LAW OFFICE OF
JEFFREY M. WINTON PLLC
1725 I STREET, N.W., SUITE 300
WASHINGTON, D.C. 20006

JEFFREY M. WINTON
jwinton@jmwinton.com

202.413.2779 (DIRECT)
202.478.1996 (TELECOPY)

June 23, 2014

Docket No. 140318257-4257-01
Total Pages: 92

BY ELECTRONIC FILING

Paul Piquado,
Assistant Secretary for Import Administration,
Room 1870,
Department of Commerce,
14th Street & Constitution Ave. NW.,
Washington, DC 202300

Public Document

Re: Differential Pricing Analysis

Dear Assistant Secretary Piquado:

On May 9, 2014, the Department published a notice in the *Federal Register* soliciting comments on the "differential pricing analysis" the Department has used in recent proceedings. My comments are set forth in the enclosed submission. I should note that these comments are filed on my own behalf, in furtherance of my personal interest in the fair and mathematically-sound implementation of the relevant provisions of the U.S. antidumping statute and the WTO Antidumping Agreement, and are not intended to represent the views of any companies that I have represented (or may in the future represent) in proceedings before the Department.

Please do not hesitate to contact me if you have any questions.

Respectfully submitted,

/s/ Jeffrey M. Winton

Jeffrey M. Winton

RESPONSE TO THE DEPARTMENT'S MAY 9 REQUEST FOR COMMENTS ON ITS "DIFFERENTIAL PRICING ANALYSIS" (DOCKET NO. 140318257-4257-01)

Submission by

LAW OFFICE OF JEFFREY M. WINTON PLLC
1725 I Street, N.W. Suite 300
Washington, DC 20006
(202) 413-2779

June 23, 2014

Table of Contents

INTRODUCTION 1

SUMMARY 4

A. The Definition of "Pattern" 4

B. The Need for Rigorous Testing 4

C. Failure to Satisfy Conditions Needed for Use of Cohen's d Test 5

D. Arbitrary and Improper Thresholds 5

E. Inability of d Test to Distinguish Significant from Insignificant Price Differences 7

F. Failure to Account for Co-Linearity in the Data 7

G. Insufficient Explanation Why Average-to-Average Comparisons Cannot Be Used 8

COMMENTS 9

A. The Department Should Define the Concept of "Pattern" Independently of the Statistical Test Used to Identify whether Patterns Exist in the Data 9

B. The Department Should Ensure that Whatever Methodology It Develops Undergoes Rigorous Testing before It Is Applied 11

C. The Differential Analysis Depends on the Unsupportable Assumptions that the Data for Each Customer, Time Period, and Region Has a Sufficient Number of Data Points and Follows a "Normal" Gaussian Distribution 15

1. The Differential Analysis Does Not Require a Large Enough Sample in Each Sub-Group to Permit a Statistically-Meaningful Comparison 15

2. The Differential Analysis Improperly Assumes that the Data for Each Customer, Time Period, and Region Follows a "Normal" Gaussian Distribution 16

Page

a. Cohen's d Test Can Appropriately Be Used Only when the Two Data Sets Being Compared Are Approximately "Normal" 16

b. If the Conditions for Normality Are Not Met, the Results of Any Analysis Based on Means and Standard Deviations (such as Cohen's d Test) Are "Meaningless" and "Invalid" 18

c. There Is No A Priori Reason to Believe that the Data for Each Customer, Time Period, and Region Follows a "Normal" Gaussian Distribution 20

(1) Because the Differential Analysis Does Not Rely on Random Samples, the Central Limit Theorem Does Not Apply, and, as a Result, the "Normality" of the Data Sets Being Compared Cannot Be Assumed 20

(2) Application of the Statistical Tests for Normality in Past Cases Has Demonstrated that the Chances that U.S. Sales Data Follows a "Normal" Distribution Are Vanishingly Small 21

(3) The Assumption of Normality Implicit in the Differential Analysis Is Inconsistent with Ordinary Market Behavior, where the Buyers and Sellers Possess Information about Other Transactions, and the Prices for Individual Sales Are Not Independent or Random 22

(4) Empirical Evidence Confirms that Prices for Individual Transactions Are Not Set at Random and Are Not Independent of the Prices for Other Transactions 23

(5) The Differential Analysis's Use of Quantity-Weighted Prices Ensures that Data Points Will Not Be Independent 24

3. Because There Is No Reason to Expect U.S. Sales Data to Follow a "Normal" Distribution, the Department Must Utilize Statistical Tests that Will Not Be Distorted by Non-Normality in the Data 25

D. The Thresholds Employed by the Department's Differential Analysis Are Arbitrary and Improper 26

Page

1. The Thresholds Proposed by Professor Cohen for Identifying "Small," "Medium," and "Large" Differences between Data Sets Are Inherently Arbitrary and Cannot Properly Be Applied as Bright-Line Tests 26

2. The Thresholds Employed in the "Ratio Test" Element of the Differential Analysis Are Inherently Arbitrary 29

a. The Department Has Not Offered Any Justification for the 33- and 66-Percent Thresholds Employed in the "Ratio Test" 29

b. The 33- and 66-Percent Thresholds Used in the "Ratio Test" Do Not Bear Any Relationship to Statistical Properties 31

(1) In the Absence of Information about the Structure of the Data, No Conclusions Can Be Drawn About the Percentage that Can Be Expected to Satisfy the Cohen's d Test Element of the Differential Analysis 31

(2) Even if U.S. Sales Data Followed a Gaussian Distribution, Any Given Customer, Time Period, or Region Would Be Expected to Satisfy Cohen's d Test Simply by Chance More than 42 Percent of the Time 32

c. The Department's "Ratio Test" Fails to Correct for the Distortions Caused by Running Multiple Iterations of the Test by Customer, Region, and Time Period 34

3. The Thresholds Established by the Differential Analysis's "Ratio Test" Are Routinely Met by Data that Follows a Random Walk 36

a. Prices that Follow a Random Walk May Give an Illusion of Trends 36

b. The Differential Analysis Would Routinely Generate Spurious Positive Results when Applied to Random Walks Whose Apparent Trends Are Only Illusions 39

E. The Differential Analysis Is Unable to Distinguish between "Significant" and "Insignificant" Price Differences 40

Page

1. Where Price Variability Is Low, Cohen's d Test May Exaggerate Insignificant Price Differences....41

2. Where Price Variability Is High, Cohen's d Test May Mask Patterns of Significant Price Differences....42

F. The Differential Analysis Fails to Address Potential Distortions that May Be Caused by Co-Linearity in the Data....44

G. The Differential Analysis Fails to Explain Why Any Patterns of Price Differences Were Not, or Could Not Be, Taken into Account Using an Average-to-Average Comparison....46

CONCLUSION....52

ATTACHMENTS

1. Excerpts from Hogarty and. Kromrey Paper

2. Excerpts from NBER Working Paper 15743

3. Results of Application of Differential Analysis to 360-Day Random Walk Data

4. Results of Application of Differential Analysis to 180-Day Random Walk Data

LAW OFFICE OF
JEFFREY M. WINTON PLLC

RESPONSE TO THE DEPARTMENT'S MAY 9 REQUEST FOR COMMENTS ON ITS "DIFFERENTIAL PRICING ANALYSIS" (DOCKET NO. 140318257-4257-01)

These comments are submitted by, and on behalf of, the Law Office of Jeffrey M. Winton, PLLC, in response to the Department's May 9, 2014, notice requesting comments on the "Different Pricing Analysis" used in antidumping investigations and reviews.[1]

INTRODUCTION

The task before the Department when considering whether "targeted dumping" exists within the meaning of the antidumping statute is to ascertain whether there is a "pattern" of U.S. prices "that differ significantly among purchasers, regions, or periods of time." Under the "Differential Analysis," the Department has attempted to perform that task based solely on calculations that employ some common statistical tools, without, however, defining the question those tools are supposed to answer. In the end, therefore, the Department's analysis becomes impossibly tautological: A "pattern" is defined as any data that generates a positive result from the Differential Analysis, and the Differential Analysis is then found to be satisfactory because it always identifies the data that satisfies that

[1] *See Differential Pricing Analysis; Request for Comments,* 79 Fed. Reg. 26720 (May 9, 2014).

We would note that there are a number of fundamental legal issues concerning the Department's approach to "targeted dumping" in investigations and reviews that are independent of the statistical analysis employed by the Department. While we do not agree with many of the positions taken by the Department on those issues, we understand that years of litigation in the U.S. courts and before WTO dispute settlement bodies are likely to be required before those issues are resolved. Accordingly, we have limited our comments to the mathematical and statistical issues raised by the Department's use of the Differential Analysis described in its request for comments.

definition. We would hope that the absurdity of such a circular approach to statutory implementation would be obvious.

The development of a proper test to implement the statute's requirement must begin, in our view, with a better definition of the problem the test is intended to solve. In other words, there needs to be a clearer understanding both of the types of trends in data that should be considered to constitute "patterns" of significant price differences, and of the types of data that should not be considered to constitute such "patterns." Then, once there is an understanding of what is and is not a pattern, any proposed statistical test can be evaluated to ensure that the test does not routinely generate false positives (in which patterns are found when none exist) or false negatives (in which patterns that do exist are missed). But, unless and until such testing can be performed, there is no basis for assuming that whatever mixture of statistical tools the Department may whip up will actually accomplish the task set by the statute.

In any event, even without such testing, it is clear that the statistical tools embodied in the Differential Analysis are not appropriate. As a matter of mathematics, the Cohen's *d* test that is at the heart of the Differential Analysis can only be properly used when the data under consideration follows a "normal" Gaussian distribution. However, it is self-evident that prices for U.S. sales are neither random nor independent of one another, as the assumption of "normality" would require.

Furthermore, even if the assumption of "normality" were satisfied, the cut-offs used in the Differential Analysis are inherently arbitrary and incorrect. For example, even in ideal circumstances, when the conditions for "normality" are satisfied, the odds that any given customer, region, or time period will "pass" the *d* test just by chance, due to random

variation, are greater than 42 percent. If the tests of customers, regions, and time periods are independent of one another, the chances that any sale would be found to "pass" at least one of the three tests by chance would be more than 80 percent. And, if the assumption of "normality" is not satisfied, the odds may become even worse. In these circumstances, the fact that more than 33 percent of an exporter's sales are found to "pass" the *d* test provides no indication that the apparent price differences represent a "pattern" within the meaning of the statute. Purely random data would be expected to surpass that threshold in virtually every case.

It is clear, therefore, that the Differential Analysis does not constitute a reliable test for identifying "patterns" in the U.S. sales data. This conclusion should not be surprising: Proper statistical analysis is quite difficult. But, if errors are understandable, failing to correct them is not. Especially on an issue that can have a dramatic impact on the results of the Department's investigations and reviews, the Department has a fundamental obligation to get it right.

Given the technical complexities, this may be an area where it makes sense to outsource development of a proper test to outside academic experts who specialize in statistical analysis. We would hope that such experts would be able to advise the Department as to whether a statistical test to identify "patterns" that meet the statutory definition is even possible and, if so, to develop an approach that is conceptually rigorous and mathematically sound. And, by delegating the development of the test to recognized experts, the Department would demonstrate its good faith to reviewing courts, international panels, and the public — and incidentally make it much more difficult for self-interested advocates to challenge the results.

SUMMARY

A. The Definition of "Pattern"

As a recent text on *Best Practices in Quantitative Methods* has observed, "a statistical analysis assumes a statistical model, which must fit the setting, design of the experiment, and research questions sufficiently well." As a consequence, if the model is not correct or does not fit the research question, "the outcome of the analysis is likely to be misleading."

In developing the Differential Analysis, however, the Department has failed to define the "research question." Instead, it has defined the statistical analysis, and used the output of that analysis to define the question that the analysis is supposed to answer. Such an approach is clearly improper, and it makes it impossible to conclude that the results generated by the Differential Analysis are meaningful, and not misleading.

B. The Need for Rigorous Testing

In order to ensure that the Differential Analysis works in practice as well as in theory, the Department should subject its analysis to rigorous tests to confirm that the methodologies and thresholds it proposes to use are not overly sensitive or inappropriately insensitive. Such testing is an intrinsic part of any proper statistical analysis. But, as far as we know, however, the Department has not explained whether or how it tested the Differential Analysis for such problems.

The limited testing we have been able to perform suggests that the Differential Analysis does not work in practice. For example, in one recent case, we provided an analysis of ten randomized data sets, in which our client's actual U.S. sales prices were replaced with purely random numbers, while the other information concerning the sales (such as customer, date of sale, and destination) were left unchanged. When we ran this

data through the Differential Analysis, positive results were generated for 5 of the 10 randomized databases. In other words, there was roughly a 50 percent chance that purely random data (which, by definition, has no patterns in it) would nevertheless be found by the Differential Analysis to have a "pattern of export prices … that differ significantly among purchasers, regions, or periods of time." While there may be limits on what can be extrapolated from a single case, the fact that the Differential Analysis repeatedly found patterns in purely random data undoubtedly raises questions that merit further analysis.

C. Failure to Satisfy Conditions Needed for Use of Cohen's d Test

As a matter of statistics, the Cohen's *d* test used in the Department's Differential Analysis can only be used when certain fundamental conditions are met. Most importantly, there must be a sufficient number of data points in the two groups being compared, and the data being analyzed must constitute a "normal" (or, more precisely, "Gaussian") distribution, with random, mutually independent, and identically distributed data. If those conditions are not met, then the results generated by the *d* test are, in the words of the National Institute for Standards and Technology, "meaningless" and "invalid." Unfortunately, both theory and evidence confirm that the conditions needed to justify the application of the *d* test to U.S. sales price data do not (and, indeed, cannot) exist.

D. Arbitrary and Improper Thresholds

Professor Cohen has admitted that the cut-off he proposed for determine whether a *d* coefficient is "large" is inherently arbitrary. The relevance of the data he considered —

such as IQs, high-school grades, and heights of teenaged girls — to the identification of patterns of significant differences in market price data is not immediately clear.

The cut-offs used in the "ratio test" element of the Differential Analysis are also arbitrary and inadequate. As far as we can tell, the Department has never explained why its 33- and 66-percent cut-offs are appropriate. At the same time, mathematical analysis confirms that these cut-offs are far too low. Even when the conditions for "normality" are satisfied, the odds that any given customer, region, or time period will "pass" the *d* test just by chance, due to random variation, are greater than 42 percent. And, if the tests of customers, regions, and time periods are independent of one another, the chances that any sale would be found to "pass" at least one of the three tests by chance would be more than 80 percent.

With data that does not satisfy the "normality" assumption, the odds are even worse. . For example, when prices follow a random walk from day to day, the odds of a positive *d* test result for any quarterly time period would be around 72 percent for an analysis encompassing only two quarters (as in a non-market-economy investigation). When the period is extended to four quarters (as in a market-economy investigation), the odds of a positive *d* test result for any quarterly time period are around 55 percent — and the odds of having at least one quarterly period each year with a positive result are more than 90 percent. In these circumstances, the fact that more than 33 percent of an exporter's sales are found to "pass" the *d* test provides no indication that the apparent price differences represent a "pattern" within the meaning of the statute. Purely random data would be expected to surpass that threshold in virtually every case.

E. Inability of d Test to Distinguish Significant from Insignificant Price Differences

The *d* test values used in the Differential Analysis are calculated by dividing the difference between the means of the "test" and "base" groups by the "pooled" standard deviation of the two groups. When price variability is low, the "pooled" standard deviation will also be low, and the *d* value will be relatively high. By contrast, when price variability is high, the "pooled" standard deviation will also be high, and the *d* value will be relatively low.

