Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling

Joseph V. Terza [a,*], Anirban Basu [b,1], Paul J. Rathouz [c,2]

[a] *Department of Epidemiology and Health Policy Research and Department of Economics, University of Florida, 1329 SW, 16th Street, Room 5130, Box 100147,Gainesville, FL 32610-0147, United States*
[b] *Section of General Internal Medicine, Department of Medicine, Center for Health and the Social Sciences, University of Chicago, 5841 South Maryland Avenue, MC-2007, AMD B226C, Chicago, IL 60637, United States*
[c] *Department of Health Studies, University of Chicago, 5841 South Maryland Avenue, MC-2007, Room W-264, Chicago, IL 60637, United States*

Abstract

The paper focuses on two estimation methods that have been widely used to address endogeneity in empirical research in health economics and health services research—*two-stage predictor substitution* (2SPS) and *two-stage residual inclusion* (2SRI). 2SPS is the rote extension (to nonlinear models) of the popular linear two-stage least squares estimator. The 2SRI estimator is similar except that in the second-stage regression, the endogenous variables are not replaced by first-stage predictors. Instead, first-stage residuals are included as additional regressors. In a generic parametric framework, we show that 2SRI is consistent and 2SPS is not. Results from a simulation study and an illustrative example also recommend against 2SPS and favor 2SRI. Our findings are important given that there are many prominent examples of the application of inconsistent 2SPS in the recent literature. This study can be used as a guide by future researchers in health economics who are confronted with endogeneity in their empirical work.
© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Endogeneity of regression predictors is a common problem in many areas of applied economics, including health economics and health services research, as these fields rely heavily on observational data. Endogeneity arises owing to problems such as omitted confounder variables, simultaneity between a predictor and the outcome, and errors in regression covariates. Instrumental variables (IV) methods form a common body of approaches to handing such endogeneity. The theoretical and methodological literature guiding the use of IVs in linear regression models is large and serves as the basis for most practitioners' understanding of the assumptions and implementation of IV models. In many

* Corresponding author. Tel.: +1 352 265 0111x85068; fax: +1 352 265 7221.
  E-mail addresses: jvt@ichp.ufl.edu (J.V. Terza), abasu@medicine.bsd.uchicago.edu (A. Basu), prathouz@uchicago.edu (P.J. Rathouz).
[1] Tel.: +1 773 834 1796; fax: +1 773 834 2238.
[2] Tel.: +1 773 834 1970; fax: +1 773 702 1979.

health economics and health services research problems, however, linear regression models are now being replaced by nonlinear regression models, including generalized linear models, as these models are often more appropriate for limited-dependent variables, count variables and skewed distributions such as healthcare costs.

Despite the growing use and appreciation of nonlinear models among empirical researchers in health economics and health services research, there appears to be some confusion surrounding the applications of IV methods in the context of these models. The goal of the present paper is to address this concern while at the same time unifying earlier results under a common nonlinear modeling framework. We carefully examine two instrumental variables (IV)-based approaches to correcting for endogeneity bias in nonlinear models – two-stage residual inclusion (2SRI) and two-stage predictor substitution (2SPS)—focusing especially on a class of nonlinear models that have been widely exploited in empirical health economics and health services research. We show the consistency of the 2SRI estimator in this class of models and reemphasize the inconsistency of the alternative 2SPS approach. Our goal is to demonstrate the superiority of the 2SRI method, to guide applied researchers in carrying out 2SRI estimation when they are trying to address endogeneity in nonlinear models, and to help them understand why they should steer away from the popular 2SPS approach.

2SPS is the rote extension to nonlinear models of the popular linear two-stage least squares (2SLS) estimator. In the first-stage of 2SPS, auxiliary (reduced form) regressions are estimated, and the results are used to generate predicted values for the endogenous variables. The second-stage regression is then conducted for the outcome equation of interest after replacing the endogenous variables with their predicted values. The 2SRI estimator has the same first-stage as 2SPS. In the second-stage regression, however, the endogenous variables are not replaced. Instead, the first-stage residuals are included as additional regressors in second-stage estimation. This method was first suggested by Hausman (1978) in the linear context as a means of testing for endogeneity. We focus on these two methods because both have been applied in empirical studies in health economics and health services research. Indeed, these models can be easily implemented using any modern statistical software package.

We begin, in the next section with detailed descriptions of the methods within a unified modeling framework. This framework extends the two-stage least squares (2SLS) linear modeling approach for instrumental variables to nonlinear outcome and/or auxiliary models, encompassing many parametric nonlinear models that are commonly used in empirical health economics and health services research. The statistical properties of the two alternative methods are more formally examined in Section 3. There, we note that although the two methods produce identical results in the fully linear model (a special case of the broader class of models we consider), they do not coincide in the generic nonlinear model. Moreover, we show why 2SRI is generally statistically consistent in this broader class, but 2SPS is not. In Section 4, we compare the methods using simulated data in the context of two interesting nonlinear models involving endogenous regressors. The results reflect the theoretical consistency of 2SRI and the lack thereof for 2SPS. Further comparisons are drawn between the methods in Section 5, wherein we re-estimate Mullahy's (1997) exponential regression model of the effect of prenatal smoking on birthweight using a more flexible functional form. The 2SPS and 2SRI estimates differ substantially. The final section summarizes and concludes. The theoretical consistency of 2SRI, the results of the simulation analyses, and the findings from re-estimation of Mullahy's (1997) model all support the use of 2SRI over 2SPS.

2. The modeling framework and estimation

2.1. The model

We employ the following nonlinear modeling framework. The main, and minimal, assumption of the model is that the conditional mean of the outcome (y) is of the form:

$$E[y|x_e, x_o, x_u] = M(x_e\beta_e + x_o\beta_o + x_u\beta_u) \qquad (1)$$

where $M(\cdot)$ is a known nonlinear function, and we distinguish among three types of regressor: $x_e = [x_{e1} \quad x_{e2} \quad \cdots \quad x_{eS}]$ denotes a $1 \times S$ vector of endogenous regressors; $x_o = [x_{o1} \quad x_{o2} \quad \cdots \quad x_{oK}]$ is a $1 \times K$ vector of observable exogenous regressors (*observable confounders*); and $x_u = [x_{u1} \quad x_{u2} \quad \cdots \quad x_{uS}]$ is a $1 \times S$ vector of *unobservable confounder* latent variables (omitted variables) that influence the outcome y and are correlated with the endogenous variables. Correspondingly, β_e and β_u are $S \times 1$ vectors, and β_o is a $K \times 1$ vector of unknown regression parameters. Letting $\beta' = [\beta'_e \quad \beta'_o \quad \beta'_u]$ be the corresponding column vector, the regression model corresponding to

(1) is

$$y = M(x_e\beta_e + x_o\beta_o + x_u\beta_u) + e \tag{2}$$

where e is the random error e, tautologically defined as $e = y - M(x_e\beta_e + x_o\beta_o + x_u\beta_u)$ so that $E[e|x_e, x_o, x_u] = 0$.

