INFERENCE USING SAMPLE MEANS OF PARAMETRIC NONLINEAR DATA TRANSFORMATIONS

(Supplementary Appendix)

by

Joseph V. Terza
Department of Economics
Indiana University Purdue University Indianapolis
Indianapolis, IN 46202
Phone: 317-274-4747
Fax: 317-274-0097
Email: jvterza@iupui.edu

(November, 2015)

DO NOT CITE OR DISTRIBUTE WITHOUT AUTHOR'S PERMISSION

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).

The most commonly applied parametric estimators reside in the class of *optimization estimators (OEs)* – statistical methods that produce estimates as optimizers of well specified objective functions.[1] Model design or computational convenience often dictates that an OE be implemented in two stages. In such cases the parameter vector of interest is partitioned as $\omega' = [\delta' \quad \tau']$ and conformably estimated in two-stages. First, an estimate of δ is obtained as the optimizer of an appropriately specified first-stage objective function

$$\sum_{i=1}^{N} q_1(\delta, V_{1i}) \qquad \text{(A-1)}$$

where V_{1i} denotes the relevant subvector of the observable data for the ith sample individual (i = 1, ..., N). Next, an estimate of τ is obtained as the optimizer of

$$\sum_{i=1}^{N} q_2(\hat{\delta}, \tau, V_{2i}) \qquad \text{(A-2)}$$

where V_{2i} denotes the second-stage analog to V_{1i} and $\hat{\delta}$ is the first-stage estimate of δ.

It is well established that under general conditions, this two-stage optimization estimator (2SOE) is consistent and asymptotically normal.[2] Our interest here is in detailing the asymptotic covariance matrix of $\hat{\tau}$ -- the second-stage estimator obtained from (A-2). Before proceeding we establish the following notational conventions:

-- q_1 is shorthand notation for $q_1(\delta, V_{1i})$ as defined in (A-1)

-- q_2 is shorthand notation for $q_2(\hat{\delta}, \tau, V_{2i})$ as defined in (A-2)

[1] Sometimes called M-estimators.
[2] See Newey and McFadden (1994) or White (1994) for details.

-- $\nabla_s q_j$ denotes the gradient of q_j $(j = 1, 2)$ with respect to parameter subvector s – a row vector.

-- $\nabla_{st} q_j$ denotes the matrix whose typical element is $\partial^2 q_j / \partial s_r \partial t_c$ -- its row dimension corresponds to that of its first subscript and the column dimension to that of its second subscript.

From Newey and McFadden (1994) or White (1994) we have that the correct asymptotic covariance matrix of $\hat{\tau}$ is

$$\begin{aligned}\text{AVAR}(\hat{\tau}) = \text{E}\left[\nabla_{\tau\tau} q_2\right]^{-1} \Big[& \text{E}\left[\nabla_{\tau\delta} q_2\right] \text{AVAR}(\hat{\delta}) \text{E}\left[\nabla_{\tau\delta} q_2\right]' \\ & - \text{E}\left[\nabla_\tau q_2' \nabla_\delta q_1\right] \text{E}\left[\nabla_{\delta\delta} q_1\right]^{-1} \text{E}\left[\nabla_{\tau\delta} q_2\right]' \\ & - \text{E}\left[\nabla_{\tau\delta} q_2\right] \text{E}\left[\nabla_{\delta\delta} q_1\right]^{-1} \text{E}\left[\nabla_\tau q_2' \nabla_\delta q_1\right]' \Big] \text{E}\left[\nabla_{\tau\tau} q_2\right]^{-1} \\ & + \text{E}\left[\nabla_{\tau\tau} q_2\right]^{-1} \text{E}\left[\nabla_\tau q_2' \nabla_\tau q_2\right] \text{E}\left[\nabla_{\tau\tau} q_2\right]^{-1} \qquad \text{(A-3)}\end{aligned}$$

where $\text{AVAR}(\hat{\delta})$ denotes the asymptotic covariance matrix of $\hat{\delta}$.

In the present context, $[\hat{\beta}' \quad \hat{\gamma}]$ can be cast as a 2SOE with β and γ playing the roles of δ and τ, respectively. Here we have $q_1(\beta, X)$ as the relevant objective function for first-stage estimation of the deep parameters β, and

$$q_2(\beta, \gamma, X_i) = -(g(\beta, X) - \gamma)^2 . \qquad \text{(A-4)}$$

as the relevant second-stage objective function for the estimation of γ. Let's first examine the term $\text{E}\left[\nabla_\gamma q_2' \nabla_\delta q_1\right]$ in (A-5) whose essential component is

$$\nabla_\gamma q_2 = 2(g(\beta, X) - \gamma) \tag{A-5}$$

so

$$E\left[\nabla_\gamma q_2' \nabla_\beta q_1\right] = 2E\left[(g(\beta, X) - \gamma)\, E[\nabla_\beta q_1 \mid X]\right].$$

Typically

$$E[\nabla_\beta q_1 \mid X] = 0\,. \tag{A-6}$$

For example, when $\hat{\beta}$ is obtained via the nonlinear least squares (NLS) method we have

$$q_1(\beta, V_1) = -(Y - J(\beta, X))^2$$

where $J(\beta, X)$ denotes the relevant nonlinear regression function, $V_1 = [Y \quad X]$

$$J(\beta_0, X) = E[Y \mid X] \tag{A-7}$$

and β_0 denotes the true value of the parameter vector. It follows that

$$\nabla_\beta q_1 = 2(Y - J(\beta, X))\nabla_\beta J$$

so

$$E[\nabla_\beta q_1 \mid X] = 2E[E[(Y - J(\beta, X)) \mid X]\nabla_\beta J] \tag{A-8}$$

because $\nabla_\beta J$ is a function only of X. From (A-7) we get

$$E[(Y - J(\beta, X)) \mid X] = 0$$

so (A-6) holds.

