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

Joseph V. Terza
Department of Economics
Indiana University Purdue University Indianapolis
Indianapolis, IN 46202
Email: jvterza@iupui.edu

(November, 2014)

Please cite:

Terza, J.V. (2014): "Simpler Standard Errors for Multi-Stage Regression-Based Estimators: Illustrations in Health Economics," Keynote Address, Mexican Stata Users Group Meeting, Mexico City.

Motivation

-- Focus here is on two-stage optimization estimators (2SOE)

-- Asymptotic theory for 2SOE (correct standard errors) available for many years

-- Both stages are maximum likelihood estimators (MLE)

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.**

-- More general cases

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.**

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

Motivation (cont'd)

-- Textbook treatments of the subject

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

Greene (2008): *Econometric Analysis, 6th Edition*, Upper Saddle River, NJ: Pearson, Prentice-Hall.

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

-- Nonetheless, applied researchers often implement bootstrapping methods or ignore the two-stage nature of the estimator and report the uncorrected outputs from packaged statistical software.

Motivation (cont'd)

-- With a view toward easy software implementation (in Stata), we offer the practitioner a 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.

-- We cast the discussion in the context of regression models involving endogeneity – a sampling problem whose solution often requires a 2SOE.

Motivation (cont'd)

-- Examples of relevant methodological contexts involving endogeneity:

1) The two-stage residual inclusion (2SRI) estimator suggested by Terza et al. (2008) for nonlinear models with endogenous regressors

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

2) The two-stage sample selection estimator (2SSS) developed by Terza (2009) for nonlinear models with endogenous sample selection

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

3) Causal incremental and marginal effects estimators proposed by Terza (2014).

Terza, J.V. (2014): "Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight," Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

Motivation (cont'd)

-- In this presentation we will discuss (1) and (3) – 2SRI and Causal Effects

-- We will detail the analytics and Stata code for our simplified standard error formulae for both of these and give illustrative examples.

2SOE and Their Asymptotic Standard Errors

-- The parameter vector of interest is partitioned as $\omega' = [\delta' \quad \gamma']$ and estimated in two-stages:

-- First, an estimate of δ is obtained as the optimizer of an appropriately specified first-stage objective function

$$\sum_{i=1}^{n} q_1(\delta, V_{1i}) \qquad (1)$$

where V_{1i} denotes the relevant subvector of the observable data for the ith sample individual (i = 1, ..., n).

2SOE and Their Asymptotic Standard Errors (cont'd)

-- Next, an estimate of γ is obtained as the optimizer of

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

where V_{2i} denotes the relevant subvector of the observable data for the ith sample individual, and $\hat{\delta}$ denotes the first-stage estimate of δ.

-- Under fairly general conditions it can be shown that:

$$D^{-\frac{1}{2}}\sqrt{n}\left(\begin{bmatrix}\hat{\delta}\\ \hat{\gamma}\end{bmatrix}-\begin{bmatrix}\delta\\ \gamma\end{bmatrix}\right) = D^{-\frac{1}{2}}\sqrt{n}\left(\hat{\omega}-\omega\right)\xrightarrow{d} N(0,I)$$

i.e., $\hat{\omega} = [\hat{\delta} \quad \hat{\gamma}]$ is asymptotically normal with asymptotic covariance matrix D.

2SOE and Their Asymptotic Standard Errors: Some Notation

-- Rewriting the asymptotic covariance matrix of $\hat{\omega} = [\hat{\delta} \quad \hat{\gamma}]$ in partitioned form we get

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{12}' & \mathbf{D}_{22} \end{bmatrix} \quad (3)$$

where

$\mathbf{D}_{11} = \text{AVAR}^*(\hat{\delta})$ denotes the asymptotic covariance matrix of $\hat{\delta}$

$\mathbf{D}_{22} = \text{AVAR}(\hat{\gamma})$

$\mathbf{D}_{12}$ is left unspecified for the moment.

$\hat{\delta}$ and $\hat{\gamma}$ are the first and second stage estimators, respectively

and the "*" denotes the matrix two which the relevant "packaged" asymptotic covariance matrix estimator converges (by "packaged" we mean that which would be obtained from Stata ignoring the two-stage nature of the estimator.)

2SOE and Their Asymptotic Standard Errors (cont'd)

-- It is incorrect to ignore the two-stage nature of the estimator and use the "packaged" standard errors from the second-stage [i.e., the packaged estimator of D_{22} in (3) with D_{12} set equal to 0].

-- The problem is that the expressions for the correct asymptotic covariance matrix of the generic 2SOE found in textbooks [Cameron and Trivedi (2005), Greene (2012), and Wooldridge (2010)] are daunting.

-- As a result, applied researchers opt for approximation methods like bootstrapping, or ignore the need for correction and report "packaged" results.

-- In the following, we offer a substantial simplification of the correct form of D (and its relevant partitions) that we hope will be useful to practitioners.

