SIMPLIFIED STANDARD ERRORS FOR MULTI-STAGE REGRESSION-BASED ESTIMATORS

Joseph V. Terza
Department of Economics
University of North Carolina at Greensboro
Greensboro, NC 27402-6165
Phone: (336) 334-4892
E-mail: jvterza@uncg.edu

(July, 2012)

Please cite:

Terza, J.V. (2012a): "Correct Standard Errors for Multi-Stage Regression-Based Estimators: A Practitioner's Guide with Illustrations," Unpublished manuscript, Department of Economics, University of North Carolina at Greensboro.

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 resampling 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)

-- In the paper -- we detail our simplified covariance specifications (standard errors) for three very useful estimators in applied 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.

Motivation (cont'd)

and

3) Causal incremental and marginal effects estimators proposed by Terza (2012b).

Terza, J.V. (2012b): "Health Policy Analysis via Nonlinear Regression Methods: Estimation and Inference from a Potential Outcomes Perspective, Unpublished manuscript, Department of Economics, University of North Carolina at Greensboro.

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

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

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

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

where V_i denotes the relevant subvector of the observable data for the ith sample individual (i = 1, ..., n); e.g., if the first-stage implements the nonlinear least squares (NLS) method

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

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

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

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

where Z_i is the full vector of observable data, and $\hat{\delta}$ denotes the first-stage estimate of δ; e.g., if the second-stage is MLE

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

with $f(Y_i | X_{pi}, W_i; \hat{\delta}, \gamma)$ being the relevant conditional density of the dependent variable Y_i.

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.

-- Practitioners often opt for resampling methods like bootstrapping, or in the case of "effect" estimation, the approach suggested by Krinsky, I. and Robb L. (1986, 1990, 1991).

-- A possible reason for this is that the expressions for the correct asymptotic covariance matrix of the generic 2SOE found in textbooks are daunting.

-- In the following, we offer a substantial and legitimate simplification that may make implementation of the correct asymptotic standard error formulations more accessible to practitioners.

2SOE and Their Asymptotic Standard Errors: Some Notation

-- The correct asymptotic covariance matrix of $\hat{\omega}$ is

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

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.

-- The devil, of course, is in the "D"-tails.

2SOE and Their Asymptotic Standard Errors: More Notation

-- 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. 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

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

-- $\nabla_{st}q$ 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 and Their Asymptotic Standard Errors (cont'd)

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

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

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

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

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

-- We seek simple estimators of D_{12} and D_{22}.

Simple Standard Error Formulae – MLE

-- In the paper we show that when the second stage estimator is MLE the correct formulations 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

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

$\hat{\delta}$ and $\tilde{\gamma}$ denote the first and second stage estimators, respectively, 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 the correct formulations simplify as

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

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

where

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

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

Simple Standard Error Formulae (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.

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 (pseudo) 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 -- hence, the "pseudo" modifier.

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 second stage.

-- Therefore the relevant version of $q(\hat{\delta}, \hat{\gamma}, Z)$ is

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

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 (2012b) 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.

Multi-Stage Causal Effect Estimators (cont'd)

As counterparts to the *average treatment effect* in the POF, Terza (2012b) defines the *average incremental effect* and the *average marginal effect*, respectively, as

$$\mathrm{AIE}\big(\Delta(X_{p1})\big) = E[Y_{X_{p1}+\Delta(X_{p1})}] - E[Y_{X_{p1}}] \quad \text{and} \quad \mathrm{AME} = \lim_{\Delta \to 0} \frac{\mathrm{AIE}(\Delta)}{\Delta}$$

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

$Y_{X_p^*}$ 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^*.