As a result, the outcome of the Differential Analysis will depend in large measure on the variability of the prices in the "test" and "base" groups. When variability is low, even insignificant price differences (such as those due to rounding) could generate a positive *d* test result. By contrast, when variability is high, even clear and consistent price differences (for example, where a seller always sells at a lower price to one customer on any given day) might generate a negative *d* test result. In these circumstances, the *d* test simply does not provide a reasonable basis for determining whether patterns of significant price differences actually exist.

F. Failure to Account for Co-Linearity in the Data

Where prices may are a function of quantities or level of trade, but the sales in larger quantities or at each level of trade are not evenly distributed among customers, time periods and regions, the Differential Analysis may find an apparent correlation between prices and the groups of customers, time periods and regions — even though the prices for sales in the same quantities or at the same level of trade were consistent across all of these groups. The use of quantity-weighted prices exacerbates this problem, because it gives

excessive weight to sales in larger quantities that have lower prices precisely because they are made in larger quantities.

A finding of targeted dumping that is driven entirely by different distributions of quantities or levels of trade across groups of customers, time periods or regions would not be consistent with the statutory provisions concerning targeted dumping. Such a finding would also be inconsistent with the provisions of the statute that specifically recognize the potential impact of quantities and level of trade on prices, and direct the Department to make appropriate adjustments to avoid finding dumping simply due to the fact that sales at different quantities or at different levels of trade have different prices.

G. Insufficient Explanation Why Average-to-Average Comparisons Cannot Be Used

Even when patterns of significant price differences are found to exist, the antidumping statute permits the Department to depart from the normal average-to-average comparison methodology only if it "explains why such differences cannot be taken into account using" an average-to-average or transaction-to-transaction calculation methodology. The Department has claimed that it satisfies this requirement by demonstrating that the dumping margins calculated using the alternate "targeted dumping" methodology differ in a "meaningful" manner from the dumping margins calculated using the standard methodology. But the mere existence of different results is plainly insufficient, by itself, to satisfy the statutory requirements.

There is, in fact, no reason to believe that the price differences that give rise to a finding of "targeted dumping" are always the cause of the different results from the different comparison methodologies. Instead, it is more likely that any differences in the

results are primarily a function of the different treatment of negative dumping margins under the Department "standard" methodology (where "zeroing" is not used) and its "alternate" methodologies (where negative margins are "zeroed"). While the mechanics are somewhat different in investigations and reviews, due to the differences in the Department's comparison methodologies, the net effect is the same: Differences in dumping margins generated by the application of "zeroing" are not the same as differences in dumping margins caused by patterns of price differences by customer, region, or time period.

COMMENTS

A. The Department Should Define the Concept of "Pattern" Independently of the Statistical Test Used to Identify whether Patterns Exist in the Data

In order to comply with the statutory targeted-dumping provision, the Department's methodology must, *inter alia,* identify 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."[2] But the Department has never attempted to define what the term "pattern" actually means. Instead, the Department has promulgated a series of computer programs — first the so-called "Nails" test and then the Differential Analysis — and taken the position that any sales which generate positive results from whichever test is currently in fashion must constitute a "pattern."

As noted in a recent text on *Best Practices in Quantitative Methods*, "a statistical analysis assumes a statistical model, which must fit the setting, design of the experiment,

[2] *See* Tariff Act § 777A(d)(1)(B); 19 U.S.C. § 1677f-1(d)(1)(B) (emphasis added).

and research questions sufficiently well; otherwise, *the outcome of the analysis is likely to be misleading*."[3] In its targeted-dumping analysis, however, the Department has failed to define the "research question." Instead, it has defined the statistical analysis, and used the output of that analysis to define the question that the analysis is supposed to answer. Under the Department's tautological approach, it is impossible to assess whether the "Nails" test and Differential Analysis are actually able to identify "patterns," because they assume that a "pattern" is anything that generates a positive result under the "Nails" test or Differential Analysis.[4]

In our view, there are certain factual situations that, intuitively, should be considered "patterns." For example, a "pattern" should be found to exist when a seller targets particular customers, regions, or time periods in order to wrest market share from competitors. On the other hand, no "pattern" should be found when different prices reflect changes in input costs, or *random* movements in, *inter* alia, gross prices, adjustments, or

---

[3] *See* F. Fidler and G. Cumming, "The New Stats: Attitudes for the 21st Century," in BEST PRACTICES IN QUANTITATIVE METHODS, ed. J. Osborne (2008), at 2 (emphasis added).

[4] In its notice requesting comments on the Differential Analysis, the Department has suggested that its goal in developing the "Nails" test and Differential Analysis is to address "potentially hidden or masked dumping that can occur when the Department determines weighted-average dumping margins using the average-to-average comparison method. *See Differential Pricing Analysis; Request for Comments,* 79 Fed. Reg. at 26722. That goal may serve as a starting point for defining the concept of "patterns," but it is not sufficient, by itself, to establish that definition. In order to determine whether dumping margins have been "hidden" or "masked," it is necessary first to have a concept of what the correct dumping margins should be. In this regard, the fact that an average-to-average comparison provides a different result than a transaction-to-average comparison does not mean that the transaction-to-average margins are correct and the average-to-average margins are not. It means only that different comparisons give different results. A definition of "pattern" is needed to allow the Department to choose which of these methods is more correct. It is not appropriate to allow the desired outcome (such as a preference for transaction-to-average comparisons) to define when a "pattern" will be found.

exchange rates.[5] While these intuitions are not sufficient to provide an over-arching definition of "pattern" that covers all possible cases, they can, at least, provide a starting point for evaluating the statistical tests the Department proposes to employ to identify "patterns." A test that consistently fails to find a "pattern" when the evidence demonstrates that targeting exists should be considered unreliable. A test that routinely finds "patterns" in purely random data — or in data that follows a random walk — should also be considered unfit.

B. The Department Should Ensure that Whatever Methodology It Develops Undergoes Rigorous Testing before It Is Applied

As discussed more fully below, we believe that the statistical tools and numerical cut-offs used in the Differential Analysis suffer from serious conceptual flaws. But even if they had a sound theoretical basis, that alone would not mean that they were suitable for use. Instead, in order to ensure that the Differential Analysis works in practice as well as in theory, the Department should subject its analysis to rigorous tests to confirm that the methodologies and thresholds it proposes to use are not overly sensitive (generating "false positive" findings of patterns where none in fact exist) or inappropriately insensitive (generating "false negative" findings that there are no patterns where they do in fact does exist). Such testing for "Type I" and "Type II" errors is an intrinsic part of any proper

[5] In this regard, one dictionary defines the term "random" to mean "lacking a definite plan, purpose, or *pattern*." *See* MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY (11th ed. 2003) at 1029 (emphasis added).

statistical analysis.[6] As far as we know, however, the Department has not explained whether or how it tested the Differential Analysis for such problems.

As mentioned, the recent *Best Practices in Quantitative Methods* text has noted that "a statistical analysis assumes a statistical model, which must fit the setting, design of the experiment, and research questions sufficiently well; otherwise, *the outcome of the analysis is likely to be misleading*." It is necessary, therefore, "to ask how robust the statistical technique is to departures from strict compliance with the assumptions of the statistical model.[7] For these purposes, Monte Carlo simulations — in which the proposed

---

[6] *See, e.g.,* J. Cohen, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988) at 4. In this regard, Professor Cohen offered the following observations regarding the importance of Type I and Type II errors in behavioral science,

> The view offered here is that more often than not, the behavioral scientist will decide that Type I errors, which result in false positive claims, are more serious and therefore to be more stringently guarded against than Type II errors, which result in false negative claims. The notion that failure to find is less serious than finding something that is not there accords with the conventional scientific view.
>
> It is proposed here as a convention that, when the investigator has no other basis for setting the desired power value, the value 0.80 be used. This means that **b** {the Type II error rate} is set at 0.20. This arbitrary but reasonable value is offered for several reasons. The chief among them takes into consideration the implicit convention for **a** {the Type I error rate} of 0.5. The **b** of 0.20 is chosen with the idea that the general relative seriousness of these two kinds of errors is of the order of .02/.05, *i.e.,* that Type I errors are of the order of four times as serious as Type II errors. This 0.8 desired power convention is offered with the hope that it will be ignored whenever an investigator can find a basis in his substantive concerns in his specific research investigation to choose a value *ad hoc.*

Id. 56 (citation omitted).

[7] *See* F. Fidler and G. Cumming, "The New Stats: Attitudes for the 21st Century," in BEST PRACTICES IN QUANTITATIVE METHODS, ed. J. Osborne (2008), at 2 (emphasis added).

statistical test is applied to numerous sets of random data — can be applied "to investigate how widely it is reasonable to apply a statistical technique."[8]

While we do not have access to nearly as many data sets as the Department does, our own experience suggests that the Differential Analysis cannot survive such testing. In the current investigation of *Oil Country Tubular Goods ("OCTG") from Vietnam*, we provided an analysis of ten randomized data set, in which our client's actual U.S. sales prices were replaced with purely random numbers, while the other information concerning the sales (such as customer, date of sale, and destination) were left unchanged. We then ran the ten randomized data through the computer program the Department uses to implement the Differential Analysis. For 5 of the 10 randomized databases, the Differential Analysis generated positive results. In other words, we found that there was roughly a 50 percent chance that purely random data would nevertheless be found by the Differential Analysis to have a "pattern of export prices … that differ significantly among purchasers, regions, or periods of time."[9]

Of course, there are limits on what can be extrapolated from a single case. It is possible that there were particular aspects of our client's sales data that confounded the Differential Analysis. It is also possible that the six-month investigation period employed in that case (under the Department's practice for investigations of non-market-economy

---

[8] *Id.*

[9] Because this analysis was based on our client's actual sales data, it includes business proprietary information that cannot be included in a public submission. Further details concerning the analysis we employed in the *OCTG* case are available in our June 6, 2014, case brief, including a non-proprietary declaration by the computer consultant who generated the random data and ran it through the Department's computer program implementing the Differential Analysis.

countries) exacerbated certain tendencies in the Differential Analysis's examination of price differences by time period. Nevertheless, the fact that the Differential Analysis repeatedly found patterns in purely random data undoubtedly raises questions that merit further analysis.

In order to assess the reliability of the Differential Analysis as a general matter, the Department should run tests that are capable of evaluating the Differential Analysis's susceptibility to false positives (Type I errors) and false negatives (Type II errors). For example, the Department should run its Differential Analysis on a wide variety of data sets containing random prices with different structures (in terms of customers, date of sale, and destination), in order to gauge how often it will find "patterns" in data that has no patterns in it. In addition, the Department should create additional data sets in which patterns that are known to meet the statutory criteria are embedded, and then run that data through the Differential Analysis to gauge whether and how often it fails to find such patterns in the data. Such testing will allow the Department to judge whether the Differential Analysis has an acceptable rate of false positives and false negatives. In the absence of such testing, there is no basis for the Department to conclude that the Differential Analysis actually works as intended.

C. The Differential Analysis Depends on the Unsupportable Assumptions that the Data for Each Customer, Time Period, and Region Has a Sufficient Number of Data Points and Follows a "Normal" Gaussian Distribution

1. The Differential Analysis Does Not Require a Large Enough Sample in Each Sub-Group to Permit a Statistically-Meaningful Comparison

It is widely recognized that a statistical analysis has little power if the number of data points being analyzed is not sufficiently large. It is not surprising, then, that Professor Cohen has indicated that the *d* test provides valid results only when there are two samples and the "sample sizes increase *above 20 to 30 cases*." And, even when the samples meet that requirement, the results may (according to Professor Cohen) also be unreliable if there are "substantially unequal sample sizes (whether small or large)."[10]

The Department has not, however, restricted the application of the Differential Analysis to situations in which the groups being compared have substantially equal sample sizes in the range of 20 to 30 observations. Instead, the Differential Analysis "calculates the Cohen's *d* coefficient with respect to comparable merchandise if the test and comparison groups of data each have at least *two observations*, and if the sales quantity for the comparison group accounts for at least *five percent* of the total sales quantity of the comparable merchandise."[11]

A comparison of samples encompassing just two sales obviously does not conform to Professor Cohen's assessment that the *d* test named after him yields valid results only as

---

[10] *Id.* at 19-20.

[11] *See Differential Pricing Analysis; Request for Comments,* 79 Fed. Reg. 26720, 26722.

"sample sizes increase *above 20 to 30 cases*." By the same token, a comparison involving one sample that has only five percent of the data of the other sample does not meet Professor Cohen's requirement that the sample sizes be "substantially" equal. In short, the Department's implementation of the Differential Analysis allows comparisons of samples that are not large enough or sufficiently well-matched to permit a proper statistical analysis.

2. *The Differential Analysis Improperly Assumes that the Data for Each Customer, Time Period, and Region Follows a "Normal" Gaussian Distribution*

a. *Cohen's d Test Can Appropriately Be Used Only when the Two Data Sets Being Compared Are Approximately "Normal"*

Cohen's *d* test is an example of a "parametric" test — which can appropriately be applied only when the data being tested conforms to a defined distribution.[12] As a practical matter, this means that the underlying data must approximately follow a "normal" (or, more precisely, "Gaussian") distribution.[13] In other words, tests based on means and

---

[12] *See, e.g.,* SAS Institute Inc., *SAS/STAT 9.2 User's Guide* (2008), Chapter 16, "Introduction to Nonparametric Analysis," at 291 ("In statistical inference, or hypothesis testing, the traditional tests are called *parametric* tests because they depend on the specification of a probability distribution (such as the normal) except for a set of free parameters. Parametric tests are said to depend on distributional assumptions. *Nonparametric* tests, on the other hand, do not require any strict distributional assumptions.").

[13] *See, e.g.,* R. Grissom and J. Kim, Effect Sizes for Research: Univariate and Multivariate Applications (2d ed. 2012) at 66 ("When the distribution of scores of a comparison population is not normal, the usual interpretation of a d_G or d in terms of estimating the percentile standing of the average-scoring members of another group with respect to the supposed normal distribution of the comparison group's scores would be invalid. Also, because standard deviations can be very sensitive to a distribution's shape, … nonnormality can greatly influence the value of a standardized-mean-difference effect size and its estimate. The problem of normality has led many to prefer measures of effect size that do not require normality…."). *See also id.*, at 62, 68-69

(footnote continued on following page)

standard deviations, such as the Cohen's *d* effect-size test used in the Differential Analysis, are appropriate only when the data follows the familiar bell-shaped curve.[14]

Professor Cohen, who proposed the "*d* test" that is central to the Differential Analysis, has also confirmed the importance of such assumptions to the use of the test that bears his name. In the chapter of his textbook describing the *"d* test," Professor Cohen explains that *d* is simply an effect size index for "*t* tests."[15] A *t* test, in turn, is a "power analysis in the case where two samples, each of *n* cases, have been *randomly and independently drawn from normal populations*, and the investigator wishes to test the null hypothesis that their respective population means are equal…."[16] In other words, Cohen's *d* is an attempted mathematical measure of the power of a statistical analysis to identify statistically meaningful differences between the average values of two random samples that follow a "normal" Gaussian distribution.[17] It is not intended as a measure of patterns of significant

---

(footnote continued from previous page)
(describing Cohen's *d* as one method of calculating the standardized difference between means that assumes normality).

[14] "The familiar bell-shaped curve which has been mathematically described by the mathematician Gauss … is called the Gaussian or normal distribution." *See, e.g.,* J.K. Taylor and C. Cihon, Statistical Techniques for Data Analysis (2d ed. 2004) at 27.

[15] *See* J. Cohen, "Statistical Power Analysis for the Behavioral Sciences" (2d ed. 1988) at 20.

[16] *Id.* at 19 (emphasis added).

