

UNITED STATES COURT OF INTERNATIONAL TRADE

THE STANLEY WORKS (LANGFANG)
FASTENING SYSTEMS CO., LTD. and
STANLEY BLACK & DECKER, INC.,

Plaintiffs,

v.

UNITED STATES,

Defendant,

and

MID CONTINENT STEEL & WIRE, INC.,

Defendant-Intervenor.

Before: Richard K. Eaton, Judge

Court No. 17-00071

OPINION

[United States Department of Commerce’s final results are sustained.]

Dated: August 13, 2018

Lawrence J. Bogard, Neville Peterson, LLP, of Washington, DC, argued for plaintiffs. With him on the brief was *Peter J. Bogard*.

Sosun Bae, Trial Attorney, Commercial Litigation Branch, Civil Division, U.S. Department of Justice, of Washington, DC, argued for defendant. With her on the brief were *Chad A. Readler*, Acting Assistant Attorney General, *Jeanne E. Davidson*, Director, and *Patricia M. McCarthy*, Assistant Director. Of counsel on the brief was *Jessica R. DiPietro*, Attorney, Office of the Chief Counsel for Trade Enforcement & Compliance, U.S. Department of Commerce, of Washington, DC.

Ping Gong, The Bristol Group PLLC, of Washington DC, argued for defendant-intervenor. With her on the brief was *Adam H. Gordon*.

Eaton, Judge: Before the court is The Stanley Works (Langfang) Fastening Systems Co., Ltd. and Stanley Black & Decker, Inc.’s (collectively, “Stanley” or “plaintiff”) motion for

judgment on the agency record challenging the final results of the United States Department of Commerce (“Commerce” or the “Department”) in *Certain Steel Nails From the People’s Republic of China*, 82 Fed. Reg. 14,344 (Dep’t Commerce Mar. 20, 2017), P.R. 290, bar code 3551507-01, ECF No. 34 (“Final Results”), *as amended by* 82 Fed. Reg. 19,217 (Dep’t Commerce Apr. 26, 2017), P.R. 307, bar code 3566359-01, ECF No. 34 (“Amended Final Results”), and accompanying Issues and Decision Memorandum, P.R. 289, bar code 3551476-01, ECF No. 34 (“Final I&D Memo”).

Stanley objects to the Final Results on three grounds, claiming that (1) Commerce contravened 19 C.F.R. § 351.414(f) (2008) by, among other things, self-initiating a targeted dumping analysis; (2) the differential pricing analysis manifests an unreasonable interpretation of 19 U.S.C. § 1677f–1(d)(1)(B) primarily because the Cohen’s *d* test is not reasonably used to evaluate targeted dumping and is incorrectly calculated; and (3) the World Trade Organization (“WTO”) Appellate Body has held that the differential pricing analysis contravenes U.S. obligations under the antidumping agreement, thereby calling into question Commerce’s arguments regarding the reasonableness of its differential pricing analysis. *See* Pls.’ Mem. Supp. Mot. J. Admin. R., ECF No. 29-1 (“Pls.’ Br.”) 2-3, 46.

Defendant, the United States (the “Government” or “defendant”), on behalf of Commerce, argues that (1) 19 C.F.R. § 351.414(f) (2008) does not apply to administrative reviews; (2) many of Stanley’s arguments have been foreclosed by the Federal Circuit; and (3) Stanley’s WTO argument notwithstanding, Commerce was reasonable in interpreting the relevant statute and regulations when conducting its differential pricing analysis to reach the conclusion that an alternative comparison method should be used to calculate Stanley’s dumping margin. *See* Def.’s Resp. Opp’n Pls.’ Mot. J. Agency R., ECF No. 31 (“Def.’s Br.”) 4-5.

For its part, Defendant-Intervenor, Mid Continent Steel & Wire, Inc., argues that Commerce's implementation of the differential pricing analysis is reasonable and adds that "[t]he WTO decision . . . is not binding on the United States unless and until Congress and the Administration implement it pursuant to the statutory scheme." Def.-Int.'s Resp. Br., ECF No. 30 ("Def.-Int.'s Br.") 2, 4.

The court has jurisdiction pursuant to 28 U.S.C. § 1581(c) (2012). For the reasons set forth below, Commerce's Final Results are sustained.

LEGAL FRAMEWORK

In an administrative review of an antidumping duty order, Commerce determines the amount of any antidumping duty by first determining "the normal value^[1] and export price^[2] (or

¹ Normal value is:

the price at which the foreign like product is first sold (or, in the absence of a sale, offered for sale) for consumption in the exporting country, in the usual commercial quantities and in the ordinary course of trade and, to the extent practicable, at the same level of trade as the export price or constructed export price

19 U.S.C. § 1677b(a)(1)(B)(i) (2012).

² Export price is:

the price at which the subject merchandise is first sold (or agreed to be sold) before the date of importation by the producer or exporter of the subject merchandise outside of the United States to an unaffiliated purchaser in the United States or to an unaffiliated purchaser for exportation to the United States, as adjusted under subsection (c) of this section.

19 U.S.C. § 1677a(a).

constructed export price³⁾ of each entry of the subject merchandise” and then calculates “the dumping margin for each such entry.” 19 U.S.C. § 1675(a)(2)(A)(i)-(ii) (2012). A “dumping margin” is “the amount by which the normal value exceeds the export price or constructed export price of the subject merchandise.” 19 U.S.C. § 1677(35)(A). In an antidumping investigation, there are three methods by which Commerce may compare normal value with export price to determine whether merchandise is being sold for less than fair value (*i.e.*, whether it is being dumped). *See* 19 U.S.C. § 1677f-1(d). Generally, Commerce uses one of two methods: (1) a comparison of the weighted-average of an exporter’s normal values to the weighted-average of its export prices for comparable merchandise (the “A-A” method), or (2) a comparison of the normal values of an exporter’s individual transactions to the export prices of an exporter’s individual transactions for comparable merchandise (the “T-T” method).⁴ *See* 19 U.S.C. § 1677f-1(d)(1)(A)(i)-(ii).

³ Constructed export price is:

the price at which the subject merchandise is first sold (or agreed to be sold) in the United States before or after the date of importation by or for the account of the producer or exporter of such merchandise or by a seller affiliated with the producer or exporter, to a purchaser not affiliated with the producer or exporter, as adjusted under subsections (c) and (d) of this section.

19 U.S.C. § 1677a(b). The export price or constructed export price is sometimes referred to as the U.S. price.

⁴ Although § 1677f-1(d)(1)(A) lists both the A-A and T-T methods as Commerce’s general methods for comparing normal value with export price to determine whether merchandise is being dumped, in actual practice, Commerce’s regulations specify that T-T will be rarely used. *See* 19 C.F.R. § 351.414(c)(1)-(2) (2015) (“In an investigation or review, [Commerce] normally will use the [A-A] method unless [Commerce] determines another method is appropriate in a particular case. . . . [Commerce] will use the [T-T] method only in unusual situations . . .”).

If Commerce finds, however, that there is evidence of targeted dumping, *i.e.*, that “there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time,” and “explains why such differences cannot be taken into account using” the A-A or T-T methods, it may use an alternative method and compare “the weighted average of the normal values to the export prices (or constructed export prices) of individual transactions” (the “A-T” method). 19 U.S.C. § 1677f-1(d)(1)(B).⁵

⁵ 19 U.S.C. § 1677f-1(d)(1)(A), provides:

In an investigation under [19 U.S.C. § 1673], [Commerce] shall determine whether the subject merchandise is being sold in the United States at less than fair value—

- (i) by comparing the weighted average of the normal values to the weighted average of the export prices (and constructed export prices) for comparable merchandise, or
- (ii) by comparing the normal values of individual transactions to the export prices (or constructed export prices) of individual transactions for comparable merchandise.

19 U.S.C. § 1677f-1(d)(1)(A). Section 1677f-1(d)(1)(B) (targeted dumping) provides:

[Commerce] may determine whether the subject merchandise is being sold in the United States at less than fair value by comparing the weighted average of the normal values to the export prices (or constructed export prices) of individual transactions for comparable merchandise [*i.e.*, by using the A-T method], if—

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

19 U.S.C. § 1677f-1(d)(1)(B).

Commerce has promulgated a targeted dumping regulation to flesh out the statute, 19 C.F.R. § 351.414(f) (2008). *See Antidumping Duties; Countervailing Duties*, 62 Fed. Reg. 27,296, 27,373-76 (Dep't Commerce May 19, 1997) (“Final Rule”). The salient elements of this regulation are:

(f)(1) [Commerce] may apply the [A-T] method . . . in an antidumping investigation if:

- (i) As determined through the use of, among other things, *standard and appropriate statistical techniques*, there is targeted dumping in the form of a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time . . . [§ 351.414(f)(1)(i)]
- (2) [Commerce] normally will limit the application of the [A-T] method to those sales that constitute targeted dumping . . . [§ 351.414(f)(2) (2008) (*i.e.*, the Limiting Rule)].
- (3) [Commerce] normally will examine only targeted dumping described in an allegation Allegations must include all support factual information, and an explanation as to why the [A-A] or [T-T] method could not take into account any alleged price differences [§ 351.414(f)(3) (2008)].

19 C.F.R. § 351.414(f)(1)-(3) (2008) (emphasis added). Notably, by their plain language, the statute and the regulation only address antidumping investigations. 19 U.S.C. § 1677f-1(d)(1)(A)-(B) (“In an investigation . . . [Commerce] may determine whether subject merchandise is being sold in the United States at less than fair value by comparing the weighted average of the normal values to the export prices (or constructed export prices) of individual transactions for comparable merchandise”); 19 C.F.R. § 351.414(f) (2008) (“[Commerce] may apply the [A-T] method . . . in an antidumping investigation”).⁶

⁶ Commerce attempted to withdraw this regulation in 2008, but the Federal Circuit later invalidated the withdrawal. *See Withdrawal of the Regulatory Provisions Governing*

As to administrative reviews, although the statute and regulations give Commerce a framework for determining whether, in antidumping investigations, merchandise is being sold at less than fair value, or whether targeted dumping may be occurring, the section of the code addressing reviews (§ 1677f-1(d)(2)) does not specify which comparison method it must use. *See* 19 U.S.C. § 1677f-1(d)(2).⁷ Commerce’s regulations, however, state that it will apply the A-A method in both investigations *and* reviews “unless [Commerce] determines another method is appropriate in a particular case.” 19 C.F.R. § 351.414(c)(1) (2015). To determine whether another method is appropriate, Commerce’s practice, where there appears to be targeted dumping, is to use the same approach in administrative reviews that it does in investigations. *See JBF RAK LLC v. United States*, 790 F.3d 1358, 1364 (Fed. Cir. 2015). Thus, in an administrative review, Commerce will apply the A-T method when it (1) finds that there is evidence of targeted dumping, *i.e.*, “a pattern of export prices (or constructed export prices) for comparable

Targeted Dumping in Antidumping Duty Investigations, 73 Fed. Reg. 74,930 (Dep’t Commerce Dec. 10, 2008); *see also Mid Continent Nail Corp. v. United States*, 846 F.3d 1364, 1368 (Fed. Cir. 2017) (“Commerce violated the requirements of the APA in withdrawing the regulation, leaving the regulation in force . . .”). Thus, the Limiting Rule (*i.e.*, the provision of the regulation directing Commerce to limit its application of the A-T method to those sales that constitute targeted dumping) remained in force for investigations following the attempted withdrawal. In *Apex Frozen Foods Private Ltd. v. United States*, however, the Federal Circuit found that this provision did not apply to administrative reviews. *See Apex Frozen Foods Private Ltd. v. United States*, 862 F.3d 1322, 1336 (Fed. Cir. 2017).

