Journal of Business and Economics, ISSN 2155-7950, USA
May 2016, Volume 7, No. 5, pp. 836-870
DOI: 10.15341/jbe(2155-7950)/05.07.2016/011

http://www.academicstar.us

Rising Standards for Class Certification: Implications for Economic Analysis

Gordon C. Rausser[1], Gareth J. Macartney[2]

(1. University of California, USA; 2. OnPoint Analytic, Inc., USA)

Abstract: Legal standards for class certification in class-action suits have risen steadily, as exemplified by the U.S. Supreme Court's recent decisions to reject certification in Wal-Mart v. Dukes and Behrend et al. v. Comcast Corp. et al. The resulting higher standards present challenges for economic experts, who present complex econometric and statistical analysis to courts. Some experts have presented analyses that would not pass peer review (and thus have no place in court). Moreover, the new standards have exposed a tension between the economic and statistical methods that academic economists accept and the requirements of proof that arise in litigation. We document unacceptable methods of class-certification analysis and present methodologies consistent with modern economics. Sound hypotheses must be formulated, and rigorous empirical testing must be conducted. We validate the use of objective summary statistics, regression analyses, and reject the pervasive use of "data mining". Further, analyses of liability and impact should be conducted prior to any implementation of a class-wide damages framework. Given the non-experimental nature of economic data, tests for commonality across the proposed class should be soundly motivated.

Key words: class certification; competition; econometric methods; hypothesis testing; statistical methods

JEL codes: K21, K41, C11, C12, C18

1. Introduction

On June 20, 2011, the US Supreme Court reversed a lower court's decision to grant plaintiffs' motion for class certification in *Wal-Mart Stores, Inc. v. Dukes et al.* (131 S.Ct. 2541) (hereinafter *Wal-Mart*). The plaintiffs in this employment discrimination case claimed that although the company's policy was to allow local managers to exercise their own discretion over pay and promotions, its corporate culture resulted in women receiving lower pay and fewer promotions. The plaintiffs had chosen to pursue the suit as a class action on behalf of a nationwide class of some 1.6 million female employees. When the Supreme Court refused to certify the class, the media devoted considerable attention to the decision, since it presumably implied a higher bar for gender-discrimination lawsuits.

The *Wal-Mart* decision was also significant for class-action practice. One observer described the ruling as "a capstone — and an exclamation point — to the trend initiated in the lower courts towards more rigorous certification standards" in class-action suits (James F. Speyer, Ronald C. Redcay & Kelly A. Welchans, 2011).[1]

Gordon C. Rausser, Robert Gordon Sproul Distinguished Professor, University of California, Berkeley; research areas/interests: law and economics, econometrics, public policy, statistical decision theory, industrial organization. E-mail: rausser@berkeley.edu.
Gareth J. Macartney, Ph.D., Director of Competition Analytics, OnPoint Analytics, Inc..

[1] The authors note that this is the Supreme Court's first significant class-action decision since Amchem Prods., Inc. v. Windsor, 521

This trend has been, in part, a reaction to the lower standards that courts had historically accepted.[2] For example, it was considered permissible simply to assume the merits of the plaintiffs' allegations (under what was known as "the Eisen Rule") and, on the basis of this assumption, to assume further (based on arguments relating to market-equilibrium pricing) that all class members were affected by the alleged violation of law (this became known as "the Bogosian shortcut").[3,4] In addition, the plaintiffs' economic expert was required only to *describe* the class-wide analysis that he intended to perform at trial to prove impact and quantify damages. He was not required to *demonstrate*, at the class-certification stage, that his proposed analysis would actually work.

The standards that have more recently been emerging from U.S. Courts of Appeals and in the *Wal-Mart* decision require that the plaintiffs' economic expert actually perform an analysis of commonality at the class-certification stage. Courts are now encouraged to consider fully the reports and testimony of both the plaintiffs' and the defendants' economic experts (hereafter, "PEs" and "Des", respectively). Courts were also charged with resolving the factual and legal disputes that are relevant to certification, even if these disputes overlap with the allegations that are central to the case (i.e., the "merits") (Michael D. Hausfeld, Gordon C. Rausser, Gareth J. Macartney, Michael P. Lehmann & Sathya S. Gosselin, 2014, pp. 77-133).

We seek to clarify class-certification standards and define the practices that economic experts should adopt when presenting testimony regarding whether a proposed class meets those standards. We begin by outlining the main elements of this type of litigation and the certification criteria that are defined in the Federal Rules of Civil Procedure (Section II). Next, we briefly present the important legal decisions that reach beyond *Wal-Mart*, including *Hydrogen Peroxide* and *Comcast* that have framed the evolving class-certification standards (Section III). We proceed by reviewing the most relevant economics literature on class certification (Section IV) and the type of data generally available in class certification matters (Section V). Subsequently, we document examples of the misuse of data analysis and econometrics in class certification (Section VI). Section VII structures the correct lens through which an economist should assess class certification, whether serving as PE or DE (Section VII). Finally, Section VIII offers some concluding remarks.

2. The Class-certification Landscape

It is important to recognize from the outset that the certification decision often determines the outcome of class actions. Since plaintiffs are often so numerous that a class action is the only cost-effective means of recovering damages, losing their motion to grant class certification can mean the end of any chance for plaintiffs to be compensated for any alleged wrongful conduct. For defendants, losing their motion to deny class certification exposes them to a trial and possibly ruinous liability. This is especially true in punitive-damages litigation, where, by law, damage awards are treble the calculated class-wide damages.

Rule 23 of the Federal Rules of Civil Procedure mandates the factors to be considered in the class-certification decision. Rule 23(a) sets out four conditions, *all* of which must be met in order for a class to be certified. The class must be numerous ("numerosity"); there must be questions of law and fact common to the class ("commonality"); the proposed class representatives must be typical of the class ("typicality"); and those

U.S. 591 (1997).

[2] Amendments made to Rule 23 of the Federal Rules of Civil Procedure in 2003 have also influenced the level of rigor required in the class-certification decision. *See* In re Hydrogen Peroxide Antitrust Litigation, 552 F.3d 305 (2008).

[3] *See* Eisen v. Carlisle & Jacquelin, 417 U.S. 156, 94 S.Ct. 2140, 40 L.Ed. 732 (1974).

[4] *See* Bogosian v. Gulf Oil Corp (561 F.2d 434).

representatives must fairly protect the interests of the class ("adequacy of representation"). After Rule 23(a) has been satisfied, Rule 23(b) sets out three further conditions, *one* of which must be met for a class to be certified. Plaintiffs can choose which one of the three parts of Rule 23(b) the court should apply when considering their motion for class certification. Of these three conditions, the one that is most often at issue for antitrust litigations is Rule 23(b)(3), which states that a class action is appropriate only if "the court finds that the questions of law or fact common to class members *predominate* over any questions affecting only individual members, and that a class action is *superior* to other available methods for fairly and efficiently adjudicating the controversy [emphasis added]."[5] This "predominance" and "superiority" requirement provides fertile ground for economic experts to use the tools of their trade to support (or reject) arguments made by opposing experts.

2.1 Liability, Impact, and Damage Analyses

Liability, common impact, and damages are factually and logically connected. The common-impact and damage analyses both depend upon either a presumption or a demonstration (by a preponderance of the evidence) of liability. The damage analysis must measure the type of injury posited to have occurred, given the common-impact analysis. However, although liability, common impact, and damages are related, they are also separable and subject to different forms of proof.

Since the Eisen ruling (1974), courts appear to have recognized the inherent conditionality of liability, common impact, and damages and the potential implications of each on the other two. From an economic perspective, liability must be evaluated first, because identifying wrongful conduct must precede the analysis of how that conduct may have affected putative class members. The next step is impact analysis. Here, the goal is to test whether, as a result of the identified wrongful conduct, all or most of the class would actually have suffered economic harm (regardless of its quantum). The class-wide damages that result from that common impact can then be quantified using a reliable statistical model or some other methodology.

At each of the three stages, the economic expert operates within the boundaries of accepted paradigms,[6] such as the assumptions that agents act rationally and that the laws of supply and demand apply (Gregory Mankiw, 2014). The paradigm helps to define a set of possible *maintained hypotheses,* from which those most consistent with economic logic can be selected. Since not all hypotheses are testable, given the limitations of real-world data, it is also necessary to identify one or more *testable hypotheses*.[7] In this setting, the PE may formulate a maintained hypothesis that the defendants acted as alleged and in a fashion that would be expected to cause injury to the putative class. Given this maintained hypothesis of liability, the expert can then assess empirically the hypothesis that such behavior had a common impact on the putative class. If the results of the inquiry support the

[5] The other two conditions are designated as 23(b)(1) and 23(b)(2). Rule 23(b)(1) allows a class to be maintained where "prosecuting separate actions by or against individual class members would create a risk of" either "(A) inconsistent or varying adjudications," or "(B) adjudications . . . that, as a practical matter, would be dispositive of the interests of the other members not parties to the individual adjudications or would substantially impair or impeded their ability to protect their interests." Rule 23(b)(2) applies when "the party opposing the class has acted or refused to act on grounds that apply generally to the class, so that final injunctive relief or corresponding declaratory relief is appropriate respecting the class as a whole."

[6] In economics, a paradigm "defines the type of relationships to be investigated and the methods and abstractions which are regarded as legitimate within a particular problem area," *see* Gordon Rausser (1973), *Discussion*, *American Journal of Agricultural Economics*, Vol. 55, pp. 271-279 [hereinafter Rausser, 1973]. Thomas Kuhn originally proposed the notion of a paradigm for the natural sciences, but it has now been fashionable in economics for several decades, *see* Thomas Kuhn (1970), *The Structure of Scientific Revolutions* (2nd ed.).

[7] "The selected maintained hypotheses … isolate a still smaller set, which represents the testable hypotheses," *see Rausser* (1973), p. 275.

hypothesis, then the expert can implement a damage analysis framework and quantify class-wide damages.[8]

2.2 Investigating Liability through Common Evidence

The exact analysis required to investigate whether liability can be demonstrated through common evidence will depend on the facts of the industry and the nature of the allegations against the defendants. It will also be limited by the discovery available to the experts at the class-certification stage. Experts who seek to use a common analysis in order to provide testimony regarding whether a large group of customers has been damaged by defendants' alleged actions will generally use econometric concepts to analyze market definition, product substitutability, and price dispersion.

In their academic work, professional economists commonly make inferences based on the identification of average effects or the location of a probability distribution (e.g., the average effect of an economic policy on a group).[9] Given the inherent randomness of markets, empirical economists recognize that models cannot achieve perfect explanatory power. For most economists, this lack of perfection does not mean that models, in conjunction with sound economic principles, are incapable of proving, for example, that an economic policy or some specific wrongful market behavior damaged all members of a group in similar ways. But legal standards in the U.S. require that impact and damages be proven for each and every class member, so that the inferences economists commonly accept are not necessarily accepted by courts as sufficient. Amidst the battles of the experts, courts have faced obstacles distinguishing unsound economic arguments from sound ones and have not always judged similar economic analyses consistently.[10]

Among the factors a court must naturally consider in assessing common evidence of liability are the following:

(1) Can common evidence on industry characteristics be used to support the factual foundations of the PE's model and hypothesis?

(2) Is it economically rational that the defendants would act in a fashion that was common across the proposed class, so that the plaintiffs' model can be appropriately applied on a common basis?

(3) Can a common economic model (or other acceptable methodology) be constructed that explains how the alleged behavior by the defendants was rational? Specifically, would the potential benefits to the defendants of their alleged actions outweigh the associated costs?

(4) Are there any implications of such a model that are inconsistent with the accepted economic principles of rational agents acting to serve their own best interest? Such implications would undermine the validity of the model.

(5) Are the analytical inferences from the documentary record and the transaction data aligned and if an alignment is found, does the analysis support the allegation that the defendants acted in a common fashion across

---

[8] The most commonly used methods for computing class-wide antitrust damages rely on the quantification of a price "delta", which is computed as the difference between prices class members actually paid for the class products and prices that would have been paid in the alternative, which are understood to be unaffected by the alleged misconduct. *See* Jonathan Baker & Daniel Rubinfeld, "Empirical methods in antitrust litigation: Review and critique", *American Law and Economics Review*, 1999, Vol. 1, pp. 386-435 at 391. *See also* James A. Brander & Thomas W. Ross, *Estimating Damages from Price-Fixing*, Litigating Conspiracy: An Analysis of Competition Class Actions (Stephen Pitel ed. 2006), at p. 351: "Estimation of reduced-form price equations is the preferred and most commonly applied method for damage estimation by economists in price-fixing cases."

[9] *See* David Evans (2009), *The New Consensus on Class Certification: What it Means for the Use of Economic and Statistical Evidence in Meeting the Requirements of Rule 23*, Antitrust Chronicle.

[10] DEs have prevailed in some class-certification cases by simply documenting price dispersion for the relevant product(s) and market(s). They have erroneously cited this dispersion as evidence in and of itself that a common economic model cannot be used to explain the full range of variation in market prices — and therefore, that class-wide impact cannot be proven.

the putative class that is consistent with the PE's model and hypotheses?

(6) Does the economic model account for all (or at least the predominant) common influences on price and/or output?

(7) Does the economic model systematically account for material factors or peculiarities of the market and individual transactions?

3. Rising Legal Standards

We do not attempt to comprehensively analyze class-certification decisions and individual court rulings. Instead, we present the fundamental principles that have driven the standard of economic proof required for class certification.

3.1 The Old Standard: Eisen and Bogosian

In the past, district courts often relied on two rulings to simplify their task of evaluating class certification (Richard A. Ripley & Mark J. Glueck, February 2009). *Eisen*[11] was interpreted as denying district courts the authority to resolve factual or legal disputes related to the merits in its evaluation of the requirements for Rule 23.[12] What became known as the *Eisen Rule* came to dominate class action practice (Geoffrey P. Miller, 2004, p. 2). The strongest form of the *Eisen Rule* is simply to assume that plaintiffs' allegations are valid. Judges typically invoked this strong form of the rule as a shortcut to certifying a class (Miller, 2004, p. 16).

The second historical standard that is most important for class certification is known as the "Bogosian shortcut". This standard followed the Third Circuit's 1977 decision in Bogosian v. Gulf Oil Corp (561 F.2d 434). The pertinent passage of *Bogosian* essentially states that if plaintiffs who are claiming that a violation of antitrust law has occurred can prove that the prices in an industry exhibit a "price structure", it is legitimate to infer that all class members have suffered a common impact or some damage. A price structure would exist if, for example, prices for most members of the class were consistently higher (even if prices differed across regions or class members) than would have been the case under competitive conditions. Like the *Eisen Rule*, *Bogosian* provided courts with a simple assessment of certifying classes. Often, a PE has had only to argue that the products purchased by class members were homogenous and part of the same market, that the defendants possessed market power, and that the products' prices rose and fell together. Armed with this *Bogosian* pricing-structure opinion, plaintiffs then appealed to the *Eisen Rule*, asserting (or assuming) that the defendants collectively agreed to raise the prices of those products. If these allegations were proved, all class members were judged to have suffered economic damage. The only other element necessary to certification was a description (but not implementation) of a feasible methodology for calculating damages.

