Turning *Daubert* on Its Head:
Efforts to Banish Hypothesis Testing in Antitrust Class Actions

BY LAILA HAIDER, JOHN H. JOHNSON, AND GREGORY K. LEONARD

If the confrontation of economic theories with observable phenomena is the objective of empirical research, then hypothesis testing is the primary tool of analysis. To receive empirical verification, all theories must eventually be reduced to a testable hypothesis.[1]

PLAINTIFFS IN RECENT ANTITRUST class actions have sought to exclude expert evidence and opinions offered by defendants' expert economists with the argument that the experts' statistical testing of plaintiffs' proposed methodology does not satisfy the *Daubert* standard for the admissibility of expert testimony. The statistical testing at issue relates to the determination of whether purported effects from plaintiffs' regression models estimated on the proposed class as a whole hold when the models are estimated separately for different subsets or different members of a proposed class. Plaintiffs in *In re Air Cargo Shipping Services Antitrust Litigation*,[2] for example, sought to exclude testimony from defendants' experts related to this form of testing.[3] The recent decision in *In re Processed Egg Products Antitrust Litigation* sheds further light on plaintiffs' criticisms of such testing.[4] Meanwhile, the very recent decision in *Food Lion LLC v. Dean Foods Co.*[5] highlights the importance of such testing at the class certification stage.

Expert economists testifying on behalf of plaintiffs often propose a statistical model for assessing the impact of defendants' conduct on all or virtually all members of the proposed class and to calculate proposed class members' damages. Typically, the statistical model is a regression model that is estimated by pooling together the sales transaction data of proposed class members. Further, plaintiffs claim that the proposed regression model provides a "common" method for the reliable determination of classwide impact and damages. Their assertion that a proposed "common" regression model provides a reliable determination of injury for all or virtually all class members is a testable hypothesis, with the alternative hypothesis being that reliable determination of impact and damages requires not a common but individual regression models for some or all proposed class members.

An economist may inform this inquiry by conducting statistical tests to determine whether the hypothesized common effects from plaintiffs' regression models in fact hold for different subsets of proposed class members or even for individual members of a proposed class. Plaintiffs in recent antitrust class actions have sought to exclude this form of statistical testing.

The particular form of statistical testing that has been challenged relates to the assessment of whether a plaintiff has met the predominance requirement under Federal Rule of Civil Procedure 23(b)(3). For a proposed class to be certified under this rule, it must be the case that "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."[6] Using statistical testing to assess whether the predominance requirement is met is inherently neutral, favoring neither the class action plaintiff nor the defendant, and does not make assumptions either way about the propriety of class treatment. Moreover, conducting statistical testing is consistent with recent court decisions, including Supreme Court rulings that require a rigorous analysis to verify that all prongs of Rule 23 are satisfied before certifying a class.[7] As a result, courts have relied upon the statistical testing we describe at the class certification stage to determine whether the Rule 23(b)(3) condition is met.[8]

Laila Haider is a Partner at Edgeworth Economics LLC and an Associate Editor of Antitrust. John H. Johnson is the President and CEO of Edgeworth Economics, LLC. Gregory K. Leonard is a Partner at Edgeworth Economics LLC. Dr. Haider was the consulting expert for defendants in In re Photochromic Lens Antitrust Litigation; *Dr. Johnson provided expert testimony in* In re Wholesale Grocery Products Antitrust Litigation *and in* In re Class 8 Transmissions Indirect Purchaser Antitrust Litigation; *and Dr. Leonard provided expert testimony in* In re Photochromic Lens Antitrust Litigation. *The authors thank Dan Ding for research assistance. The opinions expressed are those of the authors and do not necessarily reflect the views of the firm or its clients.*

Yet, plaintiffs have sought to exclude statistical testing that can be useful for conducting the predominance inquiry. Plaintiffs' efforts to banish these methods are not only inconsistent with recent court decisions emphasizing the importance of rigorous analysis, but also seek to turn *Daubert* on its head. Statistical testing, such as the type plaintiffs have sought to exclude, is at the heart of the scientific method and is routinely used in scholarly economics empirical research.

