A-583-854
Remand
Slip Op. 23-45
POI: 04/01/2013 – 03/31/2014
Public Document
E&C/OVI: CV/GM

***Mid Continent Steel & Wire, Inc. v. United States*, Consol. Court No. 15-213 (CIT April 3, 2023)**
Certain Steel Nails from Taiwan

FINAL RESULTS OF REDETERMINATION PURSUANT TO COURT REMAND

I. SUMMARY

The U.S. Department of Commerce (Commerce) has prepared these final results of redetermination pursuant to the opinion and remand order of the U.S. Court of International Trade (CIT) in *Mid Continent Steel & Wire, Inc. v. United States*, 628 F. Supp. 3d 1316 (CIT April 3, 2023) (*Mid Continent VI*). These final results of redetermination concern one principal issue in the less-than-fair-value (LTFV) investigation of certain steel nails (nails) from Taiwan: Commerce's use of a simple average rather than a weighted average when calculating the denominator of the effect size (*i.e.*, the "Cohen's *d* coefficient"), a part of the Cohen's *d* test used in Commerce's differential pricing analysis.[1]

In *Mid Continent VI*, the CIT remanded Commerce's third final results of redetermination concerning the *Final Determination*. Specifically, the CIT found that Commerce did not comply with the direction of the U.S. Court of Appeals for the Federal Circuit (Federal Circuit) "to provide reasonable justification for departing from what the acknowledged literature teaches."[2] The CIT found that Commerce misinterpreted the Federal Circuit's mandate and that

[1] *See Certain Steel Nails from 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 VI* at 15 (citing *Mid Continent Steel & Wire, Inc. v. United States*, 31 F.4th 1367, 1381 (Fed Cir. 2022) (*Mid Continent V*)).

Commerce's practical justifications were "unsupported but not unsupportable."[3] As explained below, on remand, we have complied with *Mid Continent VI* by providing further detailed explanations to support Commerce's use of the simple average and why a simple average is reasonable to use in Commerce's analysis.

On July 7, 2023, Commerce released the Draft Redetermination to interested parties.[4] On July 10, 2023, the Taiwan Plaintiffs requested a one-week extension to submit comments in response to Commerce's Draft Redetermination.[5] On July 13 and 14, 2023, Commerce granted the Taiwan Plaintiffs' extension request, in full.[6] On July 21, 2023, Mid Continent[7] and the Taiwan Plaintiffs[8] submitted comments.[9] Commerce addresses the comments from the interested parties below.

---

[3] *See Mid Continent VI* at 14 (citing *Mid Continent V*, 31 F.4th at 1379 ("Commerce has not offered an adequate explanation of why {equal rationality and genuineness} support {} the particular step Commerce must justify … {a}nd in any event, Commerce has not provided a reasonable explanation for this predictability assertion")).

[4] *See* Draft Results of Redetermination Pursuant to Court Remand, Mid Continent Steel & Wire, Inc. v. United States, Consol. Court No. 15-00213, Slip Op. 23-45 (CIT April 3, 2023), dated July 7, 2023 (Draft Redetermination).

[5] *See* "Taiwan Plaintiffs' Request for Extension of Time to File Comments on Draft Redetermination Pursuant to Court Remand, Court 15-00213, Slip Op. 23-45 (CIT April 3, 2023)," dated July 10, 2023.

[6] *See* Commerce's Letter, "Extension of Deadline for Submission of Comments on Draft Results of Redetermination," dated July 13, 2023; and Commerce's Letter, "Second Extension of Deadline for Submission of Comments on Draft Results of Redetermination," dated July 14, 2023.

[7] Mid Continent Steel & Wire, Inc. (Mid Continent) is a domestic interested party, and was the petitioner in the LTFV investigation.

[8] The Taiwan Respondents, foreign producers or exporters of the subject merchandise, are PT Enterprise Inc., Pro-Team Coil Nail Enterprise Inc., Unicatch Industrial Co., Ltd., WTA International Co., Ltd., Zon Mon Co., Ltd., Hor Liang Industrial Corp., President Industrial Inc., and Liang Chyuan Industrial Co., Ltd. (collectively, Taiwan Respondents or Taiwan Plaintiffs). PT Enterprises, Inc. (PT) was a mandatory respondent in the less-than-fair-value investigation.

[9] *See* Mid Continent's Letter, "Comments on Draft Results of Redetermination Pursuant to Court Remand in Mid Continent Steel & Wire, Inc. v. United States, Slip Op. 23-45 (CIT 2023)," dated July 21, 2023 (Mid Continent Comments); Taiwan Respondents' Letter, "Comments of Taiwan Plaintiffs on Draft Results of Redetermination Pursuant to Court Remand Mid Continent Steel & Wire, Inc. v. United States, Consol. Court No. 15-213, Slip Op. 23-45 (April 3, 2023), Antidumping Duty Investigation on Certain Steel Nails from Taiwan (A-583-854)," dated July 21, 2023 (Taiwan Respondents' Comments).

II. BACKGROUND

On May 20, 2015, Commerce published its *Final Determination*, in which it applied a differential pricing analysis to determine whether to use an alternative comparison method to calculate each respondent's estimated weighted-average dumping margin as permitted by section 777A(d)(1)(B) of the Tariff Act of 1930, as amended (the Act).[10] As part of this analysis, Commerce performed a "Cohen's *d* test" to determine whether U.S. prices differ 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 denominator (*i.e.*, the "pooled standard deviation") 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; (b) there is no statutory directive with respect to how Commerce should determine whether a pattern of prices that differ significantly exists; and (c) it is a reasonable approach that affords predictability.[11] 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."[12]

On March 23, 2017, the CIT sustained Commerce's use of a simple average in the *Final Determination*.[13] On October 3, 2019, the Federal Circuit vacated and remanded the CIT's

---

[10] *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 at 10-12. The *Final Determination* conclusively implemented the analysis set forth in the *Preliminary Determination*.
[11] *See Final Determination* IDM at Comment 2, pp. 28-29.
[12] *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."

judgment sustaining Commerce's calculation of the Cohen's *d* coefficient, with instructions to remand to Commerce for further explanation regarding Commerce's decision to use a simple average to calculate the denominator of the Cohen's *d* coefficient.[14] On December 3, 2019, the CIT remanded the case to Commerce in accordance with *Mid Continent III*.

On June 16, 2020, Commerce issued its Second Redetermination after addressing the comments and new factual information placed on the record by Commerce and the interested parties, Taiwan Respondents and Mid Continent.[15] To support its continued reliance on a simple average, Commerce explained that a simple average provided predictability, the pricing behavior of each group was equally rational and equally genuine, and weighting would give more inappropriate weight to the pricing behavior of one group over the other. On January 8, 2021, the CIT sustained Commerce's *Second Redetermination* in *Mid Continent IV*.[16] In particular, the CIT held that Commerce's choice to use a simple average for the pooled standard deviation was reasonable.

The Taiwan Respondents appealed the CIT's judgment in *Mid Continent IV* to the Federal Circuit. On April 21, 2022, the Federal Circuit vacated *Mid Continent IV* and remanded the issue to Commerce, finding that Commerce had not adequately justified its adoption of a simple average to calculate the denominator of the Cohen's *d* coefficient.[17] On November 10, 2022, Commerce issued its Third Redetermination after addressing the Federal Circuit's remand

---

[13] *See Mid Continent Steel & Wire, Inc. v. United States*, 219 F. Supp. 3d 1326 (CIT 2017) (*Mid Continent I*). The CIT remanded the calculation of PT's general and administrative (G&A) expense ratio, which Commerce recalculated in the first redetermination. The CIT affirmed Commerce's recalculation of PT's G&A expense ratio in *Mid Continent Steel & Wire, Inc. v. United States*, 273 F. Supp. 3d 1161 (CIT 2017) (*Mid Continent II*).
[14] *See Mid Continent Steel & Wire, Inc. v. United States*, 940 F. 3d 662 (Fed Cir. 2019) (*Mid Continent III*).
[15] *See Final Results of Redetermination Pursuant to Court Order Mid Continent Steel & Wire, Inc. et al. v. United States*, Court No. 15-00213 (CIT December 3, 2019), dated June 16, 2020 (Second Redetermination), available at https://access.trade.gov/resources/remands/15-00213.pdf.
[16] *See Mid Continent Steel & Wire, Inc. v. United States*, 945 F. Supp. 3d 1298 (CIT 2021) (*Mid Continent IV*).
[17] *See Mid Continent V*.

order in *Mid Continent V*,[18] and comments submitted by the interested parties, Taiwan Respondents and Mid Continent.[19] On April 3, 2023, the CIT remanded Commerce's Third Redetermination for further explanation or reconsideration consistent with *Mid Continent V*, addressed herein.

III. ANALYSIS

Section 777A(d)(1)(B) of the Act provides that Commerce may resort to a comparison method based on the average-to-transaction comparison method when two requirements have been met: (1) there exists a pattern of prices that differ significantly for comparable merchandise among purchasers, regions or time periods (the pattern requirement); and (2) one of the standard comparison methods under section 777A(d)(1)(A) of the Act cannot account for such differences (the meaningful difference requirement). To examine these two requirements, Commerce introduced a differential pricing analysis in 2013.[20] In its examination of the pattern requirement, Commerce uses first the "Cohen's *d* test" and then the ratio test. The Cohen's *d* test examines whether the sale prices to a given purchaser, region, or time period differ significantly from the sale prices of comparable merchandise to other purchasers, regions, or time periods, respectively. The ratio test, which is not at issue in this litigation, assesses the extent of the prices which are found to differ significantly and, thus, determines whether there exists a pattern of prices that differ significantly.

The Cohen's *d* test is based on a measure of effect size, the concept of which was expounded by Dr. Jacob Cohen in his textbook on *Statistical Power Analysis for the Behavior*

[18] *See Mid Continent Steel & Wire, Inc. v. United States*, Consol. Court No. 15-00213 (CIT 2022).
[19] *See Final Results of Redetermination Pursuant to Court Order Mid Continent Steel & Wire, Inc. v. United States*, Court No. 15-00213 (CIT June 14, 2022), dated November 10, 2022 (Third Redetermination), available at https://access.trade.gov/resources/remands/21-1747.pdf.
[20] *See, generally, Mid Continent V*, 31 F.4th at 1370-73.

Sciences.[21] Effect size is a measure of the practical significance of the difference in two means.[22] The effect size, the "Cohen's *d* coefficient," is the ratio of the difference in the means, divided by the "standard deviation," *i.e.*, the variance in the underlying data.[23] It is this denominator of the Cohen's *d* coefficient, the "standard deviation," that is the subject of this litigation.[24]

In the *Final Determination*, Commerce calculated the denominator of the effect size as the "simple average" of the standard deviations[25] of the test group[26] and the comparison group.[27] In its challenges to Commerce's approach, the Taiwan Respondents have argued that Commerce must use a weighted average rather than a simple average to calculate the denominator of the effect size. Commerce has rejected the arguments made by the Taiwan Respondents, although the Federal Circuit has twice found Commerce's explanations to be inadequate to justify reliance on a simple average. Most recently, in *Mid Continent V*, the Federal Circuit stated:

> We hold that Commerce has not adequately justified its adoption of simple averaging for the Cohen's *d* denominator. Commerce has departed from the methodology described in all the cited statistical literature governing Cohen's *d*, but it has not justified that departure as reasonable.[28]

---

[21] *See, generally,* Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Publishers (1988) (*Cohen*), 19-74. The first two chapters of *Cohen* are included in Appendix II to Commerce's Draft Results of Redetermination Pursuant to Court Order, *Mid Continent Steel & Wire, Inc. et al. v. United States*, Court No. 15-00213, (CIT 2019) (Second Draft Redetermination).

[22] *See* Ellis, Paul D., *The Essential Guide to Effect Sizes*; Cambridge University Press (2010) (*Ellis*), at 3-4 ("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."); *see* also 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 (September 2002) (*Coe*), at 5 ("Effect 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."), The first two chapters of *Ellis* are included in Appendix I to the Second Draft Redetermination; *Coe* is included in Appendix III to the Second Draft Redetermination.

[23] *See Cohen* at 20 (equations 2.2.1 and 2.2.2).

[24] *See Mid Continent V*, 31 F.4th at 1377 ("Commerce recognized that the function of the denominator in the Cohen's *d* coefficient is to be a "yardstick to gauge the significance of the difference of the means" of the sales prices of the test and comparison groups.").

[25] Although this has been presented as a "simple average," Dr. Cohen describes the precise formula as "the root mean square of σ_A and σ_B." *See Cohen* at 44.

[26] The "test group" includes all sale prices of comparable merchandise to a given purchaser, region, or time period during the period of investigation (or review).

[27] The "comparison group" includes all sale prices of comparable merchandise during the period of investigation (or review) to all other purchasers, regions, or time periods.

[28] *See Mid Continent V*, 31 F. 4th at 1377.

* * *

> In this situation, Commerce needs a reasonable justification for departing from what the acknowledged literature teaches about Cohen's *d*. It has departed from those teachings about how to calculate the denominator of Cohen's *d*, specifically in deciding to use simple averaging when the groups differ in size. And its explanations for doing so fail to meet the reasonableness threshold (a deferential one, in recognition of expertise) for the reasons we have set forth.[29]

The Federal Circuit concluded that "Commerce must either provide an adequate explanation for its choice of simple averaging or make a different choice, such as use of weighted averaging or use of the standard deviation for the entire population."[30]

In the *Third Redetermination*, Commerce reexamined the academic literature and found that the academic literature supports the use of a simple average when the underlying data encompassed the full populations of sale prices rather than sample sale prices drawn from a larger universe of sales.[31] Commerce's explanation encompassed Dr. Cohen's presentation of effect size, and specifically focused on equation 2.3.2, where the denominator of the Cohen's *d* coefficient is calculated as the "simple average" of the standard deviations of two populations being compared when "there is no longer a common within-population σ."[32] Commerce explained that the sample size limitation did not affect the effect size because sample size is not relevant when the full populations of sale prices are used,[33] but instead are relevant only to the ***t***-test and power analysis which are based on sampled data.[34] Further, Commerce explained that the weighted-average formula from the literature applied to sampled data and not to

[29] *Id.*, 31 F. 4th at 1381.
[30] *Id.*
[31] *See Third Redetermination* at 8-12, 43, and 57.
[32] *See Cohen* at 44 (referencing *Cohen* at 20 and equations 2.2.1 and 2.2.2).
[33] Although an effect size may be estimated based on sampled data, Dr. Cohen explicitly states that effect size is a property of a population, and measures "the degree to which the phenomenon is present in the population."[33] *See* Third Redetermination at 8 (citing *Cohen* at 9).
[34] *See* Third Redetermination at 14-15 ("Thus, sample sizes are an input for the ***t***-test and the determination of whether the results of the analysis are statistically significant." (internal citations omitted)).

populations.[35] Finally, Commerce also explained that the Federal Circuit's suggested "use of the standard deviation for the entire population" would not be appropriate as this was not the same as the within population standard deviation as included in *Cohen* equations 2.2.1 and 2.2.2.[36]

In *Mid Continent VI*, the CIT found unsupported that Commerce's further explanation that the academic literature supports the use of a simple average to calculate the denominator of the Cohen's *d* coefficient. According to the CIT, the Federal Circuit held "that Commerce needed to justify its departure from the established statistical practice."[37] The CIT further found that even if the Federal Circuit had left Commerce with the option "to offer an explanation {} of its view of the literature," Commerce's explanation "fail{s} to support its position."[38] The CIT did not address either the weighted average approach advocated by the Taiwan Respondents or a single standard deviation of all prices in both the test and comparison groups, as identified as an option by the Federal Circuit.[39]

In *Mid Continent VI*, the CIT opined that "Commerce's non-academic arguments were unsupported – not unsupportable." Additionally, the CIT found:

> Commerce's reference to equations (2.2.1) and (2.2.2) as "explicitly" calculating effect size based on actual populations seems inconsistent, given that Cohen used these equations to generate ***d*** values to create his power tables, not as stand-alone tests. A test for full populations in the context of {a} power analysis would be redundant on its face, as there would be no question of statistical significance to analyze. Thus, Commerce does not explain, and it is not discernable why Commerce believes that equations (2.2.1) and (2.2.2)—still less equation (2.3.2), which expressly implicates sample size—are intended for testing full populations.

