Evaluating the Alleged Misconceptions about the In-Sample Prediction Approach to Assessing Common Impact

BY DANIEL RAMSEY AND CELESTE SARAVIA

A KEY QUESTION IN CLASS CERTIFICATION decisions in the United States is whether class-wide evidence can be used to determine whether the challenged conduct resulted in harm to all or almost all proposed class members.[1] One popular approach, which the plaintiffs' experts sometimes call the "in-sample prediction approach," claims to estimate the impact of the challenged conduct on *each* transaction. In a price fixing class action, the in-sample prediction approach purports to provide an estimate of the overcharge on each transaction, which we refer to as a transaction-specific overcharge.[2] The transaction-specific overcharge estimates could be, for example, $3.2 for the first transaction, $1.1 for the second transaction, and so on for every transaction. The expert would then use these estimates of the transaction-specific overcharges to determine the percentage of class members who were harmed on at least one transaction. In every class action we identified in which a plaintiff's expert used this approach, the expert found that almost all class members were harmed, and the judge certified the class (see Table 1).

In a recent article in this magazine, McClave and McClave argue that there are "some common misconceptions" regarding this approach, which they refer to as a "two-step econometric method."[3] In their article, they attempt to "provide clarifying details" about the approach and "address three of these misconceptions."[4] Similar to other plaintiffs' experts who have used the approach, McClave and McClave argue that the approach can estimate "individual transaction overcharges," and thus the approach can determine whether each "transaction is … impacted by the conduct," and "the percentage of class members impacted on at least one transaction."[5]

However, their article does not demonstrate that the approach is a reliable method of making these determinations.

Daniel Ramsey is a Principal at Cornerstone Research and Celeste Saravia is a Vice President at Cornerstone Research. The opinions expressed are those of the authors and do not necessarily represent the views of Cornerstone Research.

Table 1. Cases in which a plaintiff's expert used the in-sample prediction approach

#	Case	Class Certified	Class Cert Decision Date	Percent of Class Members Found Harmed by Plaintiff Expert
1	In re Broiler Chicken Grower Antitrust Litig. (E.D. Okla)	Yes	2024	95–100
2	Le v. Zuffia, LLC (D. Nev.)	Yes	2023	96.3–98.8
3	In re Pork Antitrust Litig. (D. Minn.)	Yes	2023	99
4	In re Broiler Chicken Antitrust Litig. (N.D. Ill.)	Yes	2022	93
5	In re Packaged Seafood Prods. Antitrust Litig. (S.D. Cal.)	Yes	2019	94.5
6	In re Capacitors Antitrust Litig. (N.D. Cal.)	Yes	2018	98
7	In re Domestic Drywall Antitrust Litig. (E.D. Pa.)	Yes	2017	98
8	In re Korean Ramen Antitrust Litig. (N.D. Cal.)	Yes	2017	98
9	In re Air Cargo Shipping Servs. Antitrust Litig. (E.D.N.Y.)	Yes	2014	96
10	In re Chocolate Confectionery Antitrust Litig. (M.D. Pa.)	Yes	2012	98

Source: (1) Mem. Dec. & Order Granting Pls.' Mot. for Class Cert. & Denying Def.'s Mot. to Exclude at 40, *In re* Broiler Chicken Grower Antitrust Litig. (No. II), No. 6:20-md-02977-RJS-CMR (E.D. Okla. May 8, 2024); (2) Order, Le v. Zuffa, LLC, No. 2:15cv01045RFBBNW (D. Nev. Aug. 9, 2023); Expert Report of Hal J. Singer, Ph.D. at Table 8, Le v. Zuffa, LLC, No. 2:15cv01045RFBBNW (D. Nev. Aug. 27, 2019); (3) Mem. Op. & Order Granting Class Pls.' Mots. for Class Cert. & Denying Defts.' Mots. to Exclude Expert Test. at 56, *In re* Pork Antitrust Litig., 665 F. Supp. 3d 967 (D. Minn. 2023); Singer Report at 132–133; (4) Mem. Op. & Order at 32, *In re* Broiler Chicken Antitrust Litig., No. 1:16-cv-08637 (N.D. Ill. May 27, 2022); (5) Order Granting Mots. for Class Cert. at 16, *In re* Packaged Seafood Prods. Antitrust Litig., 332 F.R.D. 308 (S.D. Cal. 2019); Olean Wholesale Grocery Cooperative, Inc. v. Bumble Bee Foods LLC, 31 F.4th 651 (9th Cir. 2022) (en banc); (6) Order re DPPs' Class Cert. Mot. & Defs.' Daubert Mots. to Exclude Expert Ops. at 11, *In re* Capacitors Antitrust Litig. (No. III), No. 17md02801JD (N.D. Cal. Nov. 14, 2018); Expert Report of James T. McClave, Ph.D. at 9, *In re* Capacitors Antitrust Litig., No. 3:14-cv-03264-JD (N.D. Cal. Feb. 24, 2017); (7) Mem. re DPPs' Mot. for Class Cert. at 52, *In re* Domestic Drywall Antitrust Litig., No. 13-MD-2437 (E.D. Pa. Aug. 23, 2017); (8) Order Granting Mots. for Class Cert. at 10, *In re* Korean Ramen Antitrust Litig., No. 13-cv-04115-WHO (N.D. Cal. Jan. 19, 2017); (9) R. & R. at 19, *In re* Air Cargo Shipping Servs. Antitrust Litig., No. 06-md-1775 JG (VVP) (E.D.N.Y. Oct. 15, 2014); (10) Mem. at 42–43, *In re* Chocolate Confectionery Antitrust Litig., No. 08-md-1935 (M.D. Pa. Dec. 7, 2012)

