Comment on “On the Informativeness of Descriptive Statistics for Structural Estimates”

Andres Santos
Department of Economics
UCLA
andres@econ.ucla.edu

July 10, 2020

1 Introduction

Andrews et al. (2018), henceforth AGS, make important progress on a challenging and fundamental question. The authors consider a setting in which a researcher reports an estimate $\hat{c}_n$ that is unbiased under the researcher's model. The researcher also reports a statistic $\hat{\gamma}_n$ and a measure of informativeness of $\hat{\gamma}_n$ for $\hat{c}_n$, denoted by Δ. AGS shows Δ can be related to properties of the distribution of $\hat{c}_n$ under models larger than the one employed by the researcher. As a result, Δ can play a role in reassuring a reader who is concerned about some aspects of the researcher's model.

I illustrate the properties of Δ with two examples. The first example is based on the classic selection problem of Heckman (1979) and portrays a setting in which a high Δ is not reassuring. The example shows that if a researcher has assumed random assignment and the reader is concerned about it, then the bias of $\hat{c}_n$ under the reader's model can be arbitrarily large regardless of Δ. The second example is based on the discussion of Attanasio et al. (2012) in AGS. This example illustrates that if $\hat{c}_n$ and $\hat{\gamma}_n$ are estimators of a common parameter, then Δ can be linked to the properties of a Hausman (1978) test. As a result, $\sqrt{1-\Delta}$ can be shown to equal the (normalized) largest bias of $\hat{c}_n$ under the reader's model and, in this sense, a high Δ is reassuring. I show this conclusion holds generally in a classical local asymptotic framework (Bickel et al., 1993). The two examples underscore that, as emphasized by AGS, whether a high Δ is reassuring to the reader can crucially depend on the particulars of the researcher's and reader's models.

I employ the notation in AGS with some minor modifications. The researcher's model is indexed by $\eta \in H$. The reader entertains a larger model indexed by $(\eta, \zeta) \in H \times Z$. I focus on the perspective of the reader, who views the parameter of interest as a function of (η, ζ) that we denote by $c(\eta, \zeta)$. Under the reader's model, the distribution of the data is determined by the value of (η, ζ). To emphasize this dependence, I write $E_{(\eta,\zeta)}[\hat{c}_n]$ to denote the expectation of $\hat{c}_n$ under (η, ζ). The property that $\hat{c}_n$ is unbiased under the researcher's model is equivalent to $E_{(\eta,0)}[\hat{c}_n] = c(\eta, 0)$ for any $\eta \in H$, while the bias under (η, ζ) in the reader's model is given by $E_{(\eta,\zeta)}[\hat{c}_n] - c(\eta, \zeta)$.

As in AGS, I consider an analysis local to proper specification, meaning we study (η, ζ) that are close to some $(\eta_0, 0)$. Since $E_{(\eta_0,0)}[\hat{c}_n] = c(\eta_0, 0)$ due to $\hat{c}_n$ being unbiased under the researcher's model, the bias of $\hat{c}_n$ under any (η, ζ) in the reader's model equals

$$\underbrace{E_{(\eta,\zeta)}[\hat{c}_n] - c(\eta, \zeta)}_{\text{Bias}(\eta,\zeta)} = \underbrace{E_{(\eta,\zeta)}[\hat{c}_n] - E_{(\eta_0,0)}[\hat{c}_n]}_{\Lambda_{\text{E}}(\eta,\zeta):=\text{Change in Expectation}} + \underbrace{c(\eta_0, 0) - c(\eta, \zeta)}_{\Lambda_{\text{P}}(\eta,\zeta):=\text{Change in Parameter}} . \quad (1)$$

AGS shows a close link between Δ and the properties of $\Lambda_{\text{E}}(\eta, \zeta)$, which in turn implies a close to link to the properties of $\text{Bias}(\eta, \zeta)$ over neighborhoods on which $\Lambda_{\text{P}}(\eta, \zeta) = 0$. The examples below explore to what extent Δ is reassuring when $\Lambda_{\text{P}}(\eta, \zeta) \neq 0$.

