

© Health Research and Educational Trust  
DOI: 10.1111/1475-6773.12494  
EDITORIAL

## Editorial

# Inference Using Sample Means of Parametric Nonlinear Data Transformations

In empirical HSR, statistics of key analytic interest are often of the following general form

$$\hat{\gamma} = \sum_{i=1}^N \frac{g(\hat{\beta}, X_i)}{N} \quad (1)$$

where  $\gamma = E[g(\beta, X)]$  is the parameter of ultimate interest to be estimated by equation (1),  $g(\cdot)$  is a known (possibly nonlinear) transformation,  $\hat{\beta}$  is a pre-estimate of  $\beta$ —a vector of “deeper” model parameters, and  $X_i$  denotes a vector of observed data on  $X$  for the  $i$ th member of a sample of size  $n$  ( $i = 1, \dots, N$ ). The three most commonly encountered formulations of equation (1)—average treatment effect (ATE), average marginal effect (AME), and average incremental effect (AIE)—correspond to the following formulations of  $g(\cdot)$ , respectively

$$m(\beta, 1, X_o) - m(\beta, 0, X_o) \quad (2)$$

$$g(\beta, X) = \frac{\partial m(\beta, X_p, X_o)}{\partial X_p} \quad (3)$$

$$m(\beta, X_p + \Delta, X_o) - m(\beta, X_p, X_o) \quad (4)$$

where  $m(\beta, X_p, X_o) = E[Y | X_p, X_o]$  is a regression function written so as to highlight the distinction between a policy-relevant regressor of interest,  $X_p$ , and a vector of regression controls,  $X_o$ ;  $X = [X_p \ X_o]$ ;  $\beta$  is a vector of regression parameters, and  $\Delta$  is a known exogenous (usually policy-driven) increment to  $X_p$ . After the regression parameter estimates are obtained [e.g.,  $\hat{\beta}$ —estimated via the nonlinear least (NLS) method], under fairly general

conditions, in conjunction with equation (1), the formulations in equation (2), (3), and (4), respectively yield consistent estimators of the ATE when  $X_p$  is binary; the AME when  $X_p$  is continuous and interest is in the effect attributable to an infinitesimal policy change; and the AIE when  $X_p$  is discrete or continuous and the relevant policy increment is  $\Delta$ . In this note, we focus on the specification and computation of the correct “*t*-statistic” for equation (1) as derived from standard asymptotic theory. This *t*-statistic has the following general form

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

where  $\gamma^\dagger$  is the relevant “null” value of  $\gamma$  (as in a test of the null hypothesis  $H_0: \gamma = \gamma^\dagger$ ), and  $se(\hat{\gamma})$  is the asymptotic standard error of equation (1) defined as  $se(\hat{\gamma}) \equiv \sqrt{\widehat{avar}(\hat{\gamma})}$ , with  $\widehat{avar}(\hat{\gamma})$  being a consistent estimator of the asymptotic variance of  $\hat{\gamma}$ . Under slightly stronger conditions than those required for the consistency of equation (1), it can be shown that equation (5) is asymptotically standard normal distributed. In the remainder of this note, we take the consistency and asymptotic normality of  $\hat{\gamma}$  as given, and concentrate on the correct formulation of  $se(\hat{\gamma})$  as derived from standard asymptotic theory. In the Appendix, we show that for most (if not all) of the useful forms of equation (1)

$$\widehat{avar}(\hat{\gamma}) = A + B \tag{6}$$

where  $A = \left( \frac{\sum_{i=1}^N \nabla_{\beta} g(\hat{\beta}, X_i)}{N} \right) \widehat{AVAR}(\hat{\beta}) \left( \frac{\sum_{i=1}^N \nabla_{\beta} g(\hat{\beta}, X_i)}{N} \right)'$

$$B = \frac{\sum_{i=1}^N (g(\hat{\beta}, X_i) - \hat{\gamma})^2}{N}$$

$\nabla_{\beta} g(\hat{\beta}, X_i)$  (a row vector) denotes the gradient of  $g(\beta, X)$  evaluated at  $X_i$  and  $\hat{\beta}$ , and  $\widehat{AVAR}(\hat{\beta})$  is an estimator of the asymptotic covariance matrix of  $\hat{\beta}$ . Dowd, Greene, and Norton (2014) opine that inclusion of  $B$  in equation (6) “seems incorrect to us” and exclude it from the suggested formulation of the asymptotic standard error of  $\hat{\gamma}$  given in their equation (18). That the inclusion

---

Address correspondence to Joseph V. Terza, Ph.D., Department of Economics, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202; e-mail: [jvterza@iupui.edu](mailto:jvterza@iupui.edu).