In this article, we find that the in-sample approach is unreliable because it fails two fundamental tests of reliable econometric methods.[6] First, it is not a consistent estimator, meaning that its estimate will not converge to the true value of the overcharge as the data grows large. For example, if the true overcharge on a transaction was $1, the approach's estimate could be quite far from $1 even with a very large number of observations. Second, it has a false positive error rate far beyond typical thresholds, meaning that it will find that a substantial percentage of class members were harmed even if they were not harmed. Additionally, the approach is contradicted by a large academic literature. We also find that the three alleged misconceptions identified by McClave and McClave are irrelevant to our analysis of whether the in-sample approach is reliable.

Overview of the In-Sample Approach and An Illustrative Example

Overview of the approach. The in-sample approach purports to assess which proposed class members were harmed (i.e., suffered impact) in two steps.[7] In the first step, the plaintiff's expert seeks to determine the *average* class-wide impact, if any, of the challenged conduct. For example, in a price fixing case, the expert seeks to estimate the average overcharge paid by class members. The expert might purport to find that, for example, the challenged conduct caused the prices paid by class members to be $2.0 higher on average.

The expert typically performs this first step by using transaction data to estimate a regression equation. The dependent variable is a measure of price, and the explanatory variables include multiple control variables that can explain variation in prices across transactions and an indicator variable that indicates whether the transaction occurred during the alleged conspiracy.[8] The plaintiff's expert's claim that this regression provides an estimate of the average overcharge is only valid if several assumptions hold. However, even if the expert finds that the conduct had an impact on average prices, this does not imply that all class members are harmed, nor does it show which class members are harmed.

In the second step, the plaintiff's expert uses the first-step regression results to calculate which transactions were affected and which class members were harmed. The following equation provides the purported transaction-specific overcharges:[9]

(1) Transaction-Specific Overcharge = Average Overcharge + Residual

For example, if the first-step regression estimates an average overcharge of $2 across all transactions, and the first transaction has a residual of $1 from that regression, then the technique would find that the overcharge on the first transaction is $3. The residual is the portion of the outcome that is not explained by the model. For example, if a transaction's price was $12, but the first-step regression model predicted it to be $11, then the residual is $1. A plaintiff's expert using this approach typically does not require the transaction-specific overcharge to meet any statistical threshold. The plaintiff's expert typically concludes that a class member is harmed if the class member was harmed on *at least one transaction*, rather than requiring the class member to be harmed on *net* across all of their transactions.[10]

Stylized example that illustrates the approach's fundamental flaws. We illustrate the approach with a simple example in which there are four transactions. These data are constructed such that the conduct has *no impact* on the average price or on the price of any transaction.[11] Figure 1 depicts these four transactions. Two of the transactions are subject to the challenged conduct (the "treatment group," which are depicted on the right side of the figure), and these have prices of $12 and $10. The other two transactions are not subject to the challenged conduct (the "control group," which are depicted on the left side of the figure), perhaps because they occurred in a different region or year, and these also have prices of $12 and $10.