[17] In his textbook, Professor Cohen explained the impact of departures from these assumptions as follows:

> In the formal development of the *t* distribution for the difference between two independent means, the assumption is made that the populations sampled are normally distributed and that they are of homogeneous (*i.e.,* equal) variance. *Moderate* departures from these

(footnote continued on following page)

differences in an entire population whose data is not random, has not been sampled, and does not follow a "normal" distribution.

b. If the Conditions for Normality Are Not Met, the Results of Any Analysis Based on Means and Standard Deviations (such as Cohen's d Test) Are "Meaningless" and "Invalid"

If the underlying data distribution is not approximately Gaussian, the output of any parametric test is inherently unreliable. As a handbook published by the National Institute of Standards and Technology ("NIST") explains,

> There are four assumptions that typically underlie all measurement processes; namely, that the data from the process at hand "behave like":
>
> 1. random drawings;
> 2. from a fixed distribution;
> 3. with the distribution having fixed location; and
> 4. with the distribution having fixed variation….[18]

(footnote continued from previous page)

> assumptions, however, have generally negligible effects on the validity of both Type I and Type II error calculations. This is particularly true for nondirectional tests and as sample sizes increase *above 20 to 30 cases*. The only noteworthy exception to the above is under the condition of *substantially unequal variances together with substantially unequal sample sizes (whether small or large)*.

Id. at 19-20.

In short, while Professor Cohen has asserted that his *d* test might survive some "moderate" departures from the assumption of normality, he agreed that the *d* test is not robust when the sample size is small (under 20 to 30 for any test), or when the samples being compared are substantially unequal in size or have substantially unequal variances.

[18] *See* NIST/SEMATECH, *e-Handbook of Statistical Methods,* Section 1.2.1 (available at « www.itl.nist.gov/div898/handbook/eda/section2/eda21.htm »).

> The randomness assumption is the most critical but the least tested.
>
> If the randomness assumption does not hold, then
>
> 1. All of the usual statistical tests are *invalid*.
> 2. The calculated uncertainties for commonly used statistics become *meaningless*.
> 3. The calculated minimal sample size required for a pre-specified tolerance becomes *meaningless*.
> 4. The simple model: y = constant + error becomes invalid.
> 5. The parameter estimates become *suspect and non-supportable*.[19]

As the NIST handbook concludes, "Because the validity of the final scientific/engineering conclusions is inextricably linked to the validity of the underlying univariate assumptions, it naturally follows that there is a real necessity that each and every one of the above four assumptions be routinely tested."[20] If the assumptions are not tested and shown to be accurate, the results of the analysis are, in the words of the NIST handbook, "invalid" and "meaningless."

Simulation-testing of *Cohen's d* and other parametric "effects size" tests have confirmed that the results are unreliable when applied to data that does not follow a "normal" Gaussian distribution. For example, a 2001 study using a Monte Carlo simulation concluded that "the most frequently used effect size estimates (Cohen's d and Hedges' g) evidence non-trivial sensitivity to violations of normality and homogeneity of

---

[19] *Id.,* Section 1.2.5.1 (available at « www.itl.nist.gov/div898/handbook/eda/section2/eda251.htm »).

[20] *Id.,* Section 1.2.5.1 (emphasis added) (available at « www.itl.nist.gov/div898/handbook/eda/section2/eda23.htm »).

variance…."[21] In other words, the analysis of random data trials using Cohen's *d* demonstrated distortions when the data under consideration did not follow a "normal" Gaussian distribution. Consequently, before Cohen's *d* test can properly be used, it is essential first to confirm that the assumption of "normality" is appropriate.

c. There Is No A Priori Reason to Believe that the Data for Each Customer, Time Period, and Region Follows a "Normal" Gaussian Distribution

(1) Because the Differential Analysis Does Not Rely on Random Samples, the Central Limit Theorem Does Not Apply, and, as a Result, the "Normality" of the Data Sets Being Compared Cannot Be Assumed

In many academic or scientific contexts in which random *samples* are being analyzed or compared, "normality" may be assumed as a result of the application of the "Central Limit Theorem.[22] However, the Differential Analysis does not rely on random samples from the underlying data. Instead, the Department simply takes the entire data set encompassing all U.S. sales, and runs that data through its analysis.

But, when raw, unsampled data is being analyzed, the Central Limit Theorem does not apply, and the existence of a Gaussian distribution cannot be assumed. A "normal"

[21] *See* K. Hogarty, J. Kromrey, "We've Been Reporting Some Effect Sizes: Can You Guess What they Mean?," paper presented at April 2001 Meeting of American Educational Research Association, at 9-10. (A copy of the relevant pages from this paper is provided in Attachment 1.)

[22] As a matter of mathematics, it is possible to show that with a sufficiently large sample, random data that is mutually independent and identically distributed will tend to approximate a normal Gaussian distribution. *See, e.g.,* O. Ibe, *Fundamentals of Applied Probability and Random Processes* (2005) at 227.

distribution" is a "rather special" category.[23] And, as discussed more fully below, there is no basis for assuming that data sets consisting of prices by customer, time period, or region fall within it. Instead, the extent to which the data conforms to the Gaussian distribution is an empirical matter.

(2) Application of the Statistical Tests for Normality in Past Cases Has Demonstrated that the Chances that U.S. Sales Data Follows a "Normal" Distribution Are Vanishingly Small

There are a number of statistical tests that can be run to determine whether a particular data set conforms to a "normal" Gaussian distribution. For example,

- Skewness tests measures the extent to which the actual data departs from the symmetry of a normal distribution.
- Kurtosis tests measure whether the actual data set is steeper or flatter than a normal distribution would be.
- The Shapiro-Wilk test evaluates the null hypothesis that a sample $x_1, ..., x_n$ came from a normally distributed population.
- The Kolmogorov–Smirnov test is a goodness of fit test that compares a sample of data to a presumed distribution (such as the "normal" Gaussian distribution).

[23] As one textbook warns,

> Normal distributions play a large role in statistics, but they are rather special and not at all "normal" in the sense of being average or natural.

D. Starnes, D. Yates, and D. Moore, *Statistics through Applications* (2005) at 116.

- The Cramér–von Mises criterion can be used to evaluate the goodness of fit of a cumulative distribution function (such as the "normal" Gaussian distribution) compared to a given empirical distribution (the actual data).
- The Anderson–Darling test can be used to evaluate whether a given sample of data (*e.g.,* the actual price data) is drawn from a given probability distribution (*e.g.,* the "normal" Gaussian" distribution).

These tests can be run automatically in SAS using the procedure PROC_UNIVARIATE. When we have run them in the past for our clients' sales data, they have confirmed that the chances that the sales data follows a "normal" Gaussian distribution are vanishingly small.[24] It would be surprising if the application of these tests to the U.S. sales data for other companies gave a different result. At a minimum, the Department should run them before employing an analysis that requires that the data follow a "normal" distribution.

(3) The Assumption of Normality Implicit in the Differential Analysis Is Inconsistent with Ordinary Market Behavior, where the Buyers and Sellers Possess Information about Other Transactions, and the Prices for Individual Sales Are Not Independent or Random

Standard economic theory assumes that buyers and sellers have perfect information, and thus are aware of the prices for all sales of the same product in the market. In reality,

[24] We have submitted the results of the SAS "univariate" tests in our case briefs (1) in a recent review of the antidumping order on *Copper Tubing from Mexico,* and (2) in the investigation of *OCTG from Vietnam.* In both cases, the Shapiro-Wilk test indicated that the probability that the test statistic W could have come from a "normal" Gaussian distribution was less than 0.0001; the Kolmogorov–Smirnov test indicates that the probability that the test statistic D could have come from a "normal" distribution was less than 0.01; the Cramér–von Mises criterion indicated that the probability that the test statistic W-Sq could have come from a "normal" Gaussian distribution was less than 0.005., and the Anderson–Darling test indicated that the probability that the test statistic A-Sq could have come from a Normal distribution was less than 0.005.

such perfect information cannot exist. Nevertheless, it is clear that a buyer has access to information regarding its own prior purchases from all of its suppliers, a seller has access to information regarding its own prior sales to all of its customers, and there are other sources of information (such as industry publications and indices) that allow both buyers and sellers additional insight.

Given the availability of this information to buyers and sellers, it is not plausible to assume that prices in the market are set entirely by random processes. While there may be some random *element* to price movements, the prices to a given customer on a given date will also undoubtedly be affected, at least to some extent, by prices to the same customer on a previous date. Furthermore, while the prices charged to different customers on the same date may not be identical, those prices will be affected by the same market forces and the same market information, and therefore will not be wholly independent of one another. Given these correlations, it is not reasonable to assume that the requirements will be met for a "normal" Gaussian distribution, in which each data point is random and independent of the other data points.

(4) Empirical Evidence Confirms that Prices for Individual Transactions Are Not Set at Random and Are Not Independent of the Prices for Other Transactions

The realities of market competition suggest, therefore, that the prices for transactions on any given date will be a function of other transactions at the same time or on the previous day. This conclusion is reinforced by economic studies suggesting that price movements for commodity products are indistinguishable from a "random walk." For example, a recent paper from the National Bureau of Economic Research concluded that "across commodity indices we cannot generate forecasts that are, on average, structurally

more accurate and robust than those based on a random walk or autoregressive specifications."[25]

In a random walk, the prices for any given day depend on the prices for the previous day plus or minus a random amount. And, by definition, data that follows a random walk does not follow a "normal" Gaussian distribution — because it violates the requirement that each data point be independent of the others.

(5) The Differential Analysis's Use of Quantity-Weighted Prices Ensures that Data Points Will Not Be Independent

As discussed above, market realities and empirical evidence both confirm that prices for individual sales transactions are unlikely to follow a "normal" Gaussian distribution. But, even if there were some reason to expect the prices for individual sales to be completely independent, the calculations embodied in the Differential Analysis would still violate "normality." In its analysis, the Department does not treat each sale as a single data point to be given equal weight. Instead, when calculating the standard deviation for each group, the Department uses a quantity-weighted analysis. And, the quantity-weighting, by definition, would transform even completely uncorrelated sales-prices into a highly correlated data set.

In effect, the Department's quantity-weighted calculation of standard deviation treats each quantity unit (for instance, each ton sold) as if it were a discrete data-point. A sale of

[25] *See* J. Groen and P. Pesenti, "Commodity Prices, Commodity Currencies, and Global Economic Developments," NBER Working Paper 15743 (Feb. 2010) at 3 ("across commodity indices we cannot generate forecasts that are, on average, structurally more accurate and robust than those based on a random walk or autoregressive specifications.").

10 tons at a price of $1,000 is given the same weight as 10 separate data-points each at a price of $1,000. A second sale of 20 tons at a price of $1,200 is given the same weight as 20 separate data-points each at a price of $1,200. Each individual sale, therefore, gives rise to a number of data points (equal to the sale quantity) at the same price and with the same characteristics (in terms of price, region, and time period).

Such an analysis violates the fundamental requirement of a "normal" Gaussian distribution that each data-point be independent. Because all of the units sold in a single transaction are necessarily sold to the same customer at the same destination at the same time and at the same price, treating each quantity-unit as a separate data point necessarily violates the independence requirement and distorts the statistical results. Thus, even if the prices for individual sales followed a "normal" Gaussian distribution, the quantity-weighted calculations employed by the Differential Analysis would transform the data into a non-normal distribution. In the words of the NIST, the results of the Differential Analysis are, therefore, "invalid" and "meaningless."

3. Because There Is No Reason to Expect U.S. Sales Data to Follow a "Normal" Distribution, the Department Must Utilize Statistical Tests that Will Not Be Distorted by Non-Normality in the Data

There are a number of "non-parametric" statistical tests that can be employed even when the data does not conform to a "normal" Gaussian distribution.[26] Further

[26] *See, e.g.,* SAS Institute Inc., *SAS/STAT 9.2 User's Guide* (2008), Chapter 16, "Introduction to Nonparametric Analysis," at 291 ("In statistical inference, or hypothesis testing, the traditional tests are called *parametric* tests because they depend on the specification of a probability distribution (such as the normal) except for a set of free parameters. Parametric tests are said to depend on distributional assumptions. *Nonparametric* tests, on the other hand, do not require any strict distributional assumptions.").

investigation is required to determine whether those tests can properly be used to determine whether a pattern of significant price differences within the meaning of the antidumping statute exists within the U.S. sales data. However, any difficulties that might arise from the use of non-parametric tests would not justify the continued use of a parametric test (like an analysis based on Cohen's *d*) when the conditions for using a parametric test are not satisfied.

D. The Thresholds Employed by the Department's Differential Analysis Are Arbitrary and Improper

1. The Thresholds Proposed by Professor Cohen for Identifying "Small," "Medium," and "Large" Differences between Data Sets Are Inherently Arbitrary and Cannot Properly Be Applied as Bright-Line Tests

Professor Cohen , who proposed the numerical cut-offs used to determine whether a Cohen's *d* coefficient was "small, "medium," or "large," has cautioned that the cut-offs he proposed are necessarily arbitrary:

> For each statistical test's ES {Effect Size} index, the author proposes, *as a convention,* ES values to serve as "operational definitions of the qualitative adjectives "small," "medium," and "large." This is an operation fraught with many dangers. The definitions are *arbitrary*, such qualitative concepts as "large" are sometimes understood as absolute, sometimes as relative; and thus they run the risk of being misunderstood.[27]

As a result, Professor Cohen accompanied his proposal with "invitations not to employ them" when better alternatives are available:

> {T}hese proposed conventions were set forth throughout with much diffidence, qualifications, and invitations not to employ them if

[27] *See, e.g.,* J. Cohen, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988) at 12.

> possible. The values chosen had no more reliable a basis than my own intuition. They were offered as conventions because they were needed in a research climate characterized by a neglect of attention to issues of magnitude…. But there are difficulties and much room for misunderstanding.[28]

And, finally, Professor Cohen noted that "{t}he terms 'small,' 'medium,' and 'large' are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation."[29]

Not surprisingly, the usefulness of Professor Cohen's classifications has been the subject of some controversy. As explained in a recent *Guide to Effect Sizes* text,

> {T}he interpretation of results using Cohen's criteria remains a controversial practice. Noted scholars such as Gene Glass, one of the developers of meta-analysis, have vigorously argued against classifying effects into "t-shirt sizes" of small, medium, and large:
>
> > There is no wisdom whatsoever in attempting to associate regions of the effect size metric with descriptive adjectives such as 'small,' 'moderate,' 'large,' and the like. Dissociated from a context of decision and comparative value, there is little inherent value to an effect size of 3.5 or 2….
>
> Reliance on arbitrary benchmarks such as Cohen's hinders the researcher from thinking about what the results really mean. Thompson takes the view that Cohen's cut-offs are "not generally useful" and notes the risk that scholars may interpret these numbers with the same mindless rigidity that has been applied to the $p = .05$ level in statistical significance testing. Shaver agrees: "Substituting

[28] *Id.* at 532.

[29] .*Id.* at 25.

> sanctified effect size conventions for the sanctified .05 level of statistical significance is not progress."[30]

In light of these concerns, the *Guide to Effect Sizes* text classifies Cohen's benchmarks as a "last resort," to be employed only when the "interpret{ing} the practical significance of … research results by grounding them in a meaningful context … is problematic."[31] Similarly, the *Best Practices in Quantitative Methods* text describes the "unfortunate practice" of using Cohen's cut-offs as fixed benchmarks that might be applied with rigidity as being "stupid in another metric."[32]

These criticisms do not mean that Professor Cohen's cut-offs and classifications are never useful. But they do mean that, when his classifications are employed, they must be used with care.

