie}_i = -\beta_u[\exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)]\xi^{\left(\frac{1}{\eta}-\frac{1}{2}\right)}W$$

$$\nabla_{\eta}\widehat{aie}_i = \beta_u[\exp(X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)$$
$$- \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + X_u(W_i,\hat{\delta})\hat{\beta}_u)]r(W,\delta)\left[\frac{1}{\eta^2}\ln(\xi) - \frac{W\alpha}{\eta\xi^{\frac{1}{2}}}\right]$$

$$\widehat{AVAR}([\hat{\tau}\ \hat{\alpha}]) = \widehat{AVAR}([\hat{\beta}_p\ \ \hat{\beta}_o\ \ \hat{\beta}_u\ \ \hat{\alpha}\ \ \hat{\eta}])$$

$$\xi = \left(\frac{\eta}{2}(W\alpha)+1\right)^2$$

and $\widehat{AVAR}([\hat{\beta}_p\ \ \hat{\beta}_o\ \ \hat{\beta}_u\ \ \hat{\alpha}\ \ \hat{\eta}])$ denotes the estimated asymptotic covariance matrix of the 2SRI estimator of $[\hat{\beta}_p\ \ \hat{\beta}_o\ \ \hat{\beta}_u\ \ \hat{\alpha}\ \ \hat{\eta}]$. The results for (32) are reported in the third row of Table 2.

As baselines for comparison, we estimate the linear and exponential models by applying OLS and NLS to the following restricted versions of (17) and (7), respectively

$$Y = X_p\beta_p + X_o\beta_o + e^o \tag{33}$$

$$Y = \exp(X_p\beta_p + X_o\beta_o) + e^o \tag{34}$$

where e^o is the regression error term and $E[e^o \mid X_p, X_o] = 0$.[28] The unobservable confounder X_u is not included in (33) and (34) which, therefore, ignore the potential endogeneity of X_p. Here, as for the IV, GMM, and 2SRI estimates discussed above, we take the relevant pre- version of the policy variable (X_p^*) to be X_p^{exog}. For the linear model in (33), the AIE is

[28] The OLS results for (33) and the NLS results for (34) are given in the fifth and sixth columns of Table 3, respectively.

$$\mathrm{AIE}(\Delta) = \mathrm{E}[-X_p^{exog}\beta_p] \tag{35}$$

and is estimated as

$$\widehat{\mathrm{AIE}(\Delta)} = -\hat{\beta}_p \sum_{i=1}^{n} \frac{1}{n} X_{pi} \tag{36}$$

where $\hat{\beta}_p$ is the OLS estimate of β_p and X_{pi} is the observed smoking level for the ith individual. The correct asymptotic standard error for (36) is obtained from (21) with $\hat{\beta}_p$ and $\mathrm{a\,var}(\hat{\beta}_p)$ replaced by $\hat{\beta}_p$ and $\mathrm{a\,var}(\hat{\beta}_p)$, respectively. The results for (36) are displayed in the fourth row of Table 2.

For the exponential model, we again take X_p^{exog} as the relevant value of X_p^* which, combined with (4), implies that the relevant version of the AIE For the exponential model in (34) is

$$\mathrm{AIE}(\Delta) = \mathrm{E}\left[\exp(X_o\beta_o) - \exp(X_p^{exog}\beta_p + X_o\beta_o)\right] \tag{37}$$

which can be consistently estimated using

$$\widehat{\mathrm{AIE}}(\Delta) = \sum_{i=1}^{n} \frac{1}{n}\left\{\exp(X_{oi}\hat{\beta}_o) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o)\right\} \tag{38}$$

where $\hat{\beta}_o$ denotes the NLS estimate of β_o. The correct asymptotic properties of (38) follow (15) and (16) with

$$\widehat{aie}_i = \exp(X_{oi}\hat{\beta}_o) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o)$$

$$\nabla_{[\tau\ \delta]}\widehat{aie}_i = [\nabla_{\beta_p}\widehat{aie}_i \quad \nabla_{\beta_o}\widehat{aie}_i]$$

$$\nabla_{\beta_p}\widehat{aie}_i = -\exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o)X_{pi}$$

$$\nabla_{\beta_o}\widehat{aie}_i = [\exp(X_{oi}\hat{\beta}_o) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o)]X_{oi}$$

$$\widehat{AVAR}([\hat{\tau}\ \hat{\alpha}]) = \widehat{AVAR}([\hat{\beta}_p \quad \hat{\beta}_o])$$

and $\widehat{AVAR}([\hat{\beta}_p \quad \hat{\beta}_o])$ denotes the estimated asymptotic covariance matrix of the NLS estimator of $[\hat{\beta}_p \quad \hat{\beta}_o]$. The results for (38) are given in the fifth row of Table 2.

