

THE ECONOMICS OF COMMON IMPACT IN ANTITRUST CLASS CERTIFICATION

PAUL A. JOHNSON*

Recent U.S. courts of appeal decisions have implemented a fundamental paradigm shift in the standards applied to the analysis of class certification in antitrust litigation.¹ These decisions reject explicitly or implicitly the past status quo whereby courts could presume that class certification is appropriate in antitrust matters, and/or leave “merits” disputes unresolved, including disputes among experts so long as their opinions are not “fatally flawed.”²

* Principal, Bates White Economic Consulting, Washington D.C. I thank Ai Deng, Richard Higgins, Chip Miller, Ed Steegmann, Scott Thompson, Keith Waehrer, three anonymous referees, and the editors of the *Antitrust Law Journal* for their comments. All opinions and any errors are my own and not those of partners or employees of Bates White.

¹ See, e.g., *In re Hydrogen Peroxide Antitrust Litig.*, 552 F.3d 305, 307 (3d Cir. 2008); *In re Initial Public Offering Sec. Litig.*, 471 F.3d 24, 40 (2d Cir. 2006) (“Obviously, we can no longer continue to advise district courts that ‘some showing,’ of meeting Rule 23 requirements will suffice and that ‘findings’ are required, or that an expert’s report will sustain a plaintiff’s burden so long as it is not ‘fatally flawed,’ and that the plaintiff must prove Rule 23 requirements.”) (citations omitted); *Blades v. Monsanto Co.*, 400 F.3d 562, 575 (8th Cir. 2005) (“We have stated that in ruling on class certification, a court may be required to resolve disputes concerning the factual setting of the case. This extends to the resolution of expert disputes concerning the import of evidence concerning the factual setting—such as economic evidence as to business operations or market transactions.”); *Gariety v. Grant Thornton, LLP*, 368 F.3d 356, 366 (4th Cir. 2004) (“Thus, while an evaluation of the merits to determine the strength of plaintiffs’ case is not part of a Rule 23 analysis, the factors spelled out in Rule 23 must be addressed through findings, even if they overlap with issues on the merits.”); *O’Sullivan v. Countrywide Home Loans, Inc.*, 319 F.3d 732, 737–38 (5th Cir. 2003) (“The party seeking certification bears the burden of demonstrating that the requirements of rule 23 have been met.”); *Newton v. Merrill Lynch, Pierce Fenner & Smith, Inc.*, 259 F.3d 154, 187–88 (3d Cir. 2001) (“In a sworn declaration, plaintiffs’ expert provided no model formula, but instead projected that he could devise a formula that would measure damages among the class and serve as a plan for allocation. We are not convinced.”); *Szabo v. Bridgeport Machs., Inc.*, 249 F.3d 672, 675 (7th Cir. 2001) (“The proposition that a district judge must accept all of the complaint’s allegations when deciding whether to certify a class cannot be found in Rule 23 and has nothing to recommend it.”).

² The presumption that class certification is appropriate in antitrust matters appears to have arisen from the Third Circuit’s decision in *Bogosian v. Gulf Oil Corp.*, 561 F.2d 434, 455 (3d Cir. 1977) (“If, in this case, a nationwide conspiracy is proven, the result of which was to increase prices to a class of plaintiffs beyond the prices which would obtain in a competitive regime, an individual plaintiff could prove fact of damage simply by proving that the free market prices would be lower than the prices paid and that he made some purchases at the higher

The Third Circuit's opinion in *In re Hydrogen Peroxide Antitrust Litigation*³ encapsulates the rejection of the prior practice:

First, the decision to certify a class calls for findings by the court, not merely a "threshold showing" by a party, that each requirement of Rule 23 is met. Factual determinations supporting Rule 23 findings must be made by a preponderance of the evidence. Second, the court must resolve all factual or legal disputes relevant to class certification, even if they overlap with the merits—including disputes touching on elements of the cause of action. Third, the court's obligation to consider all relevant evidence and arguments extends to expert testimony, whether offered by a party seeking class certification or by a party opposing it.⁴

This new standard, which requires courts to delve into substantive issues and resolve expert disputes, has two main, obvious implications for economic expert testimony in class certification. First and foremost, if courts may no longer resolve often complicated fact issues with a presumption, but must decide among irreconcilable expert opinions, expert economic analysis is likely to become a central, contentious issue in class certification proceedings.⁵ Second, the new standard is moving district courts toward a "preponderance" standard and away from a "threshold showing" standard.⁶ And while

price."). The reluctance to leave "merits" disputes unresolved appears to have derived from the Supreme Court's decision in *Eisen v. Carlisle & Jacquelin*, 417 U.S. 156, 177 (1974) ("We find nothing in either the language or history of Rule 23 that gives a court any authority to conduct a preliminary inquiry into the merits of a suit in order to determine whether it may be maintained as a class action."). Lower courts interpreted *Eisen* to hold that an expert opinion was acceptable so long as it was not "fatally flawed" or if it demonstrated "some showing" that the Rule 23 requirements were met. *See, e.g.*, *Caridad v. Metro-North Commuter R.R.*, 191 F.3d 283, 293 (2d Cir. 1999) ("In deciding a certification motion, district courts must not consider or resolve the merits of the claims of the purported class. Here, the District Court credited Metro-North's expert evidence over that of the Class Plaintiffs. Such a weighing of the evidence is not appropriate at this stage in the litigation.").

³ 552 F.3d 305 (3d Cir. 2008).

⁴ *Id.* at 307. As another example, the Second Circuit rejected the notion that an expert's opinion may be acceptable "simply by being not fatally flawed" and called on district judges "to assess all of the relevant evidence admitted at the class certification stage and determine whether each Rule 23 requirement has been met, just as the judge would resolve a dispute about any other threshold prerequisite for continuing a lawsuit." *In re Initial Public Offering Sec. Litig.*, 471 F.3d 24, 42 (2d Cir. 2006).

⁵ *See generally* John M. Majoras, *Opening the Curtain: Why Economics Is Taking Center Stage in Class Certification Battles in Antitrust Cases*, GLOBAL COMPETITION POL'Y, Spring 2008, Vol. 6, No. 2, at 1, 7 (noting that recent circuit court decisions recognize the value of economic evidence); Steven E. Bizar, Thomas P. Manning & Landon Y. Jones, *Bogosian, Linerboard After Hydrogen Peroxide*, COMPETITION L. 360, Feb. 10, 2009, at 3 (noting that *Hydrogen Peroxide* rejected the notion of not considering in depth opinions of defense expert economists).

⁶ *E.g.*, *Hydrogen Peroxide Litigation*, 552 F.3d at 326 ("Applying a presumption of impact based solely on an unadorned allegation of price-fixing would appear to conflict with the 2003 amendments to Rule 23, which emphasize the need for a careful, fact-based approach, informed, if necessary, by discovery.").

there may be debate about whether the new standard requires class certification analysis to be as rigorous as it is at other, later phases of litigation, the ante clearly has been raised for litigants given that outcomes at the class certification stage are often highly influential in determining litigation outcomes. Widely divergent views, however, still exist as to what constitutes probative economic analysis for purposes of class certification.⁷

This article informs this debate over the appropriate economic standard by presenting an economic foundation for the study of the legal requirement of “common impact”—a necessary element for class certification under Federal Rule of Civil Procedure 23.⁸ At the most basic level, I argue that, in antitrust cases, the legal standard should be analyzed (at least in part) with the assistance of economic analysis addressing the following question: Are prices paid by putative class members determined in a common way?⁹

Stating a standard is far easier than applying it, however. Determining common impact, at least from an economic perspective, can be complicated. This Article addresses three points relevant to its study with regard to class certification: (1) “predominance” of common factors should be determined by analyzing whether all economically significant determinants of price are common; (2) it is useful to recognize in studying common impact that determinants of price can be either related or unrelated to the defendant conduct that plaintiffs allege violated antitrust law; and (3) the examination of which “facts” are relevant for common impact in antitrust matters may be hindered

⁷ For example, Dr. John Beyer, who frequently testifies in support of class certification, has offered a list of factors that “tend to support a conclusion that all members of the class would have been adversely affected.” See John C. Beyer, *The Role of Economics in Class Certification and Class-Wide Impact*, in *LITIGATING CONSPIRACY: AN ANALYSIS OF COMPETITION CLASS ACTIONS*, 325 (Stephen G.A. Pitel ed., 2006). Other experts argue that the implementation of this “prototypical plaintiffs’ argument” usually lacks any intellectual or scientific coherence. See John H. Johnson & Gregory K. Leonard, *Economics and the Rigorous Analysis of Class Certification in Antitrust Cases*, 3 J. COMPETITION L. & ECON., 341, 345 (2007).

⁸ In federal courts, where most antitrust cases are brought, antitrust class certification often turn on whether the requirements of Federal Rule of Civil Procedure 23(b)(3) are met, i.e., if “questions of law or fact common to the members of the class predominate over any questions affecting only individual members” In legal parlance, this concept is typically termed “common impact.”

⁹ Plaintiffs in antitrust matters also may allege anticompetitive conduct that does not affect price but that does affect service or quality. To the extent that the alleged conduct manifests through different levels of service or quality, the interpretation offered here need not change, but it may become substantially more complex. The key challenge in cases where conduct does not affect price is that the conduct and its effect on putative class members needs to be objectively measured. For example, suppose that a cartel agrees to restrict advertising. While advertising but for the alleged conduct might be estimated in a way similar to how but-for pricing is estimated, the analysis also needs to consider whether consumers benefit from advertising in similar ways. In other examples, such as coordination on one aspect of product quality, it may be challenging to assess how the conduct affected overall quality. For example, if software manufacturers agree to restrict development of their product related to Internet connectivity, it might be difficult to assess how this conduct affected the total value consumers attach to the product.

by three common pitfalls: confusing average impact with common impact, using antitrust market definition to inform an analysis of common impact, and assuming the alleged conduct to study common impact. Understanding these points leads to three types of empirical tests that can be useful to courts in their determination of common impact. Each of the tests focuses upon differences in prices paid by the putative class members. By focusing on the commonality of pricing in an economically consistent way, the goal is to provide courts with workable, consistent, and meaningful measurements of the legal requirement of commonality.

I. AN ECONOMIC FRAMEWORK FOR ANTITRUST CLASS CERTIFICATION

Federal Rule of Civil Procedure 23(b)(3) sets a legal standard for antitrust class certification. Among other requirements, it requires that “questions of law or fact common to the members of the class predominate over any questions affecting only individual members.” I start by considering the economic implications of this legal requirement. First, I offer an economic interpretation of “predominance” by defining common and individual factors. Second, I propose that one can gain insights into the meaning of common impact by distinguishing “conduct” from “non-conduct” factors, with conduct factors defined as those affected by the allegedly unlawful behavior. Third, I discuss which “facts” are relevant for common impact in antitrust matters and discuss three pitfalls that may hinder an examination of common impact.

