SIMPLER STANDARD ERRORS FOR MULTI-STAGE REGRESSION-BASED ESTIMATORS: ILLUSTRATIONS IN HEALTH ECONOMICS

by

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

(May, 2014)

PLEASE DO NOT QUOTE WITHOUT AUTHOR'S PERMISSION

This research was supported by a grant from the Agency for Healthcare Research and Quality (R01 HS017434-01). Presented at the 2012 Stata Conference in San Diego CA, July 26-27, 2012.

Abstract

With a view towards lessening the analytic and computational burden faced by practitioners seeking to correct the standard errors of two-stage estimators, we offer a heretofore unexploited simplification of the conventional formulation for the most commonly encountered cases in empirical application – two-stage estimators that, in either stage, involving maximum likelihood estimation or the nonlinear least squares method. Also with the applied researcher in mind, we cast the discussion in the context of nonlinear regression models involving endogeneity – a sampling problem whose solution often requires two-stage estimation. We detail our simplified standard error formulations for three very useful estimators in applied contexts involving endogeneity in a nonlinear setting (endogenous regressors, endogenous sample selection, and causal effects). The analytics and Stata/Mata code for implementing our simplified formulae are demonstrated with illustrative real-world examples and simulated data.

1. Introduction

Asymptotic theory for the two-stage optimization estimator (2SOE) (in particular, correct formulation of the asymptotic standard errors) has been available to applied researchers for decades [see Murphy and Topel (1985) for cases in which both stages are MLE; and Newey and McFadden (1994) and White (1994) for more general classes of 2SOE]. Despite textbook treatments of the subject [Cameron and Trivedi (2005), Greene (2012), and Wooldridge (2010)], when conducting statistical inference based on two-stage estimates, applied researchers often implement bootstrapping methods or ignore the two-stage nature of the estimator and report the uncorrected second-stage outputs from packaged statistical software. In the present paper, with a view toward easy software implementation (in Stata), we offer the practitioner a heretofore largely unexploited simplification of the textbook asymptotic covariance matrix formulations (and their estimators – standard errors) for the most commonly encountered versions of the 2SOE -- those involving MLE or the nonlinear least squares (NLS) method in either stage. In addition, and perhaps more importantly from a practitioners standpoint, we cast the discussion in the context of regression models involving endogeneity – a sampling problem whose solution often requires a 2SOE.

We detail our simplified covariance specifications for three estimators that can be in applied in empirical contexts involving endogeneity -- the two-stage residual inclusion (2SRI) estimator suggested by Terza et al. (2008) for nonlinear models with endogenous regressors; the two-stage sample selection estimator (2SSS) developed by Terza (2009) for nonlinear models with endogenous sample selection; and causal incremental and marginal effects estimators as discussed by Terza (2012). The analytics and Stata code for implementing our simplified formulae for correcting the asymptotic standard errors of each of these estimators are demonstrated with specific illustrative real-world examples.

The remainder of the paper is organized as follows. In the next section, we review the asymptotic theory of 2SOE and give the conventional textbook formulation of the corresponding correct asymptotic covariance matrix. We also show how this formulation can be simplified when the second stage of the estimator implements either NLS or MLE. In section 3, we detail the 2SRI, 2SSS, and causal effect estimators and, in light of the discussion in section 2, we derive their correct (and simplified) asymptotic standard errors. Specific illustrations of the estimators given in section 3 (and their corrected asymptotic standard errors) are detailed in section 4, complete with corresponding Stata code and applications to real data. The final section summarizes and concludes. Technical details are given in appendices that will be supplied upon request.

2. Two-Stage Optimization Estimators and Their Asymptotic Standard Errors

The vast majority of estimators implemented in empirical health economics and health services research are *optimization estimators (OEs)* – statistical methods that produce estimates as optimizers of well specified objective functions. The most prominent OE examples are the maximum likelihood estimator (MLE) and the nonlinear least squares (NLS) method. Model design or computational convenience often dictates that an OE be implemented in two stages. In such cases the parameter vector of interest is partitioned as $\omega' = [\delta' \quad \gamma']$ and conformably estimated in two-stages. First, an estimate of δ is obtained as the optimizer of an appropriately specified first-stage objective function

$$\sum_{i=1}^{n} q_1(\delta, V_i) \tag{1}$$

where $q_1(\ )$ corresponds to a specific type of OE and V_i denotes the relevant subvector of the observable data for the ith sample individual (i = 1, ..., n). Next, an estimate of γ is obtained as the optimizer of

$$\sum_{i=1}^{n} q(\hat{\delta}, \gamma, Z_i) \qquad (2)$$

where $q(\ )$ defines the relevant single-stage OE, Z_i is the full vector of observable data, and $\hat{\delta}$ denotes the first-stage estimate of δ.

It is well established that under general conditions, this two-stage optimization estimator (2SOE) is consistent and asymptotically normal.[1] Our interest here is in simplifying the formulation of the corresponding asymptotic covariance matrix of $\hat{\omega}' = [\hat{\delta}' \ \ \hat{\gamma}']$, where $\hat{\gamma}$ denotes the second-stage estimator obtained from (2). For future reference and notational convenience, this matrix is denoted

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix}$$

where $D_{11} = \text{AVAR}(\hat{\delta})$ denotes the asymptotic covariance matrix of $\hat{\delta}$, $D_{22} = \text{AVAR}(\hat{\gamma})$, and D_{12} is left unspecified for the moment. For cases in which the ultimate estimation objective is γ, only D_{22} is of interest. In most cases, however, the full vector of parameter estimates $\hat{\omega}$ will be needed for an additional estimation step. We will discuss one such example (causal effect estimation) later in this paper. Hence our interest is in simplifying the details of the full formulation of D. Before proceeding we establish the following notational conventions:

[1] See Newey and McFadden (1994) or White (1994) for details.

-- q_1 is shorthand notation for $q_1(\delta, V)$ as defined in (1)

-- q is shorthand notation for $q(\delta, \gamma, Z)$ as defined in (2)

-- $\nabla_s q$ denotes the gradient of q with respect to parameter subvector s – a row vector.