2SOE and Their Asymptotic Standard Errors: More Notation

-- q_1 is shorthand notation for $q_1(\delta, V_1)$ as defined in (1)

-- q is shorthand notation for $q(\delta, \gamma, V_2)$ as defined in (2)

-- $\nabla_s q$ denotes the gradient of q with respect to parameter subvector s. This is a row vector whose typical element is $\partial q / \partial s_j$; the partial derivative of q with respect to the jth element of s

-- $\nabla_{st} q$ denotes the Jacobian of $\nabla_s q$ with respect to t. This is a matrix whose typical element is $\partial^2 q / \partial s_j \partial t_m$; the cross partial derivative of q with respect to the jth element of s and the mth element of t – the row dimension of $\nabla_{st} q$ corresponds to that of its first subscript and the column dimension to that of its second subscript.

2SOE: An Example

-- For example, suppose the vector of observable data for the ith sample individual is $Z = [Y \quad X_p \quad X_o \quad W^+]$ where

$Y \equiv$ the outcome of policy interest

$X_p \equiv$ the policy variable of interest

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

$W^+ \equiv$ a vector of identifying instrumental variables.

-- Suppose our objective is to estimate the regression (broadly defined) of Y on $[X_p \quad X_o]$ purged of bias due to the potential endogeneity of X_p.

2SOE: An Example (cont'd)

-- A 2SOE of the following form might be appropriate:

***<u>First Stage</u>*: Consistently estimate δ via the nonlinear least squares (NLS) method.**

For example, we might use

$$q_1(\delta, V_{1i}) = -(X_{pi} - r(W_i\delta))^2$$

where $V_1 = [X_p \quad X_o \quad W^+]$, $W = [X_o \quad W^+]$ and r( ) is a known function.

***<u>Second Stage</u>*: Consistently estimate γ via a maximum likelihood estimator (MLE).**

For example, we might use

$$q(\hat{\delta}, \gamma, V_{2i}) = \ln\ f(Y_i | X_{pi}, W_i; \hat{\delta}, \gamma)$$

with $V_2 = Z$ and $f(Y \mid X_p, W; \delta, \gamma)$ being the relevant conditional density of Y.

2SOE and Their Asymptotic Standard Errors (cont'd)

-- The devil is, of course, in the "D"-tails (seek simple estimators of D_{12} and D_{22})

-- The typical textbook rendition of the "D"-tails is something like the following

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

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

where $AVAR^*(\hat{\delta})$ is the "packaged" and legitimate asymptotic covariance matrix of $\hat{\delta}$, and $AVAR^*(\hat{\gamma})$ is "packaged" but incorrect covariance matrix of $\hat{\gamma}$.

-- No need to define any of the components of this mess at this point. Just wanted to make a point.

Simple Standard Error Formulae – MLE

-- When the second stage estimator is MLE the correct (and practical) formulations of the estimators of D_{12} and D_{22} simplify as

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

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

where

$$\tilde{E}^{*}\left[\nabla_{\gamma} q' \nabla_{\delta} q\right] = \sum_{i=1}^{n} \nabla_{\gamma} q(\hat{\delta},\tilde{\gamma},V_{2i})' \nabla_{\delta} q(\hat{\delta},\tilde{\gamma},V_{2i})$$

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.

Simple Standard Error Formulae – NLS

-- When the second stage estimator is NLS such that

$$q(\delta, \gamma, V_{2i}) = -(Y_i - \mu(\delta, \gamma, V_{3i}))^2$$

where Y is a scalar element of V_2 and V_3 is a subvector of V_2 (not including Y), the correct formulations of the estimators of D_{12} and D_{22} simplify as

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

$$\hat{D}^*_{22} = \hat{E}^*\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}^*\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR}^*(\hat{\delta})\hat{E}^*\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} + \widehat{AVAR}^*(\hat{\gamma}) \qquad (4)$$

where

$$\hat{E}^*\left[\nabla_{\gamma\delta} q\right] = \sum_{i=1}^{n} \nabla_\gamma \mu(\hat{\delta}, \hat{\gamma}, V_{3i})' \nabla_\delta \mu(\hat{\delta}, \hat{\gamma}, V_{3i})$$

$$\hat{E}^*\left[\nabla_{\gamma\gamma} q\right] = \sum_{i=1}^{n} \nabla_\gamma \mu(\hat{\delta}, \hat{\gamma}, V_{3i})' \nabla_\gamma \mu(\hat{\delta}, \hat{\gamma}, V_{3i}).$$

Simple Standard Error Formulae – NLS (cont'd)

-- So, for example, the "t-statistic" $(\hat{\gamma}_k - \gamma_k)/\sqrt{\hat{D}^*_{22(k)}}$ 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, where $\hat{D}^*_{22(k)}$ denotes the kth diagonal element of $\hat{D}^*_{22}$.

Example: Two-Stage Residual Inclusion (2SRI)

-- 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 (an unobservable confounder).

Example: 2SRI (cont'd)

-- To formalize this, we follow Terza et al. (2008), and assume that

$$E[Y \mid X_p, X_o, X_u] = \mu(X_p, X_o, X_u; \beta) \quad \text{and} \quad X_p = r(W, \alpha) + X_u$$

[outcome regression] [auxiliary regression]

X_o denotes a vector of observable confounders (variables that are possibly correlated with both Y and X_p)

X_u is a scalar comprising the unobservable confounders

β and α are parameters vectors

$W = [X_o \quad W^+]$

W^+ is an identifying instrumental variable, and

$\mu(\quad)$ and $r(\quad)$ are known functions.

