A-583-854
Remand – CAFC 18-1229
Public Document
E&C/OVI: CV/FB

Final Results of Redetermination Pursuant to Court Order
Mid Continent Steel & Wire, Inc. et al. v. United States
Court No. 15-00213 (December 3, 2019)

I. SUMMARY

The Department of Commerce (Commerce) has prepared these final results of redetermination pursuant to the remand order of the United States Court of International Trade (CIT) in *Mid Continent Steel & Wire, Inc. et al. v. United States.* Consol. Court No. 15-00213 (December 3, 2019) (*Remand Order*). These final results of redetermination address one issue in the less-than-fair-value (LTFV) investigation of certain steel nails (nails) from Taiwan.[1] The Court of Appeals for the Federal Circuit (CAFC) affirmed, in part, and vacated, in part, the judgment of the CIT.[2] Specifically, the CAFC vacated and remanded the CIT's judgment upholding Commerce's use of a simple average when calculating the pooled standard deviation, a part of the Cohen's *d* test in Commerce's differential pricing analysis. As we explain below, on remand, we have complied with the CIT's *Remand Order* by providing further explanation of our methodology as requested in the *CAFC Holding.*

II. BACKGROUND

On May 20, 2015, Commerce published its *Final Determination*, in which it applied a differential pricing analysis to determine whether it could use an alternative comparison method

[1] *See Certain Steel Nails from the Republic of Taiwan: Final Determination of Sales at Less Than Fair Value*, 80 FR 28959 (May 20, 2015) (*Final Determination*), and accompanying Issues and Decision Memorandum (IDM).
[2] *See Mid Continent Steel & Wire, Inc. et al. v. United States*, 940 F. 3d 662 (Fed Cir. 2019) (*CAFC Holding*).

to calculate each respondent's estimated weighted-average dumping margin pursuant to section 777A(d)(1)(B) of the Tariff Act of 1930, as amended (the Act).[3] As part of this analysis, Commerce performed a "Cohen's *d* test" to determine whether prices differed significantly among purchasers, regions, or time periods. In response to comments from interested parties concerning whether a simple average instead of a weighted average should be used to calculate the pooled standard deviation in the calculation of the Cohen's *d* coefficient, Commerce explained in the *Final Determination* that the calculation of the pooled standard deviation based on a simple average of the variances determined for the test and comparison groups was appropriate because: (a) it is consistent with our normal practice; and (b) there is no statutory directive with respect to how Commerce should determine whether a pattern of prices that differ significantly exists, and it is a reasonable approach that affords predictability.[4] Moreover, Commerce further found that the use of a simple average was reasonable because the respondent's pricing behavior to each group would be weighted equally, and the magnitude of the sales to one group would not "skew the outcome."[5]

On March 23, 2017, the CIT sustained Commerce's use of a simple average to calculate the pooled standard deviation in the *Final Determination*.[6] On October 3, 2019, the CAFC vacated and remanded the CIT's judgment sustaining Commerce's calculation of the pooled standard deviation within the Cohen's *d* test, with instructions to remand to Commerce for

[3] *See Certain Steel Nails from Taiwan: Negative Preliminary Determination of Sales at Less Than Fair Value and Postponement of Final Determination*, 79 FR 78053 (December 29, 2014) (*Preliminary Determination*), and accompanying Preliminary Decision Memorandum (PDM) at 10-12. The *Final Determination* conclusively implemented the analysis set forth in the *Preliminary Determination*.
[4] *See Final Determination* IDM at Comment 2, pp. 28-29.
[5] *Id*. Commerce stated that … "{T}he Department finds it reasonable to use a simple average of the variances, in which the respondent's pricing behavior to each group will be weighted equally, and the magnitude of the sales to one group does not skew the outcome."
[6] *See Mid Continent Steel & Wire, Inc. v. United States, et al.*, Court No. 15-00213, Slip Op. 17-31 (CIT March 23, 2017).

further explanation regarding Commerce's decision to use a simple average to calculate the pooled standard deviation in the calculation of the Cohen's *d* coefficient.[7] On December 3, 2019, the CIT remanded to Commerce to provide further explanation consistent with the *CAFC Holding*.[8]

On March 3, 2020, Commerce released its draft results of redetermination and requested comments from interested parties.[9] On March 19, 2020, Commerce received comments and new factual information from PT Enterprise Inc., Pro-Team Coil Nail Enterprise Inc., Unicatch Industrial Co., Ltd., WTO International Co., Ltd., Zon Mon Co., Ltd., Hor Liang Industrial Corp., President Industrial Inc. and Liang Chyuan Industrial Co., Ltd. (collectively, PT);[10] and Commerce received comments from Mid Continent Steel & Wire, Inc. (MidContinent).[11] On April 9, 2020, Commerce provided parties an opportunity to submit factual information to rebut, clarify, or correct the factual information contained in the March 19, 2020 submission.[12] On May 4, 2020, MidContinent submitted new factual information in response to PT Comments and associated comments.[13] In these final results of redetermination, we address both PT's and MidContinent's comments.

[7] *See CAFC Holding* at 21.
[8] *See Remand Order.*
[9] *See* "Draft Results of Redetermination Pursuant to Court Order Mid Continent Steel & Wire, Inc. et al. v. United States Court No. 15-00213 (December 3, 2019)" dated March 3, 2020 (Draft Results of Redetermination).
[10] *See* Letter from PT, "Comments of Taiwan Plaintiffs on Draft Redetermination Pursuant to Court Remand, Court 15-00213 (CAFC No. 18-1229, October 3, 2019)" dated March 19, 2020 (PT Comments).
[11] *See* Letter from MidContinent, "*Mid Continent Steel & Wire, Inc. v. United States, et al.*, Consol. Court No. 15-00213 (CIT December 3, 2019): Comments on Draft Remand Redetermination," dated March 19, 2020 (MidContinent Comments).
[12] *See* Letter to Interested Parties, "Certain Steel Nails from the Republic of Taiwan - Information to Rebut, Clarify, or Correct," dated April 9, 2020.
[13] *See* Letter from MidContinent, "*Certain Steel Nails from Taiwan*: Factual Information Information {sic} To Rebut, Clarify, Or Correct the New Factual Information Contained in The Taiwan Plaintiffs' March 19, 2020 Submission," dated May 4, 2020 (MidContinent Rebuttal).

III. ANALYSIS

As an initial matter, we note that there is nothing in section 777A(d) of the Act that mandates how Commerce measures whether there is a pattern of prices that differs significantly or explains why the average-to-average (A-to-A) method or the transaction-to-transaction (T-to-T) method cannot account for such differences. Therefore, interpreting and implementing those statutory provisions[14] is fully within Commerce's authority under the Act.[15] As explained in the *Preliminary Determination*, as well as in various other proceedings,[16] Commerce's differential pricing analysis is reasonable, including the use of the Cohen's *d* test as a component in this analysis, and is in no way contrary to the law.

Section 777A(d)(1)(B) of the Act provides that Commerce may resort to an alternative comparison method based on average-to-transaction (A-to-T) comparisons if Commerce finds that two requirements are satisfied:

> (B) Exception. The administering authority may determine whether the subject merchandise is being sold in the United States at less than fair value by comparing the weighted average of the normal values to the export prices (or constructed export prices) of individual transactions for comparable merchandise, if--

[14] *See Koyo Seiko Co., Ltd. v. United States*, 20 F. 3d 1156, 1159 (Fed. Cir. 1994) ("The purpose of the antidumping statute is to protect domestic manufacturing against foreign manufacturers who sell at less than fair market value. Averaging U.S. prices defeats this purpose by allowing foreign manufacturers to offset sales made at less-than-fair value with higher priced sales. Commerce refers to this practice as 'masked dumping.' By using individual U.S. prices in calculating dumping margins, Commerce is able to identify a merchant who dumps the product intermittently—sometimes selling below the foreign market value and sometimes selling above it. We cannot say that this is an unfair or unreasonable result." (internal citations omitted)).

[15] *See Chevron U.S.A., Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837, 842-43 (1984) (*Chevron*) (recognizing deference where a statute is ambiguous and an agency's interpretation is reasonable); *see also Apex Frozen Foods Private Ltd. v. United States*, 37 F. Supp. 3d 1286, 1302 (*Apex*) (applying *Chevron* deference in the context of Commerce's interpretation of section 777A(d)(1) of the Act).

[16] *See, e.g.*, *Welded Line Pipe From the Republic of Korea: Final Determination of Sales at Less Than Fair Value*, 80 FR 61366 (October 13, 2015) (*Line Pipe from Korea*), and accompanying IDM at Comment 1; *Circular Welded Non-Alloy Steel Pipe From the Republic of Korea: Final Results of Antidumping Duty Administrative Review; 2012-2013*, 80 FR 32937 (June 10, 2015) (*CWP from Korea*), and accompanying IDM at Comments 1-2; and *Welded ASTM A–312 Stainless Steel Pipe From the Republic of Korea: Final Results of Antidumping Duty Administrative Review; 2013–2014*, 81 FR 46647 (July 18, 2016), and accompanying IDM at Comment 4.

> (i) there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time, and
> (ii) the administering authority explains why such differences cannot be taken into account using a method described in paragraph (1)(A)(i) or (ii).

The term "prices that differ significantly" connotes different prices where the difference has meaning, where it has or may have influence or effect, where it is noticeably or measurably significant, and where it may be beyond something that occurs by chance. In order to determine whether there is a pattern of prices that differ significantly among purchasers, regions, or time periods, Commerce employed its Cohen's *d* and ratio tests.[17] The purpose of the Cohen's *d* test is to examine whether the respondent's pricing behavior, as exhibited in its reported prices in the U.S. market, differed significantly between purchasers, regions, or time periods. This difference in prices is quantified through the calculation of the Cohen's "*d*" coefficient, which is one of a number of measures of "effect size."[18]

Dr. Jacob Cohen has employed effect size as part of his power analysis and defined the calculation of his *d* coefficient along with thresholds with which researchers could examine their application of the Cohen's *d* coefficient.[19] Dr. Paul Ellis explains the role of effect size by asking the question "So what? Why do this study? What does it mean for the man on the street?"[20] Dr. Ellis continues:

> A statistically significant result is one that is unlikely to be the result of chance. But a practically significant result is meaningful in the real world. It is quite

[17] *See Preliminary Determination* PDM at 11.
[18] *See*, generally, Ellis, Paul D., *The Essential Guide to Effect Sizes*; Cambridge University Press (2010) (*Ellis*), generally at 3-24. The first two chapters of Dr. Ellis' text are attached as Appendix I to the Draft Results of Redetermination.
[19] *See* Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Publishers (1988) (*Cohen*). The first two chapters of Dr. Cohen's text are attached as Appendix II to the Draft Results of Redetermination.
[20] *See Ellis* at 3; *see also Certain Frozen Warmwater Shrimp from the Socialist Republic of Vietnam: Final Results of Antidumping Duty Administrative Review, 2014-2015*, 81 FR 62717 (September 12, 2016), and accompanying IDM at 16-17.

