Some statistical aspects of the Department's use of Cohen's D in measuring differential pricing in Anti-Dumping cases that should be considered before it is formally adopted.

Submitted by Professors Joseph L. Gastwirth, Reza Modarres and Qing Pan[1]

Motivated by the Department of Commerce's request for comments on is "differential pricing" analysis, the authors have investigated the statistical properties of the first phase of the Department's method, which uses Cohen's D statistic and a "ratio" test to determine the method, average to average or average to transaction (with or without zeroing) will be used to calculate the dumping margin.

Differential Pricing Analysis relies on the effect size measure of the difference between the means (averages) of two groups and uses Cohen's D statistic to measure the size of an effect. It is defined as the standardized difference between the means or averages of between two groups. Unlike significance tests such as the two-sample t-test, Cohen's D statistic is independent of sample sizes. This measure was popularized by Psychologist Jacob Cohen in his book "Statistical Power Analysis for the Behavioral Sciences" (1969, 1988). When the spreads or variability of the data within each of the two groups are the similar, Cohen's D is the ratio of the difference in means to a commonly used measure of spread, the standard deviation. For example, suppose the standard deviation of scores on a standard math achievement test is about 15. If a new teaching method is superior to the existing one, after a year, students taught by it might have an average score of 80, while students taught with the existing method might have an average score of 75. The effect size is the ratio of the difference of five points to 15, or one-third.

In this comment we discuss the current way DOC uses Cohen's D to detect "hidden dumping" and demonstrate that it has desirable statistical properties when the observation (prices during the period under review) can be considered as a random sample from a large population of prices. In particular, each sale price is statistically independent of the others. This means that knowing that the price of one sale is above (or below) average does not provide information about whether another sales price is above (or below) average. Then it will be shown that if this assumption is *not* satisfied, the current procedure can *erroneously classify* a firm that is not dumping but produces a product whose price structure is affected by seasonal patterns, e.g. Christmas items, or a sudden increase due to a shock to the system, e.g. a serious weather or political event, as dumping its product.

Section 1 of the comment describes the way the DOC currently uses Cohen's D statistic in its procedure for classifying dumping. Section 2 presents the results of our simulation and recommendations and suggestions for further study of the method before it is formally adopted

[1] The authors are Professors in the Department of Statistics of the George Washington University, Washington DC, 20052. The analysis reported in the comment, the inferences made from them and views expressed are our own and not the University's.

are given in the last section. An Appendix provides more background on Cohen's D statistic and other statistical topics mentioned in the comment.

1. Review of the use of Cohen's D statistic and "ratio test"

Before examining the statistical properties the DOC procedure it is useful to review the method. The prices of relevant sales are categorized in three ways: by quarter, region and purchaser. When the Department reviews whether there was dumping in a particular quarter, say, it carries out a three-step procedure. First Cohen's D is calculated comparing each quarter's sales with those of the other three quarters. Then it compares each purchasers sales prices to the prices paid by the other purchasers and finally it compares the prices in each region to those of the remaining regions. Here, we describe the quarterly comparisons as the others are similar from a statistical viewpoint. The commonly thresholds for determining whether the effect size (difference in average price in the test quarter to that of the other three), is small, medium or large, .2. .5 and .8, respectively are used to classify whether the difference in averages is "significant". The Department uses the large threshold (.8) to determine whether the sales of the test group pass Cohen's D test. If the proportion of sales with at least one Cohen's D statistic value greater or equal to 0.8 is 0.66 or more then the Department will apply the average to transaction method to determine the dumping margin. If the proportion of sales that "pass" the .8 threshold is between .33 and .66, then the Department considers using the average to transaction method to estimate the dumping margin for those sales that "pass" the Cohen's D criteria. If no more than 33% of the value of total sales "pass" the Cohen's criteria, then the Department will not consider an alternative method to the average-to-average method of determining a dumping margin.

It is important to note that each sale is tested numerous times during the DOC method as it is included in the quarter to remaining quarter comparisons, region vs. remaining regions comparisons and in the purchaser vs. all other purchaser comparisons. Furthermore, each sale is in one test group for each of the three analyses (quarter, region, purchaser) but is in the "control" or comparison group for the other comparisons based on Cohen's D. This makes the statistical comparisons dependent, which ordinarily would complicate the derivation of the probability a particular comparison would meet the Cohen's D is at least .8 criteria. Therefore, the next section reports the results of a simulation to assess the statistical characteristics of the DOC method using Cohen's D and "ratio" tests.