Hypothesis Testing Is Consistent with Requirements Placed by *Daubert* on Expert Testimony

The essence of the scientific method is formulating a "hypothesis," identifying observable implications of the hypothesis, and comparing those implications against real world outcomes. A hypothesis is deemed to be false if its implications are inconsistent with actual outcomes. When outcomes are subject to statistical noise, statistical hypothesis testing methods are used by researchers in scientific disciplines to assess whether an apparent inconsistency between a hypothesis's implications and outcomes is due to the falsity of the hypothesis or, instead, can be explained by statistical variation.[9]

Statistical hypothesis testing is a cornerstone of empirical economic research, as it is in other empirical scientific disciplines. Economists regularly formulate hypotheses based on theoretical economic models or economic reasoning and test the implications of the hypotheses against relevant economic data using statistical hypothesis testing methods. Indeed, in refereed papers published in scholarly academic journals, it is rare to find an econometric result that has not been subject to statistical hypothesis testing of some kind.

Daubert and its progeny, in interpreting the requirements of Federal Rule of Evidence 702, require that an expert "[be] as careful as he would be in his regular professional work outside his paid litigation consulting"[10] and "employ[] in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."[11] If applying the same intellectual rigor as she would apply in a scholarly publication, an economist who offers an econometrics opinion in a litigation matter generally would apply various types of statistical hypothesis testing to assess the validity of the opinion offered.

Air Cargo, Eggs, and Plaintiffs' Attempts to Exclude or Challenge Such Testing

In *Air Cargo*, a putative class of persons and entities that purchased airfreight shipping services directly from defendant airlines for shipments both to and from the United States alleged that defendants participated in a price-fixing scheme resulting in customers paying allegedly supracompetitive prices.[12] In response, the defendants' experts tested the plaintiffs' methodology by estimating the plaintiffs' proposed regression model "using only the data for specific subsets of the class" in order to establish whether the plaintiffs' experts' purported results "held true across the class."[13] The plaintiffs

took issue with this particular form of testing and sought to exclude the defendants' experts' testimony on the grounds that "running subsets of data 'with no economic rationale,' instead of forming and testing a hypothesis" does not follow "the scientific method."[14] The plaintiffs also argued that it is inappropriate to apply their proposed "globally-specified" regression models to "specific or local subsets of data."[15] These arguments are wrong from a scientific point of view.

The recent decision in *Eggs* sheds further light on plaintiffs' criticisms of this form of statistical testing and highlights plaintiffs' misconceptions about these methods. In *Eggs*, a putative class of direct purchasers of eggs or egg products alleged that the defendants conspired to control and limit the supply of eggs, which resulted in their paying higher prices for these products.[16] The defendants' expert conducted statistical testing of the plaintiffs' proposed regression model to determine whether the plaintiffs' claim of classwide impact held across customers of different defendant suppliers. The plaintiffs challenged the testing on the grounds that it constitutes "inappropriate 'data mining,' which 'involves applying a model to arbitrary subsets of the transactional data without an economic theory for selecting such subsets."[17]

Though the district court in *Air Cargo* was not persuaded by the results of the defendants' experts' testing and certified the proposed class, the district court did not exclude the defendants' experts' testimony on such testing. The magistrate judge in *Air Cargo* recommended that doing so would be "draconian" and the court "need not resort to a gag order" on the defendants' experts.[18]

Similarly, the district court in *Eggs* did not find the results of the defendants' expert's testing "sufficiently convincing to derail class certification at this time" and certified the proposed class in part.[19]

In several other antitrust class actions (including *Food Lion*), however, the results of similar statistical testing have derailed class certification on the basis that the Rule 23(b)(3) requirement was not met. Accordingly, it is instructive to explain the relevance of such testing for class certification and to clarify plaintiffs' attempts to turn *Daubert* on its head by claiming that testing violates the scientific method.