-- $\nabla_{st} q$ denotes the matrix whose typical element is $\partial^2 q / \partial s_j \partial t_m$ -- its row dimension corresponds to that of its first subscript and the column dimension to that of its second subscript.

We now turn to the details of the elements of D. We first note that D_{11} warrants no discussion, because neither its formulation nor its estimation are affected by the two-stage nature of the estimator -- γ does not appear in (1). Therefore, the correct standard errors of, and other inferential statistics pertaining to, $\hat{\delta}$ can be obtained from the "packaged" output of the software used for first-stage estimation. By the same token, we need only consider how the choice of method for the second-stage determines the formulation and estimation of D_{12} and D_{22}. Because MLE and NLS are the most commonly implemented OEs, we focus on 2SOEs that implement these methods in the second stage. Using the results of Murphy and Topel (1985) it is easy to show that when the second stage is MLE we have[2]

$$D_{12} = E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_\gamma q' \nabla_\delta q_1\right]' \mathrm{AVAR}^*(\tilde{\gamma}) + \mathrm{AVAR}(\hat{\delta}) E\left[\nabla_\gamma q' \nabla_\delta q\right]' \mathrm{AVAR}^*(\tilde{\gamma})$$

$$\begin{aligned} D_{22} = \mathrm{AVAR}^*(\tilde{\gamma}) \Big\{ & E\left[\nabla_\gamma q' \nabla_\delta q\right] \mathrm{AVAR}(\hat{\delta}) E\left[\nabla_\gamma q' \nabla_\delta q\right] \\ & - E\left[\nabla_\gamma q' \nabla_\delta q_1\right] E\left[\nabla_{\delta\delta} q\right]^{-1} E\left[\nabla_\gamma q' \nabla_\delta q\right]' \\ & - E\left[\nabla_\gamma q' \nabla_\delta q\right] E\left[\nabla_{\delta\delta} q\right]^{-1} E\left[\nabla_\gamma q' \nabla_\delta q_1\right]' \Big\} \mathrm{AVAR}^*(\tilde{\gamma}) + \mathrm{AVAR}^*(\tilde{\gamma}) \end{aligned} \quad (3)$$

[2] An appendix detailing this result will be supplied upon request.

where $\tilde{\gamma}$ denotes the second stage MLE estimate of γ, and AVAR*( ) is the matrix to which the "packaged" asymptotic covariance estimator of the second stage converges in probability.[3] Likewise, using the results of White (1994), we can show that when the second stage is NLS we have[4]

$$D_{12} = E\left[\nabla_{\delta\delta}q_1\right]^{-1} E\left[\nabla_\gamma q'\nabla_\delta q_1\right]' E\left[\nabla_{\gamma\gamma}q\right]^{-1} - AVAR(\hat{\delta})E\left[\nabla_{\gamma\delta}q\right]' E\left[\nabla_{\gamma\gamma}q\right]^{-1}$$

$$\begin{aligned} D_{22} = E\left[\nabla_{\gamma\gamma}q\right]^{-1} \Big\{ & E\left[\nabla_{\gamma\delta}q\right]AVAR(\hat{\delta})E\left[\nabla_{\gamma\delta}q\right]' \\ & - E\left[\nabla_\gamma q'\nabla_\delta q_1\right]E\left[\nabla_{\delta\delta}q\right]^{-1} E\left[\nabla_{\gamma\delta}q\right]' \\ & - E\left[\nabla_{\gamma\delta}q\right]E\left[\nabla_{\delta\delta}q\right]^{-1} E\left[\nabla_\gamma q'\nabla_\delta q_1\right]'\Big\} E\left[\nabla_{\gamma\gamma}q\right]^{-1} + AVAR^*(\hat{\gamma}). \end{aligned} \tag{4}$$

We can, however, also show that when the second stage estimator is MLE or NLS[5]

$$E\left[\nabla_\gamma q'\nabla_\delta q_1\right] = 0. \tag{5}$$

This allows us to greatly simplify (3) and (4), respectively, as

$$D_{12} = AVAR(\hat{\delta})\, E\left[\nabla_\gamma q'\nabla_\delta q\right]' AVAR^*(\tilde{\gamma})$$

$$D_{22} = AVAR^*(\tilde{\gamma})E\left[\nabla_\gamma q'\nabla_\delta q\right]AVAR(\hat{\delta})E\left[\nabla_\gamma q'\nabla_\delta q\right]'AVAR^*(\tilde{\gamma}) + AVAR^*(\tilde{\gamma}) \tag{6}$$

when the second stage is MLE, and

$$D_{12} = -\,AVAR(\hat{\delta})E\left[\nabla_{\gamma\delta}q\right]' E\left[\nabla_{\gamma\gamma}q\right]^{-1}$$

---

[3] By "packaged" we mean that which would be obtained from any econometrics computer package for the second stage estimator of γ, ignoring the two-stage nature of the estimator.

[4] An appendix detailing this result will be supplied upon request.

[5] An appendix detailing this result will be supplied upon request.

$$D_{22} = E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma\delta} q\right] AVAR(\hat{\delta}) E\left[\nabla_{\gamma\delta} q\right]' \, E\left[\nabla_{\gamma\gamma} q\right]^{-1} + AVAR*(\hat{\gamma}) \tag{7}$$

when the second stage is NLS.

