RIGOROUS ANALYSIS OF CLASS CERTIFICATION COMES OF AGE

John H. Johnson
Gregory K. Leonard*

Since 2007, significant legal changes have reshaped the standards courts apply for the certification of classes in antitrust litigation.[1] Two articles in this issue of the *Antitrust Law Journal* provide some valuable new thinking on empirical economic approaches to class certification in response to this new environment.[2] In *The Economics of Common Impact in Antitrust Class Certification*, Paul Johnson seeks to "inform[ ] this debate over the appropriate economic standard by presenting an economic foundation for the study of the legal requirement of 'common impact.'"[3] He suggests a taxonomy for the factors affecting prices, according to which a factor is categorized as either common or not common (hereafter "noncommon") and either conduct-related or non-conduct-related. He then proposes a test for identifying common impact that does not require individualized data or analysis. He also explores the

* John H. Johnson is President and CEO of Edgeworth Economics, an Affiliated Professor at Georgetown University's Public Policy Institute in Washington, D.C., and an Assistant Editor of this *Journal*. Gregory K. Leonard is a Senior Vice President at NERA Economic Consulting in San Francisco, CA, and an Associate Editor of this *Journal*. Dr. Johnson was a consultant to Arkema, Inc. in both *In re Hydrogen Peroxide Antitrust Litigation* and *In re Plastics Additives Antitrust Litigation.* We thank Fei Deng and Laila Haider for helpful discussions and Justin Henderson for helpful research assistance.

[1] *See In re* Hydrogen Peroxide Antitrust Litig., 552 F.3d 305 (3d Cir. 2008); *see also In re* Plastics Additives Antitrust Litig., No. 03-CV-2038, 2010 U.S. Dist. LEXIS 90135, (E.D. Pa. Aug. 31, 2010); *In re* Graphics Processing Units Antitrust Litig., 253 F.R.D. 478, 483 (N.D. Cal. 2008). For an economist's view on the changing standards, see John H. Johnson & Gregory K. Leonard, *Economics and the Rigorous Analysis of Class Certification in Antitrust Cases,* 3 J. Competition L. & Econ. 341 (2007).

[2] Other recent articles by economists that discuss issues related to class certification include Steven R. Peterson & Andrew Y. Lemon, *Rigorous Analysis to Bridge the Inference Gap in Class Certification*, 7 J. Competition L. & Econ. 93 (2011); Pierre Cremieux et al., *Proof of Common Impact in Antitrust Litigation: The Value of Regression Analysis*, 17 Geo. Mason L. Rev. 939 (2010); Richard Schmalensee, *Economic Analysis of Class Certification*, Global Competition Pol'y, Spring 2008, Vol. 6, No. 2.

[3] Paul A. Johnson, *The Economics of Common Impact in Antitrust Class Certification*, *supra* this issue, 77 Antitrust L.J. 533, 535 (2011).

usefulness of correlation and cointegration analysis in the class certification context.

In *Correlation and Regression Analysis in Antitrust Class Certification*, Michelle M. Burtis and Darwin V. Neher similarly assess the usefulness of price correlation analysis to answer the question of whether impact can be determined for each class member using a common method or common evidence.[4] They conclude that "[t]here are drawbacks associated with the use of price correlations to draw inferences about common impact."[5] In this conclusion, they part ways with Paul Johnson, who, despite acknowledging various potential issues, finds that correlation analysis and related approaches "provide[ ] a tangible, well-defined, and objective foundation on which an expert can rely."[6]

As economists, we are quite optimistic about the ability of empirical testing to provide useful guidance to courts in class certification proceedings.[7] To advance the debate, in this Comment we attempt to incorporate the ideas

---

[4] Michelle M. Burtis & Darwin V. Neher, *Correlation and Regression Analysis in Antitrust Class Certification*, *supra* this issue, 77 ANTITRUST L.J. 495, 497–98 (2011).

[5] *Id.* at 497. Burtis and Neher also critique the use of regression models as a common method for determining impact. *Id.* at 498–99.

[6] Johnson, *supra* note 3, at 562.

[7] In this regard, we believe that Paul Johnson has misinterpreted our prior work when he claims that it paints "a somewhat bleak picture for empirical analysis" in class certification. *Id.* at 550. Quite to the contrary, our view is now and always has been that empirical tests are necessary for rigorous analysis of class certification. *See* Johnson & Leonard, *supra* note 1, at 350–51 (describing tests to determine whether common evidence can be used to determine impact for all class members). Paul Johnson also misinterprets our views when he states that we would "requir[e] a legal standard of proof of common impact that is not possible to achieve . . . ." Johnson, *supra* note 3, at 550. To put it in Paul Johnson's terms, we are simply proposing that courts require tests of the null hypothesis of common impact that have reasonable power to reject the null hypothesis if it were false (rather than "presuming" the null hypothesis is true without requiring any rigorous testing at all), and we point out that, to have reasonable power, such tests often require individualized analysis. This harks back to the famous question in statistics of how strongly one should embrace the null hypothesis when it has not been rejected by a statistical test. Certainly one should not feel confident about a null hypothesis that survives a test when the test had little power to reject the null hypothesis in the first place. Paul Johnson's argument that requiring plaintiff experts to "prove" common impact is asking too much because "common impact cannot be proved in a scientific sense" (*see, e.g.*, Johnson, *supra* note 3, at 548) is off-target because it fails to consider the wider decisionmaking framework within which empirical testing plays a role. As the statisticians Jerzy Neyman and Egon Pearson recognized, the outcome of a statistical test often forms the basis for someone to make a decision or take an action. J. Neyman & E.S. Pearson, *The Testing of Statistical Hypotheses in Relation to Probabilities a priori*, 29 MATHEMATICAL PROC. CAMBRIDGE PHIL. SOC'Y 492 (1933). Thus, while a statistical test that fails to reject the null hypothesis does not "prove" the null hypothesis to be true, a decisionmaker (having to go one way or the other) may, given the test result, proceed as if the null hypothesis were true. For example, a test that fails to reject the null hypothesis of common impact may lead a court to decide to certify a class. Thus, as we noted earlier, it is important that a court consider whether the test has reasonable power to reject the null hypothesis. An expert can help the court assess the power of a test.

