Econometric Methods for Assessing Impact in Antitrust Class Actions

James T. McClave, Ph. D. and Jamie McClave, Ph. D.[1]

Abstract

Multiple regression analysis is widely used and well-accepted for calculating damages in price-fixing antitrust litigations. Multiple regression models are used to determine whether prices were elevated by the alleged anticompetitive conduct and, if so, to quantify plaintiffs' damages. If the case is a class action, class certification may depend on whether classwide methods exist to estimate damages and whether the challenged behavior impacted all or nearly all class members. Two standard types of models are often used to estimate the aggregate, classwide damages: forecasting and dummy variable models.[2] We show that both model types can also be used to assess the extent of classwide impact by applying the fundamental premise of comparing actual and estimated "but-for" competitive prices.

1.0 Introduction

Econometric analyses are commonly employed in antitrust class actions. At the class certification stage plaintiffs typically need to establish that (1) classwide methods exist to estimate damages, and (2) the challenged behavior impacted all or nearly all proposed class members.[3] Since *Hydrogen Peroxide* the bar has been raised from showing that such methods *are likely* to exist to showing such methods *actually* exist.[4] In our experience the requirement now means conducting specific analyses that provide at least preliminary estimates of aggregate classwide damages. This analysis typically addresses the question of whether the alleged conspiracy had the effect of increasing prices to a level above that which would have prevailed absent the conspiracy. If the analysis demonstrates elevated prices, the extent of the elevated prices' impact across the proposed class may still be an open issue.

---

[1] The authors would like to thank Allison Zhou, Brooke Harmer, Devon Myers, Jodie Newman, and Wei Zhou for their contributions to this paper.

[2] Rubinfeld, D.L. (2011). "Reference Guide on Multiple Regression," Reference Manual on Scientific Evidence, Third Edition, pp. 303-357; McCrary, J. and Rubinfeld, D. (2014). "Measuring Benchmark Damages in Antitrust Litigation," Journal of Econometric Methods, 3(1): pp. 63-74.

[3] Federal Rules of Civil Procedure, Rule 23(b)(3). For purposes of this article we assume that plaintiffs are alleging a price-fixing conspiracy that violates Section 1 of the Sherman Act.

[4] *In Re Hydrogen Peroxide Antitrust Litigation,* 552 F.3d 305 (3rd Cir. 2008) ("The evidence and arguments a district court considers in the class certification decision call for rigorous analysis. A party's assurance to the court that it intends or plans to meet the requirements is insufficient"; "proper analysis under Rule 23 requires rigorous consideration of all the evidence and arguments offered by the parties. It is incorrect to state that a plaintiff need only demonstrate an "intention" to try the case in a manner that satisfies the predominance requirement").

Multiple Regression Analysis (MRA) is the most common methodology used to address the first issue, the estimation of aggregate damages across the class.[5] In some cases the MRA has been extended to address the second issue, the determination of the extent of impact across the class.[6] This article discusses two standard MRA methodologies often used to estimate classwide damages, and how both can be extended to address the impact issue.

2.0 <u>Assumptions</u>

We consider a price-fixing class action that alleges antitrust violations involving multiple defendants supplying multiple products at issue to a large number of multiple direct purchaser class members. Since the issues of damages and impact are only relevant if the trier-of-fact determines that the challenged behavior occurred, that is, that defendants are liable for the alleged antitrust conduct, we assume that the defendants will be found liable. We do not, however, assume that the challenged behavior elevated prices or that any impact of the conspiracy was classwide. The questions of damages and impact are, in fact, the subject of the econometric analysis.[7]

We assume that MRA will be used to compare prices during the alleged class period (or "conspiracy period") to prices during a benchmark period believed to be free of anticompetitive behavior. For purposes of our discussion we assume that the benchmark period consists of a time prior to the onset of the conspiratorial behavior, although the methodologies discussed herein can be extended to benchmarks that consist of times both before and after the conspiracy period, only after the conspiracy period, and even to "yardstick" benchmarks defined as comparable but unaffected markets concurrent with the conspiracy period. The data in both the benchmark and class periods are assumed to consist of numerous transactions between defendants and plaintiffs. In simple terms, the MRA consists of an equation that relates a variable of interest, the "dependent variable," to a set of variables, the "explanatory variables," thought to be related to the dependent variable. The dependent variable in MRAs involving antitrust allegations is almost always the natural logarithm of the unit price of

---

[5] See, e.g., Rubinfeld, D.L. (2011). "Reference Guide on Multiple Regression," Reference Manual on Scientific Evidence, Third Edition, pp. 303-357 at p. 306: "antitrust violations" listed among types of cases for which "regression analysis has been used most frequently."

[6] See, e.g., *In re Air Cargo Shipping Services Antitrust Litigation*, 2014 WL 7882100 (E.D.N.Y., 2014); *In re Korean Ramen Antitrust Litigation*, 2017 WL 235052 (N.D.Cal., 2017); *In re Capacitors Antitrust Litigation (No. III)*, 2018 WL 5980139 (N.D.Cal., 2018); *In re Packaged Seafood Antitrust Litigation*, 332 F.R.D. 308 (S.D.Cal., 2019); *In re Disposable Contact Lens Antitrust*, 329 F.R.D. 336 (M.D.Fla., 2018); *In re Peanut Farmers Antitrust Litigation*, Case No. 2:19-cv-00463-RAJ-LRL (E.D.Vir., 2020); Wortman v. Air New Zealand, 326 F.R.D. 549 (N.D.Cal., 2018); *In re Broiler Chicken Antitrust Litigation*, 2022 WL 1720468 (N.D. Ill., 2022); *In re Pork Antitrust Litigation*, Civil No. 18-1776 (JRT/JFD), MDL No. 21-2998 (MN, 2023); *In re Turkey Antitrust Litigation*, Civil No. 19 C 8318 (SRH), (N.D.II, 2024); *In re HDD Suspension Assembly Antitrust Litigation*, Civil No. 19-md-02918-MMC (N.D. Cal., 2024).

