Econometrica Supplementary Material

SUPPLEMENT TO "A COMMENT ON: *'On the Informativeness of Descriptive Statistics for Structural Estimates'*"

(*Econometrica*, Vol. 88, No. 6, November 2020, 2259–2264)

STÉPHANE BONHOMME
Department of Economics, University of Chicago

S1. BIAS CALCULATIONS

LET ME FIRST BRIEFLY RECALL the bias expressions in the paper. For simplicity, throughout I will fix a value η of the base model parameter. Let $f_{\eta,\zeta}$ be the density of the data D under the reader's model. Let f_η denote the density of D under the base model. Given a quantity of interest $c(\eta)$, the unrestricted bias of $\widehat{c}$ is

$$\sup\left\{\left|\mathbb{E}_{f_{\eta,\zeta}}(\widehat{c})-c(\eta)\right| : \zeta\in Z, 2\int \log\left(\frac{f_{\eta,\zeta}}{f_\eta}\right)f_{\eta,\zeta}\leq \mu^2\right\}, \tag{S1}$$

where I use twice the KL divergence for $r(\eta,\zeta)^2$.

Assume that $\mathbb{E}_{f_\eta}(\widehat{c})=c(\eta)$, and let $h_{\widehat{c}}$ denote the influence function of $\widehat{c}$ under the base model. The unrestricted bias can be expanded for small μ as

$$b_N=\mu\sqrt{\operatorname{Var}\left[h_{\widehat{c}}(D)\right]}+o(\mu),$$

where the variance is evaluated under the base model. To derive the restricted bias, one adds the constraint $\mathbb{E}_{f_{\eta,\zeta}}[h_{\widehat{\gamma}}(D)]=0$, where $h_{\widehat{\gamma}}$ denotes the influence function of $\widehat{\gamma}$, and obtains

$$b_{RN}=\mu\sqrt{\operatorname{Var}\left[\operatorname{res}\left(h_{\widehat{c}}(D),h_{\widehat{\gamma}}(D)\right)\right]}+o(\mu).$$

Let me now describe the approach that I have adopted in the discussion, which I have borrowed from Bonhomme and Weidner (2019). Let $\pi\in\Pi$ be a density. Let $f_{\eta,\pi}$ be the density of the data D under the reader's model. Let π_η denote the base value of π, and let $f_\eta=f_{\eta,\pi_\eta}$ denote the density of D under the base model. Given a quantity of interest $c(\eta,\pi)$, I define the unrestricted bias of $\widehat{c}$ as

$$\sup\left\{\left|\mathbb{E}_{f_{\eta,\pi}}(\widehat{c})-c(\eta,\pi)\right| : \pi\in\Pi, 2\int \log\left(\frac{\pi}{\pi_\eta}\right)\pi\leq \mu^2\right\}. \tag{S2}$$

There are two differences between (S1) and (S2). First, now the quantity of interest is $c(\eta,\pi)$. This allows for misspecification of the quantity of interest, even when η is known. Second, now the KL divergence is expressed in terms of the (infinite-dimensional) parameter π, not in terms of the density of the data. This allows one to cover settings, such as the second example, where π is not identified, while being able to add structure to the neighborhoods (e.g., independence assumptions) in a tractable way.

Stéphane Bonhomme: sbonhomme@uchicago.edu

 https://doi.org/10.3982/ECTA18789

Suppose that $\mathbb{E}_{f_\eta}(\widehat{c}) = c(\eta, \pi_\eta)$. Bonhomme and Weidner (2019) showed that the bias in (S2) can be expanded under suitable regularity conditions as

$$\mu\sqrt{\operatorname{Var}\bigl(\mathbb{E}_{f_\eta}\bigl[h_{\widehat{c}}(D)\nabla_\pi \log f_\eta(D)\bigr] - \nabla_\pi c(\eta, \pi_\eta)\bigr)} + o(\mu),$$

where ∇_π denote (Gâteaux) derivatives, and the variance is evaluated under the base model.

Consider the first example. In this case, $D = Y$, and π is the density of Y under the reader's model. Moreover, f_η is a normal density with mean m and variance σ^2. The quantity of interest is $c(\eta, \pi) = \mathbb{E}_\pi[\mathbf{1}\{Y \le a\}]$. Note that, since π is the density of the data, the difference in the quantity of interest is the only reason why (S1) and (S2) differ in this example. In this case, $\mathbb{E}_{f_\eta}[h_{\widehat{c}}(D)\nabla_\pi \log f_\eta(D)]$ can be represented by $h_{\widehat{c}}(Y)$. In addition, $\nabla_\pi c(\eta, \pi_\eta)$ can be represented by $\mathbf{1}\{Y \le a\} - \mathbb{E}(\mathbf{1}\{Y \le a\})$, where the expectation is evaluated under the base model. This gives the following bias expression:

$$b_N^{\text{mod}} = \mu\sqrt{\operatorname{Var}\bigl[h_{\widehat{c}}(Y) - \mathbf{1}\{Y \le a\}\bigr]} + o(\mu).$$