3.2 Raising the Bar: Recent Rulings

In recent years, the tide has shifted away from shortcuts such as those prompted by *Eisen* and *Bogosian* (Miller, 2004, p. 17; Steig D. Olson, 2009, pp. 935-978).[13] Three rulings in particular have typified this sea

---

[11] "There is "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." *Eisen*, p. 177.

[12] In contrast to Eisen, the Supreme Court's 1978 ruling in Coopers & Lybrand v. Livesay (437 U.S. 463, 469 n.12, 1978) appears to encourage analysis of the underlying merits at class certification: "'Evaluation of many of the questions entering into determination of class action questions is intimately involved with the merits of the claims. . . The more complex determinations required in Rule 23(b)(3) class actions entail even greater entanglement with the merits,'" *see* Charles A. Wright et al., Federal Practice and Procedure (1976), at p. 485, n. 45. Nonetheless, Eisen's looser standards appear to have dominated the legal standards for class certification.

[13] Some lawyers and judges argue that a desire to protect defendants against weak actions has motivated the recent shift, rather than

change in the standards required for class certification: *Hydrogen Peroxide*, *Wal-Mart* and *Comcast*.

3.2.1 Hydrogen Peroxide

Many U.S. Courts of Appeals rulings have clarified how courts should determine the suitability of class treatment, but the ruling that is recognized as defining most clearly the extent of fact finding required under Rule 23 is the Third Circuit's 2008 ruling in *Hydrogen Peroxide* (Ripley & Glueck, 2009, p. 1).[14] In granting plaintiffs' motion for class certification, the district court had applied both *Eisen* and *Bogosian*. But the Court of Appeals overturned the decision and clarified the responsibilities of district courts in evaluating class certification emphasizing that: a) courts must resolve all legal and factual disputes relevant to class certification, even if they overlap with the merits; b) courts must make a "rigorous assessment" of the models and methods that plaintiffs propose to use in order to prove impact and measure class-wide damages; c) factual determinations supporting Rule 23 must be made by a preponderance of the evidence; and d) the court must fully consider expert evidence from both plaintiffs and defendants.

The Court of Appeals also ruled that the district court had applied *Bogosian* inappropriately. The facts showed that the price of hydrogen peroxide was lower at the end of the alleged conspiracy than at the start. In addition, the economic experts did not agree about the price structure of different types of the product. Therefore, the Court of Appeals ruled, it was not possible to presume that the alleged conspiracy had impacted class members by raising all of their prices. From the standpoint of sound methodology, the most important criticism the appellate court made was to chastise the PE for glossing over important price differences by conflating average prices for a wide range of products. Further, although the PE had proposed a method by which to calculate class-wide damages, he had not performed any analysis to show that the method was workable. Unconvinced that the PE could produce a workable damage analysis at trial, the court found that any formulaic approach to calculating damages would have to include such a multitude of product characteristics and price determinants that it would defeat any reasonable notion of proof common to the class.

3.2.2 Wal-Mart

The Supreme Court's decision in *Wal-Mart* further increased the rigor with which class certification is treated. In this ruling, the Supreme Court transferred much of the burden that formerly belonged to the "predominance" and common-impact requirements of Rule 23(b)(3) to Rule 23(a)(2). Although the district court had certified the class under Rule 23(a) and Rule 23(b)(2) and the U.S. Court of Appeals for the Ninth Circuit had upheld the decision, the Supreme Court reversed the judgment, ruling that the proposed class did not meet the criteria set out by either Rule 23(a)(2) or Rule 23(b)(2). The court did not specify whether *all* members of a class must have been injured in order for the class to be certified. But because the plaintiffs' claims rested upon proof that "Wal-Mart engages in a *pattern or practice* of discrimination",[15] the court concluded that the *reasons* for each employment decision would have to be determined, and that this task could not be undertaken on a class-wide basis.

The Supreme Court did specify two methods that the plaintiffs might have used to gain class-action status. One would have been to show that a single instrument, such as a biased employment test, had been applied to all employees and applicants, and that they had all been subjected to its outcome. The second would have been to

a determination to analyze closely whether class treatment is the best format for a possible trial.

[14] The authors cite three other Courts of Appeals rulings that have sought to clarify these responsibilities: the Seventh Circuit's 2001 ruling in Szabo vs. Bridgeport Machines (249 F.3d 672), the Second Circuit's 2006 ruling in In Re Initial Public Offerings (471 F.3d 24), and the First Circuit's 2008 ruling in In Re New Motor Vehicles Canadian Export Antitrust Litigation (553 F.3d 1).

[15] *See Wal-Mart* at p. 11.

provide "significant proof" of a "general policy of discrimination".[16] In the present instance, the Court ruled, the PE's analysis had failed to consider that local differences might have led to the disparity in promotions between women and men. The Court also rejected the PE's anecdotal evidence regarding discrimination as too sparse and not sufficiently reliable to justify drawing inferences about a company-wide policy or practice.

Incorporating into Rule 23(a) the issues of predominance and common impact traditionally dealt with under Rule 23(b)(3) means that courts must resolve these issues during class-certification, regardless of which part of Rule 23(b) plaintiffs cite. The Court also made it clear that full-blown evidentiary hearings were appropriate at the class-certification stage (rather than being reserved for trial, as had previously been the case). As in *Hydrogen Peroxide*, the court was unconvinced by an analysis purporting to prove discrimination that was based on averages, in this case average regional wages. The court also implied that explicit evidence of a policy of wrong-doing is required and that an appeal to a "culture" of such acts is not adequate.

3.2.3 Comcast

A third decision of particular importance in recent years, is the Supreme Court's reversal of class certification in *Behrend et al. v. Comcast Corp. et al*. In *Comcast* class plaintiffs presented four theories of impact, one of which was an "overbuilder" theory: "Comcast's activities reduced the level of competition from 'overbuilders', companies that build competing cable networks in areas where an incumbent cable company already operates."[17] It had been ruled that three of the other plaintiff's theories were not suitable for certification, leaving only the "overbuilder" theory. However, PE had "designed a regression model comparing actual cable prices in the Philadelphia DMA with hypothetical prices that would have prevailed but for petitioners' allegedly anticompetitive activities. The model calculated damages of $875,576,662 for the entire class. .... As the PE acknowledged, however, the model did not isolate damages resulting from any one theory of antitrust impact."[18]

The Court stated that "We start with an unremarkable premise. If respondents prevail on their claims, they would be entitled only to damages resulting from reduced overbuilder competition, since that is the only theory of antitrust impact accepted for class-action treatment by the District Court. It follows that a model purporting to serve as evidence of damages in this class action must measure only those damages attributable to that theory. If the model does not even attempt to do that, it cannot possibly establish that damages are susceptible of measurement across the entire class for purposes of Rule 23(b)(3)."[19] As a result, the Court reversed the judgment of the Court of Appeals for the Third Circuit to uphold the district court's certification of the plaintiff class. The majority decision of the Supreme Court, however, was criticized by the dissenting opinion of Justices Ginsburg, Breyer, Sotomayor and Kagan, which stated that: "[t]he oddity of this case, in which the need to prove damages on a class-wide basis through a common methodology was never challenged by respondents... is a further reason to dismiss the writ as improvidently granted. The Court's ruling is good for this day and case only. In the mine run of cases, it remains the 'black letter rule' that a class may obtain certification under Rule 23(b)(3) when liability questions common to the class predominate over damages questions unique to class members."[20] Regardless, Comcast did not change substantive antitrust law, only requiring that a PE's class-wide damage analysis be directly connected with the challenged conduct or plaintiffs' theory of liability.

[16] *See id.* at p. 13.
[17] *See Comcast*, at p. 3.
[18] *See id.* at p. 4.
[19] *See id.* at p. 7.
[20] *See id.*, dissenting opinion, at p. 5.

In summary, as exemplified by *Hydrogen Peroxide* and *Wal-Mart*, it is no longer possible to satisfy class-certification standards by performing analyses based on inference and simple averages and, following *Comcast*, there is now more emphasis on the need for expert analysis to closely correspond to plaintiffs' allegations. However, if courts are to analyze the economic methods of both plaintiffs and defendants more rigorously, it is important to distinguish between sound and unsound methodology.

4. Economic Literature

The economic literature identifies the challenges faced by an expert witness who testifies in litigation matters. These challenges include recognizing the difference in knowledge and expertise between economists and lawyers and recognizing the difference between objective, independent research and advocacy. As Fisher pointed out in 1986, the gap in skills and understanding between economists and lawyers can tempt expert witnesses to simplify their exposition so much that it loses precision and even obfuscates: "It is always tempting to seek certainty in quantification rather than to study all aspects of a truly complex problem." (Franklin Fisher, 1986, pp. 277-286). Moreover, sometimes a temptation arises to steer the analysis toward a desired outcome: the expert "can slip little by little . . . from true objectivity".[21]

The heightened legal standards have made the challenges Fisher identified nearly 30 years ago even more important today. A serious issue that economists and lawyers face, and that has been brought to the fore with the recently heightened legal standards, is that economic and statistical methods tend to focus on central, average tendencies rather than on the experience of each and every individual in a market. That said, econometric methods are available to assess commonality and test for impact and damages. More than twenty years ago, a paper by Hartman and Doane noted that judicial decisions concerning class certification "appear inconsistent and idiosyncratic", largely because of the difficulty of proving that class members have been "commonly and uniformly damaged by defendant's illegal actions." (Raymond S. Hartman & Michael J. Doane, 1987, pp. 351-372). As a method of proving commonality, the authors proposed the hedonic regression methodology, which can be used to control for (or "factor out") inessential attributes. The authors note that this method's "usefulness for class certification procedures lies in its ability to identify and measure the commonality in a group of apparently heterogeneous products, services, or individuals. When such commonalities are identified and measured, the courts can support class certification and can calculate the common effect of illegal actions, correcting for any apparent heterogeneity." (Raymond S. Hartman & Michael J. Doane, 1987, p. 352). This early paper provides a good example of a reasonable regression-based assessment of common impact.

More recent papers have sought to extend the use of regression analysis to test more thoroughly for commonality, in response to the recent changes in the case law. However, the tests proposed by these papers may well prove draconian (one would be hard pressed to find any regression model applied to real world data that could pass their proposed commonality tests) and often appear devolved from sound economic theory. As a case in point, Paul Johnson seeks to offer an "economic interpretation of 'predominance'" and argues that "'predominance' of common factors should be determined by analyzing whether all economically significant determinants of price are common." (Paul A. Johnson, 2011, pp. 533-567). He distinguishes between common *conduct* factors and common *non-conduct* factors (Johnson, 2011, p. 538). Common *conduct* factors are effects on prices that result from the alleged illegal action. Common *non-conduct* factors are product and market

[21] *See id.*, at p. 285.

characteristics that affect price but are not related to the alleged illegal action. In order to test for common *conduct* factors, Johnson proposes using a regression-based model to calculate but-for prices for each proposed class member and compare them to the actual prices paid. If, for some significant number of proposed class members, the difference between the actual price paid and the estimated but-for price suggests no harm, then, Johnson suggests that it is fair to conclude that the alleged conduct did not have a common impact.[22]

This high bar for the PE exceeds the legal standard and is subject to but-for prediction errors (see Section VII.D). Under this rule, the PE would need to do more than show that a common method to calculate class-wide damages is feasible. He would also have to accurately calculate individual but-for prices for every class member. Given accepted error rates in regression analysis, Johnson's method would almost certainly turn up some number of class members who would appear not to have been harmed. But drawing an inference from these apparent exceptions that these class members were *in fact* left unharmed by the conspiracy is both unwarranted and tantamount to "data mining". If this inference were considered acceptable, defendants could search at will for pricing patterns that they could use to invalidate the PE's analysis. Absent any formal formulation of a hypothesis that some class members were unharmed because of their individual economic circumstances (e.g., they had alternative goods on offer by which to escape the defendants' alleged conduct), and given the presence of random variation in pricing data, the author's proposed test falls short of the standards for accurate statistical inference.[23]

Johnson goes further and proposes a similar test for common *non-conduct* factors. He argues that the economist should formulate a model and use it to predict prices. Then, the expert should compare those predicted prices to actual prices in order to determine whether any differences between predicted and actual are "uniform across putative class members." (Johnson, 2011, p. 556). According to Johnson, a lack of uniformity among class members in these regression residuals would constitute evidence of omitted variable bias and "cause estimated effects of the conduct to be biased and unreliable."[24] But Johnson wrongly assumes that any price variation that cannot be explained by the regression model automatically disproves common impact and precludes damages. This erroneous assumption, which is at odds with standard econometric practice, is all too common among DEs (Pierre Cremieux, Ian Simmons & Edward A. Snyder, 2010, pp. 939-968; John H. Johnson & Gregory K. Leonard, 2007, pp. 341-356).

Similarly, John Johnson and Gregory Leonard propose a test for commonality that is essentially a choice between two overarching "models" of price formulation. Their first regression model estimates a set of common coefficients on supply and demand explanatory variables and a conspiracy variable using all of the available customer data (Johnson & Leonard, 2007, p. 349). Their second model (or more accurately, a series of customer-specific models which could number in the thousands, depending on the size of the putative class), estimates customer-specific coefficients on the same supply and demand explanatory variables and conspiracy variable used in their first model, using individual, customer-specific regressions based on transaction data for only each customer in question (Johnson & Leonard, 2007, p. 350). Johnson and Leonard then propose that an

22 Johnson (2011) describes a third test that investigates co-movement of market prices. This is well-covered ground.

23 A statistician investigates whether an empirical estimate is statistically significant, given the variation in the population of data points from which he has chosen a sample. For example, is the average price for Group A higher than the average price for Group B only because of sample selection, or does it constitute evidence of a true difference that is characteristic of the population as a whole? The result is said to be statistically significant at the 95% confidence level, that is, when the econometrician is 95% confident that the result is evidence of a true relationship and not solely due to sample selection and/or random variation in the data. This confidence level becomes meaningless with repeated scouring of the data.

24 *See ibid.*

appropriate test for commonality would use "standard statistical tests" to ascertain whether the individual customer-specific coefficients estimated from their second "model" (really, the series of customer-specific regressions) are different across customers and different from the common coefficients estimated in their first, common model. If they are different, then a common model "would be inappropriate and it would not be possible to use common proof to establish impact on all class members." (Johnson & Leonard 2007, p. 351). However, the authors interpret the need for their proposed test of commonality as a death knell for the whole concept of class actions. They write that because their "test of whether common proof can be used requires that the customer-specific regressions be run…one must run individualized analyses to determine whether class certification is appropriate, thereby seemingly defeating the 'efficiency' justification for allowing class actions." (Johnson & Leonard, 2007, p. 351). The authors refer to this as the "common proof paradox". This appears to us as a circular, nonsensical point and one that becomes moot when we consider, as we shall in Section VI, that their proposed test of commonality lacks sound economic or statistical foundation.

