Parametric Regression and Health Policy Analysis: Survey-Based

Estimation and Inference in the Presence of Endogeneity

by

Joseph V. Terza*
Department of Epidemiology and Health Policy Research
and Department of Economics
University of Florida
Gainesville, FL 32610

(March, 2008)

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. Such policy analytic studies typically use nonexperimental (survey) data and focus 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 decision-making entity. The goal is the estimation of the effect that a prospective exogenous change in x_p would have on a targeted outcome of interest (y) ó the *policy effect*. In the present paper we propose a pseudo-difference-of-means (PDOM) framework for policy effect estimation. The policy effect typically takes one of three forms (incremental, marginal, or treatment) depending on the nature of the policy variable (discrete, continuous, or binary), and on the type of change in x_p under policy consideration. The full development of the PDOM estimator in each of its three forms is presented -- from its intuitive conceptual foundations to the details of its practical implementation. In the design of the PDOM estimator, particular attention is paid to accommodating the potential endogeneity of the policy variable. The asymptotic properties of the PDOM are derived after showing that it can be cast as a special case of a general class of estimators that have been thoroughly studied in the econometrics literature. Three empirical applications of the PDOM are discussed -- one for each of the alternative forms of the PDOM (incremental, marginal, treatment). In each of these examples the potential endogeneity of the policy variable was tested and found to be both statistically and substantively significant.

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. Typically, the 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. The key policy analytic objective is estimation of the effect that a prospective exogenous change x_p would have on a specified outcome of interest (y) [the *policy effect (PE)*]. Consider estimation of the effect that a change in the policy variable would, on average, have on the outcome. In the ideal, to ensure that the results would be causally interpretable, the data for such a study would come from an experiment in which values of x_p can be mandated for each sample member. Formally, for the case in which x_p is not binary (i.e. it is count-valued or continuous) and the prospective discrete policy change (a change in x_p from x_{p1} to x_{p2}) is known, the estimation objective is the following incremental or õarcö policy effect

$$PE_{ARC}(\ x_p) = \frac{E[y_{x_{p2}} - y_{x_{p1}}]}{x_p} \tag{1}$$

where $x_p = x_{p2}$ ó x_{p1}, and $y_{x_p^*}$ is the random variable representing the value of the outcome for the case in which the value of the policy variable is mandated to be x_p^* for everyone in the relevant population. If x_p is continuous and the specific prospective policy change is not defined (i.e., x_{p1} and x_{p2} have not been specified) the relevant version of (1) is the following marginal (partial) policy effect

2

$$PE_{MARG} = \lim_{x_p \to 0} PE_{ARC}(\ x_p) = \frac{\partial E[y_{x_p}]}{\partial x_p}. \tag{2}$$

Finally, if x_p is binary [i.e., $x_{p1}=0$ and $x_{p2}=1$] (1) becomes the following treatment effect measure

$$PE_{TE} = E[y_1 - y_0]. \tag{3}$$

This paper focuses on estimation of such policy effects.

In the next section, we detail the conceptual background and formal development of a unifying and consistent pseudo difference-of-means (PDOM) estimation framework for (1), (2), and (3) when: a) only survey data is available for estimation; b) a parametric conditional mean regression form can be used to represent $E[y_{x_p}]$; and c) survey observations on x_p are sampled endogenously. Because (2) and (3) are special cases of (1), most of the introductory exposition is cast in the arc policy effect context. In section 3, we show that the PDOM method can be represented as a special case of a general class of estimators whose statistical properties are well-established in the econometrics literature ó two-stage optimization estimators. The relevant asymptotic inferential statistics for hypothesis testing and confidence interval estimation are, therefore, easily derived from the general theory for this class of estimators. In section 4, to illustrate the implementation of the PDOM method and corresponding asymptotic inferential statistics, we discuss empirical applications of the estimator. Three examples are detailed ó one for each of the relevant forms of the PDOM as it pertains to estimation of arc (incremental), marginal, and treatment effects ó expression (1), (2) and (3), respectively. The final section of the paper summarizes and concludes.

1

2. Policy Effects, Endogeneity and the Pseudo-Difference of Means Estimator

We begin by defining some important terms and concepts. We place the discussion in the context of the arc policy effect defined in (1) [henceforth the *policy effect*] but all of the concepts discussed are easily extended to marginal and treatment effects as given in (2) and (3). *The policy* is defined as a *mandated* change in the value of the policy variable (x_p) from x_{p1} to x_{p2}. *The policy effect (PE)*, as defined in (1) measures the per unit of x_p expected amount by which $y_{x_p^*}$ will change as a result of the policy, recalling that $y_{x_p^*}$ is the random variable representing the potential outcomes under the mandated condition that $x_p = x_p^*$ for everyone in the population. As such, $y_{x_p^*}$ is *counterfactual* from the perspective of survey data in that its value can only be observed for individuals whose sampled values of x_p are equal to x_p^*. For example, let

     y the number of yearly visits to the physician

     x_p the per visit copay

and suppose a change in the copay from \$15 to \$25 is proposed. Consider the hypothetical population of 6 individuals shown in Figure 1.

Figure 1

Person	y_{15} (x_p^* = \$15)	y_{25} (x_p^* = \$25)
1	3	3
2	5	4
3	2	1
4	3	3
5	1	3
6	3	2

The policy effect as defined in (1) in this case is

$$PE = \frac{E[y_{25} - y_{15}]}{10} = \frac{2.667 - 2.833}{10} = -.0166 .\qquad(4)$$

Expression (4) measures the change in the expected number of doctor visits per dollar change in the copay. If we could sample experimentally, each sample members' behavior (manifested as the number of times they visit the physician - $y_{x_p^*}$) could be observed for both of the relevant values of x_p ($x_{p1} = \$15$ and $x_{p2} = \$25$, respectively) and a causally interpretable estimate of (4) would be obtained using the following simple difference-of-means (DOM)

$$\frac{1}{10}\left(\frac{\sum_{i=1}^{n} y_{25(i)}}{n} - \frac{\sum_{i=1}^{n} y_{15(i)}}{n} \right)\qquad(5)$$

where $y_{x_p^*(i)}$ denotes the experimental value of the outcome for the ith sampled individual (i = 1, ..., n) under the mandated policy value x_p^* (= \$15 or \$25), and n is the sample size. Unfortunately, in the survey sampling context y_{15} and y_{25} are counterfactual in the sense that neither is fully observable in the population -- i.e. some individuals will be observed with $x_p = 15$, others with $x_p = 25$, and still others with x_p equal to neither \$15 nor \$25. In our example, suppose that only individuals 1 and 2 actually have a copay of \$15 (i.e., the observed value of x_p for persons 1 and 2 is 15). For these individuals, y_{15} is observable [in the form of (y | $x_p = 15$)] but y_{25} is not -- i.e. for individuals 1 and 2, $y_{x_p^*}$ is a counterfactual entity when $x_p^* \neq 15$. An analogous assessment holds for individuals 3 and 4. For these two individuals y_{25} is observable as (y | $x_p = 25$) but the random variable $y_{x_p^*}$ is a counterfactual entity for them when

3

$x_p^* \neq 25$. Finally, for individuals 5 and 6, $y_{x_p^*}$ is counterfactual when either $x_p^* = 15$ or $x_p^* = 25$.

Therefore, the DOM given in (5) is not feasible. Moreover, the following version of the DOM, based on the observable data, will be biased

$$\frac{1}{10}\left(\frac{\sum_{i=1}^{n_{25}}(y_i \mid x_{pi} = 25)}{n_{25}} - \frac{\sum_{i=1}^{n_{15}}(y_i \mid x_{pi} = 15)}{n_{15}} \right) \qquad (6)$$

where $(y_i \mid x_{pi} = x_p^*)$ denotes the observed value of the outcome conditional on the fact that the

ith individual was sampled from the subpopulation for whom $x_p = x_p^*$, and $n_{x_p^*}$ denote the size of

the corresponding subsample. To see this, consider the version of Figure 1 shown in Figure 2

Figure 2

Person	y_{15} (x_p^* = \$15)	y_{25} (x_p^* = \$25)
(observable x_p = \$15) 1 2	**A (observable y at x_p = \$15)** 3 5	**D** 3 4
(observable x_p = \$25) 3 4	**B** 2 3	**E (observable y at x_p = \$25)** 1 3
(\$15 ≠ observable x_p ≠ \$25) 5 6	**C** 1 3	**F** 3 2

Note that we distinguish observed values of the policy variable from mandated values using a

"*" superscript -- x_p denotes observed values, x_p^* denotes mandated values. The population data

4

on the outcome y for individuals whose observed copay under survey sampling would be $15 (i.e. persons 1 and 2) are given in cells A and D of Figure 2. In cell A are given the observable (factual) values of y for these individuals. The counterfactual (nonobservable) values of y as they would be for these individuals if the copay were mandated to be $25 are given in block D. Likewise the population data for those who actually have a $25 copay (i.e. persons 3 and 4) are given in cells B and E. The latter displaying the values of the outcome (doctor visits) that are observable via survey sampling. The former containing the values of y pertaining to the counterfactual scenario in which a $15 copay is mandated. Cells C and F hold the potential doctor visit outcomes for those individuals whose copays are neither $15 nor $25. These cells correspond to the counterfactual scenarios in which the manadated value of x_p is either $15 or $25, respectively. The population data in both of these cells is, of course, counterfactual. The counterfactual cells of the population distributions of the random variables y_{15} and y_{25} are indicated by shading. The observable cells of the population distributions of the counterfactual random variables y_{15} and y_{25} are unshaded (A and E, respectively). If we simply conduct a survey (i.e. draw a sample from blocks A and E) and apply the version of the DOM estimator given in (6), the result will be biased for (4) because (6) is unbiased for[1]

$$\frac{E[y \mid x_p = 25] - E[y \mid x_p = 15]}{10} = \frac{2-4}{10} = -.2 \qquad (7)$$

which greatly overstates (in absolute value) the true policy effect given in (4) [-.0166].

