Comment on "Identification of Nonseparable Triangular Models with Discrete Instruments" by D'Haultfœuille and Février

Alexander Torgovitsky[*]

June 25, 2015

Abstract

The May 2015 issue of *Econometrica* contains two papers published back-to-back that purport to show essentially the same results about identification in nonseparable instrumental variables models. These two papers are Torgovitsky (2015) and D'Haultfœuille and Février (2015). This comment discusses the differences between the two papers, and corrects some of the false statements that D'Haultfœuille and Février (2015) make about Torgovitsky (2015). In clarifying the differences, I argue that the results in D'Haultfœuille and Février (2015) are weaker than those in Torgovitsky (2015) for all empirically relevant cases. Specifically, the only identification result in D'Haultfœuille and Février (2015) that is not implied by results in Torgovitsky (2015) requires all of the following four conditions to hold: (i) either the endogenous variable is scalar and its support is completely unbounded, or it is bounded and the structural function is well-behaved everywhere except at the boundary of the support of the endogenous variable; (ii) the instrument is discrete with at least 3 points of support; (iii) a high-level "'non-periodicity" condition is satisfied; and (iv) none of the conditional distribution functions of the endogenous variable intersect one another on the interior of its support. In all other cases, the results in Torgovitsky (2015) are either not covered by those in D'Haultfœuille and Février (2015), or are obtained under weaker conditions.

---
[*]Department of Economics, Northwestern University, `a-torgovitsky@northwestern.edu`.

1 Introduction

This comment clarifies the differences between Torgovitsky (2015, "T" in the following) and D'Haultfœuille and Février (2015, "DF" in the following). I will assume that the reader has briefly read both papers and I will maintain the notation used in T in the following discussion. Specifically, Y is the continuously distributed outcome variable, X is the continuously distributed endogenous variable, ε is the scalar unobservable in the outcome equation, g is the structural function of interest which maps realizations of X and ε into Y, $g^\star$ is the "true value" of g, i.e. the value assumed to generate the data, and Z is the excluded, exogenous instrument. The primary goal of both papers is to establish point identification of the function $g^\star$, which is assumed in both papers to be continuous in X and ε and strictly increasing in ε, at least on the interior of the support of (X, ε). In the next section, I explain the differences between the two papers, and argue that the results in T are stronger than those in DF in all empirically relevant cases. In Section 3, I correct some false claims made by DF about T that appear in their published paper.

2 The Differences Between DF and T

Before discussing the differences between the two papers, first consider the similarities. DF and T study the same nonseparable triangular model under the same assumptions on the exogeneity of the instrument, and on the dimension of heterogeneity in the outcome and first stages.[1] The key observation, which was first made in the job market paper version of T (Torgovitsky, 2010; see equations (4) and (5)), is that establishing point identification of $g^\star$ in these models is tantamount to solving a functional fixed point problem. The equations that capture this observation are the two unnumbered insets on pg. 1189 of T, which are given as equation (3.2) on pg 1202 of DF. Both DF and T use the same argument to obtain this equation; compare the discussion on pp. 1189-1190 of T and the second inset of the proof of Theorem 1 on pg. 1205 of DF.

Both DF and T then proceed to find sufficient conditions under which this fixed point problem admits a unique solution, thereby ensuring point identification of $g^\star$. This is where the two papers differ in their technical approach. T uses direct differential and sequencing arguments to show that certain conditions on the dependence between X and Z ensure that the fixed point problem admits a unique solution. DF observe that the collection of

---

[1]The job market paper version of T (Torgovitsky, 2010), emphasized an equivalent interpretation to these assumptions in terms of the copulas of $\{(X, \varepsilon)|Z = z\}_{z \in \mathcal{Z}}$. This was removed in during the publication process in order to comply with space constraints.

transformations relevant to the fixed point equation can be viewed as a group generated by the action of function composition (Theorem 1 of DF). They then consider sufficient conditions that imply that the orbit of this group generates all support points of X, thereby ensuring that the fixed point problem admits a unique solution.

The main differences between DF and T are therefore in the sufficient conditions used to ensure the uniqueness of a solution to the functional fixed point problem that characterizes point identification of $g^\star$. Both papers provide several sets of conditions that apply in different situations, depending on the properties of (and the relationship between) the endogenous variable (X) and the instrument (Z). Both papers discuss the case in which X is scalar in their main texts, and treat the case of a vector-valued X (which raises severe complications) in their supplemental appendices. Figure 1 provides a schematic description of the differences between the main results in the two papers. The following discussion clarifies and expands upon the content of Figure 1.