In a subsequent paper, Johnson and Leonard describe again their method of individual customer-specific regressions as the basis for a test of commonality in class certification analysis (John H. Johnson & Gregory K. Leonard, 2011, pp. 570-586). This second paper refers to the literature on measuring treatment effects to justify the authors' approach concerning individual customer regressions. The treatment effects literature aims to provide methods for estimating the effect of a policy (or "treatment") on a group of economic agents (such as individuals, households or firms). We agree with Johnson and Leonard that there are some parallels between estimating treatment effects and estimating the effect of allegedly illegal actions on a class, where the latter is indeed a form of "treatment". However, there is nothing in the treatment effects literature that explicitly supports running customer-specific regressions. Johnson and Leonard write: "a fundamental premise of the treatment effects literature is that the effects of a treatment may be heterogeneous across subjects, and any econometric technique for estimating a treatment effect must take this heterogeneity into account." (Johnson & Leonard, 2011, p. 574; Guido W. Imbens & Jeffrey M. Wooldridge, 2009, pp. 5-86). Certainly, Imbens and Wooldridge recognize that it is desirable for the theoretical setup that precedes estimation of treatment effects to "allow[] for general heterogeneity in the effects of the treatment." (Guido W. Imbens & Jeffrey M. Wooldridge, 2009, p. 7). But the methods proposed do not measure individual-specific treatments, instead: "Most of these estimands are average treatment effects, either for the entire population or for some subpopulation, although some correspond to other features of the joint distribution of potential outcomes." (Guido W. Imbens & Jeffrey M. Wooldridge, 2009, p. 15).

Indeed, anyone familiar with the treatment effects literature will be aware of the ubiquity of the term *average* treatment effect. Allowing for heterogeneity in treatments enables one to model different average effects for different subpopulations, but Imbens and Wooldridge certainly do not argue that it is possible to reliably estimate treatment effects for specific individuals using only data for each individual. The simple reason is that any estimate of a treatment effect requires comparison of a treated group to a control group (Guido W. Imbens & Jeffrey M. Wooldridge, 2009, p. 6). If data are only available for an individual *upon* treatment, which is often the case, then estimation of a treatment effect using only data for that individual is precluded. Along similar lines, Johnson and Leonard's customer-specific regressions could never be applied to any customer who only makes purchases during a conspiracy and not before. But even with data on one individual both before and after a treatment, the effect of the treatment using that individual's data alone is confounded if other, potentially unobservable factors, could affect outcomes for that individual. Such problems can be mitigated when groups of individuals are compared, which is why estimation of treatment effects concerns the estimation of distributional effects, such as average treatment effects and quantile treatment effects.

Bret Dickey and Daniel Rubinfeld describe a variant of the Johnson and Leonard approach, but with some qualifications (Bret M. Dickey & Daniel L. Rubinfeld, 2010-2011, pp. 459-486). Dickey and Rubinfeld recommend a possible decomposition of the class into potential subclasses, which may be distinguished by way of customer size or geographic location or other relevant dimensions ((Bret M. Dickey & Daniel L. Rubinfeld, 2010-2011, p. 464). The model is then estimated for each subclass and the estimated coefficients recorded. If the variance in those coefficients is small, then the authors recommend that the original class be certified. Unlike Johnson and Leonard, Dickey and Rubinfeld appear to accept that there will inevitably be some variance in the estimates of coefficients on different regression sub-samples. If that variance is large, then perhaps well identified subclasses should be certified. If there are a very large number of subclasses identified by examining the variances in coefficient estimates across different potential subclasses, then class certification may be altogether inappropriate. The authors argue that: "[t]he appropriate number of subclasses would be resolved by trading off any decrease in variance within classes as the number of subclasses increases against the resulting increases in the cost of litigation." (Bret M. Dickey & Daniel L. Rubinfeld, 2010-2011, p. 465). However, in contrast to Johnson and Leonard as well as Paul Johnson, all of whom argue that any differences in coefficient estimates for *any* explanatory variables across individual class members or potential subclasses should defeat class certification, Dickey and Rubinfeld argue that it is the coefficient used to estimate per-unit damages is what really matters. They write: "defendants should not be able to defeat class certification by showing only that there is substantial variation in one or more explanatory variables. Rather, defendants should be expected to explain why the variation in the demand and supply variables is likely to lead to variation in per-unit damages." (Bret M. Dickey & Daniel L. Rubinfeld, 2010-2011, p. 467).

With respect to these recent papers, it is critically important to keep in mind that, given the non-experimental primary transaction data normally produced by defendants, the use of these decomposition methods can lead to highly unreliable results and conclusions. Without careful theory formulation and evidence outside of the data itself to justify the decomposition groups, these methods amount to nothing more than unscientific data mining. Also, particularly with Johnson and Leonard's recommendation for customer-specific regressions, often many class members have too few transactions to accurately estimate the effect of supply and demand factors on their prices.[25] Further, the arguments and methods proposed by these and other "reformers" ignore the theoretical foundations of reduced-form pricing models. These models are not plucked out of thin air; they are the result of a formal derivation that takes as its starting point the structural *aggregate* supply and demand equations for *the market*. As we describe in Section VI.B.1, these structural equations themselves are an *aggregation* of the individual customers' demand functions and the individual suppliers' cost functions. Accordingly, it is wholly inappropriate to run the derived reduced-form pricing equation for an individual customer and claim to have identified the individual customer's demand function. Further, as noted earlier, it is unreasonable to expect these regressions to produce the same outcome for every cut or subset of the data.

In summary, much of this recent literature has abandoned the reasonable, well-structured hypothesis tests of common impact formulated by Hartman and Doane. Instead, it has advocated methods that invite data mining and ignore the fundamental underpinnings of reduced-form pricing regressions. Given the complexity of this type of

[25] This type of faulty analysis did manage to convince one judge to deny class certification: In re Plastic Additives Antitrust Litigation, 2010 U.S. Dist. LEXIS 90135, *63 (E.D. Pa.) [hereinafter *Plastic Additives 2010*].

analysis, there is much room for the obfuscation feared by Fisher, particularly for lawyers and (ultimately) jurors who lack training in statistics.

5. Alignment of Economic Theory, Discovery Record, and Transaction Data

In their role as scientists, economists pursue the scientific method, by focusing on objective development and testing of theories of how markets and institutions work. This necessarily involves the specification of an internally consistent theory, the collection of data, and the use of statistical methodologies to verify or refute their theories. A principal obstacle faced in the pursuit of this scientific method is that in most class certification matters, we only have non-experimental data available. As econometrician Jeffrey Wooldridge emphasized "rarely can we run a controlled experiment" to uncover a causal relationship between one economic variable and another (Jeffrey Wooldridge, 2002, p. 3). Instead, economic data that are generated and recorded as part of real world interactions consist of two basic components: a systematic signal component that represents a causal relationship between measureable economic variables (for instance, the effect of increased income levels on price, through increased demand), and a component generated by a process of unobserved disturbances — the random noise component. The random noise component can result from measurement error or idiosyncratic variations that are not readily explained by any economic model, and thus are not relevant to an economic expert's analysis. Professional economists recognize that "not all possible variables that might influence the dependent variable can be included…some cannot be measured and others might make very little difference." (Daniel L. Rubinfeld, 2011, p. 314). "As a result, no model could hope to encompass the myriad essentially random aspects of economic life." (William H. Greene, 2011, p. 6).

Since World War Two, much of the economic profession has focused on the development of econometric methodologies, integrating economic theory, data and statistical methodology to analyze observations on economic phenomenon that are not generated by controlled experiments. Standalone analysis of economic data without a clear articulation of the underlying theory is contrary to the fundamental principles of the scientific method. *A priori* information, sourced with the discovery record and fundamental economic principles, must proceed any statistical analysis of the available data on individual transactions between class member and defendants. As noted in one of the seminal books on econometrics, Goldberger (1964) emphasized sound analysis "is not simply a matter of fitting curves to data or 'measurement without theory'" but; "rational methods of measuring economic relationships must be grounded in a specification of the probabilistic mechanisms that link economic observations to economic theory."

In all investigations and analyses of non-experimental economic data, it is the identification of causality, not just correlation that is the core challenge. Theoretical structures generate hypotheses that are subject to common proof for how defendants may or may not have coordinated their actions. In the final analysis, the weight given to any statistical or econometric analysis of the available data must be evaluated in terms of the underlying theory and the documentary evidence that is produced through discovery. As is well known, correlations do not imply casualty. As the court described in *In Re: High-Tech Employee Antitrust Litigation*, economists "analyze correlations which are routinely used…to draw causal conclusions when supported by compelling frameworks and complementary information." (Gregory J. Werden & Luke M. Froeb, 1993, pp. 329-353) [26] The

[26] "A wide-ranging critique of price correlation analysis can be found in Werden and Froeb, who illustrate with a numerical example that 'price correlation can even get things exactly backwards'".

complementary information and compelling frameworks must align with documentary evidence produced through discovery on the observations of coordinated actions and the actual data; one by itself is not sufficient.

Aside from the need to align economic theory, documentary evidence, and any statistical analysis that might be performed, issues about the underlying data that might be produced by defendants can prove to be a gating factor to probative analysis. In class certification matters, data generally come in one of two forms: secondary or primary data. Secondary data, as the name suggests, are not based on primary transactions between the defendants and the members of the class. For most industries, there are one or more data sources that monitor the market performance and report average or prevailing prices. These secondary sources generally collect their data through surveys of industry participants. However, their experimental design for aggregating the average prices that they report and whether the survey is random, representative, and how it might be stratified, is generally treated as proprietary information. The survey data that are summarized are not based on actual transactions. Theoretically, it has been demonstrated that such surveys are subject to strategic exaggerations that serve the interests or exposure positions of the respondents (Gordon C. Rausser, Leo Simon, & Jinhua Zhao, 2016). Such data can become the only source of actual data analysis if and only if the defendants, for whatever reasons, do not end up producing primary transaction data. In a major case, *In Re: Blomkest Fertilizer Inc. v. Potash Corp. of Saskatchewan, 2003, F.3D1028 (8th cir. 2000)* and the lower court decision in this matter, *In Re: Potash Antirust Litigation 954F.FUPP.1334 (D.MINN.1977)* only secondary data were available. The available secondary data used in this particular class certification matter were not sufficiently rich to provide the necessary variation in both the systematic signal as well as the random disturbances to identify the potential causality of two major events that the trier of fact found probative. The secondary data source was referred to as "Green Sheets", which used telephone communications and faxes to obtain survey data as a basis for prices quoted by potash sellers. These data were then aggregated to report average prices to their subscribers (Richard A. Posner, 2009, p. 93).[27]

Even in disputes where primary data detailing actual transactions between defendants and members of the class are available through discovery, it does not follow that all potential forces that drive the value of transactions have in fact been produced along with the transaction database. As a result, it is critically important that PEs be given the opportunity to identify the potential causal determinants of pricing and seek their inclusion in the production of the primary transaction database. For example, in *In Re: High-Tech Employee Antitrust Litigation*, had data on age, education, number of months in the company, gender, location, title, and employer not been produced, there would have been limited opportunities to statistically isolate the attributes and characteristics of each member of the class, and how these attributes and characteristics would have influenced the actual level of compensation and how it might have varied over time. In contrast, in another antitrust litigation involving nurses' wages and allegations of monopsonization (*Reed v. Advocate Health Care, USDC N.D. Illinois, Case No. 06C3337, Sept. 28, 2009*), primary data on individual nurse compensation were produced by the defendants, but various non-conspiratorial causal variables, such as the number of years of professional experience, the educational degree and certifications, was not. The lack of these potential causal variables explaining much of the variation in wages was unavailable and the Court in *Reed* criticized the statistical analysis performed by the PE for not explaining sufficient variation in nurses' wages. The unacceptable share of the systematic variation explaining nurses' wages might have been mitigated had causal information on the characteristics and attributes of nurses

[27] Posner briefly examines this court ruling, suggesting that in price fixing matters, this is an example of disputes that are not "always handled well".

(experience, education, certifications) been produced along with wages and other compensations paid by the defendants. The lesson is that the potential explanatory or causal variables that drive compensation or the values of transactions must be included with the production of the primary data transactions.

6. Misuse of Data Analysis and Econometrics in Class-certification Analysis

Unacceptable uses of data analysis and econometrics fall into two broad categories: a) analysis related to price variation; and b) the misuse of regression models. This section addresses both issues.

6.1 Price Variation

Arguments concerning price variation have become central to most class certifications. This centrality has resulted partly from the *Bogosian* ruling that if plaintiffs can prove that the law of one price holds in the market, then the alleged actions will necessarily have increased that market price and affected all class members. The statement in Rule 23(b)(3), requires that questions common to class members should predominate over questions specific to individual class members, also focuses attention on price variation.

Price-variation analysis has often been performed in a manner that would not pass professional peer review (John H. Johnson & Gregory K. Leonard, 2008, p. 108). Examples of unsound analysis of price-variation include instances in which a PE averages all price variation without considering relevant distinguishing factors, and instances in which a DE relies on endless bivariate scatter plots designed to exaggerate small and insignificant variations.

A problem arises when a PE relies on visual inspections of average price series to argue that prices for all proposed class members move in common. Figure 1, which was presented by the PE in a price-fixing case, is an example of such an analysis. Figure 1 shows average prices for five different strengths of a product sold by defendants, who, according to the plaintiffs, engaged in a price-fixing cartel. From this and similar charts, the PE argued that the market for these products demonstrated a “pricing structure”, and therefore, that the alleged cartel would have resulted in a common price increase for all class members.

