

A COMMENT ON:

“On the Informativeness of Descriptive Statistics for Structural Estimates”
by Isaiah Andrews, Matthew Gentzkow, and Jesse M. Shapiro

STÉPHANE BONHOMME

Department of Economics, University of Chicago

THIS IS AN EXCELLENT PAPER on an important topic. In a first contribution (Andrews, Gentzkow, and Shapiro (2017)), the authors developed measures of sensitivity of estimated parameters to estimation moments. In this new paper, they propose a measure of informativeness. These two papers provide new tools to facilitate the interpretation of empirical research.

Suppose the researcher has estimated a structural model and produced an estimate \hat{c} of a target parameter c . In addition, suppose she has computed a vector $\hat{\gamma}$ of descriptive statistics, such as some moments of the data. The informativeness measure Δ that the authors propose to report is equal to the share of the asymptotic variance of \hat{c} that is explained by $\hat{\gamma}$. In large samples, this quantity can be estimated using least squares, by regressing the influence function of \hat{c} on that of $\hat{\gamma}$, and computing the R^2 coefficient in the regression. The formula is a simple function of the elements of the asymptotic variance matrix of $(\hat{c}, \hat{\gamma})$. To provide a framework to interpret Δ , the authors allow for the possibility that the researcher’s model is misspecified. They rely on a local asymptotic approach where the degree of misspecification is proportional to sampling variability. Their examples nicely illustrate how the values that Δ take relate to the empirical context.

In my discussion, I will focus on the interpretation of Δ . I will make two points, using simple examples.

Misspecification. Let me start with a simple illustrative example, which will allow me to discuss the role of misspecification in the analysis of informativeness. Suppose the researcher is interested in estimating the cumulative distribution function $\Pr(Y \leq a)$ of a scalar random variable Y at a point a . She wishes to measure the informativeness of two descriptive statistics: the sample mean \hat{m} of Y and its sample variance $\hat{\sigma}^2$. To answer this question using the approach of the paper, one needs to specify a base model, and an estimator \hat{c} of the quantity of interest that is asymptotically unbiased under that model. Let me choose a particular base model, $Y \sim \mathcal{N}(m, \sigma^2)$. In addition, let me choose a particular estimator, $\hat{c} = \Phi(\frac{a-\hat{m}}{\hat{\sigma}})$, for Φ the standard normal c.d.f. All the calculations in this example will rely on this base model and estimator. Let me consider as descriptive statistics the vector $\hat{\gamma} = (\hat{m}, \hat{\sigma}^2)'$. In this example, it is easy to see that $\Delta = 1$ at all points a , since \hat{c} is a non-stochastic function of $\hat{\gamma}$. I plot Δ as a function of a in the upper-left graph in Figure 1.

To understand why $\Delta = 1$ in this case, it is useful to recall that $\sqrt{1 - \Delta}$ is the ratio of two biases. The unrestricted bias is the supremum, in a local neighborhood of the normal base model, of the difference $|\mathbb{E}(\hat{c}) - c|$, where c is the quantity of interest. Let me assume, and this is an important assumption that I will discuss below, that the quantity of interest is $c(m, \sigma^2) = \Phi(\frac{a-m}{\sigma})$, where m and σ^2 are the mean and variance of Y under the base

Stéphane Bonhomme: sbonhomme@uchicago.edu

I thank Martin Weidner and the authors for comments.

