The Stata Journal (2017)
17, Number 4, pp. 939–961

Causal effect estimation and inference using Stata

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

Abstract. Terza (2016b, *Health Services Research* 51: 1109–1113) gives the correct generic expression for the asymptotic standard errors of statistics formed as sample means of nonlinear data transformations. In this article, I assess the performance of the Stata `margins` command as a relatively simple alternative for calculating such standard errors. I note that `margins` is not available for all packaged nonlinear regression commands in Stata and cannot be implemented in conjunction with user-defined-and-coded nonlinear estimation protocols that do not make a `predict` command available. When `margins` is available, however, I establish (using a real-data example) that it produces standard errors that are asymptotically equivalent to those obtained from the formulations in Terza (2016b) and the appendix available with this article. This result favors using `margins` (with its relative coding simplicity) when available. In all other cases, use Mata to code the standard-error formulations in Terza (2016b). I discuss examples, and I give corresponding Stata do-files in appendices.

Keywords: st0506, margins, causal effect estimation, causal inference

1 Introduction

Terza (2016b) gives the correct generic expression for the asymptotic standard errors of statistics formed as sample means of nonlinear data transformations. In this article, I offer guidance to empirical researchers for implementing such statistics (and their standard errors) in Stata for causal inference—for example, the estimation of the average treatment effect (ATE), average incremental effect (AIE), or average marginal effect (AME). I discuss (and detail in the appendixes) simple Stata and Mata commands that can be used to calculate these causal estimators and their correct standard errors in any nonlinear modeling context. I also explore the potential use of `margins` as a simpler alternative for standard-error calculation. Pursuant to this, I ask and answer the following:

> Q1: For which nonlinear regression estimation protocols coded in Stata is `margins` available (or unavailable)?

and

> Q2: For the cases in which `margins` is available, are there relevant versions of `margins` (RM) that will return the correct standard errors as obtained using the Terza (2016b) formulations (TF) coded in Mata?

 st0506

The answer to Q1 can be viewed as defining necessary conditions for using `margins`. Meanwhile, the existence and correctness of RM, as defined by the answer to Q2 if affirmative, supplies a sufficient condition under which `margins` would be appropriate. In the answer to Q2, we find that in any particular empirical study—conditional on availability as in Q1— although the values of the standard errors returned by RM are not identical to those produced by TF, the two approaches are asymptotically equivalent (that is, virtually the same in large samples). Therefore, when `margins` is available, one should use the relatively easy-to-code RM. In all other cases, one can use TF.[1]

The remainder of the article is organized as follows: In the next section, I discuss availability of `margins` for Stata-based causal inference in nonlinear empirical settings (therefore answering Q1). In section 3, I review the TF—the generic formulation of the asymptotically correct standard errors in this context suggested by Terza (2016b). I apply the TF to data from Fishback and Terza (1989) and estimate the following causal effects on employment (the binary outcome variable of interest) and their standard errors: the ATE of disability (a binary variable), the AME of income (a continuous variable), and the AIE of an additional year of work experience (a continuous variable). I then detail the RM that should be used in this empirical context, for which `margins` is available. By answering Q2, I re-estimate the ATE for disability, the AME for income, and the AIE for experience using the RM and the Fishback and Terza (1989) data. The TF and RM standard-error results differ but are shown to be asymptotically equivalent (theoretically the same in large samples). In section 4, I consider estimation of the AIE of smoking during pregnancy on infant birthweight using a model akin to that of Mullahy (1997). Because of the endogeneity of the smoking variable, unconventional regression methods are required, for which `margins` is either unavailable or incorrect. In this example, one must revert to the TF to calculate standard errors. In section 5, I summarize and conclude the article.

2 The availability of the Stata margins command

This section focuses on the necessary condition (as reflected in Q1) for the appropriate use of `margins` in the present context—availability. `margins` may be used after most of Stata's packaged estimation commands to estimate an ATE, AME, or AIE. However, there are important cases in which `margins` is not available. For example, attempting to invoke `margins` with the `biprobit` command (bivariate probit) will produce the following error message:

```
"default prediction is a function of possibly stochastic quantities other than
> e(b)"
```

Moreover, `margins` cannot be used in cases where the asymptotic covariance matrix of the underlying parameter estimation protocol requires community-contributed programming code (for example, covariance matrices calculated via community-contributed

1. Our focus on causal estimation and inference does not substantially limit the generality of the results presented below, because we can think of no interesting empirical applications of TF or RM in empirical econometrics that would not aim to produce causally interpretable results.

Mata code), such as two-stage estimation protocols (Terza 2016a). A prominent example is the two-stage residual inclusion (2SRI) estimator (see Terza, Basu, and Rathouz [2008] and Terza [Forthcoming]); although 2SRI parameter estimates can be obtained using packaged Stata commands that would otherwise permit `margins`, the estimated covariance matrices output by those commands would be incorrect and therefore should not be used as inputs to the standard-error calculations in `margins`.[2] As Terza (Forthcoming) shows, estimation of the correct asymptotic covariance matrix of the 2SRI estimator is one such case that requires special programming. There are, of course, many similar cases.

3 Calculation of the asymptotically correct standard errors of causal effects estimators

Statistics aimed at estimating causal effects of interest often use the following general form,

$$\widehat{\gamma} = \sum_{i=1}^{N} \frac{g(\widehat{\boldsymbol{\theta}}, \mathbf{X}_i)}{N} \tag{1}$$

where $\gamma = E\{g(\boldsymbol{\theta}, \mathbf{X})\}$ is the parameter of ultimate interest to be estimated by (1), $g(\cdot)$ is a known (possibly nonlinear) transformation, $\widehat{\boldsymbol{\theta}}$ is a preestimate of $\boldsymbol{\theta}$—a vector of "deeper" model parameters—and $\mathbf{X}_i$ denotes a vector of observed data on $\mathbf{X}$ for the ith member of a sample of size n $(i = 1, \ldots, N)$. The three most common versions of (1)—ATE, AME, and AIE—correspond to the following formulations of $g(\cdot)$, respectively,

$$g(\boldsymbol{\theta}, \mathbf{X}) = m(\boldsymbol{\theta}, \mathbf{1}, \mathbf{X}_o) - m(\boldsymbol{\theta}, \mathbf{0}, \mathbf{X}_o) \tag{2}$$

$$g(\boldsymbol{\theta}, \mathbf{X}) = \frac{\partial m(\boldsymbol{\theta}, \mathbf{X}_p, \mathbf{X}_o)}{\partial \mathbf{X}_p} \tag{3}$$