Why do (4) and (7) differ? They differ because of the existence of variables that influence the value of y and are correlated with x_p ó so-called *confounders*. For example,

---

[1] $E[y \mid x_p=25]$ and $E[y \mid x_p=15]$ are obtained as the average values within blocks A and E, respectively.

suppose sicker individuals, who are apt to visit a physician, are less likely to be observed with lower copays due to individual insurance plan choice. In this case, health status is a confounder that will cause the (7) to overstate (in absolute value) the true (albeit counterfactual) policy effect (4). Specifically, due to the health status confounder, on average, individuals observed with low copays (the sicker individuals in cell A) visit the doctor more ($E[y \mid x_p = 15] = 4$) than would be manifested if all individuals in the population faced a \$15 copay ($E[y_{15}] = 2.833$ -- the average value in the second full column of Figure 2). By a similar argument, those observed with high copays (the healthier individuals in block E) visit the doctor less ($E[y \mid x_p = 25] = 2$) than would be observed if all individuals in the population faced a \$25 copay ($E[y_{25}] = 2.667$ -- the average value in the third full column of Figure 2). To summarize, in general, (1) and[2]

$$\frac{E[y \mid x_p = x_{p2}] - E[y \mid x_p = x_{p1}]}{x_p}$$

are likely to differ because

$$E[y_{x_p*}] \neq E[y \mid x_p = x_p*]. \tag{8}$$

Moreover, (8) is a direct consequence of the existence of confounders and, as we have seen, is the reason for the biasedness of the simple difference of means estimator given in (6) which is based on observable (factual) data only.

We seek a version of (6) that requires only observable data but is unbiased. What is needed here is a way to express the counterfactual expectation on the left-hand side of (8) [i.e., the full-column average in Figure 2] in terms of the expected value of an aspect of the

---

[2] The non-equality in (8) can be characterized in the second and third columns of Figure 2 as the likely difference between the average value in the unshaded block and the full column average.

distribution of the outcome that is observable (factual) – akin to the right-hand side of (8) [i.e., the average of y in the unshaded block in a particular column]. With this in mind, by the iterated expectations rule (Wooldridge, 2003, pp. 18-22), we note that for any confounder v we can rewrite the left-hand side of (8) as

$$E[y_{x_p*}] = E\left[E[y_{x_p*} \mid v]\right]. \tag{9}$$

This prompts us to consider the relationship between $E[y_{x_p*} \mid v]$ and $E[y \mid x_p = x_p^*, v]$. The former is the key counterfactual component of the right-hand side of (9), and the latter is accessible (estimable) via observable (factual) data. To explore this relationship, let us return to the example. Suppose the individuals in the population characterized in Figures 1 and 2 are either õillö or õwellö and that everyone is well except persons #2 and #5. If we condition the on only the well subpopulation (i.e. if we drop persons #2 and #5), we obtain Figure 3.

Figure 3
Conditioned on Health Status (only those who are well)

Person	y_{15} (x_p^* = \$15, well)	y_{25} (x_p^* = \$25, well)
(observable x_p = \$15, well) **1**	**A (observable y at** **x_p = \$15, well)** 3	**D** 3
(observable x_p = \$25, well) **3** **4**	**B** 2 3	**E (observable y at** **x_p = \$25, well)** 1 3
(\$15 ≠ observable x_p ≠ \$25, well) **6**	**C** 3	**F** 2

From Figure 3 we find that the discrepancy between

7

$$\frac{E[y_{25} \mid \text{well}] - E[y_{15} \mid \text{well}]}{10} = \frac{2.25 - 2.75}{10} = -.05$$

and

$$\frac{E[y \mid x_p = 25, \text{well}] - E[y \mid x_p = 15, \text{well}]}{10} = \frac{2-3}{10} = -.1$$

[viz., .05] is smaller than the difference between (4) and (7) [viz., .183]. This exercise illustrates how conditioning on confounders can serve to bring $E[y_{x_p*} \mid v]$ and $E[y \mid x_p = x_p^*, v]$ closer to equality. Indeed, if such conditioning is *comprehensive*, in the sense that all possible confounders are included in the analysis, we obtain

$$E[y_{x_p*} \mid x] = E[y \mid x_p = x_p^*, x] \tag{10}$$

where x denotes the comprehensive vector of confounders.

In the remainder of this section, we develop a consistent two-stage policy effect estimator based on (10) that is designed to accommodate cases in which the survey data on the policy variable of interest is *endogenously* sampled ó i.e. cases in which the comprehensive vector of confounders includes both observable (x_o) and unobservable (x_u) components. We begin by assuming that $y_{x_p^*}$ (the counterfactual outcome at the fixed mandated value x_p^*) follows a parametric random process of the form

$$y_{x_p^*} = H(x_p^*, x_o, x_u, \ , \ ) \tag{11}$$

where H( ) is a known (possibly nonlinear) function, is the random error term, and is a vector of unknown parameters. Moreover, is defined such that

8

$$E[y_{x_p^*} \mid x_o, x_u] = M(x_p^*, x_o, x_u, \;) = \int H(x_p^*, x_o, x_u, \;, \;) \, f(\;\mid x_o, x_u) \, d \tag{12}$$

where $f(\;\mid x_o, x_u)$ denotes the conditional pdf of given x_o and x_u. The form of $M(\;)$ will often be similar to that of $H(\;)$. For example, suppose $H(\;)$ is linear and $E[\;\mid x_o, x_u] = 0$ so that (11) becomes

$$y_{x_p^*} = x_p^* {}_p + x_o {}_o + x_u {}_u + \tag{13}$$

and (12) yields

$$M(x_p^*, x_o, x_u, \;) = x_p^* {}_p + x_o {}_o + x_u {}_u \tag{14}$$

where $= [{}_p \quad {}_o \quad {}_u]$ is the vector of unknown parameters. For count-valued and other nonnegative outcomes one might assume that

$$y_{x_p^*} = \exp(x_p^* {}_p + x_o {}_o + x_u {}_u + \;) \tag{15}$$

where $E[\exp(\;) \mid x_o, x_u] = 1$. In this case, (12) yields

$$M(x_p^*, x_o, x_u, \;) = \exp(x_p^* {}_p + x_o {}_o + x_u {}_u) \tag{16}$$

where is defined as in (14). As a final illustration, consider the case in which (11) is of the form

$$y_{x_p^*} = I(x_p^* {}_p + x_o {}_o + x_u {}_u + \; > 0) \tag{17}$$

9

where $(\ | x_o, x_u)$ is standard normal distributed, and I(A) denotes the indicator function that takes

on the value if 1 of condition A is true, and 0 otherwise. From (12) we obtain

$$M(x_p^*, x_o, x_u,) = (x_p^*{}_p + x_o{}_o + x_u{}_u).$$
(18)

where $(\ )$ denotes the standard normal cumulative distribution function and   is defined as in

(14).

Using (9) and (12) we can now rewrite (1), (2), and (3) respectively as[3]

$$PE_{ARC}(\ x_p) = \frac{1}{x_p} E[M(x_{p2}, x_o, x_u,) - M(x_{p1}, x_o, x_u,)]$$
(19)

$$PE_{MARG} = \frac{\partial E[M(x_p, x_o, x_u,)]}{\partial x_p} = E\left[\frac{\partial M(x_p, x_o, x_u,)}{\partial x_p}\right]$$
(20)

and

$$PE_{TE} = E[M(1, x_o, x_u,) - M(0, x_o, x_u,)].$$
(21)

Consistent estimation of each of (19) through (21) requires a consistent estimate of . Terza et

al. (2008) discuss a two-stage residual inclusion (2SRI) method that can be used in all three

cases, and Terza (2008) suggests full information maximum likelihood (FIML) and two-stage

method of moments (TSM) approaches that are appropriate only in the context of (21).[4] Both

2SRI and TSM are based on following regression specification which is implied by the

comprehensiveness of $x = [x_o \quad x_u]$, and equations (10) and (12)

---

[3] The second equality in (20) holds under fairly general conditions (see Bierens, 1994, p. 25).
[4] The TSM methods discussed in Terza (2008) are applicable to a broader class of models that encompasses those discussed here and may offer efficiency gains relative to 2SRI, although this claim has not been formally validated.

$$E[y \mid x_p, x_o, x_u] = M(x_p, x_o, x_u, \ ). \tag{22}$$

For 2SRI estimation, we note that (22) implies the following sampling model

$$y = M(x_p, x_o, x_u, \ ) + e \tag{23}$$

where $e = y - M(x_p, x_o, x_u, \ )$ is the regression error term. In addition we assume the existence of the following auxiliary equation

$$x_p = r(w, \ ) + x_u \tag{24}$$

where r is a known (possibly nonlinear) function, $w = [x_o \quad w^+]$, $E[x_u \mid w] = 0$, is a vector of unknown parameters, and w^+ denotes a vector of observable *identifying instrumental variables* that are correlated with x_p and are neither included in x_o nor correlated with e. Under these assumptions, the following two-stage (2SRI) estimator is consistent for :[5]

First Stage

Consistently estimate by applying the nonlinear least squares (NLS) method to (24).

Second Stage

Estimate by applying NLS to the following version of (23)

$$y_i = M(x_{pi}, x_{oi}, \ddot{x}_{ui}, \ ) + e_i \tag{25}$$

where $\ddot{x}_{ui} = x_{pi} - r(w_i, \text{``})$, `` denotes the first-stage estimate of , i = 1, ..., n indicates the ith sample member, and n is the sample size.

_____

[5] See Terza, Basu, and Rathouz (2008) for details.

Given the 2SRI estimates, consistent estimators of (19) and (20) can be obtained from the following pseudo difference-of-means (PDOM)

$$\widehat{PE}_{ARC}(x_p) = \frac{1}{x_p} \sum_{i=1}^{n} \frac{1}{n} \left\{ M(x_{p2}, x_{oi}, \ddot{x}_{ui}, \text{''}) - M(x_{p1}, x_o, \ddot{x}_{ui}, \text{''}) \right\} \tag{26}$$

$$\widehat{PE}_{MARG} = \sum_{i=1}^{n} \frac{1}{n} \frac{\partial M(x_{pi}, x_{oi}, \ddot{x}_{ui}, \text{''})}{\partial x_p} \tag{27}$$

where `` denotes the 2SRI estimate of .

For the special case in which x_p is binary, Terza (2008) considers models in which (24) is replaced by

$$x_p = I(w + x_u > 0) \tag{28}$$

and the distribution of $(x_u \mid w)$ is known. For example, we may assume that $(x_u \mid w)$ is standard normal distributed so that (28) defines a conventional probit model. Based on this assumption and (22), the TSM estimator suggested by Terza (2008) yields a consistent estimate of , call it $\tilde{}$.[6] Therefore, (21) can be consistently estimated using the following PDOM

$$\widetilde{PE}_{TE} = \sum_{i=1}^{n} \frac{1}{n} \left\{ \int_{x_u} \left[M(1, x_{oi}, x_u, \tilde{}) - M(0, x_{oi}, x_u, \tilde{}) \right] g(x_u \mid w) dx_u \right\} \tag{29}$$

where $g(x_u \mid w)$ denotes the conditional probability density function of x_u given w. If, in addition to (12) and (22), we have that

---

[6] See Terza (2008) for details.

$$f(y_{x_p^*} \mid x_o, x_u) = f(y \mid x_p^*, x_o, x_u) = g(y, x_p^*, x_o, x_u, \ )$$

where g( ) is a known probability density function, then Terza (2008) shows that a FIML method can be used to consistently estimate .[7] Applications of the TSM and FIML methods suggested by Terza (2008) can be found in Coulson et al. (1995), Kenkel and Terza (2001), Koc (2005), McGeary and French (2000), Neslusan et al. (1999), Pryor and Terza (2002), Terza (1994ª, 1994b,1998, 1999, 2002), Terza, Kenkel, et al. (2008), and Treglia et al. (1999).