The first distinction between DF and T arises in the case in which the instrument Z continuous. T provides a result that establishes point identification of $g^\star$ by exploiting continuity in Z, while imposing only weak relevance conditions; see Theorem S1 in the supplemental material of T. DF claim in their introduction (footnote 2, on pg. 1199) to be able to apply their analysis to the continuous Z case, however they provide no additional details.[2] It is possible that what DF mean is that they could *discretize* a continuously distributed Z and then apply the same analysis as they do for discrete Z. However, this necessarily does not exploit continuity in Z. Theorem S1 of T shows that when Z is continuously distributed, point identification of $g^\star$ can be achieved under a relatively weak dependence condition between X and Z by exploiting smooth variation in Z. Moreover, this result also immediately generalizable to the case in which X is a vector, see Section S1 of T.

The more surprising result of both papers is that point identification of $g^\star$, which is an infinite-dimensional object, can be obtained when Z is discrete or even just binary. Let $F_{X|Z}(\cdot|z)$ denote the conditional distribution function of X, given $Z = z$. The analysis of the discrete Z case for both papers depends on whether and to what extent $F_{X|Z}(\cdot|z^a)$ intersects $F_{X|Z}(\cdot|z^b)$ for some values $z^a \neq z^b$. It also depends on whether these intersections occur in the interior or at the boundary of the support of the endogenous variable X.

For any z, let $\operatorname{supp}(X|Z = z)$ denote the support of $X|Z = z$, and note that supports are always closed sets by definition. Let int denote the interior of a set. Consider the following

---

[2]Allowing Z to be continuous would fundamentally change the structure of the technical analysis in DF, because the group that they consider would have an infinite number of elements in this case.

assumption:

> **Assumption R:** For every z, $\mathrm{supp}(X|Z = z)$ is a (closed) interval, and there exists a z^a and z^b such that $\{x \in \mathrm{int}\left[\mathrm{supp}(X|Z = z^a) \cap \mathrm{supp}(X|Z = z^b)\right] : F_{X|Z}(x|z_a) = F_{X|Z}(x|z_b)\}$ is finite and nonempty.

Assumption R says that there exist two values of the instrument for which the conditional distribution functions of X intersect a finite and non-zero number of times on the *interior* of the support of X. DF refer to this situation as an instance of the "non-free" case. Both papers establish point identification of $g^\star$ in this case, although DF require the additional and empirically relevant condition that $\mathrm{supp}(X|Z = z)$ does not vary with z, i.e. that the support of the endogenous variable is unaffected by variation in the instrument.[3] Hence, T's results are stronger than DF's whenever Assumption R holds. In principle, DF's non-free case can also occur without Assumption R, however they do not discuss general sufficient conditions for this.[4] Unless one assumes the non-freeness property directly, T's results are therefore stronger than DF's for all non-free cases.

In addition to allowing for Assumption R as stated, Theorem 2 in T also allows the intersections in Assumption R to occur at a boundary point of $\mathrm{supp}(X|Z = z^a) \cap \mathrm{supp}(X|Z = z^b)$. In order to do this, T requires the assumption that the continuity and monotonicity of $g^\star$ is satisfied on both the interior *and* the boundary of the support of X, whereas DF only impose these conditions on the interior of the support.[5] These seemingly minor differences between the interior and the boundary are actually quite important in this setting. The reason is that whenever $\underline{x} \equiv \inf \mathrm{supp}(X) > -\infty$, i.e. whenever X is bounded from below, there will *always* exist a $z^a \neq z^b$ such $F_{X|Z}(\underline{x}|z^a) = F_{X|Z}(\underline{x}|z^b) = 0$, i.e. such that these

---

[3]The reason that DF need this additional condition is a consequence of their group theoretic strategy. Without this assumption, functions that comprise the members of their group would not necessarily be well-defined everywhere on the support of X. DF describe this condition as a "regularity" condition, but provide no justification for this label. DF also claim on pg. 1201 of their main text to show in the supplemental appendix that "...several of our results still hold if the intersections of conditional supports are large enough." I was unable to find any such results in their supplemental appendix.