Example: 2SRI (cont'd)

-- The true causal regression model in this case is

$$Y = \mu(X_p, X_o, X_u; \beta) + e$$

where e is the random error term, tautologically defined as

$$e = Y - \mu(X_p, X_o, X_u; \beta).$$

-- The β parameters are not directly estimable (e.g. by NLS) due to the presence of the unobservable confounder X_u.

Example: 2SRI (cont'd)

The following 2SOE is, however, feasible and consistent.

<u>First Stage</u>: Obtain a consistent estimate of α by applying NLS to the auxiliary regression and compute the residuals as

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

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

<u>Second Stage</u>: Estimate β by applying NLS to

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

where e^{2SRI} denotes the regression error term.

Example: 2SRI (cont'd)

-- In order to detail the asymptotic covariance matrix of this 2SRI estimator, we cast it in the framework of the generic 2SOE discussed above with α and β playing the roles of δ and γ, respectively.

-- This version of the 2SRI estimator implements NLS in its first and second stages so the relevant versions of $q_1(\delta, V_1)$ and $q(\hat{\delta}, \gamma, V_2)$ are

$$q_1(\alpha, V_1) = -(X_p - r(W, \alpha))^2$$

and

$$q(\hat{\alpha}, \beta, V_2) = -\left(Y - \mu(X_p, X_o, \hat{X}_u; \beta)\right)^2$$

where $V_1 = [X_p \quad W]$ and $V_2 = [Y \quad X_p \quad W]$.

Smoking and Birthweight: Parameter Estimation via 2SRI

-- Re-estimate model of Mullahy (1997) using 2SRI

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.

Y = infant birthweight in lbs
X_p = number of cigarettes smoked per day during pregnancy

Outcome Regression

$$E[Y \mid X_p, X_o, X_u] = \mu(X_p, X_o, X_u; \beta) = \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)$$

Auxiliary Regression

$$X_p = \exp(W\alpha) + X_u \;\Rightarrow\; X_u = X_p - \exp(W\alpha)$$

$$q_1(\alpha, V_{1i}) = -(X_{pi} - \exp(W_i\alpha))^2$$

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

Smoking and Birthweight: Parameter Estimation via 2SRI (cont'd)

X_o = [1 PARITY WHITE MALE]
W = [X_o W^+]

W^+ = [EDFATHER EDMOTHER FAMINCOM CIGTAX88]

PARITY = birth order

WHITE = 1 if white, 0 otherwise

MALE = 1 if male, if female

EDFATHER = paternal schooling – yrs.

EDMOTHER = maternal schooling – yrs.

FAMINCOM = family income ($\times 10^{-3}$)

CIGTAX99 = per pack state excise tax on cigarettes.

Smoking and Birthweight: Parameter Estimation via 2SRI (cont'd)

-- Obtain the parameter estimates in both stages using the Stata "glm" command with "link(log)", "family(gaussian)" and "vce(robust)" options.

-- Using results for the case in which the 2nd stage of the 2SOE is NLS

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

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

and

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

Smoking and Birthweight: Parameter Estimation via 2SRI (cont'd)

where

$$\hat{E}^*\left[\nabla_{\beta\alpha} q\right] = -\sum_{i=1}^{n} \hat{\beta}_u \exp(X_i\hat{\beta})^2 \exp(W_i\hat{\alpha}) X_i{}' W_i$$

$$\hat{E}^*\left[\nabla_{\beta\beta} q\right] = \sum_{i=1}^{n} \exp(X_i\hat{\beta})^2 X_i{}' X_i$$

$X_i = [X_{pi} \quad X_{oi} \quad \hat{X}_{ui}]$, and $\widehat{AVAR}^*(\hat{\alpha})$ and $\widehat{AVAR}^*(\hat{\beta})$ are the estimated covariance matrices obtained from first and second stage GLM estimation, respectively.

2SRI Estimation -- Notes on Stata Implementation

-- Use MATA for calculation of the estimated asymptotic covariance matrix.

-- Use the st_matrix MATA command immediately after first and second stage GLM estimations to save $\widehat{AVAR}^*(\hat{\alpha})$ and $\widehat{AVAR}^*(\hat{\beta})$ as MATA matrices, e.g.

2SRI Estimation -- Notes on Stata Implementation: 1st Stage GLM

Stata Code

```
/**********************************************
** 2SRI Estimation begins here.              **
**********************************************/
/**********************************************
** First-stage NLS estimation of the auxiliary **
** exponential regression (via GLM).          **
** Conduct Wald test of joint significance of  **
** the instruments.                           **
** Save xuhat and the predicted values from the**
** regression                                 **
**********************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88, ///
family(gaussian) link(log) vce(robust)
test (EDFATHER = 0) (EDMOTHER = 0) (FAMINCOM = 0) (CIGTAX88 = 0)
predict xuhat, response
predict expwalpha, mu

/**********************************************
** Load the coefficient vector and covariance  **
** matrix from first-stage GLM into MATA      **
** matrices.                                  **
**********************************************/
mata: alpha=st_matrix("e(b)")'
mata: v1=st_matrix("e(V)")
```