> possible, and unfortunately quite common, for a result to be statistically significant and trivial. It is also possible for a result to be statistically nonsignificant and important. Yet scholars, from PhD candidates to old professors, rarely distinguish between the statistical and the practical significance of their results.[21]

In order to evaluate whether such a practically significant result is meaningful, Dr. Ellis states that this "implies an estimation of one or more effect sizes."[22]

> An effect size refers to the magnitude of the result as it occurs, or would be found, in the population. Although effects can be observed in the artificial setting of a laboratory or sample, effect sizes exist in the real world.[23]

Dr. Robert Coe[24] observes the same purpose for effect size, stating that "{e}ffect size is simply a way of quantifying the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference."[25] In order to answer Dr. Ellis' question "So What?", effect size is a widely used concept to understand the practical meaning of the difference between two groups of data.

As is clear from *Ellis*, there are numerous measures of effect size depending on the data to be evaluated. Dr. Ellis states:

> When we compare groups on continuous variables (*e.g.*, age, height, IQ), the usual practice is to gauge the difference in the average or mean scores of each group.[26]

> To calculate the difference between two groups we subtract the mean of one group from the other (M1 - M2) and divide the result by the standard deviation (SD) of the population from which the groups were sampled.[27]

---

[21] *See Ellis* at 3-4.
[22] *Id.* at 4.
[23] *Id.* at 4-5.
[24] *See* Coe, Robert, "It's the Effect Size, Stupid: What effect size is and why it is important," Paper presented at the Annual Conference of British Educational Research Association (Sept. 2002) (*Coe*) at 5, http://www.leeds.ac.uk/educol/documents/00002182.htm. Dr. Coe's paper is attached as Appendix III to the Draft Results of Redetermination.
[25] *Id.* at 1.
[26] *See Ellis* at 9.
[27] *Id.* at 10; *see also Cohen* at 20; and *Coe* at 2.

As stated by Dr. Cohen,[28] his *d* coefficient is defined as

$$d = \frac{m_A - m_B}{\sigma}$$

Thus, the significance of the difference is gauged using "the amount of variation found within a group," and that we can then "use this as a yardstick against which to compare the difference."[29] Dr. Ellis further indicates that the "only tricky part in this calculation is figuring out the population standard deviation."[30]

Of the numerous types of effect size discussed by Dr. Ellis, he identifies three such measures that are based on a difference of the means. Each of these differ in the calculation of the denominator, *i.e.*, the yardstick to gauge the significance of the difference of the means, which can be calculated as a "pool {of} the two standard deviations to calculate a Cohen's *d* index of effect size."[31] As discussed by Dr. Ellis, the yardstick used to gauge the significance of the differences in the means is based on the particular circumstances that are present in a given investigation. Although in Cohen's abstract model, where the "standard deviation" is "the standard deviation of either population (since they are assumed equal),"[32] the "standard deviation" can be an average of the standard deviations for the two groups (Cohen), the standard deviation of the control group (Glass), or a weighted average of the standard deviations (Hedge).[33] Dr. Cohen further provides that when the standard deviations of the two groups are not equal,[34]

[28] *See Cohen* at 20, where m_A and m_B are the mean values for groups A and B, respectively, and σ is the standard deviation of the value for the population.
[29] *See Coe* at 2.
[30] *See Ellis* at 10.
[31] *Id.*
[32] *See Cohen* at 20.
[33] *See Ellis* at 10.
[34] *See Cohen* at 43-44.

> …the formula requires the root mean square {of the standard deviations for each group}, that is, the square root of the mean of the two variances:[35]
>
> $$\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$
>
> The unequal variability need not affect the conception of *d* developed in Section 2.2. Given that there is a difference between σ_A and σ_B, we merely are using a kind of average within-population standard deviation to standardize the difference between means.[36]

Thus, even though examining the difference in the means is a widely accepted approach in measuring effect size, there are many "yardsticks" by which this difference may be gauged, including the square root of the simple average of the variances within each group.

As noted above, the purpose of Commerce's Cohen's *d* test is to determine whether U.S. prices differ significantly among purchasers, regions, or time periods – *i.e.*, do prices to each purchaser, region, or time period differ significantly from all other prices of the comparable merchandise. Although these are all prices in the U.S. market made by the respondent, this analysis requires that these prices be subdivided into separate distinct groups to consider separately whether the respondent's pricing behavior for sales to one specific group differs from its pricing behavior for all other sales. In other words, these prices, all of which are used to evaluate: 1) a respondent's pricing behavior in the U.S. market; and 2) whether the respondent is dumping, are now considered to represent two distinct pricing behaviors which may differ significantly. For the purpose of this particular analysis, Commerce finds that these two distinct pricing behaviors are separate and equally rational, and each is manifested in the individual

[35] The variance is the square of the standard deviation (*i.e.*, $v = \sigma^2$). Variance is defined as $v = \frac{\sum_{k=1}^{n}(x_k - \bar{x})^2}{n}$ where v is the variance, *k* is an index from 1 to *n*, the number of observations, x_k is the value of *x* for each observation (*i.e.*, the price), and $\bar{x}$ is the mean of x.

[36] *See Cohen* at 44.

prices within each group. Therefore, each warrants an equal weighting when determining the "standard deviation" used to gauge the significance of the difference in the means of the prices of comparable merchandise between these two groups. Because Commerce finds that each of these pricing behaviors are equally genuine when considering the distinct pricing behaviors between a given purchaser, region, or time period and all other sales, an equal weighting is justified when calculating the "standard deviation" of the Cohen's *d* coefficient. To do otherwise and use an average weighted by sales volume, sales value, or number of transactions would give preference to one pricing behavior over the other, and therefore would bias the "yardstick" by which Commerce measures the observed difference in prices between the test and comparison groups.

PT alleges that when calculating the "standard deviation," the use of a simple average (equal weighting) produces "skewed" results, but it is the alternative, a weighting based on a weighted average, which would produce unbalanced, non-skewed results.[37] Commerce disagrees. To be sure, the use of a simple versus weight average yields very different results. Nonetheless, it is not the results which validate the methodology, but the methodology which validates the results, and the methodology advocated by PT is not valid, and therefore, neither is the outcome which results from that methodology.

PT is correct that in calculating a weighted-average dumping margin, Commerce generally uses weighted averages. When averaging data within a specific group, a weighted average is usually appropriate, such as when calculating a weighted-average U.S. price, a weighted-average normal value, a weighted-average cost of manufacture, or a weighted-average dumping margin. Each of these examples is to calculate a mean value for sales which are part of the same, common group:

[37] *See* ECF No. 53, "Opening Brief of Defendants-Cross-Appellants" in *Mid Continent Steel & Wire Inc. v. United States*, CAFC No. 2018-1229, -1251 (filed August 16, 2018) (PT CAFC Brief) at 44-46.

- For U.S. prices and the normal value used in an A-to-A comparison,[38] the weighted average is for each *given* product[39] sold in the United States and the corresponding identical or most similar foreign like product when normal value is based on comparison market sale prices.
- The cost of manufacture is the weighted average of the individual input costs (*e.g.*, steel wire rod) for all individual products which share the same physical characteristics, and which define a *given* CONNUM in a proceeding.
- The weighted-average dumping margin is the weighted average of the individual comparison results for all of the U.S. sales of a *given* respondent.

As these examples clearly demonstrate, the calculation of a weighted average is bounded by the range of data for a given purpose for the calculation. One would not calculate a single weighted average of the U.S. sale prices of all products to make a comparison with normal value because those prices are determined based on factors for which there is no accounting for the differences which those factors create (*e.g.*, differences in the physical characteristics and corresponding costs of manufacture) which would skew, or distort, the result of that calculation.

In remanding the issue, the *CAFC Holding* refers to the numerical illustrations that PT provided in its brief to the CAFC,[40] in which PT included two examples to support its assertion that using a simple average rather than a weighted average distorts, or skews, the results. PT claims that the calculation is distorted on the basis that the sales in the test group change from being at significantly different prices from the comparison group to not being significantly

[38] *See* section 777A(d)(1)(A) of the Act; and 19 CFR 351.414(b)(1).
[39] Such "product" is defined by the product control number (CONNUM) which is defined by the physical characteristics of the merchandise under examination. A given CONNUM may in fact encompass numerous individual products which share the same physical characteristics.
[40] *See CAFC Holding* at 20; *see also* PT CAFC Brief at 44-46.

different. In PT's hypothetical calculation,[41] the standard deviation of the test group is 0.04 and the standard deviation of the comparison group is 0.4, resulting in a Cohen's *d* coefficient that is "passing" when using a simple average and "no passing" when using a weighted average (*i.e.*, the Cohen's *d* coefficient using a simple average is greater than the Cohen's *d* coefficient using a weighted average). Thus, PT asserts, the simple average results are skewed because the difference is found to be significant (*i.e.*, it passes the Cohen's *d* test threshold) when the weighted average results are not skewed because they do not find that the difference being significant. PT states that this is caused because "the lower the quantity, the bigger impact price changes within that low quantity group will have"[42] when Commerce uses a simple average. Commerce agrees with this statement because, as in PT's hypothetical example, the weight for the test (*i.e.*, small) group changes from 0.50 to 0.01.[43] The weight based on volume for the smaller group will always be less than 0.50, the weight when using a simple average. Further, when each of the two standard deviations in PT's example are switched with one another, then the relationship between the resulting Cohen's *d* coefficients also reverses, demonstrating that the change from a simple average to a weighted average does not result in a larger Cohen's *d* coefficient.