5. Discussion

We first note that the 2SRI first-stage estimated value of η (1.404) was statistically significantly different from 2 (test of H_o: $\eta = 2$; tstat = 2.011, p-value = .044). This result, combined with the fact that the apparently appropriate second stage (outcome) model is exponential (recall, the estimated value of γ approaches 0), points away from the linear model [specified in (17)] and the conventional IV estimator. Between the two other endogeneity correcting methods – GMM and 2SRI, one might prefer the former because it does not require the specification and estimation of an auxiliary regression (reduced form) akin to (v) in (24). On the other hand, 2SRI dominates all methods with regard to goodness-of-fit based on the sum of squared errors criterion (see the last row of Table 3). Moreover, there are other important reasons for using the 2SRI method in this (and similar) contexts. First, the 2SRI estimates afford a straightforward test for the endogeneity of X_p – viz. a simple t-test of the null hypothesis that

β_u, the coefficient of X_u, in (7) is equal to zero. As can be seen in the third column of Table 3, $\hat{\beta}_u$ is statistically significant at nearly the 5% level, indicating that X_p is indeed endogenous. Secondly, we can use the first-stage 2SRI estimates to conduct a Wald-type test of the joint statistical significance of the IVs [EDFATHER, EDMOTHER, FAMINCOM, CIGTAX88 – the same IVs used by Mullahy (1997)]. Based on the results shown in the fourth column of Table 3, we find these IVs to be only marginally jointly significant (Wald-statistic = 7.283, p-value = .122). Finally, we note that if the 2SRI model specification comprising (29) and (7) as its first and second stages, respectively, is indeed the true model, then based on the results in Table 2 the estimated bias for conventional linear IV is 21% [$(\widehat{AIE(\Delta)}_{IV} - \widehat{AIE(\Delta)}_{2SRI}) / \widehat{AIE(\Delta)}_{2SRI} \times 100$], and for GMM is 30% [similarly computed]. The corresponding estimated average birth weight increases attributable to the hypothesized smoking prevention/cessation policy are (in ounces): IV (2.6); GMM (2.29); and 2SRI (3.26). To place these results in perspective, are we focused on the subsample of mothers with low birth weight (LBW) infants – i.e. those with birth weight less than 88 ounces. LBW has been found to be contributory to perinatal morbidity, learning disabilities, and delayed motor and social development (Centers for Disease Control, 2012). For the LBW subsample, using the 2SRI regression results, we estimated the AIE of our hypothetical prevention/cessation policy to be 5.94 oz. Adding this increment to the birth weights of the infants of each of the 20 smokers in the LBW subsample would be enough to move 6 of them out of the LBW category.

6. Summary and Conclusion

This paper offers a generic and unified framework for empirical policy analysis via NR estimation. The discussion begins with a clear conceptual PO framework for specifying the policy-relevant estimation objective. This framework accommodates any type of policy variable – binary, discrete or continuous – and does not require that either the policy values of interest or prospective policy increments be fixed in value across the relevant population. Moreover, the approach we propose is designed to incorporate the use of NR methods that account for the potential endogeneity of the policy variable. As a case in point, we consider the analysis of potential gains in infant birth weight that may result from effective prenatal smoking prevention and cessation policy. Here, the policy of interest, if fully effective, would maintain zero levels of smoking for the non-smokers (prevention) and convince the smokers to quit before becoming pregnant (cessation). Clearly, the policy variable of interest is likely to be endogenous -- unobserved health behavioral factors that are correlated with smoking during pregnancy may also affect infant birth weight. The relevant pre-policy version of X_p is not fixed in value – it is the random variable representing the pre-policy distribution of smoking levels across the population of pregnant women. By the same token, the policy-driven increment required to bring individual prenatal smoking levels to zero must vary across the population. We follow this example throughout the discussion.

