© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2007.00807.x
METHODS ARTICLE

The Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note

Joseph V. Terza, W. David Bradford, and Clara E. Dismuke

Objective. To investigate potential bias in the use of the conventional linear instrumental variables (IV) method for the estimation of causal effects in inherently nonlinear regression settings.
Data Sources. Smoking Supplement to the 1979 National Health Interview Survey, National Longitudinal Alcohol Epidemiologic Survey, and simulated data.
Study Design. Potential bias from the use of the linear IV method in nonlinear models is assessed via simulation studies and real world data analyses in two commonly encountered regression setting: (1) models with a nonnegative outcome (e.g., a count) and a continuous endogenous regressor; and (2) models with a binary outcome and a binary endogenous regressor.
Principle Findings. The simulation analyses show that substantial bias in the estimation of causal effects can result from applying the conventional IV method in inherently nonlinear regression settings. Moreover, the bias is not attenuated as the sample size increases. This point is further illustrated in the survey data analyses in which IV-based estimates of the relevant causal effects diverge substantially from those obtained with appropriate nonlinear estimation methods.
Conclusions. We offer this research as a cautionary note to those who would opt for the use of linear specifications in inherently nonlinear settings involving endogeneity.

Key Words. Econometrics, nonlinear models, health economics

When analyzing data with the goal of informing health policy, the ability to draw true causal inference from the estimation results is of paramount importance. The typical health policy analysis focuses on identifying the effect that a variable, over which there exists some degree of policy control (x_p—henceforth the *policy variable*), has on an outcome of some policy interest (y—the *outcome variable*). Empirical analyses that offer estimates of mere associations between x_p and y are of little value to policy makers. Estimating the desired causal effect of x_p on y is not straightforward and is particularly difficult in the context of nonexperimental (survey) data. In observational surveys respondent behavior, as manifested in the value of y, can be influenced by a myriad of stimuli aside from the policy variable x_p. Such alternative influences on y will obfuscate causal inference if they are also correlated with x_p. If not properly taken into account, this will lead to bias in conventional estimation of causal effects.[1] For example, Mullahy (1997) posits that an individual's *habit stock*, accumulated over previous periods of smoking, will have an effect on current cigarette demand.[2] We might observe, for instance, that individuals with higher habit stocks have greater demands for cigarettes. Interpreting such an observation, as indicative of the causal effect of habit stock on current cigarette smoking will, however, be upward biased if "health-minded" individuals tend to smoke less both currently and in the past.

Some confounding influences (henceforth *confounders*) are observable and can be controlled through the use of regression analysis or matching methods. Others, such as an individual's overall health-mindedness in the above example, are unobservable and their influences cannot, therefore, be controlled by standard estimation methods. In such cases, the policy variable of interest is said to be *endogenous*.[3] A commonly implemented method that is designed to deal with endogeneity is the instrumental variables (IV) method.[4] The conventional IV method is based on the assumption that the regression of the outcome of interest (y) on the policy variable (x_p) and the observable confounders is linear. The IV method has, nonetheless, been implemented by many applied researchers in health services research and health economics contexts that are inherently nonlinear; e.g., binary response models, count data models, and limited dependent variable models. In the present paper we show, via simulation, that such applications of the conventional IV method, in which nonlinearity in the specification of the regression of y on x_p and the confounders (both observable and unobservable) is ignored, can lead to bias in the estimation of the causal effect of x_p on y. We conduct the simulation analyses in two common nonlinear regression contexts—binary response models (e.g., probit analysis and logistic regression) and models with non-negative-dependent variables (e.g., count and duration models).[5] We find the

Address correspondence to Joseph V. Terza, Department of Epidemiology and Health Policy Research and Department of Economics, University of Florida, 1329 SW 16th Street, Room 5130, PO Box 100147, Gainesville, FL 32610-0147. W. David Bradford is with the Department of Health Administration and Policy, Center for Health Economic and Policy Studies, Medical University of South Carolina, Charleston, SC. Clara E. Dismuke is with the Center for Health Economic and Policy Studies, Medical University of South Carolina, Charleston, SC.

potential bias to be substantial. This point is further illustrated in the context of two applied examples using actual data. We first revisit Mullahy's (1997) model of the effect of habit stock (a potentially endogenous continuous variable) on cigarette demand (a count variable). We find, using data obtained from the author, that there is substantial divergence between the linear IV estimate of the habit stock effect and that which we obtained under a flexible specification that accounts for both endogeneity and the inherent nonlinearity of the count regression model. As a second example, we estimate the effect of substance abuse (a potentially endogenous binary variable) on employment status (a binary variable—employed versus not employed). The model is a binary analog to the one used by Mullahy and Sindelar (1996) and Terza (2002) for the estimation of the effect of alcohol abuse on a three-category employment outcome (out of the labor force, unemployed, and employed). We find a substantial difference between the IV and bivariate probit (BVP) estimates of the substance abuse effect. Moreover, we find that, the former is statistically insignificant while the latter is highly significant.

The remainder of the paper is organized as follows: in the next section, we review the health services research and health economics literatures, focusing on applications of the linear IV method in inherently nonlinear modeling contexts. We then describe the sampling design for, and results from, our simulation study in which substantial evidence of potential IV bias is demonstrated. Our analyses of the effect of habit stock on cigarette demand and the impact of substance abuse on employment are discussed in the following section. Comparisons of the IV and nonlinear regression estimates of the causal effects in both cases comport with the results of our simulation study. The final section summarizes the results and recommends caution when using the conventional IV estimator in inherently nonlinear settings.

