CPI COMPETITION POLICY INTERNATIONAL

CPI Antitrust Chronicle

July 2013 (2)

Not So Natural Experiments

Gregory K. Leonard
Edgeworth Economics LLC

Not So Natural Experiments

Gregory K. Leonard[1]

I. INTRODUCTION

Over the last 20 years, the economics profession has moved away from the estimation of structural economic models and toward the use of experiments, either controlled experiments or so-called "natural" experiments, for the purposes of estimating the causal effects of programs, policies, and other interventions.[2] While industrial organization economists have been slower to adopt the experimental approach, it is increasingly commonly seen in industrial organization settings.[3] This is particularly so in the area of antitrust analysis. For example, experimental approaches have been used to analyze the effect of a firm's presence in a local market on another firm's prices in that market in the context of the proposed Staples/Office Depot and Whole Foods/Wild Oats mergers, the effects of consummated mergers on prices, and overcharges associated with price-fixing conspiracies.[4]

However, the experimental approach recently has been subject to somewhat of a backlash in the economics literature.[5] Among the criticisms are that experiments, and particularly natural experiments, frequently answer a very limited question that is not the question of real interest and that the conditions required for the reliability of the results are often not well-articulated and, more importantly, are often not satisfied.

In light of the prominence of experimental approaches in economics, the increasing application of such approaches in industrial organization, and the emerging criticisms, this area of economic research is likely to have an impact on the way that antitrust analyses are performed. In this paper, I review the use of natural experiments in the antitrust context.

II. REVIEW OF THE ECONOMETRICS LITERATURE ON THE ESTIMATION OF TREATMENT EFFECTS

Over the course of the last 30 years, a literature has developed in economics and statistics that addresses methods for estimating the causal effect of a "policy intervention" (a "treatment") on economic outcomes.[6] Such methods have been applied, for example, to determine the effect of

---

[1] Partner, Edgeworth Economics LLC, gleonard@edgwortheconomics.com. I thank Laila Haider for helpful comments.

[2] *See*, e.g., Joshua D. Angrist & Jorn-Steffen Pischke, *The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics*, 24 J. ECON. PERSPECTIVES 3 (2010).

[3] *See*, e.g., Angrist & Pischke, *supra* note 2; Aviv Nevo & Michael D. Whinston, *Taking the Dogma Out of Econometrics: Structural Modeling and Credible Inference*, 24 J. ECON. PERSPECTIVES 69 (2010).

[4] *See*, e.g., Orley Ashenfelter & Daniel Hosken, *The effect of mergers on consumer prices: Evidence from five mergers on the enforcement margin*, 53 J. L. & ECON. 417 (2010).

[5] *See*, e.g., Michael P. Keane, *Structural vs. Atheoretic Approaches to Econometrics*, 156 J. ECONOMETRICS 3 (2010); Angus Deaton, *Instruments, Randomization, and Learning About Development*, 48 J. ECON. LIT. 424 (2010); James J. Heckman & Sergio Urzua, *Comparing IV With Structural Models: What Simple IV Can and Cannot Identify*, 156 J. ECONOMETRICS 27 (2010).

[6] An excellent review of the treatment effects literature is provided by Guido W. Imbens & Jeffrey Wooldridge, *Recent Developments in the Econometrics of Program Evaluation*, 47 J. ECON. LIT. 5 (2009). *See also* James J.

a job-training program on a worker's wage, the effect of an educational initiative on a student's educational achievement, and the effect of an economic development program on a region's economic growth.

The starting point for this "treatment effects" literature is the familiar controlled experimental design framework, in which a researcher applies an experimental "treatment" to a group of "subjects," called the "treated group." A second group of subjects, called the "control group," does not receive the treatment. Perhaps the most familiar example of random experimental design occurs in the clinical testing of pharmaceuticals, where groups of patients are randomly assigned to take either the drug or a placebo.

Economists have increasingly been designing and conducting experiments of this nature, where assignment of the treatment is under the control of the researcher so that randomized assignment is possible.[7] However, in many other situations, economists have little or no control over the assignment of the treatment. The treatment effects literature has focused on when the magnitude of treatment effects can be reliably estimated in this latter context.

The treatment effects literature uses an explicit "counterfactual" framework, in which the treatment effect for a given subject is defined as the difference between two "potential outcomes": the outcome with treatment and the outcome without treatment. This framework allows for the possibility of heterogeneity among subjects in the magnitude of the treatment effect. In the context of a drug, for example, it may be the case that application of the drug results in a profound improvement in the relevant health outcome for some subjects, a moderate improvement for others, and no improvement at all for still others.