In the special case in which x_o is comprehensive (i.e. there are no unobservable confounders), (26), (27) and (29) reduce to the estimators considered by Wooldridge (2003), Basu and Rathouz (2005) and many others. These estimators are, of course, not consistent when x_p is endogenously sampled.

3. The PDOM as a Two-Stage Optimization Estimator: Asymptotic Inference

In the previous section, we developed versions of the PDOM estimator [(26), (27) and (29)] for each of the three versions of the policy effect defined in (1), (2) and (3). We now turn to the asymptotic properties of these estimators. Of particular interest here is the derivation of the correct asymptotic standard errors of these estimators. To summarize the results of the previous section, if we have a consistent estimator for = [ø ø]ø (say ̄), then we can consistently estimate the relevant policy effect using

$$\overline{PE} = \sum_{i=1}^{n} \frac{\overline{pe_i}}{n} \tag{30}$$

where

---

[7] See the last illustrative example in section 4.

$$\overline{pe}_i = \begin{cases} pe(x_{p1}, x_{p2}, w_i, \overline{}) & \text{for (26) and (29)} \\ \quad\quad\text{or} \\ pe(x_{pi}, w_i, \overline{}) & \text{for (27)} \end{cases}$$

$$pe(x_{p1}, x_{p2}, w_i,) = \begin{cases} \dfrac{1}{x_p}\left\{ M(x_{p2}, x_o, x_u,) - M(x_{p1}, x_o, x_u,) \right\} & \text{for (26)} \\ \quad\quad\text{or} \\ \displaystyle\int_{x_u} \left[M(1, x_o, x_u,) - M(0, x_o, x_u,) \right] g(x_u \mid w)\, dx_u & \text{for (29)} \end{cases}$$

and $\quad pe(x_p, w,) = \dfrac{\partial M(x_p, x_o, x_u,)}{\partial x_p}.$ $\quad\quad\quad$ for (27)

In order to derive the asymptotic properties of (30) we can cast it as a special case of the class of two-stage optimization (2SOPT) estimators discussed by White (1994), Newey and McFadden (1994), and Wooldridge (2003).[8] 2SOPT estimators are characterized by two objective functions: $Q()$, a full information objective function whose optimizer is a consistent estimator of all parameters of the model;[9] and $Q_1()$, a first stage objective function whose optimizer is a consistent estimator of a subvector of the full set of parameters of interest. In the 2SOPT protocol: Q_1 is optimized to obtain an estimate of the relevant subvector of parameters, then Q is optimized with the first-stage parameters fixed at their estimated values. The estimator of the generic policy effect (PE) in (30) can be represented as a 2SOPT estimator by specifying the full information objective function as

$$Q(, PE) = \sum_{i=1}^{n} q(, PE, u_i) \quad\quad\quad\quad (31)$$

---

[8] These authors extend the results Murphy and Topel (1985) for two-stage maximum likelihood estimators to the more general class of two-stage optimization estimators.

[9] Here we use the term "full information" to indicate that $Q()$ takes account of all of the available nonsample information. This does not imply that full information maximum likelihood estimation is possible.

where and $pe(x_{pi}, w_i, \ )$ are defined as in (30)

$$q(\ , PE, u_i) = q_1(\ , u_i) - \left(pe(x_{pi}, w_i, \ ) - PE\right)^2 \tag{32}$$

$u_i = [y_i \quad x_{pi}, \ w_i]$ and $Q_1(\ ) = \sum_{i=1}^{n} q_1(\ , u_i)$ denotes the appropriate first-stage objective function

(2SRI, TSM or FIML) for consistent estimation of . The first stage of the 2SOPT protocol in

this case yields an estimate of (2SRI, TSM, or FIML) as discussed in the previous section. In

the second stage an estimate of PE is obtained by optimizing $Q(\bar{\ }, PE)$ where $\bar{\ }$ denotes the first-

stage estimate of . Given the specification of Q in (31), this second stage is tantamount to

optimizing

$$-\sum_{i=1}^{n} \left(pe(x_{pi}, w_i, \bar{\ }) - PE\right)^2 \tag{33}$$

with respect to PE. It is, however, easy to show that the optimizer of (33) is (30). Therefore, our

2SOPT characterization of (30) is valid.

Because (30) can be cast as a special case of the generic 2SOPT estimator, its asymptotic

properties can be easily derived from the general theory. First we define some notational

conventions. For a scalar function õsö of two vector arguments j and t (i.e. s = s(j, t) where s is a

scalar and j and t are vectors) we define:

$$\nabla_j s = \frac{\partial s}{\partial j} = \text{the gradient vector of s with respect to the elements of j}$$

and

$$\nabla_{jt} s = \frac{\partial^2 s}{\partial j \partial t} = \text{the matrix of cross-partial derivatives of s with respect to the elements of j}$$

15

and t.

We also assume that the former is a row vector, and the latter is a matrix with row dimension equal to that of the first subscript on ∇, and column dimension equal to that of the second subscript. Under the regularity conditions given in Theorem 6.11 of White (1994), (30) is consistent and

$$\sqrt{\frac{n}{a\,var\left(\overline{\overline{PE}}\right)}}(\overline{\overline{PE}} - PE) \xrightarrow{\ d\ } n(0,1) \tag{34}$$

where $\overline{\overline{PE}}$ is defined in (30), $a\,var\left(\overline{\overline{PE}}\right)$ is the asymptotic variance of $\overline{\overline{PE}}$, $\xrightarrow{\ d\ }$ denotes convergence in distribution, n(0, 1) represents the standard normal variate. In Appendix A we show that

$$a\,var\left(\overline{\overline{PE}}\right) = E[\nabla\ pe]AVAR(\bar{\ })E[\nabla\ pe]' + E\left[(pe - PE)^2\right]. \tag{35}$$

Expression (35) directly applies in the context of 2SRI. Estimates of   are not, however, involved in the in the TSM version of (30). Therefore the relevant form of (35) in this case is

$$a\,var\left(\overline{\overline{PE}}\right) = E[\nabla\ pe]AVAR(\tilde{\ })E[\nabla\ pe]' + E\left[(pe - PE)^2\right]. \tag{36}$$

The asymptotic variance given in (35) can be consistently estimated using

$$\overline{a\,var}\left(\overline{PE}\right) = \left(\frac{\sum_{i=1}^{n}\nabla\ \overline{pe}_i}{n}\right)\left(n\ \overline{AVAR}(\overline{\ })\right)\left(\frac{\sum_{i=1}^{n}\nabla\ \overline{pe}_i}{n}\right)' + \left(\frac{\sum_{i=1}^{n}(\overline{pe}_i - \overline{PE})^2}{n}\right) \quad (37)$$

where $\nabla\ \overline{pe}_i$ denotes ∇ pe evaluated at x_{pi}, w_i and $\overline{\ }$; and $\overline{AVAR}(\overline{\ })$ is the estimated asymptotic

covariance matrix of $\overline{\ }$. It follows from (37) that

$$\sqrt{\frac{n}{\overline{a\,var}\left(\overline{PE}\right)}}(\overline{PE} - PE) \xrightarrow{\ d\ } n(0,1). \quad (38)$$

The asymptotic standard error derived by Basu and Rathouz (2005) for the version of (30) in

which there are no unobservable confounders is easily seen to be a special case of (37).

4. Examples

The following three examples illustrate estimation of each of the three different policy

effect measures given in (1), (2) and (3). They are, respectively: the incremental or õarcö policy

effect -- relevant in cases in which the policy variable is not binary and the details of the

prospective policy (x_{p1} and x_{p2}) are known; the marginal policy effect ó relevant when the policy

variable is continuous and the prospective policy change is not defined; and treatment effects

relevant in cases in which the policy variable is binary.

4.1 The Arc Policy Effect of Habit Stock on Smoking

Terza, Bradford and Dismuke (2007) re-estimated Mullahyøs (1997) model of cigarette

consumption (y) focusing on the effect of *habit stock* (x_p). Habit stock is a measure of the

accumulated effects of past smoking on present consumption. Mullahy's model has implications for the potential effectiveness of smoking cessation policies. Only policies that can sustain smoking cessation long enough to drive habit stock to zero are likely to be long-term effective. Recent research by Volpp et al. (2006) suggests, for instance, that individuals may be induced to stop smoking for 75 days with cash payments of $200 per person. Such modest payments mean that encouraging cessation for periods approaching those needed to reduce the habit stock to zero would be feasible for many employers and health systems ó and potentially highly cost effective.

The Centers for Disease Control and Prevention (CDC) reports that one of its national objectives for 2010 is to decrease the prevalence of cigarette smoking among adults from the current 20% level to less than 12% (CDC, 2006). Consider accomplishing this goal through effective anti-smoking and cessation programs that prevent smoking initiation and forestall relapse for periods sufficient to drive enough individuals' habit stocks to zero. In this context the key question is... What level of exogenous (policy driven) across-the-board habit stock depreciation would likely lead to the targeted prevalence rate? In theory, the answer can be obtained by solving the following version of the generic arc policy effect given in (1) for $x_{p12\%}$

$$\text{population } 88^{\text{th}} \text{ \%-tile of daily smoking} = E[y_0 - y_{x_{p12\%}}] \qquad (39)$$

Here x_{p2} [as defined in (1)] is set equal zero by definition because we are considering here only those policies that are *effective* ó i.e., only those that sustain abstinence long enough for individual's habit stocks to depreciate to zero.[10] As a practical matter, we can find the value of $x_{p12\%}$ at which the following version of (26) is satisfied

---

[10] In expression (40) we forego division of both sides by x_p. It is superfluous in this context.

$$\text{sample } 88^{\text{th}} \text{ \%-tile of daily smoking } = \sum_{i=1}^{n} \frac{1}{n} \left\{ M(0, x_{oi}, \ddot{x}_{ui}, \text{``}) - M(x_{p12\%}, x_{oi}, \ddot{x}_{ui}, \text{``}) \right\}$$

(40)