Two levels of estimation are detailed. First, assuming an appropriate NR specification and the existence of corresponding consistent parameter estimates, we show how (and under what conditions) the assumed NR model and results can be used to formulate and estimate the policy effect of interest. Secondly, in the context of the birth weight/smoking example, we

demonstrate the use of two NR modeling and estimation methods – GMM and 2SRI. For comparison, we also discuss the conventional linear IV model and estimator in this context. Full details of the implementation of these methods is given; complete with correct formulations for the asymptotic standard errors. The empirical analyses are conducted using the date from Mullahy (1997). The results favor the use of GMM or 2SRI over IV in this context and, on the basis of goodness-of-fit and diagnostic testing capabilities, one might opt for 2SRI. All methods yielded positive and statistically significant estimates of the effect of the hypothesized smoking prevention/cessation policy on birth weight. Based on the 2SRI results we estimate that the policy, if fully effective, would reduce the percentage of low birth weight babies born by mothers who smoke (and quit due to the policy) by 30%.[29]

The use of NR methods by empirical health policy researchers abounds. Studies that offer clear policy-relevant interpretations of NR results are, however, rare. In this paper, we offer a comprehensive PO-based policy analytic framework within which the applied researcher can: 1) clearly define the policy-relevant estimation objective; 2) consistently estimate that objective using NR methods designed to account for possible endogeneity; 3) conduct correct asymptotic inference; and 4) offer policy-relevant interpretations of the estimation and inferential results. It is hoped that this work will serve as a useful guide to applied health policy analysts.

[29] 2SRI yielded the largest impact estimate of the three.

References

Abrevaya, J. (2002): Computing Marginal Effects in the Box–Cox Model, *Econometric Reviews,* 21, 383–393.

Abrevaya, J. and Hausman, J.A (2004): "Response Error in a Transformation Model with an Application to Earnings-Equation Estimation," *Econometrics Journal*, 7, 366–388.

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

Arvin-Rad, H. (1997): "Comparison of Deterministic and Stochastic Predictors in Nonlinear Systems When the Disturbances are Small," *Econometric Theory*, 13, 368-391.

Basu, A. and Rathouz, P.J. (2005): "Estimating Marginal and Incremental Effects on Health Outcomes Using Flexible Link and Variance Function Models," *Biostatistics*, 6, 93-109.

Basu, A., (2005): "Extended Generalized Linear Models: Simultaneous Estimation of Flexible Link and Variance Functions, *The Stata Journal*, 5, 501-516.

Basu, A., Arondekarc, B.V., and Rathouz, P.J. (2006): "Scale of Interest Versus Scale of Estimation: Comparing Alternative Estimators for the Incremental Costs of a Comorbidity," *Health Economics*, 15, 1091–1107.

Berndt Ernst R.. Showalter Mark H, and Wooldridge Jeffrey M. (1990): "A Theoretical and Empirical Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative," Massachusetts Institute of Technology, Working paper -- WP#3187-90-EFA.

Blackburn, M.L. (2007): "Estimating Wage Differentials Without Logarithms," *Labour Economics*, 14, 73–98.

Box, G.E.P. and Cox D.R. (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society,* Series B 26, 211-252.

Centers for Disease Control (2012): http://ephtracking.cdc.gov/showRbLBWGrowthRetardationEnv.action

Gencay, R. and Yang, X (1996a): "A Forecast Comparison of Residential Housing Prices by Parametric versus Semiparametric Conditional Mean Estimators," *Economics Letters*, 52, 129-135.

Gencay, R. and Yang, X (1996b): "Forecast Comparisons of Residential Housing Prices by Parametric and Semiparametric Regression," *The Canadian Journal of Economics*, 29, Special Issue: Part 2, S515-S519.

Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

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

Kenkel, D., Terza, J. (2001): "The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect," Journal of Applied Econometrics, 16, 165-184.

Machadoa, J. A. F., and Mata, J. (2000): "Box-Cox Quantile Regression and the Distribution of Firm Sizes," *Journal of Applied Econometrics*, 15, 253-274.

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.

Murphy, K.M., and Topel, R.H. (1985): "Estimation and Inference in Two- Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370-379.

Newey, W.K. and McFadden, D. (1994): Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Engle, R.F., and McFadden, D.L., Amsterdam: Elsevier Science B.V., 2111-2245, Chapter 36.

Rubin (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701

Rubin (1977): "Assignment to a Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1-26.

Showalter, M.H. (1994): "A Monte Carlo Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative," *Review of Economics and Statistics*, 76, pp. 560-570.