The "average treatment effect" ("ATE") is defined as the average of the treatment effects over the population from which the subjects are drawn. In the usual case, for each subject, only one of the two potential outcomes is observed. Either a subject receives treatment or not. Therefore, it is not possible generally to simply calculate the treatment effect for each subject and take an average over subjects to obtain an estimate of the ATE. A different approach must be used. One idea is to use the treated group subjects to estimate the average outcome with treatment and use the control group subjects to estimate the average outcome without treatment. Then, the ATE could be estimated by taking the difference between the average outcome with treatment and the average outcome without treatment. For example, suppose the effectiveness of a weight loss drug is being studied. The treatment group receives the drug and the control group receives a placebo. If the treatment group on average lost four pounds and the control group on average gained one pound, the estimate of the ATE would be five pounds of weight loss.

This approach works when the assignment of treatment is randomized.[8] In that case, by construction there is no systematic factor that explains why the treatment group subjects ended

---

Heckman, *Building Bridges Between Structural Program Evaluation Approaches to Evaluating Policy*, 48 J. ECON LIT. 356 (2010).

[7] *See*, e.g., Esther Duflo, Rachel Glennerster, & Michael Kremer, *Using Randomization in Development Economics Research: A Toolkit*, 4 HANDBOOK OF DEVELOPMENT ECONOMICS 3895 (2007).

[8] Of course, randomization of treatment does not solve all problems. For example, subjects generally have free will and thus can, for example, drop out of the experiment. If the probability of attrition is related to the effect of the treatment, the result can be "attrition bias" in the estimate of the effect of the treatment (*see*, e.g., Jerry A. Hausman

up receiving the treatment and similarly why the control group subjects ended up not receiving the treatment. Therefore, there is nothing "special" about the treatment group subjects—they are "representative" of the entire population. Accordingly, the average of the treated group's with-treatment outcomes provides a good estimate of the average with-treatment outcome for the whole population. Similarly, the average outcome for the control group provides a good estimate of the average without-treatment outcome for the whole population. The difference provides a good estimate of the ATE.[9]

However, as discussed above, in many circumstances the economist does not have control over the assignment of the treatment and therefore there is no assurance that the assignment of the treatment has been random. Indeed, there may be reason to think that the assignment was not random, but instead related to the characteristics of the subjects. This is particularly true in situations where subjects are making economic decisions with their own best interests in mind.

In a non-controlled setting, it is useful to think about the assignment of treatment arising through one of three possible mechanisms. First, treatment may be assigned by some "natural" process that mimics randomized assignment ("natural randomization"). For example, suppose the research question involves the effects of capacity on price in a given industry. Some production facilities may be subject to shutdowns for extended periods as a result of factors, such as weather-related events or equipment breakdowns, that are completely exogenous to the outcomes being studied. This could provide a natural experiment, where the treated group consists of market transactions occurring during periods in which a production facility was shut down for such an exogenous reason. With "natural randomization," the ATE can be estimated using the same straightforward comparison of the average outcomes for the treated and control groups.

The second mechanism is where, in addition to some element of natural randomization, the assignment of treatment is related to observable (to us) characteristics of the subjects that are also related to the subjects' outcomes.[10] In this mechanism, the existence of unobserved (to us) characteristics that are both related to assignment and also affect outcomes (conditional on the observable characteristics) is explicitly ruled out. Such a mechanism is called "selection on observables" or "unconfoundedness."

With selection on observables, the difference between the average outcomes for the treated group and the control group does not provide a valid estimate of the ATE in general because the treated group will differ from the control group not only in terms of having received treatment, but also because of differences in the observed characteristics that affect outcomes. Accordingly, the average outcome will differ between the two groups both because of the treatment and because of the observed characteristics. However, it is nevertheless possible to

& David A. Wise, *Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment*, 47 ECONOMETRICA 455 (1979)). Moreover, randomization does not necessarily allow measurement of parameters of interest (*see*, e.g., Heckman & Urzua, *supra* note 5). For example, the ATE may not be of policy interest given the way that the treatment would actually be applied.

[9] By a "good" or "valid" estimate, in technical terms, I mean one that is "consistent."

[10] If assignment is based on observed characteristics that are not related to outcomes, the assignment can be considered to be of the first type.

control for the observed characteristics using more sophisticated econometric procedures and still consistently estimate the ATE in the case of selection on observables.[11]

The third mechanism is where treatment is assigned through a combination of natural randomization, observed characteristics that are also related to outcomes, and unobserved characteristics that are also related to outcomes (conditional on the observed characteristics). By definition, unconfoundedness does not hold under this assignment mechanism.