2 Two Examples

The first example illustrates a setting in which Δ should not reassure the reader.

Example 2.1. Consider a binary treatment model with potential outcomes $(Y(0), Y(1))$. The researcher observes a sample $\{Y_i(D_i), D_i\}_{i=1}^{2n}$ and postulates $Y(0), Y(1)$ are normally distributed with unknown means $\eta := (\beta(0), \beta(1))$ and known variances $(\sigma^2(0), \sigma^2(1))$. The parameter of interest is the average treatment effect (ATE). We observe an equal number n of treated and untreated observations, and let $\bar{Y}_n(d)$ denote the sample average among observations with $D_i = d$. The researcher assumes random assignment and reports $\hat{c}_n \equiv \bar{Y}_n(1) - \bar{Y}_n(0)$ as the estimate of the ATE. Further suppose that it is uncontroversial that $Y(0)$ is independent of D. To reassure a reader concerned about selection the researcher therefore sets $\hat{\gamma}_n = \bar{Y}_n(0)$ and reports Δ, which here equals

$$\Delta = \frac{\sigma^2(0)}{\sigma^2(0) + \sigma^2(1)}.$$

The reader is comfortable assuming $Y(0)$ is independent of D, but she is concerned that $Y(1)$ may not be independent of D. Under her model $(\bar{Y}_n(0), \bar{Y}_n(1))$ instead satisfy

$$\begin{pmatrix} \bar{Y}_n(0) \\ \bar{Y}_n(1) \end{pmatrix} \sim N\left(\begin{bmatrix} \beta(0) \\ \beta(1) + \zeta \end{bmatrix}, \begin{bmatrix} \sigma^2(0)/n & 0 \\ 0 & \sigma^2(1)/n \end{bmatrix} \right), \tag{2}$$

where ζ is the unknown selection bias.[1] Suppose the researcher reports that $\hat{\gamma}_n = \bar{Y}_n(0)$ is highly informative for $\hat{c}_n = \bar{Y}_n(1) - \bar{Y}_n(0)$. Should the reader be reassured about $\hat{c}_n$ since she believes $\hat{\gamma}_n$ is unbiased for $E[Y(0)]$? Intuitively, the answer should be no: Δ is determined by the variances $(\sigma^2(0), \sigma^2(1))$, which need not be related to the possible presence of selection into treatment. In fact, in this example, the bias of $\hat{c}_n$ under the reader's model can be arbitrarily large despite Δ being arbitrarily close to one.

For a formal analysis, note that under model (2) the ATE equals $c(\eta, \zeta) = \beta(1) - \beta(0)$. Set a point $(\eta_0, 0) = ((\beta_0(0), \beta_0(1)), 0)$ in the researcher's model and following AGS let

$$N_\mu := \left\{ (\eta, \zeta) : \frac{(\beta(0) - \beta_0(0))^2}{\sigma^2(0)} + \frac{(\beta(1) + \zeta - \beta_0(1))^2}{\sigma^2(1)} \leq \frac{\mu^2}{n} \right\}$$

be the corresponding local neighborhood. Under any (η, ζ) in the reader's model, the bias of $\hat{c}_n$ equals $E_{(\eta,\zeta)}[\hat{c}_n] - c(\eta, \zeta) = \zeta$ and hence the largest bias on N_μ is given by

$$\sup_{(\eta,\zeta) \in N_\mu} \text{Bias}^2(\eta, \zeta) = \sup_{(\eta,\zeta)} \zeta^2 \text{ s.t. } \frac{(\beta(0) - \beta_0(0))^2}{\sigma^2(0)} + \frac{(\beta(1) + \zeta - \beta_0(1))^2}{\sigma^2(1)} \leq \frac{\mu^2}{n}. \tag{3}$$

[1] While it is possible to devise a selection mechanism that delivers (2), note standard selection models such as Heckman (1979) deliver asymptotic (but not finite sample) normality. Here, I impose normality directly for the sake of transparency in the calculations that follow.