The CIT's opinion further cites *Mid Continent V* where the Federal Circuit held:

> Commerce used a "pooled standard deviation," pooling the standard deviations for each pair of test and comparison groups. As discussed above, it used simple

[35] *Id.* at 15-16 (citing *Cohen* at 66-67 and equations 2.5.1 and 2.5.2).
[36] *Id.* at 17-23.
[37] *See Mid Continent VI*, Consol. Court No. 15-00213, Slip Op. 23-45 at 11.
[38] *Id.*, Consol. Court No. 15-00213, Slip Op. 23-45 at 15.
[39] *Id.*, at 7, footnote 3.

averaging to do the pooling—even where the test and comparison groups have different sizes. In making that choice to use simple averaging, however, Commerce departed from, rather than followed, the cited statistical literature. As we have described above, Commerce's formula for the denominator,

$$\sqrt{\frac{\sigma_A^2+\sigma_B^2}{2}}$$

comes from a section of *Cohen* that addresses a situation in which the two groups at issue are of the same size … ({*i.e.*,} "CASE 2: $\sigma_A \neq \sigma_B$, $n_A = n_B$"). By contrast, when the sampled groups have unequal sizes, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging.[40]

Thus, the Federal Circuit found that the simple average, Dr. Cohen's equation 2.3.2, applies only when the sample sizes of the two groups are equal, *i.e.*, where "n_A=n_B."

Commerce respectfully disagrees with the understanding that Dr. Cohen's equation 2.3.2 does not define the effect size of the populations underlying a statistical analysis, but instead applies only to an analysis of sampled data with equal sample sizes.[41] Nonetheless, accepting the Federal Circuit's finding that in Dr. Cohen's scenario, the simple averaging formula applies only when the sample sizes of the two groups are equal, we find that Commerce's use of the simple average in the Cohen's *d* test is reasonable when the data under analysis are the full populations of sale prices in the test group and of sale prices in the comparison group.

In particular, below, Commerce provides reasonable justification to support departing from the Federal Circuit's understanding of the academic literature's use of a simple or weighted average. First, we explain the general principle that sample size is an indicator of reliability; similarly, because Commerce's use of the Cohen's *d* test is based on the full universe of sale

[40] *See Mid Continent V*, 31 F.4th at 1378.

[41] Commerce notes that the effect size, *e.g.*, the Cohen's *d* coefficient, is one of four independent parameters that determine statistical inference. *See Cohen* at 14. Further, Dr. Cohen states that the effect size is "'the *degree* to which the phenomenon is present in the population,' or 'the degree to which the null hypothesis is false;' … {it} *is some specific nonzero value in the population* … {where the} larger this value, the greater the *degree* to which the phenomenon under study is manifested." *Id.* at 9-10 (emphasis in original). Thus, Commerce understands that the effect size (*i.e.*, the Cohen's *d* coefficient) is a parameter which measures "the degree to which the phenomenon {*i.e.*, difference in prices} is present in the population{s}."

prices in the test group and in the comparison group, the test and comparison groups are equally reliable such that the use of a simple average is reasonable for purposes of calculating the denominator of the Cohen's *d* coefficient. Next, we explain why Commerce finds that the use of a single standard deviation of all of the prices of comparable merchandise is not reasonable in the denominator of Commerce's Cohen's *d* coefficient for properly identifying significant price differences.

A. The Simple Average, or Equal Weighting, Incorporates the Equal Reliability of the Calculated Standard Deviations and is Reasonable in Commerce's Cohen's *d* Test

In his presentation of the parameters of the statistical power analysis, Dr. Cohen describes the "reliability of sample results and sample size":

> 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 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.[42]

Dr. Cohen further notes that:

> The nature of the dependence of reliability upon n {*i.e.*, sample size} is obvious from the illustrative formulas, and, indeed, intuitively. The larger the sample size, other things being equal, the smaller the error and the greater the reliability or precision of the results.[43]

Indeed, when Dr. Cohen defines the four parameters of statistical inference, "sample size (n)" represents the reliability of the sample results.[44] Accordingly, the sample size is a gauge of the reliability of sample results as part of Dr. Cohen's power analysis. The larger the sample size vis-à-vis the population, the more reliable the sample results.

[42] *See Cohen* at 6 (emphasis in original).
[43] *Id.* at 7.
[44] *Id.* at 14 ("Four parameters of statistical inference have been described: power, significance criterion (a), sample size (n), and effect size (ES).").

As discussed above, the Federal Circuit understands that the use of the simple average of differing standard deviations, Dr. Cohen's equation 2.3.2, applies to an analysis involving sampled data because it is part of a power analysis which involves sampled data, including the use of a ***t***-test to evaluate the statistical significance of the results.[45] Furthermore, Dr. Cohen's equation 2.3.2 applies only when the sample sizes are equal, *i.e.*, $n_A = n_B$.[46] Because the sample sizes are equal in size and reliability, the estimated standard deviation for each of the sampled groups also has the same "reliability" (or precision) of a sample value {which} is the closeness with which it can be expected to approximate the relevant population value."[47] Consequently, a simple average of the standard deviations of the two groups is appropriate because the reliability of each value of the standard deviation is equal. In other words, when the sample sizes of the two groups are equal, then the reliability of the estimates of the standard deviations are the same, and it is appropriate to give equal weights, *i.e.*, a simple average, when averaging the two standard deviations to calculate the denominator of the Cohen's *d* coefficient.

In contrast, "when the sampled groups have unequal sizes {*i.e.*, $n_A \neq n_B$}, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging."[48] With the weighted average, the standard deviation of the group with the larger sample size (*i.e.*, sales volume) is given more weight than the group with the smaller sample size.[49] If the sample size of group A is larger than the sample size of group B, then the reliability of the standard deviation of group A will be greater than the reliability of group B. In such a

[45] *Id.* at 43-44.
[46] *See Mid Continent V*, 31 F.4th at 1378 ("{Equation 2.3.2, the simple average,} comes from a section of Cohen that addresses a situation in which the two groups at issue are of the same size. ('CASE 2: $\sigma_A \neq \sigma_B$, $n_A = n_B$')." (internal citations omitted, emphasis added)).
[47] *See Cohen* at 6.
[48] *See Mid Continent V*, 31 F.4th at 1378 (referencing *Cohen* at 67; *Ellis* at 26-27; *Coe* at 6); *see* also *Mid Continent VI*, Consol. Court No. 15-00213, Slip Op. 23-45 at 16 ("{T}he Court of Appeals has already held that the literature does not suggest simple averaging for unequal-sized groups." (citing *Mid Continent V*, 31 F.4th at 1380)).
[49] *See, e.g.*, *Coe* at 6 (equation 4).

situation, the standard deviation of group A has more reliability and is given more weight than the standard deviation of group B when calculating the denominator of the Cohen's *d* coefficient. Because the group with the larger sample size has greater reliability, the weights reflect the relative reliability of the standard deviations from the two groups.

In Commerce's application of the Cohen's *d* test, Commerce uses the full populations of data, *i.e.*, all prices of comparable merchandise to a given purchaser, region, or time period (*i.e.*, the test group) and all prices of comparable merchandise to all other purchasers, regions, or time periods (*i.e.*, the comparison group). As a result, the standard deviations calculated for the test and comparison groups each have a reliability of 100 percent, *i.e.*, "the closeness with which {the calculated value} can be expected to approximate the relevant population value."[50] In other words, the reliability of the calculated standard deviations based on the full population of sale prices to each group is identical. Because the reliability of the standard deviations based on full populations is equal, to calculate the denominator of the Cohen's *d* coefficient, Commerce finds that it is reasonable to weight these standard deviations equally, *i.e.*, a simple average, as presented in Dr. Cohen's equation 2.3.2, just as when the reliability is equal for standard deviations based on sampled data with equal sample sizes.

Commerce's use of the simple average to calculate the denominator of the Cohen's *d* coefficient is reasonable. The parameters calculated in the Cohen's *d* test which are used to calculate the Cohen's *d* coefficient (*i.e.*, the standard deviation and mean of each group), reflect a 100 percent reliability that they represent the parameters of the population because the Cohen's *d* test includes all sale prices in the test and comparison groups. With sampled data, Dr. Cohen provides the use of a simple average, *i.e.*, equation 2.3.2, when sample sizes are equal and

[50] *See Cohen* at 6.

standard deviations differ.[51] The use of the simple average when the sample sizes are equal reflects that the calculated parameters used to calculate the Cohen's *d* coefficient are equally reliable. Therefore, and because the reliability of these values is also equal when the calculated parameters are based on the full population of U.S. sale prices, it is reasonable to combine the standard deviations using a simple average to calculate the denominator of the Cohen's *d* coefficient.

B. The Use of a Single Standard Deviation is Not Reasonable for Commerce's Cohen's *d* Test

As explained above, Commerce's use of full populations when applying the Cohen's *d* test is significant and informs Commerce's decision to rely on a simple average when calculating the denominator of the Cohen's *d* coefficient. However, we recognize that the Federal Circuit in *Mid Continent V* held that:

> {t}he cited literature makes clear that one way to form the more general data-pool dispersion figure for the denominator—seemingly the preferred way if the full set of population data is available—is to use the standard deviation for the entire population.[52]

The Federal Circuit further held that:

> Commerce did not use the standard deviation of all the data for its denominator. It made that choice even while recognizing that it had the full set of data for U.S. sales for the period Commerce was reviewing.[53]

---

[51] *See Mid Continent VI*, Consol. Court No. 15-00213, Slip Op. 23-45 at 16 ("A test for full populations in the context of power analysis would be redundant on its face, as there would be no question of statistical significance to analyze. Thus, Commerce does not explain, and it is not discernable why Commerce believes that equations (2.2.1) and (2.2.2)—still less equation (2.3.2), which expressly implicates sample size—are intended for testing full populations {noting in footnote 12 that, in} *Mid Continent V*, the Court of Appeals discusses the use of equation (2.3.2) with sample groups, rather than full populations, implicitly recognizing that the equation does not apply only to full populations." (internal citations omitted)).
[52] *See Mid Continent V*, 31 F.4th at 1377.
[53] *Id.*, 31 F.4th at 1378.

The Federal Circuit concluded that:

> {i}ndeed, when the entire population is known, the cited literature points toward using the standard deviation of the entire population as the denominator in Cohen's d—which Commerce has not done.[54]

Consequently, the Federal Circuit indicated that Commerce may choose, on remand, to "use … the standard deviation for the entire population" in the denominator of the Cohen's *d* coefficient in lieu of a simple average.[55]

Therefore, we are addressing this option identified by the Federal Circuit and we explain below why we do not consider that option reasonable in the context of Commerce's differential pricing analysis. At the outset, we clarify that Commerce's methodology uses the standard deviations for the full populations. That is, Commerce's application of the Cohen's *d* test includes the full population of sale prices to the test group and the full population of sale prices to the comparison group. These two groups of data are full, separate populations, as recognized in Dr. Cohen's general formulation of the effect size, where the denominator, σ, is defined as either the standard deviation of either population A or the standard deviation of population B when the standard deviation of population A is assumed to be equal to the standard deviation of population B.[56] Therefore, the denominator, σ, in Dr. Cohen's equations 2.2.1 and 2.2.2, is *either* the standard deviation of population A *or* the standard deviation of population B, *but it is not* the standard deviation of populations A and B combined together.

We find that the option that the Federal Circuit identified, of using the standard deviation of all sale prices in the test and comparison groups as the denominator, is not appropriate for purposes of Commerce's Cohen's *d* test. Under this formulation, one standard deviation would

[54] *Id.*, 31 F.4th at 1380.
[55] *Id.*, 31 F.4th at 1381.
[56] *See Cohen* at 20 and 27 ("$\sigma_A = \sigma_B = \sigma$").

be calculated for populations A and B as a single population based on commingled sale prices. Professor Coe describes the effect size as the difference in the means divided by the "standard deviation":

> {t}he "standard deviation" is a measure of the spread of a set of values. Here it refers to the standard deviation of the population from which the different treatment groups were taken. In practice, however, this is almost never known, so it must be estimated either from the standard deviation of the control group, or from a "pooled" value from both groups (see question 7, below, for more discussion of this).[57]

Under question 7, "Which 'standard deviation'?," Professor Coe first proposes using the standard deviation of the control group, as with Glass' Δ.[58] Alternatively, given difficulties in selecting a control group, Professor Coe states that, *with sampled data*,

> it is often better to use a 'pooled' estimate of standard deviation. The pooled estimate is essentially an average of the standard deviations of the experimental and control groups (Equation 4).[59]

Thus, in identifying options for calculating the standard deviation in instances where full population data are not available, Professor Coe recognizes that there are two groups of data, each with its own standard deviation. Indeed, in describing the calculation of a pooled estimate of standard deviation, Professor Coe distinguishes a pooled average of the standard deviations of the experimental and control groups from a single "pooled" standard deviation:

> {n}ote that this is not the same as the standard deviation of all the values in both groups 'pooled' together. If, for example each group had a low standard deviation but the two means were substantially different, the true pooled estimate (as calculated by Equation 4) would be much lower than the value obtained by pooling all the values together and calculating the standard deviation.[60]

---

[57] *See Coe* at 2.
[58] *Id.* at 6; *see also Ellis* at 10.
[59] *See Coe* at 6.
[60] *Id.*

The cause for this overestimation is that the standard deviation within each group is calculated based on the mean within each group, whereas the standard deviation for both groups together would be the mean of all observations in both groups. Further, as recognized by Professor Coe, as the difference in the means increases between the two groups, the standard deviation of all observations in both groups will also increase rather than remain constant when based on the standard deviation of the observations within each group.

To illustrate the differences in the calculations, when the standard deviations in *Coe* Equation 4 is expanded, the equation can be restated as,

$$SD_{pooled} = \sqrt{\frac{(N_E - 1)\frac{\sum(X - \bar{X}_E)^2}{N_E - 1} + (N_C - 1)\frac{\sum(X - \bar{X}_C)^2}{N_C - 1}}{N_E + N_C - 2}}$$

which simplifies to:

$$SD_{pooled} = \sqrt{\frac{\sum(X - \bar{X}_E)^2 + \sum(X - \bar{X}_C)^2}{N_E + N_C - 2}}$$

Note that this is the same equation as *Cohen* equation 2.5.2 as well as the equations for the denominator for Cohen's *d* and Hedges' *g* in *Ellis*.[61] Even with sampled data, each of the formulas used to estimate the denominator of the effect size maintain the separate group of data and do not commingle all of the observations to calculate a single standard deviation for all of the data combined. The standard deviation for each group is based on the square of the difference between each observation within the group and that group's mean. The standard deviation of each group, whether sampled or full populations, is centered on the mean of each group.

[61] *See Ellis* at 26-27. Note that $\sum(X - \bar{X}_E)^2/N_E - 1$ and $\sum(X - \bar{X}_C)^2/N_C - 1$ above are the standard deviation for each group, *SD_A* and *SD_B*, respectively, from Dr. Ellis' equation for Hedges' g.

The equation for a "single standard deviation" of all observations combined together differs substantially:

$$SD_{single} = \sqrt{\frac{\sum(X - \bar{X}_{E\&C})^2}{N_E + N_C - 1}}$$

In the equation for the proposed SD_{single}, the standard deviation is based on the sum of the square of the difference of each observation from the single mean of the commingled observations in both groups. Accordingly, whereas the pooled standard deviation reflects only the variation in the data within each group, the "single standard deviation" not only reflects the variation of the data within each group, but also the difference in the means between the two groups. In application, this means that as the difference in the means between the test and comparison groups increases, the "single standard deviation" will also increase despite there being no change in the variances, *i.e.*, the dispersion, in the data within each of the two groups. Accordingly, the value of a pooled standard deviation (SD_{pooled}) will remain constant because it is based on the relationship (or spread) of the data within each group, rather than the value of a single standard deviation (*i.e.*, SD_{single}) increasing as the difference in the means between the two groups increases.

Therefore, the option to use a single standard deviation of all data when the data are explicitly separated into two separate populations is not a reasonable approach for Commerce's Cohen's *d* test. The single standard deviation causes the denominator of the Cohen's *d* coefficient to reflect not just the dispersion of the data within each group, but also the dispersion of the data between the two groups. Commerce uses effect size, the result of the Cohen's *d* test, to examine the difference in the mean prices to each group relative only to the dispersion of prices within both groups. The significance in the difference in the mean prices cannot be accurately gauged when that difference in the prices between the two groups is part of the "yardstick" used to assess that difference as achieved with a single standard deviation (*i.e.*, SD_{single}).