BACKGROUND

Since McClellan, McNeil, and Newhouse (1994) appeared in the *Journal of the American Medical Association*, it has been fairly common practice in applied health services research and health economics to use the conventional linear IV method to estimate causal effects from survey data. In many of these applications, however, the underlying regression model is inherently nonlinear. In such cases, to accommodate the linear IV method in the presence of endogeneity, researchers have replaced: probit models with linear probability models; Poisson regressions with linear regressions; and conventional nonlinear duration formulations with linear approximations. Recent examples include Brooks et al. (2003, 2006), Stone et al. (2006), Beegle, Dehejia, and Gatti (2005), Lo Sasso and Buchmeuller (2004), Grabowski and Hirth (2003), Frances et al. (2000), Malkin, Broder, and Keeler (2000), Gowrisankaran and Town (1999), and Mullahy and Sindelar (1996).[6] It appears that in many instances this approach is adopted due to an incorrect perception that appropriate and practical nonlinear alternatives to conventional IV do not exist. For instance, in a duration modeling context, Gowrisankaran and Town (1999) explain that "the reason that we use a linear probability model instead of a more common Weibull or Lognormal specifications for the hazard is that it is extremely difficult to use nonlinear models such as these with endogenous variables" (p. 754).[7] In fact, methods designed to cope with endogeneity in nonlinear modeling are readily available for the vast majority of cases because most regression studies in health services research and health economics fall into one of two categories: (1) those with binary dependent variables; and (2) those with count, duration, or otherwise nonnegative, dependent variables. For models in the former category, the regression specification should be restricted to the unit interval. This restriction is typically imposed via a cumulative distribution function formulation of the regression function, and methods for dealing with endogeneity in such probit and logit type contexts have been offered by Ashford and Sowden (1970), Blundell and Smith (1989, 1993), Lee (1979, 1981), Newey (1987), Rivers and Vuong (1988), and Smith and Blundell (1986).[8] Appropriate applications of these methods in health services research and health economics can be found in Bao, Duan, and Fox (2006), Alexandre and French (2001), Averett et al. (2004), Bollen, Guilkey, and Mroz (1995), Mroz et al. (1999), Norton, Lindrooth, and Ennett (1998), Rees et al. (2001), Ribar (1994), and Sen (2002). In nonnegative dependent variable models, the exponential regression specification is typically employed. Estimators that correct for endogeneity in the exponential regression framework have been suggested by Mullahy (1997), Terza (1994, 1998), and Wooldridge (1997, 2002). These methods have been applied in health services research and health economics by Dickie (2005), Koc (2005), Neslusan et al. (1999), and Treglia, Neslusan, and Dunn (1999). Wooldridge (2002) suggests the use of a two-stage residual inclusion (2SRI) method for count data models. Terza, Basu, and Rathouz (2007) explore the use of the 2SRI method for models with more general forms of nonlinearity. Examples of the use of the 2SRI method in health services research and health economics can be found in Basur et al. (2004), DeSimone (2002), Gibson et al. (2006), Norton and van Houtven (2006), Moran and Simon (2006), Shea et al. (2007), Harold van

Houtven and Norton (2007), and Stuart, Doshi, and Terza (2007). Authors who have applied 2SRI methods in other fields of study are Alvarez and Glasgow (1999), Burnett (1997), McGarrity and Sutter (2000), and Petrin and Train (2006).

In the following section, we explore the potential bias in using the IV method in a linearized version of an inherently nonlinear model. We simulate outcomes data in the two commonly encountered nonlinear contexts discussed above—nonnegative regression models and binary response models. For the former, we simulate outcomes data using a flexible nonlinear functional form—a variant of the inverse Box and Cox (1964) (IBC) model proposed by Wooldridge (1992). The sampling design incorporates a continuous endogenous regressor. For each simulated sample, the conventional linear IV method is used to estimate the relevant marginal effect of the endogenous regressor. Results are compared with the true marginal effect and the average absolute bias is reported for each of six different sample sizes (1,000; 5,000; 10,000; 50,000; 100,000; and 500,000). To each sample, we also apply a consistent nonlinear 2SRI method that is based on the IBC regression model used to generate the data. The average absolute bias of the corresponding 2SRI-based estimated marginal effect estimates (relative to the true value) are evaluated for each of the sample sizes. In an analogous simulation experiment, we generated data on a binary outcome based on BVP sampling design in which one of the regressors in the outcome equation is itself binary and endogenous. For each sample, conventional IV and BVP estimates of the treatment effect of the endogenous regressor are obtained. The estimates are summarized and compared based on average absolute bias, as in the IBC-based simulation exercise for the nonnegative outcome case.

SIMULATION STUDY

In the sequel, we examine the consequences of inappropriate application of the linear IV estimator to cases in which the underlying data generating process for the outcome of interest (y) follows a nonlinear regression specification. As we saw in the literature review, two nonlinear cases in which linear IV is commonly applied are models in which y is nonnegative, and models in which y is binary.

