Econometrica, Vol. 88, No. 6 (November, 2020), 2265–2269

A COMMENT ON: *"On the Informativeness of Descriptive Statistics for Structural Estimates" by Isaiah Andrews, Matthew Gentzkow, and Jesse M. Shapiro*

YUICHI KITAMURA
Department of Economics, Cowles Foundation, Yale University

1. ANDREWS, GENTZKOW, AND SHAPIRO'S Δ-MEASURE

IN THEIR EXCELLENT AND SIMULATING PAPER, Andrews, Gentzkow, and Shapiro (2020; hereafter AGS20) propose a new measure of informativeness of descriptive statistics. Suppose the researcher considers a structural model parameterized by η. The parameter of interest c, which can be a counterfactual value, is a function of η. The researcher estimates c by an estimator $\hat{c}$, and also reports a descriptive statistic $\hat{\gamma}$. To be more precise, for a value η_0, consider a sequence $\eta_n(h) := \eta_0 + \frac{1}{\sqrt{n}}h$, which yields an expansion

$$c\big(\eta_n(h)\big) = c(\eta_0) + \frac{1}{\sqrt{n}}c^\star(h) + o\left(\frac{1}{\sqrt{n}}\right). \tag{1.1}$$

The same setting also applies to γ, leading to a similar expansion. The values $c(\eta_n(h))$ and $\gamma(\eta_n(h))$ are the "estimand" within the paper's local asymptotic framework, that is, the parameter values to be learned by $\hat{c}$ and $\hat{\gamma}$. Under AGS20's Assumption 2, we have

$$\sqrt{n}\begin{pmatrix} \hat{c} - c\big(\eta_n(h)\big) \\ \hat{\gamma} - \gamma\big(\eta_n(h)\big) \end{pmatrix} \to_d \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \Sigma_{c\gamma} \\ \Sigma_{\gamma c} & \Sigma_{\gamma\gamma} \end{pmatrix}\right).$$

This means that the estimator $\hat{c}$ is asymptotically unbiased for the parameter $c(\eta_n(h))$, up to the order of $1/\sqrt{n}$ when the model is correctly specified. In view of (1.1), for all purposes it is more convenient to consider $c^\star(h)$ as the parameter of interest, as the authors suggest.

Having set up the local asymptotic framework as above, the authors proceed to introduce misspecification into it. This, of course, is the main theme of the paper. It considers (sequences of) misspecified probability distributions such that $\hat{c}$ tends to $c(\eta_0) + \frac{1}{\sqrt{n}}\bar{c}$ up to the order of $\frac{1}{\sqrt{n}}$. The difference between $\bar{c}$ and $c^\star$ is then the misspecification bias.

The key elements of the paper are two sets of sequences of CDFs: for each fixed value $\underline{c}^\star \in \mathbf{R}$ it defines $\mathcal{S}^{RN}(\underline{c}^\star)$, which is the set of infinite sequences of probability distributions such that its each element $\tilde{S}$ has another sequence S satisfying the following conditions (see AGS20 for precise statements, in slightly different notation):

- (i-a) S is consistent with the structural model (i.e., the model is correct under S),
- (i-b) the value of $c^\star$ corresponding to S is equal to $\underline{c}^\star$,
- (ii-a) $\tilde{S}$ and S are close to each other,
- (ii-b) the limiting values of $\hat{\gamma}$ under S and $\tilde{S}$ are equal (up to the order of $1/\sqrt{n}$).

Yuichi Kitamura: yuichi.kitamura@yale.edu

 https://doi.org/10.3982/ECTA18788

Dropping the requirement (ii-b) defines $\mathcal{S}^N(\underline{c}^\star)$. The asymptotic bias of $\hat{c}$ over $\bigcup_{\underline{c}^\star}\mathcal{S}^{RN}(\underline{c}^\star)$ is naturally no greater than that with $\bigcup_{\underline{c}^\star}\mathcal{S}^N(\underline{c}^\star)$. AGS20 advocate the use of

$$\Delta = \frac{\Sigma_{c\gamma}\Sigma_{\gamma\gamma}^{-1}\Sigma_{\gamma c}}{\sigma_c^2},$$

which is the ratio of these two maximal asymptotic biases of $\hat{c}$. The measure Δ gives the degree to which the extra restriction (ii-b) limits the misspecification bias of $\hat{c}$. [An aside: From the perspective of robust statistics, it seems natural to regard S as an "ideal" CDF, or, more precisely, an ideal sequence of CDFs. The researcher does not have access to data from S, however, and instead observes data generated from $\tilde{S}$, which is subject to measurement error or other causes of data contamination. Such a stand is, for example, taken by Kitamura, Otsu, and Evdokimov (2013). A possible advantage of this interpretation is that it is easy to define what the true parameter value is.]