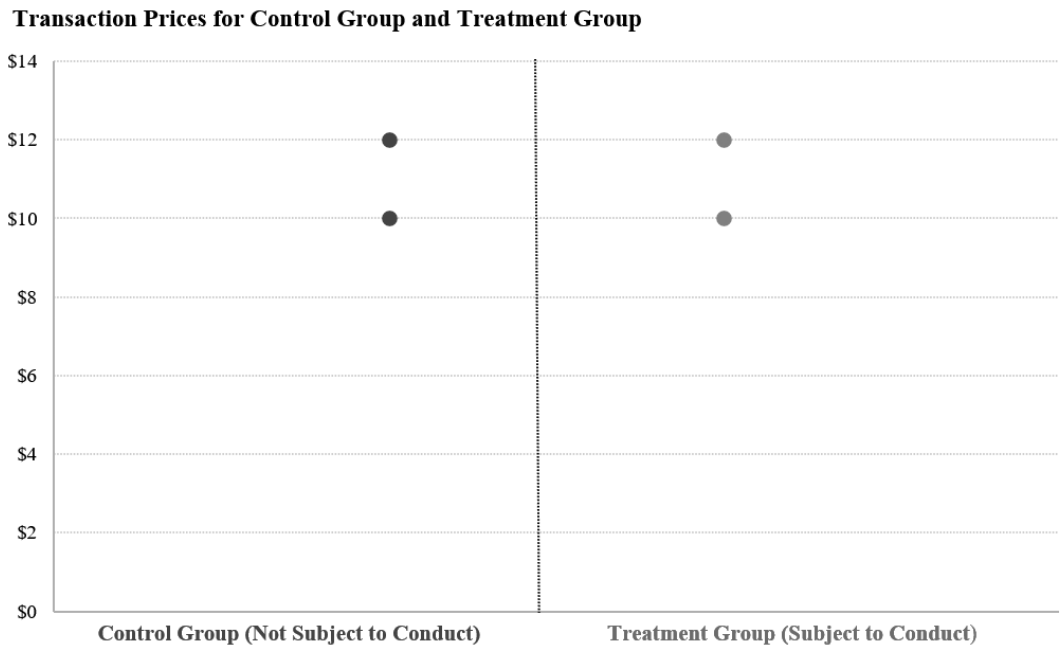
The in-sample approach applied to these data would correctly determine that the average impact was zero, but it would incorrectly determine that one of the two transactions subject to the challenged conduct was harmed. In particular, the approach would find that the $12 transaction would have been $11 but for the challenged conduct, meaning it had an overcharge of $1. And the approach would find that the $10 transaction benefited from the challenged conduct by $1. Thus, even though the prices are *identical* in the control group and treatment group, the approach somehow finds that both treatment group transactions were impacted by the conduct (one harmed and one benefitted).[12] It obtains these results because the residual is $1 for the $12 transaction, and the residual is -$1 for the $10 transaction (which it adds to the average impact of 0).

This example illustrates two fundamental flaws of the in-sample approach. First, it does not correctly identify the overcharge for any transaction. Here, the true impact was zero for each transaction, but the approach found an overcharge of $1 on one transaction and -$1 on the other transaction. Second, the approach has a substantial false positive rate, meaning that it finds a high percentage of class members harmed even if they were not harmed (50 percent in this example). In the next section, we explain that these flaws are not limited to this example but are fundamental properties of the approach.

Evaluation of the Reliability of the In-Sample Approach

Evaluation of Whether the Approach is a Consistent Estimator. We assess whether the approach meets a minimal requirement of a reliable econometric approach: whether it is a *consistent* estimator of the transaction-specific overcharges.[13] A consistent estimator will converge to the true value as the size of the data increases. For example, if the true overcharge on a given transaction is $1.0, will the approach's estimate be close to $1.0 with sufficiently large data? If it is

Figure 1. Example Transaction Prices



Note: This figure depicts the four transactions in the example. Of the two transactions in the treatment group, the in-sample approach would find the $12 transaction to have a $1 overcharge and the $10 transaction to have a $1 undercharge.

not a consistent estimator, the approach could provide an estimate that is substantially different from $1.0 even as the size of the data increases to infinity, and thus, the approach would not reliably determine the harm incurred by individual transactions or class members.