presented by the other articles in this issue into a parsimonious econometric framework. Using the heterogeneous treatment effects literature in econometrics as a starting point, we synthesize ideas in a way that is consistent with the current legal standard and provides valuable testable propositions in the class certification context. This econometric framework is consistent with our 2007 article and our long-held view that empirical work is a valuable tool in analyzing class certification. It also provides a way to reconcile the approach of our 2007 article with the approaches suggested by the two articles in this issue.[8]

I. THE CHANGED LEGAL STANDARD

When we wrote our 2007 article we were motivated, in part, by an emerging trend in judicial decisions suggesting "that courts are starting to take a harder look at whether classes should be certified in Section 1 cases."[9] Four years later, in fact, the standards applied in antitrust class actions have dramatically changed. A series of influential decisions have rejected a longstanding presumption of injury in antitrust price-fixing cases and courts now regularly conduct a "rigorous" analysis of evidence at the class certification stage to determine whether the requirements of the Federal Rules of Civil Procedure are met. The Third Circuit decision in *In re Hydrogen Peroxide Antitrust Litigation* reshaped the standard for antitrust class actions in the price-fixing context.[10]

The core question in class certification is whether common impact can be established for each class member using common evidence. Despite its widespread usage, there continues to be confusion over the precise meaning of this standard. A plaintiff in the class certification phase seeks to show whether or not the same evidence can be put forward for all (or substantially all) class members if the case were to move forward to a liability and damages phase. This inquiry—whether evidence common to all class members predominates over individual issues—corresponds to the Rule 23(b)(3) predominance requirement.[11] In *Hydrogen Peroxide*, the Third Circuit recognized that the class certification phase is an evidentiary proceeding, and, accordingly, it emphasized the appropriateness of weighing competing economic expert testimony (including the resolution of factual disputes between experts) in ultimately

---

[8] While disagreeing with Paul Johnson on several important points discussed in detail below, we note that we also agree with a number of his points. For example, we agree that it is important to avoid confusing average impact with impact to every class member. *Compare* Johnson, *supra* note 3, at 544–45, *with* Johnson & Leonard, *supra* note 1, at 350. We also agree that it is important not to assume impact while assuming plaintiffs' allegations are true. *Compare* Johnson, *supra* note 3, at 546–47, *with* Johnson & Leonard, *supra* note 1, at 343 n.6.

[9] Johnson & Leonard, *supra* note 1, at 341.

[10] *In re* Hydrogen Peroxide Antitrust Litig., 552 F.3d 305 (3d Cir. 2008).

[11] FED. R. CIV. P. 23(b)(3).

determining whether impact can be demonstrated on a classwide basis using common evidence.[12]

Shortly after its *Hydrogen Peroxide* decision, the Third Circuit remanded another antitrust price-fixing case—*In re Plastics Additives Antitrust Litigation*—to the district court for reconsideration in light of the new standard elucidated in *Hydrogen Peroxide*.[13] *Plastics Additives* is one of the first applications of the *Hydrogen Peroxide* standard and is important for the guidance it provides with respect to the practical economic analysis that courts will look to when weighing the appropriateness of class certification.

In *Plastics Additives*, the district court focused its attention on a few key types of economic evidence. First, the court ruled that simply referencing list prices or price increase announcements that purportedly applied to all class members was not sufficient to demonstrate antitrust impact to all class members. Instead, the court weighted heavily the actual prices paid by some customers, which "[had] no relationship with Defendants' price increases."[14] Second, the court rejected the plaintiffs' analysis of "market structure" based on the interchangeability of various products at issue and the defendants' combined large market share in the industry. Here, the court ruled that these characteristics did not actually exist in the markets for the products at issue, citing economic evidence that "class members were able to obtain lower prices from Defendants by threatening to buy from non-Defendant producers" and actual evidence that the defendants "lowered prices to compete for sales."[15]

The court also considered a wide range of regression evidence. In particular, the court explained:

> [I]n resolving the question of whether common issues predominate on the element of impact, we need go no further than expose the regressions' fundamental failure: Plaintiffs must show that individual injury is capable of proof common to the class, but . . . regressions tell us nothing about individual class member experience.[16]

Here, the plaintiffs' expert's classwide regression methods were not applicable to any individual class members and the plaintiffs' expert did not run any regressions at the individual class-member level to determine if the classwide

[12] For a description of the potential implications of the *Hydrogen Peroxide* decision, see John H. Johnson, *The Standards for Class-Certification Expert Testimony After* In re Hydrogen Peroxide Antitrust Litigation, 2009 ANTITRUST REP., no. 2, at 3, 7.