IV. INTERESTED PARTY COMMENTS

Dr. Cohen presented his original formulation of his *d* coefficient as

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

for a one-tailed case, or as

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

for a two-tailed case, where m_A and m_B are the "population means" and σ is "the standard deviation of either population (since they are assumed equal)."[62] Recognizing that this is a general, theoretical presentation of Dr. Cohen's *d* coefficient where the standard deviations of the two populations will not, with any reasonable expectations, be equal, the question at issue in this litigation is how to define the denominator of the Cohen's *d* coefficient. The Federal Circuit in *Mid Continent V* provided three options to address the court's concerns:[63]

1) A simple average of the standard deviation of the prices in the test group and the standard deviation of the prices in the comparison group;
2) A single standard deviation of the prices in both the test group and the comparison group;
3) A weighted average of the standard deviation of the prices in the test group and the standard deviation of the prices in the comparison group.

---

[62] *See Cohen* at 20, equations 2.2.1 and 2.2.2.
[63] *See Mid Continent V*, 31 F.4th at 1381 ("Commerce must either provide an adequate explanation for its choice of simple averaging or make a different choice, such as use of weighted averaging or use of the standard deviation for the entire population.").

Below, Commerce addresses the comments from the Taiwan Respondents and Mid Continent on the Draft Redetermination for each of these options. Finally, Commerce address other comments concerning the Draft Redetermination.

COMMENT 1: COMMERCE'S DRAFT REDETERMINATION DOES NOT PROVIDE AN ADEQUATE EXPLANATION FOR ITS CHOICE OF A SIMPLE AVERAGE

Simple Average and Equal Reliability

Taiwan Respondents' Comments:

- Commerce asserts that it is justified in using the simple average because the test group and comparison group are of equal 100 percent reliability.[64]
- "The fact that Commerce is comparing 100% reliable full populations, and can rely on {simple average} when sample sizes are equal, does not allow Commerce to rely on SA when comparing full populations of different sizes. Full populations having different sizes are distinctly different from samples having the same size. Using a simple average can make sense when sample sizes are identical, since simple averaging of two groups of equal sizes yields the same results as weighted averaging the two groups. In contrast, when full populations are of unequal sizes, simple averaging results in different results than weighted averaging."[65]
- Moreover, the results using the simple average "are fundamentally flawed because of the distortions inherent in simple averaging."[66]
- "{T}he fact that two samples may be equally reliable because they are the same {sample} size does not mean that they are '100% reliable' in the same manner as full populations are 100% reliable."[67]
- "{T}hat two groups are equally reliable or 100% reliable has nothing to do with creating an appropriate denominator yardstick of the Cohen's *d* analysis."[68]
- Commerce's "equally reliable" explanation is virtually identical to its "equally rational" and "equally genuine" explanations which were rejected in *Mid Continent V*.

[64] *See* Taiwan Respondents' Comments at 11 (citing Draft Redetermination at 12 ("The parameters calculated in the Cohen's *d* test which are used to calculate the Cohen's *d* coefficient (*i.e.*, the standard deviation and mean of each group), reflect a 100 percent reliability that they represent the parameters of the population because the Cohen's *d* test includes all sale prices in the test and comparison groups. With sampled data, Dr. Cohen provides the use of a simple average, *i.e.*, equation 2.3.2, when sample sizes are equal and standard deviations differ. The use of the simple average when the sample sizes are equal reflects that the calculated parameters used to calculate the Cohen's *d* coefficient are equally reliable. Therefore, and because the reliability of these values is also equal when the calculated parameters are based on the full population of U.S. sale prices, it is reasonable to combine the standard deviations using a simple average to calculate the denominator of the Cohen's *d* coefficient." (internal citation omitted))).
[65] *See* Taiwan Respondents' Comments at 12.
[66] *Id.*
[67] *Id.*
[68] *Id.*

- Commerce's analysis "conflates equality of reliability with equality of size."[69]
- Commerce's argument is made of a series of "four propositions based on erroneous and fluid interpretations of Cohen's d…"[70]
- Proposition 1: "Commerce is correct that when a researcher contemplates taking *independent* random *samples* of equal sizes from two *large* groups assumed to have equal *unweighted* standard deviations and is interested in the difference in *unweighted* group averages, {simple average} can be used for pooling," where all of the equations in the academic literature "apply either to entire populations or to independent random samples that are for all practical purposes negligibly small compared to the population size."[71] "{Simple average} is used because *in these special circumstances* … it is the sole correct measure of standard deviation that is not affected by the difference of the averages and accounts properly for all variation in the combined population."[72] Commerce's inference that simple average is used because "the reliability of each value of the standard deviation is equal" is incorrect because whether the use of a simple average is appropriate has nothing to do with reliability.[73]
- Proposition 2: "When the researcher contemplates taking independent random samples of *different* sizes from two (large) groups assumed to have equal standard deviations (none of these conditions generally hold in a DP analysis), a version of {weighted average} is used *with weights based on the sample sizes*."[74] Commerce "confuses the *standard error* of a sample … with the *standard deviation* of the population from which the sample is obtained."[75]
- The standard error of the mean is equal to the square root of "the usual unbiased estimate (from the random sample) of the population variance of X {*i.e.*, s^2}… {divided by} the number of independent units in (*i.e.*, the size of) the sample {*i.e.*, n}."[76]
- "The appearance of a pooled standard deviation in the denominator of Cohen's d is motivated by the key role of the standard error in statistical power analyses, *but it involves only the standard deviation s* and not the sample size n. This must be the case, because Cohen's d is a property of the population (or an estimate of it from a sample), whereas the standard error is a *sample* property that varies strongly with sample size – a factor irrelevant to the population."[77]
- Proposition 3: Commerce "confuses standard deviations with standard errors and misrepresents Cohen's meaning of 'reliable.'"[78] The "*standard error of the mean* computed from a full census of a (finite) population indeed can be considered zero ('100

[69] *Id.* at 13.
[70] *Id.*
[71] *Id.* and footnote 4 (emphasis in original); *see also* footnote 5 ("In *Cohen* the averages are never weighted; in Commerce's {differential pricing} analysis, the averages are always weighted by transaction quantity").
[72] *Id.* at 14 (emphasis in original).
[73] *Id.* 13-14 (internal citation omitted).
[74] *Id.* at 14 (emphasis in original).
[75] *Id.* and at footnote 5 ("standard errors and standard deviations should not be confused and are not interchangeable.").
[76] *Id.* at 15 (citing Cohen at 6-7). Dr. Cohen defines the standard error as $SE_{\bar{X}} = \sqrt{s^2/n}$.
[77] *Id.* (emphasis in original).
[78] *Id.*

percent reliable'); in contrast, the standard deviation of the population is definitely not zero unless all the values in the population are equal."[79]

- Proposition 4: "Commerce follows with a mathematical argument based on misunderstanding the meaning of zero and continuing to confuse standard deviations with 'reliability.'"[80] Commerce equates reliability with standard error, but then draws an unsupported conclusion concerning standard deviation. "In this 'full populations' context, both standard errors are zero {but just because} any two zeros are always equal, it does not follow that the corresponding standard deviations should be combined with equal weights."[81]
- "In sum, all that Commerce has demonstrated in this section of its {Draft Redetermination} is that in a full census of a population of two groups, the standard deviations of the groups are known with certainty. However, contrary to Commerce's conclusion, this certainty determines nothing about how to pool those standard deviations for computing an effect size."[82]

Mid Continent's Comments:

- "Mid Continent agrees with {Commerce's} analysis."[83]

Commerce's Position:

Although the Taiwan Respondents appear to agree with certain of Commerce's individual statements, they deny the relationship between these statements and Commerce's conclusion that the simple average is reasonable given the understanding of the academic literature. To summarize these statements:

1) The academic literature applies to analyses of sampled data, including Dr. Cohen's equation 2.3.2 which provides for the simple average of the standard deviations of two groups of data where the standard deviations differ but the sample sizes are equal.

2) The reliability of an estimated parameter based on sampled data may be dependent on many factors, but it is always dependent on the number of observations in the sampled data, *i.e.*, the sample size (n).

3) As Dr. Cohen's equation 2.3.2 requires that sample sizes be equal, this results in the reliability of the estimated standard deviations in equation 2.3.2 to be equal.

[79] *Id.* at 15-16 (emphasis in original).
[80] *Id.* at 16 (citing Draft Redetermination at 11-12 ("Because the reliability of the standard deviations based on full populations is equal, to calculate the denominator of the Cohen's *d* coefficient, Commerce finds that it is reasonable to weight these standard deviations equally…")).
[81] *Id.*
[82] *Id.*
[83] *See* Mid Continent Comments at 3.

Accordingly, the estimated standard deviations are combined using a simple average, *i.e.*, an average with equal weights.

4) When the sample sizes differ, then the reliability of the estimated standard deviations differs, and it is appropriate to weight average the estimated standard deviations to calculate the pooled standard deviation. The standard deviation for the population with the greater sample size will be accorded the larger weight.

5) When the analysis is based on data where each group is the full universe of data for that group (*i.e.*, the full population), then the reliability of the standard deviations for each group is 100 percent. The calculated standard deviations are the actual values of the standard deviation of each group.

6) Therefore, the average of the standard deviations to calculate the denominator of the d coefficient is reasonably unweighted, *i.e.*, a simple average, because, as with the situation where the sample sizes are equal and the reliability of the estimated standard deviations are equal, the reliability of the standard deviations calculated for the full populations of data are equal.

The Taiwan Respondents assert that:

> The fact that Commerce is comparing 100% reliable full populations, and can rely on {simple average} when sample sizes are equal, does not allow Commerce to rely on {simple average} when comparing full populations of different sizes.[84]

The Taiwan Respondents agree that the value of the standard deviations calculated using the full populations of data results in 100% reliable values and, thus, the reliability of these values is equal. Further, the Taiwan Respondents agree that using a simple average where the sample sizes are equal is reasonable, but state that a simple average is not reasonable with "full populations having different sizes."[85] However, the Taiwan Respondents fail to understand, as set forth above, that because "Commerce is comparing 100% reliable full populations" that it is the reliability of the calculated-from-population values that make the use of a simple average reasonable, as with the calculated-from-samples-with-equal-sample-sizes values that make the use of a simple average. The reliability of a value calculated for a population depends on the fact

[84] *See* Taiwan Respondents' Comments at 12.
[85] *Id*.

that the underlying data include the full universe of data. The number of observations in the population does not affect the reliability of a value of a parameter (*e.g,* the standard deviation) of the population.

The Taiwan Respondents believe that the use of a simple average "can make sense when sample sizes are identical since simple averaging of two groups of equal sizes yields the same results as weighted averaging the two groups."[86] We agree with this statement, yet the Taiwan Respondents' justification is just another example of their reliance on an arithmetic tautology to support their argument. Certainly, weight averaging two values where the weights, which, here, are the sample sizes, are equal will yield the same result as a simple average. However, Taiwan Respondents fail to address Commerce's explanation of reliability and why weight averaging is reasonable when sample sizes are unequal. "{A}n agency is not duty-bound to follow published literature when, *e.g.*, the literature is inapplicable to the specific problem before the agency." [87] Here, Commerce must go beyond the academic literature because Commerce uses populations, and the courts' understanding of the academic literature assumes sampling for all options to calculate the denominator of the Cohen's *d* coefficient.[88] As noted by the Federal Circuit, "when the sampled groups have unequal sizes, the cited literature uniformly teaches use of a pooled

---

[86] *Id*.
[87] *See Mid Continent V*, 31 F.4th at 1381.
[88] *See Mid Continent VI*, 628 F. Supp. 3d at 1325 ("Commerce's assertion that equation (2.5.2) requires estimation from a sample while equation (2.3.2) does not require estimation from a sample, appears inconsistent with the literature. *Id.* at 15 (citing *Cohen* at 66-67). Although Commerce identifies σ_A and σ_B in equation (2.3.2) as representing standard deviations of full populations, it fails to consider that the σ values themselves seem to be used by Cohen as pre-test estimates of the full population value, which will later be calculated with sampling. *See Cohen* at 44 (stating that equation (2.3.2) is accurate "provided that sample sizes are about equal"); *see also Ellis* at 10, 10 n.8 (stating of Cohen's *d* test "{i}f {the standard deviations} of both groups are roughly the same then it is reasonable to assume that they are estimating a common population standard deviation"). Thus, Commerce's assertion that sampling is not implicated in equation (2.3.2) is unsupported, as Cohen seems to use this equation in calculating statistical power."); *Cohen* at 67; *Ellis* at 10; *Coe* at 6-7.

standard deviation estimate that involves weighted averaging."[89] As Commerce explains, the use of a weighted average in that scenario is reasonable as it reflects the relative reliability of the estimated standard deviations used to calculate the pooled standard deviation.

Dr. Cohen demonstrates, through the standard error, that reliability is dependent upon, indeed inversely related to, the sample size. Dr. Cohen explains that the reliability of the results based on sample data "is always dependent upon the size of the sample."[90] As raised by the Taiwan Respondents in their comments, this can be seen in Dr. Cohen's definition of the standard error[91] of the estimated mean of a group of sampled data:

$$SE_{\bar{X}} = \sqrt{\frac{s^2}{n}}$$

where s^2 is the square of the estimated standard deviation of the sampled data and n is the sample size.[92] As the sample size, n, increases, the standard error decreases, and the reliability of the estimated statistic increases.

As discussed in the Draft Redetermination, because the reliability of an estimated parameter increases as the sample size increases, it is reasonable that when averaging the estimated parameters that the values of these parameters be weighted to reflect the relative reliability of the two values being averaged. If one value is more reliable than another, then the more reliable value logically warrants more weight than a less reliable value. This is reflected in the equations in the academic literature, as cited by the Federal Circuit, where the estimated

---

[89] *See Mid Continent V*, 31 F.4th at 1378, citing *Cohen* at 67 (equation 2.5.2, "Note that we have defined ***s*** quite generally so that it will old for all cases involving two independent samples, whether or not sample sizes are equal."), *Ellis* at 26-27 (where two equations for the pooled standard deviation, one restating Dr. Cohen's equation 2.5.2 and the second, by Dr. Hedges, where the equation is restated in terms of the estimated standard deviations of each group), and *Coe* at 6 (Equation 4 which restates Dr. Hedges equation from *Ellis*).
[90] *See* Draft Redetermination at 9 (citing *Cohen* at 6).
[91] *See Cohen* at 6 ("one conventional means for assessing the reliability of a statistic is the standard error (SE) of the statistic").
[92] *Id.* at 6-7.

values of the standard deviations of each group are weighted by the sample size of each group on which each value is calculated.[93] Commerce does agree, based on the reliability of the estimated standard deviations, that using a simple average is reasonable when the sample sizes are equal; further, the weighted average is used to average the estimated standard deviations from sampled data. However, because Commerce is using full populations and not samples, Commerce must determine a reasonable method of calculating the denominator of the Cohen's *d* coefficient, and based upon reliability, Commerce considers that a simple average is reasonable to calculate the denominator for Commerce's Cohen's *d* test, when the values of the standard deviations are based on the full populations of the data within each group (*i.e.*, within the test group and within the comparison group) independent of the number of observations in each population. To follow the logic to its conclusion, the academic literature teaches that weight averaging is appropriate when the reliability of the samples is different; the literature also teaches that the reliability increases as the sample size increases. Therefore, when using a full population, *i.e.,* a "sample size" of 100%, and regardless of the number of observations in the sample, the reliability reaches 100%, and it is appropriate to weight the values equally, *i.e.*, to calculate a simple average.