As a consequence, the methods that work in the case of selection on observables generally fail to provide valid estimates of the ATE in the case of selection on unobservables. Instead, an instrumental variables ("IV") technique must be used. IV requires that one or more of the observable characteristics that are related to assignment of treatment be independent of the outcomes. Such a characteristic is called an "instrument" and, because changes in its value result in exogenous changes in the assignment of treatment and thereby outcomes, an estimate of an average effect of treatment can be teased out for a particular subset of the population by comparing the outcomes of subjects with different values of the instrument. However, in general, the IV approach is not able to estimate the ATE for the overall population.[12]

III. ADVANTAGES AND DISADVANTAGES OF "NATURAL EXPERIMENTS" AND APPLICATION IN ANTITRUST

The treatment effects approach, applied to natural experiments, has several advantages relative to the "structural model" approach traditionally used in the economics. For example, the treatment effects approach does not generally require the specification of a full structural model of the economic phenomenon in question, which may involve a number of equations and modeling choices. With the treatment effects approach, one need only specify a "reduced form" equation for the outcome and, if unconfoundedness is thought not to hold, identify an instrument.

Thus, the treatment effects approach has the potential to be more robust than structural models. Moreover, the treatment effects approach requires much less complex econometric techniques, as compared to structural models that often require substantial econometric

[11] The basic idea of the more sophisticated techniques is as follows. Suppose there is a single observed characteristic X that takes on the value of HIGH or LOW. Both outcomes and the incidence of treatment tend to be higher for subjects with X=HIGH than those with X=LOW. For the subset of subjects with X=HIGH, calculate the difference between the average outcome for the treated subjects and the average outcome for the control subjects. This procedure provides a good estimate of the average treatment effect for the subset of subjects with X=HIGH, call it ATE(X=HIGH), because it controls for the effects of X by conducting the treated versus control comparison within the subset of subjects for which X has the same value. The overall ATE can be obtained by taking a weighted average of ATE(X=HIGH) and ATE(X=LOW) (obtained in a similar fashion), using the relative frequency of X=HIGH and X=LOW among subjects as the weights. Under certain assumptions, this technique can be implemented in the familiar regression framework where the dependent variable is the outcome and the explanatory variables include the observed characteristics and a treatment indicator variable.

[12] For example, suppose the instrument takes on two values. IV using this instrument estimates the average treatment effect for the subset of the population whose treatment assignment would change if the value of the instrument were changed from one of its values to the other. In general, because of the assumption of heterogeneity, IV in the treatment effects context has different properties than in the context of "traditional" structural simultaneous equations models (*see*, e.g., James J. Heckman & Edward Vytlacil, *Structural Equations, Treatment Effects, and Econometric Policy Evaluation* 1, 73 ECONOMETRICA 669 (2005)).

knowledge and experience to implement. For these reasons, and because of the appeal of the analogy to randomized experiments, the treatment effects approach has become widespread in the economics profession.

Recently, however, some economists have pointed out various disadvantages of the treatment effects approach relative to the structural model approach. First, because it is a reduced form rather than structural, the treatment effects approach generally is not informative about the nature of the underlying structural economic relationships that determine outcomes. This limits the usefulness of the results of a treatment effects analysis, making it much more *sui generis*. For example, the treatment effects approach may provide a useful way to estimate the amount by which prices changed after a particular merger. However, the results would not be helpful for understanding the deeper structural mechanisms governing demand and supply that determined why and by how much prices changed.

Without a structural model, it would be difficult to extrapolate from the experience represented by the particular natural experiment that was analyzed (e.g., a consummated merger) to another hypothetical treatment (e.g., a contemplated merger in the same industry). In contrast, a structural model, because it specifies the entire nature of the economic system, can be used to predict effects of a hypothetical treatment that has not yet been experienced in practice.

For similar reasons, the result from a natural experiment may not be directly on point for the research question of interest. As an example, consider the *Staples* and *Whole Foods* cases. A key question was whether retailers within a given category ("office superstores" in the case of *Staples* and "organic food supermarkets" in the case of *Whole Foods*) were sufficiently differentiated from other retailer categories (e.g., Wal-Mart in the case of *Staples* or regular supermarkets in the case of *Whole Foods*) that a merger between two retailers within the category would decrease competition sufficiently to allow the merged firm to increase prices.

To address this question, analyses were performed to determine whether prices in a given local market were lower when there were, say, two within-category competitors present in the local market as compared to one. This can be thought of as a natural experiment, where the "subjects" were the local markets, the "treatment" was having a second within-category competitor present, and the "outcome" was the price. Applying the treatment effects approach in this context may have provided a good estimate of the average effect on price of such entry. However, the research question was how much prices would increase in a local market after the proposed merger between two within-category firms. As numerous commentators have pointed out, while related, the natural experiment of entry does not line up directly with the research question of the effect of the proposed merger.[13] Specifically, after the proposed merger, the merged firm may not close down one of its stores in each overlap local market and, even if it did, the effect of such closure may not be to reverse of the effect of the previous entry.