PT's hypothetical example also rests on the assumption that the standard deviation increases as the size of the group increases. PT claims that Commerce's rejection of this assumption is not reasonable, and the CAFC appears to have agreed with PT's claims in its holding on the basis of PT's examples;[44] however, the record demonstrates that in fact standard

[41] *See* PT CAFC Brief at 45.
[42] *See* PT CAFC Brief at 44.
[43] For a simple average, the weights for the test and comparison group will be 0.50 and 0.50, respectively (where the sum of the weights sum axiomatically to one). For PT's hypothetical example, the weight based on volume will be 10/1010 (1010 is the total volume for both groups) which is 0.01.
[44] *See CAFC Holding* at 20.

deviation really does not necessarily increase as the size of a group increases. First, as a general observation with all other things being equal, the variance and standard deviation will be smaller as the number of observations increases, as is evident in the definition of the variance noted above in footnote 30. Second, in the simple example where all prices are the same, the standard deviation will be zero, whether there are five sales or five thousand sales. Third, PT's own U.S. sales data support the proposition that the magnitude of the standard deviation of the price is not determined by the size of the group of sales. As shown by PT's own sale data in Appendix IV of the Draft Results of Redetermination, unchanged in these final results of redetermination and adopted herein, which reflects the calculations for first step of the Cohen's *d* test for PT's sales data and the standard deviation of the prices for each group of sales of the same comparable merchandise, the group with the highest standard deviation, 1.11069, only includes three sales, whereas the group with the largest number of sales has a standard deviation of 0.07473. In examining the data included in Appendix IV, which is sorted by increasing standard deviation values, there is no pattern between the magnitude of the standard deviations and either volume, value, or number of transactions. Thus, Commerce's rejection of PT's assumption underlying PT's hypothetical example purportedly demonstrating the distortion introduced into Commerce's analysis because of its use of a simple average is supported by the record, and Commerce continues to find PT's hypothetical, results-oriented example to be misleading and irrelevant.

Similarly, PT's example, based on its own data,[45] is cherry-picked to demonstrate the same meritless pattern as PT's flawed hypothetical example. This example is presented in Appendix V of the Draft Results of Redetermination, unchanged in these final results of redetermination and adopted herein, along with Commerce's calculations where the standard

[45] *See* PT CAFC Brief at 46.

deviations for the test and comparison groups are reversed. The difference in the calculation of the pooled standard deviation for these two results is the pairing of individual standard deviations and weights (*i.e.*, volume). The pooled standard deviation from PT's original calculations is 0.1107 using a simple average and 0.1327 using a weighted average. When the standard deviations (or the weights) are reversed, the pooled standard deviation is 0.1107 using a simple average and 0.0830 using a weighted average. Thus, irrespective of the quantity sold or the variation of the prices within the group, the importance of the pricing behavior to one group vis-à-vis all other groups remains constant. This illustrates that a simple average is predictable, whereas a weighted average introduces a variable relationship between the pricing behaviors of the two groups.

We note, though, when Commerce examines actual price data in its dumping analysis, switching the standard deviations or the weights is not something that would be done in the Cohen's *d* test. The data are what the data are. However, when examining PT's sales pricing data across different products, the same instability arises.

A careful examination of PT's example based on its own data also illustrates Commerce's argument that using a simple average provides "predictability" to the analysis. Appendix VI summarizes the PT sales data used in the Cohen's *d* test for a limited number of products. This summary includes the observation for a specific product and region used by PT in its example on page 46 of PT's brief to the CAFC, as discussed above, where the test group has a weight of 0.13 and the comparison group has a corresponding weight of 0.87.[46] For the data in

[46] *See* Appendix VI of the Draft Results of Redetermination, unchanged in these final results of redetermination and adopted herein, under "Weights" and "Volume" which lists the weights which PT propose to be used in the calculation of the pooled standard deviation based on the total sale volume (kg) for each of the test and comparison groups. Because the weights in Appendix VI are calculated relative to total sales volume, these two values will axiomatically sum to one.

this table, which includes approximately only five percent of the products sold by PT in the U.S. market, the weights for the test group for this same region range from 0.04 to 0.59, with the corresponding weights for the comparison group ranging from 0.96 to 0.41. Using a weighted average will mean that for one product, PT's pricing behavior to this specific group will determine 59 percent of the "yardstick" and for a different product PT's pricing behavior to that same group will determine only four percent of the "yardstick" with the remainder of the "yardstick," 41 percent and 96 percent, respectively, being determined by PT's pricing behavior to all other groups.

When Commerce attempts to compare PT's pricing behavior to a specific purchaser, region, or time period to its pricing behavior to all other purchasers, regions, or time periods, the basis for determining whether that behavior differs significantly should be consistent (imagine the resulting inconsistencies if every speed camera in a city were calibrated to a different scale, whereby if a car were traveling at 35 miles per hour (MPH), the various speed cameras would record a wide variation in speeds, *e.g.*, 27 MPH, 42 MPH, or 51 MPH). As with the example in Appendix V using PT's data, the use of a simple average of the variances to calculate the pooled standard deviation provides a consistent and predictable measure to determine whether the differences in the mean prices between two groups are significant.

The purpose of the Cohen's *d* test is to determine whether there exists a pattern of prices that differ significantly among purchasers, regions, or time periods. Commerce has used the widely accepted concept of "effect size" and the Cohen's *d* coefficient to determine whether the respondent's pricing behavior for the sales to a particular group differs significantly from the pricing behavior for sales discussed above. Weighting, by volume, the average of the variances for the test and comparison groups creates a wide variation, from 0.04 to 0.59, as to the

Filed By: Fred Baker, Filed Date: 6/16/20 4:42 PM, Submission Status: Approved

importance of the pricing behavior of the given group vis-à-vis all other groups when each pricing behavior is equally valid. In contrast, a simple average does not introduce such wide swings in the predominance of one of the pricing behaviors over the other, and is predictable because the importance given to each pricing behavior will be the same for all products. Thus, Commerce's use of a simple average addresses Commerce's expressed concern to use a consistent, predictable approach, where each pricing behavior is equally taken into account when gauging the significance of the difference in the mean prices of the test and comparison groups. Use of a weighted average, however, would inject an unpredictable, widely varying and seemingly random accounting of the two pricing behaviors when each of these pricing behaviors are equally representative of the prices to a given purchaser, region, or time period and the prices to all other purchasers, regions, or time periods.

We find that Commerce's use of a simple average is not only a reasonable approach but a more accurate and consistent measurement to use as a "yardstick" to determine if an exporter's pricing behavior as to a certain purchaser, region, or period of time differs significantly from that exporter's pricing behavior as to all other purchasers, regions, or periods of time. The choice of the method used by Commerce to determine the "yardstick," *i.e.*, the pooled standard deviation, is within the agency's discretion, as long as the method it selects is reasonable, and supported by substantial evidence on the record.[47] A simple average is a reasonable approach to calculate the pooled standard deviation for Commerce's differential pricing analysis because a simple average ensures that an exporter's pricing behavior in the test group is equally important to its pricing behavior in the comparison group, and, thus, whether an exporter's pricing behavior as to a

[47] *See Tri Union Frozen Prods. V. United States*, 163 F. Supp. 3d 1255, 1300 (CIT 2016). Commerce's methodological choice for determining whether a respondent has engaged in a particular pricing behavior warranting application of an alternative comparison methodology is precisely the type of determination where Commerce has been afforded significant discretion. .

certain purchaser, region, or period of time differs significantly from that exporter's pricing behavior as to all other purchasers, regions, or periods of time. Using a weighted average (whether by volume, value, or number of transactions) would improperly give preference to one pricing behavior over another, and this preference would vary wildly for the same purchaser, region or time period for different products. Commerce's approach removes this bias and instability, and ensures the consistency and objectivity in evaluating the pricing differences between purchasers, regions, or time periods, consistent with the purpose of the Cohen's *d* test.

Under different, albeit not entirely divergent, facts, the CIT affirmed Commerce's use of a simple average in its application of the Cohen's *d* test in *Stanley Works Langfang Fastening Sys. Co. v. United States.*[48] In its holding, the CIT found that Commerce's decision to use a simple average in Commerce's final determination was reasonable, since the aim in examining "targeted" dumping is to thwart unlawful pricing behavior.[49] The CIT held that a pooled standard deviation that gives equal weight to the pricing behavior toward a certain purchaser, or in a certain region or period of time, is a reasonable means to create a "benchmark" by which to measure the differences in a certain group of sales to the overall range of differences in the test and comparison groups.[50] The CIT observed that any price differences found using Commerce's Cohen's *d* test are relative to the variance of prices within the two groups, and thus are tailored to the individual pricing behaviors at issue.[51] The CIT then explained that on a case-by-case basis, "Commerce is able to contextualize the magnitude of the pricing differences between the test and

[48] *See Stanley Works Langfang Fastening Sys. Co. v. United States*, 333 F. Supp. 3d 1329, 1348 (CIT August 13, 2018), (citing *Mid Continent Steel & Wire, Inc. v. United States, et al.*, 219 F. Supp. 3d 1326, 1342 (CIT March 23, 2017)).
[49] *Id.*
[50] *Id.*
[51] *Id.*

comparison groups, which helps it to determine whether there is a pattern of prices that differ significantly among purchasers, regions, or periods of time."[52] As a result, the CIT held that "the pricing behavior in each group is of equal importance, and therefore, using a simple average to calculate the pooled standard deviation (thereby giving equal weight to the standard deviations in both groups) is reasonable."[53]

Similarly, in this litigation, we continue to find on remand that using a simple average (*i.e.*, equal weights) to calculate the pooled standard deviation is appropriate to give equal weight to each of the pricing behaviors as manifested in the prices included in the test group and in the prices included in the comparison group. To use an average weighted by sales volume, sales value, or number of transactions would give preference to one pricing behavior over the other, and therefore would bias the "yardstick" by which Commerce measures the observed difference in prices between the test and comparison groups, and this bias would vary dramatically for each product sold to a given purchaser, region, or time period.

Therefore, in accordance with the *Remand Order*,[54] which stems from the *CAFC Holding* vacating the CIT's judgment and directing the CIT to remand this issue to Commerce, for these final results of redetermination, Commerce is complying with the Court's *Remand Order* to provide further explanation of our calculation of the Cohen's *d* coefficient.

[52] *Id.*
[53] *Id.*
[54] *See Remand Order* at 18-21.