Professor Cohen has explained that he based his cut-off for "large" effect sizes on the consideration of such items as "the mean IQ difference estimated between holders of the Ph.D. degree and typical college freshmen, or between college graduates and persons with only a 50-50 chance of passing an academic high school curriculum," as well as "the mean difference in height between 13- and 18-year-old girls."[33] Before accepting his

---

[30] P. Ellis, THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS (2010) at 41-42 (citations omitted).

[31] *Id.* at 42.

[32] *See* B. Thompson, "Computing and Interpreting Effect Sizes, Confidence Intervals, and Confidence Intervals for Effect Sizes," in BEST PRACTICES IN QUANTITATIVE METHODS, ed. J. Osborne (2008), at 258 ("As noted elsewhere, 'if people interpreted effect sizes {using fixed benchmarks} with the same rigidity that $\alpha = .05$ has been used in statistical testing, we would merely be being stupid in another metric.") (editorial insertion in original)).

[33] *See, e.g.,* J. Cohen, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988) at 27.

classification, therefore, it is necessary to consider the context. Unless there is some reason to believe that market price data will follow the same distribution as IQ, high-school grades, and childhood growth curves, there is no basis for applying a cut-off that Professor Cohen derived from such phenomenon to the entirely distinct task of identifying patterns of significant price differences.

As a separate matter, it is also necessary to consider the likelihood that a difference that Professor Cohen described as "large" might happen just by chance. Even if the Cohen correctly conclude that the difference in average heights of 13- and 18-year old girls is large, there still may be a significant number of 13-year-old girls who are taller than a significant number of 18-year-old girls. Such overlaps may not be important when evaluating efficacy of a new medical treatment, or when developing correlations between child-care interventions and health. But they are critical to the assessment whether average prices to a particular customer, region, or time period that differ from the overall average should be considered something unusual.

2. The Thresholds Employed in the "Ratio Test" Element of the Differential Analysis Are Inherently Arbitrary

a. The Department Has Not Offered Any Justification for the 33- and 66-Percent Thresholds Employed in the "Ratio Test"

The second stage of the Differential Analysis employs a "ratio test" to determine which alternate dumping margin calculation methodology will be considered. If the sales that "pass" the Cohen's *d* test constitute more than 33 percent but less than 66 percent of total sales, the Department will consider an alternate comparison methodology that uses a transaction-to-average methodology (with zeroing of negative dumping margins) only for

the sales that have been found to "pass" the Cohen's *d* test. If the sales that "pass" the Cohen's *d* test constitute 66 percent or more of total sales, the Department will consider an alternate comparison methodology that uses a transaction-to-average methodology (with zeroing of negative dumping margins) for all U.S. sales.[34]

As far as we can tell, the Department has never explained why the thresholds used in the "ratio test" should be 33 and 66 percent, and not some other numbers (such as 40 and 80 percent, 50 and 90 percent, or any two other numbers between 0 and 100). Nor, for that matter, has the Department explained why a ratio between 33 and 66 percent calls for consideration of the transaction-to-average methodology only for the sales that "pass" the Cohen's *d* test, while a ratio of 66 percent or more calls for the application of the transaction-to-average methodology for all sales.

Of course, a numerical threshold that is adopted without any justification is inherently arbitrary. If the Department is to apply numerical cut-offs in its ratio tests, it must explain why it has selected the thresholds it intends to use, and it must also explain how those thresholds relate to the statutory criteria. In the absence of such explanation, the Differential Analysis cannot be upheld.[35]

---

[34] *See Differential Pricing Analysis; Request for Comments,* 79 Fed. Reg. at 26722-23.

[35] We recognize that the decision by the Court of International Trade in the *Mid Continent Nail* case affirmed the Department's adoption of the 33- and 5-percent thresholds used in the "Nails" test, even though the Department appears to have provided only a conclusory justification for using them. *See Mid Continent Nail v. United States,* 712 F. Supp. 2d 1370; (May 4, 2010). In our view, however, the CIT's decision in that case applied an extremely cursory review that is unlikely to have been affirmed by the Court of Appeals. Furthermore, it is inconceivable to us that a dispute-settlement panel under the NAFTA or the WTO Agreements would follow a similarly lax approach.

b. The 33- and 66-Percent Thresholds Used in the "Ratio Test" Do Not Bear Any Relationship to Statistical Properties

(1) In the Absence of Information about the Structure of the Data, No Conclusions Can Be Drawn About the Percentage that Can Be Expected to Satisfy the Cohen's d Test Element of the Differential Analysis

We are not aware of any statistical properties of standard deviations and Cohen's *d* calculations that suggest that there is anything magical about 33- and 66-percent thresholds, even if one makes heroic assumptions about how the underlying data might be distributed. And, if no assumptions about the shape of the distribution are made, it is not possible to identify meaningful cut-offs for determining whether the proportion of sales "passing" the Cohen's *d* test is significant or not.

We recognize that it is possible to conclude as a matter of formal mathematics that no more than 125 percent of the data can be more than 0.8 standard deviations from the mean.[36] But that conclusion has no meaning: When each data point is mapped to a single location, no more than 100 percent of the data can be located anywhere in the distribution.

[36] This result is derived from Chebyshev's inequality, which holds, in general that, in an arbitrary distribution, no more than $1/k^2$ of the data may be located more than k standard deviations from the mean. For any value of k that is less than or equal to 1, Chebyshev's inequality does not generate meaningful results (since the value of $1/k^2$ will in such cases always be greater than or equal to 100 percent). However, this inequality does have some predictive power for identifying the number of data points that are beyond some higher number of standard deviations from the mean.

For example, Chebyshev's inequality tells us that no more than 25 percent of the data can be more than two standard deviations from the mean (since k in that case will be 2, and $1/k^2$ will be one-fourth); no more than 11 percent of the data can be more than three standard deviations from the mean (since k in that case will be 3, and $1/k^2$ will be one-ninth); no more than 6.25 percent of the data can be more than four standard deviations from the mean (since k in that case will be 4, and $1/k^2$ will be one-sixteenth); and so on. *See, e.g.,* C. Grinstead and J. Snell, Introduction to Probability (2d ed. 1997) at 306.

It does not permit an assessment whether customers, regions, or time periods with average prices that fall more than 0.8 standard deviations from the mean are in any way unusual for the actual distribution that is being examined.

(2) Even if U.S. Sales Data Followed a Gaussian Distribution, Any Given Customer, Time Period, or Region Would Be Expected to Satisfy Cohen's d Test Simply by Chance More than 42 Percent of the Time

As demonstrated above, actual sales data is unlikely to have the characteristics needed to permit application of any parametric test — since the data is not random and independent, and it does not follow a normal distribution. But even if the data were more suitable, the very nature of the Department's analysis — where each individual customer is tested against the aggregate of all other customers — would render the results unreliable.

As Professor Cohen has indicated, his *d* test can be used appropriately only when there are two independent samples of roughly equal size. It does not work if a small subset of the data is compared against the rest. As a mathematical matter, a comparison of a single data point (such as the average price to one customer) against the rest of the data (the average prices to all other customers) will find substantial differences in a large percentage of cases, even if the data follows a normal distribution and has no patterns in it.

A Cohen's *d* value of 0.8 means only that the difference between the mean of the target group and the mean of the base group is at least 0.8 standard deviations. While 0.8 standard deviations may have seemed "large" to Professor Cohen, a range that includes all data that is plus or minus 0.8 standard deviations from the mean in a "normal" Gaussian

distribution only encompasses about 57.6 percent of the data.[37] Roughly 42.4 percent of the data will fall more than 0.8 standard deviations from the mean. This means that, if we assume that average prices by customer, region, or time period follow a "normal" Gaussian distribution, and if the number of data points is large enough to permit statistical analysis, then 42.4 percent of the customers, regions, or time periods can be expected to have an average price that is at least 0.8 standard deviations from the mean of the data set.[38]

With at least 42 percent of customers, regions, or time periods "passing" the Cohen's *d* test purely by chance, it would be virtually certain that there would be a "pass" rate above the 33 percent threshold the Differential Analysis uses to decide whether or not to consider the "mixed" comparison methodology. The only question, then, would be whether the exporter would be subject only to the mixed methodology, or if it would be

---

[37] The basic "empirical rule" that applies to Gaussian distributions holds that roughly 68 percent of the data will fall within one standard deviation of the mean. As a more general case, the number of data points that fall within x standard deviations of the mean can be calculated using the "error function." In particular, the percentage of the data that will fall within x standard deviations (Φ) is given by the formula:

$$\Phi = \text{erf}\left(\frac{x}{\sqrt{2}}\right)$$

The "error function" can be accessed in Microsoft Excel using the "ERF" formula. Setting x to equal 0.8 gives a result of 57.63 — which means that 57.63 percent of the data will fall within 0.8 standard deviations of the mean and, by process of subtraction, 42.37 percent of the data will fall more than 0.8 standard deviations from the mean.

[38] If each customer is separately analyzed against the data for the rest of the population, the percentage of customer falling 0.8 standard deviations or more from the mean may actually increase: If the average price for the specific customer being analyzed is equal to the mean, the removal of that customer from the base population will not affect the mean, but will increase the standard deviation of the base population. If the average price for the specific customer being analyzed is not equal to the mean, the removal of that customer from the base population will shift the mean in the opposite direction from the average price for that customer — making it more likely that the average price for that customer will be more than 0.8 standard deviations from the mean.

unlucky enough to rise above the 66 percent threshold that would prompt the Department to consider a transaction-to-average-with-zeroing methodology for all U.S. sales.

c. The Department's "Ratio Test" Fails to Correct for the Distortions Caused by Running Multiple Iterations of the Test by Customer, Region, and Time Period

As discussed above, if price data follows a "normal" Gaussian distribution, then 57.6 percent of the customers, regions, or time periods can be expected to have an average price that is no more than 0.8 standard deviations from the mean of the data set, and the remaining 42.4 percent of the customers, regions, or time periods can be expected to have an average price that is at least 0.8 standard deviations from the mean of the data set. However, because the Differential Analysis runs the calculations three separate times — once by customer, once by region, and once again by time period — the chances of a positive result from the Cohen's *d* test are even greater.

If, for a given sale, there is a 57.6 percent chance that the Cohen's *d* test will give a negative result under one of the tests, and if the analyses by customer, region, and time period are completely independent, then the chances that there will be a negative result for that sale on all three tests would be only 19.1 percent (0.576 x 0.576 x 0.576). In other words, there would be an 80.9 percent chance that the sale would generate a positive result for a particular sale under one of the three tests.

Of course, it may not be entirely reasonable to assume that the tests by customer, region, and time period are completely independent. For example, a single low-priced sale is likely to contribute to positive *d* test results under all three analyses, whether viewed as part of an average price by customer, region, or time period. But, by the same token, the results of the separate tests for a given sale will not always be the same. It is possible, for

example, that a sale that is close to the average price will be found to generate a positive *d* test result when analyzed by time period, but not when analyzed by customer, because the sale has been averaged with a low-priced sale to a different customer during the same time period. It is equally possible that a second sale that is also close to the average price will be found to generate a positive *d* test result when analyzed by customer, but not when analyzed by time period, because the sale has been averaged with a low-priced sale in a different period to the same customer. A third sale that is close to the average price might be found to generate a positive *d* test result when analyzed by region, but not when analyzed by customer or time period, because it has been averaged with a low-priced sale in a different period to a different customer in the same region.

In these circumstances, it is almost certain that the results of the Cohen's *d* element of the Differential Analysis will be inflated by the repeated testing of the same sales by customer, region, and time period. In the worst case, if the test results are completely independent, there would be more than an 80 percent chance that any single sale would generate a positive *d* test result simply by chance. It would not be surprising, then, for a "pass" rate in excess of 66 percent to be found in a substantial number of cases in which the data follows a "normal" Gaussian distribution, and does not actually have any patterns in it.

Such a result demonstrates the unsuitability of the Differential Analysis even when applied to a data set that does conform to the assumptions needed to permit the use of Cohen's *d* test. A statistical test that routinely finds "patterns" in random data that follows a "normal" Gaussian distribution is a statistical test that is badly in need of revision.

3. The Thresholds Established by the Differential Analysis's "Ratio Test" Are Routinely Met by Data that Follows a Random Walk

a. Prices that Follow a Random Walk May Give an Illusion of Trends

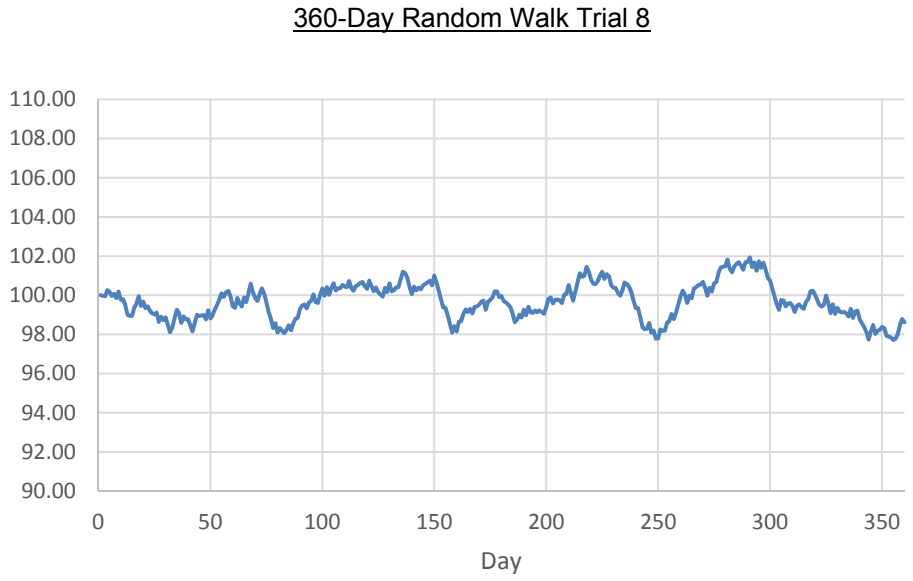
As discussed above, empirical evidence suggests that price movements for commodity products are indistinguishable from a "random walk."[39] It is possible for a random walk to contain an underlying trend, where the prices over time are expected to respond to broader market forces, even if day-to-day price changes are random. But it is also possible that completely random changes in prices from day to day, with no underlying trend, might nevertheless appear to reflect some underlying pattern. In other words, even if the random component is defined to have an expected value of zero, it is possible for the random walk to show sustained increases or sustained decreases, and to end at a point quite different from the initial starting value.

To illustrate the potential divergences in prices that are defined by a random walk, we have, in past cases, created simple models of an investigation or review period in which the prices for each day were set to equal the previous day's price *plus* a random number ranging from 0.5 to -0.5.[40] In some cases, the random-walk prices fluctuated around the initial starting point, with no apparent trend, as illustrated by the following graph.

---

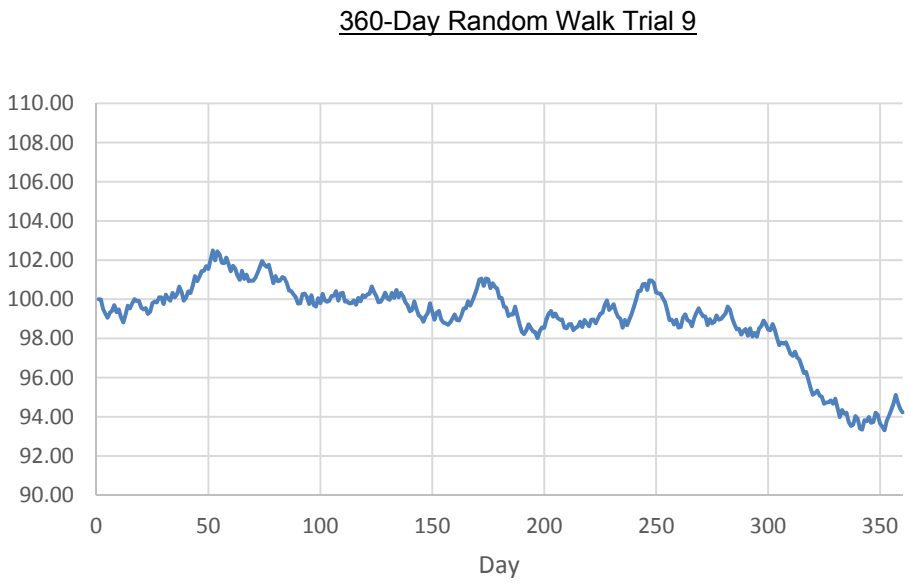
[39] *See* J. Groen and P. Pesenti, "Commodity Prices, Commodity Currencies, and Global Economic Developments," NBER Working Paper 15743 (Feb. 2010) at 3. (This paper is available at « www.nber.org/papers/w15743 ». A copy of the relevant pages is provided, for the Department's reference, in Attachment 2.)