⁷ Title 19 U.S.C. § 1677f-1(d)(2) states:

In a review under section 1675 of this title [*i.e.*, in an administrative review of an antidumping duty order, countervailing duty order, or a notice of suspension of liquidation], when comparing export prices (or constructed export prices) of individual transactions to the weighted average price of sales of the foreign like product, [Commerce] shall limit its averaging of prices to a period not exceeding the calendar month that corresponds most closely to the calendar month of the individual export sale.

19 U.S.C. § 1677f-1(d)(2).

merchandise that differ significantly among purchasers, regions, or periods of time,” and (2) explains “why such differences cannot be taken into account using [the A-A or A-T methods].” 19 U.S.C. § 1677f-1(d)(1)(B)(i)-(ii).

In both investigations and reviews, when determining whether targeted dumping may be occurring in both investigations and reviews, and therefore, whether Commerce may apply the A-T method, Commerce uses the differential pricing analysis. *See Timken Co. v. United States*, 40 CIT __, __, 179 F. Supp. 3d 1168, 1173 (2016); *see also Certain Steel Nails From the People’s Republic of China*, 81 Fed. Reg. 62,710 (Dep’t Commerce Sept. 12, 2016) (“Preliminary Results”), and accompanying Preliminary Issues and Decision Memorandum, P.R. 256, bar code 3503883-01, ECF No. 34 (“Preliminary I&D Memo”) at 19. The differential pricing analysis is a two-stage process involving three separate “tests.” In the first stage, Commerce uses what it calls the “Cohen’s *d* test”⁸ together with the “ratio test” to determine whether there is “a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time.” 19 U.S.C. § 1677f-1(d)(1)(B)(i); *see* Preliminary I&D Memo at 20.

If the results of these tests do not suggest that there is a pattern of prices that differ significantly for comparable merchandise among purchasers, regions, or periods of time, then Commerce may not consider the application of the A-T method. *See* Preliminary I&D Memo at 20-21. If, however, the results of these tests reveal that such a pattern exists, that is, that targeted dumping may be occurring, Commerce will move to the second stage of the differential pricing analysis, and use the “meaningful difference test” to determine whether the price differences can

⁸ As will be seen, labeling the formula Commerce uses as a “Cohen’s *d* test” has raised questions as to its appropriateness for identifying differential pricing.

be taken into account using the A-A method. *See* Preliminary I&D Memo at 20-21; *Timken*, 179 F. Supp. 3d at 1173-74; *Apex Frozen Foods Private Ltd. v. United States*, 40 CIT ___, ___, 144 F. Supp. 3d 1308, 1331 (2016), *aff'd*, 862 F.3d 1337 (Fed. Cir. 2017) (“*Apex I*”) (“Once Commerce establishes that there is a pattern of significant price differences, Commerce’s practice in reviews requires it to explain whether A-A cannot account for such price differences before deciding to apply A-T. Commerce has chosen to answer whether A-A cannot account for such price differences by engaging in its meaningful differences analysis, which is the second stage of the differential pricing analysis.”). Thus, Commerce uses the Cohen’s *d* test to determine whether targeted dumping may be occurring, the ratio test to see if any potential targeted dumping matters, and the meaningful difference test to determine whether the A-A method can account for any pricing differences found, *i.e.*, whether the A-A method can “unmask” targeted dumping.

As currently applied, Commerce’s differential pricing analysis is product specific and is performed at the level of individual product control numbers (*i.e.*, “CONNUMs”⁹), net of adjustments to gross U.S. selling price. Before Commerce begins its differential pricing analysis, it (1) disaggregates sales data collected from respondents and then (2) sorts the sales of each CONNUM into sales made to particular purchasers, geographic regions, or time periods. A group of CONNUM sales specific to one particular purchaser, region, or time period will form a “test” group, while the CONNUM’s remaining sales (*i.e.*, sales to all other purchasers, regions, or from all other time periods) will form a “comparison” or “base” group. *See* Preliminary I&D Memo at

⁹ A CONNUM is a product control number, or “a numerical representation of a product consisting of a series of numbers reflecting characteristics of a product in the order of their importance used by Commerce to refer to particular merchandise.” *Tri Union Frozen Prods., Inc. v. United States*, 40 CIT ___, ___, 163 F. Supp. 3d 1255, 1301 n.28 (2016).

19-20. The differential pricing analysis serially analyzes prices to each purchaser, region, and time period as a test group, and then reuses those prices when forming other comparison groups for that particular CONNUM.

As to the purpose of the first test, the so called Cohen's *d* test, Commerce seeks to measure the "effect size" between two groups.¹⁰ That is, this test measures the extent to which "the net prices to a particular purchaser, region, or time period differ significantly from the net prices of all other sales of comparable merchandise" by taking the difference between the weighted-average net prices of the test and comparison groups, divided by the "pooled" standard deviation of the net prices of the two groups.¹¹ Final I&D Memo at 18. The resulting coefficient is then categorized as either falling within a "small," "medium," or "large" threshold.¹² Preliminary I&D Memo at 20. Notably, Commerce does not consider whether a test group's weighted-average price is higher or lower than the comparison group's weighted-average price in determining the effect size.

Of these thresholds, Commerce has concluded that the "large" threshold (a 0.8 standard deviation or greater) indicates a significant difference between the two groups. Thus, if the resulting coefficient meets or exceeds the "large" threshold (*i.e.*, if the weighted-averages of the

¹⁰ Commerce describes "effect size" as "quantify[ing] the size of the difference between two groups, and may therefore be said to be a *true measure of the significance of the difference.*" Final I&D Memo at 10 (quoting *Xanthan Gum From the People's Republic of China*, 78 Fed. Reg. 33,351 (Dep't Commerce June 4, 2013) and accompanying Issues and Decision Mem., Cmt. 3).

¹¹ To calculate the pooled standard deviation, Commerce takes the square root of: the sum of the square of the comparison group's standard deviation and the square of the test group's standard deviation, divided by two.

¹² These thresholds were developed, and used by, Dr. Jacob Cohen himself. See Stanley Submission of Factual Material, P.R. 230, bar code 3483603-01, Attach. A, ECF No. 34 ("Robert Coe, *It's the Effect Size, Stupid*") at 5.

comparison group and the test group differ by at least 0.8 standard deviations), the sales within that test group are considered to have “passed” the Cohen’s *d* test. Commerce has further determined that sales “passing” the test differ significantly from all other sales for that particular CONNUM. *See* Preliminary I&D Memo at 20. Commerce then performs the same analysis on a different CONNUM test group and continues until it has cycled through all of a respondent’s sales.

Following the Cohen’s *d* test, Commerce uses the “ratio test” to “assess[] the extent of significant price differences for all sales measured by the Cohen’s *d* test.” Preliminary I&D Memo at 20. Under the ratio test, if the value of sales to certain purchasers, regions, and time periods that “pass”¹³ the Cohen’s *d* test account for 66 percent or more of the value of a respondent’s total sales, then Commerce considers there to be an “identified pattern of prices that differ significantly” such that it may consider the application of the A-T method to all sales. Preliminary I&D Memo at 20. If the value of passing sales accounts for only 33 percent or less of the value of a respondent’s total sales, however, then the results do not support the consideration of the application of the A-T method to any of respondent’s sales. If the value of passing sales is more than 33 percent but less than 66 percent of the value of a respondent’s total sales, then Commerce may consider the application of the A-T method for all passing sales, but the A-A method will be used for all remaining sales. Preliminary I&D Memo at 20.

In those instances where the Cohen’s *d* test and the ratio test have found evidence that targeted dumping may be occurring, *i.e.*, where passing sales represent more than 33 percent of

¹³ As described above, a sale “passes” the Cohen’s *d* test if the Cohen’s *d* coefficient falls within the “large” classification threshold, *i.e.*, if the Cohen’s *d* test results in a 0.8 or higher standard deviation.

the value of a respondent's total sales, Commerce then moves on to the second stage of its analysis. In the second stage of Commerce's differential pricing analysis, Commerce seeks to determine "whether using only the [A-A method] can appropriately account for such differences" found in the previous stage by applying what is known as the "meaningful difference test." Preliminary I&D Memo at 20. Under this test, Commerce first calculates the dumping margin that would result by applying the A-A method to all sales and then calculates dumping margins using the A-T method based on the results of the Cohen's *d* and ratio tests described above (*i.e.*, by (1) applying the A-T method to all passing sales and the A-A method to the remaining sales, and (2) applying the A-T method to all sales). Preliminary I&D Memo at 20. Commerce then compares the A-A margin with the appropriate A-T margin to determine if there is a "meaningful difference" between the two. Commerce considers there to be a "meaningful difference" when the comparison demonstrates: (1) where both margins calculated are above the *de minimis* threshold, that there is a 25 percent relative change in the margins; or (2) where the margin calculated using the A-A method is *de minimis*, that the A-T method generates a dumping margin that crosses the *de minimis* threshold. If a meaningful difference exists, Commerce infers that the A-A method is unable to account for the price differences to particular purchasers, regions, or in particular periods of time (*i.e.*, that the A-A method would not "unmask" observed pricing differences which evidence targeted dumping). *See* Preliminary I&D Memo at 20-21.

BACKGROUND

In August 2008, Commerce published an antidumping duty order covering certain steel nails from China. *See Certain Steel Nails From the People's Republic of China*, 73 Fed. Reg. 44,961 (Dep't Commerce Aug. 1, 2008) (order). In October 2015, following a request by, among others, Stanley, Commerce initiated the seventh administrative review of the order for the period

of August 1, 2014, through July 31, 2015 (the “POR”). *Initiation of Antidumping and Countervailing Duty Admin. Review*, 80 Fed. Reg. 60,356, 60,360 (Dep’t Commerce Oct. 6, 2015). Stanley was named as a mandatory respondent in the review and submitted responses to all of Commerce’s initial and supplemental antidumping questionnaires. Selection of Respondents for Individual Review Mem. (Dec. 16, 2015), P.R. 76, bar code 3426396-01, ECF No. 34; Stanley Section A-D Questionnaire Resp., P.R. 90, bar code 3433013-01, P.R. 110, bar code 3442643-01, P.R. 117, bar code 3442681-01, ECF No. 34; Stanley Suppl. Section A, C, and D Questionnaire Resp., P.R. 198, bar code 3472991-01, ECF No. 34.

During the course of the review, the Department, on its own initiative, considered whether targeted dumping was present during the POR. Commerce published the preliminary results of its seventh administrative review in the Federal Register on September 12, 2016, employed its differential pricing analysis, and, having found evidence of targeted dumping, preliminarily calculated a weighted-average dumping margin of 5.90 percent for Stanley. Preliminary Results, 81 Fed. Reg. at 62,711; *see also* Preliminary I&D Memo 19-20. As part of its analysis, Commerce concluded that there was a pattern of export prices for comparable merchandise that differed significantly among purchasers, regions, or time periods. Preliminary I&D Memo at 21. Specifically, the Department found that 77.8 percent of the value of Stanley’s U.S. sales “passed” the Cohen’s *d* test, “confirm[ing] the existence of a pattern of prices that differ significantly among purchasers, regions, or time periods.” Preliminary I&D Memo at 21.