[13] *In re* Plastics Additives Antitrust Litig., Nos. 07-2159 & 07-2418, 2009 U.S. App. LEXIS 2177, at *2 (3d Cir. Jan. 27, 2009).

[14] *In re* Plastics Additives Antitrust Litig., No. 03-CV-2038, 2010 U.S. Dist. LEXIS 90135, at *24 (E.D. Pa. Aug. 31, 2010).

[15] *Id.* at *45–46.

[16] *Id.* at *55–56.

regressions were broadly applicable.[17] The court's conclusion—that the plaintiffs' expert needed to demonstrate through empirical testing that his class-wide regression model reliably represented the "experience" of individual class members—epitomizes the trend towards more rigorous analysis that courts have taken over the past several years.

II. APPLYING LEARNING FROM THE TREATMENT EFFECTS LITERATURE

In the last twenty years, the "treatment effects" literature in economics has focused on methods for determining the causal effect of a "policy intervention" (the "treatment") on economic outcomes.[18] Such methods have been applied, for example, to determine the effect of 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.[19] One key insight in this literature is that the ability to identify reliably a causal effect requires careful econometric specification and a mechanism to distinguish between causation and correlation.

The treatment effects literature uses as a starting point the familiar experimental design framework, where an experimental "treatment" is applied to a group of "subjects," called the "treated group." The effect of the treatment is determined by comparing the outcomes for the treated group to the outcomes for a second group of subjects, called the "control group," to whom the treatment was not applied. If subjects have been randomly assigned to the treatment and control groups, this comparison in general will provide a valid method for identifying the causal effect of the treatment.

Perhaps the most familiar example of random experimental design occurs in clinical testing of pharmaceuticals, where groups of patients are randomly assigned to take either the drug or a placebo. Economists rarely have the luxury of randomly assigning individuals to a "treatment" group to measure causal effects of interest, however. The methods developed in the treatment effects literature seek to overcome the complications caused by the inability to randomly assign subjects to treatment and control groups.

A price-fixing conspiracy fits well within the treatment effects paradigm: the price-fixing conspiracy is the "treatment," the customers are the "subjects," the "economic outcome" is a customer's purchase price, and the "effect of the treatment" is the "impact" of the conspiracy on the customer's purchase

---

[17] *See id.* at *57.

[18] *See generally* Guido W. Imbens & Jeffrey M. Wooldridge, *Recent Developments in the Econometrics of Program Evaluation*, 47 J. ECON. LIT. 5 (2009).

[19] *See id.* at 6–7.

price. In fact, the methods commonly used to estimate the impact of a price-fixing conspiracy, such as the "before-after" method, have parallels in the treatment effects literature.[20]

How can the learning from the treatment effects literature be applied to the issues that arise in a class certification proceeding? To start, a fundamental premise of the treatment effects literature is that the effects of a treatment may be heterogeneous across subjects, and any econometric technique for estimating a treatment effect must take this heterogeneity into account.[21] Analogously, in the language of class certification, the impact of a conspiracy may well be heterogeneous across customers, and an econometric technique proposed by a plaintiffs' expert for determining impact and damages must adequately account for this heterogeneity. Allowing for heterogeneity in the effect of a treatment is inherent in the "counterfactual" framework that has been widely adopted in the treatment effects literature.[22] In this framework, the effect of the treatment on the outcome p for an individual subject i is written as:

$$p_i(1) - p_i(0)$$

where $p_i(1)$ is the outcome for subject i if it receives the treatment and $p_i(0)$ is the outcome for subject i if it does not receive treatment. For example, in the price-fixing setting, $p_i(1)$ would be customer i's price given the conspiracy, and $p_i(0)$ would be customer i's price absent the conspiracy (i.e., the "but-for" price). Note that, as economists, we never observe both an individual customer's price in a world where a price-fixing conspiracy occurred and in the world where a price-fixing conspiracy did not occur. Thus, we must rely upon econometric analysis to determine the but-for price.

The counterfactual framework allows the treatment effect to vary across subjects. Indeed, the treatment effect may be zero for some subjects, positive for other subjects, and negative for others. For example, a job training program may have no effect for some workers, benefit other workers, and may harm still others (if, for example, their skills decline while receiving useless training).[23]

---

[20] ABA SECTION OF ANTITRUST LAW, PROVING ANTITRUST DAMAGES: LEGAL AND ECONOMIC ISSUES 57–58 (2d ed. 2010).

[21] Imbens & Wooldridge, *supra* note 18, at 7.

[22] *Id.* The counterfactual framework is largely conceptually the same as the "but-for versus actual" framework used to estimate antitrust damages.

[23] In some cases where treatment effects methods have been applied, only post-treatment outcomes are observed. In that case, the estimation methods involve comparing outcomes from the treated group to the outcomes from the control group, with adjustments made for differences in characteristics between the two groups. In this situation, the effect of the treatment on an individual subject cannot be determined. Only the "average treatment effect" or the "average treatment effect on the treated" can be determined. When pre- and post-treatment outcomes are observed

Explicitly allowing for heterogeneity is not limited to the treatment effects literature. For example, in the consumer demand literature, models have been developed where the effects of product attributes, including price, on consumers' demands are allowed to be heterogeneous over consumers.[24] Empirically, such heterogeneity has been found to be quite important. For example, demand studies have found that some customers exhibit very little or no sensitivity to price, while others are very price-sensitive.[25]

Economic science in the last twenty years has made great strides in explicitly accounting for heterogeneity in econometric analysis of the effects of treatments and, as an empirical matter, has found such heterogeneity to be an important feature of markets. As a result, presumptions of classwide impact that courts have traditionally applied are not appropriate. In fact, it would seem that the burden of establishing the *absence* of heterogeneity in a price-fixing case (i.e., the existence of common impact) should rest on plaintiffs. Be that as it may, a class should not be certified without a rigorous investigation into whether class members exhibit substantial heterogeneity with respect to the impact of the alleged conduct.