-- Terza (2012b) shows that under primitive regression assumptions (e.g. the outcome and auxiliary models in 2SRI), if we can consistently estimate the parameters of the model (e.g. $\tau = [\alpha' \quad \beta']$ in the above 2SRI setup) and can find an appropriate way to proxy X_u, AIE and AME can be consistently estimated using

$$\widehat{AIE}(\Delta(X_{p1i})) = \sum_{i=1}^{n} \frac{1}{n} \left\{ \mu(X_{p1i} + \Delta_i(X_{p1i}), X_{oi}, \hat{X}_{ui}; \hat{\tau}) - \mu(X_{p1i}, X_{oi}, \hat{X}_{ui}; \hat{\tau}) \right\}$$

$$\widehat{AME} = \sum_{i=1}^{n} \frac{1}{n} \frac{\partial \mu(X_{p1i}, X_{oi}, \hat{X}_{ui}; \hat{\tau})}{\partial X_{p1i}}$$

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 ith observation in a sample of size n (i = 1, …, n).

Multi-Stage Causal Effect Estimators (cont'd)

-- We now turn to the asymptotic properties of these estimators.

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

$$\widehat{PE} = \sum_{i=1}^{n} \frac{\widehat{pe}_i(\hat{\alpha}, \hat{\beta})}{n} \qquad \widehat{pe}_i(\hat{\alpha}, \hat{\beta}) \text{ being shorthand for } pe(X_{p1i}, X_{oi}, \hat{X}_{ui}(\hat{\alpha}, W_i), \hat{\beta})$$

where

$$pe(X_{p1}, X_o, X_u(\alpha, W), \beta) =$$

$$\mu(X_{p1} + \Delta(X_{p1}), X_o, X_u(\alpha, W), \beta) - \mu(X_{p1}, X_o, X_u(\alpha, W), \beta) \quad \text{for AIE}$$

or

$$\frac{\partial \mu(X_{p1}, X_o, X_u(\alpha, W), \beta)}{\partial X_{p1}} \qquad \text{for AME}$$

and $\hat{X}_{ui}(\hat{\alpha}, W_i) = X_{pi} - r(W_i, \hat{\alpha})$.

$\widehat{PE}$ as a 2SOE

-- We can cast $\widehat{PE}$ as a 2SOE:

-- First stage... consistent estimation of α and β (e.g. via 2SRI).

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

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

where

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

$Z_i = [Y_i \quad X_{p1i} \quad W_i]$ and $[\hat{\alpha}' \quad \hat{\beta}']$ is the first-stage estimator of $[\alpha' \quad \beta']$.

$\widehat{PE}$ as a 2SOE – Asymptotic Standard Error

-- Because $\widehat{PE}$ is virtually NLS, using the above results, its correct standard error is

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

where

$\sum_{i=1}^{n} \nabla_{[\alpha' \ \ \beta']} \widehat{pe}_i(\hat{\alpha}, \hat{\beta})$ **denotes** $\nabla_{[\alpha' \ \ \beta']} pe(X_{p1}, X_o, X_u(\alpha, W), \beta)$ **evaluated at** X_{pi}, X_{oi}, W_i,

and $[\hat{\alpha}' \ \ \hat{\beta}']$ **and**

$\widehat{AVAR}([\hat{\alpha}' \ \ \hat{\beta}'])$ **is the estimated asymptotic (2SRI?) covariance matrix of** $[\hat{\alpha}' \ \ \hat{\beta}']$**.**

$\widehat{PE}$ as a 2SOE – Asymptotic t-stat

-- So, for example, the "t-statistic" $\sqrt{n}(\widehat{PE} - PE) / \sqrt{\widehat{a\,var}\left(\widehat{PE}\right)}$ 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.

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

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: $X_{p1} = X_p$

-- Post-policy version of the policy variable: $X_{p2} = X_p + \Delta(X_p)$ where $\Delta(X_p) = -X_p$