$$g(\boldsymbol{\theta}, \mathbf{X}) = m(\boldsymbol{\theta}, \mathbf{X}_p + \boldsymbol{\Delta}, \mathbf{X}_o) - m(\boldsymbol{\theta}, \mathbf{X}_p, \mathbf{X}_o) \tag{4}$$

where $m(\boldsymbol{\theta}, \mathbf{X}_p, \mathbf{X}_o) = E(\mathbf{Y}|\mathbf{X}_p, \mathbf{X}_o)$ is a regression function written to highlight the distinction between a policy-relevant regressor of interest, $\mathbf{X}_p$, and a vector of regression controls, $\mathbf{X}_o$; $\mathbf{X} = [\mathbf{X}_p\ \mathbf{X}_o]$; $\boldsymbol{\theta}$ is a vector of regression parameters; and $\boldsymbol{\Delta}$ is a known exogenous (usually policy-driven) increment to $\mathbf{X}_p$. After the regression parameter estimates are obtained (for example, $\widehat{\boldsymbol{\theta}}$—estimated via the nonlinear least [NLS] method), under fairly general conditions, in conjunction with (1), the formulations in (2), (3), and (4), respectively, yield consistent estimators of the ATE when $\mathbf{X}_p$ is binary; the AME when $\mathbf{X}_p$ is continuous and interest is in the effect attributable to an infinitesimal policy change; and the AIE when $\mathbf{X}_p$ is discrete or continuous and the relevant policy increment is $\boldsymbol{\Delta}$. The focus of this article is not only estimation of these causal effects but also attendant inference, usually drawn from the value of a "t statistic" for (1) as derived from standard asymptotic theory. Such a t statistic has the following general form,

2. In section 4, I give an example of such an incorrect covariance matrix in the 2SRI context.

$$\frac{\sqrt{N}(\widehat{\gamma}-\gamma^{\dagger})}{\text{se}(\widehat{\gamma})} \tag{5}$$

where $\gamma^{\dagger}$ is the relevant "null" value of γ (as in a test of the null hypothesis $H_0 : \gamma = \gamma^{\dagger}$) and $\text{se}(\widehat{\gamma})$ is the asymptotic standard error of (1) defined as $\text{se}(\widehat{\gamma}) \equiv \sqrt{\widehat{\text{a var}}(\widehat{\gamma})}$, with $\widehat{\text{a var}}(\widehat{\gamma})$ being a consistent estimator of the asymptotic variance of $\widehat{\gamma}$. Under slightly stronger conditions than those required for the consistency of (1), it can be shown that (5) is asymptotically standard normal distributed.

The key to using (5) for inference in this context is correct specification of $\text{se}(\widehat{\gamma})$ [that is, correct specification of $\widehat{\text{a var}}(\widehat{\gamma})$]. Terza (2016b) shows that

$$\widehat{\text{a var}}(\widehat{\gamma}) = A + B \tag{6}$$

where

$$A = \left\{ \frac{\sum_{i=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\theta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_i\right)}{N} \right\} \widehat{\text{AVAR}}\left(\widehat{\boldsymbol{\theta}}\right) \left\{ \frac{\sum_{i=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\theta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_i\right)}{N} \right\}' \tag{7}$$

$$B = \frac{\sum_{i=1}^{N} \left\{ g\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_i\right) - \widehat{\gamma} \right\}^2}{N} \tag{8}$$

$\boldsymbol{\nabla}_{\boldsymbol{\theta}} g(\widehat{\boldsymbol{\theta}}, \mathbf{X}_i)$ (a row vector) denotes the gradient of $g(\boldsymbol{\theta}, \mathbf{X})$ evaluated at $\mathbf{X}_i$ and $\widehat{\boldsymbol{\theta}}$, and $\widehat{\text{AVAR}}(\widehat{\boldsymbol{\theta}})$ is an estimator of the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$. Equations (1) through (8) and their supplemental discussion constitute what we referred to above (in Q2) as the TF.

As an example, and for the purpose of comparing results obtained from the TF with those produced via `margins`, we consider the following model of the likelihood of employment. Let $\mathbf{Y}$ be the indicator of an individual's employment status such that

$\mathbf{Y}$(`employed`) $\equiv$ `1` if the individual is employed, `0` if not

and

$\mathbf{X}_p \equiv$ the particular employment determinant of interest

We will, in turn, estimate the causal effect of each of the following three employment determinants using the relevant versions of the statistic given in (1).

ATE with (1) defined as in (2)

$$\mathbf{X}_{p1}(\texttt{disabil}) \equiv \texttt{1} \text{ if the individual has a disability, } \texttt{0} \text{ if not (binary variable)} \quad (9)$$

AME with (1) defined as in (3)

$$\begin{aligned}\mathbf{X}_{p2}(\texttt{othhinc}) &\equiv \text{other household income}\\ &= \text{income earned by others in the household}\\ &\quad \text{(continuous variable)}\end{aligned} \quad (10)$$

and

AIE with (1) defined as in (4) with $\Delta = 1$

$$\mathbf{X}_{p3}(\texttt{exper}) \equiv \text{years of work experience} = \text{age} - \text{grade} - 6 \text{(count variable)} \quad (11)$$

The elements of the vector of additional control variables ($\mathbf{X}_o$) are[3]

`male` ≡ `1` if male, `0` if not
`black` ≡ `1` if black, `0` if not
`grade` ≡ years of schooling completed
`inschool` ≡ `1` if enrolled in school in 1980, `0` if not
`vet` ≡ `1` if veteran of military service, `0` if not
`neverm` ≡ `1` if never married, 0 otherwise
`income4` ≡ interest, dividend, and rental income
`worlft75` ≡ `1` if worked less than full time in 1975, `0` if not
`spanish` ≡ `1` if of hispanic descent, `0` if not
`indian` ≡ `1` if Native American, `0` if not
`foreignb` ≡ `1` if foreign born, `0` if not
`nonengl` ≡ `1` if speaks English poorly or not at all, `0` if not
`smsa` ≡ `1` if resides in a Standard Metropolitan Area, `0` if not
`regso` ≡ `1` if resides in Southern Census Region, `0` if not
`regwe` ≡ `1` if resides in Western Census Region, `0` if not
`regnc` ≡ `1` if resides in North Central Census Region, `0` if not
`evermus` ≡ `1` if current state differs from state of birth, `0` if not
`npershh` ≡ number of persons in the household
`othhinc` ≡ income earned by others in the household

3. In each case, the nonfeatured causal variables will be included among the elements of $\mathbf{X}_o$ (for example, when estimating the ATE of `disabil`, `exper` and `othhinc` will be included in $\mathbf{X}_o$).