model. In a Kullback–Leibler (KL) neighborhood of radius μ , the unrestricted bias can then be approximated for small μ by¹

$$\mu\sqrt{\text{Var}[h_{\hat{c}}(Y)]}, \tag{1}$$

where $h_{\hat{c}}(Y)$ denotes the influence function of \hat{c} , and the variance is computed under the base model. In turn, the restricted bias is equal to the worst-case bias of \hat{c} in a smaller neighborhood that consists of distributions that are local to the base model, and in addition are such that $\hat{\gamma}$ is asymptotically unbiased for $(m, \sigma^2)'$. That is, under any distribution in this neighborhood, the sample mean and variance are asymptotically unbiased for these parameters. Hence, in the restricted neighborhood, we have $c = \Phi(\frac{a-m}{\sigma})$ and $\gamma = (m, \sigma^2)'$, similarly as under the base model. The restricted bias is

$$\mu\sqrt{\text{Var}[\text{res}(h_{\hat{c}}(Y), h_{\hat{\gamma}}(Y))]}, \tag{2}$$

where $h_{\hat{\gamma}}(Y)$ denotes the influence function of $\hat{\gamma}$, and $\text{res}(h_{\hat{c}}(Y), h_{\hat{\gamma}}(Y))$ is the least-squares residual in the population regression of $h_{\hat{c}}(Y)$ on $h_{\hat{\gamma}}(Y)$ and a constant. In the example, $h_{\hat{c}}(Y)$ is a linear combination of the two components of $h_{\hat{\gamma}}(Y)$. As a result, the restricted bias is equal to zero, so $\sqrt{1 - \Delta} = 0$ and $\Delta = 1$.

So far, I have defined the quantity of interest as $c(m, \sigma^2) = \Phi(\frac{a-m}{\sigma})$. Yet, this may not be the most natural target quantity if one worries that Y is not normally distributed. Alternatively, consider the quantity $c(\pi) = \mathbb{E}_{\pi}[\mathbf{1}\{Y \leq a\}]$, where π denotes the density of Y . This is a functional of the density of the data, and I am now going to assume that the researcher and the reader are interested in this quantity. When Y is not normally distributed, $|c(\pi) - \Phi(\frac{a-m}{\sigma})|$ will be non-zero in general. As a result, even when m and σ^2 are known, $c(\pi)$ is subject to model misspecification. Let me now describe a measure of informativeness that is motivated by a framework where $c(\pi)$ is the quantity of interest. To do so, I follow the approach of [Bonhomme and Weidner \(2019\)](#), and I detail the bias calculations in the Supplemental Material ([Bonhomme \(2020\)](#)). In a KL neighborhood of the normal base model of radius μ , the worst-case bias of \hat{c} as an estimate of $c(\pi)$ is

$$\mu\sqrt{\text{Var}[h_{\hat{c}}(Y) - \mathbf{1}\{Y \leq a\}]}. \tag{3}$$

Note that, in contrast with (1), in (3) the bias formula depends on the indicator function $\mathbf{1}\{Y \leq a\}$, whose presence arises due to the change in the quantity of interest. For the restricted bias, I compute the worst-case bias of \hat{c} as an estimate of $c(\pi)$, subject to the restriction that $\hat{\gamma}$ is asymptotically unbiased for $(m, \sigma^2)'$. I obtain

$$\mu\sqrt{\text{Var}[\text{res}(h_{\hat{c}}(Y) - \mathbf{1}\{Y \leq a\}, h_{\hat{\gamma}}(Y))]} \tag{4}$$

I then define the following modified measure of informativeness:

$$\Delta^{\text{mod}} = 1 - \frac{\text{Var}[\text{res}(h_{\hat{c}}(Y) - \mathbf{1}\{Y \leq a\}, h_{\hat{\gamma}}(Y))]}{\text{Var}[h_{\hat{c}}(Y) - \mathbf{1}\{Y \leq a\}]}, \tag{5}$$

¹In this discussion, I focus on (twice) KL, and I abstract from smaller-order terms in the bias expressions.

