A-583-854
Remand
CAFC 21-1747
POI: 04/01/2013 – 03/31/2014
Public Document
E&C/OVI: CV/GM

Mid Continent Steel & Wire, Inc. v. United States
CAFC Case 21-1747 (CAFC April 21, 2022)
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 of the U.S. Court of Appeals for the Federal Circuit (CAFC) in *Mid Continent Steel & Wire, Inc. v. United States*, 31 F.4th 1367 (Fed. Cir. April 21, 2022) (*Mid Continent V*), and the subsequent remand order issued by the U.S. Court of International Trade (CIT) in *Mid Continent Steel & Wire, Inc. v. United States*, Consol. Court No. 15-00213 (CIT June 14, 2022) (*Remand Order*). These final results of redetermination concern one issue in the less-than-fair-value (LTFV) investigation of certain steel nails (nails) from Taiwan.[1]

In *Mid Continent V*, the CAFC vacated and remanded the CIT's judgment[2] sustaining Commerce's use of a simple average when calculating the denominator of the effect size (*i.e.*, the "Cohen's *d* coefficient"), a part of the Cohen's *d* test in Commerce's differential pricing analysis in the LTFV investigation of nails from Taiwan. As we explain below, on remand, we have complied with *Mid Continent V* by providing further explanation of Commerce's methodology.

[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 Steel & Wire, Inc. v. United States*, 495 F. Supp. 3d 1298 (CIT 2021) (*Mid Continent IV*).

On September 13, 2022, Commerce released the Draft Results of Redetermination to interested parties.[3] On September 30, 2022, Mid Continent and the Taiwan Respondents[4] submitted comments.[5] On October 25, 2022, Commerce identified new factual information included in the comments from the Taiwan Respondents and requested that the Taiwan Respondents redact the new factual information outlined in Commerce's Letter of October 25, 2022. On October 26, 2022, Taiwan Respondents requested that Commerce reconsider its decision to reject certain portions of their comments from the record.[6] On October 27, 2022, the Taiwan Respondents resubmitted their redacted comments on the Draft Results of Redetermination.[7] On November 1, 2022, Commerce issued a letter to the Taiwan Respondents, allowing them to refile certain additional portions of their rejected comments.[8] On November 3, 2022, the Taiwan Respondents submitted comments incorporating additional text that Commerce listed in its letter that reconsidered rejection.[9]

---

[3] *See* Draft Results of Redetermination Pursuant to Court Remand, *Mid Continent Steel & Wire, Inc. v. United States*, CAFC Case 21-1747 (CAFC April 21, 2022), dated September 13, 2022 (Draft Redetermination).
[4] The comments are submitted on behalf of 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, the Taiwan Respondents or Taiwan Plaintiffs). PT Enterprises, Inc. (PT) was a mandatory respondent in the less-than-fair-value investigation.
[5] *See* Mid Continent's Letter, "Certain Steel Nails from Taiwan: Comments on Draft Results of Redetermination Pursuant to Court Remand in Mid Continent Steel & Wire, Inc. v. United States, CAFC Case 21-1747 (CAFC Apr. 21, 2022)," dated September 30, 2022 (Mid Continent's Comments); *see also* the Taiwan Respondents' Letter, "Comments of Taiwan Plaintiffs on Draft Redetermination Pursuant to Court Remand, (CAFC No. 21-1747, April 21, 2022)," dated September 30, 2022 (TR Comments).
[6] *See* Taiwan Respondents' Letter, "Taiwan Respondents Request for Reconsideration of Commerce's Rejection of Comments on Draft Redetermination Pursuant to Court Remand, (CAFC No.21-1747, April 21, 2022)," dated October 26, 2022.
[7] *See* Taiwan Respondents' Letter, "Taiwan Respondents Resubmission of Comments on Draft Redetermination Pursuant to Court Remand, (CAFC No. 21-1747, April 21, 2022)," dated October 27, 2022 (TR Resubmitted Comments). On October 27, 2022, Commerce issued a letter instructing the Taiwan Respondents to submit the redacted version of the TR Comments by the stated deadline of 5:00 pm on October 27, 2022, while Commerce's decision regarding reconsideration of its rejection of the TR Comments was pending. On October 27, 2022, Mid Continent submitted a letter opposing the TR Respondents' request for reconsideration. *See* Mid Continent's Letter, "Certain Steel Nails from Taiwan: Opposition to Request for Reconsideration," dated October 27, 2022.
[8] *See* Letter to Taiwan Respondents, "CAFC 21-1747, Certain Steel Nails from Taiwan: Reconsideration of Comments of Taiwan Plaintiffs on Draft Results of Redetermination," dated November 1, 2022.
[9] *See* Taiwan Respondents' Letter, "Taiwan Respondents Second Resubmission of Comments on Draft Redetermination Pursuant to Court Remand, (CAFC No. 21-1747, April 21, 2022), dated November 3, 2022 (TR Second Resubmitted Comments).

Commerce addresses the comments from the interested parties below.

II. BACKGROUND

On May 20, 2015, Commerce published its *Final Determination*, in which we applied a differential pricing analysis to determine whether we could use an alternative comparison method to calculate each respondent's estimated weighted-average dumping margin pursuant to section 777A(d)(1)(B) of the Tariff Act of 1930, as amended (the Act).[10] As part of this analysis, Commerce performed a "Cohen's *d* test" to determine whether prices differed significantly among purchasers, regions, or time periods. In response to comments from interested parties concerning whether a simple average instead of a weighted average should be used to calculate the "pooled standard deviation" in the calculation of the Cohen's *d* coefficient, Commerce explained in the *Final Determination* that the calculation of the pooled standard deviation based on a simple average of the variances determined for the test and comparison groups was appropriate because: (a) it is consistent with our normal practice; and (b) there is no statutory directive with respect to how Commerce should determine whether a pattern of prices that differ significantly exists, and it is a reasonable approach that affords predictability.[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]

---

[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 (p. 28-29).
[12] *Id*. (stating "{Commerce} 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.").

On March 23, 2017, the CIT sustained Commerce's use of a simple average to calculate the pooled standard deviation in the *Final Determination*.[13] On October 3, 2019, the CAFC vacated and remanded the CIT's judgment sustaining Commerce's calculation of the pooled standard deviation within the Cohen's *d* test, with instructions to remand to Commerce for further explanation regarding Commerce's decision to use a simple average to calculate the pooled standard deviation in the calculation of the Cohen's *d* coefficient.[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*.[16] In particular, the CIT held that Commerce's choice to use a simple average for the pooled standard deviation is reasonable.

---

[13] *See Mid Continent Steel & Wire, Inc. v. United States*, 219 F. Supp. 3d 1326 (CIT 2017) (*Mid Continent I*)*; 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, available at https://access.trade.gov/resources/remands/17-31.pdf (*First Redetermination*). The CIT remanded the calculation of the general and administrative (G&A) expenses ratios, which Commerce recalculated in the *First Redetermination*. The CIT affirmed Commerce's recalculation of the G&A ratios in *Mid Continent Steel & Wire, Inc. v. United States*, 273 F. Supp. 3d 1161 (CIT 2017).

[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 Remand, Mid Continent Steel & Wire, Inc. et al. v. United States*, Court No. 15-00213 (CIT December 3, 2019), dated June 16, 2020, available at https://access.trade.gov/resources/remands/15-00213.pdf (*Second Redetermination*). The interested parties that submitted comments were PT Enterprise Inc.; Pro-Team Coil Nail Enterprise Inc.; Unicatch Industrial Co., Ltd.; WTO International Co., Ltd.; Zon Mon Co., Ltd., Hor Liang Industrial Corp.; President Industrial Inc. and Liang Chyuan Industrial Co., Ltd. foreign producers or exporters of the subject merchandise); and Mid-Continent Steel & Wire, Inc. (Mid-Continent is a domestic interested party and the petitioner in the LTFV investigation).

[16] *See Mid Continent IV*.

The Taiwan Respondents appealed the CIT's second judgment to the CAFC. On April 21, 2022, the CAFC vacated *Mid Continent IV* and remanded the issue back 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]

III. ANALYSIS

Section 777A(d)(1)(B) of the Act provides that Commerce may resort to a comparison method based on average-to-transaction comparisons (wherein Commerce compares weighted-average normal values with the export prices or constructed export prices of individual sales) 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.[18] In its examination of the pattern requirement, Commerce first uses the "Cohen's *d* test" and then uses 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. The ratio test, which is not at issue in this litigation, is used to assess the extent of the significant price differences for all sales as measured by the Cohen's *d* test and whether there exists a pattern of prices that differ significantly.

[17] *See Mid Continent V*.
[18] *See generally Mid Continent V*, 31 F.4th at 1370-73.

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,[19] and which is a measure of the practical significance of the difference in two means.[20] 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. It is this denominator of the Cohen's *d* coefficient, the "standard deviation," that is the subject of this litigation.[21]

In the *Final Determination* of the LTFV investigation, Commerce calculated the denominator of the effect size as the "simple average" of the standard deviations[22] of the test group[23] and the comparison group.[24] 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 rejected each of the Taiwan Respondents' arguments by finding that the nature of the pattern requirement and Commerce's use of the Cohen's *d* test supported the reasonableness of using a simple average. However, the CAFC has twice found that Commerce's explanations, in the *Final Determination* and in the

---

[19] *See generally* Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Publishers (1988) (*Cohen*), at 19-74. The first two chapters of *Cohen* are included in Appendix II to Commerce's Second Draft Redetermination. *See* Draft Results of Remand Redetermination, *Mid Continent Steel & Wire, Inc. et al. v. United States*, Court No. 15-00213, dated December 3, 2019 (Second Draft Redetermination).

[20] *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.

[21] *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.").

[22] The "simple average" is really the square root of the simple average of the variances of the test group and the comparison group, as discussed below.

[23] 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).

[24] 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.

Second Redetermination did not adequately justify reliance on a simple average. Most recently, in *Mid Continent V*, the CAFC stated:

> {w}e 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.[25]
>
> . . .
>
> 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.[26]

The CAFC found that rationality and genuineness of the seller's pricing choices have no evident connection to the undisputed purpose of the denominator, which is to provide a dispersion figure for the more general pool that serves as a yardstick for deciding on the significance of the difference in mean prices of the two groups.[27] The CAFC further found that Commerce's explanation that a simple average provides predictability was inadequate because the mathematical formulas for simple average, weighted average, or full population standard deviation have no identified elements of discretion, or other components, that distinguish them with respect to prediction.[28] The CAFC 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."[29]

---

[25] *See Mid Continent V*, 31 F. 4th at 1377.
[26] *Id.*, 31 F. 4th at 1381.
[27] *Id.*, 31 F. 4th at 1379.
[28] *Id.*, 31 F. 4th at 1380.
[29] *Id.*, 31 F. 4th at 1381.

A. The Academic Literature Has Support For A Simple Average For The Denominator Of The Effect Size

As detailed above, the CAFC's decision in *Mid Continent V* is premised on a finding that Commerce departed from academic literature in relying on a simple average in the Cohen's *d* denominator. Accordingly, we have reevaluated the academic literature on the record of this remand proceeding and the circumstances in which the Cohen's *d* test is performed in Commerce's differential pricing analysis in light of the CAFC's decision. Based on this reevaluation, and as described below, we now find that the literature does have support for Commerce's reliance on a simple average when sampling is not used, the standard deviations of the full populations are known, and the standard deviations of both populations are not equal.

Dr. Cohen presented effect size as part of his concept of power analysis,[30] where effect size is one element of Dr. Cohen's power analysis and represents "the degree to which the phenomenon is present in the population."[31] In Dr. Cohen's general formulation of "the effect size (ES) we wish to detect," he defines the "*d*" coefficient as the "standardizing of the raw effect size as expressed in the measurement unit of the dependent variable {*i.e.*, the difference in the means} by dividing it by the (common) standard deviation of the measures in their respective populations, the latter also in the original measurement unit."[32] Mathematically, Dr. Cohen expressed the effect size as,

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

for a one-tailed case, or as

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

[30] *See Cohen* at 1 ("The purpose of this book is to provide a self-contained comprehensive treatment of statistical power analysis from an 'applied' viewpoint.").
[31] *Id.* at 9.
[32] *Id.* (referencing *Cohen* at 20).

for a two-tailed case,[33] where m_A and m_B are the "population means" and σ is "the standard deviation of either population (since they are assumed equal)."[34] Dr. Cohen repeated this definition of effect size for a population in his discussion of the "power tables," where "σ is the common within-population standard deviation (*i.e.*, $\sigma_A = \sigma_B = \sigma$)."[35] Thus, the common within-population standard deviation is defined in the academic literature as equal to the standard deviation of population A or the standard deviation of population B, which are equal.