The data for this illustration were drawn from the database analyzed by Fishback and Terza (1989). The analysis sample comprises 31,507 observations. The descriptive statistics of the analysis sample are given in table 1.

Table 1. Descriptive statistics

Variable	Mean	Min	Max
`employed`	0.68	0.00	1.00
`disabil`	0.10	0.00	1.00
`exper`	30.41	9.00	59.00
`othhinc`	13231.87	0.00	74839.90
`male`	0.47	0.00	1.00
`black`	0.07	0.00	1.00
`grade`	12.01	0.00	20.00
`inschool`	0.01	0.00	1.00
`vet`	0.27	0.00	1.00
`neverm`	0.05	0.00	1.00
`income4`	543.80	0.00	62005.00
`worlft75`	0.37	0.00	1.00
`spanish`	0.04	0.00	1.00
`indian`	0.00	0.00	1.00
`foreignb`	0.07	0.00	1.00
`nonlengl`	0.02	0.00	1.00
`smsa`	0.81	0.00	1.00
`regso`	0.30	0.00	1.00
`regwe`	0.18	0.00	1.00
`regnc`	0.27	0.00	1.00
`evermus`	0.37	0.00	1.00
`npershh`	3.40	1.00	16.00

We assume a probit regression specification for the employment outcome, so the parameters of the model can be estimated using the Stata `probit` command—for which `margins` is available. Using the TF, we estimated the ATE, AME, and AIE as formalized in (2)–(4) and detailed in (9)–(11), respectively.

For the purpose of illustration, consider the formulation and Stata coding of the TF standard errors and asymptotic t statistics for the estimated ATE of disability. In this example, the estimated causal effect of interest is given by the following versions of (1) and (2),

$$\widehat{\gamma} = \sum_{i=1}^{N} \frac{g\left(\widehat{\boldsymbol{\beta}}, \mathbf{X}_i\right)}{N}$$

with

$$g(\boldsymbol{\beta}, \mathbf{X}_p, \mathbf{X}_o) = m(\boldsymbol{\beta}, \mathbf{1}, \mathbf{X}_o) - m(\boldsymbol{\beta}, \mathbf{0}, \mathbf{X}_o)$$

where $\widehat{\boldsymbol{\beta}}' = [\widehat{\boldsymbol{\beta}}_p \; \widehat{\boldsymbol{\beta}}_o']$ is the probit estimate of $\boldsymbol{\beta}' = [\boldsymbol{\beta}_p \; \boldsymbol{\beta}_o']$,

$$m(\boldsymbol{\theta}, \mathbf{X}_p, \mathbf{X}_o) = \Phi(\mathbf{X}_p \boldsymbol{\beta}_p + \mathbf{X}_o \boldsymbol{\beta}_o)$$

and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. The asymptotically correct standard error is given in (6) with

$$A = \left\{ \frac{\sum_{i=1}^{N} \nabla_{\boldsymbol{\beta}} g\left(\widehat{\boldsymbol{\beta}}, \mathbf{X}_i\right)}{N} \right\} \widehat{\mathrm{AVAR}}\left(\widehat{\boldsymbol{\beta}}\right) \left\{ \frac{\sum_{i=1}^{N} \nabla_{\boldsymbol{\beta}} g\left(\widehat{\boldsymbol{\beta}}, \mathbf{X}_i\right)}{N} \right\}' \tag{12}$$

$$B = \frac{\sum_{i=1}^{N} \left\{ g\left(\widehat{\boldsymbol{\beta}}, \mathbf{X}_i\right) - \widehat{\gamma} \right\}^2}{N} \tag{13}$$

and

$$\nabla_{\boldsymbol{\beta}} g(\boldsymbol{\beta}, \mathbf{X}) = \varphi(\mathbf{X}^1 \boldsymbol{\beta}) \mathbf{X}^1 - \varphi(\mathbf{X}^0 \boldsymbol{\beta}) \mathbf{X}^0 \tag{14}$$

where $\mathbf{X}^1 = [\mathbf{1} \; \mathbf{X}_o]$ and $\mathbf{X}^0 = [\mathbf{0} \; \mathbf{X}_o]$. The Stata and Mata do-file used to obtain the estimate of the ATE of disability on employment and calculate its TF-based asymptotically correct standard error using (12)–(14) is

```
/*************************************************
** This do-file estimates ATE of disability on
** employment, along with its correct asymptotic
** standard error.
** Stata probit procedure used to fit the model.
** Fishback-Terza data are used.
** Program implements Mata, not margins.
*************************************************/
/*************************************************
** Preliminary Stuff.
*************************************************/
clear mata
clear matrix
clear
set more off
capture log close

/*************************************************
** Set the default directory.
*************************************************/
cd <PATH FOR THE DEFAULT DIRECTORY>

/*************************************************
** Set up the output file.
*************************************************/
log using ATE-Simple-disabil.log, replace
```

```
/*************************************************
** Read in the data.
*************************************************/
use fishback-terza-male-female-data-clean.dta

/*************************************************
** Compute descriptive statistics.
*************************************************/
summarize employed disabil exper othhinc male black grade  ///
 inschool vet neverm income4 worlft75 spanish indian       ///
 foreignb nonlengl smsa regso regwe regnc                  ///
 evermus npershh

/*************************************************
** Probit.
*************************************************/
probit employed disabil male black grade inschool       ///
 vet neverm income4 worlft75 spanish indian foreignb    ///
 nonlengl smsa regso regwe regnc evermus npershh exper othhinc

/*************************************************
** Start mata.
*************************************************/
mata:

/*************************************************
** Save probit coefficient estimates and covariance
** matrix estimate.
*************************************************/
beta=st_matrix("e(b)")´
AVARbeta=st_matrix("e(V)")

/*************************************************
** End mata.
*************************************************/
end

/*************************************************
** Post data into mata.
*************************************************/
putmata employed disabil neverm male grade exper        ///
 vet worlft75 foreignb nonlengl black indian            ///
 spanish regnc regso regwe smsa inschool evermus        ///
 npershh othhinc income4

/*************************************************
** MATA Start-up.
*************************************************/
mata:

/*************************************************
** Sample size.
*************************************************/
N=rows(employed)
```

```
/*************************************************
** Define the matrix of regressors, including
** Xp, Xo and a constant term.
*************************************************/
X=disabil, male, black, grade, inschool, vet, neverm,   ///
 income4, worlft75, spanish, indian, foreignb,          ///
 nonlengl, smsa, regso, regwe, regnc, evermus,          ///
 npershh, exper, othhinc, J(N,1,1)

/*************************************************
** Create the X sup 1 vector (Xp=1).
*************************************************/
Xsup1=X
Xsup1[.,1]=J(N,1,1)

/*************************************************
** Create the X sup 0 vector (Xp=0).
*************************************************/
Xsup0=X
Xsup0[.,1]=J(N,1,0)

/*************************************************
** Compute the index vector for (Xp=1).
*************************************************/
Xsup1Beta=Xsup1*beta

/*************************************************
** Compute the index vector for (Xp=0).
*************************************************/
Xsup0Beta=Xsup0*beta

/*************************************************
** Compute the estimated effect for each
** individual in the sample.
*************************************************/
g=normal(Xsup1Beta)-normal(Xsup0Beta)

/*************************************************
** Compute the average treatment effect.
*************************************************/
ATE=mean(g)

/*************************************************
** Compute the gradient of g with respect
** to beta.
*************************************************/
pbetag=normalden(Xsup1Beta):*Xsup1:-normalden(Xsup0Beta):*Xsup0

/*************************************************
** Average the gradient of g with respect to beta.
*************************************************/
pbetag=mean(pbetag)

/*************************************************
** Compute the estimated asymptotic variance.
*************************************************/
varATE=pbetag*(N:*(AVARbeta))*pbetag´:+mean((g:-ATE):^2)
```

```
/************************************************
** The corresponding standard error.
************************************************/
seATE=sqrt(varATE/N)

/************************************************
** Compute the relevant asymptotic t-statistic.
************************************************/
tstatATE=ATE/sqrt(varATE/N)

/************************************************
** Compute the corresponding p-value.
************************************************/
pvalue=2:*(1:-normal(abs(tstatATE)))

/************************************************
** Display effect results obtained via Mata.
************************************************/
header="ATE","asy-se","asy-t-stat","p-value" \ "","","",""
results=ATE,seATE,tstatATE,pvalue
resview=strofreal(results)
"disabil"
header \ resview

/************************************************
** MATA end.
************************************************/
end

log close
```