Nonnegative y and Continuous x_p

In order to maintain conformity with the corresponding illustrative example discussed in the next section, we implement a model in which both the outcome of interest and the endogenous policy variable (x_p) are continuous. We used the following regression specification to simulate the nonnegative outcomes data

$$\begin{aligned} E[y|x_p, x_o, x_u] &= k(x_p\beta_p + x_o\beta_o + x_u\beta_u, \gamma) \\ &= \left\{ \left(\left(\frac{\gamma}{2}\right) \left(x_p\beta_p + x_o\beta_o + x_u\beta_u\right) + 1 \right)^2 \right\}^{\frac{1}{\gamma}} \end{aligned} \quad (1)$$

where $E[a|b]$ denotes the conditional expectation of *a* given *b*; x_o and x_u represent observable and unobservable confounders, respectively; and the βs and γ are parameters to be estimated. The value of γ is unrestricted (i.e., $-\infty < \gamma < \infty$). Equation (1) is a variant of the inverse of the flexible form suggested by Box and Cox (1964). The IBC conditional mean regression specification was first suggested by Wooldridge (1992) and later implemented by Kenkel and Terza (2001) and Basu and Rathouz (2005). The IBC functional form approaches the exponential model as $\gamma \rightarrow 0$. 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.[9] In simulating the data, we designated the *average marginal effect* (AME)

$$E\left[\frac{\partial\, k(x_p\beta_p + x_o\beta_o + x_u\beta_u, \gamma)}{\partial\, x_p}\right]_{x_p = x_p^*} \quad (2)$$

as both the estimation objective and the basis for evaluation of the performance of the IV estimator, where $E[\ ]_{x_p = x_p^*}$ denotes mathematical expectation evaluated at $x_p = x_p^*$. We simulated 1,000 samples of various sizes (1,000; 5,000; 10,000; 50,000; 100,000; and 500,000) with the value of γ fixed at an intermediate value between 0 and 2—viz., $\gamma = 1.765$.[10] We applied the conventional linear IV estimator to each of the generated samples. The IV estimate of the AME, at any value of x_p (x_p^*), is the IV estimated coefficient of that variable. For each sample size, we computed the mean percentage absolute difference between the IV estimated coefficient of x_p ($\widehat{\text{IV}}$) relative to the true value of the average marginal effect, which is measured as

$$\frac{\text{abs}\left(\widehat{\text{IV}} - \text{AME}\right)}{\text{abs}(\text{AME})} \times 100 \text{ percent} \quad (3)$$

where abs(•) is the absolute value function and AME denotes the true value of the average marginal effect.

For the purpose of comparison, to each sample we also applied a consistent nonlinear 2SRI method based on the IBC regression specification in equation (1).[11,12] Using the 2SRI-IBC results, we consistently estimated AME using the following sample analog to equation (2)

$$\widehat{\mathrm{AME}} = \sum_{i=1}^{N_j} \frac{1}{N_j} \frac{\partial k(x_{\mathrm{p}}^{*}\hat{\beta}_{\mathrm{p}} + x_{\mathrm{o}i}\hat{\beta}_{\mathrm{o}} + \hat{x}_{\mathrm{u}i}\hat{\beta}_{\mathrm{u}}, \gamma = 1.765)}{\partial x_{\mathrm{p}}} \tag{4}$$

where the N_j denotes the jth sample size (N_j = 1,000; 5,000; 10,000; 50,000; 100,000; and 500,000), the "^s" denote the 2SRI estimates, and $\hat{x}_{\mathrm{u}i}$ is consistent estimate of x_{u} obtained as a byproduct of 2SRI estimation. Note that both equations (2) and (4) require that x_{p} be fixed at a known (policy relevant) value (x_{p}^{*}). As in the case of the IV estimates, for each sample size, we computed the average of the percentage absolute bias for $\widehat{\mathrm{AME}}$, which is measured as[13]

$$\frac{\mathrm{abs}(\widehat{\mathrm{AME}} - \mathrm{AME})}{\mathrm{abs}(\mathrm{AME})} \times 100 \text{ percent} \tag{5}$$

To investigate variation in IV bias across the range of the policy variable, for each sample size, we computed equations (3) and (5) at each of the quartiles of the simulated distribution of x_{p}. The results in Table 1 demonstrate that for all three quartiles of x_{p}, substantial bias can result from inappropriate application of the conventional linear IV estimator to an inherently nonlinear model. Moreover, the bias is not attenuated as the sample size increases. The fact that the values of equation (5) in the third, fifth, and seventh columns of Table 1 decrease monotonically as the sample size increases, supports the theoretical consistency (large-sample unbiasedness) of the 2SRI-IBC estimator.

Binary y and Binary x_p

For the case in which y and x_{p} are both binary, we conducted a similar simulation analysis using the following regression specification to generate the

Table 1: Average Percent Absolute Bias (γ = 1.765)

	x_{p}^{*} = 1st Quartile of x_p		x_{p}^{*} = 2nd Quartile of x_p		x_{p}^{*} = 3rd Quartile of x_p	
γ	*IV (%)*	*IBC (%)*	*IV (%)*	*IBC (%)*	*IV (%)*	*IBC (%)*
1,000	48.84	7.16	49.14	7.04	53.10	6.93
5,000	47.39	2.99	47.70	2.93	51.77	2.86
10,000	48.61	2.05	48.91	2.00	52.88	1.97
50,000	49.36	0.98	49.65	0.95	53.57	0.93
100,000	49.37	0.69	49.67	0.68	53.58	0.66
500,000	49.01	0.30	49.30	0.29	53.25	0.29

IV, instrumental variables; IBC, inverse Box–Cox.

