Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Transparency in Structural Research

Isaiah Andrews, Matthew Gentzkow & Jesse M. Shapiro

To cite this article: Isaiah Andrews, Matthew Gentzkow & Jesse M. Shapiro (2020) Transparency in Structural Research, Journal of Business & Economic Statistics, 38:4, 711-722, DOI: 10.1080/07350015.2020.1796395

To link to this article: https://doi.org/10.1080/07350015.2020.1796395

Transparency in Structural Research

Isaiah Andrews
Department of Economics, Harvard University, Cambridge, MA; NBER, Cambridge, MA

Matthew Gentzkow
Department of Economics, Stanford University, Stanford, CA; NBER, Cambridge, MA (*gentzkow@stanford.edu*)

Jesse M. Shapiro
Department of Economics, Brown University, Providence, RI; NBER, Cambridge, MA

We propose a formal definition of transparency in empirical research and apply it to structural estimation in economics. We discuss how some existing practices can be understood as attempts to improve transparency, and we suggest ways to improve current practice, emphasizing approaches that impose a minimal computational burden on the researcher. We illustrate with examples.

KEY WORDS: Exploratory data analysis; Multivariate analysis; Sensitivity analysis.

1. INTRODUCTION

Structural empirical research can sometimes look like a black box. Once upon a time, a structural article might begin with an elaborate model setup containing dozens of assumptions, present a similarly complex recipe for estimation, and then jump immediately to reporting model estimates and counterfactuals that answer the research question of interest. A reader who accepted the full list of assumptions could walk away having learned a great deal. A reader who questioned even one of the assumptions might learn very little, as they would find it hard or impossible to predict how the conclusions might change under alternative assumptions.

Modern research articles taking a structural approach often look very different from this caricature. Many devote significant attention to descriptive analysis of important facts and relationships in the data. Many provide detailed discussions of how these descriptive statistics relate to the structural estimates, connecting specific data features to key parameter estimates or conclusions. Such analysis has the potential to make structural estimates more transparent, helping skeptical readers learn from the results even when they do not fully accept all the model assumptions.

In this article, we consider the value of transparency in structural research. We propose a formal definition of the transparency of a statistical report. We argue that our definition provides a rationale for many current practices, and suggests ways these practices can be improved. We discuss these potential improvements, emphasizing those that impose a minimal computational burden on the researcher.

Our definition of transparency follows the one proposed in Andrews, Gentzkow, and Shapiro (2017).[1] We situate it in a model of scientific communication based on Andrews and Shapiro (2020). In the model, a researcher observes data informative about a quantity of interest c. The researcher reports an estimate $\hat{c}$ of c along with auxiliary statistics $\hat{t}$ to a set of readers indexed by r. Under the researcher's maintained assumptions a_0, $\hat{c}$ is valid, for example, in the sense that it is asymptotically normal and unbiased. Not all readers accept a_0, however, and different readers may entertain different alternative assumptions $a \neq a_0$. After receiving the report $(\hat{c}, \hat{t})$, each reader updates their prior beliefs, selects an estimate d_r of c, and realizes a quadratic loss $(d_r - c)^2$. For a given reader, we define the *transparency* of the report to be the reduction in expected loss from observing $(\hat{c}, \hat{t})$, relative to the reduction from observing the full data. In other words, research is transparent to the extent that it makes it easy for readers to reach the same inference about c that they would reach by analyzing the data in full under their preferred assumptions. We show that transparency is distinct from other econometric desiderata such as efficiency and robustness.

After describing our model and definition of transparency in Section 2, we discuss several practices that we believe can improve the transparency of structural estimation. We illustrate throughout with stylized examples drawn from our model and real-world examples drawn from the literature.

Section 3 discusses descriptive analysis, which we interpret as including in $\hat{t}$ statistics $\hat{s}$ that are either directly informative about the parameter of interest c, or informative about the plausibility of the assumptions a_0. We argue that descriptive statistics of both kinds can aid transparency.

Section 4 discusses the analysis of identification. Although transparency is a distinct property from model identification, we argue that clear discussion of identification can improve transparency by sharpening readers' beliefs about the appropriateness of the researcher's assumptions.

Section 5 discusses ways to improve the transparency of the estimator. We argue that transparency is improved when

[1] See also the discussions of transparency in Angrist and Pischke (2010) and Heckman (2010).

Journal of Business & Economic Statistics
October 2020, Vol. 38, No. 4
DOI: 10.1080/07350015.2020.1796395

$\hat{c}$ depends to a large degree on some interpretable statistics $\hat{s}$ and when the form of the relationship between the two is made clear to the reader. We suggest this as a rationale for targeting descriptive statistics directly in estimation. Building on work by Andrews, Gentzkow, and Shapiro (2017, 2020), we discuss how local approximations can be used to clarify the relationship between $\hat{c}$ and $\hat{s}$.

Section 6 discusses sensitivity analysis, which we take to encompass a range of approaches for demonstrating the sensitivity of conclusions to alternative assumptions. When readers are concerned about a small number of known alternative assumptions a, the researcher can improve transparency by reporting estimates $\hat{c}_a$ that are valid under these alternatives, as in a traditional sensitivity analysis. When readers are concerned about a richer or unknown set of alternatives a, then it is no longer practical to report an estimate corresponding to each of these. Building on work by Conley, Hansen, and Rossi (2012) and Andrews, Gentzkow, and Shapiro (2017), we discuss how including in $\hat{t}$ statistics based on local approximations can help readers assess a larger set of assumptions. For cases where a qualitative conclusion (e.g., the direction of a causal effect or welfare change) is important, we also discuss the value of reporting features of alternative realizations of the data that would lead to a conclusion different from the researcher's.

2. TRANSPARENCY IN A MODEL OF SCIENTIFIC COMMUNICATION

2.1. Setup

A researcher observes data $D \in \mathscr{D}$. The researcher makes a set of assumptions a_0 under which $D \sim F(a_0, \eta)$ for $\eta \in H$ an unknown parameter. The researcher computes a point estimate $\hat{c} = \hat{c}(D)$ of a scalar quantity of interest $c(a_0, \eta)$, along with a vector of auxiliary statistics $\hat{t} = \hat{t}(D)$. The latter may include descriptive evidence, sensitivity analysis, and various auxiliary statistics as discussed in Sections 3–6.[2]

The researcher reports $(\hat{c}, \hat{t})$ to readers $r \in \mathscr{R}$ who do not have access to the underlying data. In most applications, researchers and readers focus on statistics of much lower dimension than the raw data (though researchers might also make the data available), so we will primarily consider $\dim(\hat{t}) \ll \dim(D)$ and ask what readers learn from $(\hat{c}, \hat{t})$. Readers are concerned that the researcher's model may be misspecified, and they consider assumptions $a \in \mathscr{A}$ that may be different from a_0. Under assumption $a \in \mathscr{A}$, $D \sim F(a, \eta)$ where η is again unknown, and the quantity of interest is $c(a, \eta)$.[3] Each reader r has a prior π_r on the assumptions a and model parameter η, and aims to estimate $c(a, \eta)$, choosing a decision $d_r \in \mathbb{R}$ and incurring quadratic loss $L(d_r, c(a, \eta)) = (d_r - c(a, \eta))^2$.

Following Andrews and Shapiro (2020), we define reader r's *communication risk* from $(\hat{c}, \hat{t})$ as their ex-ante expected loss from taking their optimal action based on $(\hat{c}, \hat{t})$. Under squared error loss this optimal action is simply r's posterior mean for c given $(\hat{c}, \hat{t})$, so

$$E_r\left[\min_{d_r} E_r\left[(d_r - c)^2 \,|\hat{c}, \hat{t}\right]\right] = E_r\left[\mathrm{var}_r\left(c|\hat{c}, \hat{t}\right)\right].$$

Here, $E_r[\cdot]$ and $\mathrm{var}_r(\cdot)$ denote the expectation and variance under π_r, respectively, and we write c as shorthand for $c(a, \eta)$. Reader r's risk from observing the full data is $E_r[\mathrm{var}_r(c|D)] \leq E_r\left[\mathrm{var}_r\left(c|\hat{c}, \hat{t}\right)\right] \leq \mathrm{var}_r(c)$, with equality in the first comparison only if r's posterior mean based on $(\hat{c}, \hat{t})$ is almost surely the same as that based on the full data.

We define the *transparency* of $(\hat{c}, \hat{t})$ *for* r as the reduction in communication risk from observing $(\hat{c}, \hat{t})$, relative to the reduction from observing the full data

$$T_r\left(\hat{c}(\cdot), \hat{t}(\cdot)\right) = \frac{\mathrm{var}_r(c) - E_r\left[\mathrm{var}_r\left(c|\hat{c}, \hat{t}\right)\right]}{\mathrm{var}_r(c) - E_r[\mathrm{var}_r(c|D)]},$$

and define transparency to be one when the denominator is zero. Thus, the transparency of $(\hat{c}, \hat{t})$ for r lies between zero and one, is equal to one when observing $(\hat{c}, \hat{t})$ yields the same risk for r as observing the full data, and is equal to zero when observing $(\hat{c}, \hat{t})$ yields no reduction in risk, while observing D would yield some reduction.

It is sometimes straightforward to construct fully transparent reports, that is, reports with transparency equal to one. If $\hat{t}$ is sufficient for (a, η), for instance, then $(\hat{c}, \hat{t})$ is fully transparent for all readers. When it is infeasible to report a sufficient statistic, we can still construct a fully transparent report for reader r by reporting that reader's posterior mean $\hat{t} = E_r[c|D]$. Note, however, that in this case $(\hat{c}, \hat{t})$ need not be transparent for readers r' with $\pi_{r'} \neq \pi_r$. Heterogeneity in π_r across readers is thus central to the study of transparency.

