Rejoinder to Comments on

On the Informativeness of Descriptive Statistics for Structural Estimates

Isaiah Andrews, *Harvard University and NBER*[*]

Matthew Gentzkow, *Stanford University and NBER*

Jesse M. Shapiro, *Brown University and NBER*

August 2020

We greatly appreciate the opportunity to exchange ideas with, and learn from, the discussants. Their comments open up new areas for exploration, and deepen our understanding. We thank them for their thoughtful engagement with our work.

The comments include many insightful points, and for brevity we do not address them all. Instead, we focus here on elaborating some aspects of our approach in relation to some of the points and examples in the comments.

The primitive elements of our setting are the distribution of the data F, the quantity of interest c, and the estimator $\hat{c}$. A researcher's goal is to estimate c using a sample of observations from F.

We are especially interested in settings in which c can be connected to F only through economic structure. An example is the application to Gentzkow (2007) that we pursue in Section 6.2 of the article. There, c is the effect on print readership of the *Washington Post* of completely eliminating the paper's online edition. One can imagine idealized experiments that recover this quantity, but such experiments have not been run, and Gentzkow (2007) instead uses a model of newspaper demand to extrapolate c from survey data on individual news reading choices.

Because the quantity of interest c is connected to F only through economic assumptions, c will not in general be point identified once we admit the possibility of misspecification. We take as our starting point an extreme version of this situation where, absent further restrictions, any value of c

[*]E-mail: iandrews@fas.harvard.edu, gentzkow@stanford.edu, jesse_shapiro_1@brown.edu.

is consistent with any value of F.[1]

This differs from some of the settings discussed in the comments (Bonhomme, Misspecification Example; Kitamura, Section 2; Santos, Example 2.2) in which we can uncontroversially define c as a functional of the distribution F. Though not our focus here, such settings are very important. For example, Gentzkow (2007) may wish to know the median number of hours per week spent consuming the news, or a labor economist may wish to calculate the difference in wages between the 90th and 10th percentile workers. Given knowledge of the true distribution F of the data (on time spent consuming the news, on earnings), no economic assumptions are needed to learn c in these cases.

By contrast, when c is the effect of a counterfactual policy or decision, or a model-defined object like consumer surplus, we cannot uncontroversially define c as a functional of F. Instead, we learn about c from F through contestable economic assumptions. Gentzkow's (2007) model defines a particular functional $c^0(F)$ such that, under the model, $c = c^0(F)$.[2] Gentzkow then constructs an estimator $\hat{c}$ based on the observed data, where $\hat{c}$ is asymptotically unbiased for c under the model. For ease of exposition let us abstract from the approximation error, so $\mathrm{E}_F[\hat{c}] = c^0(F)$ and hence under the model $\mathrm{E}_F[\hat{c}] = c$ for all $F \in \mathcal{F}^0(c)$.

We suspect that few, if any, readers of Gentzkow (2007) think that his analysis gets the counterfactual c exactly right, even given perfect knowledge of F. But we also suspect that many readers would find the estimate $\hat{c}$ in Gentzkow (2007) useful in learning about c. Thus, we suspect that the reality for many readers is in between the situation of total non-identification of c and the situation of uncontroversial point identification.

We formalize this in-between situation by considering a reader who believes the true model is in a statistical neighborhood of the researcher's model. Specifically, we suppose that for the true value of c, the distribution F lies in a neighborhood $\mathcal{N}\left(\tilde{F}\right)$ of a distribution $\tilde{F}$ consistent with c under the researcher's model.[3] Hence, the true (c, F) pair may lie outside of those allowed by the researcher's model, but is "close" in a sense determined by the neighborhoods $\mathcal{N}$. Correspondingly, a given distribution F may now be consistent with a set $\mathcal{C}^N(F)$ of possible values of c.[4] Importantly, though not in general a singleton, for bounded $\mathcal{N}$ the set $\mathcal{C}^N(F)$ will typically be limited in size.

This relaxation of the researcher's model corresponds to the intuition that if Gentzkow's (2007)

[1]Specifically, under the assumptions in Section 2 of the article, the set of (c, F) pairs $\{(c(\eta), F(\eta, \zeta)) : \eta \in H, \zeta \in Z\}$ is the Cartesian product $\{c(\eta) : \eta \in H\} \times \{F(\eta, \zeta) : \eta \in H, \zeta \in Z\}$, so the identified set for c under any $F \in \{F(\eta, \zeta) : \eta \in H, \zeta \in Z\}$ is $\mathcal{C}(F) = \{c(\eta) : \eta \in H\}$. Bonhomme and Santos consider settings where the quantity of interest is $\tilde{c}(\eta, \zeta)$ and so depends on both η and ζ. Since the foundation of our approach is the set of possible (c, F) pairs, so long as $\{(\tilde{c}(\eta, \zeta), F(\eta, \zeta)) : \eta \in H, \zeta \in Z\}$ remains a Cartesian product, equal to $\{c(\eta) : \eta \in H, \zeta \in Z\} \times \{F(\eta, \zeta) : \eta \in H, \zeta \in Z\}$, our analysis remains unchanged.