[4]DF define the non-free case in their Definition 1 as a property of the group generated by composing the functions relevant to the fixed point equation that characterizes identification of $g^\star$. The fact that Assumption R implies their non-freeness definition is immediate. In Section S1.1.2 of their supplemental appendix, DF provide a very specific numerical example in which non-freeness occurs without Assumption R. However, they provide no discussion of sufficient conditions for this to occur more generally.

[5]DF express contradictory views about the importance of this difference. On pg. 1201, they assert that allowing for discontinuity at the boundary of the support is "important." However, in the very next paragraph, they refer to requiring continuity and strict monotonicity at the boundary as "slightly reinforcing" the assumption that the same conditions hold on the interior. DF do not provide any discussion of empirical examples in which requiring continuity and/or strict monotonicity on the entire support would be substantively more restrictive than requiring it only on the interior of the support.

two conditional distribution functions intersect at $\underline{x}$. Of course, this is also true (with 1 instead of 0) if X is bounded from above by a finite $\overline{x} \equiv \sup \mathrm{supp}(X)$. Hence, Theorem 2 in T applies whenever X is bounded, as long as the shape restrictions on $g^\star$ are assumed to extend to this boundary point.[6]

In contrast, DF only consider intersection points on the *interior* of the support, as in Assumption R. In addition to the non-free case discussed above, in which there is an intersection on the interior (Assumption R), DF also define a "free" case in which there are no intersections on the interior. Their result for this so-called free case (Theorem 2) requires Z to have 3 or more points of support (see the discussion following Assumption 4). It also imposes a high-level "non-periodicity" condition (see Assumption 4) which is not assumed in any of the analysis in T.

As discussed above, T's Theorem 2 applies to cases in which there are no intersections on the interior of the support, and hence also covers most instances of what DF refer to as the free case. In particular, Theorem 2 of T establishes point identification of $g^\star$ when X is bounded from either above or below and $g^\star$ is assumed to be well-behaved at this boundary, even if the conditional distribution functions do not intersect on the interior of the support. In DF's terminology, this would be the free case with the added assumption that $g^\star$ is well-behaved at the boundary point. In this case, DF's Theorem 2 is weaker than T's Theorem 2, since the former requires both the non-periodicity condition, and the assumption that Z has 3 or more points of support.

The subset of the free cases in which DF's Theorem 2 is not covered by T's Theorem 2 occur when either X is completely unbounded, or $g^\star$ is not well-behaved at the boundary of the support of X. These cases still require the non-periodicity condition and the assumption that Z has 3 or more points of support. Since T's results are also stronger than DF's for the non-free case, the implication is therefore that T's results are stronger than DF's unless each of the following four conditions are true: (i) either the support of X is unbounded, or it is bounded but $g^\star$ is not assumed to be well-behaved at the boundary; (ii) Z is discretely distributed with 3 or more points of support; (iii) the non-periodicity condition (Assumption 4 of DF) holds; and (iv) this is the free case, so that the conditional distribution functions of X do not intersect on the interior of the support of X.

Conditions (i)–(iv) do not represent an empirically relevant case. Most economic variables of interest in economic applications are either positive or negative, which implies a natural

---

[6]It is also required that the variation between Z and X is sufficiently strong in the sense of the second requirement of Assumption R that the conditional distribution functions do not intersect an infinite number of times. However, this aspect is common to both DF and T.

lower or upper bound of 0.[7] Hence, as long as extending the properties of $g^\star$ to this boundary is not objectionable—and DF provide no claim of why it would be—condition (i) already represents an unusual case. The addition of (iv) to (i) then imposes a very specific and unusual structure on the relationship between X and Z. In addition, Assumption (ii) rules out common binary instruments, such as the intent-to-treat. Moreover, the non-periodicity condition (iii) required for DF's result is high-level and therefore difficult to interpret and/or verify. Combined, (i)–(iv) are sufficiently restrictive as to represent a case of empirical irrelevance. Since T's results are stronger than DF's for all other cases, it follows that T's results are stronger than DF's for all empirically relevant cases.

