© Health Research and Educational Trust
 DOI: 10.1111/1475-6773.12493
 REPLY

Reply

Response to “Inference Using Sample Means of Parametric Nonlinear Data Transformations”

We thank Joseph Terza for his careful read (Terza 2016) of our paper (Dowd, Greene, and Norton 2014). There is no disagreement about how to evaluate the asymptotic distribution of a nonlinear function ($\hat{\gamma}$) of estimated parameters ($\hat{\beta}$) or about the first term (A) in Terza’s equation 6. The discussion is entirely about the second term (B) in equation 6.

The function in question is $\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^n \hat{\gamma}_{ik}$, for example, the partial effect of the k^{th} explanatory variable on the probability that a binary-dependent variable is equal to 1 in a logit or probit model, averaged across the observations, X_{ik} in the sample. The subscript i refers to the individual observation in a sample of size n . Our discussion of the standard error of $\hat{\gamma}_k$ was based on the equation $\hat{\gamma}_{ik} = g(\hat{\beta})|X_i$, consistent with the assumption that the data are “fixed in repeated samples.” This means that it is possible to collect additional samples of data with the same values of X . Different samples may produce different values of the dependent variable conditional on X , but the difference will be due *solely* to the observations having different error terms.

The expression $\hat{\gamma}_{ik} = g(\hat{\beta})|X_i$ is easy to understand when $g(\hat{\beta})|X_i$ is the partial effect of X_k evaluated for a specific type of individual whose values of X are equal to X^o . Asymptotic theory has a clear interpretation in this case. In ever larger samples of people with $X = X^o$, $\hat{\beta}$ converges to the true population β ; the variance of $g(\hat{\beta})|X = X^o$ shrinks to zero; and $g(\hat{\beta})|X = X^o$ converges to the true value for the entire population *for whom* $X = X^o$.

Our discussion was based on the assumption that the sample of observations that produced the estimates of $\hat{\beta}$ is a collection of individuals, each with their own value of $X = X_i^o$ and moving from $\hat{\gamma}_k = g(\hat{\beta})|X = X^o$ for a subset of

the sample to $\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^n g(\hat{\beta}) | X_i = X_i^o$ for the full sample changes nothing about the assumptions underlying the explanatory variable values, or the interpretation of asymptotic theory—that is, that the effect of larger samples is to add more error terms to each value of $X_i = X_i^o$.

Terza takes issue with the assumption of fixed X . If the overarching assumption is changed to random sampling from the population of (y_i, X_i) , then a second term that accounts for the sampling variation of the average over X_i enters the asymptotic variance of $\hat{\gamma}_k$. This is the ‘B’ term in equation 6, which, after accommodating the common definition of the asymptotic variance, is $\frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^n \left[g(\hat{\beta}, X_i) - \bar{g}(\hat{\beta}, X) \right]^2 \right)$.

Researchers can come to different conclusions on this issue. Terza’s approach has intuitive appeal because when every X_i truly is an element of a random sample, estimates of the population mean of functions of X_i , that is, $\frac{1}{n} \sum_{i=1}^n g(\beta, X_i)$, converge in probability to the population mean of $g(\beta, X_i)$ as the sample size increases. However, in our discussion, the analogy to larger samples is an increasing number of samples drawn from the same values of X_i as in the original estimation sample, that is, bootstrap samples. As we showed in our original article, bootstrapping produces standard error estimates that are virtually identical to those produced by the delta method without Terza’s term B, and to those produced by the method of Krinsky and Robb.

Analysts will decide which approach best suits their purposes. We do note the following about recent common practice. If the analyst requests $\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^n \hat{\gamma}_{ik}$ from ‘margins’ in Stata or ‘PARTIALS’ in NLOGIT, the software will compute the standard error of $\hat{\gamma}_k$ using the delta method, that is, $\frac{1}{n} \sum_{i=1}^n g(\hat{\beta}) | X_i = X_i^o$, term A in Terza’s equation 6, rather than $\frac{1}{n} \sum_{i=1}^n g(\hat{\beta}, X_i)$, that is, both terms A and B.

Address correspondence to Bryan E. Dowd, Ph.D., School of Public Health, Division of Health Policy and Management, University of Minnesota, Box 729 MMC, Minneapolis, MN 55455; e-mail: dowdx001@umn.edu. William H. Greene, Ph.D., is with the Stern School of Business, New York University, New York, NY. Edward C. Norton, Ph.D., is with the Department of Health Management and Policy and the Department of Economics at the University of Michigan, Ann Arbor, MI.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: No financial support.

Disclosures: None.

Disclaimers: None.

William H. Greene

Bryan E. Dowd

Edward C. Norton

REFERENCES

- Dowd, B. E., W. H. Greene, and E. C. Norton. 2014. "Computation of Standard Errors." *Health Services Research* 49 (2): 731–50.
- Terza, J. V. 2016. "Inference Using Sample Means of Parametric Nonlinear Data Transformations." *Health Services Research* 51 (3 Pt 1): 1109–13.

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.