In Dr. Cohen's general formulation of effect size, the denominator of the ratio, *i.e.*, the "standard deviation," is the standard deviation of population A or the standard deviation of population B, which are assumed to be identical. Thus, when the standard deviations of population A and population B are equal, either of the standard deviations of the two populations is used as the denominator. However, when the standard deviations of population A and population B are not equal:[36]

> the definition of *d* will be slightly modified. Since there is no longer a common within-population σ, *d* is defined as above (formulas (2.2.1) and (2.2.2)), but instead of σ in the denominator, the formula requires the root mean square of σ_A and σ_B, that is, the square root of the mean of the two variances:
>
> $$\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$

In other words, when the standard deviations of the two populations are not equal, then the denominator of the effect size should be the simple average of the two, unequal standard deviations of population A and population B. In this scenario, there is no common within-population standard deviation. Moreover, unlike a common within-population standard

[33] *Id*. (referencing *Cohen* at 20 (equations 2.2.1 and 2.2.2, respectively)).
[34] *Id*. (referencing *Cohen* at 20).
[35] *See Cohen* at 27.
[36] *Id.* at 43-44 and equation 2.3.2.

deviation where one of the population standard deviations is used as the denominator, the denominator in this scenario is defined as the root mean square, *i.e.*, the simple average, of the standard deviations of population A and population B.[37] Throughout *Cohen*, when the standard deviations of the two populations are known, the denominator of the effect size is either the common population standard deviation when the standard deviations of the two populations are equal,[38] or the root square mean of the two standard deviations when the standard deviations of the two populations are unequal.[39]

Consistent with Dr. Cohen's general formulation of effect size based on means and standard deviations of two populations, Dr. Ellis recognized that:

> {t}he best way to measure an effect is to conduct a census of an entire population but this is seldom feasible in practice. Census-based research may not even be desirable if researchers can identify samples that are representative of broader populations and then use inferential statistics to determine whether sample-based observations reflect population-level parameters.[40]

However, given Dr. Cohen's general formulation of effect size and the *d* coefficient where the denominator of the ratio was defined as the "standard deviation," Dr. Ellis observed:

> {t}he only tricky part in this calculation is figuring out the population standard deviation. If this number is unknown, some approximate value must be used instead. When he originally developed this index, Cohen (1962) was not clear on how to solve this problem, but there are now at least three solutions. These solutions are referred to as Cohen's *d*, Glass's delta or Δ, and Hedges' *g*. As we can see from the following equations, the only difference between these metrics is the method used for calculating the standard deviation:

---

[37] *Id.* at 44-45 ("Note that this value is not the standard deviation of either the population of men workers or that of women workers, but the root mean square of their respective population standard deviations, σ' (formula (2.3.2)).").
[38] *Id.* at 20 and 27.
[39] *Id.* at 44, 60 ("The inequality of population σ values results only in a standardization of the difference in population means by the root mean square of the population variances (formula (2.3.2)) instead of the common population standard deviation."), 61 ("Since she is assuming that $\sigma_S^2 \neq \sigma_C^2$, the standardizing unit cannot be the common within-population standard deviation, but is instead the square root of the mean of the two variances, *i.e.*, $\sqrt{(\sigma_S^2 + \sigma_C^2)/2}$ (formula (2.3.2))."), 63 ("Note that d_4' is simply the $m_P - m_C$ difference, standardized by the common within-population standard deviation (or, if $\sigma_P^2 \neq \sigma_C^2$, their root mean square, σ', formula (2.3.2))."), and 65 ("where σ is either the common population standard deviation or σ' from formula (2.3.2)").
[40] *See Ellis* at 5.

$$Cohen's\ d = \frac{M_1 - M_2}{SD_{pooled}}$$

$$Glass's\ \Delta = \frac{M_1 - M_2}{SD_{control}}$$

$$Hedges'\ g = \frac{M_1 - M_2}{\mathrm{SD}\ *_{pooled}}$$

> Choosing among these three equations requires an examination of the standard deviations of each group.[41]

Thus, when the standard deviations of the two populations are unknown, Dr. Ellis and other academic authors provide alternatives with which to estimate the denominator of the effect size. As noted in the equations above, Dr. Ellis provides different formulations for the "pooled standard deviation" as an estimate for the denominator of the effect size:

For Cohen's *d*:[42]

$$SD_{pooled} = \sqrt{\frac{\sum(X_A - \bar{X}_A)^2 + \sum(X_B - \bar{X}_B)^2}{n_A + n_B - 2}}$$

For Hedges' *g*:[43]

$$SD\ *_{pooled} = \sqrt{\frac{(n_A - 1)SD_A^2 + (n_B - 1)SD_B^2}{n_A + n_B - 2}}$$

In each of these equations, the variable *n* represents the sample size of each group of data.

When based on sampled data, Dr. Cohen stated that "{g}enerally, we can define the effect size *in the sample* (ES_s) using sample statistics in the same way as we define it for the population, and a statistically significant ES_s is one which exceeds an appropriate criterion

[41] *Id.* at 10.
[42] *Id.* at 26.
[43] *Id.* at 27.

value."[44] Dr. Cohen also provides an estimation of effect size when the analysis is based on sampled data:[45]

> {a}ccordingly, we redefine our ES index, *d*, so that its elements are sample results, *rather than population parameters*, and call it d_s. For all tests of the difference between means of independent samples,
>
> $$d_s = \frac{\bar{X}_A - \bar{X}_B}{s}$$
>
> where $\bar{X}_A$ and $\bar{X}_B$ = the two sample means, and
> s = the usual pooled within sample estimate of the population standard deviations, that is,
>
> $$s = \sqrt{\frac{\Sigma(X_A - \bar{X}_A)^2 + \Sigma(X_B - \bar{X}_B)^2}{n_A + n_B - 2}}$$

The equation to estimate the denominator of the effect size based on sampled data, the "pooled" standard deviation, is identical to that included by Dr. Ellis for the Cohen's *d* coefficient, *i.e.*, the "pooled standard deviation." This is not the equation which Commerce uses in the Cohen's *d* test because it is based on the use of sampled data; Commerce's analysis encompasses the full population of data, *i.e.*, sale prices and, thus, it is appropriate for Commerce to use Dr. Cohen's simple average of the standard deviations of the test and comparison groups.

Commerce recognizes that in our prior remand proceedings, we used the term "pooled standard deviation" to denote the denominator of the "Cohen's *d* coefficient" used in the Cohen's *d* test. We clarify that our reference to a "pooled standard deviation" is not consistent with the use of that term in the academic literature and may have caused confusion with the courts. The "pooled standard deviation," as used by the academic authors, references some of the approaches to estimate the denominator of the effect size based on the actual standard deviations of the

[44] *See Cohen* at 17 (emphasis in the original).
[45] *Id.* at 66-67 and equations 2.5.1 and 2.5.2 (emphasis added).

populations when such actual values are not known. Commerce has not used the "pooled standard deviation" as the term is meant in the academic literature to calculate the denominator of the Cohen's *d* test. Rather, Commerce has used the simple average of the actual standard deviations of the populations of the test and comparison groups as set forth in Dr. Cohen's equation 2.3.2. Commerce notes that if the two standard deviations are equal, then *Cohen* equation 2.3.2 simplifies into the identity $\sigma' = \sigma_A = \sigma_B = \sigma$, as used in Dr. Cohen's initial formulation of effect size in *Cohen* equations 2.2.1 and 2.2.2.

Professor Coe's discussion of effect size is consistent with that of Dr. Cohen and Dr. Ellis:

> {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.[46]

In his discussion of "Which 'standard deviation'?," Professor Coe presents different arguments for and against using different approaches to provide the "best estimate of standard deviation." One option is the standard deviation of a "control group," *i.e.*, Glass Δ as presented by Dr. Ellis. A second option is a "'pooled' estimate of standard deviation," which is "essentially an average of the standard deviations of the experimental and control groups (Equation 4)."[47] Each of Professor Coe's approaches is an estimate of the actual standard deviation, σ, of Dr. Cohen's general formulation of effect size, and rely on sampled data rather than on the actual standard deviations of the populations for which the difference in the means is tested.

In sum, the academic literature allows for Commerce's use of the simple average, *i.e.*, *Cohen* equation 2.3.2, as the denominator of the effect size, *i.e.*, the Cohen's *d* coefficient, when

[46] *See Coe* at 2.
[47] *Id.* at 6-7. Equation 4 is identical to the SD^*_{pooled} for Hedges' *g* in *Ellis* at 27.

the actual standard deviation of each population is known and they are unequal. Commerce's calculation of the effect size in the Cohen's *d* test is based on the full population of sale prices of comparable merchandise to a given purchaser, region, or time period and the full population of all other sale prices of comparable merchandise (*i.e.*, the test and comparison groups, respectively).[48] Accordingly, Commerce's calculation of the Cohen's *d* coefficient is based on the actual means and standard deviations of the test and comparison groups. Commerce's calculation of the Cohen's *d* coefficient is not based on sampled data, and there is no estimation of the actual mean and standard deviation of the test group and of the comparison group. The academic literature provides for the use of a weighted average as a possible approach when estimating the denominator of the effect size when the actual standard deviations are not known, which is not the situation with Commerce's application of the Cohen's *d* test. Therefore, the academic literature allows for the use of the simple average to calculate the denominator of the effect size, and it does not necessarily support the use of a weighted average.

B. Sample Sizes Do Not Limit The Use Of A Simple Average

In Commerce's use of the Cohen's *d* test, the standard deviations of population A (test group) and population B (comparison group) are known, but the standard deviations of population A and population B are not equal. Consequently, we formulate the denominator of the effect size using equation 2.3.2, where the standard deviations of population A and population B are known and not equal.

The CAFC described Dr. Cohen's alternative formula set forth in equation 2.3.2 as "designed to be applied when the two groups, though of the same size, have different standard

[48] *See Mid Continent V*, 31 F.4th at 1378 ("Indeed, in each test-group/comparison-group pair, the test and comparison groups together make up 'the entire universe, *i.e.*, population, of the available data,' because for each test group, the comparison group is all other sales data." (internal citation omitted)).

deviations."[49] However, after reevaluation of the academic literature, we find that the sample size limitation does not prevent the use of equation 2.3.2. Dr. Cohen does not apply the limitation of equal sample sizes, *i.e.*, $n_A = n_B$, in his description of equation 2.3.2 to calculate the denominator of the effect size. Rather, the sample size, n, is an important factor in the determination of the statistical significance of an analysis result based on a sample:

> {t}he 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.[50]

Thus, sample sizes are an input for the t-test and the determination of whether the results of the analysis are statistically significant.[51] When the effect size is based on sampled data, the sample size is also an input into the calculated effect size as can be seen in the equations presented above from *Cohen*, *Ellis*, and *Coe* for calculating a pooled standard deviation. Indeed, Dr. Cohen, with reference to the pooled standard deviation used to estimate the denominator of the effect size,[52] notes that "we have defined *s* quite generally so that it will hold for all cases involving two independent samples, whether or not sample sizes are equal."[53] If the effect size were based on sampled data, then it would be determined by equations 2.5.1 and 2.5.2 where the sample sizes are an input into the estimation of the effect size of the full populations of data.[54]

---

[49] *See Mid Continent V,* 31 F.4th at 1372 (referencing *Cohen* at 44 (equation 2.3.2)).
[50] *See Cohen* at 6.
[51] *Id.* at 19 and 43.
[52] *Id.* at 67 (equation 2.5.2).
[53] *Id.* at 67. Note that "*s*" is Dr. Cohen's calculation of the denominator of the effect size based on sampled data, whereas "σ" is the denominator of the effect size based on the full populations. The formula for "*s*," *Cohen* equation 2.5.2, is included above.
[54] *Id.* at 66-67.