The distinction between the boundary and interior for the discrete Z results of T and DF can be most clearly observed in the simple example discussed by T on pg. 1191. In this example, X has half-bounded support $[\xi, \infty)$ with ξ finite (e.g., $\xi = 0$), and $Z \in \{0, 1\}$ is a binary instrument. The conditional distribution functions are assumed to satisfy $F_{X|Z}(x|1) > F_{X|Z}(x|0)$ for all $x > \xi$. DF would call this situation the free case, since $F_{X|Z}(\cdot|1)$ and $F_{X|Z}(\cdot|0)$ do not intersect on the interior of the support of X, i.e. on (ξ, ∞). Their identification result for the free case (Theorem 2) requires Z to be discrete with 3 or more points of support. As a result, it does not apply to this binary Z case. In contrast, T shows how to use a simple sequencing argument to establish point identification for this case. The argument given by T takes less than a half of a page, and uses only basic properties of sequences and continuity. The fact that DF's analysis does not apply to even this most basic case suggests that their attempt to apply group theory to this identification problem is fundamentally misguided.

There are also several smaller differences between the papers. T provides a sharp characterization of the identified set (Theorem 1) in terms of the distribution of the observable random variables. This directly provides an *equivalent* condition for the model to be correctly specified, that involves only the distribution of observables. In contrast, DF only derive a testable implication, but do not determine whether this implication can also be satisfied when the model is misspecified (pp. 1204–1205), which would lead to an inconsistent test. T's Theorem 1 also directly suggests a minimum distance estimator, which has been developed and studied in detail in Torgovitsky (2013). DF provide no discussion of

---

[7]Exceptions to this include transformations such as the log or difference of a variable. However, even for these exceptions, it is difficult to think of a variable of interest that is more appropriately modeled as unbounded rather than having compact support. Moreover, in the nonparametric framework considered by DF and T, these transformations could be absorbed into the specification of the admissible g functions, in which case the underlying variable X would again have a natural lower or upper bound of 0.

estimation, and it is not clear how their identification analysis could be used to construct a reasonable estimator, except perhaps by following T's approach. Finally, T provides an extension of the analysis with a discrete instrument to the vector X case (Theorem S2). This requires stronger, but still interpretable conditions, on the support of X. (In fact, the key assumption that T imposes in this extension is that $\text{supp}(X|Z = z)$ does not vary with z. This assumption is maintained by DF throughout the entirety of their analysis.) DF provide two point identification results for the vector X case (Theorems S1 and S2), both of which impose additional restrictions on the functional relationship between X and Z, as well as additional high-level rank conditions.

3   Correction of False Claims

DF make the following inaccurate claims about T in the course of their published paper.

1. On pg. 1199, DF state (emphasis mine):

   *"... we prove ... that **unless the instrument is binary**, and Z has a strictly monotonic effect on X, g [$g^\star$] is fully identified under ... "*

   This statement is false or misleading, depending on how one interprets the sentence. As shown in T and discussed in the previous section, point identification of $g^\star$ can still be shown when Z is binary under quite general conditions. The statement is correct (but misleading) if it is interpreted as saying that *their analysis* does not establish identification when Z is binary, rather than saying that such results are not possible.

2. On pg. 1201, DF state:

   "He [Torgovitsky (2015)] imposes rather that this support [the support of $X|Z = z$] is bounded either from above or below."

   This statement is false. The conditions of Theorem 2 of T can be satisfied when the support of X is unbounded, as discussed in the previous section. This is clearly evident from the statement and discussion of Theorem 2 in T.

3. On pg. 1204, DF state:

   "Note that Theorem 2 does not cover the case of a binary instrument with freeness, since Assumption 4 implies $K \geq 3$. This is a case for which

> *Theorem 1 does not apply. [. . .]. In this case (see D'Haultœuille and Février (2011)), we can show that the model is not identified. . ."*