In this discussion, however, it is important to keep in mind that I am describing an economic framework for assessing common impact and not a legal one. The legal framework is established in Rule 23 and the case law that interprets the Rule. What I am attempting to do is clarify the legal rule by analyzing its economic implications in a way that may not be appreciated fully in the case law.

A. PREDOMINANCE AND COMMON FACTORS

Rule 23(b)(3) requires that “common” factors predominate over “individual” factors. Therefore, a threshold issue to understanding the predominance standard is to define the categories of common factors and individual factors. From an economic perspective, common and individual factors may be defined as:

- A *common* factor is 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.
- An *individual* factor is a determinant of price that affects price but at the same time is not “common.”

Common factors are often factors that are used in a similar fashion within an industry and are available at similar prices. For example, input costs like fuel, labor, and capital costs are usually viewed as common factors when they are used as inputs into a common production process. For example, costs across airlines may be affected in similar ways by an increase of 10 percent in jet fuel prices.¹⁰

Other factors, however, that affect all or most defendants or putative class members may not be so obviously common, even if the factor has wide relevance. Defendants, for example, may have shipped products varying distances to putative class members. But whether variable shipping distance had a common effect on prices paid is an important part of the investigation of common impact. For example, putative class members located in a particular geography could have had access to non-cartelized supply. Thus, a necessary first step for economic analysis is to identify all important factors that affect price in order to test for common impact.

To do so, economic experts should provide a qualitative review of the industry and the allegations that identifies economically significant factors. They should use that review to examine whether common factors are the only economically significant determinants of price, and they should perform a series of rigorous empirical tests that are suggested by the qualitative review.¹¹ The qualitative review should also help the court understand why the empirical tests are relevant. The empirical analysis should meaningfully test common impact and outputs of the tests should be explained in terms that are clear to non-economists and compelling to the court.¹² I discuss three types of such tests in the second half of this article.

Common factors are likely present in most cases (e.g., the price of fuel paid by defendants in a price-fixing context). Nonetheless, an expert must avoid focusing only on the effects of such factors without also examining potentially economically significant individual factors. The qualitative review may help illuminate whether economically significant factors have been overlooked. From the defense perspective, it is likely that individual factors will affect

¹⁰ There is usually a presumption that markets provide such factors in a non-discriminatory way to producers so that if fuel costs, for example, increase to one defendant then other defendants face similar increases. This presumption obviates the need to collect such variables from individual defendants.

¹¹ An expert may interpret the qualitative review of the industry and facts as sufficiently convincing so that no quantitative analysis is undertaken. Such a choice should not invalidate the arguments made by the expert, but that expert should assume that the other side will conduct a thorough quantitative analysis.

¹² For example, when confronted with variation in prices, either side's expert may conduct a regression analysis and report that a measure called R^2 is 60 percent. The other side's expert may question the probative value of such an analysis. It is unclear whether a court would, or should, view as helpful such a "battle of the experts" over purely technical issues without more context.

most prices to some extent; the focus should be on whether these effects are economically significant. Again, the qualitative review should (1) narrow and specify the design of empirical tests that economists can use to examine whether prices reacted in a common way to hypothesized common factors and (2) determine whether these factors comprise all of the economically significant determinants of prices.

Assessing “predominance” of common factors entails an inherent trade-off between judicial efficiency and accuracy. Thus, economists should not view the appropriate standard as whether an analysis that uses common factors exclusively is just as accurate as an analysis that uses both common and individual factors.¹³ In other words, a defense that points out statistically valid but economically unimportant exceptions to an analysis that relies exclusively on common evidence is not likely to persuade a legal audience. Rather, economists who analyze common impact should interpret “predominance” to require that the only *economically significant* factors are those that are common to members of the putative class and individual factors are economically insignificant.¹⁴

B. CONDUCT AND NON-CONDUCT FACTORS

Antitrust class certification proceedings often focus on the common impact of the defendants’ alleged wrongful conduct. For purposes of economic analysis, however, it is also important to understand the impact of factors other than the conduct alleged.

Therefore, in analyzing common impact, it is helpful to divide factors that affect price into two mutually exclusive sets: conduct factors and non-conduct factors.

- *Conduct* factors are determinants of price that are influenced by the defendant conduct that plaintiffs allege to have violated antitrust law.
- *Non-conduct* factors are determinants of price that are unrelated to the defendant conduct alleged.

¹³ An analysis that uses both common and individual evidence can always perfectly mimic an analysis that uses common evidence exclusively. This result does not mean, however, that any analysis using common and individual evidence will be more accurate than analysis using common evidence exclusively.

¹⁴ Both experts and courts necessarily exercise judgment in determining economic significance. Thus, one subjective term (only common factors have economic significance) is admittedly used to define another subjective term (predominance of common factors). However, determination of economic significance is central (at least implicitly) to any work an economic expert does. Thus, the exercise of expert judgment is hardly uniquely applied in class certification cases.

To illustrate the concept of conduct and non-conduct factors, consider two prototypical price fixing examples. In the first example, consider a cartel for trucking services between two cities where the cartel has agreed to coordinate pricing only to “large” customers. Assume that this trucking service is homogeneous (i.e., trucking provided by one firm is a perfect substitute for trucking provided by another firm), and consumers cannot resell the services to arbitrage away price differences. In the second example, suppose the cartel produces various formulations of an industrial chemical. Also suppose that some formulations are only produced by a single cartel member, and certain consumers must consume particular formulations, because of the particularities of their production process. The chemical formulations are heterogeneous and so is consumer demand for the formulations.

In each case, courts may not certify the class when the class is defined to include all customers, *but this result occurs for different reasons*. In the first case, impact is not common, because the effect of the alleged conduct only targeted a subset of class members (i.e., “large” customers).¹⁵ Thus, a conduct factor (price-fixing efforts) differs across class members to negate common impact. In the second case, impact is not common, because the effect of the alleged conduct on customers who consume products sold by a single cartel member must be zero; they would purchase from a monopoly supplier with or without a cartel. Thus, non-conduct factors (consumer demand and producer supply) differ across class members to negate common impact.¹⁶

An economically coherent treatment of common impact should consider both conduct factors and non-conduct factors. As these examples suggest, the distinction between conduct and non-conduct factors can be important. Despite this importance, however, the notion of conduct and non-conduct factors does not appear to have been widely appreciated in other treatments of class

¹⁵ By defining subclasses, of course, one could change conclusions about whether conduct factors are common. However, adequately defining a subclass may not always be easy; in the example above, defining a subclass of “large” customers is too vague, and defining a subclass that contains injured customers is tautological and unhelpful. For an example of a case where the court viewed subclasses as necessary for a finding of common impact, see *Carol McDonough v. Toys “R” Us, Inc.*, 638 F. Supp. 2d 461, 473 (E.D. Pa. 2009). In that case, retailer Babies “R” Us, Inc., was accused of coercing a number of manufacturers into ceasing discounting by other retailers. It was recognized that each manufacturer may have been coerced to differing degrees, and subclasses were appropriate to allow for these differences.

¹⁶ Both of these examples are purposefully simplified. For example, the first example assumes customers cannot arbitrage away price differences among themselves (e.g., by reselling). This assumption rules out the possibility that small customers might be affected by coordinated behavior that is focused on large customers. The second example assumes perfectly inelastic substitution between products for certain consumer producers (at least over a limited range of prices). This assumption immediately rules out a price effect to these customers.

certification, which have mainly focused on whether all putative class members were injured.¹⁷

Another nuance that does not seem to have been well understood by reported cases is the difference between reliably calculating damages and assessing common impact. The predominance of common conduct factors implies that all putative class members were targeted by the alleged conduct, but, as the above example shows, all class members may not have been injured.

However, the predominance of common non-conduct factors has a different implication for economists, namely, that an econometric model can calculate reliable plaintiff-specific damages regardless of the predominance of common conduct factors. In other words, a reliable class member-specific estimate of damages exists, but that estimate may be zero for some class members. Specifically, the econometrician can use a common framework to estimate reliably what prices would have been absent the alleged conduct (“but-for prices”) because all determinants of price, except potentially the alleged conduct, were common factors. In this respect, non-conduct factors are relevant to a reliable damages model, and conduct factors are not. The significance of this point, however, is largely economic and not legalistic, because whether courts will certify a class when a reliable damages methodology is available but injury is not common appears to be a legal issue that has not been resolved at the circuit level.¹⁸

To better illustrate this dynamic, I diagram and discuss the four possible scenarios that can arise in considering conduct and non-conduct factors that are either common or individual.

¹⁷ See, e.g., Johnson & Leonard, *supra* note 7, at 342 (“[F]or a class to be certified, plaintiffs must establish that each class member can be shown to have been harmed by the alleged conduct”); ABA SECTION OF ANTITRUST LAW, *ECONOMETRICS* 183 (2005) (“At the certification stage, courts determine whether the plaintiffs will be able to use common evidence in their attempt to prove that the defendants’ alleged antitrust violations affected each member of the class.”).

¹⁸ The Seventh Circuit seems to suggest that if some (but not a “great many”) putative class members are uninjured, class certification may be appropriate. See *Kohen v. Pac. Inv. Mgmt. Co.*, 571 F.3d 672 (7th Cir. 2009) (“What is true is that a class will often include persons who have not been injured by the defendant’s conduct; indeed this is almost inevitable Such a possibility or indeed inevitability does not preclude class certification A related point is that a class should not be certified if it is apparent that it contains a great many persons who have suffered no injury”) (citations omitted). Other courts suggest a stricter requirement. See, e.g., *Newton v. Merrill Lynch, Pierce Fenner & Smith, Inc.*, 259 F.3d 154, 188 (3d Cir. 2001) (“The ability to calculate the aggregate amount of damages does not absolve plaintiffs from the duty to prove each investor was harmed by the defendants’ practice.”).

		Non-conduct factors	
		Common	Individual
Conduct factors	Common	All class members targeted. Reliable but-for prices may be calculated.	All class members targeted. Reliable but-for prices may not be calculated.
	Individual	Not all class members targeted. Reliable but-for prices may be calculated.	Not all class members targeted. Reliable but-for prices may not be calculated.

1. Both Conduct and Non-Conduct Factors Are Predominantly Common

In this case, if one transaction is impacted by the alleged conduct, other transactions are also impacted. This result necessarily follows because the economically significant determinants of price for one class member are identical to those of any other class member. While the degree to which consumers can substitute to other goods, for example, will be an important factor in understanding whether and by how much the alleged conduct actually raised price (ultimately a merits issue to be determined by the fact finder), it should have no bearing on class certification. If price determinants are predominantly common, one class member could avoid harm from an anticompetitive common price increase by substituting to another good only inasmuch as other class members would also be able to do so.