The relevant output from this do-file is

```
. probit employed disabil male black grade inschool
> vet neverm income4 worlft75 spanish indian foreignb
> nonlengl smsa regso regwe regnc evermus npershh exper othhinc
Iteration 0:   log likelihood = -19823.769
Iteration 1:   log likelihood =  -14122.42
Iteration 2:   log likelihood = -14057.314
Iteration 3:   log likelihood = -14057.182
Iteration 4:   log likelihood = -14057.182
Probit regression                               Number of obs     =     31,507
                                                LR chi2(21)       =   11533.17
                                                Prob > chi2       =     0.0000
Log likelihood = -14057.182                     Pseudo R2         =     0.2909

------------------------------------------------------------------------------
    employed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     disabil |  -.7658884   .0284741   -26.90   0.000    -.8216966   -.7100803
        male |  -.0182789   .0264679    -0.69   0.490    -.0701551    .0335973
       black |   .1075914   .0351067     3.06   0.002     .0387836    .1763992
       grade |  -.0024472   .0034459    -0.71   0.478     -.009201    .0043066
    inschool |   .3675846   .0800549     4.59   0.000     .2106799    .5244893
         vet |   .1636881   .0269121     6.08   0.000     .1109414    .2164349
      neverm |   .0455236   .0441799     1.03   0.303    -.0410674    .1321145
     income4 |   -.000053   4.29e-06   -12.35   0.000    -.0000614   -.0000446
    worlft75 |  -1.396906   .0209017   -66.83   0.000    -1.437872   -1.355939
     spanish |  -.0472764   .0467021    -1.01   0.311    -.1388108     .044258
      indian |   .0268261   .1385675     0.19   0.846    -.2447613    .2984135
    foreignb |  -.0268723   .0379451    -0.71   0.479    -.1012433    .0474986
    nonlengl |   .0947007   .0685965     1.38   0.167    -.0397459    .2291473
        smsa |   .1789895   .0217065     8.25   0.000     .1364455    .2215335
       regso |  -.0358343    .023932    -1.50   0.134    -.0827402    .0110717
       regwe |  -.0157091   .0275966    -0.57   0.569    -.0697974    .0383792
       regnc |   .0244632   .0240862     1.02   0.310     -.022745    .0716714
     evermus |   .0064609   .0192114     0.34   0.737    -.0311928    .0441146
     npershh |  -.0137427   .0062981    -2.18   0.029    -.0260868   -.0013986
       exper |  -.0275907   .0011352   -24.30   0.000    -.0298157   -.0253657
     othhinc |  -5.72e-06   7.64e-07    -7.48   0.000    -7.21e-06   -4.22e-06
       _cons |   2.015496    .081264    24.80   0.000     1.856222     2.17477
------------------------------------------------------------------------------

: header \ resview
                 1            2            3            4
    +-------------------------------------------------------+
  1 |          ATE       asy-se   asy-t-stat        p-value |
  2 |                                                       |
  3 |    -.216556     .0086262    -25.10448              0 |
    +-------------------------------------------------------+
```

The ATE estimate is displayed in the header over the first two columns of table 2, and its TF-based asymptotic standard error and t statistic are shown in the first two columns of the first row of that table. The analogous TF for the AME of household income and the AIE of experience were derived and coded in Stata and Mata. The effect estimates are given in the respective headers of table 2, and the TF-based standard errors and t statistics are correspondingly displayed in the first row of that table.[4]

4. The full do-files, available from the *Stata Journal* software package, are given in appendix A.

Table 2. ATE, AME, and AIE estimates and their standard errors

	Disability $\widehat{\gamma}=\widehat{\text{ATE}}=0.217$		Other household income $\widehat{\gamma}=\widehat{\text{AME}}=1.43\text{e–}06$		Experience $\widehat{\gamma}=\widehat{\text{AIE}}=0.007$	
Standard-error calculation method	Asymptotic standard error	Asymptotic t statistic	Asymptotic standard error	Asymptotic t statistic	Asymptotic standard error	Asymptotic t statistic
TF	0.0086262	−25.10	1.91e–07	−7.49	0.0002813	−24.64
RM	0.0086215	−25.12	1.88e–07	−7.61	0.0002831	−24.48
ATF	0.0086215	−25.12	1.88e–07	−7.61	0.0002831	−24.48