Therefore, the bias of $\hat{c}_n$ under the reader's model is unbounded on any neighborhood N_μ unless the domain of (η, ζ) is restricted – e.g., in (3) set $\beta(0) = \beta_0(0)$, $\beta(1) = \beta_0(1) - \zeta$, and let ζ be arbitrary. These calculations simply reflect that the ATE is not identified under the reader's model (Manski, 1989, 1990) – i.e., in (1) it is possible to make $\text{Bias}(\eta, \zeta)$ arbitrary by changing $\Lambda_{\text{P}}(\eta, \zeta)$ without affecting $\Lambda_{\text{E}}(\eta, \zeta)$.

To appreciate the role Δ plays in this example, define the restricted neighborhood $N_\mu^R := \{(\eta, \zeta) \in N_\mu : \beta(0) = \beta_0(0)\}$ – i.e., N_μ^R consists of the local distributions satisfying $E_{(\eta,\zeta)}[\bar{Y}_n(0)] = E_{(\eta_0,0)}[\bar{Y}_n(0)]$. Proposition 1 in AGS then implies that

$$\frac{\sup_{(\eta,\zeta)\in N_\mu^R} \Lambda_{\text{E}}^2(\eta, \zeta)}{\sup_{(\eta,\zeta)\in N_\mu} \Lambda_{\text{E}}^2(\eta, \zeta)} = \frac{\mu^2 \text{Var}\{\hat{\gamma}_n - \hat{c}_n\}}{\mu^2 \text{Var}\{\hat{c}_n\}} = \frac{\sigma^2(1)}{\sigma^2(0) + \sigma^2(1)} = 1 - \Delta. \tag{4}$$

Thus, Δ characterizes the ratio of the largest change in the expectation of $\hat{c}_n$ over N_μ^R to the largest change in expectation of $\hat{c}_n$ over N_μ. However, Δ does not offer an analogous connection to the bias of $\hat{c}_n$ under the reader's model. This conclusion may be overturned if we additionally impose that the ATE be constant on N_μ and N_μ^R– i.e., in defining N_μ and N_μ^R also require $\beta(1) - \beta(0) = \beta_0(1) - \beta_0(0)$, which implies $\Lambda_{\text{P}}(\eta, \zeta) = 0$. The resulting neighborhoods are akin to those employed by AGS and implicitly bound the selection bias as a function of μ – e.g., when $\mu = 0$, the selection bias can be unbounded on N_μ but equals zero on the subset of N_μ satisfying $\beta(1) - \beta(0) = \beta_0(1) - \beta_0(0)$. ■

Example 2.1 suggests that, whenever $c(\eta, \zeta)$ is not identified under the reader's model, a high Δ may not be reassuring. The next example illustrates that, whenever $c(\eta, \zeta)$ is identified under the reader's model, a close connection between Δ and the bias of $\hat{c}_n$ under the reader's model is guaranteed under an appropriate choice of $\hat{\gamma}_n$.

Example 2.2. Consider a randomized evaluation of a cash transfer program in which households are randomized into subsidy levels $s \in \{0, 1, 2\}$. Each household has three potential outcomes $(Y(0), Y(1), Y(2))$ and we observe $Y(S)$ where S is the subsidy level assigned to the household. The researcher postulates that $(Y(0), Y(1), Y(2))$ are normally distributed with known covariance matrix and means satisfying

$$E[Y(s)] = \alpha + \beta s \tag{5}$$

for some unknown $\eta := (\alpha, \beta)$. Let $\bar{Y}_n(s)$ denote the mean outcome for the group assigned subsidy s and assume for simplicity that the variance of $\bar{Y}_n(s)$ is the same for all s. The efficient estimator for (α, β) is then the OLS estimator $(\hat{\alpha}_n, \hat{\beta}_n)$. Suppose the parameter of interest is $E[Y(2)]$ and the researcher reports $\hat{c}_n = \hat{\alpha}_n + 2\hat{\beta}_n$. The researcher also sets $\hat{\gamma}_n = \bar{Y}_n(2)$ and reports the informativeness of $\hat{\gamma}_n$ for $\hat{c}_n$.