2.2. Example and Comparison to Other Econometric Properties

A linear IV example helps to fix ideas and clarify the difference between transparency and other econometric properties. Suppose that the data $D = \{(Y_i, X_i, Z_i)\}_{i=1}^n$ consist of observations of an outcome Y_i, an endogenous regressor X_i, and a candidate instrument Z_i, all of which are scalar. Readers believe that the data follow

$$Y_i = X_i c + Z_i a + \varepsilon_i, \tag{1}$$

$$X_i = Z_i \gamma + V_i, \tag{2}$$

where the instruments Z_i are fixed. The reduced-form error from regressing Y_i on Z_i is $U_i = cV_i + \varepsilon_i$. We assume the errors (U_i, V_i) are iid normal across i, $(U_i, V_i) \sim N(0, \Xi)$, with Ξ commonly known, so the parameter is $\eta = (c, \gamma) \in \mathbb{R}^2$. Suppose that $\mathscr{A} = \mathbb{R}$, so that assumptions $a \in \mathscr{A}$ correspond to the coefficient on Z_i in (1), and that the researcher's assumption is $a_0 = 0$. Under assumption a_0, Z_i is a valid instrument in the regression of Y_i on X_i, while under $a \neq 0$ the exclusion restriction fails. Denote the usual IV estimate by $\hat{c} = \sum Z_i Y_i / \sum Z_i X_i$ and the first-stage coefficient by $\hat{\gamma} = \sum Z_i X_i / \sum Z_i^2$.

The report $\hat{c}$ may not be fully transparent. For example, consider a reader r who has a degenerate prior on $a \neq a_0$

[2] In settings where c is partially identified under the assumptions a_0, one could instead take $\hat{c}$ to report an estimate for the identified set. Some of our analysis (particularly in Section 5) would need to be adapted to this case. See Tamer (2010) and Molinari (2020) for overviews of the partial identification literature.

[3] We assume a common parameter η for simplicity, but one could more generally have different model parameters η_a for each $a \in \mathscr{A}$. Alternatively, one can view a as just another unknown parameter, though in many interesting cases (a, η) will not be jointly identified.

but a continuous joint prior on (c, γ). Note that for n large, $\hat{c}$ converges in probability to $c + a/\gamma$ under mild conditions. Because the reader is uncertain about the value of γ, however, they cannot infer the value of c from the estimate $\hat{c}$ even in a large sample. By contrast, with access to the full data they would learn the value of γ, and thus be able to infer c. Thus, as $n \to \infty$ the transparency of $\hat{c}$ for such a reader is bounded away from one. In contrast, the report $(\hat{c}, \hat{\gamma})$ has transparency $T_r = 1$ for all readers r, as $(\hat{c}, \hat{\gamma})$ is sufficient for the unknown parameters (c, γ). Reporting the auxiliary statistic $\hat{\gamma}$ can thus improve transparency in this example.

Transparency is distinct from a number of other properties discussed in the econometrics literature. For example, estimators are often evaluated based on their efficiency in mean squared error, where the mean squared error of $\hat{c}$ under (a, η) is $E_{F(a,\eta)}\left[\left(c - \hat{c}\right)^2\right]$. The estimator $\tilde{c}$ dominates the estimator $\hat{c}$ in mean squared error under the assumptions a if it achieves a lower mean squared error for all η, with strict inequality for some η. Efficiency and transparency can imply substantially different rankings of estimators. To illustrate, continue with the instrumental variables example and suppose along the lines of Andrews and Shapiro (2020) that all readers believe c lies between values c_L and c_U with probability one, $\Pr_r\{c \in [c_L, c_U]\} = 1$ for all r. Let $\hat{c}$ again denote the IV estimator, and let $\tilde{c}$ denote the IV estimator censored to lie in $[c_L, c_U]$, $\tilde{c} = \max\left\{c_L, \min\left\{\hat{c}, c_U\right\}\right\}$. The estimator $\tilde{c}$ dominates $\hat{c}$ in mean squared error (indeed, the mean squared error of $\hat{c}$ is infinite whenever Ξ has full rank).[4] At the same time, since $\tilde{c}$ is a non-invertible transformation of $\hat{c}$, the report $\hat{c}$ is weakly more transparent than the report $\tilde{c}$ for all readers r, and the report $(\hat{c}, \hat{\gamma})$ achieves full transparency ($T_r = 1$) for all readers r, while the report $(\tilde{c}, \hat{\gamma})$ does not.

While we allow the possibility that the readers and the researcher contemplate different assumptions, transparency is also distinct from traditional measures of robustness. To illustrate, note that in our instrumental variables example with $\mathscr{A} = \mathbb{R}$, the report $(\hat{c}, \hat{\gamma})$ is fully transparent, but all estimators $\bar{c}$ of c have infinite worst-case mean squared error over $(c, a) \in \mathbb{R}^2$ for any γ, $\sup_{(c,a)\in\mathbb{R}^2} E_{F(a,\eta)}\left[(c - \bar{c})^2\right] = \infty$, and so are non-robust in that sense.

Finally, transparency is distinct from identification. In our instrumental variables example, $(\hat{c}, \hat{\gamma})$ is fully transparent, but c is unidentified under $\mathscr{A}$ absent further restrictions, in the sense that any distribution for D allowed by the model is consistent with any value of c.

2.3. Relationship to Other Recent Discussions of Transparency and Interpretability

We define transparency as a property of the statistics that the researcher reports. The usefulness of a given statistical report depends on the reader's understanding of the process that generated the statistic. Transparency as we define it is thus related to growing literature on research transparency in economics, which emphasizes issues such as clear documentation of experimental procedures (see, e.g., Christensen and Miguel 2018). One focus of that literature is the open sharing of research data, which achieves transparency equal to one for any reader r with the capacity to analyze the full data D.

We focus on applications to structural research in economics, where statistical methods are often derived from explicit assumptions about economic primitives. Applications of machine learning, by contrast, often focus on exploring relationships among observed variables without an explicit causal framework. A recent literature considers ways to improve the interpretability of the models used in machine learning, and of the resulting estimates (see, e.g., Murdoch et al. 2019). Some of the approaches emphasized in that literature, such as the highlighting of data features important for a given prediction, seem related in spirit to those we discuss below.

2.4. Routes to Improved Transparency

The remaining sections of the article discuss practical approaches to improving transparency in structural estimation. We emphasize alternative assumptions a that we think are likely to be of most interest to readers of structural research. Likewise, we limit attention to reporting strategies that we view as reasonable, ruling out for instance that researchers encode the full data in the decimal expansion of $\hat{t}$. Finally, because working with nonlinear structural models is often computationally expensive, we emphasize approaches that impose a minimal additional computational burden on the researcher.

3. DESCRIPTIVE ANALYSIS

The first element that can contribute to transparent structural research is descriptive analysis. In our framework, a descriptive analysis takes the auxiliary statistics $\hat{t}$ to include some statistics $\hat{s}$ that are either directly informative about c or informative about the plausibility of the assumptions a_0. Examples include summary statistics, data visualization, or correlations illustrating key causal relationships. Such evidence is sometimes described as "model-free," in the sense that it has a meaningful interpretation that does not rely explicitly on the assumptions of the structural model.[5] Pakes (2014) formalized the role of descriptive analysis in providing a set of facts that the structural model should rationalize.[6]

Our framework suggests two ways that such descriptive analysis can improve transparency. First, descriptive statistics $\hat{s}$ may provide evidence about c that is informative under a wider range of assumptions than a_0. This would be true, for example, if $\left|\text{corr}_r\left(c, \hat{s}\right)\right|$ is large under many priors π_r, including those that do not put much mass on a_0.[7]

A leading case is where $\hat{s}$ includes convincing experimental or quasi-experimental estimates of treatment effects closely related to c. Autor et al. (2019), for example, presented quasi-experimental evidence on the effects of disability insurance (DI) receipt in Norway on outcomes including total income,

[4] The absolute deviation of $\tilde{c}$ from c is weakly smaller than that of $\hat{c}$ for all realizations of the data, and strictly smaller for some, so $\tilde{c}$ also dominates $\hat{c}$ in many other senses, for example, as measured by quantiles of the absolute deviation.

[5] See, for example, Polyakova (2016) and Rossi and Chintagunta (2016).

[6] See also the discussion in Lewbel (2019, sec. 5.1).

[7] In particular note that for scalar $\hat{s}$, $E_r\left[\text{var}_r\left(c|\hat{s}\right)\right] \leq \text{var}_r(c)\left(1 - \text{corr}_r\left(c, \hat{s}\right)^2\right)$, so a large correlation directly bounds the average posterior variance.

consumption expenditure, and transfer income, using random assignment of DI judges as a source of exogenous variation. They then estimated a structural model that allows them to back out the welfare effects of DI awards. A reader who is skeptical of the structural model's assumptions might still learn a lot about the welfare effects based on the descriptive evidence alone. For example, such a reader might update positively on the welfare effects to the extent that DI substantially increases consumption or update negatively to the extent that it crowds out other transfer income.

Similarly, Attanasio, Meghir, and Santiago (2012) presented treatment-control differences from a randomized evaluation of the PROGRESA conditional cash transfer that show how the program affected school enrollment of children in various age groups. They then estimated a dynamic model of the school enrollment decision that allows them to simulate alternative policies such as one that reallocates grant funding from younger children to older children. The observed treatment-control differences do not speak directly to the effect of this reallocation because it was not part of the original experiment. A reader who does not accept all of the assumptions of the structural model might nevertheless learn a fair amount about the likely effects of the reallocation from comparing the treatment effects on older and younger children.