Another drawback of the treatment effects approach is that, even if the natural experiment itself is closely related to the research question, the treatment effects estimation techniques do not always allow estimation of a quantity that is directly relevant to the research

[13] *See*, e.g., Gregory K. Leonard & Lawrence Wu, *Assessing the Competitive Effects of a Merger: Empirical Analysis of Price Differences Across Markets and Natural Experiments*, 22 ANTITRUST 96 (2007).

question. For example, as discussed above, IV techniques generally produce an average treatment effect for a subset of the population, not the whole population. This subset may not be of particular interest and it may not be appropriate to extrapolate the result for the subset to the whole population.

On the other hand, sometimes the quantity that can be estimated turns out to be of interest. For example, consider again the *Staples/Whole Foods* analyses. For reasons discussed below, it may be that the treatment effects approach allows estimation of the average treatment effect only for the subset of local markets where entry had already occurred. However, this subset constitutes the set of local markets where the merger would have a competitive effect, at least in the short run, because, in other local markets, entry had not yet occurred and therefore prices would not change as a result of the merger.[14] Thus, in this particular instance, the limitation to a subset of the population aligns well with the research question.

A somewhat more subtle point is related to the heterogeneity among subjects in the magnitude of the treatment effect. One can think of there being a distribution of treatment effects in the population. The treatment effects approach at best is generally informative only about the average of this distribution. However, other characteristics of the distribution may be of interest. For example, in an antitrust class action, a relevant question during the class certification phase is whether all proposed class members were injured, i.e., whether the treatment effect associated with the alleged conspiracy was greater than zero for all proposed class members. The average treatment effect (which is what can be estimated by the treatment effects approach) does not address the relevant question because, for example, the average treatment effect could be greater than zero, while a substantial fraction of the proposed class may have had zero treatment effects.

Finally, the assumptions required for the treatment effects approach to provide valid estimates may not be justified in many cases. For example, unconfoundedness is a strong assumption. It holds only when there are no unobserved characteristics of subjects that are related both to the assignment of treatment and the outcomes (conditional on the observed characteristics).

And when subjects themselves make choices that determine or influence the assignment of treatment, the possibility that the size of the treatment effect for a subject influences the subject's choices must be taken seriously. For example, a retailer deciding which local markets to enter would likely consider the effect of its entry on market prices, preferring to enter in those markets where the post-entry price would remain high. In making its entry decisions, the retailer may well consider information not observable to the economist. In this situation, uncounfoundedness likely does not hold because the assignment of the treatment (entry) to subjects (the local markets) is done on the basis of unobservables that also influence the outcomes (prices). IV techniques may then be used.

However, it may be difficult to justify that a particular observable characteristic is a valid instrument, i.e., is both related to the assignment of treatment and unrelated to the outcomes

[14] In the longer run, when one or the other of the firms might enter into other local markets, there could be competitive effects in these local markets as well.

(conditional on the other observable characteristic). In the case of the *Staples/Whole Foods* analyses, a potential instrument might be how far a local market is from the retailer's headquarters.[15] This characteristic would be a valid instrument if it influenced the retailer's decision to enter a local market, but did not influence its prices.[16]

Often it is not straightforward to assess whether unconfoundedness holds or whether a variable is a valid instrument. Ironically, in making this assessment, one often needs to think about the underlying structural economic model. Therefore, application of the treatment effects approach does not generally allow one to avoid having to think about the underlying economic principles.

IV. CONCLUSION

The treatment effects approach has led to significant changes in the way economics research is conducted. However, recently critical re-evaluations of the approach have appeared in the literature bringing greater clarity regarding its disadvantages. Antitrust analysis based on natural experiments should pay heed to the warnings that have come out of this re-evaluation. In particular, there is no escaping the need to understand the underlying economics. Only when evaluated in the context of the pertinent economics can natural experiments provide useful and reliable information relevant to the research question at hand.

[15] *See*, e.g., Panle Jia, *What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry*, 76 ECONOMETRICA 1263 (2008).

[16] The idea would be that a local market on the boundary of the existing footprint would have both lower fixed costs of entry and lower distribution costs after entry occurred. The distribution cost is a marginal cost that likely affects pricing, while the fixed costs of entry affect entry, but not subsequent pricing. Thus, the boundary variable likely would be a valid instrument only if distribution cost was separately included as an explanatory variable for price.