Two-Stage Residual Inclusion Estimation: A Practitioners Guide to Stata Implementation

by

Joseph V. Terza

Department of Economics
Indiana University Purdue University Indianapolis
Indianapolis, IN 46202

(July, 2016)

Motivation: Smoking and Infant Birth Weight

-- As an example, we revisit the regression model of Mullahy (1997) in which

Y = infant birth weight in lbs.

X_p = number of cigarettes smoked per day during pregnancy.

-- We seek to regress Y on X_p with a view toward the estimation of (and drawing inferences regarding) the causal effect of the latter on the former.

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.

Motivation: Smoking and Infant Birthweight

-- Two complicating factors:

-- the regression specification is nonlinear because Y is non-negative.

-- X_p is likely to be *endogenous* – correlated with unobservable variates that are also correlated with Y.

-- For example, unobserved unhealthy behaviors may be correlated with both smoking and infant birth weight.

-- If the endogeneity of X_p is not explicitly accounted for in estimation, effects on Y due to the unobservables will be attributed to X_p and the regression results will not be causally interpretable (CI).

Remedy: Two-Stage Residual Inclusion

-- In the generic version of the above model

$Y \equiv$ dependent variable

and the covariates include:

$X_p \equiv$ endogenous regressor (usually a policy-relevant variable)

$X_o \equiv$ vector of observable exogenous (non-endogenous) regressors

and

$X_u \equiv$ unobservable variable that is correlated with X_p but not correlated with X_o.

-- The presence of X_u in the model embodies the endogeneity of X_p.

Two-Stage Residual Inclusion (cont'd)

-- Following Terza et al. (2008), we posit the following model

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

$$= \mu(X; \beta) + e \qquad \text{[outcome regression]} \qquad (1)$$

and

$$X_p = r(W; \alpha) + X_u \qquad \text{[auxiliary regression]} \qquad (2)$$

where β and α are the parameter vectors to be estimated

$$X = [X_p \quad X_o \quad X_u]$$

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

W^+ is a vector of identifying instrumental variables (IV)

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

Two-Stage Residual Inclusion (cont'd)

and e is the random error term, tautologically defined as

$$e = Y - \mu(X; \beta)$$

so that $E[e \mid X] = 0$.

Two-Stage Residual Inclusion (cont'd)

-- The auxiliary regression specification in (2) implies that X_u can be written as the following function of W and α

$$X_u(W; \alpha) = X_p - r(W; \alpha). \qquad (3)$$

-- Given (3), an alternative and equivalent, representation of (1) is

$$Y = \mu(X_p, X_o, X_u(W; \alpha); \beta) + e. \qquad (4)$$

-- The β parameters in expression (1) are not directly estimable [e.g. via the nonlinear least squares method (NLS)] because X_u is unobservable.

Two-Stage Residual Inclusion (cont'd)

-- Terza et al. (2008) show that the following two-stage protocol is consistent.

***First Stage*: Obtain a consistent estimate of α by applying NLS to (2) and compute the *residual* as the following estimated version of (3)**

$$\hat{X}_u = X_p - r(W; \hat{\alpha}) \qquad (5)$$

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

***Second Stage*: Consistently estimate β by applying NLS to**

$$Y = \mu(X_p, X_o, \hat{X}_u; \beta) + e^{2SRI} \qquad (6)$$

where e^{2SRI} denotes the regression error term that is not identical to e due to the replacement of X_u with the residual $\hat{X}_u$.

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.

Two-Stage Residual Inclusion – Alternatives to NLS

-- It is not necessary that NLS be implemented in either or both of the stages of 2SRI. Any consistent estimator will do.

-- For instance, a maximum likelihood estimator (MLE) can be used in either, or both, of the stages.

-- For MLE in the first stage, specify 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 $r(W; \alpha)$ in (2) {the relevant conditional mean, i.e. $r(W; \alpha) = E[X_p \mid W]$}.

-- In this case, the 2SRI first stage estimator would be the MLE of α.