At issue here is the correlation between x_e and x_u—this is the essence of the endogeneity problem. To formalize the relationship between x_e and x_u, and thereby provide a means for dealing with endogeneity bias through the use of instrumental variables (IV), we define the following set of (possibly) nonlinear auxiliary (or reduced form) equations:

$$x_{es} = r_s(w\alpha_s) + x_{us} \quad \text{for } s = 1, \ldots, S \tag{3}$$

where $w = [x_o \quad w^+]$, $w^+ = [w_1^+, \ldots, w_{S^+}^+]$ is a $1 \times S^+$ vector of identifying IV, and α_s is a $(K + S^+) \times 1$ column vector of parameters. The elements of w^+ must satisfy the following three conditions: (1) they cannot be correlated with x_u; (2) they must be sufficiently correlated with x_e (i.e. they must not be "weak"); and (3) they can neither have a direct influence on y nor be correlated with the error term in (2). Also there must be at least as many elements in w^+ as there are endogenous regressors in (2) (i.e. $S^+ \geq S$).[3]

Theoretical and applied econometric studies that implement the nonlinear model specification (2) abound in the health economics and health services research literatures.[4] In the remainder of this section, we examine two alternative ways to estimate the parameters of the model in (2) and (3): *two-stage predictor substitution* (*2SPS*) *and two-stage residual inclusion* (*2SRI*).[5] Both methods extend the familiar linear two-stage least squares (2SLS) method to the nonlinear model defined in (2) and (3), and both have been applied in the health economics and health services research literature.

2.2. Two-stage predictor substitution

The two-stage predictor substitution (2SPS) method is straightforward and simple to implement in the context of the model defined by (2) and (3). In the first-stage, obtain consistent estimates of the vectors α_s ($\hat{\alpha}_s$) by applying the nonlinear least squares method (NLS) or any other consistent estimation technique to auxiliary Eq. (3). Next, compute the "predictors" of $x_{e1}, \ldots, x_{eS}$ as

$$\hat{x}_{es} = r_s(w\hat{\alpha}_s) \quad \text{for } s = 1, \ldots, S \tag{4}$$

Correspondingly, define the "residuals" in these models as

$$\hat{x}_{us} = x_{es} - r_s(w\hat{\alpha}_s) \quad \text{for } s = 1, \ldots, S \tag{5}$$

In the second-stage estimate $\gamma' = [\gamma_e' \quad \gamma_o']$ apply NLS to

$$y = M(x_o\gamma_o + \hat{x}_e\gamma_e) + e^{2SPS} \tag{6}$$

where $\hat{x}_e = [\hat{x}_{e1} \hat{x}_{e2}, \ldots, \hat{x}_{eS}]$, γ_e is $S \times 1$, γ_o is $K \times 1$, and e^{2SPS} denotes the regression error term.

Consider, for instance, the linear case in which (1) becomes:

$$E[y|x] = x_e\beta_e + x_o\beta_o + x_u\beta_u \tag{7}$$

---

[3] The model can be generalized in a number of ways. First, it can be extended to allow: the outcome (*y*) to be vector-valued; multiple regression indexes (i.e. $x\beta$ is a vector); and multiple auxiliary regression indexes (i.e. $w\alpha$ is a vector). Secondly, this nonlinear framework constitutes the most parametrically parsimonious specification of the regression of *y* on *x*—i.e. only the key conditional mean regressions are specified ((2) and (3)). The model can, of course, be more fully specified by positing higher-order conditional moment restrictions (heteroskedasticity, skewness, kurtosis, etc.). Indeed, the model can be made fully parametric by specifying the joint probability density function of (*y*|*x*) and (*x_e*|*w*). A more general version of the model, incorporating these features, can be found in an appendix that will be supplied upon request. All results presented here for the simple model in Eqs. (2) and (3) are valid for, and can easily be extended to, the general model.

[4] See Terza (1994a,b), Coulson et al. (1995), Neslusan et al. (1999), Treglia et al. (1999), McGeary and French (2000), Kenkel and Terza (2001), Terza (1998,1999,2002), DeSimone (2002), Pryor and Terza (2002), Baser et al. (2004), Norton and Van Houtven (2006), Gibson et al. (2006), Terza (2006a,b), Terza and Tsai (2006), Shin and Moon (2007), Shea et al. (2007), Stuart et al. (2007), Lindrooth and Weisbrod (2007), and Terza (2007).

[5] Terza (2006a) shows that the GMM cannot, in general, be directly applied to the model defined in Eq. (2).

In addition, assume that the auxiliary Eq. (3) are linear, i.e.

$$x_{es} = w\alpha_s + x_{us} \quad \text{for } s = 1, \ldots, S \tag{8}$$

In this case, 2SPS is identical to the popular two-stage least squares (2SLS) [or linear instrumental variables (IV)] method and is, therefore, consistent (Greene, 2003). For a heuristic understanding of how 2SLS works, consider the linear regression model obtained by combining (2) and (7). The problem with naïve OLS regression of y on $[x_e \quad x_o]$ (ignoring the unobserved confounders x_u) is that the error term in this regression is equal to $(x_u\beta_u + e)$; this regression error is correlated with x_e, thereby introducing bias in the OLS estimators of the coefficients of $[x_e \quad x_o]$. 2SLS works by replacing $x_e\beta_e$ in Eq. (2) with $(\hat{x}_e\beta_e + \hat{x}_u\beta_e)$ and then relegating $\hat{x}_u\beta_e$ to the error term. The error thus becomes $(\hat{x}_u\beta_e + x_u\beta_u + e)$ which is easily shown to be uncorrelated with the regressors $\hat{x}_e$ and x_o.