Although Taiwan Respondents attempt to distinguish the reliability of a value estimated on sampled data from the reliability of a value calculated on a full population, stating "the fact that two samples may be equally reliable because they are the same size does not mean that they

[93] Commerce recognizes that the academic literature addresses weighting defined as the sample size of the underlying data. Further, the Taiwan Respondents have argued that the appropriate weights for averaging the values for the pooled standard deviation is the total sales quantity to the test group and to the comparison group, whereas Mid Continent has argued for the use of the sample size of the two groups as the appropriate weights. Both the Federal Circuit in *Mid Continent V* ("we do not have before us for review a choice of one basis of weighting rather than another.") and the CIT in *Mid Continent VI* ("Because Commerce has not elected to use {either a weighted average or the single standard definition}, the court does not reach the relative merits of using either a single standard deviation or PT's proposed {weighted average} equation") did not opine on this question. *See Mid Continent V*, 31 F.4th at 1381, footnote 6; *Mid Continent VI* at 7, footnote 3. Commerce, in this redetermination, also does not adopt one specific basis for weighting as we continue to find that the use of a simple average is reasonable.

are '100% reliable' in the same manner as full populations are 100% reliable."[94] However, Commerce never claimed that the reliability of a value estimated on sampled data is "100% reliable." Even though an estimated value is not 100% reliable, it does not follow that two equally reliable samples, which are equally reliable because of equal sample size, should not be equally weighted. Commerce is not stating that the reliability of the values based on sampled data has to be equal to the reliability of the values based on a population. What Commerce said is that when the sample sizes are equal, then the reliability of the two estimated standard deviations are equal, and that when these values are to be averaged it is reasonable to use equal weights. Further, when the sample sizes are not equal, then the reliability of the two estimated standard deviations are not equal, and when these values are averaged, then it is reasonable to use weights, for example, based on the sample sizes or on the volume of sales within each group.

The Taiwan Respondents argue that Commerce's "equally reliable" argument in the Draft Redetermination is "virtually identical" to Commerce's "'equally rational' and 'equally genuine' argument{s} that the Federal Circuit expressly rejected in *Mid Continent V*."[95] Commerce's "equally rational" and "equally genuine" arguments explained how a respondent's pricing behavior was not dependent upon the volume of sales to a given purchaser, region or time period; and, thus, justified using a simple average, *i.e.*, equal weights, when averaging the standard deviations of the prices from test and comparison groups. As the Taiwan Respondents point out, the Federal Circuit found that "Commerce has not adequately justified, through its central rationale, its departure from the statistical literature's description of the Cohen's *d* coefficient".[96] However, in the Draft Redetermination, that estimated values from sampled data with equal

[94] *See* Taiwan Respondents' Comments at 12.
[95] *Id.* at 12-13.
[96] *Id.* at 13 (citing *Mid Continent V*, 31 F.4th at 1379).

samples sizes are "equally reliable," or that actual values from populations are "equally reliable," is based on the characteristic of the data rather than the pricing behavior of the respondents. In fact, "equally reliable" is not a function of the type of data under examination. The Taiwan Respondents' assumption that these terms are linked simply because they both include the word "equally" is meritless.

Under "Proposition One," the Taiwan Respondents assert that a simple average's use is limited to the "*special circumstances* (of comparing independent samples of large groups employing unweighted averages and unweighted standard deviations)"[97] or perhaps, alternatively, "when a researcher contemplates taking *independent* random *samples* of equal sizes from two *large* groups assumed to have equal *unweighted* standard deviations."[98] These "special circumstances," Taiwan Respondents assert have nothing to do with "equal reliability." Taiwan Respondents' claim is not supported by the record, and is inconsistent with the understanding of the Federal Circuit. The Federal Circuit stated:

> As we have described above, Commerce's formula for the denominator, {equation 2.3.2}, comes from a section of *Cohen* that addresses a situation in which the two groups at issue are of the same size. *Cohen* at 43-44; *id.* at 43 ("CASE 2: $\sigma_A \neq \sigma_B$, $n_A = n_B$"). By contrast, when the sampled groups have unequal sizes, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging.[99]

Taiwan Respondents do not explain how the limitations for the use of a simple average beyond those identified by the Federal Circuit under "Case 2" are relevant, and thus how they discredit the findings made in the Draft Redetermination.

Under "Proposition Two," the Taiwan Respondents argue that Commerce confuses the

[97] *Id.* at 14 (emphasis in original).
[98] *Id.* at 13 (emphasis in original) (where, in footnote 4, the Taiwan Respondents claim that "{a}ll the formulas in *Cohen*, *Coe*, or *Ellis* apply either to entire populations or to independent random samples that are for all practical purposes negligibly small compared to the population size.).
[99] *See Mid Continent V*, 31 F.4th at 1378 (citing *Cohen* at 67; *Ellis* at 26-27; *Coe* at 6.)

> *standard error* of a sample, a measure of "reliability" of the sample average … with the *standard deviation* of the population from which the sample is obtained.[100]

The Taiwan Respondents also state that:

> The appearance of a pooled standard deviation in the denominator of {the} Cohen's *d* {coefficient} is motivated by the key role of the standard error in statistical power analyses, *but it involves only the standard deviation s* and not the sample size *n*. This must be the case, because {the} Cohen's *d* {coefficient} is a property of the population (or an estimate of it from a sample), whereas the standard error is a *sample* property that varies strongly with sample size – a factor irrelevant to the population.[101]

Commerce disagrees that it "confuses" the standard deviation with the standard error. Commerce did not mention standard error in its Draft Redetermination, and standard error is not part of the calculation of the Cohen's *d* coefficient. The Taiwan Respondents confuse these two measures by equating "the role of the standard error in statistical power analyses" with the use of the standard deviation as the gauge by which the significance of the difference in the means is measured.

Standard error is not part of the calculation of the Cohen's *d* coefficient. The Cohen's *d* coefficient uses the standard deviation of the population (or the estimated standard deviation and the sample size when based on sampled data). The standard error is based on the estimated standard deviation and the sample size of the sampled data. Simply because both values use the same variable inputs does not mean that the standard error is part of the Cohen's *d* coefficient. The Taiwan Respondents have not explained how the calculation of the Cohen's *d* coefficient includes the value of the standard error, or how this invalidates Commerce's use of a simple average based on the equal reliability of the calculated standard deviations. The Taiwan Respondents' argument that Commerce has misunderstood the relationships between the

---

[100] *See* Taiwan Respondents' Comments at 14 (emphasis in original, internal citations omitted).
[101] *Id.* at 15 (emphasis in original).

standard error and the standard deviation, and how that invalidates Commerce's argument that a single average is reasonable remains unexplained and unsupported by the record evidence.

Under "Proposition Three," the Taiwan Respondents again assert that Commerce has confused the "standard deviation" with the "standard error," when in fact it is the Taiwan Respondents which are confusing the two concepts. The Taiwan Respondents state that Commerce claims that when the full populations are used, "the standard deviations calculated for the test and comparison groups each have a reliability of 100 percent."[102] We agree, yet the Taiwan Respondents extend this statement to argue that this is evidence that Commerce "confuses standard deviation with standard errors and misrepresents {Dr.} Cohen's meaning of 'reliable'."[103] The Taiwan Respondents appear to support their claim of Commerce's confusion with the statements that the standard error of a mean based on the full population "can be considered zero ('100 percent reliable')" whereas the "standard deviation of the population is not zero unless all values in the population are equal."[104] We also agree with these two statements.[105] Yet, despite reciting these statements, the Taiwan Respondents fail to explain how these statements demonstrate Commerce's confusion. The confusion lies with the Taiwan Respondents and their conflation of standard deviation and standard error, apparently because the standard error for both the mean and the standard deviation are both a function of the estimated standard deviation of the *sampled data*.

Under "Proposition Four," because Commerce "{misunderstands} the meaning of zero and {continues} to confuse standard deviations with 'reliability''" the Taiwan Respondents argue

---

[102] *Id.*
[103] *Id.*
[104] *Id.* at 16.
[105] When calculated for a full population, both the mean and the standard deviation are the actual values of those population parameters, there is zero standard error, and the values are "100 percent reliable." Further, the standard deviation of a population is zero when all of the values in the population are equal, and the standard deviation of a population, *i.e.*, the dispersion of the data, is non-zero when the values in the population are not equal.

that "Commerce has used 'reliability' synonymously with standard error but then draws a conclusion about standard deviations, one that is not supported."[106] The Taiwan Respondents support this statement by stating that, with full populations, "both standard errors are zero," and because "any two zeros are always equal, it does not follow that the corresponding standard *deviations* should be combined with equal weights."[107] Commerce agrees with the situation described by the Taiwan Respondents, *i.e.*, standard deviations calculated using the full population results in zero standard error, which means that the reliability of the two standard deviations are equal, which supports the use of a simple average. However, the Taiwan Respondents fail to identify what is wrong with this, and as such, we find that this argument is without merit.

We find that the arguments presented by the Taiwan Respondents are not persuasive regarding Commerce's explanation that the equal reliability of the actual values of the standard deviations based on full populations and the equal reliability of estimated values of the standard deviations based on equal sample sizes. Accordingly, we continue to find that the equal reliability of the standard deviations based on equal sample sizes reasonably supports the use of a simple average when the full populations of data are used to calculate the actual values of the standard deviation which also have equal, 100 percent, reliability.

COMMENT 2: INDEPENDENCE FROM THE DIFFERENCES IN SALE VOLUMES

Taiwan Respondents' Comments:

- Commerce also asserts that it is justified in using a simple average because the difference in the prices between the test group and the comparison group in the numerator of the Cohen's *d* coefficient is not dependent on the volume of sales to the test and comparison groups, . . . the yardstick which is used to measure this difference in the means should reasonably be defined by the standard deviation of each group of prices (*i.e.*, the variance

[106] *See* Taiwan Respondents' Comments at 16.
[107] *Id.* (emphasis in original).

or dispersion of individual prices) on an equal basis, independent of the volume of sales in the test group and comparison group."[108]

- Commerce's assertion is flawed because "the numerator of the Cohen's *d* coefficient is *not* independent of the volume of sales in each group" because the average price of each group is weighted by the sale quantities in each group.[109] "It simply is false that '{e}ach of the weighted-average prices and the difference between them does not depend upon the relative volume of sales (either the number of observations or the sale quantity),'"[110] Because the numerator does take into account the sale volumes with the weighted-average prices, and "the numerator *does* depend on the volumes of sales, which may be *unequal*, the 'yardstick' {*i.e.*, the denominator} must also reflect the sales volumes."[111]
- Moreover, simply because the difference in the means does not consider the sales volumes does not mean that a simple average is reasonable to calculate the denominator of the Cohen's *d* coefficient. If this logic were reasonable, then a simple average would always be used, but the Federal Circuit has found that "the academic literature does not allow reliance on {simple average} when units are sampled."[112]
- Commerce has not "explained why relying on {simple average} leads to reasonable results."[113] The courts have already found that use of a simple average is not appropriate, and the additional examples previously provided the Taiwan Respondents and repeated in these comments support the courts' conclusions.

Mid Continent's Comments:

- "Mid Continent agrees with {Commerce's} analysis."[114]

Commerce's Position:

We agree with the Taiwan Respondents' argument that if Commerce's logic is applied when the difference in the means (*i.e.*, the numerator) is independent of the total sale volumes to each group, then "{Commerce} should rely on {simple average} in all cases, whether measuring samples or full populations."[115] The Taiwan Respondents continue that "the courts already have concluded that the academic literature does not allow reliance on {simple average} when units

[108] *Id.* at 16-17 (citing Draft Redetermination at 14).
[109] *Id.* at 17 (emphasis in original).
[110] *Id.* at 18 (citing Draft Redetermination at 13); *see also* Taiwan Respondents' Comments at footnote 6 ("{T}he example tabulated in Section IC {} concerns two groups of transactions at the same unit prices but with different quantities. Their weighted means and weighted standard deviations both differ.").
[111] *See* Taiwan Respondents' Comments at 19 (emphasis in original).
[112] *Id.* at 17.
[113] *Id.*
[114] *See* Mid Continent Comments at 3.
[115] *See* Taiwan Respondents' Comments at 17.

are sampled."[116] Certainly, the courts have stated that the academic literature provides for a weighted average when the sample sizes differ (and when the sample sizes are equal, the weighted average becomes a simple average). Thus, on further consideration, Commerce is no longer relying on argument (B) in the Draft Redetermination, and this argument has been removed from the final results of this redetermination as support that the use of a simple average is reasonable. However, because Commerce's explanation regarding reliability still supports the use of a simple average, and Commerce continues to use a simple average to calculate the denominator of the Cohen's *d* test.

COMMENT 3: SINGLE STANDARD DEVIATION

Taiwan Respondents' Comments:

- "{O}ne conclusion {that} Commerce draws is correct: 'whereas the pooled standard deviation reflects only the variation in the data within each group, the 'single standard deviation' not only reflects the variation of the data within each group, but also the difference in the means between the two groups.'"[117]
- That Commerce concludes that the use of a single standard deviation[118] is not "reasonable in the context of Commerce's differential pricing analysis" suggests that the use of a simple average is also not reasonable.[119]
- "There is no dispute" that the "only standard deviation that is relevant and meaningful for Cohen's *d* is the one employed to compute the standard error of the *weighted* mean," *i.e.*, "the quantity-weighted standard deviation."[120] However, despite the fact of "Commerce's insistence that it is analyzing a population," Commerce's explanation is

---

[116] *Id.*
[117] *Id.* at 20 (citing Draft Redetermination at 18).
[118] *See Mid Continent V*, 31 F.4th at 1377 ("The cited literature makes clear that one way to form the more general data-pool dispersion figure for the denominator—seemingly the preferred way if the full set of population data is available—is to use the standard deviation for the entire population."), at 1380 ("Indeed, when the entire population is known, the cited literature points toward using the standard deviation of the entire population as the denominator in Cohen's *d*—which Commerce has not done."), and at 1381 ("Commerce must either provide an adequate explanation for its choice of simple averaging or make a different choice, such as use of weighted averaging or *use of the standard deviation for the entire population*." (emphasis added)). Commerce has stated that the prices to the test group represent one full population, and the prices to the comparison group represent a second, full population. Thus, the "entire population" refers to either the test group or the comparison group. Commerce respectfully disagrees with the apparent understanding that all prices together from the test and comparison groups represent a single population. Further, as stated by Professor Coe, and discussed in the Draft Redetermination, combining the data from the two groups into a single pool of data is not appropriate.
[119] *See* Taiwan Respondents' Comments at 19-20 (citing Draft Redetermination at 15).
[120] *Id.* at 20 (emphasis in original).

based on "samples {where} no weights are used in computing the averages or the standard deviations."[121] The equations in the Draft Redetermination are irrelevant.[122]

- Commerce's conclusion that "the pooled standard deviation reflects only the variation in the data within each group … is a *defining characteristic* of a pooled standard deviation and so it must be true of any correct formula."[123]
- Commerce's "thought experiment" where prices in one group are increased which does not impact the standard deviations with each group, but does impact the single standard deviation of all of the prices as if in a single group.[124] Under this approach, the pooled standard deviation will not change "provided the transaction quantities do not change."[125] "Equivalently," Commerce proposes a second approach "to perform pooling" which is to "adjust the overall standard deviation (of the combined populations) by an amount determined by the difference in average values between the two groups."[126] "Commerce's point is that the amount subtracted in (2) {i.e., the value of Taiwan Respondents' Formula***} is usually positive, and therefore the overall standard deviation proposed by the Federal Circuit is generally too large (which would decrease Cohen's *d*, making it more difficult for data to "pass" the test)."[127]
- "The squares of the standard deviations (the group variances) enjoy the simplest possible relationship: the (weighted) variance of the combined population is the sum of (A) the (weighted) variance of group A, (B) the (weighted) variance of group B, and (C) the (weighted) variance associated with the difference in the (weighted) group averages. To adjust the overall standard deviation, then, the rules of arithmetic state it will suffice to subtract (C) from the overall variance. Subtraction is the sole, universally applicable mathematical procedure that will make the result independent of (C), as required by Commerce."[128]
- The key formula to calculate the "overall standard deviation from the difference in means is the product of three quantities: (1) the relative weight of group A, (2) the relative weight of group B, and (3) the square of the difference in the means" where the "relative weight" is the proportion of the sales quantity of each group to the total sales quantity.[129] This formula is illustrated by the examples provided by the Taiwan Respondents that quantifies the difference in the weighted variance between the weighted average of the standard deviations of the test and comparison groups and the single standard deviation of all of the data combined into a single group.[130] "*This recipe for pooling variances (and therefore standard deviations, upon taking square roots) unites* all *the formulas for Cohen's* d *that appear in* Cohen, Coe, *and* Ellis *and generalizes them to the case of weighted means needed to perform a DP analysis.*"[131]

---

[121] *Id.* (emphasis in original).
[122] *Id.* (citing Draft Redetermination at 17 where equations from *Cohen* and *Ellis* are presented).
[123] *Id.* (emphasis in original).
[124] *Id.* at 21.
[125] *Id.*
[126] *Id.* at 21-22.
[127] *Id.* at 22.
[128] *Id.*
[129] *Id.* at 23.
[130] *Id.* at 23-30.
[131] *Id.* at 29 (emphasis in original).