Taylor, L.O. (2008): "Theoretical Foundations and Empirical Developments in Hedonic Modeling," Chapter 1 in *Hedonic Methods in Housing Markets*, A. Baranzini et al. (eds.), Springer Science + Business Media.

Terza, J. (1998): "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects," *Journal of Econometrics*, 84, 129-154.

_____ (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.

_____ (2009): "Parametric Nonlinear Regression with Endogenous Switching," *Econometric Reviews*, 28, 555-580.

_____ (2012): "Correct Standard Errors for Multi-Stage Regression-Based Estimators: A Practitioner's Guide with Illustrations," Unpublished Manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

_____, Basu, A. and Rathouz, P. (2008): "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling," *Journal of Health Economics*, 27, 531-543.

_____, Bradford, W.D. and Dismuke, C.E. (2008): "The Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note," *Health Services Research*, 43, 1102-1120.

White, H. (1994): *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

Wooldridge, J.M. (1992): "Some Alternatives to the Box-Cox Regression Model," *International Economic Review*, 33, 935-955.

Wooldridge, J.M. (1994): "A Simple Specification Test for the Predictive Ability of Transformation Models," *Review of Economics and Stat*istics, 76, 59-65.

Wooldridge, J.M. (2010): *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed. Cambridge, MA: MIT Press.

Table 1: Birth Weight Model: Variable Definitions and Means

Variable	Definition	Mean
Outcome Variable (Y)		
BIRTHWT	birth weight (lbs.)	7.419
Potentially Endogenous Policy Variable (X_p)		
CIGSPREG	cigarettes smoked per day during pregnancy	2.087
Observable Confounders (X_o)		
PARITY	birth order	1.633
WHITE	1 if white, 0 otherwise	.746
MALE	1 if male, 0 otherwise	.521
Instrumental Variables (W^+)		
EDFATHER	years of education of father	11.324
EDMOTHER	years of education of mother	12.927
FAMINCOM	family income (/1000)	29.027
CIGTAX88	per pack state excise tax on cigarettes (cents)	19.553

Table 2: AIE Estimates

Method	$\widehat{AIE(\Delta)}$	t-statistic	p-value
Linear IV	0.162	3.372	0.001
GMM	0.143	3.135	0.002
2SRI	0.204	2.569	0.010
OLS	0.068	5.330	< .001
NLS	0.071	5.014	< .001

Table 3: Parameter Estimation*

Method / Variable	Linear IV	GMM	2SRI 2nd Stage	2SRI 1st Stage	OLS	NLS
CIGSPREG	**-0.077** (-3.491) [<.001]	**-0.010** (-3.460) [0.001]	**-0.013** (-2.937) [0.003]		**-0.033** (-5.840) [<.001]	**-0.005** (-5.620) [<.001]
PARITY	**0.129** (3.324) [0.001]	**0.018** (3.330) [0.001]	**0.018** (3.292) [0.001]	**0.215** (0.873) [0.383]	**0.109** (2.900) [0.004]	**0.014** (2.990) [0.003]
WHITE	**0.399** (4.772) [<.001]	**0.054** (4.440) [<.001]	**0.054** (4.379) [<.001]	**0.679** (1.377) [0.169]	**0.408** (5.010) [<.001]	**0.056** (4.750) [<.001]
MALE	**0.194** (2.910) [0.004]	**0.027** (2.950) [0.003]	**0.027** (2.930) [0.003]	**0.104** (0.339) [0.734]	**0.194** (2.900) [0.004]	**0.026** (2.900) [0.004]
CONSTANT	**6.956**	**1.939**	**1.944**	**4.794**	**6.889**	**1.932**
IBC-2SRI 1st Stage Residual $\hat{X}_u$	--	--	**0.008** (1.949) [0.051]	--	--	--
EDFATHER	--	--	--	**-0.087** (-1.455) [0.146]	--	--
EDMOTHER	--	--	--	**-0.297** (-2.220) [0.026]	--	--
FAMINCOM	--	--	--	**-0.026** (-2.232) [0.026]	--	--
CIGTAX88	--	--	--	**0.030** (1.400) [0.162]	--	--
Box-Cox Parameter $\hat{\eta}$	--	--	--	**1.402** (4.717) [<.001]	--	--
Sum of Squared Residuals	**2127.312**	**2185.027**	**2120.781**	--	**2244.577**	**2131.658**

*t-statistics in parentheses and p-values in square brackets.