[40] The model was created using a spreadsheet that contained ten trials at a time. We re-calculated the spreadsheet with new random numbers ten times to create 100 trials. In the review of *Copper Tubing from Mexico,* the model encompassed 100 trials of a "year" consisting of 360 days divided into four 90-day quarters. In the investigation of *OCTG from Vietnam*, the model encompassed 100 trials of 180-day "half-year" period divided into two 90-day quarters.



In other cases, the random-walk prices ended substantially higher than the initial starting point, with an apparent upward trend, as illustrated by the following graph.

And, in still other cases, the random-walk prices ended substantially lower than the initial starting point, with an apparent downward trend, as illustrated by the following graph.



Each graph, when viewed in isolation, might appear to reveal some "pattern" of price movements.[41] But, when viewed together, it is clear that there is no consistent pattern at all. And, since all three graphs were created using the same formulas, and the price movements occurred entirely at random, any "pattern" that might be seen is simply an illusion.

[41] The random-walk data depicted in these graphs was set forth in our January 23, 2014, case brief in the administrative review of the antidumping order on *Copper Tubing from Mexico* for the 2011-12 review period, which is available for public inspection on the IA Access web-site. The methods used to create the random walk data were also described in that brief, and a copy of the Microsoft Excel file used to generate the random-walk data was also submitted to the Department and is available for public inspection on the IA Access web-site. In order to avoid excessive bulk, we have not reproduced that material in this submission. However, we would be happy to provide the data and any additional documentation the Department may need upon request.

b. The Differential Analysis Would Routinely Generate Spurious Positive Results when Applied to Random Walks Whose Apparent Trends Are Only Illusions

Although the price movements depicted above are illusionary, the Differential Analysis would easily be taken in by them. The weakness of the Department's analysis can be seen by dividing our 100 random-walk full-year and half-year trials into 90-day quarters, and calculating Cohen's *d* for each.

When we ran this test using the 360-day trials, 94 of the 100 "years" tested had at least one quarter that "passed" the Cohen's *d* test. Furthermore, of the 400 quarters tested (4 quarters per year for 100 years), 220 had a Cohen's *d* that was greater than 0.8. In other words, application of Cohen's *d* test to prices that follow a random walk generated a positive result when tested for quarterly price differences a full 55 percent of the time.

The results when we ran this test using the 180-day trials were similar. Out of the 100 "half-years" tested, 72 had quarterly prices for which Cohen's *d* that was greater than 0.8.[42] In other words, application of Cohen's *d* test to prices that follow a random walk generated a positive result when tested for quarterly price differences a full 72 percent of the time.[43]

---

[42] Because the 180-day period encompassed only two quarters,

[43] For the Department's reference, a summary of the results of our testing of the random walk pricing data under the 360-day and 180-day trials is provided in Attachments 3 and 4. The day-to-day random-walk data for the 360-day trial was set forth in our January 23, 2014, case brief in the administrative review of the antidumping order on *Copper Tubing from Mexico* for the 2011-12 review period, which is available for public inspection on the IA Access web-site. The day-to-day random-walk data for the 180-day trial was set forth in our June 6, 2014, case brief in the antidumping investigation of *Oil Country Tubular Goods from Vietnam*, which is also available for public inspection on the IA Access web-site. The methods used to create the random walk data were described in those briefs, and copies of the Microsoft Excel files used to generate the random-walk data were also submitted to the Department and are available for public inspection on the IA Access web-site. In order to avoid excessive bulk, we have not reproduced that material in this

(footnote continued on following page)

As the analysis of random-walk data confirms, there is a high likelihood that the "ratio test" thresholds of 33- and 66-percent that are used in the Differential Analysis will be satisfied by price data that simply follows a random walk. But, any patterns that may appear to exist in random-walk data are entirely an illusion — the cumulative result of random daily fluctuations. Once more, a statistical test that routinely finds such fluctuations to constitute "targeted dumping" is a statistical test that is badly in need of serious revision.

E. The Differential Analysis Is Unable to Distinguish between "Significant" and "Insignificant" Price Differences

As the Department is aware, the Cohen's *d* figure used in the Differential Analysis is calculated by dividing the sales into a target group and a base group, calculating the weighted-average mean price for each of those groups, and then dividing the difference in the means by the "pooled" standard deviation for the two groups.[44] This means that, when the pooled standard deviation is high, Cohen's *d* will be lower. When the pooled standard deviation is low, Cohen's *d* will be higher. The variability of the prices in each group will, therefore, have an inverse impact on Cohen's *d*. As a result, the Differential Analysis will fail to identify clear patterns of price differences in some cases, and it will exaggerate the differences in others.

(footnote continued from previous page)
submission. However, we would be happy to provide the data and any additional documentation the Department may need upon request.

[44] The "pooled" standard deviation is calculated by squaring the standard deviation for each group, adding the figures together and dividing the sum by two, and then taking the square root of the resulting figure.

1. Where Price Variability Is Low, Cohen's d Test May Exaggerate Insignificant Price Differences

As mentioned, when price variability is low, the "pooled" standard deviation will also be low. And, because Cohen's *d* varies inversely with the standard deviation, the *d* value calculated for a data set with high variability will tend to be inflated. As a result, it is possible that insignificant price differences (such as those due to rounding) could generate a positive *d* test result.

Suppose, for example, that, in order to avoid any hint of "targeted dumping," Company X insists on selling its product to Customers A and B at the same net price, $10 per unit, on any given day. Suppose, further, that due to minor variations in exchange rates or freight costs or other adjustments, the average prices differ very very slightly — by only a fraction of a cent — perhaps $9.999 on day 1, $10.000 on days 2 and 3, and $9.999 on day 4. Finally, suppose that Customer A makes purchases every day, while Customer B only makes purchases every third day (on day 1 and day 4). Under this scenario, the average price to Customer A will be $9.9995, while the average price to Customer B will be $9.9990. The difference in the means would be only $0.0005 — less than a tenth of a cent. But the pooled standard deviation would be an even smaller 0.0003536. As a result, the Cohen's *d* for each customer group would be 1.414 (that is, $0.0005 divided by 0.0003536). Although the prices to Customers A and B were the same on any given day, and although there is almost no difference in the prices charged from one day to the next, the Department will find a "high" Cohen's *d* value:

Sample Cohen's d Calculation
Where Prices Are Effectively the Same, But Variability Is Low

	Customer A	Customer B	Combined
Day 1 Price	9.999	9.999	
Day 2 Price	10.000		
Day 3 Price	10.000		
Day 4 Price	9.999	9.999	
Mean	9.9995	9.999	
Difference in Mean			0.0005
Standard Deviation	0.0005	0.000	0.0003536
Cohen's d			1.414

In short, a positive finding from the Differential Analysis may only reflect the variability of the data within the target and base groups, and not a meaningful pattern of price differences between the groups.

2. Where Price Variability Is High, Cohen's d Test May Mask Patterns of Significant Price Differences

At the same time, because Cohen's *d* varies inversely with variability, relatively high variability will tend to result in lower *d* values. As a result, the Differential Analysis may give a negative result even when there is an obvious and clear pattern of significant price differences.

Suppose, for example, that Company Y sells its products to Customers A and B with consistently different prices — for example, the price to customer A is always $1 less than the price to customer B. Although a clear pattern of price differences does exist, the Department's calculations do not always generate a *d* value in excess of 0.8. Instead, the outcome of the Differential Analysis depends on the variability of the prices.

This effect can be seen using the following numerical examples. First, assume that the price to Customer A is $9 on the first day and $10 on the second, and that the price to

Customer B is $10 on the first day and $11 on the second. Under the Department's Differential Analysis, the average price for Customer A would be $9.5, the average price for Customer B would be $10.5, the difference in the means would be $1, the pooled standard deviation would be 0.5, and the Cohen's *d* for each customer group would be 2 (that is, $1 divided by 0.5). In such circumstances, the sales to each customer would "pass" the Differential Analysis, as shown in the following calculation.

Sample Cohen's *d* Calculation
Where Prices to One Customer Are Always Lower than Prices to the Other

Lower Variability Case

	Customer A	Customer B	Combined
Day 1 Price Day 2 Price	9.0 10.0	10.0 11.0	
Mean	9.5	10.5	
Difference in Mean			1.0
Standard Deviation	0.5	0.5	0.5
Cohen's *d*			2.0

But, now suppose that although the price differences between the customers are the same, the prices are more volatile — for example, the price to Customer A is $8 on day 1 and $11 on day 2, while the price to Customer B is $9 on day 1 and $12 on day 2. As in the previous example, the price to Customer A is always $1 below the price to Customer B on the same day. The average price to each customer is also the same as in the previous example — $9.5 for Customer A and $10.5 for Customer B. But, because of the greater variability in the prices, the "pooled" standard deviation rises to 1.5, and the calculated Cohen's *d* drops to 0.667 (that is, $1 divided by 1.5), as shown in the following table.

Sample Cohen's *d* Calculation
Where Prices to One Customer Are Always Lower than Prices to the Other

Higher Variability Case

	Customer A	Customer B	Combined
Day 1 Price Day 2 Price	8.0 11.0	9.0 12.0	
Mean	9.5	10.5	
Difference in Mean			1.0
Standard Deviation	1.5	1.5	1.5
Cohen's *d*			0.667

In short, even though the pattern of price differences is the same in the two examples, the movement of prices within each group is not, and, as a result, the sales to each customer are found to "pass" the Differential Analysis in one example and not in the other.

As these examples demonstrate, a positive Cohen's *d* test result does not necessarily mean that there actually was a pattern of significant price differences. At the same time, a negative Cohen's *d* test result does not necessarily mean that there was no pattern of significant differences. Whatever Cohen's *d* may measure, it is not a suitable test for determining whether the statutory criteria for finding "targeted dumping" have been met.

F. The Differential Analysis Fails to Address Potential Distortions that May Be Caused by Co-Linearity in the Data

There are many characteristics of a sale that may be correlated with the price and that may also be unevenly distributed among customers, time periods and regions. Prices may, for example, be a function of quantities or level of trade. The sales in larger quantities or at each level of trade may not be evenly distributed among customers, time periods and regions. When the prices are analyzed, the Differential Analysis may find an apparent correlation between prices and the groups of customers, time periods and regions — even

though the prices for sales in the same quantities or at the same level of trade were consistent across all of these groups.[45]

The use of quantity-weighted prices exacerbates this problem, because it gives excessive weight to sales in larger quantities that have lower prices precisely because they are made in larger quantities. If these larger-quantity sales are not evenly distributed among the groups being analyzed, a spurious pattern of "targeted dumping" may be found.

In this regard, it should be noted that the antidumping statute specifically recognizes the potential impact of both quantities and level of trade on prices, and directs the Department to make appropriate adjustments to avoid finding dumping simply due to the fact that sales at different quantities or at different levels of trade have different prices.[46] A finding of targeted dumping that is driven entirely by different distributions of *quantities or level of trade* across groups of customers, time periods or regions is not consistent with the statutory provisions concerning targeted dumping (which require differences in prices across groups) or with the statutory provisions concerning adjustments for differences in quantities and level of trade (which indicate that dumping should not be found just because prices vary by quantity or level of trade).

---

[45] This problem of "co-linearity" is often demonstrated through the well-known example of the relationship between crime and ice cream consumption. It is said that there is a strong apparent correlation between high levels of consumption of ice cream and high crime rates. However, this does not mean that eating ice cream turns one into a criminal. Instead, the apparent correlation reflects the fact that both ice cream consumption and crime are actually correlated with another factor — the temperature. In other words, to the extent that crime rates are higher in hot weather and ice-cream consumption is higher in hot weather, then an analysis that considers only ice cream and crime will find that they are correlated, even though it is actually the weather that is independently driving both.

[46] *See, e.g.,* Tariff Act of 1930, § 773(a)(6)(C)(i) and 773(a)(7); 19 U.S.C. § 1677b(a)(6)(C)(i) and 1677(a)(7).

G. The Differential Analysis Fails to Explain Why Any Patterns of Price Differences Were Not, or Could Not Be, Taken into Account Using an Average-to-Average Comparison

As mentioned, the statute permits the Department to depart from the normal average-to-average comparison to account for targeted dumping (in investigations) only if it "explains why such differences cannot be taken into account using" an average-to-average or transaction-to-transaction calculation methodology.[47] The Department has claimed that it satisfies this requirement by demonstrating that the dumping margins calculated using the alternate "targeted dumping" methodology differ in a "meaningful" manner from the dumping margins calculated using the standard methodology. But the mere existence of different results is plainly insufficient, by itself, to satisfy the statutory requirements.

There is, in fact, no reason to believe that the price differences that give rise to a finding of "targeted dumping" would be the cause of the different results from the different comparison methodologies. Instead, the different results are primarily a function of the different treatment of negative dumping margins under the Department "standard" methodology (where "zeroing" is not used) and its "alternate" methodologies (where negative margins are "zeroed"). While the mechanics are somewhat different in investigations and reviews, due to the differences in the Department's comparison methodologies, the net effect is the same: Differences in dumping margins generated by the application of "zeroing" are not the same as differences in dumping margins caused by patterns of price differences by customer, region, or time period.

[47] *See* Tariff Act § 777A(d)(1)(B); 19 U.S.C. § 1677f-1(d)(1)(B).

Consider, for example, a situation in which a product is sold at the same price in both the United States and the comparison market, but in which there were monthly spikes in prices (due, for example, to a global imbalance between supply and demand) that affected both markets equally. Under the Differential Analysis, such price spikes may well result in a finding that there was a pattern of significant price differences by time period. Furthermore, the price spike may also result in a finding there were patterns of significant price differences by customer and region, if certain customers located in certain regions happened to purchase a disproportionately large or small quantity of the product at times that were affected by the price spike.

But, despite appearance of pattern of price differences, the spike in prices should not have an impact on the dumping margins. If the price spike is truly global, the U.S. price and normal value should move in tandem. At the same time, if there was no dumping in the absence of the price spike, it is difficult to understand why dumping should be found to exist simply because prices in both markets moved in tandem. An average-to-average comparison comports with that intuition — since the average prices in both markets (whether calculated monthly or for the entire period) should be equally affected.

On the other hand, a transaction-to-average comparison, in which negative margins are "zeroed," will distort the analysis. In any market, unless all sales are made at the same price, there will always be some sales at prices above the average and some at prices below the average. Consequently, when an individual U.S. sale is compared to an average normal value, there will almost certainly be some sales for which the U.S. price is lower than the average, and others for which the U.S. price is higher than the average — even though the

prices in both markets moved in tandem and the average prices across the markets were the same.