III. ECONOMETRIC MODELS AND COMMON EVIDENCE

In his approach to class certification, Paul Johnson proposes categorizing factors that affect price as either "common" or "individual."[26] He defines a "common" factor as "a determinant of price that, if it affects the price paid by one putative class member in a certain way, necessarily affects prices paid by all putative class members in the same or similar way."[27] In contrast, he defines an individual factor as "a determinant of price that affects price but at the same time is not 'common.' "[28] He also proposes categorizing factors depending on whether they are "conduct" factors—factors that are within the control of the defendant—or "non-conduct" factors—factors that are not controlled by the defendant.[29] Paul Johnson argues that this taxonomy of factors

---

for all subjects, however, it is possible (under certain assumptions) to determine the treatment effect on an individual subject by applying methods similar to the before-after method used in price-fixing cases. But, in general, determining the treatment effect for each individual subject requires individualized analysis.

[24] *See, e.g.*, Jerry A. Hausman & Gregory K. Leonard, *Economic Analysis of Differentiated Products Mergers Using Real World Data*, 5 GEO. MASON L. REV. 321 (1997).

[25] *See, e.g.*, Aviv Nevo, *Measuring Market Power in the Ready-to-Eat Cereal Industry*, 69 ECONOMETRICA 307, 329 (2001).

[26] *See, e.g.*, Johnson, *supra* note 3, at 536.

[27] *Id.*

[28] *Id.*

[29] *Id.* at 538.

can be used to set up statistical tests based on "natural experiments" and correlation (and cointegration) analysis.[30]

In implementing Paul Johnson's approach, one must go one step farther and, as the treatments effects literature underscores, distinguish between the *factors* themselves and the *effects* of those factors. A *factor* may be common (the same for all class members, e.g., the cost of an input to production) or individual (in that it differs across customers, e.g., a downstream demand driver that is different for customers in different downstream industries). But an *effect* of a factor may also be common or individual. Even a factor that is common to all customers (e.g., the cost of an input) may well have an effect on price that varies across customers (e.g., different customers' prices respond differently to a change in the cost of the input). Conversely, a factor that varies across customers, for example, the customer's size, may have an effect (defined as the change in the customer's price for a one unit change in the level of the factor for that customer) that is equal in magnitude for all customers.

A. Determinants of a Customer's But-For Price

To clarify these ideas, it is helpful to express them in mathematical terms. Suppose the but-for price is generated according to the following equation:[31]

$$p_{it}(0) = X_{it}\beta_i + Z_t\gamma_i + \mu_i + \varepsilon_{it} \tag{1}$$

The left-hand-side variable is price (or perhaps the logarithm of price) for customer i at time t in the absence of a conspiracy.[32] Note that the presence of an "i" subscript indicates that the factor itself differs across customers, whereas the absence of an "i" subscript indicates that the factor is the same for all customers.

The but-for price is a function of both observed and unobserved factors. The observed factors include individual factors that change over time X_{it}, such as the customer's size and available substitutes, and common factors that change over time Z_t, such as input costs and GNP.

The effects of these observed factors are represented by the coefficients β_i and γ_i, indicating that the effect of the factor may differ across customers. For example, assume a key input in the production of a consumer product is a rubber polymer. Producers sell the consumer product to retailers. Even though

---

[30] *Id.* at 544–66.

[31] We discussed a similar equation and the implications for class certification in our 2007 article. *See* Johnson & Leonard, *supra* note 1, at 349.

[32] We have used the "counterfactual" framework of the treatment effects literature, indicating that this is the equation for the but-for price with the "0" in parentheses.

all producers use the rubber polymer in producing the consumer product, the price each retailer pays to a producer for the consumer product may change differently in response to a change in the cost of the rubber polymer. For example, a retailer that has a great deal of buying power (e.g., Walmart) may refuse to accept any price increase from producers following an increase in the cost of the rubber polymer price. On the other hand, a small retailer with limited buying power may see its price for the consumer product increase by the full amount of the increase in the rubber polymer cost.

Price is also a function of the effects of unobserved factors. The effects of unobserved individual factors that are constant over time are represented by μ_i, and the effects of unobserved individual and common factors that change over time are combined into the term ε_{it}.[33]

Again using the counterfactual framework of the treatment effects literature, let the impact of the conspiracy for customer i at time t be δ_{it}, which is defined as

$$\delta_{it} = p_{it}(1) - p_{it}(0) \tag{2}$$

where $p_{it}(1)$ is customer i's actual price at time t, as indicated by the "1" in parentheses (i.e., the price the customer paid during the alleged conspiracy). Note that this formulation allows for the possibility that each customer (and, indeed, time period) sustained a different degree of injury, which may be zero or even negative (in which case there is no impact).[34] Again, allowing for this heterogeneity of effects is consistent with the treatment effects literature.