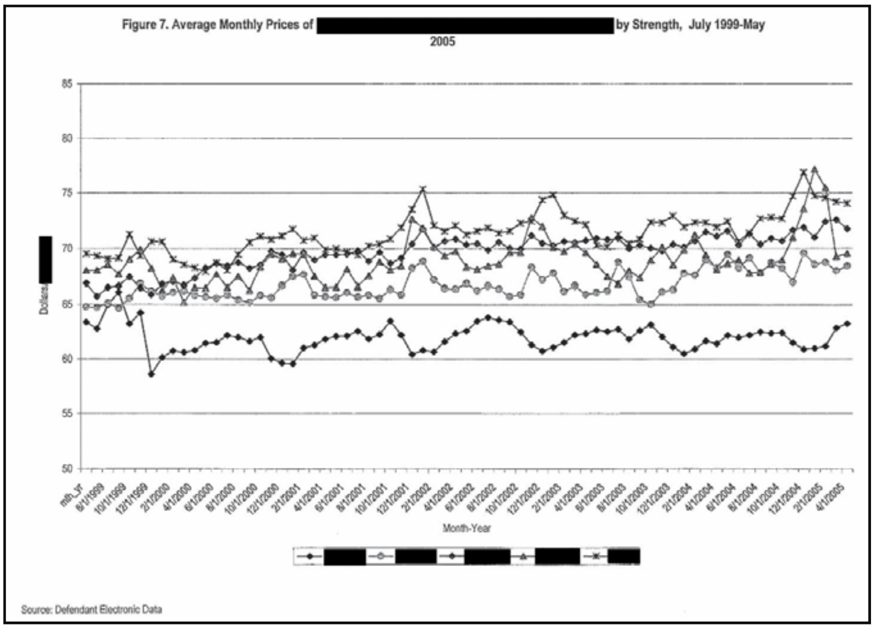

Figure 1 A Faulty Visual Investigation of Pricing Structure

Although analysis of this type has been submitted in U.S. litigations very recently (e.g., in 2010, in *In Re: Plastic Additives Antitrust Litigation*), economists would not consider Figure 1 to constitute sound analysis. The graph does not clearly reveal a common relationship among these price series.[28] It is well understood in economics that spurious correlations can make non-stationary price series appear related. Academic economists use rigorous co-integration tests to investigate such relationships, but PEs often fail to implement this methodology.[29] Some authors have noted that any analysis of price co-movement, be it via a cointegration test or a simple correlation analysis, must also involve "an investigation, or analysis, of the reasons for the observed correlation [or equally, co-integration]." (Michelle M. Burtis & Darwin V. Neher, 2011, pp. 495-532). Although the authors argue that there are many reasons why two price series may exhibit some correlation or co-integration, such as a common input cost, that are not probative regarding common impact, such tests are an essential part of the antitrust economist's toolkit. To be sure, such tests should not be used in isolation from record evidence and economic theory. But in conjunction with other evidence, perhaps strong support that the two products in question are demand or supply-side substitutes, co-integration analysis is probative concerning market definition and whether or not an attempt at a market-wide price increase will likely increase prices commonly across the class.

As well as the unreliable use of visual inspection, averaging prices to show a pricing structure has also come under fire (most notably — and properly — in *Hydrogen Peroxide*). But it is unfortunate that "average" has become something of a dirty word in antitrust litigation.[30] It is true that when the issue under debate is whether individual effects predominate over common effects, care must be taken not to average over, and thereby minimize significant differences among individual effects. But in all markets, individual price variations occur that may not be relevant to the analysis being undertaken. In fact, in order to test a particular hypothesis regarding pricing observations, econometric analysis uses averages specifically to "average out" irrelevant individual differences. Therefore, it is wrong to argue that averaging should not be used in class-certification analysis. The appropriate question is, what kind of averaging is appropriate for investigating the formulated hypothesis? It should be obvious that price differences that are relevant to the analysis (such as prices for different grades of a product that have different end uses) should be factored into the PE's analysis. But it would be correct for the PE to construct an average price series for each grade of product and then investigate whether they are related. This method will "average out" individual price differences that are irrelevant to the specific inquiry (e.g., regional price differences). The latter price differences can then be investigated separately and serve as a basis for comparison.

Thus, even though averages can be used misleadingly in analyses of price variation, economists and lawyers should not completely reject the use of averages and other summary statistics. In fact, doing so could lead to incomprehensible and misleading exaggerations of price variation. Accurate inferences generally require that summary statistics of individual transaction prices (the probability distribution, the mean, variance, and higher moments, if they exist) be assessed. Of course, the PE must still collect and analyze individual price data, but inspecting every single price is neither meaningful nor necessary. Indeed, the statistical principles that econometricians rely on cannot be employed when conducting an individual inspection. For instance, it is not possible to establish whether the differences in price paid by two types of customer are *statistically significant*

[28] In *Hydrogen Peroxide,* the court heavily criticized the PE's visual inspection of price series. In *Plastic Additives,* the judge ruled that the plaintiffs' graphs on pricing structure did not even superficially show that prices moved together.

[29] For a discussion of co-integration analysis, *see* James D. Hamilton (1994), *Time Series Analysis*, chap. 19.

[30] *See*, for example, Reed vs. Advocate Health Care 268 F.R.D. 573 (N.D. Ill. 2009).

without using the full sample of pricing data to capture the magnitude of random variation in pricing. Similarly, it is not possible to establish whether such price differences are *economically significant* without comparing them to the average prices paid in the industry (and the variation in those prices).

Nonetheless, some DEs have rejected the accepted economic methods that are routinely used to evaluate price variation. Perhaps relying on Johnson & Leonard's claim that accurate analysis of price variation is an inherently illogical concept (Johnson & Leonard, 2007, pp. 34, 38), they present only scatter plots of raw data, in a manner that is just as misleading as Figure 1. Specifically, Figure 2 presents a chart presented by a DE in a price-fixing case. The chart shows prices paid by five selected customers for the same strength of a product over a six-year period.

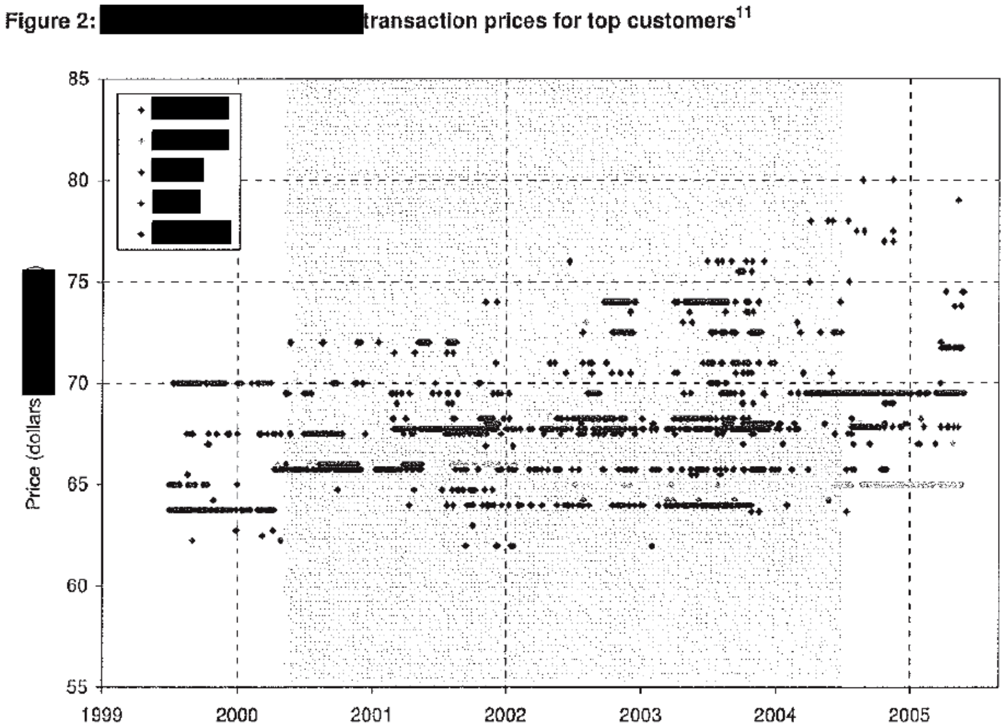

Figure 2 A Faulty Scatter-Plot Investigation of Price Dispersion

To most untrained observers, Figure 2 appears to demonstrate extensive price dispersion. The eye is drawn to the difference between the high and low data points, and since many data points are placed on top of one another toward the center of the chart, it is difficult to assess the density of the distribution. Based on this and similar scatter plots ostensibly showing that class members paid widely dispersed prices, the DE in this case argued that the evidence could not be used to prove common impact.

However, visual analyses such as the one shown in Figure 2 are flawed for at least three reasons. First, Figure 2 relies on subjective visual inspection, just as Figure 1 does. One analyst could decide that wide variation is present. Another analyst, observing that the weight of the observations centers around the middle of the chart, could conclude that dispersion is limited, save for a few outlying observations (which could result from measurement error or some other specific cause). Second, since it is impossible to benchmark this amount of dispersion to any other industry, it is also impossible to say whether this dispersion is typical of a homogenous good or evidence of heterogeneity among class members. The third reason in which Figure 2 is deficient is that factors that explain the scatter may exist that are not typically represented in a two-dimensional scatter plot. Figure 2 conditions on products of similar strength and on similar class members ("top" customers). But many other factors that help determine prices could vary continuously across the observations (e.g., delivery costs and/or changes in supply and demand conditions). These potential explanatory factors cannot be represented in a two-dimensional scatter plot. However, they can easily be evaluated by a multivariate statistical analysis.

Economists can use summary statistics to make objective sense of Figure 2, and economic experts in legal cases should do the same. One summary statistic that professional economists widely accept is the coefficient of variation (the standard deviation divided by the mean or a measure of the second moment divided by the first moment of the probability distribution). This statistic, like the standard deviation, is a representation of the probability distribution of observations around the average. It is superior to the standard deviation since it is normalized by the mean. The resulting statistic is, therefore, a percentage that facilitates comparisons to price dispersion in other industries. Such comparisons, along with other analyses, can assist the court in assessing whether the price dispersion under study indicates intra-class heterogeneity (which would be problematic for class treatment) or is typical of even homogenous products (and so has no real import). Perhaps more importantly, the presence of price dispersion should not hinder class certification if the variations are "systematic and reasonably can be controlled for in any but-for analysis." (James F. Nieberding & Robin Cantor, 2007, pp. 61-84). With sufficiently detailed data, differences in transaction characteristics can be included in regression analysis to explain the systematic components of price variation. Naturally, to demonstrate this at the class certification stage requires detailed discovery.

6.2 Misuse of Regression Models

Regression models are also misused in many class certifications. As noted earlier, the former legal standard for the class-certification stage required the PE to show only that a feasible method of assessing class-wide damages existed. But since *Hydrogen Peroxide*, in which the court ruled that a mere assurance from plaintiffs that they can perform such an analysis is insufficient for class certification, PEs have increasingly relied on multivariate regression analysis of impact and/or damages in order to demonstrate that their statistical methods will actually work. Meanwhile, the DE often tries to pick the regression analysis apart often without any underlying testable hypotheses. A major concern is that although such analyses have influenced court rulings, many of the analyses used heretofore by both sides could not survive an objective peer review standard.

6.2.1 Incorrect Decomposition of Regression Samples

As described in our literature review in Section IV, recent papers on the economic analysis for class certification have recommended that the damage regression model offered by PE should be tested for commonality by running it for each class member individually and comparing the coefficient estimates. The primary advocates of this method, Johnson & Leonard, argue that the mere fact that this test for commonality requires individual class member regression runs defeats the notion that class certification is a more efficient means of adjudication than individual analysis. They call this the "common proof paradox" (Johnson & Leonard, 2007, p. 351). We do not judge this argument worthy of serious consideration, other than to say that with modern computing the application of the same regression model to hundreds or thousands of class members can be achieved in an automated fashion in a matter of hours. Efficiency is not the problem with the "Johnson & Leonard" method. Rather, the problem is that the method ignores the foundations of reduced-form pricing models and the principles of scientifically reliable hypothesis formulation. To fully explain this perspective, we describe the reduced-form regression models typically used by PEs.

A PE will often include in his expert report a trial run of a regression model designed to quantify class-wide damages. The model that is the workhorse of antitrust litigation is the reduced-form pricing model (Baker & Rubinfeld, 1999, p. 391; James F. Nieberding, 2006, pp. 361-380). This model regresses price on a set of transaction characteristics deemed to be important price determinants.[31] Measures of input costs and demand control for differences in supply and demand factors between a competitive benchmark and the class members' transactions. PEs often capture the effect of the defendants' actions by using a dummy variable set to 1 for the class members' transactions and set to 0 for the benchmark transactions. A typical reduced-form damages model is specified as:

$$p = \beta_0 + \gamma D + \beta_1 z_1 + \beta_2 z_2 + \varepsilon_0 \tag{1}$$

Where p is the price paid for a good; β_0 is a constant to be estimated by the regression; γ is the coefficient on a distinguishing variable D that distinguishes class transactions from the competitive benchmark transactions; β_1 is a vector of coefficients on the exogenous demand factors contained in z_1; and β_2 is a vector of coefficients on the exogenous supply factors contained in z_2. The coefficient of interest is γ — the effect of the alleged illegal action on prices. We will refer to γ as "the overcharge" (which it would be if the case were a monopolistic price-fixing conspiracy).[32] The PE may estimate (1) and cite the estimated overcharge, $\hat{\gamma}$, as evidence that his model works and finds positive damages for the class.

Either the PE or the DE may then decide to assess the commonality across class members of the overcharge estimated by the model described in equation (1). Acceptable methods for performing such an assessment exist. For example, a PE or DE can hypothesize that some group of class members was able to bargain its way out of the price-fix. In order to investigate this hypothesis, either expert can identify such a group in the transaction data, create a distinguishing variable for that group's transactions, and interact that distinguishing variable with D in order to establish whether the estimated overcharge for that group is statistically and economically different from the overcharge for the rest of the class.

A method that is not professionally acceptable (but has been used in antitrust litigations and is advocated by some of the papers described in our literature review) is to run (1) individually for each class member and estimate γ for each case. In Plastic Additives, for example, the DE took the PE's class-wide damage regression and ran it for each class member, one at a time. The court noted that of the "115 [class members that he did this for,] 81 show[ed] no evidence of a statistically significant increase in prices as a result of the alleged conduct."[33] The PE protested, arguing that the individual regressions were not well specified, as was evidenced by two facts: the coefficients on supply and demand variables changed with each regression run, and they were sometimes statistically insignificant or had a sign inconsistent with economic theory. But the court was not convinced by the PE's argument, which the court characterized as "based solely on economic theory."[34] Instead, the court agreed with the DE that although "in a hypothetical textbook example…[an] increase in cost will increase prices[,]…in the real world[,]…there are many reasons why…the relationship between the two variables is not positive."[35]

The court's faith in DE's analysis was misplaced, because the DE misinterpreted the critical assumptions underlying model specification in equation (1). The coefficients β_1 and β_2 on the exogenous demand and supply

[31] These characteristics are sometimes determined through a separate "common-factors" regression analysis.

[32] Strictly speaking, if the model was specified in logarithms, the percentage overcharge would be equal to $\exp\{\gamma\}-1$.

[33] *See Plastic Additives 2010*, at *61.

[34] *See id.*, at *67.

[35] *See id.*, at *69.

factors z_1 and z_2 are correctly interpreted as measuring the relationship between *market* supply and demand factors and the *market*-equilibrium price.[36] They do not measure the willingness to pay for a particular customer (class member) of the product or that buyers' individual demand relationships.[37] In essence, even if it is implemented for an individual consumer, the reduced-form regression does not identify a consumer's individual demand function.