Two-Stage Residual Inclusion – Alternatives to NLS (cont'd)

-- Similarly for MLE in the second stage, specify a known form for the conditional density of $(Y \mid X_p, W, X_u)$, say $f(Y \mid X_p, W, X_u; \alpha, \beta)$.

-- The second stage estimator would then be the MLE of β.

-- In the vast majority of applied settings, the 2SRI estimates of α and β are very easy to obtain via standard regression commands offered by Stata.

Back to the Example: Smoking and Infant Birth Weight

To the above smoking and birth weight model we add

$$X_o = [PARITY \quad WHITE \quad MALE]$$

$$W^+ = [EDFATHER \quad EDMOTHER \quad FAMINCOM \quad CIGTAX]$$

where

PARITIY = birth order

WHITE = 1 if white, 0 otherwise

MALE = 1 if male, 0 otherwise

EDFATHER = paternal schooling in years

EDMOTHER = maternal schooling in years

FAMINCOME = family income

and

CIGTAX = cigarette tax.

Smoking and Infant Birth Weight (cont'd)

-- Mullahy's (1997) regression model can be written as the following version of (1)

[see Terza (2006)]

$$Y = \exp(X_p\beta_p + X_o\beta_o + X_u\beta_u) + e$$

$$= \exp(X\beta) + e \qquad (7)$$

where and $\beta' = [\beta_p \quad \beta_o' \quad \beta_u]$.

Terza, J. (2006): "Estimation of Policy Effects Using Parametric Nonlinear Models: A Contextual Critique of the Generalized Method of Moments," *Health Services and Outcomes Research Methodology*, 6, 177-198.

Smoking and Infant Birth Weight (cont'd)

-- In the original study, the model was estimated via a GMM procedure that does not require specification of an auxiliary regression for X_p.

-- Mullahy's GMM method, though very clever, does not permit identification and estimation of β_u.

-- This precludes a direct test of endogeneity because, under the assumed regression specification in (7), X_p is exogenous is iff $\beta_u = 0$.

-- Such a test is, however, supported in the 2SRI estimation framework.

-- We specify the relevant auxiliary regression as the following version of (2)

$$X_p = \exp(W\alpha) + X_u. \qquad (8)$$

Smoking and Infant Birth Weight (cont'd)

-- In this context the 2SRI protocol is:

***First Stage*: Consistently estimate α by applying NLS to (8) and save the residuals as defined in (5). In this case**

$$\hat{X}_u = X_p - \exp(W\hat{\alpha}) \qquad (9)$$

where $\hat{\alpha}$ is the NLS estimate of α.

In Stata use

```
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER ///
   FAMINCOM CIGTAX88, ///
   family(gaussian) link(log) vce(robust)
predict Xuhat, response
```

Smoking and Infant Birth Weight (cont'd)

```
------------------------------------------------------------------------------
             |               Robust
    CIGSPREG |      Coef.   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
```

Smoking and Infant Birthweight (cont'd)

***Second Stage*: Consistently estimate β by applying NLS to this version of (6)**

$$Y = \exp(X_p\beta_p + X_o\beta_o + \hat{X}_u\beta_u) + e^{2SRI} \qquad (10)$$

In Stata use

```
glm BIRTHWTLB CIGSPREG PARITY WHITE MALE Xuhat, ///
   family(gaussian) link(log) vce(robust)
```

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

Standard Errors in a 2SRI Setting: Bootstrapping

-- The standard errors (t-z-statistics, p-values) of the estimates of the elements of $\hat{\beta}$ (the 2SRI elements of β) as displayed in the above Stata output are not correct (i.e. cannot be used to estimate asymptotic confidence intervals or to conduct asymptotic hypothesis tests).

-- Bootstrapping can be used to approximate the asymptotically correct standard errors (ACSE) for $\hat{\beta}$ (500 replications).