Mean and variance

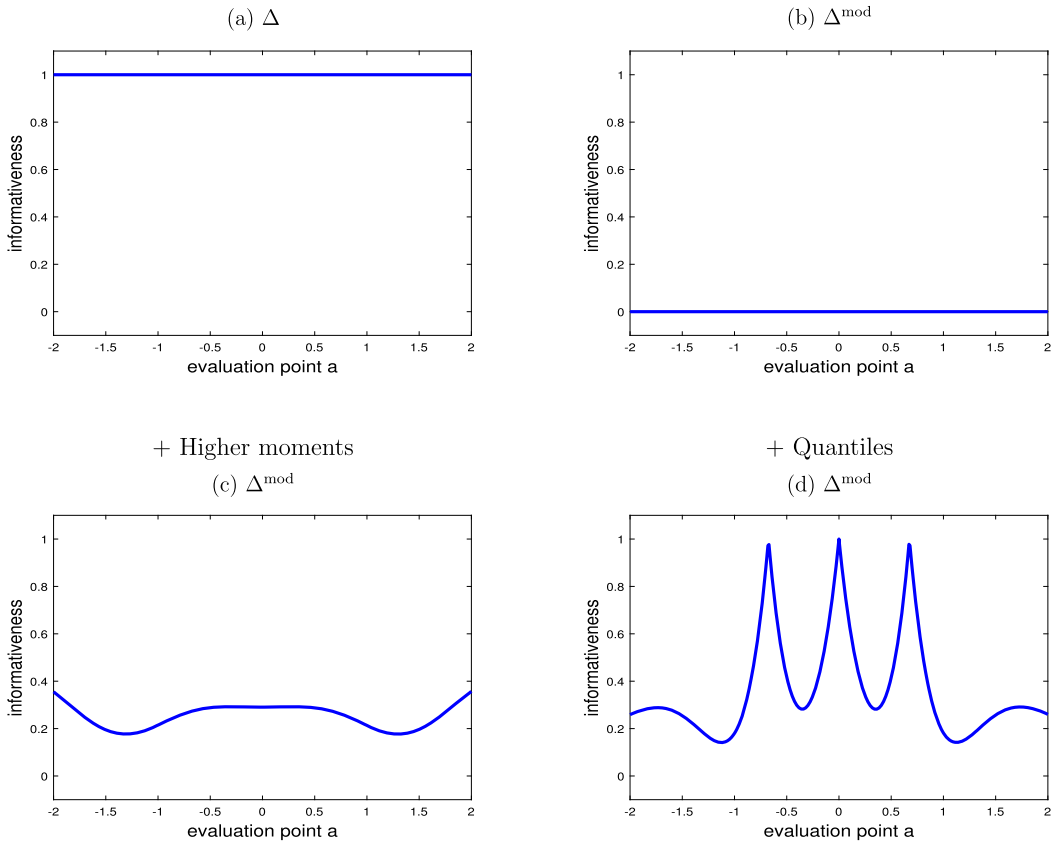


FIGURE 1.—Informativeness of descriptive statistics for a probability estimate. *Notes:* Informativeness of various descriptive statistics $\hat{\gamma}$ for $\hat{c} = \Phi(\frac{a-\hat{m}}{\hat{\sigma}})$ in the first example. Graph (a) shows the original measure Δ , using mean and variance as $\hat{\gamma}$. Graphs (b) to (d) show the modified measure Δ^{mod} when taking $\hat{\gamma}$ to be mean and variance, the first four moments, and mean, variance, and three quantiles, respectively. The evaluation point a is reported on the x -axis.

where $\sqrt{1 - \Delta^{\text{mod}}}$ quantifies the reduction in the bias of \hat{c} as an estimate of $c(\pi)$, when one imposes that $\hat{\gamma}$ is asymptotically unbiased for $(m, \sigma^2)'$.