[2]To connect this notation to that in the article, write $c^0(F) = \left\{c : F \in \mathcal{F}^0(c)\right\}$.

[3]As Kitamura (Section 1) highlights, we can think of $\tilde{F}$ as an "ideal" distribution which is then perturbed to obtain F.

[4]In the notation of the article, $\mathcal{C}^N(F) = \left\{c : F \in \mathcal{F}^N(c)\right\}$.

assumptions are approximately right, we are still able to learn something from his analysis. This contrasts with the setting in Santos, Example 2.1, which considers an alternative relaxation of the researcher's model under which there is no relationship between c and F so, as Santos notes, c is entirely unidentified. This alternative relaxation is not nested by the one we define based on bounded neighborhoods $\mathcal{N}$.

The set $\mathcal{C}^N(F)$ tells us which values of c are consistent with the complete distribution of the observed data. In reality, most readers of Gentzkow's (2007) paper only know the data features reported in the published article. Imagining a reader who knows only the mean of the estimator, we can define the larger identified set $\mathcal{C}^N(\mathrm{E}_F[\hat{c}]) \supseteq \mathcal{C}^N(F)$ of values of c consistent with that mean.[5]

The last step in our analysis quantifies the benefits of re-imposing certain model-derived assumptions. Gentzkow's (2007) model makes a large number of assumptions. Two potentially important ones are timing restrictions, namely that shocks to utility for different editions of the *Post* are independent over time for a given respondent, and exclusion restrictions, namely that certain variables shift the utility for reading the online edition without directly affecting the utility for reading the print edition.

Some readers may find the timing restrictions more palatable than the exclusion restrictions. Others may feel the opposite. Both types of readers may wish to know which restriction is more important for Gentzkow's (2007) conclusions. One approach to answering that question is to ask how much the set $\mathcal{C}^N(F)$ narrows if we impose that only one of these restrictions is correct. The measure Γ^{ind} in Bonhomme (Structure Example) illustrates this approach for an independence assumption not unlike the exclusion restriction in Gentzkow's (2007) application.

Our approach instead leverages the fact that, because Gentzkow's (2007) assumptions imply a relationship between F and c, they also imply a relationship between $\gamma(F)$ and c, for $\gamma(\cdot)$ any functional of F. If we can find a functional $\gamma(\cdot)$ that is related to c mainly through one of the two restrictions, then we can ask how much preserving its relationship to c shrinks $\mathcal{C}^N(\mathrm{E}_F[\hat{c}])$ by considering the size of the set $\mathcal{C}^{RN}(\mathrm{E}_F[\hat{c}]) \subseteq \mathcal{C}^N(\mathrm{E}_F[\hat{c}])$ that results when we exclude from consideration any models that disagree with Gentzkow (2007) on the relationship between $\gamma(F)$ and c.[6] In the normal setting in Section 3 of the article, or in our local asymptotic analysis, the size of the set $\mathcal{C}^{RN}(\mathrm{E}_F[\hat{c}])$ relative to that of $\mathcal{C}^N(\mathrm{E}_F[\hat{c}])$ is $\sqrt{1-\Delta}$, for Δ the informativeness measure that we define.

[5]This is
$$\mathcal{C}^N(\mathrm{E}_F[\hat{c}]) = \left\{c : \mathrm{E}_F[\hat{c}] = \mathrm{E}_{\tilde{F}}[\hat{c}] \text{ for some } \tilde{F} \in \mathcal{F}^N(c)\right\}.$$
Under the local misspecification asymptotics we consider in the article, ζ impacts only the mean of $\hat{c}$. When the same is true in finite samples, the identified set based on the mean of $\hat{c}$ is equivalent to the identified set based on the full distribution of $\hat{c}$.

[6]In the notation of the article, $\mathcal{C}^{RN}(F) = \{c : F \in \mathcal{F}^{RN}(c)\}$.

We think our approach is likely to be especially useful in settings where there are functionals $\gamma(F)$ that are naturally tied to c via focal assumptions of the researcher's model. One reason to think such settings are important is the common practice in structural research papers of reporting descriptive statistics and discussing the sense in which they "drive" the researcher's estimator (see also the discussion in Andrews et al. Forthcoming). In settings where such functionals are not available, an approach along the lines of Bonhomme's Γ^{ind} seems more suitable.

Kitamura (Section 2) and Santos (Example 2.2) observe that if, unlike in Gentzkow (2007), c is an uncontroversial functional of F, then Δ governs the scope for improved efficiency by imposing the model-implied restrictions on the relationship between c and $\gamma(F)$. This observation highlights that Δ may have interesting interpretations other than those we emphasize, and that, as Kitamura notes, there are deep connections between questions of efficiency and robustness.

We again thank the discussants for their extremely thoughtful and insightful comments.

References

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. Forthcoming. Transparency in structural research. *Journal of Business and Economic Statistics.*

Gentzkow, Matthew. 2007. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* 97(3): 713-744.