Stata Code for Bootstrapping

```
/*************************************************
** Begin Stata program for bootstrapping.      **
*************************************************/
program twosri, eclass
tempname b V
capture drop Xuhat

/*************************************************
** Apply GLM for the 2SRI first stage.         **
*************************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88, ///
family(gaussian) link(log) vce(robust)

/*************************************************
** Save the first stage residuals.             **
*************************************************/
predict Xuhat, response

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

/*************************************************
** End Stata program for bootstrapping.          **
*************************************************/
matrix `b' = e(b)
ereturn post `b'
end
```

Stata Code for Bootstrapping (cont'd)

```
/************************************************
** Bootstrap.                                  **
************************************************/
bootstrap _b, reps(3000) seed (10101) nodots nowarn: twosri
```

2SRI Results (n = 1,388; 500 replications)

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
CIGSPREG	-.0140086	.0038255	-3.66	0.000	-.0215063	-.0065108
PARITY	.0166603	.005216	3.19	0.001	.0064372	.0268835
WHITE	.0536269	.0133074	4.03	0.000	.0275449	.079709
MALE	.0297938	.0094097	3.17	0.002	.0113511	.0482364
Xuhat	.0097786	.0038694	2.53	0.011	.0021947	.0173625
_cons	1.948207	.0170106	114.53	0.000	1.914867	1.981547

Standard Errors in a 2SRI Setting: ACSE

-- How good are the bootstrapped standard errors (BSE)? To evaluate this we need the true ACSE.

-- Underlying the ACSE is the estimated asymptotically correct covariance matrix of $\hat{\beta}$ -- EACCM($\hat{\beta}$).

-- The ACSE are the square roots of the diagonal elements of the EACCM($\hat{\beta}$).

-- Terza (2016) shows that the exact form of the EACCM($\hat{\beta}$) depends on the method implemented in the second stage of 2SRI – NLS or MLE.

Terza, J.V. (2016): "Simpler Standard Errors for Two-Stage Optimization Estimators," the *Stata Journal*, 16, 368-385.

EACCM($\hat{\beta}$) When 2nd Stage of 2SRI is NLS

$$\left(B_{\beta}^{NLS\prime}B_{\beta}^{NLS}\right)^{-1}\left(B_{\beta}^{NLS\prime}B_{\alpha}^{NLS}\right)\left(n\,\widehat{AVAR}*(\hat{\alpha})\right)\left(B_{\beta}^{NLS\prime}B_{\alpha}^{NLS}\right)'\left(B_{\beta}^{NLS\prime}B_{\beta}^{NLS}\right)^{-1}$$
$$+\left(n\,\widehat{AVAR}*(\hat{\beta})\right) \qquad (11)$$

where $\widehat{AVAR}*(\hat{\alpha})$ and $\widehat{AVAR}*(\hat{\beta})$ are the estimated covariance matrices obtained from the first and second stage packaged regression outputs, respectively

$B_{\alpha}^{NLS} \equiv (n \times K_{\alpha})$ matrix whose typical (ith) row (i = 1, …, n) is (12)

$\nabla_{\alpha}\mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta})$

$B_{\beta}^{NLS} = (n \times K_{\beta})$ matrix whose typical (ith) row (i = 1, …, n) is (13)

$\nabla_{\beta}\mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta})$

K_{α} and K_{β} are the dimensions of α and β, respectively

$\nabla_t s(t)$ is the gradient of the scalar function, s, with respect to the vector t.

and i denotes the ith observation in a sample of size n.

EACCM($\hat{\beta}$) When 2nd Stage of 2SRI is MLE

$$\left(n\,\widehat{AVAR}*(\hat{\beta})\right)\left(B_{\beta}^{MLE\prime}B_{\alpha}^{MLE}\right)\left(n\,\widehat{AVAR}*(\hat{\alpha})\right)\left(B_{\beta}^{MLE\prime}B_{\alpha}^{MLE}\right)'\left(n\,\widehat{AVAR}*(\hat{\beta})\right)$$
$$+\left(n\,\widehat{AVAR}*(\hat{\beta})\right)$$

where

$B_{\alpha}^{MLE} \equiv (n \times K_{\alpha})$ matrix whose typical (ith) row (i = 1, …, n) is

$\nabla_{\alpha} f(Y_i \mid X_{pi}, W_i, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta})$

$B_{\beta}^{MLE} = (n \times K_{\beta})$ matrix whose typical (ith) row (i = 1, …, n) is

$\nabla_{\beta} f(Y_i \mid X_{pi}, W_i, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta})$.