Hypothesis Testing for the Assessment of Classwide Impact to Direct Purchasers

Use of Regression Analysis in Antitrust Class Actions. For any given customer, direct information on the actual price paid typically is available. The critical question for assessing whether the customer was injured (or impacted) is the price the customer would have paid but for the defendants' conduct. Thus, an assessment of the impact of this conduct necessarily requires the plaintiff to construct the hypothetical "counterfactual" world that would have existed absent the allegedly anticompetitive conduct. The difference between the counterfactual "but-for" price and the actual price paid by the customer is the "overcharge" paid by the customer (if any). A positive overcharge indicates that the

customer was injured by paying an artificially high price. The size of the overcharge indicates how much the price was inflated, and this would be offered as a measure of the amount of damages suffered by the customer.

Regression analysis is a statistical methodology that is frequently used in antitrust cases to estimate but-for prices and thereby to evaluate injury and damages suffered by class members resulting from the alleged anticompetitive conduct. Regression analysis, under certain conditions, can identify and measure the effects of a set of economic factors on an economic outcome. In the context of an antitrust class action, regression analysis may be used to determine whether the alleged conduct resulted in a customer (or customers) paying an overcharge and to estimate the magnitude of the overcharge (i.e., damages).

But regression analysis can yield an unreliable and misleading result if, for example, the regression model is based upon assumptions that are inconsistent with the underlying economic workings of the marketplace in question. Thus, when plaintiffs assert that their proposed regression model provides a "common" method that can reliably be used at trial to assess classwide impact and damages, this claim should be tested against economic data on actual market outcomes using statistical hypothesis testing.

A "Common" Regression Model Versus Individual Regression Models. Consider a regression model where the dependent variable is the price paid for a product by a given customer to a given supplier in a given transaction. The explanatory variables included in the model should account for factors that economic principles suggest might affect the price the customer paid to the supplier, such as the cost of manufacturing the product, characteristics of the product, the extent of competition faced by the supplier, demand condi-

tions in the marketplace, and characteristics of the customer. In addition, one or more variables that account for the alleged conduct may be included as explanatory variables.

The regression model described here would apply to the specific customer-supplier combination for which it was designed and might not reliably apply to a different customer-supplier combination. For example, the set of explanatory variables that it is appropriate to include in the regression model may differ across customer-supplier combinations, or the coefficients on the explanatory variables (i.e., their effects) may differ across customer-suppliers. These differences can be a result of the differences in the supply and demand conditions the different customer-supplier pairs faced. For example, for a customer that had alternative sources of supply, it might be important to include explanatory variables that accounted for the competition provided by these alternative sources, while for a customer without alternatives there would be no need for such explanatory variables. Similarly, the coefficient for an explanatory variable representing the alleged conduct may differ across customers (including being zero for some customers) as a result of customers' susceptibility to an overcharge differing based on their individual economic circumstances.

It has become standard in modern empirical economics to recognize and, where possible, account for, possible heterogeneity across economic agents (e.g., customers and suppliers) in their responses to changes in economic factors.[20] For example, the well-known "BLP" approach frequently used in demand analysis allows for the possibility that different consumers have different levels of price sensitivity (including possibly zero sensitivity).[21] Thus, for a defendant's economist to raise the possibility that a regression model may differ across customer-suppliers is consistent with the current state-of-the-art in economics research.

If the set of explanatory variables or the effects of the explanatory variables differ across customer-supplier combinations for a proposed class of customers, in general there is not a single common regression model that applies to all. Put differently, the individual regression model for one proposed class member is not informative about the individual regression models necessary to assess impact for other proposed class members. Instead, individual regression models would need to be developed and run separately for each proposed class member.

Similarly, there may not be a single regression model that applies to all products to the extent supply and demand conditions differ across products purchased by the proposed class. Thus, it may be necessary from an econometric point of view to have separate regression models for each proposed class member-supplier-product combination.

In summary, whether there is a single, common regression model applies to all proposed class members (and products) or individual regression models must be used instead, is a critical empirical question that directly relates to the predominance inquiry that is required by Rule 23(b)(3). Assertions

about the applicability of a single, common regression model and the existence of common classwide effects should not be taken at face value—these assertions represent hypotheses that are amenable to and should be subjected to testing. Further, as discussed below, there are well-accepted statistical tests that can be utilized to do so.