*2. Non-Conduct Factors Are Predominantly Common,
But Conduct Factors Are Individual*

In this case, even though not all class members were targeted by the alleged conduct, the existence of predominantly common non-conduct factors implies that a damages methodology that uses common evidence exclusively can reliably calculate damages to all class members. Said differently, only common evidence need be consulted to determine pricing but for the alleged conduct. In particular, a damages methodology could be proposed that correctly assigns zero damages to transactions not impacted by the conduct.

Reconsider the example of colluding trucking companies that produce a homogeneous good but which only target “large” customers. In this case, an econometric model could estimate a relationship between trucking prices and non-conduct factors (e.g., fuel prices, demand for trucking services) over a time period absent any alleged conduct and forecast but-for prices in the period subject to the alleged conduct. Customers who were not targeted (and hence did not suffer injury) would have (in expectation) but-for prices equal to their actual prices; customers who were successfully targeted (and hence did suffer injury) would have (in expectation) but-for prices that are less than

their actual prices paid.¹⁹ Again, while this observation has implications for when a reliable common framework for damages may exist, it does not speak to whether courts would certify a class where substantial numbers of putative class members were not injured.

3. *Conduct Factors Are Predominantly Common But Non-Conduct Factors Are Individual*

In this case, for example, some customers might have access to non-cartelized supply only revealed by an individualized inquiry. In general, all economically significant determinants of price are critical for a reliable damages methodology. And in this case, the existence of individual non-conduct factors precludes that exclusive use of common evidence will estimate damages reliably. Suppose a plaintiffs' expert proposes an econometric model that uses only common evidence to calculate damages. But if individual non-conduct factors affect pricing but do not appear as explanatory variables, this econometric model will likely yield unreliable estimates of but-for prices.

In sum, the presence of non-conduct factors that are not predominantly common has serious consequences for damages awards in class certification proceedings. In addition to leading to an unreliable estimation of total damages, these factors may also lead to an unreliable allocation of damages among class members. The district court explicitly recognized this logic in *Piggly Wiggly Clarksville v. Interstate Brands Corp.*,²⁰ a matter that involved allegations of horizontal price-fixing among bakeries. In that case, differences in procurement strategies, geographic markets, delivery costs, and other services included with each purchase implied that "it will be impossible to present evidence in a common manner as to the price each Plaintiff would have paid but for the alleged conspiracy."²¹ Thus, the court took the view that an intractable damages methodology was sufficient to render classwide treatment undesirable. It is noteworthy that this view does not depend on identification of putative class members who were not impacted. The Fifth Circuit upheld the decision.²²

¹⁹ The commonly referenced "dummy variable" econometric model is usually not used in a way that allows for this possibility, however. In common usage, that model estimates a single *average* measure of damages that is *assumed* to apply equally to all putative class members. Other econometric models do not suffer from this shortcoming. For example, a predictive model might be used to generate but-for prices that are then compared to actual prices paid to generate a transaction-specific measure of overcharge. Used in this way, the transaction-specific overcharge will not generally be equal across all observations.

²⁰ 215 F.R.D. 523 (E.D. Tex. 2003).

²¹ *Id.* at 531.

²² *Piggly Wiggly Clarksville, Inc. v. Interstate Brands Corp.*, 100 F. App'x 296 (5th Cir. 2004).

4. *Both Conduct and Non-Conduct Factors Are Predominantly Individual*

Finally, there is the situation in which conduct and non-conduct factors are sufficiently individualized to make an analysis of predominance and damages unworkable on a broad basis: the alleged conduct did not target all putative class members, and ignoring individual evidence leads to unreliable estimates of but-for prices. An instructive example is *Rodney v. Northwest Airlines, Inc.*,²³ a case that involved allegations of monopolization by Northwest Airlines on seventy-four routes. The district and appellate courts both noted that evidence of monopoly power specific to one route was independent of evidence on other routes.²⁴ In other words, (conduct and non-conduct) supply factors were route-specific and not common to all the members of the class. Questions about (non-conduct) demand factors were viewed as route-specific as well: “an analysis of whether bus travel from Detroit to Toledo is reasonably interchangeable with a flight between those two cities will not help to define the market for travel between Minneapolis and Los Angeles.”²⁵

This observation is not to suggest that no common factors affect price—defendant’s capital, labor, and jet fuel costs may have affected prices paid in a common fashion. Rather, it suggests that an analysis that considers common demand and supply factors to the exclusion of individual factors will necessarily miss economically significant factors. By discussing such disparate demand and supply factors, the decision is a clear example of a case where common facts do not predominate.²⁶ Of course, the route-specific logic of *Rodney* suggests that the courts might have been amenable to route-specific subclasses.

C. PITFALLS IN ANALYZING COMMON IMPACT

Classifying factors as either common or individual, or conduct or non-conduct is necessary, but not sufficient, for a coherent and informative analysis of common impact. This section, without pretending to be exhaustive, discusses three pitfalls that might ensnare an analysis of common impact beyond such a classification: confusing average impact with common impact, using antitrust

²³ 146 F. App’x 783 (6th Cir. 2005).

²⁴ *Id.* at 789 (“[P]roof that Northwest exercised monopoly power over the route that Rodney took does not establish that Northwest exercised monopoly power over the other 73 routes.”).

²⁵ *Id.* at 787.

²⁶ For similarly clear discussions of how differences in demand and supply factors are central to class certification, see *Heerwagen v. Clear Channel Communications*, 435 F.3d 219, 227–29 (2d Cir. 2006) (considering whether demand and supply factors for rock concerts were common across the United States, thereby justifying class treatment); and *Blades v. Monsanto Co.*, 400 F.3d 562, 570 (8th Cir. 2005) (considering whether supply and demand factors for seeds were highly individualize thereby requiring individualized evidence).

market definition to inform an analysis of common impact, and assuming the alleged conduct to study common impact.

1. *Common Impact and Average Impact*

Irrespective of whether the analysis of common impact concerns conduct or non-conduct factors, such analysis must focus on differences in supply and demand factors across transactions and not on characterizations of the average transaction.²⁷ Litigants frequently do not appreciate this point and confound what is essentially an analysis of average impact with an analysis of common impact.²⁸ Claims about average impact concern the likelihood and effectiveness of the alleged conduct and may be studied by measuring, for example, the conduct's effect on *average* price. Common impact is fundamentally different. It concerns whether the conduct's effect on *individual* prices can be characterized by relying exclusively on common factors.

When a common impact analysis focuses on, say, the average effect of a factor, it misses the point. In particular, while an average may be useful for the fact finder's merits determination (e.g., how likely and effective the alleged conduct was on average), it is completely unrelated to class certification. For example, suppose one half of class members suffered a 20 percent overcharge and that the other half was not impacted or suffered a 0 percent overcharge. The average overcharge of 10 percent exists despite the fact that substantial numbers of class members were not impacted.²⁹

²⁷ This point is forcefully made in Richard Schmalensee, *Economic Analysis of Class Certification*, GLOBAL COMPETITION POL'Y, Spring 2008, Vol. 6, No. 2, at 1, 2 ("But the focus in class certification is on *differences* . . .").

²⁸ A recent district court decision analyzed common impact from alleged collusion by hospitals to lower wages to registered nurses. The plaintiffs' expert, Professor Gordon Rausser, used an average effect on wages to demonstrate common impact. The judge rightly recognized the fallacy of this approach and noted that "[m]easuring average base wage suppression does not indicate whether each putative class member suffered harm from the alleged conspiracy. In other words, it is not a methodology common to the class that can determine impact with respect to each class member." *Reed v. Advocate Health Care*, 268 F.R.D. 573, 591 (N.D. Ill. 2009).

²⁹ As another example of the confusion over average impact, consider one factor that is routinely cited in class certification proceedings: the elasticity of demand. To the extent that the *average* demand elasticity is low (i.e., purchasers are not particularly price sensitive, on average), coordination among competitors would tend to increase *average* price more than if average demand elasticity were high. But this observation exclusively addresses the average damages resulting from plaintiffs' allegations (i.e., the effectiveness of the alleged conduct); it is irrelevant for the analysis of common impact. In particular, average demand elasticity could be low, but substantial numbers of individual purchasers may have high demand elasticity (perhaps because they have different alternatives). Thus, to the extent that a factor, such as demand elasticity, is studied at the class certification phase, courts should pay particular attention to differences and give little weight to arguments that characterize the average effect of that factor throughout an industry.

One example of this confusion is found in *Caridad v. Metro-North Commuter Railroad*.³⁰ In that case, plaintiffs' expert opined that "being black" had a common impact as a result of alleged discrimination. He offered a regression model that reportedly showed that being black significantly reduced the likelihood of promotion.³¹ But a regression approach reports the *average* effect of being black and does not study whether all black *individuals* were impacted by discriminatory practices. While the appellate court's reluctance to decide among "statistical dueling" by experts is much cited,³² the more troubling aspect of the decision was that the court confounded issues related to average impact (i.e., whether black employees suffered from discrimination, on average) with common impact (i.e., whether discrimination affected all black employees). The average effect on all black employees could be positive without all black employees having suffered impact.³³

2. Common Impact and Antitrust Market Definition

Practitioners should not confuse commonality of demand and supply conditions with antitrust market definition, a widely used and accepted concept defined in the Horizontal Merger Guidelines.³⁴ Such confusion is possible because the concept of "market" is often a key issue in antitrust matters, but is often used more loosely in class certification than it is in other areas of antitrust.³⁵ Common impact entails stronger conditions than finding a Guidelines market, so finding that putative class members made purchases within a single antitrust market is necessary, but not sufficient, for finding common impact.

³⁰ 191 F.3d 283 (2d Cir. 1999). For a more detailed discussion of this case, see Robert G. Bone & David S. Evans, *Class Certification and the Substantive Merits*, 51 DUKE L.J. 1251, 1256–57, 1320–28 (2002).

³¹ *Caridad*, 191 F.3d at 286, 288–89.

³² *Id.* at 292.

³³ *Caridad* is by no means alone in confusing average and common impact. Another prominent example is the district court opinion in *Linerboard*, which involved allegations of coordinated output restrictions. *In re Linerboard Antitrust Litig.*, 203 F.R.D. 197 (E.D. Pa. 2001), available at <http://www.paed.uscourts.gov/documents/opinions/01D0719P.pdf>. The court's opinion mistook a damages analysis, which measured average impact, as being probative of common impact. *Id.* at 218 ("Plaintiffs in this case have advanced econometric models to be used to establish impact."); *id.* at 218 n.14 ("These models can also be used to show a method of calculating the damages, the amount of the overcharge."); see also *id.* at 220 ("The Court notes that the econometric methods plaintiffs have proposed for proving impact can also be used to calculate damages.").