Back to the Smoking and Birth Weight Example

-- In this case

$$\mu(X_p, X_o, [X_p - r(W; \alpha)]; \beta)$$

$$= \exp(X_p\beta_p + X_o\beta_o + [X_p - \exp(W\alpha)]\beta_u)$$

and because the 2nd stage of our 2SRI estimator is NLS, the following versions of (12) and (13) are relevant:

$$\nabla_\alpha \mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta}) = -\hat{\beta}_u \exp(X_i\hat{\beta})\exp(W_i\hat{\alpha})W_i \qquad (14)$$

$$\nabla_\beta \mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta}) = \exp(X_i\hat{\beta})X_i \qquad (15)$$

where $X_i = [X_{pi} \quad X_{oi} \quad \hat{X}_{ui}]$ and $\hat{\beta}' = [\hat{\beta}_p \quad \hat{\beta}_o' \quad \hat{\beta}_u]$.

Back to the Smoking and Birth Weight Example: Walking Through the Stata Code

-- After the 2SRI first stage, use the following to save the vector of first-stage coefficient estimates and its corresponding estimated covariance matrix so that they are accessible in Mata:

```
mata: alphahat=st_matrix("e(b)")'
mata: Valphahat=st_matrix("e(V)")
```

The first statement yields $\hat{\alpha}$.

The second statement yields $\widehat{\text{AVAR}}^{*}(\hat{\alpha})$.

Back to the Example: Walking Through the Stata Code (cont'd)

-- After the 2SRI second stage, use the following to save the vector of second-stage coefficient estimates and its corresponding estimated covariance matrix so that they are accessible in Mata (also single out $\hat{\beta}_u$):

```
mata: betahat=st_matrix("e(b)")'
mata: Vbetahat=st_matrix("e(V)")
mata: Bu=betahat[5]
```

The first statement yields $\hat{\beta}$.

The second statement yields $\widehat{AVAR}^*(\hat{\beta})$.

The third statement yields $\hat{\beta}_u$.

Back to the Example: Walking Through the Stata Code (cont'd)

-- Construct **X** and **W** matrices, where **X** is the matrix whose columns are **Xp**, **Xo** and a constant term (a column vector of 1s); and **W** has columns **Xo**, **Wplus** and a constant term. Make sure that the ordering of the columns of X and W (including the constant term) conforms to the ordering of the estimated coefficients in $\hat{\beta}$ and $\hat{\alpha}$.

```
putmata CIGSPREG BIRTHWTLB PARITY WHITE MALE EDFATHER ///
      EDMOTHER FAMINCOM CIGTAX88 Xuhat
mata: X=CIGSPREG, PARITY, WHITE, MALE, ///
      Xuhat, J(rows(PARITY),1,1)
mata: W=PARITY, WHITE, MALE, EDFATHER, EDMOTHER, ///
      FAMINCOM, CIGTAX88, J(rows(PARITY),1,1)
```

Back to the Example: Walking Through the Stata Code (cont'd)

-- Use $\hat{\beta}$, $\hat{\beta}_u$, $\hat{\alpha}$, `X`, and `W` to construct the two gradient matrices needed to calculate the correct standard errors for $\hat{\beta}$ -- based on (14) and (15):

```
mata: Bbeta=exp(X*betahat):*X

mata: Balpha=-Bu:*exp(X*betahat):*exp(W*alphahat):*W
```

The first yields B_{β}^{NLS} based on (15).

The second yields B_{α}^{NLS} based on (14).