As we discuss formally in Section 3, the consistency property of 2SLS in the fully linear models does not generally extend to the use of 2SPS in nonlinear models. The reason for this hinges on the fact that, in the nonlinear case, neither $x_u\beta_u$ nor $\hat{x}_u\beta_e$ is additive in the model (i.e. they are "inside" the function $M(\cdot)$). Therefore, neither of these terms can simply be moved "outside" of $M(\cdot)$ to become part of the error term e in (2). Applications of inconsistent 2SPS methods in nonlinear health econometric contexts can be found in many papers in health economics.[6,7] The 2SRI method, discussed next, addresses the limitation of 2SPS in nonlinear models through an alternative approach to the problem that is equivalent to 2SPS in the fully linear model.

2.3. Two-stage residual inclusion

An alternative implementation of the two-stage IV approach in nonlinear models is the two-stage residual inclusion (2SRI) method. The first-stage of the 2SRI estimator is identical to that of 2SPS. The second-stage of the estimator applies NLS to the following version of (2):

$$y = M(x_e\beta_e + x_o\beta_o + \hat{x}_u\beta_u) + e^{2SRI} \tag{9}$$

where e^{2SRI} is the regression error term, and where $\hat{x}_u$ is as defined in Section 2.2. Note that the actual observed value of the endogenous regressors x_e are maintained in the second-stage regression model while the residuals from the auxiliary regressions are substituted for the unobserved confounders x_u.[8] The reason that 2SRI works is simple: if the auxiliary regression parameters (α_s) were known then, by (3), the values of x_u would also be known and could be included among the observable controls in NLS estimation. In short, and the endogeneity of x_e would cease to exist. Although we do not know α, we *can* consistently estimate it and thereby obtain very good estimates of the true x_u's.

In the linear case defined by (7) and (8), like 2SPS, 2SRI is identical to the popular two-stage least squares (2SLS) (or linear instrumental variables (IV)) method and is, therefore, consistent. The 2SRI method is not new. It was first proposed by Hausman (1978) as a means of directly testing for endogeneity in the model defined in Eqs. (7) and (8). In this case, because 2SRI is identical to 2SLS, it is consistent for β_e and β_o. It is also easy to show that 2SRI yields a consistent estimate of β_u in the linear case.[9] Consistent 2SRI methods for specific nonlinear models have been developed by Blundell and Smith (1989, 1993), Newey (1987), Rivers and Vuong (1988), and Smith and Blundell (1986). Wooldridge (1997, 2002) suggests the use of the 2SRI method for count data models. Examples of the use of the 2SRI method in health economics can be found in DeSimone (2002), Baser et al. (2004), Norton and Van Houtven

---

[6] See Burgess et al. (2002), Cawley (2000), Ettner et al. (1999), Fox (2002), French et al. (2001), Holmes and Deb (1998), Howard (2000), Lu and McGuire (2002), Meer and Rosen (2004), Register and Williams (1992), and Savage and Wright (2003). Applications of questionable 2SPS methods outside of health economics can be found in Greene (1998) and Gramm (2003).

[7] Lee (1979,1981) has developed a 2SPS method that is consistent under special conditions when the outcome is binary and the endogenous variable is continuous. Bollen et al. (1995), Norton et al. (1998), and Mroz et al. (1999) apply this method.

[8] The 2SRI estimator can be cast as a special case of the conventional generic two-stage optimization estimator. Therefore, its asymptotic properties (in particular correct asymptotic standard errors) follow directly from the discussions found in Newey and McFadden (1994), White (1994, Chapter 6), or Wooldridge (2002, Chapter 12). Alternatively, the standard errors of the 2SRI estimator can be obtained via bootstraping.

[9] Note that this feature of 2SRI also holds true in the nonlinear case. Therefore, the exogeneity of x_e can be tested via a conventional Wald-type statistic for H_0: $\beta_{u1} = \beta_{u2} = \cdots = \beta_{uS} = 0$.

(2006), Shea et al. (2007), Stuart et al. (2007), Gibson et al. (2006), Shin and Moon (2007), and Lindrooth and Weisbrod (2007).[10]

In what follows, we formally show that 2SRI is generally consistent in the broad class of models with endogenous regressors characterized by (2) and (3).

3. Formal treatment of the consistency properties of the estimators

As discussed earlier, in the linear model 2SLS = 2SPS = 2SRI. Therefore, all three methods are consistent. These identities do not, however, hold in the generic nonlinear case so the consistency of each method must be individually examined. To prove the consistency of 2SRI, we cast it as a special case of the generic two-stage optimization estimator (OE) (see Newey and McFadden, 1994; White, 1994, Chapter 6; or Wooldridge, 2002, Chapter 12). For simplicity of exposition, let us assume that x_e and x_u each have only one element (i.e. the model involves only one endogenous variable [$S = 1$]); extension to the higher-dimensional case is straightforward. In this case, there is one auxiliary regression equation and only one vector of auxiliary parameters (α) to be estimated (i.e. $x_e = r(w\alpha) + x_u$). In what follows, we will ignore estimation of the auxiliary regression parameters (α) because, given (3), we can obtain consistent estimates of α via NLS. Abstracting from the estimation of α, and couching the discussion in the OE context, it follows that the second-stage NLS 2SRI estimator is consistent for the value of $\lambda = [\lambda_e \quad \lambda_o' \quad \lambda_u]$ that optimizes:

$$E[(y - M(x_e\lambda_e + x_o\lambda_o + (x_e - r(w\alpha))\lambda_u))^2] \tag{10}$$

Therefore, establishing the consistency of the 2SRI estimator for β, as defined in (2), amounts to showing that β is equal to the specific value of λ that optimizes (10). Now, given the properties of the identifying IV (w^+) we can replace (1) with:

$$E[y|x_e, w, x_u] = M(x_e\beta_e + x_o\beta_o + x_u\beta_u) \tag{11}$$

Therefore, β is the parameter value at which $M(\cdot)$ is equal to the conditional mean of y given x_e, x_o, x_u, and w^+. Following Goldberger (1991, p.53) it follows that β is best predictor parameter vector value under the mean squared error criterion. In other words, β is the optimizer of (10). The consistency of NLS 2SRI for β is thus established.