³⁴ U.S. Dep't of Justice & Fed. Trade Comm'n, Horizontal Merger Guidelines (2010) [hereinafter Guidelines], available at <http://ftc.gov/os/2010/08/100819hmg.pdf>.

³⁵ For example, *Heerwagen v. Clear Channel Communications*, 435 F.3d 219, 224 (2d Cir. 2006), turned on whether the "relevant market" for concert tickets was national or local; in *Blades v. Monsanto Co.*, 400 F.3d 562, 570 (8th Cir. 2005), whether the "markets" for genetically modified seeds were individualized based on geographic location was an important component of the analysis; and in *Rodney v. Northwest Airlines, Inc.*, 146 F. App'x 783, 787 (6th Cir. 2005), "relevant markets" was seen as the central question.

A first reason that common impact is a stronger requirement than the Guidelines' approach to market definition is that the latter explicitly ignores issues related to supply side substitution and entry.³⁶ To illustrate, suppose price-fixing allegations are brought against a set of producers of an industrial chemical X whose precise formulation does not permit any substitution among end users so that X is a Guidelines antitrust market. But suppose there are a number of non-defendant producers that primarily produce a slightly different chemical but can easily produce chemical X. In such a situation, one should not conclude that because chemical X is a Guidelines antitrust market the cartel injured all putative class members; doing so ignores potential supply from non-defendant producers who may have sold chemical X to some putative class members.

A second reason that common impact is a stronger requirement than Guidelines antitrust market definition is that the latter only addresses issues of demand-side substitution by asking whether a hypothetical monopolist would find a small price increase above competitive or prevailing levels to be profitable. Even if such a price increase were profitable, some customers might purchase other products, implying that demand conditions are not common across purchasers. For example, suppose plaintiffs allege collusion among suppliers of a given product and that the product in question is in a single antitrust market. However, if some purchasers substitute to non-cartelized supply when faced with a small price increase, impact may not be common.

3. *Common Impact and Assuming the Allegations*

In investigating common impact, practitioners will often assume the allegations about anticompetitive conduct to be true. This approach can be helpful in an investigation of common impact because it focuses attention away from assessing fact or magnitude of damages, which are merits questions answered at a later stage of litigation. However, the expert should take care that, by assuming the allegations, he or she does not also effectively assume common impact. For instance, if an allegation asserts common impact, it is hardly useful to assume the allegations to test for common impact. Thus, an analysis of

³⁶ The Guidelines discuss supply-side substitution and entry but do so separately from market definition. See Guidelines, *supra* note 34, § 4 ("Market definition focuses solely on demand substitution factors, i.e., on customers' ability and willingness to substitute away from one product to another in response to a price increase. . . . The responsive actions of suppliers are also important in competitive analysis. They are considered in these Guidelines in the sections addressing the identification of market participants, the measurement of market shares, the analysis of competitive effects, and entry."). For a persuasive discussion of why Guidelines market definition ignores supply-side substitution and entry, see Jonathan B. Baker, *Market Definition*, in 1 ISSUES IN COMPETITION LAW AND POLICY 315 (Wayne Dale Collins ed., 2008).

common impact that essentially assumes that the impact is common is tautological and not useful.³⁷

While the absurdity of “assuming the conclusion” may be clear in the stark form presented above, a more subtle form of assuming the conclusion may be more common in practice. At the very least, a casual approach to the particularities of the allegations may mask important testable implications. For example, a complaint may allege that the conspiracy raised prices to all customers. But, if evidence indicates that the conspirators explicitly did not raise prices either to customers in a particular geography or to customers of a particular set of products, there is no value in ignoring such evidence on the basis that an examination of common impact assumes the allegations. Instead, it would seem useful to exploit the insights offered by a review of the evidence in constructing empirical tests.

II. EMPIRICAL ANALYSIS IN ANTITRUST CLASS CERTIFICATION

Economists agree that empirical analysis of data is useful, if not critical, because it tests hypotheses suggested by a theory.³⁸ As noted above, courts have moved from simply asking whether an expert’s analysis in class certification is “fatally flawed” to requiring a more rigorous analysis.³⁹ From a scientific point of view, this change is fortunate, because empirical “analysis” that does not follow the standards of the discipline can be molded to fit any desired conclusion. In class certification, however, there is not yet general agreement about what constitutes probative empirical analysis. This section uses the notions of common, individual, conduct, and non-conduct factors described in the previous section to inform this debate by describing three types of empirical tests that can be probative in many situations. A common theme is that these tests exploit transaction-level data to study how prices in different transactions react to economically significant determinants of price.⁴⁰

³⁷ This point was made forcefully in *Szabo v. Bridgeport Machines, Inc.*, 249 F.3d 672 (7th Cir. 2001). As the court noted, “The proposition that a district judge must accept all of the complaint’s allegations when deciding whether to certify a class cannot be found in Rule 23 and has nothing to recommend it.” *Id.* at 675.

³⁸ This notion is at the heart of the “scientific method” and applies more broadly than just to economics. See generally David Goodstein, *How Science Works*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 67–82 (2d ed. 2000) (discussing how the scientific method involves empirically testing a hypothesis).

³⁹ See *supra* note 2.

⁴⁰ A related issue is the availability of sufficient information to make a reliable analysis of common impact. Obviously, expert opinion relies critically on fact discovery to generate data and documents, and some have noted that expert opinion in class certification must be developed with “limited discovery of relevant information.” Beyer, *supra* note 7, at 326. While discovery at the class certification phase may indeed be limited, fairly rich data are frequently available (for example, Beyer graphs rich data developed from three class certification matters.) *Id.* at 330–32. And even in cases where it is not, courts’ increased attention may incentivize litigants to pursue more robust discovery at the class certification stage. Nevertheless, the appropriate response of a

Before discussing empirical testing, it is important to consider what constitutes sufficient “proof” of common impact. Of course, the scientific method can never prove a hypothesis; it can only reject, or falsify, a hypothesis.⁴¹ The Supreme Court, in *Daubert*, explicitly recognized the importance of falsification in science and made it a required property of expert testimony in legal proceedings.⁴² Because common impact is a restriction on how prices are determined, the appropriate hypothesis to falsify is that of common impact. Critically, common impact cannot be proved in a scientific sense because its converse (“individual impact”) is not a falsifiable hypothesis due to its more general nature (i.e., common impact is a special case of individual impact). Thus, both plaintiff and defense experts must proceed by examining whether evidence rejects the hypothesis of common impact. In this sense, a defense expert has a lighter burden than does a plaintiff expert because showing the existence of substantial exceptions to common impact effectively disproves a general claim about all or substantially all transactions. Conversely, the plaintiff expert cannot reach a general conclusion by examining only a few examples (even if they are important ones).⁴³ Instead, the plaintiff expert must

class certification expert to limited or imperfect data or documents is to apply standards accepted within the discipline; adoption of a less stringent standard in response to limited discovery has no appeal from a scientific point of view.

⁴¹ See Goodstein, *supra* note 38, at 70 (noting that a theory can never be proved right by agreement with observation; it can only be proved wrong by disagreement with observation).

⁴² *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993) (“Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested. ‘Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified . . .’”).

⁴³ Sampling, in some circumstances, may be an acceptable alternative to examining all transactions. By definition, if a sample of transactions is representative of the population as a whole, then a characterization of the sample also characterizes the population (subject to sampling error that can be statistically quantified). Thus, in theory, an affirmative opinion about common impact could be supported with analysis of a representative sample by, for example, applying a common framework to estimate but-for prices and assess whether substantially all class members within the sample were injured. Moreover, some courts may be amenable to analysis of a random sample of putative class members to determine injury. See, e.g., *Kohen v. Pac. Inv. Mgmt. Co.*, 571 F.3d 672, 679 (7th Cir. 2009) (discussing the possibility of deposing a random sample of class members to determine how many benefited from the alleged conduct).

While, the appropriate legal and statistical standards that govern the use of sampling techniques in class certification are beyond the scope of this article, some remarks are in order.

First, if electronic transaction data are available, then (usually) processing data for the entire population is just as easy as analysis of a representative sample; with powerful and relatively inexpensive computing capabilities, computer code can often operate as easily on 50 million records as it can on 50 thousand records. Thus, in some instances the use of sampling may not be merited or necessary. This same observation may be true less often when it comes to analyzing the data, because the computing burden of some types of analysis grows quickly with the number of observations. Analysis on representative samples may be valuable in such cases.

Second, suppose that electronic transaction data are only available at the class certification stage from some defendants. While these data represent a sample of the population of transactions, they clearly are not a representative sample, and any conclusion based on such data may not extend to the population. An open question for courts is how much weight to give conclusions that tend to support common impact in such circumstances.

convince the court that a set of sufficiently powerful tests has failed to reject the hypothesis of common impact. It is through this process that common impact is proved in a legal sense.

The question remains, however, what exactly should constitute empirical legal proof of common impact if empirical methods can only reject the hypothesis? The obvious answer is that courts, informed by the adversarial process, make this determination. I take no position about how much proof courts do or should require before finding common impact.⁴⁴ Two prominent economists active in class certification, however, have taken the normative position that legal proof of common impact should require characterizing all or substantially all transactions—potentially including individual investigations to collect class member-specific data.⁴⁵ These economists have termed this requirement the “common proof paradox,” because conducting an individualized investigation defeats the “efficiency” justification for allowing class actions.⁴⁶ And while these economists propose an empirical test of common impact, they also note its inapplicability—hence, the paradox.

The paradox is seen most clearly in their proposed test of a “fixed pricing structure.”⁴⁷ That test requires conducting individual investigations by collect-

Finally, sampling techniques may be particularly valuable when electronic transaction data are not available. But, while in theory such an approach could be probative of common impact courts would not only have to determine whether experts’ statistical tests were convincing, but they would also have to determine whether the sample on which such tests were based was representative.

⁴⁴ However, as noted in the introduction (*supra* note 1) the positive question (i.e., how much proof *do* courts require) is partially answered by the fact that courts are requiring more than they required in the past.

⁴⁵ Johnson & Leonard, *supra* note 7, at 344 (“[W]e believe that the requirement of common proof of antitrust injury should usually present a substantial hurdle for plaintiffs. How can one establish that individual data are not needed without actually looking at individual data to see if such data are needed? Put another way, how can one rule out the importance of a set of information without first examining it?”) These authors do not explicitly draw the distinction between legal and scientific proof. However, I understand them to be proposing a legal standard because, as discussed in the text, one cannot “prove” in a scientific sense the hypothesis of common impact.