To define M( ) and , and to facilitate regression estimation, we assume [as did Terza, Bradford and Dismuke (2007)] that

$$y_{x_p^*} = H(x_p^*, x_o, x_u, \ , \ ) = \begin{cases} \left(\left(\frac{}{2} \left(x_p^* {}_p + x_o {}_o + x_u {}_u \right) + 1 \right)^2 \right)^{\frac{1}{}} \exp(\ ) & \text{if} \quad \neq 0 \\[2em] \exp\left(x_p^* {}_p + x_o {}_o + x_u {}_u + \ \right) & \text{if} \quad = 0 \end{cases}$$

(41)

where $= [{}_p \quad {}_o \quad {}_u, \ ]$, is a scalar parameter with unrestricted in range; and $E[\exp(\ ) \mid x_o, x_u] = 1$. This a variant of the inverse of the flexible functional form suggested by Box and Cox (1964). The inverse Box-Cox (IBC) conditional mean regression specification was first considered and implemented by Wooldridge (1992). The IBC functional form approaches the exponential model as 0. When $= 2$ and $x_p {}_p + x_o {}_o + x_u {}_u > -1$, it reduces to a simple linear regression model.[11] Using (12) we obtain

$$E[y_{x_p^*} \mid x_o, x_u] = M(x_p^*, x_o, x_u, \ ) = \begin{cases} \left(\left(\frac{}{2} \left(x_p^* {}_p + x_o {}_o + x_u {}_u \right) + 1 \right)^2 \right)^{\frac{1}{}} & \text{if} \quad \neq 0 \\[2em] \exp\left(x_p^* {}_p + x_o {}_o + x_u {}_u \right) & \text{if} \quad = 0 \end{cases}$$

(42)

---

[11] When $= 2$, equation (1) becomes $E[y \mid x_p, x_o, x_u] = g(z) = |z+1|$, where |a| denotes the absolute value of a and $z = x_p {}_p + x_o {}_o + x_u {}_u$. 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 equation (1) becomes the simple linear regression model.

19

The model includes x_u to allow for the potential endogeneity of the habit stock variable x_p. Habit stock may be endogenous because it is determined by past smoking, and the unobservable determinants of past smoking are likely to be correlated with present smoking. Terza, Bradford and Dismuke (2007) estimated using Mullahyøs data and a version of the 2SRI method detailed in section 2 [equations (23) ó (25)]. In the first stage they estimated the linear version of (24) [$x_p = w + x_u$] via OLS. Then, in the second stage, NLS was applied to the following version of the sampling model based on (42)

$$y_i = \left(\left(\frac{1}{2}\left(x_{pi}\,_p + x_{oi}\,_o + \ddot{x}_{ui}\,_u\right)+1\right)^2\right)^{\frac{1}{2}} + v_i \tag{43}$$

where $\ddot{x}_{ui} = x_{p1} - w$ " is the first-stage OLS residual, " is the first-stage OLS estimate of , v_i is the regression error term, $i = 1, ..., n$ denotes the ith sample member, and n is the sample size.[12] The definitions of all of the variables included in the analysis are given in Table 1, and Table 2 shows the descriptive statistics of the sample. As a basis for comparison, we applied NLS to the variant of (43) with $_u$ set equal to zero (no unobservable confounders ó i.e. x_p assumed exogenous). The results are displayed in the first three columns of Table 3. Columns 4 through 6 contain the second-stage 2SRI results obtained by Terza, Bradford and Dismuke (2007) for equation (43), and the 2SRI first stage results are shown in the remaining three columns. The null hypothesis that x_p is exogenous can be tested based on the coefficient of the first stage residual [H_o: $_u = 0$]. As can be seen in the fifth column of Table 3, the t-stat for that coefficient is -3.01 which leads to rejection of the exogeneity null at less than a .01 significance level.

---

[12] As can be seen in the fourth column of Table 3, the estimated value of is 1.77 which is significantly different from zero (t = 28.22). Therefore, the exponential version of (39) is not relevant.

The 88th %-tile daily smoking rate in the sample is 1.25 (25 cigarettes per day). Combining the 2SRI-IBC estimates in the fourth column of Table 3 with (40) and (43) we find that the target level of habit stock ($x_{p12\%}$) is 270 [accurate to the third decimal ó i.e. the right-hand side of (40) at x_{p1}=270 is equal to -1.257]. Placing the discussion in the notation of section 3, when $x_{p12\%}$ =270, the right-hand side of (40) can be rewritten as

$$\widehat{PE}_{ARC} = \sum_{i=1}^{n} \frac{\widehat{pe}_i}{n} = 1.257 \qquad (44)$$

where

$$\widehat{pe}_i = pe(270, 0, w_i, \text{``})$$

$$pe(270, 0, w_i,) = M(0, x_{oi}, (x_{pi} - w_i),) - M(270, x_{oi}, (x_{pi} - w_i),)$$

and `"` denotes the 2SRI estimate of = [   ]. Assuming a habit stock depreciation rate of 10% per day [the rate used by Mullahy (1979)], this would mean that prevention and cessation policies that keep people from smoking for at least 120 days would be required to achieve the CDC 2010 12% smoking prevalence goal. Asymptotic inference for (44) can be drawn from the following version of (38)

$$\sqrt{\frac{n}{\widehat{a \, var}(\widehat{PE}_{ARC})}} (\widehat{PE}_{ARC} - PE_{ARC}) \xrightarrow{\text{d}} n(0,1) \qquad (45)$$

where

$$PE = E[y_0 - y_{270}]$$

$$\widehat{a \, var}(\widehat{PE}_{ARC}) = \left(\frac{\sum_{i=1}^{n} \nabla \, \widehat{pe}_i}{n}\right) \left(n \, \widehat{ACOV}(\text{``})\right) \left(\frac{\sum_{i=1}^{n} \nabla \, \widehat{pe}_i}{n}\right)' + \left(\frac{\sum_{i=1}^{n} \left(\widehat{pe}_i - \widehat{PE}_{ARC}\right)^2}{n}\right) \quad (46)$$

$\widehat{ACOV}(\text{``})$ is the asymptotic covariance matrix of the 2SRI estimator of

$$\nabla \, \widehat{pe}_i = [\nabla \, \widehat{pe}_i \quad \nabla \, \widehat{pe}_i \quad \nabla \, \widehat{pe}_i]$$

and the details of $\nabla \, \widehat{pe}_i$, $\nabla \, \widehat{pe}_i$ and $\nabla \, \widehat{pe}_i$ are given in Appendix B. To test the conventional null hypothesis for PE_{ARC} [i.e., H_o: $PE_{ARC} = 0$] we computed the relevant asymptotic t-stat [based on (45)] and obtained

$$\frac{\widehat{PE}_{ARC}}{\sqrt{\dfrac{\widehat{a \, var}(\widehat{PE}_{ARC})}{n}}} = \frac{-1.257}{.229} = -5.49 \, .$$

The conventional null is rejected at any reasonable significance level.

For the purpose of comparison, we conducted a similar analysis using the NLS-IBC results, which are not corrected for the potential endogeneity of the habit stock variable (shown in column 1 of Table 3). Using these uncorrected estimates, we found that in order to bring smoking prevalence down to the CDC 2010 target, smoking prevention and cessation policy would have to be effective enough to reduce habit stock to zero from a level of 324. This amounts to a 20% bias relative to the endogeneity corrected estimate (270). Full depreciation of a habit stock equal to 324 would require sustained abstinence for a minimum of 140 days.

4.2 The Marginal Effect of Prescription Drug Use on In-Patient Hospital Expenditure

Stuart et al. (2008) estimated a two-part model of the effect of prescription drug (Rx) utilization on inpatient hospital expenditure. They implemented the 2SRI method (Terza et al., 2008) to accommodate the potential endogeneity of Rx use. In the first part of the model (the *hurdle* component), a probit-type specification, as defined in equations (17) and (18), was used to regress a binary indicator of any in-patient hospitalization during the year on Rx utilization and confounder controls (both observable and unobservable). In this illustrative example, we focus on the second part of the model (the *levels* component) in which in-patient hospital expenditure (y) was regressed on Rx utilization (x_p) and confounder controls (x_o and x_u) using the data from the subsample of individuals who experienced at least one hospitalization during the year. For this part of the analysis, Stuart et al. (2008) used the exponential regression specification defined in equations (15) and (16). Although the authors estimated the marginal effect of Rx utilization in the full two-part model and found that each additional prescription drug used hospital spending reduced by $104 (p-value < .001), they did not report separate effects for the hurdle and levels components of the two-part model. As an illustration of the estimation of a marginal policy effect [as defined in (2)] and the use of the methods presented in this paper, we focus only on the estimation of the effect of Rx us on inpatient hospital expenditure given that the individual has been hospitalized at least once during the year ó i.e. the levels part of the two-part expenditure model. In addition to serving its expositional purpose, this analysis may shed light on the potential differences in Rx drug effects between those who are hospitalized vs. those who are not. This may be of interest because it is likely that adherence to Rx regimens is more closely monitored in an inpatient setting. This would lead to higher observed cost offsets for those who are hospitalized vs. those who are not.

23

We estimated the marginal policy effect of Rx use on hospital expenditure for the hospitalized subpopulation using the following version of (30)

$$\widehat{PE}_{MARG} = \sum_{i=1}^{n_1} \frac{\widehat{pe}_i}{n_1}$$
(47)

where

$$\widehat{pe}_i = \frac{\partial \exp(x_{pi}``_p + x_{oi}``_o + \ddot{x}_{ui}``_u)}{\partial x_p} = ``_p \exp(x_{pi}``_p + x_{oi}``_o + \ddot{x}_{ui}``_u)$$

$\ddot{x}_u = x_p - \exp(w``)$; ``, ``$_p$, ``$_o$ and ``$_u$ are the 2SRI estimates obtained by Stuart et al. (2008); and n_1 denotes the size of the hospitalized subpopulation. The relevant versions of (24) and (25) for 2SRI estimation are, respectively

$$x_p = \exp(w\ ) + x_u$$
(48)

and

$$y_i = \exp(x_{pi}\ _p + x_{oi}\ _o + \ddot{x}_{ui}\ _u) + e_i\ .$$
(49)

Definitions of the variables used in the regression analyses are given in Table 4 and the descriptive statistics of the sample are in Table 5. The 2SRI results are displayed in columns 4 through 9 of Table 6 ó first stage results for in columns 7 through 9, and second stage results for $_p$, $_o$ and $_u$ in columns 4 through 6. The null hypothesis that x_p is exogenous can be tested based on the coefficient of the first stage residual [H_o: $_u = 0$]. As can be seen in the fifth column of Table 6, the t-stat for that coefficient is -2.56 which leads to rejection of the exogeneity null at the .01 significance level. The 2SRI estimates yielded a marginal policy effect estimate of -339.14 indicating that a filling one more prescription leads to a \$339 reduction in

24

inpatient hospital expenditure. As expected this is much higher than the \$104 effect found by Stuart et al. (2008).