-- AIE estimator is

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

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

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

AIE of Smoking on Birthweight – Asymptotic Standard Error

$$\widehat{\text{a var}}(\widehat{\text{PE}}) = \left(\frac{\sum_{i=1}^{n} \nabla_{\beta} \widehat{\text{pe}}_i(\hat{\beta})}{n} \right) \widehat{\text{COV}}_{\text{GMM}} \left(\frac{\sum_{i=1}^{n} \nabla_{\beta} \widehat{\text{pe}}_i(\hat{\beta})}{n} \right)' + \frac{\sum_{i=1}^{n} \left(\widehat{\text{pe}}_i(\hat{\beta}) - \widehat{\text{PE}} \right)^2}{n}$$

$$\nabla_{\beta} \widehat{\text{pe}}_i(\hat{\beta}) = [\nabla_{\beta_p} \widehat{\text{pe}}_i(\hat{\beta}) \quad \nabla_{\beta_o} \widehat{\text{pe}}_i(\hat{\beta})]$$

$$\nabla_{\beta_p} \widehat{\text{pe}}_i(\hat{\beta}) = \exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o)[X_{pi} + \Delta(X_{pi})] - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o)X_{pi}$$

$$\nabla_{\beta_o} \widehat{\text{pe}}_i(\hat{\beta}) = \left[\exp([X_{pi} + \Delta(X_{pi})]\hat{\beta}_p + X_{oi}\hat{\beta}_o) - \exp(X_{pi}\hat{\beta}_p + X_{oi}\hat{\beta}_o) \right] X_{oi}$$

and $\widehat{\text{COV}}_{\text{GMM}}$ is the GMM estimated asymptotic covariance matrix of $\hat{\beta}^{+}$.

AIE Asymptotic Standard Error -- Practical Notes on Stata Implementation

-- MATA code for calculating the AIE estimator

$\hat{\boldsymbol{\beta}}_{\mathbf{p}}$ `= BXpGMMC = BGMMC[1]`

$\left[\mathcal{X}_{\mathbf{p}} + \Delta(\mathcal{X}_{\mathbf{p}}) \vdots\ \mathcal{X}_{\mathbf{o}}\right]$ `= XNULL= J(rows(Xo),1,0),Xo`

$\mathcal{X}_{\mathbf{p}}\hat{\boldsymbol{\beta}}_{\mathbf{p}} + \mathcal{X}_{\mathbf{o}}\hat{\boldsymbol{\beta}}_{\mathbf{o}}$ `= XBGMMC = XC*BGMMC`

$[\mathcal{X}_{\mathbf{p}} + \Delta(\mathcal{X}_{\mathbf{p}})]\hat{\boldsymbol{\beta}}_{\mathbf{p}} + \mathcal{X}_{\mathbf{o}}\hat{\boldsymbol{\beta}}_{\mathbf{o}}$ `= XBNULLGMMC=XNULL*BGMMC`

$\widehat{\mathbf{pe}}(\hat{\boldsymbol{\beta}}) =$ `peAIEGMM = exp(XBNULLGMMC):-exp(XBGMMC)`

$\widehat{\mathbf{AIE}}$ `= PEAIEGMM = mean(peAIEGMM)`

Practical Notes on Stata Implementation (cont'd)

-- MATA code for calculating the asymptotic standard error estimate

$\nabla_{\beta_p}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta})=$ `ppebpAIEGMM=exp(XBNULLGMMC):*XpC`

$\nabla_{\beta_o}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta})=$`peboAIEGMM=(exp(XBNULLGMMC):-exp(XBGMMC)):*Xo`

$\nabla_{\beta}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta})$ `= ppebAIEGMM=ppebpAIEGMM,ppeboAIEGMM`

$\widehat{a\,var}(\widehat{AIE})=$

```
avarPEAIEGMM=mean(ppebAIEGMM*n:*COVGMMC)*mean(ppebAIEGMM)'
        +mean((peAIEGMM:-PEAIEGMM):^2)
```

where $\mathcal{X}_o$ and $\nabla_{\beta_o}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta})$ are n × K matrices; $\mathcal{X}_p$, $\Delta(\mathcal{X}_p)$, $\widehat{pe}(\hat{\beta})$ and $\nabla_{\beta_p}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta}^{+})$ are n × 1 vectors; $\nabla_{\beta}\widehat{pe}(\mathcal{X}_p,\mathcal{X}_o,\hat{\beta})$ is an n × (K+1) matrix; $\widehat{AIE}$ and $\widehat{a\,var}(\widehat{AIE})$ are scalars; and K is the dimension of X_o.

Results for Smoking and Birthweight Model

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

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