Commerce also preliminarily found that the A-A method could not account for such differences because the differences in the weighted-average dumping margins were meaningful, *i.e.*, Stanley’s margin crossed the *de minimis* threshold when calculated using the A-T method. Preliminary Results Analysis Memorandum for Stanley (Sept. 6, 2016), P.R. 259, bar code

3504519-01, ECF No. 34 (“Preliminary Analysis Memorandum”) at 16. In other words, Commerce determined that the A-A method could not account for the observed differences in prices among purchasers, regions, or periods of time. Thus, in accordance with the ratio test, because the value of passing sales represented 66 percent or more of Stanley’s total U.S. sales value, Commerce applied the A-T method to all of Stanley’s sales and calculated a 5.90 percent dumping margin. *See* Preliminary Analysis Memorandum at 16.

On March 20, 2017, Commerce issued its Final Results, which were amended on April 26, 2017, for a ministerial error. *See* Final Results, 82 Fed. Reg. at 14,344; Amended Final Results, 82 Fed. Reg. at 19,217. In its Final Results, Commerce again employed its differential pricing analysis and all of its elements. In so doing, Commerce quoted two academic articles in support of the use of the Cohen’s *d* test: *It’s the Effect Size, Stupid*,¹⁴ by Robert Coe, and *Difference Between Two Means*,¹⁵ by David Lane. Final I&D Memo at 10, 11 n.70. Based on the results of its differential pricing analysis, Commerce calculated a final dumping margin for Stanley of 5.78 percent. Amended Final Results Analysis Memorandum for Stanley (Apr. 19, 2017), P.R. 305, bar code 3565149-01, ECF No. 34 (“Amended Final Results Analysis Memo”) at 2. Had Commerce not applied the A-T method, Stanley’s dumping margin would have been zero. *See* Amended Final Results Analysis Memo at 2.

¹⁴ Robert Coe, *It’s the Effect Size, Stupid*.

¹⁵ Stanley Submission of Factual Material, P.R. 230, bar code 3483603-01, Attach. B, ECF No. 34 (“David Lane, *Difference Between Two Means*”).

STANDARD OF REVIEW

“The court shall hold unlawful any determination, finding, or conclusion found . . . to be unsupported by substantial evidence on the record, or otherwise not in accordance with law.” 19 U.S.C. § 1516a(b)(1)(B)(i).

DISCUSSION

I. The “Allegation” and “Appropriate Statistical Techniques” Requirements of 19 C.F.R. § 351.414(f) and Their Application to Administrative Reviews

In 1997, Commerce promulgated regulations dealing with its procedures and standards for determining whether a respondent in an investigation is engaged in targeted dumping. *See* Final Rule, 62 Fed. Reg. at 27,373-76. As a procedural matter, since the regulation dealt with investigations, Commerce was directed to “normally . . . examine only targeted dumping described in an allegation” that included “all supporting factual information, and an explanation as to why the [A-A] or [T-T] method could not take into account any alleged price differences.” 19 C.F.R. § 351.414(f)(3) (2008).

Additionally, the regulations directed Commerce to (1) use “standard and appropriate statistical techniques” when determining whether there is a pattern of prices that differ significantly, and (2) “limit the application of the [A-T] method to those sales that constitute targeted dumping” (*i.e.*, the Limiting Rule). 19 C.F.R. §§ 351.414(f)(1)(i), (f)(2) (2008). In *Apex Frozen Foods Private Ltd.*, the Federal Circuit found that the Limiting Rule only applied to antidumping investigations, not administrative reviews. *See Apex Frozen Foods Private Ltd. v. United States*, 862 F. 3d 1322, 1336 (Fed. Cir. 2017). Stanley argues, however, that the Final Results violate the remaining sections of the 1997 targeted dumping regulation—in particular,

the “allegation” requirement and the “appropriate statistical techniques” requirement—which, Stanley notes, the Federal Circuit did not specifically address in *Apex*.¹⁶ Pls.’ Br. 16-17.

A. The “Allegation” Requirement Does Not Apply to Administrative Reviews

As to the “allegation” requirement found in § 351.414(f)(3) (2008), Stanley claims that Commerce acted unlawfully by initiating a differential pricing analysis without an allegation by an interested party that Stanley was engaged in targeted dumping (*i.e.*, by self-initiating a targeted dumping analysis). Pls.’ Br. 16. According to Stanley, Commerce previously “recognized the substantive importance of requiring a petitioner to allege targeting” when Commerce promulgated its targeted dumping regulation, but failed to explain why here it “no longer needs a petitioner’s ‘intimate knowledge’ and ‘expertise’ to ‘focus appropriately any analysis of targeted dumping.’” Pls.’ Br. 16 (quoting Final Rule, 62 Fed. Reg. at 27,296). Therefore, plaintiff maintains that Commerce’s *sua sponte* initiation of its differential pricing analysis in this review was unlawful.

Stanley’s argument is unconvincing because it ignores the differences in the manner in which investigations and reviews are commenced. Investigations, in nearly every case, begin with the filing of a petition by a domestic interested party (normally a manufacturer or labor union). *See* 19 C.F.R. § 351.201. These petitions may be hundreds of pages long and must contain reasonably available data supporting the allegations of dumping. *See* 19 C.F.R. § 351.202.

¹⁶ Plaintiff additionally claims that the Final Results contravene the Limiting Rule of § 351.414(f)(2), but concedes that the Federal Circuit has found that the Limiting Rule applies only to antidumping investigations. Pls.’ Br. 16 (“The Final Results also contravene the ‘limiting rule’ in § [351.414(f)(2)]. However . . . the [Federal Circuit] recently concluded that the limiting rule only applies to antidumping investigations.”).

A request for a review, on the other hand, is a far less detailed affair. Indeed, a request need not contain any allegations or data at all. All that is required is that the interested party requesting a review provide a reason why a review should be commenced. *See* 19 C.F.R. § 351.213(b)(1). Moreover, any interested party, including a foreign manufacturer or exporter, may request a review. *See* 19 C.F.R. § 351.213(b)(1) (“Each year during the anniversary month of the publication of an antidumping or countervailing duty order, a domestic interested party or an interested party . . . may request in writing that [Commerce] conduct an administrative review . . . of specified individual exporters or producers covered by an [antidumping] order . . .”). Indeed, these requests are typically a letter of one or two pages that contain no more specific claim than that dumping may have been occurring or that a company wishes to have an accurate dumping margin for the period of review. Given the differences in commencing these two proceedings, it is not reasonable that the “allegation” requirement be retained in administrative reviews.

In addition, the court notes that the “allegation” requirement specifically states that a targeted dumping allegation must be “filed within the time indicated in § 351.301(d)(5),” a subsection that, by its own terms, applies only to investigations. 19 C.F.R. § 351.414(f)(3) (2008); *see* 19 C.F.R. § 351.301(d)(5) (2008); *see also* Final Rule, 62 Fed. Reg. at 27,336 (“[Section] 351.301(d)(5) sets forth the time limit for a targeted dumping allegation in an [antidumping] investigation.”). Therefore, the court finds that the “allegation” requirement of § 351.414(f)(3) (2008) does not apply to administrative reviews, and therefore, Commerce did not act unlawfully by self-initiating its targeted dumping analysis.

B. The “Appropriate Statistical Techniques” Requirement Applies to Administrative Reviews

Next, Stanley claims that the Final Results violate the “appropriate statistical techniques” requirement of 19 C.F.R. § 351.414(f)(1)(i) (2008) because “the Cohen’s *d* [test] is not appropriately used in a targeted dumping context.” Pls.’ Br. 16-17.

In response, the Government argues that the “appropriate statistical techniques” requirement does not apply to administrative reviews. Def.’s Br. 11 (“Stanley fails its heavy burden of showing that Commerce’s interpretation of its own regulation, 19 C.F.R. § 351.414(f), as not applying to administrative reviews, such as the one presently at issue, is not entitled to deference. As such, the Court should sustain Commerce’s final results.”).

Even considering Commerce’s sometimes extravagant claims for deference, stating that it need not comply with the requirement that it use an appropriate statistical technique to determine if targeted dumping may be present in a review, is surprising. Having chosen to employ the same method to ferret out targeted dumping in reviews as in investigations,¹⁷ the Department cannot willy-nilly decide to use portions of the regulations that lay out the method and discard others. Using a statistical technique that is not appropriate would simply not be reasonable. In fact, it would be an abuse of discretion to use an inappropriate statistical technique. *See Impact Steel Can. Corp. v. United States*, 31 CIT 2065, 2074, 533 F. Supp. 2d 1298, 1305 (2007). Therefore, Commerce must comply with the “appropriate statistical techniques” part of its regulation. As shall be seen, however, the court further finds that an appropriate statistical technique was used here.

¹⁷ Commerce first used the Cohen’s *d* test in the antidumping investigation *Xanthan Gum From the People’s Republic of China*, 78 Fed. Reg. 33,351 (Dep’t Commerce June 4, 2013) (final determination).

II. Differential Pricing is a Reasonable Interpretation of the Statute

Stanley argues that “[a]ll three elements [of differential pricing] manifest an unreasonable interpretation of the statute and do not effectuate the statute’s purpose.” Pls.’ Br. 18.

A. The Cohen’s *d* Test

Stanley’s first argument against the use of Commerce’s differential pricing analysis is that the Cohen’s *d* test “contravenes both congressional guidance and Commerce’s obligation to calculate dumping margins as accurately as possible.” Pls.’ Br. 18-19 (citation omitted). According to Stanley, this is primarily because the Cohen’s *d* measures the effect of an intervention, and not just the difference between two groups or sets of data, and therefore its use is inappropriate in the targeted dumping context. Pls.’ Br. 19.

As an initial matter, Stanley’s claims, taken as a whole, invite the court to answer the question as to whether the Cohen’s *d* test, as used by Commerce, together with the ratio test constitute a reasonable way of determining if differential pricing is present. In other words, the question is whether Commerce’s method is fit for the purpose to which it is put. While it may be that, were the question whether the Cohen’s *d* statistic, as originally envisioned by Dr. Cohen, is a reasonable way of identifying a pattern of prices that differ significantly among purchasers, regions, or periods of time, then Stanley’s arguments would have some purchase.¹⁸ Because,

¹⁸ Commerce stated in its Final Results that it “has relied upon . . . a specific approach developed by Jacob Cohen called the ‘*d*’ statistic or, as the Department has labeled it, the ‘Cohen’s *d* coefficient.’” Final I&D Memo at 9. As shall be seen, while there are some differences in how Commerce calculates the Cohen’s *d* and the method generally used in the social sciences to determine the effect size of a particular intervention, Commerce’s calculation is nevertheless based on the method developed by Dr. Cohen himself, and any differences do not make the test unrecognizable, but instead, appear to be the result of Commerce’s ultimate purpose for conducting the test, *i.e.*, determining whether prices for comparable merchandise differ significantly by purchaser, region, or period of time.

however, the court is tasked with determining whether Commerce's method, as actually applied, is a reasonable interpretation of the statute (as distinct from, for instance, a reasonable interpretation of Dr. Cohen's work) it must look at what Commerce has actually done, not what the Cohen's *d* has been used for in other contexts.