Further, as described by Dr. Cohen, the sample size is also an input into Dr. Cohen's power tables:

> {n}ote that if $\sigma_A \neq \sigma_B$ and it is also the case that $n_A \neq n_B$, the nominal values for **t** and power at a given significance criterion, **a**, may differ greatly from the true values. Under these conditions ($\sigma_A \neq \sigma_B$ and $n_A \neq n_B$, simultaneously), the values in Tables 2.3 {*i.e.*, "Power of t test") may be greatly in error.[55]

The condition that the sample sizes be equal for this aspect of Dr. Cohen's power analysis is also unrelated to the measure of effect size. Effect size of the population, along with sample size, *n*, significance criteria, *a*, and Dr. Cohen's power are the four parameters of statistical inference.[56]

Therefore, even though equal sample sizes are a requirement for Dr. Cohen's "Case 2" for the t-test and power analysis, the sample size is not relevant to the measure of the effect size that continues to be presented as representing the effect size of the full populations of data. If the effect size were based on the sampled data, then Dr. Cohen's equation 2.5.2 would be the basis to calculate the denominator of the effect size and not Dr. Cohen's equation 2.3.2.

C. The Use Of Taiwan Respondents' Proposed Weighted Average Is Not Supported By The Academic Literature

The Taiwan Respondents have proposed using an alternative calculation for the denominator of the effect size where the denominator is equal to:[57]

$$denominator = \sqrt{\frac{W_a}{W_a + W_b}\sigma_a^2 + \frac{W_b}{W_a + W_b}\sigma_b^2}$$

"{Taiwan Respondents} pointed to sections of *Cohen* (at 67), of *Coe* (at 6), and of *Ellis* (at 10, 26, 27), all of which set forth formulas that clearly use weighted averages when comparing groups that have both different sizes and different standard deviations (and hence variances)."[58]

---

[55] *See Cohen* at 44.
[56] *Id.* at 14.
[57] *See Mid Continent V*, 31 F.4th at 1374.
[58] *Id.*

However, none of Taiwan Respondents' citations to the academic literature, many of which have been discussed above, support its proposed alternative calculation of the denominator of the effect size. Although a weighted average is discussed in Taiwan Respondents' citations, each of their citations to *Cohen*, *Ellis*, and *Coe* relate to the estimation of the denominator based on sampled data, whereas their proposed equation purports to be based on the standard deviations of the full populations of A and B. After reevaluating the academic literature, the Taiwan Respondents' proposed equation appears nowhere in the academic literature for when the standard deviations of both populations are known, and finds no support therein. Further, if the standard deviations of both populations are unknown, *i.e.*, the situation where the academic literature does include a weighted average, then the Taiwan Respondents' proposed equation is equally unsupported because the standard deviations, σ_a and σ_b, in the Taiwan Respondents' proposed equation are the actual population parameters of the two populations. As explained above, variables which represent population values and sample values use different symbols in the academic literature, and, unlike equation 2.3.2, Taiwan Respondents cite to equations that use variables which represent sample values, not variables that represent population parameters. However, Commerce's calculation of the Cohen's *d* coefficient, and generally the distinction between results based on the full populations of data vis-à-vis sampled data, finds support in the academic literature.

D. The Use Of A Single Standard Deviation Is Not Contemplated By The Academic Literature

As discussed above, the preferred approach to calculating the denominator of the effect size is "to conduct a census of an entire population."[59] In *Mid Continent V*, the CAFC 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

[59] *See Ellis* at 5.

> of population data is available—is to use the standard deviation for the entire population.[60]

The CAFC 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.[61]

With the CAFC concluding 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.[62]

Consequently, the CAFC 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.[63]

At the outset, we clarify that Commerce's methodology does use 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 the standard deviation of either population A or the standard deviation of population B when the standard deviation of population A is equal to the standard deviation of population B.[64] As discussed above, when the standard deviations of populations A and B are not equal, *i.e.*, when "there is no

[60] *See Mid Continent V*, 31 F.4th at 1377.
[61] *Id.*, 31 F.4th at 1378.
[62] *Id.*, 31 F.4th at 1380.
[63] *Id.*, 31 F.4th at 1381.
[64] *See Cohen* at 20 and 27 ("$\sigma_A = \sigma_B = \sigma$").

longer a common within-population σ,"[65] then Dr. Cohen provides for the alternative calculation of the denominator, as used by Commerce in the Cohen's *d* test.[66]

Based on Commerce's reevaluation of the academic literature, we find that the option that the CAFC identified of using the standard deviation of all sale prices as the denominator of all sale prices of the comparable merchandise, *i.e.*, of all sale prices in the test and comparison groups, is not appropriate for purposes of Commerce's differential pricing analysis. Under this formulation, Commerce would calculate the standard deviation of populations A and B as a single population of commingled sale prices. However, Commerce's reevaluation of the academic literature demonstrates that Dr. Cohen delineates between the two distinct populations that are the source of the means whose difference is being assessed.[67] In Dr. Cohen's general formulation (*i.e.*, equations 2.2.1 and 2.2.2), the denominator of the effect size is "the (common) standard deviation of the measures in their respective populations," or in other words "the standard deviation of either population (since they are assumed equal)."[68] When "there is no longer a common with-in population σ," then Dr. Cohen provides that "*d* is defined as above (formulas (2.2.1) and (2.2.2)), but instead of σ in the denominator, the formula requires the root mean square of σ_A and σ_B."[69] In Dr. Cohen's words, there may be a ***common value of the standard deviations*** of two populations, but Dr. Cohen does not provide for ***a single standard deviation*** in which the two populations are combined into one set of data. For Dr. Cohen, there are two separate populations of data, each of whose standard deviation is part of the calculation of effect size, but there is not a single standard deviation based on commingled data.

[65] *See Cohen* at 44.
[66] *Id.* at 44 (equation 2.3.2).
[67] *Id.* at 20, 27, and 44. Even when the analysis is based on sampled data, Dr. Cohen, as well as Dr. Ellis and Professor Coe, maintain the distinction between the two groups of data in their formulas to estimate the "standard deviation," *i.e.*, the denominator, of the effect size. *Id.* at 67; *Ellis* at 26-27; and *Coe* at 6.
[68] *See Cohen* at 20.
[69] *Id.* at 44.

As with Dr. Cohen, 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).[70]

Under question 7, "Which 'standard deviation'?," Professor Coe first proposes using the standard deviation of the control group, as with Glass' Δ.[71] 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).

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 populations 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.[72]

Coe "Equation 4" is discussed above and is identical to Dr. Ellis' equation for the pooled standard deviation for Hedges *g*.[73] The cause for this overestimation is that the standard

---

[70] *See Coe* at 2.
[71] *Id.* at 6; *see also Ellis* at 10.
[72] *See Coe* at 6.
[73] *Compare* Coe at 6 *with* Ellis at 27.

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*.[74] 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 population, is centered on the mean of each group.

[74] *See Ellis* at 26-27.

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 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. Further, as the difference in the means of the two groups increases, the "single standard deviation" will also increase even though there is no change in the variances, *i.e.*, the dispersion, in the data within each of the two groups. Accordingly, the value of SD_{pooled} will remain constant because it is based on the relationship of the data within each group, however, the value of SD_{single} will increase 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 contemplated in the academic literature and we do not consider it to be a reasonable approach for Commerce's Cohen's *d* test because the results of the calculation reflect not just the dispersion of the data within each group, but also the dispersion of the data between the two groups, the precise aspect, *i.e.*, the difference in prices, that the effect size is meant to quantify. Moreover, as detailed above, the methodology of simple averaging the standard deviations for the two groups comports with the academic literature under the circumstances in which Commerce is applying the Cohen's *d* test in which both populations are not equal and samples are not used. The academic literature, whether the effect size is based on population parameters or estimates based on statistical samples, bases the calculations on the standard deviations of each group of data, test and comparison group, experimental and control group, and not on an overall, single, commingled group of data.

IV. INTERESTED PARTY COMMENTS

Mid Continent's Comments:

- Mid Continent supports Commerce's determination and the additional explanation it provided to reinforce its decision.
- Commerce's use of a simple average of the variances in calculating the denominator of the effect size, as part of the Cohen's *d* test in Commerce's differential pricing analysis, is consistent with the statute, has support in the academic literature, and is reasonable in its examination of whether a pattern of prices existed pursuant to section 777A(d)(1)(B)(i) of the Act. Moreover, it is appropriate given the circumstances and statutory objective.
- The Cohen's *d* coefficient is a difference of means divided by a measure of variance to calculate a ratio that describes the dispersion in a group of data, whether a sample or an entire population.[75]
- Mid Continent outlines Dr. Cohen's general formula and the version used by Commerce in its differential pricing analysis.[76] Specifically, the form in the first equality of the equation identified by Mid Continent is directly from Dr. Cohen's book, the second equality defines σ as Commerce calculates it, and the third equality shows the structure of the shared variability in the denominator.[77]
- Commerce's purpose in using the Cohen's *d* coefficient is to calculate a standardized expression of difference in the means to measure the effect by which the two populations' average prices differ, which is also the main purpose for the development of the *d* coefficient as described by Dr. Cohen himself: measuring the effect to which the two populations' average values differ.[78]
- For Commerce's purposes, the two groups to be compared are determined for each discrete category of products (CONNUM) of subject merchandise by selecting the sale prices to an individual purchaser that are then compared to the sale prices to all other purchasers, the sale prices in a specific region compared to the sale prices in all other regions, or the sale prices in a specific time period compared to the sale prices in all other time periods. The analysis is repeated for all purchasers, regions, and time periods to produce different versions of the *d* coefficient that are used to evaluate whether patterns of "significantly"[79] differing prices exist in a complete population of U.S. sales made by the foreign producer or exporter. All sale prices of each CONNUM are used in each of

[75] *See* Mid Continent's Comments at 2.
[76] *Id*. at 3.
[77] *Id*. at 3.
[78] *Id*. at 3.
[79] Mid Continent indicates that this is significance in the plain language sense, not in the statistical sense, as no distributional test is performed.

these analyses.[80] Therefore, this is a population-to-population comparison, and no statistical sampling, or estimation, is performed or needed.

- The form of the *d* coefficient above is exactly the correct method for the application performed by Commerce for two main reasons. First, the groups of data examined are populations, not samples.[81] Second, group sizes are already considered when calculating the means of the test and comparison groups (*i.e.*, the numerator of the calculation) in order to account for different sized sales quantities within each group. Furthermore, in terms of nomenclature, the label "Cohen's *d*" is used to refer to a number of related techniques, rather than a single, strictly defined formula.
- Inasmuch as the CAFC has taken issue with the way in which Commerce has applied the Cohen's *d* test, this may be remedied by a straightforward distinction of nomenclature by stating that the differential pricing methodology in fact uses an academically supported variation of the Cohen's *d* coefficient given the circumstances of the data and the specific purpose of the analysis.[82] Commerce reiterates this point when it cites Cohen's discussion of how to approach a situation when the standard deviations of populations A and B are not equal.[83]
- It is imperative to note that the formula offered by Cohen in equation 2.3.2 reflects the variances of the two populations without reference to the relative size (*i.e.*, "n") of either one.[84]
- Commerce has demonstrated that there is clear support in the academic literature for the approach it took in its differential pricing analysis. More generally, the idea of comparing means and somehow standardizing them to find the difference between mean price values should be the goal rather than simply following certain pre-defined instructions that are applicable to sample-based studies, but not to the situation facing Commerce, where it has complete population data at its disposal and seeks to fulfill a statutorily defined objective.[85]
- The Taiwan Respondents' suggestion to weight by group volume is not appropriate because this approach produces exactly the opposite of what Commerce is trying measure.[86]
- Mid Continent presents two hypothetical scenarios where the mean price difference (m_A and m_B) and the within-group standard deviations (σ^2_A and σ^2_B) remain the same but the volume (WA and WB) for the higher priced group sales is increased. This illustrates that a supplier can manipulate sales volume to alter the value of *d* when this approach is used (WA and WB) for the higher priced group sales is increased.[87]
- Commerce should consider adopting the 0.5 effect size threshold in its differential pricing methodology, especially if the Court mandates changes to other discretionary choices

[80] Mid Continent states that Commerce then sums the volume of the sales whose Cohen's *d* coefficient exceeds 0.8 and uses that to determine whether the volume of sales whose prices differ significantly requires changes to the methodology used to calculate the margin of dumping.
[81] *See* Mid Continent's Comments at 4.
[82] *Id*. at 6.
[83] *Id*. at 6, citing the Draft Redetermination at 8-9, *citing Cohen* at 43-44.
[84] *Id*. at 7.
[85] *Id*. at 8.
[86] *Id*. at 8.
[87] *Id*. at 9.