IV. SUMMARY AND ANALYSIS OF INTERESTED PARTIES' COMMENTS ON THE DRAFT RESULTS

Comments from PT:

PT asserts that Commerce has failed to fulfill the CAFC's direction to further explain its use of a simple average to calculate the pooled standard deviation, which cannot be defended on any basis "*i.e.*, reasonable, factual, actual, statistical or mathematical."[55] Accordingly, Commerce must calculate the pooled standard deviation based on a weighted average of the variances of the test and comparison groups, which is "objective, consistent, effective and fair" and "specifically accounts for any undue influence that could occur due to the price differences between the two groups."[56]

PT states that Commerce's goal "is to determine whether the difference between the mean of the test group and the mean of the comparison groups {sic} 'differ significantly from the prices of *all other sales of the comparable merchandise*.'"[57] However, PT asserts that Commerce's approach fails to "accomplish Commerce's goals and accordingly, is unreasonable."[58] PT argues that "simple averaging produces distortive and inaccurate results" and that "simple averaging does not reflect the pricing behavior of the CONNUM, Commerce's stated goal."[59] Further, PT claims that Commerce's revised example to counter its own example from its brief to the CAFC[60] does not disprove PT's point.

PT argues that Commerce has failed to follow the various formulas to calculate the pooled standard deviation presented in *Cohen*, *Ellis* and *Coe*. These references, PT claims,

[55] *See* PT Comments at 2.
[56] *Id.*
[57] *Id.* at 4 (citations omitted, emphasis in original).
[58] *Id.* at 5.
[59] *Id.* at 6.
[60] *Id.* at 3-6.

support its argument that the pooled standard deviation should be calculated as a weighted average, rather than a simple average, of the variances of the test and comparison groups.

In *Cohen*, the general formulation of the *d* coefficient is defined as:

$$d = {(M_A - M_B)}/{SD}$$

where M_A and M_B are the population means and SD is the standard deviation of either population (since they are assumed equal).[61] PT then states:

> Cohen explained that his basic formula was based on the following assumption:
>
> > In the formal development of the t distribution for the difference between two independent means, the assumption is made that the populations sampled are normally distributed and that they are of homogeneous (*i.e.*, equal) variance.
>
> As such, Cohen's basic formula for computing the Cohen's d value is applicable to a situation where the test and control groups have: (1) the same standard deviation or variance; and (2) the same size.[62]

PT then highlights two exceptions to Dr. Cohen's assumptions when (1) variances differ but the sample size is the same and when (2) both the variances and sample sizes differ. Under the first exception, *Cohen* provides that the pooled standard deviation would be the simple average of the variances of the two groups; however, this still requires that the sizes of the two groups be identical.[63] Under the second exception, *Cohen* provides that the pooled standard deviation would be an average of the variances weighted by the number of observations in each group and reflecting two fewer degrees of freedom (*i.e.*, the denominator is the sum of the number of observations minus two).[64] PT concludes that "where both variances and sizes of test

[61] *Id*. at 10.
[62] *Id*. (citing *Cohen* at 19 and 20).
[63] *See* PT Comments at 11 (citing *Cohen* at 44).
[64] *See* PT Comments at 11-12 (citing *Cohen* at 67).

and control groups are different, Dr. Cohen himself reasoned that the statistical method for computing the pooled standard deviation should be based upon a weight averaging instead of simple averaging of the two variances."[65]

PT extends its interpretation of *Cohen* to weighting in general where generic weights take the place of the number of observations[66] from *Cohen*:

$$S_{pooled} = \sqrt{\frac{W_a \times S_a^2 + W_b \times S_b^2}{W_a + W_b - 2}}$$

where PT emphasizes that in "the present context of sale transactions of goods, *the analogue for numerical strength or count of people is the total weight of goods encompassed by the test and control groups*."[67] Thus, PT asserts that its extension of Dr. Cohen's formula results in the PT's proposed formula:

$$S_{pooled}^2 = \frac{W_a}{W_a + W_b} \times S_a^2 + \frac{W_b}{W_a + W_b} \times S_b^2$$

Thus, PT's proposed approach is supported by *Cohen* whereas Dr. Cohen "expressly rejects the simple averaging pooling methodology used by Commerce."[68]

PT asserts that *Coe* agrees with *Cohen*. *Coe* presents the equation when both the variances and sample sizes differ:[69]

$$SD_{pooled} = \sqrt{\frac{(N_e - 1) \times SD_e^2 + (N_c - 1) \times SD_c^2}{N_e + N_c - 2}}$$

[65] *Id.* at 12.
[66] *Id.*
[67] *Id.* (emphasis in the original).
[68] *Id.* at 13.
[69] *Id.* (citing *Coe* at 6).

PT concludes that *Coe*, like *Cohen*, supports weighting the variances by each group's respective sample size, and extends the term for sample size, N_e and N_c, to denote "the respective weight of the two groups."[70]

In *Ellis*, PT finds three alternatives for calculating the denominator of a measure of effect size, two of which – Cohen's *d* and Hedges' *g* – use a pooled standard deviation.[71] *Ellis* repeats Dr. Cohen's formula noted above where both the standard deviations and sample sizes differ. PT also notes Ellis description of Hedges' *g*,[72] which "Ellis expressly recommends that 'if the groups are dissimilar in size,' Hedges' g formula that weight each group's standard deviation by its sample size' is preferable…"[73]

PT summarizes that Commerce has failed to follow the specific formulas presented in *Cohen*, *Coe* and *Ellis* in its application of the Cohen's *d* test. For data "characterized by disparate variances and sizes of the test and control groups," all three references "expressly prescribe weighing the individual variances by their respective sizes."[74]

Further, the testimonies provided by Dr. Sobel and Dr. Hubel support PT's proposed approach of averaging the variances given the "disparate variances and sizes of the two groups."[75] Thus, all five sources are in agreement that it is appropriate to weight average the variances in the calculation of the denominator of the measure of effect size.

PT also asserts that weighting the average of the variances is supported by algebraic and statistical principles. PT sets forth the idea that when a population is composed of two distinct

[70] *Id.* at 14.
[71] *Id.* (citing Ellis at 10). The third method, Glass' *d*, defines the denominator as the standard deviation of the comparison group.
[72] *Id.* at 15. The formula for Hedges' *g* is the same as the formula presented above from *Coe*.
[73] *Id.* (citing Ellis at 10).
[74] *Id.* at 16.
[75] *Id.*

groups of data, that the variance of the population is derived from two components: (1) the variance attributed to differences within each group and (2) the variance attributed to the difference in the means of the two groups. PT argues that the second component should be eliminated from the analysis because otherwise the results would be "distorted by circularity."[76] PT hypothesizes that this is what Commerce has viewed as "one group skewing the other."[77] However, PT's proposed formula eliminates this distortion by limiting the component of the populations variance that is due to the difference in the means of the two groups.

PT notes that Dr. Huber's testimony supports the use of a weighted average of the variances because the underlying average prices and variances used in the calculation of the Cohen's *d* coefficient are themselves weighted averages.[78] Further, the weights used in these calculations are appropriately the sales quantities when the variable under examination is the unit price of each sale.[79] PT further provides "public domain literature" supporting the use of a weighted average.[80]

PT argues that Commerce has failed to provide an objective rationale for use of a simple average of the variances to calculate the pooled standard deviation.[81] The formula used by Commerce is limited to the situation where the sample sizes, $n_1$and n_2, are equal. When the sample sizes are unequal, Commerce must use the alternative formula noted above that weights the variances by the size of each sample, or in the situation of the use of sales, weights the variances by the sale quantities of each group.

[76] *Id.* at 17.
[77] *Id.* (internal citations omitted).
[78] *Id.* at 18-19.
[79] *Id.* at 18.
[80] *Id.* at 19.
[81] *Id.* at 20-21.

Filed By: Fred Baker, Filed Date: 6/16/20 4:42 PM, Submission Status: Approved

PT rejects Commerce's statement that the pricing behavior, for the purposes of the differential pricing behavior, is "separate and equally rational"[82] to justify equal weighting of the variances in the calculation of the pooled standard deviation.[83] Commerce has failed to define what is meant by "equally rational," especially when the purpose of the Cohen's *d* test is to determine whether the respondent's pricing behavior differs significantly. Accordingly, Commerce has prejudged the issue by equating "disparately distributed (in terms of variance and size) groups"[84] and has potentially distorted the results of the analysis.

Further, even if such pricing behavior is "separate and equal," it is unclear how this supports equal weighting of the variances. PT posits that:

> …in {the} ordinary course of sales, equal pricing behavior should attach to individual unit of sales (Kg) and not to group of sales made to myriad customers and time periods in the context of a regional test group / comparison group and likewise for the other two kinds of test / comparison groups.
>
> When each Kg of sales is treated on an equal footing in pooling their variances, the resulting pooled standard deviation accords with a weighted averaging methodology.[85]

Otherwise, PT claims, "the Cohen's d values are internally inconsistent."[86]

Thus, Commerce's "separate and equal" rational does not support its use of a simple average of the variances, which is "inherently unequal, insofar as they have different sizes and different standard deviations."[87] Commerce's approach is no fairer "than claiming the school children of different colors should be educated in 'separate and equal' facilities."[88]

As demonstrated in PT's examples summarized from the proceeding before the CAFC:

---

[82] *Id.* at 22.
[83] *Id.* at 22-24.
[84] *Id.* at 22.
[85] *Id.* at 23.
[86] *Id.* at 24.
[87] *Id.* at 25.
[88] *Id.* at 24.

> …in a factual scenario in which the test group and control group are composed of significantly different quantities, and significantly different standard deviations in prices, reliance of a simple average to determine whether the pricing difference is significant makes no sense and is objectively unreasonable and inaccurate.[89]

The CAFC agreed that Commerce's approach "led to facially absurd and inaccurate results"[90] and remanded the issue to Commerce. Commerce cannot rebut the absurdity of the results and simply claimed that the results are irrelevant. First, based on *Cohen*, *Ellis* and *Coe* all agree that use of a simple average is not appropriate when the sample sizes differ. Second, the testimony from Dr. Sobel and Dr. Huber also state that Commerce's use of a simple average does not apply in the situation where the sample sizes differ. Dr. Huber asserts that Commerce's approach "is flawed because it is mathematically incorrect, inconsistent, arbitrary, ineffective at achieving its objectives, unpredictable, and (therefore) unreliable."[91] Dr. Huber supports this conclusion with his declaration that:

> The consistency principle (¶ 24) leads ineluctably to the conclusion that *the sales volume-weighted standard deviation of the group residuals is the only legitimate value of the standard unit to use in the Cohen's d calculation* for the purpose of meeting Commerce's statutory objective of finding significant patterns of price variation.[92]

Third, and lastly, Commerce's "pronouncement that 'it is not the results which validate the methodology, but the methodology which validates the results" fails to consider that non-sensical results cannot be valid. Dr. Huber's analysis has confirmed that Commerce's approach does not make sense and the CAFC agreed with PT's example. Form cannot be elevated over substance when results are absurd.