⁴⁶ “[T]he test of whether common proof can be used requires that the customer-specific regressions be run. Thus, one must run individualized analyses to determine whether class certification is appropriate, thereby seemingly defeating the ‘efficiency’ justification for allowing class actions.” *Id.* at 351. But even this approach cannot “prove” a hypothesis of common impact in a scientific sense because one could always claim that the individualized analyses have not been sufficiently thorough. Thus, a useful plaintiff analysis must test the hypothesis with a “sufficient” amount of rigor. To illustrate, consider testing the hypothesis “all birds can fly” with an empirical examination of birds that winter in Washington, D.C. This empirical examination is clearly not sufficiently thorough due to its restriction to a certain geography. On the other hand, one cannot “prove” Darwin’s theory of natural selection in a scientific sense no matter how much confirmatory evidence is gathered. However, application of the scientific method over a century and a half has made this theory a cornerstone of biology: the scientific consensus is that the examination of Darwin’s theory has been sufficiently thorough to apply the theory generally.

⁴⁷ *See id.* at 348–51.

ing and analyzing “a set of supply and demand variables specific to [each] customer.”⁴⁸ It is unclear how large the set of customer-specific variables needs to be to achieve a level of sufficient legal proof; potentially, the set could be very large and, thus, require a very high threshold for legal proof. But no matter how rich the set of customer-specific variables collected is, such a test, like all others, cannot scientifically “prove” or “determine” common impact because one can always point to additional customer-specific models that have not been tested.⁴⁹

Requiring collection of a potentially very large set of class member-specific information to test sufficiently the hypothesis of common impact paints a somewhat bleak picture for empirical analysis because empirical legal proof of common impact will likely be extremely difficult, if not impossible, to assemble. This does not mean, however, that a defendant expert need conduct individual inquiries to reject the hypothesis of common impact: rejecting common impact without resorting to individualized evidence should be particularly compelling. However, requiring a legal standard of proof of common impact that is not possible to achieve in most, if not all, circumstances effectively precludes the possibility of class treatment.

I argue, however, that empirical analysis should be viewed in a more optimistic light. 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). The disadvantage of these tests is that they are less “powerful” than tests that also exploit individual factors, meaning that they are less likely to reject correctly the hypothesis of common impact so necessarily establish a lower level

⁴⁸ *Id.* at 349.

⁴⁹ Technically, customer-specific variables that are omitted from a regression model are captured in the regression error term. Econometric theory specifies the conditions that regression error terms must have for estimated regression coefficients to have desirable properties. While econometric tests exist to examine the properties of regression error terms, like all tests, they cannot “prove” the hypothesis that the regression errors have desirable properties—they can only reject that hypothesis. In this sense, pointing to a potentially limitless set of omitted customer-specific variables is the same as pointing out that the regression has an error term that is not identically zero.

of “proof.”⁵⁰ However, a main goal of the description of the tests that follows is to show how they can be implemented to have the most power possible to reject the hypothesis of common impact respecting the constraint that individual investigations are not conducted. Again, I take no normative position on whether failing to reject common impact with these tests should constitute legal proof of common impact. I note, however, that requiring examination of potentially countless individual factors involves endless attempts to falsify the hypothesis of common impact that is the common proof paradox.

To the extent that reaction to a hypothesized common factor is not common, one may infer that either (1) the hypothesized common factors are not, in fact, common, or (2) there are other important determinants of price that are individual, or (3) both. Thus, the hypothesis of common impact is rejected, or falsified. The converse is also true to the extent that one can only reject the hypothesis of common impact: if reaction is common, the hypothesized common factors may indeed be common, and other economically significant factors may indeed also be common. In the latter case, the fact finder must make a decision whether such evidence represents a “rigorous enough” analysis to constitute legal proof.

A. PRICE RESPONSES TO HYPOTHESIZED COMMON CONDUCT FACTORS

In some cases, allegations are specific enough to generate testable hypotheses. An example of such a hypothesis is that all class members were similarly impacted after the date that a cartel agreed to implement price increases. In this case, class certification differs from a typical damages analysis because the average effect of a cartel’s attempt to increase price is irrelevant. Rather, it is the uniformity of the price responses to the cartel that is germane. Examination of price responses to these conduct factors that are hypothesized to be common can produce direct and useful indicators of the appropriateness of class treatment for a wide variety of cases.

⁵⁰ That these tests may be less powerful can be seen in the following simple hypothetical. Suppose that individual factors are important determinants of price so that common impact should not be found. Additionally, suppose that omission of these individual factors causes biases in the estimated parameters. In theory, it could be the case that these biases are manifest in such a way that one finds, for example, that substantially all putative class members paid prices that were higher after the onset of a conspiracy. But as described at length in the text, such an outcome does not “prove” common impact in a scientific sense—it only fails to reject the hypothesis of common impact.

I take no opinion on whether this approach generally establishes legal proof. I do take the position, however, that if these tests falsify the hypothesis of common impact, then this establishes scientific proof against the hypothesis and should be relevant to a determination of legal proof.

A price reaction to a conduct factor is an example of a “natural experiment” that is widely used in antitrust analysis.⁵¹ In analyzing a natural experiment, the expert identifies a control group and a treatment group of transactions. The control group comprises transactions that are not affected by the alleged conduct, and the treatment group comprises transactions that are affected by the alleged conduct. The control and treatment groups are then compared to estimate a measure of the effect of the alleged conduct.⁵² To the extent that the alleged conduct is common to all class members, this comparison should reveal a fairly uniform measure of effect across class members.

In cases where the allegations are not specific (e.g., they do not specify the set of products affected or the time period during which the conduct took place), they may generate few, if any, testable implications. Conversely, allegations that are specific may present the researcher with a number of different natural experiments against which to test common impact.⁵³ Examples are, therefore, specific to particular allegations but are likely available in a large number of cases. In a price-fixing matter, for example, obvious examples include a comparison of prices before and after a cartel has allegedly implemented a price increase, customer allocation, or other action designed to raise prices. Alternatively, one can also compare prices in the time period following cartel detection with prices in the time period before detection.⁵⁴

Here, an economist could examine prices paid by putative class members to gain insight into whether an allegedly coordinated price increase had a common impact. The economist might use prices paid prior to the alleged coordination as a benchmark to assess the effects of the alleged coordination. As a simple example, customer and product-specific price changes may be reported just before and just after the onset of the alleged coordination; if substantial numbers of price changes were positive and substantial numbers were negative, then this would be evidence for rejecting the hypothesis of common im-

⁵¹ Mary Coleman & James Langenfeld, *Natural Experiments*, in 1 ISSUES IN COMPETITION LAW AND POLICY, *supra* note 36, at 743–73.

⁵² Ideally, the only difference between the treatment and control group would be the alleged conduct. But because economics does not (usually) benefit from a controlled laboratory setting, such an ideal condition is rarely attained. Instead, the economist takes steps (via econometric methods) to hold constant factors other than the alleged conduct.

⁵³ In this sense, the “factual enhancement” requirements of *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544, 557 (2007), suggest that allegations should be detailed and rich enough to have testable implications.

⁵⁴ Such analysis could be carried out by comparing product and customer-specific average prices in conduct and non-conduct periods. Unlike an average reaction to the alleged conduct measured across all customers, this average does not mask customer-specific reactions to the alleged conduct. However, it does, to some extent, mask within-customer dispersion in pricing. An expert economist should consider this dispersion to the extent that it is substantial.

pact.⁵⁵ In some cases, demand and supply factors unrelated to the alleged conduct beyond customer and product effects may make a simple comparison of prices before and after the alleged coordination inapposite. For example, raw material prices may have increased substantially causing a simple comparison of customer and product-specific prices to have little power to reject the hypothesis of common impact. Faced with such a situation, an economist may use econometric techniques to control for price changes caused by these factors. For example, an economist could specify an econometric model that explains prices paid as a function of all economically significant common factors beyond the control of the alleged cartel in addition to customer and product controls.⁵⁶ This econometric model serves as a hypothesized common damages framework. Heuristically, the 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.⁵⁷ The predicted prices represent those that would have obtained if the nature of the competition between firms had been unaffected by the alleged coordination; in other words, they represent prices but for the alleged coordination.

In an attempt to falsify common impact, the economist would calculate differences between these predicted prices and actual prices paid. What distinguishes an analysis of common impact from, say, a damages analysis, is that the economist focuses on the uniformity of these differences across putative class members. For example, this approach could identify the proportion of transactions whose but-for price was equal to or below the actual price (i.e.,

⁵⁵ It would be incorrect to criticize this simple (but potentially powerful) methodology because it fails to use econometric techniques to control for confounding factors. In fact, it controls for two important factors: product and customer identity. Moreover, such an analysis is numerically equivalent to a regression analysis that uses fixed effects for customer-product combinations to test the homogeneity of the effect of the cartel. It does not, however, control for other factors (e.g., raw material price increases) that may confound the effects of the alleged coordination.

⁵⁶ An analysis that uses factors under the influence of the alleged cartel as controls is not useful. To illustrate why, suppose the econometric model uses a measure of capacity utilization to control for prices paid by putative class members. Here the alleged conduct of the cartel (e.g., agreeing to increase prices) will, in most cases, cause quantity sold (and therefore capacity utilization) to fall. In this sense, that model will not isolate the alleged cartel's effects. For a more comprehensive discussion, see Halbert White, Robert Marshall & Pauline Kennedy, *The Measurement of Economic Damages in Antitrust Civil Litigation*, ECON. COMM. NEWSL. (ABA Section of Antitrust Law), Spring 2006, at 17–22, available at <http://www.abanet.org/antitrust/at-committees/at-econ/newsletter.html>.

⁵⁷ This approach would preclude estimation of the effect on customers who bought a particular product only after initiation of the conduct. To the extent that a substantial volume of commerce is covered by such instances, the power of the test to reject the hypothesis of common impact is diminished.

where the hypothesized common damages framework indicates no injury).⁵⁸ With this approach, three outcomes are possible:

- *Substantially all transactions were affected by the event.* This result is consistent with the hypothesis of common impact: that both the conduct factors in question had a common impact and individual factors are not economically significant.⁵⁹ Such a finding should be viewed as an element of legal proof supporting the hypothesis of common impact, but does not “prove” the hypothesis in a scientific sense.
- *A substantial proportion of transactions were affected by the event and a substantial proportion of transactions were not affected by the event.* This result rejects the hypothesis of common impact: that either the conduct factors in question did not have a common effect on price, the hypothesized common damages framework ignored important individual factors, or both. Such a finding should be viewed as an element of legal proof against the hypothesis of common impact.
- *Substantially all transactions were unaffected by the event.* This result is consistent with the hypothesis of common impact but is also evidence that a particularly weak test has been conducted. For example, the event in question may have been economically unimportant (e.g., the defendants had already raised prices to collusive levels) or the alleged coordination was ineffectual. Such a finding would suggest that other evidence be used to attempt to falsify the hypothesis of common impact. This other evidence may include examination of non-conduct factors, which is examined next.