2. DEFINING INFORMATIVENESS: ROBUSTNESS, MISSPECIFICATION, AND EFFICIENCY

I find the measure Δ very neat, and it is practical. Even without referencing to the theoretical derivation through misspecification analysis in AGS20, one can say that it has an intuitive form. To begin, it holds that $1/(1-\Delta) = \{\Sigma^{-1}\}_{11}\{\Sigma\}_{11}$, where $\{\cdot\}_{ij}$ denotes the (i, j) element of a matrix. Or, alternatively, if there exists a vector of estimators $(\tilde{c}, \tilde{\gamma}')'$ such that they are $\mathrm{N}(0, \Sigma)$ in finite samples, then $\Delta = [(\text{variance of } \tilde{c}) - (\text{variance of } \tilde{c} \text{ conditional on } \tilde{\gamma})]/(\text{variance of } \tilde{c})$. The latter fact in particular seems to suggest a potential connection between AGS20's Δ and estimation efficiency. In what follows, we explore such relationships.

As done in AGS20, the first step is to introduce a restriction on the set of permissible data distributions implied by the descriptive statistic $\hat{\gamma}$. Instead of considering three sequences of distributions $\tilde{S} \in \mathcal{S}^N(\underline{c}^\star)$, $\tilde{S}^{\mathrm{r}} \in \mathcal{S}^{RN}(\underline{c}^\star)$, and S, we use the empirical distribution of data $\hat{F}_n$, its "restricted version" $\hat{F}_n^{\mathrm{r}}$ (to be defined shortly), and the data generating distribution F_0 that is consistent with the structural model. The three distributions stay together within a $\sqrt{n}$ neighborhood by the functional CLT, effectively satisfying the restriction (ii-a). Misspecification plays no role here, however: the researcher observes i.i.d. data drawn from F_0.

Next, I reformulate $\hat{c}$ as a statistical functional in von Mises' sense, as in, for example, Fernholz (2012). To this end, let us introduce a functional $C : \mathcal{F} \to \mathbf{R}$, where the symbol $\mathcal{F}$ denotes the set of distribution functions, so that it is equal to the estimator $\hat{c}$ when evaluated at the empirical distribution $\hat{F}_n$, so that $\hat{c} = C(\hat{F}_n)$. Moreover, evaluated at the true DGP F_0, it provides the true value of the parameter of interest, thus $c_0 = C(F_0)$. Many estimators can be interpreted as statistical functionals of the empirical distribution; see, for example, Ichimura and Newey (2015). Assume that $C(\cdot)$ is Hadamard differentiable at F_0 with respect to $\|\cdot\|_\infty$:

$$\frac{\left|C(F^t) - C(F_0) - \dot{C}(F_0, F^t - F_0)\right|}{t} \to 0 \quad \text{for all } \{F^t\}$$

such that

$$\left\|t^{-1}(F^t - F_0) - H\right\|_\infty \to 0 \quad \text{for some } H,$$

where $\dot{C}(\cdot,\cdot)$ is continuous and linear in the second argument. Letting $\phi_c(D) := \dot{C}(F_0, \mathbf{1}\{D \le \cdot\} - F_0)$ denote the influence function in accordance to the asymptotic linear representation in Assumption 1 of AGS20, a simple application of the Functional Delta Method yields

$$\sqrt{n}(\hat{c} - c_0) = \sqrt{n}\big[C(\hat{F}_n) - C(F_0)\big] \to_d \mathrm{N}\big(0, \sigma_c^2\big),$$

where $\sigma_c^2 = \int \phi_c^2 \, dF_0$. This is what Huber (2004, p. 37) called "a one line asymptotic normality proof."

We now seek for a similar asymptotic normality result, but this time with an additional constraint analogous to (ii-b) used in the bias analysis in AGS20. We will introduce a constrained empirical CDF $\hat{F}_n^{\mathrm{r}}$ and consider a plug-in estimator $\hat{c}^{\mathrm{r}} = C(\hat{F}_n^{\mathrm{r}})$. To this end, once again I use statistical functional notation and write

$$\hat{\gamma} = \Gamma(\hat{F}_n), \qquad \gamma_0 = \Gamma(F_0).$$

To develop a parallelism with AGS20, we use an appropriate analogue of (ii-b) to define a set of restricted CDFs:

$$\mathcal{F}^{\mathrm{r}} := \big\{F \in \mathcal{F} : \Gamma(F) = \Gamma(F_0)\big\}. \tag{2.1}$$