Hypothesis Testing of a Proposed Common Regression Model. The dummy overcharge model frequently is put forward by plaintiffs' experts in antitrust class actions as the regression model to be used to assess impact and damages. The plaintiffs' experts in *Air Cargo* and *Eggs*, as well as in various other class actions, proposed regression models of this form. A typical feature in this model is that it assumes a single set of explanatory factors across all proposed class members, i.e., a single set of supply and demand conditions and a single variable capturing the effect of the alleged conduct. Further, the model is estimated by pooling together sales transactions for all customers. In doing so, the approach assumes that a single set of supply and demand conditions explain prices paid by all customers. Further, it assumes that the supply and demand factors had the *same* or *uniform* effect (i.e., resulting in the same change in prices paid) across all customers. These are testable assumptions and may or may not be valid depending upon the facts of a given case.

Similarly, the standard dummy variable approach assumes that the alleged conduct at issue resulted in the same overcharge for each customer. But use of a single dummy variable coefficient for all customers assumes away the very issue that is at the heart of the inquiry in the class certification phase. That is, the dummy variable approach does not allow for the possibility that some members of the proposed class were not impacted by the alleged conduct; instead, it imposes the assumption that either all members were impacted or none were. The validity of such an assumption should be subjected to hypothesis testing. Otherwise, the dummy variable model could mask substantial underlying variation in responses of different customers to the alleged conduct. Put another way, a finding of a positive overcharge, *on average*, across all proposed class members cannot be taken as proof that all the underlying individual overcharges are also positive.

This particular failure of the dummy overcharge regression model has been widely discussed by antitrust practitioners.[22] The ABA Section of Antitrust Law publication on the application of econometrics to antitrust issues explains the critical assumption underlying this type of model: "The reduced-form pricing equation [with a single dummy variable for all class members] *assumes that a conspiracy has the same effect on every purchaser* and focuses on an average effect, which may hide variation across class members."[23] Accordingly, a number of courts have rejected regression approaches that estimate an average effect of the alleged conduct on members of the proposed class precisely on the basis of this issue.

For example, the court in *Food Lion* recently rejected the plaintiffs' expert's regression approach that calculated an average overcharge across members of the proposed class who were located in various zip codes. The defendants' expert applied the plaintiffs' regression model to different subsets of proposed class members (i.e., to class members "on a zip code by zip code basis") and illustrated "no evidence of injury" for about a quarter of class member purchases.[24] In light of this evidence, the court stated that "[d]espite . . . conclusory claims that his model does not rely on averaging, the Court finds, regardless of the nomenclature used, that the model does in fact assume, largely by the coefficients assigned to competitive variables . . . common impact, and does employ averaging, in the ordinary sense of the word, to find impact within zip codes where data is otherwise insufficient to draw the conclusion of impact or where the available data for that particular zip code shows positive benefits or neutral impact."[25]

Further, the magistrate judge in *In re Photochromic Lens Antitrust Litigation* recommended denial of the proposed indirect purchaser class and rejected the plaintiffs' expert's regression approach that calculated annual average overcharges across customers, on the grounds that the approach was not "a workable methodology to gauge impact."[26] Similarly, in *In re Wholesale Grocery Products Antitrust Litigation*, the court rejected plaintiffs' expert's damages calculation that relied on averages, stating "[t]hat profits may have increased on *average*, does not mean that monopolist profits were extracted from each class member," and concluding that the plaintiffs' expert's tests "cannot establish that prices or upcharges or profits *actually* increased for *each* class member."[27]