made as a matter of policy in Commerce's carefully balanced methodology, such as the use a simple average when calculating the denominator of the Cohen's *d* analysis.[88]

- Using the standard deviation of the entire population in the denominator is not reasonable given the circumstances and statutory objective.[89]
- Mid Continent references the CAFC's opinion that Commerce has not explained why the basic choice of weight averaging of unequal-size groups fails to apply to the present context.[90] Mid Continent argues that, in defining Cohen's *d*, the Court is assuming that sigma, σ, is "common" (equal) in both populations, which is a significant assumption to make, and whether the population standard deviation is indeed the "common" measure of variation across all populations needs to be questioned. If two populations are different in terms of their observation values, the population standard deviation will be skewed in favor of the group with a 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 simply comparing prices regardless of the number or quantity of sales.[91]
- The proposal by the CAFC to use the standard deviation of the entire population is not appropriate given the context of Commerce's analysis. 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 deviation. This would undermine the ultimate goal of Commerce's evaluation of the data, which is to determine whether the means of the two groups individually are different enough such that they cross a preselected effect size threshold (currently 0.8), and are therefore, differentially priced.[92]
- It was implied in *Mid Continent V* that the quantity/population size is ignored if the square root of the average of variances is used instead of the pooled standard deviation. That is not correct because the population size is used in the calculation of each mean and standard deviation and hence in the calculation of Cohen's *d*.[93] Mid Continent identifies the equation used by Commerce for the Cohen's *d* measure, stating that the population sizes are considered and used in the calculation of the Cohen's *d* measure used by Commerce, and by using this version, instead of the pooled version, the population size information is not lost.[94]
- Based on the conditions presented to it, Commerce applied the correct form of the Cohen's *d* measure in its differential pricing methodology. If anything, Commerce was overly conservative in using the arbitrary "large" effect size threshold of 0.8 when it would be perfectly justified to use the "medium" threshold of 0.5. Commerce identified support in the academic literature for its approach of using a simple average of variances in calculating the denominator of the effect size, and we have offered additional explanation herein as to why this is justified. Commerce also correctly rejected the respondents' suggestion to use volume weighting, as well as the Federal Circuit's

---

[88] *Id*. at 10.
[89] *Id*. at 11, citing *Mid Continent V*.
[90] *Id*. at 11.
[91] *Id*. at 11-12.
[92] *Id*. at 12.
[93] *Id*. at 12-13.
[94] *Id*. at 14.

proffered idea to use the single standard deviation of the entire population, as both of these proposals are without merit.[95]

Taiwan Respondents' Comments:

As a general note, in its comments on the Draft Redetermination, the Taiwan Respondents reference Commerce's Second Redetermination as the "first determination" in these series of judicial challenges. However, the First Redetermination was in response to the Court's first opinion and remand order in *Mid Continent I*. The First Redetermination was affirmed by the Court in *Mid Continent II*. The Second Redetermination was in response to the CAFC opinion and remand in *Mid Continent III*. The Second Redetermination was affirmed in *Mid Continent IV*. The instant third redetermination is in response to the CAFC's opinion and remand in *Mid Continent V*. In addressing the TR Second Resubmitted Comments, Commerce uses "Second Redetermination" when referencing the Taiwan Respondent's use of "first redetermination."

- Commerce's Draft Redetermination fails to "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."[96]
- Commerce simply repeats that the academic literature supports its use of the simple average which the CAFC already had rejected in *Mid Continent V*.
- Using a simple average "is mathematically, economically, and statistically unsound, and leads to unreasonable, and at times, absurd results."[97]
- Using a weighted average "is objective, consistent, effective, predictable and fair, and leads to reasonable results as required by law."[98]
- The CAFC rejected Commerce's arguments to use a simple average because (1) the pricing behaviors to the to two groups are "equally rational" and "equally genuine", (2) the simple average provides "predictability" and (3) an "abstract effect." [99] and
- Commerce "has not defended its reasons which were rejected by the Federal Circuit" and has presented no "reasons why simple averaging should be preferred over weighted averaging." [100] Commerce's analysis "is based solely on its new interpretation of the statistical literature (primarily Cohen) which, according to Commerce, supports the proposition that the 'literature does have support for Commerce's reliance on a simple average when sampling is not used, the standard deviations of the full populations are known, and the standard deviations of both populations are not equal.'" [101] Commerce failed to conclude that use of the simple average is reasonable and only found that the academic literature "*allows for the use* of a simple average … and *does not necessarily support the use* of a weighted average."[102] Based on such logic, Commerce

[95] *See* Mid Continent's Comments at 14-15.
[96] *See* TR Second Resubmitted Comments at 1 (quoting *Mid Continent V*, 31 F.4th at 1381).
[97] *Id.* at 2.
[98] *Id.* at 2.
[99] *Id.* at 6-8.
[100] *Id.* at 8.
[101] *Id.* at 8-9, citing Draft Redetermination at 7.
[102] *Id.* at 13 (quoting from Draft Redetermination at 13) (emphasis added by Taiwan Respondents).

cannot conclude, based on the facts of this investigation, that a simple average is reasonable and that a weighted average is not reasonable.[103]

- Commerce's analysis of the academic literature is wrong and Commerce cannot use a simple average unless it is "supported by substantial evidence and is reasonable as applied to the circumstances of this case."[104]
- Commerce should use a weighted average rather than a simple average for multiple reasons, which include:
  - (1) a weighted average leads to reasonable results whereas a simple average does not;
  - (2) the academic literature support using a weighted average "when comparing actual populations of unequal sizes and unequal standard deviations"[105]
  - (3) "{w}eighted averaging is supported by basic principles of economics, mathematics and statistics {whereas} simple averaging is not;"[106]
  - (4) Commerce uses weights when calculating the means and standard deviations of the test and comparison groups, but then "inconsistently relying on a {simple average} to calculate the {pooled standard deviation}";[107]
  - (5) Commerce "cannot reasonably disassociate the {weighted average} price of each group from its quantity, since quantity is a critical factor throughout Commerce's determination of whether merchandise is sold to the United States at less than fair value;"[108]
  - (6) "{a simple average} skews critical factors influencing pricing behavior, most notably quantities sold and spreads between prices;"[109]
  - (7) "{a simple average} results in a particular sale having one weight when assigned to one group (in an economically arbitrary manner) and a second, totally different weight when assigned to another group;"[110] and
  - (8) "Judicial precedent holds that {a simple average} cannot be used when data needed to calculate a {weighted average} is readily available."[111]
- The five examples previously provided by the Taiwan Respondents in their briefs to the courts "illustrate how and why weight averaging is reasonable, while simple averaging is not." [112]

---

[103] *See* TR Second Resubmitted Comments at 13, citing to *Yangzhou Bestpak Gifts & Crafts Co. v. United States*, 716 F.3d 1370, 1378 (Fed. Cir. 2013) (*Bestpak*) ("Nevertheless, '{w}hile various methodologies are permitted by the statute, it is possible for the application of a particular methodology to be unreasonable in a given case.' 'Form should be disregarded for substance and the emphasis should be on economic reality.' This court finds that this case presents that situation. Although Commerce may be permitted to use a simple average methodology to calculate the separate rate, the circumstances of this case renders a simple average of a *de minimis* and AFA China-wide rate unreasonable as applied. Similarly, a review of the administrative record reveals a lack of substantial evidence showing that such a determination reflects economic reality." (citations omitted)).
[104] *Id.* at 14.
[105] *Id.*
[106] *Id.*
[107] *Id.*
[108] *Id.*
[109] *Id.*
[110] *Id.*
[111] *Id.*
[112] *Id.* at 15.

- Taiwan Respondents reiterate a hypothetical example which was included in the CAFC's opinion in *Mid Continent III* where the CAFC stated that "we cannot conclude that Commerce's methodology was a reasonable exercise of its agency discretion in light of the statutory constraints and policies."[113] When graphically presented, the prices to the test group "are consistent" with the prices in the comparison group, yet the Cohen's *d* coefficient using a simple average results in a value of 0.88 (*i.e.*, a significant difference) whereas the Cohen's *d* coefficient using a weighted average results in a value of 0.63 (*i.e.*, a non-significant difference). This "correct value of Cohen's *d*," *i.e.*, 0.63, "reflects the visually obvious fact that the Test Group prices are wholly consistent with the Comparison Group prices."[114]
- Taiwan Respondents reiterate four examples which they had included in their comments to the CIT in response to Commerce's draft redetermination in the second remand segment. These examples are based on certain of PT's U.S. sale prices for which Taiwan Respondents calculated the Cohen's *d* coefficients using both a simple average and a weighted average.
  - Two examples involve test groups which change from pass to no-pass, which are graphically represented from Taiwan Respondents previous submissions. These graphs demonstrate that the sales in the test group "are sold at substantially similar prices" that "fall squarely within the range of the sale prices" in the comparison group, which "suggests that there is no significant difference between prices in the Test Group and Comparison Group." In both examples, the "unwarranted pass" of the results of the Cohen's *d* test based on a simple average is caused by a small number of sales and a small variance in prices in the test group.[115]
  - Two examples involve test groups which change from no-pass to pass, which is caused by the opposite situation where there is a large variance in prices within the test group with a small number of sales. This large variance "is a pattern characteristic of 'masked dumping' {where} an unusually low unit price of $5.31 is counterbalanced by another sale at an unusually high unit price of $5.81." As a result, Commerce's use of a simple average fails to "detect{} the apparent dumping pattern" as "evidence of targeted dumping" and "to correctly account for the difference in spreads of Comparison Group and Test Group {prices}."[116]
- "These examples demonstrate how the use of a {simple average} gives undue weight to the smaller group."[117] When the smaller group has a smaller variance in prices, then this results in an incorrect increase in the value of the Cohen's *d* coefficient and an "unreasonable finding of significant price differences." When the smaller group has a larger variance in prices, then this results in a failure to identify significant price differences that can mask dumping.

---

[113] *See* TR Second Resubmitted Comments at 15, quoting *Mid Continent III*, 940 F.3d at 674.
[114] *Id.* at 17.
[115] *Id.* at 19-24.
[116] *Id.* at 24-28.
[117] *Id.* at 28.

- In sum, using a weighted average "in these examples makes sense {whereas} a {simple average} does not." "The reason for these differences is found in the linchpin of {Taiwan Respondents'} argument; that is, {a simple average} methodology may lead to unreasonable results when the {standard deviation} of a small sized group is relatively large (unnaturally increasing the Cohen's *d* denominator, potentially turning a pass into a no-pass) or relatively small (unnaturally decreasing the Cohens *d* denominator, potentially turning a no-pass into a pass)."[118]
- "The Department has never challenged the veracity of the data in, or the results of, {Taiwan Respondents'} examples, and has never argued that these examples incorrectly turn passes into no-passes, and no-passes into passes. The Department also has not proffered its own examples showing that {a simple average} leads to more reasonable results than {a weighted average}."[119]
- A methodology is validated by the results of the analysis and confirms "that a methodology is reasonable in the real world."[120] Absurd results cannot be sustained. Although a methodology may appear to be reasonable, the methodology must be applied to facts to determine "whether theoretical reasonableness leads to reasonable results," as reasoned by the CAFC in *Bestpak*.[121]
- Further, Commerce must apply the Cohen's *d* test as intended by Dr. Cohen, "that the results are 'appropriate to the data, test, and statistical model employed.'"[122] Use of a simple average leads to unreasonable results, whereas use of a weighted average does not, as recognized by the CAFC in *Stupp*.[123]
- Commerce's previously articulated reason that the use of a simple average is reasonable because most of the results between a simple average and a weighted average do not change is wrong.[124] The calculated Cohen's *d* coefficient differs between the simple and weighted average unless the quantities and standard deviations between the test and comparison groups are identical. Further, the use of the weighted average "results in PT's margin being reduced from 2.16% percent {sic} to *de minimis*."[125] Thus, the use of a simple average has "a material