The expressions in (6) and (7) are of practical use in that they served to highlight the covariance matrix components that can be directly obtained from packaged econometric software vs. those that require special programming. It is clear that software implementation of the corrected covariance formulation is simpler in the second-stage MLE case. Here the only component that must be analytically derived is $E\left[\nabla_{\gamma} q' \nabla_{\delta} q\right]$. A consistent estimator of this component is

$$\tilde{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q\right] = \frac{\sum_{i=1}^{n} \nabla_{\gamma} q(\hat{\delta}, \tilde{\gamma}, Z_i)' \nabla_{\delta} q(, \tilde{\gamma}, Z_i)}{n} \tag{8}$$

where $\hat{\delta}$ and $\tilde{\gamma}$ denote the first and second stage estimators, respectively. Therefore, when the second stage is MLE, a consistent estimator of D is

$$\tilde{D} = \begin{bmatrix} \tilde{D}_{11} & \tilde{D}_{12} \\ \tilde{D}_{12}' & \tilde{D}_{22} \end{bmatrix}$$

where

$$\tilde{D}_{11} = \widetilde{AVAR}(\hat{\delta})$$

$$\tilde{D}_{12} = \widetilde{AVAR}(\hat{\delta}) \, \tilde{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q\right]' \, \widetilde{AVAR}*(\tilde{\gamma})$$

$$\tilde{D}_{22} = \widetilde{AVAR}*(\tilde{\gamma}) \, \tilde{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q\right] \widetilde{AVAR}(\hat{\delta}) \, \tilde{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q\right]' \widetilde{AVAR}*(\tilde{\gamma}) + \widetilde{AVAR}*(\tilde{\gamma}) \tag{9}$$

and $\widetilde{AVAR}(\hat{\delta})$ and $\widetilde{AVAR}*(\tilde{\gamma})$ are the estimated covariance matrices obtained from the first and second stage packaged regression outputs, respectively. So, for example, the "t-statistic"

$(\tilde{\gamma}_k - \gamma_k) / \sqrt{\tilde{D}_{22(k)}}$ for the kth element of γ is asymptotically standard normally distributed and can be used to test the hypothesis that $\gamma_k = \gamma_k^0$ for γ_k^0, a given null value of γ_k.

On the other hand, when second stage is NLS, $q(\delta, \gamma, Z_i) = -(Y_i - J(\delta, \gamma, V_i))^2$ and $E\left[\nabla_{\gamma\delta} q\right]$ and $E\left[\nabla_{\gamma\gamma} q\right]$ can be consistently estimated using

$$\hat{E}\left[\nabla_{\gamma\delta} q\right] = \frac{\sum_{i=1}^{n} \nabla_\gamma J(\hat{\delta}, \hat{\gamma}, V_i)' \nabla_\delta J(\hat{\delta}, \hat{\gamma}, V_i)}{n} \tag{10}$$

and

$$\hat{E}\left[\nabla_{\gamma\gamma} q\right] = \frac{\sum_{i=1}^{n} \nabla_\gamma J(\hat{\delta}, \hat{\gamma}, V_i)' \nabla_\gamma J(\hat{\delta}, \hat{\gamma}, V_i)}{n}. \tag{11}$$

respectively, where V_i denotes the ith observation on V, and $\hat{\delta}$ and $\hat{\gamma}$ denote the first and second stage estimators, respectively. Therefore, when the second stage is NLS, a consistent estimator of D is

$$\hat{D} = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}_{12}' & \hat{D}_{22} \end{bmatrix}$$

where

$$\hat{D}_{11} = \widehat{AVAR}(\hat{\delta})$$

$$\hat{D}_{12} = - \widehat{AVAR}(\hat{\delta}) \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$\hat{D}_{22} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR}(\hat{\delta}) \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} + \widehat{AVAR}^*(\hat{\gamma}) \tag{12}$$

and $\widehat{AVAR}(\hat{\delta})$ and $\widehat{AVAR}^*(\hat{\gamma})$ are the estimated covariance matrices obtained from the first and second stage packaged regression outputs, respectively.

3. Some Useful Two-Stage Optimization Estimators

Here we discuss a few 2SOE that can be used in empirical contexts involving endogeneity. These methods are designed to correct for endogeneity bias and, therefore, allow for causal interpretation of regression results. These methods are particularly useful for retrospective and prospective empirical analysis of health policy because they produce results that are causally interpretable.

3.1 Two-Stage Residual Inclusion

Suppose the researcher is interested in estimating the effect that a policy variable of interest X_p has on a specified outcome Y. Moreover, suppose that the data on X_p is sampled endogenously – i.e. it is correlated with an unobservable variable X_u that is also correlated with Y. To formalize this, we follow Terza et al. (2008), and assume that the data generating process has the following components

$$E[Y \mid X_p, X_o, X_u] = \mu(X_p, X_o, X_u; \beta) \quad \text{[outcome regression]} \tag{13}$$

and

$$X_p = r(W, \alpha) + X_u \quad \text{[auxiliary regression]} \tag{14}$$

where X_o denotes a vector of observable confounders (observable variables that are possibly correlated with both Y and X_p), β and α are parameters vectors, $W = [X_o \;\; W^+]$, W^+ is an identifying instrumental variable, and μ( ) and r( ) are known functions. Because the set of confounders (X_o and X_u, respectively), is comprehensive (i.e. includes all possible confounders), we can show that as a special case of the extended potential outcomes framework developed by Terza (2012), the model in (13) and (14) can be used for causal analysis. The true causal regression model corresponding to (13) is[6]

$$Y = \mu(X_p, X_o, X_u; \beta) + e \tag{15}$$

where e is the random error term, tautologically defined as $e = Y - \mu(X_p, X_o, X_u; \beta)$. The β parameters in expression (15) are not directly estimable (e.g. by NLS) due to the presence of the unobservable confounder X_u. The following 2SOE is, however, feasible.

First Stage: Obtain a consistent estimate of α by applying NLS to (14) and compute the residuals as

$$\hat{X}_u = X_p - r(W, \hat{\alpha}) \tag{16}$$

where $\hat{\alpha}$ is the first-stage estimate of α.

Second Stage: Estimate β by applying NLS to

$$Y = \mu(X_p, X_o, \hat{X}_u; \beta) + e^{2SRI} \tag{17}$$

where e^{2SRI} denotes the regression error term. Terza et al. (2008) call this method two-stage residual inclusion (2SRI).