[89] *Id.*
[90] *Id.*
[91] *Id.* at 25 (quoting Dr. Huber Testimony at ¶ 74).
[92] *Id.* at 26 (quoting Dr. Huber Testimony at ¶ 35) (emphasis in the original).

Commerce's rejection of PT's hypothetical example is fundamentally flawed. First, PT rejects Commerce's assertion that PT's analysis is skewed because of the results, *i.e.*, where the simple average finds that the difference in the prices for PT's example are significant, but that use of a weighted average finds that the difference is not significant. PT states that the results have nothing to do with its position, rather PT claims that it is Commerce's methodology that is skewed. The use of a simple average to calculate the pooled standard deviation places undue weight for the variance for small sale quantities, rather than treating all quantities equal, and produces "absurd" results, "*i.e.*, aggregated difference in prices based on a simple average of variances of the test group and control group cannot be reliable when such pooled standard deviation is vastly different than the population standard deviation, which indicates normal differences in prices in the CONNUM."[93]

Second, Commerce's attempt to refute this example (*i.e.*, "when each of the two standard deviations in PT's example are switched with one another"[94]) is meaningless. Commerce's calculations are wrong and contrary to the actual results. Third, Commerce is wrong in its reason to reject PT's example because the standard deviation does not necessarily increase as the size of the group increases. Commerce's statement that the magnitude of the standard deviation is not dependent upon the volume, value or number of observations has nothing to do with PT's position, but it is a reason to use a weight average so that "the resulting pooled variance accurately mirrors the contributions of its constituent groups."[95] Fourth, Commerce is wrong in its claim that PT "cherry picked" the example from its own data in this investigation, and PT has

[93] *Id.* at 27.
[94] *Id.* (quoting the Draft Results of Redetermination).
[95] *Id.* at 29-30.

demonstrated that Commerce simply does not understand the meaning of its switching the standard deviations between the test and comparison groups.

Commerce has failed to explain how varying weights for the test and comparison groups based on region for various CONNUMs are "unfair or skewed."[96] These weights are directly proportional to the sale quantities of the test and comparison groups, and the two pricing behaviors can thus not simply be assumed to be equal.

PT rejects Commerce's assertion that use of a simple average ensures consistency, predictability and objectivity and that the weighted average biases or skews the resulting pooled standard deviation. Commerce's goal of equality "is contrary to its goal of calculating an accurate pooled variance"[97] because a test group of 10 kilograms is not equal to a comparison group of 1000 kilograms. Rather than the weighted average approach treating these two groups as equal biases the smaller group, creates instability "and ensures inconsistency and the absence of consistency."[98]

Dr. Sobel concurs, stating that "Commerce seems to mean anything other than equal weighting they wish to use, and Commerce never says that they mean by bias or skew, nor does their usage have anything to do with the well-defined meanings of these terms in the statistical literature."[99] Dr. Sobel continues:

> In statistics, bias refers to whether the average value of an estimator of a population quantity has the same value as the population quantity or not. In the former instance, the estimator is said to be "unbiased", otherwise it is said to be "biased". The term "skew" in statistics is used to refer to the asymmetry of a distribution, with the normal distribution, which is symmetric about its mean (and where the mean equals the median equals the mode) often used as a baseline for

[96] *Id.* at 31.
[97] *Id.* at 33.
[98] *Id.*
[99] *Id.* at 34 (quoting Dr. Sobel Testimony).

> comparison. It is difficult to see how the term "skewed" that Commerce uses has anything to do with its meaning in statistics.[100]

Further, the results of Commerce's analysis for all of PT's U.S. sales validates PT's proposed use of a weighted average to calculate the pooled standard deviation. PT delineates small and large groups by the threshold of 20 observations, a commonly used measure for group size in statistics. The results of PT's analysis show that a larger proportion of small test groups pass the Cohen's d test than large test groups. This proves that smaller group sizes create larger passing rates than large group sizes. Further, PT claims that "the normal expectation in statistics is that larger test groups should be better able to demonstrate an effect than a smaller group,"[101] contrary to the results of Commerce's analysis based on the use of a simple average to calculate the pooled standard deviation.

Comments from MidContinent:

MidContinent supports Commerce's continued use of a simple average of the variances of the test and comparison groups to calculate the pooled standard deviation.[102] MidContinent avers that use of a simple average rather than a weighted average removes the instability and bias introduced by using weights that vary and give preferences to one pricing behavior over another. As a result, Commerce has ensured "consistency and objectivity" in its analysis.

MidContinent rejects PT's arguments that Commerce must use a weighted average to calculate the pooled standard deviation. MidContinent emphasizes that an analysis must account for the context in which the analysis is being made.[103] MidContinent provides examples where

[100] *Id.* (quoting Dr. Sobel Testimony).
[101] *Id.* at 36.
[102] *See* MidContinent Comments at 2-3.
[103] *See* MidContinent Rebuttal at 3-6

the selection of the statistic used to characterize a given set of data, *e.g.*, the distinction between using the mean versus the median of a population.

MidContinent rejects PT's assertion that using quantity as the weight in averaging the variances is appropriate. It states that nowhere does Dr. Cohen and other researchers rely on sale quantities, but rather on the number of observations present in each group.[104] MidContinent rejects PT's logic that elsewhere Commerce uses sale quantity as the weight and, as here, "the rules of arithmetic determine the answer."

MidContinent highlights the possible distortions which could arise from the use of quantity rather than number of observations as a weight for averaging the variances in the calculation of the pooled standard deviation. MidContinent describes in detail the roll of the number of observations and sample size in the analysis,[105] emphasizing that "sample sizes always matter."[106] To refute the relevance of PT's approach, MidContinent compares the equations used by Dr. Cohen and proposed by PT to calculate the Cohen's *d* coefficient, as well as the importance of the sample size in the power analysis developed by Dr. Cohen. MidContinent concludes that PT's substitution of sale quantity for sample size irrationally changes the calculations of statistical significance and effect size, including also correction for bias inherent within the Cohen's *d* coefficient.

MidContinent frames the instant methodological issue as two decisions.[107] First, if a researcher determines that weighting is appropriate, then it is appropriate to weight the variances by the number of observations in each sample group. Second, if the number of observations and the variances differ, then it is appropriate to not weight the variances and to calculate a simple

[104] *Id.* at 6-8.
[105] *Id.* at 8-16.
[106] *Id.* at 12.
[107] *Id.* at 16.

Filed By: Fred Baker, Filed Date: 6/16/20 4:42 PM, Submission Status: Approved

average (*i.e.*, to weight the variances equally). MidContinent emphasizes that there are two options for Commerce, either (1) to weight the variances based on the number of observations in each group, or (2) to calculate a simple average of the variances. MidContinent posits that the preference based on Dr. Cohen's own reference is to rely on a simple average of the variances because the mean prices of the two groups constitute "an abstract effect" of each group that is independent of other characteristics of each group, such as the frequency of the sales associated with each group.[108]

MidContinent also distinguishes Commerce's analysis as an "observational" study as opposed to an "experimental" study.[109] MidContinent explains that in an experimental study, the researcher controls the groups under analysis, whereas in an observational study, the characteristics of the data under examination define the groups in the study. MidContinent notes that in a study of cigarette smokers, whether the participants are in the smoker or non-smoker group is determined based on the behavior of the individual and is not under the control of the researcher. Similarly, the test and comparison groups of pricing data used by Commerce in the Cohen's *d* test is determined by the data itself and is not defined by Commerce as part of its analysis. Accordingly, the weighting of the variance will shift importance to the group which happens to have the larger frequency of sales when the focus of Commerce's analysis is the difference in the mean prices of the two groups. MidContinent concludes that this shifting of weighting from the group with the smaller number of sales to the group with the larger number of sales should be describe as "over-weighting," or what has been labeled thus far as "skewed" results.[110]

[108] *Id.* at 20.
[109] *Id.* at 20-23.
[110] *Id.* at 23.

Lastly, MidContinent disagrees with PT's rejection of Commerce's use of the terms "rational" and "genuine" to describe the pricing behavior observed to each group defined in the Cohen's *d* test.[111] MidContinent argues that both terms are accepted terms understood in an economic context. Citing Paul Samuelson and William Nordhaus, MidContinent asserts that "rational consumer behavior is that which maximizes the consumer's benefit or well-being."[112] Further, MidContinent asserts that a "rational" firm is a company which maximizes profits, stating that:

> …the neoclassical model that has dominated economic thought over the last several decades "assumes that firms maximize their profits from producing and selling goods and services, and households maximize their utility (or satisfaction) from consuming goods and services. Economic actors are assumed to be self-interested and 'rational', meaning that people generally make logical decisions that produce the best outcomes for themselves."[113]

Furthermore, the term "genuine," as in a genuine price, and by extension a genuine pricing behavior, "reflects its cost, quality and/or competitive environment."[114] Thus, Commerce's use of the concepts "rational" and "genuine" are well defined and well supported in the study of economics.

Commerce's Position:

For purposes of these final results, as discussed herein, we have provided extensive analysis and explanation of our decision to use a simple average instead of a weighted average to

[111] *Id.* at 23-26.
[112] *Id.* at 25.
[113] *Id.* (quoting Neva Goodwin, Jonathan M. Harris, Julie A. Nelson, Brian Roach, and Mariano Torras. Microeconomics in Context, 4th ed. Routledge (sample available at http://www.bu.edu/eci/files/2019/05/MIC_4e_Ch7.pdf)). MidContinent did not provide screenshots or a "last visited" date of this source in its comments.
[114] *Id.* at 26 (quoting Mark Armstrong and Yongmin Chen, Discount Pricing, Oxford University Discussion Paper Series, #605, May 2012 at 2 (available at https://www.economics.ox.ac.uk/materials/papers/5819/paper605.pdf)). MidContinent did not provide screenshots or a "last visited" date of this source in its comments.

determine the pooled standard deviation to calculate the Cohen's *d* coefficient as part of the differential pricing analysis in the *Final Determination* of the LTFV investigation.