[118] *See* TR Second Resubmitted Comments at 29.
[119] *Id.* at 29.
[120] *Id.* at 29.
[121] *Id.* at 29, citing to *Bestpak*, 716 F.3d at 1378 ("Nevertheless, while various methodologies are permitted by the statute, it is possible for the application of a particular methodology to be unreasonable in a given case … Although Commerce may be permitted to use a simple average methodology to calculate the separate rate, the circumstances of this case renders a simple average of a *de minimis* and AFA China-wide rate unreasonable as applied." (internal citations deleted).)
[122] *Id.* at 31 (no citation provided).
[123] *Id.* at 31, citing *Stupp Corp. v. United States*, 5 F.4th 1341, 1359 (Fed. Cir. 2021) (*Stupp*) ("{T}he problem in that situation is a function of Commerce's use of the simple average pooled standard deviation").
[124] Taiwan Respondents provide an analysis of the impact on all Cohen's *d* coefficients between using a simple average and a weighted average. When using a weighted average instead of a simple average, the test groups for 27 periods, 20 regions and one purchaser change from passing to not passing the Cohen's *d* test, and four periods, four regions and two purchasers change from not passing to passing the Cohen's *d* test. Further, the vast majority of the Cohen's *d* test do not change from passing or not passing whether based on a simple average or a weighted average. *Id.* at 18-19, citing "Consolidated Plaintiffs Taiwan Respondents Comments on Final Results of Redetermination," dated July 28, 2020 (Appx 1122-1373) at 17-18.
[125] *Id.* at 32.

impact on the results of the less-than-fair-value investigation" and "a 'material difference' in PT's margin."[126]

- The CAFC, in *Mid Continent V*, correctly reasoned that the academic literature (*i.e.*, *Cohen*, *Ellis* and *Coe*) supports that a weighted average, and not a simple average, of the standard deviations should be used as the basis for the denominator of the Cohen's *d* coefficient.[127] While Commerce is correct that the academic literature generally focuses on the analysis of sampled data as distinguished from the 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."[128]
    - With both sampled data and the actual populations, the size of each group, the standard deviation of each group, and the total quantity of each group are critical factors. "Group size matters, whether the group is a sample or the entire population."[129]
    - "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 factor in sample size in calculating the result, and in rejecting the Cohen's *d* reliance on pooled standard deviations, Commerce has 'cherry picked' Cohen's *d* principle which lead to its desired result, and rejected those that do not."[130]
    - Dr. Cohen stated that "when all data (rather than merely a sample) are available, weighting is required where the size of the groups differ."[131]
        - "When … the populations {A and B} are concrete and unequal collections of cases, the inequality should figure in the assessment of the degree of relationship…"[132] Dr. Cohen's equation 2.2.7 explicitly includes the "relative weights" of the two populations.
        - "{E}qual-sampling … is not objectionable if the investigator wishes to consider membership in a given … group as an abstract effect quite apart from the relative frequency with which that effect … occurs in the population, but it clearly cannot be referred to the natural population with its varying group frequencies."[133]
    - "The size of samples of two or more groups may differ for two reasons, depending on how the samples are selected. {W}hen a population (consisting of two or more groups) is either randomly sampled or completely sampled (a "census"), the expected sample sizes of each subgroup will be proportional to its proportion of the population. Those proportions often differ."
    - Cohen also discusses a human population comprised of religious groups of which Jews are a minority. To better characterize such a small subpopulation, Cohen contemplates a strategy in which the investigator 'may advantageously

[126] *See* TR Second Resubmitted Comments at 32.
[127] *Id.* at 32.
[128] *Id.* at 33.
[129] *Id.* at 33.
[130] *Id.* at 33.
[131] *Id.* at 34, citing to various internet references and academic papers included in Taiwan Respondents' "Comments of Taiwan Plaintiffs on Draft Redetermination Pursuant to Court Remand, Court 15-00213 (CAFC No. 18-1229, October 3, 2019), dated March 19, 2020.
[132] *Id.* at 34, quoting *Cohen* at 23-24.
[133] *Id.* at 34, citing Exhibit Three.

oversample Jews by having their *n* equal (or even draw a larger sample ...) ... and thus increase his power {that is, his chance of detecting a difference between Jews and all others}." According to Cohen, the definition and interpretation of "effect size" depends on the selected sampling method.
  - The phrase "not objectionable" in Cohen's text is limited to samples. This case, however, does not involve samples. The data in this case consist of the entire (*i.e.*, natural) population. Cohen's statement does not apply to a census." [134]
  - "Cohen explicitly states that such an effect size is 'artificial.' By this he means it does not characterize a property of the population." [135]
  - "In sum, Cohen concludes that {a simple average} does not apply to a natural population, and, even if it did, {a simple average} would yield 'artificial' results that do not reflect the population."[136]
- The use of a weighted average is supported by the "basic principles of economics, mathematics and statistics."[137] "Weighted-average data generally lead to more accurate results than simple-averaged data."[138]
  - The court rejected Commerce's use of a simple average instead of more accurate weighted-average data.[139]
  - The court rejected Commerce's use of a simple average to calculate world benchmark prices.[140]
  - The court held that that relying on a simple average, rather than a weighted average, to calculate an "all-others" rate was "unreasonable in light of the statute's clear preference for the accuracy-enhancing value of weight-averaging and the particular facts of this case." [141]
  - Commerce has stated that using a weighted average is superior.[142]
  - In the Cohen's *d* test, Commerce uses a weighting to calculate the mean and standard deviation of the test and comparison groups. It is inconsistent for

---

[134] *See* TR Second Resubmitted Comments at 36.
[135] *Id.*
[136] *Id.*
[137] *Id.* at 13.
[138] *Id.* at 36.
[139] *Id.* at 36, citing *Allied Tube & Conduit Corp. v. United States*, 132 F. Supp.2d 1087, 1096 (Ct. Int'l Trade 2001) (*Allied Tube*).
[140] *Id.* at 36, citing *RZBC Group Shareholding Co. v. United States*, 100 F. Supp.3d 1288, 1309 (Ct. Int'l Trade 2015) (*RZBC Group*) ("{a} simple average, unlike a weighted average, gives equal weight to all prices regardless of the quantities sold").
[141] *Id.* at 37, citing *MacLean-Fogg Co. v. United States*, 100 F. Supp. 3d 1349, 1359–64 (CIT 2015) (*MacLean-Fogg*).
[142] *Id*. at 37, 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) (*PET Film from China*), and accompanying IDM at 8 ("using a simple average to increase the impact of lower volume exporters necessarily distorts the margin by inflating the effect of a smaller amount of data."); *Fresh Garlic from the People's Republic of China: Final Rescission of New Shipper Reviews*, 76 FR 52315 (August 22, 2011) (*Garlic from China*), and accompanying IDM at Comment 8 ("In Chenhe, the court concluded that 'it was reasonable for Commerce to conclude that the small quantity of {the respondent's} sale would not be indicative of typical future transactions' .... We likewise, here, find the quantity to be low, and not typical of other garlic transactions.").

Commerce to then use a simple average of the standard deviations to calculate the denominator of the Cohen's *d* coefficient.[143]

- Previously, Commerce wrongly argued that it uses a simple average "so that the 'pricing behavior to each group will be weighted equally, and the magnitude of the sales to one group does not skew the outcome.'"[144] Commerce calculates weighted standard deviations, yet then "re-characterize{s} 'pricing behavior' without regard to quantity." Commerce provides no explanation why the sales quantities for the test and comparison groups should be ignored.[145]
- By using a simple average in its calculation of the denominator, Commerce has impermissibly elevated the "power" of low quantity sales to have an impact incongruent with their actual size.[146]
- Commerce can rely on simple averaging only as a last resort, when data needed to rely on "accuracy-enhancing" weighted averaging are unavailable. And even then, simple averaging is not permissible if the result does not conform to economic reality and is not supported by substantial evidence.[147]
- To give each transaction its *appropriate* importance in the Cohen's *d* calculation, each transaction must be weighted by quantity, which is exactly as Commerce already does for characterizing pricing behaviors within each group, and for calculating a weighted-average dumping margin.[148]
- Commerce's calculations are internally inconsistent. "Commerce uses weights to compute average prices and {standard deviations} of all Test Groups and Comparison Groups. But then, for determining the denominator of the Cohen's *d* equation, Commerce ignores the relationship of each subgroup {standard deviation} to the whole."[149]
- "Simple averaging to pool the Test Group and Comparison Group {standard deviations} creates the very inequality (in the statistical treatment of pricing behavior) that Commerce claims it wishes to eliminate. This is the reason why simple averaging of prices is inconsistent with the accuracy enhancing value of weight-averaging."[150]
- "The {simple averaging} methodology leads to results which are directly contrary to the statutory mandate to determine whether price differences are 'significant'" because it "accords different weights to one sale, depending on whether that sale falls within the Test Group or Comparison Group. The {simple averaging} methodology ignores judicial precedent holding that weighted average analysis must be used when the necessary underlying data is available. The {simple averaging} methodology ignores Cohen's mandate that '{Effect sizes} must be indexed or measured in some defined unit

[143] *See* TR Second Resubmitted Comments at 38, citing *Mittal Canada, Inc. v. United States*, 461 F. Supp. 2d 1325, 1330 (CIT 2006) ("Where an agency's interpretation of a statute is internally inconsistent, its claim to reasonableness is obviously compromised"); *see also* TR Second Resubmitted Comments at 45.
[144] *Id.* at 38, quoting *Final Determination* IDM at 28-29.
[145] *Id.* at 38.
[146] *Id.* at 39.
[147] *Id.* at 40.
[148] *Id.* at 40.
[149] *Id.* at 42.
[150] *Id.* at 43.

appropriate to the data, test, and statistical model employed.' In contrast, the {weighted averaging} methodology does not suffer from any of these fundamental defects."[151]

- "For the {differential pricing} analysis, Commerce always weights the unit prices by quantities when computing the spread of any group of transactions." Taiwan Respondents outline the basic Analysis of Variance (ANOVA) relation in several equations and state that the "dispute in this case concerns the meaning of the '+' in {the first equation it presented, in terms of} how are the test and comparison spreads to be combined mathematically (by simple averaging or weighted averaging)? This question can be answered by using the equivalent {second equation presented}, since there is no choice as to how to remove an effect from the overall spread: the spreads must be expressed as consistently weighted sums of squares. Also, the minus sign in {the second equation} requires subtraction in the usual numerical sense."[152]
  - o Commerce's simple average methodology "creates an inconsistency in the mathematics, the statistics, and the meanings of the quantities used in the Cohen's *d* calculations. Whatever the {simple average} standard deviation 'yardstick' might be, it does *not* correspond to anything in Cohen, no matter what Commerce elects to call it. Thus, it does not produce a correct value of Cohen's *d*.[153]
  - o In the Draft Redetermination at 11, "Commerce acknowledged that its prior reference to a 'pooled standard deviation is not consistent with the use of that term in the academic literature and may have caused confusion with the courts.' Commerce then sates {sic} that it 'has not used the 'pooled standard deviation' as that term is meant in the academic literature to calculate the denominator of the Cohen's d test.' *Id*. at 12. Commerce appears to believe that by stating that it is not calculating a {pooled standard deviation}, its failure to comply with ANOVA no longer is relevant."[154]
  - o 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."[155]
  - o "Commerce has failed to explain why the absence of considering the between spread leads to a reasonable result when applying a {simple averaging} methodology and an unreasonable result when applying a {weighted averaging} methodology." "{W}hile 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."[156]
- "Rational economic behavior requires that Commerce consider quantity in determining the pooled {standard deviation}."[157]
  - o "In this case mean prices represent dollars per kilogram ($/kg), variances are weighted by kilograms and {standard deviations} are weighted by square roots of

[151] *See* TR Second Resubmitted Comments at 44.
[152] *Id*. at 45-47.
[153] *Id*. at 48 (emphasis in the original).
[154] *Id*. at 49.
[155] *Id*. at 50.
[156] *Id*. at 51.
[157] *Id*. at 52.

kilograms. Thus, the fact that the analysis involves known quantities requires that it consider quantities sold. By ignoring quantities, Commerce distorts its analysis as to whether prices in the two groups are significantly different."