Second, descriptive statistics $\hat{s}$ may provide evidence that helps readers evaluate the researcher's assumptions a_0. Allcott et al. (2019) estimated a structural model of grocery demand that allows them to decompose sources of nutritional inequality in the United States. To estimate price sensitivity, the authors instrumented for the price of a product in a given store with the price of the same product in other stores in the same chain. The exclusion restriction is that the variation in prices due to the composition of chains in a particular market is orthogonal to unobserved preference differences. In their descriptive analysis, the authors support the plausibility of this assumption by showing that this variation in prices is orthogonal to observed demographics that predict choices.

Agarwal et al. (2018) used an estimated structural model of bank lending to predict the extent to which credit expansions are passed on to borrowers. A key assumption of the model is that borrowers' unobserved characteristics are smooth around a set of credit score thresholds where credit limits change discontinuously. The authors' descriptive analysis confirms the "first stage" effect of the discontinuities on credit limits and then shows that observed borrower characteristics are smooth around the discontinuities, increasing the plausibility of the assumption that unobserved characteristics are smooth as well.

An important strength of descriptive analysis is that it permits a wider range of robustness and sensitivity analysis than is typically possible for computationally demanding structural estimates. Considering many alternative sets of controls, isolating variation along discontinuities, or adding highly saturated fixed effects are often not possible in complex models. Performing such checks is typically easier for the statistics reported in a descriptive analysis, and reporting them can strengthen confidence in model assumptions.

Descriptive statistics $\hat{s}$ can improve a reader's ability to evaluate the researcher's model even if they do not directly test its formal assumptions. For example, if an important assumption in the model is that a jurisdiction-level policy variable is assigned independently of unobservables, providing a map illustrating the spatial distribution of the policy can be very helpful to a reader.[8] Because many readers will have prior beliefs on the spatial distribution of unobservables, such a map can complement more formal balance tests that evaluate the correlation of the policy variable with observable characteristics of the jurisdiction. In a similar way, many types of summary statistics and data visualization can help to sharpen readers' priors on the researcher's assumptions and thus aid readers' interpretation of the estimator $\hat{c}$, in the sense that $\text{corr}_r\left(c, \hat{c}|\hat{s}\right)$ is much larger than $\text{corr}_r\left(c, \hat{c}\right)$ for some realizations of $\hat{s}$.

4. IDENTIFICATION

A second element that can contribute to transparent research is explicit discussion of identification. Such discussion is now common in much empirical research including structural research. Angrist and Pischke (2010) called "a conceptual framework that highlights specific sources of variation" one of the "hallmark[s] of contemporary applied microeconomics" (p. 12). Kleven (2018) showed that the share of NBER working papers in public economics discussing identification has risen from roughly 0% in 1980 to almost 50% today. Of the 123 structural articles published in the *American Economic Review*, *Econometrica*, the *Quarterly Journal of Economics,* and the *Journal of Political Economy* between January 2018 and November 2019, 80% included explicit discussion of identification.[9]

Formally, a quantity c is *identified* in the researcher's model if $c\left(a_0, \eta\right) \neq c\left(a_0, \eta'\right)$ implies $F\left(a_0, \eta\right) \neq F\left(a_0, \eta'\right)$ (Matzkin 2013; Lewbel 2019). In other words, distinct values of c correspond to distinct distributions of the data under the researcher's maintained assumptions. A quantity c is *identified by* a specific vector of statistics $\hat{s}$ if $c\left(a_0, \eta\right) \neq c\left(a_0, \eta'\right)$ implies distinct distributions of $\hat{s}$ under $F\left(a_0, \eta\right)$ and $F\left(a_0, \eta'\right)$.

There is a disconnect between this formal econometric definition and the discussions of identification that appear in some empirical articles. Keane (2010, p. 6) wrote,

> What is meant by "identified" [by some authors] is subtly different from the traditional use of the term in econometric theory.... Here, the phrase "how a parameter is identified" refers... to a more intuitive notion that can be roughly phrased as follows: What are the key features of the data, or the key sources of (assumed) exogenous variation in the data, or the key a priori theoretical or statistical assumptions imposed in the estimation, that drive the quantitative values of the parameter estimates, and strongly influence the substantive conclusions drawn from the estimation exercise?

There are two important differences between the formal definition of identification and the "intuitive notion" Keane (2010) described. First, point identification is formally a binary property. A quantity of interest c either is or is not identified by a

[8] See, for example, Fetter and Lockwood (2018, Figure 3), Bernard, Moxnes, and Saito (2019, Figure 8), and Hackmann (2019, Figure 3).

[9] Here, we define "structural" broadly to include any article that explicitly estimates the parameters of an economic model.

statistic $\hat{s}$. It is not clear in what meaningful sense a particular feature or source of exogenous variation could be the "key" source of identification. Second, identification is a property of a model, not a property of an estimator. Whether or not c is identified by $\hat{s}$ need not be related to whether or not $\hat{s}$ "drive[s] the quantitative values of the parameter estimates." Indeed, it is possible that c is identified by $\hat{s}$ yet the estimator $\hat{c}$ does not depend on $\hat{s}$ at all.

Many discussions of identification in recent structural work fit Keane's (2010) description. A number refer to particular quantities as "primarily," "mainly," or "largely" identified by particular data features.[10] Some focus on properties of estimators rather than models, using "identification" as essentially a synonym for "estimation."[11] Many acknowledge that they are departing from formal statements by saying they discuss identification "intuitively" or "loosely," or by noting explicitly that they discuss relationships of individual parameters to specific statistics even though all parameter estimates are determined jointly.[12]

We believe that clear and precise discussions of identification have an important role to play in making structural research transparent. In our framework, such discussions can be understood as a way to communicate and clarify the implications of the baseline assumptions a_0 and the space of relevant alternatives $a \neq a_0$, allowing readers to form more precise priors π_r. Focusing on partial identification, Tamer (2010) wrote: "[The partial identification approach] links conclusions drawn from various empirical models to sets of assumptions made in a transparent way. It allows researchers to examine the informational content of their assumptions and their impacts on the inferences made." We believe clear discussions of point identification can likewise increase transparency.

Such clarifying discussions would of course be unnecessary if readers could fully evaluate all of a model's assumptions directly. In reality, doing so is difficult. The abstract mathematical space in which assumptions are stated is often not one in which readers have well-formed intuitions. An assumption that sounds innocuous may in fact be highly restrictive, while another that sounds obviously unrealistic may in fact be a reasonable approximation. Identification discussions can illuminate the way the model's assumptions map the distribution of observables to the key quantity c. This is often a valuable complement to direct inspection of the assumptions in mathematical terms.

To illustrate what we mean, suppose a discrete-choice demand model assumes that the utility of a consumer i for a good j contains an additive error ε_{ij} which is iid Type I extreme value. How would a reader unfamiliar with such models evaluate the distributional assumption on the error term? Mathematically, the assumption is that the CDF of ε_{ij} is $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. Plotting the implied CDF or PDF would show that this is a single-peaked distribution not too different from a normal. It seems challenging to judge by introspection whether either the formula or the plot is a reasonable representation of the distribution of consumer utility, or under what circumstances it would be a better or worse approximation.

Studying the implications of the extreme value assumption for identification turns out to be instructive. As is now well understood (e.g., Anderson, De Palma, and Thisse 1992), imposing this form on the errors can mean that the share of consumers choosing each good j is alone sufficient to identify: (i) the relative own-price elasticities and markups of any two goods j and k; (ii) how consumers reallocate if any good is removed from the choice set; (iii) relative consumer welfare under different choice sets. An unfamiliar reader who learned these implications might update in the direction of thinking the distributional assumption is stronger than they thought and worth additional scrutiny. The reader might also be able to form new intuitions about what alternative assumptions a are most relevant to consider—for example, alternative error distributions that decouple substitution patterns from market shares (Berry, Levinsohn, and Pakes 1995).

We suggest that two principles should guide discussions of identification. First, these discussions should be precise, with the verb "identify" used only in its formal econometric sense. It is best to avoid quantitative modifiers like "primarily identifies" or "mainly identifies" with uncertain meaning. Like any other theoretical statement, statements about identification that are not immediately obvious should either be accompanied by formal proof or introduced explicitly as conjectures.

The statement that a quantity c "is identified by" a particular vector of statistics $\hat{s}$ should mean that the distribution of $\hat{s}$ is sufficient to infer the value of c under the model. If this statement applies only given knowledge of some other parameters, then this should be made explicit. Looking over cases in the recent literature where authors claim something is "identified by" specific features of the data, one sees three common structures of argument.[13] The first structure is to prove identification as a formal proposition.[14] The second structure is an informal "triangular" argument. In the case where the object of interest is a parameter vector η, this might show that η_1 is identified by a statistic $\hat{s}_1$ alone, η_2 is identified by $\hat{s}_2$ once the value of η_1 is known, η_3 is identified by $\hat{s}_3$ once the values of

[10] For example, Beraja et al. (2018) wrote, "Any empirical measure of refinancing elasticities to interest rate reductions will always be *primarily identified* from recession periods" (p. 156, emphasis added). Fu and Gregory (2019) wrote, "The dispersion… is thus *identified mainly* from the size of RDD parameter" (p. 407, emphasis added). Crawford et al. (2018) wrote, "The procompetitive effects of vertical integration are *largely identified* from the degree to which RSN carriage is higher for integrated distributors" (p. 893, emphasis added).

[11] See, for example, the subsection titled "Identification Strategy" in Harasztosi and Lindner (2019, p. 2701), the section titled "Identification of Structural Parameters" in Head and Mayer (2019, p. 3095), and the section titled "Estimation and Identification" in Hackmann (2019, p. 1702).