At the outset, we repeat the basic principles discussed above regarding Dr. Cohen's basic expression, in which he defines his *d* coefficient as the standard deviation of the underlying population of data. As discussed above, the significance of the difference is gauged using "the amount of variation found within a group," and we can then "use this as a yardstick against which to compare the difference."[115] As discussed by Dr. Ellis, the yardstick used to gauge the significance of the differences in the means is based on the particular circumstances that are present in a given investigation. Although in Cohen's abstract model, where the "standard deviation" is "the standard deviation of either population (since they are assumed equal),"[116] the "standard deviation" can be an average of the standard deviations for the two groups (Cohen), the standard deviation of the control group (Glass), or a weighted average of the standard deviations (Hedge).[117] Dr. Cohen further provides that when the standard deviations of the two groups are not equal,[118]

> …the formula requires the root mean square {of the standard deviations for each group}, that is, the square root of the mean of the two variances:
>
> $$\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$
>
> The unequal variability need not affect the conception of *d* developed in Section 2.2. Given that there is a difference between σ_A and σ_B, we merely are using a kind of average within-population standard deviation to standardize the difference between means.[119]

[115] *See Coe* at 2.
[116] *See Cohen* at 20.
[117] *See Ellis* at 10.
[118] *See Cohen* at 43-44.
[119] *See Cohen* at 44.

Thus, even though examining the difference in the means is a widely accepted approach in measuring effect size, there are many "yardsticks" by which this difference may be gauged, including the square root of the simple average of the variances within each group.

As noted above, the purpose of Commerce's Cohen's *d* test is to determine whether U.S. prices differ significantly among purchasers, regions, or time periods – *i.e.*, do prices to each purchaser, region, or time period differ significantly from all other prices of the comparable merchandise. Although these are all prices in the U.S. market made by the respondent, this analysis requires that these prices be subdivided into separate distinct groups to consider separately whether the respondent's pricing behavior for sales to one specific group differs from its pricing behavior for all other sales. In other words, these prices, all of which are used to evaluate: 1) a respondent's pricing behavior in the U.S. market; and 2) whether the respondent is dumping, are now considered to represent two distinct pricing behaviors which may differ significantly. For the purpose of this particular analysis, Commerce finds that these two distinct pricing behaviors are separate and equally rational, and each is manifested in the individual prices within each group. Therefore, each warrants an equal weighting when determining the "standard deviation" used to gauge the significance of the difference in the means of the prices of comparable merchandise between these two groups. Because Commerce finds that each of these pricing behaviors are equally genuine when considering the distinct pricing behaviors between a given purchaser, region, or time period and all other sales, an equal weighting is justified when calculating the "standard deviation" of the Cohen's *d* coefficient. To do otherwise and use an average weighted by sales volume, sales value, or number of transactions would give preference to one pricing behavior over the other, and therefore would bias the "yardstick" by which Commerce measures the observed difference in prices between the test and comparison groups.

As discussed herein, Commerce has considered two possible approaches to calculate the pooled standard deviation as part of the Cohen's *d* test: (1) use an average of the variances weighted by the sale quantity for the test and comparison groups, as proposed by PT; or (2) use a simple average of the variances as in the draft results of redetermination and the underlying *Final Determination*. Commerce finds that the first option, as proposed by PT, is not supported by the context of Commerce's analysis, and it appears that the only support for this option offered by PT over that used in the Draft Results of Redetermination is that the results are beneficial to PT in reducing the number of test groups that pass the Cohen's *d* test, thus lowering PT's estimated weighted-average dumping margin. Option (2) reflects the application of Dr. Cohen's basic concept to the context of Commerce's Cohen's *d* test and which is supported by concepts presented by both PT and MidContinent. As discussed below, for these final results of redetermination, Commerce has continued to calculate the pooled standard deviation as the simple (*i.e.*, equally weighted) average of the variances of the test and comparison groups to calculate the pooled standard deviation in the Cohen's *d* test.

Commerce recognizes that "context" as discussed by MidContinent is an inescapable factor in any analysis, and therefore is also relevant to our analysis. Indeed, *Cohen*, *Ellis* and *Coe* are replete with examples of the relevance of context in data analysis. Without context, the results of an analysis may be misinterpreted contrary to the underlying purpose of the analysis.

The overall context of Commerce's use of the Cohen's *d* test, and indeed its entire dumping analysis, is based on market economy principles. PT superficially dismisses Commerce's discussion of "rational" and "genuine" behavior by claiming that Commerce has failed to define these terms. MidContinent rebuts that the terms "rational" and "genuine" are

widely understood in economics.[120] As a foundation of economic theory, an individual's behavior is governed by decisions which maximizes the benefits to that individual and that such behavior is rational. Commerce agrees with MidContinent and recognizes that such benefits are governed by a company's goals which may include such benefits as profit, market share or shareholder return. As such, a company, including PT, pursues defined, established, corporate goals, and the purpose of its decisions, including its pricing behavior, is to maximize the realization of these goals. General economic theory holds that such goals are rational and that the company's behavior is a genuine reflection of such goals.

The statute directs Commerce in an LTFV investigation to determine "whether the subject merchandise is being, or is likely to be, sold in the United States at less than its fair value"[121] and to "determine the estimated weighted average dumping margin for each exporter and producer individually investigated."[122] Accordingly, Commerce has calculated a single estimated weighted-average dumping margin for PT which takes into account all of PT's sales in the U.S. market during the period of investigation. As such, PT's pricing behavior is summarized by a single result which quantifies the extent of its dumping. Such a determination is based on the factual record of the investigation, and Commerce is not required to divine the reasons behind such factual information, such as the reasons behind its pricing behavior or the ultimate benefit accruing to the company. In other words, Commerce accepts that the company acts rationally based on the economic principles described above, and conducts its dumping analysis accordingly based on the factual record manifested by that rational behavior.

[120] *See* MidContinent Rebuttal at 23-26.
[121] *See* section 735(a)(1) of the Act.
[122] *See* section 735(c)(1)(B)(i) of the Act.

In the context of section 777A(d)(1)(A) of the Act, the statute requires Commerce in an LTFV investigation to "determine whether the subject merchandise is being sold in the United States at less than fair value" by comparing normal value with U.S. price using either the A-to-A method or the T-to-T method, where the A-to-A method is generally applied,[123] as in this investigation. When applying the A-to-A method, one concern is that dumping will be masked (or "targeted") such that dumping will be undercounted.[124] Such masking can arise either explicitly when higher U.S. prices offset lower U.S. prices or implicitly within the average U.S. prices of the A-to-A method, or explicitly when aggregating individual comparison results and offsets are granted for non-dumped sales.[125] In order to take account of masked dumping, the statute provides that Commerce may use an alternative comparison method based on the A-to-T method.[126]

In the context of an LTFV investigation, section 777A(d)(1)(B) of the Act permits Commerce to use an alternative comparison method, based on the A-to-T method, when two criteria have been satisfied, the first being that "there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time." Interpreting and implementing this statutory requirement is fully within Commerce's discretion.[127] Commerce's differential pricing analysis is objective and reasonable, including the use of the Cohen's *d* test as a component in this analysis, and is in no way contrary

[123] *See* 19 CFR 351.414(c)(1).
[124] *See* Statement of Administrative Action Accompanying the Uruguay Round Agreements Act, H.R. Doc. 103-316, vol 1 (1994) (SAA) at 842-843; *see also United States Steel Corp. v. United States*, 621 F. 3d 1351, 1362–63.
[125] *See Union Steel v. United States*, 713 F. 3d 1101, 1108-1109; *see also Certain Hot-Rolled Steel Flat Products From Japan: Final Determination of Sales at Less Than Fair Value and Final Affirmative Determination of Critical Circumstances*, 81 FR 53409 (August 12, 2016) and accompanying IDM at 30-33.
[126] *See* section 777A(d)(1)(B) of the Act.
[127] *See Chevron* (recognizing deference where a statute is ambiguous and an agency's interpretation is reasonable); *see also Apex*, 37 F. Supp. 3d 1286, 1302 (applying *Chevron* deference in the context of Commerce's interpretation of section 777A(d)(1) of the Act).

to the law. As part of its differential pricing analysis, Commerce uses the Cohen's *d* and ratio tests to evaluate whether there exists a pattern of prices that differ significantly.

As discussed in the Draft Results of Redetermination, the purpose of the Cohen's test is to determine whether the difference in U.S. prices for comparable merchandise between a given purchaser, region or time period and all other purchasers, regions or time periods, respectively, is significant. The Cohen's *d* coefficient, a measure of effect size, quantifies the "practical significance"[128] of the difference in the means of two groups of data. As prescribed by the statute, Commerce must examine the respondent's pricing behavior between purchasers, regions and time periods. Thus, even though the respondent's pricing behavior rationally reflects that respondent's established goals to maximize its benefits, in the context of section 777A(d)(1)(B), Commerce is directed to consider the respondent's pricing behavior to each purchaser, region or time period separately from its pricing behavior to other purchasers, regions or time periods. The U.S. prices to each purchaser, region or time period are the product of the respondent's rational pricing behavior, and equally and genuinely reflect the respondent's overall goals to maximize its own benefits. Simply because Commerce subdivides the respondent's U.S. prices into separate and distinct groups in the context of the Cohen's *d* test does not diminish the assumption that each of these U.S. prices equally reflect the genuine pursuit of the goals of the respondent. Therefore, in the context of the Cohen's *d* test, the U.S. prices to each purchaser, region or time period separately and equally represent the respondent's pricing behavior, which itself is determined by the respondent's rational economic goal of maximizing the benefits accruing to the respondent.

[128] *See Ellis* at 3-4; and *Coe* at 1.

PT asserts that Commerce's reasoning for "'separate and equally rational' fairs no better than claiming that school children of different colors should be educated in 'separate and equal' facilities."[129] Commerce fails to find any relevance in the relationship asserted by PT between the context of the Cohen's *d* test and the abhorrent "separate *but* equal" practice of racial discrimination in education in the United States.

Dr. Cohen, in his text *Statistical Power Analysis for the Behavioral Sciences*, incorporates effect size and statistical analysis as components of his power analysis.[130] Similarly, as posited by Dr. Ellis, a researcher must be as interested in the practical significance of the researcher's findings as well as in the statistical significance of the data on which that research is based.[131] Further, Dr. Coe's paper, as well as the arguments in PT Comments and MidContinent Rebuttal address aspects of both effect size and statistical significance. However, in the context of the Cohen's *d* test and Commerce's dumping analysis as a whole, Commerce's analysis is based on all of the U.S. sales data for the respondent to each purchaser, region and time period during the period of investigation.[132] Commerce does not sample the respondent's U.S. sales data used in the Cohen's *d* test, and the calculated means and variances of the U.S. prices are the actual values of the entire population of U.S. sales and are not estimates of those values. [133] As such, Commerce has appropriately described its Cohen's *d* test as not involving sample data from a larger population, but as an analysis of the entire universe of data used in the dumping analysis. Although the sampling method is critical[134] to establish the validity of a

[129] *See* PT Comments at 24.
[130] *See Cohen* at 4.
[131] *See Ellis* at 3-4.
[132] *See Ellis* at 5 ("The best way to measure an effect is to conduct a census of an entire population").
[133] *See Cohen* at 2 ("Results from a random sample drawn from a population will only approximate the characteristics of the population.").
[134] *See*, *e.g.*, *Cohen* at 6 ("The reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value. It is necessarily an estimated value in practice, since the

researcher's results, the statistical significance and inferences, including sample size, associated with research based on sampled data are not relevant in Commerce's Cohen's *d* test.