To illustrate, I compute Δ^{mod} for different values of a , using parameter values $m = 0$ and $\sigma^2 = 1$. For computation, I rely on simulations (see the Supplemental Material). In this simple nonlinear example, I obtain very different results when comparing Δ^{mod} to Δ . While the latter is equal to 1, the former is equal to zero irrespective of a ; see the upper-right graph in Figure 1. It is intuitive that Δ^{mod} is low in this case. Indeed, $c(\pi) = \mathbb{E}_{\pi}[\mathbf{1}\{Y \leq a\}]$ does not directly depend on m or σ^2 . In addition, m and σ^2 are estimated under the normal base model. As a result, in this example, knowing that \hat{m} and $\hat{\sigma}^2$ estimate those parameters correctly has no effect on the bias of \hat{c} as an estimate of $c(\pi)$. In the Supplemental Material, I provide a formal argument, which does not rely on normality, to show that $\Delta^{\text{mod}} = 0$. This contrasts with the previous setup, where knowing m and σ^2 provided all the information needed to know $c(m, \sigma^2)$. I then compute Δ^{mod} for other vectors of descriptive statistics. In the lower-left graph in Figure 1, I show the results when adding third- and fourth-order sample moments of Y to $\hat{\gamma}$. The results show that Δ^{mod} is not zero

in this case. This is intuitive since knowing that π has zero skewness, say, restricts the type of deviations from the normal distribution that are taken into account in the bias calculation. In the lower-right graph in Figure 1, I show the results when using mean, variance, and three quantiles of Y (25%, 50%, and 75%) in $\hat{\gamma}$. The results show that $\Delta^{\text{mod}} = 1$ at the three quantiles, and that $\Delta^{\text{mod}} < 1$ outside. By contrast, when including other descriptive statistics in $\hat{\gamma}$ in addition to mean and variance, the original informativeness measure remains $\Delta = 1$.

In this simple example, I have attempted to measure informativeness in a setting where the researcher is interested in a particular quantity $c(\pi)$ that may be misspecified even when (m, σ^2) are known. Of course, the fact that $\Delta = 1$ in the example is an artefact of the choice of base model and estimator, and different choices would give different Δ values. Yet the example shows that, for the same base model and estimator, the modified measure Δ^{mod} may differ substantially from the one proposed in the paper.

Structure. I now want to focus on another example, which is still simple but is closer in spirit to the structural applications that the authors use as motivation. This will allow me to discuss the role of neighborhoods and model assumptions in the construction of the informativeness measure. Consider the following binary choice model:

$$Y = \mathbf{1}\{X'\eta \geq \varepsilon\}, \quad X \in \{x_1, \dots, x_K\}.$$

This model is related to the regression example in the paper, with the difference that here the outcome variable is binary. The researcher’s goal is to estimate the probability of $Y = 1$ occurring if X were exogenously set to some value \tilde{x} . She postulates the base model $(\varepsilon | X) \sim \mathcal{N}(0, 1)$, and uses Probit for estimation.² As in the first example, the definition of informativeness depends on what one considers to be the quantity of interest. For example, one could focus on $c(\eta) = \Phi(\tilde{x}'\eta)$: however, this would ignore that, even if η were known, the misspecification of the normal base model might still affect the quantity of interest.

Consider instead the alternative quantity $c(\eta, \pi) = \mathbb{E}_\pi[\mathbf{1}\{\tilde{x}'\eta \geq \varepsilon\}]$, where π denotes the density of unobservables ε . This is a functional of the parameter and the density of unobservables, and I will assume the researcher and the reader are interested in it. Note that here $c(\eta, \pi)$ depends on η . It is natural to expect that when $c(\eta, \pi)$ is the quantity of interest, $\hat{\eta}$ will be informative about \hat{c} , albeit not perfectly. More generally, in a structural model, $c(\eta, \pi)$ may depend on both structural parameters η and distributions π of unobservables. In the present case, computing worst-case biases in neighborhoods of densities π around the base density gives the following value for the modified informativeness measure (see the Supplemental Material):

$$\Delta^{\text{mod}} = 1 - \frac{\text{Var}[\text{res}(h_{\hat{c}}(Y, X) - \mathbf{1}\{\tilde{x}'\eta \geq \varepsilon\}, h_{\hat{\gamma}}(Y, X))]}{\text{Var}[h_{\hat{c}}(Y, X) - \mathbf{1}\{\tilde{x}'\eta \geq \varepsilon\}]}$$