Notwithstanding the origin of the Cohen's *d* as generally for use in the social sciences, Commerce states that the test "may be instructive for purposes of examining whether to apply an alternative comparison method in this administrative review" because it "is a generally recognized statistical measure of the extent of the difference between the mean . . . of a test group and the mean of . . . a comparison group." Preliminary I&D Memo at 19-20. Although Stanley argues that using the Cohen's *d* test is inappropriate in the targeted dumping context, plaintiff points to no evidence demonstrating why the test cannot be used in a "business" or "finance" context or should be restricted to the social sciences. Moreover, it is not the case, as Stanley argues, that effect size may only be used to quantify the effectiveness of a particular intervention. *See, e.g.,* Robert Coe, *It's the Effect Size, Stupid* at 1. As Commerce notes:

The difference in two prices, such as the difference in the mean prices for two groups (*e.g.*, ten dollars), has no inherent meaning unless it is relevant to a given benchmark. For example, a ten dollar difference in the price of two cars is substantially different than a ten dollar difference in the price of a hamburger. . . . For the Cohen's *d* coefficient, this examination of the price differences between test and comparison groups is relative to the "pooled standard deviation." The use of a simple average in determining the pooled standard deviation equally weighs a respondent's pricing practices to each group and the magnitude of the sales to one group does not skew the outcome. . . . The pooled standard deviation reflects the dispersion, or variance, of prices within each of the two groups. . . . When the difference in the weighted-average sale prices between the two groups is measured relative to the pooled standard deviation, then this value is expressed in standardized units based on the dispersion of the prices within each group. This is the concept of an effect size, as represented in the Cohen's *d* coefficient.

Final I&D Memo at 11-12. Thus, as used by Commerce, the Cohen's d test performs a task frequently performed by statistical analysis by converting absolute differences to standardized variations from a mean. Here, Commerce hopes to find whether there is a "pattern of export prices" for comparable merchandise that "differ significantly" among purchasers, regions, or periods of time, as required by the statute. *See* 19 U.S.C. § 1677f-1(d)(1)(B)(i). The purpose of the Cohen's d test is to help determine whether the difference between two groups is significant enough to be of practical importance. *See, e.g.,* Robert Coe, *It's the Effect Size, Stupid*, at 2. In other words, Cohen's d can contextualize the difference between two means by using the variation found within each group of sales as a yardstick to compare the differences in prices to certain purchasers, regions, or periods of time. By looking at the results of the test, Commerce can determine how far apart the means of the two sales groups are in standardized units, which, when combined with Cohen's general interpretation conventions, allows Commerce to contextualize the magnitude of that difference, and whether that difference is large enough to matter (*i.e.*, whether Commerce should consider the application of the A-T method).

This, to the court, is a reasonable way to determine whether prices "differ significantly" as required by the statute, particularly because, as Commerce emphasizes, simply finding a difference between the groups in terms of a dollar amount does not necessarily inform Commerce about the magnitude of that difference (*i.e.*, whether it is "significant"). Commerce has supplied an adequate explanation as to why it is useful to use a statistical analysis, such as the Cohen's d test (as applied by Commerce), as distinct from an arithmetical comparison. Stanley has supplied no reason why Commerce's use of the Cohen's d is not an appropriate statistical technique and the court cannot find one. Therefore, the court finds that Commerce's

use of the Cohen's d test as used in Commerce's targeted dumping analysis is reasonable, adequately explained, and therefore, lawful and supported by substantial evidence.

Next, Stanley argues that Dr. Cohen's classification of effect sizes as small, medium, and large is "arbitrary" and the classifications are "neither fixed nor defined by Cohen's d ," but are "merely conventions . . . that Jacob Cohen himself acknowledge[d] the danger of using . . . out of context." Pls.' Br. 22 (internal quotations marks omitted) ("Commerce defended [Dr. Cohen's classifications] by asserting that 'the large threshold provides the strongest indication that there is a significant difference between the means of the test and comparison groups.' This rationale merely relies on the obvious: something 'large' is bigger than something 'small.' It fails to explain *why* any of Cohen's classifications are appropriately used to analyze nail prices or why price differences that are a fraction (0.8) of a standard deviation mean anything at all in selling nails.").

The court is unconvinced, however, that Commerce's use of the "small," "medium," and "large" thresholds is not reasonable. First, as Commerce stated, its classifications are "generally accepted thresholds for the Cohen's d test" which "have been widely adopted" by practitioners using the Cohen's d coefficient. Final I&D Memo at 11 (internal quotation marks omitted); *see also* David M. Lane, *Difference Between Two Means* at 2. The articles referenced by Stanley¹⁹ demonstrate as much. *See, e.g.,* Robert Coe, *It's the Effect Size, Stupid*, at 5 ("Another way to interpret effect sizes is to compare them to the effect sizes of differences that are familiar. For

¹⁹ Stanley submitted several academic articles for the record of this review, including: *It's the Effect Size, Stupid: What Effect Size Is and Why It Is Important* by Robert Coe, and *Difference Between Two Means* by David M. Lane. *See* Stanley Submission of Factual Material (July 1, 2016), P.R. 230, bar code 3483603-01, Attachs. A, B, ECF No. 34.

example, Cohen . . . equates [an effect size of 0.8] to the difference between the heights of 13 year old and 18 year old girls.”).

Moreover, Commerce does not apply the chosen thresholds in an arbitrary manner: only the “large” threshold (which Cohen generally described as a “grossly perceptible [effect size] and therefore large” and has also equated it to the difference in IQ between a Ph.D.²⁰ degree holder and a typical college freshman) becomes the touchstone measure of a “significant” difference in prices. Robert Coe, *It’s the Effect Size, Stupid*, at 5; see Final I&D Memo at 11-12. Keeping in mind that the Cohen’s *d* does not identify dumping, but rather a pattern of export prices for comparable merchandise that differ significantly among purchasers, regions, or periods of time, the use of a grossly perceptible standard is reasonable. Accordingly, the court finds that Commerce lawfully used these thresholds to help it determine which sales “pass” its Cohen’s *d* test.

Stanley then argues that the Cohen’s *d* is “a form of statistical inference” which should not be “used when the entire data population is known” and must generally be accompanied by a “confidence interval,”²¹ which Commerce failed to provide. Pls.’ Br. 23-24. In addition, Stanley

²⁰ While it may be that only the holder of a Ph.D. such as Dr. Cohen would have used this example, the point is well taken.

²¹ In statistics, determining how well a sample statistic (*i.e.*, when the entire population is not known) estimates the underlying population value can be addressed by using a confidence interval which provides a range of values likely to contain the population parameter of interest. In *It’s the Effect Size Stupid*, Coe explains how a confidence interval may be used in the context of determining effect size:

Clearly, if an effect size is calculated from a very large sample it is likely to be more accurate than one calculated from a small sample. This ‘margin for error’ can be quantified using the idea of a ‘confidence interval’, which provides the same information as is usually contained in a significance test: using a ‘95%

(footnote continued . . .)

claims that Commerce must account for “statistical significance” in conducting its differential pricing analysis. Pls.’ Br. 25.

Stanley’s complaints about the use of a form of the Cohen’s d test when the entire population is known are a bit puzzling. As Commerce notes

the data upon which the statistical measure of effect size is based are not random samples, but rather the entire population of data (*i.e.*, the U.S. sales to each purchaser, region, and time period). Stanley has reported all of its sales of subject merchandise in the U.S. market during the [POR], and it is this data upon which the Department is basing its analysis consistent with the requirements of [19 U.S.C. § 1677f–1(d)(1)(B)], just as it has when calculating Stanley’s weighted-average dumping margin. Accordingly, the Department’s calculation of the Cohen’s d coefficient includes no noise or sampling error as the underlying means and variances used to calculate the Cohen’s d coefficient are not estimates, but the actual values based on the complete U.S. sales data as reported by Stanley in this review.

Final I&D Memo at 10-11.

This is an important observation, as normally the Cohen’s d is used to make inferences from samples. Then, another test, a statistical significance test, is used to determine whether the findings were likely due to chance. Statistical significance and effect size are difference concepts: the former demonstrates that there is a difference between groups that is probably not

confidence interval’ is equivalent to taking a ‘5% significance level’. To calculate a 95% confidence interval, you assume that the value you got (e.g. the effect size estimate of 0.8) is the ‘true’ value, but calculate the amount of variation in this estimate you would get if you repeatedly took new samples of the same size (i.e. different samples of 38 children). For every 100 of these hypothetical new samples, by definition, 95 would give estimates of the effect size within the ‘95% confidence interval’. If this confidence interval includes zero, then that is the same as saying that the result is not statistically significant. If, on the other hand, zero is outside the range, then it is ‘statistically significant at the 5% level’. Using a confidence interval is a better way of conveying this information since it keeps the emphasis on the effect size – which is the important information – rather than the p-value.

Robert Coe, *It’s the Effect Size, Stupid*, at 8.

the result of chance, while the latter says something about the size of the difference. *See, e.g.*, Robert Coe, *It's the Effect Size, Stupid*, at 8 (“It is important to know the statistical significance of a result, since without it there is a danger of drawing firm conclusions from studies where the sample is too small to justify such confidence. However, statistical significance does *not* tell you the most important thing: *the size of the effect.*”). Because the Cohen’s *d* test, as used by Commerce, employs the entire universe of data, there is no need to test for statistical significance. That is, no inference is being made from a sample. *See* Final I&D Memo at 10-11. Thus, since the entire data population is available, the concerns that normally require a finding of statistical significance using a second test and an accompanying confidence interval are not present in Commerce’s differential pricing analysis.

Moreover, simply because the Cohen’s *d* has traditionally been applied as a form of statistical inference (*i.e.*, a test used when only samples of a population are available), plaintiff points to no evidence tending to suggest that it cannot be used when the entire population is known. As with many statistical tests, the appropriateness of a particular formula depends on how the problem is defined. Where, as here, Commerce has defined the problem as determining whether the magnitude of the difference among sales is worth paying attention to (and knowing that the pricing data is not merely a sample, but represents the entire population), using the Cohen’s *d* test is not unreasonable. *See* Final I&D Memo at 10-11. The Cohen’s *d* has been described as the “standardised mean difference between two groups,” and as such, can be useful to Commerce in finding whether there is a pattern of prices that differ significantly, as required by the statute. *See* Robert Coe, *It's the Effect Size, Stupid*, at 3. Put simply, the results of the Cohen’s *d* test, where 100 percent of the sales are known, are likely to be more reliable because they do not rely on inference.

For these reasons, the court finds that Commerce's use of the Cohen's d test in the context of a targeted dumping evaluation is not unreasonable and that it aids in Commerce fulfilling its obligation to calculate dumping margins as accurately as possible.

B. Commerce's Calculation of the Cohen's d

Next, Stanley argues that "[e]ven if it were reasonable to use the Cohen's d statistic in a targeted dumping context, the Final Results would nevertheless be unlawful because Commerce incorrectly calculates the Cohen's d statistic, which inflates the Cohen's d coefficients and the resulting [Cohen's d test] 'pass' rates." Pls.' Br. 26. Stanley makes three arguments to support its position.