Similarly, PT objects to Commerce's use of the terms "skew" and "bias," arguing that these terms have specific meanings in statistics.[135] Commerce agrees that these terms have defined meaning in statistical studies. However, as statistical terms, "skew" and "bias," as defined by PT, are not relevant to Commerce's analysis, just as with sample size, statistical significance and statistical inference, because the data in the analysis is not sampled. Alternatively, MidContinent uses the term "over-weighted" to describe the imbalance created when each variance is not weighted equally when averaged. This is only half of the story in that PT's proposed weighting using sale quantity involves not only over-weighting but also corresponding under-weighting. Commerce finds that each of these terms, in generic usage, describe the question of the appropriate weights to use when averaging the variances to calculate the pooled standard deviation given the context of the Cohen's *d* test. Thus far in this proceeding, the inappropriate approach is "skewed," "biased," "over-weighted," or "absurd and inaccurate;"[136] the appropriate approach is "objective, consistent, effective and fair."[137]

The purpose of Commerce's Cohen's *d* test establishes the context and the appropriate arithmetic for the analysis. Commerce agrees with PT's assertion that "arithmetic matters," and that this is related to the context of the analysis. Whether to weight average or simple average data in a given analysis is generally an arithmetic issue that is also subject to the context and the

population value is generally unknown. Depending upon the statistic in question, and the specific statistical model on which the test is based, reliability may or may not be directly dependent upon the unit of measurement, the population value, and the shape of the population distribution. However, it is always dependent upon the size of the sample.").

[135] *See* PT Comments, Exhibit 1 at 5 ("In statistics, bias refers to whether the average value of an estimator of a population quantity has the same value as the population quantity or not." "The term 'skew' in statistics is used to refer to the asymmetry of a distribution, with the normal distribution, which is symmetric about its mean…").

[136] *See* PT Comments at 24.

[137] *See* PT Comments at 2.

underlying purpose of the calculation. For example, an average U.S. price for the A-to-A comparison method is the average of individual unit prices (*i.e.*, sale value divided by sale quantity), weighted by sale quantity, for a group of *n* sales is calculated as the sum of the product of each unit price and the sale quantity, divided by the sum of the sale quantities. Thus, the numerator is the sum of individual sale values, and the numerator and the denominator resolve to the total sale value and the total sale quantity. Further, based on the purpose of this average (*i.e.*, for comparison to normal value), the individual sales are limited by the factors related to the A-to-A comparison. Similarly, the weighted-average dumping margin[138] is calculated as the average of individual dumping margins (individual amount of dumping divided by individual U.S. sale value) weighted based on the U.S. value of each sale. Here the context and the arithmetic is determined based on the statutory definition.

For these two weighted average calculations, the arithmetic supports the use of an appropriate weight to average the underlying values because the product of the averaged values and the weight represents the extended, total value of sales and the extended, total amount of dumping, respectively, and the weighted-average value is defined as the total value of sales or the total amount of dumping, divided by the total amount of the weights, *i.e.*, total sale quantity or total sale value, respectively. In the context of calculating an average U.S. price for a comparison with normal value using the A-to-A method, first, the statute requires a weighted average, and given the arithmetic, weighting by sale quantity is appropriate. The context also supports the use of a weighted average dumping margin because that is its statutory definition.[139] Further, first for the calculation of the average prices, and then for the averaging of the dumping margins (which includes the U.S. prices), the underlying U.S. sales are each associated with the

[138] *See* section 771(35)(B) of the Act.
[139] *See* section 771A(35)(B) of the Act.

same pricing behavior of the respondent and thus the other parameters associated with the individual U.S. sales are relevant to each other, including each of the sale quantities or the U.S. prices.

In contrast, in other situations, a simple average may be more appropriate for the context to reflect the purpose of the analysis. For example, a different context in which Commerce calculates an average of individual dumping margins is the calculation of a rate for all-other producers and exporters in an LTFV investigation. Consistent with section 735(c)(5) of the Act, when all of the estimated weighted-average dumping margins determined for the mandatory respondents are zero, *de minimis*, or based entirely on facts available, then Commerce's practice is to calculate the "all-others" rate as the simple average of the individual dumping margins alleged in the underlying petition which gave rise to the investigation.[140]

In this situation, the context is that the dumping margins alleged in the petition each individually represents evidence of dumping, separate from any of the factors which were used to determine these individual dumping margins. As MidContinent explains, a parameter for a "natural population" can be viewed as an "abstract effect"[141] where representation in a larger "super population" is based on its presence in the larger group rather than on other information from the "natural population," and this membership in the "super population" is weighted equally in summarizing the information for the larger population. Similarly, the dumping

[140] *See Acetone From Spain: Preliminary Affirmative Determination of Sales at Less Than Fair Value, and Preliminary Determination of No Shipments*, 84 FR 37990 (August 5, 2019) ("Commerce has preliminarily determined the estimated weighted-average dumping margin for the individually examined respondent {} entirely under section 776 of the Act. Consequently, pursuant to section 735(c)(5)(B) of the Act, Commerce's normal practice under these circumstances has been to calculate the all-others rate as a simple average of the alleged dumping margins from the petition."); unchanged in *Acetone From Spain: Final Determination of Sales at Less Than Fair Value, and Final Determination of No Shipments*, 84 FR 56166 (October 21, 2019) ("As discussed in the Preliminary Determination, Commerce based the selection of the all-others rate on the simple average of the two dumping margins calculated for subject merchandise from Spain alleged in the petition, in accordance with section 735(c)(5)(B) of the Act").

[141] *See* MidContinent Rebuttal at 18-20.

margins alleged in the petition equally contribute the evidence of dumping during the period of investigation, and as such these "abstract" rates are equally weighted to calculate the weighted-average dumping margin for all other producers and exporters rather than calculating this rate based on a weighted average of the dumping margins alleged in the petition. Thus, in this situation, the context differs from the context when calculating a weighted-average dumping margin based on a respondent's cumulative U.S. sales and the aggregated results of comparing U.S. prices to normal values because the factors underlying the petition dumping margins were not relevant to one another.

PT argues that *Cohen*, *Ellis* and *Coe* support its proposal to calculate the pooled standard deviation as the average the variances of the test and comparison groups, weighted by the relevant sale quantities of each group. Commerce disagrees with PT's conclusion because PT has failed to take into account the context of each of the academic references vis-a-vis Commerce's Cohen's *d* test.

PT begins its argument based on the equations in *Cohen*,[142] *Ellis*,[143] and *Coe*,[144] which are presented in different forms, but are in fact consistent (for shorthand, we refer to these equations as a single "reference equation").[145] Commerce agrees that the reference equation is informative for the calculation of the pooled standard deviation in the context of the Cohen's *d* test. However, PT does not recognize that the context of the reference equation is an analysis of sampled data, as evidenced by the fact that the "weights" represent the sample sizes of the data,

[142] *See Cohen* at 67 (equation 2.5.2).

[143] *See Ellis* at 26 (endnote 8) and 27 (endnote 9). These two equations are identical in that $SD = \frac{\sum(x-\bar{x})^2}{(n-1)}$.

[144] *See Coe* at 6.

[145] PT's restatements of *Cohen*'s equation 2.5.2 at the top of page 12 and Ellis' equation in endnote 8 at the top of page 15 each erroneously omits the numerators the reduction in the degrees of freedom associated with sampled data while each of the denominators reflects this reduction. Nonetheless, *Cohen*'s equation 2.5.2, correctly "expressed," and Ellis' equation in endnote 8, correctly "reduced," is identical to the equation in *Coe*.

and by the fact that these weights, representing the degrees of freedom, are each one less than the number of observations in each sample because each variance is measured relative to the estimated sample mean and not the actual population mean.

PT then extrapolates the reference equation to an equation analogous to the simple average used in the Cohen's *d* test, but which explicitly includes weights, *i.e.*, PT's proposed approach to calculate the pooled standard deviation in the Cohens' *d* test. PT nominally recognizes that there is some kind of discrepancy between the reference equation and its proposed approach, and, perhaps unknowingly, PT erroneously dismisses this difference between the reference equation and the context of the Cohen's *d* test:

> Coe's pooled standard deviation formula weighs the variance by a factor of (N-1), which for large sizes typically involved in behavioral sciences experiments approximates N.[146]

Certainly, as PT states, the value of (N-1) will approach the value of (N) as (N) increases in value; however, this does not explain the difference between the reference equation and PT's proposed weighted formula. This is not an issue involving the limits of an independent variable (*e.g.*, where $N \rightarrow \infty$) but a question concerning degrees of freedom which arises from a principle of mathematics. The difference is that the context of the reference equation and the discussions in *Cohen*, *Coe*, and *Ellis* reflect sampled data where the calculation of the (estimated) variance includes the dependent variable of the estimated mean, and the context of PT's proposed weighted formula and the Cohen's *d* test where the calculation of the (actual) variance includes the independent variable of the actual mean. PT again clearly fails to recognize the context of Commerce's Cohen's *d* test and the basis for that analysis by not understanding the role of sampling in *Cohen*, *Coe* and *Ellis*.

[146] *See* PT Comments at 14.