The reader is concerned about the linearity assumption in (5) and instead entertains

$$E[Y(s)] = \alpha + \beta s + 1\{s = 2\}\zeta,$$

while agreeing with all other aspects of the researcher's model. Thus, under the reader's model $\bar{Y}_n := (\bar{Y}_n(0), \bar{Y}_n(1), \bar{Y}_n(2))'$ is normally distributed with unknown mean satisfying $E_{(\eta,\zeta)}[\bar{Y}_n(s)] = \alpha + \beta s + 1\{s = 2\}\zeta$ and known covariance matrix, which we denote by Ω. Should the reader be reassured if $\hat{\gamma}_n$ is highly informative for $\hat{c}_n$? Intuitively, the answer should be yes: Because $\hat{\gamma}_n = \bar{Y}_n(2)$ is the estimator the reader would have preferred, a high correlation between $\hat{c}_n$ and $\hat{\gamma}_n$ lends credence to $\hat{c}_n$ under the reader's model.

A formal result follows from the fundamental insights of Hausman (1978). Under the reader's model, the parameter of interest $E[Y(2)]$ may be expressed as $c(\eta, \zeta) = \alpha + 2\beta + \zeta$. Set $\hat{r}_n := \hat{\gamma}_n - \hat{c}_n$ and note that $\hat{c}_n$ and $\hat{r}_n$ are uncorrelated due to $\hat{c}_n$ and $\hat{\gamma}_n$ being efficient and inefficient under the researcher's model. Further define

$$N_\mu := \{(\eta, \zeta) : \mathtt{D}(\eta, \zeta)'\Omega^{-1}\mathtt{D}(\eta, \zeta) \le \mu^2\} \quad \text{where} \quad \mathtt{D}(\eta, \zeta) := E_{(\eta,\zeta)}[\bar{Y}_n] - E_{(\eta_0,0)}[\bar{Y}_n]$$

as the neighborhood around $(\eta_0, 0)$. Since $\hat{\gamma}_n$ is an unbiased estimator for $c(\eta, \zeta)$ and $E_{(\eta_0,0)}[\hat{r}_n] = 0$ due to $\hat{c}_n$ being unbiased under the researcher's model, it follows that

$$\sup_{(\eta,\zeta)\in N_\mu} |E_{(\eta,\zeta)}[\hat{c}_n] - c(\eta, \zeta)| = \sup_{(\eta,\zeta)\in N_\mu} |E_{(\eta,\zeta)}[\hat{r}_n] - E_{(\eta_0,0)}[\hat{r}_n]| = \mu\sqrt{\mathrm{Var}\{\hat{r}_n\}} \tag{6}$$

$$\sup_{(\eta,\zeta)\in N_\mu} |c(\eta_0, 0) - c(\eta, \zeta)| = \sup_{(\eta,\zeta)\in N_\mu} |E_{(\eta_0,0)}[\hat{\gamma}_n] - E_{(\eta,\zeta)}[\hat{\gamma}_n]| = \mu\sqrt{\mathrm{Var}\{\hat{\gamma}_n\}}, \tag{7}$$

where the final equalities in (6) and (7) hold by Proposition 1 in AGS. Hence, for $\mathrm{Bias}(\eta, \zeta)$ and $\Lambda_{\mathrm{P}}(\eta, \zeta)$ as defined in (1), $\hat{c}_n$ and $\hat{r}_n$ being uncorrelated imply that