Consider the second example, where $D = (Y, X)$, and π is the density of (ε, X). Then f_η is the product of the conditional density of Y given X, which is a Bernoulli with probability $\Phi(X'\eta)$, and the density f_X of X, which I assume is not subject to misspecification. The quantity of interest is $c(\eta, \pi) = \mathbb{E}_\pi[\mathbf{1}\{\widetilde{x}'\eta \ge \varepsilon\}]$. In this case, $\mathbb{E}_{f_\eta}[h_{\widehat{c}}(D)\nabla_\pi \log f_\eta(D)]$ can be represented by $h_{\widehat{c}}(Y, X) - \mathbb{E}(h_{\widehat{c}}(Y, X) \mid X)$. In addition, $\nabla_\pi c(\eta, \pi_\eta)$ can be represented by $\mathbf{1}\{\widetilde{x}'\eta \ge \varepsilon\} - \mathbb{E}(\mathbf{1}\{\widetilde{x}'\eta \ge \varepsilon\} \mid X)$. This gives the following bias expression:[1]

$$b_N^{\text{mod}} = \mu\sqrt{\operatorname{Var}\bigl[h_{\widehat{c}}(Y, X) - \mathbf{1}\bigl\{\widetilde{x}'\eta \ge \varepsilon\bigr\}\bigr]} + o(\mu).$$

Continuing with the second example, and still focusing on the quantity $c(\eta, \pi)$, but now adding independence, π is the density of ε, independent of X. Then, for any function g, $\nabla_\pi \mathbb{E}_\pi[g(Y, X)]$ can be represented by $\mathbb{E}[g(Y, X) \mid \varepsilon] - \mathbb{E}[g(Y, X)]$. Hence the bias becomes

$$b_N^{\text{ind}} = \mu\sqrt{\operatorname{Var}\bigl[\mathbb{E}\bigl(h_{\widehat{c}}(Y, X) \mid \varepsilon\bigr) - \mathbf{1}\bigl\{\widetilde{x}'\eta \ge \varepsilon\bigr\}\bigr]} + o(\mu).$$

Last, the restricted bias analogs to b_N^{mod} and b_N^{ind} are obtained by imposing the constraint $\mathbb{E}_{f_{\eta,\pi}}[h_{\widehat{\gamma}}(Y)] = 0$ in the first example, and $\mathbb{E}_{f_{\eta,\pi}}[h_{\widehat{\gamma}}(Y, X)] = 0$ in the second example. The bias formulas in the main text, and the associated informativeness measures, follow.

S2. NUMERICAL APPROXIMATIONS

I draw S observations from the normal base model, and compute the moments using the simulated draws. I take $S = 500{,}000$ in the first example (to achieve numerical precision in Figure 1(c)), and $S = 20{,}000$ in the second example.

S3. $\Delta^{\text{mod}} = 0$ IN THE FIRST EXAMPLE

Suppose one wants to estimate $c(\pi) = \mathbb{E}_\pi(w(D))$, such as $\mathbb{E}_\pi(\mathbf{1}\{Y \le a\})$ in the first example. Let $\widehat{c} = c(\pi_{\widehat{\eta}})$, where $\widehat{\eta}$ is the maximum likelihood estimator of η under the

[1] I have used that $\mathbb{E}(h_{\widehat{c}}(Y, X) - \mathbf{1}\{\widetilde{x}'\eta \ge \varepsilon\} \mid X)$ is approximately constant in a local asymptotic.

base model. Starting from the identity $\mathbb{E}_{\pi_\eta}(w(D)) = c(\pi_\eta)$ and η-differentiating it (under sufficient regularity) gives

$$\mathbb{E}_{\pi_\eta}\big(w(D)\nabla_\eta \log \pi_\eta(D)\big) = \nabla_\eta c(\pi_\eta),$$

from which one can check that, to first order,

$$\mathbb{E}_{\pi_\eta}\big[\big(w(D) - \widehat{c}\big)\nabla_\eta \log \pi_\eta(D)\big] = 0.$$

From this, it follows that $\Delta^{\text{mod}} = 0$ when using $\widehat{\gamma} = \widehat{\eta}$ as a vector of descriptive statistics, whereas $\Delta = 1$ since $\widehat{c}$ is a non-stochastic function of $\widehat{\eta}$.

REFERENCES

BONHOMME, S., AND M. WEIDNER (2019): "Minimizing Sensitivity to Model Misspecification," Preprint, arXiv:1807.02161. [1,2]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 2 July, 2020; final version accepted 3 August, 2020; available online 12 August, 2020.