B. PRICE RESPONSES TO HYPOTHESIZED COMMON NON-CONDUCT FACTORS

Rule 23(b)(3) does not restrict “common questions of fact” to the defendant’s conduct that a plaintiff alleges violated antitrust law. In this manner the legal and economic points of view are consistent as an economically coherent

⁵⁸ While this description is entirely heuristic, formal means of testing for a common response to an event are well known in econometrics. The standard method involves estimating two econometric models: the first model restricts the reaction to the event to be common across putative class members; the second model does not make this restriction. The performance of these two models is then compared to assess whether the restriction is supported by the data. This type of testing procedure has existed in the econometrics literature for at least half a century. See, e.g., Gregory C. Chow, *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*, 28 *ECONOMETRICA* 591 (1960).

⁵⁹ It is crucial that the econometric model consider all important common factors. For example, suppose that prices increase due to increases in the cost of raw materials, but the econometric model ignores these factors. In this case, any estimate of the effect of the event in question will be confounded with the effect of increased cost of raw materials.

analysis of common impact investigates whether all economically significant determinants of prices are common. This section proposes a type of test that is similar in execution to the price-response test described above but replaces testing for the common effect of conduct factors with testing for the common effect of non-conduct factors. Once again, it is important to reiterate that these tests are useful because they provide additional opportunity to reject the hypothesis of common impact. They cannot “prove” common impact in a scientific sense because, for example, a non-conduct factor may have a common impact, but conduct factors do not. However, to the extent the court believes these tests powerful, they may be elements of legal proof supporting the hypothesis of common impact

The methodology differs in that the natural experiment is not identified from the allegations but from a qualitative review that reveals economically significant determinants of price. The qualitative review is important because these tests are likely more informative in that they are more powerful if they exploit changes in economically significant factors as opposed to economically insignificant factors. In particular, a change in an economically insignificant factor should not cause significant changes in price, by definition. Thus, a test that aims to detect differences in the distribution of individual reactions in such a case will have no power to reject the hypothesis of common impact. To illustrate, suppose that yearly contracts in an industry are signed in January and these contracts call for constant prices throughout the year. A test of price changes between June and July, for example, will reveal common (zero) effects regardless of whether prices are determined predominantly by common factors.

As in the case with natural experiments based on conduct factors, examples will be specific to the particular industries and allegations in question. For example, entry or exit of competitors and/or new products may provide an opportunity to falsify the hypothesis of a common reaction to an important non-conduct factor. Additionally, changes in price lists may provide the means for a natural experiment if evidence indicates that, for example, a cartel attempted to raise prices through means other than changing list prices (e.g., limiting discounts from list price). Other examples might include changes in regulation or changes in marketing practices by either defendant or non-defendant firms.

To take a specific example, suppose plaintiffs allege price fixing in an industry in which delivery costs are an important component of pricing. Moreover, suppose that delivery distance is contained in a database so an individualized investigation is not necessary to construct such a variable. Customers’ (varying) locations are clearly unrelated to the price-fixing allegations, but a methodology that accurately estimates prices but for the alleged price fixing must control for delivery distance. At the class certification stage,

the germane question is whether prices paid respond in a common fashion to delivery distance. To answer this question, the economist may: (1) specify an econometric model that controls for important demand and supply factors, including delivery distance; (2) estimate the model and produce predicted values of price; and (3) compare actual prices paid with predicted prices to study whether these differences are uniform across putative class members.⁶⁰ Evidence that the hypothesized common framework does not adequately predict prices for some customers is a rejection of the hypothesized common framework. Conversely, evidence that the hypothesized common framework does predict prices adequately is consistent with common impact because the hypothesis has not been falsified.

It bears repeating that this type of analysis does not test for the alleged conduct's common impact; it only tests whether other important, non-conduct, determinants affect class members in a common way. Nevertheless, these tests are useful in attempting to falsify the hypothesis of common impact because common impact requires that both conduct and non-conduct factors have a common impact on prices. If important determinants of price do not affect class members in a common way, the trier of fact will not, without innumerable mini-trials for each class member, be able to draw reliable conclusions on what prices class members would have paid absent the alleged conduct. In this case, an econometric damages methodology that exploits only common evidence will ignore the effects of important individual factors; in slightly more technical jargon, the econometric model omits relevant variables. Econometricians recognize that omitting relevant variables will generally cause estimated effects of the conduct to be biased and unreliable. This error occurs because such omission prevents the econometrician from controlling for economically significant factors—the very motivation for the use of econometric techniques in the first place.⁶¹

C. CO-MOVEMENT OF PRICES

The tests considered to this point require one to identify specific events in order to compare a treatment set of prices with a control set of prices. The tests described in this section do not require identification of such events, so they may be valuable to complement analysis based on natural experiments or if no natural experiments exist. Instead, they examine the co-movement of prices in order to ascertain whether prices are affected by similar demand and supply factors. This test is useful for falsifying the hypothesis of common

⁶⁰ Again, this description is a heuristic of such a testing procedure. *See supra* note 58.

⁶¹ A special case involves lack of correlation between omitted and included variables. In this special case, some econometric models may yield unbiased, but inefficient, estimates of damages.

impact because prices that are not affected by similar demand and supply factors may diverge over time. However some methods of analyzing co-movement have very low power to reject the hypothesis of common impact. This section discusses methods that have higher power and so may be more useful in attempting to falsify the hypothesis and, hence, constitute more compelling legal proof of common impact if the hypothesis is not falsified.

Economists often analyze the co-movement of prices in class certification proceedings and sometimes refer to this analysis as a test for an underlying “pricing structure.”⁶² Unfortunately, this analysis, in practice, often relies solely on the expert’s subjective visual inspection.⁶³ Visual inspection may provide an initial clue about price co-movement, but a court should view with great skepticism a conclusion made solely on that basis.⁶⁴ In economics, visual inspection of empirical patterns in data (without objective statistical testing) is not used to draw definitive conclusions. That ground should be sufficient by itself to disqualify a “stare-and-compare” methodology given courts’ rejection of a lower scientific standard as an appropriate standard at the class certification stage.

To illustrate the drawback of subjective visual inspection, consider Figure 1 that depicts hypothetical prices paid by three putative class members for a well-defined hypothetical product.

⁶² See Johnson & Leonard, *supra* note 7, at 348; Beyer, *supra* note 7, at 328–11; James F. Nieberding & Robin A. Cantor, *Price Dispersion and Class Certification in Antitrust Cases: An Economic Analysis*, 14 J. LEGAL ECON. 61, 67 (2007). However, the economic literature does not use the term “pricing structure” in analyzing co-movement of prices, and the use of the term, in this sense, seems to be restricted to class certification.

The term has taken on some legal significance, however. In an early judicial adoption of the notion, the court in *Bogosian v. Gulf Oil Corp.*, 561 F.2d 434 (3d Cir. 1977), noted:

If the price structure in the industry is such that nationwide the conspiratorially affected prices at the wholesale level fluctuated within a range which, though different in different regions, was higher in all regions than the range which would have existed in all regions under competitive conditions, it would be clear that all members of the class suffered some damage, notwithstanding that there would be variations among all dealers as to the extent of their damage.

Id. at 455.

⁶³ See Johnson & Leonard, *supra* note 7, at 348 (“Far too often, we have seen assertions that a pricing structure can simply be ‘observed’ or ‘visually determined’ by review of pricing graphs and patterns.”).

⁶⁴ The parallels to *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999), are too strong to go unmentioned. In *Kumho Tire*, the Court affirmed the inadmissibility of expert testimony that relied on visual inspection to establish that manufacturing defects instead of other factors caused a tire failure. *Id.* at 153–55. See generally Margaret A. Berger, *The Supreme Court’s Trilogy on the Admissibility of Expert Testimony*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 9 (2d ed. 2000) (generally discussing admissibility of expert testimony).

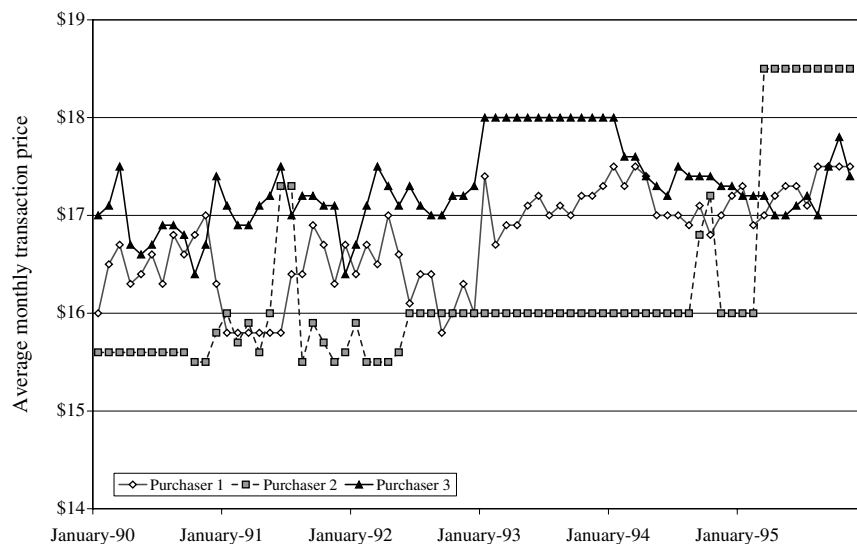


FIGURE 1: AVERAGE MONTHLY TRANSACTION PRICES FOR THREE PURCHASERS

Figure 1 is truly worth a thousand words and subjective visual inspection can find support for and against co-movement. For example, most transaction prices are in a fairly tight range and are generally higher at the end of the period than at the beginning. But there are numerous instances where prices obviously do not move together and where some purchasers pay constant prices over fairly long periods of time while other purchasers pay prices that vary month to month.

Fortunately, a number of objective methods are available to assess co-movement of prices. Two of the most widely accepted of these methods are correlation and cointegration.⁶⁵ Correlation analysis is a standard tool for this purpose.⁶⁶ Cointegration analysis is so widely used and accepted that in 2003 Clive Granger was awarded economics' top honor, the Nobel Prize, "for methods of analyzing economic time series with common trends (cointegra-

⁶⁵ This observation does not mean that other methods do not exist or are not useful. A comprehensive treatment of mathematical measures of co-movement of random variables, however, is beyond the scope of this article.

⁶⁶ See, e.g., George J. Stigler & Robert A. Sherwin, *The Extent of the Market*, 28 J.L. & ECON. 555 (1985) (discussing empirical methods to assess the extent of a "market"); Michael J. Doane & Daniel F. Spulber, *Open Access and the Evolution of the U.S. Spot Market for Natural Gas*, 37 J.L. & ECON. 477 (1994) (applying methods to assess how deregulation affected the scope of the "market" for natural gas in the United States).

tion).”⁶⁷ It would seem appropriate, and perhaps even necessary, for economic experts to consider these types of tests in studying price co-movement.