The in-sample approach is *not* a consistent estimator. As shown in Equation (1) above, the approach obtains the transaction-specific overcharge by adding the residual to the average overcharge. The approach provides no basis for assuming the residual is entirely due to the conduct, and such an assumption is contradicted by econometric principles. Regression residuals are the portion of the outcome (price in this example) that is unexplained by the model, and are determined by *numerous* factors, such as relevant variables omitted from the model, data error (for example, a variable is measured with error), or random noise. Essentially every econometric model will have regression residuals since it is typically not possible to perfectly predict every transaction's price.[14] Since the approach includes the residual in its estimate of the transaction-specific overcharge, the overcharge estimate errantly includes factors that are unrelated to the overcharge, such as omitted variables or data error. For example, assume that in our example above there were omitted variables, such as whether the customer negotiated the price and the customer's negotiating skill, that explain some of the variation in prices. The in-sample approach would falsely attribute the impact of any omitted variables to the conduct. This error will persist no matter how many observations are used in the analysis, and thus, the approach is not a consistent estimator.[15]

Evaluation of the False Positive Rate of the Approach. Next, we assess whether the in-sample approach has a false positive rate within the typical thresholds. False positives are instances when the true impact of the conduct is zero, but the approach finds an impact.[16] Common econometric and statistical approaches typically have false positive error rates of 5 percent or less.[17] If an approach has a high false positive rate, it is typically not a reliable approach.[18]

The in-sample approach has a false positive rate well above 5 percent, and it can approach 100 percent. Equation (1) above shows why the in-sample approach has a high false positive rate. Consider an example in which the challenged conduct had no impact on any transaction, and thus, the average impact is also zero. As the number of transactions increases, the estimated average impact will approach zero and the approach's estimated impact on a given transaction is just the regression residual. In any regression, the regression residuals sum to zero, and thus there are both positive and negative residuals.[19] Thus, since as discussed above, the approach typically does not require the estimated transaction-specific overcharge to meet any threshold of statistical significance, if the residuals are fairly evenly split between positive and negative values, the approach would incorrectly conclude that roughly half of transactions are harmed (the positive residuals in a price fixing case) even if the conduct had no impact on any transaction.

This false positive issue is exacerbated by the fact that each class member typically has many transactions, sometimes even hundreds of transactions.[20] The probability the approach finds that a class member is harmed on at least one transaction

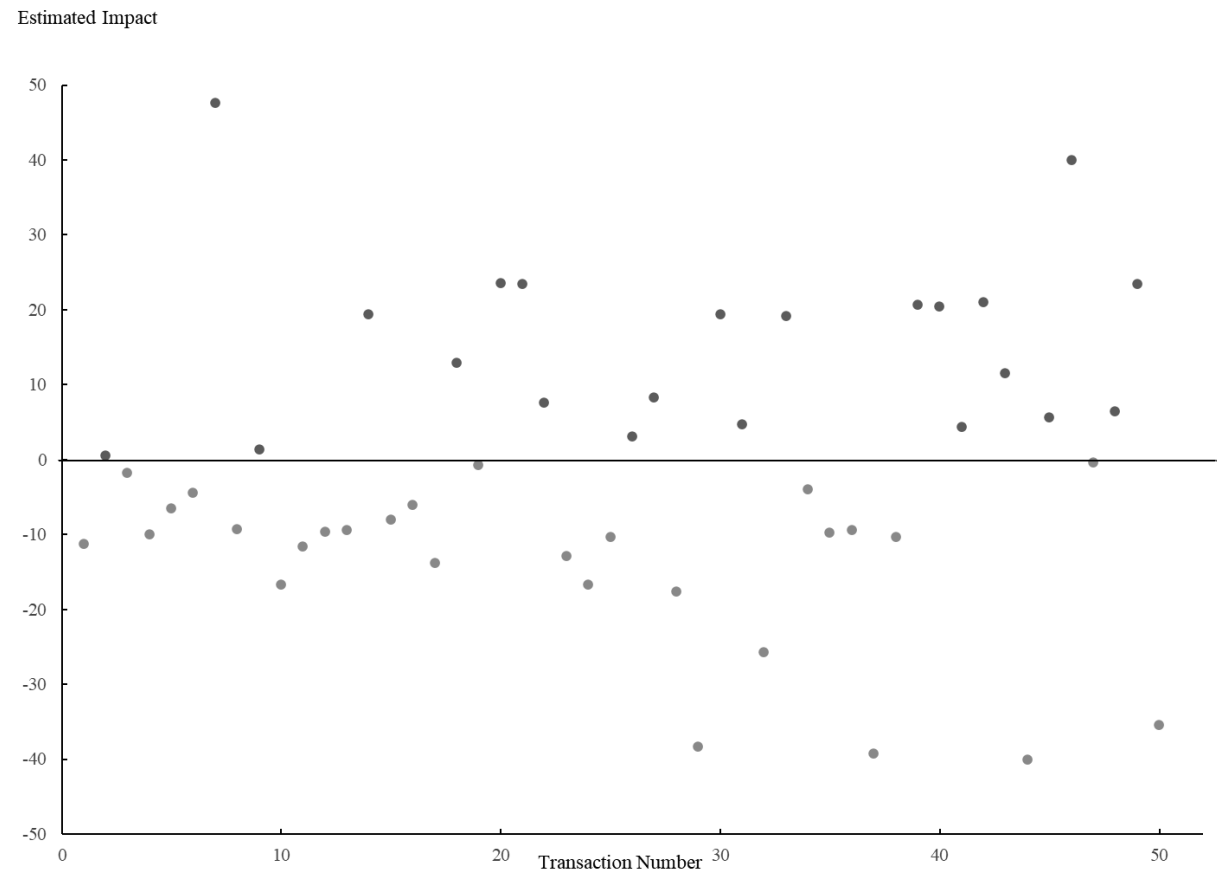
increases with the number of transactions of that class member. If the true impact of the conduct is zero, and a class member has two transactions, the probability the approach finds that the class member is harmed on at least one transaction is 75 percent.[21] If the true impact of the conduct is zero, and a class member has five transactions, then the probability the approach finds that the class member is harmed on at least one transaction is 97 percent. Thus, even though the true impact of the challenged conduct is zero in this example, the approach would find that almost all class members were harmed.