Section 2: Statistical properties of the DOC procedure

As the price data from actual cases are not available and usually redacted from legal decisions, we simulated data fixing the numbers of sales in each region, quarter and purchaser. Three data configurations (mean +SD structure) are generated. We assume sales prices are like a random sample from a normal distribution with the same mean $4.08 and standard deviation (STD) $0.48; the prices are truly independent of the region, quarter and purchaser. This scenario

is consistent with *no dumping* as the distribution of sales prices is the same for any combination of region, quarter or season or purchaser. In contrast to the independent sampling scenario, the second data configuration reflects the sales record of an item with regional differences between the mean sale prices in the three regions. The sales in all three regions have the same standard deviation $0.48. However, the first region's mean price is $3.67, 10% less than the mean in the second region $4.08. While the third region's mean price is $4.49, 10% more than that of the second region. This scenario reflects *regional dumping*. Last, we consider a seasonal data pattern, where each quarter has its own mean and SD and they differ from each other. Therefore, the prices are correlated with quarter and the variances across quarters are not the same (technically inhomogeneous). We do not consider this situation as dumping because *seasonal fluctuations* normally occur.

First, the standard DOC procedure described previously is carried out on each of the three data configurations. As the use of the DOC method identified a large majority of sales as "passing" Cohen's D statistic when there were different means for the seasons or regions, we explored the following "adjusted" method for the seasonal data configuration where the quarterly specific means are price fluctuations due to a seasonal pattern, rather than dumping. In the mean adjusted version, we use the data centered at quarter means. That is, within each quarter, the quarter-specific mean is deducted from the observed price of each sale. After adjusting for quarter-specific mean, the sample averages in each quarter and in the combined data of several quarters (e.g., the control group) are always zero, leading to Cohen's D equaling zero. The quarter-specific method essentially only calculates Cohen's D for the region and purchaser categories.

One thousand random data sets were generated from each of the three data configurations and the DOC procedures was applied to each dataset; with the exception that when a known seasonal pattern was adjusted for, the adjustment was made before the DOC method was applied to the data. The percent of all sales being flagged by the 0.33 and 0.66 thresholds for the "ratio" test are reported in the Table below.

Table 1: Results of the Simulation Study

Data Configuration	Procedure	% passing 0.33 threshold	% passing 0.66 threshold
Independent region, quarter, purchaser	DOC	6.5%	0.4%
10% change in regional mean	DOC	98.7%	96.4%
Change in quarterly mean and SD	DOC	99.9%	56.8%
	Quarter-Mean Adjusted	2.8%	0.9%

The first line of Table 1 shows that when there is no difference in transaction prices during the relevant period and the price data is like a random sample from a common normal distributions true, the false positive rates are in the acceptable range, i.e., the proportion of sales exceeding the 0.66 threshold is very low (0.4%) so it does not target a producer who is not dumping. The second line, however, shows that when the price of the product under scrutiny varies from region to region due to dumping, almost all (>95%) of the sales will be "flagged" as meeting the .33 and .66 criteria. The third line shows that when there are normal seasonal patterns that create different price levels in the different quarters, the false positive rate becomes unacceptably high. When we apply the standard DOC procedure to the third data configuration, 999 out of the 1000 datasets are flagged at the 0.33 threshold while more than half are flagged using the 0.66 threshold.

The last line in Table 1 illustrates the importance of adjusting for normal price patterns. When one adjusts for the seasonal variation by deducting the quarter-specific means from each sale price, only 5% of the sales would meet the 0.33 threshold, which corresponds to the standard .05 level criterion for statistical significance. This "false positive" rate should be acceptable. Moreover, only 1% of the sales now meet the 0.66 threshold. Thus, if the Department adjusted for seasonal patterns before applying the procedure, it would be reasonable as it would not identify firms that did not dump as firms that were dumping. Currently, however, the DOC does not adjust the sales price data for seasonal variation or other known patterns, which may arise from a known shock to the system. If the current system is maintained, and no adjustments for known patterns are made then many firms that are not dumping will be classified as dumping.

3. Summary and Discussion

Although Cohen's D statistic is a measure of "effect size" rather than a formal statistical test to determine whether a difference between two means is statistically significant, its use by DOC as a "test" might be questioned. When used in conjunction with the Department's ratio test, the simulation shows that the procedure has reasonable statistical properties when the prices being investigated can be considered as a random sample from an underlying population of prices. Of course, each product will have its own price distribution but that should only have a minor effect the statistical properties of the method, unless the price distribution is very unusual.