Alternatively, the ATE, AME, and AIE estimates and their standard errors (asymptotic t statistics) can be obtained via `margins`. Note that there are two options in `margins` for calculating standard errors: a) `vce(delta)`, which is only appropriate for cases in which the matrix of observations on the independent variables (say, $\boldsymbol{\chi}$) is assumed to be fixed in repeated samples; and b) `vce(unconditional)`, which is used in all other cases. Generally, $\boldsymbol{\chi}$ is not assumed to be fixed in repeated samples.[5] Therefore, the `vce(unconditional)` option is the most relevant to the current discussion. StataCorp (2017, 1455–1466) describes the three cases in which the `vce(unconditional)` option is applicable: case I—you have a representative sample and have not `svyset` your data; case II—you have a weighted sample and have not `svyset` your data; and case III—you have `svyset` your data. Case I is relevant here. In this case, the `margins` command for the ATE and AME [as defined in (2) and (3)] is

```
margins, dydx(varname) vce(unconditional)                        (a)
```

and for the AIE [as defined in (4)] with $\Delta = 1$, the proper use of `margins` is

```
margins, at((asobserved)_all)                         ///
    at(varname=generate(varname+1))                   ///
    contrast(atcontrast(r)) vce(unconditional)                   (b)
```

where *varname* is the name of the causal variable. The Stata statements in (a) and (b) constitute what we referred to above (in Q2) as the RM. We estimated the ATE (disability), AME (other household income), and AIE (experience) using the RM in (a) and (b), accordingly. For instance, the Stata do-file used to estimate the ATE of disability on employment using the RM is[6]

5. Dowd, Greene, and Norton (2014) derive standard-error formulations under the fixed-in-repeated-samples assumption. This is a very restrictive and unrealistic assumption that is usually invalid for empirical econometrics.
6. Note that, to invoke the `vce(unconditional)` option in `margins`, you must use the `vce(robust)` option in the relevant Stata estimation command.

```
/*************************************************
** This program estimates ATE of disability on
** employment, along with its correct asymptotic
** standard error.
** Stata probit procedure used to fit the model.
** Fishback-Terza data are used.
** Program implements margins, not Mata.
*************************************************/
/*************************************************
** Preliminary Stuff.
*************************************************/
clear mata
clear matrix
clear
set more off
capture log close

/*************************************************
** Set the default directory.
*************************************************/
cd <PATH FOR THE DEFAULT DIRECTORY>

/*************************************************
** Set up the output file.
*************************************************/
log using ATE-MARGINS-disabil.log, replace

/*************************************************
** Read in the data (Full dataset -- Males and
** Females).
*************************************************/
use fishback-terza-male-female-data-clean.dta

/*************************************************
** Compute descriptive statistics.
*************************************************/
summarize employed disabil exper othhinc male ///
 black grade inschool vet neverm             ///
 income4 worlft75 spanish indian foreignb     ///
 nonlengl smsa regso regwe regnc             ///
 evermus npershh

/*************************************************
** Probit.
*************************************************/
probit employed i.disabil male               ///
 black grade inschool vet neverm             ///
 income4 worlft75 spanish indian foreignb     ///
 nonlengl smsa regso regwe regnc             ///
 evermus npershh exper othhinc, vce(robust)

/*************************************************
** Calculate and store the margins results.
*************************************************/
margins, dydx(disabil) vce(unconditional)

log close
```