- "As shown in these examples, the correct pooled variance *according to Commerce 's argument (C)* is 1750. This agrees with what {weighted average} pooling calculates, but it differs materially from the {simple average} result of 2050. Moreover, because a pooled variance is supposed to reflect *only part* of the population variance, *it cannot ever exceed the population variance*. Both the first and last version of this example exhibit transactions with population variances (of 1875 and 1750) that are *exceeded* by the {simple average} value of 2050. These conclusions demonstrate {simple average} cannot be correct, *according to Commerce's own characterization of pooling*."[132]

Mid Continent's Comments:

- "Mid Continent agrees with {Commerce's} analysis."[133]
- Nonetheless, "Commerce should expand its reasoning to consider the following points and practical example."[134]
  - The Federal Circuit's proposal "assumes that the population standard deviation is common or the same in both populations," and this assumption needs to be questioned.[135]
  - "If two populations are different in terms of their observation values, the population standard deviation will be skewed in favor of the group with the larger amount of variation among its observations. The pooled variance will be affected by the larger population size, which is contrary to Commerce's goal of comparing *prices* independent of the number or quantity of sales. Commerce uses the average of both variances to make it *representative of both populations*."[136]
  - As an example provided by Mid Continent, "the larger one population size is, the more weight will be given to the variance of the larger group, and it will directly affect the values of the pooled standard deviation."[137]
  - "Commerce's calculation already incorporates different group sizes in the calculation of the means used in the numerator of the effect size calculation. Additionally group sizes were considered in the calculation of standard deviation values used to calculate the denominator of the effect size."[138]
  - The example shows that the proposal by the Federal Circuit to use the standard deviation of the entire population is not appropriate; one group will almost always have a large number of sales and, therefore, have an outsize influence on the calculation of the overall standard deviation.
  - This undermines Commerce's goal, to determine whether the means of two groups individually are different enough that they cross a particular effect size threshold, indicating that they are "differentially priced."[139]
  - Moreover, contrary to the implication in *Mid Continent V*, that "the quantity/ population size is ignored if the square root of the average of variances is used instead

[132] *Id.* at 29-30 (emphasis in original).
[133] *See* Mid Continent Comments at 3.
[134] *Id*. at 7.
[135] *Id.*
[136] *Id.* (emphasis in original).
[137] *Id.*
[138] *Id.*
[139] *Id.*

> of the pooled standard deviation," Commerce does use "the population size … in the calculation of each mean and standard deviation, and therefore in the calculation of Cohen's *d*."[140]

Commerce's Position:

Commerce agrees with the Taiwan Respondents' assessment that Commerce correctly concluded that the use of a single standard deviation is not supported by the academic literature for use as the denominator of the Cohen's *d* coefficient.[141] Following Professor Coe's explanation, a pooled standard deviation (*i.e.*, based on a weighted average of the estimated standard deviations of the sampled data in each group) includes only the variances in the data within each group.[142] However, the single standard deviation includes not only the variances of the data within each group but also includes the difference in the means between the two groups. Accordingly, the single standard deviation would distort the measure of the effect size.[143]

However, Commerce disagrees with the Taiwan Respondents that this analysis also leads to the conclusion that "*only the* {*weighted average*} *formula will work*" to calculate the denominator of the Cohen's *d* coefficient.[144] The Taiwan Respondents conclude that Commerce's explanation concerning the single standard deviation "is correct" and

> is a *defining characteristic* of a pooled standard deviation and so must be true of any correct formula. But, unfortunately for Commerce's argument, the {simple average} formula for pooling does *not* enjoy this property: it *does* reflect variation in the difference of weighted group means. **Only the {weighted average} formula will work in the {differential pricing} analysis, as will now be shown.**"[145]

* * *

[140] *Id.*
[141] *See* Taiwan Respondents' Comments at 20.
[142] *See* Draft Redetermination at 16 (citing *Coe* at 6 (with sample data "it is often better to use a 'pooled' standard deviation. The pooled estimate is essentially an average of the standard deviations of the experimental {*i.e.*, the test} and control {*i.e.*, the comparison} groups (Equation 4)."); *see also Ellis* at 10).
[143] *See* Draft Redetermination at 17-18.
[144] *See* Taiwan Respondents' Comments at 20 (emphasis in original).
[145] *Id.* (emphasis in original).

> Moreover, because a pooled variance is supposed to reflect *only part* of the population variance, *it cannot ever exceed the population variance*. …These conclusions demonstrate {simple average} cannot be correct, *according to Commerce's own characterization of pooling*.[146]

First, the Taiwan Respondents provide an equation which purports to quantify the difference between the weighted average of the estimated standard deviations and the single standard deviation.[147] According to the Taiwan Respondents, "the pooled variance {*i.e.*, the weighted average} is found by subtracting the variance of the mean difference {*i.e.*, the single standard deviation} from the population variance {*i.e.*, the value from Formula***}."[148] Second, the Taiwan Respondents state that the simple average of the estimated standard deviations does not follow the same pattern (*i.e.*, that the single standard deviation is greater than the (weighted) average of the estimated standard deviations): instead, the simple average is greater than the single standard deviation.[149] The Taiwan Respondents provide three examples which ostensibly support their conclusions.[150] In these examples, the Taiwan Respondents allegedly demonstrate that the value of the weighted average is the value of the single standard deviation less the value

[146] *Id.* at 29-30 (emphasis in original)
[147] *Id.* at 23 ("The contribution to the overall standard deviation from the difference in means is a product of three quantities: (1) the relative weight of group A, (2) the relative weight of group B, and (3) the square of the difference in means" where the "'relative weight' of a group is its proportion of the total transaction quantity.") This formula (*i.e.*, the product of (1), (2) and (3) as described by the Taiwan Respondents) can be expressed as $\left(\frac{w_A}{w_A+w_B}\right) \times \left(\frac{w_B}{w_A+w_B}\right) \times (\text{difference of the means})^2$ -- *i.e.*, the "Formula***"). Indeed, the Taiwan Respondents use this formula to calculate the value of the weighted average based on the value of the single standard deviation. *See* Taiwan Respondents' Comments at 26 ("The pooled variance {i.e., the weighted average} is found by subtracting the variance of the mean difference {*i.e.*, the value of Formula***} from the population variance {i.e., the single standard deviation}." (emphasis omitted)).
[148] *Id.* at 26.
[149] *Id.* at 29-30 ("because a pooled variance {*i.e.*, standard deviation} is supposed to reflect *only part* of the population variance {*i.e.*, the single standard deviation}, *it cannot ever exceed the* {*single standard deviation*}. Both the first and the last {examples} exhibit {single standard deviations} that are *exceeded* by the {simple average} value {}. These conclusions demonstrate that {simple average} cannot be correct, *according to Commerce's on characterizations of pooling*." (emphasis in original)).
[150] *Id.* at 23-30 ("**confirming the validity of this analysis**" (emphasis in original)). In the appendix to these final results of redetermination, Commerce has replicated the examples as presented by the Taiwan Respondents, and labelled them according to the page numbers on which each example starts in the Taiwan Respondents' Comments. Commerce has also generated three additional examples following the Taiwan Respondents' approach of changing the prices within either of the price groups by an equal amount so that the standard deviations within each group do not change but that the difference in the means does change. *See also* Taiwan Respondents' Comments at 21.

of Formula***. Further, for two of the three examples, the value of the simple average exceeds the value of the single standard deviation, which proves that the simple average does not follow Commerce's, and Professor Coe's, "defining characteristic" that the single standard deviation includes an amount for the difference in the means, *i.e.*, something extra, a "premium," that is not part of the pooled standard deviation calculated using a weighted average of the estimated standard deviations of the sampled data in each group.

The Taiwan Respondents' arguments and logic are unpersuasive for several reasons. Foremost, the Taiwan Respondents appear not to understand the framework, logic, and principles of Professor Coe's explanation of the distortion caused by the use of a single standard deviation as the denominator of Dr. Cohen's *d* coefficient. Second, the Taiwan Respondents' Formula*** does not quantify the difference between the single standard deviation and the weighted average of the estimated standard deviations of the two groups. Third, because of the Taiwan Respondents' misunderstanding of Professor Coe's analysis, their conclusion that the value of the simple average can never exceed the value of the single standard deviation is not valid. Fourth, the three examples provided by the Taiwan Respondents fail to support their conclusions, but, rather, are consistent with the principles outlined by Professor Coe and as repeated by Commerce in the Draft Redetermination.

As an initial matter, Commerce agrees with the Taiwan Respondents that "all averages, standard deviations, and variances are weighted."[151] *Within each group*, whether generically identified as "A" and "B", or specifically identified as "test" and "comparison" in the Cohen's *d* test, the mean (or averages in the numerator) and the standard deviation in the denominator (the variance is equal to the square of the standard deviation) are weighted by the sale quantity of

[151] *See* Taiwan Respondents' Comments at 22.

each transaction. This within group weighting is different than the question of weighting when averaging the standard deviations of the two groups to calculate the denominator of the Cohen's *d* coefficient.[152]

<u>Professor Coe's Explanation:</u> Professor Coe's paper only concerns effect size as used with sampled data, and his analysis of the appropriateness of using a single standard deviation is limited to the comparison of the value of the single standard deviation with the value of a weighted average of the estimated standard deviations. As recognized in *Mid Continent V*, when sampled data are the basis for the analysis, the academic literature provides for a weighted average.[153] Professor Coe explains that the single standard deviation will include an amount for the difference in the means, which is not a part of the weighted average (*i.e.*, the pooled standard deviation). Thus, the value of the single standard deviation will *exceed* the value of the pooled standard deviation. Professor Coe *does not* address the simple average of the estimated standard deviations as an approach to calculate the denominator. Professor Coe simply states that the pooled standard deviation (*i.e.*, the weighted average of the estimated standard deviations) will remain constant, *i.e.*, independent of the difference in the means.[154] Further, following Professor Coe's analysis, the single standard deviation will increase (or decrease) as the difference in the means increases (or decreases). Accordingly, Professor Coe concludes that the single standard

[152] As noted above, such weighting is the subject of this litigation, whether equal weighting (simple average) or weighted using some measure of each of the two groups (*e.g.*, number of transactions, or the total sale quantity).
[153] *See Mid Continent V*, 31 F.4th at 1378 ("When the full population data set is unavailable, all of the cited literature points to use of a 'pooled standard deviation' of the two particular groups at issue to form the denominator … when the *sampled groups* have unequal sizes, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging." (citing *Cohen* at 67; *Ellis* at 26-27; *Coe* at 6) (emphasis added)). Following the understanding of the Federal Circuit, when the sample sizes are equal, *i.e.*, $n_A = n_B$, then the simple average of *Cohen* equation 2.3.2 may be used. *See Mid Continent V*, 31 F.4th at 1378 ("Commerce's formula for the denominator, {equation 2.3.2} comes from a section of *Cohen* that addresses a situation in which the two groups at issue are of the same size. *Cohen* at 43-44; *id.* at 43 ('CASE 2: $\sigma_A \neq \sigma_B$, $n_A = n_B$').").
[154] *See* Draft Redetermination at 18 ("the value of a pooled standard deviation (SD_{pooled}) will remain constant because it is based on the relationship of the data within each group").

deviation is inappropriate to use as the denominator of the *d* coefficient as it fluctuates with the difference in the means as opposed to remaining a constant by which the "effect" (or significance) of that difference is measured. These three principles – (1) the value of the weighted average remains constant with changes in the difference in the means, (2) the single standard deviation increases (or decreases) as the single standard deviation increases (or decreases), and (3) the value of the single standard deviation will exceed the value of the weighted average (or be equal when the difference in the means is zero) – are the result of Professor Coe's analysis. Professor Coe makes *no statement or conclusion* concerning the relationship between the value of a simple average of the estimated standard deviations when the group sizes are unequal and the value of the single standard deviation or the value of the weighted average.

Taiwan Respondents' Formula***: The Taiwan Respondents present a formula with no explanation of its derivation to quantify the difference in the values between a weighted average and the single standard deviation. As noted in the Draft Redetermination, Commerce presented the equations to calculate both the pooled standard deviation (*i.e.*, weighted average) and the single standard deviation;[155] however, Commerce has not derived an equation that expresses the

[155] *Id.* at 17 and 18. Note that the Taiwan Respondents appear to not understand that the two equations on page 17 for the pooled standard deviation, SD_{pooled}, are based on a weighted average and not a simple average. *See* Taiwan Respondents' Comments at 20 ("The formulas appearing in *Cohen* or *Ellis*, as noted at page 17 of the {Draft Redetermination}, are irrelevant, and their superficial similarity to {the simple average} is meaningless and does not necessarily generalize to any other situation."). When the equation for the standard deviation is inserted in the first equation (*e.g.*, $SD_E = \sqrt{\frac{\Sigma(X-\bar{X}_E)^2}{N_E-1}}$, or, as the square of the standard deviation, $SD_E^2 = \frac{\Sigma(X-\bar{X}_E)^2}{N_E-1}$), then the first equation on page 17 reduces to $SD_{pooled} = \sqrt{\frac{(N_E-1)SD_E^2+(N_C-1)SD_C^2}{N_E+N_C-2}}$. This equation is the formula for the weighted average of the estimated standard deviations (for groups E and C), as advocated for by the Taiwan Respondents, and *is not* the formula for the simple average, as rejected by the Taiwan Respondents.

difference between these two calculations.[156] Further, in its three examples, the Taiwan Respondents used Formula*** to calculate the value of the weighted average based on the value of the single standard deviation.[157]

However, the Taiwan Respondents' Formula*** is too simplistic to define that difference given the equations for the weighted average and the single standard deviation. Indeed, this is demonstrated by the examples included in the Appendix to this remand redetermination. Although the three self-selected examples provided by the Taiwan Respondents result in that difference being equal to the value calculated using the Formula***, the three additional examples provided by Commerce, as discussed below, do not demonstrate that the difference between the values of the pooled standard deviation and the single standard deviation are equal to the value resulting from the use of the Formula***. Therefore, besides illustrating that there is a difference between the values of the pooled standard deviation and the single standard deviation, the Taiwan Respondents' Formula*** is meritless.

Value of the Simple Average: As noted above, Professor Coe only discussed a pooled standard deviation that is calculated as a weighted average of the estimated standard deviations within each group. As such, the assumption by the Taiwan Respondents that there is an expected relationship between the simple average when group sizes are unequal and the single standard deviation is not part of the academic literature and is not supported by the record. Accordingly, the conclusion that the Taiwan Respondents draw, *i.e.*, the simple average is illogical and the weighted average is the only approach that is reasonable, is also unsupported.

---

[156] From the Draft Redetermination, the difference between the single standard deviation and the weighted average would be $\sqrt{\frac{\sum(X-\bar{X}_{E\&C})^2}{N_E+N_C-1}} - \sqrt{\frac{(N_E-1)\frac{\sum(X-\bar{X}_E)^2}{N_E-1}+(N_C-1)\frac{\sum(X-\bar{X}_C)^2}{N_C-1}}{N_E+N_C-2}}$.

[157] *See* Taiwan Respondents' Comments at 26 ("The **pooled variance** {*i.e.*, the value of the weighted average} is found by subtracting the variance of the mean {*i.e.*, the value of Formula***} from the population variance {*i.e.*, the value of the single standard deviation}" (emphasis in original)).

As illustrated in the examples provided by the Taiwan Respondents, as discussed below, the value of the simple average is constant and independent of the difference in the means just as the value of the weighted average is constant and independent of the difference in the means.

Examples: The Appendix includes the three examples provided by the Taiwan Respondents,[158] as well as three additional examples added by Commerce which modify the first of the Taiwan Respondents' examples pursuant to the principles included in the Taiwan Respondents' Comments.[159] These examples demonstrate that the Taiwan Respondents' Formula*** is erroneous. For the three examples provided in Taiwan Respondents' Comments. For the three examples provided in Taiwan Respondents' Comments, Formula*** does provide the value of the difference between the value of the weighted average and the value of the single standard deviation, as colored in **green** in the Appendix. However, beyond the Taiwan Respondents' self-selected data, their Formula*** fails to accurately calculate the difference between the weighted average and the single standard deviation, as colored in **red** for the last three examples in the Appendix. Thus, the Taiwan Respondents' Formula*** fails to quantify the difference in the two values and is meaningless.