2SRI Estimation -- Notes on Stata Implementation: 1st Stage GLM (cont'd)

Stata Output

CIGSPREG	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
PARITY	.0413746	.0740355	0.56	0.576	-.1037323	.1864815
WHITE	.2788441	.244504	1.14	0.254	-.200375	.7580632
MALE	.1544697	.1801299	0.86	0.391	-.1985785	.5075179
EDFATHER	-.0341149	.0184968	-1.84	0.065	-.070368	.0021381
EDMOTHER	-.0991817	.0296607	-3.34	0.001	-.1573155	-.0410479
FAMINCOM	-.0183652	.0069294	-2.65	0.008	-.0319465	-.0047839
CIGTAX88	.0190194	.0132204	1.44	0.150	-.0068922	.0449309
_cons	2.043192	.3649598	5.60	0.000	1.327884	2.7585

```
. test (EDFATHER = 0) (EDMOTHER = 0) (FAMINCOM = 0) (CIGTAX88 = 0)

 ( 1)  [CIGSPREG]EDFATHER = 0
 ( 2)  [CIGSPREG]EDMOTHER = 0
 ( 3)  [CIGSPREG]FAMINCOM = 0
 ( 4)  [CIGSPREG]CIGTAX88 = 0

           chi2(  4) =   49.33
         Prob > chi2 =    0.0000
```

2SRI Estimation -- Notes on Stata Implementation: 2nd Stage GLM

Stata Code

```
/**************************************************
** Apply GLM for the 2SRI second stage.            **
**************************************************/
glm BIRTHWTLB CIGSPREG PARITY WHITE MALE xuhat, ///
family(gaussian) link(log) vce(robust)
```

Stata Output

```
------------------------------------------------------------------------------
             |               Robust
   BIRTHWTLB |      Coef.   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
------------------------------------------------------------------------------
```

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation

-- MATA code for calculating the estimated asymptotic covariance matrix

$$\hat{D}^* = \begin{bmatrix} \hat{D}^*_{11} & \hat{D}^*_{12} \\ \hat{D}^{*\prime}_{12} & \hat{D}^*_{22} \end{bmatrix}$$

where

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

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

and

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

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

$$\hat{E}^*\left[\nabla_{\beta\beta}q\right] = \sum_{i=1}^{n}\exp(X_i\hat{\beta})^2\, X_i{}'X_i$$

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation (cont'd)

```
/************************************************
** Use the Stata "putmata" command to send
** Stata data variables into Mata vectors.
************************************************/
putmata CIGSPREG BIRTHWTLB PARITY WHITE MALE EDFATHER ///
 EDMOTHER FAMINCOM CIGTAX88 xuhat expwalpha
.
.
.
/************************************************
** MATA Start-up.                              **
************************************************/
mata:
.
.
.

/************************************************
** Load the coefficient vector and covariance  **
** matrix from second-stage GLM into MATA      **
** matrices.                                   **
************************************************/
beta=st_matrix("e(b)")'
v2=st_matrix("e(V)")
.
.
.
```

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation (cont'd)

```
/**********************************************
** Load the coefficient vector and covariance  **
** matrix from second-stage GLM into MATA      **
** matrices.                                   **
**********************************************/
beta=st_matrix("e(b)")'
v2=st_matrix("e(V)")
.
.
.
/**********************************************
** Load the W-variables for the rhs of the      **
** first stage GLM equation into a MATA matrix **
** -- don't  include the policy variable or xuhat**
** -- do include the IVs.                      **
** -- do include a constant term               **
**********************************************/
W=PARITY, WHITE, MALE, EDFATHER, EDMOTHER, FAMINCOM, ///
 CIGTAX88, J(rows(PARITY),1,1)

/**********************************************
** Load the X-variables for the rhs of the      **
** second stage GLM equation into a MATA matrix**
** -- don't  include the policy variable or the**
** IVs.                                            **
**********************************************/
X=PARITY, WHITE, MALE, xuhat
```

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation (cont'd)

```
/**************************************************
** Generate 2 matrices:
** X0 does not include the policy variable xp
** X1 does include the policy variable xp
** Appending a constant term to the end of each
** matrix.
**************************************************/
X0=X,J(rows(X),1,1)
X1=xp,X,J(rows(X),1,1)

/**************************************************
** Compute x1b1 multiplying the matrix         **
** of exogenous variables (X1) by the          **
** coefficient vectors.                            **
**************************************************/
x1b1=X1*beta
```

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation (cont'd)

```
/***************************************************
** Compute the asymptotic covariance matrix of
** the 2SRI estimates (See Appendix A).
***************************************************/
paMu=-bxu*exp(x1b1):*expwalpha:*W
pbMu=exp(x1b1):*X1
pbaq=pbMu'*paMu
```

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

```
pbbq=pbMu'*pbMu
```

$$\hat{E}*\left[\nabla_{\beta\beta}q\right] = \sum_{i=1}^{n}\exp(X_i\hat{\beta})^2\,X_i{}'X_i$$