1. *Correlation Analysis*

A correlation coefficient is a statistical measure of the co-movement between two series, such as two prices. It always lies between one and negative one: a correlation coefficient near one indicates a strong positive correlation (i.e., a change in one series coincides with a proportional change in the other series), a correlation coefficient near negative one indicates a strong negative correlation (i.e., a change in one series coincides with a proportional, but opposite, change in the other series), and a correlation coefficient of zero indicates no correlation (i.e., a change in one series is not associated with any systematic change in the other series). In this sense, a high correlation supports a finding of common impact, and a low correlation does not. Figure 2 shows strong positive, strong negative, and zero correlations between two variables, *X* and *Y*.

Expert witnesses face two obstacles when undertaking a correlation analysis to study common impact. The first concerns the difference between germane and non-germane common trends that can substantially reduce the power of the analysis of co-movement to reject the hypothesis of common impact. The second concerns the determination of what constitutes “substantial” correlation.⁶⁸

Economists have recognized the well-known statistical result that if prices are substantially affected by a factor that affects many things in an economy, such as inflation, two products subject to different demand and supply conditions may have prices that are highly correlated. This phenomenon has been termed a “nonsense correlation.”⁶⁹ The danger in conducting an analysis without heeding the possibility of nonsense correlations is that an important common factor could dominate the effects of supply and demand factors particular

⁶⁷ Nobel Found., *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003*, http://nobelprize.org/nobel_prizes/economics/laureates/2003/.

⁶⁸ One should not confuse the use of correlation analysis to study class certification with the use of correlation analysis to delineate Guidelines antitrust markets. See generally Gregory J. Werden & Luke M. Froeb, *Correlation, Causality, and All that Jazz: The Inherent Shortcomings of Price Tests for Antitrust Market Delineation*, 8 REV. INDUS. ORG. 329 (1993) (showing the nonequivalence of Guidelines antitrust markets and correlation). Two products can be in the same antitrust market but have low correlation. Two products can be highly correlated and in disparate antitrust markets. The fact that antitrust market definition focuses exclusively on demand-side substitutability and ignores supply-side substitutability is important to appreciate this result.

⁶⁹ See Stigler & Sherwin, *supra* note 66, at 573 (noting study showing “nonsense correlations” between (1) saws and granulated bulk salt, (2) plows and cotton yarn, and (3) Texas hides and medium salt).

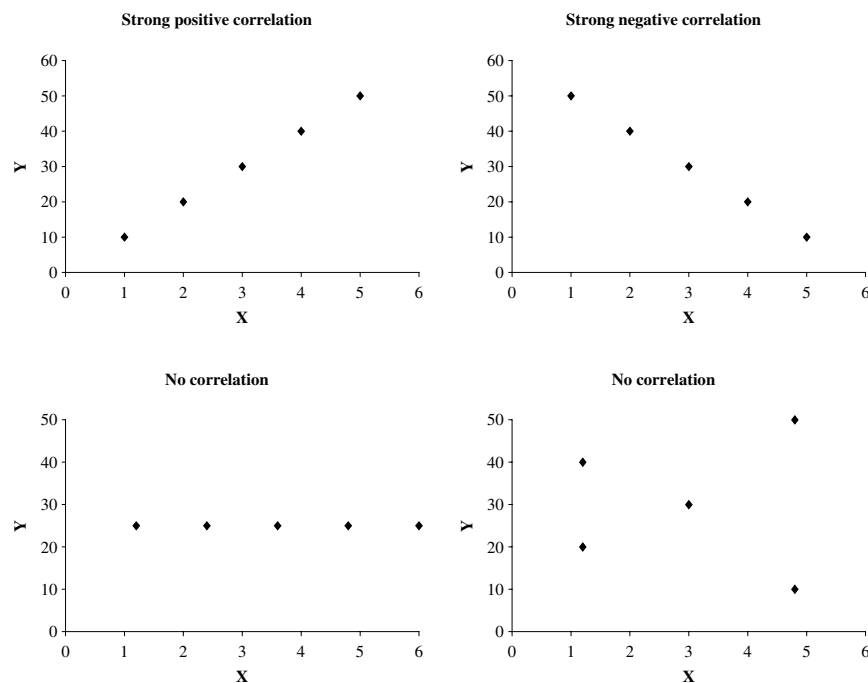


FIGURE 2: EXAMPLES OF POSITIVE, NEGATIVE, AND ZERO CORRELATION

to each product so that the test has no power to reject the hypothesis of common impact. This dynamic could lead to a false conclusion that prices are subject to the same demand and supply factors. To remedy this problem, economists compute the correlation coefficient between series after, for example, the effects of inflation have been removed. Other factors, such as raw material costs, may impart dominant trends on prices that mask the effects of individual factors and may be treated similarly. Specifically, the series in question might be regressed on a producer or consumer price index, and correlations between the resulting residuals are calculated. This technique can be substantially more powerful at rejecting the hypothesis that the series in question are subject to common demand and supply factors and, hence, of common impact as demonstrated by the following example.⁷⁰

⁷⁰ One may be tempted to employ the simpler method of dividing each price by the CPI or PPI series to calculate correlation between real prices. This approach is not used in the literature and is not helpful, because dividing each series by CPI or PPI will, by construction, add a common trend to the data. Adding a common trend to data that are to be tested for a common trend is both tautological and unhelpful. As an easy illustration, when one divides two constant series (whose correlation is undefined) by any PPI or CPI, perfect correlation results.

Consider Figure 3, which graphs hypothetical prices paid by two putative class members over time. Both price series tend to increase systematically over time and, in fact, the correlation coefficient of 0.92 between the two series indicates high correlation, seemingly implying that the hypothesis of common impact should not be rejected. However, one may expect that over a ten-year period, the effects of inflation may be the primary driver of this high correlation. Additionally, more careful examination may cause one to doubt whether these series are in fact driven by common factors: for example, in 1999 Customer 1's price was more than double Customer 2's, but at the end of the time period, Customer 1's price was approximately 15 percent lower than Customer 2's. In fact, when controlling for the effects of inflation in the fashion described above, the correlation coefficient of -0.13 is substantially lower and actually negative. This is evidence that would weigh for rejecting the hypothesis of common impact. The weak power of calculating simple correlations is demonstrated by the fact that these two series are actually indexed prices of gold ore and Brent crude oil—two commodities that are not determined by the same demand and supply factors.⁷¹

Measures exist to determine whether a correlation coefficient is *statistically* significant, but there is no generally accepted level above which one can conclude that series are *economically* significant.⁷² In practice, the lack of a recognized threshold means that certain empirical results may be insufficiently clear to reject the hypothesis that prices are predominantly determined by common demand and supply factors. For example, a correlation of 0.5 is probably not a strong enough result to reject categorically the hypothesis of common impact. Moreover, this conclusion does not depend on whether the estimated correlation coefficient is statistically different from zero. In this sense, a correlation coefficient's consistency with common impact differs from more familiar treatment of whether an estimate is statistically different from zero. However, the empirical results of a correlation analysis may provide sufficiently clear evidence in other circumstances. For example, suppose that data permit calculation of a large number of correlation coefficients, and even the smallest is statistically larger than zero and greater than some indisputably "large" level, say 0.8 or 0.9.⁷³ In this circumstance, if additional quantitative and qualitative review indicated that prices were predominantly affected by common demand and supply factors, then an economist could

⁷¹ The Brent Crude oil series is available from the U.S. Energy Information Administration <http://tonto.eia.doe.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=M>. The Gold Ore series is available from the Bureau of Labor Statistics (Gold ore producer price index, series PCU2122212122210).

⁷² In this discussion, the problem of nonsense correlations will be assumed to have been eliminated.

⁷³ If one is analyzing time series of prices paid for the same product by n customers, a correlation analysis will calculate $(n^2-n)/2$ correlations.

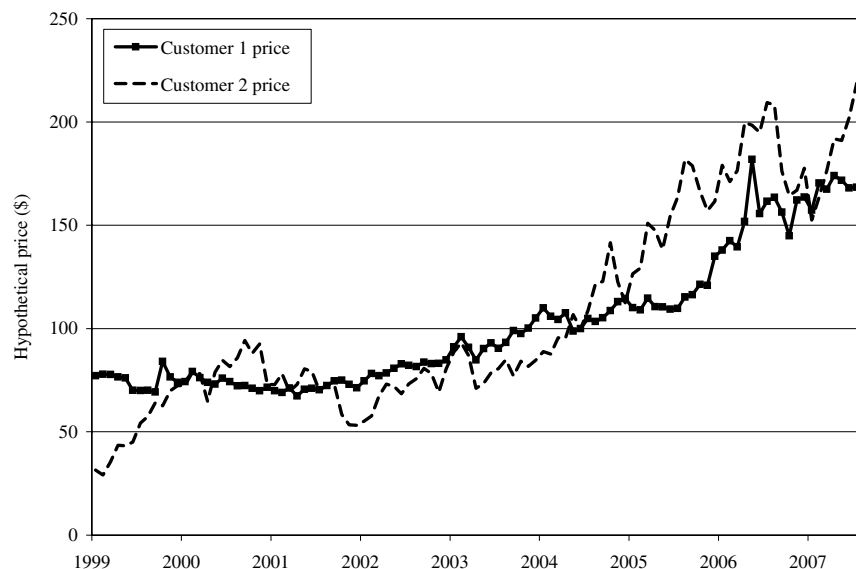


FIGURE 3: HYPOTHETICAL PRICES PAID BY TWO PUTATIVE CLASS MEMBERS OVER TIME

refer to this test as one that could not reject the hypothesis of common impact and courts may view it as an element of legal proof supporting common impact.

At the other extreme, suppose the analysis finds approximately one half of the correlations to be negative and one half to be positive. Such a result is identical to what one would predict if inherently unrelated series are analyzed. In this case (of course subject to additional qualitative and quantitative review), an economist would likely be justified in referring to this test as rejecting the hypothesis of common impact and courts may view it as an element of legal proof against common impact.

While some judgment is necessary in interpreting a correlation analysis, economic experts should make their basis for a conclusion about co-movement as clear as possible. Calculating correlation coefficients provides a tangible, well-defined, and objective foundation on which an expert can rely. By relying exclusively on visual inspection or other subjective or nebulous criteria, an expert does not provide a similarly clear basis for others to understand his or her conclusions.

2. Cointegration Analysis

Cointegration analysis, from a high-level perspective, is similar to correlation in that it is a statistical method of testing the hypothesis that multiple series co-move. Cointegration is firmly within the modern economist's toolkit,⁷⁴ but it does rely on advanced statistical properties of time series, so it is not as accessible as correlation analysis. For the purposes of this Article, I describe, at a general level, cointegration analysis.⁷⁵ I then explain how it might be applied successfully in the class certification context. It is important to stress that the techniques I outline do not require additional data or discovery. Rather, they are complementary steps an expert can and should take to make an analysis of co-movement both more consistent with commonly accepted scientific standards and more revelatory of common impact.