---

[158] The three examples provided by the Taiwan Respondents are labelled based on the page number on which each example originates in the Taiwan Respondents' Comments.

[159] *See* Taiwan Respondents' Comments at 21 ("One mechanism whereby this difference of (weighted) means can be changed is to add a constant value (positive or negative) to every price in one of the groups, without changing the transaction quantities. For instance, when all the unit prices in the transaction group are increased by one dollar, the difference in weighted means between that group and the comparison group is also increased by one dollar, but the standard deviations in each group remain the same. {footnote 7}"). In footnote 7, the Taiwan Respondents assert that "Conventional formulas for standard deviations appear to involve the mean and therefore look like they might be affected by changes in the mean. However, those formulas can be rewritten solely in terms of differences between the values in a group. When the same number is added to all those values, none of their differences is altered, ergo the standard deviation (weighted or not) will not change." Although Taiwan Respondents' conclusion that "When the same number is added to all those values, none of their differences is altered, ergo the standard deviation (weighted or not) will not change" is accurate, Taiwan Respondents' statements that the "standard deviation *appear* to involve the mean" and "*look like they might* be affected by changes in the mean" are incorrect. The standard deviation *is* a function of the mean and the standard deviation *is* affected by changes in the mean. *See* Draft Redetermination at 18. Taiwan Respondents' statement that "those formulas can be rewritten solely in terms of the differences between the values in a group" is unsupported, just as Taiwan Respondents' failure to explain the derivation of their Formula***.

Nonetheless, these six examples do illustrate the three principles described by Professor Coe[160] in addressing the differences in a denominator defined as a weighted average of the estimated standard deviations and in a denominator (erroneously) defined as a single standard deviation. First, the weighted average values, colored in **blue**, are the same value (*i.e.*, **1750**), independent of the value of the difference in the mean. Second, the values of the single standard deviation, colored in **brown**, vary with the value of the difference in the means, increasing or decreasing as the difference of the means increases or decreases; and each value of the single standard deviation is greater than the single value of the weighted average. Third, the value of the single standard deviation is greater than the value of the weighted average, as the value of the single standard deviation includes an amount that represents the difference in the means between the two groups. As demonstrated above, the Taiwan Respondents' Formula*** does not quantify this difference.

As noted above, Professor Coe does not address the value of a simple average of the estimated standard deviations or its relationship with the value of the single standard deviation. Thus, the conclusions drawn by the Taiwan Respondents concerning the inappropriateness of using the simple average based on its extension of the analysis of the single standard deviation is inapposite.

Nonetheless, the value of the simple average also follows a similar pattern as that for the weighted average, where the simple average is just the special situation where the weights are equal. For the six examples in the Appendix, the values of the simple average, colored in

[160] Professor Coe's three principles are: (1) the value of the weighted average remains constant as the difference in the means changes, (2) the single standard deviation increases (or decreases) as the difference in the means increases (or decreases), and (3) the value of the single standard deviation is greater than the value of the weighted average, except that the two values are equal under the special circumstance when the difference of the means is zero (see the example from page 29 in the Taiwan Respondents' Comments).

purple, are constant (*i.e.*, **2050**) independent of the difference in the means. Thus, the simple average follows the first characteristic addressed by Professor Coe.

Commerce continues to find that a single standard deviation is not consistent with the academic literature, as agreed to by the Taiwan Respondents. However, the argument set forth by the Taiwan Respondents that this also leads to the support of a weighted average and the discredit of simple average based on their hypothesized Formula*** and self-selected examples as supporting evidence fails.

We disagree with Mid Continent's assertion that Commerce should question the assumption in the academic literature that there is a common standard deviation for the two populations. As part of Dr. Cohen's general formulation of his *d* coefficient, he defined the denominator, σ, as "the standard deviation of either population (since they are assumed equal)."[161] Mid Continent appears to conflate this common, within-population standard deviation for each group, *i.e.*, Group A and Group B, with a single standard deviation for all of the data in Group A and Group B together. As noted by Dr. Cohen, $\sigma_A = \sigma_B = \sigma$, and this, as discussed in the Draft Redetermination, is different from the standard deviation of the data from Groups A and B together.[162] Commerce sees no reason to question the assumptions underlying Dr. Cohen's presentation and formulation of his *d* coefficient, or its reasoning in the Draft Redetermination that the use of a single standard deviation is appropriate.

Mid Continent also argues that the use of the single standard deviation is not appropriate "due to the nature of the data being analyzed, one group will almost always have a much larger number of sales, and thus have an outsized influence on the calculation of the overall standard

[161] *See Cohen* at 20 and 27 ("σ is the common within-population standard deviation (*i.e.*, $\sigma_A = \sigma_B = \sigma$).")
[162] *See* Draft Redetermination at 14-19.

deviation."[163] We disagree. As discussed above, the inappropriateness of the single standard deviation is because it includes variation of the difference in the means as well as the variance of the data within both groups. It is not due to the difference in the weights (whether measured as the number of observations or the total sale quantities) between the two groups.

Further, Mid Continent misconstrues the role of weighting within each group to calculate the mean and the standard deviation, and the weighting that reflects the difference in the total sales volume between each group. As concluded above, the within group weighting is completely different than the question of weighting when averaging the standard deviations of the two groups to calculate the denominator of the Cohen's *d* coefficient.

COMMENT 4: WEIGHTED AVERAGING IS A REASONABLE METHODOLOGY FOR CALCULATING THE DENOMINATOR OF THE COHEN'S *D* COEFFICIENT; A SIMPLE AVERAGING IS NOT

***Cohen*, *Ellis*, and *Coe* Support Reliance on a Weighted Average Methodology**[164]

Taiwan Respondents' Comments

- The relevant chapters of *Cohen*, *Coe*, and *Ellis* on the record support the Taiwan Plaintiffs' position that "the cited literature nowhere suggests simple averaging for unequal-size groups."
- While *Cohen*, *Coe*, and *Ellis* focus on samples as distinguished from actual populations, what Commerce has not done, because it cannot do, is to explain why an analysis of all available data should be treated differently than an analysis of sample data.
- By relying on the Cohen's *d* methodology as the basis for its differential pricing analysis, but then rejecting the Cohen's *d* requirement to consider quantity in calculating the result, and in rejecting Cohen's reliance on pooled standard deviations, Commerce has "cherry picked" those parts of Cohen's *d* principle which lead to its desired result, and rejected those that do not.
- Cohen stated that when all data (rather than merely a sample) are available, weighting is required where the sizes of the groups differ.
- Because Cohen's qualifying "but" clause emphasizes that the (equal-sampling) effect size does not describe the population ("cannot be referred to the natural population with its varying group frequencies"), the application of the simple averaging methodology is explicitly limited and rendered inapplicable to the targeted dumping analysis.

[163] *See* Mid Continent Comments at 8.
[164] *See* Taiwan Respondents' Comments at 30-33.

- In *Mid Continent V*, the Federal Circuit resolved the dispute as to the meaning of Cohen's analysis.
- Cohen concludes that simple averaging does not apply to a natural population, and, even if it did, simple averaging would yield "artificial" results that do not reflect the population.

Commerce's Margin Calculation Methodology Supports Reliance on Weighted Average[165]

Taiwan Respondents' Comments

- Commerce's comparison of prices in two distinct groups of sales in two distinct markets (*i.e.*, home market and U.S. market) to determine a respondent's dumping margin is similar to Commerce's comparison of prices in two distinct markets (*i.e.*, test group and comparison group) to determine whether that respondent is target dumping.
- In the same manner as Commerce relies on weight averaging prices to calculate a dumping margin, Commerce should weight average the standard deviation of each group based on the quantity of U.S. sales to determine the pooled standard deviation.
- The judicial and administrative precedent[166] which require Commerce to rely on a weighted average over a simple average in calculating margins (as well as other similar calculations) are equally applicable to Commerce's calculation of the Cohen's *d* denominator.
- The courts have repeatedly recognized that weighted-average data generally lead to more accurate results than simple-averaged data. Weighted averages are preferred because simple averages inflate and distort the impact of small-quantity transactions.[167]
- Commerce itself has recognized the superiority of quantity-weighted averaging.[168]
- Commerce relies on weighted averaging for all other calculations required to complete the differential pricing (DP) analysis, but then pools the two weighted standard deviations without using any weighting. This methodology is inconsistent with the remainder of the Cohen's *d* formula.[169]
- As is the case with its margin calculations, Commerce conducts the identical "comparison" between "two distinct groups of data" whether each group encompasses a full population (as is normally the case) or a sample. If the meaningful comparison in a sample uses weighted-averaging, then the meaningful comparison in the population will use weighted-averaging as well.
- In the Cohen's *d* analysis, Test Groups and Comparison Groups are not distinct groups with distinct pricing patterns. These groups do not have an independent existence and as such are endowed with little or no special economic meaning.

---

[165] *Id.* at 33-37.
[166] *Id.* at 34 (citing *Diamond Sawblades Manufacturers' Coal. v. United States*, 37 ITRD 2191 (CIT 2015), *aff'd* 866 F.3d 1304 (Fed. Cir. 2017) and *MacLean-Fogg Co. v. United States*, 100 F. Supp. 3d 1349, 1359–64 (CIT 2015)).
[167] *Id.* at 34 (citing *Allied Tube & Conduit Corp. v. United States*, 132 F. Supp. 2d 1087, 1096 (CIT 2001) and *RZBC Group Shareholding Co. v. United Stat*es, 100 F. Supp. 3d 1288, 1309 (CIT 2015)).
[168] *Id.* at 35 (citing *Polyethylene Terephthalate Film, Sheet, and Strip from the People's Republic of China: Final Results of the 2009-2010 Antidumping Duty Administrative Review of the Antidumping Duty Order*, 77 FR 14493 (March 12, 2012), and accompanying IDM at 8 and *Fresh Garlic from the People's Republic of China: Final Rescission of New Shipper Reviews*, 76 FR 52315 (August 22, 2011), and accompanying IDM at Comment 8.
[169] *Id.* at 35-36 (citing *Mittal Canada, Inc. v. United States*, 461 F. Supp. 2d 1325, 1330 (CIT 2006)).

- In the weighted-average methodology, each kilogram is treated equally relative to all other kilograms sold for every one of the Period, Regional, and Purchaser tests. Because each kilogram of sales is correctly treated on an equal basis, regardless of whether it falls within a Test Group or a Comparison Group, relying on a weighted pooled standard deviation (SD) makes sense.

Weighted-Averaging Leads to Reasonable Results Based on the Record in this This Case; Simple-Averaging Does Not [170]

Taiwan Respondents' Comments

- Results differ when weight averaging, rather than simple averaging, is used to calculate the denominator in the Cohen's *d* equation. The existence of this difference, which Taiwan Respondents have identified throughout these proceedings, is not in dispute. The disagreement between the parties is whether simple averaging leads to a reasonable result.
- As Commerce recognizes, when the Test Group and Comparison Group have different quantities, the group with the larger quantity will have a greater impact on the ultimate result. What Commerce fails to recognize, however, is that this result makes sense; in this manner, each kilogram (or each nail) has an equal impact on the result, whether in the Test Group or the Comparison Group.
- In contrast, if each group is given equal weight, a kilogram in the smaller group will have a disproportionately larger impact on the results than a kilogram in the larger group. And the same kilogram, in the same sale, will have a different impact depending whether the sale falls within the Test Group or the Comparison Group.
- The numerator—whose calculation is not disputed—is the difference between two *quantity weighted* means. Thus, in Commerce's simple-averaging method, each kilogram affects the numerator of Cohen's *d* in one way but affects the denominator in a different way, undermining the validity of the resulting ratio. In contrast, the weighted averaging methodology uses consistent weights for each kilogram sold in both the numerator and the denominator, yielding a meaningful ratio.
- The Taiwan Respondents provide five examples of data showing the differences between relying on the simple averaging and weighted averaging methodology. The first example shows hypothetical data. The other four show actual sales reported by PT in its database, as analyzed by Commerce.
- These examples illustrate a predictable tendency for weighted averaging to make sense and produce a reliable determination, whereas simple-averaging does not. Simple averaging also fails to achieve the objective of Commerce's DP analysis; that is, of identifying evidence of targeted dumping.
- Reliance on simple averaging methodology leads to unreasonable results when the SD of a small-sized group is relatively large (unnaturally increasing the Cohen's *d* denominator, potentially turning pass into no-pass) or relatively small (unnaturally decreasing the Cohen's *d* denominator, potentially turning no-pass into pass).

[170] *See* Taiwan Respondents' Comments at 37-51.

- These results show that weighted average methodology is superior to the simple-average methodology. Examining the accuracy of results is the best method of confirming that a methodology is reasonable in the real world, in light of its statutory purpose. In contrast, a methodology which leads to absurd results cannot be sustained.
- The results of relying on weighted average or simple average only are identical (*i.e.*, the same *d*) when the quantity and SD of the Test Group and Comparison Group are identical; otherwise the *d* differs by varying amounts depending on the difference in weighted-average means, quantities, and SDs in each group. Commerce's simple average methodology would be flawed even if the ultimate results (*i.e.*, the dumping margin) were the same. However, the results are not the same, and result in a meaningful difference in the dumping margin.
- Relying on a weighted pooled standard deviation comports with the general purpose of Commerce's DP analysis. Commerce's goal is to ensure that high priced sales with negative dumping are not "masking" dumping of low-priced sales of comparable merchandise. Because dumping margins are calculated on a weighted-average basis, the extended margin of low quantity high-priced sales cannot offset the extended margin of higher quantity dumped sales. By using a simple average in the denominator, Commerce has impermissibly elevated the power of low quantity sales to have an impact incongruent with their actual size.
- Factoring quantity into the analysis is necessary to evaluate whether a price difference is significant. Miniscule sales quantities could create an affirmative determination of significant price differences unrelated to the significance of the quantities sold, which is unreasonable.
- The Court should require that Commerce calculate the pooled standard deviation in order to properly place the emphasis on economic reality. Simple averaging is not permissible if the result does not conform to economic reality and is not supported by substantial evidence.
- Commerce uses weights to compute average prices and SDs of all Test Groups and Comparison Groups but ignores the relationship of each subgroup SD to the whole when determining the denominator of the Cohen's *d* equation.

The "Statutory Context" Favors Weighted Averaging[171]

Taiwan Respondents' Comments

- The context of the statutory requirement supports reliance on weighted averaging. Commerce's differential pricing analysis is designed to determine whether "there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions or periods of time." 19 U.S.C. § 1677f-1 (d)(1)(B)(i).
- The "context" of Commerce's test requires that it apply a reasonable methodology to determine whether a price difference is significant, which leads to reasonable results, and which is supported by substantial evidence.

[171] *Id.* at 52-53.

- The simple averaging methodology applied by Commerce does not meet the "context" standard whereas a WA methodology does. The simple averaging methodology leads to results which are directly contrary to the statutory mandate to determine whether price differences are "significant."
- The simple averaging methodology gives different weights to one sale, depending on whether that sale falls within the Test Group or the Comparison Group. The simple averaging ignores Cohen's mandate that "{Effect sizes} must be indexed or measured in some defined unit appropriate to the data, test, and statistical model employed." In contrast, the WA methodology does not suffer from any of these fundamental defects.
- The courts may consider acceptance of Commerce's application of the Cohen's *d* test if Commerce employs a weighted average methodology to calculate the "significant difference" denominator.
- Citing *Stupp*,[172] the Taiwan Respondents reference an example of the Federal Circuit's discussions of the relationship of weight averaging and the overall reasonableness of Cohen's *d*. As demonstrated by this example, the context of the statute supports reliance on a weighted average.