We can rearrange to express the actual price as being equal to the but-for price plus the effect of the conspiracy:

$$p_{it}(1) = p_{it}(0) + \delta_{it} \tag{3}$$

Given the but-for price equation (1) and the assumption of heterogeneous effects in equation (2), reliable estimation of the impact for each customer (and time period) can be achieved using the following steps:

1. Estimate the but-for price equation for each customer i using its data from outside the conspiracy period;[35]

[33] There is nothing gained for exposition purposes by writing the effect of an unobserved factor as the product of a coefficient (the effect) and the value of the factor. The inability to observe the values taken on by a factor means that the coefficient and the factor cannot be separately determined. Instead, we represent the unobserved factor and its effect with a single variable that is implicitly the product of a coefficient and the value of the factor.

[34] Impact is binary—either the class member was impacted (injured) by the conduct or not. Damages are the extent of the injury. There is impact if and only if damages are greater than zero.

[35] Fitting the but-for price equation to data from outside the alleged conspiracy period provides a reliable estimate of the but-for price provided that no structural change unrelated to the

2. Use the but-for price model for customer *i* to predict customer *i*'s but-for prices during the conspiracy period; and
3. Take the difference between customer *i*'s actual prices during the conspiracy period and the predicted but-for prices.

The procedure described above is inherently individualized. It requires that the but-for price be modeled separately for each customer, with the potential for different sets of individual factors to be included in the models for different customers. For example, the but-for price models for customers in different downstream industries may include different sets of demand factors. Moreover, this procedure requires that the coefficients on the observed factors be allowed to vary across customers. As described above, this focus on individual customer level regressions was discussed by the court in the *Plastics Additives* decision.[36]

B. Situations that Potentially Allow for Use of a "Common" But-For Price Model to Determine Impact for Each Class Member

We now consider possible situations where the but-for price is determined in a more simplified way than equation (1) and, as a consequence, possibly would allow a "common" approach to modeling the but-for price.[37] Laying out these situations provides a basis to better understand the various tests that have been proposed for when a common approach to determining impact can be used. The first possible situation is if the but-for price is based only on observed common factors with the effects of those factors the same for all customers.[38] In this case, the but-for price is generated according to the equation

$$p_{it}(0) = Z_t\gamma + \varepsilon_{it} \tag{4}$$

Note that, in equation (4), the coefficient γ on the common factors does not vary across individuals or over time.

---

alleged conduct occurred in the price determination process during the alleged conspiracy period. Otherwise, the period outside the alleged conspiracy period may not provide a reliable guide to what should have happened during the alleged conspiracy period.

[36] *In re* Plastics Additives Antitrust Litig., No. 03-CV-2038, 2010 U.S. Dist. LEXIS 90135, at *55–72 (E.D. Pa. Aug. 31, 2010).

[37] One additional situation in which the impact question can be answered using a common regression model is if the impact (or lack thereof) is known to be the same for each class member. In that case, only one class member (or the "average" class member) needs to be analyzed in order to determine whether all of the class members were impacted. However, as with the other situations discussed here, a test of the assumption of equal impact for all class members generally requires an individualized analysis, which leads back to the common proof paradox.

[38] This case would be included in Paul Johnson's "common non-conduct factors" case, where, again, he has not distinguished between factors and effects. *See supra* Part III; *see also* Johnson, *supra* note 3, at 541–42.

The second possible situation is if the but-for price is based on observed factors, both common and individual, with the effects of those factors the same for all customers.[39] In this case, the but-for price is generated according to the equation

$$p_{it}(0)=X_{it}\beta+Z_t\gamma+\varepsilon_{it} \tag{5}$$

Note that, in equation (5), coefficients γ and β do not vary across individuals or over time.

The third possible situation is if the but-for price depends on observed factors with the same effects across customers plus unobserved individual factors that are constant over time.[40] In this case, the but-for price is generated according to the equation

$$p_{it}(0)=X_{it}\beta+Z_t\gamma+\mu_i+\varepsilon_{it} \tag{6}$$

The reason that each of these situations allows for the possibility of a "common" approach to the but-for price is that in each case a but-for price model could be estimated using one customer (or group of customers) and then be used to determine impact for another customer, although often with some additional individualized analysis specific to that customer.[41] This is not true if the but-for price is determined by the more general equation (1). In that case, the but-for model estimated on one customer cannot be applied reliably to determine impact for another customer because each customer has a different set of observed individual factors and different coefficients (i.e., effects of the factors).

It should be noted that even if the but-for price can be modeled in a common fashion, as long as effects are heterogeneous it is possible that some class members might have been injured and some might not have been. If true, this

[39] This case would also be included in Paul Johnson's "common non-conduct factors" case. Johnson, *supra* note 3, at 541–42.

[40] This case does not appear to easily fit within Paul Johnson's framework. In particular, as discussed *infra* note 41, even if properly specified, this but-for price model would not provide a common method for determining impact for any proposed class member that made purchases only within the alleged conspiracy period.