The following example may help to demonstrate this problem. Suppose that a producer sells its product to two customers in both the United States and the comparison market. Each customer pays the same average price in any given month, although the prices for individual transactions vary by as much as $0.50 from the average monthly price. Suppose further that Customer A in each market purchases equal quantities in each quarter, while Customer B only purchases in the fourth quarter. Finally, suppose that the price for the product is equal to $3 per unit for the first nine months of the period (January to September), but then spikes to $4 per unit in the fourth quarter (October to December). The following tables summarize the monthly pricing information in each market, as well as the calculation of the monthly- and period-average U.S. price and normal value.

Hypothetical U.S. Price Data

	Customer A			Customer B			Average U.S. Price
	Min.	Avg.	Max.	Min.	Avg.	Max.	
January	2.5	3.0	3.5				3.0
February	2.5	3.0	3.5				3.0
March	2.5	3.0	3.5				3.0
April	2.5	3.0	3.5				3.0
May	2.5	3.0	3.5				3.0
June	2.5	3.0	3.5				3.0
July	2.5	3.0	3.5				3.0
August	2.5	3.0	3.5				3.0
September	2.5	3.0	3.5				3.0
October	3.5	4.0	4.5	3.5	4.0	4.5	4.0
November	3.5	4.0	4.5	3.5	4.0	4.5	4.0
December	3.5	4.0	4.5	3.5	4.0	4.5	4.0
Total	33.0	39.0	45.0	10.5	12.0	13.5	
Period Average		3.25			4.0		3.4

Hypothetical Comparison Market Price Data

	Customer A			Customer B			Average Normal Value
	Min.	Avg.	Max.	Min.	Avg.	Max.	
January	2.5	3.0	3.5				3.0
February	2.5	3.0	3.5				3.0
March	2.5	3.0	3.5				3.0
April	2.5	3.0	3.5				3.0
May	2.5	3.0	3.5				3.0
June	2.5	3.0	3.5				3.0
July	2.5	3.0	3.5				3.0
August	2.5	3.0	3.5				3.0
September	2.5	3.0	3.5				3.0
October	3.5	4.0	4.5	3.5	4.0	4.5	4.0
November	3.5	4.0	4.5	3.5	4.0	4.5	4.0
December	3.5	4.0	4.5	3.5	4.0	4.5	4.0
Total	33.0	39.0	45.0	10.5	12.0	13.5	
Period Average		3.25			4.0		3.4

Due to the assumed price spike in the fourth quarter, this data shows apparently significant differences in average prices by quarter (an average price of 3 in the first three quarters, and 4 in the fourth quarter), and also in average prices by customer (an average price of 3.25 to Customer A, and 4.0 to Customer B). However, these differences in average prices are the same in both markets. Furthermore, the apparent difference in average prices by customer is actually a reflection of the different timing of purchases by each customer — since the prices for contemporaneous purchases by the two customers are always the same.

Under these assumptions, an analysis using the normal average-to-average comparison methodology employed in investigations will find no dumping, because the average U.S. and comparison market prices for the period are the same. Similarly, an analysis using the monthly average-to-average comparisons employed in reviews will also find no dumping, because the monthly average prices in the two markets are the same.

However, if a transaction-to-average comparison is used with zeroing, dumping margins will be created. In an investigation, the comparisons will be as follows:

Dumping Calculation for Investigation Using
Transaction-to-Average Comparison (with Zeroing) for Entire Period

	U.S. Sales						Average Normal Value	Aggregate Dumping Margin
	Customer A			Customer B				
	Min.	Avg.	Max.	Min.	Avg.	Max.		
Jan.	2.5	3.0	3.5				3.4	1.3
Feb.	2.5	3.0	3.5				3.4	1.3
Mar.	2.5	3.0	3.5				3.4	1.3
Apr.	2.5	3.0	3.5				3.4	1.3
May	2.5	3.0	3.5				3.4	1.3
June	2.5	3.0	3.5				3.4	1.3
July	2.5	3.0	3.5				3.4	1.3
Aug.	2.5	3.0	3.5				3.4	1.3
Sep.	2.5	3.0	3.5				3.4	1.3
Oct.	3.5	4.0	4.5	3.5	4.0	4.5	3.4	
Nov.	3.5	4.0	4.5	3.5	4.0	4.5	3.4	
Dec.	3.5	4.0	4.5	3.5	4.0	4.5	3.4	
Total	33.0	39.0	45.0	10.5	12.0	13.5		11.7
Average		3.25			4.0			

An overall average dumping margin of 7.65 percent would therefore be created using the transaction-to-average investigation methodology, even though the prices in both markets were the same at any given time.[48]

Similarly, the results using the transaction-to-average methodology for reviews, in which the U.S. prices are compared to a monthly-average normal value, would be as follows:

[48] The dumping margin rate of 7.65 percent was calculated by dividing the aggregate dumping margin (11.7) by the total price for all sales (33 + 39 +45 +10.5 +12 + 13.5). The aggregate dumping margin for each month was calculated by comparing the average normal value for the period to the three prices in each month (minimum, average, and maximum), adding the positive margins together, and ignoring any negative margins. Thus, for the month of January, the aggregate dumping margin was calculated by subtracting the minimum price (2.5) and average price (3.0) from the average normal value (3.4), while ignoring the maximum price (3.5) because it was higher than the average normal value, to find an aggregate margin of 1.3 (that is, (3.4 - 2.5) + (3.4 – 3.0)).

Dumping Calculation for Review Using Transaction-to-Average
Comparison(with Zeroing) with Monthly-Average Normal Values

	U.S. Sales						Monthly Normal Value	Aggregate Dumping Margin
	Customer A			Customer B				
	Min.	Avg.	Max.	Min.	Avg.	Max.		
Jan.	2.5	3.0	3.5				3.0	0.5
Feb.	2.5	3.0	3.5				3.0	0.5
Mar.	2.5	3.0	3.5				3.0	0.5
Apr.	2.5	3.0	3.5				3.0	0.5
May	2.5	3.0	3.5				3.0	0.5
June	2.5	3.0	3.5				3.0	0.5
July	2.5	3.0	3.5				3.0	0.5
Aug.	2.5	3.0	3.5				3.0	0.5
Sep.	2.5	3.0	3.5				3.0	0.5
Oct.	3.5	4.0	4.5	3.5	4.0	4.5	4.0	0.5
Nov.	3.5	4.0	4.5	3.5	4.0	4.5	4.0	0.5
Dec.	3.5	4.0	4.5	3.5	4.0	4.5	4.0	0.5
Total	33.0	39.0	45.0	10.5	12.0	13.5		6.0
Average		3.25			4.0			

An overall average dumping margin of 3.92 percent would therefore be created using the transaction-to-average methodology for reviews.[49]

As these examples show, a finding that a transaction-to-average comparison with zeroing generates higher dumping margins than the normal average-to-average comparison does not mean that the results of the average-to-average comparison are somehow incorrect. Instead, the higher dumping margins found from the transaction-to-average comparison may simply reflect the distortions caused by zeroing. The Differential

[49] The dumping margin rate of 3.92 percent was calculated by dividing the aggregate dumping margin (6.0) by the total price for all sales (33 + 39 +45 +10.5 +12 + 13.5). The aggregate dumping margin for each month was calculated by comparing the average normal value for the period to the three prices in each month (minimum, average, and maximum), adding the positive margins together, and ignoring any negative margins. Thus, for the month of January, the aggregate dumping margin was calculated by subtracting the minimum price (2.5) from the average normal value for that month (3.0), while ignoring the average price (3.0) and the maximum price (3.5) because they were equal to or greater than the average normal value, to find an aggregate margin of 0.5 (that is, (3.0 - 2.5).

Analysis's exclusive focus on the difference in results between the two comparisons does not, therefore, satisfy the Department's obligation, under the statute, to explain why the alleged price differences cannot be taken into account using an average-to-average or transaction-to-transaction comparison.

CONCLUSION

As the foregoing discussion demonstrates, the technical issues involved in the identification of patterns of significant price differences within the meaning of the statute are highly complex. That complexity does not, however, justify the use of statistical tools when the conditions for their use have not been met, or when the results they generate are not meaningful. It is clear that the Differential Analysis described in the Department's request for comments needs serious rework. It is our hope that the comments set forth above will assist the Department in undertaking that task.

Respectfully submitted,

/s/ Jeffrey M. Winton

Jeffrey M. Winton

January 23, 2013

Attachment 1

Excerpts from Hogarty and. Kromrey Paper

We've Been Reporting Some Effect Sizes: Can You Guess What They Mean?

Kristine Y. Hogarty
Jeffrey D. Kromrey
University of South Florida

Paper presented at the annual meeting of the American Educational Research Association, April 10-14, 2001, Seattle.

We've Been Reporting Some Effect Sizes: Can You Guess What They Mean?

INTRODUCTION

Over the years, there has been a concerted effort aimed toward encouraging researchers to standardly provide some indication of effect size along with or in place of the results of statistical significance tests. Effect sizes have been viewed as consistent with null hypothesis significance testing and as an important compliment. Yet, despite urgings for the consistent reporting of effect size, these measures are seldom found in published reports, and are seemingly still far from becoming standard practice (Kirk, 1996, Thompson & Snyder, 1997, 1998). Carver (1993) contends, "statistical significance tells us nothing directly relevant to whether the results we found are large or small, and it tells us nothing with respect to whether the sampling error is large or small. This problem can be eliminated by reporting both effect size and standard errors" (p. 291). Additionally, the reporting of effect sizes assists researchers in planning future research (i.e., the determination of sample size for subsequent experimentation) as well as facilitating comparison of results across studies through the use of meta-analytic techniques.

One practical impediment to the use and reporting of effect sizes may stem from poor understanding. Choosing among the various possible effect-size estimates is not always apparent (Rosenthal, 1991), and opinions vary regarding the merits of the various possibilities (Crow, 1991; Gorsuch, 1991: McGraw, 1991; Parker, 1995; Rosenthal, 1991; Strahan, 1991). An important consideration is the extent to which indices of effect size calculated from a sample provide information about the magnitude of effect in the population from which the sample was drawn. That is, the statistical bias and sampling error associated with sample effect size indices are attributes that must be taken into account in developing accurate interpretations of observed effect sizes. Further, the valid interpretation of sample effect sizes must include a consideration of the sensitivity of effect size indices to differences in population distribution shape or differences in population variances. The research reported here is intended to elucidate the differences between both parametric and non-parametric indices of effect size and expand upon the utility of the indices under conditions of population non-normality and heterogeneous variances.

Effect Size Indices

Traditional measures of effect size (Cohen's d or Hedge's g) may be used to describe differences in means relative to an assumed common variance. Cohen's d is given by

consistent pattern of sampling error was also witnessed for the CL effect size (Figure 13), but with a slight increase in error with non-normal populations. The distributions of standard errors for the trimmed-d (Figure 14) bore a striking resemblance to the pattern witnessed for Cohen's d, with little variability across population shapes, but multiple conditions with extreme estimates. The naïve estimator of *W* (Figure 15) evidenced little change across distribution shapes except for the most extreme condition (skewness = 2, kurtosis = 6) in which greater variability was observed. For the Q632 bootstrap estimate (Figure 16), considerable variability in the standard errors was evidenced within each population shape, yet the distributions remained relatively consistent regardless of distribution shape. For Cliff's d (Figure 17), the standard error remained relatively consistent across distribution shapes as did the sampling error of its close relative effect size A (Figure 18).

Marginal standard error estimates by research design factors are presented in Tables 7 and 8. These tables present the average standard error across factors included in the design of the simulation study. For the normally distributed populations (Table 7), the effect of sample size was most evident for the trimmed-d effect size, which showed pronounced sampling error with small samples (standard error = 1.89 for samples of size 5), but a substantial reduction when larger samples were used (bias = .42 with samples of size 20). The trimmed-d effect size estimate consistently evidenced the greatest amount of sampling error across variance heterogeneity and population effect size. In contrast, the Q632 estimate was the most consistent estimator, with a standard error of approximately .04 across all conditions. Surprisingly, variance heterogeneity and population effect size were not seen to influence the sampling error associated any of the other effect size estimates.

For the most extreme non-normality examined in this study, the marginal results evidenced a similar pattern, but the sampling error was more pronounced for some of the conditions. Once again the trimmed-d evidenced the greatest amount of sampling error. An increase in sampling error was also apparent for both Cohen's d and Hedges' g across most conditions. Surprisingly, more consistency across estimates was witnessed for the Gamma index for the non-normal distributions, regardless of sample size, variance heterogeneity or population effect size. Once again, the CL, Q, Q632, Cliff's d and A were slightly influenced by sample size, but relatively unaffected by heterogeneity of variance or magnitude of effect size.

DISCUSSION

The results of this research suggest that the most frequently used effect size estimates (Cohen's d and Hedges' g) evidence non-trivial sensitivity to violations of normality and homogeneity of variance, confirming the concerns that have been raised in the literature about the appropriateness of

these indices for describing effects in such populations (Kraemer & Andrews, 1982 ;Wilcox & Muska, 1999). Surprisingly, the use of trimmed means and Winsorized variances did not provide a substantial improvement in robustness, and the resulting trimmed-d effect size evidenced extreme bias in small sample applications. Such bias was not noted in previous research on this statistic's use in meta-analytic tests of homogeneity (Hogarty & Kromrey, 1999) because the bias in the index did not adversely affect the test of differences in effect size across classes of studies (e.g., the differences in values of equally biased indices provided a valid test of population difference).

Several alternative non-parametric indices (i.e., gamma, CL and Q) appeared less sensitive than Cohen's d or Hedges' g but evidenced either substantial bias when applied to small samples or non-trivial sensitivity to population distribution shape or variance heterogeneity. The best performances in producing relatively unbiased estimates of the relevant parameters with consistent standard errors were provided by Cliff's d and the measure of stochastic superiority ($\hat{A}$) proposed by Vargha and Delaney (2000). These indices are closely related, as each is a simple function of the proportion of sample occurrences of dominance or superiority. In addition to providing unbiased estimates of the population parameters and stable standard errors across a variety of distribution shapes and levels of variance heterogeneity, these indices share an advantage of not requiring interval-level measurement of the dependent variable. Additional research on these indices (e.g., a delineation of small, medium and large effect sizes) is certainly called for. Further, the work by both Cliff (1993, 1996) and Vargha and Delaney (2000) on the sampling properties of these indices suggest readily available extensions into meta-analytic tests.

In summary, the use of effect sizes has grown in popularity in recent years (although such application remains far from universal). Renewed debates regarding the over-reliance on hypothesis testing, emphasizing the often misleading nature and inappropriate use of such tests (Nickerson, 2000), may be partially responsible for this increased interest. Because the use of effect sizes, in many instances, provides useful information to supplement traditional inferential statistics, advocacy for their use is appropriate. As the reporting and interpretation of effect sizes become more commonplace, researchers must remain mindful of the limitations of certain indices. For example, Wilcox and Muska (1999) have pointed out the important distinction among indices that reflect differences in location and those that represent more global differences in distributions. Further, Fern and Monroe (1996) have delineated the variety of research factors (e.g., designs, operational details, measurement reliability, sample characteristics) that must be considered in the appropriate interpretation of observed effect sizes.

Attachment 2

Excerpts from NBER Working Paper 15743

NBER WORKING PAPER SERIES

COMMODITY PRICES, COMMODITY CURRENCIES, AND GLOBAL ECONOMIC DEVELOPMENTS

Jan J. J. Groen
Paolo A. Pesenti

Working Paper 15743
http://www.nber.org/papers/w15743

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2010

We thank Kalok Chan, Takatoshi Ito, Warwick McKibbin, Roberto Mariano, John Romalis, Andrew Rose, two reviewers, and conference participants at the EASE-20 conference in Hong Kong for many helpful suggestions, as well as Spencer Amdur for excellent research assistance. The views expressed here are those of the authors, and do not necessarily reflect the position of the Federal Reserve Bank of New York, the Federal Reserve System, any other institution with which the authors are affiliated, or the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Jan J. J. Groen and Paolo A. Pesenti. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

walk process. The three approaches include a model in which forecasts are based only on the information embedded in observed past movements of commodity currencies, as in CRR, and two variants of a factor-augmented regression model that makes use of information from a relatively large data set, as described below. The purpose of our exercise is ultimately to provide an agnostic but reasonably systematic look at the global roots of commodity price dynamics. Rather than attempting to answer questions such as "why are commodity prices so high or so low "and "how long are they going to stay where they are", our contribution has the more modest purpose of providing an empirical assessment of the extent to which information embedded in indicators of global economic developments may help in predicting movements of commodity prices, by improving upon the naive statistical benchmarks or the CRR approach.