A collection of economic time series is said to be cointegrated if the underlying series are "non-stationary" but the difference between the series is "stationary."⁷⁶ Roughly speaking, a series is stationary if it has a tendency to revert back to some long-run average following a change; that is, if price disruptions are temporary. In contrast, a non-stationary series does not have a tendency to revert back to some average following a change. The remarkable characteristic of cointegrated series is that, while individual series may be highly unstable (i.e., they do not revert back to a long-run average), differences among cointegrated series are stable. The stability of the differences between the series does not allow one to reject the hypothesis that one price is affected by conduct and non-conduct factors inasmuch as other prices are affected. The behavior of stock market indices is a good example of this behavior. Individually, the Dow Jones Industrial Average and the S&P 500 index may be highly unstable, but the difference between these two series is substantially more stable. In other words, the determinants of stock prices composing the Dow Jones Industrial Average and the S&P 500 are similar, so that if one index increases, the other is likely to increase as well.

Two non-stationary series can have a high correlation, even though they are economically unrelated. Cointegration analysis was developed, in part, to address this drawback, and it is the main reason that cointegration analysis may be of value. Economists use the term "spurious correlation" to describe high

⁷⁴ For example, cointegration is sometimes taught even at the undergraduate level. Additionally, as noted *supra* in the text accompanying note 67, the Nobel Prize was recently awarded for cointegration. Such a distinction is only given to techniques that have thoroughly pervaded the discipline.

⁷⁵ For a non-technical and entertaining overview of these concepts, see Granger's Nobel Prize lecture. Clive W.J. Granger, Nobel Lecture: Time Series Analysis, Cointegration, and Applications (Dec. 8, 2003), available at http://nobelprize.org/nobel_prizes/economics/laureates/2003/granger-lecture.pdf.

⁷⁶ The technical condition is that a "linear combination" of the series be stationary.

correlation between two unrelated non-stationary series. While the intuition underlying this result is not easy to grasp, its statistical foundations are well understood by economists.⁷⁷ Spurious correlation is highly relevant to the study of co-movement because it implies that a correlation analysis can be extremely misleading when the underlying series are all non-stationary, in the sense that it has little power to reject the hypothesis of co-movement.

Cointegration, more so than correlation, is a method that can be hampered by situations where customers make irregular and/or a limited number of purchases during a given time period.⁷⁸ This observation may be particularly relevant for analysis of common impact because prices averaged over customers and products may not be used to fill in missing observations. In this sense, cointegration analysis may be less informative in such situations.

Beyond this caveat, two additional, albeit technical, caveats are required for cointegration analysis to have any power to reject the hypothesis of common impact.

First, prices must all be individually non-stationary.⁷⁹ Fortunately, economists have developed a vast set of tests for non-stationarity. A finding that some prices are stationary and others are not is strong evidence that these series have fundamentally different statistical properties as a result of being affected by different demand and supply factors. Such fundamentally different statistical properties suggest that they should not be analyzed in a common framework. In addition to this observation, an economist who performs a cointegration analysis without verifying that all prices are individually non-stationary runs a serious risk of conducting a "test" with absolutely no power to reject the hypothesis of co-movement. Specifically, application of some cointegration tests will have the feature that inclusion of *any* stationary series in a test for cointegration will lead to an apparent (but incorrect) conclusion that the series are cointegrated. This result is not a flaw in the theory of cointegration; rather, it is a fundamental misapplication of the theory.⁸⁰

⁷⁷ This notion has been appreciated by economists since at least the mid-1970s. See C.W.J. Granger & P. Newbold, *Spurious Regressions in Econometrics*, 2 J. ECONOMETRICS 111 (1974). More recent research provided the relevant statistical theory that underlies this phenomenon. See P.C.B. Phillips, *Understanding Spurious Regressions in Econometrics*, 33 J. ECONOMETRICS 311 (1986).

⁷⁸ The reason for cointegration's sensitivity to missing data is that the formal tests typically require a number of lags of the series in question. Thus, one missing data will affect multiple observations.

⁷⁹ This notion is universally appreciated in the econometric literature. See, e.g., JAMES D. HAMILTON, *TIME SERIES ANALYSIS* 571 (1994) (noting that cointegration analysis requires each of the individual series to be non-stationary).

⁸⁰ This result is intuitive to see recalling the definition of cointegration: non-stationary series are cointegrated if there exists a linear combination that is stationary. If one ignores the first requirement of the definition (that the series are non-stationary), one can trivially find a linear

Second, and perhaps less appreciated, is the fact that cointegration among any number of variables only requires a stationary difference between only *two* of the variables. However, the question of relevance for class certification is whether *all* (or substantially all) series co-move. Therefore, for class certification, it appears that a stronger condition than cointegration is needed for tests to have any power to reject the hypothesis that they all co-move. In particular, cointegration between every pair of series is the relevant criterion. To illustrate, consider the case of analyzing three series: the price of Berkshire Hathaway A shares, the price of Berkshire Hathaway B shares, and the price of crude oil.⁸¹ While the prices of A and B shares are undeniably determined by common demand and supply factors, the price of crude oil is determined by quite different factors. However, according to the definition of cointegration, the collection of these three prices will be cointegrated because a stable relationship exists between the A and B shares.⁸² An obvious method to resolve this problem involves checking for cointegration between each pair of prices.⁸³

As a further illustration of this latter concept, consider Figure 4, which graphs prices paid by three putative class members over time (it adds prices for a third customer to the prices reflected in Figure 3). In particular, unlike Customer 1 and Customer 2, the relationship of the difference in prices between Customer 2 and Customer 3 appears stable over time. And, in fact, formal tests of prices paid by Customer 2 and Customer 3 find cointegration, and cointegration is not found between Customer 1's price and prices paid by either Customer 2 or 3. But critically, cointegration tests of *the three prices collectively* reveal the presence of cointegration. Such "collective" testing for cointegration masks the fact that Customer 1's prices appear to be determined by factors that do not affect the other customers' prices. Such an approach clearly should be avoided when assessing common impact as it has no power

combination that is stationary: the combination that puts a weight of one on the stationary series and puts a weight of zero on the other series.

⁸¹ Until recently, Berkshire Hathaway A shares were immediately convertible to 30 B shares. Thus, "most of the time, the demand for the B will be such that it will trade at about 1/30th of the price of the A." Memo from Warren E. Buffet, Chairman, Comparative Rights and Relative Prices of Berkshire Class A and Class B Stock (Feb. 2, 1999; updated July 3, 2003), available at <http://www.berkshirehathaway.com/compab.html>.

⁸² This conclusion presupposes that the three component series have been found to be non-stationary, although the economic theory of asset prices strongly suggests this to be the case.

⁸³ Econometricians have also introduced the concept of "irreducible cointegration" that is relevant. An irreducible cointegrating relation is "one from which no variable can be omitted without the loss of cointegration property." See James Davidson, *Structural Relations, Cointegration and Identification: Some Simple Results and Their Application*, 87 J. ECONOMETRICS 87, 87 (1998).

to reject the hypothesis of co-movement; pairwise cointegration testing is the preferred method.⁸⁴

Analyzing co-movement of prices via cointegration analysis provides for a useful test of the hypothesis of common impact. But, as is the case even for correlation analysis, correct application of the technique requires a certain degree of technical knowledge that courts may not always have. While, in theory, the adversarial process may be illuminating, in practice, courts may find themselves confronted by diametrically opposed opinions and approaches. Given the new requirement to “resolve all factual or legal disputes relevant to class certification,”⁸⁵ courts may consider a wider role for independent court-appointed experts to better understand technical issues at the class certification stage.

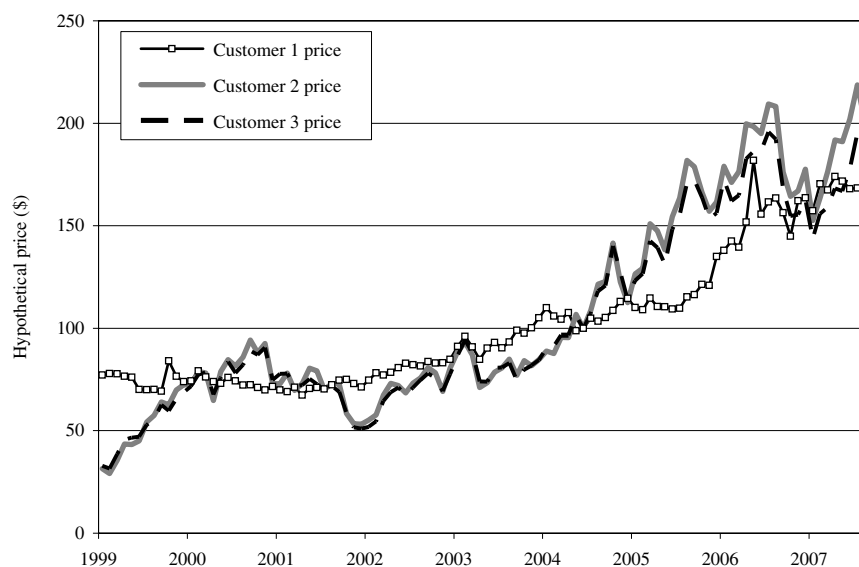


FIGURE 4: HYPOTHETICAL PRICES PAID BY THREE PUTATIVE CLASS MEMBERS OVER TIME

III. CONCLUSION

The jurisprudence of class certification is rapidly changing. Courts are moving away from the practice of assuming common impact and toward a

⁸⁴ Similar examples can be constructed to illustrate the danger of conducting cointegration tests when some of the underlying price series are stationary. *See supra* note 80.

⁸⁵ *In re Hydrogen Peroxide Antitrust Litig.*, 552 F.3d 305 (3d Cir. 2008).

more stringent standard for economic evidence. One can expect, therefore, that the economic analysis of common impact will receive renewed and deserved attention.

With these changes, litigants and courts will face new questions that have heretofore not been thoroughly considered. One of the most important is whether a class should be certified if a reliable common framework to study the effects of the alleged conduct is available but where there is no guarantee that substantially all putative class members were injured. This question is best appreciated by careful consideration of the conduct and non-conduct factors relevant to a particular case: predominance of common non-conduct factors implies that a reliable common framework can be used to predict but-for prices, but presence of individual conduct factors implies that not all putative class members were targeted. As a simple example, suppose that a court believes that a formulaic damages framework that ignores individual evidence is reliable but that this framework allows for the possibility that substantial numbers of (or even all) putative class members suffered no injury (i.e., zero damages). Would (and should) that court certify the class? This article does not pretend to answer that question, but hopefully provides a useful framework to consider that, and other, questions in the future.