The court in *In re Plastics Additives Antitrust Litigation* highlighted the plaintiffs' expert's "admission that his regressions produce only single, industry-wide estimates that do not help determine whether each class member suffered any impact," and went on to point out the "unrefuted evidence" which showed that the single estimates produced by these regressions are "in fact not representative of individual class member experience."[28] Similarly, in *In re Graphics Processing Units Antitrust Litigation*, the court noted that plaintiffs' expert "mysteriously chose to *average* certain products and purchases with one another and then correlate instead of correlating disaggregated data for individual products and particular customers."[29] In essence, this exercise "evaded the very burden that he was supposed to shoulder—i.e., that there is a common methodology to measure impact across individual products and specific direct purchasers."[30]

Indeed, economists have long recognized the existence of variation across economic agents in their responses to economic factors and the need to test for such variations before applying a single regression model to a large group of customers or suppliers. Statistical tests can be employed to determine whether the effects of the explanatory variables, including the effect of the alleged conduct, vary across customers. The most widely used such test is known as a Chow test.[31] According to a leading econometrics textbook, "The Chow test—which is simply an F test—can be used to determine whether a multiple regression function differs across two

groups."[32] Similarly, another leading econometrics textbook explains that "one of the popular methods for testing for differences between two (or more) regressions is the Chow test."[33] If statistical testing reveals as false the hypothesis that the effect of the alleged conduct or the effects of the supply and demand factors included in the regression model are the same across customers in the proposed class, then the a priori assumption of commonality should be rejected.

Claim #1: It is Inappropriate to Apply a Regression Model Specified for All Customers to Subsets of Customers. In recent cases, plaintiffs have asserted that it is inappropriate for defendants' experts to apply purported classwide models specified for all customers to subsets of customers. The basis for this criticism is that such testing results in the inappropriate pairing of a "global-specified" or "market-wide" set of factors with "specific or local subsets of data" for which the model was not designed.[34] This assertion makes no economic sense for several reasons.

First, the hypothesis being posed by plaintiffs at the class certification stage is that the price outcomes for individual class members are sufficiently well explained by the "market-wide" model and that impact and damages for any individual class member can be reliably assessed at trial using that model. If this hypothesis is true, the market-wide model applied to a subset of customers should do a good job explaining the price outcomes for those customers. If, on the other hand, the hypothesis is false, the market-wide model will do a poor job of explaining price outcomes for some subsets of customers. It is the ability (or lack thereof) of the market-wide model to explain the price outcomes for subsets of customers that forms the basis for the statistical test of the hypothesis. Put another way, if it is the case that supply and demand factors have common effects across proposed class members, the statistical testing will show that allowing for individual effects is not necessary.[35] Similarly, if the alleged conduct has a common effect across the members of the proposed class, the statistical testing described here will not provide a basis to doubt that hypothesis.

Second, the claim that a market-wide model cannot be applied to subsets of customers because then it would no longer be a market-wide model is based on circular reasoning that (likely intentionally) forecloses the possibility of scientific testing. But, in that event, the hypothesis is not capable of being verified or falsified based on scientific testing and thus would not meet the definition of what constitutes a scientific hypothesis. *Daubert*, for example, requires that expert opinions be based on falsifiable hypotheses.[36]

Claim #2: Estimating a Regression Model for Subsets of Customers Is "Statistical Trickery" Because of Issues Related to Sample Size. Some commentators have referred to applications of hypothesis testing to determine whether customers can be pooled together in a single regression as "statistical trickery."[37] The stated concern is that when the regression is run separately for subsets of customers, the number of observations in each subset is so small that statistical

tests of the hypothesis of zero overcharge have little statistical power to reject this hypothesis even if the overcharge is actually large. However, this argument fails to recognize that separate regressions for the subsets of customers are necessary (and appropriate) only if the hypothesis of a single model has already been tested and rejected.[38] Once that hypothesis has been rejected, use of a single model (and the pooling of customer experiences) is not scientifically valid even if tests run using separate models for subsets of customers had little statistical power to reject the hypothesis of zero overcharge.

