The Stata Journal (2023)
23, Number 4, pp. 1057–1061 DOI: 10.1177/1536867X231212445

Simpler standard errors for two-stage optimization estimators revisited

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

Abstract. "Simpler standard errors for two-stage optimization estimators" (Terza, 2016a, *Stata Journal* 16: 368–385) offers an analytic simplification of the daunting textbook formulations of the asymptotic variance–covariance matrix of a class of two-stage optimization estimators. Here I revisit that simplification and show that it applies to a much broader class of estimators than was originally considered. I also offer a correction that further enhances the generality of this asymptotic variance–covariance matrix formulation. These points are illustrated via a real-data application.

Keywords: st0736, two-stage optimization estimators, standard errors, asymptotic theory, endogeneity, two-stage residual inclusion, sandwich estimator

1 Introduction

In a previous article (Terza 2016a), I offered an analytic simplification of the daunting textbook formulation of the asymptotic variance–covariance matrix (AVAR) of the class of two-stage optimization estimators (2SOE) for which the second-stage objective function has either a least-squares or a log-likelihood formulation (I henceforth refer to this class of estimators as LS-LL-2SOE). An important condition established in that article holds for a much more general class that subsumes LS-LL-2SOE. I also correct a subtle but potentially important error in the AVAR formulation of Terza (2016a); the correction serves to enhance the generality of the AVAR simplification.

2 A generalization and a correction

The null condition given in (10) of Terza (2016a) greatly simplifies the AVAR formulation for LS-LL-2SOE as detailed in (11) and (12). For the remainder of this discussion, we use the notation and equation numbers from Terza (2016a). In appendix A, which will be made available by the *Stata Journal* as online supplementary material, I show that this null condition holds for a broader class of two-stage estimators in which the relevant objective functions, q_1 and q_2, are formulated as conditional random functions.

Terza (2016a) considered applications of LS-LL-2SOE via packaged commands in Stata, with supplementary Mata code for the calculation of the estimates of relevant

AVAR and the implied asymptotic standard errors (ASE). The two highlighted cases were those in which the second-stage estimator is implemented using a nonlinear least-squares command (for example, `glm` with the appropriate options) or a maximum likelihood (ML) command (for example, `probit` or `poisson`). The discussion of the former offered in Terza (2016a) warrants no comment, revision, or correction. The explication therein of the latter is, however, flawed.

Terza (2016a) claims that, when the second-stage estimator is implemented via an ML command, the following conditions hold:

$$\mathbf{E}\left(\boldsymbol{\nabla}_{\beta\beta}\mathbf{q}_2\right) = -\mathbf{E}\left(\boldsymbol{\nabla}_{\beta}\mathbf{q}_2'\boldsymbol{\nabla}_{\beta}\mathbf{q}_2\right) \tag{7}$$

$$\mathbf{E}\left(\boldsymbol{\nabla}_{\beta\alpha}\mathbf{q}_2\right) = -\mathbf{E}\left(\boldsymbol{\nabla}_{\beta}\mathbf{q}_2'\boldsymbol{\nabla}_{\alpha}\mathbf{q}_2\right) \tag{8}$$

These conditions would provide a further simplification in the implementations of (11) and (12) in that the researcher would only be required to supply first-order analytic derivatives and coding for the estimation of the related components of the relevant AVAR estimator [see (13) and (14)]. Unfortunately, although $\mathbf{q}_2$ may take the form of a log-likelihood function, it is not necessarily either the true joint full-information log likelihood for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ or the true conditional log likelihood for $\boldsymbol{\beta}$ (Vuong 1984). Therefore, (7) and (8) do not necessarily hold. In which case, (11) and (12) would be replaced by

$$\mathbf{D}_{12} = \mathbf{AVAR}^*\left(\widehat{\boldsymbol{\alpha}}\right)\mathbf{E}(\boldsymbol{\nabla}_{\beta\alpha}\mathbf{q}_2)'\mathbf{AVAR}^*\left(\widehat{\boldsymbol{\beta}}\right) \tag{11'}$$