Weighted Averaging Complies With ANOVA; Simple Averaging Does Not[173]

Taiwan Respondents' Comments

- The "Analysis of Variance" (ANOVA) concerns the relationship among data "spreads" and related quantities. For its DP analysis, the Taiwan Respondents assert that Commerce always weights the unit prices by quantities when computing the spread of any group of transactions.
- ANOVA exploits mathematical relationships between the spread of a batch of data (whether considered a "population," "sample," or something else) and the spreads within subgroups of that batch. Commerce computes the overall spread of the data by weighting unit prices by their quantities. The between spread is a unique, predictable function of the difference in weighted means of the Test Group and the Comparison Group. Commerce already computes and uses that difference in the numerator of its Cohen's *d* formula. Consequently, Commerce has no discretion as to how to compute the pooled spread.
- An example as to how ANOVA applies to this case is discussed with respect to a certain CONNUM. Using this example, the Taiwan Respondents assert that Commerce's simple averaging methodology results in a situation in which the three component SSes no longer sum to the SS for the entire group. That artifice creates an inconsistency in the mathematics, the statistics, and the meanings of the quantities used in the Cohen's *d* calculations. Thus, whatever the SA standard deviation "yardstick" might be, it does not correspond to anything in Cohen, no matter what Commerce chooses to call it. Thus, it does not produce a correct value of Cohen's *d*.
- This decomposition of the overall spread into three separate, identifiable, meaningful spreads achieved by ANOVA explains why weighted pooling is a valid methodology to determine whether there is a significant difference between Test Group and Comparison

[172] *See Stupp Corp. v. United States*, 5 F.4th 1341, 1359 (Fed. Cir. 2021) (*Stupp*).
[173] *See* Taiwan Respondents' Comments at 52-53.

Group prices, and simple averaging is not. This analysis resolves the question this Court posed in *Mid Continent III*: "Commerce gave no explanation of why weighting should not be done by "quantity" of units sold—here measured in kilograms—and what if any manipulation/predictability concern there would be if that path were followed." Appx656. It also establishes why relying on the simple averaging methodology leads to unpredictable, arbitrary and statistically unreasonable results.

Commerce's Rejection of a Pooled Standard Deviation Undermines its Analysis[174]

Taiwan Respondents' Comments

- Commerce appears to believe that by stating that it is not calculating a pooled standard deviation (PSD), its failure to comply with ANOVA no longer is relevant. The status of the SA method, then, is that it is a purely *ad hoc*, statistically unjustified, theoretically unsupported, and not systematically tested decision procedure.
- In the *Mid Continent VI Draft Results*, Commerce does not discuss whether it is relying on a PSD to calculate the Cohen's *d* denominator.
- Commerce's simple averaging methodology is not supported by any academic literature, leads to odd results, is contrary to economic reality and judicial precedent and is not reasonable.
- Commerce cannot have it both ways. Specifically, Commerce cannot totally abandon the academic literature when its methodology is not supported by the literature and then turn around and reject an alternative methodology because the academic literature does not support that methodology. Commerce's treatment of the literature underscores the deficiencies of its Draft Results analysis.
- Using a single standard deviation when weight averaging the quantity in a Test Group and Comparison Group leads to a much more reasonable result than using a single standard deviation when taking a simple average of the SDs of each group. Thus, while the single standard deviation solution proposed by the *Mid Continent V* court does not strictly conform to academic literature, it arguably could be deemed reasonable under law.
- Commerce fails to consider whether its interpretation leads to reasonable results. It also fails to consider whether the results of its methodology conform to the statutory mandate of determining whether there is a significant difference in prices between the two groups.

Rational Economic Behavior Requires That Commerce Consider Quantity in Determining the Pooled Standard Deviation[175]

Taiwan Respondent's Comments

- In this case, mean prices represent dollars per kilogram ($/kg), variances are weighted by kilograms and SDs are weighted by square roots of kilograms. The fact that the analysis involves known quantities requires that it consider quantities sold. By ignoring

[174] *Id.* at 58-59.
[175] *Id.* at 59-61.

quantities, Commerce distorts its analysis as to whether prices in the two groups are significantly different.

- Taiwan Plaintiffs discuss how prices are based on supply and demand and assert that a price without a quantity is as meaningless for economic analysis as a quantity without a price.
- The issue of interest in this case is the significance of the difference in weighted average prices per kilogram in a Test Group and a Comparison Group. This significance cannot be determined without considering quantity.
- The weighted averaging methodology treats each kilogram of nails equally regardless of whether the kilogram appears in a Test Group or a Comparison Group. In contrast, under Commerce's simple averaging methodology a particular transaction may have a relatively large weight and another time have a relatively small weight depending on whether the transaction falls within a Test Group or (one of several possible) Comparison Groups.
- Weighted averaging is the only pooling method that assures consistency in all possible cases, no matter how the transactions might be split into groups for comparison. For this reason, Commerce's differential pricing methodology must use weighted averaging pooling.

Weighted Averaging Should be Based on Quantity Sold [176]

Taiwan Respondent's Comments

- The focus in this case is not the number of transactions (*i.e.*, number of sales); rather, it is the quantity of kilograms sold. The mean prices of the Test Group and the Comparison Group reflect prices per kilogram, rather than the total amounts of each sale.
- The difference in mean prices depends on the price per kilogram, and not on a price per sale. Commerce also weights the SDs of each group to reflect quantities sold. Thus, the SDs are not based solely on the number of sales.
- The Taiwan Plaintiffs acknowledge that neither Cohen, Ellis, nor Coe discuss the issue as to whether a weighted average should be based on quantity or number of sales and assert that it is not surprising, because Cohen's *d* and its progeny were designed to be used by social scientists to measure differences among persons, in which each observation (*i.e.*, person) represents one unit, and not to measure whether two prices are significantly different, in which each observation (*i.e.*, sale) represents multiple units (of varying multiplicities).
- Relying on a weighted averaging methodology based on kilograms to calculate a pooled SD is a reasonable way to achieve a statistically and mathematically valid comparison between the numerator and denominator.

---

[176] *Id.* at 61-62.

Commerce's Position:

Use of a Simple Average Is Reasonable

In general, the Taiwan Plaintiffs assert that the use of a weighted average leads to reasonable results and the use of a simple average leads to unreasonable results. The Taiwan Plaintiffs use five examples to demonstrate that the results after using a weighted average are reasonable. For the first three examples, which include the hypothetical example and the two examples based on PT's U.S. sale prices where the change from a simple average to a weighted average causes the Cohen's *d* test to change from pass to no-pass, Taiwan Plaintiffs conclude, based on the graphical presentation of the prices, that the prices in the test and comparison groups appear similar which shows that there is no significant difference. For the two examples based on PT's U.S. sale prices where the change from a simple average to a weighted average causes the Cohen's *d* test to change from no-pass to pass, the large variance in the small group is "characteristic of 'masked dumping'" which Commerce's use of a simple average has failed to detect. The Taiwan Plaintiffs conclude that these erroneous results are caused by the use of a simple average which gives too much weight to the smaller group of sales. The Taiwan Plaintiffs insist that the results of the application of a specific methodology must be reasonable, and the results of Commerce's Cohen's *d* test based on PT's U.S. prices in this investigation, based on these examples, are absurd.[177]

Commerce disagrees with the Taiwan Plaintiffs that the proffered examples demonstrate that the use of a weighted average is reasonable and that the use of a simple average is not. The only differences are the results themselves, and the arithmetic logic that different outcomes result when different weights are used to combine the standard deviations in the denominator of the

[177] *Id.* at 37-49.

Cohen's *d* coefficient. Contrary to the claims by the Taiwan Plaintiffs, the graphical representations of the test and comparison groups of prices do not demonstrate a conclusion that a given set of compared prices represents a Cohen's *d* coefficient that is larger or smaller than 0.8, *i.e.*, that the differences in prices are significant or not. In other words, there is no visual distinction between any of the graphical representations of the test and comparison group prices which would lead a reasonable observer to recognize that one difference in prices pass the Cohen's *d* test and another difference in prices does not pass the Cohen's *d* test, irrespective of whether a simple average or a weighted average is used. The only visually recognizable pattern in the Taiwan Plaintiffs' graphical representations is whether the small group, *i.e.*, the group with few observations, has a larger or smaller variance in the prices. As the Taiwan Plaintiffs describe the consequences of the larger or smaller variance in prices for the smaller group of prices relative to the larger group of prices:

> The accurate visual display of spreads and weights reveals how SA overweights the smaller group (the one with lower total quantity; *i.e.*, kilograms sold). When the smaller group has a small spread, the {simple averaging} methodology decreases the pooled standard deviation ("PSD"). Because that PSD is the denominator of the Cohen's d formula, the result is an incorrect increase in Cohen's *d*. This can cause a low "no-pass" value of *d* to exceed Commerce's threshold of 0.80, incorrectly resulting in a "pass." [178]

The Taiwan Plaintiffs' argument is simply an arithmetic tautology. When the weights for averaging two values change from being identical (*e.g*., one) to being non-equal values, the results will change. When the weights are based on the sales quantities of each group, the smaller group will have less weight than the larger group, and the value being average (*i.e.*, the standard deviation) will have a smaller impact on the calculated average, and conversely the value of the larger group will have a larger impact. If the standard deviation of the smaller group

[178] *Id.* at 46.

is small, then the calculated average will be larger and the Cohen's *d* coefficient will be smaller. If the standard deviation for the smaller group is larger, then the calculated average will be smaller and the Cohen's *d* coefficient will be larger. Reliance on such arithmetic logic to invent support for the reasonableness of a weighted average is without merit.

Taiwan Plaintiffs quote and agree with the above paragraph but insist that Commerce fails to recognize that the result makes sense.[179] However, as Commerce has reiterated multiple times, this is results-oriented, and the results does not make a methodology reasonable. The Taiwan Plaintiffs merely describe how weighting would work.[180]

Although the Taiwan Plaintiffs imply that using a simple average in the Cohen's *d* denominator is "contrary to economic reality,"[181] the Taiwan Plaintiffs do not enumerate what that "economic reality" is for PT or how Commerce has failed to address it with Commerce's use of a simple average instead of a weighted average. The Taiwan Plaintiffs' presumption appears to be that "economic reality" is that its U.S. sale prices do not differ significantly, and that "economic reality" would inform Commerce that the use of a simple average is unreasonable and the use of a weighted average is appropriate. However, the Taiwan Plaintiffs' presumption lacks a foundation in the record.

In sum, the Taiwan Plaintiffs' claim that the use of a weighted average is reasonable is based, in part, on the results which it generates. The fact that the results of the proposed methodology benefit the proposer of the methodology does not provide support for the reasonableness of the methodology. As the Taiwan Plaintiffs state "Reliance on {a weighted average}, rather than {a simple average}, results in PT's margin being reduced from 2.16 %

---

[179] *Id.* at 38.
[180] *Id.*
[181] *Id.* at 3, 50-51, 59.

percent to *de minimis*. Thus, Commerce's choice of methodology has 'a material impact on the results of the less-than-fair-value investigation.'"[182]

Moreover, outside of the context of the First Amendment, judicial challenges that are based on hypotheticals are disfavored.[183] Because the facts of a hypothetical example are developed for a set purpose, an alternative hypothetical example, based on different facts, could be proposed which would demonstrate the opposite, contrary result.

Accuracy of Results

The Taiwan Plaintiffs assert that "{w}eighted-average data generally lead to more accurate results than simple-averaged data."[184] We find that the judicial opinions and Commerce determinations that the Taiwan Plaintiffs cite for this proposition are not relevant to the issue in this investigation. In those cases, the Court and Commerce found in the particular circumstances of each case that use of a weighted average was more accurate than using a simple, unweighted, equally weighted average. However, none of the cited cases involved the advanced statistical concepts at issue in this case.

In general, Commerce agrees that the use of a weighted average is appropriate in many situations, such as in the calculation of a weighted-average U.S. or comparison market price, a period-wide weighted-average cost of production, and a weighted-average dumping margin. Each of these calculations provides a single measure which aggregates a given value, *e.g.*, price, for a group of observations. As noted by the Taiwan Plaintiffs, Commerce also uses a weighted

[182] *Id.* at 49.
[183] *See*, *e.g.*, *Wash. State Grange v. Wash. State Republican Party*, 552 U.S. 442, 450 (2008) (stating in the context of a facial challenge to a statute "we must be careful not to go beyond the statute's facial requirements and speculate about 'hypothetical' or 'imaginary' cases.").
[184] *See* Taiwan Respondents' Comments at 34.

average to calculate the mean price of the test and comparison groups, as well as a weighted standard deviation of the prices within each group.

However, here in the Cohen's *d* test, Commerce is comparing the prices to a given purchaser, region or time period with the prices of comparable merchandise. The purpose of the Cohen's *d* test is to determine whether the difference in prices between these two groups is significant. As such, Commerce is directed to compare the prices to each given purchaser, region and time period to prices to all other purchasers, regions, or time periods. In the analysis of the difference in the means, two distinct groups of data are compared to determine "the degree to which the phenomenon is present in the population" such that the two groups of data being compared do not share the same characteristics.

Contrary to Taiwan Plaintiffs' suggestion, this calculation does not allow Commerce "to determine whether that respondent is target {sic} dumping."[185] To determine whether dumping exists requires both a U.S. price and a normal value (*i.e.*, home market price, third country price, or constructed value). Therefore, Commerce cannot determine whether dumping exists based on the Cohen's *d* test.

As explained above, for Commerce's Cohen's *d* test, Commerce found that a simple average is reasonable to calculate the denominator. Commerce accepts that academic literature assumes sampling for all the different ways the denominator of the Cohen's *d* coefficient can be calculated. However, Commerce uses the full populations in Commerce's Cohen's *d* test. Dr. Cohen's equation 2.3.2 required equal sample sizes, and sample sizes represent reliability. Because the reliability of the groups is equal when full populations are used, Commerce finds

---

[185] *Id.* at 33, 49.

that using a simple average to calculate the denominator of Commerce's Cohen's *d* test is appropriate.

Nonetheless, the Taiwan Plaintiffs insist that there is no distinction between an analysis based on sampled data and an analysis based on full populations of data. The Taiwan Plaintiffs claim that Commerce has failed "to explain why an analysis of all available data should be treated differently than an analysis of sample data."[186] However, Commerce has explained that sample size represents reliability, Dr. Cohen's equation 2.3.2 requires equal sample sizes, and full populations have equal reliability.

<u>Populations and Group Size</u>

Taiwan Plaintiffs argue that "{g}roup size matters, whether the group is a sample or the entire population."[187] Commerce has considered the group size limitations in the academic literature. The academic literature does not discuss the size of the population when the full population is used for the Cohen's *d* test. However, for samples, the academic literature discusses how the size of the sample is related to reliability. The academic literature's limitations on sample size for the different ways of calculating the denominator of the Cohen's *d* test are whether the sample sizes are equal or unequal. Because the groups are equally reliable when the sample sizes are equal and because when full populations are used, they too are equally reliable, Commerce finds it appropriate to use Dr. Cohen's equation 2.3.2 when full populations are used.

Further, the Taiwan Plaintiffs allege that "Commerce has 'cherry picked' {the} Cohen's *d* principle which {led} to its desired result, and rejected those that do not."[188] As an initial

[186] *Id.* at 31.
[187] *Id.*
[188] *Id.*

matter, none of the academic texts on the record here discuss Commerce's "Cohen's *d* test." Neither Dr. Cohen, nor Dr. Ellis, nor Professor Coe opined on the application of the concept of effect size to examine whether prices differ significantly among purchasers, regions or time periods under the antidumping statute. Nor could one reasonably expect an academic author to be omniscient and describe all possible applications of his or her concepts, including the situation addressed by Commerce in the use of its Cohen's *d* test. Similarly, these academic authors do not know the myriad situations in which their concepts may be applied. Such expectations are unrealistic that any applications must be preordained by an academic author rather than their concepts being adapted and applied in situations unimagined by the original authors. Nonetheless, these academicians did describe the general principles behind both the concept of effect size and its place in research and data analysis which Commerce has applied in its differential pricing analysis. Commerce has followed these principles in conceptualizing and applying the Cohen's *d* test.

<u>Correlation Coefficient Is Not Relevant</u>

The Taiwan Plaintiffs argue that Dr. Cohen states that when "populations {A and B} are concrete and unequal collections of cases, the inequality should figure in the assessment of the degree of the relationship." as reflected in Dr. Cohen's equation 2.2.7.[189] However, Dr. Cohen's discussion involves "***d*** in terms of correlation and proportion of variance" where "membership in the A or in the B population may be considered to be a simple dichotomy or a two point scale."[190] This differs considerably from Commerce's use of effect size in the context of a difference in means. First, this involves an analysis of data that is a "simple dichotomy or a two point scale," such as a yes or no, or "for example, 0 for membership in A and 1 for membership

[189] *Id.* at 32.
[190] *See Cohen* at 23.

in B (the values assigned are immaterial)."[191] In Commerce's application of effect size and use of Dr. Cohen's *d* statistic, the data is of a continuous variable (*i.e*., the price).[192] In contrast, where effect size is based on dichotomous variables,[193] the effect size is based on the relationship between probability of one or the other value of the two point scale.[194] Dr. Cohen's equations 2.2.6 (for "equally numerous" populations) and 2.2.7 (for "unequal collections") simply states the relationship between the *d* coefficient and "Pearson product-moment correlation coefficient (***r***)."[195] This aspect of effect size is unrelated to the concept used by Commerce as the basis for the Cohen's *d* test, and, therefore, the argument by the Taiwan Respondents is not relevant to the issue of whether Commerce's use of a simple average in its application of the Cohen's *d* test is reasonable.