PT concludes that:

> …for situations characterized by disparate variances and sizes of the test and control groups, the Cohen, Ellis and Coe papers attached to the Draft Results do not support {Commerce}'s draft decision to calculate the pooled variance by taking the simple average of the variances of the test group and the control group. To the contrary, in this specific context, all three experts – Cohen, Coe and Ellis – expressly prescribe weighing the individual variances by their respective sizes.[147]

Commerce disagrees because PT has failed to account for the context of Commerce's Cohen's *d* test, which differs significantly from the context of the discussion in *Cohen* and reference equation sourced from *Cohen*, *Coe* and *Ellis*. First and foremost, PT fails to account for the context of the data used in Commerce's Cohen's *d* test. The Cohen's *d* test used in the differential pricing analysis does not involve a sample of a respondent's U.S. sales but rather the entire universe, *i.e.*, population, of the available data. As such, sample size is not a relevant parameter in the context of the Cohen's *d* test and PT's reliance on it in rejecting the simple average is misplaced. Therefore, as discussed above, the discussion of the concepts in *Cohen*, *Coe* and *Ellis* are theoretically applicable and informative to Commerce's use of effect size in the Cohen's *d* test; however, to the extent that specific formulas and criteria involve statistical sampling and associated inferences, these references are not directly contextually equated with the analysis in the Cohen's *d* test. PT's rejection of Commerce's use of the simple average based on *Cohen*'s use of a simple average when sample sizes are equal is contextually without merit.

Second, PT fails to account for the context of the statute, which requires that Commerce examine whether prices differ significantly between purchasers, regions or time periods. How to weight data in any analysis is a critical decision and must reflect the context of the analysis and be arithmetically reasonable. The appropriate question is what weight should be used in any calculation, not whether one should weight the data as posed by PT. Equal, unitary weighting

[147] *See* PT Comments at 16.

(*i.e.*, a simple average), weighting using an associated variable, or even asymmetrical weighting (as in Glass' Δ) are all potential approaches.

As discussed above, the calculation of an average sale price is weighted by the sale quantity because arithmetically this results in the ratio of the total value and the total quantity. Further, in the context of a dumping analysis, this average is calculated on a product-specific basis to effectuate the most accurate comparison of U.S. price with normal value. Also discussed above is the calculation of a weighted-average dumping margin for a respondent. Again, arithmetically, this calculation results in the total amount of dumping being divided by the total value of U.S. sales, and the context of this calculation is all U.S. sales by the respondent because each of those U.S. sales are perceived to be part of the respondent's pricing behavior as a whole.

The context of the Cohen's *d* test is that Commerce examines the prices to individual purchasers, regions or time periods as directed by the statute. As such, Commerce calculates an average price and variance of the prices for each group that is weighted by the sale quantity. Based on the context of the statute, Commerce considers the pricing behavior to each purchaser, region or time period for the purpose of the Cohen's *d* test as distinct and independent pricing behaviors, *i.e.*, separate, and equally genuine in that each independent pricing behavior reflects the respondent's pursuit of rational economic goals. Accordingly, calculating an average price and a variance of the price is based on individual transactions which within each group are defined to be associated with one another and reflective of the same pricing behavior, and thus other characteristics of the individual sales, such as sale quantity are relevant within the group. However, because the statute creates separate groups that are to be examined independently, for the Cohen's *d* test the only parameters which relate between these groups are the mean and variance of the underlying individual sales. Other characteristics of the individual sales in one

group are not relevant to the individual sales in a different group. As such, the mean and variance of the price in each group constitute an abstract effect, as described by MidContinent, which exclusively define the independent pricing behavior in each group, and the additional details of the underlying transactions are not relevant between different groups for this analysis. Similarly, when comparing price between different purchasers, regions or time periods in the context of section 777A(d)(1)(B)(i) of the Act, the CAFC has found that Commerce does not need to understand the intent underlying the pricing behavior within each of the test and comparison groups in the context of examining whether there exists a pattern of prices that differ significantly.[148]

This is analogous to the situation discussed above where dumping margins alleged in a petition are simple-averaged when calculating the all-others rate in an LTFV investigation. In this situation, the parameters surrounding the individual dumping margins alleged in the petition are not relevant and the alleged dumping margins themselves are seen each as an "abstract effect." The quantity associated with the U.S. price is not relevant in the context of section 735(c)(5)(B) of the Act. Nor are such factors as the producer, exporter, U.S. customer relevant. Thus, it is appropriate to base this calculation on a simple average rather than on a weighted average basis.

MidContinent argues that "Cohen himself recognizes instances where a researcher may prefer to use a simple average over a weighted average."[149]

> Cohen draws a distinction between an experiment, where the creation of each condition (*e.g.*, experiment and control arms) is the "systematic artificial creation

[148] *See JBF RAK LLC v. United States*, 790 F. 3d 1358, 1368 (CAFC June 2015) (stating that for a targeted dumping allegation that "Section 1677f-1(d)(1)(B) does not require Commerce to determine the reasons why there is a pattern of export prices for comparable merchandise that differs significantly among purchasers, regions, or time periods, nor does it mandate which comparison methods Commerce must use in administrative reviews."); and *Borusan Mannesmann Boru Sanayi Ve Ticaret A.S. v. United States*, 608 Fed. Appx. 948 (CAFC June 2015) (agreeing with *JBF RAK LLC*).

[149] *See* MidContinent Rebuttal at 18.

> of the experimenter" and a study of natural populations, whose "properties as populations include their relative sizes in their combined superpopulation."
>
> In the context of natural populations, such as the universe of U.S. sales analyzed in an antidumping proceeding, Cohen explains that the researcher may choose to weight each group according to their frequency in the superpopulation, or to weight groups equally.[150]

PT argues that a weighted average is appropriate, given the examples provided by Commerce, "because both the test and control groups fall under the same CONNUM and therefore are part of the 'same, common group.'"[151] Commerce agrees that each group is defined by comparable merchandise, as explicitly stated in the description of the Cohen's d test.[152] However, PT continues to ignore the statutory context that each group in the Cohen's *d* test is also defined by purchaser, region or time period. As such, it is the intersection of comparable merchandise and purchaser, region or time period which defines the groups used in the Cohen's *d* test.

PT also posits that "{Commerce}'s equal pricing behavior rationale behind attaching equal weight to variances, if applied properly, actually supports the weighted averaging methodology because in ordinary course of sales, equal pricing behavior should attach to individual unit of sales (Kg)…."[153]

However, PT's argument that "equal pricing behavior should attach to individual unit of sales," actually gets to the heart of the statutory context of the Cohen's *d* test:

> Test and comparison groups are created artificially for differential pricing analysis and do not have an independent existence of their own in practical terms.[154]

[150] *See* MidContinent Rebuttal at 18 (internal citations omitted); *see also* MidContinent Rebuttal, Exhibit 1 at ¶¶ 35-39.
[151] *See* PT Comments at 21.
[152] *See Certain Steel Nails from Taiwan: Negative Preliminary Determination of Sales at Less Than Fair Value and Postponement of Final Determination*, 79 FR 78053 (December 29, 2014), and the accompanying PDM at 10-11.
[153] *See* PT Comments at 23.
[154] *Id*.

Filed By: Fred Baker, Filed Date: 6/16/20 4:42 PM, Submission Status: Approved

As discussed above, the statutory context of the Cohen's *d* test is to examine whether prices differ significantly between purchasers, regions or time periods. Although a respondent's rational pricing behavior determines its U.S. prices, such prices may vary between these statutory groupings. As such, these groupings can be seen as arbitrary, as well as abstract, and exist only to consider whether there exists a pattern of prices that differ significantly. It is because of this statutory context that the U.S. price means and variances exist as abstract parameters.

At various points in its comments, PT has labelled the results of Commerce's use of a simple average of the variances to calculate the pooled standard deviations as "absurd,"[155] "internally inconsistent, making them arbitrary and unpredictable,"[156] and that "{Commerce}'s methodology led to facially absurd and inaccurate result{s} was a critical factor in determining whether the methodology is reasonable."[157] Additionally, PT summarizes its invalidation of the simple average, stating that "a methodology which yields results which do not make sense cannot be valid."[158] PT supports this claim with numerous examples from Dr. Huber's testimony, emphasizing that "{Commerce} cannot elevate form over substance."[159] Based on its hypothetical examples, PT also asserts that the results of using a weighted average are "accurate and reasonable."[160] Further, in response to Commerce's counterfactual examples, PT parrots Dr. Sobel's statement that the "examples themselves do not constitute a real justification for Commerce's preferences."[161]

However, the same can be said with respect to PT's examples. Thus, PT must agree with Commerce's rational that "it is not the results which validate the methodology, but the

[155] *Id.* at 21.
[156] *Id.* at 24.
[157] *Id.* at 24-25.
[158] *Id.* at 26.
[159] *Id.*
[160] *Id.* at 21.
[161] *Id.* at 34.

methodology which validates the results." Even if PT's examples were to be considered to be determinative as to the reasonableness or unreasonableness of the calculation of the Cohen's *d* coefficient, PT has failed to define any quantified or qualified threshold or meaning that the results based on a simple average are absurd, inaccurate, inconsistent or arbitrary. The only qualitative factor which PT cites is whether the results for a given test group pass or do not pass the Cohen's *d* test. In other words, whether the calculated Cohen's *d* coefficient is greater than, or less than, the "large" threshold established by Dr. Cohen, and which the CAFC has held to be reasonable.[162] Further, it is not even that the calculated Cohen's *d* coefficient calculated using a simple average of the variances appears to be unusually large, *i.e.*, some multiples larger than the large threshold, or that PT's calculated alternative is substantially lower than the large threshold. In PT's hypothetical example,[163] the calculated Cohen's *d* coefficients are 0.8795 and 0.62809 based on a simple and weighted average, respectively. Neither result appears "facially absurd or inaccurate" except that each lies on the opposite side of the 0.8 large threshold. Therefore, not even PT's examples can support that the results of Commerce's use of a simple average are absurd, arbitrary or inaccurate.

Accordingly, Commerce continues to find that it is reasonable and appropriate in the context of the Cohen's *d* test to the use of a simple average of the variances of the test and comparison groups to calculate the pooled standard deviation, the denominator of the Cohen's *d* coefficient. Commerce has addressed the concerns expressed by the CAFC and provided further explanation to support its use of a simple average. Further, PT has failed to demonstrate that its proposed use of a weighted average is more appropriate or "objective, consistent, effective and fair."

[162] *See CAFC Holding*, 940 F. 3d at 673.
[163] *See* PT Comments at 3.

V. FINAL RESULTS OF REDETERMINATION

Pursuant to the CIT's *Remand Order*, Commerce provided a more extensive analysis and explanation of our decision to use a simple average instead of a weighted average to determine the pooled standard deviation to calculate the Cohen's *d* coefficient as part of the differential pricing analysis in the *Final Determination* of the LTFV investigation.

6/16/2020

X

Signed by: JEFFREY KESSLER
Jeffrey I. Kessler
Assistant Secretary
for Enforcement and Compliance

Filed By: Fred Baker, Filed Date: 6/16/20 4:42 PM, Submission Status: Approved