$$\begin{aligned}\mathbf{D}_{22} = {} & \mathbf{AVAR}^*\left(\widehat{\boldsymbol{\beta}}\right)\mathbf{E}(\boldsymbol{\nabla}_{\beta\alpha}\mathbf{q}_2)\mathbf{AVAR}^*\left(\widehat{\boldsymbol{\alpha}}\right)\mathbf{E}(\boldsymbol{\nabla}_{\beta\alpha}\mathbf{q}_2)'\mathbf{AVAR}^*\left(\widehat{\boldsymbol{\beta}}\right) \\ & + \mathbf{AVAR}^*\left(\widehat{\boldsymbol{\beta}}\right)\mathbf{E}(\boldsymbol{\nabla}_{\beta}\mathbf{q}_2'\boldsymbol{\nabla}_{\beta}\mathbf{q}_2)\mathbf{AVAR}^*\left(\widehat{\boldsymbol{\beta}}\right)\end{aligned} \tag{12'}$$

By the same token, however, this discussion clarifies that (11') and (12') [without imposing (7) and (8)] are valid for a much broader class of estimators that subsumes LS-LL-2SOE.

3 Example: Smoking and infant birthweight

Here I consider the same model (regression of infant birthweight on endogenous smoking during pregnancy) data and variables (Mullahy 1997) and estimator (two-stage residual inclusion [2SRI]—Terza, Basu, and Rathouz [2008]) discussed in Terza (2016a) with one change: in this treatment, I implement an ML command (the `streg` command with the `distribution(lognormal)` option). In this version of the model, the unobservable confounder (X_u) is included as a regressor in the specification of the location parameter of the lognormal. See Terza (2016a) for a discussion of X_u in the present context.

I consider two different approaches to estimating the AVAR of the 2SRI second-stage estimator, $\widehat{\boldsymbol{\beta}}$: a) with the incorrect conditions (7) and (8) imposed [so the AVAR estimator is based on (14) in Terza (2016a)], and b) without (7) and (8) imposed [with

the AVAR estimator based on (12') above]. The `.do` and `.dta` files for both estimation protocols are published as part of the online *Stata Journal* package for this article. The suffixes `incorrect` and `correct` in the titles of the do-files are indicative of estimation approach (a) versus (b), respectively, for the AVAR of the 2SRI second-stage estimator. The requisite matrix calculus formulations for estimation of the components of (14) and (12') are given in appendix B, which will be made available by the *Stata Journal* as online supplementary material. The results are shown in table 1.

Table 1. 2SRI second-stage estimates

Variable	Parameter estimate	Correct 2SRI [(7) and (8) not imposed]		Incorrect 2SRI [(7) and (8) imposed]		Raw `streg` output	
		ASE	t stat	ASE	t stat	ASE	t stat
		1	2	3	4	5	6
`CIGSPREG`	−0.014	0.00441	−3.169	0.00424	−3.30	0.00385	−3.63
`PARITY`	0.018	0.00585	3.034	0.00598	2.97	0.00568	3.12
`WHITE`	0.060	0.01473	4.045	0.01385	4.30	0.01221	4.88
`MALE`	0.030	0.01031	2.872	0.01078	2.75	0.01009	2.93
X_u	0.010	0.00447	2.218	0.00432	2.29	0.00396	2.51
Constant	1.926	0.01991	96.744	0.01832	105.17	0.01701	113.25
$\ln(\sigma)$	−1.683	0.04501	−37.392	0.02158	−78.00	0.01898	−88.68

ASEs and t statistics for the universally correct approach [(12'), which does not impose (7) and (8)] are shown in columns 1 and 2. Columns 3 and 4 display the results obtained using (14), which incorporates conditions (7) and (8). Results from the raw `streg` output are shown in columns 5 and 6. As can be seen from comparing columns 1 and 2 with 3 and 4, dropping conditions (7) and (8) makes a difference, albeit an apparently small one in this application. As expected, the raw `streg` results yield uniformly smaller (larger [in absolute value]) standard errors (t statistics) compared with the correct results (columns 3 and 4).

Without further comprehensive study, it is difficult to assess (generally) how, in application, results and conclusions may differ because of the use of the universally correct approach based on (12'). It is even less clear how that difference might weigh against the simplification in the implementation of (14) that conditions (7) and (8) afford. These are topics for future research.