To understand why equation (1) does not make it possible to identify customer-specific demand factors, we can consider the underlying structural equations for the case of a competitive regime benchmark.[38] Underlying equation (1) are the following market demand and supply equations and market-equilibrium condition:

$$q_D = a_0 - a_1 p + a_2 z_1 + \varepsilon_1 \tag{2}$$

$$q_S = b_0 + b_1 p + b_2 z_2 + \varepsilon_2 \tag{3}$$

$$q_D = q_S \tag{4}$$

The reduced form is derived using (2), (3), and the market-equilibrium condition (4), so that the coefficients in (1) are:

$$\beta_0 = \frac{a_0 - b_0}{a_1 + b_1},\ \ \beta_1 = \frac{a_2}{a_1 + b_1},\beta_2 = -\frac{b_2}{a_1 + b_1},\ \ \varepsilon_0 = \frac{\varepsilon_2 - \varepsilon_1}{a_1 + b_1} \tag{5}$$

Consider, β_1, the coefficient on the exogenous demand factors in the reduced-form equation (1). From (5), we see that this coefficient is composed of: a_2 — the shift in demand caused by a change in the exogenous demand factor z_1; a_1 — the slope of the market demand curve, and b_1 — the slope of the supply curve. This is all fundamental, text-book econometrics.[39] More fundamental and a key insight is the way the parameters a_1, a_2, and b_1 relate to the characteristics of the individual market participants.

Let us consider a simple illustrative example, that of a market for washing machines. Each consumer's individual demand is discrete: each person only wants one washing machine and will buy one when the price falls below his reservation price. Because consumers have different tastes and incomes, the distribution of reservation prices is continuous and results in a continuous demand function. For each price p, the demand curve gives the number of consumers whose reservation price is at least p. The distribution of reservation prices determines the slope of the demand curve, a_1. Suppose that the market-demand shifter, z_1, is average income per capita. Consumers' reservation prices will increase as their income increases; the exact amount of each consumer's increase will be determined by the specifics of the consumer's utility function and budget constraint. Thus, an increase in average income per capita will shift the distribution of reservation prices upwards, so that a higher

---

36 *See* Hal R. Varian (1992), *Microeconomic Analysis* (3rd ed.) [hereinafter *Varian 1992*], at 202-203, which describes how the reduced form model is derived from the "demand and supply system" that defines the "demand and supply for some good…its price…[in terms of]…variables that affect supply and demand….The reduced-form parameters can be used to predict how the ***equilibrium price*** will change as the [supply and demand] variables change," (emphasis added). Economists define the ***equilibrium price*** as the "price at which the quantity of a good supplied [by the market] is equal to the quantity demanded [by the market]", John Black, Oxford Dictionary of Economics (2nd ed., 2002), at 150.

37 *See* Varian (1992), at 219, which describes how the equilibrium price, which Varian previously established is the focus of a reduced form model, is determined by ***industry*** supply and demand: "The industry supply function measures the total output supplied at any price. The **industry demand function** measures the total output demanded at any price. An **equilibrium price** is a price where the amount demanded equals the amount supplied," (emphasis in the original).

38 In a typical damage analysis, equation (1) will be performed on stacked data of transactions from both a benchmark competitive regime and a conspiratorial regime. For the sake of simplicity, we can consider the derivation of the reduced form under the competitive regime, where the market equilibrium is determined by the intersection of the market demand and supply curves. Our points also hold for the conspiracy regime, where firms have conspired so that the market equilibrium is determined by the interaction of the marginal revenue and supply curves.

39 Specifically, the notation used here (and the following washing-machine example) are adapted from *Varian 1992*.

quantity of washing machines will be demanded for a given price. The extent of this shift in the quantity demanded at each price is given by a_2, and the aggregation of the effects on consumers' reservation prices (for those consumers who received an increase in income) will determine a_2. For the sake of simplicity, assume that washing-machine manufacturers are all identical and have the same marginal-cost curves. Market supply is then a simple aggregation of those marginal-cost curves, with slope b_1. The demand-shift increase that results from the hypothetical increase in income per capita leads to higher output, as manufacturers move up their marginal-cost curves and are able to cover increased incremental costs at the new, higher prices. The new market-equilibrium price from the shift in z_1 is given by equation (1) and depends on β_1.

Clearly, estimating equation (1) (with z_1 equal to average income per capita) for an individual consumer could never make it possible to recover the specifics of that consumer's demand function or of how the consumer's reservation price varies with income. If such a consumer-specific regression were estimated, and it showed β_1 to be negative or zero when it was positive for the full sample of consumers, it would not be valid to claim that this anomalous result derived from the specific consumer's different response to increased income.

Essentially, the framework does not allow the identification of the specifics for an individual consumer's demand function. Moreover, that one consumer's demand function alone does not determine the price that the consumer pays. Rather, the intersection of the *market* aggregate demand and supply curves determines the price that *all* consumers pay.[40] Indeed, a buyer of a washing machine may never need to pay his reservation price, if the equilibrium price is always below the maximum that he is willing to pay. Thus, even the data on prices paid by that individual consumer may contain no information on his individual demand function or budget constraint. Also, the finding of β_1 to be negative or zero in the consumer-specific regression estimation is likely due to insufficient transaction data, and the results from such an individual regression run would be unreliable.

As in the case of decompositions across individual members of the class, DEs frequently decompose complete samples and investigate individual groupings of class members. Generally, they then compare the results of the estimated impacts for the individual group to the full sample analysis of the empirical model results that the PE has presented. Generally they refer to this as sensitivity analysis. Unfortunately, this is *not* reliable sensitivity analysis. Instead, without alignment of the discovery record and the underlying economic theory, it is nothing more than "data mining". This distinction is documented in a recent book published on quantitative techniques in antitrust analysis by Davis and Garcés who state "when the regression is run on a sample composed by different groups on distinct time frames it is useful to test whether the results are robust to the exclusion of some of the groups or time intervals. Since the regression on the whole sample gives us an average of the effect over the sample population, we want to make sure this average is representative of the effect and is not the average of widely different magnitudes. If excluding a group (such as a country or a firm) or a time period drastically affects the results, this fact should be reported. Particularly, this type of robustness check will help detect whether the results are driven by one small part of the sample as opposed to by the whole sample." (Davis, Peter, & Eliana Garcés, 2009, p. 118). As this statement clearly establishes, the focus of sensitivity analysis should be to verify whether the average effects of the complete sample are driven by values of a particular group. The average effect then should be recomputed by removing a group, or groups, with potentially extreme values and comparing the

[40] *See* Varian (1992), at 152, which describes how the "aggregate demand for [a] good" is derived from a summation of the individual consumers' demand functions; *see also* Andreu Mas-Colell, Michael D. Whinston & Jerry R. Green (1995), *Microeconomic Theory*, at 105-109, which describes how aggregate demand is determined by a sum of consumers' individual demand functions.

result to the original full sample estimation. Some DEs find this unacceptable and instead of focusing on the results found in the main sample when one group is removed, they focus on the results found when the model is run on that one group in isolation. This does not qualify as sensitivity analysis and, without documentary evidence or underlying economic theory to support the isolation of the group in question, such an analysis is unreliable data mining.

6.2.2 “False Positives”

Another approach that DEs have used to attack PEs’ regression models and deny common impact is the so-called “false-positives analysis”. This analysis deserves special mention because it is sufficiently complicated to confuse courts, counsel, and experts alike. For example, the PE in both *Reed vs. Advocate Health Care* and *In Re: Rail Freight Fuel Surcharge Antitrust Litigation* used a common-factors regression to demonstrate that most of the price variation across individual transactions could be explained by a set of pricing factors common to all class members. The PE then calculated a class-wide overcharge, using a damage analysis that was based on a higher level of aggregation than the common-factors regression but which included the same set of common transaction factors. This damage analysis, which compared class members’ transaction prices to benchmark transaction prices (as in equation (1)), also had a different model specification than did the common-factors regression: it included demand and supply factors to control for economic differences between class members’ transactions and the benchmark period transactions.[41]

In response, in order to calculate “predicted but-for prices”, the DE took the estimated class-wide overcharge found by the PE’s damage regression and subtracted it from each of the predicted transaction prices found by the common-factors regression. Then, in order to calculate each class member’s overcharge, he compared these “predicted but-for prices” to the actual prices paid by the class members. The DE first cited instances in which this individual overcharge was zero or negative as evidence that some class members did not suffer harm. He then set out to estimate the probability that a given transaction was not affected by the conspiracy (even though the PE’s damage regression had estimated that the same transaction *was* affected). The DE defined as “false positives” instances in which the *predicted* but-for price was lower than the actual price (i.e., instances in which the transaction was damaged, according to the PE’s model), at the same time that the *actual* but-for price was really higher than the actual price (i.e., an instance in which the transaction was actually not damaged). However, since the “*actual* but-for price” occurs only in the counterfactual “but-for” world, in which the alleged conspiracy did not occur, it cannot be observed directly. As a result, it can only be simulated. The DE did this by adding random draws from a standard statistical distribution to the *predicted* but-for price.

To understand the DE’s method more fully, let Y_a denote the actual price in the conspiratorial world (observed from the transaction data); let Y_p denote the predicted price in the conspiratorial world (predicted by the common-factor regression); and let d denote the average overcharge (estimated from the damage model). The DE then constructs Y_{pb}, the predicted but-for price, and Y_{ab}, the actual but-for price, as follows:

$$Y_{pb} = Y_p - d \tag{6}$$

$$Y_{ab} = Y_{pb} + \epsilon_b \tag{7}$$

Where ϵ_b is a simulated term drawn randomly from a normal distribution.

A “false positive” is defined as a transaction for which $Y_a > Y_{pb}$, but in reality $Y_a < Y_{ab}$. In other words,

[41] This step was not necessary for the common-factors regression. The latter did not include a benchmark because its purpose was to investigate price dispersion rather than to estimate any overcharge.

although the actual price is greater than the but-for price predicted by the model (i.e., $Y_a > Y_{pb}$, which suggests an overcharge for this transaction), in reality the actual price is less than the but-for price (i.e., $Y_a < Y_{ab}$, meaning that in reality there was no overcharge for this transaction). We can re-write $Y_a < Y_{ab}$ as follows:

$$Y_a < Y_{ab} = Y_{pb} + \epsilon_b = Y_p - d + \epsilon_b \tag{8}$$

In constructing Y_{pb} in equation (6), the DE uses both the damage model (for d) and the common-factors regression (for Y_p). The use of the two models in this fashion is the first error in the DE's analysis. His analysis conflates two entirely different regression models, each of which has a different specification, level of aggregation, and sample selection.[42] Furthermore, by applying the overcharge from the damage regression to each class member's transaction price, the DE committed a second error in assuming that all transactions experienced the same overcharge. In contrast, the PE's damage regression model does not impose this restriction. It simply measures the *average* (class-wide) overcharge across the class.

The third error in the DE's analysis is that, because it relies on draws of ϵ_b from an *unbounded* normal distribution, it is inevitable that at least some "false positives" will be found. The common-factor model that generates the predicted conspiratorial price, Y_p, is equal to the actual price, Y_a, plus an error term, ϵ, that includes factors omitted from the model. This means that equation (8) can be written as

$$Y_a < Y_a + \epsilon - d + \epsilon_b, \text{ or } d < \epsilon + \epsilon_b \tag{9}$$

Equation (9) indicates that a "false positive" will occur if the damages are outweighed by the sum of two error terms, where ϵ derives from the conspiratorial world and ϵ_b is simulated via a draw from a normal distribution that uses the root mean square error (RMSE) from the common-factor regression as its standard deviation. This formulation establishes the argument that the DE is trying to make: the higher the error rates in the regressions, the more likely the model will falsely find that a transaction is damaged when in fact it was not. However, the fact that the simulated error term, ϵ_b, is drawn from an *unbounded* normal distribution *guarantees* that "false positives" will be found.

The tautological nature of this "false-positives" argument is mathematically apparent once we specify a probability distribution for $\epsilon + \epsilon_b$. Since the simulation assumes independent normal distributions for ϵ and ϵ_b, their sum will also have a normal distribution. By definition, the probability of d being smaller than the normally distributed sum $\epsilon + \epsilon_b$ will be

$$\Pr(d < \epsilon + \epsilon_b) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\epsilon+\epsilon_b} e^{-\frac{x^2}{2\sigma^2}} dx \tag{10}$$

Which is always positive, guaranteeing that the simulation will always find some "false positives," whether or not they actually exist.

A fourth problem occurs when the DE uses the root-mean-square error (RMSE) from the common-factor regression to specify the standard deviation of his normal distribution. When the standard deviation is large, there is a higher probability of drawing "false positives." Using the RMSE statistic from the common-factor regressions overestimates the "false-positive" rate because these regression models include neither conspiratorial covariates

[42] The common-factor regression, which focuses on transaction characteristics, does not always contain variables for economic supply and demand factors or for the variables distinguishing the alleged conspiracy. Accordingly, it cannot accurately predict but-for prices in the fashion that the DE attempted. Further, the common-factor regression is often run on a sample of transactions both in and outside of the class definition. But many of these transactions are not expected to be damaged at all. The purpose of the PE's common-factor regression is to determine whether *common transaction characteristics predominate* for both transactions in the class definition *and* transactions used as a benchmark. Thus, the common-factor and damage-regression models have different purposes and are not comparable in any practical sense.

nor, in many cases, supply and demand characteristics. (This is because the purpose of the common-factors regression is to identify the amount of variation in prices explainable by observable transaction characteristics alone.) Equation (11) can be used to show that decreasing σ to σ_1 ($\sigma > \sigma_1$) means that the probability of a false positive would decrease:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\epsilon+\epsilon_b} e^{-\frac{x^2}{2\sigma^2}} dx > \frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{\epsilon+\epsilon_b} e^{-\frac{x^2}{2\sigma_1^2}} dx \tag{11}$$

A fifth problem with the analysis is that in order to obtain a meaningful number of "false positives", one must assume that ϵ and ϵ_b need to be assumed independent of each other (otherwise, the errors compound, and if their distributions are highly correlated, many "false positives" are erroneously found). Given that one error is the difference between the *actual price* and the *predicted price* in the common-factor regression, and the second error is the difference between the *actual but-for price* and the *predicted but-for price,* it is likely that the errors would be related. Conceptually, the second error term results from a hypothetical run of the common-factors regression in the but-for world. The DE generates this result by drawing the second error from the normal distribution and constraining it to be independent of the first error term. But it is highly unlikely that the two errors are unrelated. More likely, the extent to which the common-factors model under- or over-predicts a consumer's price (for example, because that consumer has better information than other consumers, a typical theoretical reason for price dispersion and residuals) will be the same in both worlds. At a minimum, a DE would need to convince courts otherwise, by arguing that the fact of the conspiracy somehow affects the distribution of residuals across consumers.