First, Stanley claims that the Cohen's d test is incorrectly calculated because Commerce "calculated the pooled standard deviation^[22] in the Cohen's d statistic," which gives equal weight to the squared standard deviations of the test and comparison price groups, "despite irrefutable evidence that the test groups for Stanley were much smaller in volume and had smaller standard deviations than the comparison groups." Pls.' Br. 26-27. To bolster its argument, Stanley looks to the Robert Coe article it submitted, *It's the Effect Size, Stupid* (often cited by Commerce), which the company claims "is clear that where either the size or the variability of the test and comparison groups is different, the correct calculation of the pooled standard deviation in the Cohen's d statistic requires that the standard deviations must be weighted by size." Pls.' Br. 27 ("The use of a pooled estimate of standard deviation depends on the assumption that the two calculated standard deviations are estimates of *the same* population value,' and '[i]nterpretation of effect-size generally depends on the assumptions that 'control' and 'experimental' group

²² The pooled standard deviation is an aggregate measure of the distribution of prices (that is, the variances) within the test and comparison groups.

values are normally distributed and *have the same standard deviations.*” (quoting Robert Coe, *It's the Effect Size, Stupid*, at 6, 9)). Thus, Stanley claims that, by not weighting the standard deviations of the groups, Commerce's approach effectively assumed the test and comparison groups for Stanley's CONNUMs were of equal population values with equal standard deviations from the mean. For Stanley, because the test and comparison groups are not of equal population value and do not have the same variances, Commerce's method is unreasonable.

Commerce's calculation of its Cohen's *d* test is reasonable. Stanley's argument is essentially that what Commerce calls the “Cohen's *d* test” is not actually the Cohen's *d* test, and that Commerce's tinkering with the test has resulted in an unreasonably high number of “passing” sales. It is possible that Commerce's insistence that it is applying the Cohen's *d*, rather than a variation of it, has caused some mischief. While it may be that the Department concluded that affixing a famous name to its calculations would enhance its claim that it was satisfying the injunction found in the regulation that it use “standard and appropriate statistical techniques,” attaching the Cohen's *d* name has opened a world of possibilities to talented lawyers. The court reiterates, however, that the appropriateness of any statistical formula depends on how the problem is defined. Indeed, even the Coe paper, relied on by Stanley, demonstrates that there are different ways to calculate a Cohen's *d* statistic depending on population sizes and type of intervention.²³ See, e.g., Robert Coe, *It's the Effect Size, Stupid*, at 10-11.

²³ It bears repeating that here the entire universe of sales is known, and there is no intervention.

Here, the calculation of the pooled standard deviation is important because a smaller standard deviation can result in small price differences²⁴ having a “large” effect size (and therefore, “passing” the Cohen’s *d* test). Stanley is correct in noting that the test group will likely have a smaller number of observations (and variance) than the comparison group,²⁵ and that in these circumstances, using a simple average of the groups’ standard deviations would result in a lower pooled standard deviation than would a pooled standard deviation based on a weighted-average of the groups’ standard deviations. Commerce, however, has stated that the pooled standard deviation should reflect the average pricing behavior for the two groups, and not necessarily an average of all individual sales. *See* Final I&D Memo at 12 (“The use of a simple average in determining the pooled standard deviation equally weighs *a respondent’s pricing practices to each group* and the magnitude of the sales to one group does not skew the outcome.”) (emphasis added).

Commerce’s decision to use a simple average is reasonable in the targeted dumping context where the nature of the problem is to ferret out certain unlawful pricing behavior, *i.e.*, that higher priced sales are being used to mask other dumped sales. Accordingly, a standard deviation that gives equal weight to the pricing behavior toward a certain purchaser, or in a certain region or period of time, is a reasonable way to create a benchmark by which to measure the differences in a certain group of sales to the overall range of differences in the test and comparison groups. *See Mid Continent Steel & Wire, Inc. v. United States*, 41 CIT __, __, 219 F.

²⁴ Price differences, in this case, refer to differences in the weighted-average net prices of the test and comparison groups.

²⁵ And indeed, the specific numbers given by Stanley show that this was the case here. Pls.’ Br. 27-28.

Supp. 3d 1326, 1342 (2017) (“It is discernible from Commerce’s explanations that Commerce views the pooled standard deviation as an average reflective of the respondent’s average pricing behavior for these two groups, rather than an average reflective of all of the individual prices.”).

In the Final Results, Commerce states that its goal is to determine if an exporter’s pricing behavior as to a certain purchaser, region, or period of time differs significantly from that exporter’s pricing behavior as to all other purchasers, regions, or periods of time, and thus, that an exporter’s pricing behavior in a “test” group is equally important to its pricing behavior in a “control” group. *See* Final I&D Memo at 12. Because of this, Commerce reasonably found that using a simple average achieved this balance:

The pooled standard deviation reflects the dispersion, or variance, of prices within each of the two groups. When the variance of prices is small within these two groups, then a small difference between the weighted-average sale prices of the two groups may represent a significant difference, but when the variance within the two groups is larger (*i.e.*, the dispersion of prices within one or both of the groups is greater), then the difference between the weighted-average sale prices of the two groups must be larger in order for the difference to perhaps be significant. When the difference in the weighted-average sale prices between the two groups is measured relative to the pooled standard deviation, then this value is expressed in standardized units based on the dispersion of the prices within each group. This is the concept of an effect size, as represented in the Cohen’s *d* coefficient.

Final I&D Memo at 12. In other words, any price differences found using Commerce’s Cohen’s *d* test are relative to the variance of prices within the two groups, and thus are tailored to the individual pricing behavior at issue. *See* Final I&D Memo at 12; *see also Soc Trang Seafood Joint Stock Co. v. United States*, 42 CIT __, __, Slip Op. 18-75, at 17 (June 21, 2018) (“Commerce’s [Cohen’s *d* test] evaluates whether the price variance is significant as compared to the actual prices at issue, and not as compared to some other set of prices. The statute allows Commerce to look at individual pricing behavior.”). The court finds this explanation reasonable because Commerce is able to contextualize the magnitude of the pricing differences between the

test and comparison groups, which helps it to determine whether there is a pattern of prices that differ significantly among purchasers, regions, or periods of time. That is, notwithstanding the difference in population and variance between the two groups, the pricing behavior in each group is of equal importance, and therefore, using a simple average to calculate the pooled standard deviation (thereby giving equal weight to the standard deviations in both groups) is reasonable.

Plaintiff's second argument is that there is an "upward bias" in Commerce's Cohen's *d* test calculations which is "systemic." Pls.' Br. 29. Stanley argues that Commerce's use of the Cohen's *d* test in the targeted dumping context, together with its method of calculating the pooled standard deviation, results in a test meant to lead to high pass rates. *See* Pls.' Br. 30. To support its position, Stanley references a chart attached to its initial case brief that reviews the preliminary results of Commerce's proceedings from March 2013 (its first use of the Cohen's *d* test in *Xanthan Gum From the People's Republic of China*) through September 30, 2016 (shortly after Commerce published the Preliminary Results of this review). Pls.' Br. 29 (citing Stanley Case Br., Addendum C, P.R. 269, bar code 3518140-01). For Stanley, the chart demonstrates that "Commerce's incorrect calculations of the Cohen's *d* coefficient generate 'pass' rates that exceed the Department's 33 percent threshold for using the A-T method in over three-quarters of the decisions."²⁶ Pls.' Br. 29. This upward bias, according to plaintiff, "leads to an unreasonably

²⁶ Specifically, Stanley claims:

As of September 30, 2016, Commerce had issued preliminary decisions with respect to 279 respondents that exported a wide variety of merchandise ranging across an array of industries. Of these 279 respondents, the Department found only 25 not to have any sales that "passed" [Cohen's *d* test] and only 45 more to have [Cohen's *d* test] "pass" rates below the 33 percent threshold. The remaining 209 respondents included 95 respondents with [Cohen's *d* test] "pass" rates over 66 percent and three respondents with [Cohen's *d* test] "pass" rates of 100

(footnote continued . . .)

frequent use of the ratio and meaningful difference tests,” which “[does] not effectively protect respondents from the bias inherent in the [Cohen’s *d* test]” and ultimately results in an inappropriate use of the A-T method. Pls.’ Br. 30-31.

Commerce’s use of the Cohen’s *d* test in the targeted dumping context is not “systemically biased” toward finding passing sales. The court has previously explained its view as to the reasonableness of using the Cohen’s *d* test in the targeted dumping context as well as Commerce’s calculation of the pooled standard deviation. *See supra* Part II.A, B. As to the chart cited by Stanley purporting to show an upward bias in its calculation method, the court agrees with defendant that the data fails to establish “that a bias exists in Commerce’s application of the Cohen’s *d* test.” Def.’s Br. 22. Commerce states:

The data show that 207 of the 276 cases cited involved a sufficient percentage of sales passing the Cohen’s *d* test to consider the application of an alternative comparison methodology. Of these, the Department only applied the [A-T] method to either a portion or all of a respondent’s sales in 85 of these 207 determinations. Accordingly, relying upon Stanley’s own data, there does not exist a bias in the Department’s application of the differential pricing analysis, including the Cohen’s *d* test, based on the use of a simple average in determining the pooled standard deviation. *Around one-third of the cases to which Stanley cites resulted in the application of an alternative comparison methodology, representing less than one-half of the cases in which there existed a pattern of prices that differ significantly pursuant to the Cohen’s *d* and ratio tests.*

percent. In other words, Commerce has concluded that 45 percent of the respondents in preliminary decisions each targeted more than two-thirds of their sales – and that three respondents targeted *every* sale. It makes no economic or financial sense for any one company to “target” the majority of its sales. It is unreasonable to conclude that almost half of all investigated companies do so, particularly when those companies sell a wide variety of products under an equally wide variety of market dynamics. Moreover, Commerce’s conclusions that three companies targeted *all* of their sales is simply illogical – if all of a company’s sales are “targeted,” then none can be.

Pls.’ Br. 29-30.

Stanley states that the data show 95 respondents with [Cohen's *d* test] "pass" rates of over 66 percent, and three with "pass" rates of 100 percent. Stanley avers that this demonstrates the unreasonableness of differential pricing because it makes no economic sense for any one company to "target" the majority of its sales, and because if *all* sales are "targeted," then none can be. This line of reasoning demonstrates a misunderstanding of how the Department determines the existence of a pattern of export prices that differs significantly among purchasers, regions, or time periods. Indeed, the focus is not on "targeting" and economic decision-making, but on the difference between export prices. For example, consider two purchasers, A and B. If the prices to purchaser A are found to differ significantly from the prices to purchaser B, then it follows that the prices to purchaser B differ significantly from the prices to purchaser A. Here, it is reasonable to conclude that *all* prices differ significantly. Similarly, if the prices to purchaser A do not differ significantly from the prices to purchaser B, then it follows that the prices to purchaser B do not differ significantly from the prices to purchaser A. Here, it is reasonable to conclude that *none* of the prices differ significantly. While Stanley pointed to three instances where all of the respondent's sales prices differed significantly, there are also 25 cases in the data where none of the sales prices differed significantly. This demonstrates that the Department's approach is reasonable and does not exhibit a bias; the phenomenon to which Stanley points as proof of bias is greatly outweighed by the opposite result, *i.e.* that no sales pass the Cohen's *d* test. *Accordingly, Stanley's own data demonstrate that, if anything, there is a tendency against finding a pattern of prices that differ significantly across purchasers, regions, or time periods.*

Final I&D Memo at 14-15 (emphasis added). In addition, Stanley's own numbers show that the ratio test and meaningful difference test weed out circumstances in which the A-T method need not be applied (*i.e.*, circumstances in which there is not sufficient evidence that targeted dumping may be occurring). Therefore, since less than half of the cases cited in Stanley's numbers resulted in an application of the A-T method, it is apparent that there is no unreasonable, or biased, result in Commerce's use of the Cohen's *d* test.