$$\frac{\sup_{(\eta,\zeta)\in N_\mu} \mathrm{Bias}^2(\eta, \zeta)}{\sup_{(\eta,\zeta)\in N_\mu} \Lambda^2_{\mathrm{P}}(\eta, \zeta)} = \frac{\mathrm{Var}\{\hat{r}_n\}}{\mathrm{Var}\{\hat{\gamma}_n\}} = 1 - \frac{\mathrm{Var}\{\hat{c}_n\}}{\mathrm{Var}\{\hat{\gamma}_n\}} = 1 - \frac{\mathrm{Cov}^2(\hat{c}_n, \hat{\gamma}_n)}{\mathrm{Var}\{\hat{c}_n\}\mathrm{Var}\{\hat{\gamma}_n\}} := 1 - \Delta, \tag{8}$$

where the final equality follows by definition of Δ. In words, Δ determines the largest bias that $\hat{c}_n$ may attain on N_μ under the reader's model, normalized by the largest change that the parameter of interest may take over N_μ under the reader's model.

To connect (8) to Example 2.1 and the analysis in AGS, define $N^R_\mu := \{(\eta, \zeta) \in N_\mu : \gamma(\eta, \zeta) = \gamma(\eta_0, 0)\}$. By Proposition 1 in AGS we then obtain

$$\frac{\sup_{(\eta,\zeta)\in N^R_\mu} \Lambda^2_{\mathrm{E}}(\eta, \zeta)}{\sup_{(\eta,\zeta)\in N_\mu} \Lambda^2_{\mathrm{E}}(\eta, \zeta)} = \frac{\mu^2\mathrm{Var}\{\hat{r}_n\}\mathrm{Var}\{\hat{c}_n\}/\mathrm{Var}\{\hat{\gamma}_n\}}{\mu^2\mathrm{Var}\{\hat{c}_n\}} = 1 - \Delta, \tag{9}$$

where $\Lambda_{\mathrm{E}}(\eta, \zeta)$ was defined in (1). Thus, as in Example 2.1, Δ is closely linked to the possible changes in the expectation of $\hat{c}_n$ under the reader's model. However, unlike in Example 2.1, setting $\hat{\gamma}_n$ to be the efficient estimator under the reader's model additionally ensures that Δ fully determines the (normalized) largest bias that $\hat{c}_n$ may attain under the reader's model. In fact, the ratio of the change in expectations of $\hat{c}_n$ (i.e. (9)) is exactly equal to the largest normalized bias (i.e. (8)). ■

The arguments underlying Example 2.2 readily extend to a local asymptotic framework. I establish such a result by closely following Section 3.4 in Chen and Santos (2018), which proposes a generalization of the incremental J-test of Eichenbaum et al. (1988). Suppose $\{V_i\}_{i=1}^n$ is an i.i.d. sample, let $\mathbf{P}$ and $\mathbf{M}$ denote the set of distributions of V that agree with the researcher's and the reader's model, and assume $\mathbf{P} \subseteq \mathbf{M}$. I conduct an analysis local to a distribution $P_0 \in \mathbf{P}$ that satisfies the assumptions of the researcher's model. Specifically, for $T(P_0)$ the tangent space of $\mathbf{M}$ at P_0, let $V \sim P_{1/\sqrt{n},g}$ for $t \mapsto P_{t,g} \in \mathbf{M}$ a parametric submodel with score $g \in T(P_0)$. Further assume the parameter of interest is identified from the distribution of V under the reader's model. To make such dependence explicit, I write $\theta(P)$ for the unique value of the parameter that is compatible with V having distribution $P \in \mathbf{M}$. Provided $\hat{c}_n$ is asymptotically linear, it is possible to define the analogues to $\text{Bias}(\eta, \zeta)$ and $\Lambda_{\text{E}}(\eta, \zeta)$ in (1) through