A sixth problem is more conceptual. Essentially, the DE argued that the lack of precision in the PE's proposed regression models meant that the models could show, falsely, that a purchaser had been damaged by the alleged conspiracy when the purchaser was not, in fact, damaged by any conspiracy. The DE claimed that he could calculate (and had calculated) the probability that the PE's analysis would produce just this type of false result. However, the DE typically does not calculate the correct conditional probability. *The probability that a customer was not damaged, given that the model estimated that it was damaged* is often set equal to *the probability that the model estimates damages for a customer, given that it was not damaged*. This well-known logical mistake is often referred to as the "Inversion Fallacy". Put simply, the probability of event A, given that B is true, is not generally equal to the probability of event B, given that A is true. By Bayes' theorem,

$$P(A \mid B) = P(B \mid A).P(A) / P(B) \tag{12}$$

The probability of A, given B, is equal to the probability of B, given A, and the probability of A, divided by the probability of B. In Bayesian terms, one has to consider not only the error rate of the damage model but also the circumstances under which a consumer finds itself in the damage model in the first place (i.e., the probability that the consumer has been affected, regardless of what the damage model says). The correct conditional probability is given by

P (consumer was not damaged | model estimates that the consumer was damaged)
*= P (model estimates that the consumer was damaged | consumer was not damaged) **
P (consumer was not damaged) / P(model estimates that the consumer was damaged) (13)

In other words, the chance that a consumer was not damaged *even though* the damage model found that he was, is higher when the probability that the consumer was not damaged is high in the first place.

In legal cases, this failure to consider this conditional probability correctly has been called the Defense

Attorney's Fallacy.[43] When estimating a damage model in an antitrust matter, a PE does not select consumers randomly from across U.S. industry in order to determine whether those consumers have been affected by a price-fixing conspiracy. Rather, the consumers who find themselves in the damage model are the subject of a litigation, and by the time he performs the damage regression, a PE should have already provided strong evidence that these consumers were, in fact, affected by the alleged conspiracy. This step creates a prior probability of harm, which must be considered in order for an analysis of "false positives" to be the basis for sound inferences.

In a more general setting, DEs often attempt to demonstrate the presence of false positives by running PE's model on transactions that are believed to be unaffected by the alleged conduct. For illustrative purposes, consider the case of a biomedical company that develops a new test for a disease and wishes to determine whether the test reliably identifies affected patients. The biomedical company can test a group of patients who are known not to have the disease. If the test regularly indicates that some numbers of these unexposed and healthy patients have the disease, the test generates false positive results thus making it unreliable for diagnosing disease. Similarly, if a model finds "damages" in data where the alleged collusive conduct is known to be absent, these too are false positive results. However, in order for such results to be truly false positives, one must be very certain based on evidence and coherent economic theory, that such transactions are, in fact, unaffected by the alleged conduct. Or, returning to our example, that the group of patients really do not have the disease.

6.3 Unnecessary and Erroneous "Disentanglement" of Defendants' Acts

The Supreme Court's ruling in *Comcast* has been used by many DEs to argue that where defendants are alleged to have engaged in various acts to manipulate prices (perhaps a combination of bid-rigging, price agreements with supposed competitors, and foreclosure of other would-be competitors) that any reliable damages model must be able to disentangle and estimate the effect on prices of each act in isolation, rather than estimate the totality of their effect using standard benchmark regression techniques. To be sure, PE in *Comcast* designed a model to estimate the combined effect of four alleged acts, and then fell afoul of a ruling that removed three of those four alleged wrongful acts from the class action. But typically, the correct economic approach is to investigate the totality of the effect of acts designed to manipulate the competitive environment. Indeed, many cartels and firms engaged in anti-competitive behavior use a set of mechanisms in order to manipulate prices. This has been well documented in the economics literature.[44] The set of such mechanisms are often complementary, meaning that defendants jointly implement them in such a way that their combined effect on prices is greater than the sum of their parts. This is perfectly logical and consistent with standard economic concepts. This means that the marginal effect of each supply control mechanism on prices depends on the implementation and effectiveness of all the other acts and vice versa. This sort of complementarity, including how the marginal effect of one action depends on the presence or absence of other actions, is well understood and often studied by economists.[45]

[43] Consider a person on trial for murder whose rare blood type is found to match the blood found at the scene of the crime but is also shared with 1% of the population. The defense might argue that in a city of one million, 10,000 people could be guilty of the crime. However, those people do not find themselves on trial. When evaluating the probability of guilt based on this blood evidence, the jury also needs to consider the prior probability of the defendant's guilt, based on the evidence that put him on trial in the first place. *See* Walter F. Abbot & John Batt, *A Handbook of Jury Research*, The American Law Institute, 1999, pp. 3-8.

[44] *See* George A. Hay and Daniel Kelley, *An Empirical Survey of Pricing Fixing Conspiracies*, *Journal of Law and Economics*, Vol. 17, 1974, pp. 13-38, Appendix at pp. 29-38; John Asker and Heski Bar-Isaac, *Raising Retailers' Profits: On Vertical Practices and the Exclusion of Rivals*, *The American Economic Review*, Vol. 104, 2014, pp. 672-686.

[45] "If, in a maximization problem, the objective reflects a complementarity between an endogenous variable and an exogenous parameter, in the sense that having more of one increases the marginal return to having more of the other, then the optimal value of the former will be increasing in the latter. In the case of multiple endogenous variables, then all of them must also be complements in order to guarantee that their increases are mutually reinforcing. This conclusion follows directly from the underlying

Accordingly, an analysis of the combined effect of these acts on prices — e.g., the overall effect of an alleged conspiracy on prices — is often appropriate, unless there is a strong legal motivation to do otherwise, as was ultimately the case in *Comcast*.

7. Sound Methods of Class-certification Analysis

Since class-certification analysis is generally empirical, great care must be taken in drawing inferences. The preceding section described types of analysis that are unacceptable because they are misleading or simply false. In this section, we describe acceptable approaches to class certification that rest on sound economic theory. We begin by discussing the principles of scientific practice that must be adopted in order to avoid unreliable inferences. These concepts and practices are (or at least, should be) well known to practicing economists, but they can sometimes remain obscure in the litigation arena.

7.1 Formulating a Hypothesis

It is tempting for experts engaged in certification analysis to comb through the available data to find support for their positions. But this approach is not consistent with sound scientific practice. The latter requires the investigator *first* to specify a theory with testable predictions and *then* to test those predictions with empirical observation. As we have discussed in prior work (Michael D. Hausfeld, Gordon C. Rausser, Gareth J. Macartney, Michael P. Lehmann, & Sathya S. Gosselin, pp. 77-133), the process of formulating hypotheses has been formalized in science and economics over the centuries. The 18th century philosopher David Hume gave structure to the scientific process in *An Enquiry Concerning Human Understanding,* which argued the necessity of reasoning, rather than merely observation (David Hume, 1748). Karl Popper almost 200 years later wrote that “[a] scientist, whether theorist or experimenter, puts forward statements, or systems of statements, and tests them step by step. In the field of the empirical sciences, more particularly, he constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment.” (Popper K., 1935, p. 3) Popper writes that “a hypothesis can only be empirically *tested* — and only *after* it has been advanced.”[46] As the economist Earl Swanson writes, one starts with a theory and goes from it to “observations instead of vice versa.”(Earl R. Swanson, 1960, pp. 1483-1486). This sentiment has been echoed elsewhere in the literature, “We are never justified in inferring theories from empirical observations.” (Carl F. Christ and Gordon C. Rausser, 1973, pp. 271-279). Patterns are readily observed in the real world, and paying casual attention to them, rather than following a formal process of logical theory, model specification and estimation, may lead to spurious inferences (Harold T. Shapiro, 1973, pp. 250-258).

The process of searching data (which can include running multiple diverse regression models) to find a hypothesis (rather than forming one first) is often referred to as “data mining”. In one study on the dangers of data mining, consider the supposedly substantial evidence that stock market returns are higher on certain days of the week, weeks of the month, months of the year and so on (Ryan Sullivan, Allan Timmerman, & Halbert White, 2001, pp. 249-286). Although these patterns are found to be statistically significant using standard statistical tests,

complementarity relationship and is thus independent of the aforementioned superfluous assumptions. It thus holds even if there are multiple optimal values of the endogenous variable(s),” Amir, Rabah, “Supermodularity and complementarity in economics: An elementary survey”, *Southern Economic Journal*, Vol. 71, No. 3, 2005, pp. 636-660. “The term strategic complementarity means that the optimal action of one agent is an increasing function of the actions of other,” John Haltiwanger and Russel Cooper, *Evidence on Macroeconomic Complementarities*, *The Review of Economics and Statistics*, Vol. 78, No. 1, 1996, pp. 78-93.

[46] *Popper*, at 7 (emphasis in the original).

the authors demonstrate through 100 years of daily data that this statistical significance disappears once the distortions due to data mining are accounted for. The authors conclude that the traditional statistical tests do not take into account the relentless searching that has gone into finding patterns in stock market returns that are merely the product of chance.

Tests of statistical significance involve a margin of error, often reported as a “confidence level”. This confidence level records the probability that the test result is found purely because of random variation in the data, and is not evidence of a real relationship between two variables. With relentless testing on the same dataset, it is inevitable that results will be found that fall into this error margin and, although reported as statistically significant, such results are, in fact, due merely to random variation in the data.[47] Therefore, caution should be exercised by the courts in the certification process to avoid attributing any unwarranted significance to economic conclusions based on data mining tactics.

Scientifically reliable analysis in class certification should avoid data mining and instead apply hypothesis formulation to each of the three elements of liability, impact, and damages. An expert’s analysis at each element may investigate a relevant *testable* hypothesis. The result of that test may then serve as the foundation of the subsequent element as what is known in economics and statistics as a *maintained hypothesis*.[48] For instance, PE may start by testing the liability hypothesis — i.e., did the defendants engage in the alleged actions? If the liability hypothesis is supported by the evidence, it becomes a *maintained hypothesis* in the impact analysis. The PE can then test the common impact hypothesis that the defendants’ behavior *caused* injury to the class, taking the maintained hypothesis as a fact. If such causation is also demonstrated, the PE will then have, consistent with the rigorous application of scientific principles, constructed *maintained* hypotheses sourced with both liability and common impact. With these two pillars as a foundation, the expert can proceed to assess the damages hypothesis, i.e., whether there is or is not a quantum of damages suffered by the class.

7.2 Causality

Economists try to capture causal relationships. As noted earlier, however, correlation does not imply causality. This critical distinction is often overlooked in class-certification analysis. The most obvious example occurs when an expert relies excessively on univariate analysis, in which pricing is investigated along only one dimension, rather than multivariate analysis, which attempts to control for each of the observable characteristics that would likely influence price in a meaningful way. In order to identify the *reasons* for observed price movements, the economic expert must, at a minimum, use multivariate analysis to control for various economic and transaction characteristics that vary over time and would otherwise make it impossible to isolate any wrongful price manipulation. However, even multivariate regression analysis is limited, because all of the factors that influence price cannot necessarily be observed in the available data. For example, consider the case of an allegedly conspiratorial tax applied to the “base” prices paid by a proposed class of consumers for a product. An individual customer might manage to achieve greater base-price discounts for any number of reasons not reported in the data,

---

[47] Traditional statistical tests, such as a t-test, report a confidence level. An estimated statistical effect of an explanatory variable on a dependent variable is said to be statistically significant at a 95% level of confidence if there is only a 5% probability that the chosen sample exhibits the effect (due to random variation in the data), even though in reality there is no true effect. With many repeated statistical testing of the same dataset (in this case the dataset of stock returns), researchers will eventually find statistical variations that fall into this 5% margin of error, but will be reported by the t-test and the researcher as statistically significant (this is known as a Type I error).

[48] Gordon C. Rausser, “The validity and verification of complex system models”, *American Journal of Agricultural Economics*, Vol. 55, No. 2, 1973, pp. 271-279: “[T]he selected maintained hypotheses … isolate[s] a still smaller set which represents the testable hypotheses.”

such as the customer's acquisition of an interest in a related business that would add to volume purchases or election to leadership in an influential trade association. Such facts could, however, have been the same in the absence of the collusively imposed tax and would not necessarily mean that the customer was able to escape the effects of the tax.

Most industries exhibit some price bargaining and discounting, but this alone does not necessarily mean that collusive — and effective — price manipulation cannot also occur.[49] It is possible to manipulate a particular component of price without manipulating — or triggering offsetting adjustments in — the other components. For example, in an industry in which negotiation over the base prices paid by customers has always taken place, the imposition of a uniform tax or supplemental fee on top of those base prices may be an effective means for horizontal competitors to collusively implement across-the-board price increases. By itself, the fact that some class members experience declining base prices during the class period does not justify an inference that they have successfully countered the tax or supplemental fee that has been added to their base prices. Indeed, the customers' base prices might well have declined in the same way if the tax had not been imposed. In order to empirically support the claim that certain customers have been able to avoid a supplemental tax, it would be necessary to establish a causal link between the imposition of the tax and the downward adjustment in these particular customers' base prices.

This approach is consistent with the academic literature. As Wooldridge emphasizes, simply finding that two variables are correlated is rarely enough to conclude that a change in one variable causes a change in another.[50] Economic variables are often correlated because they are simultaneously determined by some other factor. Wooldridge continues, "The notion of *ceteris paribus* — that is, holding all other (relevant) factors fixed — is at the crux of establishing a causal relationship . . . [a] first course in econometrics teaches students how to apply …[multivariate] regression analysis to estimate *ceteris paribus* effects."[51]

Econometricians seeking to identify a causal relationship in the presence of unexplained price variation use an approach referred to in the profession as an "identification strategy." Ideally, an identification strategy involves a natural experiment, in which a treated group is compared to an otherwise identical or near-identical untreated group in order to identify the effect of the treatment. For example, if a conspiratorial tax was sequentially administered to different segments of the class, the expert could compare the base prices paid by those subjected to the tax in each period with the base prices for those to whom the tax had not yet been applied. Provided sufficient controls were incorporated to minimize any differences between the two groups, a decrease in base prices paid by class members who suffered the tax relative to those who, during the same period, did not suffer the tax would strongly support the contention that some discounting was being implemented to offset the tax.

Such identification strategies — a key part of any model design — require support from information outside of the model and pricing data. This is standard practice in economics, and in litigation, a responsible PE should use the documents produced in discovery to support his choice of identification strategy, competitive benchmark, and model design. This standard practice is often ignored by DEs, who strive to argue that any persuasive

[49] Indeed, in industries with extensive price negotiation, the publication of collusively set price schedules may merely adjust the point at which those negotiations begin, artificially moving the entire schedule up or down while preserving the individual price differences that would otherwise exist. A conspiracy that manipulates those list prices by adjusting the starting point for negotiations was recognized in *In Re: Ethylene Propylene Diene Monomer (EPDM) Antitrust Litigation, 256 F.R.D. 82, 90 (U.S.D.C. Conn. 2009)*.

[50] *See* Wooldridge 2002, p. 3.