Finally, Stanley claims that it is "unreasonably difficult" for a respondent to revise its pricing to avoid high "pass" rates "because the standard deviations of the test and comparison groups comprising the pooled standard deviation become smaller as any differences in a respondent's prices for that CONNUM are eliminated." Pls.' Br. 31. Plaintiff then argues that Commerce's calculation fails to account for "decreases in the size of price variances that result

from a respondent's efforts to eliminate differences in its prices." Pls.' Br. 32. For plaintiff, because "smaller price differences render smaller pooled standard deviations" in Commerce's application of the Cohen's *d* test, Cohen's *d* coefficients will fall into the "large" category (and thus, "pass" the Cohen's *d* test) even if a respondent attempts to attain price homogeneity. Pls.' Br. 32.

Stanley's argument appears to misunderstand the relation of the Cohen's *d* test to the statute. The Cohen's *d* test does not determine whether Commerce will calculate a dumping margin using the A-T method, but rather, is only one of two tests²⁷ used to determine whether prices differ significantly, *i.e.*, whether there is a pattern of differing prices for comparable merchandise among purchasers, regions, or periods of time. Indeed, under the ratio test, before Commerce can even consider applying the A-T method to any of Stanley's sales, more than 33 percent of its total sales value must pass the Cohen's *d* test. In addition, even if Commerce's Cohen's *d* and ratio tests suggest there is a pattern of export prices that differ significantly among purchasers, regions, or periods of time, such that Commerce may consider the application of the A-T method, it still must explain why the A-A method cannot account for these differences.

As the Department noted, "[a] company may sell subject merchandise in the United States market at significantly different prices, yet none of these sales are priced at less than normal value," and that in such situations, "the [A-A] method will be able to account for such differences" because there are no dumped sales. Final I&D Memo at 15. Moreover, in the hypothetical suggested by plaintiff, where an exporter has changed its pricing practices to attain near homogeneity, there will likely not be a "meaningful difference" between the margin

²⁷ The other test is the ratio test.

calculated using the A-A method and that calculated using the A-T method. This is because, under such circumstances, the weighted-average export price (*i.e.*, the export price calculated using the A-A method) would be very close to the price of individual transactions in the United States, and therefore, the A-A method would be deemed able to account for such differences. *See infra* Part II.C.ii. Thus, high Cohen's *d* pass rates do not automatically lead to the application of the A-T method. In any event, all that is required of Commerce under the statute at this stage in its analysis is to determine whether "there is a pattern of export prices (or constructive export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time." 19 U.S.C. § 1677f-1(d)(1)(B)(i). Commerce's calculation of the Cohen's *d* test, in conjunction with its ratio test, is a reasonable method for making this determination.

C. Differential Pricing Does Not Contravene the Statute

i. The Ratio Test

Following the Cohen's *d* test, Commerce uses the "ratio test" to "assess[] the extent of the significant price differences for all sales as measured by the Cohen's *d* test." Preliminary I&D Memo at 20. If the value of sales to certain purchasers, regions, and time periods that "pass" the Cohen's *d* test accounts for 66 percent or more of the value of a respondent's total sales, then, for Commerce, "the identified pattern of prices that differ significantly supports the consideration of the application of the [A-T method] to all sales" Preliminary I&D Memo at 20. If the value of passing sales accounts for 33 percent or less of the value of a respondent's total sales, however, then the results do not support the application of the A-T method to any of the respondent's sales. If the value of passing sales is between 33 and 66 percent of the value of a respondent's total sales, then Commerce may consider the application of the A-T method for all

passing sales, but the A-A method will be used for all remaining sales. *See* Preliminary I&D Memo at 20.

Stanley argues that the differential pricing analysis fails to meet either of the two preconditions necessary before Commerce may apply the A-T method under 19 U.S.C. § 1677f-1(d)(1)(B). Pls.' Br. 32. That is, for Stanley, the differential pricing analysis does not identify a "pattern" of prices that differ significantly among purchasers, regions, or periods of time, nor does it explain why the A-A method cannot account for such differences. Stanley asserts that this is because (1) the "ratio" test merely "stratifies Cohen's *d* test pass rates," it does not describe a pattern; and (2) the meaningful difference test fails to explain why Commerce cannot account for a perceived price difference using the A-A method. Pls.' Br. 33, 35.

Defendant responds that "Commerce explained in the final results how the stratification of pass rates under the Cohen's *d* test identifies a pattern of prices that differ significantly." Def.'s Br. 26. According to defendant, Commerce uses the ratio test to "complete its determination of whether there exists a pattern of prices that differ significantly by purchaser, region, or period of time" because, even if sales for one or more groups of comparable merchandise may pass the Cohen's *d* test, "it does not necessarily follow that, in relation to the total volume of a respondent's export sales, there is sufficient evidence that a pattern of prices exists that differ significantly." Def.'s Br. 26. In other words, for Commerce, the ratio test completes Commerce's determination of whether a pattern of prices exists that differ significantly by "assess[ing] the extent of the significant price differences for all sales as measured by the Cohen's *d* test." Preliminary I&D Memo at 20.

Commerce has reasonably explained how the ratio test, in conjunction with the Cohen's *d* test, satisfies 19 U.S.C. § 1677f-1(d)(1)(B)(i) (*i.e.*, how the tests identify a "pattern of export

prices” for comparable merchandise that “differ significantly among purchasers, regions, or periods of time.”). Here, Commerce has found that, when the value of a respondent’s U.S. sales that “pass” the Cohen’s *d* test accounts for more than 33 percent of the value of its total sales, this indicates a pattern of price differences exists such that Commerce may consider applying the A-T method to a limited amount of the respondent’s sales. *See* Final I&D Memo at 18. Likewise, Commerce maintains that when the value of a respondent’s U.S. sales that “pass” the Cohen’s *d* test accounts for 66 percent or more of the value of its total sales, this indicates there exists a pattern of price differences such that Commerce may consider applying the A-T method to all of the respondent’s sales. *See* Final I&D Memo at 17-18. By creating these thresholds, Commerce reasonably identified when price differences are more than just random occurrences, *i.e.*, when a “pattern” exists. Indeed, in order for Commerce to apply A-T to all of a respondent’s sales, most of the respondent’s sales (roughly two thirds) must have “passed” the Cohen’s *d* test, a threshold unlikely to be the result of chance.

This method is a reasonable one for meeting the prerequisite of § 1677f-1(d)(1)(B)(i), particularly since the statute gives no guidance as to how Commerce should make its determination. 19 U.S.C. § 1677f-1(d)(1)(B); *see also* Final I&D Memo at 17 (“Neither the statute nor the SAA^[28] provide any guidance in determining how to apply the [A-T] method once the requirements of [19 U.S.C. § 1677f-1(d)(1)(B)(i)] and (ii) have been satisfied. Accordingly, the Department has reasonably created a framework to determine how the [A-T] method may be

²⁸ Statement of Administrative Action accompanying the Uruguay Round Agreements Act (“SAA”), H.R. Doc. No. 103-316, vol. 1, at 842-43, *reprinted in* 1994 U.S.C.C.A.N. 4040, 4177-78. The SAA “shall be regarded as an authoritative expression by the United States concerning the interpretation and application of the Uruguay Round Agreements and this Act in any judicial proceeding in which a question arises concerning such interpretation or application.” 19 U.S.C. § 3512(d).

considered as an alternative to the standard [A-A] method based on the extent of the pattern of prices that differ significantly as identified with the Cohen's *d* test.”). Commerce was faced with the task of creating a method for determining when it should use the A-T method. Stanley has failed to show that Commerce's method does not do what it is supposed to do. Accordingly, the court finds that Commerce's use of the ratio test is a reasonable interpretation of § 1677f-1(d)(1)(B)(i).

ii. The Meaningful Difference Test

Under the meaningful difference test, Commerce first calculates the dumping margin that would result by applying the A-A method to all sales, *i.e.*, Commerce calculates a dumping margin the same way that it would absent any targeted dumping procedures. Commerce then calculates two additional dumping margins: (1) by applying the A-T method to all sales that passed the Cohen's *d* test and the A-A method to the remaining sales, and (2) by applying the A-T method to all sales.²⁹ Preliminary Analysis Memorandum at 16. Depending on the results of the ratio test,³⁰ Commerce then compares (1) the margin calculated under its normal method (*i.e.*,

²⁹ While Commerce states that “the Department tests whether using an alternative comparison method, based on the results of the Cohen's *d* and ratio tests described above, yields a meaningful difference in the weighted-average dumping margin as compared to that resulting from the use of the [A-A] method only,” Preliminary I&D Memo at 20, the Amended Final Results Analysis Memo shows that Commerce actually calculated three margins: (1) by applying the A-A method to all sales; (2) by applying the A-T method to those sales that passed the Cohen's *d* test and the A-A method to all remaining sales; and (3) by applying the A-T method to all sales. *See* Amended Final Results Analysis Memo at 2. The Department then, based on the results of the ratio test, selects the appropriate A-T method and compares that margin to the margin calculated using the A-A method. Amended Final Results Analysis Memo at 2.

³⁰ As described above, the sales to which Commerce will apply the A-T method (provided a “meaningful difference” is found) depends on the results the ratio test. If the results of the ratio test indicate that passing sales represent 66 percent or more of a respondent's total sales value, Commerce will use the margin calculated by applying A-T to all sales for its “meaningful difference” comparison. If the passing sales represent more than 33 percent and less

using the A-A method), and (2) the dumping margin calculated using the A-T method, to determine if there is a “meaningful difference” between the two. Preliminary I&D Memo at 20. Commerce considers there to be a “meaningful difference” when the comparison demonstrates (1) that there is a 25 percent relative change in the weighted-average dumping margin between the A-A method and the appropriate A-T method where both margins are above the *de minimis* threshold; or (2) that the A-T method generates a dumping margin that crosses the *de minimis* threshold when compared to the A-A method. If a meaningful difference exists, Commerce infers that the A-A method is unable to account for the price differences among particular purchasers, regions, or in particular periods of time (*i.e.*, that the A-A method would not “unmask” observed pricing differences which evidence targeted dumping). *See Apex Frozen Foods Private Ltd. v. United States*, 862 F.3d 1337, 1348 (Fed. Cir. 2017) (“*Apex II*”) (“Commerce’s meaningful difference analysis—comparing the ultimate antidumping rates resulting from the A-A methodology, without zeroing; and the A-T methodology, with zeroing—was reasonable.”).

Notwithstanding the Federal Circuit’s approval of Commerce’s meaningful difference test³¹ (applied and explained in the same manner as Commerce has done so here), Stanley argues that the Court has not addressed its argument, which is that the meaningful difference test is “flawed methodologically” because Commerce performs its A-A and A-T comparison “based on Stanley’s *total sales* even though it performed the [Cohen’s *d* test] based on *sales of individual CONNUMs*.” Pls.’ Br. 37, 39-40 (“By separating the basis for its determination of a

than 66 percent of a respondent’s sales, then Commerce will use the margin calculated using the A-T method on passing sales and the A-A method on remaining sales.