In order to detail the asymptotic covariance matrix of this 2SRI estimator, we cast it in the framework of the generic 2SOE discussed above. This version of the 2SRI estimator implements NLS in its second stage. Therefore, expressions (10) through (12) are relevant, with α and β playing the roles of δ and γ, respectively, and $q(\hat{\delta}, \hat{\gamma}, Z_i)$ replaced by

---

[6] See Terza (2012) for the strict definition of *true causal model*.

$$q(\hat{\alpha}, \beta, Y, X_p, W) = -\left(Y - \mu(X_p, X_o, \hat{X}_u; \beta)\right)^2. \tag{18}$$

Specific illustrations of (13) through (18) and (10) through (12) in this context will be given in the next section.

It should be noted here that MLE can be implemented in either of the stages of the 2SRI method. For MLE to be implemented in the first stage, the primitive in (14) must be replaced by an assumption which specifies a known form for the conditional density of $(X_p \mid W)$, say $g(X_p \mid W; \alpha)$. Such an assumption would, of course, imply a formulation for the conditional mean $E[X_p \mid W]$, say $r(W, \alpha)$. Therefore, in this case, the first stage of the estimator would consist of maximizing (1) with $q_1(\delta, V_i)$ replaced by $\ln[g(X_{pi} \mid W_i; \alpha)]$ and subsequently computing the residuals as in (16). For MLE to be implemented in the second stage, the primitive in (13) must be replaced by an assumption which specifies a known form for the conditional density of $(Y \mid X_p, X_o, X_u)$, say $f(Y \mid X_p, X_o, X_u; \beta)$. The second stage of the estimator would then consist of maximizing (2) with $q(\hat{\delta}, \gamma, Z_i)$ replaced by $\ln[f(Y_i \mid X_{pi}, X_{oi}, \hat{X}_{ui}; \beta)]$. To obtain the correct asymptotic covariance matrix, the expressions in (6), (8) and (9) would be appropriately specified to accommodate the log-likelihood form of $q(\ )$.

3.2 A Two-Stage Estimator for Nonlinear Models Involving Endogenous Sample Selection

Here again, we suppose the researcher is interested in estimating the effect that a policy variable of interest X_p has on a specified outcome Y. In this case, structure of the model is nearly the same as that developed in section 3.1 above. There are, however, two important differences. First, the observability of the outcome of interest (Y) for each member of the relevant population is assumed to be determined by a binary sample selection variable, X_s, that is endogenous (correlated with the unobservable confounder X_u) and does not appear in the true causal regression specification for the outcome conditional on the confounders. The outcome regression in (13) is, therefore, replaced with

$$E[Y \mid X_p, X_o, X_u, X_s] = \mu(X_p, X_o, X_u; \tau) \tag{19}$$

where τ is a vector of unknown parameters. Secondly, we formalize the correlation between X_s and X_u as

$$X_s = I(W\theta + X_u > 0) \tag{20}$$

where $W = [X_p \quad X_o \quad W^+]$, W^+ is a vector of identifying instrumental variables, and $(X_u \mid W)$ has a known distribution. Note that X_p is included among the instruments here because it is assumed to be exogenous (the source of endogeneity in this case is X_s). Terza (2009) shows that (19) and (20) imply

$$E[Y \mid W, X_s = 1] = \frac{\int_{-W\theta}^{\infty} \mu(X_p, X_o, X_u; \tau) g(X_u \mid W)\, dX_u}{1 - G(-W\theta \mid W)} \tag{21}$$

where g( ) and G( ) denote the pdf and cdf of $(X_u \mid W)$, respectively. This motivates the following consistent two-stage estimator:

First Stage:

Estimate θ by applying appropriate MLE to $X_s = I(W\theta + X_u > 0)$ using the full sample.

Second Stage:

Estimate τ by applying NLS to the following nonlinear regression model motivated by (21) using the subsample of observations for whom $X_s = 1$

$$Y = \frac{\int_{-W\hat{\theta}}^{\infty} \mu(X_p, X_o, X_u; \tau) g(X_u \mid W)\, dX_u}{1 - G(-W\hat{\theta} \mid W)} + \upsilon \tag{22}$$

where $\hat{\theta}$ is the first-stage estimate of θ and υ is the regression error term.[7]

In order to detail the asymptotic covariance matrix of this estimator, we cast it in the framework of the generic 2SOE discussed above. Because NLS is implemented in the second stage, expressions (10) through (12) are relevant, with θ and τ playing the roles of δ and γ, respectively, and $q(\hat{\delta}, \hat{\gamma}, Z_i)$ replaced by

$$q(\hat{\theta}, \tau, Y, X_p, W) = -\left(Y - \hat{E}[Y \mid W, X_s = 1]\right)^2 \tag{23}$$

where $\hat{E}[Y \mid W, X_s = 1]$ is the same as (21) with θ replaced with $\hat{\theta}$. Specific illustrations of expressions (10) through (12) in this context will be given in the next section. Here, as for the 2SRI estimator, the second stage can be MLE. In this case, (19) must be replaced by an assumption which specifies a known form for the conditional density of $(Y \mid X_p, X_o, X_u, X_s)$, say $h(Y \mid X_p, X_o, X_u; \beta)$. The second stage of the estimator would then consist of maximizing (2) with $q(\hat{\delta}, \gamma, Z_i)$ replaced by the appropriate log-likelihood form based on h(|). To obtain

[7] The requisite integral for (20) can be evaluated using quadrature or simulation approximation.

the correct asymptotic covariance matrix, the expressions in (6), (8) and (9) would be appropriately specified to accommodate the log-likelihood form of q( ).