To further illustrate the false positive issue, we simulate data where the impact of the conduct on each transaction is assumed to be zero. There are 100 individuals, each of whom makes one transaction. We assume that the price varies across transactions, and the average price is $100.[22] We randomly assign 50 individuals to the group subject to the challenged conduct ("class members") and 50 to the control group not subject to the challenged conduct. Due to the random assignment, the average impact of the challenged conduct should be near 0 and statistically insignificant, which we confirm to be the case.

We then apply the in-sample approach to these data. Figure 2 depicts the impact estimated by the in-sample approach for each transaction subject to the conduct. It shows that roughly half of the estimated transaction-specific overcharges are positive (the dots above the zero line) and roughly half are negative (the dots below the zero line). That is, even though the data are constructed such that there is no impact on any transaction, the approach finds that almost half (22 of 50) of the transactions are harmed.

To illustrate how the false positive rate increases with the number of transactions per class member, we extend the prior simulation such that each individual has four additional transactions subject to the same assumptions.[23] The approach finds that 44 percent of the class members are harmed on their first transaction (which is shown in row one of Figure 3), about 70 percent are harmed on at least one of their first two transactions (row two), and 100 percent are harmed on at least one of their five transactions (row five). In summary, this simulation shows that the in-sample approach has a false positive rate that is much greater than 5 percent and can be as high as 100 percent. Postal and Terner in this issue perform different simulations and also find that the false positive rate can approach 100 percent if the number of transactions per class member is sufficiently large.

Figure 2. Estimated transaction-specific impacts when the true impact is zero



Note: The dots are the transaction-level impacts estimated by the in-sample approach. The approach would find all dots above zero to be harmed if it is a price fixing case.

Figure 3. Share of class members that the approach finds harmed when the true impact is zero

Transactions Per Class Member	Share of Class Members Found Harmed at Least Once
1	44%
2	70%
3	88%
4	98%
5	100%

This analysis studies the key metric typically used by the in-sample approach: the percent of class members overcharged on at least one transaction. McClave and McClave also put forth a second metric, the percent of class members with a net average overcharge across their transactions.[24] This would involve calculating the transaction-specific overcharges as described above, and then averaging them for each class member. However, this alternate metric has similar flaws. For example, if there was no overcharge on any transaction, and the residuals are symmetrically distributed, then roughly half of class members would have a positive net average residual, and the approach would incorrectly conclude that roughly half of class members were harmed. Consistent with this logic, the simulation in Figure 3 would find that 48 percent of class members with three transactions were harmed on net.