  - "Prices are based on the relationship of supply and demand." "A price without a quantity is as meaningless for economic analysis as a quantity without a price." "{B}y ignoring quantities, Commerce removes the linchpin of rational economic behavior from the comparison."[158]
  - "The {weighted averaging} methodology treats each kilogram of nails equally and consistently, 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 both a relatively large weight and a relatively small weight depending on whether the transaction falls within a Test Group or (one of several possible) Comparison Groups."[159]
  - "The violation of this natural, intuitive consistency demonstrates that {a simple average} lacks objectivity; it depends on the arbitrary, varying sizes of the multiple comparison groups. {A weighted average} is the only pooling method that assures consistency in all possible cases, no matter how the transactions might be split into groups for comparison."[160]
- "Commerce's decision must be supported by substantial evidence based on the facts in this particular case."[161] "Commerce should attempt to reach the most reasonable, accurate and fair result possible, which will be affirmed by the Federal Circuit, rather than a result that can be defended solely because courts normally defer to agency decisions."[162]
- Weighting should be based on the sales quantity and not on the number of observations. "The focus of this case is not on the number of transactions (i.e., number of sales); rather, it is the quantity of kilograms sold."[163] "In *Mid Continent V*, the {CAFC} agreed with the {Taiwan Respondents'} analysis."[164]

Commerce's Position:

At the heart of the CAFC's opinion in *Mid Continent V* is its holding that "Commerce needs a reasonable justification for departing from what the acknowledged literature teaches about Cohen's *d*."[165] Contrary to Taiwan Respondents' assertion that Commerce has "presented {no} new reasons why simple averaging should be preferred over weighted-averaging,"[166] in its

[158] *See* TR Second Resubmitted Comments at 52.
[159] *Id.* at 53.
[160] *Id.* at 53.
[161] *Id.* at 54, citing *Bestpak*, 716 F.3d at 1378.
[162] *Id.* at 55.
[163] *Id.* at 55.
[164] *Id.* at 56, citing *Mid Continent V*, 31 F.4th at 1381, n.6.
[165] *See Mid Continent V*, 31 F.4th at 1381.
[166] *See* TR Second Resubmitted Comments at 8.

Draft Redetermination, Commerce discussed in detail the academic literature, and after re-evaluating the academic literature, explained how the academic literature validates Commerce's use of a "simple average" of the standard deviations[167] to calculate the denominator of the Cohen's *d* coefficient. Commerce's application of the Cohen's *d* test is based on all of the U.S. sales data, *i.e.*, the full, actual population, and is not based on sampled data.[168] Taiwan Respondents do not contest the fact that Commerce's application of the Cohen's *d* test is based on all U.S. sale prices to each test group and to each comparison group, *i.e.*, the full populations of price data, but rather the Taiwan Respondents merely complain that Commerce did not "explain why an analysis of all available data should be treated differently than an analysis of sample data."[169] Dr. Cohen explicitly presents equations to calculate the effect size based on a population.[170] In direct contrast, Dr. Cohen and other authors in the academic literature present equations to calculate the effect size based on sampled data to estimate the effect size in the actual population.[171] These approaches to estimate the actual value of the effect size in the populations may include weighting. However, because Commerce's application of the Cohen's *d* test is based on the actual full populations of U.S. sale prices in each of the test group and comparison group, such estimation of the Cohen's *d* coefficient is not necessary and reliance on weighting to calculate the denominator of the Cohen's *d* coefficient is not warranted.

The Taiwan Respondents fail to rebut the substance of Commerce's discussion of the academic literature that explicitly supports the use of a simple average. Instead, the Taiwan

---

[167] The "simple average" is "the square root of the {simple} mean of the two variances {of the two populations}" where the variance of each population is the square of the standard deviation of each population. *See Cohen* at 44.
[168] *See Mid Continent V*, 31 F.4th at 1380 ("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.")
[169] *See* TR Second Resubmitted Comments at 33.
[170] *See Cohen* at 20, equations 2.2.1 and 2.2.2, and at 43-44, equation 2.3.2.
[171] *See Cohen* at 66-67, equations 2.5.1 and 2.5.2; *Ellis* at 10, 26-27; *Coe* at 2, 6.

Respondents include aspects of *Cohen* that are not relevant to Commerce's calculation of the Cohen's *d* coefficient, and present prior arguments, which rebut Commerce's prior reasoning in the Second Redetermination. The CAFC has already opined in *Mid Continent V* that these prior arguments are unpersuasive to support that the use of a simple average is reasonable, and now Commerce has taken a new approach which focusses on the academic literature consistent with the CAFC's opinion in *Mid Continent V*.

In its current comments, the Taiwan Respondents have failed to demonstrate that the use of a simple average is "mathematically, economically, and statistically unsound, and leads to unreasonable, and at times, absurd results" or that the use of a weighted average is "statistically and mathematically sound, has express support in the literature, is objective, consistent, effective, predictable and fair, and leads to reasonable results."[172]

Use of a Simple Average Is Reasonable

In general, the Taiwan Respondents 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 Respondents use five examples to demonstrate that the results after using a weighted average are reasonable. For the hypothetical example as depicted on a "dot" chart, Taiwan Respondents state that the "correct value of the Cohen's *d* coefficient" is based on the weighted average and "reflects the visually obvious fact" that the prices in the test group are "wholly consistent" with the prices in the comparison group.[173] 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 pass to no-pass, Taiwan Respondents similarly conclude, based on the graphical presentation of the prices, that the prices in the test and comparison groups appear "substantially

[172] *See* TR Second Resubmitted Comments at 2.
[173] *Id.* at 17.

similar" which "*suggests* that there is no significant difference."[174] 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.[175] Taiwan Respondents 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, and leads to "unreasonable results." Citing to *Bestpak*, where the CAFC found the use of a simple average was unreasonable based on case-specific facts, Taiwan Respondents insist that even if the statute permits different methodologies, 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, is unreasonable.

Commerce disagrees with the Taiwan Respondents that the proffered examples demonstrate that the use of a weighted average is reasonable and that the use of a simple average is not. In fact, 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 Cohen's *d* coefficient. Contrary to the claims by the Taiwan Respondents, the graphical representations of the test and comparison groups of prices[176] do not demonstrate a "visually obvious" 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

[174] *Id.* at 22.
[175] *See* TR Second Resubmitted Comments at 25.
[176] *Id.* at 15-28, Figures 1-6.

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 Respondents' 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 Respondents describe the consequences of the larger or smaller variance in prices for the smaller group of prices relative to the larger group of prices:

> These examples show how {a simple average} gives undue weight to the smaller group (the one with lower total quantity sold). When the smaller group has a small spread, this incorrectly *decreases* the pooled standard deviation. Because that {pooled standard deviation} is the denominator of Cohen's *d*, 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." … The corresponding {simple average} Cohen's *d* leads to an unreasonable finding of significant price differences (pass).[177]

The Taiwan Respondents' "unreasonable finding" 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 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.

[177] *See* TR Second Resubmitted Comments at 28 (emphasis in original).

Further, the Taiwan Respondents' reliance on *Bestpak* is misplaced. In *Bestpak*, the CAFC found that

> {T}he circumstances of this case render a simple average of a *de minimis* and AFA China-wide rate unreasonable as applied. Similarly, a review of the administrative record reveals a lack of substantial evidence showing that such a determination reflects economic reality.[178]

As an initial point, the issue before the CAFC in *Bestpak* was not the reasonableness of a simple average as compared with a weighted average. Instead, at issue in *Bestpak* was whether the weighted-average dumping margins determined for two mandatory respondents could be used as the basis for the weighted-average dumping margin for a non-examined, separate-rate company, namely Bestpak, or whether only one of these rates should be the basis for Bestpak's rate. The CAFC in *Bestpak* was not addressing whether a simple average was appropriate but whether it was appropriate to use both rates and average them to determine the weighted-average dumping margin for Bestpak. Therefore, the way in which *Bestpak* addressed the use of a simple average is irrelevant to the issue presented in this litigation.

Moreover, citing *Bestpak*, the Taiwan Respondents argue that Commerce must ensure that its chosen methodology conforms to "economic reality." In *Bestpak*, the CAFC questioned whether the circumstances in that case regarding Bestpak's "economic reality" supported Bestpak's weighted-average dumping margin as reasonable. The CAFC found that circumstances, such as the fact that Bestpak had established its separateness from the government and the China-wide entity, and that Commerce relied on incomplete information to inadequately link Bestpak's rate to Bestpak's "economic reality," rendered the rate assigned to Bestpak unreasonable. In contrast, in this proceeding, all of PT's U.S. sale prices are included in Commerce's Cohen's *d* test, as well, all of PT's information is in Commerce's dumping analysis

[178] *See* TR Second Resubmitted Comments at 13 and 30, quoting *Bestpak*, 716 F.3d at 1378.

as a whole. Further, the information on which Commerce relies is PT's own data and not information from another source. This information thus reflects PT's economic reality and is the basis for Commerce's dumping analysis.

Although Taiwan Respondents imply that using a simple average in the Cohen's *d* denominator is "contrary to economic reality," the Taiwan Respondents 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 Respondents' 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 Respondents' presumption lacks a foundation in the record.

In sum, Taiwan Respondents' 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. For this particular argument, the Taiwan Respondents have provided no reason that the weighted average is reasonable and the simple average is not reasonable beyond the results-oriented outcome which benefits the Taiwan Respondents. As the Taiwan Respondents state "Reliance on {a weighted average}, rather than {a simple average}, results in PT's margin being reduced from 2.16% percent to *de minimis*. Thus, Commerce's choice of methodology has 'a material impact on the results of the less-than-fair-value investigation.'"[179]

[179] *See* TR Second Resubmitted Comments at 32 (internal citation omitted).

Moreover, outside of the context of the First Amendment, judicial challenges that are based on hypotheticals are disfavored.[180] Because the facts of a hypothetical example are developed for a set purpose, an alternative set of facts could be proposed which would demonstrate the opposite, contrary result.

<u>Accuracy of Results</u>

The Taiwan Respondents assert that "{w}eighted-average data generally lead to more accurate results than simple-averaged data."[181] We find that the judicial opinions and Commerce determinations that the Taiwan Respondents cite for this proposition are not relevant to 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 Respondents, Commerce also uses a weighted 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

---

[180] *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.").

[181] *See* TR Second Resubmitted Comments at 36.

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 all other prices and *not combine*, *i.e.*, commingle, these prices into a single group. 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.

Notwithstanding the CAFC finding that the qualitative factors which Commerce has previously presented to support the use of a simple average were not persuasive, the academic literature does contain support for the use of a simple average as discussed above. Dr. Cohen explicitly prescribes the simple average when "there is no longer a common within-population {standard deviation}."[182] Certainly the academic literature, as discussed above, does provide for alternative calculation formulas which may involve the use of weights, however, these calculations are limited to calculations based on sampled data when the actual population parameters are unknown.[183]

<u>Use of a Simple Average Is Supported by the Academic Literature</u>

In *Mid Continent V*, the CAFC opined

> 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.[184]

In response to the CAFC direction 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

[182] *See Cohen* at 43-44 and equation 2.3.2.
[183] *See*, *e.g.*, *Ellis* at 10.
[184] *See Mid Continent V*, 31 F.4th at 1381.

or use of the standard deviation for the entire population."[185] In the Draft Redetermination and re-examining the academic literature that is on the record of this and prior remand segments, Commerce reviewed the calculation of effect size, and specifically Dr. Cohen's *d* coefficient, *i.e.*, the effect size for the difference in the means between two populations or sampled populations. Based on that review, Commerce finds that a simple average is the approach that Dr. Cohen prescribes to calculate the *d* coefficient when (as here) the full populations of data are the basis for this calculation. Distinct from the calculation of Dr. Cohen's *d* coefficient based on the full populations of data, the academic literature also prescribes *various alternative approaches* to estimate Dr. Cohen's *d* coefficient when the calculation is based on sampled data.

In their comments on the Draft Redetermination, the Taiwan Respondents state that

> While Commerce is correct that Cohen/Coe/Ellis' analyses generally focus on sample 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. In both cases, the size of each subgroup, the {standard deviation} of the subgroups, and the total transaction quantities in the group are critical factors in characterizing the group's pricing patterns. Group size matters, whether the group is a sample or the entire population.[186]

The assertion by the Taiwan Respondents that these "critical factors" apply, whether the compared groups of data are sampled or entire populations, is not supported by the academic literature.