[7] If a reliable econometric analysis finds that prices were elevated above competitive levels, the analysis provides empirical support for plaintiffs' allegations of anticompetitive conduct.

transactions, which we represent by P and call “price” without repeating “logarithm of” each time.[8] We also assume that the econometrician has correctly identified relevant explanatory variables impacting prices that are unrelated to the alleged conspiratorial behavior, represented by the matrix X. The explanatory variables consist of both time-related variables, like cost and demand, and cross-sectional variables, like customer and product characteristics. We will refer to the variables in X as the “competitive explanatory variables” (CEVs). We assume that the MRA models in this exposition are well-specified, and Ordinary Least Squares (OLS) is used to provide estimates of the relationship between P and X.[9] We understand that model specification will likely be challenged during class certification. This article, however, focuses on the MRA methodology for assessing damages and impact across the class; challenges to model specification go to the validity of the model specification rather than the underlying methodology.

The following sections consider each of the steps of an econometric analysis intended to address the issues of damages and impact across the putative class. We compare the forecasting and dummy variable methods for accomplishing each step of the analysis.

3.0 <u>Model Specification</u>

3.1 <u>Forecasting Model</u>

The forecasting method uses data during only the benchmark period to estimate the relationship between price and the CEVs X. The forecasting model specification can be written as:

$$P_{t,ij} = \beta X_{t,ij} + \varepsilon_{t,ij} \tag{1}$$

where $P_{t,ij}$ is (the logarithm of) the price paid by customer i for product j at time t; $X_{t,ij}$ represents the values of the CEVs corresponding to that transaction; β is a vector of coefficients that relate price to the CEVs; and $\varepsilon_{t,ij}$ is the “error term,” which is the difference between the actual price and the regression model’s characterization of the price effects of the CEVs.[10] We assume that the errors are random variables with mean

[8] The use of the logarithmic transformation of prices in econometric models is common both in legal applications and more generally. Gujarati, D.N., and Porter, D.C. (2009). Basic Econometrics, 5th ed., pp. 159-162. The methods described herein are also applicable to models of price in (untransformed) levels.
[9] Our discussion can be expanded to include other estimation techniques, such as weighted and generalized least squares, with an appropriate expansion of assumptions necessary for their validity.
[10] Transactions are sometimes aggregated over some time period, such as weeks or months for specific customers, products, and other transaction classifications. The model is estimated using aggregated observations rather than individual transactions. We use “transaction” to refer to the observations used to estimate the model, whether individual or aggregated.

zero and are orthogonal to X.[11] Both $P_{t,ij}$ and $X_{t,ij}$ are observable; β and $\varepsilon_{t,ij}$ are not observable and must be estimated.

3.2 Dummy Variable Model

Unlike the forecasting method which uses only transactions during the benchmark period to estimate the CEVs' relationships with price, the dummy variable method uses transaction data during both the benchmark and class periods to estimate those relationships. The model specification is otherwise analogous to that for the forecasting model. During the benchmark period the model is the same as forecasting equation:

$$P_{t,ij} = \beta X_{t,ij} + \varepsilon_{t,ij} \tag{1}$$

During the class period, however, the model must allow not only for the price effects of the CEVs, the model must include an additional factor to account for potential price effects of the conspiracy:

$$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} + \varepsilon_{t,ij} \tag{2}$$

where $\tau_{t,ij}$ is the true overcharge associated with the (t,ij) transaction in the class period. It is included in the model specification to enable the determination of whether the transaction was overcharged and, if so, to what extent.

Because transactions during both periods will be used to estimate the dummy variable model, we combine the benchmark and class period equations into a single model specification:

$$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} D_C + \varepsilon_{t,ij} \tag{3}$$

where D_C is equal to one for transactions in the class period, and equal to 0 for transactions in the benchmark period. The variable D_C is referred to as an "indicator" or "dummy" variable, which has the effect of only including the potential overcharge for transactions during the class period.

4.0 Estimation of the Aggregate Classwide Overcharge

4.1 Forecasting Method

The forecasting method for estimating damages uses equation (1) to forecast prices during the class period. Since the forecasting model is estimated using only

[11] We understand that, like model specification, the error term assumptions may be challenged. Such challenges often relate to model specification, such as violation of independence between X and ε due to claimed endogeneity of certain explanatory variables. Such challenges are distinct from the evaluation of the methodology's reliability.

transactions during the benchmark period, the forecasted prices during the class period are often referred to as "but-for" prices, meaning that the forecasted prices are an estimate of the prices that would have occurred during the class period in the absence of – "but-for" – the conspiracy.

Actual transaction prices during the class period can be represented using the same class period specification as for the dummy variable method:

$$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} + \varepsilon_{t,ij} \quad (4)$$

where $\tau_{t,ij}$ is included to enable the determination of the existence and extent of the true overcharge, if any, associated with the (t,ij) transaction in the class period. Note that the relationship between the price $P_{t,ij}$ and the CEVs, represented by $\beta X_{t,ij}$, is the same during the class period as during the benchmark period. This equivalence is a fundamental assumption underlying the forecasting model.

We next determine whether the class period transactions were elevated in aggregate. The true classwide overcharge τ is defined as the mean of the individual class period transaction's overcharges:

$$\tau = \sum_C \tau_{t,ij} \div N_C \quad (5)$$

where $\sum_C$ adds the individual transactions overcharges over the N_C class period transactions. Using equation (4) to substitute for the individual transaction overcharges, we have:

$$\tau = \sum_C (P_{t,ij} - \beta X_{t,ij} - \varepsilon_{t,ij}) \div N_C \quad (6)$$

Because both the CEVs coefficients β and the error terms ε are unobservable, τ must be estimated.