Back to the Example: Walking Through the Stata Code (cont'd)

-- Calculate the EACCM of $\hat{\beta}$ based on (11).

```
mata: B1=Bbeta'* Bbeta

mata: B2=Bbeta'*Balpha

mata: EACCM=invsym(B1)*B2*Valphahat*B2'*invsym(B1)+Vbetahat
```

-- Calculate the vector of asymptotically correct standard errors for `betahat`

```
mata: ACSE=sqrt(diagonal(Dhat))
```

Back to the Example: Walking Through the Stata Code (cont'd)

-- Calculate the vector of asymptotically correct t-statistics to be used to test the conventional null hypothesis regarding the elements of β (viz., $H_o : \beta_k = 0$, where β_k denotes the kth element of β)

```
mata: tstats=betahat:/ACSE
```

Back to the Example: Results ACSE vs. BSE

2SRI Second Stage, GMM and NLS Estimates

	2SRI				GMM		OLS	
Variable	Estimate	Correct t-stat	Bootstrp t-stat (500reps)	Raw t-stat	Estimate	t-stat	Estimate	t-stat
CIGS	-0.01	-3.68	-3.66	-4.08	-0.01	-3.46	0.00	-5.62
PARITY	0.02	3.18	3.19	3.41	0.02	3.33	0.01	2.99
WHITE	0.05	4.22	4.03	4.55	0.05	4.44	0.06	4.75
MALE	0.03	3.13	3.17	3.35	0.03	2.95	0.03	2.90
X_u	0.01	2.56	2.53	2.83	--	--	--	--
Constant	1.95	117.64	114.53	123.74	1.94	121.71	1.93	133.70

n = 1,388

Back to the Example: Bootstrapping Results

n = 1,388
CPU Time for ACSE = 0.618 (secs.)

Replications	% Avg. Absolute Bias	% Max. Absolute Bias	CPU Time (secs.)
100	7.10%	10.20%	16.967
250	5.20%	8.30%	42.390
500	1.80%	4.70%	83.691
1000	1.20%	2.70%	218.160
2000	1.90%	3.50%	340.284
3000	1.11%	3.58%	489.870

ACSE vs. BSE: Caveats

-- ACSE requires special programming (e.g. in Mata) but so does BSE (Stata programming).

-- BSE is only an approximation to ACSE, not clear how good that approximation is in a particular empirical context. Some sense of convergence must be achieved but this can be time consuming.

-- Elapsed computation time for ACSE in this example (n = 1,388) was less than a second.

ACSE vs. BSE: Caveats (cont'd)

-- Elapsed computation time for BSE in this example (n = 1,388; 500 replications) was 1.5 minutes, for analytic samples in health econ and health services research of sizes in the 10s of thousands, this may be an issue.

-- Convergence issues. For unstable estimation routines due to data or modeling issues, BSE may be additionally biased.

The Example: Alternative Specification

-- A large proportion of the analysis sample are non-smokers.

-- For the auxiliary regression we used the modified two-part model of Mullahy (1998).

-- The two-parts of the auxiliary regression are (i.e. the first stage of 2SRI):

Part 1: Estimate α_1 by regressing $I(X_p>0)$ on W using probit analysis and the full sample, where I(C) is the index function = 1 if condition C holds, 0 otherwise.

Part 2: Estimate α_2 by applying NLS to

$$(X_p > 0) = \exp(W\alpha) + v$$

using the subsample of smokers (i.e., those for whom $X_p > 0$.

Mullahy, J. (1998). "Much ado about two: reconsidering retransformation and the two-part model in health econometrics." *Journal of Health Economics* 17(3): 247-281.

The Example: Alternative Specification (cont'd)

-- Second stage of 2SRI:

Consistently estimate β by applying NLS to this version of (6)

$$Y = \exp(X_p\beta_p + X_o\beta_o + \hat{X}_u\beta_u) + e^{2SRI} \qquad (10)$$

with $\hat{X}_u = X_p - \Phi(W\hat{\alpha}_1)\exp(W\hat{\alpha}_2)$ **-- the residuals from the first-stage two-part model, where Φ( ) is normal cdf.**

-- Note that in the two-part model for the auxiliary regression

$$E[X_p \mid W] = \Phi(W\alpha_1)\exp(W\alpha_2).$$

Alternative Specification: Walking Through the Stata Code

-- First part of 2SRI first stage:

```
/*********************************************
** Generate the binary smoking variable.     **
*********************************************/
gen ANYCIGS=CIGSPREG>0

/*********************************************
** 2SRI first stage first part probit estimates.**
*********************************************/
probit ANYCIGS PARITY WHITE MALE EDFATHER EDMOTHER ///
   FAMINCOM CIGTAX88
predict CIGPROB
test EDFATHER EDMOTHER FAMINCOM CIGTAX88
```

Alternative Specification: Walking Through the Stata Code (cont'd)

-- After the first part of the 2SRI first stage, use the following to save the vector of first part first-stage coefficient estimates and its corresponding estimated covariance matrix so that they are accessible in Mata:

```
mata: alpha1hat=st_matrix("e(b)")'
mata: Valpha1hat=st_matrix("e(V)")
```

The first statement yields $\hat{\alpha}_1$.