[12] For example, Allcott et al. (2019) wrote, "*Loosely*, the first set of moments identify the β parameters…" (p. 1827, emphasis added). Autor et al. (2019) wrote, "While the mapping between model parameters and sample moments is less direct for the disutility parameters, there are data moments that *intuitively* provide identifying information. While all parameters are estimated simultaneously, it can be instructive to focus on one parameter at a time" (p. 2644, emphasis added). Fu and Gregory (2019) wrote, "Although all of the structural parameters are identified jointly, we provide a sketch of identification here by describing which auxiliary models are most informative about certain structural parameters" (p. 407).

[13] See also Gentzkow, Shapiro, and Sinkinson (2014, secs. V.A and VI.A).

[14] See, for example, Agarwal and Somaini (2018), Bonhomme, Lamadon, and Manresa (2019), and Chiappori et al. (2019).

η_1 and η_2 are known, and so on.[15] When formalized, this can of course be a valid method of proof that η is identified by $\hat{s}$. The final structure is an elementwise argument where saying that η_j is identified by $\hat{s}_k$ means this is true given that all other elements of η are known. Many "heuristic" or "informal" discussions of identification seem to take this form, though sometimes without making explicit the requirement that the other elements of η are known.[16] Enumerating such relationships for all η_j does not establish identification of η from $\hat{s}$. In many models, the statement that η_j is identified by $\hat{s}_k$ in this sense will be true for many different statistics $\hat{s}_k$. Such discussions may nevertheless provide some useful intuition about the model.

Some authors support discussions of identification with simulations showing how the distributions of some statistics $\hat{s}$ change when each parameter is varied in turn, holding all other parameters constant.[17] A statistic is then sometimes said to "identify a parameter" if the distribution of the statistic responds strongly as the parameter varies. Note that this amounts to a version of the third argument structure above, establishing elementwise relationships that do not imply formal identification of the model as a whole. We see this kind of simulation as valuable provided it is clear that it speaks to identification of the parameter of interest only if the values of the other parameters are known.

The second principle we would recommend is that discussion of model identification be clearly distinguished from discussion of estimation. How $\hat{s}$ and c are related under the model is distinct from how $\hat{s}$ is related to the specific estimator $\hat{c}$, and the statement "c is identified by $\hat{s}_j$" need not imply that $\hat{s}_j$ is an important determinant of $\hat{c}$. How to clarify the data features that actually do drive $\hat{c}$ is the topic we take up in Section 5. As we note there, transparency is often improved when the discussion of identification elucidates the same relationships that turn out to be important in estimation.

In some cases, discussion of the construction of an estimator can itself constitute a heuristic proof of identification. For example, it may be that estimation consists of a series of plug-in or linear estimators for parameters whose identification is well understood.[18] Such cases may explain how "identification strategy" has come to be used in some of the literature as a synonym for "estimation strategy."[19]

5. ESTIMATION

Descriptive analysis and discussion of identification can together help readers understand how the researcher's model maps features of the data to conclusions about c, and assess the validity of the assumptions underlying this mapping. Research is most transparent when readers can use this and other information to interpret the structural estimates $\hat{c}$ taking account of the forms of misspecification they find most relevant.

To do so, the reader needs to understand how the specific estimator $\hat{c}$ depends on the data D. There are often many distinct vectors of intuitive statistics $\hat{s}$ that each identify c under the researcher's model, and in over-identified settings there are many different transformations of a given vector $\hat{s}$ that estimate c. Identification discussions can at best clarify the sets of possible statistics and transformations.

Knowing the form of the estimator $\hat{c}$ is essential to transparency for two reasons. First, for reasons related to the discussion in Section 4, the statistics $\hat{s}$ on which the estimator depends—the statistics that "drive" the estimator in common parlance—influence which violations of assumptions matter most. Second, as we elaborate below, knowing *how* the estimator depends on these statistics can allow a reader to judge the likely bias induced by specific violations.

In this section, we consider how to make estimation more transparent. We focus on the value of both highlighting a specific vector of statistics $\hat{s}$ that determine the estimator $\hat{c}$ either exactly or approximately, and making the form of the relationship between the two clear to the reader. To fix ideas, without loss of generality we can write

$$\hat{c} = h\left(\hat{s}\right) + \nu_h,$$

where $h(\cdot)$ is some function and ν_h is a residual whose structure depends on $h(\cdot)$. The first approach we discuss is to choose an estimator $\hat{c}$ such that $\nu_h = 0$ and then characterize the form of the function $h(\cdot)$. The second approach we discuss is to choose an estimator $\hat{c}$ such that $\nu_h \neq 0$ and then demonstrate that ν_h is small in an appropriate sense so that $\hat{c} \approx h\left(\hat{s}\right)$. In Section 6, we discuss how a reader can assess specific forms of misspecification in the context of such estimators.

5.1. Target Descriptive Statistics in Estimation

The first approach is to target $\hat{s}$ directly in estimation, so that $\hat{c} = h\left(\hat{s}\right)$. This is of course only sensible when c is identified by $\hat{s}$. If c is identified by the population value s of $\hat{s}$ and the implied relationship $c = \Gamma(s, a)$ does not vary across the alternatives a of interest (i.e., $c = \Gamma(s)$ for all $a \in \mathscr{A}$), the plug-in estimator $\hat{c} = h\left(\hat{s}\right) = \Gamma\left(\hat{s}\right)$ may have high transparency for all readers. Related ideas appear in the literature as a justification for basing model estimation and testing on matching certain statistics of interest. (See, e.g., Dridi, Guay, and Renault 2007; DellaVigna 2018; Nakamura and Steinsson 2018.) Even when $\Gamma(s, a)$ depends on a, an estimator of the form $\hat{c} = h\left(\hat{s}\right) = \Gamma\left(\hat{s}, a_0\right)$ may still be reasonably transparent if the form of the relationship $c = \Gamma(s, a_0)$ is made clear.

In practice, estimation based on targeting a vector of descriptive statistics $\hat{s}$ is often implemented via some form of minimum distance estimation that chooses parameters to match the observed $\hat{s}$ to the value predicted under the model. Transparency provides a potential justification for choosing such estimators even when more efficient estimators, such as the MLE, are available.

The literature contains numerous examples of estimators that target descriptive statistics. Gourinchas and Parker (2002)

[15] See, for example, Eckstein, Keane, and Lifshitz (2019, pp. 235–236), Fréchette, Lizzeri, and Salz (2019, p. 2976), and Fu and Gregory (2019, pp. 407–409).

[16] Two articles that make this requirement explicit are Autor et al. (2019, pp. 2644–2645) and David and Venkateswaran (2019, pp. 2548–2549).

[17] See, for example, Autor et al. (2019, Figure 4) and David and Venkateswaran (2019, sec. III.C).

[18] See, for example, the section titled "Identification of Structural Parameters" in Head and Mayer (2019, p. 3095).

[19] The "Identification Strategy" section in Harasztosi and Lindner (2019, p. 2701) is a recent example.

estimated a lifecycle model of consumption and savings with a precautionary motive. After estimating properties of the income process in a first step, Gourinchas and Parker (2002) estimated structural preference parameters in a second step by minimizing the distance between the observed age profile of consumption and the profile predicted by their economic model. De Nardi, French, and Jones (2010) likewise estimated a model of consumption and savings by retirees by targeting observed median asset profiles for different groups of individuals. See also Goettler and Gordon (2011), DellaVigna, List, and Malmendier (2012), Nikolov and Whited (2014), and Autor et al. (2019).

These examples have in common that the estimator is (at least asymptotically) a function of descriptive statistics that a reader might find intuitively related to the parameters of interest. Part of the reason for this choice of estimator may be computational, for example, due to the difficulty of computing the likelihood. However, in some cases the authors invoke non-computational considerations in justifying their choice of moments to match, some of which seem related to the considerations we discuss here.[20]

To formalize the value of targeting descriptive statistics, we consider a variant of the instrumental variables example introduced in Section 2.2.

Example. Suppose that the underlying data consist of n iid draws $\left\{\left(Y_i, X_i, Z_{i,1}, \dots Z_{i,J}\right)\right\}_{i=1}^{n}$ for $Z_i = \left(Z_{i,1}, \dots Z_{i,J}\right)$ a vector of J mutually orthogonal and mean-zero instruments proposed by the researcher. For $c \in \mathbb{R}$ and $a \in \mathscr{A} = \mathbb{R}^J$, the data follow

$$Y_i = X_i c + Z_i' a + \varepsilon_i,$$

where we now treat the instruments Z_i as random and allow the error ε_i to be nonnormal. Let G denote the joint distribution of $(Z_i, X_i, \varepsilon_i)$ and assume all readers believe that $G \in \mathscr{G}$ for some class of distributions with $E_G[Z_i \varepsilon_i] = 0$ for all $G \in \mathscr{G}$.

The instruments are valid under the researcher's assumption $a_0 = 0$, but readers suspect they may in fact be invalid. Suppose that each instrument j has a nonzero first-stage coefficient, $E_G\left[Z_{i,j} X_i\right] \neq 0$ for all $G \in \mathscr{G}$. Under distribution G, true parameter value c, and assumptions a, the probability limit of the instrumental variables estimator based on the jth instrument alone, $\hat{c}_j = \sum Z_{i,j} Y_i / \sum Z_{i,j} X_i$, is

$$\frac{E_{(G,a,c)}\left[Z_{i,j} Y_i\right]}{E_G\left[Z_{i,j} X_i\right]} = c + \frac{E_G\left[Z_{i,j}^2\right]}{E_G\left[Z_{i,j} X_i\right]} a = c + \frac{a_j}{\gamma_j},$$

for $\gamma_j = E_G\left[Z_{i,j} X_i\right] / E_G\left[Z_{i,j}^2\right]$ the first stage coefficient on the jth instrument.