of  $B$  is not “incorrect” is proven by the derivation in an appendix that is included among the supplementary materials for this paper.<sup>1,2</sup>

So how does our derivation of equation (6) differ from the approach taken by Dowd, Greene, and Norton (2014) in deriving their equation (18)? To answer we must take a closer look at the sampling assumptions underlying the respective derivations. Unlike Dowd, Greene, and Norton (2014), we assume that the sample observations for all the relevant variables, including  $X$ , are drawn randomly from the relevant joint distribution for the population of interest. We impose no sampling restriction on  $X$  and we allow it to be random—the same assumption that we make for the other elements of the data vector. For example, for the case in which  $\hat{\beta}$  is obtained by regressing  $Y$  on  $X$  (based on a correctly specified nonlinear model), we treat *both  $Y$  and  $X$*  as random in sampling. What we have described here is simple random sampling (SRS), which is clearly the most commonly encountered type of sampling in empirical HSR. Moreover, we adopt the conventional approach to deriving the asymptotic properties of  $\hat{\gamma}$  (in particular, its asymptotic standard error). Conventional asymptotic theory assumes SRS and focuses on the limiting properties of estimators as the sample size ( $N$  in our case) approaches  $\infty$ . (The analysis assumes that the same sample is used to estimate  $\beta$  and the mean effect on  $m(\cdot)$ , and that the objective of the analysis involves generalization to a population that is potentially large compared to the sample; equation (6) can be modified for alternative assumptions).

Dowd, Greene, and Norton (2014), on the other hand, supplant SRS with an unrealistic fixed-in-repeated-sampling assumption (FIRS) in which the matrix of observations on  $X$  (say  $\chi$  with  $N$  rows and  $K$  columns;  $K$  being the number of regressors in  $X$ ) is fixed (nonrandom) so that, in sampling, only the observations on  $Y$  are randomly drawn. Moreover, they assume that increases in the sample size are not effected by increasing  $N$  but instead by holding it and  $\chi$  fixed and drawing repeated observations on  $Y$  for each of the fixed rows of  $\chi$ . Denote the number of such FIRS observations on  $Y$  as  $N^*$ . Their formulation of the asymptotic standard error of  $\hat{\gamma}$  is obtained while fixing  $\chi$  and  $N$  and allowing  $N^*$  to approach  $\infty$ . Both the FIRS and its attendant asymptotics are unrealistic and irrelevant in the present context because there are no empirical contexts in HSR for which this assumption could be reasonably maintained. The FIRS assumption is characteristic of an experiment where no generalization is intended beyond the specific designed distribution of  $X$ , rather than of an analysis of a random sample drawn from and intended to be representative of a larger population.

Given that (a) the formulation of the asymptotic standard error of  $\hat{\gamma}$  in equation (6) is realistic, relevant, and indeed correct; (b) the practical significance of including B can only be conclusively evaluated in the context of each particular empirical application after it has been estimated; and (c) the calculation of B imposes only minimal marginal computational burden, there remains no reasonable justification for excluding it from equation (6) as Dowd, Greene, and Norton (2014) recommend in their equation (18).

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* This research was supported by a grant from the Agency for Healthcare Research and Quality (R01 HS017434-01) and by grants from the National Institutes of Health (NIH-1 R01 CA155329-01 and NIH-1RC4AG038635-01).

*Disclosures:* None.

*Disclaimers:* None.

Joseph V. Terza

## NOTES

1. Basu and Rathouz (2005) and Wooldridge (2010) derive equation (6) via standard asymptotic theory. See Appendix C of Basu and Rathouz (2005), and problem 12.17 of Wooldridge (2010), the solution for which is on pp. 184–186 of Wooldridge (2011).
2. In the Supplementary Appendix, we also show that for the most general version of equation (1), an additional term would need to be included in equation (6). Such general versions of equation (1) coincide with models that do not afford causal interpretation of the estimates of  $\beta$  obtained there from and are, therefore, of very limited empirical analytic interest. See the Supplementary Appendix for details.

## REFERENCES

- Basu, A., and P. J. Rathouz. 2005. "Estimating Marginal and Incremental Effects on Health Outcomes Using Flexible Link and Variance Function Models." *Biostatistics* 6: 93–109.
- Dowd, B. E., W. H. Greene, and E. C. Norton. 2014. "Computation of Standard Errors." *Health Services Research* 49: 731–50.

- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd Edition. Cambridge, MA: MIT Press.
- . 2011. *Solutions Manual and Supplementary Materials for Econometric Analysis of Cross Section and Panel Data*, 2nd Edition. Cambridge, MA: MIT Press.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1: Derivation of Equation (6).