Ordinary Least Squares (OLS) can be used to estimate the price effects, β, of the CEVs, X, using only transactions during benchmark period, producing the estimation equation:

$$p_{t,ij} = bX_{t,ij} \quad (7)$$

where $p_{t,ij}$ is the forecasting model's estimated price for customer i who purchased product j at time t. The vector of coefficients b is the estimate of the relationship between price and each of the CEVs. Note that the OLS estimate for the error term $\varepsilon_{t,ij}$ is its mean value, zero. Substituting the OLS estimates into (6) produces an estimate of τ:

$$E = \sum_C (P_{t,ij} - bX_{t,ij}) \div N_C \quad (8)$$

The estimate E is therefore the mean of the differences between the actual prices and the forecasting model's estimated but-for prices across all transactions in the class period. Under the assumptions underlying the forecasting model, E has optimal statistical properties for estimating τ, and is the well-accepted method for estimating the aggregate classwide damages using the forecasting method.[12] An important characteristic of this aggregate classwide overcharge estimate is that to the extent any class period transaction was not affected by the alleged price fixing, it does not contribute to the damages estimate. Thus there is no concern about inflating the damages estimate by including the unimpacted class members' transactions.[13]

4.2 Dummy Variable Method

Like the forecasting method, the application of the dummy variable method typically begins with estimating the classwide aggregative overcharge. To accomplish this, remembering that the model will be estimated using transactions during both the benchmark and damages periods, we incorporate the true aggregate overcharge τ in the dummy variable model specification (2):

$$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} D_C + \varepsilon_{t,ij} = \beta X_{t,ij} + \tau D_C + \delta_{t,ij} \qquad (9)$$

where

$$\delta_{t,ij} = (\tau_{t,ij} - \tau) D_C + \varepsilon_{t,ij} \qquad (10)$$

Note that $\delta_{t,ij}$ consists of two components for transactions during the class period: the difference between the individual transaction's overcharge and the aggregate overcharge, and the error term. Since the aggregate overcharge is the mean of the class period transactions' overcharges, the first term sums to zero over the class period. Recall that we assume that the model's CEVs, X, and the dummy variable D_C, are orthogonal to the error term ε, which has a mean value of zero.[14] The values of $P_{t,ij}$, $X_{t,ij}$, and D are observable; the values of β, τ, and $\delta_{t,ij}$ are unobservable.

---

[12] Under the model assumptions, OLS estimates are unbiased and consistent. See, e.g., McCrary, J. and Rubinfeld, D.L. (2014). "Measuring Benchmark Damages in Antitrust Litigation," Journal of Econometric Methods, 3(1): 63-74, equation (7) at 65. The equation there is a quantity weighted sum, assuming the price model is in levels (dollars) rather than logarithms (percentages). That equation and equation (8) herein share the property that the aggregate overcharge is calculated using the sum of differences between actual and forecasted prices (or logarithmic prices) over all transactions during the class period.

[13] Technically, the expected value of unimpacted transactions is zero.

[14] The inclusion of potential overcharges in the class period residuals may introduce heteroscedasticity – unequal variances – between the benchmark and class period residuals. OLS estimates remain unbiased and statistically consistent in the presence of heteroscedasticity. Tests for statistical significance can be adjusted for potential heteroscedasticity using robust standard errors. See, e.g., Kennedy, P. (2008). A Guide to Econometrics, 6th ed., pp. 113, 115.

Model (9) is estimated using OLS. which provides estimates of both the CEVs coefficients β and the classwide overcharge τ, using all transactions in both the benchmark and class periods. The estimated regression model is:

$$p_{t,ij} = bX_{t,ij} + E \times D_C \tag{11}$$

where $p_{t,ij}$ is the estimated (logarithmic) price for customer i who purchased product j at time t. Thus, the application of OLS to the dummy variable model provides not only estimates of the price effects b of the CEVs, but also the estimate E of the classwide overcharge estimate τ.

Recall that the forecasting model's estimate of E is obtained by calculating the mean difference between actual and forecasted but-for prices during the class period, shown in (8). This may at first appear to be very different from the estimate obtained from the dummy variable method that uses all transactions in both periods to simultaneously estimate β and τ by b and E, respectively. The difference, however, is more illusory than real.

In fact, the OLS calculation of the aggregate classwide overcharge estimate for the dummy variable model, like the forecasting model's estimate, consists of the mean difference between the actual and estimated but-for prices during the class period:

$$E = \sum_C (P_{t,ij} - bX_{t,ij}) \div N_C \tag{12}$$

This formula is *identical* to equation (8) used to estimate the aggregate overcharge by the forecasting method.[15] As with the forecasting method, the estimated price $bX_{t,ij}$ from the dummy variable model is the "but-for" price, meaning the estimated price absent the price effect of the conspiracy. The only difference is the data used to estimate the CEVs price effects, b: only benchmark transactions for the forecasting method, both benchmark and class period transactions for the dummy variable method. Both the dummy variable method and the forecasting method, however, estimate the aggregate classwide overcharge by the mean difference between actual and but-for prices during the class period.

4.3 Testing the Statistical Significance of the Estimated Aggregate Overcharge

For both methods, we use τ to represent the true classwide aggregate overcharge, defined as the mean difference between the actual prices paid during the class period and the "true" but-for prices. An exact determination of whether the true

[15] The proof that the dummy variable's OLS overcharge estimate E is the mean difference between actual and but-for prices during the class period is in the Appendix.

classwide overcharge τ is greater than zero is impossible, because the true but-for prices are unobservable, having been "hidden" by the conspiratorial behavior.[16]

Standard statistical methods can be used to make inferences about the true aggregate overcharge. Specifically, we can test the competing hypotheses that the true overcharge is zero, meaning that the challenged conduct did not elevate prices across the class, vs. that the true overcharge is positive, meaning that prices were elevated. In statistical parlance, we test the null hypothesis that τ is zero – no classwide overcharge – against the alternative hypothesis that τ is greater than zero – positive classwide overcharge. Fundamentals of statistical testing assume that the null hypothesis is true unless the data convincingly indicates it is not. In this application, the assumption is that the conspiracy has not elevated prices *unless and until* the class period transactions provide convincing statistical evidence to the contrary.