Asymptotic inference for (47) can be drawn from the following version of (38)

$$\sqrt{\frac{n}{a\,\widehat{var}(\widehat{PE}_{MARG})}}\,(\widehat{PE}_{MARG} - PE_{MARG}) \xrightarrow{d} n(0,1) \tag{50}$$

where

$$PE_{MARG} = E\left[\frac{\partial \exp(x_p\beta_p + x_o\beta_o + x_u\beta_u)}{\partial x_p}\right]$$

$$a\,\widehat{var}(\widehat{PE}_{MARG}) = \left(\frac{\sum_{i=1}^{n}\nabla\,\widehat{pe}_i}{n}\right)\left(n\,\widehat{ACOV}(\hat{\theta})\right)\left(\frac{\sum_{i=1}^{n}\nabla\,\widehat{pe}_i}{n}\right)' + \left(\frac{\sum_{i=1}^{n}\left(\widehat{pe}_i - \widehat{PE}_{MARG}\right)^2}{n}\right)$$

$\widehat{ACOV}(\hat{\theta})$ is the asymptotic covariance matrix of the 2SRI estimator of

$$\nabla\,\widehat{pe}_i = [\nabla_\beta\,\widehat{pe}_i \quad \nabla_\delta\,\widehat{pe}_i\,]$$

$\theta = [\beta_p \quad \beta_o \quad \beta_u]$ and the details of $\nabla_\beta\,\widehat{pe}_i$ and $\nabla_\delta\,\widehat{pe}_i$ are given in Appendix C. To test the conventional null hypothesis for PE_{MARG} [i.e., H_o: $PE_{MARG} = 0$] we computed the relevant asymptotic t-stat which follows from (50) and obtained

$$\frac{\widehat{PE}_{MARG}}{\sqrt{\frac{a\,\widehat{var}(\widehat{PE}_{MARG})}{n}}} = \frac{-339.14}{136.21} = -2.49\,.$$

The corresponding p-value is .01.

For the purpose of comparison, we conducted a similar analysis using the NLS exponential regression results, which are not corrected for the potential endogeneity of the Rx utilization variable (shown in column 1 of Table 6). This yielded a very small positive (.96) and insignificant (p-value = .96) estimate of the marginal effect, indicating that Rx utilization has virtually no effect on hospital expenditure.

4.3 The Treatment Effect of Advice on Alcohol Use Among Hypertensive Men

For illustrative purposes we revisit the study conducted by Kenkel and Terza (2001) [KT] in which the authors explore the role that physician advice plays in the prevention of alcohol-related problems. They seek to estimate the impact that physician counseling to decrease alcohol consumption has on drinking practices. Their results indicate that policies promoting such brief interventions can be effective. This is important given that, compared to alcohol taxation, physician advice is a more precisely targeted policy that does not impose extra costs on responsible drinkers. Moreover, compared to the resource costs of arresting, processing, and punishing drunk drivers, physician advice may be a lower cost policy alternative. The binary advice variable (x_p) considered by KT is based on survey respondentsø answers to the following question: õHave you ever been told by a physician to drink less?ö In their study, the estimation objective was the treatment effect of x_p on the *amount of drinking*. In the present illustration, we estimate the effect of physician advice on the *likelihood of drinking* at all. Here the outcome of interest (y) is binary (1 if a non-zero amount of drinking is observed, 0 otherwise). We assume that $y_{x_p^*}$ (the counterfactual outcome as it would be at the universally mandated value x_p^* [1 or 0]) follows the probit-type parametric process defined in (17) from which the counterfactual conditional mean regression specification in (18) follows.

26

The model includes x_u to allow for the potential endogeneity of advice. For example, health-minded individuals may have a higher than average propensity to seek advice, and a simultaneously higher than average likelihood of avoiding potentially unhealthy behaviors like drinking. On the other hand, it may be that unobservable influences on drinking are positively related to advice. For example, in the data used in the study (the 1990 National Health Interview Survey) alcohol consumption is observed alcoholism is not. Diagnosed alcoholics may be more likely to receive advice from their doctors.

Following KT and Terza (2008), we formalize the relationship between x_p and x_u by assuming that

$$x_p = I(w + x_u > 0) \tag{51}$$

where $w = [x_o \quad w^+]$, $(x_u \mid w)$ is standard normal distributed, and w^+ is the vector of identifying instrumental variables related to health insurance status, physician contacts, and health problems. If x_o and x_u are comprehensive (i.e. $E[y_{x_p} \mid x_o, x_u] = E[y \mid x_p, x_o, x_u]$), using results from Terza (2008), we can show that consistent full information maximum likelihood (FIML) estimates the parameters of the model, $= [\quad]$ can be obtained by optimizing the following log-likelihood function

$$L(\mid y, x_p, w) = \prod_{i=1}^{n} \left\{ y_i\, x_{pi} \ln(P_{11i}) + y_i\,(1 - x_{pi}) \ln(P_{10i}) + (1 - y_i)\, x_{pi} \ln(P_{01i}) + (1 - y_i)\, x_{pi} \ln(P_{00i}) \right\}$$

$$\tag{52}$$

where

$$P_{11i} = \int_{-w_i}^{\infty} \Phi(x_{pi}\beta_p + x_{oi}\beta_o + x_u\beta_u)\phi(x_u)\,dx_u$$

$$P_{10i} = \int_{-\infty}^{-w_i} \Phi(x_{pi}\beta_p + x_{oi}\beta_o + x_u\beta_u)\phi(x_u)\,dx_u$$

$$P_{01i} = \int_{-w_i}^{\infty} \left[1 - \Phi(x_{pi}\beta_p + x_{oi}\beta_o + x_u\beta_u)\right]\phi(x_u)\,dx_u$$

$$P_{00i} = \int_{-\infty}^{-w_i} \left[1 - \Phi(x_{pi}\beta_p + x_{oi}\beta_o + x_u\beta_u)\right]\phi(x_u)\,dx_u$$

and $\phi(\ )$ denotes the standard normal probability density function. The correct estimator of the treatment effect as given in (3) is, in this case, the following version of (30)

$$\widetilde{PE}_{TREAT} = \sum_{i=1}^{n} \frac{\widetilde{pe_i}}{n} \tag{53}$$

where

$$\widetilde{pe_i} = pe(0, 1, w_i, \tilde{\beta})$$

and

$$pe(0,1,w_i,\beta) = \int_{x_u} \left[\Phi(\beta_p + x_{oi}\beta_o + x_u\beta_u) - \Phi(x_{oi}\beta_o + x_u\beta_u)\right]\phi(x_u)\,dx_u. \tag{54}$$

The data for the analysis came from the 1990 Health Interview Survey -- the same data set used by KT. The definitions of all variables included in the model can be found in Table 7, and the descriptive statistics of the sample are displayed in Table 8. The FIML results for

$= [\begin{array}{ccc} _p & _o & _u \end{array}]$ and $\quad$, obtained from (52), are shown in columns 4 and 9 of Table 9, respectively.[13] Combining these results with (53) we estimated the policy effect of physician advice to be -.39. This point estimate indicates that nearly a 4 point decrease in the probability of being a drinker could be attributed to the receipt of advice from a physician to drink less. The null hypothesis that x_p is exogenous can be tested based on the coefficient of x_u [H_o: $_u = 0$]. As can be seen in the fifth column of Table 9, the t-stat for that coefficient is 3.65 which leads to rejection of the exogeneity null at less than a .01 significance level.

Asymptotic inference for (53) can be drawn from the following version of (38)

$$\sqrt{\frac{n}{\widetilde{a\,var}(\widetilde{PE}_{TREAT})}}(\widetilde{PE}_{TREAT} - PE) \xrightarrow{d} n(0,1) \tag{55}$$

where

$$\widetilde{a\,var}(\widetilde{PE}_{TREAT}) = \left(\frac{\sum_{i=1}^{n}\nabla \widetilde{pe}_i}{n}\right)\left(n\ \widetilde{ACOV}(\tilde{\ })\right)\left(\frac{\sum_{i=1}^{n}\nabla \widetilde{pe}_i}{n}\right)' + \left(\frac{\sum_{i=1}^{n}\left(\widetilde{pe}_i - \widetilde{PE}_{TREAT}\right)^2}{n}\right)$$

$\widetilde{ACOV}(\tilde{\ })$ denotes the estimated asymptotic covariance matrix of $\tilde{\ }$ -- the FIML

estimate of $\quad$ obtained from (52)

$$\nabla \widetilde{pe}_i = [\nabla_p \widetilde{pe}_i \quad \nabla_o \widetilde{pe}_i \quad \nabla_u \widetilde{pe}_i]$$

and the details of $\nabla_p \widetilde{pe}_i$, $\nabla_o \widetilde{pe}_i$ and $\nabla_u \widetilde{pe}_i$ are given in Appendix D.

---

[13] The integrals required for the likelihood function (52) can be evaluated using quadrature or simulation methods. We used the GAUSS INTQUAD1 procedure for this purpose.

For the purpose of comparison, we estimated the drinking equation with simple probit analysis ignoring the potential endogeneity of the advice variable. In this case the relevant probit model and policy effect estimator are, respectively,

$$y = I(x_p \beta_p + x_o \beta_o + \varepsilon > 0) \tag{56}$$

$$\widetilde{PE} = \sum_{i=1}^{n} \frac{\widetilde{pe}_i}{n} \tag{57}$$

where

$$\widetilde{pe}_i = pe(0,1,x_i,\tilde{\beta})$$

$$pe(0,1,x_i,\tilde{\beta}) = \Phi(\tilde{\beta}_p + x_{oi}\tilde{\beta}_o) - \Phi(x_{oi}\tilde{\beta}_o).$$

$$x_p = I(w\gamma + x_u > 0)$$

where $(\varepsilon \mid x_p, x_o)$ is standard normal distributed. The results from estimating the probit model in (56) are given in the first column of Table 9. Combining these results with (57) we estimated the policy effect of physician advice to be 0.078 with asymptotic t-stat value of 4.58 (p-value < .0001). This point estimate is counterintuitive, indicating that nearly an 8 point increase in the probability of being a drinker could be attributed to the receipt of advice from a physician to drink less.

5. Discussion

This paper offers a generic and unified framework for the use of nonexperimentally based regression results for policy analysis paying particular attention to correcting for the potential endogeneity of the policy variable. We introduce a PDOM estimator that is easy to implement

and derive its asymptotic properties as a special case of the general class of 2SOPT estimators. Empirical applications demonstrating the implementation of the incremental, marginal, and treatment effect versions of the PDOM are discussed. In all three of these examples, the policy variable is found to be endogenous. The results indicate that ignoring the presence of unobservable confounders can lead to substantial bias.