$$\sqrt{n}\{\hat{c}_n - \theta(P_{1/\sqrt{n},g})\} \overset{L_{n,g}}{\to} N(\text{Bias}(g), \text{AsyVar}\{\hat{c}_n\})$$
$$\sqrt{n}\{\hat{c}_n - \theta(P_0)\} \overset{L_{n,g}}{\to} N(\Lambda_{\text{E}}(g), \text{AsyVar}\{\hat{c}_n\}),$$

where $\overset{L_{n,g}}{\to}$ denotes convergence in distribution along $\{V_i\}_{i=1}^n \sim \bigotimes_{i=1}^n P_{1/\sqrt{n},g}$. Also set

$$\Lambda_{\text{P}}(g) := \lim_{n\to\infty} \sqrt{n}\{\theta(P_{1/\sqrt{n},g}) - \theta(P_0)\}$$

as the analogue to $\Lambda_{\text{P}}(\eta, \zeta)$ in (1). Finally, as local neighborhoods we employ $N_\mu := \{g \in T(P_0) : \|g\|_{P,2} \leq \mu\}$ and $N_\mu^R := \{g \in N_\mu : \sqrt{n}\{\theta(P_{1/\sqrt{n},g}) - \theta(P_0)\} = o(1)\}$.

Chen and Santos (2018) shows that if $\hat{c}_n$ and $\hat{\gamma}_n$ are efficient estimators under $\mathbf{P}$ and $\mathbf{M}$, then a Hausman (1978) test based on $\hat{\gamma}_n - \hat{c}_n$ aims its power at the (local) distribution in $\mathbf{M} \setminus \mathbf{P}$ that induces the largest bias on $\hat{c}_n$. The next proposition employs the same arguments to generalize the conclusions of Example 2.2. I omit a proof and regularity conditions for conciseness, though note the result allows $\mathbf{M}$ to be overidentified.

Proposition 2.1. *Let $\hat{c}_n$ and $\hat{\gamma}_n$ be asymptotically linear efficient estimators of $\theta(P_0)$ at $P_0 \in \mathbf{P} \cap \mathbf{M}$ with respect to $\mathbf{P}$ and $\mathbf{M}$. Then, under classical regularity conditions:*

$$\frac{\sup_{g\in N_\mu} \text{Bias}^2(g)}{\sup_{g\in N_\mu} \Lambda_{\text{P}}^2(g)} = \frac{\sup_{g\in N_\mu^R} \Lambda_{\text{E}}^2(g)}{\sup_{g\in N_\mu} \Lambda_{\text{E}}^2(g)} = \frac{\text{AsyVar}\{\hat{\gamma}_n - \hat{c}_n\}}{\text{AsyVar}\{\hat{\gamma}_n\}} = 1 - \Delta.$$

3 Summary

Reporting Δ can be a useful tool for applied work. However, as emphasized by AGS, a high Δ is not automatically reassuring nor is it a substitute for identification analysis. Example 2.1 illustrates that Δ may not reassure a reader whose model fails to identify $c(\eta, \zeta)$. In contrast, whenever $c(\eta, \zeta)$ is identified under the reader's model, Example 2.2 and Proposition 2.1 illustrate one may select a $\hat{\gamma}_n$ that ensures Δ is informative.

References

ANDREWS, I., GENTZKOW, M. and SHAPIRO, J. M. (2018). On the informativeness of descriptive statistics for structural estimates. Tech. rep., National Bureau of Economic Research.

ATTANASIO, O. P., MEGHIR, C. and SANTIAGO, A. (2012). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. *The Review of Economic Studies*, **79** 37–66.

BICKEL, P. J., KLASSEN, C. A., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

CHEN, X. and SANTOS, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

EICHENBAUM, M. S., HANSEN, L. P. and SINGLETON, K. J. (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, **103** 51–78.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society* 1251–1271.

HECKMAN, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, **47** 153–161.

MANSKI, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources* 343–360.

MANSKI, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, **80** 319–323.