The relevant output from this do-file is

```
. probit employed i.disabil male
>  black grade inschool vet neverm
>  income4 worlft75 spanish indian foreignb
>  nonlengl smsa regso regwe regnc
>  evermus npershh exper othhinc, vce(robust)
Iteration 0:   log pseudolikelihood = -19823.769
Iteration 1:   log pseudolikelihood =  -14122.42
Iteration 2:   log pseudolikelihood = -14057.314
Iteration 3:   log pseudolikelihood = -14057.182
Iteration 4:   log pseudolikelihood = -14057.182
Probit regression                               Number of obs     =     31,507
                                                Wald chi2(21)     =    9046.46
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14057.182               Pseudo R2         =     0.2909

------------------------------------------------------------------------------
             |               Robust
    employed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.disabil |  -.7658884   .0284655   -26.91   0.000    -.8216797   -.7100971
        male |  -.0182789   .0265917    -0.69   0.492    -.0703978    .0338399
       black |   .1075914   .0324009     3.32   0.001     .0440868     .171096
       grade |  -.0024472   .0034933    -0.70   0.484    -.0092939    .0043994
    inschool |   .3675846   .0826749     4.45   0.000     .2055448    .5296244
         vet |   .1636881   .0277747     5.89   0.000     .1092507    .2181255
      neverm |   .0455236   .0432359     1.05   0.292    -.0392172    .1302643
     income4 |   -.000053   4.80e-06   -11.03   0.000    -.0000624   -.0000436
    worlft75 |  -1.396906   .0203743   -68.56   0.000    -1.436839   -1.356973
     spanish |  -.0472764   .0466468    -1.01   0.311    -.1387024    .0441496
      indian |   .0268261   .1193102     0.22   0.822    -.2070177    .2606699
    foreignb |  -.0268723   .0392518    -0.68   0.494    -.1038044    .0500597
    nonlengl |   .0947007   .0667308     1.42   0.156    -.0360892    .2254906
        smsa |   .1789895   .0220443     8.12   0.000     .1357835    .2221955
       regso |  -.0358343     .02374    -1.51   0.131    -.0823639    .0106953
       regwe |  -.0157091   .0280164    -0.56   0.575    -.0706202     .039202
       regnc |   .0244632   .0240745     1.02   0.310    -.0227219    .0716483
     evermus |   .0064609   .0191481     0.34   0.736    -.0310687    .0439905
     npershh |  -.0137427   .0062653    -2.19   0.028    -.0260224    -.001463
       exper |  -.0275907   .0011489   -24.01   0.000    -.0298425   -.0253389
     othhinc |  -5.72e-06   7.54e-07    -7.58   0.000    -7.19e-06   -4.24e-06
       _cons |   2.015496   .0820759    24.56   0.000      1.85463    2.176362
------------------------------------------------------------------------------

. /*************************************************
> ** Calculate and store the margins results.
> *************************************************/
. margins, dydx(disabil) vce(unconditional)
Average marginal effects                        Number of obs     =     31,507
Expression   : Pr(employed), predict()
dy/dx w.r.t. : 1.disabil

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.disabil |   -.216556   .0086215   -25.12   0.000    -.2334538   -.1996582
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

We first note that the ATE estimate produced by this code is identical to that which was obtained earlier via the above Mata code (shown in the header above the first two columns of table 2). The results for the RM-based asymptotic standard error and t statistic for the ATE, displayed in the second row (first two columns) of table 2, differ from the TF-based results (albeit slightly). Similar RM-based calculations for the AME of household income and the AIE of experience were conducted.[7] As was the case for the ATE, the corresponding estimates of the AME and AIE are identical to those obtained using the TF (they are given in the respective headers of table 2). RM-based standard errors and t statistics are correspondingly displayed in the first row of that table. They also differ slightly from the TF-based values.

We suspected that, asymptotically, the difference between the TF and RM results is nil. To verify our conjecture, we note, as the appendix available with Terza (2016b) shows, that the following asymptotic variance estimator is asymptotically equivalent to (6),

$$\widehat{\text{a var}}(\widehat{\gamma}) = A + B + 2C \tag{15}$$

where the scalar C is a function of $\widehat{\boldsymbol{\theta}}$, $\mathbf{X}_i$, and $\widehat{\gamma}$. The asymptotic variance estimators in (6) and (15) are asymptotically equivalent in that C converges to zero in the limit as N approaches ∞. To assess our conjecture empirically, we re-estimated the standard errors of the three effect parameters (and their asymptotic t statistics) using (15). We call this the adjusted TF (ATF) approach to estimating the standard errors.[8] The ATF and RM results are identical. From this, we conclude that RM should be used when the `margins` command is available and correct.

4 A case for which margins is unavailable or incorrect: Causal inference in the 2SRI context

Consider the regression model of Mullahy (1997), in which the objective is to draw causal inferences regarding the effect of prenatal smoking ($\mathbf{X}_p$–`CIGSPREG`) on infant birthweight (Y–`BIRTHWTLB`), while controlling for infant birth order (`PARITY`), race (`WHITE`), and sex (`MALE`). The regression model for the birthweight outcome that he proposed can be written as

$$\mathbf{Y} = \exp(\mathbf{X}_p\boldsymbol{\beta}_p + \mathbf{X}_o\boldsymbol{\beta}_o + \mathbf{X}_u\boldsymbol{\beta}_u) + \mathbf{e} \tag{16}$$

where $\mathbf{X}_u$ comprises unobservable variables that are potentially correlated with prenatal smoking (for example, general "health mindedness" of the mother), $\mathbf{e}$ is the regression error term, $\mathbf{X}_o = [$`PARITY WHITE MALE`$]$ is a row vector of regressors that are uncorrelated with $\mathbf{X}_u$, and $\mathbf{e}$ and the $\boldsymbol{\beta}$'s are the regression parameters.[9] At issue here is the fact that there exist unobservables (as captured by $\mathbf{X}_u$) that are correlated with both $\mathbf{Y}$ and $\mathbf{X}_p$. In other words, $\mathbf{X}_p$ is endogenous. To account for said endogeneity, we

7. The full do-files, available from the *Stata Journal* software package, are given in appendix B.
8. The full do-files used to obtain the ATF estimates and their asymptotically correct standard errors, available from the *Stata Journal* software package, are given in appendix C.
9. Mullahy (1997) does not explicitly specify the model in terms of the unobservable $\mathbf{X}_u$. Nevertheless, (16) is substantively identical to Mullahy's model (see Terza [2006]).

follow the 2SRI approach suggested by Terza, Basu, and Rathouz (2008) and formalize the correlation between $\mathbf{X}_u$ and $\mathbf{X}_p$ as

$$\mathbf{X}_p = \exp(\mathbf{W}\boldsymbol{\alpha}) + \mathbf{X}_u$$

where $\boldsymbol{\alpha}$ is a vector of regression parameters $\mathbf{W} = [\mathbf{X}_o\ \mathbf{W}^+]$ and $\mathbf{W}^+$ is a vector of identifying instrumental variables specified in this case as

$$\mathbf{W}^+ = [\texttt{EDFATHER EDMOTHER FAMINCOME CIGTAX}]$$

with

$$\begin{aligned} \texttt{EDFATHER} &\equiv \text{paternal schooling in years} \\ \texttt{EDMOTHER} &\equiv \text{maternal schooling in years} \\ \texttt{FAMINCOME} &\equiv \text{family income} \end{aligned}$$

and

$$\texttt{CIGTAX} \equiv \text{cigarette tax}$$

Suppose that the ultimate objective here is estimation of the causal effect of a policy that completely prevents and eliminates smoking during pregnancy. In this example, given a consistent estimate of $\boldsymbol{\theta}' = [\boldsymbol{\alpha}'\ \boldsymbol{\beta}']$ (say, $\widehat{\boldsymbol{\theta}}$) with $\boldsymbol{\beta}' = [\boldsymbol{\beta}_p\ \boldsymbol{\beta}_o\ \boldsymbol{\beta}_u']$, the estimated causal effect of interest is given by the following versions of (1) and (4),

$$\widehat{\gamma} = \sum_{i=1}^{N} \frac{g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right)}{N} \tag{17}$$

with

$$g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) = m\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_{pi} + \boldsymbol{\Delta}_i, \mathbf{W}_i\right) - m\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_{pi}, \mathbf{W}_i\right) \tag{18}$$

where $\boldsymbol{\Delta} = \boldsymbol{\Delta}_i = -\mathbf{X}_{pi}$, $\mathbf{Z}_i = [\mathbf{X}_{pi}\ \mathbf{W}_i]$ and

$$m\left(\widehat{\boldsymbol{\theta}}, \mathbf{X}_{pi}, \mathbf{W}_i\right) = \exp\left[\mathbf{X}_{pi}\widehat{\boldsymbol{\beta}}_p + \mathbf{X}_{oi}\widehat{\boldsymbol{\beta}}_o + \left\{\mathbf{X}_{pi} - \exp\left(\mathbf{W}_i\widehat{\boldsymbol{\alpha}}\right)\right\}\widehat{\boldsymbol{\beta}}_u\right]$$

The asymptotically correct standard error is obtained from (6), with

$$A = \left\{\frac{\sum_{i=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\theta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right)}{N}\right\} \widehat{\mathrm{AVAR}}(\widehat{\boldsymbol{\theta}}) \left\{\frac{\sum_{i=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\theta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right)}{N}\right\}' \tag{19}$$

$$B = \frac{\sum_{i=1}^{N} \left\{g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) - \widehat{\gamma}\right\}^2}{N} \tag{20}$$

and

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) = \left\{\boldsymbol{\nabla}_{\boldsymbol{\alpha}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right)\ \ \boldsymbol{\nabla}_{\boldsymbol{\beta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right)\right\} \tag{21}$$

where $\mathbf{Z}_i = [\mathbf{X}_{pi}\ \mathbf{W}_i]$ and $\mathbf{W}_i = [\mathbf{X}_{oi}\ \mathbf{W}_i^+]$

$$\begin{aligned}\boldsymbol{\nabla}_{\boldsymbol{\alpha}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) &= -\widehat{\boldsymbol{\beta}}_u \exp\left(\mathbf{W}_i \widehat{\boldsymbol{\alpha}}\right) g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) \mathbf{W}_i \\ \boldsymbol{\nabla}_{\boldsymbol{\beta}} g\left(\widehat{\boldsymbol{\theta}}, \mathbf{Z}_i\right) &= \exp\left(\mathbf{V}_i^{\Delta} \widehat{\boldsymbol{\beta}}\right) \mathbf{V}_i^{\Delta} - \exp\left(\mathbf{V}_i \boldsymbol{\beta}\right) \mathbf{V}_i \\ \mathbf{V}_i &= [\mathbf{X}_{pi}\quad \mathbf{X}_{oi}\quad \{\mathbf{X}_{pi} - \exp\left(\mathbf{W}_i \widehat{\boldsymbol{\alpha}}\right)\}]\end{aligned}$$

and

$$\mathbf{V}_i^{\Delta} = [\mathbf{X}_{pi} + \boldsymbol{\Delta}\quad \mathbf{X}_{oi}\quad \{\mathbf{X}_{pi} - \exp\left(\mathbf{W}_i \widehat{\boldsymbol{\alpha}}\right)\}]$$

We first estimated $\boldsymbol{\theta}' = [\boldsymbol{\alpha}'\ \boldsymbol{\beta}']$ using the 2SRI method as detailed in section 4 of Terza (2016a). The first- and second-stage 2SRI results are shown in tables 1 and 2 of that article. We then used the 2SRI result $(\widehat{\boldsymbol{\theta}})$ and (17) to obtain a consistent estimate of the desired AIE. The relevant asymptotic standard error and t statistic were calculated using (6) and (19)–(21).

Note that (19) requires the correct expression for $\widehat{\text{AVAR}}(\boldsymbol{\theta})$, and this cannot be obtained directly from the outputs from the `glm` commands used to produce consistent estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For instance, the covariance matrix estimate of second-stage `glm` for $\boldsymbol{\beta}$ would be consistent for

$$E\left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} q\right)^{-1} E\left(\nabla_{\boldsymbol{\beta}} q' \nabla_{\boldsymbol{\beta}} q\right) E\left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} q\right)^{-1} \tag{22}$$

where $\nabla_{\boldsymbol{\beta}} q$ and $\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} q$ are shorthand notation for the gradient and Hessian of

$$q(\boldsymbol{\theta}, \mathbf{X}_p, \mathbf{W}) = -\left(\mathbf{Y} - \exp\left[\mathbf{X}_p \boldsymbol{\beta}_p + \mathbf{X}_o \boldsymbol{\beta}_o + \{\mathbf{X}_p - \exp(\mathbf{W}\boldsymbol{\alpha})\} \boldsymbol{\beta}_u\right]\right)^2$$

respectively. But (22) is incomplete—an additional term is required to account for the fact that the estimator has two stages (see Terza [2016a]). It is this incorrect covariance matrix estimate that would be used by `margins` in calculating the standard error of the AIE. Therefore, `margins` cannot be used in this case. Following Terza (2016a), we have that the correct estimated asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ (the 2SRI estimator of $\boldsymbol{\theta}$) is

$$\widehat{\text{AVAR}}(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} N \times \widehat{\text{AVAR}} * (\widehat{\boldsymbol{\alpha}}) & \widehat{\mathbf{D}}_{12} \\ \widehat{\mathbf{D}}_{12}' & N \times \widehat{\mathbf{D}}_{22}^{\dagger} \end{bmatrix}$$

where $\widehat{\text{AVAR}} * (\widehat{\boldsymbol{\alpha}})$ is the estimated covariance matrix output by `glm` for the first-stage estimate of $\widehat{\boldsymbol{\alpha}}$, and $\widehat{\mathbf{D}}_{12}$ and $\widehat{\mathbf{D}}_{22}^{\dagger}$ are as defined in (22) and (26) of Terza (2016a), respectively.

The key Stata and Mata code for calculating the AIE estimate, its correct standard error, and its t statistic is