To illustrate the value of targeting descriptive statistics in this example, let the descriptive statistics $\hat{s}$ consist of the first m single-instrument coefficients $\hat{s} = \left(\hat{c}_1, \dots, \hat{c}_m\right)$ for $m < J$. We suppose that readers have sharp priors on the bias in these estimates, in the sense that reader r believes the first m elements of the bias vector $b = (a_1/\gamma_1, \dots, a_J/\gamma_J)'$ equal a known vector b_r with probability one, $Pr_r\left\{(a_1/\gamma_1, \dots, a_m/\gamma_m)' = b_r\right\} = 1$. Thus, all readers are certain about the bias from using the first m instruments, while they may be uncertain about the remaining instruments. This could be because the potential biases from the first m instruments are especially intuitive, for instance, because these instruments are highly credible and $b_r = 0$, or because the researcher has clarified the potential biases, for example, through descriptive analysis and discussion of identification.

In this case, an estimator targeting the descriptive statistics $\hat{s}$ may be more transparent than the maximum-likelihood estimator under the researcher's assumption $a_0 = 0$. To provide a concise illustration, suppose the sample size is large enough that $\hat{c}$ is approximately normal and neglect the approximation error to obtain

$$\hat{c} = \iota c + b + \xi, \ \xi \sim N(0, \Omega), \tag{3}$$

for ι the vector of ones. In this asymptotic model $\eta = (c, \gamma, \Omega)$. Suppose further that the researcher observes only $D = \left(\hat{c}, \Omega\right)$, that c and (b, Ω) are independent under π_r for all $r \in \mathscr{R}$, and that Ω is commonly known.

We let $\hat{c}_0 = \left(\iota' \Omega^{-1} \iota\right)^{-1} \iota' \Omega^{-1} \hat{c}$ denote the maximum-likelihood estimator under the assumption $a_0 = 0$, and we let $\hat{c}_S$ denote the estimator that efficiently minimizes the distance between Sc and $\hat{s} = S\hat{c}$ for S the selection matrix such that $S\hat{c}$ picks out the first m elements of $\hat{c}$. The variance of $\hat{c}_0$ given c under π_r is

$$\mathrm{var}_r\left(\hat{c}_0 | c\right) = \left(\iota' \Omega^{-1} \iota\right)^{-1} + \left(\iota' \Omega^{-1} \iota\right)^{-2} \iota' \Omega^{-1} \mathrm{var}_r(b)\, \Omega^{-1} \iota$$

where the first term is the sampling variance of the MLE and the second term reflects instrument invalidity. By contrast, the variance of $\hat{c}_S$ given c under π_r is simply the sampling variance of $\hat{c}_S$. When reader r is very uncertain about instrument validity (in the sense that the variance of the last $J - m$ elements of b is large), $\hat{c}_S$ may be more transparent than $\hat{c}_0$.[21] This is intuitive in the case where $b_r = 0$, so reader r believes that the first m instruments are valid. Note, however, that it remains true even when $b_r \neq 0$. Hence, what is important for transparency in this setting is not that the first m instruments are valid, but that readers have precise beliefs about the bias these instruments induce.

While we have motivated (3) as an asymptotic approximation to over-identified instrumental variables regression with potentially invalid instruments, it is equivalent to some other important problems. For instance, (3) can be interpreted as a regression model for $Y = \hat{c}$ with omitted variable b. As discussed in Armstrong and Kolesár (2019), this model can also be understood as an asymptotic approximation to GMM under local misspecification.

Knowing that the estimator has the form $\hat{c} = h\left(\hat{s}\right)$ means that readers know the estimator depends on the data only

[20] For example, De Nardi, French, and Jones (2010) wrote that "Because our underlying motivations are to explain why elderly individuals retain so many assets and to explain why individuals with high income save at a higher rate, we match median assets by cohort, age, and permanent income" (p. 47). Nikolov and Whited (2014) wrote that "The success of [the approach to estimation] relies on model identification, which requires that we choose moments that are sensitive to variations in the structural parameters ... We now describe and rationalize the ... moments that we match" (p. 1899). Autor et al. (2019) included certain moments "...to discipline the model to recover our estimates of the causal effects of" a policy variable of interest (p. 2645).

[21] For such a reader, if we hold $\mathrm{var}_r(c)$ fixed and take $\mathrm{var}_r\left(\hat{c}_0 | c\right) \to \infty$ by taking $\mathrm{var}_r(b) \to \infty$, we have that $\mathrm{var}_r\left(c | \hat{c}_0\right) \to \mathrm{var}_r(c)$.

through the statistics $\hat{s}$, but they may not know the nature of the dependence. As noted above, there are often many different functions $h(\cdot)$ that constitute valid estimators under the model—in particular, whenever different subsets or transformations of s are each sufficient to identify c. Some of these functions $h(\cdot)$ may be convincing to a large set of readers, in the sense that $\Gamma(s,a) \approx h(s)$ for many $a \in \mathscr{A}$, while others may only be convincing to readers who accept a_0. In such cases, communicating the geometry of $h(\cdot)$ can help readers evaluate the estimator and so improve transparency.

In practice, research articles typically provide a formal definition of the estimator. Even a precise definition may not make the geometry of $h(\cdot)$ obvious, however. In linear models, for instance, recent articles characterized regression discontinuity estimators (Gelman and Imbens 2019) and two-way fixed effect estimators (e.g., Athey and Imbens 2018; Goodman-Bacon 2019; Sun and Abraham 2020; de Chaisemartin and D'Haultfoeuille 2020; Imai and Kim 2020) and in some cases argued that these estimators use the data in ways that may be unanticipated and undesired by many readers and researchers.

In nonlinear models, such characterizations may be even more difficult to come by and therefore even less obvious *ex ante*. One solution could be to fully describe $h(\cdot)$ by brute-force enumeration, but this is often infeasible. For example, Gourinchas and Parker (2002) summarized the age profile of consumption with the mean adjusted log consumption at each of the 40 ages from 26 through 55. As even a single estimation step may be computationally demanding, computing and visualizing Gourinchas and Parker's (2002) estimator on a 40-dimensional domain seems daunting.

Andrews, Gentzkow, and Shapiro (2017) proposed to focus on the local *sensitivity* of the estimator to the statistics targeted in estimation. Sensitivity corresponds (in a sense made precise in Andrews, Gentzkow, and Shapiro 2017) to the derivatives of $h(\cdot)$ when $h(\cdot)$ is differentiable. It is possible to approximate this derivative numerically, for example, by evaluating the estimator at perturbations of the form $\hat{s} + \epsilon e_j$ for ϵ a small number and e_j the jth standard basis vector, and then computing the numerical derivative $\left(h\left(\hat{s} + \epsilon e_j\right) - h\left(\hat{s}\right)\right)/\epsilon$. Repeated estimation may be computationally demanding, but Andrews, Gentzkow, and Shapiro (2017) showed that in many applications (including Gourinchas and Parker 2002) repeated estimation is unnecessary if the reader is willing to focus on the asymptotic value of the derivative.

Andrews, Gentzkow, and Shapiro (2017) plotted the local sensitivity of the estimators of key structural parameters in Gourinchas and Parker (2002) with respect to the statistics $\hat{s}$ targeted in estimation. They argue that the local properties of $h(\cdot)$ revealed by this exercise make qualitative sense in light of the economic analysis and discussion in Gourinchas and Parker (2002), and that knowledge of sensitivity could be useful to a reader who wishes to learn from $\hat{c}$ but is concerned about misspecification of the assumptions a_0.

Example (continued). Continue to suppose that the first m elements of b are known under π_r, but now suppose that Ω may not be commonly known. The estimator $\hat{c}_S$ can be written as

$$\hat{c}_S = \left(\iota' S' \left(S\Omega S'\right)^{-1} S\iota\right)^{-1} \iota' S' \left(S\Omega S'\right)^{-1} \hat{s} = \Lambda_S \hat{s}$$

for $\hat{s} = S\hat{c}$, where Λ_S is the sensitivity of $\hat{c}_S$ to $\hat{s}$ as defined by Andrews, Gentzkow, and Shapiro (2017). Reporting $\left(\hat{c}_S, \sigma_S, \Lambda_S\right)$—that is, taking $\hat{c} = \hat{c}_S$ and $\hat{t} = (\sigma_S, \Lambda_S)$ for σ_S the standard error of $\hat{c}_S$—is weakly more transparent for all readers r than reporting $\left(\hat{c}_S, \sigma_S\right)$ alone. To see that it may be strictly more transparent, suppose that reader r has a normal prior on c, $c \sim N\left(0, \omega_r^2\right)$. The average posterior variance for reader r based on the full data is then bounded below by $E_r\left[\mathrm{var}_r\left(c|D,b\right)\right] = E_r\left[\left(\omega_r^{-2} + \sigma_0^{-2}\right)^{-1}\right]$ for σ_0 the usual standard error of $\hat{c}_0$, while the average posterior variance based on observing $\left(\hat{c}_S, \sigma_S, \Lambda_S\right)$ is $E_r\left[\left(\omega_r^{-2} + \sigma_S^{-2}\right)^{-1}\right]$. This bounds the transparency of reporting $\left(\hat{c}_S, \sigma_S, \Lambda_S\right)$ from below. By contrast, the transparency of $\left(\hat{c}_S, \sigma_S\right)$ alone may be small when b_r is large. Consider, for instance, priors which imply that σ_S is fixed while $\Lambda_S b_r$ is uniformly distributed on some interval. The transparency of $\left(\hat{c}_S, \sigma_S\right)$ goes to zero as $\mathrm{var}_r\left(\Lambda_S b_r\right) \to \infty$. Intuitively, even if the reader knows the bias b_r of estimates based on the first m instruments, to infer the bias of $\hat{c}_S$ they must also know how these m estimates are combined to form $\hat{c}_S$. Absent such knowledge, uncertainty about how the bias in $\hat{s}$ translates to bias in $\hat{c}_S$ renders $\hat{c}_S$ uninformative when b_r is large.