2SRI Asymptotic Standard Errors -- Notes on Stata Implementation (cont'd)

```
D11=v1
```

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

```
D12=v1*pbaq'*invsym(pbbq)
```

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

```
D22= invsym(pbbq)*pbaq*v1*pbaq'* invsym(pbbq)+v2
```

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

```
D=D11, D12 \ D12', D22
```

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

2SRI -- Notes on Stata Implementation: Results

	1	2	3	4	5
1	variable	estimate	t-stat	wrong-t-stat	p-value
2					
3	CIGSPREG	-.0140086	-3.678995	-4.07594	.0002342
4	PARITY	.0166603	3.180623	3.410309	.0014696
5	WHITE	.0536269	4.217293	4.545233	.0000247
6	MALE	.0297938	3.130267	3.3546	.0017465
7	xuhat	.0097786	2.557676	2.830723	.0105374
8	constant	1.948207	117.6448	123.7389	0

Multi-Stage Causal Effect Estimators

-- Here the focus is on the evaluation of the anticipated or past effect of a specified policy on the value of an economic outcome of interest (Y) – the *outcome.*

-- The policy in question is typically defined in terms of a past or proposed change in a specified variable (X_p) – the *policy variable.*

-- For example, consider the analysis of potential gains in infant birth weight (Y) that may result from effective prenatal smoking prevention and cessation policy.

-- Here, X_p represents smoking during pregnancy and the policy of interest, if fully effective, would maintain zero levels of smoking for non-smokers (prevention) and convince smokers to quit before becoming pregnant (cessation).

Multi-Stage Causal Effect Estimators: The Potential Outcomes Framework

-- 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 (2014) 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. See also Angrist and Pischke (2009), pp. 13-15 and 52-59.

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

Terza, J.V. (2014): "Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight," Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

Multi-Stage Causal Effect Estimators: Review of the POF

-- Defining $Y_{X_p^*}$ to be the random variable representing the distribution of *potential outcomes* as they would manifest if the policy variable were exogenously mandated (ceteris paribus) to be X_p^* – as in a fully effective policy intervention like the smoking and birthweight example described above.

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

-- Here we draw the distinction between X_p, the observable or factual version of the policy variable, and X_p^*, its unobservable (hypothetically mandated) or *counterfactual* version.

-- Likewise we use Y to denote the observable version of the outcome, while $Y_{X_p^*}$ is the policy-relevant counterfactual.

-- Note that the symbols X_p and Y are doing notational double duty in that they are used as generic conceptual representations of the policy variable and outcome, respectively, and are also used denote their observable versions.

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

--Clearly, the only measures of the effects of changes in the policy variable on the outcome that are policy relevant are those that are causally interpretable (CI).

-- We take as axiomatic that an effect measure is CI only if it is defined in terms of changes in the relevant potential outcome -- e.g., a change from $Y_{X_p^{pre}}$ to $Y_{X_p^{post}}$ that would be caused by a policy-induced exogenous change in X_p from pre-policy to post policy (say from X_p^{pre} to X_p^{post}).

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

-- This ensures that such a measure represents outcome effects that can be exclusively attributed to exogenously mandated (ceteris paribus) changes in the policy variable.

-- Without loss of generality we write $X_p^{post} = X_p^{pre} + \Delta$ (Δ being the mandated policy change).

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

-- In generic terms, the estimation objective here is the difference between the distributions of $Y_{X_p^{pre}}$ and $Y_{X_p^{pre}+\Delta}$ [or some particular aspect (parameter) thereof], where X_p^{pre} and $X_p^{pre}+\Delta$ represent well-defined and mandated pre- and post-intervention versions of the policy variable, respectively.

-- For example, in a number of empirical policy analytic contexts, the following *average incremental effect (AIE)* is of interest

$$AIE(\Delta) = E[Y_{X_p^{pre}+\Delta}] - E[Y_{X_p^{pre}}]. \qquad (5)$$

-- Terza (2014) shows how the AIE and other counterfactual causal measures can be estimated using nonlinear regression methods and observable (factual) data.

Review of the POF: Back to the Example

-- In our birth weight/smoking example

-- X_p^{pre} would denote the pre-policy prenatal smoking distribution

-- $\Delta = -X_p^{pre}$ is the policy-induced change in prenatal smoking

-- $Y_{X_p^{pre}}$ is the random variable representing the pre-policy distribution of birth weights

-- $Y_{X_p^{post}}$ is the random variable representing the post-policy potential birth weight outcomes.

Review of the POF: Back to the Example (cont'd)

-- The relevant version of the AIE in this example is

$$\mathrm{AIE}(\Delta) = \mathrm{E}[\mathrm{Y}_0] - \mathrm{E}[\mathrm{Y}_{\mathrm{X}_p^{pre}}]. \qquad (6)$$

-- As this example demonstrates, in the general potential outcomes (PO) policy analytic framework, both X_p^{pre} and Δ can be random variables.

-- Expression (5) is, in fact, quite general. For example, when the policy variable is binary, if we set $X_p^{pre} = 0$ and $\Delta = 1$ then (5) measures the *average treatment effect* (ATE)

$$\text{ATE} = E[Y_1] - E[Y_0]. \quad (7)$$

-- Note that in this case Δ, X_p^{pre} and $X_p^{pre} + \Delta$, are all degenerate random variables.