[41] We are not aware of a court ruling on whether the required additional analysis specific to a customer is sufficiently complex that it rises to an individualized analysis that renders class certification inappropriate. To illustrate, suppose the but-for price was governed by equation (6). Suppose further that we estimated β and γ using data for customer 1. If data on the observed factors and prices for customer 2 were available outside the conspiracy period, we could use such data (along with the estimates of β and γ) to estimate μ_2, which is needed to assess impact on customer 2. Does this step rise to an individualized analysis? Arguably, it does. Moreover, if no data were available on the observed factors and prices for customer 2 outside the conspiracy period, it would not be possible to assess impact on customer 2 using a common approach. Finally, even if data on the observed individual factors were potentially available, it might take substantial individualized inquiry to obtain such data. Again, this may render the process of determining impact individualized even if a "common" but-for price model was appropriate.

possibility would be uncovered by the application of the common but-for price model, but our understanding of the case law is that a class should not be certified if a substantial number of the class members were not harmed.[42] Thus, even where appropriate, the common but-for price model might have to be applied at the class certification stage in order to pare down the class to those individuals who were harmed.

C. Mistaken Application of a "Common" But-For Price Model

What happens if an economist estimates one of the potentially common but-for models (4), (5), or (6) when, in fact, the but-for price is generated according to equation (1)? To answer this question, we consider a simplified example (the same basic conclusions carry over to the more general case). Suppose that the but-for price is actually generated by a simplified version of equation (1):

$$p_{it}(0) = \alpha + \gamma_i Z_t + \varepsilon_{it} \tag{1'}$$

where α is the "intercept," Z_t is a common factor, for example, a raw material cost, γ_i is the effect of the raw material cost, which is assumed to differ across customers, and ε_{it} are the unobserved factors (also commonly known as the regression "error term"). Thus, while the only factor in this model is common, the effect of this common factor is heterogeneous across customers. The economist mistakenly does not allow for the possibility of heterogeneity and instead implements the common but-for model based on equation (4):

$$\tilde{p}_{it}(0) = \alpha + \gamma Z_t + \varepsilon_{it} \tag{4'}$$

This common but-for model assumes (incorrectly) that all customers have the same coefficient on the observed common factor.[43]

Suppose further that there are only two customers. Then, the γ coefficient in the common but-for price model would be a weighted average of γ_1 and γ_2,

---

[42] This test seems to be consistent with Paul Johnson's approach as well. *See* Johnson, *supra* note 3, at 541–42.

[43] Burtis and Neher make the related point that regression coefficients are averages of underlying coefficients for individual class members or individual transactions. They go on to say that "there is no way to associate particular coefficient results with particular transactions or particular proposed class members." Burtis & Neher, *supra* note 4, at 518 n.53. We have a slight disagreement with this statement. It would seem to be going further than necessary to require that a plaintiff's expert recover coefficients for individual transactions—impact is generally assessed at the class-member level, not the transaction level. Second, it is possible to recover coefficients at the class-member level—by estimating individual class-member regression models. Indeed, Burtis and Neher seem to agree that this is feasible, at least in some cases. *See id.* at 519. However, the important point is that the analysis is individualized if separate regression models need to be specified and estimated for each class member.

lying between those two values.[44] The extent of injury for each of the two customers could be estimated using the common but-for price model by taking the difference between the actual price $p_{it}(1)$ and the but-for price from the common model $\tilde{p}_{it}(0)$. To see how accurate this approach would be, we can substitute for the actual and but-for prices using equations (1′) and (4′) to obtain:

$$p_{1t}(1)-\tilde{p}_{1t}(0)=\delta_{1t}+(\gamma_1-\gamma)Z_t+\varepsilon_{1t}$$
$$p_{2t}(1)-\tilde{p}_{2t}(0)=\delta_{2t}+(\gamma_2-\gamma)Z_t+\varepsilon_{2t}$$

For simplicity, we will assume that the two regression error terms ε_{1t} and ε_{2t} are both equal to zero.[45]

This equation shows that taking the difference between the actual price and the but-for price based on the incorrect common but-for price model leads to an assessment of the injury for each customer i (where i = 1 or 2) that is biased by an amount $(\gamma_i-\gamma)Z_t$. This bias, in turn, could lead to the mistaken conclusion that all customers were impacted when some were not. Suppose, for example, that Z_t is a raw material cost and has the value of \$0.50; that the effect of the raw material cost on the price paid is $\gamma_1=1$ for customer 1 and $\gamma_2=0.5$ for customer 2 (so that $\gamma=0.75$); and that customer 1 was not impacted by the conspiracy, that is, $\delta_{1t}=\$0$, but customer 2 was injured by an amount $\delta_{2t}=\$0.20$ (and thus was impacted). Under these assumptions, the differences between the actual prices and the respective but-for prices based on the common model for the two customers is:

$$p_{1t}(1)-\tilde{p}_{1t}(0)=0+(1-0.75)0.50=\$0.125$$
$$p_{2t}(1)-\tilde{p}_{2t}(0)=0.2+(0.50-0.75)0.50=\$0.075$$

Because the actual price exceeds the but-for price for both customers, the mistaken application of the common but-for model leads to the false conclusion that both customers were impacted, when in fact customer 1 was not impacted at all. We term this a "false positive"—the use of the incorrect common but-for price model leads to the mistaken conclusion of common impact.

It is also possible that the mistaken use of the common but-for price model can lead to a "false negative"—the mistaken conclusion that there is no com-

---

[44] The situation can, in fact, be considerably more complex than this when there are many customers, both common and individual factors in the model, correlations among these factors, and different numbers of observations for different customers, making mistaken inferences potentially even more likely.