A similar line of reasoning for the 2SPS estimator breaks down. In the OE framework, the NLS 2SPS estimator is consistent for the value of $\delta = [\delta_e \quad \delta_o' \quad \delta_u]$ that optimizes:

$$E[(y - M(r(w\alpha)\delta_e + x_o\delta_o))^2] \tag{12}$$

At issue here is whether β_e and β_o as defined in Eq. (11) are equal to the optimizing values of δ_e and δ_o, respectively, for (12). We could in this case make the "best predictor" argument, as we did for 2SRI, if we could establish that $E[y|w] = M(r(w\alpha)\beta_e + x_o\beta_o)$. But this is not possible, in general. The problem is that, although by definition $E[x_u|w] = 0$ (one of the properties of the IV), the term $x_u\beta_u$ cannot, in general, be eliminated from (11) by conditioning on w because it is "inside" the nonlinear function $M(\cdot)$.[11]

The general inconsistency of 2SPS in the context of the model defined in (2) and (3) leads one to speculate about the possibility of specifying a similar nonlinear parametric model for which 2SPS *would be* consistent in the presence of endogeneity. We were unable to come up with an equally general specification and, given the analyses in the present section (and the supplementary appendices available from the first author upon request), do not believe that such a specification exists.[12]

---

[10] Other applications of 2SRI outside of health economics can be found in Burnett (1997), Alvarez and Glasgow (1999), McGarrity and Sutter (2000), and Petrin and Train (2006).

[11] A formal proof of the general inconsistency of 2SPS is available from the authors upon request. In this supplementary appendix, we also discuss a special case in which 2SPS is consistent.

[12] One might, for instance, consider the nonsymmetric additive model wherein $= H(x_e\beta_e + x_o\beta_o) + x_u\beta_u + e$s a candidate for consistent 2SPS estimation. Kelejian (1971), however, proves that 2SPS is generally inconsistent in such models.

4. Simulation analysis

As a follow-up to the discussion in the previous section, we explore potential biases from 2SPS relative to 2SRI estimation using simulated data in a few interesting nonlinear models involving endogenous regressors. Each of these examples is inspired by a published study in the health economics literature.

4.1. A Duration model with multinomial endogenous treatments

Gowrisankaran and Town (1999) seek to estimate the effect of hospital choice on patient mortality hazard rates. In order to deal with the potential endogeneity of the hospital choice variables, they linearize both the hazard model and the multinomial model of inpatient choices, and implement the conventional linear IV method. They justify this approach by stating that "The reason that we use a linear probability model instead of a more common Weibull or Lognormal specifications for the hazard model is that it is extremely difficult to use nonlinear models such as these with endogenous variables" (p. 754). Such a model can be specified as a special case of our modeling framework, and consistently estimated via the 2SRI method.[13]

To illustrate this point, we constructed a simulation design in which the data generating process for y is assumed to be Weibull distributed, conditional on x_e, w, and x_u; and $x_e = [\, x_{e1} \quad x_{e2} \,]$ where x_{ej} is binary with value 1 if and only if the jth multinomial logit (MNL) alternative is manifested ($x_{ej} = 0$ otherwise).[14] Note that there are actually three possible MNL outcomes in this model ($j = 0, 1, 2$) but under the usual identification restriction, the model has been normalized on one of the outcomes ($j = 0$). Using this sampling design, we generated 1000 samples of size $n = 5000$ and to each of them applied four different estimators: (1) True Model—maximum likelihood estimation (MLE) based on the actual model used to simulate the data; (2) Naïve Model—MLE ignoring the unobservable confounders (x_u); (3) 2SPS; and (4) 2SRI.[15] Using the results from each of these models, we estimated the following average treatment effects as a basis for comparison of estimator performance[16]:

$$E[y_{x_{e\omega}=1}] - E[y_{x_{e\xi}=1}] \equiv \text{effect on survival of exogenously imposing treatment option } \omega \text{ vs. treatment option } \xi$$

$$\text{for } \omega, \xi = 0, 1, 2 \text{ and } \xi \neq \delta \tag{13}$$

where $y_{x_{e\omega}=1}$ denotes the random variable representing the counterfactual outcome as it would be under the exogenously imposed $x_{e\omega} = 1$ scenario.[17]

The results of our simulation analysis are displayed in Table 1. For each of the True, Naïve, 2SPS, and 2SRI estimators we give the percent absolute bias of the estimated value of (13) relative to the corresponding true values across the 1000 simulated samples. The differences between 2SPS and 2SRI in the estimation of the average effects in (13) are striking. The mean overall bias for the 2SPS estimated average effects (the simple average of the first three entries in the 2SPS column of Table 1) is 36% as compared to 8% for 2SRI (as a baseline for comparison, note that this value is 8% for the True Model). To investigate the theoretical prediction that the 2SRI estimates should converge toward the true values as the sample size increases, we also simulated 1000 replicates of size 20,000 and repeated the estimations. These results, given in the bottom half of Table 1, comport with the consistency of the 2SRI estimator and the lack thereof for 2SPS. Huge differences between 2SPS and 2SRI with regard to bias in the estimation of the average effects persist despite the increase in sample size. The overall bias for 2SPS (the simple average of the last three entries in the 2SPS column of Table 1) remains high at 27% but drops to 1.3% for 2SRI (the analogous value for the True Model is 1.3%).

---

[13] We note that Gowisankaran and Town later teamed with Geweke and re-estimated their model in an appropriately designed Bayesian framework (see Geweke et al., 2003).

[14] We refer here to the multinomial logit model introduced by McFadden (1973).

[15] The complete details of the sampling design will be supplied upon request.

[16] Recall that the treatment options are mutually exclusive and collectively exhaustive. Therefore, if $x_{e\omega} = 1$ then $x_{e\xi} = 0$ for all $\omega \neq \xi$ and $\omega, \xi = 0, 1, 2$.

[17] Details of the estimators for the average treatment effects in Eq. (13) will be supplied upon request.