Put differently, this argument is not a valid scientific justification to adopt the single model or the pooling of customer experiences.[39] If anything, it suggests that plaintiffs would be unable to prove their case using regression analysis. As a general matter, there is no guarantee that an empirical economic study will turn out to have strong statistical power against a particular hypothesis of interest.[40]

Claim #3: Estimating a Regression Model for Subsets of Customers Is an Exercise in Data Mining. Another concern that has been expressed is that testing plaintiffs' model on subsets of customers is an exercise in data mining.[41] This characterization is incorrect. It would be data mining to churn through a large number of various possible subsets of the data, defined without any economic basis, and perform a statistical test for each subset, until a subset is found that (nominally) rejects the hypothesis of commonality. However, this procedure is not what we propose, nor our understanding of what has actually been done by defendants' experts in antitrust class actions. Instead, a particular partition of the proposed class is identified ex ante based on economic principles applied to the facts of the industry in question. Partitioning of this kind is exactly the "alternative hypothesis" against which the "null hypothesis" of commonality should be tested, given the inquiry mandated by Rule 26(b)(3). For example, customers might be partitioned by size, geography, supplier, or end use, or each customer may form its own subset. The statistical test is then performed on the identified partition. This is not data-mining because the partitioning is not chosen ex post based on the results of the statistical testing; the partitioning is chosen ex ante and the testing is performed to determine whether the regression model in fact differs across the subsets.

Relevance of Hypothesis Testing in Indirect Purchaser and Consumer Products Class Actions

The hypothesis testing we describe is relevant not only to antitrust class actions involving direct purchasers who make repeat purchases and leave a trail of detailed information about them. Hypothesis testing is equally relevant to antitrust class actions involving indirect purchaser claims or consumer products where less information is known about customers purchasing the products at issue.

Plaintiffs in indirect purchaser class actions typically propose a regression model and claim that it provides a common method for the assessment of impact and damages on a class-

wide basis. However, as indirect purchasers, and in keeping with the requirements of Rule 23(b)(3), their methodology must be capable of showing pass-through of an overcharge upstream to all or virtually all proposed class members.[42]

Similarly, plaintiffs in consumer product cases (involving claims of say, false advertising) typically propose a regression model and assert that it provides a common method for the assessment of impact and damages on a classwide basis. While data availability issues may make it impossible to test the proposed common model against separate models for each individual purchaser, for example, it may still be possible to test the proposed common model against separate models for subgroups of customers based on geographic area, distribution channel, supplier, or demographic characteristics if data exist for such subgroups. If, for example, statistical testing reveals as false the hypothesis that the effect of the alleged conduct or the effects of the explanatory factors included in the regression model are the same across different types of products sold or across different sellers, then the a priori assumption of a common method is invalid.

Conclusion

The hypothesis testing discussed here is consistent with courts' demands for more rigor at the class certification stage and the widespread use of statistical hypothesis testing methods in scholarly economics research. Yet some participants in antitrust class actions seek to banish such testing.

The specific facts of a case will determine whether results from the statistical hypothesis testing we describe will be persuasive to a court's decision with respect to class certification. It is crucial, however, to recognize that plaintiffs' efforts to banish such hypothesis testing altogether seek to turn *Daubert* on its head. Courts should refuse to go down such an anti-scientific path. ■

1   Robert F. Engle, Wald, *Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics*, *in* 2 Zvi Griliches & Michael D. Intriligator, Handbook of Econometrics 775 (1984).

2   *In re* Air Cargo Shipping Services Antitrust Litig., No. 06-md-1775, 2014 U.S. Dist. LEXIS 180914 (E.D.N.Y. Oct. 15, 2014), *report and recommendation adopted*, 2015 U.S. Dist. LEXIS 90402 (E.D.N.Y. July 10, 2015).

3   *See id.* at *261.

4   *In re* Processed Egg Products Antitrust Litig., 312 F.R.D. 171 (E.D. Pa. 2015).