To simplify our argument, assume further that Γ is a functional such that $\hat{\gamma} = \Gamma(\hat{F}_n) = \gamma_0 + A(\hat{F}_n) \int \phi_\gamma \, d\hat{F}_n$, where ϕ_γ corresponds to the influence function in Assumption 1 of AGS20 and $A(\cdot)$ is a square matrix. This does not seem too restrictive in our context, as descriptive statistics such as sample means, OLS, and just identified linear IV estimators belong to this category. We next obtain an appropriate "restricted version" of $\hat{F}_n$ by introducing a divergence measure between CDFs. This is again consistent with the use of the Cressie–Read family in Section 4.3 in AGS20. To use a concrete example, we work on the Kullback–Leibler divergence $K(F, G) = \int \log \frac{f}{g} \, dF$ for CDFs F and G with density functions f and g. Other divergence measures, such as other members of the Cressie–Read family, or even a larger class of divergence family, can be used; see Imbens, Spady, and Johnson (1998), Kitamura (2007), and Newey and Smith (2004) for related discussions. In any event, consider the minimizer of the following problem:

$$\text{minimize } K(\hat{F}, F) \quad \text{subject to } F \in \mathcal{F}^{\mathrm{r}}.$$

Assuming that $A(F_0)$ is nonsingular, under weak restrictions, the convexity of the set $\{F \in \mathcal{F} : \int \phi_\gamma \, dF = \gamma_0\}$ guarantees a unique solution given by

$$\hat{F}_n^{\mathrm{r}}(z) = \sum_{i=1}^n \mathbf{1}\{D_i \le z\} \hat{p}_i, \qquad \hat{p}_i = \frac{\exp\big(\theta_n' \phi_\gamma(D_i)\big)}{\sum_{i=1}^n \exp\big(\theta_n' \phi_\gamma(D_i)\big)},$$

$$\theta_n = \operatorname*{argmin}_\theta \sum_{i=1}^n \exp\big(\theta' \phi_\gamma(D_i)\big),$$

for large n. This corresponds to exponential tilting (see, e.g., Kitamura and Stutzer (1997)). Moreover, the empirical process $\sqrt{n}(\hat{F}^{\mathrm{r}} - F_0)$ converges to a Gaussian process: $\sqrt{n}(\hat{F}^{\mathrm{r}} - F_0) \rightsquigarrow \mathbb{G}$, where $\mathbb{G}$ has the covariance kernel of the form $F_0(t \wedge s) - F_0(t)F_0(s) -$

$\mathrm{Cov}(\mathbf{1}\{D \le t\}, \phi_\gamma)\,\mathrm{Var}(\phi_\gamma)^{-1}\,\mathrm{Cov}(\phi_\gamma, \mathbf{1}\{D \le s\})$. See Sheehy (1988). Let $\hat{c}^{\mathrm{r}} := C(\hat{F}^{\mathrm{r}}_n)$, which incorporates the restriction implied by the descriptive statistic $\hat{\gamma}$ on the CDF to be plugged in for the statistical functional $C(\cdot)$. With the Hadamard differentiability assumption on the functional $C(\cdot)$, we have, once again by the Functional Delta Method (see, in particular, Van der Vaart (2000, Theorem 20.8)),

$$\begin{aligned}
\sqrt{n}\big(\hat{c}^{\mathrm{r}} - c_0\big) &= \sqrt{n}\big[C\big(\hat{F}^{\mathrm{r}}_n\big) - C(F_0)\big] \\
&= \sqrt{n}\dot{C}\big(F_0, \hat{F}^{\mathrm{r}}_n - F_0\big) + o_p(1) \\
&= \sqrt{n}\dot{C}\left(F_0, \sum_{i=1}^{n} \hat{p}_i \mathbf{1}\{D_i \le \cdot\} - F_0\right) + o_p(1) \\
&= \sqrt{n}\sum_{i=1}^{n} \hat{p}_i \dot{C}\big(F_0, \mathbf{1}\{D_i \le \cdot\} - F_0\big) + o_p(1) \\
&= \sqrt{n}\sum_{i=1}^{n} \hat{p}_i \phi_c(D_i) + o_p(1) \\
&\to_d \mathrm{N}\big(0, \mathrm{Var}(\phi_c) - \mathrm{Cov}(\phi_c, \phi_\gamma)\,\mathrm{Var}(\phi_\gamma)^{-1}\,\mathrm{Cov}(\phi_\gamma, \phi_c)\big).
\end{aligned}$$