Table 1
Simulation results for Weibull outcome with multinomial endogenous variables average absolute %bias

	Average effects			
	True model (%bias)	Naïve model (%bias)	2SPS model (%bias)	2SRI model (%bias)
Based on 1000 replicates of size $n = 5000$				
$E[y_{x_{e1}=1}] - E[y_{x_{e0}=1}]$	16	205	28	16
$E[y_{x_{e2}=1}] - E[y_{x_{e0}=1}]$	2	99	38	2
$E[y_{x_{e2}=1}] - E[y_{x_{e1}=1}]$	5	34	42	5
Based on 1000 replicates of size $n = 20,000$				
$E[y_{x_{e1}=1}] - E[y_{x_{e0}=1}]$	2	205	2	2
$E[y_{x_{e2}=1}] - E[y_{x_{e0}=1}]$	2	97	27	2
$E[y_{x_{e2}=1}] - E[y_{x_{e1}=1}]$	0	34	51	0

The value in a particular cell of the table is the average percentage absolute bias, over the 1000 simulated samples, for a particular (estimator-q, average effect-t, sample size-j) combination, and is measured as $\sum_{m=1}^{1000}((1/1000)(\text{abs}(\widehat{AE(t)}_{qrm} - AE(t))/\text{abs}(AE(t)))) \times 100\%$, where $AE(t)$ denotes the true value of the tth effect, $\widehat{AE(t)}_{qrm}$ is its estimated value obtained by applying the qth method to mth sample of the rth sample size, with, $q = $ true MLE, naïve, 2SPS, 2SRI, $t = E[y_{x_{e1}=1}] - E[y_{x_{e0}=1}], E[y_{x_{e2}=1}] - E[y_{x_{e0}=1}], E[y_{x_{e2}=1}] - E[y_{x_{e1}=1}], r = 5000, 20,000$.

4.2. Ordered logit with a count-valued endogenous treatment

Lu and McGuire (2002) seek to estimate the effect of treatment on subsequent substance abuse. The ordinal outcome in their model can be represented by the vector: $y = [\, y_1 \quad y_2 \quad y_3 \quad y_4 \,]$, where

- $y_1 = 1$ if the client got worse at the time of discharge (used more drugs than at admission), 0 otherwise;
- $y_2 = 1$ if the client's drug use frequency stayed the same, 0 otherwise;
- $y_3 = 1$ if the client got better but did not achieve abstinence, 0 otherwise;
- $y_4 = 1$ if abstinence was achieved, 0 otherwise.

The endogenous variable in the regression is $x_e = $ log of number of visits the client makes during the episode of treatment.

They model y as an ordinal logit regression, and estimate the parameters using 2SPS. As discussed above, this estimator is not consistent. To illustrate this point we generated 1000 replicates each of sample size $n = 10,000$ (in line with the Lu and McGuire (2002) sample size of 13,362) based on an ordered logit sampling design.[18] Using the appropriate ordered logit MLE, for each sample we estimated the True (actual value of x_u used) and Naïve (x_u not included) versions of the model. We also applied the 2SPS estimator used by Lu and McGuire (2002), and the 2SRI ordered logit MLE. The results from each of these models were then used to estimate the following average probability effects attributable to specified exogenous changes in x_e:

$$P(y_{j(x_e=b)} = 1) - P(y_{j(x_e=a)} = 1) \equiv \text{the effect of an exogenous change in } x_e \text{ from } a \text{ to } b \text{ on the probability of}$$

the individual being in outcome category j for $j = 1, \ldots, 4$ (14)

where $y_{j(x_e=c)}$ denotes the random variable representing the counterfactual outcome as it would be under the endogenously imposed $x_e = c$ scenario.[19]

The simulation results are displayed in Table 2. For samples of size $n = 10,000$, 2SRI clearly dominates 2SPS with regard to estimation of the estimated probability effects (14). Taking the simple average of the first four entries in the 2SPS column of Table 2 yields an overall mean bias of 26.5%, while the analogous measure for 2SRI is at 7% (the baseline measure from the True Model is 5%). To track the behavior of the estimators with an increase in sample size, we reran the analysis with $n = 20,000$. The size of the 2SPS bias remained virtually unchanged, while the bias in 2SRI

---

[18] Details of the simulation design will be supplied upon request.
[19] Details of the estimators for the probability effects in Eq. (14) will be supplied upon request.

Table 2
Simulation results for ordered categorical outcome with count endogenous variable average absolute %bias

	Average effects			
	True model (%bias)	Naïve model (%bias)	2SPS model (%bias)	2SRI model (%bias)
Based on 1000 replicates of size $n = 10,000$				
$P(y_{1(x_e=\ln(4))} = 1) - P(y_{1(x_e=\ln(2))} = 1)$	0	20	28	4
$P(y_{2(x_e=\ln(4))} = 1) - P(y_{2(x_e=\ln(2))} = 1)$	1	67	29	1
$P(y_{3(x_e=\ln(4))} = 1) - P(y_{3(x_e=\ln(2))} = 1)$	15	367	44	19
$P(y_{4(x_e=\ln(4))} = 1) - P(y_{4(x_e=\ln(2))} = 1)$	3	191	5	4
Based on 1000 replicates of size $n = 20,000$				
$P(y_{1(x_e=\ln(4))} = 1) - P(y_{1(x_e=\ln(2))} = 1)$	0	20	28	0
$P(y_{2(x_e=\ln(4))} = 1) - P(y_{2(x_e=\ln(2))} = 1)$	0	67	29	0
$P(y_{3(x_e=\ln(4))} = 1) - P(y_{3(x_e=\ln(2))} = 1)$	0	367	37	0
$P(y_{4(x_e=\ln(4))} = 1) - P(y_{4(x_e=\ln(2))} = 1)$	0	191	8	0

The value in a particular cell of the table is the average percentage absolute bias, over the 1000 simulated samples, for a particular (estimator-q, average effect-t, sample size-j) combination, and is measured as $\sum_{m=1}^{1000}((1/1000)(\text{abs}(\widehat{AE(t)}_{qrm} - AE(t))/\text{abs}(AE(t)))) \times 100\%$, where $AE(t)$ denotes the true value of the tth effect, $\widehat{AE(t)}_{qrm}$ is its estimated value obtained by applying the qth method to mth sample of the rth sample size, and q = true MLE, naïve, 2SPS, 2SRI $t = P(y_{1(x_e=\ln(4))} = 1) - P(y_{1(x_e=\ln(2))} = 1)$, $P(y_{2(x_e=\ln(4))} = 1) - P(y_{1(x_e=\ln(2))} = 1)$, $P(y_{3(x_e=\ln(4))} = 1) - P(y_{3(x_e=\ln(2))} = 1)$, $P(y_{4(x_e=\ln(4))} = 1) - P(y_{4(x_e=\ln(2))} = 1)$, $r = 10,000, 20,000$.

estimation all but disappears. Here, as in the Weibull duration analysis, the simulation results support the theoretical consistency of the 2SRI estimator.