Because economic statistics typically reflect seasonal as well as cyclical patterns, when these are not accounted for, i.e. appropriate adjustments to the data are made before applying the DOC methodology, many companies that are not dumping and are obeying the law, will be erroneously identified as dumping. It should also be noted that both the U.S. Bureau of the Census and the Bureau of Economic Analysis, which are components of the Department of Commerce, release economic statistics, such as retail sales, after incorporating seasonal adjustment, it should be possible for the DOC to call on their in-house experts to create appropriate adjustment methodology to account for legitimate economic differences affecting the prices of a product in various seasons or regions of the country.

Our simulations assumed the prices followed a normal distribution and it would be helpful if the Department examined the prices from actual investigations to see whether they were approximately normal. A reference for a graphical check and statistical tests is provided after the appendix. If the Department finds that the typical data examined in these cases differs noticeably from a normal distribution, it should repeat the simulation from distributions that are similar to the ones that occur in practice.

Appendix: Properties of Cohen's D statistic

The population effect size δ of interest to DOC is

$\delta = \frac{\mu_1 - \mu_2}{\sigma}$, where μ_1 is the mean of the first group, μ_2 is the mean of the second group and σ is the standard deviation of either group. Cohen's D statistic is defined as

$D = \frac{\bar{x}_1 - \bar{x}_2}{s_p}$, where $\bar{x}_1$ and $\bar{x}_2$ denote the group means with n_1 and n_2 observations, respectively, and s_p is a sample estimate of σ. The DOC procedure assumes that the test and control groups are independent with equal variances $V_1$and V_1 and uses $(V_1 + V_2)/2$ to estimate s_p^2. Notice this variance estimator is different from the variance estimator in the commonly used two-sample t-test, which is a weighted average of the two sample variances$\frac{V_1(n_1-1)+V_2(n_2-1)}{n_1+n_2-2}$ where n_1 and n_1 are sample sizes in the test and control groups, respectively.

Cohen suggests the (0.2, 0.5, 0.8) cut-off values to classify the effect size as small, medium, or large. This rule is to a large extent subjective and the cut-off values should vary depending on the application. Cohen explicitly noted that these cut-off values ought not to be viewed as absolute. Penfield (2006) assesses coverage of confidence intervals for Cohen's D. Furthermore, Assumptions of independence of the two groups and the homogeneity of the variances (Miller, 1986) must hold for Cohen's D statistics. When the assumption of variance homogeneity (equal variability of the observations in the two groups) is violated, Cohen's D maybe a biased measure of the effect size (Kelley, 2005).

Because Cohen's D statistic was originally proposed for the analysis of data occurring in educational and psychological testing, where the data often follows a "bell curve" or normal distribution, the study described here also assumed prices were normally distributed. There are many tests to check whether data is consistent with the assumption of normality, e.g. the Wilk-Shapiro, the Jarque-Bera test as well as a graphical procedure, see Gel, Miao and Gastwirth (2010) for a description of the Robust Quantile-Quantile (RQQ) plot. The graph also indicates how the data differs from the normal. For example, it might be skewed to the right, like income data or it might have "heavier tails" implying more observations are further than 2 standard deviations from the average than one expects in a sample from a normal curve. If the RQQ plot actual price data occurring in a particular investigation deviates noticeably from what is expected

in a sample from a normal or “bell curve”, then one can use the plot to guide the choice of a more appropriate distribution for the data. Then DOC could replicate the study described earlier by simulating data from that distribution and making the same calculations that are reported in our study.

References:

1) Cohen, J. (1969) Statistical Power Analysis for the Behavioral Sciences. NY: Academic Press.
2) Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates.
3) Gel, Y., Miao, W. and Gastwirth, J.L (2005). The importance of checking the assumptions underlying statistical analyses: graphical methods for assessing normality. *Jurimetrics Journal***, 46,** 3-29. (cited in *Mid-Continent Nail Corp. v. Unites States* (C.I.T., 2010).
4) Kelley, K. (2005). The Effects of No-nnormal Distributions on Confidence Intervals around the Standardized Mean Difference: Bootstrap and Parametric Confidence Intervals. Educational and Psychological Measurement, Vol. 65, 51-69.
5) Miller, R. G. (1986). Beyond ANOVA. John Wiley. New York.
6) Penfield, R. D. (2006). Confidence Interval Coverage for Cohen's Effect Size Statistic. Educational and Psychological Measurement, Vol. 66, p. 945-960.