The main conclusions of the paper can be summarized as follows. We are able to provide some corroboration, albeit rather mild, for the CRR results. For one specific commodity index, at the shortest forecasting horizons (up to one-quarter ahead), the predictions of an exchange rate-based model are significantly better than those based on a random walk, although they do not outperform an autoregressive specification; at the one-year ahead horizon, the performance is reverted, as the CRR model significantly outperforms the autoregressive benchmark but not the random walk. When other indices are considered, the results are nuanced. We also find that a model encompassing principal components extracted from a panel of global economic explanatory variables generally performs poorly. We obtain more promising results when we replace the principal components approach with a different methodology (a partial least squares factor-augmented model), suggesting that information from a larger set of macrovariables can have some predictive power. However, across commodity indices we cannot generate forecasts that are, on average, structurally more accurate and robust than those based on a random walk or autoregressive specifications.

The paper is organized as follows. Section 2 provides a synthetic survey of the different arguments used to rationalize and predict shifts in commodity prices. Section 3 describes the methodology used in constructing our exchange rate-based and factor-augmented regression models and assessing their forecasting properties against the naive statistical benchmarks. Section 4 reports and discusses our results. Section 5 concludes.

Data Description

Variable	Countries	Source	Transform	Indices
Australian Dollar Exchange Rate	-	Bloomberg	1	vwxyz
Canadian Dollar Exchange Rate	-	Bloomberg	1	vwxyz
New Zealand Dollar Exchange Rate	-	Bloomberg	1	vwxyz
South African Rand Exchange Rate	-	Bloomberg	1	vwxyz
Chilean Peso Exchange Rate	-	Bloomberg	1	w
Baltic Dry Index (BDI)	-	Bloomberg	1	vwxyz
Industrial Production	abcdefg	OECD	1	vwxyz
Industrial Production	j	OECD	1	w
Nominal Short Term Interest Rates (3 Month)	abceg	OECD	2	vwxyz
Nominal Short Term Interest Rates (3 Month)	df	OECD	2	wxyz
Real Short Term Interest Rates (3 Month)	abceg	OECD	2	vwxyz
Real Short Term Interest Rates (3 Month)	df	OECD	2	wxyz
Long Term Interest Rates (10 Year)	abcfgk	OECD	2	vwxyz
Long Term Interest Rates (10 Year)	e	OECD	2	w
Business Confidence Indicator	g	OECD	5	vwxyz
Business Confidence Indicator	e	OECD	5	wxyz
Business Confidence Indicator	bcdeflm	OECD	5	w
Consumer Confidence Indicator	bcdf	OECD	5	wxyz
Consumer Confidence Indicator	fl	OECD	5	wxy
Consumer Confidence Indicator	a	OECD	5	w
Unemployment	aefg	OECD	2	vwxyz
Unemployment	b	OECD	2	wxy
Unemployment	h	OECD	2	wy
Retail Trade Volume	acefg	OECD	1	vwxyz
Retail Trade Volume	b	OECD	1	wxyz
Retail Trade Volume	k	OECD	1	wxy
Retail Trade Volume	d	OECD	1	w
Hourly Earnings in Manufacturing	dfg	OECD	1	vwxyz
Hourly Earnings in Manufacturing	e	OECD	4	vwxyz
Hourly Earnings in Manufacturing	a	OECD	1	vxyz
Goods Exports	abcgefghk	OECD	1	vwxyz
Goods Exports	ij	OECD	1	w
Goods Imports	abcgefghk	OECD	1	vwxyz
Goods Imports	ij	OECD	1	w

Term Slope Structure (Long term - short term rates)	abcg	OECD	5	vwxyz
Term Slope Structure (Long term - short term rates)	f	OECD	5	wxy
Term Slope Structure (Long term - short term rates)	e	OECD	5	w
Core CPI	abcdefgl	OECD	3	vwxyz
Core CPI	m	OECD	3	vz
Broad Money (M3)	agk	OECD	3	vwxyz
Broad Money (M3)	i	OECD	3	wxyz
Broad Money (M3)	el	OECD	3	wy
Broad Money (M3)	f	OECD	3	w
Narrow money (M1)	aegl	OECD	3	vwxyz
Narrow money (M1)	ik	OECD	3	wxyz
Narrow money (M1)	f	OECD	3	w
LME Copper Warehouse Stocks	-	EIA	1	vwxyz
LME Lead Warehouse Stocks	-	EIA	1	vwxyz
LME Zinc Warehouse Stocks	-	EIA	1	vwxyz
LME Aluminum Warehouse Stocks	-	EIA	1	wxy
LME Nickel Warehouse Stocks	-	EIA	1	wxy
LME Tin Warehouse Stocks	-	EIA	1	w
Crude Oil Stocks, Non-SPR (Strategic Petrol Reserve)	-	EIA	1	vwxyz
Crude Oil Stocks, Total	-	EIA	1	vwxyz
Crude Oil Stocks, SPR	-	EIA	1	wxy
Jet Fuel Stocks	-	EIA	1	vwxyz
Motor Gasoline Stocks	-	EIA	1	vwxyz
Residual Fuel Oil Stocks	-	EIA	1	vwxyz
Other Petroleum Products Stocks	-	EIA	1	vwxyz
Total Petroleum Stocks	-	EIA	1	vwxyz
United States Crude Oil Production	-	EIA	1	vwxyz
Non-OPEC Crude Oil Production	-	EIA	1	vwxyz
World Crude Oil Production	-	EIA	1	vwxyz
OPEC Crude Oil Production	-	EIA	1	vwxyz
Total World Coal Stocks	-	EIA	1	vwxyz
Distillate Fuel Oil Stocks	-	EIA	1	wxyz
Propane/Propylene Stocks	-	EIA	1	wxyz
Liquefied Petroleum Gases Stocks	-	EIA	1	wxyz
Natural Gas in Underground Storage - Working Gas	-	EIA	1	wyz
Natural Gas in Underground Storage - Total	-	EIA	1	wyz
Currency: Banknotes and Coin	f	Bank of England	1	vwxyz
Wheat Futures Price (1, 3, and 6 month)	-	Bloomberg	1	w*
Corn Futures Price (1 and 3 month)	-	Bloomberg	1	w*
Hogs Futures Price (1, 3, and 6 month)	-	Bloomberg	1	w*
WTI Futures Price (1, 3, 6, and 12 month)	-	Bloomberg	1	w*
Heating Oil Futures Price (1, 3, 6, and 12 month)	-	Bloomberg	1	w*
Brent Crude Futures Price (3 and 6 month)	-	Bloomberg	1	w*
Copper Futures Price (3 and 6 month)	-	Bloomberg	1	w*
Gold Futures Price (3, 6, and 12 month)	-	Bloomberg	1	w*
Silver Futures Price (3, 6, and 12 month)	-	Bloomberg	1	w*
Lead Forward Price (3 and 15 month)	-	Bloomberg	1	w*
Copper Forward Price (3 month)	-	Bloomberg	1	w*
Nickel Forward Price (3 month)	-	Bloomberg	1	w*
Tin Forward Price (3 month)	-	Bloomberg	1	w*
Zinc Forward Price (3 month)	-	Bloomberg	1	w*
Aluminum Forward Price (3 month)	-	Bloomberg	1	w*
Dependent Variables				
Reuters/Jefferies Commodity Price Index	-	CRB	1	v
CRB Industrial Metals Price Index	-	CRB	1	v
S&P/Goldman Sachs Commodity Price Index	-	Goldman Sachs	1	v
Dow Jones/AIG Commodity Price Index	-	Dow Jones	1	w
DJAIG Energy Commodity Price Index	-	Dow Jones	1	w
DJAIG Industrial Metals Price Index	-	Dow Jones	1	w
IMF Global Commodity Price Index	-	IMF	1	x
IMF Industrial Metals Price Index	-	IMF	1	x
S&P/Goldman Sachs Energy Commodities Price Index	-	Goldman Sachs	1	y
S&P/Goldman Sachs Industrial Metals Price Index	-	Goldman Sachs	1	z

*These are only used in the DJAIG models for Tables 11-12.

Attachment 3

Results of Application of Differential Analysis
to 360-Day Random Walk Data

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 1 to 10

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
1	97.09	1.12	1	97.10	1.11	1	99.89	0.53	0	101.91	1.71	1	1
2	101.43	1.07	1	102.46	0.48	0	102.50	0.46	0	106.84	2.00	1	1
3	103.28	0.19	0	106.69	1.76	1	103.26	0.20	0	101.20	1.38	1	1
4	101.27	0.19	0	102.76	1.11	1	101.97	0.62	0	97.85	1.91	1	1
5	104.33	1.33	1	107.68	1.07	1	107.03	0.61	0	105.70	0.35	0	1
6	101.03	1.15	1	101.39	1.32	1	96.47	0.92	1	95.06	1.56	1	1
7	103.12	0.40	0	103.67	0.26	0	102.60	1.02	1	104.43	1.16	1	1
8	99.26	0.62	0	100.01	0.47	0	99.70	0.02	0	99.77	0.13	0	0
9	100.59	1.04	1	99.89	0.58	0	99.20	0.12	0	96.38	1.74	1	1
10	102.08	0.83	1	98.87	1.50	1	100.27	0.49	0	102.54	1.16	1	1
	101.35		6	102.05		6	101.29		2	101.17		8	9

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 11 to 20

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
11	99.57	0.77	0	98.48	0.73	0	99.11	0.13	0	98.89	0.17	0	0
12	99.58	1.43	1	98.64	1.00	1	94.90	0.70	0	92.64	1.73	1	1
13	98.85	0.41	0	97.63	0.85	1	97.88	0.75	0	105.62	2.01	1	1
14	100.77	0.20	0	100.25	0.60	0	100.58	0.09	0	100.95	0.49	0	0
15	101.13	1.01	1	100.16	0.17	0	100.51	0.48	0	98.05	1.65	1	1
16	98.07	1.96	1	94.02	0.01	0	91.60	1.19	1	92.47	0.77	0	1
17	96.92	0.85	1	96.31	1.27	1	99.00	0.60	0	100.32	1.52	1	1
18	100.74	0.62	0	102.42	1.15	1	101.80	0.50	0	100.34	1.04	1	1
19	101.86	1.61	1	103.70	0.10	0	103.82	0.21	0	104.98	1.29	1	1
20	99.83	0.58	0	100.41	0.29	0	100.08	0.21	0	100.55	0.50	0	0
	99.73		5	99.20		4	98.93		1	99.48		6	7

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 21 to 30

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
21	102.55	1.01	1	101.45	1.42	1	107.66	0.87	1	109.55	1.56	1	1
22	103.30	1.19	1	101.12	1.10	1	102.44	0.29	0	101.79	0.39	0	1
23	100.79	1.97	1	104.44	0.76	0	104.45	0.76	0	104.03	0.45	0	1
24	102.50	1.44	1	103.71	0.84	1	106.31	0.47	0	108.98	1.81	1	1
25	100.36	0.38	0	101.20	1.24	1	100.26	0.28	0	98.14	1.90	1	1
26	101.61	1.60	1	104.38	0.84	1	110.23	0.76	0	113.62	1.68	1	1
27	97.28	1.18	1	98.25	0.05	0	99.26	1.14	1	98.37	0.10	0	1
28	98.83	1.42	1	97.52	0.32	0	95.52	1.35	1	96.68	0.39	0	1
29	100.61	0.49	0	100.08	0.36	0	100.26	0.07	0	100.27	0.06	0	0
30	97.46	0.85	1	97.32	0.94	1	99.10	0.20	0	101.30	1.59	1	1
	100.53		8	100.95		6	102.55		3	103.27		5	9

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 31 to 40

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
31	99.57	1.40	1	98.04	0.42	0	98.44	0.06	0	97.51	1.04	1	1
32	98.91	1.16	1	98.97	1.13	1	103.40	0.47	0	107.15	1.82	1	1
33	100.42	2.02	1	96.90	0.19	0	95.49	1.07	1	95.98	0.76	0	1
34	102.01	0.99	1	100.13	0.83	1	101.70	0.69	0	100.12	0.84	1	1
35	100.63	1.93	1	105.98	0.08	0	106.71	0.36	0	109.73	1.49	1	1
36	99.02	0.83	1	100.09	1.53	1	96.08	1.08	1	95.76	1.29	1	1
37	98.16	1.93	1	103.86	0.07	0	107.92	1.26	1	106.35	0.74	0	1
38	95.79	1.78	1	91.97	0.33	0	92.11	0.26	0	90.42	1.19	1	1
39	97.40	0.05	0	95.12	1.33	1	96.81	0.31	0	99.94	1.58	1	1
40	98.15	1.95	1	93.35	0.25	0	89.73	1.04	1	89.40	1.16	1	1
	99.01		9	98.44		4	98.84		4	99.24		8	10

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 41 to 50

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
41	98.80	1.92	1	95.15	0.64	0	95.46	0.42	0	94.82	0.87	1	1
42	97.15	1.70	1	94.60	0.07	0	93.31	0.96	1	93.72	0.67	0	1
43	98.79	1.36	1	98.04	0.83	1	95.95	0.65	0	94.68	1.55	1	1
44	99.80	0.78	0	96.80	1.38	1	97.93	0.57	0	100.34	1.17	1	1
45	97.36	1.84	1	94.03	0.25	0	92.96	0.92	1	93.36	0.67	0	1
46	100.67	0.57	0	101.73	1.52	1	99.13	0.81	1	98.60	1.28	1	1
47	100.98	1.62	1	103.39	0.51	0	103.87	0.93	1	103.03	0.19	0	1
48	101.56	1.59	1	103.36	0.73	0	106.43	0.72	0	108.30	1.60	1	1
49	99.91	0.95	1	100.15	0.54	0	100.94	0.75	0	100.94	0.75	0	1
50	102.10	0.97	1	102.13	0.96	1	103.55	0.02	0	106.34	1.91	1	1
	99.71		8	98.94		4	98.95		4	99.41		6	10

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 51 to 60

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
51	99.02	1.82	1	96.51	0.37	0	94.66	0.70	0	93.29	1.49	1	1
52	99.44	0.03	0	101.86	1.76	1	98.30	0.88	1	98.34	0.84	1	1
53	100.06	1.69	1	98.19	0.54	0	95.92	0.85	1	95.06	1.38	1	1
54	99.83	0.23	0	100.11	0.37	0	101.80	1.23	1	95.75	1.83	1	1
55	98.16	1.03	1	98.10	1.07	1	101.93	1.22	1	101.38	0.89	1	1
56	100.76	1.15	1	100.37	0.89	1	98.69	0.25	0	96.42	1.79	1	1
57	99.04	0.48	0	101.10	1.63	1	95.87	1.29	1	96.68	0.83	1	1
58	98.23	0.43	0	96.99	1.36	1	98.61	0.14	0	101.35	1.92	1	1
59	98.10	1.17	1	99.06	0.51	0	100.23	0.30	0	101.80	1.38	1	1
60	101.13	1.07	1	100.83	1.27	1	103.84	0.76	0	105.05	1.58	1	1
	99.38		6	99.31		6	98.98		5	98.51		10	10