5. Mullahy's birthweight model revisited

To demonstrate the potential differences that might arise in actual practice between the 2SPS and 2SRI estimates, we re-estimated Mullahy's (1997) model of the effect of prenatal cigarette smoking on birthweight using data supplied by the author. Mullahy (1997) suspects that maternal smoking during pregnancy may be correlated with the unobservable determinants of birthweight, so he specifies a nonlinear conditional mean regression model, which can be viewed as a special case of (1). In Mullahy's model, birthweight (y) is the following function of prenatal smoking (x_e), other observable determinants (x_o) and a scalar representing the unobservable birthweight determinants that are correlated with prenatal smoking (x_u)[20]:

$$y = \exp(x_e\beta_e + x_o\beta_o + x_u) + e \tag{15}$$

where x and β are defined canonically, and e is the random error term which is tautologically defined as $e = y - \exp(x_e\beta_e + x_o\beta_o + x_u)$, so that $E[e|x_e, x_o, x_u] = 0$. Mullahy (1997) demonstrates that, given a vector of instrumental variables $w = [x_o w^+]$, if the following conditions hold:

$$\begin{aligned} E[\exp(x_u)|w] &= 1 \\ E[y|x_e, w, x_u] &= E[y|x_e, x_o, x_u] \end{aligned} \tag{16}$$

then β_e and β_o can be consistently estimated via a generalized method of moments (GMM) estimator that does not require explicit specification of an auxiliary regression of x_e on w.[21]

We implemented a flexible functional form for which GMM is not feasible (see Terza, 2006a) and 2SPS is not consistent. Specifically, we replaced (15) with the following variant of the inverse of the Box and Cox (1964) model originally suggested by Wooldridge (1992) for nonlinear models that do not involve endogeneity:

$$E[y|x_e, x_o, x_u] = k(x_e\beta_e + x_o\beta_o + x_u\beta_u, \tau) \tag{17}$$

---

[20] Mullahy (1997) sets the coefficient of x_u (β_u) equal to 1. It is not identified and is therefore irrelevant in his model.
[21] For a discussion of the GMM see Hansen (1982).

Table 3
Descriptive statistics of sample for re-analysis of Mullahy's birthweight model

Variable	Mean	Max	Min
BIRTHWT (oz.)	7.42	16.94	1.44
CIGSPREG	2.09	50.00	0.00
PARITY	1.63	6.00	1.00
WHITE	0.78	1.00	0.00
MALE	0.52	1.00	0.00
EDFATHER	11.32	18.00	0.00
EDMOTHER	12.93	18.00	0.00
FAMINCOM	29.03	65.00	0.50
CIGTAX88	19.55	38.00	2.00

where

$$k(a, \tau) = \begin{cases} \left(((\tau/2)a + 1)^2\right)^{1/\tau} & \tau \neq 0 \\ \exp(a) & \tau = 0 \end{cases}$$

and $0 \leq \tau \leq 2$. This version of the inverse Box–Cox (IBC) model maintains the desired positivity of the regression function (regardless of the values of τ and $x_e\beta_e + x_o\beta_o + x_u\beta_u$), and possesses all of the essential properties of Wooldridge's (1992) IBC formulation. In particular, $k(a, \tau)$ subsumes the linear model when $\tau = 2$, and $k(a, \tau) \to \exp(a)$ as $\tau \to 0$. We estimated the parameters of (17) using both the 2SRI and 2SPS estimators. Following Mullahy (1997) who states "... a linear reduced form for CIGARETTES may not be unreasonable" we specify the auxiliary regression for prenatal cigarette consumption (x_e) as in Eq. (8).

We used the same variables as did Mullahy: y = the newborn's weight measured in pounds; x_e = number of cigarettes smoked per day during pregnancy, x_o = [1 PARITY WHITE MALE], w^+ = [EDFATHER EDMOTHER FAMINCOM CIGTAX88]; PARITY = birth order; WHITE = 1 if white, 0 otherwise; MALE = 1 if male, 0 otherwise; EDFATHER = paternal schooling (years); EDMOTHER = maternal schooling (years); FAMINCOM = family income ($\times 10^{-3}$); CIGTAX99 = per pack state excise tax on cigarettes. The descriptive statistics of the sample are given in Table 3.

For 2SRI estimation, we applied OLS to the linear auxiliary equation, and used NLS to estimate β and τ in the following version of (9):

$$y = k(x_e\beta_e + x_o\beta_o + \hat{x}_u\beta_u, \tau) + e^{2SRI} \tag{18}$$

where $\hat{x}_u = x_e - w\hat{\alpha}$ and $\hat{\alpha}$ denotes the first-stage OLS estimator of α.[22] For 2SPS estimation, we implemented the same first-stage estimator of α but in the second-stage applied NLS to

$$y = k(\hat{x}_e\delta_e + x_o\delta_o, \tau) + e^{2SPS} \tag{19}$$

where $\hat{x}_e$ denotes the first-stage OLS predictor of x_e. The first-stage OLS estimates are given in Table 3. The 2SRI and 2SPS results are shown in Table 4. As in the simulation analyses of the previous section, the ultimate estimation objective is the causal effect of an exogenous change in prenatal smoking frequency on birthweight. For instance, consider:

$$E[y_0] - E[y_{20}] \equiv \text{the average effect of an exogenously imposed decrease in prenatal smoking from one pack}$$

$$\text{(20 cigarettes) per day to zero usage} \tag{20}$$

where $y_{x_e^*}$ is the random variable representing birthweight as it would be under the exogenously imposed prenatal smoking level x_e^*. Terza (2006b) shows that under general conditions we can rewrite (20) as $E\left[E[y|x_e = 0, x_o, x_u] - E[y|x_e = 20, x_o, x_u]\right]$ which, when combined with (16) yields:

$$E[y_0] - E[y_{20}] = E\left[k(x_o\beta_o + x_u\beta_u, \tau) - k(20\beta_e + x_o\beta_o + x_u\beta_u, \tau)\right] \tag{21}$$

---

[22] We performed a line search on τ to find its sum of squared residuals minimizing value.