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

-- When the policy variable is continuous and no specific policy increment (Δ) has been defined (in which case it is typically assumed that Δ approaches 0), then the *average marginal effect* (AME) of an infinitesimal change in the policy variable is measured as

$$\mathrm{AME} = \lim_{\delta \to 0} \frac{\mathrm{AIE}(\delta)}{\delta} \qquad (8)$$

where $\mathrm{AIE}(\delta)$ is defined as in (5) and δ is a constant (a degenerate random variable).

Multi-Stage Causal Effect Estimators: Review of the POF (cont'd)

-- The measures defined in (5), (7) and (8) are logical targets for health policy analysis.

-- Moreover, they are CI because they are PO-based.

-- Which of them is apropos a particular policy context will depend on the support of the policy variable in question and whether or not the policy increment (Δ) is known.

Specification and Estimation AIE, ATE and AME via Regression Modeling

-- The *expected potential outcome* ($E[Y_{X_p^*}]$), can be rewritten in a way that facilitates the specification [estimation] of (5), (7) and (8) via nonlinear regression (NR) models [methods].

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

where

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

X_o is a vector of observable confounders for X_p

and

X_u is a scalar comprising all unobservable confounders for X_p.

Specification and Estimation AIE, ATE and AME via Regression Modeling (cont'd)

-- Using (9), the AIE, ATE and AME can be rewritten as:

$$\mathbf{AIE}(\Delta) = \mathbf{E}_{X_p^{pre}+\Delta, X_o, X_u}\left[\mu(X_p^{pre}+\Delta, X_o, X_u, \tau)\right] - \mathbf{E}_{X_p^{pre}, X_o, X_u}\left[\mu(X_p^{pre}, X_o, X_u, \tau)\right] \tag{10}$$

$$\mathbf{ATE} = \mathbf{E}_{X_o, X_u}\left[\mu(1, X_o, X_u, \tau)\right] - \mathbf{E}_{X_o, X_u}\left[\mu(0, X_o, X_u, \tau)\right] \tag{11}$$

and

$$= \mathbf{E}_{X_p^{pre}, X_o, X_u}\left[\left.\frac{\partial \mu(X_p^{*}, X_o, X_u; \tau)}{\partial X_p^{*}}\right|_{X_p^{*}=X_p^{pre}}\right]. \tag{12}$$

Specification and Estimation AIE, ATE and AME via Regression Modeling (cont'd)

-- Assuming that we have a consistent estimator for τ (say $\hat{\tau}$) and an appropriate observable proxy value for the unobservable X_u [say $\hat{X}_u(W,\hat{\tau})$ -- note that we have already mentioned such a proxy in the context of 2SRI estimation, viz., the first-stage residual], consistent estimators for (10), (11) and (12) are, respectively:

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ \mu(X_{pi}^{pre} + \Delta_i, X_{oi}, \hat{X}_u(W_i,\hat{\tau});\hat{\tau}) - \mu(X_{pi}^{pre}, X_{oi}, \hat{X}_u(W_i,\hat{\tau});\hat{\tau}) \right\} \tag{13}$$

$$\widehat{\mathrm{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \left\{ \mu(1, X_{oi}, \hat{X}_u(W_i,\hat{\tau});\hat{\tau}) - \mu(0, X_{oi}, \hat{X}_u(W_i,\hat{\tau});\hat{\tau}) \right\} \tag{14}$$

$$\widehat{\mathrm{AME}} = \sum_{i=1}^{n} \frac{1}{n} \left. \frac{\partial \mu(X_p^*, X_{oi}, \hat{X}_u(W_i,\hat{\tau});\hat{\tau})}{\partial X_p^*} \right|_{X_p^* = X_{pi}^{pre}} . \tag{15}$$

Asymptotic Properties of $\widehat{\text{AIE}(\Delta)}$, $\widehat{\text{AIE}(\Delta)}$ and $\widehat{\text{AME}}$

-- We use the notation "PE" to denote the relevant policy effect [AIE, ATE or AME] and rewrite AIE, ATE and AME in generic form as

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i}{n} \tag{16}$$

where $\widehat{pe}_i$ is shorthand notation for $pe(X_{pi}^{pre}, \Delta_i, X_{oi}, \hat{X}_u(W_i, \hat{\tau}), \hat{\tau})$, $\hat{\tau}$ is the consistent estimator of τ and

$$pe(X_p^{pre}, \Delta, X_o, X_u(W,\delta), \tau) = \mu(X_p^{pre} + \Delta, X_o, X_u(W,\tau), \tau)$$

$$- \mu(X_p^{pre}, X_o, X_u(W,\tau), \tau) \quad \text{for (13)}$$

Asymptotic Properties of $\widehat{\mathrm{AIE}(\Delta)}$, $\widehat{\mathrm{AIE}(\Delta)}$ and $\widehat{\mathrm{AME}}$ (cont'd)

$$\mathbf{pe}(X_p^{pre}, \Delta, X_o, X_u(W,\tau), \tau) = \mu(1, X_o, X_u(W,\tau); \tau) - \mu(0, X_o, X_u(W,\tau); \tau) \qquad \text{for (14)}$$

$$\mathbf{pe}(X_p^{pre}, \Delta, X_o, X_u(W,\tau), \tau) = \left. \frac{\partial \mu(X_p^*, X_o, X_u(W,\tau); \tau)}{\partial X_p^*} \right|_{X_p^* = X_p^{pre}} . \qquad \text{for (15)}$$