The best statistical estimate of the hidden true classwide overcharge is E as specified above for both the forecasting model and the dummy variable model.[17] The test is conducted by comparing E to the variability of the estimate. We omit the technical details of the tests, focusing on the potential results of the test. If the data fail to provide sufficient evidence to support the alternative hypothesis, the conclusion is that E is not "statistically significant." If the statistical test does provide sufficient evidence to support the alternative hypothesis that prices were elevated during the class period, E is said to be "statistically significant."[18]

If the test result is that E is not statistically significant, the question of the extent of classwide impact may be considered moot.[19] On the other hand, if the test supports the hypothesis that τ is positive – E is statistically significant –, it provides statistical

---

16 See *Story Parchment Co. v. Paterson Parchment Paper Co.*, 282 U.S. 555 (1931); *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442 (2016).

17 Here "best" means that E is the unbiased estimate of τ with the smallest variance.

18 "Statistical significance" in hypothesis testing is measured by the likelihood that the estimated overcharge E would have been observed by chance when the true overcharge τ is zero, that is, when the null hypothesis is true. This likelihood is referred to as the significance level of the test, and the smaller the significant level, the more convincing the evidence for the alternative hypothesis and against the null hypothesis. One of the most common significance levels used in testing is .05, meaning that the alternative hypothesis is accepted only if the data indicate less than a 5% probability that the size of the overcharge estimate E would have occurred if the true overcharge τ were zero. In that case, we refer to E as "statistically significant" at the 5% level of significance. Technical details of testing hypotheses can be found in any standard statistics or econometrics texts. See, e.g., McClave, J.T., Benson, P.G., and Sincich, T. (2018). Statistics for Business and Economics, 13th ed., Chapter 7.

19 Failure to find statistical significance does not, standing alone, indicate that the conspiracy did not inflate prices. Statistical tests have the possibility of failing to support the alternative hypothesis even when it's true. If the true classwide overcharge is positive but the test failed to find a statistically significant effect, the test resulted in a "false negative." If the true classwide aggregate overcharge is zero and the test has correctly found no statistical significance, no further assessment of classwide impact is warranted.

evidence that prices were elevated during the class period to an extent that is consistent with having been caused by the conspiracy and inconsistent with having been caused by the non-conspiratorial explanatory factors – the CEVs –, or by other idiosyncratic random effects on prices. In some cases the statistical significance of the classwide overcharge estimate could, if accompanied by other economic and documentary evidence, be found sufficient to support the inference of classwide impact.[20] For purposes of this discussion, however, we assume that further econometric analysis is conducted to assess classwide impact.

5.0 Econometric Assessment of Impact

Statistical significance of E supports the inference that the true classwide overcharge, τ, is positive. τ is the mean of individual class period transactions overcharges,$\tau_{t,ij}$,. Like τ, however, $\tau_{t,ij}$ is not observable; it has to be estimated because the conspiracy concealed the prices that would have been paid "but for" its existence. Just as E provides an estimate of the true overcharge E, the transactional components of E provide estimates of the true overcharge for each transaction. The comparisons of actual prices and but-for prices based on either the forecasting or dummy variable method associated with each class period transaction are:

$$E_{t,ij} = P_{t,ij} - bX_{t,ij} \tag{13}$$

where $E_{t,ij}$ is the overcharge estimate for transaction (t,ij) during the class period.

For both methods class period prices consist of price effects of the CEVs, the potential overcharge, and random error:

$$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} + \varepsilon_{t,ij} \tag{14}$$

Substituting this expression into (1) expresses the individual transaction's estimated which, when substituted into the individual overcharge equation (13), produces:

$$E_{t,ij} = (\beta\text{-}b)X_{t,ij} + \tau_{t,ij} + \varepsilon_{t,ij} \tag{15}$$

---

[20] For example, proof of a pricing structure can be used to infer common impact. See, e.g., *In re High-Tech Employee Antitrust Litigation*, 985 F.Supp.2d 1167, at 1206 (N.D. Cal. 2013). Order Granting Plaintiffs' Supplemental Motion For Class Certification ("Plaintiffs noted that Dr. Leamer's approach followed a roadmap widely accepted in antitrust class actions that uses evidence of general price effects plus evidence of a price structure to conclude that common evidence is capable of showing widespread harm to the class."); Memorandum Opinion and Order (Judge Sunil R. Harjani), *In re Turkey Antitrust Litigation*, Civil No. 19 C 8318 (SRH), (N.D.Il., Jan. 22, 2025) ("the Court determines that the evidence put forward by the DPPs, including Williams' market structure analysis, overcharge regression model, supplemented by the in-sample prediction method, robustness checks on the overcharge regression analysis, correlation analyses, production regression, and record evidence, when combined, is sufficient to show common questions predominate as to common impact. *Kleen*, 831 F.3d at 925").

As with the forecasting method, we rely on the fact that b is an unbiased estimate of β. Since the expected value of $\varepsilon_{t,ij}$ is zero, it follows that $E_{t,ij}$ is an unbiased estimate of the true overcharge $\tau_{t,ij}$ for transaction (t,ij).