The second statement yields $\widehat{\text{AVAR}}^*(\hat{\alpha}_1)$.

Alternative Specification: Walking Through the Stata Code (cont'd)

-- Second part of 2SRI first stage:

```
/*********************************************
** 2SRI first stage second part probit NLS      **
** estimates.                                  **
*********************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER ///
   FAMINCOM CIGTAX88 if ANYCIGS==1, ///
   family(gaussian) link(log) vce(robust)

predict CIGMEAN

test EDFATHER EDMOTHER FAMINCOM CIGTAX88

/*********************************************
** Generate the first-stage residuals.        **
*********************************************/
gen Xuhat=CIGSPREG-CIGPROB*CIGMEAN
```

Alternative Specification: Walking Through the Stata Code (cont'd)

-- After the second part of the 2SRI first stage, use the following to save the vector of second part first-stage coefficient estimates and its corresponding estimated covariance matrix so that they are accessible in Mata:

```
mata: alpha2hat=st_matrix("e(b)")'
mata: Valpha2hat=st_matrix("e(V)")
```

The first statement yields $\hat{\alpha}_2$.

The second statement yields $\widehat{\text{AVAR}}^*(\hat{\alpha}_2)$.

Alternative Specification: Walking Through the Stata Code (cont'd)

-- 2SRI second stage:

```
/*************************************************
** 2SRI second stage NLS estimates.             **
*************************************************/
glm BIRTHWTLB CIGSPREG PARITY WHITE MALE Xuhat, ///
   family(gaussian) link(log) vce(robust)
```

Alternative Specification: Walking Through the Stata Code (cont'd)

-- After the 2SRI second stage, use the following to save the vector of second-stage coefficient estimates and its corresponding estimated covariance matrix so that they are accessible in Mata (also single out $\hat{\beta}_u$):

```
mata: betahat=st_matrix("e(b)")'
mata: Vbetahat=st_matrix("e(V)")
mata: Bu=betahat[5]
```

The first statement yields $\hat{\beta}$.

The second statement yields $\widehat{AVAR}^*(\hat{\beta})$.

The third statement yields $\hat{\beta}_u$.

Alternative Specification: Walking Through the Stata Code (cont'd)

-- In this case the relevant version of μ( ) in (1) is

$$\mu(X_p, X_o, [X_p - r(W; \alpha)]; \beta)$$

$$= \exp(X_p\beta_p + X_o\beta_o + [X_p - \Phi(W\alpha_1)\exp(W\alpha_2)]\beta_u)$$

and the requisite gradients for the EACCM and ACSE (i.e. for B_α^{NLS} and B_β^{NLS}) are

$$\nabla_\alpha \mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta})$$

$$= -[\hat{\beta}_u \exp(X_i\hat{\beta})\exp(W_i\hat{\alpha}_2)\varphi(W\hat{\alpha}_1)W_i \quad \hat{\beta}_u \exp(X_i\hat{\beta})\exp(W_i\hat{\alpha}_2)\Phi(W\hat{\alpha}_1)W_i]$$

and

$$\nabla_\beta \mu(X_{pi}, X_{oi}, [X_{pi} - r(W_i; \hat{\alpha})]; \hat{\beta}) = \exp(X_i\hat{\beta})X_i.$$

Alternative Specification: Walking Through the Stata Code (cont'd)

-- Use $\hat{\beta}$, $\hat{\beta}_u$, $\hat{\alpha}$, X, and W to construct the two gradient matrices needed to calculate the correct standard errors for $\hat{\beta}$ -- based on (14) and (15):

```
/*****************************************
** Set up the two gradient matrices for the ACSE**
*****************************************/
mata: Bbeta=exp(X*betahat):*X
mata:Balpha1=-Bu:*exp(X*betahat):*normalden(W*alpha1hat)
      :*exp(W*alpha2hat):*W
mata: Balpha2=-Bu:*exp(X*betahat):*normal(W*alpha1hat)
      :*exp(W*alpha2hat):*W
mata: Balpha=Balpha1,Balpha2
```

The first yields B_{β}^{NLS} based on (15).