[45] The ε_{it} terms do not induce any bias because they have zero mean conditional on the X variables. While they could induce false positives or false negatives in the impact assessment for a given class member in a given sample, this source of error would exist even if the correct individualized but-for price model was used. Thus, this source of error, unlike the $(\gamma_i-\gamma)Z_t$ terms, is not due to mistaken use of the common but-for model.

mon impact. To see this, consider the same example as above, but with both customers actually injured by the same amount of \$0.10 (i.e., $\delta_{1t} = \delta_{2t} = \$0.10$) (so that they were both impacted). Then, the comparison of the actual prices and the but-for prices based on the incorrect common model would yield the following results:

$$p_{1t}(1) - \tilde{p}_{1t}(0) = \$0.10 + (1 - 0.75)\$0.50 = \$0.225$$
$$p_{2t}(1) - \tilde{p}_{2t}(0) = \$0.10 + (0.50 - 0.75)\$0.50 = -\$0.025$$

Because the actual price is greater than the but-for price for customer 1 but less than the but-for price for customer 2, the mistaken application of the common but-for model leads to the incorrect conclusion that there was no common impact in this example.

In short, when a common but-for price model is not appropriate (i.e., does not fit the facts of the case), it does not provide a reliable method for determining whether class members were impacted. If individualized factors or effects are important, a common but-for model is econometrically misspecified and does not provide a valid estimate of the effects of the observed factors. Worse, however, is the potential that the misspecified model can produce incorrect answers about whether individual class members were impacted and thus incorrect conclusions about whether a class should be certified.

In light of the foregoing discussion, we now turn to Paul Johnson's suggested approach for testing common impact:

> In fact, a number of tests that do not require individual evidence (beyond electronic transaction-level data that are often available at the class certification phase) are naturally suggested by an understanding of common impact. Instead of studying how prices react to *individual* factors, these tests analyze how prices react to hypothesized *common* factors, such as whether all putative class members pay higher prices after the formation of a cartel or whether all putative class members pay lower prices after entry by a noncolluding firm. Thus, the advantage of these tests is that they eliminate the need for an analysis of individual factors (the fact of an alleged cartel meeting is common to all putative class members, as is entry of a noncolluding firm).[46]

He goes on to describe a specific version of his proposed test in more detail as follows:

> [A]n economist could specify an econometric model that explains prices paid as a function of all economically significant common factors . . . . [T]he economist could estimate the model over the time period when there was no alleged conduct and then use the estimate to predict prices into the time period when there was alleged conduct. . . . In an attempt to falsify common impact, the economist would calculate differences between these predicted prices and the actual prices paid . . . [and] identify the proportion of transac-

[46] Johnson, *supra* note 3, at 550.

> tions whose but-for price was equal to or below the actual price . . . . [If] [s]ubstantially all transactions were affected by the [alleged conduct]. . . [s]uch a finding should be viewed as an element of legal proof supporting the hypothesis of common impact.[47]

The notable aspect of Paul Johnson's approach is that he proposes estimating a common but-for model based only on common factors and then seeing whether the application of this model leads to estimates of the effects of the conspiracy that are of the same sign for all class members. If so, he would conclude that impact is common (and the common but-for model can appropriately estimate damages); if not, he would conclude that impact is not common or that individualized factors are too significant for a common approach to work.

Note, however, that Paul Johnson's proposed approach is identical to the one that we demonstrated above and can easily lead to false positives and false negatives. Estimating a common but-for model and comparing the resulting but-for prices to actual prices can give the wrong answer if the common but-for model is misspecified.

D. AVOIDING MISTAKEN APPLICATION OF A "COMMON" BUT-FOR PRICE MODEL

The discussion above demonstrates why the application of a common but-for price model when it does not fit the facts of the case can lead to unreliable conclusions concerning impact on a given class member. Thus, it is important to determine when a common but-for price model fits the facts of the case.

Note that each of the common but-for models (4)–(6) is a special case of equation (1), with particular parameter restrictions imposed. For example, equation (5) is a special case of equation (1) with the restrictions $\beta_i = \beta$, $\gamma_i = \gamma$, and $\mu_i = 0$ for all i. As we discussed in our 2007 article, these parameter restrictions can be tested using generally accepted econometric methods as follows.[48] First, estimate each customer's but-for price equation based on equation (1). Second, ask whether the restrictions implicit in equation (5) are consistent with the parameters that were estimated in the first step. If not, it is not possible to use the "common" but-for price model (5). The issue with

[47] *Id.* at 553–54.

[48] Johnson & Leonard, *supra* note 1, at 350–51. An alternative formulation of this test seems to avoid the common proof paradox, but does not. This alternative formulation involves first running the restricted model based on equation (5) and then asking whether the restrictions are tightly binding (the "LM" form of the test). However, for this formulation to have power, it is necessary to identify the factors for each class member for which the restrictions might potentially be binding. This involves the same individualized inquiry as running the separate but-for price models for each class member. Thus, use of the LM formulation (in a way that seeks to have good statistical power) does not avoid the common proof paradox.

performing this test is that it requires each customer's but-for price equation to be separately estimated. This is the common proof paradox we identified in our 2007 article.[49] However, while individualized inquiry is needed, it does not mean that econometric testing is unable to provide useful information to a court. To the contrary, it is precisely the empirical tests we have proposed that allow one to determine whether a common approach to estimating the but-for price will be reliable.[50] Notably, our tests are more accurate than the test proposed by Paul Johnson because they are based on properly specified models.