data on y

$$E[y|x_p, x_o, x_u] = \Phi(x_p\lambda_p + x_o\lambda_o + x_u\lambda_u) \tag{6}$$

where $\Phi(\bullet)$ denotes the standard normal distribution function, the λs are parameters to be estimated, x_o and x_u represent observable and unobservable confounders, respectively; and x_p is binary and generated via the following probit process

$$x_p = I(w\alpha + x_u > 0) \tag{7}$$

in which w is a vector of instrumental variables, α is a vector of parameters to be estimated, and x_u is assumed to be standard normally distributed. In simulating the data, we designated the *average treatment effect* (ATE)

$$E\left[\Phi(\lambda_p + x_o\lambda_o + x_u\lambda_u) - \Phi(x_o\lambda_o + x_u\lambda_u)\right] \tag{8}$$

as both the estimation objective and the basis for evaluation of estimator performance. The sample sizes and number of replications were the same as for the nonnegative outcome simulations discussed in the previous subsection. Here again, we applied the conventional linear IV estimator to each of the generated samples. Similar to the previous simulation study, the IV estimate of the ATE is the IV estimated coefficient of that variable x_p. For each sample size, we computed the mean percentage absolute difference between the IV estimated coefficient of x_p ($\widetilde{\text{IV}}$)and the true value of the ATE, which is measured as

$$\frac{\text{abs}\left(\widetilde{\text{IV}} - \text{ATE}\right)}{\text{abs}(\text{ATE})} \times 100 \text{ percent} \tag{9}$$

where ATE denotes the true value of the average treatment effect.

For each sample, we also estimated the parameters of equations (6) and (7) via BVP analysis.[14] Using the BVP results, we consistently estimated the ATE using the following sample analog to equation (8):

$$\widetilde{\text{ATE}} \sum_{i=1}^{N_j} \frac{1}{N_j} \int_{-\infty}^{\infty} \left\{\Phi(\tilde{\lambda}_p + x_{oi}\tilde{\lambda}_o + x_u\tilde{\lambda}_u) - \Phi(x_{oi}\tilde{\lambda}_o + x_u\tilde{\lambda}_u)\right\}\varphi(x_u)\,\mathrm{d}x_u \tag{10}$$

where $\varphi(\bullet)$ denotes the standard normal density function, the "$\sim$ s" denote the BVP estimates, and N_j is defined as in equation (4).[15] We computed the average of the percentage absolute bias for $\widetilde{\text{ATE}}$ as[16]

$$\frac{\text{abs}\left(\widetilde{\text{ATE}} - \text{ATE}\right)}{\text{abs}(\text{ATE})} \times 100 \text{ percent} \tag{11}$$

The results in Table 2 are similar in implication to those of Table 1. The IV estimate of ATE is biased and remains so as the sample size increases, while the bias in the BVP estimator monotonically decreases as the sample size increases. This latter result reflects the theoretical consistency of the BVP estimator.

Bhattacharya, Goldman, and McCaffrey (2006), as part of their simulation study, examine IV bias when y and x_p are binary. Overall, their results support the use of BVP over IV in this type of model. They show that IV performs poorly relative to BVP: (1) in their empirical example; (2) if the observations on y are substantially imbalanced toward 0 or 1 [i.e., $p(y=1)$ is $<.16$ or $>.84$]; and (3) when the joint probability distribution underlying the model is not bivariate normal. The only aspect of their analysis in which the IV method appears to hold its own relative to BVP is when the true model has a BVP structure like in the one specified in equations (6) and (7) and $p(y=1)$ is fixed at .5, although even in this case they find that BVP is still less biased than IV. They find that, although IV bias is larger than that of BVP, the difference is not substantial. This is not at odds with the results displayed in Table 2. In their Monte Carlo study, Bhattacharya, Goldman, and McCaffrey (2006) generate samples of only one size ($N=5{,}000$) and do not investigate estimator performance as the sample size increases. Instead, they focus on differences in bias as the true treatment effect increases. In the context of our simulation study, a sample size of 5,000 is relatively small and, as can be seen in Table 2, for the smaller sample sizes we too find the bias differences between IV and BVP to be unremarkable. As Table 2 demonstrates, however, the bias problem with the use of the IV method (relative to the nonlinear method) in this context clearly emerges as the sample size is increased—the difference between IV and BVP bias widens substantially as the BVP estimate approaches the true value and the IV estimate remains virtually unchanged.

Table 2: Average Percent Absolute Bias IV versus Bivariate Probit (BVP)

Sample Size	*IV (%)*	*BVP (%)*
1,000	16.13	15.00
5,000	10.65	6.89
10,000	11.75	4.75
50,000	11.30	2.13
100,000	11.17	1.57
500,000	11.40	0.70

IV, instrumental variables.