Table 4
First-stage OLS estimates of auxiliary regression in the re-analysis of Mullahy's birthweight model

Variable	Estimate	t-stat
CONSTANT	6.74	6.49
PARITY	0.30	1.72
WHITE	0.78	1.89
MALE	−0.04	−0.13
EDFATHER	−0.12	−3.14
EDMOTHER	−0.33	−4.37
FAMINCOM	−0.02	−2.01
CIGTAX88	0.03	1.43

Second column gives estimates of the elements of α in $x_e = w\alpha + x_u$.

Table 5
Estimation results (IBC version of Mullahy's birthweight model

Variable	2SPS IBC		2SRI IBC	
	Coeff.	t-stat	Coeff.	t-stat
CONSTANT	58.96	22.07	1.27	215.33
CIGSPREG	−1.92	−3.47	−0.004	−3.52
PARITY	3.23	3.26	0.01	3.39
WHITE	9.64	4.85	0.02	4.64
MALE	4.64	2.81	0.01	2.93
First-stage residual (x_u)	–	–	0.024	2.03
Box–Cox parameter τ	5.20	–	−0.95	–
$E[y_0] - E[y_{20}]$	1.85		1.41	

Second and fourth columns, respectively, give the estimates of the parameters in Eqs. (18) and (19). Last row shows the 2SPS and 2SRI estimated effects of an exogenously imposed one-pack per day reduction in smoking, respectively.

We estimated (21), using the 2SRI results, as

$$\sum_{i=1}^{n} \frac{1}{n} \left\{ k(x_{oi}\hat{\beta}_e + \hat{x}_{ui}\hat{\beta}_u, \hat{\tau}) - k(20\hat{\beta}_e + x_{oi}\hat{\beta}_e + \hat{x}_{ui}\hat{\beta}_u, \hat{\tau}) \right\} \tag{22}$$

where $\hat{x}_u$ denotes the first-stage OLS residual, and the "~s" indicate the 2SRI estimates. Alternatively, we estimated (21) with the 2SPS results using:

$$\sum_{i=1}^{n} \frac{1}{n} \left\{ k(x_{oi}\tilde{\delta}_e, \tilde{\tau}) - k(20\tilde{\delta}_e + x_{oi}\tilde{\delta}_e, \tilde{\tau}) \right\} \tag{23}$$

where the "~s" indicate the 2SPS estimates. The values of (22) and (23) are given in the last row of Table 5. As is shown therein, the predicted effects of an exogenous reduction in smoking from a pack per day to abstinence differ substantially between the two methods. Given the consistency of the 2SRI estimates, the results imply that 2SPS overstates (in absolute terms) the effect on birthweight by approximately 7 oz. which is about 6% of the sample mean birthweight.

6. Discussion

We have examined two estimation methods that are commonly used in health economic applications involving nonlinear models with endogenous regressors—two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI). The discussion begins with a detailed description of the estimators in an intuitively appealing nonlinear regression framework that explicitly accounts for endogeneity (i.e. the presence of unobservable confounders). Within that framework we show that the 2SRI estimator is generally consistent while the 2SPS approach is not. To assess the potential extent of the bias in 2SPS estimation, we conducted simulation analyses based on two studies found in

the recent health economics literature—one in which an inherently nonlinear duration model involving endogeneity was "linearized" in order to apply conventional linear instrumental variables (IV) methods; the other in which 2SPS was applied in order to deal with an endogenous regressor in an ordered logit model. The results reveal that potential bias from the use of 2SPS can be substantial and that such bias is not attenuated as the sample size increases. As a follow-up to the simulation analyses, and to examine the possible differences between 2SRI and 2SPS estimates in a real-world estimation setting, we revisited the study of the effect of prenatal cigarette smoking on birthweight conducted by Mullahy (1997). Using a flexible functional form we re-estimated the model and found there to be considerable differences between the estimates obtained via 2SPS vs. 2SRI.

Our theoretical results, combined with those from the simulation and replication studies, favor the use of 2SRI for the econometric estimation of nonlinear models with endogenous regressors. We hope that this work will serve as a guide to applied researchers in health economics and health services research.

Acknowledgements

References

Alvarez, R.M., Glasgow, G., 1999. Two-stage estimation of nonrecursive choice models. Political Analysis 8, 147–165.

Baser, O., Bradley, C.J., Gardiner, J.C., Given, C., 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.

Blundell, R.W., Smith, R.J., 1989. Estimation in a class of simultaneous equation limited dependent variable models. Review of Economics and Statistics 56, 37–58.

Blundell, R.W., Smith, R.J., 1993. Simultaneous microeconmetric models with censored or qualitative dependent variables. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, vol.2. North Holland Publishers, Amsterdam, pp. 1117–1143.

Bollen, K.A., Guilkey, D.K., Mroz, T.A., 1995. Binary outcomes and endogenous explanatory variables: tests and solutions with an application to the demand for contraceptive use in Tunisia. Demography 32, 111–131.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society, Series B 26, 211–252.

Burgess, S., Gregg, P., Propper, C., Wasgbrook, E., ALSPAC Study Team, 2002. Maternity Rights and Mothers' Return to Work. Working Paper - CMPO, 02/055, 00.

Burnett, N.J., 1997. Gender economics courses in liberal arts colleges. Journal of Economic Education 28, 369–377.

Cawley, J., 2000. An instrumental variables approach to measuring the effect of body weight on employment disability. Health Services Research 35, 1159–1179.

Coulson, E., Neslusan, C., Stuart, B., Terza, J., 1995. Estimating the moral hazard effect of supplemental medical insurance in the demand for prescription drugs by the elderly. American Economic Review - Paper and Proceedings 85, 122–126.

DeSimone, J., 2002. Illegal drug use and employment. Journal of Labor Economics 20, 952–977.