McClave and McClave might contend that our analysis of the false positive rate is mistaken since McClave and McClave contend that a plaintiff's expert would only perform the second step of the approach if the first step finds a statistically significant average overcharge.[25] However, such a contention would be incorrect. The purpose of the analysis in this section—which performs the second step even though the first step finds zero average overcharge—is not to determine the percent of class members who are truly harmed, but to assess the reliability of the in-sample approach by measuring its false positive rate. This type of analysis is often referred to as a falsification test—which measures a method's false positive rate by applying the method to situations where there is no impact—and such tests are a standard way to assess the reliability of a method.[26]

Evaluation of Whether the Approach is Consistent with the Academic Literature. There is a large academic literature on causal inference, which discusses various approaches to estimating the impact of policies or events, such as alleged collusion. This literature concludes that it is *not* generally possible to do what the in-sample approach claims to do, which is to estimate *individual* causal effects such as transaction-specific overcharges.[27] For that reason, the academic literature typically estimates *average* effects, which in the case of a price fixing matter would be the average overcharge, rather than transaction-specific overcharges. Moreover, while there is a recent academic literature that attempts to go beyond estimating average effects, it does not propose a method like the in-sample approach.[28]

Evaluation of McClave and McClave's Three Alleged Misconceptions

The three alleged misconceptions identified by McClave and McClave are irrelevant to our analysis of whether the in-sample approach is reliable. Putting aside the merits of the alleged misconceptions, they do not demonstrate that the approach is reliable or undermine our reasons for finding that it is unreliable.

Misconception One. McClave and McClave argue that some defendants' experts incorrectly claim that the in-sample approach assumes that "all class members had the same average overcharge" and thus assumes that all class members are harmed.[29] We agree that the approach does not assume each transaction has the same overcharge. Instead, the approach makes a fundamentally flawed assumption about the transaction-specific overcharges.

As discussed above, the approach obtains the transaction-specific overcharge by adding the residual to the average overcharge. The approach provides no basis for assuming the residual is entirely due to the conduct, and such an assumption is contradicted by econometric principles. As discussed above, since the approach includes the residual in its estimate of the transaction-specific overcharge, the overcharge estimate errantly includes factors that are unrelated to the overcharge, such as omitted variables or data error. This error will persist no matter how many observations are used in the analysis, and thus, the approach is not a consistent estimator.

Misconception Two. McClave and McClave claim that some defendants' experts incorrectly try to rebut the in-sample approach by using a different econometric approach that attempts to "estimate a separate overcharge for each" class member.[30] The defendants' experts then claim that a relatively large percent of class members did not have statistically significant overcharges, which contradicts the in-sample approach's finding of class-wide harm. McClave and McClave claim this alternate approach is unreliable because it will often "have little [statistical] power and [is] likely to produce false negatives" especially for "small class members," and thus it does not undermine the in-sample approach.[31] However, this argument is flawed and does not establish the reliability of the in-sample approach.

First, McClave and McClave's statement that some defendants' experts have put forward an analysis that McClave and McClave view as flawed or unreliable (which we do not assess) does not imply that the in-sample approach is reliable. Even if one approach is unreliable, it does not imply that a different and unrelated approach is reliable.

Second, even if McClave and McClave were correct that some defendant experts have "estimate[d] a separate overcharge for each" class member with "little [statistical power]," this would not imply that estimating separate overcharges for different class members or groups of class members is necessarily unreliable. To the contrary, it is standard in econometrics to estimate different coefficients (such as

overcharges) for different groups of individuals for which the effect is hypothesized to differ, and there are standard approaches for doing so.[32] For example, if there is a reason to expect that class members with long-term contracts were differentially impacted than those who purchased without contracts, then standard approaches could be used to estimate different overcharge coefficients for the two groups. Moreover, McClave and McClave appear to recognize that it is reasonable and appropriate to statistically test whether different overcharge coefficients should be estimated for different class members or groups of class members.[33]

In conclusion, while the in-sample approach is contradicted by the academic literature, there is academic literature that establishes reliable approaches to estimating different parameters (such as overcharges) for different groups.

Misconception Three. McClave and McClave claim that some defendants' experts incorrectly conclude that class members are unharmed if the in-sample approach finds them to be unharmed. McClave and McClave claim that in certain instances, one should overrule the in-sample approach's finding that certain class members are *unharmed* and conclude that they were actually *harmed*.[34] In particular, they claim that if the approach identifies "only a relatively few small customers … as uninjured," but finds "extensive impact across the remainder of the class," then there is "a basis to infer that these small customers were also impacted."[35] However, this claim is flawed and does not establish the reliability of the approach.