I compute Δ and Δ^{mod} in a numerical experiment where $\tilde{x} = 4$, X takes values 1, 2, 3 with equal probability, and $\eta = (-1, \frac{1}{2})'$ where the first coefficient corresponds to the intercept and the second one to the coefficient of X . I contrast two descriptive statistics: the sample mean \hat{p} of Y when $X = 1$, and the Probit maximum likelihood estimate $\hat{\eta}$. I again resort to simulations to compute Δ^{mod} , and I report the results in Table I. For

²That is, $\hat{c} = \Phi(\tilde{x}'\hat{\eta})$, where $\hat{\eta} = \text{argmax}_\eta \sum_{i=1}^n Y_i \log[\Phi(X'_i\eta)] + (1 - Y_i) \log[1 - \Phi(X'_i\eta)]$.

TABLE I
INFORMATIVENESS OF DESCRIPTIVE
STATISTICS IN THE BINARY CHOICE EXAMPLE^a

Desc. Stat.	Δ	Δ^{mod}	Δ^{ind}	Γ^{ind}
\widehat{p}	0.196	0.286	0.028	–
$\widehat{\eta}$	1	0.839	0.385	–
None	–	–	–	0.795

^aInformativeness of various descriptive statistics $\widehat{\gamma}$ for $\widehat{c} = \Phi(\widetilde{x}'\widehat{\eta})$ in the second example. Δ is the original measure of informativeness, Δ^{mod} is the modified measure, Δ^{ind} is the modified measure under independence between ε and X , and Γ^{ind} measures the informativeness of the independence assumption for \widehat{c} .

the first descriptive statistic, I find $\Delta = 0.19$ and $\Delta^{\text{mod}} = 0.28$. Both measures suggest that the mean outcome at $X = 1$ provides limited information about the potential value of the outcome if X was exogenously set to 4. For the second descriptive statistic, I find $\Delta = 1$, which reflects the fact that \widehat{c} is a non-stochastic function of $\widehat{\eta}$. In addition, I find $\Delta^{\text{mod}} = 0.83$. In this case, and in contrast with the first example, Δ and Δ^{mod} do not behave very differently.

I want to use this example to discuss the role of neighborhoods. Indeed, in the spirit of the paper, until now I have considered all possible conditional distributions of $(\varepsilon | X)$ local to the base model when computing worst-case biases. In particular, this allows for unrestricted (local) endogeneity of X . However, in many cases, the researcher may be willing to make certain assumptions about the distributions of unobservables and observables, even though she thinks the base model is probably wrong. To illustrate, suppose the researcher is willing to assume that X has been assigned exogenously, independently of ε . How does this assumption modify her analysis of informativeness? This example is relevant to situations where the researcher uses experimental data, or empirical designs such as difference-in-differences or instrumental variable methods that rely on substantive assumptions on the data generating process. A case in point is [Attanasio, Meghir, and Santiago \(2012\)](#), which the authors use as an example. Under independence between ε and X , computing worst-case biases of \widehat{c} as an estimate of $c(\eta, \pi)$ in neighborhoods of densities π around the base density, I obtain the following informativeness measure:

$$\Delta^{\text{ind}} = 1 - \frac{\text{Var}[\text{res}(\mathbb{E}(h_{\widehat{c}}(Y, X) | \varepsilon) - \mathbf{1}\{\widetilde{x}'\eta \geq \varepsilon\}, \mathbb{E}(h_{\widehat{\gamma}}(Y, X) | \varepsilon))]}{\text{Var}[\mathbb{E}(h_{\widehat{c}}(Y, X) | \varepsilon) - \mathbf{1}\{\widetilde{x}'\eta \geq \varepsilon\}]}$$

Numerically, I find $\Delta^{\text{ind}} = 0.02$ for the first statistic (the frequency \widehat{p} of Y at $X = 1$), and $\Delta^{\text{ind}} = 0.38$ for the second one (the estimate $\widehat{\eta}$). Informativeness measures are thus quite different, and lower, when one restricts the analysis to an environment where ε and X are independent.