5   Food Lion LLC v. Dean Foods Co., No. 2:07-CV-188, 2016 U.S. Dist. LEXIS 8152 (E.D. Tenn. Jan. 25, 2016).

6   Fed. R. Civ. P. 23(b)(3).

7   *See* Comcast Corp. v. Behrend, 133 S. Ct. 1426 (2013); Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338 (2011); *In re* Hydrogen Peroxide Antitrust Litig., 552 F.3d 305 (3d Cir. 2008). In *Hydrogen Peroxide*, the Third Circuit noted that "[t]he evidence and arguments a district court considers in the class certification decision call for *rigorous analysis*." *Hydrogen Peroxide*, 552 F.3d at 318 (emphasis added). Moreover, in vacating the class certification order, the court of appeals specifically stated that "a party's assurance to the court that it intends or plans to meet the requirements [of Rule 23] is insufficient." *Id.*

8   We discuss several court decisions that have relied on this form of statistical testing.

9   "Significance testing" was formalized and popularized by R.A. Fisher. R.A. Fisher, Statistical Methods for Research Workers (1925). "Hypothesis testing" was developed by Jerzy Neyman and Egon Pearson. Jerzy Neyman & Egon S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, 231 Philosophical Transactions of the Royal Society. Series A, Containing Papers of a Mathematical or Physical Character, 289 (1933). Modern statistical hypothesis testing is essentially a hybrid of these two approaches.

10 Sheehan v. Daily Racing Form, Inc., 104 F.3d 940, 942 (7th Cir. 1997).

11 Kumho Tire Co. v. Carmichael, 526 U.S. 137, 152 (1999).

12 *Air Cargo*, 2014 U.S. Dist. LEXIS 180914, at *97.

13 *Id.* at *140. The magistrate judge referred to this form of testing as "sub-regressions."

14 *Id.* at *142.

15 *Id.* at *141.

16 *Eggs*, 312 F.R.D. 171.

17 *Id.* at 189.

18 *Air Cargo*, 2014 U.S. Dist. LEXIS 180914, at *144.

19 The *Eggs* court certified the shell eggs subclass and not the egg products subclass. *Eggs*, 312 F.R.D. at 204.

20 *See, e.g.*, James J. Heckman & Edward Vytlacil, *Policy-Relevant Treatment Effects*, 91 Am. Econ. Rev. 107 (2001) ("Accounting for individual-level heterogeneity in the response to treatment is a major development in the econometric literature on program evaluation. A substantial body of empirical evidence demonstrates that econometric models fit on individual-level data manifest heterogeneity in treatment effects that is present even after conditioning on observables.") (footnote omitted).

21 Steven Berry, James Levinsohn & Ariel Pakes, *Automobile Prices in Equilibrium*, 63 Econometrica 841 (1994).

22 *See, e.g.*, ABA Section of Antitrust Law, Econometrics: Legal, Practical, and Technical Issues 357 (Lawrence Wu et al. eds., 2d ed. 2014) [hereinafter Econometrics 2014]; Laila Haider & Muneeza Alam, *Sub-Regressions: A Rigorous Test for Antitrust Class*, Law360 (Dec. 5, 2014); John H. Johnson & Gregory K. Leonard, *Economics and the Rigorous Analysis of Class Certification in Antitrust Cases*, 3 J. Comp. L. & Econ. 341 (2007); John H. Johnson & Gregory K. Leonard, *Rigorous Analysis of Class Certification Comes of Age*, 77 Antitrust L.J. 569 (2011); Bret M. Dickey & Daniel L. Rubinfeld, *Antitrust Class Certification: Towards an Economic Framework*, 66 NYU Annual Survey of Am. Law 459 (2011); Pierre-Yves Cremieux, Ian Simmons & Edward A. Snyder, *Proof of Common Impact in Antitrust Litigation: The Value of Regression Analysis*, 17 Geo. Mason L. Rev. 939 (2010); Michelle Burtis & Darwin Neher, *Correlation and Regression Analysis in Antitrust Class Certification*, 77 Antitrust L.J. 495 (2011).

23 ABA Section of Antitrust Law, Econometrics: Legal, Practical, and Technical Issues 222 (2005).

24 *Food Lion*, 2016 U.S. Dist. LEXIS 8152, at *52.

25 *Id.* at *53.