³¹ *Apex II*, 862 F.3d at 1348.

meaningful difference from the specific products that displayed significant price differences Commerce failed to meet its statutory burden to explain why [the A-A method] could not account for those price differences”). Therefore, Stanley claims that “the methodological error that is fatal to the meaningful difference test was not at issue” in *Apex II*. Pls.’ Br. 37; *see also* Pls.’ Reply Br., ECF No. 32, 12 (“While the Federal Circuit was explicit in approving Commerce’s rationale . . . it has not addressed . . . the question Stanley has raised here concerning whether Commerce’s specific implementation of the meaningful difference test contravenes 19 U.S.C. § 1677f-1(d)(1)(B)(ii).”).

For Stanley, the absence of a “reasonable nexus” between the meaningful difference test and the Cohen’s *d* test not only “produce[s] distorted results,” but also represents an unreasonable interpretation of 19 U.S.C. § 1677f-1(d)(1)(B). Pls.’ Br. 37. Stanley’s argument is based on its reading of the “such differences” language found in § 1677f-1(d)(1)(B)(ii)’s requirement that Commerce “explain why *such differences* cannot be taken into account using [the A-A] method” 19 U.S.C. § 1677f-1(d)(1)(B)(ii) (emphasis added).³² Stanley claims

³² Section 1677f-1(d)(1)(B) provides:

[Commerce] may determine whether the subject merchandise is being sold in the United States at less than fair value by comparing the weighted average of the normal values to the export prices (or constructed export prices) of individual transactions for comparable merchandise, if—

- (i) there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time, and
- (ii) [Commerce] explains why *such differences* cannot be taken into account using [the A-A method]

19 U.S.C. § 1677f-1(d)(1)(B) (emphasis added).

that the “such differences” language references the “prices” portion of the “pattern of export *prices* for comparable merchandise that *differ* significantly” language found in the statute. Pls.’ Br. 37 (citing 19 U.S.C. § 1677f–1(d)(1)(B)(i) (emphasis added)); Transcript of Oral Argument, ECF No. 40 at 6-7. Thus, because Commerce found significant pricing differences using a CONNUM-specific approach (the Cohen’s *d* test), Stanley argues that Commerce must also conduct its meaningful difference test on a CONNUM-specific basis, *i.e.*, by applying the A-A method to sales of individual CONNUMs, rather than to Stanley’s overall sales.

Although the Federal Circuit did not specifically address the argument raised by Stanley, its holding nonetheless directs the court to find for the Government. As the *Apex II* Court noted, “Commerce devised its meaningful difference test, in which antidumping rates—*as they would ultimately be applied for the A-A methodology versus an alternative*—are compared, across *all sales*,” and concluded that “there is no basis (statutory or otherwise) for demanding a distinction between the meaningful difference analysis and the ultimate margin calculation.” *Apex II*, 862 F.3d at 1346, 47 (emphasis added). Thus, the Federal Circuit was fully aware of the method by which the meaningful difference test was conducted and approved its use. Also, in “assess[ing] whether Commerce’s reading of the statute was permissible and whether its implementation was otherwise . . . unreasonable,” the Federal Circuit specifically found that the meaningful difference test, that is, “comparing the ultimate antidumping rates resulting from the A-A methodology” with the appropriate A-T method, “was reasonable.” *Id.* at 1348.

Here, as Commerce states, “finding that there exists a pattern of prices that differ significantly means only that the Department will consider whether the standard comparison methodology can account for such differences,” *i.e.*, whether using the A-A method as it would ultimately be applied could account for the pattern of price differences found using the Cohen’s

d test. Final I&D Memo at 15. For Commerce, “comparing the weighted-average dumping margins calculated using the two comparison methods allows the Department to quantify the extent to which the [A-A] method cannot take into account different pricing behaviors exhibited by the exporter in the U.S. market.” Final I&D Memo at 13. The court agrees. The meaningful difference test fulfills the statutory requirement that Commerce explain why the A-A method cannot account for the perceived pattern of pricing differences. Moreover, the Federal Circuit has noted that “[u]nder a plain reading of the statute [19 U.S.C. § 1677f-1(d)(1)(B)(ii)], the use of ‘such differences’ does not, in itself, manifest Congress’s intent to dictate how Commerce is to make the determination whether the A-A method[] can account for potential targeted or masked dumping.” *Apex II*, 862 F.3d at 1345. Thus, Commerce’s approach has been approved by the Federal Circuit, and the court therefore finds that it was also reasonable here.

Accordingly, the court finds the meaningful difference test, as applied, to be lawful under 19 U.S.C. § 1677f-1(d)(1)(B)(ii).

D. Differential Pricing Does Not Contravene Congressional Intent as Expressed in the Legislative History

In the Final Results, Commerce found that 77.8 percent of Stanley’s U.S. sales “passed” the Cohen’s *d* test, and therefore, using the ratio test,³³ applied the A-T method to all of Stanley’s sales for the POR. Amended Final Results Analysis Memorandum at 2. Notably, Commerce deemed sales to have “passed” the Cohen’s *d* test whether they passed because the test group’s

³³ As discussed above, the ratio test provides that if the value of sales to certain purchasers, regions, and time periods that “pass” the Cohen’s *d* test account for 66 percent or more of the value of a respondent’s total sales, then Commerce considers there to be an “identified pattern of prices that differ significantly” such that it may consider the application of the A-T method to all sales. Preliminary I&D Memo at 20.

sales were higher priced than the comparison group or lower priced than the comparison group, with no inquiry into whether passing sales were actually dumped.³⁴ Final I&D Memo at 16. Stanley argues that “Commerce’s failure to limit its targeting analysis to sales that ‘pass’ the [Cohen’s *d* test] with ‘low’ prices conflicts with the SAA’s express statement that ‘targeted dumping’ comprises prices that are *both* dumped and below prices ‘to other customers.’” Pls.’ Br. 42 (“[T]he standard described in the SAA is prices ‘to other customers,’ not a price to ‘any other customer,’ evidencing Congress’ intent that the possibility of targeted dumping is to be measured in relation to prices below the general norm.”). Thus, for plaintiff, “[b]y embracing higher than normal price sales as evidence of ‘targeting,’” the differential pricing analysis “contravenes Congress’s intent as to what comprises the problem—targeted dumping—that Commerce is authorized to address.” Pls.’ Br. 42. Stanley thus argues that Commerce’s approach does not properly address targeted dumping, as it is supposed to, because Commerce considers sales that are sold at a higher price than other sales to be evidence of targeted dumping.

Stanley then claims that “embracing higher than normal prices as evidence of ‘targeting’ is conceptually absurd.” Pls.’ Br. 43. Stanley reasons that because “[t]he only rational reason to ‘target’ is to gain sales,” a seller cannot “successfully gain sales by charging the allegedly ‘targeted’ customer a higher price than it charges other customers for identical merchandise.” Pls.’ Br. 43. Therefore, Stanley claims that the Final Results are unlawful because they ignore the intent of the statute as articulated in the SAA to focus only on sales that were lower than the norm. Pls.’ Br. 43.

³⁴ That is, as long as there was a 0.8 standard deviation difference between the test and comparison groups, Commerce considered the sales to have passed the Cohen’s *d* test.

The court is not persuaded that the differential pricing analysis runs counter to congressional intent. As an initial matter, the statute does not specify whether prices must “differ” by being priced lower or higher than comparison sales. *See* 19 U.S.C. § 1677f-1(d)(1)(B). Thus, Commerce has not violated the plain language of the statute. Moreover, as the Department emphasized, “higher priced sales will offset lower priced sales, either implicitly through the calculation of a weighted-average sale price for a U.S. averaging group, or explicitly through the granting of offsets when aggregating the [A-A] comparison results, that can mask dumping.” Final I&D Memo at 16. Therefore, when Commerce calculates the weighted-average export price (or constructed export price) for sales included in a particular averaging group,³⁵ higher priced sales may drive the averaging group’s export price up, potentially concealing dumped sales within the group. In addition, when aggregating the results of the averaging groups to determine the weighted-average dumping margin, higher priced sales could result in averaging groups for which the weighted-average export price exceeds the weighted-average normal value, which would offset the results of any averaging groups for which the weighted-average export price is less than the weight-average normal value. Therefore, higher priced sales are relevant to Commerce’s analysis. This is consistent with the SAA’s description of “concealed” targeted dumping, which, according to the text, occurs when “an exporter may sell at a dumped price to particular customers or regions, while selling at higher prices to other customers or regions.” SAA at 842, 1994 U.S.C.C.A.N. at 4177-78. Thus, considering that the purpose of applying the

³⁵ An averaging group consists of “subject merchandise that is identical or virtually identical in all physical characteristics and that is sold to the United States at the same level of trade.” 19 C.F.R. § 351.414(d)(2).

A-T method is to unmask targeted dumping, Commerce's consideration of "higher priced" sales (which may mask lower priced, or dumped, sales) is reasonable.

As to Stanley's argument that the SAA links "targeting" with "dumping," the court is also not convinced that the only sales relevant when determining whether prices differ significantly are those that are lower priced than the comparison group. First, the SAA mentions that the targeted dumping statute (19 U.S.C. § 1677f-1(d)(1)(B)) will provide a comparison method in situations where the A-A or T-T method cannot account for a pattern of prices that differ significantly among purchasers, regions, or time periods, *i.e.*, "where targeted dumping *may* be occurring." SAA at 843, 1994 U.S.C.C.A.N. at 4178 (emphasis added). This statement does not, on its face, confine Commerce's method to solely analyzing sales at less than fair value, nor does it require Commerce to make an affirmative finding of targeted dumping. *See Stanley Works (Langfang) Fastening Sys. Co. v. United States*, 41 CIT __, __, 279 F. Supp. 3d 1172, 1191 (2017). As has been previously stated, the Cohen's *d* test in no way measures dumping—it only identifies a pattern of differing prices. In fact, every sale used to reach a finding that there was such a pattern could be dumped or not dumped. That is, merely because a sale is high in relation to the mean does not tell Commerce anything about whether or not it is a sale at less than fair value (*i.e.*, "dumped"). At the initial stage of its analysis, Commerce is only tasked with determining whether there is a pattern of prices that differ significantly. If such a pattern is found, Commerce will consider whether the A-A method can account for these differences, and if it cannot, the SAA considers this to be evidence that targeted dumping may be occurring.

In addition, the SAA itself anticipates that targeted dumping encompasses "situations [in which] an exporter may sell at a dumped price to particular customers or regions, while selling at

higher prices to other customers or regions” and thus, explicitly considers higher priced sales to be relevant. SAA at 842, 1994 U.S.C.C.A.N. at 4177-78 (emphasis added). Thus, not only does the SAA contemplate considering higher prices in the targeted dumping context, but also, as the Department states, by “considering all sales, higher priced sales and lower priced sales, the Department is able to analyze an exporter’s pricing practice and to identify whether there is a pattern of prices that differ significantly” by purchaser, region, or period of time. Final I&D Memo at 16. As this Court has found, “[a]ll sales are subject to the differential pricing analysis because its purpose is to determine to what extent a respondent’s U.S. sales are differentially priced, not to identify dumped sales,” and therefore, “Commerce is not restricted in what type of sales it may consider in assessing the existence of such a pattern so long as its methodological choice enables Commerce to reasonably determine whether application of A-T is appropriate.” *Apex I*, 144 F. Supp. 3d at 1330.