Let $s^2_{\hat{c}^{\mathrm{r}}} := \mathrm{Var}(\phi_c) - \mathrm{Cov}(\phi_c, \phi_\gamma)\,\mathrm{Var}(\phi_\gamma)^{-1}\,\mathrm{Cov}(\phi_\gamma, \phi_c)$; then, denoting the asymptotic variance of $\hat{c}$ by $s^2_{\hat{c}} = \sigma^2_c$, we obtain

$$\Delta = \frac{s^2_{\hat{c}} - s^2_{\hat{c}^{\mathrm{r}}}}{s^2_{\hat{c}}}.$$

This provides an alternative expression of the measure Δ as the relative asymptotic efficiency gain for the differentiable statistical functional C from imposing the invariance restriction with respect to Γ in (2.1). This is not a coincidence; local robustness and (local) asymptotic efficiency are naturally intertwined concepts. See, for example, Rieder (2014) for an insightful discussion. In the context of GMM, Honoré, Jørgensen, and de Paula (2020) pointed out that a sensitivity statistic proposed by Andrews, Gentzkow, and Shapiro (2017)—a companion paper of AGS20—is related to measuring how each moment condition contributes to estimation precision in GMM. The latter, in turn, is connected to an earlier literature on informativeness/redundancy of moment conditions (see, e.g., Breusch, Qian, Schmidt, and Wyhowski (1999)). The approach here is somewhat different, as I make explicit use of the restricted space of CDFs $\mathcal{F}^{\mathrm{r}}$ to explore its impact on the asymptotic variance of the statistical functional C, paralleling AGS20's analysis of asymptotic biases through restrictions on the space of CDFs.

Last, the result I outlined above that uses $\hat{F}^{\mathrm{r}}_n$ in the possibly nonlinear statistical functional $C(\cdot)$ is closely related to results concerning estimation of moments in the presence of other moment restrictions via "re-weighting": see, for example, Back and Brown (1993), Brown and Newey (2002), and Hellerstein and Imbens (1999). Note that for these problems, the estimand can be viewed an exactly linear functional of the underlying distribution. Further developments in von Mises calculus with side constraints would be of great interest, if one wishes to generalize the type of argument presented here in various directions.

With all that said, I strongly believe the issue of robustness should be a central concern for modern econometricians, and I find AGS20's local misspecification analysis illuminating, compelling, and useful. Their statistic Δ is a practical and valuable measure; it will be a useful tool for applied economists for years to come.

REFERENCES

ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132 (4), 1553–1592. [2268]

BACK, K., AND D. P. BROWN (1993): "Implied Probabilities in GMM Estimators," *Econometrica*, 61 (4), 971–975. [2268]

BREUSCH, T., H. QIAN, P. SCHMIDT, AND D. WYHOWSKI (1999): "Redundancy of Moment Conditions," *Journal of econometrics*, 91 (1), 89–111. [2268]

BROWN, B. W., AND W. K. NEWEY (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference," *Journal of Business & Economic Statistics*, 20 (4), 507–517. [2268]

FERNHOLZ, L. T. (2012): *Von Mises Calculus for Statistical Functionals*, Vol. 19. Springer Science & Business Media. [2266]

HELLERSTEIN, J. K., AND G. W. IMBENS (1999): "Imposing Moment Restrictions From Auxiliary Data by Weighting," *Review of Economics and Statistics*, 81 (1), 1–14. [2268]

HONORÉ, B. E., T. H. JØRGENSEN, AND Á. DE PAULA (2020): "The Informativeness of Estimation Moments," *Journal of Applied Econometrics* (forthcoming). [2268]

HUBER, P. J. (2004): *Robust Statistics*, Vol. 523. John Wiley & Sons. [2267]

ICHIMURA, H., AND W. K. NEWEY (2015): "The Influence Function of Semiparametric Estimators," arXiv preprint, arXiv:1508.01378. [2266]

IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66 (2), 333–357. [2267]

KITAMURA, Y. (2007): "Empirical Likelihood Methods in Econometrics: Theory and Practice," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge: Cambridge University Press. [2267]

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica: Journal of the Econometric Society*, 65 (4), 861–874. [2267]

KITAMURA, Y., T. OTSU, AND K. EVDOKIMOV (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81 (3), 1185–1201. [2266]

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72 (1), 219–255. [2267]

RIEDER, H. (2014): "Neighborhoods as Nuisance Parameters? Robustness vs. Semiparametrics," arXiv preprint, arXiv:1412.1711. [2268]

SHEEHY, A. (1988): "Kullback–Leibler Constrained Estimation of Probability Measures," Preprint, Dept. of Statist., Stanford University. [2268]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*, Vol. 3. Cambridge University Press. [2268]

Co-editor Ulrich K. Müller handled this manuscript.

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