[51] *See id.*, p. 4.

documentary evidence of liability is somehow beyond the expertise of an economist to analyze. Not so. Economists routinely analyze information that goes beyond pricing data. Courts have recognized the importance in using documentary support for the correlations that are found in empirical analysis. For instance, the court ruled in *In Re*: *High-tech Employee Antitrust Litigation* that: "[i]n the instant case, the compelling documentary evidence along with Plaintiffs' expert theories and correlation analyses are capable of demonstrating causation on a classwide basis."[52]

This is the proper method for a PE to use when investigating whether common evidence can be used to support a maintained hypothesis of liability and common impact that could generate a quantification of damages. Both the PE and the DE should formulate hypotheses and identify causal relationships according to accepted practice, recognizing that alignment of the discovery record, variable relationships, and the available sample data. However, at the class-certification stage, the PE should not need to do more than demonstrate that this investigation can be viably performed, using admissible evidence common to the proposed class.

7.3 Economic versus Statistical Significance

Not surprisingly, dollar amounts matter in litigation. This fact makes the distinction between *economically* significant effects and *statistically* significant effects meaningful. Since econometric analysis is conducted by selecting a sample of data from a population, and statistical precision increases with sample size, the availability of very large proprietary transaction-level datasets means that it is easy to find a *statistically* significant effect. The large amount of data will allow any small price difference between groups of proposed class members to be detected.[53] As a result, after finding a statistically significant effect, the expert should ascertain whether or not that effect is also *economically significant*. To be significant, the effect is usually sizeable in dollar terms. Economically small effects are of no importance and should be ignored by experts and the court, even if they are statistically significant (Deirdre N. McCloskey & Stephen T. Ziliack, 2008, pp. 39-55).

7.3.1 Individual Damages and Prediction Errors

Given the increased emphasis in rulings such as *Hydrogen Peroxide* on proving impact for all or virtually all class members, experts have often focused on what a class-wide damages model implies about individual class member damages. Although such models do contain useful information for individuals, they are primarily designed to estimate aggregate class-wide damages. However, the application of these models to compute individual damages faces serious statistical obstacles, namely the errors associated with any point estimates of but-for prices. Mechanically, these but-for predicted prices are subtracted from realized actual prices, weighted by all the quantities purchased by a particular class member to compute that individual class member's damages.[54]

7.4 Misinterpreting Prediction Error

In applying PE's class-wide damage models to individual transactions and/or to specific class members, DEs can often find transactions or class members with "negative damages". They then infer that the PE's model is unreliable. DEs also often seek to interpret these findings to mean certain class members were in fact not injured by the conspiracy and that the conspiracy therefore did not have a common impact across the class. Typically,

[52] *In Re*: *High-tech Employee Antitrust Litigation*, Order Granting Plaintiffs' Supplemental Motion for Class Certification, filed October 24, 2013.

[53] With a very large amount of data, some academics even question the usual threshold for assessing statistical significance. *See* George G. Judge, William E. Griffiths, R. Carter Hill, Helmut Lutkepohl & Tsoung-Chao Lee, *The Theory and Practice of Econometrics* (2nd ed.), 1985, p. 131.

[54] This process is described in the academic literature. See for example, James A. Brander and Thomas W. Ross, *Estimating Damages from Price-Fixing*, Litigating Conspiracy: An Analysis of Competition Class Actions, ed. by Stephen Pitel (2006), at 352.

neither conclusion is correct. No workable regression model would be expected to show damages on all transactions and to all class members, and the fact that PE's class-wide model does not do so does not necessarily impugn its reliability. To be sure, only a zero estimate of damages for an individual class member would imply that a particular class member was not affected by the conspiracy. Technically, "negative damages" imply that the model results would tend to suggest (absent an alternative explanation) that the transaction or class member actually *benefited* from the conspiracy. It is rare that there is any basis for concluding — based on the sample data evidence alone — that *any* class member would have benefited from a conspiracy. Instead, such findings, reflect the prediction error inherent in any statistical regression model, and do not support the conclusion that any class members were not impacted by — let alone benefited from — the conspiracy.

All statistical regression models are subject to some degree of prediction error,[55] sourced with the variation between actual prices and "but-for prices" that is not explained by the model.[56] Because any unexplained variation is centered above *and* below the prices predicted by the model and it will not cause a bias in the estimate of overall, class-wide damages.[57] The estimation of "negative damages" for some percentage of class members, can be entirely consistent with an acceptable rate of prediction error at the individual level. There are at least four compelling reasons for this possible explanation.

First, there are often no findings of *zero* damages, only findings of either positive damages or negative damages. As previously noted, it is very rare that there will be something in the record evidence suggesting some class members may have benefited from the conspiracy. Without a factual basis or any economic explanation for a findings of "negative damages", it is unlikely to truly represent class members benefitting from the conspiracy. More likely, a DE's findings of "negative damages" are due to prediction error — a well-accepted statistical fact of all regression models. This means that such findings would not undermine the maintained hypotheses of liability and common impact which should serve as PE's foundation for the class-wide damages model.

Second, often the class members for which DE finds "negative damages" represent only a small portion of

[55] *See* Jeffrey M. Wooldridge, *Introductory Econometrics* (4th ed.), South-Western, 2009, [hereinafter *Wooldridge 2009*], p. 208, describes how even though the "best prediction" of a dependent variable (say, price) is determined by the explanatory variables in the model, there is a "prediction error" in using those explanatory variable to predict the actual value of the dependent variable. *See also*, *Greene*, at 81, which describes how prediction error is comprised of sampling error, which decreases as the sample size increases to the full population (i.e., as more data is used) and a constant term that describes the variation in the population. This latter term is persistent, even as the sample size increases towards the population size, so that "no matter how much data we have, we can never predict perfectly."

[56] "If all points lie on the straight line [from the estimated regression equation], each predicted value of price (sometimes called fitted value) will equal the actual price. However, when more than two data points are given, it is possible and even likely that some points associated with predicted values will not lie on the straight line. The mistake, prediction error, or deviation which results is given by $P_i - \hat{P}_i$.... At this point it is necessary to define what makes one estimated equation better than another. Plainly, if one equation leads to smaller prediction errors than another, it is the better method. But what is meant by smaller prediction errors, and what if there is more than one technique to choose from? On the basis of simplicity and statistical advantage, most econometricians choose the best fitting equation by determining parameter estimates which minimize the sum of the squared deviations between the predicted P [price] and the actual P [price]. The sum is taken over all observations on the variables which are available. Least-squares regression [such as that often used in a class-wide damages model], as this technique is called, is by far the most popular form of estimation available. It receives almost universal usage and is therefore the obvious technique for legal applications. In fact, the least-squares technique is appropriate in most cases. It has substantial advantages over most other alternative estimation techniques," Dan L. Rubinfeld and Peter O. Steiner, *Quantitative Methods in Antitrust Litigation*, Law and Contemporary Problems (1984), at 93.

[57] *See Rubinfeld 2011*, at 325: "It is useful to view the cumulative effect of all of these sources of modeling error as being represented by an additional variable, the error term, in the multiple regression model. An important assumption in multiple regression analysis is that the error term and each of the explanatory variables are independent of each other. (If the error term and an explanatory variable are independent, they are not correlated with each other.) To the extent this is true, the expert can estimate the parameters of the model without bias; the magnitude of the error term will affect the precision with which a model parameter is estimated, but will not cause that estimate to be consistently too high or too low."

class revenue. Or equivalently it is class members with a small number of transactions that are typically found to have overall negative damages or to have only negatively damaged transactions. Once again this is consistent with prediction error, not with escaping common impact. No model can precisely estimate each and every price in a market.[58] In describing the use of regression modeling to estimate damages in price-fixing cases, econometricians report that "[i]t is not assumed that the explanatory factors account perfectly for all movements of price, but only that they represent the principal influences."[59] It is unreasonable to expect any statistical model, when used to estimate each individual but-for price on the millions of class member transactions, to find each and every transaction to have a positive overcharge. Prediction error is more likely to overwhelm the actual damages that were incurred for those class members with very few transactions and thus, such statistical results should not shake our confidence in the reliability of a class-wide damage model. Given that there is likely to be an approximately constant probability of any single transaction to be found with a negative overcharge due to prediction error, a class member with very few transactions is much more likely than a class member with many transactions to be assigned negative estimated damages. Attributing these results to prediction error is all the more reasonable because a finding of a negative damages for a class member with a small number of transactions is the opposite of what an economist would expect if these patterns were a result of class members escaping common impact. It would be large, frequent transaction class members who were more likely to have the bargaining power to escape the defendants' collusive scheme, not small ones.

Third, any class-wide damage model based on statistical modeling is faced with a margin of error. When concerned with the estimation of the model's parameters (such as the overcharge), this is often referred to as the "confidence interval".[60] When concerned with the prediction of individual prices, this is often reflected in the "prediction interval".[61] If we have a 95% prediction interval, then by definition 2.5% of the sample will be outside that interval on the downside, with low or negative individual overcharges, and 2.5% will be outside that interval on the upside, with higher individual overcharges. In this application, that margin of error will predominantly reside with the class members that have a small number of transactions.[62] This is unsurprising and is the very nature of prediction error for a given level of confidence. DE's often focus on the tails of the probability distribution for predictions, emphasizing *only* at the lower tail end of the distribution.

Fourth, often there is nothing systematic in a DE's findings of "negative damages", confirming that the results are due to prediction error, rather than a lack of common impact. DEs often contend that a prediction error at the individual transaction level that is larger than the relevant average overcharge for all transactions (or all transactions at a given time) is problematic because one cannot tell whether we are observing just normal random

---

[58] *Greene*, at p. 6: "No model could hope to encompass the myriad essentially random aspects of economic life;" *Fisher*, at p. 278: "Statisticians are used to the idea that regression equations do not generally fit the data perfectly. They understand that although the systematic part of a regression equation involves the most important variables that theory or common sense suggests influence the dependent variable, the effects of minor or particular influences is left to the random error term."

[59] Michael O. Finkelstein and Hans Levenbach, *Regression Estimates of Damages in Price-Fixing Cases*, 46 Law and Contemporary Problems, No. 4 (1983) [hereinafter *Finkelstein & Levenbach*], at 155.

[60] *See Greene*, at 206: a confidence interval of 95% for a model parameter determines the range of values around the estimated parameter within which there is a 95% confidence that the true parameter value resides.

[61] Literature describes how, because of prediction error, there is a "prediction interval" (i.e., a range of predicted prices) within which it can be considered, with a specified degree of confidence, the actual price resides. *See Greene*, pp. 80-81; *Wooldridge 2009*, pp. 208-209; Henry Theil, *Principles of Econometrics*, Wiley, pp. 134-135.

[62] *See Greene*, p. 81: "…the width of a confidence interval (i.e., a **prediction interval**) depends on the distance of the elements of [the explanatory variables] from the center of the data. Intuitively this idea makes sense; the farther the forecasted point is from the center of our experience, the greater is the degree of uncertainty." (emphasis in the original).

prediction error or we are observing something real regarding common impact. In terms of reliably estimating an overcharge effect, all that matters is that the random unexplained variation is not correlated with the explanatory variables.[63] Emphasis on the size of that unexplained variation as the only indicator of the reliability is inconsistent with the economics literature.[64]

In sum, due to the inevitable presence of prediction error, if negative individual damages are found, any resulting inferences concerning common impact should not be made in a vacuum.[65] Instead, such inferences must be validated by corroborating evidence outside of the data itself. In particular, a group of transactions or class members must be identified with distinct economic characteristics, for which credible economic principles would predict that the conspiracy would have no impact and for which the model does not find overcharges. Lacking any economic explanation for the findings of negative damages, based on class member characteristics, market conditions facing certain class members or other explanatory factors, it is necessary to look for an alternative explanation. That alternative explanation is likely prediction error–not a lack of common impact.

7.5 Estimating Individual Damages

Because prediction errors have no bearing on the model's ability to assess common impact and class-wide damages, a DE may naturally use "negative damages" estimates as a gating factor for the apportionment or decomposition of class-wide damages to individual class members. As a result of prediction error, there is likely to be some proportion of class members for whom the model yields overall net negative damages. For instances of actual wrongful conduct and economic harms, such class members are responsible for a small number of transactions and revenue, and as described in the previous section are more likely to be subject to prediction error.

Plaintiffs will seek a class-wide damages award based on their PE's model. To the extent there will be a need to allocate that award among the class members, that model can provide substantial–and sufficient–information that can be used for such an allocation process, even in the presence of prediction error. For example, a class-wide damages award could be apportioned among the individual class members based on applying the average overcharge for the entire class period to the transaction volumes of individual class members, and this is often the approach adopted for allocation. This would reflect a reasonable allocation of a class-wide damages award in light of the compelling evidence that any class members for which the model produces negative damages were not, in fact, uninjured as a result of defendants' conspiracy.[66]

Further, a class-wide damage model is also capable of providing the requisite information to permit a

[63] *Rubinfeld 2011*, p. 314.
[64] *See Fisher*, p. 279 "Associated with the tendency to believe that a model must fit all cases is the tendency to judge models purely on how well they fit the data, with goodness of fit being measured in the grossest sense of R^2. There is a natural view that models are supposed to do nothing other than predict, and therefore a model that superficially appears to predict well must be believed….the danger here is that the laypersons involved will be impressed by poor work to the detriment of better models that do not fit or predict quite so well but are in fact informative about the phenomena being investigated."
[65] "[I]inferences about economic concepts like market power or efficiencies, whether quantitative or qualitative, are never made in a vacuum. Rather, these inferences are necessarily predicated on assumptions that permit estimation of the magnitude of the effects…. Antitrust policy-making, whether conducted by enforcers or courts, invariably takes place under conditions of uncertainty using what we term 'local' information. Information is local for the obvious reason that the record in any investigation or case is necessarily limited in scope. Litigants have neither infinite time nor infinite resources to gather information…. The need to make decisions when information is local has another implication for investigations and litigation: it heightens the importance of using all available evidence, whether quantitative or qualitative, in antitrust decision-making. For this reason we give qualitative and quantitative evidence equal attention below when we discuss identification with respect to the market definition and market power inquiries," Jonathan B. Baker and Timothy Bresnahan, *Economic Evidence in Antitrust: Defining Markets and Measuring Market Power*, Stanford Law and Economics Olin Working Paper No. 328 (2006), p. 4.
[66] *See Finkelstein & Levenbach*, pp. 152-153: "a computation of average damages for a class has been accepted as an appropriate basis for individual damage awards to class members."

class-wide damages award to be allocated among individual class members based on what the model reveals for each individual class member. Under such an approach, both negative and positive prediction errors can be adjusted for the purposes of allocating any class-wide damage award computed based on the model. Such adjustments are appropriate if the evidence shows that negatively damaged class members were not immune from the effects of the conspiracy and were, in fact, damaged. There are methods available to adjust prediction errors, including matching class members with fewer transactions to class members with similar attributes and a sufficient number of transactions in order to counter the prediction errors that are more problematic for class members with fewer transactions. Such a procedure would enable an estimation of individual overcharges (and damages) in a manner that accounts for both positive and negative prediction errors.