5.2. Show How Much the Estimator Depends on the Descriptive Statistics

Basing estimation directly on $\hat{s}$ may not be feasible or desirable. For example, it may be that even though $\hat{s}$ are intuitive statistics closely related to c, their distribution is not sufficient for identification. It may be that $\hat{s}$ identifies c but that identification using these statistics alone is weak. Or, it may be that the share of readers who accept the researcher's exact parametric assumptions is large enough that the efficiency gain for these readers from learning the MLE outweighs the loss of transparency for those who are more skeptical.

In these cases, $\hat{c} = h\left(\hat{s}\right) + \nu_h$ for ν_h not necessarily equal to zero. Then, making clear to readers the magnitude of ν_h as well as the form of $h(\cdot)$ can improve transparency. An example is where at least some readers believe that $c = \Gamma(s,a)$, in which case they may find $\hat{c}$ especially informative when $\nu_h \approx 0$ and $h(\cdot) \approx \Gamma(\cdot, a)$.

Characterizing the finite-sample relationship between $\hat{c}$ and $\hat{s}$, either analytically or numerically, can be difficult. For example, numerical exploration by repeatedly drawing data D from one or more data-generating processes and then computing the implied values of $\hat{c}$ and $\hat{s}$ may be very computationally demanding.

Andrews, Gentzkow, and Shapiro (2020) showed that, under asymptotic conditions related to those considered in Andrews, Gentzkow, and Shapiro (2017), many common estimators can be represented in the form

$$\hat{c} \approx \text{constant} + \Lambda\hat{s} + \nu$$

for Λ an analogue of the local sensitivity defined in Andrews, Gentzkow, and Shapiro (2017), and ν asymptotically uncorrelated with $\hat{s}$. Andrews, Gentzkow, and Shapiro (2020) proposed

to measure the size of ν by the *local informativeness* of $\hat{s}$ for $\hat{c}$, which is given by

$$\Delta = \frac{\text{Avar}\left(\Lambda\hat{s}\right)}{\text{Avar}\left(\hat{c}\right)} = 1 - \frac{\text{Avar}\left(\nu\right)}{\text{Avar}\left(\hat{c}\right)}$$

for $\text{Avar}\left(V\right)$ the asymptotic variance of a random variable V. When local informativeness $\Delta = 1$, Andrews, Gentzkow, and Shapiro (2020) showed that (under the maintained conditions) $\nu = 0$ and the setting collapses to that considered in Section 5.1. When local informativeness $\Delta = 0$, $\hat{c}$ is asymptotically independent of $\hat{s}$, in which case a reader believing that $c = \Gamma\left(s, a\right)$ may not find $\hat{c}$ to be very informative about c.

Andrews, Gentzkow, and Shapiro (2020) showed that it is often possible to approximate local sensitivity and local informativeness without the need for additional simulation or estimation of the structural model. Moreover, although both local sensitivity and local informativeness can depend on the data-generating process, Andrews, Gentzkow, and Shapiro (2017, 2020) showed that the approximations they consider hold under local violations of the researcher's assumptions, meaning that these objects can be interpreted even if the reader does not have complete confidence in the researcher's model. Andrews, Gentzkow, and Shapiro (2020) also established conditions under which a greater informativeness Δ corresponds to a larger reduction in the worst-case bias of the estimator $\hat{c}$ from accepting the model-implied relationship between c and the population value of the descriptive statistics $\hat{s}$.

Andrews, Gentzkow, and Shapiro (2020) applied their framework to Hendren's (2013) study of the market for long-term care insurance. They took $\hat{c}$ to be the maximum likelihood estimator for the minimum pooled price ratio, a quantity that determines the range of preferences for which insurance markets cannot exist, and $\hat{s}$ to be statistics summarizing the joint distribution of individuals' subjective beliefs about the likelihood of needing long-term care and their eventual need for such care. Andrews, Gentzkow, and Shapiro (2020) estimated that these descriptive statistics have an informativeness of 0.68 for the estimator $\hat{c}$, implying that the descriptive statistics can explain (in an R^2 sense) 68% of the variation in the estimator under the joint asymptotic distribution of the estimator and descriptive statistics, and that (under given conditions) accepting the model-implied relationship between the minimum pooled price ratio and the population value of the descriptive statistics reduces the worst-case bias of the estimator by a factor of $\sqrt{1 - 0.68} \approx 0.43$.

Example (continued). The asymptotic results of Andrews, Gentzkow, and Shapiro (2020) hold exactly in the linear instrumental variables example that we consider in this section. To illustrate the value of informativeness calculations, let us again suppose that under π_r the first m elements of b are known to equal b_r, and $c \sim N\left(0, \omega_r^2\right)$. Let us further suppose that reader r thinks the degree of misspecification is bounded relative to sampling uncertainty, in the sense that $Pr_r\left\{\sqrt{b'\Omega^{-1}b} < \mu^2\right\} = 1$ for some constant μ. In this case, one can show that for $\hat{c}_0$ again the maximum likelihood estimator, the increase in reader r's average posterior variance from observing $\left(\hat{c}_0, \sigma_0, \Lambda_0\right)$ for σ_0 the standard error of $\hat{c}_0$ and Λ_0 the sensitivity of $\hat{c}$ to $\hat{s}$, rather than the full data, is bounded above by

$$E_r\left[\left(\frac{\omega_r^2}{\omega_r^2 + \sigma_0^2}\right)^2 \sigma_0^2 \mu^2\left(b_r\right)\left(1 - \Delta\right)\right]$$

for $\mu\left(b_r\right) = \sqrt{\mu^2 - b_r'\left(S\Omega S'\right)^{-1} b_r}$.[22] Thus, when reader r is confident about the impact of misspecification on $\hat{s}$ and the informativeness Δ of $\hat{s}$ for $\hat{c}_0$ is high, observing $\hat{c}_0$ is nearly as good as observing the full data.

6. SENSITIVITY ANALYSIS

A key premise of the model in Section 2 is that different readers may accept different assumptions. Ideally the researcher would report the estimator that is optimal for each reader. When the set of assumptions entertained by the readers is small enough, this ideal may be achievable. This situation is one way to understand the sensitivity analysis that is common in research articles, showing how the key conclusions of the analysis change under a small set of assumptions different from those on which most of the analysis is based. When the set of assumptions entertained by the readers is rich, however, such an approach has limits, and it is desirable to help each reader assess how their own ideal estimator differs from the one reported by the researcher. In cases where a key conclusion of the researcher's analysis is qualitative, it may be useful, in addition, to report the properties of a data realization that would have led to a different qualitative conclusion.

6.1. Show the Conclusion Under Specific Alternative Assumptions

Suppose that under each set of assumptions $a \in \mathscr{A}$ there is a natural estimator $\hat{c}_a$ (e.g., maximum likelihood or efficient GMM). If the researcher knows that all (or many) readers entertain only a limited set of assumptions, in the sense that each prior π_r puts mass on a single $a \in \mathscr{A}$ and the number of distinct elements $|\mathscr{A}|$ is small, then it is natural to report the estimate $\hat{c}_a$ under each of element of $\mathscr{A}$.

For example, in their study of automobile demand Berry, Levinsohn, and Pakes (1995, Table IX) reported how a key conclusion—the markup associated with each of a set of vehicle models—changes under six different alternative models, each of which corresponds to a modification of the cost or utility function specified in the baseline model. A reader who believes in one of these alternative specifications a_j is therefore able to learn about c from an estimator that is asymptotically valid under a_j. Tables reporting estimates of key parameters of interest under alternative assumptions are a common feature of

[22] Reader r's average posterior variance from observing D is bounded below by that from observing both D and b. In the latter case, r's posterior mean is $E_r\left[c|D,b\right] = \frac{w_r^2}{w_r^2+\sigma_0^2}\left(\hat{c}_0 - \widetilde{\Lambda}b\right)$, for $\widetilde{\Lambda} = \left(\iota'\Omega^{-1}\iota\right)^{-1}\iota'\Omega^{-1}$ the sensitivity of $\hat{c}_0$ to $\hat{c}$. When r observes only $\left(\hat{c}_0, \sigma_0, \Lambda_0\right)$ they cannot construct $E_r\left[c|D,b\right]$, but can construct $c^* = \frac{w_r^2}{w_r^2+\sigma_0^2}\left(\hat{c}_0 - \Lambda_0 b_r\right)$, and the results of Andrews, Gentzkow, and Shapiro (2020) showed that if $\sqrt{b'\Omega^{-1}b} < \mu^2$, then $\left(\widetilde{\Lambda}b - \Lambda_0 b_r\right)^2 \leq \sigma_0^2\mu^2\left(b_r\right)\left(1 - \Delta\right)$.

many applied research articles (see, e.g., Gourinchas and Parker 2002, Table V).[23]

It is useful to contrast such a sensitivity analysis with a bounds analysis that reports the set of estimates $\{\hat{c}_a : a \in \mathscr{A}\}$ without specifying which estimate corresponds to which assumption (see, e.g., Leamer 1981). In some cases the mapping from assumptions to elements of the set of estimates may be obvious, at least for extreme points (e.g., the upper and lower bounds). In other cases, however, a bounds analysis can be less transparent than a sensitivity analysis with respect to the same set of assumptions. Similar considerations can apply to a partial identification-robust analysis that ensures validity under all $a \in \mathscr{A}$.[24] At the same time, bounds and identification-robust analyses often remain feasible with large sets of assumptions (again, see, e.g., Leamer 1981).