Further, the Taiwan Respondents allege that "Commerce has 'cherry picked' {the} Cohen's *d* principle which {led} to its desired result, and rejected those that do not."[187] As an initial 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

---

[185] *See Mid Continent V*, 31 F.4th at 1381.
[186] *See* TR Second Resubmitted Comments at 33.
[187] *Id.* at 33.

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 of 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. Contrary to the Taiwan Respondents' claim that Commerce has cherry-picked from the academic literature, Commerce has followed these principles in conceptualizing and applying the Cohen's *d* test.

The purpose of Dr. Cohen's text, *Statistical Power Analysis for the Behavioral Sciences*, is "to provide a self-contained comprehensive treatment of statistical power analysis from an 'applied' viewpoint" where the "power of a statistical test is the probability that it will yield statistically significant results."[188] In general, the result that is sought is based on a test of the null hypothesis, "*i.e.*, 'the hypothesis that the phenomenon to be demonstrated is in fact absent" but whereas a researcher "typically hopes to 'reject this hypothesis and thus 'prove' that the phenomenon in question is in fact present."[189] "In circumstances where two populations are being compared, the null hypothesis usually takes the form 'the difference in the value of the relevant parameters is zero,' a specific value."[190]

[188] *See Cohen* at 1.
[189] *Id.* at 1 (internal citation omitted).
[190] *Id.* at 10.

"*The power of a statistical test of a null hypothesis is the probability that it will lead to the rejection of the null hypothesis* … {and} depends upon three parameters: the significance criterion, the reliability of the sample results, and the 'effect size,' that is, the degree to which the phenomenon exists."[191] The first of these parameters, the significance criteria, "represents the standard of proof that the phenomenon exists, or the risk of mistakenly rejecting the null hypothesis."[192] The second parameter, "{t}he reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value … {and} 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*."[193] The third parameter, the "effect size," indicates "'the *degree* to which the phenomenon is present in the population,' or 'the degree to which the null hypothesis is false.'"[194]

"To this point, *the phenomenon in the population* under statistical test was considered as either absent (null hypothesis true) or present (null hypothesis false). The absence of the phenomenon implies some *specific value for a population parameter*."[195] "By the above route, it can now readily be made clear that when the null hypothesis is false, it is false to some specific degree, *i.e.*, the effect size (ES) is some specific nonzero *value in the population*. The larger this value, the greater the degree to which the *phenomenon under study* is manifested."[196]

When approaching a statistical power analysis and research planning (*i.e.*, the purpose of Dr. Cohen's text), where effect size is one of the parameters used in the acceptance or rejection

[191] *See Cohen* at 4 (emphasis in the original).
[192] *Id.* at 4.
[193] *Id.* at 6 (emphasis added).
[194] *Id.* at 9-10 (emphasis in the original).
[195] *Id.* at 8 (emphasis added).
[196] *Id.* at 10 (emphasis added).

of the null hypothesis, Dr. Cohen prompts the researcher to respond to the question "How large an effect do I expect *exists in the population*?"[197] The researcher may not be able to provide a specific value for the effect size index, and may need to rely on general terms, such as "small" or "large." Even though this convention is arbitrary, "the proposed conventions will be found to be reasonable by reasonable people."[198]

Dr. Cohen concludes his general presentation of power analysis with the recognition that "we can define the effect size in the sample (ES_S) using sample statistics *in the same way as we define it for the population*, and a statistically significant ES_S is one which exceeds an appropriate criterion value."[199] Only at the conclusion of Dr. Cohen's introduction of the statistical power analysis, which includes as one of the parameters the "effect size" as "the *degree* to which the phenomenon is present in the population," does Dr. Cohen note that a researcher may need to estimate the effect size as it exists in a population based on data sampled from that population.

Dr. Ellis and Professor Coe present similar descriptions of the concept of effect size. Dr. Ellis asks the question "So what?" to prompt a researcher into considering the meaning, the practical significance of the results of an analysis.[200] To answer this question, Dr. Ellis distinguishes between the traditional understanding of statistical significance, *i.e.*, whether an observed result is the "result of chance," and the understanding of effect, *i.e.*, practical significance in the real world.[201] Dr. Ellis also delineates between an effect size within a

---

[197] *See Cohen* at 12 (emphasis added).
[198] *Id.* at 13 (emphasis added).
[199] *Id.* at 17 (emphasis added).
[200] *See Ellis* at 3-6.
[201] *Id.* at 4-5 ("An effect size refers to the magnitude of the result as it occurs, or would be found, in the population. Although effects can be observed in the artificial setting of a laboratory or sample, effect sizes exist in the real world.")

population and an effect size which is estimated based on sampled (*i.e.*, incomplete) data.[202] Professor Coe also describes the concept of effect size, in that it "quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference."[203]

It is precisely this concept of "effect size" that Commerce has used to determine whether the prices differ significantly among purchasers, regions or time periods. Further, Commerce Cohen's *d* test applies the concept of effect size to all of the U.S. prices within each test and comparison group, *i.e.*, the full populations of data. Thus, Commerce's Cohen's *d* test quantifies the actual "*degree* to which the phenomenon is present in the population," *i.e.*, the significance of the difference in the prices in the test and comparison groups. Further, as discussed in the Draft Redetermination, the academic literature prescribes various approaches to estimate the effect size, the phenomenon in the population, when an analysis is based on sampled, incomplete data rather than the full, complete populations of data.

<u>Populations and Sampled Data Are Different</u>

Nonetheless, the Taiwan Respondents insist that there is no distinction between an analysis based on sampled data and an analysis based on full populations of data. The Taiwan Respondents claim that Commerce has failed "to explain why an analysis of all available data should be treated differently than an analysis of sample data."[204] The Taiwan Respondents even creatively fabricate the idea that a full population of data is just "completely sampled (a 'census')" to propagate the fiction that there is no difference between sampled data and the full

[202] *See Ellis* at 5 ("The best way to measure an effect is to conduct a census of an entire population, but this is seldom feasible in practice.")
[203] *See Coe* at 5.
[204] *See* TR Second Resubmitted Comments at 33.

population of data.[205] The Taiwan Respondents' conflation of sample data and population is not supported in the academic literature. As discussed above, both in the Draft Redetermination and in the response to comments, Dr. Cohen and Dr. Ellis clearly distinguish between an effect size as a phenomenon in a population which may be quantified based on "a census of an entire population" using Dr. Cohen's equations,[206] or an estimated effect size based on sampled data[207] using alternative equations.[208]

Group Size Does Not Matter With Populations

The academic literature does not support the Taiwan Respondents' demand that "{g}roup size matters, whether the group is a sample or the entire population."[209] As discussed above, Dr. Cohen identifies the three elements that are constituent components of a power analysis: (1) the significance criteria, (2) the reliability of the results and sample size, and (3) the effect size.[210] "Group size," which the Taiwan Respondents insist must be part of an analysis of effect size whether a full population or sampled, only impacts the reliability of the analysis results based on sampled data.[211] The effect size, "the *degree* to which the phenomenon is present in the population,"[212] is not dependent upon the statistical criteria, including the "group size," when the full population (as opposed to sample results) is analyzed.[213] Dr. Ellis and Professor Coe also

---

[205] *See* TR Second Resubmitted Comments at 35, appearing to reference *Ellis* at 5 ("The best way to measure an effect is to conduct a census of an entire population…").
[206] *See Cohen* at 20, equations 2.2.1 and 2.2.2, and at 43-44, equation 2.3.2.
[207] *Id.* at 17 ("we can define the effect size in the sample (ES_S) using sample statistics in the same way as we define it for the population, and a statistically significant ES_S is one which exceeds an appropriate criterion value."); *see also Ellis* at 10.
[208] *See Cohen* at 66-67, equations 2.5.1 and 2.5.2; *Ellis* at 10, 26-27; *Coe* at 2, 6.
[209] *See* TR Second Resubmitted Comments at 33.
[210] *See Cohen* at 4-14.
[211] *Id.* at 6-7 ("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.").
[212] *Id.* at 9 (emphasis in the original).
[213] If the measurement of effect size would be based on sampled data, then, as noted by Dr. Cohen, the estimated effect size would also be subject to an analysis of the reliability of that estimate. *Id.* at 17 ("a statistically significant ES_S is one which exceeds an appropriate criterion value.").

admonish the researcher not to confuse practical significance and statistical significance.[214] The Taiwan Respondents' assertion that the sample size must be considered when the results of an analysis are based on the full population conflates population and sampled data, as well as practical significance (*i.e.*, effect size) and statistical significance (*i.e.*, the reliability of the results), and is not supported by the academic literature.

Further, the purpose of Commerce's use of the Cohen's *d* test is to determine whether the prices to a given purchaser, region or time period differ significantly from the prices of comparable merchandise, as required in the statute. As such, Commerce is comparing the prices to each purchaser, region and time period to all other prices and not combining, *i.e.*, commingling, these prices into a single group. As discussed in the Draft Redetermination, Dr. Cohen's first equations for quantifying the *d* coefficient[215] have as the denominator "the standard deviation of either population (since they are assumed equal)."[216] Thus, Dr. Cohen defines the denominator of the *d* coefficient as the standard deviation of either population, not of both populations, assuming that the standard deviation of one population is the same as the standard deviation of the other population. The "group size" of either of the populations is not a factor when using the standard deviation of either of these populations; Dr. Cohen's only stipulation for equations 2.2.1 and 2.2.2 is that "they are assumed equal."[217]

If the standard deviations of the two populations are not equal, then Dr. Cohen prescribes the "root mean square of σ_A and σ_B,"[218] also with no reference to the "group size" of either of the populations.

---

[214] *See Coe* at 1 ("Effect size emphasizes the size of the difference rather than confounding this with sample size"); *see* also *Ellis* at 3-6.
[215] *See Cohen* at 20, equations 2.2.1 and 2.2.2.
[216] *See Cohen* at 20.
[217] *Id.* at 20.
[218] *Id.* at 44, and equation 2.3.2; *Id.* at 27.

> The unequal variability need not affect the conception of ***d*** developed in Section 2.2. Given that there is a difference between σ_A and σ_B, we merely are using a kind of average within-population standard deviation to standardize the difference between means. [219]

As part of "Case 2," Dr. Cohen includes the condition that the sample sizes are assumed to be equal, but this pertains only to the t-test and the power analysis which are both part of the discernment of whether the estimated parameters are reliable.[220] However, Dr. Cohen's discussion of effect size here, as in section 2.2, relates to the effect size present in the population and not as estimated in sampled data. "Group size" continues to not be relevant with respect to the measure of effect size. As discussed above, the results of Commerce's application of the Cohen's *d* test are based on the full populations of data, *i.e.*, of all U.S. prices in the test group and all U.S. prices in the comparison group, such that there is no estimation of the calculated parameters, and equation 2.3.2 here, as with equations 2.2.1 and 2.2.2 in section 2.2, are used to calculate the Cohen's *d* coefficient in Commerce's analysis. Further, in Commerce's analysis there is no relevance for statistical inference using a t-test and power analysis to examine whether the results reliably represent the population parameters.

Only when the "within-population standard deviations" are not known, and the actual population standard deviations must be estimated based on sampled data, does weighting potentially become a factor to calculate the denominator of the *d* coefficient.[221] As discussed above, Dr. Cohen, Dr. Ellis and Professor Coe each present alternatives for estimating the effect

---

[219] *See Cohen* at 44.
[220] *Id.* at 6 ("*The reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value.* It is necessarily an estimated value in practice, since the 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*." (emphasis added)).
[221] *See Ellis* at 10. For example, with Glass' Δ, the denominator is the standard deviation of the control (i.e., comparison) group, with the effect that the test group has a weight of zero and the comparison group has a weight of one.

size based on sampled data. These approaches may include weighting to estimate Dr. Cohen's "standard deviation" in the denominator of the *d* coefficient. Nonetheless, Commerce does not need to rely on these alternative approaches to estimating the denominator of the *d* coefficient because Commerce calculates the actual standard deviations of the test and comparison groups in its application of the Cohen's *d* test.

Therefore, based on the academic literature, the Taiwan Respondents' assertion that "group size" matters may pertain only to when the analysis is based on sampled data. However, contrary to the Taiwan Respondents' assumption, the academic literature draws a distinct line between population and sampled data. The Taiwan Respondents' arguments that "group size" matters to Commerce's application of the Cohen's *d* test, which is based on the full populations of data in the test and comparison groups, are without merit.