In the end, plaintiff’s argument appears to conflate passing the Cohen’s *d* test with the application of the A-T method and ultimately “unmasking” targeted dumping. The latter, however, requires not only a finding of a pattern of prices that differ significantly among purchasers, regions, or periods of time, but also an explanation as to why the A-A method cannot account for such differences and a finding of dumping using A-T. These are separate analyses, and a high result in the first does not necessarily determine the result of the second. Therefore, the court finds that the differential pricing analysis is not inconsistent with congressional intent, and Commerce reasonably considered both higher priced sales and lower priced sales in evaluating whether there exists a pattern of export prices that differ significantly among purchasers, regions, or periods of time.

E. Commerce's Implementation of the Differential Pricing Analysis is Reasonable

Next, Stanley argues that the procedure Commerce uses to form comparison groups in its differential pricing analysis also results in high Cohen's *d* test pass rates, and therefore, is an unreasonable interpretation of the statute. According to Stanley, this is because Commerce includes sales from test groups that "pass" the Cohen's *d* test in its base (or "comparison") groups, thereby causing other sales to "pass" the Cohen's *d* test that otherwise would not have passed. Pls.' Br. 44-45. Plaintiff thus argues that "when Commerce finds a sale in a test group to pass the [Cohen's *d* test], it nevertheless includes the anomalous price of that sale in the comparison (*i.e.*, base) group used to evaluate the prices of other test groups," which results in "passing" sales that would otherwise not pass. Pls.' Br. 45. Therefore, plaintiff argues, Commerce is double-counting irregular sales prices.

Plaintiff then maintains that the problem is exacerbated because of Commerce's "refusal to consider any of the many circumstances of sale that cause net prices to vary" such as movement costs, credit costs, or warranty costs. Pls.' Br. 45. As a result, plaintiff argues, even if a respondent sells products having the same CONNUM to all customers at the same gross price, adjustments to the U.S. selling price could nonetheless cause a sale to "pass" the Cohen's *d* test. Pls.' Br. 45-46. For Stanley, it is unreasonable for Commerce to conduct the Cohen's *d* test at a net price level because "the antidumping statute overtly recognizes the potential for different circumstances of sale to distort the calculation of dumping margins," and therefore, "expressly directs Commerce to correct for such distortions by adjusting normal values." Pls.' Br. 46 (citing

19 U.S.C. § 1677b(a)(6)(C)³⁶). Stanley thus claims that “[i]t is unreasonable for Commerce to account for differences in circumstances of sale when calculating dumping margins^[37] but not when determining whether such dumping was targeted.” Pls.’ Br. 46.

The court finds that Commerce’s method is reasonable. As to plaintiff’s double-counting theory, the court agrees with this Court’s analysis in *Timken*:

The purpose of Commerce’s [differential pricing] analysis is to find a pattern of prices that differ significantly Under Commerce’s methodology, even if some sales are included in a test group and later in a comparison group, their value is counted only once in the numerator of the ratio [test] if they pass Cohen’s *d*.

Timken, 179 F. Supp. 3d at 1178-79. Put simply, in determining whether the total value of sales that “pass” the Cohen’s *d* test is such that Commerce might consider the application of the A-T method (*i.e.*, whether the value of passing sales is greater than 33 percent of a respondent’s total sales value), Commerce counts the value of any particular passing sale only once in the numerator.

Moreover, to remove passing sales from subsequent comparison groups because they are, as Stanley suggests, “anomalous” would lead to inconsistent results. As Commerce stated:

³⁶ Title 19 U.S.C. § 1677b(a)(6)(C) provides, in pertinent part, that the normal value shall be

increased or decreased by the amount of any difference (or lack thereof) between the export price or constructed export price and the price described in paragraph (1)(B) (other than a difference for which allowance is otherwise provided under this section) that is established to the satisfaction of [Commerce] to be wholly or partly due to . . . other differences in the circumstances of sale.

19 U.S.C. § 1677b(a)(6)(C)(iii).

³⁷ As noted above, to calculate a dumping margin, Commerce determines the difference between the export price (or constructed export price) and the normal value of the product.

If the weighted-average price to purchaser A differs significantly from the weighted-average price to purchaser B, then the weighted-average price to purchaser B also differs significantly from the weighted-average price to purchaser A. Stanley's suggestion, that once the Department finds that the weighted-average price to purchaser A differs significantly from the weighted-average price to purchaser B, then the sales prices to purchaser A should be excluded henceforth from the analysis, is illogical. This would result in no comparison being made for the weighted-average price to purchaser B. Further, if purchaser B's sales were tested first, then purchaser A's sales would not be tested. Such an approach would lead to arbitrary and unpredictable results that would depend upon the order in which purchasers, regions or time periods were examined.

Final I&D Memo at 18-19. Similarly, if sales from purchaser A to purchaser B were found not to have passed the Cohen's *d* test, then so too will the sales from purchaser B to purchaser A, and the value of both will be included in the denominator of the ratio test. *See Timken*, 179 F. Supp. 3d at 1178-79. Stanley's argument does not make Commerce's rationale unreasonable.

In addition, the court finds that the use of net prices in the differential pricing analysis is a reasonable interpretation of the statute. As the Department states, its "analysis is to determine whether the [A-A] method is appropriate to measure the amount of dumping for a respondent" and that to "calculate a weighted-average dumping margin . . . , the Department uses net U.S. prices" Final I&D Memo at 13. Therefore, Commerce considered the use of net prices "consistent with the view that discounts, rebates and similar price adjustments are not expenses, but instead form part of the price itself." Final I&D Memo 13. This interpretation is reasonable as it appears to implement the intent of the statute (*i.e.*, to determine whether the A-A method is the appropriate tool with which to measure a respondent's dumping). Also, as Commerce emphasized, "the use of net U.S. prices would increase the variability of the sale prices within a group and thus require a larger difference in the weighted-average sale prices between the two groups" Final I&D Memo at 14. Therefore, the court finds that Commerce's use of net prices in its differential pricing analysis is a reasonable interpretation of the statute.

At bottom, plaintiff once again appears to conflate passing the Cohen's *d* test with the application of the A-T method, and ultimately, a finding that there is targeted dumping. As discussed above, (1) finding a pattern of prices that differ significantly among purchasers, regions, or periods of time, and (2) explaining why the A-A method cannot account for such differences are two separate analyses. The results of the former does not necessarily determine the result of the latter. Accordingly, the court finds that Commerce's differential pricing analysis is a reasonable interpretation of 19 U.S.C. § 1677f-1(d)(1)(B).

III. The World Trade Organization Appellate Body Decision

Finally, Stanley argues that the World Trade Organization ("WTO") Appellate Body decision in *United States—Anti-Dumping and Countervailing Measures on Large Residential Washers from Korea*³⁸ demonstrates that Commerce has interpreted and applied 19 U.S.C. § 1677f-1(d)(1)(B) in an unreasonable manner that is inconsistent with the United States' international obligations. Pls.' Br. 47. Specifically, plaintiff argues that Commerce's differential pricing analysis violates the Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994 because (1) "Commerce did not limit its 'pattern' analysis [to] sales that 'pass' the [Cohen's *d* test] because they are lower than the comparison group mean"; and (2) "Commerce employed a rote application of a series of mathematical formulae in the guise of 'tests' . . . while ignoring the nature of any factors causing price differences . . . and

³⁸ Appellate Body Report, *United States—Anti-Dumping and Countervailing Measures on Large Residential Washers from Korea*, WTO Doc. WT/DS464/AB/R (adopted Sept. 7, 2016).

thus considered only quantitative criteria.”³⁹ Pls.’ Reply Br. 18 (citing the Appellate Body Report, *United States—Anti-Dumping and Countervailing Measures on Large Residential Washers from Korea*, ¶¶ 101, 102, WTO Doc. WT/DS464/AB/R (adopted Sept. 7, 2016)). In other words, Stanley uses *Washers from Korea* to illustrate its view that Commerce’s interpretation of what constitutes “a *pattern* of export prices . . . for comparable merchandise that differ *significantly* among purchasers, regions, or periods of time” pursuant to 19 U.S.C. § 1677f-1(d)(1)(B) is unreasonable because it violates the WTO agreement. *See* Pls.’ Br. 47 (emphasis added).

This argument is unconvincing. WTO decisions are irrelevant to the interpretation of domestic U.S. law. *See* 19 U.S.C. § 3512(a)(1) (“Nothing in [the Uruguay Round Agreements Act] shall be construed . . . to amend or modify any law of the United States.”); *see also Corus Staal BV v. Dep’t of Commerce*, 395 F.3d 1343, 1348 (Fed. Cir. 2005) (“WTO decisions are ‘not binding on the United States, much less this court.’” (quoting *Timken Co. v. United States*, 354 F.3d 1334, 1344 (Fed. Cir. 2004))); *see also Corus Staal BV*, 354 F.3d at 1346 (“Commerce is not obligated to incorporate WTO procedures into its interpretation of U.S. law.”). Further, “[t]he SAA provides that ‘[r]eports issued by . . . the Appellate Body under the [WTO Dispute

³⁹ The court notes that, in its opening brief, plaintiff argued that (1) “Commerce did not limit its ‘pattern’ analysis to sales that ‘pass’ the [Cohen’s *d* test] because they are lower than the comparison group mean”; (2) “Commerce applied the A-T comparison methodology to all of Stanley’s sales”; (3) “Commerce employed a rote application of a series of mathematical formulae in the guise of ‘tests’”; and (4) “Commerce used A-T with zeroing both in the meaningful difference test and in the calculation of Stanley’s dumping margin” in contravention of the *Washers from Korea* Appellate Body decision. Pls.’ Br. 47-48. In its reply brief, however, plaintiff claims that only “[t]wo of [the Appellate Body’s] reasons [why differential pricing violates the Agreement] support a conclusion that the Final Results are unreasonable and should be remanded.” Pls.’ Reply Br. 18. Accordingly, the court will address only the two arguments that remain in plaintiff’s subsequent reply brief.

Settlement Understanding] have no binding effect under the law of the United States . . . [and] do not provide legal authority for federal agencies to change their regulations or procedures.”
Corus Staal BV v. U.S. Dep’t of Commerce, 27 CIT 388, 399, 259 F. Supp. 2d 1253, 1264 (2003) (citing SAA at 1032, 1994 U.S.C.C.A.N. at 4318).

Issues brought before WTO panels and the Appellate Body deal with whether a country is complying with the terms of the WTO Agreement. *See Corus Staal BV v. United States*, 29 CIT 777, 786, 387 F. Supp. 2d 1291, 1300 (2005). Cases brought before the Court of International Trade present questions dealing with domestic U.S. law. *Id.* (“In sum, the WTO decision-making process operates apart from the decision-making in this court. WTO decision-making starts with an international agreement, which may not match the domestic statute and which is interpreted pursuant to different principles.”). Commerce’s interpretation of a statute might well be a perfectly reasonable interpretation of U.S. law and nonetheless be found to violate the WTO Agreement, as, for instance, was the case with zeroing. *See, e.g., id.* Thus, plaintiff’s argument that the Appellate Body’s decision in *Washers from Korea* somehow shows that Commerce’s interpretation and implementation of the targeted dumping statute is unreasonable under U.S. law is far wide of the mark.

CONCLUSION

For the foregoing reasons, the court finds that Commerce's method is a reasonable one for determining if targeted dumping may be occurring and therefore denies plaintiff's motion for judgment on the agency record. Commerce's Final Results are sustained. Judgment shall be entered accordingly.

/s/ Richard K. Eaton

Richard K. Eaton, Judge

Dated: August 13, 2018
New York, New York