The next three yield B_{α}^{NLS} based on (14).

Alternative Specification: Walking Through the Stata Code (cont'd)

-- Calculate the EACCM of $\hat{\beta}$ based on (11).

```
mata: B1=Bbeta'* Bbeta

mata: B2=Bbeta'*Balpha

mata: EACCM=invsym(B1)*B2*Valphahat*B2'*invsym(B1)+Vbetahat
```

-- Calculate the vector of asymptotically correct standard errors for `betahat`

```
mata: ACSE=sqrt(diagonal(Dhat))
```

Alternative Specification: Walking Through the Stata Code (cont'd)

-- Calculate the vector of asymptotically correct t-statistics to be used to test the conventional null hypothesis regarding the elements of β (viz., $H_o : \beta_k = 0$, where β_k denotes the kth element of β)

```
mata: tstats=betahat:/ACSE
```

Alternative Specification: Bootstrapping

```
-/*************************************************
** Begin Stata program for bootstrapping.    **
*************************************************/
program twosri, eclass
tempname b V
tempvar coeff
tempvar CIGPROB
tempvar CIGMEAN
tempvar a1
tempvar a2
capture drop Xuhat

/*************************************************
** Obtain the first stage first part probit     **
** estimates.                                   **
*************************************************/
probit ANYCIGS PARITY WHITE MALE EDFATHER EDMOTHER ///
   FAMINCOM CIGTAX88
predict `CIGPROB'
matrix `a1'=e(b)
```

Alternative Specification: Bootstrapping

```
/*************************************************
** Obtain the first stage second part NLS    **
** estimates.                                 **
*************************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER ///
   FAMINCOM CIGTAX88 if ANYCIGS==1, ///
   family(gaussian) link(log) vce(robust)
predict `CIGMEAN'
matrix `a2'=e(b)

/*************************************************
** Save the first stage residuals.            **
*************************************************/
gen Xuhat=CIGSPREG-`CIGPROB'*`CIGMEAN'
```

Alternative Specification: Bootstrapping

```
/*********************************************
** Obtain the second stage NLS estimates.    **
*********************************************/
glm BIRTHWTLB CIGSPREG PARITY WHITE MALE Xuhat, ///
   family(gaussian) link(log) vce(robust)

matrix `b' = e(b)
ereturn post `b'

/*********************************************
** End Stata program for bootstrapping.      **
*********************************************/
end

*********************************************
** Bootstrap.                               **
*********************************************/
bootstrap _b, reps(3000) seed (10101) nodots nowarn: twosri
```

Alternative Specification: Results ACSE vs. BSE

	2SRI			
Variable	Estimate	Correct t-stat	Bootstrp t-stat (500reps)	Raw t-stat
CIGS	-0.01	-4.07	-3.86	-4.41
PARITY	0.02	3.36	3.45	3.66
WHITE	0.05	4.45	4.19	4.61
MALE	0.03	2.80	2.86	2.90
X_u	0.01	2.66	2.56	2.89
Constant	1.94	124.67	124.84	129.70

Alternative Specification: Bootstrapping Results

n = 1,388
CPU Time for ACSE = 0.561 (secs.)

Replications	% Avg. Absolute Bias	% Max. Absolute Bias	CPU Time (secs.)
100	4.20%	6.70%	11.372
250	3.77%	7.61%	27.384
500	3.31%	6.16%	54.973
1000	2.00%	3.63%	109.343
2000	1.90%	3.24%	218.333
3000	1.63%	2.59%	338.598