When $\hat{\beta}$ is a maximum likelihood estimator (MLE) we have that

$$q_1(\beta,\ V_1) = \ln f(Y \mid X;\ \beta) \tag{A-9}$$

where $\ln f(Y \mid X;\ \beta)$ denotes the true conditional density of Y given X. Now using (13.20) on p. 477 of Wooldridge (2010), we have that

$$E[\nabla_\beta q_1 \mid X] = E\left[\nabla_\beta \ln f(Y \mid X;\ \beta) \mid X\right] = 0\,.$$

Therefore, (A-6) holds. Given (A-6), we get

$$E\left[\nabla_\gamma q_2' \nabla_\delta q_1\right] = 0$$

so (A-3) becomes

$$\begin{aligned} \text{a var}(\hat{\gamma}) = {} & E\left[\nabla_{\gamma\gamma} q_2\right]^{-1} E\left[\nabla_{\gamma\beta} q_2\right] \text{AVAR}(\hat{\beta}) E\left[\nabla_{\gamma\beta} q_2\right]' E\left[\nabla_{\gamma\gamma} q_2\right]^{-1} \\ & + E\left[\nabla_{\gamma\gamma} q_2\right]^{-1} E\left[\nabla_\gamma q_2' \nabla_\gamma q_2\right] E\left[\nabla_{\gamma\gamma} q_2\right]^{-1}. \end{aligned} \tag{A-10}$$

From (A-5) we get

$$\nabla_{\gamma\gamma} q_2 = -2 \tag{A-11}$$

$$\nabla_{\gamma\beta} q_2 = 2\nabla_\beta g \tag{A-12}$$

and

$$\nabla_\gamma q_2 ' \nabla_\gamma q_2 = 4(g(\beta,\ X) - \gamma)^2. \tag{A-13}$$

so (A-10) can be rewritten as

$$a\,\mathrm{var}(\hat{\gamma}) = E\left[\nabla_\beta g\right] AVAR(\hat{\beta}) E\left[\nabla_\beta g\right]' + E\left[(g(\beta, X) - \gamma)^2\right]. \qquad \text{(A-14)}$$

The consistent sample analog estimator of (A-14) is expression (6) in the text.

We note that the above discussion of NLS (MLE) estimation of $\hat{\beta}$ is predicated on the assumption that the relevant conditional mean regression model (conditional probability density function and, therefore, the conditional mean regression model) is correctly specified {for example, in the context of causal effect estimation and inference [e.g., as characterized by expressions (2) - (4) in the main text] we maintain that $m(\beta, X_p, X_o) = E[Y \mid X_p, X_o]$ is correctly specified}. In the vast majority of (all?) empirical analyses in health services research, a statistic like (1) [in the main text] is used for causal inference and, therefore, this "correct conditional mean regression specification (CCM)" assumption is necessary (though not sufficient).

In so-called pseudo (quasi) maximum likelihood (PMLE) specifications, it is assumed that CCM holds despite the possibility that the posited conditional probability density specification is incorrect (see Gourieroux, Monfort and Trognon, 1984; and Gourieroux and Monfort, 1989 [GM]; section 8.4.2). The PMLE estimator of β is consistent if and only if the specified conditional probability density function is a member of the linear exponential class (see Property 8.16 of GM). It can be shown that (A-6) holds for any $q_1(\beta, X)$ [the objective function for the first-stage estimate of β] based on a member of the linear exponential class (see the discussion on pp. 242-243 of GM). Therefore, (A-14) is the correct specification for the asymptotic variance of $\hat{\gamma}$ even when $\hat{\beta}$ is PMLE.

In the interest of completeness, we note that when CCM does not hold, we can see from (A-3) that the following term must be added to (A-14) to get the correct asymptotic variance of $\hat{\gamma}$

$$-\left\{E\left[(g-\gamma)\nabla_{\beta}q_{1}\right]E\left[\nabla_{\beta\beta}q_{1}\right]^{-1}E\left[\nabla_{\beta}g\right]' + E\left[\nabla_{\beta}g\right]E\left[\nabla_{\beta\beta}q_{1}\right]^{-1}E\left[(g-\gamma)\nabla_{\beta}q_{1}\right]'\right\}.$$

As we noted earlier, however, cases in which CCM does not hold are of limited (no?) analytic interest. In any event, regardless of whether or not CCM holds we have shown that the correct specification of the asymptotic variance of $\hat{\gamma}$ must include $E\left[(g(\beta, X)-\gamma)^{2}\right]$ as in (A-14).

References

Dowd, B.E., Greene, W.H., and Norton, E.C. (2014): "Computation of Standard Errors," *Health Services Research*, 49, 731-750.

Gourieroux, C. Monfort, A., and Trognon, A. (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681-700.

______________________, (1989): *Statistics and Econometric Models: Volume I*, Cambridge: Cambridge University Press.

Newey, W.K. and McFadden, D. (1994): Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Engle, R.F., and McFadden, D.L., Amsterdam: Elsevier Science B.V., 2111-2245, Chapter 36.

White, H. (1994): *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

Wooldridge, J.M. (2010): *Econometric Analysis of Cross Section and Panel Data*, 2nd Ed. Cambridge, MA: MIT Press.