Illustrative Examples

The Effect of Habit Stock on Cigarette Demand: Nonnegative y and Continuous x_p

To further elucidate the case involving a nonnegative outcome and a continuous endogenous regressor, we reestimated Mullahy's (1997) model of cigarette demand using data supplied by the author (a sample of 6,160 men originally taken from the Smoking Supplement to the 1979 National Health Interview Survey). We focus on estimating the effect of habit stock (x_{p}) on the individual's current daily cigarette consumption (y). Habit stock is a measure of the accumulated effects of past smoking on present consumption.[17] Cigarette consumption is measured in packs (20 cigarettes). For this analysis, we implemented the IBC regression model as defined in equation (1) with x_{o} as a vector of observable confounders, x_{u} representing the unobservable confounders, the βs as the regression coefficients, and γ a freely varying unknown parameter ($-\infty < \gamma < \infty$). The specification of the outcome equation in this IBC framework encompasses that of Mullahy (1997). He estimated an exponential regression model of the demand for cigarettes. The interesting aspect of Mullahy's study is that he cleverly devised and implemented a generalized method of moments (GMM) estimator that does not require explicit regression modeling of endogenous regressor.[18] Although Mullahy's GMM approach is desirable because it can be implemented based on a weaker set of structural assumptions, Terza (2006) shows that it does not extend to the generic nonlinear framework considered in the present paper.

We first consistently estimated the parameters of the model, including γ, using the 2SRI method. We used the same variable specifications as Mullahy (1997).[19] For the purpose of comparison, we applied the conventional linear IV method to the following model

$$y = \delta_{\mathrm{p}} x_{\mathrm{p}} + \delta_{\mathrm{o}} x_{\mathrm{o}} + e \tag{12}$$

where the δs are the parameters to be estimated and e is the regression error term.[20] A few aspects of the estimation results are noteworthy. First the linear specification is rejected in the IBC context. Specifically, the null hypothesis that $\gamma = 2$ is rejected based on the typical Wald test of the estimated value of that parameter. Second, at the estimated value of γ, nearly 33 percent of the estimated values of $x_{\mathrm{p}}\beta_{\mathrm{p}} + x_{\mathrm{o}}\beta_{\mathrm{o}} + x_{\mathrm{u}}\beta_{\mathrm{u}}$ are < -1. Also, based on a comparison of the sum of squared residuals from both models, the nonlinear IBC model provides a better fit than the linear IV specification. Finally, note that the exogeneity of x_{p} is rejected at the .01 significance level (i.e., the t-statistic for the null hypothesis that $\beta_{\mathrm{u}} = 0$ is equal to -3.0062).

To characterize the divergence between the IBC and IV estimates in the present context, we computed the corresponding difference in the predicted change in the current number of packs of cigarettes smoked per year that would result from exogenously decreasing x_{p} (habit stock) to zero starting from its upper quartile value.[21] In other words, we set x_{p}^{*} at its upper quartile (178.3) value and then exogenously decreased its value to 0. This counterfactual thought experiment asks "What if all individuals in the population had a smoking habit stock equal to the upper quartile?" Starting from this hypothetical scenario, how much less would they smoke currently (on average) if habit stock were exogenously reduced to zero? The answer to this question will reveal the level of current smoking that is exclusively attributable to a given amount of past smoking.

To place a relevant policy framework on this, we have to ask what options are available to convert current smokers' habit stocks to (effectively) zero? Taking Mullahy's habit stock assumptions as a baseline (around 12 cigarettes smoked per day, with a 10 percent per day depreciation in the stock), an individual who was exogenously forced to stop smoking for approximately 90 days would have a habit stock of zero (to three significant digits). Pregnancy is one opportunity for such short-term cessation (see Bradford 2003), where a significant fraction of pregnant women who are smokers stop "cold turkey" for at least the final two-thirds of their pregnancy. Interventions on this population toward the end of their pregnancies could take advantage of the temporary absence of the habit stock pressure to smoke. Recent research by Volpp et al. (2006) suggests 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—and potentially highly cost effective.

Using our parameter estimates from Mullahy's model, at the upper quartile of the empirical distribution of habit stock, IV predicts a 409 pack decrease in yearly smoking when habit stock is counterfactually decreased to zero, whereas the IBC projected reduction is 437 packs—IV falls short of IBC by 28 packs. This difference is substantial, amounting to nearly a month's worth of smoking reduction for an individual who smokes a pack-a-day.

The Effect of Substance Abuse on Employment Status: Binary y and Binary x_p

To compare the performance of the IV method with that of BVP in a real-world context, we estimated a binary version of the model of the effect of alcohol abuse on employment status first examined by Mullahy and Sindelar (1996) and later revisited by Terza (2002). In this case, we have

$$y = \begin{array}{ll} 1 & \text{if the individual is employed full-time} \\ 0 & \text{otherwise} \end{array}$$

and

$$x_{\mathrm{p}} = \begin{array}{ll} 1 & \text{if the individual is a substance abuser} \\ 0 & \text{otherwise.} \end{array}$$

The data used in this example were taken from the National Longitudinal Alcohol Epidemiologic Survey 1992. The size of the sample is 22,107. The definitions of the variables and the specifications of x_{o} and w closely mimic those used by Mullahy and Sindelar (1996) and Terza (2002).[22] We estimated the model defined in equations (6) and (7) using BVP. The ATE as defined in equation (8) was then estimated using the BVP results and equation (10). The IV-based ATE estimate was obtained as the coefficient of δ_{p} in the relevant version of equation (12).[23] The results yield -0.16 and -0.31 as the BVP and IV estimates of the ATE, respectively. This difference is substantial (IV is 48 percent smaller than BVP) but even more importantly, the former is statistically insignificant at any reasonable level (t-stat $= -1.28$) while the p-value of the latter is $<.0001$ (t-value $= -4.02$). Note that the exogeneity of x_{p} is rejected at the .01 significance level (i.e., the t-statistic for the null hypothesis that $\beta_{\mathrm{u}} = 0$ is equal to 3.54). These results are similar to those obtained by Bhattacharya, Goldman, and McCaffrey (2006) in their illustrative example. They find that the IV estimate (-0.018) substantially underestimates the treatment effect "… (if the BVP estimate is taken as consistent)" (BVP $= -0.079$, no t-stats are given).