6.2. Show How the Conclusion Depends on Assumptions

If the set of assumptions $\mathscr{A}$ entertained by the readers is sufficiently rich, then reporting an estimator $\hat{c}_a$ associated with each assumption $a \in \mathscr{A}$ is no longer feasible. A possible alternative is to provide information about the (possibly random) function $u(a) = \hat{c}_a - \hat{c}_{a_0}$ that relates the estimator under the researcher's baseline assumption a_0 to the natural one under the reader's preferred assumption a. If all readers knew $u(\cdot)$, then each reader could adjust the baseline estimate $\hat{c}_{a_0}$ to reflect the reader's own preferred assumption a.

The omitted variables bias formula (OVBF) is perhaps the most famous tool for intuiting the properties of $u(\cdot)$. Given beliefs a about the covariance properties of an omitted regressor, the OVBF allows a reader to determine the bias in the estimator of a given coefficient resulting from the exclusion of that regressor, which might correspond to a researcher's baseline assumption a_0. The OVBF thus avoids the need to enumerate the bias implied by all possible beliefs about omitted regressors. Conley, Hansen, and Rossi (2012) generalized the OVBF to an instrumental variables setting, showing how to translate beliefs about violations of the exclusion restriction to beliefs about bias in the IV estimator. Like the OVBF, Conley, Hansen, and Rossi's (2012) approach allows different readers to reach different conclusions regarding the appropriate adjustments to the reported estimator.

Andrews, Gentzkow, and Shapiro (2017) studied the problem of translating these ideas to nonlinear structural models, where the OVBF does not apply directly. In the wide class of models that can be estimated via minimum distance, the identifying assumptions can be represented as restrictions on the population value a of a moment condition under the true value of the structural parameters. For example, in nonlinear instrumental variables estimators such as that in Berry, Levinsohn, and Pakes's (1995) study of automobile demand, the restriction is that a vector of observed instruments is orthogonal in population to a vector of unobserved structural errors. In indirect inference and other moment-matching type estimators, such as that in Gourinchas and Parker (2002), the restriction is that the population values of some statistics must match those predicted by the model.

In such settings, specific violations of the identifying assumptions can be represented as specific alternative restrictions—for example, that the covariance between the instruments and the structural error takes some specific nonzero value in population, or that the model systematically mispredicts the population value of some statistics by some specific amount. For linear models, the OVBF and its analogues tell the reader how to adjust the estimator to accommodate such perturbations to the researcher's identifying assumptions. For nonlinear models, we are not aware of a similar formula, and exhaustively checking the implications of each possible perturbation can be costly or impossible.

Andrews, Gentzkow, and Shapiro (2017) showed that if the perturbations are local—that is, small in an appropriate asymptotic sense—then the implied asymptotic bias of the estimator is given by $\Lambda(a - a_0)$, where the coefficients Λ are the local sensitivity discussed in Section 5.1, now replacing the descriptive statistics with the vector of estimation moments evaluated at the true parameter value. It is thus practical to approximate and report the coefficients of the asymptotic bias formula in many applications. In this sense, local sensitivity provides an analogue of the OVBF for general minimum distance estimators under small violations of identifying assumptions.

Andrews, Gentzkow, and Shapiro (2017) reported the local sensitivity of the estimated average vehicle markup to violations of the identifying assumptions in Berry, Levinsohn, and Pakes (1995). This analysis shows that the estimator is especially sensitive to violations of the assumption that unobserved shocks to the utility from or cost of producing a given vehicle model are orthogonal to the number of other models offered by the same or rival firms. Berry, Levinsohn, and Pakes (1995, p. 854) discussed the economic interpretation of these and other identifying assumptions. Andrews, Gentzkow, and Shapiro (2017) showed how a reader could use information on sensitivity to approximate the effect of different economically interesting violations of the identifying assumptions. Importantly, this approach does not require the researcher to know the alternative assumptions a of interest in advance, as readers can use information about sensitivity to calculate the implications of different assumptions a for themselves. Andrews, Gentzkow, and Shapiro (2017) reported a similar analysis of the sensitivity of the estimated preference parameters in Gourinchas and Parker (2002) to violations of the identifying assumptions.

[23] We focus on cases where c is point-identified under each $a \in \mathscr{A}$. When point identification may fail, the researcher can report an estimate of the identified set under each $a \in \mathscr{A}$. If the assumptions in $\mathscr{A}$ can be ranked in increasing order of strength, this approach allows the reader to see how conclusions sharpen with each incremental strengthening of the assumptions. See the discussion in, for example, Manski (2003, 2007) and Tamer (2010).

[24] To give an extreme example, consider the instrumental variables model discussed in Section 2.2. The maximum likelihood estimator for c under assumption a in this setting is $\hat{c} - a/\hat{\gamma}$, for $\hat{c}$ again the usual instrumental variables estimator, so the set of maximum likelihood estimators under $a \in \mathscr{A}$ is equal to $\mathbb{R}$ almost surely, and thus has transparency equal to zero for readers r who find the full data informative. Correspondingly, the identified set for c under $\mathscr{A}$ is equal to $\mathbb{R}$, and therefore any confidence set with coverage $1 - \alpha$ for c under all $a \in \mathscr{A}$ and $\eta \in H$ must have infinite length with probability at least $1 - \alpha$. See Dufour (1997).

6.3. Show How to Reverse the Conclusion

Some of the questions answered by structural analysis are qualitative—for example, will a given policy increase or decrease consumer surplus? Will a merger increase or decrease product quality? Empirical answers to such questions depend, by definition, on the realization of the data. To characterize this dependence, it can be helpful for researchers to discuss data realizations that would have led to the opposite conclusion. In some cases the properties of the data required for such a reversal are obvious. For example, if the effect of some policy on an outcome is estimated in a multivariate linear regression, then to reverse the researcher's conclusion about whether the policy increases the outcome requires changing the sign of the residual covariance between the policy and the outcome.

For estimators in nonlinear models, by contrast, it is sometimes not obvious what realizations of the data would lead to conclusions different from the one reached by the researcher's analysis, or even whether such realizations exist. We think that exhibiting such realizations can improve transparency, both by showing that the researcher's answer to the qualitative question is indeed an empirical one, and (in the spirit of Section 5.1) showing what sort of data realization is associated with a given conclusion.

Such an exercise might be called a *reverse sensitivity analysis*: rather than changing the inputs (e.g., data or assumptions) and investigating the effects on the outputs (conclusions), as in a traditional sensitivity analysis, here we seek to change the outputs and reverse engineer sufficient changes to the inputs. In this section, we focus on describing the required changes in the data or parameters. A complementary approach characterizes the required change in assumptions (see, e.g., Horowitz and Manski 1995; Kline and Santos 2013; Masten and Poirier 2020).

Goettler and Gordon (2011) studied whether Intel is more innovative in the production of microprocessors as a result of competition from AMD. To answer this question they estimated a dynamic model of the microprocessor industry and simulate behavior under alternative market structures. Goettler and Gordon (2011) concluded that the presence of AMD reduces the rate at which Intel innovates. They observed that their model is able to generate the opposite conclusion and exhibit parameter values for which the presence of the competitor increases the rate of innovation.

Likewise, Cuésta, Noton, and Vatter (2019) studied whether vertical integration between hospitals and insurers raises total surplus in the health care system. They found that it does. The article shows that changing the degree of consumer price sensitivity can reverse this qualitative conclusion.

Example. Suppose for simplicity that we are interested in a binary conclusion (e.g., that $\hat{c}$ is positive). Following Abadie (2020), consider how much reader r updates their beliefs about c based on learning that $\hat{c} > 0$. As noted in Abadie (2020), the law of total probability implies that for any set of values $\mathscr{C}$ for c,

$$\left|\Pr_r\{\mathscr{C}\} - \Pr_r\left\{\mathscr{C}|\hat{c}>0\right\}\right| = \frac{\Pr_r\left\{\hat{c}\le 0\right\}}{\Pr_r\left\{\hat{c}>0\right\}}\left|\Pr_r\{\mathscr{C}\} - \Pr_r\left\{\mathscr{C}|\hat{c}\le 0\right\}\right| \le \frac{\Pr_r\left\{\hat{c}\le 0\right\}}{\Pr_r\left\{\hat{c}>0\right\}}.$$

Hence, if reader r thinks it very likely that the model will generate a positive estimate, $\Pr_r\left\{\hat{c}>0\right\} \approx 1$, they barely update their beliefs when told that $\hat{c} > 0$. By contrast, if a researcher can provide evidence that plausible realizations of the data would have led to a different conclusion, learning that $\hat{c} > 0$ becomes much more informative. To formalize this in our model, suppose the report is $\left(1\left\{\hat{c}>0\right\}, \hat{t}(X,\Omega)\right)$, where $\Pr_r\left\{\hat{c}>0|\hat{t}\right\} \ll \Pr_r\left\{\hat{c}>0\right\}$. For example, $t(X,\Omega)$ might record a set of values $\mathscr{Y}$ for Y that would have led to negative estimates. Since we still have

$$\left|\Pr_r\left\{\mathscr{C}|\hat{t}\right\} - \Pr_r\left\{\mathscr{C}|\hat{c}>0,\hat{t}\right\}\right| = \frac{\Pr_r\left\{\hat{c}\le 0|\hat{t}\right\}}{\Pr_r\left\{\hat{c}>0|\hat{t}\right\}}\left|\Pr_r\left\{\mathscr{C}|\hat{t}\right\} - \Pr_r\left\{\mathscr{C}|\hat{c}\le 0,\hat{t}\right\}\right|,$$

the reader may now update substantially after learning that $\hat{c} > 0$.