One way to think about the divergence or similarity between the ASEs (or t statistics) from the two approaches is to note that, for a 2SOE protocol with an ML second stage that is designed to account for endogeneity (for example, 2SRI), the less pronounced the unobservable confounding problem is, the more likely it will be that (7) and (8) hold. In the extreme, if there is no endogeneity, then the second-stage objective function ($\mathbf{q}_2$) [see (B-8) in appendix B] is indeed the correct conditional log-likelihood function for $\boldsymbol{\beta}$ [in which case, of course, conditions (7) and (8) hold].

In our example, as is usual in the 2SRI context, we turn to the coefficient of X_u (the latent variable that comprises the unobservable confounders) to draw inference regarding the severity of the endogeneity problem. As we see in table 1, this coefficient is marginally statistically significant (p-value > 0.01). Hence, in this example, it may be the case that the $\mathbf{q}_2$ function is serving as a good approximation to the true log likelihood—that is, imposing (7) and (8) does not distort inference.

4 Conclusions

I revisited the derivation of the AVAR formulation offered in Terza (2016a) with a view toward greater generality. Moreover, I offered a correction to the AVAR formulation that serves to broaden the applicability of the analytic and computational simplification suggested by Terza (2016a). These points were illustrated via a real-data application.

To free users from the requisite Mata coding in this context, future Stata versions could include a packaged command for the generic 2SOE that incorporates the analytics, coding details, and simplifications offered in this article and in Terza (2016a). As a first step in this direction, a command implementing the widely used 2SRI estimator could be developed. In the meantime, substantial simplification or elimination of the Mata coding demands for the ASEs of 2SOEs (as detailed in the present article and in Terza [2016a]) can be achieved using the Mata `deriv()` function (Terza 2023a). Moreover, the `margins` command could be correspondingly extended to cover predictive margins (for example, causal effect parameters) and their ASEs in the 2SOE/2SRI context (Terza 2016b, 2017). Here, again, implementation of the `deriv()` function would serve to greatly reduce analytic and coding demands (Terza 2023b).

5 Acknowledgments

This research was supported by the National Institute on Alcohol Abuse and Alcoholism under subcontract to the Alcohol Research Group (Center Grant #P50 AA05595-41).

6 Programs and supplemental material

To install the software files as they existed at the time of publication of this article, type

```
. net sj 23-4
. net install st0736     (to install program files, if available)
. net get st0736         (to install ancillary files, if available)
```

7 References

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. https://doi.org/10.1162/003465397557169.

Terza, J. V. 2016a. Simpler standard errors for two-stage optimization estimators. *Stata Journal* 16: 368–385. https://doi.org/10.1177/1536867X1601600206.

———. 2016b. Inference using sample means of parametric nonlinear data transformations. *Health Services Research* 51: 1109–1113. https://doi.org/10.1111/1475-6773.12494.

———. 2017. Causal effect estimation and inference using Stata. *Stata Journal* 17: 939–961. https://doi.org/10.1177/1536867X1801700410.

———. 2023a. Simpler standard errors for two-stage M-estimators in empirical economics using numerical derivatives. Unpublished manuscript.

———. 2023b. Standard errors for regression-based causal effect estimates in economics using numerical derivatives. Unpublished manuscript.

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. https://doi.org/10.1016/j.jhealeco.2007.09.009.

Vuong, Q. H. 1984. Two-stage conditional maximum likelihood estimation of econometric models. Social science working paper, California Institute of Technology. https://resolver.caltech.edu/CaltechAUTHORS:20170919-135009248.

About the author

Joseph V. Terza is a health economist and econometrician in the Department of Economics at the Indiana University School of Liberal Arts at Indiana University–Purdue University Indianapolis. His research focuses on the development and application of nonlinear regression-based methods for empirical causal analysis in the presence of endogeneity (typically in qualitative and limited dependent variables settings). Two of his methods have been implemented as Stata commands and another as an ado-file. He was a keynote speaker at the Stata Users Group meeting in Mexico City in November 2014. His work (with coauthors) has been featured at the annual Stata Conferences in 2016 and 2018–2022.