IV. CORRELATIONS AND RELATED METHODS

In many recent cases, plaintiffs' experts attempt to establish the existence of a "price structure" whereby prices for individual class members may be different but the price differentials stay constant over time so that customers' prices move together in response to demand and supply conditions.[51] Plaintiffs' experts conclude from the existence of a "price structure" that a common method may be used to assess impact for all class members.[52]

As we discussed in our 2007 article, initially some plaintiffs' experts attempted to demonstrate the existence of a "price structure" through visual inspection of price graphs.[53] We pointed out that the visual inspection methodology was junk science and, indeed, how the very concept of "price structure" was not one that was generally accepted in economics.[54] Subsequently, plaintiffs' experts started performing correlation analyses to demonstrate the existence of a "price structure," under the presumption that such an analysis had greater scientific merit than visual inspection.[55] Burtis and Neher, however, do an excellent job pointing out the flaws in a correlation analysis. Specifically, we highlight their important reminder that "examining price correlations is not an analysis of the economics of how prices are determined."[56]

Both Paul Johnson and Burtis and Neher reject the approach, often used by plaintiffs' experts, of simply calculating correlations with no economic con-

---

[49] Johnson & Leonard, *supra* note 1.

[50] Burtis and Neher make the important point that a model such as equation (5) should not be assessed by the goodness-of-fit measure R-squared. This is because the more important are the customer-specific unobserved factors, the greater is the R-squared of a fixed effects regression: a result they term the "diversity paradox." *See* Burtis & Neher, *supra* note 4, at 498.

[51] *See, e.g.*, John H. Johnson & Gregory K. Leonard, *In the Eye of the Beholder: Price Structure as Junk Science in Antitrust Class Certification Proceedings*, ANTITRUST, Summer 2008, at 108, 109–10.

[52] *See, e.g.*, *id.*

[53] Johnson & Leonard, *supra* note 1, at 348; *see also* Johnson & Leonard, *supra* note 51, at 108.

[54] Johnson & Leonard, *supra* note 1, at 348–51.

[55] Johnson & Leonard, *supra* note 51, at 109–10.

[56] Burtis & Neher, *supra* note 4, at 509.

text to prove co-movement in prices.[57] Paul Johnson, however, gives more credence to the possibility that correlation and cointegration can provide valuable guidance. We reiterate many of the points that Burtis and Neher emphasize as cautionary in the interpretation of correlations. In particular, plaintiffs commit a basic error in logic when they attempt to infer from a correlation between two prices, often driven by one or more common factors unrelated to the alleged conduct, that the alleged conduct itself would have a common impact. For example, as a matter of logic, the fact that two customers' prices move closely together because of large swings in a common raw material cost says nothing about whether a price-fixing conspiracy would have an impact on both customers' prices.[58] Burtis and Neher draw the distinction between a common factor that drives a correlation and common impact: "In a vacuum, this is simply an empirically based version of the *Bogosian* short-cut, whereby the researcher has established certain features of the defined marketplace and then draws the conclusion that an anticompetitive act would commonly impact all participants in that marketplace."[59]

Paul Johnson and Burtis and Neher, though in different ways, attempt to emphasize that for any statistical measure to be relevant, the economic linkage underlying it must be explicit. Paul Johnson also puts forward cointegration analysis as an alternative to correlation analysis. However, cointegration for the most part does not get around any of the issues raised with respect to correlation, and it only adds a degree of complexity that is unlikely to be helpful in many cases. Paul Johnson's summary of the potential pitfalls of cointegration is useful, and we advise caution to practitioners in attempting to use these methods in the class certification context.

The solution to the problems that can arise in the use of correlation and cointegration is to attempt to embed the analysis of individual transaction data into a causal econometric framework like the one we have specified above.

V. CONCLUSION

Economic science has influenced the treatment of antitrust class certification dramatically in the last five years. The two articles in this issue provide some interesting new thinking on the topic, and we find much to agree with in both articles. However, we disagree with Paul Johnson on the issue of empirical testing. He proposes a test for common impact in which a "common" but-

---

[57] *See id.*; *see also* Johnson, *supra* note 3.

[58] Paul Johnson's suggestion that a "natural experiment" involving customer responses to a non-conduct factor may be informative as to whether the alleged conduct had a common impact (*see* Johnson, *supra* note 3, at 554–56) seems to suffer from the same flaw in logic. The fact that all customers' prices respond to a large change in raw material cost does not mean that all of their prices would respond to a price-fixing conspiracy.

[59] Burtis & Neher, *supra* note 4, at 511.

for price model is estimated and then used to measure the extent of injury and the impact for each customer. If all class members are found to have positive injury (impact), he would have the court conclude that the common but-for price model could be used to determine impact and damages. However, we find that this procedure has serious problems with accuracy. In other words, the procedure can easily generate a false positive or a false negative when the common but-for price model is misspecified. Such errors in inference arise because the common but-for price model confuses heterogeneity in the effects of the common and individual factors and the impact of the conspiracy.[60] We have demonstrated this by considering Paul Johnson's test in the treatment effects framework.

In light of recent decisions, courts will continue to be conducting rigorous analyses of class certification in antitrust cases. As this occurs, careful application of econometric methodology will continue to be of paramount importance. Further empirical work in this area, such as the two articles in this issue of the *Journal*, make a welcome contribution to the field.

[60] While a finding of "injury" of different signs for different customers would correctly lead to the conclusion that the common model was not appropriate, it would not necessarily be correct to conclude that some class members were not harmed. As discussed above, with a mistakenly applied common but-for price model, it is possible for the calculated "impacts" to be negative for some class members even though they were in fact harmed.