26 *In re* Photochromic Lens Antitrust Litig., Case No. 10-md-2173, 2013 U.S. Dist. LEXIS 186728, at *46 (M.D. Fla. Mar. 12, 2013). Notably, also, the district judge denied class certification of the proposed direct purchaser class partly on the grounds that "Direct Purchasers cannot utilize common proof to demonstrate the crucial element of antitrust impact as to *each* member of the class." *In re* Photochromic Lens Antitrust Litig., Case No. 10-md-2173, 2014 U.S. Dist. LEXIS 46107, at *104–05 (M.D. Fla. Apr. 3, 2014).

[27] *In re* Wholesale Grocery Products Antitrust Litig., Case No. 09-md-2090, 2012 U.S. Dist. LEXIS 103215, at *38 (D. Minn. Jul. 25, 2012).

[28] *In re* Plastics Additives Antitrust Litig., Case No. 03-cv-2038, 2010 U.S. Dist. LEXIS 90135, at *60 (E.D. Pa. Aug. 31, 2010). In reaching its decision to deny class certification, the court further noted that "unrefuted evidence shows that some class members suffered impact while others did not," and that it therefore could not rely on the Plaintiff expert's regression model to "demonstrate impact on a basis common to the class." *Id.* at *72.

[29] *In re* Graphics Processing Units Antitrust Litig., 253 F.R.D. 478. 493 (N.D. Cal. 2008).

[30] *Id.* Notably, when the defendants' expert presented an analysis "correlating disaggregated data for specific products and particular direct purchasers" it was found that "any supposed correlation evaporates." *Id.* at 26.

[31] The Chow test is a special case of the more general F-test.

[32] JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH 449 (4th ed. 2009). *See also* ECONOMETRICS 2014, *supra* note 22, at 358–59. Further, this ABA publication on the application of econometrics to antitrust issues explains that "[o]ther statistical tools, including additional regression specifications, may be used to test whether the average effect represented by a single coefficient from a classwide regression masks widely varying individual effects that require individualized inquiry, or whether it truly reflects common impact. One such approach is to divide the proposed class into narrowly defined subgroups and construct a series of regressions to test the stability of any estimate of average impact." ECONOMETRICS, *supra*, at 357.

[33] DAMODAR N. GUJARATI, BASIC ECONOMETRICS 443 (2d ed. 1988).

[34] *Air Cargo*, 2014 U.S. Dist. LEXIS 180914, at *229.

[35] Apart from the chance of a "false positive" finding, which is small by design if the common effects hypothesis is true. For example, the common statistical practice of using a 5% significance level means rejecting a hypotheses (such as the "common" regression model hypothesis) only if the evidence against the hypothesis is strong.

[36] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993) ("Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry.").

[37] Kenneth Flamm & Michael Naaman, *Sub-regressions in Antitrust Class Cert Can Be Unreliable*, LAW360 (Dec. 17, 2014).

[38] Clearly that test had sufficient statistical power to reject the false hypothesis of a single model.

[39] A common misconception is that applying a statistical estimation technique (such as regression) to a small sample is "unreliable." The source of the misconception is a confusion between "reliability" and "statistical precision." The "reliability" of the application of an otherwise appropriate statistical technique does not depend on the size of the sample, but the "statistical precision" of the resulting estimate may. That is, an estimate derived from a small sample typically will be less statistically precise than the estimate derived from a larger sample when all else is equal. However, the statistical precision of an estimate can be measured and thus assessed. Moreover, increasing the sample size by pooling together sales transaction data across customers is a valid approach to increasing statistical precision only if the proposed pooling has been tested using the methods described here.

[40] Of course, in this example, the sample sizes were sufficiently large to have strong power against the hypothesis that a common model applies to all subsets.

[41] *Eggs*, 312 F.R.D. at 189.

[42] *See e.g., Photochromic Lens*, 2013 U.S. Dist. LEXIS 186728, at *46; *In re* Class 8 Transmissions Indirect Purchaser Antitrust Litig., Case No. 11-00009, 2015 U.S. Dist. LEXIS 142717 (D. Del. Oct. 21, 2015).