First, McClave and McClave do not demonstrate that these ad hoc adjustments increase the reliability of the approach, and in fact, such adjustments can make the approach even less reliable. Consider the example in Figure 3 above in which each class member has three transactions, and the in-sample approach found that 88 percent of class members were harmed (even though none of them were harmed in reality). If an expert were to overrule the approach's finding that 12 percent were not harmed and instead claim that some or all of those 12 percent were harmed, this would further increase the false positive rate (from 88 percent to as much as 100 percent). As is clear from this example, the ad hoc adjustment does not identify and correct the false positives found by the approach.

Second, the fact that the approach requires this ad hoc manual adjustment is further evidence of the approach's unreliability. That is, if an expert can overrule the in-sample approach's finding for some class members in an ad hoc fashion, this casts further doubt about the reliability of the in-sample approach's finding for the other class members.

Third, McClave and McClave provide insufficient details regarding how the ad hoc adjustment should be implemented and the basis for those adjustment decisions. For example, if the in-sample approach found that 88 percent of class members were harmed and 12 percent were not harmed, is that sufficient to make an ad hoc determination that all of the 12 percent were actually harmed? McClave and McClave do not provide a methodology for making this determination.

Conclusion

The plaintiffs' experts have used the in-sample approach in more than ten recent antitrust class actions to argue that all or almost all proposed class members were harmed by the challenged conduct. In each case where the court issued a class certification decision, the court certified the class and, at least in part, credited the plaintiffs' expert's in-sample approach. In a recent article in this magazine, McClave and McClave claim that previous critiques of the approach's reliability rest on three misconceptions about the approach. In this article, we explain that the approach is unreliable because it fails two fundamental tests of reliable econometric methods and is contradicted by the academic literature. We also discuss the alleged misconceptions and show that our critiques of the approach are unrelated to the alleged misconceptions. We hope that this article spurs further research into the in-sample approach and other empirical approaches to assess whether all or almost all class members were harmed. ■

[1] Different U.S. circuit courts have taken different positions on the meaning of "all or almost all" class members. For example, some circuits require there to be less than a "de minimis" number of unharmed class members, and others allow more than a de minimis number. On June 5, 2025, the Supreme Court dismissed a writ of certiorari related to this question as improvidently granted. *Lab'y Corp. of Am. Holdings v. Davis*, No. 24-304, (U.S. June 5, 2025) (Kavanaugh, J., dissenting).

[2] The in-sample prediction approach has been used to analyze different outcomes in different cases, such as prices in a price fixing case or wages in a wage fixing case. We focus on the price fixing example (with harm measured by the price overcharge), but the same critiques would apply to the use of the in-sample method in a wage fixing case. While our article focuses on antitrust class actions, it is possible that the in-sample prediction approach has been used in other types of class actions.

[3] Jamie McClave and James McClave, "Clarifying Common Misconceptions about the Two-Step Econometric Method for Establishing Common Impact," *Antitrust Magazine* 39, no. 2 (2025), ("McClave and McClave"), 63.

[4] McClave and McClave, 63.

[5] McClave and McClave, 63.

[6] We analyze the in-sample approach further in our more technical article; see Daniel Ramsey and Celeste Saravia, "Evaluating the In-Sample Prediction Approach to Assessing Class-Wide Impact for Class Certification," *Journal of Competition Law and Economics* (2026), https://doi.org/10.1093/joclec/nhag020 ("Ramsey and Saravia").

[7] This section is based on publicly available expert reports that use the in-sample approach; see Ramsey and Saravia for further detail.

[8] The plaintiff expert typically estimates an equation like the following: $P_{it} = \beta X_{it} + \delta D_{it} + \varepsilon_{it}$, where P_{it} is a measure of the price paid for transaction i at date t. X_{it} is a vector of control variables for transaction i at date t, and β is the coefficient vector for the control variables (including a regression constant). D_{it} is a conduct indicator variable that equals 1 if transaction i at date t was allegedly subject to the challenged conduct and 0 otherwise. ε_{it} is the error term.

[9] Ramsey and Saravia show that this is an alternative mathematical expression of the in-sample approach. Plaintiff experts typically explain the approach as follows. They state that after estimating the first-step regression, they switch the conduct variable from $D_{it} = 1$ to $D_{it} = 0$, and then form regression predictions, which they call the "but-for" prediction, and then calculate the purported transaction-specific overcharge as the difference between the price and the "but-for" predicted price.

[10] While it is not the focus of this article, the economic basis for concluding that a class member is harmed if the class member was harmed on

at least one transaction (rather than being harmed on net) is unclear. McClave and McClave offer an alternative measure that identifies a class member as harmed if the mean of the class member's transaction-specific overcharges is positive. See McClave and McClave, 63. Later in this article, we show that this metric is also flawed.

