Journal of Business & Economic Statistics

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

# Thoughts on "Transparency in Structural Research"

Christopher Taber

**To cite this article:** Christopher Taber (2020) Thoughts on "Transparency in Structural Research", Journal of Business & Economic Statistics, 38:4, 726-727, DOI: 10.1080/07350015.2020.1796396

**To link to this article:** https://doi.org/10.1080/07350015.2020.1796396

Published online: 28 Sep 2020.

# Thoughts on "Transparency in Structural Research"

**Christopher TABER**
Department of Economics, University of Wisconsin-Madison, Madison, WI (*ctaber@ssc.wisc.edu*)

## 1. INTRODUCTION

I applaud Andrews, Gentzkow, and Shapiro on their work on transparency—both in this article and in their other work. This is a remarkably important problem but also very challenging. I think they have made real headway.

The article is phrased as transparency in "structural research." I think that is selling itself short as this work on transparency is important and relevant in all types of empirical work. In some sense the main difference between transparency in structural and design based work be not about the style of work but rather that many design based procedures (such as instrumental variables, difference in differences, or regression discontinuity) are familiar. As a result, well informed readers come in with a good understanding of how these methods work. Structural work has the disadvantage of both typically being more complicated and but also more novel as new structural models are usually different than the predecessors in important ways. As a result, while making progress on transparency is particularly important for structural work, it is important for all empirical work.

I would also add that I think increasing transparency is also important for the field of econometrics. Over my career I have witnessed a growing gap between theoretical econometrics and applied work. There are many reasons for this, but transparency might be an important one. When empirical researchers are presented results from both familiar methods and newly developed econometric methods, I think they tend to pay more attention to results from the familiar methods in large part because they understand them better. If theoretical econometricians can facilitate the interpretation of estimates from their methods, they would almost certainty be adopted more broadly.

All that said, this is a really hard problem. I have thought about this in my own work. I have also read many structural articles and completely understand the problem that people complain about. I have heard complaints from nonstructural people that do not want to pay attention to structural articles either because "they are poorly identified" or because "identification is a black box." I am much more sympathetic to the second complaint than the first (in fact the fact that identification is a black box makes it hard to make definitive claims on the whether it is poorly identified as not). Thus to increase the impact of structural work it is important to make it less of a black box. Doing this is extremely difficult. For much of my work, it takes me years to write an article and I have spent a lot of time experimenting with the model and the data to get a sense of how they interact to produce numbers. Communicating everything I have learned in years of exploration of the model in a few pages is impossible. We are never going to be able to make things completely transparent. Andrews, Gentzkow, and Shapiro have made an important step in this work and while we will never "solve" this problem we need to keep trying.

I want to make two comments/extensions about this work. The first is related to the difference between sensitivity and transparency. The second is thinking precisely about the central goal of the research. I will then conclude.

## 2. SENSITIVITY VERSUS TRANSPARENCY

I am not totally satisfied with the definition of transparency—though I am not going to provide an alternative definition. Using their notation Andrews, Gentzkow, and Shapiro define $c(a_0, \eta)$ to be the quantity of interest given the researchers assumptions $a_0$ and $\widehat{c}$ to be the estimate of $c(a_0, \eta)$. They then define transparency as

$$T_r\left(\widehat{c},\widehat{t}\right) = \frac{\text{var}_r(c(a,\eta)) - E_r\left[\text{var}_r\left(c(a,\eta) \mid \widehat{c},\widehat{t}\right)\right]}{\text{var}_r(c(a,\eta)) - E_r\left[\text{var}_r\left(c(a,\eta) \mid D\right)\right]}$$

where $a$ is the model assumption, $c(a, \eta)$ is the quantity of interest under alternative assumption $a$, $\widehat{t}$ is auxiliary statistics, and the expectations are taken relative to the reader's priors.

This conflates two distinct concepts. The first is that the reader might like different assumptions than the researcher $a \neq a_0$. The second is that the reader cannot observe the data but only $\widehat{c}$ and $\widehat{s}$. I would have used the term robustness for the first concept and transparency for the second. While I do not see an obvious way to decompose this into the two conceptually different issues—I would hope that future work can do this.