This statement is false. As shown in T and discussed in the previous section, point identification can still be shown when Z is binary under quite general conditions that are included in DF's free case, as long as X is bounded either from above or below and $g^\star$ is well-behaved at this boundary point.
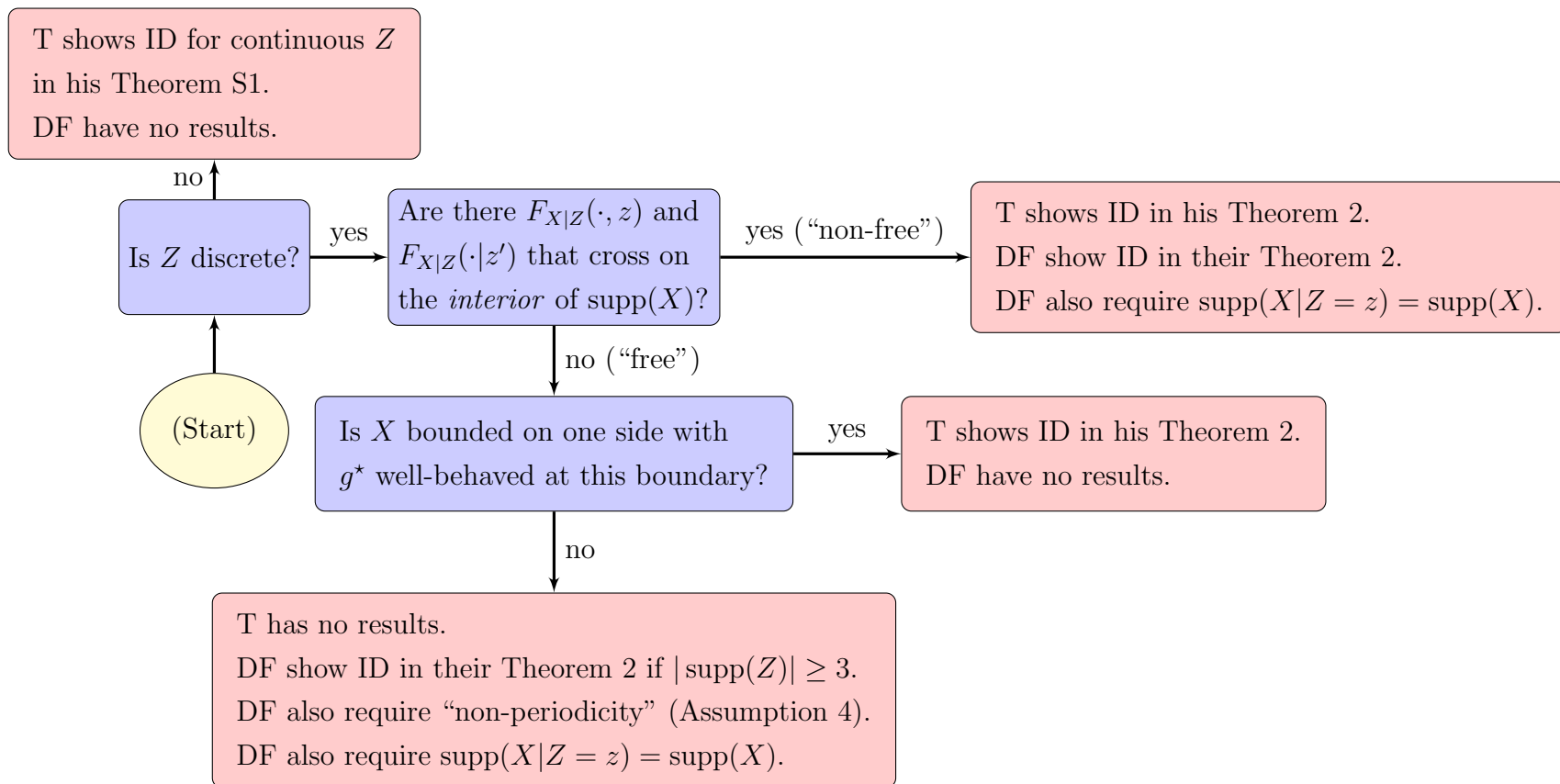
Figure 1: The differences between the conditions used to establish point identification in Torgovitsky (2015, "T") and in D'Haultfœuille and Février (2015, "DF"). DF obtain one result that is not covered by T's results. This occurs in the free case where X is completely unbounded (or bounded, but with the set of admissible g well-behaved everywhere except at this boundary point), $|\mathrm{supp}(Z)| \geq 3$, and a high-level "non-periodicity" condition holds. In all other cases, either T's results are not covered by DF, or both papers have the same essential result, but DF require the additional condition that $\mathrm{supp}(X|Z = z)$ does not vary with z.

References

D'Haultfœuille, X. and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 3, 1199–1210.

Torgovitsky, A. (2010): "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity," *Job market paper*.

——— (2013): "Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models," *Working paper*.

——— (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197.