Some have claimed that the dummy variable model's specification assumes that all class members were overcharged by the same amount, the aggregate overcharge E. As shown above this is certainly not true, since E consists of individualized comparisons of actual and but-for prices for *both* the forecasting and dummy variable methods. The difference between each transaction's estimated overcharge and the estimated aggregate overcharge is:

$$d_{t,ij} = E_{t,ij} - E \tag{16}$$

Substituting (13) for the transaction overcharge:

$$d_{t,ij} = (P_{t,ij} - bX_{t,ij}) - E = P_{t,ij} - (bX_{t,ij} + E) \tag{17}$$

The first expression in (17) is the difference between each class period transaction's overcharge and the aggregated overcharge. The second expression is the difference between the transaction's actual price and the dummy variable model's estimated price during the class period (11). Differences between actual and estimated values in a regression model are the model's "residuals."[21] Thus, the differences in (17) are the dummy variable model's residuals during the class period. Although the residuals sum to zero over the class period, there is no assumption that each residual is zero. A negative residual means that the transaction's overcharge is lower than the aggregate overcharge; a positive residual means that the transaction's overcharge is higher than the aggregate overcharge. If the residual is negative and exceeds E in absolute value, the inference is that the transaction was unimpacted by the conspiracy.

Because both methods produce unbiased estimates of each class period transaction's overcharge, we then provide econometric evidence of whether each transaction was impacted by comparing its actual and but-for price:

$$E_{t,ij} = P_{t,ij} - bX_{t,ij} > 0 \rightarrow \text{Impacted Transaction} \tag{18a}$$

$$E_{t,ij} = P_{t,ij} - bX_{t,ij} \leq 0 \rightarrow \text{Unimpacted Transaction} \tag{18b}$$

[21] See, e.g., McClave, J.T., Benson, P.G., and Sincich, T. (2018). Statistics for Business and Economics, 13th ed., p. 760.

Customer-level impact is measured by determining whether the customer was overcharged on at least one transaction during the class period.[22] For both the forecasting and dummy variable methods, this is expressed as follows for Customer i:

$$\mathrm{Max}_i(E_{t,ij}) > 0 \;\rightarrow\; \text{Impacted Customer} \tag{19a}$$

$$\mathrm{Max}_i(E_{t,ij}) \leq 0 \;\rightarrow\; \text{Unimpacted Customer} \tag{19b}$$

where Max_i denotes the calculation of the maximum value of the overcharges – the differences between actual and but-for prices -- across all of Customer i's transactions.[23]

Another measure of customer-level impact can be based on the customer's mean estimated overcharge during the class period:

$$E_i = \sum\nolimits_I (E_{t,ij}) \div N_i \tag{20}$$

where $\sum_I$ denotes the sum over all N_i of customer i's transactions during the class period, and E_i is the mean overcharge for customer i. E_i is the mean difference between Customer i's actual prices and but-for prices under *both* methods. These customer-level overcharge estimates can be used as an estimate of customer-level net impact:

$$E_i > 0 \;\rightarrow\; \text{Net Impacted Customer} \tag{21a}$$

$$E_i \leq 0 \;\rightarrow\; \text{Net Unimpacted Customer} \tag{21b}$$

The percentage of customers impacted on at least one transaction and the percentage of customers with positive mean overcharges are two measures of the degree to which overcharges are experienced throughout the class. Higher values of these percentages imply more widespread impact across the class.

---

[22] See Opinion, *In Re: Generic Pharmaceuticals Pricing Antitrust Litigation*, MDL NO. 2724 (E.D.Penn. Dec. 3, 2024), p. 22 ("Courts do not consider whether antitrust injury occurs on an overall basis—a court's inquiry on the matter is limited to determining whether a single overcharge occurred."). See also *Adams v. Mills*, 286 U.S. 397, 407 (1932) ("In contemplation of law the claim for damages arose at the time the extra charge was paid"); *Hawaii v. Standard Oil of Cal.*, 405 U.S. 251, 262 n.14 (1972) ("courts will not go beyond the fact of this injury to determine whether the victim of the overcharge has partially recouped"); and *In re Nexium Antitrust Litig.*, 777 F.3d 9, 27 (1st Cir. 2015) ("antitrust injury occurs the moment the purchaser incurs an overcharge, whether or not that injury is later offset.").

[23] Recall that this further assessment of class member impact only proceeds once an aggregate overcharge has been detected. By design, the benchmark period is assumed to be a period of competitive pricing and thus there is no aggregate overcharge; thus, there can be no assessment class member impact during a period during which there is no alleged conspiracy. The benchmark transactions are assumed to have no overcharge attributable to the conspiracy. As a result, the model's benchmark residuals will be distributed around their mean value of zero. Any attempt to interpret positive benchmark residuals as some kind of "false positive" indication of impact is a misinterpretation of a foundational statistical result.

To summarize, we proceed to impact assessment only if the estimate of aggregate classwide overcharges, E, is statistically significant.[24] This test accompanied by high percentages of class members paying at least one overcharge (19a) and of net impacted class members (21a) provide strong econometric evidence of classwide impact.[25] Defendants may argue that more evidence is required, including statistical significance not only of the aggregate overcharge, but also of individual class member's estimated overcharges. Most proposed classes, however, consist of large numbers of customers with a small number of transactions. Small customers with few transactions are less likely to produce statistically significant overcharge estimates, even if they were truly impacted by the conspiracy. Technically, the power of any tests conducted to find statistical significance even when small customers are impacted is low, meaning that the tests are likely to result in numerous false negatives, that is, tests failing to reveal impact when it is present.[26] Requiring statistical significance for each of hundreds or thousands of class members' overcharges is an artificial barrier that approaches statistical impossibility.

6.0 <u>Assessing Class Composition</u>

In the determination of classwide damages and impact, we assume that the defendants conspired. We did not assume, however, either that prices were elevated or that all or nearly all class members were impacted. We have shown that both the forecasting and dummy variable methods provide a reliable econometric methodology both to measure classwide overcharges and to determine the extent of classwide impact. If none or very few potential class members were impacted by the conspiracy, the estimate E of the aggregate overcharge is expected to fail to pass the test of statistical significance. If, however, some portion of the class is impacted but the class composition was over-specified so that another portion is unimpacted, we want the econometric analysis to reflect the difference. If the unimpacted portion is substantial, we expect both the aggregate overcharge estimate and the percentages of impacted customers to be lower using the metrics described in the previous section.