Discussion

In this study, we show that applying the conventional linear IV method when the true data generating process is nonlinear, may lead to substantial estimation bias. We examined two commonly encountered cases in empirical health services and health economics research: models with a nonnegative continuous outcome and a continuous endogenous regressor (the nonnegative/continuous case); and models with a binary outcome and a binary endogenous regressor (the binary/binary case). For the former, we used the AME as our basis for comparison, and generated data via a flexible-form nonlinear regression specification. We found that the mean absolute percentage bias from using the linear IV method to estimate the AME can be large and is not attenuated as the sample size increases. Moreover, using real-world data on the determinants of cigarette demand, we obtained substantially divergent results using linear IV versus a flexible-form estimator that takes account of the inherent nonlinearity of the model (the 2SRI-IBC estimator). In the binary/binary context, we used the ATE as the basis for comparison and simulated data using a BVP sampling design. Here again, we found that the use of the IV method can lead to substantial bias, measured in mean absolute percentage terms, especially in larger samples. As an illustrative real-world example, we estimated the effect of substance abuse on the likelihood of being employed and found the IV-based estimate of the ATE to be much smaller than that obtained via BVP analysis. These results should be taken as a cautionary note by those who would opt for the use of linear specifications and methods in inherently nonlinear settings involving endogeneity.

Acknowledgments

This work was supported by the National Institute on Drug Abuse (R01 DA013968–02) and the Substance Abuse Policy Research Program of the Robert Wood Johnson Foundation (53902). The authors are grateful for the excellent research assistance provided by F. Michael Kunz.

Disclosures: NA.

Disclaimers: NA.

Notes

1. For the purposes of this paper, we define the *causal effect of* x_{p} on y to be the direct effect of x_{p} on y, aside from all indirect or mediator effects.
2. *Habit stock*, as defined by Mullahy (1985, 1997), is an index value of depreciated smoking levels, where smoking (in terms of cigarettes per day) is depreciated at a rate of 10 percent per day, and where smokers are assumed to consume a constant level during the time they do smoke. See online Appendix B for the strict definition of the variant of the habit stock variable used by Mullahy (1997) and here.
3. This problem is known by other names in the literature—e.g., omitted variables bias, confounding, spurious correlation.
4. See Greene (2003), Chapter 5, or any other modern econometrics text for details of the IV method.
5. Nonlinearity, as we use the term in this paper, is defined as nonlinearity in the parameters. Common examples are of the form $y = M(x\pi) + v$ where x is a vector of regressors, π a vector of coefficient parameters, v a random error, and $M(\bullet)$ is a

known nonlinear function. This particular regression framework subsumes all conventionally specified qualitative and limited dependent variable models—e.g., probit, logit, tobit, Poisson, etc. The IV method is indeed appropriate for models that are linear in the parameters even though they may be nonlinear in the variables.

6. In cases involving nonnegative outcomes and endogenous regressors, the log transformation is often used to linearize the model (e.g., Cawley 2004). In this case, retransformation or smearing (Duan 1983) would be required for estimation of the marginal and incremental effects of the regressors. Retransformation is susceptible to bias because it necessitates an assumed specification for the conditional distribution of the regression error term. The properties of the smearing estimator in the presence of endogenous regressors are unknown.
7. We note that Gowisankaran and Town later teamed with Geweke and reestimated their model in an appropriately designed nonlinear Bayesian framework (see Geweke, Gowrisankaran, and Town 2003).
8. The reference to Ashford and Sowden (1970) encompasses the use of the bivariate probit model as a means of consistently estimating a probit model with an endogenous treatment effect.
9. When $\gamma = 2$, equation (1) becomes $E[y|x_p, x_o, x_u] = g(z) = |z + 1|$, where $|a|$ denotes the absolute value of a and $z = x_p\beta_p + x_o\beta_o + x_u\beta_u$. 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, equation (1) becomes the simple linear regression model.
10. We chose this particular value of γ because it is the estimated value of that parameter in our illustrative cigarette demand application (discussed later).
11. In order to gain computational speed in the simulations, we applied the 2SRI estimator to each sample with the value of γ held fixed at the known value.
12. The 2SRI method has been shown to be consistent (large-sample unbiased) by Terza, Basu, and Rathouz (2007).
13. The details of expressions (2)–(5), the simulation design, and the 2SRI method can be found in online Appendix A.
14. Terza and Tsai (2006) show that the model defined in equations (6) and (7) is identical to a bivariate probit model under a specific reparameterization of the model.
15. We used quadrature to approximate the required integral in equation (10)—specifically, the INTQUAD1 procedure in the GAUSS® programming language.
16. The details of expressions (6)–(11), the simulation design, and the BVP method can be found in online Appendix B.
17. For the strict definition of the habit stock variable see online Appendix C.
18. See Hansen (1982) for a detailed explanation of GMM estimation.
19. The definitions of the variables included in the regression specification can be found in online Appendix C.
20. For details of the 2SRI and IV modeling, including estimation results see online Appendix C.
21. We chose the upper quartile because there are no cigarette smokers in the sample at the lower quartile and median values of the habit stock variable.