Ettner, S.L., Hermann, R.C., Tang, H., 1999. Differences between generalists and mental health specialists in the psychiatric treatment of medicare beneficiaries. Health Services Research 34, 737–760.

Fox, M., 2002. Medical student indebtedness and the propensity to enter academic medicine. Health Economics 12, 101–112.

French, M.T., Roebuck, M.C., Alexandre, P.K., 2001. Illicit drug use, employment, and labor force participation. Southern Economic Journal 68, 349–368.

Geweke, J., Gowrisankaran, G., Town, R., 2003. Bayesian inference for hospital quality in a selection model. Econometrica 71, 1215–1238.

Gibson, T.B., Mark, T.L., Axelsen, K., Baser, O., Rublee, D.A., McGuigan, K.A., 2006. Impact of statin copayments on adherence and medical care utilization and expenditures. American Journal of Managed Care 12, SP11–SP19.

Goldberger, A.S., 1991. A Course in Econometrics. Harvard University Press, Cambridge, MA.

Gowrisankaran, G., Town, R., 1999. Estimating the quality of care in hospitals using instrumental variables. Journal of Health Economics 18 (6), 747–767.

Gramm, M., 2003. The case for regulatory rent-seeking: CRA based protests of bank mergers and acquisitions. Public Choice 116, 367–379.

Greene, W.H., 1998. Gender economics courses in liberal arts colleges: further results. Journal of Economic Education 29, 291–300.

Greene, W.H., 2003. Econometric Analysis, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Holmes, A.M., Deb, P., 1998. Provider choice and use of mental health care: implications for gatekeeper models. Health Services Research 33, 1263–1284.

Howard, D., 2000. The impact of waiting time on liver transplant outcomes. Health Services Research 35, 117–1134.

Kelejian, H.H., 1971. Two-stage least squares and econometric systems linear in the parameters but nonlinear in the endogenous variables. Journal of the American Statistical Association 66, 373–374.

Kenkel, D., Terza, J., 2001. The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect. Journal of Applied Econometrics 16, 165–184.

Lee, L., 1979. identification and estimation in binary choice models with limited 'Censored' dependent variables. Econometrica 47, 977–995.

Lee, L., 1981. Simultaneous equations models with discrete and censored dependent variables. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 346–364 (Chapter 9).

Lindrooth, R.C., Weisbrod, B.A., 2007. Do religious nonprofit and for-profit organizations respond differently to financial incentives? the hospice industry. Journal of Health Economics 26, 342–357.

Lu, M., McGuire, T.G., 2002. The productivity of outpatient treatment for substance abuse. Journal of Human Resources 37, 309–335.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142 (Chapter 4).

McGarrity, J.P., Sutter, D., 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.

McGeary, K.A., French, M.T., 2000. Illicit drug use and emergency room utilization. Health Services Research 35, 153–169.

Meer, J., Rosen, H.S., 2004. Insurance and the utilization of medical services. Social Science in Medicine 58, 1623–1632.

Mroz, T.A., Bollen, K.A., Speizer, I.S., Mancini, D.J., 1999. Quality, accessibility, and contraceptive use in rural Tanzania. Demography 36, 23–40.

Mullahy, J., 1997. Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. Review of Economics and Statistics 79, 586–593.

Neslusan, C.A., Hylan, T.R., Dunn, R.L., Donoghue, J., 1999. Controlling for systematic selection in retrospective analyses: an application to fluoxetine and sertraline prescribing in the United Kingdom. Value in Health 2, 435–445.

Newey, W.K., 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. Journal of Econometrics 36, 231–250.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. Elsevier Science B.V., Amsterdam, pp. 2111–2245, Chapter 36.

Norton, E.C., Van Houtven, C.H., 2006. Intervivos transfers and exchange. Southern Economic Journal 73, 157–172.

Norton, E.C., Lindrooth, R.C., Ennett, S.T., 1998. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. Health Economics 7, 439–453.

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

Pryor, C., Terza, J., 2002. Are going concern audit opinions a self-fulfilling prophecy? Advances in Quantitative Analysis of Finance and Accounting 10, 89–116.

Register, C.A., Williams, D.R., 1992. Labor market effects of marijuana and cocaine use among young men. Industrial and Labor Relations Review 45, 435–448.

Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics 39, 347–366.

Savage, E., Wright, D.J., 2003. Moral hazard and adverse selection in Australian private hospitals: 1989–1990. Journal of Health Economics 22, 331–359.

Shea, D., Terza, J., Stuart, B., Briesacher, B., 2007. Estimating the effects of prescription drug. Coverage for medicare beneficiaries. Health Services Research 43, 933–949.

Shin, J., Moon, S., 2007. Do HMO plans reduce health care expenditure in the private sector. Economic Inquiry 45, 82–99.

Smith, R.J., Blundell, R.W., 1986. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. Econometrica 54, 679–685.

Stuart, B.C, Terza, J.V., Doshi, J., 2007. Assessing the Impact of Drug Use on Hospital Costs. Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Terza, J., 1994a. 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.

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

Terza, J., 1998. Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. Journal of Econometrics 84, 129–154.

Terza, J., 1999. Estimating endogenous treatment effects in retrospective data analysis. Value in Health 2, 429–434.

Terza, J., 2002. Alcohol abuse and employment: a second look. Journal of Applied Econometrics 17, 393–404.

Terza, J., 2006a. Estimation of policy effects using parametric nonlinear models: a contextual critique of the generalized method of moments. Health Services and Outcomes Research Methodology 6, 177–198.

Terza, J., 2006b. An Econometric Framework for Analyzing Health Policy with Nonexperimental Data. Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Terza, J., 2007. Parametric Nonlinear Regression with Endogenous Switching. Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Terza, J., Tsai, W., 2006. Censored probit estimation with correlation near the boundary: a useful reparameterization. Review of Applied Economics 2, 1–12.

Treglia, M., Neslusan, C.A., Dunn, R.L., 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.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, Cambridge.

Wooldridge, J.M., 1992. Some alternatives to the Box-Cox regression model. International Economic Review 33, 935–955.

Wooldridge, J.M., 1997. Quasi-likelihood methods for count data. In: Pesaran, M., Schmidt, P. (Eds.), Handbook of Applied Econometrics, Volume II: Microeconometrics. Blackwell Publishers Ltd., Malden, MA.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.