<u>ANOVA Is Not Relevant</u>

The Taiwan Plaintiffs argue that "group size" is relevant in the context of an ANOVA analysis.[196] Certain quotations from Dr. Cohen's discussion of ANOVA[197] were initially included in an expert's statement included as an attachment to Mid Continent's comments on the Second Draft Redetermination. Based on Mid Continent's comments and the expert's statement, Commerce included Mid Continent's concept of "a 'natural population' {that} can be viewed as

---

[191] *Id*.

[192] *See Ellis* at 9 ("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.").

[193] *Id.* at 7 ("When we compare groups on dichotomous variables (*e.g*., success versus failure, treated versus untreated, agreements versus disagreements), comparisons may be based on the probabilities of group members being classified into one of the two categories."); *see also Mid Continent V*, 31 F.4th at 1378-1379 ("The discussion in that section involves ***f***, an effect size index that is related to, but not the same as, the Cohen's *d* coefficient, applicable when there are arbitrarily many groups to compare, rather than just two." (internal citation omitted)).

[194] *See Ellis* at 7-9; *see also Ellis* at 13 ("Groups compared on dichotomous outcomes" in contrast with "Groups compared on continuous outcomes").

[195] *See Cohen* at 23-24.

[196] *See* Taiwan Respondents' Comments at 53-58.

[197] *See Ellis* at 12 ("Cohen's ***f*** and $\boldsymbol{f}^2$ are used in connection with the F-tests associated with ANOVA and multiple regression (Cohen 1988). In the context of ANOVA Cohen's ***f*** is a bit like a bigger version of Cohen's *d*. While ***d*** is the standardized difference between two groups, ***f*** is used to measure the dispersion of means among three or more groups.")

an 'abstract effect'" as part of its support for using a simple average in the Second Redetermination.[198] However, the Federal Circuit found that the "abstract effect" and other arguments to be unpersuasive to support Commerce's use of a simple average,[199] and in *Mid Continent V* remanded the issue back to Commerce. In these final results of redetermination, Commerce does not rely on the concept of "natural population" or "abstract effect" because: (1) the Federal Circuit has already rejected that conceptual argument; and (2) an ANOVA analysis, itself, Dr. Cohen's *f* coefficient, as a "measure {of} the dispersion of means among three or more groups," is a distinct concept that is different from Dr. Cohen's *d* coefficient as a measure of the difference in the means of two groups. Therefore, the argument by the Taiwan Respondents based on ANOVA is not relevant to the issue of whether Commerce's use of a simple average is reasonable in its application of the Cohen's *d* test.

The Taiwan Respondents fail to explain how an ANOVA analysis is relevant to Commerce's analysis of the difference in the mean prices, either based on the academic literature or otherwise. ANOVA, or "Cohen's *f*," "quantifies the dispersion of means in three or more groups….."[200] Their claim that Commerce's analysis is somehow dependent upon ANOVA is illogical.[201]

Economic Behavior and Weight Averaging

The Taiwan Plaintiffs state that quantity is relevant when calculating the variance and standard deviations of the test and comparison groups. Further, Taiwan Plaintiffs state that prices are determined by the relationship between supply and demand. The Taiwan Plaintiffs

---

[198] *See* Second Redetermination at 40-41, 45-46.
[199] *See Mid Continent V*, 31 F.4th at 1378-79 ("Nothing in the section {i.e., Cohen at 359-361 on ANOVA and an "abstract effect"} applies simple averaging to pooled standard deviation estimates for different-size groups.").
[200] *See Ellis* at 14.
[201] *See* Taiwan Respondents' Comments at 53-58.

further argue that "the difference in weighted average prices per kilogram … cannot be determined without consideration of quantity," concluding that "Commerce's differential pricing methodology must use {weighted average} pooling."[202] However, the Taiwan Plaintiffs omit any explanation or logic which connects these disparate theoretical and applied concepts, and, therefore, these arguments provide no support for their conclusion that a weighted average must be used to calculate the denominator of the Cohen's *d* coefficient.

The Federal Circuit's Opinion in *Stupp*

In *Mid Continent V*, the Federal Circuit stated:

> Commerce observes that the cited literature discusses "sampling" from a population, whereas Commerce has the entire population data and each of its test-comparison group pairs involves the entire population. J.A. 1109. In *Stupp*, we stated that Commerce had not explained how this difference bears on the reasonableness of Commerce's use of Cohen's *d* in certain respects not at issue in the present matter. 5 F.4th at 1360.[203]

When it remanded this issue to Commerce, the Federal Circuit recognized that it has also remanded to Commerce "other aspects of Commerce's use of Cohen's *d*."[204] In their comments concerning the Draft Redetermination, the Taiwan Respondents quote from *Stupp*,[205] implying that the issue remanded in *Stupp* involves the use of a simple average and that "Commerce arguably can save Cohens *d* by replacing the {simple averaging} methodology with {a weighted averaging} methodology."[206] The issue in *Stupp* involves whether the test and comparison groups must satisfy certain statistical criteria (*i.e.*, the normality of the distributions, equal variances and sufficient sample size),[207] and not whether a simple average is reasonable to use

[202] *See* Taiwan Respondents' Comments at 59-61.
[203] *See Mid Continent V*, 31 F.4th at 1380.
[204] *Id*. at 1381.
[205] *See* Taiwan Respondents' Comments at 61-62.at 53 (citing *Stupp*, 31 F.4th at 1359).
[206] *Id.* at 31, n.14.
[207] *See Stupp*, 5 F.4th at 1360 ("We therefore remand to give Commerce an opportunity to explain whether the limits on the use of the Cohen's *d* test prescribed by Professor Cohen and other authorities were satisfied in this case or

when calculating the denominator of the Cohen's *d* coefficient. In the *Stupp* litigation, Commerce has addressed the Federal Circuit's concerns in its redetermination pursuant to remand order in that litigation and the CIT has sustained Commerce's further explanation.[208]

Commerce's Previous Clarification of a Pooled Standard Deviation Does Not Undermine its Analysis

Commerce has determined, consistent with the Court's prior decisions, that it must go beyond the academic literature to explain our use of a simple average to calculate the denominator of the Cohen's *d* coefficient and we have supported these arguments herein. As explained above, using a simple average is reasonable, while using a single standard deviation is not. Commerce's rejection of a single standard deviation supports our methodology and does not undermine our analysis.

COMMENT 5: RESPONDENTS' PROPOSED METHODOLOGY PERMITS MANIPULATION; A SIMPLE AVERAGE IS A NEUTRAL ALTERNATIVE

Mid Continent's Comments

- The Federal Circuit has twice rejected Commerce's approach, which is based on convincing the Court and the Federal Circuit that Commerce's interpretation of the academic literature is correct.
- If Commerce continues to rely on this approach, it risks being ordered to use a weighted average or the standard deviation of the entire population.
- Even if Commerce chooses to continue defending its reading of the literature, it can, and should, expand in the final remand results to explain the practical reasons why its approach is a reasonable use of its discretion.
- This discretion extends to using a simple average when calculating the denominator of the Cohen's *d* coefficient in order to minimize or eliminate respondents' ability to exploit the differential pricing analysis to mask dumping.
- The Taiwan Respondents' preferred methodology "would raise the possibility of data manipulation to "'dump-proof' a {respondent's} sales data in Commerce's calculations."[209]

whether those limits need not be observed when Commerce uses the Cohen's *d* test in less-than-fair-value adjudications.")

[208] *Stupp Corp. v. United States*, 619 F. Supp. 3d 1314 (CIT 2023), appeal docketed as Federal Circuit No. 2023-1663.

[209] *See* Mid Continent Comments at 3.

- "Commerce should support its position by reference to the reasonable, enforcement-oriented policy choices that are involved."[210]
- The Taiwan Respondents' proposal, to use a weighted average, using the quantity of sales in each group, "produces the opposite of what Commerce is trying to measure."[211]
- As examples provided in the comments show, a respondent can manipulate the sales data and the outcome of the Cohen's *d* calculations by changing the relative volume of sales in the base and comparison groups.
- Commerce's use of a simple average obviates the potential for manipulation and represents a reasonable discretionary choice to fulfill its statutory authority, enhance the effectiveness of the antidumping laws, and mitigate against manipulation.
- Commerce has the discretion to adopt a medium effect size threshold of 0.5, rather than the large effect size threshold of 0.8 that it currently uses.
- Adopting the medium effect size could further curtail impact of any manipulation.
- Commerce need not be wedded to strictly defined labels in implementing its differential pricing methodology; neither should the methodology be beholden to instructions that are applicable to sample-based studies when Commerce's analysis uses complete population data.
- Cohen's *d* refers to a number of related statistical techniques, not just one strictly defined formula.
- Commerce's differential pricing methodology uses a variation of the Cohen's *d* measure given the circumstances of the data and the specific purposes of Commerce's analysis.
- Commerce has the discretion to use a simple average to maximize the effectiveness of the antidumping methodology and to best effectuate the statutory requirement to assess whether "there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly"[212] independent of quantities, and to minimize or eliminate a party's ability to manipulate sales data to mask dumping.

Commerce Position:

Commerce disagrees with Mid Continent that Commerce has merely explained again its interpretation of the academic literature. Unlike the Third Redetermination, in which Commerce re-examined and explained its interpretation of the academic literature, in the Draft Redetermination and in these final results of redetermination, Commerce provides a "non-academic" explanation of why using a simple average to calculate the denominator of the Cohen's *d* coefficient is reasonable in light of *Mid Continent V*. However, for Commerce's

[210] *Id.* at 4.
[211] *Id.*
[212] *Id.* at 11 (citing section 777A(d)(1)(B)(i) of the Act).

differential pricing analysis, Commerce uses full populations and does not use samples. Because Commerce accepts that the academic literature assumes sampling for all options, the academic literature does not address whether a simple average or a weighted average should be used to calculate the denominator of the Cohen's *d* coefficient when using full populations. Therefore, Commerce is using general statistical principles to determine the appropriate method of calculating the denominator and these general statistical principles appear in the academic literature; however, this should not be confused with Commerce's interpretation of the academic literature, which the CIT rejected as inconsistent with *Mid Continent V*. As explained above, we agree that standard deviation and standard error, as well as equality of reliability and equality of size, though interrelated, are separate concepts. However, given their interrelationship, it is important that Commerce uses these general statistical principles to provide its practical justification sought by the Court.

Regarding Mid Continent's argument that Commerce should support its determination with additional policy-based considerations, Commerce generally agrees that the prevention of manipulation is a concern when applying its methodologies in antidumping duty proceedings. Commerce intends to remain vigilant in the prevention of opportunities for manipulation. However, as explained above, in this remand, Commerce has used general statistical principles to provide a "non-academic" justification for why using a simple average to calculate the denominator of the Cohen's *d* test is reasonable in the context of the differential pricing methodology.

V. FINAL RESULTS OF REDETERMINATION

Pursuant to the *Remand Order*, we have reconsidered the use of a simple average to calculate the denominator of the Cohen's *d* coefficient in light of *Mid Continent VI* and the

Federal Circuit's opinion in *Mid Continent V*. Based on Commerce's consideration of the concerns outlined by the Court in this redetermination, and the issues raised by the interested parties, we maintain that Commerce's use of a simple average is consistent with the statute and is reasonable in its examination of whether prices differ significantly pursuant to section 777A(d)(1)(B)(i) of the Act. Accordingly, based on the results of our analysis, the estimated weighted-average dumping margins calculated in the *First Redetermination*[213] remain unchanged.

8/31/2023

X 

Signed by: LISA WANG

Lisa W. Wang
Assistant Secretary
for Enforcement and Compliance

[213] *See Mid Continent Steel & Wire, Inc. v. United States*, 219 F. Supp. 3d 1326 (CIT 2017); *see also Final Results of Redetermination Pursuant to Court Remand, Mid Continent Steel & Wire, Inc. et al. v. United States*, Court No. 15-00213 (CIT March 23, 2017), dated June 21, 2017 (*First Redetermination*), available at https://access.trade.gov/resources/remands/17-31.pdf. The CIT remanded the calculation of PT's G&A expense ratio, which Commerce recalculated in the *First Redetermination*. The CIT affirmed Commerce's recalculation of PT's G&A expense ratio in *Mid Continent Steel & Wire, Inc. v. United States*, 273 F. Supp. 3d 1161 (CIT 2017).

APPENDIX

Page 24

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	sum wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
100	1	100	300	2	150	2500	2500	5000	2500	50				
200	1	200				2500	2500							
											1750	41.83300	2050	45.27693
100	8	800	1200	10	120	400	3200	16000	1600	40				
200	2	400				6400	12800							
100	1	100	1500	12	125	625	625	22500	1875	43.30127				
200	1	200				5625	5625							
100	8	800				625	5000							
200	2	400				5625	11250							

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	30	-equals-	125	-diff-	125

Page 27

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
130	1	130	360	2	180	2500	2500	5000	2500	50				
230	1	230				2500	2500							
											1750	41.83300	2050	45.27693
100	8	800	1200	10	120	400	3200	16000	1600	40				
200	2	400				6400	12800							
130	1	130	1560	12	130	0	0	27000	2250	47.43416				
230	1	230				10000	10000							
100	8	800				900	7200							
200	2	400				4900	9800							

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	60	-equals-	500	-diff-	500

APPENDIX

Page 29

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
70	1	70	240	2	120	2500	2500	5000	2500	50				
170	1	170				2500	2500							
											1750	41.83300	2050	45.27693
100	8	800	1200	10	120	400	3200	16000	1600	40				
200	2	400				6400	12800							

70	1	70	1440	12	120	2500	2500	21000	1750	41.83300
170	1	170				2500	2500			
100	8	800				400	3200			
200	2	400				6400	12800			

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	0	-equals-	0	-diff-	0

Page 24 **INCREASE PRICES IN GROUP A BY 100 -- TO 200 AND 300**

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	sum wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
200	1	200	500	2	250	2500	2500	5000	2500	50				
300	1	300				2500	2500							
											1750	41.83300	2050	45.27693
100	8	800	1200	10	120	400	3200	16000	1600	40				
200	2	400				6400	12800							

200	1	200	1700	12	141.6667	5625	5625	52500	4375	66.14378
300	1	300				30625	30625			
100	8	800				625	5000			
200	2	400				5625	11250			

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	130	-equals-	2347.222	-diff-	2625

APPENDIX

Page 24 **INCREASE PRICES IN GROUP B BY 30 -- TO 130 AND 230**

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	sum wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
100	1	100	300	2	150	2500	2500	5000	2500	50				
200	1	200				2500	2500							
											1750	41.83300	2050	45.27693
130	8	1040	1500	10	150	400	3200	16000	1600	40				
230	2	460				6400	12800							

100	1	100	1800	12	150	625	625	28500	2375	48.73397
200	1	200				5625	5625			
130	8	1040				25	200			
230	2	460				11025	22050			

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	0	-equals-	0	-diff-	625

Page 24 **INCREASE PRICES IN GROUP B BY 50 -- TO 150 AND 250**

price	qty	ext val	tot val	tot qty	mean	dif**2	wgt dif	sum wgt dif	var	std	wgt avg of var		simple avg of var	
a	b	c=a*b	d=sum(c)	e=sum(b)	f=d/e	g=(a-h)^2	h=g*b	i=sum(h)	j=i/e	k=sqrt(j)	m=WA sum	n=sqrt(m)	p=SA sum	q=sqrt(p)
100	1	100	300	2	150	2500	2500	5000	2500	50				
200	1	200				2500	2500							
											1750	41.83300	2050	45.27693
150	8	1200	1700	10	170	400	3200	16000	1600	40				
250	2	500				6400	12800							

100	1	100	2000	12	166.6667	625	625	42500	3541.667	59.51190
200	1	200				5625	5625			
150	8	1200				625	5000			
250	2	500				15625	31250			

	relative weight		diff means		claimed		actual
	group A	group B			diff var		diff var
FORMULA***	0.166667	0.833333	20	-equals-	55.55556	-diff-	1791.667