Appendix A

From Theorem 6.11 of White (1994), we have

$$a \, var\left(\overline{PE}\right) = E\left[\nabla_{PE\,PE}q\right]^{-1}\left[E\left[\nabla_{PE}\ q\right]AVAR(\bar{\ })E\left[\nabla_{PE}\ q\right]'\right.$$

$$- E\left[\nabla_{PE}q\nabla\ q_1\right]E\left[\nabla\ q_1\right]^{-1}E\left[\nabla_{PE}\ q\right]' - E\left[\nabla_{PE}\ q\right]E\left[\nabla\ q_1\right]^{-1}E\left[\nabla\ q_1'\nabla_{PE}q\right]'$$

$$+\ E\left[\nabla_{PE}q^2\right]\left]E\left[\nabla_{PE\,PE}q\right]^{-1}\right. \tag{A-1}$$

Moreover, $\nabla_{PE}q = 2(pe - PE)$, $\nabla_{PE\,PE}q = 2$, $E\left[\nabla_{PE}\ q\right] = -2E\left[\nabla\ pe\right]$ and

$E\left[\nabla_{PE}q^2\right] = 4E\left[(pe - PE)^2\right]$, where pe is shorthand notation for $pe(x_{pi}, w_i,\ )$. Therefore,

(A-1) can be written

$$a \, var\left(\overline{PE}\right) = E\left[\nabla\ pe\right]AVAR(\bar{\ })E\left[\nabla\ pe\right]'$$

$$+ E\left[(pe - PE)\nabla\ q_1\right]E\left[\nabla\ q_1\right]^{-1}E\left[\nabla\ pe\right]'$$

$$+ E\left[\nabla\ pe\right]E\left[\nabla\ q_1\right]^{-1}E\left[(pe - PE)\nabla\ q_1\right]'$$

$$+\ E\left[(pe - PE)^2\right] \tag{A-2}$$

Now $q_1(\ , u)$, defined in (32) in the text, can always be written as $r_1(\ _1,\ (\ _2, w, x_p), u)$, where
$(\ _2, w, x_p)$ denotes a function of a subvector of   and observable variables w and x_p that itself
becomes parametric conditional on these observable variables. Now the true value of   is
defined such that $E\left[r_1\left(\ _1,\ (\ _1, w, x_p), u\right) \mid w, x_p\right]$ is maximized with respect to $_1$ and $(\ _2, w, x_p)$.
So that

$$E\left[\nabla_{\ _1}r_1\left(\ _1,\ (\ _2, w, x_p), u\right) \mid w, x_p\right] = 0$$

32

and

$$E\left[\nabla\ r_l\left(\ _1,\ (\ _2,w,x_p),\ u\right)\mid w,x_p\right]=0.\qquad\text{(A-3)}$$

Now note that we can write

$$E\left[\nabla\ q_l(\ ,u)\mid x_p,w\right]=E\left[\nabla\ r_l\left(\ _1,\ (\ _2,x_p,w),\ u\right)\mid x_p,w\right]$$

$$=\begin{bmatrix}E\left[\nabla_{\ _1}r_l\left(\ _1,\ (\ _2,w,x_p),\ u\right)\mid x_p,w\right]\\ E\left[\nabla\ r_l\left(\ _1,\ (\ _2,w,x_p),\ u\right)\mid x_p,w\right]\nabla_{\ _2}\ (\ _2,w,x_p)\end{bmatrix}.$$

so using (A-3) we have

$$E\left[\nabla\ q_l(\ ,u)\mid x_p,w\right]=0.\qquad\text{(A-4)}$$

In the notation of (A-2), it follows from (A-4) that

$$E\left[(pe-PE)\nabla\ q_l\right]=E\left[E\left[(pe-PE)\nabla\ q_l\mid x_p,w\right]\right]$$

$$=E\left[(pe-PE)E\left[\nabla\ q_l\mid x_p,w\right]\right]$$

$$=0.\qquad\text{(A-5)}$$

In light of (A-5), (A-2) can be rewritten as

$$a\ var\left(\overline{PE}\right)=E\left[\nabla\ pe\right]AVAR(\bar{\ })E\left[\nabla\ pe\right]'+E\left[(pe-PE)^2\right].\qquad\text{(A-6)}$$

As an example, consider the nonlinear regression formulation implemented section 4.1. In this case we have

$$q_1(\ ,u) = -\left(y - M(x_p, x_o, x_p - w\ ,\ )\right)^2$$

where $M(x_p, x_o, x_p - w\ ,\ ) = E[y \mid x_p, w]$ is defined as in (43). Here $_1$ has no elements,

$$_2 = [\quad _p \quad _o \quad _u \quad]$$

$$(\ _2, w, x_p) = M(x_p, x_o, x_p - w\ ,\ ),$$

and

$$r_1\left(\ _1,\ (\ _2, w, x_p), u\right) = -\left(y - (\ _2, w, x_p)\right)^2 .$$

Because $(\ _2, w, x_p)$ is defined in this case as the mean of $(y \mid x_p, w)$, it is the optimizer of

$$E[q_1(\ ,u) \mid x_p, w] = E\left[r_1\left(\ _1,\ (\ _2, w, x_p), u\right) \mid x_p, w\right] = -E\left[\left((y \mid x_p, w) - (\ _2, w, x_p)\right)^2\right] .$$

Therefore

$$E\left[\nabla\ r_1\left(\ _1,\ (\ _2, w, x_p), u\right) \mid x_p, w\right] = 2E\left[(y \mid x_p, w) - (\ _2, w, x_p)\right] = 0$$

and

$$E\left[\nabla\ q_1(\ ,u) \mid x_p, w\right] = E\left[\nabla\ r_1\left(\ _1,\ (\ _2, x_p, w),\ u\right) \mid x_p, w\right]$$

$$= E\left[\nabla\ r_1\left(\ _1,\ (\ _2, w, x_p),\ u\right) \mid x_p, w\right] \nabla\ _2\ (\ _2, w, x_p)$$

$$= 0.$$

Appendix B

Recall

$$\widehat{pe}_i = pe(270, 0, w_i, \text{``})$$

$$pe(270, 0, w_i, \ ) = M(0, x_{oi}, (x_{pi} - w_i \ ), \ ) - M(270, x_{oi}, (x_{pi} - w_i \ ), \ )$$

`` is the 2SRI estimate of $\ = [\quad\quad]$

and

$$\nabla \ \widehat{pe}_i = [\nabla \ \widehat{pe}_i \quad \nabla \ \widehat{pe}_i \quad \nabla \ \widehat{pe}_i \ ].$$

To simplify the notation, let us write

$$M(x_p^*, x_{oi}, (x_{pi} - w_i \ ), \ ) = k(\ , \ , \ ; x_p^*) = \ (\ , \ , \ ; x_p^*)^{\left(\frac{2}{=}\right)}$$

$$(\ , \ , \ ; x_p^*) = \frac{1}{2}(x \ ) + 1$$

$$x \ = x_p^* \ _p + x_o \ _o + (x_p - w \ ) \ _u$$

$$x = [x_p^* \quad x_o \quad x_u]$$

$$x_u = x_p - w$$

and

$$' = [\ _p \quad _o' \quad _u].$$

So

$$\nabla\ pe = \nabla\ k(\ ,\ ,\ ; 0) - \nabla\ k(\ ,\ ,\ ; 270)$$

$$= -\ _u\left[\ (\ ,\ ,\ ; 0)^{\left(\frac{2}{}-1\right)} -\ (\ ,\ ,\ ; 270)^{\left(\frac{2}{}-1\right)}\right]w$$

$$\nabla\ pe = \nabla\ k(\ ,\ ,\ ; 0) - \nabla\ k(\ ,\ ,\ ; 270)$$

$$=\ (\ ,\ ,\ ; 0)^{\left(\frac{2}{}-1\right)}[0\ x_o\ x_u] -\ (\ ,\ ,\ ; 270)^{\left(\frac{2}{}-1\right)}[270\ x_o\ x_u]$$

and

$$\nabla\ pe = \nabla\ k(\ ,\ ,\ ; 0) - \nabla\ k(\ ,\ ,\ ; 270)$$

$$= k(\ ,\ ,\ ; 0)\left[\left(\frac{1}{}\right)\left(\frac{x_o\ _o + x_u\ _u}{(\ ,\ ,\ ; 0)}\right) - \left(\frac{1}{2}\right)\ln\left(\ (\ ,\ ,\ ; 0)^2\right)\right]$$

$$- k(\ ,\ ,\ ; 270)\left[\left(\frac{1}{}\right)\left(\frac{270\ _p + x_o\ _o + x_u\ _u}{(\ ,\ ,\ ; 270)}\right) - \left(\frac{1}{2}\right)\ln\left(\ (\ ,\ ,\ ; 270)^2\right)\right].$$

Recall

$$\widehat{pe}_i = {}^{"}_p \exp\left(x_{pi}{}^{"}_p + x_{oi}{}^{"}_o + \left(x_{pi} - \exp(w_i{}^{"}) \right){}^{"}_u \right)$$

$${}^{"} = [\,{}^{"}_p \quad {}^{"}_* \,] \text{ is the 2SRI estimate of } \quad = [\,_p \quad _* \,] \text{ where } \quad _* = [\,_o \quad _u \,]$$

and

$$\nabla\, \widehat{pe}_i = [\, \nabla\, \widehat{pe}_i \quad \nabla_p\, \widehat{pe}_i \quad \nabla_*\, \widehat{pe}_i \,].$$

In this case

$$\nabla\, \widehat{pe}_i = -\,_p\,_u \exp\left(x_{pi}{}^{"}_p + x_{oi}{}^{"}_o + \left(x_{pi} - \exp(w_i{}^{"}) \right){}^{"}_u \right) \exp(w_i{}^{"}) w_i$$

$$\nabla_p\, \widehat{pe}_i = \exp\left(x_{pi}{}^{"}_p + x_{oi}{}^{"}_o + \left(x_{pi} - \exp(w_i{}^{"}) \right){}^{"}_u \right)(1 + {}^{"}_p x_{pi})$$

$$\nabla_*\, \widehat{pe}_i = {}^{"}_p \exp\left(x_{pi}{}^{"}_p + x_{oi}{}^{"}_o + \left(x_{pi} - \exp(w_i{}^{"}) \right){}^{"}_u \right) x_{*i}$$

where $x_{*i} = [x_{oi} \quad \breve{x}_{ui}]$ and $\breve{x}_{ui} = x_{pi} - \exp(w_i{}^{"})$