To me the first concept is an old question we think about all the time—how robust are the results to alternative assumptions. The article makes some nice arguments about how to do this in Section 6—but in general this is very difficult. As an example an assumption that people often make in estimating discrete


Journal of Business & Economic Statistics
October 2020, Vol. 38, No. 4
DOI: 10.1080/07350015.2020.1796396

choice dynamic programming is that the flow utility can be written as

$$u_j(S_{it};\beta) + \epsilon_{ijt},$$

where $S_{it}$ is state variable for person $i$ at time $t$, $j$ is an alternative, $\beta$ is a parameter vector, and $\epsilon_{ijt}$ is an iid extreme value error term. We should worry a lot about the assumptions on the error term. However, the author is not making them because fundamentally she believes that is the true distribution, she is doing it because it is the only way to solve the problem. In reality there is no reason to expect the error term to be separable, have an extreme value marginal distribution, be uncorrelated across $j$, or is uncorrelated across $t$. Knowing how the results would vary without these assumptions is really difficult because it is not feasible to solve this alternative model. I am not sure that this article helped me think about this problem.

By contrast, I think Andrews, Gentzkow, and Shapiro's work makes a lot of progress on the second problem. This is extremely important and has gotten much less attention in economics. That is, taking the model $a_0$ as given what is the mapping between the data and the results? This to me is the most important contribution of this work. Specifically I would highlight Section 5 of the article as the most important contribution.

## 3. WHAT IS C?

I want to make an important point about my interpretation of this article which I think is important for empirical researchers. As the article states, $c$ is a scalar parameter of interest. I want to think about this in a classic structural framework. Let $c$ be some policy counterfactual of interest which is the main target of the article. The empiricist first estimates structural parameters $\widehat{\beta}$ and then uses these parameters to simulate the policy counterfactual $c$. Suppressing the assumptions and building on the notation in Section 5 of the article consider the case in which we are using indirect inference or simulated method of moments so we can write

$$\widehat{\beta} = B\left(\widehat{s}\right)$$

and then simulate the counterfactual

$$\widehat{c} = C\left(\widehat{\beta}\right).$$

This is subsumed into the notation in Section 5 of the article when

$$h\left(\widehat{s}\right) + \nu_h = C\left(B\left(\widehat{s}\right)\right).$$

Within this context there are two points I want to make. The first is that transparency about $h$ might involve transparency about $B$ and $C$ separately. For example, we might first care about the key parameters $\beta$ that are important for determining $c$. We might then want to know what the key statistics for determining each of these key parameters.

As an example consider a simple simultaneous equations model for demand and supply of a good. Suppose we want to understand the effect of increase the tax on this good (imposed on consumers). We know that the policy effect is influenced heavily by two of the parameters: the elasticity of supply and the elasticity of demand (transparency about $C(\cdot)$). We also know how they are primarily determined the variables excluded from the demand equation play a key role in determining the elasticity of demand and likewise for supply (transparency about $B(\cdot)$). To maximize transparency of this result to someone unfamiliar with the model you would presumably explain both of these relationships.

The second point I want to make is to highlight this point for writers of structural articles. Many readers go into great detail try to convey something about $B$. Often I find this a waste of time. I care less about transparency about all of the parameters. What I really care about is transparency for the key conclusion of this article. In the context of my model above it is really $h$ that I care about, not every element of $B$. One can first explain which parameters are important for $C(\cdot)$ then tell us which parts of $\widehat{s}$ are important for the important parameters I like the way that Andrews, Gentzkow, and Shapiro develop their model with that goal in mind.

## 4. CONCLUSIONS

This is a really important research agenda and I thank the authors for their work pushing it. While they have not "solved" this essentially unsolvable problem, they have made great headway. My first comment was that I think of the question of sensitivity of results to modeling assumptions as distinct from the question of transparency of structural results (taking the structural model as given). My second comment is that I would urge readers of this article to take their model with a one (or at least small) dimensional parameter of interest, $c$, seriously. Structural work will have greater if these arguments are focused on the main point of the article. I look forward to seeing more articles on this subject by these authors as well as by others who have been inspired by this important work.