3.3 Multi-Stage Causal Effect Estimators

For contexts in which the policy variable of interest (X_p) is qualitative (binary), Rubin (1974, 1977) developed the *potential outcomes framework (POF)* which facilitates clear definition and interpretation of various policy relevant treatment effects. Terza (2012) extends the POF to encompass contexts in which X_p is quantitative (discrete or continuous) and planned policy changes in X_p are incremental or infinitesimal. Correspondingly, as counterparts to the *average treatment effect* in the POF, Terza (2012) defines the *average incremental effect* and the *average marginal effect*, respectively, as[8]

$$\text{AIE}\left(\Delta(X_{p1})\right) = E[Y_{X_{p1}+\Delta(X_{p1})}] - E[Y_{X_{p1}}] \tag{24}$$

and

$$\text{AME} = \lim_{\Delta \to 0} \frac{\text{AIE}(\Delta)}{\Delta} \tag{25}$$

where X_{p1} denotes the pre-policy version if X_p (a random variable), $\Delta(X_{p1})$ denotes the policy mandated exogenous increment to the policy variable, and $Y_{X_p^*}$ denotes the potential outcome (a random variable) -- the version of the outcome that would obtain if the policy variable were exogenously and counterfactually set at X_p^*.[9]

Terza (2012) shows that under a primitive regression assumption like (13) [or (19)], if we can consistently estimate the parameters of the model (τ) and can find an appropriate (consistent)

[8] Note that $\text{AIE}(\Delta)$ is defined as in (24) with $\Delta(X_{p1}) \equiv \Delta$, a constant.

[9] For details of the extended potential outcomes framework, see Terza (2012).

way to proxy X_u then (24) and (25) can be consistently estimated using

$$\widehat{AIE}(\Delta(X_{pli})) = \sum_{i=1}^{n} \frac{1}{n} \left\{ \mu(X_{pli} + \Delta_i(X_{pli}), X_{oi}, \hat{X}_{ui}; \hat{\tau}) - \mu(X_{pli}, X_{oi}, \hat{X}_{ui}; \hat{\tau}) \right\} \tag{26}$$

$$\widehat{AME} = \sum_{i=1}^{n} \frac{1}{n} \frac{\partial \mu(X_{pli}, X_{oi}, \hat{X}_{ui}; \hat{\tau})}{\partial X_{pli}} \tag{27}$$

where $\hat{\tau}$ is a consistent estimate of τ, $\hat{X}_{ui}$ is the proxy value for X_u, and the i subscript denotes the observation for the ith individual in a sample of size n (i = 1, …, n). In (26) and (27) we assume that we can directly proxy X_u, as would be the case if we estimated the model via the 2SRI method. In the two-stage sample selection (2SSS) model detailed in section 3.2, no such direct proxy for X_u can be implemented. In the 2SSS model, however, the distribution of $(X_u \mid W)$ is assumed to be known so we can write the relevant versions of (26) and (27) as, respectively

$$\widehat{AIE}(\Delta(X_{pli})) = \sum_{i=1}^{n} \frac{1}{n} \left\{ \int_{-\infty}^{\infty} \left\{ \mu(X_{pli} + \Delta_i(X_{pli}), X_{oi}, X_u; \hat{\tau}) - \mu(X_{pli}, X_{oi}, X_u; \hat{\tau}) \right\} g(X_u \mid W)\, dX_u \right\} \tag{28}$$

$$\widehat{AME} = \sum_{i=1}^{n} \frac{1}{n} \left[\int_{-\infty}^{\infty} \left\{ \frac{\partial \mu(X_{pli}, X_{oi}, X_u; \hat{\tau})}{\partial X_{pli}} \right\} g(X_u \mid W)\, dX_u \right] \tag{29}$$

where $g(X_u \mid W)$ is the known pdf of $(X_u \mid W)$.

We now turn to the asymptotic properties of these estimators. We use the notation “PE” to denote the relevant policy effect [(24) or (25)] and rewrite (26) and (27) in generic form as

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \tag{30}$$

where $\widehat{pe}_i(\hat{\alpha}, \hat{\beta})$ is shorthand notation for $pe(X_{p1i}, X_{oi}, \hat{X}_{ui}(\hat{\alpha}, W_i), \hat{\beta})$. In cases like 2SRI, wherein X_u can be directly proxied using the first-stage estimate ($\hat{\alpha}$) and the instrumental variables (W_i), we have

$$pe(X_{p1}, X_o, X_u(\alpha, W), \beta) = \begin{cases} \mu(X_{p1} + \Delta(X_{p1}), X_o, X_u(\alpha, W), \beta) - \mu(X_{p1}, X_o, X_u(\alpha, W), \beta) & \text{for (26)} \\ \dfrac{\partial \mu(X_{p1}, X_o, X_u(\alpha, W), \beta)}{\partial X_{p1}}. & \text{for (27)} \end{cases}$$

Similarly, we rewrite (28) and (29) in generic form as

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i(\hat{\tau})}{n} \tag{31}$$

where $\widehat{pe}_i(\hat{\tau})$ is shorthand notation for $pe(X_{p1i}, X_{oi}, \hat{\tau})$ for cases like 2SSS in which X_u cannot be directly proxied and

$$pe(X_{p1}, X_o, \tau) = \begin{cases} \int_{-\infty}^{\infty} \left\{ \mu(X_{p1} + \Delta(X_{p1}), X_o, X_u; \tau) - \mu(X_{p1}, X_o, X_u; \tau) \right\} g(X_u \mid W)\, dX_u & \text{for (28)} \\ \int_{-\infty}^{\infty} \left\{ \dfrac{\partial \mu(X_{p1}, X_o, X_u; \tau)}{\partial X_{p1}} \right\} g(X_u \mid W)\, dX_u. & \text{for (29)} \end{cases}$$

In order to derive the asymptotic properties of (30) and (31) we cast them as 2SOE.