Appendix D

Recall

$$\widetilde{pe}_i = pe(0, 1, w_i, \tilde{\ })$$

$$pe(0, 1, w_i,) = \int_{x_u} \left[\ (\ _p + x_{oi}\ _o + x_u\ _u) - \ (x_{oi}\ _o + x_u\ _u) \right] \ (x_u)\, dx_u$$

$$\tilde{\ } = [\ \tilde{\ }_p \quad \tilde{\ }_o \quad \tilde{\ }_u] \text{ is the FIML estimate of } \ = [\ _p \quad _o \quad _u]$$

and

$$\nabla \widetilde{pe}_i = [\nabla_p \widetilde{pe}_i \quad \nabla_o \widetilde{pe}_i \quad \nabla_u \widetilde{pe}_i].$$

In this case

$$\nabla_p \widetilde{pe}_i = \int_{-\infty}^{\infty} \ (\tilde{\ }_p + x_{oi}\tilde{\ }_o + x_u\tilde{\ }_u) \ (x_u)\, dx_u$$

$$\nabla_o \widetilde{pe}_i = \left[\int_{-\infty}^{\infty} \left\{ \ (\tilde{\ }_p + x_{oi}\tilde{\ }_o + x_u\tilde{\ }_u) - \ (x_{oi}\tilde{\ }_o + x_u\tilde{\ }_u) \right\} \ (x_u)\, dx_u \right] x_{oi}$$

$$\nabla_u \widetilde{pe}_i = \int_{-\infty}^{\infty} \left\{ \ (\tilde{\ }_p + x_{oi}\tilde{\ }_o + x_u\tilde{\ }_u) - \ (x_{oi}\tilde{\ }_o + x_u\tilde{\ }_u) \right\} x_u \ (x_u)\, dx_u$$

References

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

Bierens, H. J. (1994): *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-section and Time Series Models*, New York: Cambridge University Press.

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 and Prevention (2006): "Cigarette Smoking Among Adults -- United States 2006," *Morbidity and Mortality Weekly Report*, 56, 1157-1161.

Coulson, E., Neslusan, C., Stuart, B., and Terza J. (1995): "Estimating the Moral Hazard Effect of Supplemental Medical Insurance in the Demand for Prescription Drugs by the Elderly,ö and *American Economic Review - Papers and Proceedings*, 85, 122-126.

Koc, C. (2005): õHealth-Specific Moral Hazard Effects,ö *Southern Economic Journal*, 72, 98-118.

McGeary, K.A., and M. T. French (2000): "Illicit Drug Use and Emergency Room Utilization," *Health Services Research*, 35, 153-169.

Mullahy, J. (1985): *Cigarette Smoking: Habits, Health Concerns, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand*, Ph.D. Dissertation, University of Virginia.

_____ (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.

Neslusan CA, Hylan TR, Dunn RL, Donoghue J. (1999): õControlling for Systematic Selection in Retrospective Analyses: An Application to Fluoxetine and Sertraline Prescribing in the United Kingdom,ö *Value in Health*, 2, 435-445.

Pryor, C., and Terza, J. (2002): "Are Going Concern Audit Opinions a Self-Fulfilling Prophecy?" *Advances in Quantitative Analysis of Finance and Accounting*, 10, 89-116.

Stuart, B.C, Terza, J.V., and Doshi, J. (2008): õAssessing the Impact of Drug Use on Hospital Costs,ö Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Terza, J. (1994a): "An Estimator for Nonlinear Regression Models with Endogenous Switching and Sample Selection," Working paper, Department of Economics, Penn State University.

_____ (1994b): "Dummy Endogenous Variables and Endogenous Switching in Exponential Conditional Mean Regression Models," Working paper, Department of Economics, Penn State University, presented at North American Summer Meeting of the Econometric Society, Quebec, Canada.

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

_____ (1999): õEstimating Endogenous Treatment Effects in Retrospective Data Analysis,ö*Value in Health*, 2, (1999), 429-434.

_____ (2002): õAlcohol Abuse and Employment: A Second Lookö *Journal of Applied Econometrics*, 17, 393-404.

_____ (2008): õParametric Nonlinear Regression with Endogenous Switching,ö *Econometric Reviews*, forthcoming.

_____, Kenkel D.S., Lin, T., and Sakata, S. (2008): "Care-Giver Advice as a Preventive Measure for Drinking During Pregnancy: Zeros, Categorical Outcome Responses, and Endogeneity," *Health Economics,* 17, 41-54.

_____, Bradford, W.D. and Dismuke, C.E. (2007): õThe Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note," *Health Services Research*, in press, published online at doi: 10.1111/j.1475-6773.2007.00807.x.

_____, Basu, A. and Rathouz, P. (2008): õTwo-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling,ö *Journal of Health Economics*, in press, published online at doi:10.1016/j.jhealeco.2007.09.009.

Treglia, M, Neslusan, C.A., Dunn R.L. (1999): õFluoxetine and Dothiepin Therapy in Primary Care and Health Resource Utilization: Evidence from the United Kingdom,ö *International Journal of Psychiatry in Clinical Practice*, 3, 23-30.

Volpp, K.A., A.G. Levy, D.A. Asch, J.A., Berlin, J.J., Murphy, A. Gomez, H. Sox, J. Zhu, and C. Lerman (2006): "A Randomized Controlled Trial of Financial Incentives for Smoking Cessation," *Cancer Epidemiology, Biomarkers, and Prevention*, 15, 12-8.

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

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

_____ (2003): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Table 1 -- Cigarette Demand Model: Variable Definitions

Variable	Definition
Outcome Variable (y)	
Number of cigarettes consumed per day	
Potentially Endogenous Policy Variable (x_p)	
HABITSTOCK*	An index of the habit-forming effects of prior cigarette consumption.
Observable Confounders (x_o)	
PRICE	After-tax cigarette tax price (statewide or regional)
REST79	State price index
INCOME	Family income (midpoint of interval) in thousands
AGE	Age at time of interview
AGE^2	Square of Age at time of interview
EDUCATION	Education completed in years at time of interview
EDUCATION^2	Square of Education completed at time of interview
FAMSIZE	Number of individuals in Household
RACE	Binary Variable=1 if individual is white
Instrumental Variables (w^+)	
RESTOCK	Interaction between state price index and habit stock
LAGPRICE	Last period after-tax cigarette price
AGE^3	Age at interview cubed
EDUCATION^3	Education completed at interview cubed
AGE×EDUCATION	Interaction between age and education level at interview

*This variable is an index based on accumulated smoking levels over time for current smokers and depreciated smoking levels for former smokers. See Mullahy (1985), Appendix 4-B for details.

Table 2 – Cigarette Demand Model: Descriptive Statistics for the Sample

Variable	Mean	Min	Max
Outcome Variable (y)			
y	0.412	0.000	4.900
Potentially Endogenous Policy Variable (x_p)			
HABITSTOCK	104.071	0.000	977.300
Observable Confounders (x_o)			
PRICE	61.180	46.300	69.800
REST79	0.241	0.000	1.000
INCOME	19047.240	500.000	30000.000
AGE	41.707	17.000	96.000
AGE^2	2041.312	289.000	9216.000
EDUCATION	12.298	0.000	18.000
EDUCATION^2	162.193	0.000	324.000
FAMSIZE	3.189	1.000	13.000
RACE	0.895	0.000	1.000
Instrumental Variables (w^+)			
RESTOCK	0.904	0.000	6.000
LAGPRICE	57.878	41.778	67.052
AGE^3	112886.638	4913.000	884736.000
EDUCATION^3	2245.499	0.000	5832.000
AGE×EDUCATION	499.563	1530.000	0.000

Table 3: Cigarette Demand Model: Nonlinear Regression Results

Variable	NLS-IBC Estimates of Cigarette Consumption Parameters (Not Corrected for Endogeneity)			2SRI-IBC Estimates of Cigarette Consumption Parameters (Corrected for Endogeneity)			OLS Estimates of Habit Stock Parameters		
	Estimate	t-stat	p-val	Estimate	t-stat	p-val	Estimate	t-stat	p-val
HABITSTOCK	0.01	29.48	<0.001	0.01	8.61	<0.001	--	--	--
PRICE	-1.8×10^{-3}	-1.26	0.21	-8×10^{-4}	-0.50	0.62	-0.76	-0.60	0.55
RESTAURANT	-0.01	-0.80	0.43	0.01	0.58	0.56	-16.54	-1.47	0.14
INCOME	1.9×10^{-7}	0.24	0.81	8.8×10^{-7}	0.96	0.34	-2.7×10^{-4}	-1.43	0.15
AGE	-0.02	-7.93	<0.001	-0.05	-4.72	<0.001	22.42	12.64	<0.001
AGE^2	1.2×10^{-4}	5.09	<0.001	4.6×10^{-4}	4.16	<0.001	-0.35	-8.71	<0.001
EDUCATION	-0.01	-1.09	0.28	-0.04	-3.13	0.002	6.70	1.12	0.26
EDUCATION^2	5.4×10^{-5}	0.16	0.87	1.9×10^{-3}	2.80	0.01	-0.41	-0.70	0.48
FAMSIZE	1.3×10^{-3}	0.32	0.75	3.1×10^{-3}	0.69	0.49	-0.75	-0.81	0.42
RACE	-0.01	-0.39	0.70	1×10^{-3}	0.04	0.97	-3.45	-0.68	0.50
RESTOCK	--	--		--	--		1.70	0.59	0.56
LAGPRICE	--	--		--	--		0.29	0.23	0.81
AGE^3	--	--		--	--		1.5×10^{-3}	5.27	<0.001
EDUCATION^3	--	--		--	--		-0.01	-0.67	0.50
AGE×EDUCATION	--	--		--	--		0.06	2.00	0.05
CONSTANT	-0.36	-3.32	<0.001	0.09	0.49	0.62	-256.72	-6.25	<0.001
First-Stage Residual	--	--		-2.4×10^{-3}	-3.01		--	--	
γ (H$_o$: γ = 0)	1.76	28.05	<0.001	1.77	28.22	<0.001	--	--	

Table 4 -- Hospital Expenditure Model: Variable Definitions

Variable	Definition
Outcome Variable (y)	
Hospital expense for the sample as a whole	
Potentially Endogenous Policy Variable (x_p)	
RXUSE	Prescription fills
Observable Confounders (x_o)	
DISABLED	Medicare entitlement status -- SSDI Disabled (<65)
DISAGED	Medicare entitlement status -- Aged/previously disabled (>65)
AGE74	Age 70-74
AGE79	Age 75-79
AGEGT80	Age 80+
MARRIED	Binary Variable=1 if individual is married
FEMALE	Binary Variable=1 if individual is female
RURAL	Urban residence
HSGRAD	Educational attainment -- High school graduate
MIDWEST	Census region -- Midwest
SOUTH	Census region ó South
WEST	Census region -- West
INC20	Annual income between $10,001 - $20,000
INC30	Annual income between $20,001 - $30,000
INCGT30	Annual income > $30,000
WHITE	Binary Variable=1 if individual is white
RISKADJ	DCG/HCC risk adjuster
Instrumental Variables (w^+)	
RXCOVRD	Full-year drug coverage