7. CONCLUSION

Estimators of nonlinear models with multiple interacting agents or sectors can be complicated functions of the data and therefore difficult for readers to understand. Yet such models form an important part of the economist's toolkit for many real-world problems. Fortunately, economists working with such models have developed many practices that aid the transparency of their research. Here, we survey those practices and suggest areas for further improvement. Many of these improvements can be adopted at little or no computational cost to the researcher.

ACKNOWLEDGMENTS

Patrick Kline, Emi Nakamura, participants in a 2020 ASSA session, discussants Stephane Bonhomme, Christopher Taber, and Elie Tamer, editor Christian Hansen, and an anonymous associate editor provided valuable comments. We thank our many dedicated research assistants for their contributions to this project.

FUNDING

We acknowledge funding from the National Science Foundation (1949047), the Brown University Population Studies and Training Center, and the Stanford Institute for Economic Policy Research (SIEPR).

[Received February 2020. Revised July 2020.]

References

Abadie, A. (2020), "Statistical Non-Significance in Empirical Economics," *American Economic Review: Insights*, 2, 193–208. [721]

Agarwal, N., and Somaini, P. (2018), "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Econometrica*, 86, 391–444. [715]

Agarwal, S., Chomsisengphet, S., Mahoney, N., and Stroebel, J. (2018), "Do Banks Pass Through Credit Expansions to Consumers Who Want to Borrow?," *Quarterly Journal of Economics*, 133, 129–190. [714]

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., and Schnell, M. (2019), "Food Deserts and the Causes of Nutritional Inequality," *Quarterly Journal of Economics*, 134, 1793–1844. [714,715]

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992), *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: MIT Press. [715]

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017), "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132, 1553–1592. [711,712,718,719,720]

——— (2020), "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica* (forthcoming). [712,718,719]

Andrews, I., and Shapiro, J. M. (2020), "A Model of Scientific Communication," NBER Working Paper No. 26824. [711,712,713]

Angrist, J. D., and Pischke, J.-S. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30. [711,714]

Armstrong, T., and Kolesár, M. (2019), "Sensitivity Analysis Using Approximate Moment Condition Models," Cowles Foundation Discussion Paper No. 2158R. [717]

Athey, S., and Imbens, G. W. (2018), "Design-Based Analysis in Difference-in-Differences Settings With Staggered Adoption," Working Paper, arXiv no. 1808.05293v3. [718]

Attanasio, O. P., Meghir, C., and Santiago, A. (2012), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA," *Review of Economic Studies*, 79, 37–66. [714]

Autor, D., Kostøl, A., Mogstad, M., and Setzler, B. (2019), "Disability Benefits, Consumption Insurance, and Household Labor Supply," *American Economic Review*, 109, 2613–2654. [713,715,716,717]

Beraja, M., Fuster, A., Hurst, E., and Vavra, J. (2018), "Regional Heterogeneity and the Refinancing Channel of Monetary Policy," *Quarterly Journal of Economics*, 134, 109–183. [715]

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019), "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, 127, 639–688. [714]

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. [715,719,720]

Bonhomme, S., Lamadon, T., and Manresa, E. (2019), "A Distributional Framework for Matched Employer-Employee Data," *Econometrica*, 87, 699–739. [715]

Chiappori, P. A., Salanié, B., Salanié, F., and Gandhi, A. (2019), "From Aggregate Betting Data to Individual Risk Preferences," *Econometrica*, 87, 1–36. [715]

Christensen, G., and Miguel, E. (2018), "Transparency, Reproducibility, and the Credibility of Economics Research," *Journal of Economic Literature*, 56, 920–980. [713]

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012), "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272. [712,720]

Crawford, G. S., Lee, R. S., Whinston, M. D., and Yurukoglu, A. (2018), "The Welfare Effects of Vertical Integration in Multichannel Television Markets," *Econometrica*, 86, 891–954. [715]

Cuésta, J. I., Noton, C., and Vatter, B. (2019), "Vertical Integration Between Hospitals and Insurers," Working Paper. [721]

David, J. M., and Venkateswaran, V. (2019), "The Sources of Capital Misallocation," *American Economic Review*, 109, 2531–2567. [716]

de Chaisemartin, C., and D'Haultfoeuille, X. (2020), "Two-Way Fixed Effects Estimators With Heterogeneous Treatment Effects," *American Economic Review* (forthcoming). [718]

De Nardi, M., French, E., and Jones, J. B. (2010), "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 118, 39–75. [717]

DellaVigna, S. (2018), "Structural Behavioral Economics," in *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 1), eds. B. D. Bernheim, S. DellaVigna, and D. Laibson, Amsterdam: Elsevier, pp. 613–723. [716]

DellaVigna, S., List, J. A., and Malmendier, U. (2012), "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 127, 1–56. [717]

Dridi, R., Guay, A., and Renault, E. (2007), "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136, 397–430. [716]

Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388. [720]

Eckstein, Z., Keane, M., and Lifshitz, O. (2019), "Career and Family Decisions: Cohorts Born 1935–1975," *Econometrica*, 87, 217–253. [716]

Fetter, D. K., and Lockwood, L. M. (2018), "Government Old-Age Support and Labor Supply: Evidence From the Old Age Assistance Program," *American Economic Review*, 108, 2174–2211. [714]

Fréchette, G. R., Lizzeri, A., and Salz, T. (2019), "Frictions in a Competitive, Regulated Market: Evidence From Taxis," *American Economic Review*, 109, 2954–2992. [716]

Fu, C., and Gregory, J. (2019), "Estimation of an Equilibrium Model With Externalities: Post-Disaster Neighborhood Rebuilding," *Econometrica*, 87, 387–421. [715,716]

Gelman, A., and Imbens, G. W. (2019), "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37, 447–456. [718]

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2014), "Competition and Ideological Diversity: Historical Evidence From US Newspapers," *American Economic Review*, 104, 3073–3114. [715]

Goettler, R. L., and Gordon, B. R. (2011), "Does AMD Spur Intel to Innovate More?," *Journal of Political Economy*, 119, 1141–1200. [717,721]

Goodman-Bacon, A. (2019), "Difference-in-Differences With Variation in Treatment Timing," NBER Working Paper No. 25018. [718]

Gourinchas, P.-O., and Parker, J. A. (2002), "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89. [716,717,718,720]

Hackmann, M. B. (2019), "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry," *American Economic Review*, 109, 1684–1716. [714,715]

Harasztosi, P., and Lindner, A. (2019), "Who Pays for the Minimum Wage?," *American Economic Review*, 109, 2693–2727. [715,716]

Head, K., and Mayer, T. (2019), "Brands in Motion: How Frictions Shape Multinational Production," *American Economic Review*, 109, 3073–3124. [715,716]

Heckman, J. J. (2010), "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48, 356–398. [711]

Hendren, N. (2013), "Private Information and Insurance Rejections," *Econometrica*, 81, 1713–1762. [719]

Horowitz, J. L., and Manski, C. F. (1995), "Identification and Robustness With Contaminated and Corrupted Data," *Econometrica*, 63, 281–302. [721]

Imai, K., and Kim, I. S. (2020), "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference With Panel Data," *Political Analysis* (forthcoming). [718]

Keane, M. P. (2010), "Structural vs. Atheoretic Approaches to Econometrics," *Journal of Econometrics*, 156, 3–20. [714,715]

Kleven, H. J. (2018), "Language Trends in Public Economics," available at *https://www.henrikkleven.com/uploads/3/7/3/1/37310663/languagetrends_slides_kleven.pdf*. [714]

Kline, P., and Santos, A. (2013), "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the US Wage Structure," *Quantitative Economics*, 4, 231–267. [721]

Leamer, E. E. (1981), "Sets of Estimators of Location," *Econometrica*, 49, 193–204. [720]

Lewbel, A. (2019), "The Identification Zoo: Meanings of Identification in Econometrics," *Journal of Economic Literature*, 57, 835–903. [713,714]

Manski, C. (2003), *Partial Identification of Probability Distributions*, New York: Springer. [720]

——— (2007), *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press. [720]

Masten, M. A., and Poirier, A. (2020), "Inference on Breakdown Frontiers," *Quantitative Economics*, 11, 41–111. [721]

Matzkin, R. L. (2013), "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486. [714]

Molinari, F. (2020), "Econometrics With Partial Identification," *Handbook of Econometrics* (forthcoming). [712]

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019), "Definitions, Methods, and Applications in Interpretable Machine Learning," *Proceedings of the National Academy of Sciences of the United States of America*, 116, 22071–22080. [713]

Nakamura, E., and Steinsson, J. (2018), "Identification in Macroeconomics," *Journal of Economic Perspectives*, 32, 59–86. [716]

Nikolov, B., and Whited, T. M. (2014), "Agency Conflicts and Cash: Estimates From a Dynamic Model," *Journal of Finance*, 69, 1883–1921. [717]

Pakes, A. (2014), "The 2013 Lawrence R. Klein Lecture: Behavioral and Descriptive Forms of Choice Models," *International Economic Review*, 55, 603–624. [713]

Polyakova, M. (2016), "Regulation of Insurance With Adverse Selection and Switching Costs: Evidence From Medicare Part D," *American Economic Journal: Applied Economics*, 8, 165–195. [713]

Rossi, F., and Chintagunta, P. K. (2016), "Price Transparency and Retail Prices: Evidence From Fuel Price Signs in the Italian Highway System," *Journal of Marketing Research*, 53, 407–423. [713]

Sun, L., and Abraham, S. (2020), "Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects," MIT Working Paper. [718]

Tamer, E. (2010), "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195. [712,715,720]