The first stage of our 2SOE characterization of (30) comprises consistent estimation of α and β (e.g. via 2SRI). The second stage of the estimator [i.e., (30) itself] is easily shown to be the optimizer of the following objective function

$$\sum_{i=1}^{n} q(\hat{\alpha}, \hat{\beta}, \text{PE}, Z_i) \tag{32}$$

where

$$q(\hat{\alpha}, \hat{\beta}, \text{PE}, Z_i) = -\left(\widehat{pe}_i(\hat{\alpha}, \hat{\beta},) - \text{PE}\right)^2 \tag{33}$$

$Z_i = [Y_i \quad X_{pli} \quad W_i]$ and $\hat{\tau}$ is the first-stage estimator of τ. Because the second stage of this 2SOE implements NLS, expressions (7) and (10) through (12) are relevant, with $[\alpha' \quad \beta']$ and PE playing the roles of δ and γ, respectively. In this case (10) and (11) become, respectively

$$\hat{E}\left[\nabla_{PE\,[\hat{\alpha} \quad \hat{\beta}]} q\right] = \frac{\sum_{i=1}^{n} \nabla_{PE[\alpha' \quad \beta']} q(\hat{\alpha}, \hat{\beta}, \text{PE}, Z_i)}{n} = \frac{-2\sum_{i=1}^{n} \nabla_{[\alpha' \quad \beta']} \widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \tag{34}$$

and

$$\hat{E}\left[\nabla_{PE\,PE} q\right] = \frac{\sum_{i=1}^{n} \nabla_{PE\,PE} q(\hat{\tau}, \widehat{PE}, Z_i)}{n} = 2\,. \tag{35}$$

where $[\hat{\alpha}' \quad \hat{\beta}']$ and $\widehat{PE}$ denote the first and second stage estimators, respectively. Note also, that in this case

$$\widehat{AVAR}^*(\widehat{PE}) = \frac{\sum_{i=1}^{n} \left(\widehat{pe}_i(\hat{\alpha}, \hat{\beta}) - \widehat{PE}\right)^2}{n}. \tag{36}$$

Combining (34) through (36) with (12) we obtain a consistent estimate of the correct asymptotic variance of (30) as

$$\widehat{\text{a var}}(\widehat{\text{PE}}) = \left(\frac{\sum_{i=1}^{n} \nabla_{[\alpha' \;\; \beta']} \widehat{\text{pe}}_i(\hat{\alpha}, \hat{\beta})}{n} \right) \widehat{\text{AVAR}}([\hat{\alpha}' \;\; \hat{\beta}']) \left(\frac{\sum_{i=1}^{n} \nabla_{[\alpha' \;\; \beta']} \widehat{\text{pe}}_i(\hat{\alpha}, \hat{\beta})}{n} \right)' + \frac{\sum_{i=1}^{n} \left(\widehat{\text{pe}}_i(\hat{\alpha}, \hat{\beta}) - \widehat{\text{PE}} \right)^2}{n} \quad (37)$$

where $\sum_{i=1}^{n} \nabla_{[\alpha' \;\; \beta']} \widehat{\text{pe}}_i(\hat{\alpha}, \hat{\beta})$ denotes $\nabla_{[\alpha' \;\; \beta']} \text{pe}(X_{p1}, X_o, X_u(\alpha, W), \beta)$ evaluated at X_{pi}, X_{oi}, W_i, and $[\hat{\alpha}' \;\; \hat{\beta}']$ and $\widehat{\text{AVAR}}([\hat{\alpha}' \;\; \hat{\beta}'])$ is the estimated asymptotic covariance matrix of $[\hat{\alpha}' \;\; \hat{\beta}']$. So, for example, the "t-statistic" $\sqrt{n}(\widehat{\text{PE}} - \text{PE}) / \sqrt{\widehat{\text{a var}}\left(\widehat{\text{PE}}\right)}$ is asymptotically standard normally distributed and can be used to test the hypothesis that $\text{PE} = \text{PE}^0$ for PE^0, a given null value of PE.[10]

We can similarly establish a consistent estimate of the correct asymptotic variance of (31) as

$$\widehat{\text{a var}}(\widehat{\text{PE}}) = \left(\frac{\sum_{i=1}^{n} \nabla_{\tau} \widehat{\text{pe}}_i(\hat{\tau})}{n} \right) \widehat{\text{AVAR}}(\hat{\tau}) \left(\frac{\sum_{i=1}^{n} \nabla_{\tau} \widehat{\text{pe}}_i(\hat{\tau})}{n} \right)' + \frac{\sum_{i=1}^{n} \left(\widehat{\text{pe}}_i(\hat{\tau}) - \widehat{\text{PE}} \right)^2}{n} \quad (38)$$

where $\nabla_{\tau} \widehat{\text{pe}}_i(\hat{\tau})$ denotes $\nabla_{\tau} \text{pe}(X_{p1}, X_o, \tau)$ evaluated at X_{p1i}, X_{oi} and $\hat{\tau}$; and $\widehat{\text{AVAR}}(\hat{\tau})$ is the estimated asymptotic covariance matrix of $\hat{\tau}$. So, for example, the "t-statistic" $\sqrt{n}(\widehat{\text{PE}} - \text{PE}) / \sqrt{\widehat{\text{a var}}\left(\widehat{\text{PE}}\right)}$ is asymptotically standard normally distributed and can be used to test the hypothesis that $\text{PE} = \text{PE}^0$ for PE^0, a given null value of PE.

---

[10] The analysis in this section encompasses cases in which X_p is either endogenous or exogenous -- the latter is characterized by the absence of X_u (no unobservable confounders). Therefore, the result obtained by Basu and Rathouz (2005) for the asymptotic standard error of the average marginal effect when X_p is *exogenous* can easily be shown to be a special case of the more general 2SOE approach taken here.

4. Illustrations

4.1 Smoking During Pregnancy and Infant Birthweight: Parameter Estimation via 2SRI

Using the 2SRI method, we re-estimate the regression model of Mullahy (1997) in which

Y = infant birthweight in lbs

X_p = number of cigarettes smoked per day during pregnancy

and show, in detail, how to obtain the correct asymptotic standard errors for the parameter estimates. In this illustration the relevant versions of the outcome and auxiliary regressions in (13) and (14) are

$$E[Y \mid X_p, X_o, X_u] = \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u) \tag{39}$$

$$X_p = \exp(W\alpha) + X_u . \tag{40}$$

We applied NLS in both of the stages of 2SRI so the first and second stage objective functions [(1) and (2)] are

$$q_1(\alpha, V_i) = -(X_{pi} - \exp(W_i\alpha))^2$$

$$q(\alpha, \beta, Z_i) = -(Y_i - \exp(X_{pi}\beta_p + X_o\beta_o + (X_{pi} - \exp(W_i\alpha))\beta_u))^2 .$$