Interestingly, in this example one can also compute 1 minus the ratio of the two unrestricted squared biases, dividing the bias under independence by the bias computed in the larger neighborhood where independence is not imposed. This gives

$$\Gamma^{\text{ind}} = 1 - \frac{\text{Var}[\mathbb{E}(h_{\widehat{c}}(Y, X) | \varepsilon) - \mathbf{1}\{\widetilde{x}'\eta \geq \varepsilon\}]}{\text{Var}[h_{\widehat{c}}(Y, X) - \mathbf{1}\{\widetilde{x}'\eta \geq \varepsilon\}]}$$

Γ^{ind} can be interpreted as a measure of the “informativeness of the independence assumption” (or “informativeness of the random assignment of X ”). It does not depend on a vector of descriptive statistics and only reflects the impact of the independence assumption on the worst-case bias of \hat{c} . I compute this measure in the example and find $\Gamma^{\text{ind}} = 0.79$. This suggests that here imposing independence between ε and X reduces bias substantially.

Final Remarks. This paper contributes to the development of econometric methods for misspecified models and sensitivity analysis, which I believe is highly relevant for economic applications. Motivated by local bias calculations, the authors have proposed a measure of informativeness that provides a useful addition to the applied researcher’s toolkit. In this discussion, my goal has been to suggest using local asymptotic approximations to construct other measures which complement the one the authors have proposed, emphasizing two features: model misspecification, and economic and design assumptions.

The local asymptotic approach has considerable appeal in terms of tractability. Of course, this simplification is not without cost. Peter Huber lucidly expressed the main concern with local asymptotics: “The crucial point is that in any practical application we have a fixed, finite sample size, and we need to know whether we are inside the range of n and $[\mu]$ for which asymptotic theory yields a decent approximation” (Huber (1997)). Developing methods allowing for non-local misspecification is important; see Christensen and Connault (2019) for a recent example. Local approaches can also be used to compute estimators which are optimal according to some risk measure, as well as confidence intervals, as in Bonhomme and Weidner (2019) and Armstrong and Kolesár (2019). However, unlike the authors’ informativeness measure and the ones I have discussed here, all of which involve ratios of local biases that do not depend on μ to first order, optimal estimation and inference require one to take a stand on the (worst-case) degree of misspecification μ .

REFERENCES

- ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *Quarterly Journal of Economics*, 132 (4), 1553–1592. [2259]
- ARMSTRONG, T. B., AND M. KOLESÁR (2019): “Sensitivity Analysis Using Approximate Moment Condition Models,” Preprint, arXiv:1808.07387. [2264]
- ATTANASIO, O. P., C. MEGHIR, AND A. SANTIAGO (2012): “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA,” *Review of Economic Studies*, 79 (1), 37–66. [2263]
- BONHOMME, S. (2020): “Supplement to ‘A Comment on: “On the Informativeness of Descriptive Statistics for Structural Estimates” by Isaiah Andrews, Matthew Gentzkow, and Jesse M. Shapiro’,” *Econometrica Supplemental Material*, 88, <https://doi.org/10.3982/ECTA18789>. [2260]
- BONHOMME, S., AND M. WEIDNER (2019): “Minimizing Sensitivity to Model Misspecification,” Preprint, arXiv:1807.02161. [2260,2264]
- CHRISTENSEN, T., AND B. CONNAULT (2019): “Counterfactual Sensitivity and Robustness,” Unpublished Manuscript. [2264]
- HUBER, P. J. (1997): “Robustness: Where Are We Now?” in *L₁-Statistical Procedures and Related Topics*. Lecture Notes—Monograph Series, Vol. 31. Hayward, CA: Institute of Mathematical Statistics, 487–498. [2264]

Co-editor Ulrich K. Müller handled this manuscript.

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