Source: 1999 and 2000 MCBS

Table 5 – Hospital Expenditure Model: Descriptive Statistics for the Sample

Variable	Mean	Min	Max
Outcome Variable (y)			
y	2097.841	95251.360	0.000
Potentially Endogenous Policy Variable (x_p)			
RXUSE	30.052	229.000	0.000
Observable Confounders (x_o)			
DISABLED	0.162	1.000	0.000
DISAGED	0.071	1.000	0.000
AGE74	0.243	1.000	0.000
AGE79	0.188	1.000	0.000
AGEGT80	0.284	1.000	0.000
MARRIED	0.484	1.000	0.000
FEMALE	0.571	1.000	0.000
RURAL	0.339	1.000	0.000
HSGRAD	0.643	1.000	0.000
MIDWEST	0.213	1.000	0.000
SOUTH	0.406	1.000	0.000
WEST	0.176	1.000	0.000
INC20	0.263	1.000	0.000
INC30	0.199	1.000	0.000
INCGT30	0.269	1.000	0.000
WHITE	0.865	1.000	0.000
RISKADJ	1.007	8.068	0.206
Instrumental Variable (w^+)			
RXCOVRD	0.773	1.000	0.000

Table 6 – Hospital Expenditure Model: Nonlinear Regression Results

Variable	NLS Exponential Regression Estimates of Hospital Expenditure Equation (Not Corrected for Endogeneity)			2SRI Exponential Regression Estimates of Hospital Expenditure Equation (Corrected for Endogeneity)			NLS Exponential Regression Estimates of Rx Usage Equation		
	Estimate	t-stat	p-val	Estimate	t-stat	p-val	Estimate	t-stat	p-val
RXUSE	9.5×10^{-5}	0.05	0.96	-0.04	-2.48	0.01	--	--	--
DISABLED	0.32	1.53	0.13	0.68	2.28	0.02	0.22	2.98	0.003
DISAGED	0.38	2.16	0.03	0.75	2.98	0.003	0.27	4.10	<0.001
AGE74	0.06	0.38	0.71	0.10	0.53	0.60	-0.04	-0.53	0.60
AGE79	0.21	1.06	0.29	0.14	0.69	0.49	-0.06	-0.87	0.38
AGEGT80	0.04	0.23	0.82	-0.12	-0.53	0.60	-0.04	-0.53	0.60
MARRIED	0.09	0.99	0.32	0.19	1.70	0.09	0.04	0.97	0.33
FEMALE	-0.25	-2.60	0.01	0.02	0.16	0.88	0.23	6.21	<0.001
RURAL	-0.01	-0.04	0.97	-2.2×10^{-3}	-0.01	0.99	0.01	0.26	0.80
HSGRAD	-0.34	-3.10	0.002	-0.43	-3.08	0.002	-0.08	-2.13	0.03
MIDWEST	-0.53	-3.48	<0.001	-0.43	-2.55	0.01	0.10	1.86	0.06
SOUTH	-0.36	-2.55	0.01	-0.31	-2.14	0.03	0.02	0.46	0.64
WEST	0.06	0.36	0.72	-3.2×10^{-3}	-0.02	0.99	-0.09	-1.60	0.11
INC20	-0.16	-1.30	0.19	-0.33	-1.94	0.05	-0.06	-1.18	0.24
INC30	-0.41	-2.71	0.01	-0.54	-2.90	0.004	-0.10	-1.68	0.09
INCGT30	0.30	2.05	0.04	0.02	0.10	0.92	-0.22	-3.88	<0.001
WHITE	0.10	0.76	0.45	0.39	2.00	0.05	0.20	3.75	<0.001
RISKADJ	0.14	3.85	<0.001	0.48	3.82	<0.001	0.19	9.48	<0.001
RXCOVRD	--	--	--	--	--	--	0.29	7.01	<0.001
CONSTANT	9.29	37.02	<0.001	9.70	26.91	<0.001	2.73	28.78	<0.001
First Stage Residual	--	--	--	0.04	2.56	0.01	--	--	--

Table 7 -- Drinking Model: Variable Definitions

Variable	Definition
Outcome Variable (y)	
Total drinks	
Potentially Endogenous Policy Variable (x_p)	
ADVICE	Drinking advice
Observable Confounders (x_o)	
EDITINC	Monthly income ($1000)
AGE40	40 < age Ö50
AGE50	50 < age Ö60
AGE60	60 < age Ö70
AGEGT70	70 < age
EDUC	Years of schooling
BLACK	Black d.v.
OTHER	Non-white, non-black
MARRIED	Married
WIDOW	Widowed
DIVSEP	Divorced or separated
EMPLOYED	Employed
UNEMPLOY	Unemployed
NORTHE	Northeast
MIDWEST	Midwest
SOUTH	South
Instrumental Variables (w^+)	
MEDICARE	Insurance through Medicare
MEDICAID	Insurance through Medicaid
CHAMPUS	Military insurance
HLTHINS	Health insurance
REGMED	Reg. source of care
DRI	See same doctor
MAJORLIM	Limits on major daily activ.
SOMELIM	Limits on some daily activ.
HVDIAB	Have diabetes
HHRTCOND	Have heart condition
HADSTROKE	Had stroke

Table 8 – Drinking Model: Descriptive Statistics for the Sample

Variable	Mean	Min	Max
Outcome Variable (y)			
y	0.786	0.000	1.000
Potentially Endogenous Policy Variable (x_p)			
DADLALC	0.279	0.000	1.000
Observable Confounders (x_o)			
EDITINC	2.579	-0.900	100.800
AGE40	0.179	0.000	1.000
AGE50	0.195	0.000	1.000
AGE60	0.183	0.000	1.000
AGE70	0.199	0.000	1.000
AGEGT70	0.122	0.000	1.000
EDUC	12.926	0.000	18.000
BLACK	0.133	0.000	1.000
OTHER	0.018	0.000	1.000
MARRIED	0.645	0.000	1.000
WIDOW	0.052	0.000	1.000
DIVSEP	0.160	0.000	1.000
EMPLOYED	0.666	0.000	1.000
UNEMPLOY	0.029	0.000	1.000
NORTHE	0.218	0.000	1.000
MIDWEST	0.275	0.000	1.000
SOUTH	0.295	0.000	1.000
Instrumental Variables (w^+)			
MEDICARE	0.252	0.000	1.000
MEDICAID	0.031	0.000	1.000
CHAMPUS	0.059	0.000	1.000
HLTHINS	0.815	0.000	1.000
REGMED	0.821	0.000	1.000
DR1	0.720	0.000	1.000
MAJORLIM	0.086	0.000	1.000
SOMELIM	0.076	0.000	1.000
HVDIAB	0.061	0.000	1.000
HHRTCOND	0.146	0.000	1.000
HADSTROK	0.036	0.000	1.000

Table 9 -- Drinking Model: Nonlinear Regression Results

Variable	Simple Probit Estimates of Drinking Equation (Not Corrected for Endogeneity)			Joint FIML Estimates of Drinking Equation (Corrected for Endogeneity)			Joint FIML Estimates of Advice Equation		
	Estimate	z-stat	p-val	Estimate	z-stat	p-val	Estimate	z-stat	p-value
DADLALC	0.29	4.31	<0.001	-1.78	-3.08	0.001	--	--	--
EDITINC	0.01	0.80	0.42	0.01	0.67	0.25	-2.5×10^{-3}	-0.43	0.66
AGE40	-0.14	-1.21	0.23	-0.02	-0.13	0.45	0.24	2.26	0.02
AGE50	-0.21	-1.81	0.07	-0.16	-0.91	0.18	0.15	1.33	0.18
AGE60	-0.25	-2.16	0.03	-0.23	-1.32	0.09	0.11	0.98	0.33
AGE70	-0.13	-1.04	0.30	-0.09	-0.48	0.32	0.10	0.79	0.43
AGEGT70	-0.03	-0.22	0.83	-0.03	-0.16	0.44	0.15	0.92	0.36
EDUC	0.04	3.54	<0.001	0.02	1.59	0.06	-0.03	-3.38	<0.001
BLACK	-2.3×10^{-3}	-0.03	0.98	0.22	1.53	0.06	0.28	3.41	<0.001
OTHER	0.11	0.46	0.64	0.31	0.91	0.18	0.22	1.10	0.27
MARRIED	-0.03	-0.27	0.78	1.8×10^{-3}	0.01	0.50	0.16	1.80	0.07
WIDOW	0.01	0.05	0.96	0.17	0.71	0.24	0.28	1.84	0.07
DIVSEP	0.01	0.06	0.95	0.14	0.83	0.20	0.30	2.85	0.002
EMPLOYED	0.26	3.14	0.00	0.22	1.87	0.03	3.2×10^{-4}	3.7×10^{-3}	1.00
UNEMPLOY	0.37	1.94	0.05	0.54	1.90	0.03	0.21	1.22	0.22
NORTHE	-0.13	-1.43	0.15	-0.12	-0.91	0.18	0.08	0.92	0.36
MIDWEST	-0.16	-1.82	0.07	-0.24	-1.83	0.03	-0.03	-0.35	0.73
SOUTH	-0.15	-1.74	0.08	-0.21	-1.65	0.05	-0.04	-0.44	0.66
Intercept	0.35	1.85	0.07	1.24	3.21	<0.001	-0.59	-3.16	<0.001
MEDICARE	--	--	--	--	--	--	-0.02	-0.15	0.88
MEDICAID	--	--	--	--	--	--	0.13	0.92	0.36
CHAMPUS	--	--	--	--	--	--	0.15	1.41	0.16
HLTHINS	--	--	--	--	--	--	-0.15	-1.99	0.05
REGMED	--	--	--	--	--	--	0.09	0.92	0.36
DR1	--	--	--	--	--	--	0.06	0.67	0.50
MAJORLIM	--	--	--	--	--	--	0.16	1.53	0.13
SOMELIM	--	--	--	--	--	--	0.13	1.41	0.16
HVDIAB	--	--	--	--	--	--	0.24	2.35	0.02
HHRTCOND	--	--	--	--	--	--	0.15	2.04	0.04
HADSTROK	--	--	--	--	--	--	0.03	0.26	0.80
x_u				1.23	3.65	<0.001			