Further analysis of transaction- and customer-level estimated overcharges can shed additional light on impact. A better alternative to extensive, low-power statistical testing at the class member level is to compare the distribution of the overcharge

[24] Failing to find E as statistically significant when in fact the classwide overcharge τ is positive (that is, the test resulted in a "false negative") would be more likely if a substantial portion of the proposed class was not impacted by the conspiracy, but still some portion was impacted. Detection of this disparate impact would require further analysis that might provide an altered class definition.
[25] See, e.g., *supra* note 5.
[26] See, e.g., Pindyck, R.S. and Rubinfeld, D.L. (1998). Econometric Models and Economic Forecasts, 4th ed., pp. 43-44; McClave, J.T., Benson, P.G., and Sincich, T. (2018). Statistics for Business and Economics, 13th ed., p. 415.

estimates in the class period to the distribution of residuals in the benchmark period. Benchmark residuals are the differences between the actual price and the model's estimated price during the benchmark period, calculated as follows for both forecasting and dummy variable models:

$$R_{t,ij} = P_{t,ij} - p_{t,ij} = P_{t,ij} – bX_{t,ij} \tag{22}$$

When b is estimated using OLS, the mean of the benchmark period residuals is equal to zero. The distribution of the benchmark period transactions' residuals will cluster around zero, with approximately half having positive values and half negative values.

The distribution of residuals can be compared to the distribution of their counterparts in the class period, the transactions' differences between the actual and "but-for" prices:

$$E_{t,ij} = P_{t,ij} - bX_{t,ij} \tag{23}$$

These transaction overcharges will be distributed around their mean, which is the aggregate overcharge estimate, E.

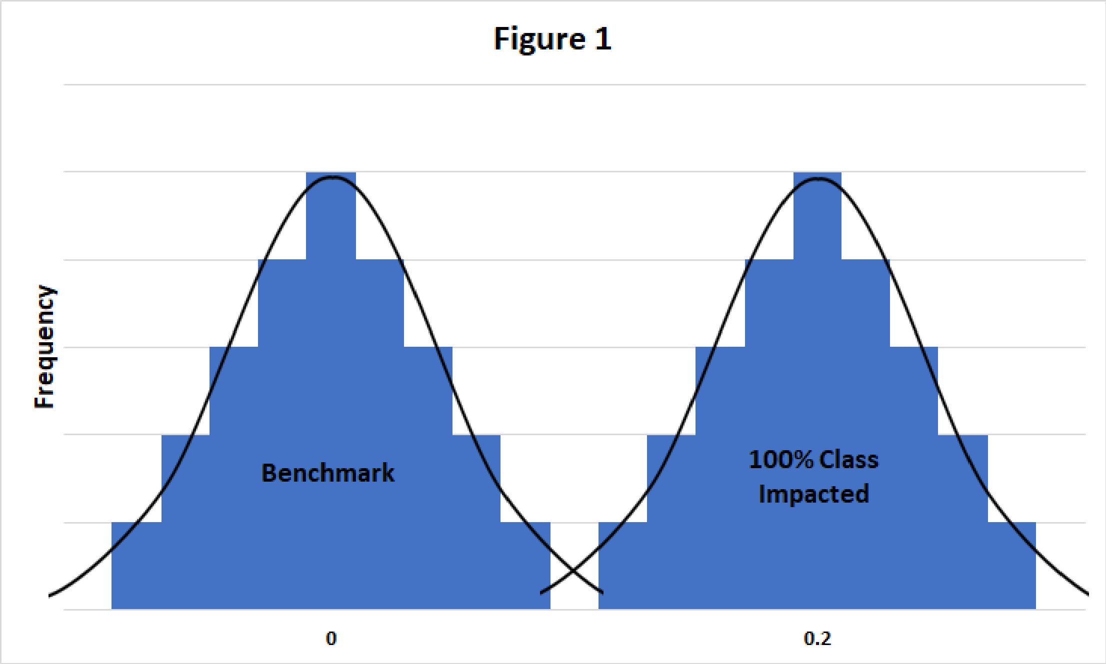
The comparison can be extended to the customer level by aggregating the residuals and the overcharge estimates across each customer's transactions. The customer-level aggregations for Customer i are:

$$R_i = \sum_i (R_{t,ij}) \div N_i \tag{24a}$$

$$E_i = \sum_i (E_{t,ij}) \div N_i \tag{24b}$$

for the benchmark transactions and class transactions, respectively. The values of R_i will be distributed around zero and the values of E_i will be distributed around E.

If the impact is classwide and relatively similar throughout the class period, the distribution of customer overcharges should be unimodal, similar in shape to the distribution of the residuals during the benchmark period but moved to the right, as demonstrated in Figure 1. If the classwide aggregate overcharge to purchases is 0.2 (an approximate 20% overcharge) and all class members were overcharged to an extent that distributes them around 0.2, the distribution of overcharge estimates could appear like that depicted on the right side of Figure 1. While the benchmark residuals cluster around zero, the class period overcharges are shifted to the right.



If, however, impact varies significantly among different groups of class members, the distribution of the customer overcharges will be multi-modal. For example, suppose some subset of class members was unimpacted by the conspiracy. This means that the true overcharge across the class consists of two distinct components:

$$\tau = [\textstyle\sum_{S1} (\tau_{t,ij}) + \sum_{S2} (\tau_{t,ij})] \div N_C \qquad (25)$$

where S1 is the subset of transactions for unimpacted customers and S2 is the subset of transactions for impacted customers. By definition, the first sum over unimpacted customers is zero, so the true total overcharge reduces to just the sum over the impacted subset:

$$\tau = [0 + \textstyle\sum_{S2} (\tau_{t,ij})] \div N_C \qquad (26)$$

Suppose, for example, that half the transactions (and customers) are in each subset. The inclusion of both subsets in the calculation will result in an underestimate of the true overcharge to the impacted subset S2 by half. That is, the divisor of τ in (26) should be $\frac{1}{2}N_C$, the number of impacted transactions, rather than N_C, the total of both unimpacted and impacted transactions.