```
/*************************************************
** Purpose: Mullahy (1997) Birth Weight model.
** Estimation of the model using the 2SRI.
** The outcome variable (birth weight) is
** non-negative and continuous, and the policy
** variable (cigarette smoking) is nonnegative
** and a count.
*************************************************/

/*************************************************
** Initial Set-up.
*************************************************/
clear mata
clear matrix
clear
set more off
capture log close

/*************************************************
** Set up default directory.
*************************************************/
cd <PATH FOR THE DEFAULT DIRECTORY>

/*************************************************
** Set up the output file.
*************************************************/
log using Mullahy-Birthweight-2SRI.log, replace

/*************************************************
** Read in the data.
*************************************************/
use mullahy-birthweight-data.dta

/*************************************************
** Descriptive Statistics.
*************************************************/
summ

/*************************************************
** First-stage NLS for alphahat and residuals.
*************************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER    ///
 FAMINCOM CIGTAX88,                                 ///
 family(gaussian) link(log) vce(robust)

/*************************************************
** Save residuals.
*************************************************/
predict Xuhat, response

/*************************************************
** Save alphahat and its covariance matrix.
*************************************************/
mata: alphahat=st_matrix("e(b)")´
mata: COValphahat=st_matrix("e(V)")
```

```
/************************************************
** Second-stage NLS for betahat.
************************************************/
glm BIRTHWT CIGSPREG PARITY WHITE MALE Xuhat,       ///
 family(gaussian) link(log) vce(robust)

/************************************************
** Save betahat, its covariance matrix and
** single out the coefficient of Xu.
************************************************/
mata: betahat=st_matrix("e(b)")´
mata: COVbetahat=st_matrix("e(V)")
mata: Bu=betahat[5]

/************************************************
** Post needed variables to Mata.
************************************************/
putmata BIRTHWT CIGSPREG PARITY WHITE MALE EDFATHER ///
 EDMOTHER FAMINCOM CIGTAX88 Xuhat

/************************************************
** Start mata.
************************************************/
mata:

/************************************************
** Set up V and W matrices.
************************************************/
V=CIGSPREG, PARITY, WHITE, MALE,                    ///
 Xuhat, J(rows(PARITY),1,1)
W=PARITY, WHITE, MALE, EDFATHER, EDMOTHER,          ///
 FAMINCOM, CIGTAX88, J(rows(PARITY),1,1)

/************************************************
** Set up the vector of rhs variable names.
************************************************/
VNAMES="CIGSPREG", "PARITY", "WHITE", "MALE",       ///
 "Xuhat", "Constant"

/************************************************
** Set up Balpha and Bbeta.
************************************************/
Ba=-Bu:*exp(V*betahat)/*
 */:*exp(W*alphahat):*W
Bb=exp(V*betahat):*V