Asymptotic Properties of $\widehat{\text{AIE}(\Delta)}$, $\widehat{\text{AIE}(\Delta)}$ and $\widehat{\text{AME}}$ (cont'd)

-- We can cast $[\hat{\tau} \quad \widehat{\text{PE}}]$ as a 2SOE:

-- First stage... consistent estimation of τ (e.g. via 2SRI).

-- Second stage... $\widehat{\text{PE}}$ itself is easily shown to be the optimizer of the following objective function

$$\sum_{i=1}^{n} q(\hat{\tau}, \text{PE}, Z_i)$$

where

$$q(\hat{\tau}, \text{PE}, Z_i) = -(\widehat{\text{pe}}_i - \text{PE})^2$$

$Z_i = [Y_i \quad X_{pi}^{pre} \quad W_i]$ and $\hat{\tau}$ is the first-stage estimator of τ.

Asymptotic Properties of $\widehat{AIE(\Delta)}$, $\widehat{AIE(\Delta)}$ and $\widehat{AME}$ (cont'd)

-- Because we can cast [$\hat{\tau}$ $\widehat{PE}$] as a 2SOE, we know that under general conditions

$$D^{-\frac{1}{2}}\sqrt{n}\left(\begin{bmatrix}\hat{\tau}\\ \widehat{PE}\end{bmatrix}-\begin{bmatrix}\tau\\ PE\end{bmatrix}\right)\xrightarrow{d} N(0,I).$$

The practical version of the consistent estimator of the partition of D that pertains to $\widehat{PE}$ is the following analog to (4):

$$\hat{D}^{*}_{22} = \hat{E}*\left[\nabla_{PE\,PE}q\right]^{-1}\hat{E}*\left[\nabla_{PE\,\tau}q\right]\widehat{AVAR}*(\hat{\tau})\,\hat{E}*\left[\nabla_{PE\,\tau}q\right]'\,\hat{E}*\left[\nabla_{PE\,PE}q\right]^{-1} + \widehat{AVAR}*(\widehat{PE}) \qquad (17)$$

Asymptotic Properties of $\widehat{AIE(\Delta)}$, $\widehat{AIE(\Delta)}$ and $\widehat{AME}$ (cont'd)

where

$$\hat{E}^*\left[\nabla_{PE\,PE}q\right] = n$$

$$\hat{E}^*\left[\nabla_{PE\tau}q\right] = \sum_{i=1}^{n}\nabla_{\tau}\widehat{pe_i}$$

$\nabla_{\tau}\widehat{pe_i}$ is shorthand notation for $\nabla_{\tau}pe(X_p^{pre}, \Delta, X_o, X_u(W,\tau),\tau)$ evaluated at $[X_{pi}^{pre} \quad \Delta_i \quad X_{oi} \quad \hat{X}_u(W_i,\hat{\tau}) \quad \hat{\tau}]$

$$\widehat{AVAR}^*(\widehat{PE}) = \sum_{i=1}^{n}(\widehat{pe_i} - \widehat{PE})^2$$

and $\widehat{AVAR}^*(\hat{\tau})$ is the correct estimator of the asymptotic covariance matrix of $\hat{\tau}$.

In summary, we rewrite (17) as

$$\hat{D}_{22}^{*} = \frac{1}{n^2}\left\{\left(\sum_{i=1}^{n} \nabla_{\tau}\widehat{pe_i}\right)\widehat{AVAR}*(\hat{\tau})\left(\sum_{i=1}^{n} \nabla_{\tau}\widehat{pe_i}\right)' + \sum_{i=1}^{n}(\widehat{pe_i} - \widehat{PE})^2\right\} \qquad (18)$$

-- So, for example, the "t-statistic" $(\widehat{PE} - PE)/\sqrt{\hat{D}_{22}^{*}}$ is asymptotically standard normally distributed and can be used to test the hypothesis that $PE = PE^0$ for PE^0, a given null value of PE.

Example: AIE of Smoking During Pregnancy on Birthweight

-- The objective is to evaluate a policy that would bring smoking during pregnancy to zero.

-- Pre-policy version of the policy variable: $X_p^{pre} = X_p$

-- Post-policy version of the policy variable: $X_p^{post} = X_p + \Delta$ where $\Delta = -X_p$

-- AIE estimator is the version of (16) in which

$$\begin{aligned} &pe(X_p^{pre}, \Delta, X_o, X_u(W,\tau), \tau) \\ &\qquad = \exp([X_p + \Delta]\beta_p + X_o\beta_o + X_u\beta_u) - \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u) \end{aligned}$$

where

$$X_u = X_p - \exp(W\alpha)$$

and $\tau = [\alpha' \quad \beta']'$.