The first and second stage 2SRI parameter estimates ($\hat{\alpha}$ and $\hat{\beta} = [\hat{\beta}_p \;\; \hat{\beta}'_o \;\; \hat{\beta}_u]$, respectively) were obtained in Stata by applying the GLM procedure with the "family(gaussian)" and "link(log)" options. After each of the stages, we then saved the parameter vectors ($\hat{\alpha}$ and $\hat{\beta}$) and their corresponding "packaged" covariance matrix estimators ($\widehat{AVAR}(\hat{\alpha})$ and $\widehat{AVAR}^*(\hat{\beta})$)

in MATA. Using MATA, we then calculated the $n \times \dim(W_i)$ matrix whose ith row is

$$\nabla_\alpha J(\hat{\alpha},\hat{\beta},Z_i) = -2\hat{\beta}_u \exp(X_i\hat{\beta})\exp(W_i\hat{\alpha})W_i$$

and the $n \times \dim(X_i)$ matrix whose ith row is

$$\nabla_\beta J(\hat{\alpha},\hat{\beta},Z_i) = 2\exp(X_i\hat{\beta})X_i$$

where dim(A) denotes the row dimension of the vector A, and $X_i = [X_{pi} \quad X_{oi} \quad \hat{X}_{ui}]$. Finally, we estimated the asymptotic covariance matrix of $\hat{\omega}' = [\hat{\alpha}' \quad \hat{\beta}]$ as

$$\hat{D} = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}'_{12} & \hat{D}_{22} \end{bmatrix}$$

where[11]

$$\hat{D}_{11} = \widehat{AVAR}(\hat{\alpha})$$

$$\hat{D}_{12} = -\widehat{AVAR}(\hat{\alpha})\hat{E}\left[\nabla_{\beta\alpha}q\right]'\hat{E}\left[\nabla_{\beta\beta}q\right]^{-1}$$

$$\hat{D}_{22} = \hat{E}\left[\nabla_{\beta\beta}q\right]^{-1}\hat{E}\left[\nabla_{\beta\alpha}q\right]\widehat{AVAR}(\hat{\alpha})\hat{E}\left[\nabla_{\beta\alpha}q\right]'\hat{E}\left[\nabla_{\beta\beta}q\right]^{-1} + \widehat{AVAR}*(\hat{\beta}).$$

$$\hat{E}\left[\nabla_{\beta\alpha}q\right] = \frac{\sum_{i=1}^{n}\nabla_\beta J(\hat{\alpha},\hat{\beta},Z_i)'\nabla_\alpha J(\hat{\alpha},\hat{\beta},Z_i)}{n} = \frac{\sum_{i=1}^{n}\hat{\beta}_u \exp(X_i\hat{\beta})^2\exp(W_i\hat{\alpha})X_i'W_i}{n} \tag{41}$$

and

$$\hat{E}\left[\nabla_{\beta\beta}q\right] = \frac{\sum_{i=1}^{n}\nabla_\beta J(\hat{\alpha},\hat{\beta},Z_i)'\nabla_\beta J(\hat{\alpha},\hat{\beta},Z_i)}{n} = \frac{\sum_{i=1}^{n}\exp(X_i\hat{\beta})^2 X_i'X_i}{n}. \tag{42}$$

The relevant lines of MATA code are:

---

[11] Expressions (41) and (42) are the relevant versions of (10) and (11).

$W_i\hat{\alpha}$: `walpha=W*alpha`

$X_i\hat{\beta}$: `xbeta=X*beta`

$\nabla_\alpha J(\hat{\alpha},\hat{\beta},Z_i)$: `pJaq=2*bu*exp(xbeta):*exp(walpha):*W`

$\nabla_\beta J(\hat{\alpha},\hat{\beta},Z_i)$: `pJbq=2*exp(xbeta):*X`

$\hat{E}\left[\nabla_{\beta\alpha}q\right]$: `pbaq=pJbq’*pJaq`

$\hat{E}\left[\nabla_{\beta\beta}q\right]$: `pbbq pJbq’*pJbq`

$\hat{D}_{11}$: `D11=avaralpha`

$\hat{D}_{12}$: `D12= avaralpha*pbaq'*luinv(pbbq)`

$\hat{D}_{22}$:

```
D22=luinv(pbbq)*pbaq*avaralpha*pbaq'*luinv(pbbq)+avarbetastar
```

$\hat{D}$: `D=D11, D12 \ D12', D22.`

The 2SRI results are given in Table 1.

Table 1: GLM Exponential Condition Mean NLS Regression w/ Corrected St. Errors

	variable	estimate	t-stat	p-value
1	variable	estimate	t-stat	p-value
2				
3	CIGSPREG	-.0140086	-3.678995	.0002342
4	PARITY	.0166603	3.180623	.0014696
5	WHITE	.0536269	4.217293	.0000247
6	MALE	.0297938	3.130267	.0017465
7	xuhat	.0097786	2.557676	.0105374
8	constant	1.948207	117.6448	0

For comparison, the second stage estimates with packaged GLM standard errors are given in Table 2.

Table 2: GLM Exponential Condition Mean NLS Regression w/ Uncorrected St. Errors

BIRTHWTLB	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
CIGSPREG	-.0140086	.0034369	-4.08	0.000	-.0207447	-.0072724
PARITY	.0166603	.0048853	3.41	0.001	.0070854	.0262353
WHITE	.0536269	.0117985	4.55	0.000	.0305023	.0767516
MALE	.0297938	.0088815	3.35	0.001	.0123864	.0472011
xuhat	.0097786	.0034545	2.83	0.005	.003008	.0165492
_cons	1.948207	.0157445	123.74	0.000	1.917348	1.979066

Note the differences in the t-statistics.