Whereas the forecasting model specification remains unchanged despite the class over-specification, the dummy variable model can be written to reflect it:

$$P_{t,ij} = \beta X_{t,ij} + (0 \times D_{C1} + \tau_{t,ij} \times D_{C2}) + \delta_{t,ij} \qquad (27)$$

where D_{C1} is an indicator ("dummy") variable equal to 1 for transactions in the unimpacted subset S_1 during the class period, and equal to 0 otherwise; and D_{C2} is an indicator variable equal to 1 for transactions in the impacted subset S_2 during the class period, and equal to 0 otherwise. Thus, the elevated prices are confined to only transactions in S_2, and there is no price elevation in S_1 transactions.

Assuming the model (9) with a single classwide indicator is estimated prior to identification of the unimpacted subset of transactions, the specification is unchanged, but the coefficient of the dummy variable is a fraction of the true overcharge experienced by the impacted subset:

$$P_{t,ij} = \beta X_{t,ij} + (q \times \tau) \times D_C + \delta_{t,ij} \qquad (28)$$

where q is the proportion of class period transactions in the impacted subset S_2 and the indicator variable D_c spans all transactions in the class period, including both S_1 and S_2. Since q is unobservable, when OLS is used to estimate the model using this specification, the result appears the same as (11):

$$p_{t,ij} = bX_{t,ij} + E \times D_C \qquad (11)$$

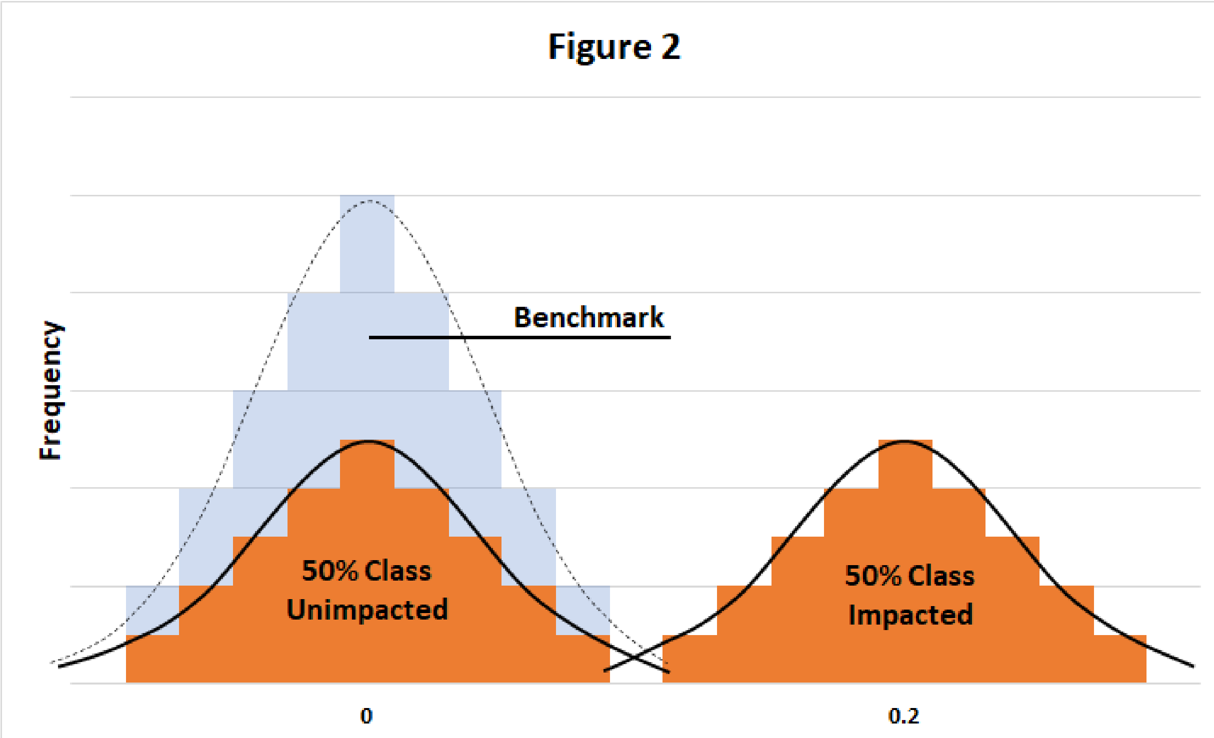
where b is the estimate of β, and E is the estimate of the aggregate classwide overcharge, $q \times \tau$. For example, if half the class transactions are impacted, the estimated classwide overcharge E will be an unbiased estimate of half the true mean overcharge across the impacted sub-class, $½\tau$.[27]

Assuming that half of the class transactions are unimpacted but with no *a priori* knowledge of which transactions and proposed class members are in the unimpacted subset, we turn once again to an analysis of the individual contributions to the aggregate overcharge estimate. This analysis for both the forecasting and dummy variable methods examines the differences between actual and but-for prices during the class period, $E_{t,ij}$, and the corresponding customer-level overcharges, E_i. The distribution of both individual and customer-level overcharge estimates will be bimodal, with half clustered around zero (mimicking the distribution of the benchmark's residuals) and half clustered around twice the estimated aggregated overcharge across the impacted sub-class, $2 \times E$.

For example, if half of purchasers were not impacted by the price-fixing conspiracy, and the half that are impacted have an aggregate overcharge of 0.2, the distribution of the class member's overcharges will be bimodal for both the forecasting and dummy variable methods, like that shown in Figure 2.