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

The Taiwan Respondents 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."[222] as reflected in Dr. Cohen's equation 2.2.7. 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."[223] 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 in B (the values assigned are immaterial."[224] In Commerce's application of effect

[222] *See* TR Second Resubmitted Comments at 34, quoting *Cohen* at 23-24.
[223] *See Cohen* at 23.
[224] *Id.* at 23.

size and use of Dr. Cohen's *d* statistic, the data is of a continuous variable (i.e., the price).[225] In contrast, effect size based on a dichotomous variables,[226] the effect size is based on the relationship between probability of one or the other value of the two point scale.[227] 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***)."[228] 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 the academic literature supports Commerce's use of a simple average in its application of the Cohen's *d* test.

ANOVA Is Not Relevant

The Taiwan Respondents argue that "group size" is relevant in the context of an ANOVA analysis. Certain quotations from Dr. Cohen's discussion of analysis of variance (ANOVA)[229] were initially included in an expert's statement included as an attachment to Mid Continent's comments on Second Draft Redetermination. Based on Mid Continents comments and the expert's statement, Commerce included Mid Continent's concept of "a 'natural population' {that} can be viewed as an 'abstract effect'" as part of its support for using a simple average in

---

[225] *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.").
[226] *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)).
[227] *See Ellis* at 7-9; *see* also *Ellis* at 13 ("Groups compared on dichotomous outcomes" in contrast with "Groups compared on continuous outcomes").
[228] *See Cohen* at 23-24.
[229] *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.")

the Second Redetermination.[230] However, the CAFC found that the "abstract effect" and other arguments to be unpersuasive to support Commerce's use of a simple average,[231] and in *Mid Continent V* remanded the issue back to Commerce. In the instant Draft Redetermination, Commerce did not rely on the concept of "natural population" or "abstract effect" because (1) the CAFC 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 the academic literature supports Commerce's use of a simple average in its application of the Cohen's *d* test.

Further, in the Draft Redetermination, and in these final results of redetermination, Commerce recognizes in its re-evaluation of the academic literature that its use of the term "pooled standard deviation" was inconsistent with the use of that term in the academic literature because "pooled standard deviation" is used to describe the estimation of the denominator of Dr. Cohen's *d* coefficient when the analysis is based on sampled data,[232] and it is not used to describe the calculation of the denominator of Dr. Cohen's *d* coefficient when calculated based on the actual values of the standard deviations of the two groups.[233] The Taiwan Respondents twist this recognition to mean that "by stating that {Commerce} is not calculating a {pooled standard deviation}, its failure to comply with ANOVA no longer is relevant." The Taiwan Respondents statement is nonsensical. Commerce's Cohen's *d* test is based on the concept of

[230] *See* Second Redetermination at 40-41, 45-46.
[231] *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.").
[232] *See Ellis* at 10.
[233] *See* Draft Redetermination at 11-12.

"effect size" when examining the difference in the means of two groups of prices. ANOVA, "Cohen's *f*," "quantifies the dispersion of means in three or more groups; commonly used in ANOVA."[234] 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. Their claim that Commerce's analysis is somehow dependent upon ANOVA is illogical.[235]

Difference in the Means and Variances Within the Groups

The Taiwan Respondents claim that the use of a simple average "arbitrarily and unpredictably combines the influences of the within-group price variation{s} with the difference in average prices, thereby either inflating or deflating Cohen's *d*" whereas the use of a weighted average "always achieves this independence" between the within-group price variations and the difference in the average prices.[236] "Consequently, the {simple average} Cohen's *d* fails to yield a valid effect size."[237] Commerce agrees, in part, that "the yardstick used in Cohen's *d*"[238] must only include the variations in the prices and exclude the differences between the average prices; however, Commerce disagrees with the Taiwan Respondents that the simple average includes some measure of the difference in the mean prices.

In *Mid Continent V*, the CAFC suggested one possible alternative approach that Commerce could take to calculate the denominator of the Cohen's *d* coefficient would be to "use the standard deviation for the entire population {*i.e.*, all of the sale prices within both the test and comparison groups}."[239] As discussed in the Draft Redetermination,[240] this approach would

---

[234] *See Ellis* at 14.
[235] *See* TR Second Resubmitted Comments at 50.
[236] *Id.* at 43.
[237] *Id.* at 43.
[238] *Id.* at 43.
[239] *See Mid Continent V*, 31 F.4th at 1381.
[240] *See* Draft Redetermination at 19-23.

"reflect not just the dispersion of the data within each group, but also the dispersion of the data between the two groups, the precise aspect, *i.e.*, the difference in prices, that the effect size is meant to quantify."[241]

Separate from the CAFC alternative approach, the Taiwan Respondents here argue that the use of the simple average introduces the same distortion as would the use of a single standard deviation of the commingled prices from the test and comparison groups. However, the Taiwan Respondents fail to explain, mathematically, how this works. As discussed above, the insertion by the Taiwan Respondents of ANOVA into an analysis based on the difference in the means is nonsensical and not supported by the academic literature. Further, the Taiwan Respondents fail to explain how the use of a simple average of the standard deviations of the test and comparison groups introduces this distortion and the use of a weighted average of the very same standard deviations of the test and comparison groups does not. The results of both averages are based on the standard deviations of the test and comparison groups, both of which only include the variations of prices within each population and do not include any measure of the differences in prices between the groups. Further, the Taiwan Respondents fail to explain how the use of equal weights (*i.e.*, the weights of one in a simple average) and unequal weights (*i.e.*, in a weighted average) could introduce the difference in the mean prices between the two groups. Accordingly, the claim by the Taiwan Respondents is illogical and has no merit.

<u>Economic Behavior and Weight Averaging</u>

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

[241] *See* Draft Redetermination at 22.

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."[242] However, the Taiwan Respondents 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, to say nothing that this is supported in the academic literature.

<u>The CAFC's Opinion in *Stupp*</u>

In *Mid Continent V*, the CAFC 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.[243]

When it remanded this issue to Commerce, the CAFC recognized that it has also remanded to Commerce "other aspects of Commerce's use of Cohen's *d*."[244] In their comments concerning the Draft Redetermination, the Taiwan Respondents quote from *Stupp*,[245] 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 an {weighted averaging} methodology."[246] The issue in *Stupp* involves whether the test and comparison groups must satisfy certain statistical criteria (*i.e.*, the normality of the distributions,

---

[242] *See* TR Second Resubmitted Comments at 52-53.
[243] *See Mid Continent V*, 31 F.4th at 1380.
[244] *Id.* at 1381.
[245] *See* TR Second Resubmitted Comments at 31, quoting Stupp, 31 F.4th at 1359 ("{T}he problem in that situation is a function of Commerce's use of the simple average pooled standard deviation.")
[246] *Id.* at 31, n.14.

equal variances and sufficient sample size),[247] and not whether a simple average is reasonable to use when calculating the denominator of the Cohen's *d* coefficient. In the *Stupp* litigation, Commerce has addressed the CAFC's concerns in its redetermination pursuant to remand order in that litigation and Commerce has addressed separately the CAFC's concerns in *Mid Continent V* in this redetermination.

<u>Use of a Simple Average Is Reasonable</u>

The Taiwan Respondents conclude that "Commerce's {simple averaging} methodology is not supported by any academic literature, leads to facially odd results, is contrary to economic reality and judicial precedent and is not reasonable."[248] Further "Commerce's newfangled reliance on its strained interpretation of academic literature 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."[249] The Taiwan Respondents have failed to demonstrate any of their claimed conclusions which remain mere unsupported assertions.

First, Commerce's use of a simple average is explicitly supported in the academic literature when the results of the analysis are based on the full populations of data that is being compares (i.e., all U.S. prices to the test group and all U.S. prices to the comparison group). The Taiwan Respondents have disregarded Commerce's analysis in the Draft Redetermination, and have now simply asserted that there is no difference between an analysis based on a population and one based on sampled data. Here, the academic literature also makes an explicit delineation

---

[247] *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 whether those limits need not be observed when Commerce uses the Cohen's *d* test in less-than-fair-value adjudications.")
[248] *See* TR Second Resubmitted Comments at 51.
[249] *Id.* at 52.

between these two situations, which further bolsters Commerce's use of a simple average in its application of the Cohen's *d* test. Alternatively, the Taiwan Respondents pointlessly cite to other parts of the academic literature which are not relevant to the difference-in-the-means analysis, and the use of the Cohen's *d* coefficient.

Further, the assertion by the Taiwan Respondents that the results of Commerce's Cohen's *d* test, based on a simple average, are "facially odd" and unreasonable is simply ineffective and unsupported. The only basis for the Taiwan Respondents' claim is that they do not like the results. For the two sets of contrasting examples which purportedly demonstrate the unreasonableness of the simple average vis-à-vis the weighted average, there is no "visually obvious" difference between the test and comparison groups when one set of comparisons supposedly exemplifies where the simple average finds a significant difference where there is none, and the other set of comparisons supposedly exemplifies where the simple average fails to find a significant difference. The only difference to be found is that the "smaller" group has a smaller or larger variation in prices relative to the "larger" group, which only demonstrates the arithmetic tautology that changing the weights when averaging the two standard deviations of the test and comparison groups will produce different results. For PT in this investigation, that change will result in a select few test groups not passing the Cohen's *d* test when a weighted average is used.

The Taiwan Respondents provide no "economic reality" which would support their arguments except for the ultimate, results-only outcome produced by the use of a weighted average. In *Bestpak*, the CAFC opinion relied on by the Taiwan Respondents, the CAFC found that the weighted-average dumping margin assigned to respondent Bestpak did not reflect Bestpak's "economic reality" because the information which Commerce used to connect that rate to Bestpak was incomplete and inadequate, and Bestpak had established its eligibility to be separate from the government-wide entity, such that its weighted-average dumping margin should not be based on the government-wide entity's information. However, in this investigation, Commerce used PT's own information which was complete and the basis for Commerce conclusions. The Taiwan Respondents do not explain how PT's own information does not reflect PT's economic reality or how for a handful of comparisons the Cohen's *d* test based on a simple average leads to distorted or absurd results.

Lastly, the Taiwan Respondents point to no judicial precedent which would support the use of a simple average when comparing two distinct groups of prices. The Taiwan Respondents do cite to several court opinions and administrative determinations where a weighted average is used to calculate a value within a given group of data. However, this is distinct from a comparison of two groups of data, which, consistent with the academic literature, when based on the complete population of prices in each of the test and comparison groups, prescribes the use of a simple average when combining the unequal, within-population standard deviations. Accordingly, the claim by the Taiwan Respondents that use of a weighted average is supported by judicial precedence is without merit.

Pursuant to Section 777A(d)(1)(B)(i) of the Act, Commerce identifies whether there is a pattern of export prices or constructed export prices that differ significantly. To examine whether prices differ significantly, Commerce has used the concept of "effect size" as a measure of "the degree to which the phenomenon exists," and specifically Dr. Cohen's *d* coefficient as the measure of the significance of the difference. The academic literature explicitly supports Commerce's calculation of the Cohen's *d* coefficient, including the use of a simple average to determine the denominator of the coefficient. Commerce has used the "large" threshold developed by Dr. Cohen as a conservative standard to determine that the difference in U.S. prices is significant. Nothing presented by the Taiwan Respondents changes the overall logic from the Draft Redetermination or detracts from the reasonableness of this approach or of the representativeness of the results for PT. Therefore, Commerce's use of a simple average in its application of the Cohen's *d* test is in accordance with the law, supported by evidence on the record, and is reasonable.

V. FINAL RESULTS OF REDETERMINATION

Pursuant to the *Remand Order*, we reconsidered the use of a simple average in the calculation of the denominator of the Cohen's *d* coefficient, as part of the differential pricing analysis, and reevaluated the academic literature on the record in light of the CAFC's opinion in *Mid Continent V*. We have determined that Commerce's use of a simple average when calculating the denominator of the effect size, as part of the Cohen's *d* test in Commerce's differential pricing analysis, is consistent with the statute, has support in the academic literature, and is reasonable in its examination of whether prices differ significantly pursuant to section 777A(d)(1)(B)(i) of the Act. Based on the results of our analysis, the estimated weighted-average dumping margins calculated in the *First Redetermination* remain unchanged.

11/10/2022

X 

Signed by: LISA WANG

Lisa W. Wang
Assistant Secretary
for Enforcement and Compliance