8. Conclusion

This paper has documented the rising legal standards required for class-certification analysis. We welcome this increased level of rigor, but we find that it presents challenges to the economic experts who are asked to present complex statistical analysis to courts. Rarely is it possible to prove that an individual customer's prices have been affected by an alleged antitrust violation without first conducting a full analysis of prices in the market in which that customer resides. Economic experts should use sound theory and professionally acceptable empirical methodologies to identify which groups of customers have been affected in common by the allegations that are the subject of the litigation. We have documented unacceptable practice and presented acceptable methodologies. We emphasize the need to form a theory-based hypothesis and then apply rigorous econometric methods to test the hypothesis empirically. Summary statistics, rather than subjective graphical techniques and anecdotal evidence, should be used to facilitate objective analysis. Meanwhile, instances of data mining that would not survive an economics seminar should be dismissed from the courtroom. Data analysis alone is neither necessary or sufficient to draw reliable inferences of liability, common impact or class-wide damages. Due to the non-experimental nature of transaction data, an alignment of the discovery record, economic theory, and the sample data must be established.

References:

Abbot W. F. and Batt J. (1999). *A Handbook of Jury Research*, The American Law Institute.

Amir R. (2005). "Supermodularity and complementarity in economics: An elementary survey", *Southern Economic Journal*, Vol. 71, No. 3, pp. 636-660.

Arthur S. Goldberg (1964). *Econometric Theory*, John Wiley & Sons, Inc. New York.

Asker J. and Bar-Isaac H. (2014). "Raising retailers' profits: On vertical practices and the exclusion of rivals", *The American Economic Review*, Vol. 25, pp. 672-686.

Baker J. and Rubinfeld D. (1999). "Empirical methods in antitrust litigation: Review and critique", *American Law and Economics Review*, Vol. 1, pp. 386-435, at 391.

Baker J. B. and Bresnahan T. (2006). "Economic evidence in antitrust: Defining markets and measuring market power", *Stanford Law and Economics Olin Working Paper*, No. 328.

Black J. (2002). *Oxford Dictionary of Economics* (2nd ed.).

Bogosian v. Gulf Oil Corp (561 F.2d 434).

Brander J. A. and Ross T. W. (2006). "Estimating damages from price-fixing", in: S. Pitel (Ed.), *Litigating Conspiracy: An Analysis of Competition Class Actions*, Irwin Law.

Bret M. Dickey and Daniel L. Rubinfeld (2010-2011). "Antitrust class certification: Towards an economic framework", *NYU Ann. Surv. Am. L.*, Vol. 66, pp. 459-486.

Burtis M. M. and Neher D. V. (2011). "Correlation and regression analysis in antitrust class certification", *Antitrust Law Journal*, Vol. 77, pp. 495-532.
Carl F. Christ and Gordon C. Rausser (1973). "Discussion on Shapiro and Fromm", *American Journal of Agricultural Economics*, Vol. 55, No. 2, pp. 271-279.
Christ C. F. and Rausser G. C. (1973). "Discussion on Shapiro and Fromm", *American Journal of Agricultural Economics*, Vol. 55, No. 2, pp. 271-279.
Coopers & Lybrand v. Livesay (1978) (437 U.S. 463, 469 n.12).
Cremieux P., Simmons I. and Snyder E. A. (2010). "Proof of common impact in antitrust litigation: The value of regression analysis", *George Mason Law Review*, Vol. 17, pp. 939-968.
Davis Peter and Eliana Garcés (2009). *Quantitative Techniques for Competition and Antitrust Analysis*, Princeton University Press, p. 118.
Daniel L. Rubinfeld (2011). *Reference Guide on Multiple Regression* (3rd ed.), Reference Manual on Scientific Evidence, p. 314.
David Hume (1748). *An Enquiry Concerning Human Understanding*.
Deirdre N. McCloskey and Stephen T. Ziliack (2008). "Signifying nothing: Reply to Hoover and Siegler", *Journal of Economic Methodology*, Vol. 15, pp. 39-55.
Dickey B. M. and Rubinfeld D. L. (2010-2011). "Antitrust class certification: Towards an economic framework", *New York University Annual Review of American Law*, Vol. 66, pp. 459-486.
Earl R. Swanson (Dec., 1960). "Are agricultural economists becoming mechanics?", *Journal of Farm Economics*, Vol. 42, No. 5, pp. 1483-1486.
Eisen v. Carlisle & Jacquelin. (1974). 417 U.S. 156, 94 S.Ct. 2140, 40 L.Ed. 732.
Evans D. (2009). "The new consensus on class certification: What it means for the use of economic and statistical evidence in meeting the requirements of rule 23", *Antitrust Chronicle*.
Finkelstein M. O. and Levenbach H. (1983). "Regression estimates of damages in price-fixing cases", *Law and Contemporary Problems*, Vol. 46, No. 4.
Franklin Fisher (1986). "Statisticians, econometricians, and adversary proceedings", *Journal of the American Statistical Association*, Vol. 81, pp. 277-286.
Geoffrey P. Miller (2004). "Review of the merits in class action certification", Working Paper 262 Berkeley Electronic Press Legal Series, available online at: http://law.bepress.com/expresso/eps/262/, p. 2.
Goldberg A. S. (1964). *Econometric Theory*, New York, NY: John Wiley & Sons, Inc.
Greene W. H. (2011). *Econometric Analysis* (7th ed.), Prentice Hall.
Gregory Mankiw (2014). *Principles of Microeconomics* (7th ed.).
Gregory J. Werden and Luke M. Froeb. (1993). "Correlation, causality and all that jazz: The inherent shortcomings of price tests for antitrust market delineation", *Review of Industrial Organization*, pp. 329-353.
Gordon C. Rausser, Leo Simon and Jinhua Zhao. "Rational exaggeration and counter-exaggeration in information aggregation games", *Economic Theory* (forthcoming).
Guido W. Imbens and Jeffrey M. Wooldridge (2009). "Recent developments in the econometrics of program evaluation", *J. ECON. LIT.*, Vol. 47, No. 5, pp. 5-86.
Haltiwanger J. and Cooper R. (1996). "Evidence on macroeconomic complementarities", *The Review of Economics and Statistics*, Vol. 78, No. 1, pp, 78-93.
Hamilton J. D. (1994). *Time Series Analysis*, Chap. 19, Princeton University Press.
Hartman R. S. and Doane M. J. (1987). "The use of hedonic analysis for certification and damage calculation in class action complaints", *Journal of Law, Economics & Organization*, Vol. 3, pp. 351-372.
Harold T. Shapiro (1973). "Is verification possible? The evaluation of large econometric models", *American Journal of Agricultural Economics*, Vol. 55, No. 2, pp. 250-258.
Hausfeld M. D., Rausser G. C., Macartney G. J., Lehmann M. P. and Gosselin S. S. (2014). "Antitrust class proceedings — Then and now", *Research in Law and Economics*, pp. 77-133.
Hay G. A. and Kelley D. (1974). "An empirical survey of pricing fixing conspiracies", *Journal of Law and Economics*, Vol. 17, pp. 13-38.
Imbens G. W. and Wooldridge J. M. (2009). "Recent developments in the econometrics of program evaluation", *Journal of Economic Literature*, Vol. 47, No. 5, pp. 5-86.
In Re: Ethylene Propylene Diene Monomer (EPDM) Antitrust Litigation (2009). 256 F.R.D. 82, 90 (U.S.D.C. Conn.).

In Re: High-tech Employee Antitrust Litigation (2013). Order Granting Plaintiffs' Supplemental Motion for Class Certification, filed October 24, 2013.

In Re: Hydrogen Peroxide Antitrust Litigation. (2008). 552 F.3d 305.

In Re: Plastic Additives Antitrust Litigation (2010). U.S. Dist. LEXIS 90135, *63 (E.D. Pa.).

James F. Speyer, Ronald C. Redcay & Kelly A. Welchans (June 2011). *Class Actions after Wal-Mart, Arnold & Porter*, LLP, Advisory.

James F. Nieberding and Robin Cantor (2007). "Price dispersion and class certification in antitrust cases: An economic analysis", *Journal of Legal Economics*, Vol. 14, pp. 61-84.

James F. Nieberding (2006). "Estimating overcharges in antitrust cases using a reduced-form approach: Methods and issues", *Journal of Applied Economics*, Vol. 9, pp. 361-380.

Jeffrey Wooldridge (2002). *Econometric Analysis of Cross Section and Panel Data*, p. 3.

Johnson J. H. and Leonard G. K. (2007). "Economics and the rigorous analysis of class certification in antitrust cases", *Journal of Competition Law and Economics*, Vol. 3, pp. 341-356.

Johnson J. H. and Leonard G. K. (2008). "In the eye of the beholder: Price structure as junk science in antitrust class certification", *Antitrust*, Vol. 22, No. 3, p. 108.

Johnson J. H. and Leonard G. K. (2011). "Rigorous analysis of class certification comes of age", *Antitrust Law Journal*, Vol. 77, pp. 570-586.

Johnson P. A. (2011). "The economics of common impact in antitrust class certification", *Antitrust Law Journal*, Vol. 77, pp. 533-567.

John H. Johnson & Gregory K. Leonard (2008). "In the eye of the beholder: Price structure as junk science in antitrust class certification", *Antitrust*, Vol. 22, p. 108.

Judge G. G., Griffiths W. E., Hill R. C., Lutkepohl H. and Lee T. C. (1985). *The Theory and Practice of Econometrics* (2nd ed.), Wiley, p. 131.

Kuhn T. (1970). *The Structure of Scientific Revolutions* (2nd ed.), University of Chicago Press.

Mankiw G. (2014). *Principles of Microeconomics* (7th ed.), South-western College Publishers.

Mas-Colell A., Whinston M. D. and Green J. R. (1995). *Microeconomic Theory*, Oxford University Press.

McCloskey D. N. and Ziliack S. T. (2008). "Signifying nothing: Reply to Hoover and Siegler", *Journal of Economic Methodology*, Vol. 15, pp. 39-55.

Miller G. P. (2004). "Review of the merits in class action certification", working paper, Berkeley Electronic Press Legal Series, Vol. 262, available online at: http://law.bepress.com/expresso/eps/262/.

Michael D. Hausfeld, Gordon C. Rausser, Gareth J. Macartney, Michael P. Lehmann and Sathya S. Gosselin (2014). *Antitrust Class Proceedings — Then and Now: Research in Law and Economics*, pp. 77-133.

Michelle M. Burtis and Darwin V. Neher (2011). "Correlation and regression analysis in antitrust class certification", *Antitrust Law Journal*, Vol. 77, pp. 495-532.

Nieberding J. F. (2006). "Estimating overcharges in antitrust cases using a reduced-form approach: Methods and issues", *Journal of Applied Economics*, Vol. 9, pp. 361-380.

Nieberding J. F. and Cantor R. (2007). "Price dispersion and class certification in antitrust cases: An economic analysis", *Journal of Legal Economics*, Vol. 14, pp. 61-84.

Olson S. D. (2009). "Chipping away: The misguided trend toward resolving merit disputes as part of the class certification calculus", *The University of San Francisco Law Review*, Vol. 43, pp. 935-978.

Paul A. Johnson (2011). "The economics of common impact in antitrust class certification", *Antitrust Law Journal*, Vol. 77, pp. 533-567

Pierre Cremieux, Ian Simmons and Edward A. Snyder (2010), "Proof of common impact in antitrust litigation: The value of regression analysis", *George Mason Law Review*, Vol. 17, pp. 939-968.

Popper K. (1935/2005). *The Logic of Scientific Discovery*, Routledge Classics, published in the Taylor & Francis e-Library.

Posner R. A. (2009). *Antitrust Law*, University of Chicago Press.

Rausser G. (1973). "The validity and verification of complex system models", *American Journal of Agricultural Economics*, Vol. 55, No. 2, pp. 271-279.

Rausser G., Simon L. and Zhao J. (2015). "Rational exaggeration and counter-exaggeration in information aggregation games", *Economic Theory*, Manuscript in preparation.

Raymond S. Hartman and Michael J. Doane (1987). "The use of hedonic analysis for certification and damage calculation in class

action complaints", *Journal of Law, Economics & Organization*, Vol. 3, pp. 351-372.
Reed vs. Advocate Health Care 268 F.R.D. 573 (N.D. Ill. 2009).
Ripley R. A. and Glueck M. J. (February 2009). "In re: Hydrogen peroxide antitrust litigation bleaches clean the class certification standard", *The Antitrust Source*.
Richard A. Ripley & Mark J. Glueck (February 2009). "In re hydrogen peroxide antitrust litigation bleaches clean the class certification standard", *The Antitrust Source*.
Richard A. Posner (2009). *Antitrust Law*, University of Chicago Press, p. 93.
Rubinfeld D. L. (2011). "Reference guide on multiple regression", in: *Reference Manual on Scientific Evidence* (3rd ed.), Washington, DC. National Academic Press, pp. 303-358.
Rubinfeld D. L. and Steiner P. O. (1983). "Quantitative methods in antitrust litigation", *Law and Contemporary Problems*, Vol. 46, pp. 69-141.
Ryan Sullivan, Allan Timmerman and Halbert White (2001). "Dangers of data mining: The case of calendar effects in stock returns", *Journal of Econometrics*, Vol. 105, pp. 249-286.
Shapiro H. T. (1973). "Is verification possible? The evaluation of large econometric models", *American Journal of Agricultural Economics*, Vol. 55, No. 2, pp. 250-258.
Speyer J. F., Redcay R. C. and Welchans K. A. (June 2011). "Class actions after Wal-Mart", *Arnold & Porter, LLP*, Advisory.
Steig D. Olson (2009), "Chipping away: The misguided trend toward resolving merits disputes as part of the class certification calculus", *The University of San Francisco Law Review*, Vol. 43, pp. 935-978.
Sullivan R., Timmerman A. and White H. (2001). "Dangers of data mining: The case of calendar effects in stock returns", *Journal of Econometrics*, Vol. 105, pp. 249-286.
Swanson E. R. (December 1960). "Are agricultural economists becoming mechanics?", *Journal of Farm Economics*, Vol. 42, No. 5, pp. 1483-1486.
Theil H. (1971). *Principles of Econometrics*, Wiley.
Varian H. R. (1992). *Microeconomic Analysis* (3rd ed.), W. W. Norton & Company.
Werden G. J. and Froeb L. M. (1993). "Correlation, causality and all that jazz: The inherent shortcomings of price tests for antitrust market delineation", *Review of Industrial Organization*, pp. 329-353.
William H. Greene (2011). *Econometric Analysis* (7th ed.), p. 6.
Wooldridge J. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.
Wooldridge J. M. (2009). *Introductory Econometrics: A Modern Approach* (4th ed.), South-Western Cengage Learning.
Wright C. A. et al. (1976). *Federal Practice and Procedure*, No. 45, p. 485.