AIE of Smoking on Birthweight – Asymptotic Standard Error

-- The estimator of the correct asymptotic variance estimator of $\widehat{PE}$ is the version of (18) with

$$\nabla_{\tau} pe = [\nabla_{\alpha} pe \quad \nabla_{\beta_p} pe \quad \nabla_{\beta_o} pe \quad \nabla_{\beta_o} pe]$$

and $\nabla_a pe$ as shorthand notation for $\nabla_a pe(X_p^{pre}, \Delta, X_o, X_u(W, \tau), \tau)$

$$[a = \alpha, \ \beta_p, \beta_o \text{ or } \beta_u].$$

-- Similarly, we use $\nabla_a \widehat{pe}_i$ as shorthand notation for $\nabla_a pe$ evaluated at

$$[X_{pi}^{pre} \quad \Delta_i \quad X_{oi} \quad \hat{X}_u(W_i, \hat{\tau}) \quad \hat{\tau}].$$

AIE of Smoking on Birthweight – Asymptotic Standard Error

-- In this example we have

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

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

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

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

where $\hat{X}_{ui} = X_{pi} - \exp(W_i\hat{\alpha})$

AIE of Smoking on Birthweight: MATA Code for AIE Estimate

```
/***********************************************
** Compute the estimated average incremental   **
** effect (the policy effect) for each             **
** individual in the sample.                   **
***********************************************/
pei=exp(x1incb1):-exp(x1b1)
```

$$pe(X_p^{pre}, \Delta, X_o, X_u(W,\tau), \tau)$$
$$= \exp([X_p + \Delta]\beta_p + X_o\beta_o + X_u\beta_u) - \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u)$$

```
/***********************************************
** Compute the AIE.                    **
***********************************************/

pe=mean(pei)
```

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i}{n}$$

AIE of Smoking on Birthweight: MATA Code for Requisite Gradient Components

```
/*************************************************
** Construct the gradient component of the asy **
** variance of the AIE that pertains to alpha. **
*************************************************/
palfa=-expwalpha:*bxu:*pei:*W
```

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

```
/*************************************************
** Construct the gradient component of the asy **
** variance of the AIE that pertains to betap. **
*************************************************/
pbetap=exp(x1incb1):*xpinc:-exp(x1b1):*xp
```

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

AIE of Smoking on Birthweight: MATA Code for Requisite Gradient Components

(cont'd)

```
/*************************************************
** Construct the gradient component of the asy  **
** variance of the AIE that pertains to betao   **
** and betau.                                   **
** NOTE THAT X0 INCLUDES XUHAT.                 **
*************************************************/
pbetao=pei:*X0
```

$$[\nabla_{\beta_o}\widehat{pe}_i \quad \nabla_{\beta_u}\widehat{pe}_i] = \left[\exp([X_{pi} + \Delta_i]\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o + \hat{X}_{ui}\hat{\beta}_u)\right] \times [X_{oi} \quad \hat{X}_{ui}]$$

AIE of Smoking on Birthweight: MATA Code for the Estimated Asymptotic Covariance Matrix

```
/************************************************
** Sum and concatenate to construct the full    **
** gradient component of the asy variance of the**
** AIE.                                         **
************************************************/
ppe=colsum(palfa),colsum(pbetap),colsum(pbetao)


/************************************************
** Compute the estimated asymptotic variance of **
** the AIE.                                       **
************************************************/
varpe=(1:/n^2):*(ppe*D*ppe':+sum((pei:-pe):^2))
```

$$\hat{D}^*_{22} = \frac{1}{n^2}\left\{\left(\sum_{i=1}^{n} \nabla_\tau \widehat{pe_i}\right) \widehat{AVAR} * (\hat{\tau}) \left(\sum_{i=1}^{n} \nabla_\tau \widehat{pe_i}\right)' + \sum_{i=1}^{n} (\widehat{pe_i} - \widehat{PE})^2\right\}$$

4 LINES OF CODE TO CALCULATE THE CORRECT ASY VARIANCE ESTIMATE FOR THE AIE ESTIMATOR

Results for Smoking and Birthweight Model

GLM Exponential Condition Mean NLS Regression

	1	2	3	4	5
1	variable	estimate	t-stat	wrong-t-stat	p-value
2					
3	CIGSPREG	-.0140086	-3.678995	-4.07594	.0002342
4	PARITY	.0166603	3.180623	3.410309	.0014696
5	WHITE	.0536269	4.217293	4.545233	.0000247
6	MALE	.0297938	3.130267	3.3546	.0017465
7	xuhat	.0097786	2.557676	2.830723	.0105374
8	constant	1.948207	117.6448	123.7389	0

AIE of Eliminating Smoking During Pregnancy

	1	2	3	4	5	6
1		%smoke-decr	incr-effect	std-err	t-stat	p-value
2						
3	2SRI-correct	100	.2300237	.0726222	3.167401	.0015381
4	c-delta-meth	100	.2300237	.0703486	3.269771	.0010763
5	2SRI-wrong	100	.2300237	.0661442	3.47761	.0005059
6	w-delta-meth	100	.2300237	.0636395	3.614479	.000301