4.2 Depression and Income for US Adults: Estimation via 2SSS

The underlying model is

$$\textit{hurdle}: \; X_s = I(X_p\beta_{p1} + X_o\beta_{o1} + X_u > 0) \tag{43}$$

$$\textit{levels}: \; Y^{\ell} = \exp(X_p\beta_{p2} + X_o\beta_{o2} + X_u\beta_{u2} + \varepsilon_2) \tag{44}$$

where

$X_s \equiv$ 1 if the individual is employed, 0 otherwise

$Y^{\ell} \equiv$ income (latent if $X_s = 0$)

$X_p \equiv$ number of depressive symptoms

$X_o \equiv$ the vector of observable control variables (observable confounders)

$X_u \equiv$ a scalar comprising the unobservable confounders

$(\varepsilon_1 \mid X_p^*, X_o, X_u) \sim n(0, 1)$

$E[\exp(\varepsilon_2) \mid X_p^*, X_o, X_u] = 1$

and I(C) denotes the indicator function whose value is 1 if condition C holds and 0 otherwise.[12]

THE REMAINDER OF THIS SECTION IS YET TO BE COMPLETED.

4.3 Average Incremental Effect of Smoking During Pregnancy on Infant Birthweight

---

[12] Note that the standard normality assumption for $(\varepsilon_1 \mid X_p^*, X_o, X_u)$ is not required. Any distributional assumption will suffice here. The normal and logistic are typical.

To follow up our analysis in section 4.1, we estimate the average incremental effect (AIE) of a policy that would cause current levels of smoking during pregnancy to fall to zero for everyone in the relevant population. In the notation of section 3.3, we have that the pre- and post-policy versions of the policy variable are $X_{p1} = X_p$ and $X_{p2} = X_p + \Delta(X_p)$, respectively, where $\Delta(X_p) = -X_p$. Moreover, using (30) we have that the AIE estimator is

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \tag{45}$$

where $\widehat{pe}_i(\hat{\alpha}, \hat{\beta})$ is $pe(X_{p1}, X_o, X_u(\alpha, W), \beta)$ evaluated at X_{pi}, X_{oi}, W_i, and $[\hat{\alpha}' \quad \hat{\beta}']$, with

$$pe(X_{p1}, X_o, X_u(\alpha, W), \beta) = \exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_o\hat{\beta}_o + X_u\hat{\beta}_u) - \exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_o\hat{\beta}_o + X_u\hat{\beta}_u).$$

Using (37), we obtain the correct asymptotic standard error of (45) as

$$\widehat{a\,var}(\widehat{PE}) = \left(\frac{\sum_{i=1}^{n} \nabla_{[\alpha' \quad \beta']} \widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \right) \hat{D} \left(\frac{\sum_{i=1}^{n} \nabla_{[\alpha' \quad \beta']} \widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \right)' + \frac{\sum_{i=1}^{n} \left(\widehat{pe}_i(\hat{\alpha}, \hat{\beta}) - \widehat{PE} \right)^2}{n}$$

$$\nabla_{[\alpha' \quad \beta']} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) = [\nabla_{\alpha} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) \quad \nabla_{\beta_p} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) \quad \nabla_{\beta_o} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) \quad \nabla_{\beta_u} \widehat{pe}_i(\hat{\alpha}, \hat{\beta})]$$

$$\begin{aligned} \nabla_{\alpha} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) = \; & -\exp(W_i\hat{\alpha})\hat{\beta}_u \Big[\exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) \\ & - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) \Big] W_i \end{aligned}$$

$$\begin{aligned} \nabla_{\beta_p} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) = \; & \exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) \\ & - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) X_{pi} \end{aligned}$$

$$\nabla_{\beta_o} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) = \Big[\exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) \Big] X_{oi}$$

$$\nabla_{\beta_u} \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) = \Big[\exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) \Big] \hat{X}_{ui}$$

.

The relevant lines of MATA code are:

$[X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u$:

```
x1incb1=X1INC*beta
```

$\widehat{pe}_i(\hat{\alpha}, \hat{\beta})$:

```
pei=exp(x1incb1):-exp(x1b1)
```

$\widehat{PE}$:

```
pe=mean(pei)
```

$\nabla_{\alpha}\widehat{pe}_i(\hat{\alpha}, \hat{\beta})$:

```
palfa=-exp(walpha):*bxu:*pei:*W
```

$\nabla_{\beta_p}\widehat{pe}_i(\hat{\alpha}, \hat{\beta})$:

```
pbetap=exp(x1incb1):*xpinc:-exp(x1b1):*xp
```

$[\nabla_{\beta_o}\widehat{pe}_i(\hat{\alpha}, \hat{\beta}) \quad \nabla_{\beta_u}\widehat{pe}_i(\hat{\alpha}, \hat{\beta})]$:

```
pbetao=pei:*X0 [NOTE THAT X0 INCLUDES Xu]
```

$\left(\frac{\sum_{i=1}^{n}\nabla_{[\alpha' \quad \beta']}\widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n}\right)$:

```
ppe=mean(palfa),mean(pbetap),mean(pbetao)
```

$\widehat{a\,var}(\widehat{PE})$:

```
varpe=ppe*(n:*D)*ppe':+mean((pei:-pe):^2).
```

The results are given in Table 3

Table 3: AIE of Eliminating Smoking During Pregnancy w/ Corrected St. Errors

	%smoke-decr	incr-effect	std-err	t-stat	p-value
1	%smoke-decr	incr-effect	std-err	t-stat	p-value
2					
3	100	.2300237	.0726222	3.167401	.0015381

The results indicate that a 100% decrease in smoking for every pregnant woman in the population would cause an average increase in birthweight of nearly a quarter of a pound.

THE REMAINDER OF THE PAPER IS YET TO BE COMPLETED.

References

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

Cameron, A.C. and Trivedi, P.K. (2005): *Microeconometrics: Methods and Applications*," New York: Cambridge University Press.

Greene (2012): *Econometric Analysis, 7th Edition*, Upper Saddle River, NJ: Pearson, Prentice-Hall.

Mullahy, J. (1997): "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics*, 79, 586-593.

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

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

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

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

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

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

_________ (2012): "Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight," unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

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

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

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