[11] See Ramsey and Saravia for an example in which the conduct had an average impact on prices.

[12] In practice, in determining whether a class member was harmed, the approach essentially discards transactions that were found to have benefited from the conduct (since the approach does not typically determine the net harm).

[13] Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach* (Mason, OH: South-Western, Cengage Learning, 5th ed., 2013), ("Wooldridge"), 169 ("…[V]irtually all economists agree that consistency is a minimal requirement for an estimator.").

[14] William H. Greene, *Econometric Analysis* (Essex, England: Pearson Education Limited, 2012), 53.

[15] The in-sample approach only identifies the transaction-specific overcharge if the *entire residual* is due to the heterogeneous impact of the conduct (i.e., how much the overcharge for each transaction varies from the average overcharge). This would require the regression to perfectly predict the data if the impact were not heterogeneous or if the transaction was not subject to the conduct, which is not plausible. See Ramsey and Saravia.

[16] Wooldridge, 779.

[17] Wooldridge, 779.

[18] Wooldridge, 779; Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119, no. 1 (2004).

[19] Wooldridge, 36.

[20] Ramsey and Saravia.

[21] Under the assumptions that the residuals are symmetric and uncorrelated across transactions for each class member, the probability the approach finds that the class member is harmed on at least one transaction is calculated as $1-(1-0.5)^n$, where n is the number of transactions.

[22] In particular, we assume that prices are distributed normally, with a mean of 100 and standard deviation of 20.

[23] We also assume that errors are uncorrelated across transactions for a given individual. It is possible that errors may be correlated across transactions for each individual. However, no case data are publicly available to perform this simulation or estimate the extent of serial correlation.

[24] McClave and McClave, 63 ("The percentage of class members impacted on at least one transaction, and the percentage of class members with positive mean overcharges, can be reported as quantitative measures of the degree to which the class experienced overcharges, that is, as measures of classwide impact.").

[25] McClave and McClave, 63–64 ("An impact assessment [for individual class members] follows only if the estimate of aggregate classwide overcharges is statistically significant.").

[26] See Andrew M. Jones, "Panel Data Methods and Applications to Health Economics," in *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, ed. Terence C. Mills and Kerry Patterson (Hampshire, England: Palgrave Macmillan, 2009), 574 ("One way to assess the robustness of an identification [estimation] strategy is to find an anti-test (or placebo test). Anti-tests provide counter-evidence by applying a model or identification strategy in a context where no effect should be detected. If an apparent 'effect' is found then the validity of the identification strategy must be called into question."); Federal Judicial Center And National Research Council of the National Academies, *Reference Manual On Scientific Evidence* (Washington, DC: National Academies Press, 2011), 13 ("First and foremost, the Court viewed science as an empirical endeavor: '[W]hether [a theory or technique] can be (and has been) tested' is the 'methodology [that] distinguishes science from other fields of human inquiry.' The Court also mentioned as indicators of good science …the existence of known or potential error rates has been determined…").

[27] Miguel A. Hernán and James M. Robins, *Causal Inference: What If* (Boca Raton, FL: Chapman & Hall/CRC, 2024), 16 ("Our discussion of randomized experiments refers to population or average causal effects because individual causal effects cannot generally be identified."); Mingzhang Yin et al., "Conformal Sensitivity Analysis for Individual Treatment Effects," *Journal of the American Statistical Association* 119, no. 545 (2024), 124 ("Therefore, the [individual treatment effect] ITE, which requires knowing both the potential outcomes, can never be observed. Furthermore, unlike population-level causal estimands, an ITE is inherently random. Even with a known joint distribution …, an ITE is not point-identifiable").

[28] See Ramsey and Saravia for further discussion.

[29] McClave and McClave, 64.

[30] McClave and McClave, 65.

[31] McClave and McClave, 65.

[32] See for example, Wooldridge, 240–45.

[33] McClave and McClave appear to recognize that it is reasonable and appropriate to statistically test whether different overcharge coefficients should be estimated for different class members or groups of class members. See McClave and McClave, 64 ("Whether formal statistical testing is appropriate for any group of estimated overcharges depends largely on two things: First, whether the test is motivated by an a priori rationale related to the case and, second, whether the sample size (number of transactions being tested) is sufficient to produce a reliable test.").

[34] McClave and McClave, 65.

[35] McClave and McClave, 66–67.