/************************************************
** Set Bbetabeta and Bbetaalpha.
************************************************/
Bbb=Bb´*Bb
Bba=Bb´*Ba

/************************************************
** Estimate the asymptotic covariance matrix.
************************************************/
D22hat=invsym(Bbb)*Bba*COValphahat*Bba´*invsym(Bbb)+COVbetahat
```

```
/************************************************
** Calculate the asymptotically correct
** standard errors.
************************************************/
ACSE=sqrt(diagonal(D22hat))

/************************************************
** Calculate the asymptotically correct
** t-stats.
************************************************/
tstats=betahat:/ACSE

/************************************************
** Compute the corresponding p-values.
************************************************/
pvalues=2:*(1:-normal(abs(tstats)))

/************************************************
** Display the results.
************************************************/
header="Variable","Coeff-Estimate","Std-errs","t-stat","p-value" \ "","","","",""
results=betahat,ACSE,tstats,pvalues
resview=strofreal(results)
header \ (VNAMES´,resview)

/************************************************
**  Computation of the AIE begins here.
************************************************/

/************************************************
** Set delta (the policy increment).
************************************************/
Xp=CIGSPREG
delta=-Xp

/************************************************
** Calculate AIE.
************************************************/
Vdelta=V
Vdelta[.,1]=Vdelta[.,1]:+delta
expVdelta=exp(Vdelta*betahat)
expV=exp(V*betahat)
expW=exp(W*alphahat)
g=expVdelta:-expV
AIE=mean(g)

/************************************************
** Calculate components of estimated asymptotic
** variance of estimated AIE.
************************************************/
gradthetag=-Bu:*expW:*g:*W,expVdelta:*Vdelta:-expV:*V
N=rows(V)
D11hat=N:*COValphahat
D22hatstar=N:*D22hat
D12hat=-D11hat*Bba´*invsym(Bbb)
Dhat=D11hat,D12hat \ D12hat´,D22hatstar
```

```
/*************************************************
** Compute the estimated asymptotic variance.
*************************************************/
avarAIEhat=mean(gradthetag)*Dhat*mean(gradthetag)´/*
*/:+mean((AIE:-g):^2)

/*************************************************
** Compute the corresponding standard error.
*************************************************/
seAIEhat=sqrt(avarAIEhat/N)

/*************************************************
** Compute the relevant asymptotic t statistic.
*************************************************/
tstatavarAIEhat=AIE/seAIEhat

/*************************************************
** Compute the corresponding p-value.
*************************************************/
pvalue=2:*(1:-normal(abs(tstatavarAIEhat)))

/*************************************************
** Display effect results obtained via Mata.
*************************************************/
header="AIEhat","asy-se","asy-t-stat","p-value" \ "","","",""
results=AIE,seAIEhat,tstatavarAIEhat,pvalue
resview=strofreal(results)
header \ resview

/*************************************************
** End Mata.
*************************************************/
end


log close
```

As shown in the header of table 3, the resultant AIE estimate is 3.68, implying that our hypothetical policy would serve to increase infant birthweight by nearly 4 ounces on average. The asymptotically correct standard errors and t statistics obtained via the TF using (19)–(21) are displayed in the first row of table 3.

Table 3. Smoking during pregnancy ($\widehat{\gamma} = \widehat{\text{AIE}} = 3.68$)

Standard-error calculation method	Asymptotic standard error	Asymptotic t statistic
TF	1.167	3.153
RM	1.046	3.517

For comparison, in table 3, we also give the incorrect results obtained using the following Stata code, which implements the RM in (b):[10]

```
/*************************************************
** First-stage NLS for alphahat and residuals.
*************************************************/
glm CIGSPREG PARITY WHITE MALE EDFATHER EDMOTHER        ///
 FAMINCOM CIGTAX88,                                     ///
 family(gaussian) link(log) vce(robust)

/*************************************************
** Save residuals.
*************************************************/
predict Xuhat, response

/*************************************************
** Second-stage NLS for alphahat and residuals.
*************************************************/
glm BIRTHWT CIGSPREG PARITY WHITE MALE Xuhat,           ///
family(gaussian) link(log) vce(robust)

/*************************************************
** Calculate and store the margins results.
*************************************************/
margins, at((asobserved) _all) at(CIGSPREG=generate(0)) ///
contrast(atcontrast(r)) vce(unconditional)
```

This code produces an incorrect value for the standard error of the AIE estimate, because `margins` implements a covariance matrix estimate that is consistent for (22), which ignores first-stage estimation of α and the concomitant requisite covariance matrix correction factor.

5 Summary and conclusion

Terza (2016b) gives the generic expression for the asymptotically correct standard errors of statistics formed as sample means of nonlinear data transformations. In this article, we assessed the performance of `margins` as a relatively simple alternative for calculating such standard errors. We noted that `margins` is not available for all packaged nonlinear

10. The full do-file for the RM estimates, available from the *Stata Journal* software package, is given in appendix D.

regression commands in Stata and cannot be implemented in conjunction with user-defined-and-coded nonlinear estimation protocols that do not make `predict` available. When `margins` is available, however, we established (using a real-data example) that it produces standard errors that are asymptotically equivalent to those obtained from the formulations in Terza (2016b). Given its relative coding simplicity, this result favors using `margins` when available. In all other cases, one can use Mata to code the standard-error formulations in Terza (2016b) [namely, (6) through (8) above].

6 References

Dowd, B. E., W. H. Greene, and E. C. Norton. 2014. Computation of standard errors. *Health Services Research* 49: 731–750.

Fishback, P. V., and J. V. Terza. 1989. Are estimates of sex discrimination by employers robust? The use of never-marrieds. *Economic Inquiry* 27: 271–285.

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.

StataCorp. 2017. *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

Terza, J. V. 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.

———. 2016a. Simpler standard errors for two-stage optimization estimators. *Stata Journal* 16: 368–385.

———. 2016b. Inference using sample means of parametric nonlinear data transformations. *Health Services Research* 51: 1109–1113.

———. Forthcoming. Two-stage residual inclusion estimation in health services research and health economics. *Health Services Research*.

Terza, J. V., A. Basu, and P. J. Rathouz. 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27: 531–543.

About the author

Joseph V. Terza is a health economist and econometrician in the Department of Economics at Indiana University–Purdue University Indianapolis. His research focuses on the development and application of methods for estimating qualitative and limited dependent variable models with endogeneity. Two of his methods have been implemented as Stata commands. He was keynote speaker at the Stata Users Group meeting in Mexico City in November 2014.