22. Definitions of the variables and details of the specifications of x_o and w^+ can be found in online Appendix D.
23. For details of the BVP and IV modeling, including estimation results see online Appendix D.

REFERENCES

Alexandre, P. K., and M. T. French. 2001. "Labor Supply of Poor Residents in Metropolitan Miami, Florida: The Role of Depression and the Co-Morbid Effects of Substance Abuse." *Journal of Mental Health Policy Economics* 4: 161–73.

Alvarez, R. M., and G. Glasgow. 1999. "Two-Stage Estimation of Nonrecursive Choice Models." *Political Analysis* 8: 147–65.

Ashford, J., and R. Sowden. 1970. "Multivariate Probit Analysis." *Biometrics* 26: 535–46.

Averett, S., B. Duncan, D. Rees, and L. Argys. 2004. "Race, Ethnicity, and Gender Differences in the Relationship between Substance Use and Adolescent Sexual Behavior." *Topics in Economic Analysis and Policy* 4: 1–32.

Bao, Y., N. Duan, and S. A. Fox. 2006. "Is Some Provider Advice on Smoking Cessation Better Than No Advice? An Instrumental Variable Analysis of the 2001 Health Interview Survey." *Health Services Research* 41: 2114–35.

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

Basur, O., C. J. Bradley, J. C. Gardiner, and C. Given. 2004. "Testing and Correcting for Non-Random Selection Bias Due to Censoring: An Application to Medical Costs." *Health Services and Outcomes Research Methodology* 4: 93–107.

Beegle, K., R. Dehejia, and R. Gatti. 2005. "Why Should We Care about Child Labor? The Education, Labor Market, and Health Consequences of Child Labor." World Bank Policy Research Working Paper #3479.

Bhattacharya, J., D. Goldman, and D. McCaffrey. 2006. "Estimating Probit Models with Self-Selected Treatments." *Statistics in Medicine* 25: 389–413.

Blundell, R. W., and R. J. Smith. 1989. "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models." *Review of Economics and Statistics* 56: 37–58.

———. 1993. "Simultaneous Microeconometric Models with Censored or Qualitative Dependent Variables." In *Handbook of Statistics*, Vol. 2, edited by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 1117–43. Amsterdam: North Holland Publishers.

Bollen, K. A., D. K. Guilkey, and T. A. Mroz. 1995. "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia." *Demography* 32: 11–131.

Box, G. E. P., and D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society-Series B* 26: 211–52.

Bradford, W. D. 2003. "Pregnancy and the Demand for Cigarettes." *American Economic Review* 93: 1752–63.

Brooks, J. M., E. A. Chrischilles, S. D. Scott, and S. S. Chen-Hardee. 2003. "Was Breast Conserving Surgery Underutilized for Early Stage Breast Cancer? Instrumental

Variables Evidence for Stage II Patients from Iowa." *Health Services Research* 38: 1385–402.

Brooks, J. M., C. P. Irwing, L. G. Hunsicker, M. J. Flanigan, E. A. Chrischilles, and J. E. Pendergast. 2006. "Effect of Dialysis Center Profit-Status on Patient Survival: A Comparison of Risk-Adjustment and Instrumental Variables Approaches." *Health Services Research* 41: 2267–89.

Burnett, N. J. 1997. "Gender Economics Courses in Liberal Arts Colleges." *Journal of Economic Education* 28: 369–77.

Cawley, J. 2004. "The Impact of Obesity on Wages." *Journal of Human Resources* 39: 452–74.

DeSimone, J. 2002. "Illegal Drug Use and Employment." *Journal of Labor Economics* 20: 952–77.

Dickie, M. 2005. "Parental Behavior and the Value of Children's Health: A Health Production Approach." *Southern Economic Journal* 71: 855–72.

Duan, N. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78: 605–10.

Frances, C. D., M. G. Shlipak, H. Noguchi, P. A. Heidenriech, and M. McClellan. 2000. "Does Physician Specialty Affect the Survival of Elderly Patients with Myocardial Infarction." *Health Services Research* 35: 1093–116.

Geweke, J., G. Gowrisankaran, and R. Town. 2003. "Bayesian Inference for Hospital Quality in a Selection Model." *Econometrica* 71: 1215–38.

Gibson, T. B., T. L. Mark, K. Axelsen, O. Basur, D. A. Rublee, and K. A. McGuigan. 2006. "Impact of Statin Copayments on Adherence and Medical Care Utilization and Expenditures." *American Journal of Managed Care* 12: SP11–19.

Gowrisankaran, G., and R. Town. 1999. "Estimating the Quality of Care in Hospitals Using Instrumental Variables." *Journal of Health Economics* 18: 747–67.

Grabowski, D., and R. Hirth. 2003. "Competitive Spillovers across Non-Profit and For-Profit Nursing Homes." *Journal of Health Economics* 22: 1–22.

Greene, W. H. 2003. *Econometric Analysis*, 5th Edition. New Jersey: Prentice Hall.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50: 1029–54.

Kenkel, D., and J. Terza. 2001. "The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect." *Journal of Applied Econometrics* 16: 165–84.

Koc, C. 2005. "Health-Specific Moral Hazard Effects." *Southern Economic Journal* 72: 98–118.

Lee, L. 1979. "Identification and Estimation in Binary Choice Models with Limited 'Censored' Dependent Variables." *Econometrica* 47: 977–95.