---

[27] Note that any claim that the presence of unimpacted customers somehow increases the overcharge is false. In fact, their presence results in a lower overcharge.



In summary, if a specific subset of the class transactions is unimpacted, the analysis described above will expose the difference between impacted and unimpacted transactions and will assist in the identification of the underlying distinction between the two. Identification of the members of each subset would normally require an investigation to determine whether they share some common property. For example, unimpacted customers might have purchased different products or might have made their purchases in a different geographic market than impacted customers. Once identified, a test of statistical significance for the difference between the means of the two subsets could be conducted to confirm that the class was over-specified. This same analysis can be extended to an appearance of multi-modal distributions attributable to subsets of class transactions with different levels of impact reflected by different aggregate overcharges. The econometric analyses of impact described in this and the previous section are likely to expose the fact that a substantial portion of the class was unimpacted.

7.0 Conclusion

The estimate of the aggregate classwide overcharge under both the forecasting method and the dummy variable method consists of comparing actual prices and the model's estimated but-for prices for all transactions during the class period. Under standard model assumptions, the difference for each transaction is a statistically unbiased estimate of the transaction's overcharge attributable to the challenged conduct. Customer-level aggregations of the differences between actual and but-for prices inform the extent of impact across the class.

Both methods offer numerous assessments of impact after the aggregate overcharge has been estimated and found to be statistically significant:

1. The calculation of the percentage of customers with at least one impacted transaction during the class period.
2. The calculation of the percentage of customers with positive mean overcharges during the class period.
3. Comparisons of the distributions of customer overcharges during the class period to the distribution of residuals during the benchmark period.
4. Examination of reasons for possible bimodal (or multi-modal) distributions of overcharge estimates for different subsets of customers, including the identification of potentially unimpacted customers.

The following table demonstrates that the application of the forecasting and dummy variable methods to measure classwide damages and to assess classwide impact is essentially identical.

Forecasting Method

Model:
$P_{t,ij} = \beta X_{t,ij} + \varepsilon_{t,ij}$

Assumptions:
$\varepsilon_{t,ij}$ mean zero, independent of X

OLS Price Estimate:
Based on Benchmark Only
$p_{t,ij} = bX_{t,ij}$

Classwide Aggregate Overcharge
Mean of Actual minus But-For Price:
$E = \sum_C (P_{t,ij} - bX_{t,ij}) \div N_C$

Test of Statistical Significance:
Compares E to its Std. Error

Assessment of Classwide Impact:

Transaction Overcharge
Actual minus But-For Price:
$E_{t,ij} = P_{t,ij} - bX_{t,ij}$

Impacted Transaction:
$E_{t,ij} > 0$

Impacted Class Member:
At least one impacted transaction:
$Max_i(E_{t,ij} > 0)$

Net overcharge positive:
$[\sum_i (E_{t,ij}) \div N_i] > 0$

Detection of Unimpacted Class Subset:
Bimodal Distribution of Overcharges
Unimpacted Subset Mode Near Zero

Dummy Variable Method

Model:
$P_{t,ij} = \beta X_{t,ij} + \tau_{t,ij} \times D_C + \varepsilon_{t,ij}$

Assumptions:
$\varepsilon_{t,ij}$ mean zero, independent of X and D_C

OLS Price Estimate:
Based on Benchmark and Class
$p_{t,ij} = bX_{t,ij} + E \times D_C$

Classwide Aggregate Overcharge
Mean of Actual minus But-For Price:
$E = \sum_C (P_{t,ij} - bX_{t,ij}) \div N_C$

Test of Statistical Significance:
Compares E to its Std. Error

Assessment of Classwide Impact:

Transaction Overcharge
Actual minus But-For Price:
$E_{t,ij} = P_{t,ij} - bX_{t,ij}$

Impacted Transaction:
$E_{t,ij} > 0$

Impacted Class Member:
At least one impacted transaction:
$Max_i(E_{t,ij} > 0)$

Net overcharge positive:
$[\sum_i (E_{t,ij}) \div N_i] > 0$

Detection of Unimpacted Class Subset:
Bimodal Distribution of Overcharges
Unimpacted Subset Mode Near Zero

Appendix: Disaggregation of the Dummy Variable Model's Classwide Overcharge Estimate

Proposition: *The OLS estimate of the aggregate classwide overcharge in a dummy variable model is equal to the mean difference between class period transactions' actual prices and the model's but-for price estimates.*

Proof:

The dummy variable model to be estimated is:

$$P_{t,ij} = \beta X_{t,ij} + \tau D_C + \delta_{t,ij}$$

Using OLS, estimates b of β and E of τ are calculated to minimize the sum of squared differences between the actual and estimated prices over the periods – benchmark and class – covered by the data:

$$OLS(b,E) = \text{Min} \left\{ \sum_{B+C} \left[P_{t,ij} - bX_{t,ij} - ED_C\right]^2 \right\}$$

A necessary condition to achieve this minimum is that the partial derivative with respect to the estimated parameters b and E must equal zero.[28] The partial derivative with respect to E is:

$$\frac{\partial[OLS(b,E)]}{\partial E} = -2 \sum_{B+C} \left[P_{t,ij} - bX_{t,ij} - ED_C\right] \times D_C = -2 \sum_{C} \left[P_{t,ij} - bX_{t,ij} - E\right] = 0$$

Rearranging to solve for E:

$$E = \frac{1}{N_c} \sum_{C} \left[P_{t,ij} - bX_{t,ij}\right]$$

This establishes that the sum of the observed residuals during the class period is zero, and that E is the mean of the class period transactions' differences between actual and but-for estimated prices.

[28] Yan, X. and Su, X. (2009). Linear Regression Analysis: Theory and Computing, pp. 10-12.