———. 1981. "Simultaneous Equations Models with Discrete and Censored Dependent Variables," Chapter 9. In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden, pp. 346–64. Cambridge, MA: MIT Press.

Lo Sasso, A. T., and T. C. Buchmeuller. 2004. "The Effect of the State Children's Health Insurance Program on Health Insurance Coverage." *Journal of Health Economics* 23: 1059–82.

Malkin, J. D., M. S. Broder, and E. Keeler. 2000. "Do Longer Postpartum Stays Reduce Newborn Readmissions? Analysis Using Instrumental Variables." *Health Services Research* 35: 1135–58.

McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality: Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272: 859–66.

McGarrity, J. P., and D. Sutter. 2000. "A Test of the Structure of PAC Contracts: An Analysis of House Gun Control Votes in the 1980s." *Southern Economic Journal* 677: 41–63.

Moran, J., and K. Simon. 2006. "Income and the Use of Prescription Drugs by the Elderly: Evidence from The Notch Cohorts." *Journal of Human Resources* 41: 411–32.

Mroz, T. A., K. A. Bollen, I. S. Speizer, and D. J. Mancini. 1999. "Quality, Accessibility, and Contraceptive Use in Rural Tanzania." *Demography* 36: 23–40.

Mullahy, J. 1985. "Cigarette Smoking: Habits, Health Concerns, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand." Unpublished 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: 586–93.

Mullahy, J., and J. Sindelar. 1996. "Employment, Unemployment, and Problem Drinking." *Journal of Health Economics* 15: 409–34.

Neslusan, C. A., T. R. Hylan, R. L. Dunn, and J. Donoghue. 1999. "Controlling for Systematic Selection in Retrospective Analyses: An Application to Fluoxetine and Sertraline Prescribing in the United Kingdom." *Value in Health* 2: 435–45.

Newey, W. K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36: 231–50.

Norton, E. C., R. C. Lindrooth, and S. T. Ennett. 1998. "Controlling for the Endogeneity of Peer Substance Use on Adolescent Alcohol and Tobacco Use." *Health Economics* 7: 439–53.

Norton, E. C., and C. H. Van Houtven. 2006. "Intervivos Transfers and Exchange." *Southern Economic Journal* 73: 157–72.

Petrin, A., and K. Train. 2006. "Control Function Corrections for Unobserved Factors in Differentiated Product Models," Working Paper.

Rees, D., L. Argys, and S. Averett. 2001. "New Evidence on the Relationship between Substance Use and Adolescent Sexual Behavior." *Journal of Health Economics* 20: 835–45.

Ribar, D. C. 1994. "Teenage Fertility and High School Completion." *Review of Economics and Statistics* 76: 413–24.

Rivers, D., and Q. H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39: 347–66.

Sen, B. 2002. "Does Alcohol-Use Increase the Risk of Sexual Intercourse among Adolescents? Evidence from the NLSY97." *Journal of Health Economics* 21: 1085–93.

Shea, D. G., J. V. Terza, B. C. Stuart, and B. Briesacher. 2007. "Estimating the Effects of Prescription Drug Coverage for Medicare Beneficiaries." *Health Services Research* 43: 933–49.

Smith, R. J., and R. W. Blundell. 1986. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." *Econometrica* 54: 679–85.

Stone, P. W., C. Mooney-Kane, E. L. Larson, D. K. Pastor, J. Zwanziger, and A. W. Dick. 2006. "Nurse Working Conditions, Organizational Climate, and Intent to Leave in ICUs: An Instrumental Variable Approach." *Health Services Research*, forthcoming.

Stuart, B. C., J. A. Doshi, and J. V. Terza. 2007. "A New Approach to Assessing the Impact of Outpatient Drug Use on Hospital Spending in the Face of Indication Bias," Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Terza, J. V. 1994. "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: 129–54.

———. 2002. "Alcohol Abuse and Employment: A Second Look." *Journal of Applied Econometrics* 17: 393–404.

———. 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: 177–98.

Terza, J. V., A. Basu, and P. Rathouz. 2007. "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics*, forthcoming.

Terza, J. V., and W. Tsai. 2006. "Censored Probit Estimation with Correlation Near the Boundary: A Useful Reparameterization." *Review of Applied Economics* 2: 1–12.

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

van Houtven, C. H., and E. C. Norton. 2007. "Informal Care and Medicare Expenditures: Testing for Heterogeneous Treatment Effects." *Journal of Health Economics*, forthcoming.

Volpp, K. G., 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.

Wooldridge, J. M. 1992. "Some Alternatives to the Box–Cox Regression Model." *International Economic Review* 33: 935–55.

———. 1997. "Quasi-Likelihood Methods for Count Data." In *Handbook of Applied Econometrics, Volume II: Microeconometrics*, edited by M. Pesaran and P. Schmidt. Malden, MA: Blackwell Publishers Ltd.

———. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:

Appendix A. Details of the Simulation Study for Nonnegative y and Continuous x_p.

Appendix B. Details of the Simulation Study for Binary y and Binary x_p.

Appendix C. Details of the Illustrative Example: The Effect of Habit Stock on Cigarette Demand—Nonnegative y and Continuous x_p.

Appendix D. Details of the Illustrative Example: The Effect of Substance Abuse on Employment Status: Binary y and Binary x_p.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2007.00807.x (this link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.