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 61 to 70

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
61	102.91	0.11	0	102.48	0.48	0	102.86	0.04	0	103.08	0.34	0	0
62	102.30	0.99	1	102.09	0.90	1	100.12	0.06	0	95.36	1.95	1	1
63	98.36	0.35	0	96.75	1.04	1	99.26	1.13	1	97.43	0.45	0	1
64	100.75	1.60	1	99.70	0.56	0	98.03	1.11	1	98.09	1.05	1	1
65	101.05	1.14	1	101.07	1.13	1	103.47	0.83	1	104.20	1.43	1	1
66	100.23	0.73	0	99.46	0.17	0	98.74	1.01	1	99.98	0.45	0	1
67	102.65	1.70	1	105.41	0.08	0	106.86	0.77	0	107.27	1.02	1	1
68	99.83	1.28	1	98.68	0.88	1	95.47	0.24	0	90.69	1.91	1	1
69	102.30	1.54	1	104.61	0.37	0	104.81	0.54	0	104.91	0.63	0	1
70	100.93	1.50	1	102.98	0.04	0	105.19	1.54	1	103.02	0.01	0	1
	101.13		7	101.32		4	101.48		5	100.40		5	9

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 71 to 80

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
71	100.13	0.86	1	101.10	0.36	0	100.40	0.52	0	101.62	1.02	1	1
72	102.13	1.24	1	103.25	0.16	0	105.03	1.57	1	103.24	0.17	0	1
73	101.24	0.14	0	100.74	0.31	0	99.81	1.17	1	102.54	1.33	1	1
74	101.82	0.59	0	101.90	0.53	0	105.00	1.72	1	101.80	0.60	0	1
75	99.23	1.71	1	101.34	0.51	0	104.38	1.22	1	104.00	1.00	1	1
76	100.83	1.12	1	101.46	0.75	0	105.29	1.49	1	103.39	0.38	0	1
77	101.31	0.52	0	103.74	1.44	1	102.63	0.54	0	100.15	1.46	1	1
78	101.73	0.93	1	102.07	0.65	0	102.70	0.13	0	104.93	1.70	1	1
79	100.08	1.09	1	98.86	0.12	0	99.44	0.46	0	97.54	1.43	1	1
80	99.23	1.00	1	99.15	0.97	1	97.14	0.10	0	92.10	2.07	1	1
	100.77		7	101.36		2	102.18		5	101.13		7	10

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 81 to 90

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
81	100.68	1.28	1	101.66	0.05	0	102.32	0.95	1	101.83	0.28	0	1
82	100.31	1.45	1	101.33	0.63	0	103.08	0.79	0	103.68	1.28	1	1
83	100.87	0.11	0	100.40	0.36	0	99.54	1.21	1	102.23	1.45	1	1
84	100.62	1.61	1	102.09	0.46	0	103.51	0.65	0	104.49	1.42	1	1
85	99.44	1.78	1	101.47	0.45	0	101.64	0.64	0	101.68	0.68	0	1
86	99.29	1.39	1	97.93	0.81	1	95.10	0.41	0	91.90	1.79	1	1
87	102.30	1.31	1	101.98	1.07	1	98.89	1.15	1	98.77	1.23	1	1
88	99.76	1.29	1	101.20	0.30	0	101.95	1.12	1	100.81	0.13	0	1
89	101.62	0.88	1	101.97	1.03	1	99.40	0.06	0	95.21	1.85	1	1
90	98.90	1.16	1	99.60	0.60	0	102.16	1.45	1	100.75	0.32	0	1
	100.38		9	100.96		3	100.76		5	100.13		6	10

Application of Cohen's *d* Test to Random-Walk Data

Summary of Results for Runs 91 to 100

	1st Quarter			2nd Quarter			3rd Quarter			4th Quarter			Total
Trial	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Average Price	Cohen's d	Pass Count	Pass Count
91	99.06	1.71	1	95.28	0.66	0	96.95	0.38	0	94.06	1.43	1	1
92	101.07	1.80	1	103.71	0.18	0	105.07	1.20	1	104.04	0.43	0	1
93	102.03	0.62	0	102.89	1.09	1	101.08	0.10	0	97.64	1.81	1	1
94	100.95	0.96	1	101.20	0.83	1	102.99	0.06	0	106.38	1.73	1	1
95	100.73	1.26	1	102.50	0.51	0	102.72	0.73	0	101.99	0.01	0	1
96	101.25	1.48	1	104.46	0.42	0	105.28	0.15	0	111.90	2.04	1	1
97	98.81	0.88	1	98.96	0.71	0	99.65	0.09	0	100.90	1.51	1	1
98	100.16	1.49	1	98.32	0.90	1	93.73	0.58	0	89.94	1.80	1	1
99	100.05	0.42	0	96.70	1.66	1	99.28	0.05	0	101.44	1.29	1	1
100	98.58	1.66	1	95.98	0.54	0	93.66	0.47	0	90.75	1.73	1	1
	100.27		8	100.00		4	100.04		1	99.90		8	10

Attachment 4

Results of Application of Differential Analysis
to 180-Day Random Walk Data

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 1 to 10

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
1	101.20	(2.46)	1.57	1.57	1	103.67	2.46	1.57	1.57	1	1
2	98.33	1.07	0.77	1.39	1	97.26	(1.07)	0.77	1.39	1	1
3	99.07	0.09	0.61	0.14	0	98.98	(0.09)	0.61	0.14	0	0
4	98.72	(2.22)	0.95	2.33	1	100.94	2.22	0.95	2.33	1	1
5	98.73	(1.09)	0.63	1.73	1	99.83	1.09	0.63	1.73	1	1
6	99.98	2.63	0.85	3.09	1	97.35	(2.63)	0.85	3.09	1	1
7	97.72	3.13	1.03	3.04	1	94.59	(3.13)	1.03	3.04	1	1
8	100.25	2.28	0.93	2.46	1	97.97	(2.28)	0.93	2.46	1	1
9	100.99	(1.40)	1.36	1.03	1	102.39	1.40	1.36	1.03	1	1
10	99.00	(0.79)	1.08	0.74	0	99.79	0.79	1.08	0.74	0	0
					8					8	8

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 11 to 20

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
11	100.08	0.59	0.82	0.72	0	99.49	(0.59)	0.82	0.72	0	0
12	98.68	0.64	0.91	0.70	0	98.05	(0.64)	0.91	0.70	0	0
13	99.86	2.05	1.16	1.76	1	97.81	(2.05)	1.16	1.76	1	1
14	97.82	(1.74)	1.02	1.71	1	99.56	1.74	1.02	1.71	1	1
15	99.79	2.34	0.73	3.21	1	97.46	(2.34)	0.73	3.21	1	1
16	98.03	2.49	1.04	2.39	1	95.55	(2.49)	1.04	2.39	1	1
17	100.15	3.64	1.25	2.93	1	96.50	(3.64)	1.25	2.93	1	1
18	96.32	4.35	1.84	2.37	1	91.97	(4.35)	1.84	2.37	1	1
19	99.81	4.84	1.01	4.80	1	94.98	(4.84)	1.01	4.80	1	1
20	98.95	(2.69)	2.17	1.24	1	101.64	2.69	2.17	1.24	1	1
					8					8	8

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 21 to 30

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
21	100.62	(1.71)	0.77	2.23	1	102.33	1.71	0.77	2.23	1	1
22	98.34	(1.34)	1.33	1.01	1	99.68	1.34	1.33	1.01	1	1
23	97.79	(2.10)	1.24	1.70	1	99.90	2.10	1.24	1.70	1	1
24	101.06	0.05	0.87	0.06	0	101.01	(0.05)	0.87	0.06	0	0
25	99.17	(3.18)	0.73	4.36	1	102.35	3.18	0.73	4.36	1	1
26	102.19	2.63	1.26	2.10	1	99.56	(2.63)	1.26	2.10	1	1
27	100.66	(4.56)	1.22	3.75	1	105.22	4.56	1.22	3.75	1	1
28	99.15	2.51	1.07	2.34	1	96.64	(2.51)	1.07	2.34	1	1
29	101.04	(3.83)	2.30	1.67	1	104.87	3.83	2.30	1.67	1	1
30	98.54	(1.44)	1.03	1.40	1	99.98	1.44	1.03	1.40	1	1
					9					9	9

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 31 to 40

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
31	98.72	4.42	1.62	2.73	1	94.31	(4.42)	1.62	2.73	1	1
32	100.69	2.77	1.05	2.64	1	97.92	(2.77)	1.05	2.64	1	1
33	96.64	1.12	1.03	1.08	1	95.52	(1.12)	1.03	1.08	1	1
34	101.59	(0.68)	1.22	0.56	0	102.27	0.68	1.22	0.56	0	0
35	98.80	1.50	0.74	2.03	1	97.30	(1.50)	0.74	2.03	1	1
36	98.79	(4.94)	1.49	3.31	1	103.73	4.94	1.49	3.31	1	1
37	96.54	(1.06)	1.82	0.59	0	97.61	1.06	1.82	0.59	0	0
38	100.43	(1.93)	0.90	2.15	1	102.36	1.93	0.90	2.15	1	1
39	98.67	1.45	1.24	1.18	1	97.22	(1.45)	1.24	1.18	1	1
40	98.80	(0.93)	0.80	1.16	1	99.74	0.93	0.80	1.16	1	1
					8					8	8

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 41 to 50

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
41	100.79	(0.09)	0.75	0.12	0	100.88	0.09	0.75	0.12	0	0
42	101.22	0.06	1.13	0.05	0	101.16	(0.06)	1.13	0.05	0	0
43	102.84	(0.47)	1.01	0.46	0	103.31	0.47	1.01	0.46	0	0
44	98.83	(0.26)	0.59	0.44	0	99.09	0.26	0.59	0.44	0	0
45	101.64	(2.58)	1.02	2.52	1	104.22	2.58	1.02	2.52	1	1
46	98.45	(1.56)	0.72	2.17	1	100.01	1.56	0.72	2.17	1	1
47	101.60	1.13	0.72	1.56	1	100.47	(1.13)	0.72	1.56	1	1
48	101.10	(0.77)	0.58	1.35	1	101.87	0.77	0.58	1.35	1	1
49	98.43	(1.54)	0.92	1.67	1	99.97	1.54	0.92	1.67	1	1
50	97.09	4.88	1.46	3.34	1	92.21	(4.88)	1.46	3.34	1	1
					6					6	6

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 51 to 60

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
51	99.18	2.93	1.23	2.38	1	96.25	(2.93)	1.23	2.38	1	1
52	99.41	(1.27)	1.47	0.87	1	100.68	1.27	1.47	0.87	1	1
53	98.22	0.41	1.31	0.31	0	97.81	(0.41)	1.31	0.31	0	0
54	102.58	(4.15)	1.37	3.03	1	106.74	4.15	1.37	3.03	1	1
55	100.88	(0.57)	1.33	0.43	0	101.45	0.57	1.33	0.43	0	0
56	99.51	0.39	0.68	0.57	0	99.12	(0.39)	0.68	0.57	0	0
57	100.36	1.29	0.97	1.33	1	99.07	(1.29)	0.97	1.33	1	1
58	98.68	(0.37)	1.30	0.29	0	99.05	0.37	1.30	0.29	0	0
59	96.88	(0.93)	1.66	0.56	0	97.81	0.93	1.66	0.56	0	0
60	97.71	5.68	1.58	3.59	1	92.03	(5.68)	1.58	3.59	1	1
					5					5	5

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 61 to 70

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
61	99.00	(2.39)	0.86	2.78	1	101.39	2.39	0.86	2.78	1	1
62	99.65	3.76	1.05	3.59	1	95.89	(3.76)	1.05	3.59	1	1
63	99.07	2.65	1.33	2.00	1	96.41	(2.65)	1.33	2.00	1	1
64	97.82	(0.19)	1.66	0.11	0	98.01	0.19	1.66	0.11	0	0
65	98.67	(0.01)	0.76	0.01	0	98.67	0.01	0.76	0.01	0	0
66	99.46	1.52	0.77	1.96	1	97.94	(1.52)	0.77	1.96	1	1
67	99.23	2.02	0.97	2.08	1	97.21	(2.02)	0.97	2.08	1	1
68	101.50	0.04	1.33	0.03	0	101.46	(0.04)	1.33	0.03	0	0
69	102.65	(2.15)	1.02	2.12	1	104.80	2.15	1.02	2.12	1	1
70	98.66	(0.93)	0.90	1.03	1	99.59	0.93	0.90	1.03	1	1
					7					7	7

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 71 to 80

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
71	97.34	0.75	1.23	0.61	0	96.59	(0.75)	1.23	0.61	0	0
72	102.94	(1.43)	1.45	0.99	1	104.37	1.43	1.45	0.99	1	1
73	101.35	1.46	1.03	1.42	1	99.89	(1.46)	1.03	1.42	1	1
74	98.45	5.33	1.71	3.13	1	93.12	(5.33)	1.71	3.13	1	1
75	102.76	(3.03)	1.75	1.73	1	105.79	3.03	1.75	1.73	1	1
76	102.96	(0.86)	1.40	0.61	0	103.82	0.86	1.40	0.61	0	0
77	101.27	(0.52)	0.82	0.63	0	101.79	0.52	0.82	0.63	0	0
78	98.61	(0.37)	1.44	0.26	0	98.99	0.37	1.44	0.26	0	0
79	98.94	0.47	1.32	0.36	0	98.47	(0.47)	1.32	0.36	0	0
80	98.71	4.85	1.34	3.62	1	93.86	(4.85)	1.34	3.62	1	1
					5					5	5

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 81 to 90

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
81	102.66	(0.97)	1.12	0.87	1	103.64	0.97	1.12	0.87	1	1
82	97.87	2.63	0.97	2.71	1	95.24	(2.63)	0.97	2.71	1	1
83	99.74	2.23	0.87	2.57	1	97.51	(2.23)	0.87	2.57	1	1
84	101.58	(4.21)	1.64	2.56	1	105.79	4.21	1.64	2.56	1	1
85	100.44	0.42	0.73	0.58	0	100.02	(0.42)	0.73	0.58	0	0
86	100.14	(3.97)	1.74	2.28	1	104.11	3.97	1.74	2.28	1	1
87	99.19	1.38	0.97	1.42	1	97.81	(1.38)	0.97	1.42	1	1
88	96.34	1.13	1.76	0.64	0	95.21	(1.13)	1.76	0.64	0	0
89	100.40	(4.09)	1.14	3.58	1	104.49	4.09	1.14	3.58	1	1
90	98.36	1.93	1.20	1.61	1	96.43	(1.93)	1.20	1.61	1	1
					8					8	8

Application of Cohen's *d* Test to Two-Quarter Random-Walk Data

Summary of Results for Runs 91 to 100

	1st Quarter					2nd Quarter					Total
Trial	Average Price	Diff. from 2d Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Average Price	Diff. from 1st Qtr	Pooled Std. Dev.	Cohen's d	Pass Count	Pass Count
91	99.64	1.06	1.04	1.02	1	98.58	(1.06)	1.04	1.02	1	1
92	99.78	(1.42)	0.81	1.75	1	101.21	1.42	0.81	1.75	1	1
93	99.42	3.96	0.91	4.33	1	95.47	(3.96)	0.91	4.33	1	1
94	100.59	(2.95)	0.70	4.23	1	103.54	2.95	0.70	4.23	1	1
95	99.62	(0.52)	0.82	0.64	0	100.14	0.52	0.82	0.64	0	0
96	101.99	0.36	0.73	0.49	0	101.63	(0.36)	0.73	0.49	0	0
97	99.43	4.53	1.43	3.16	1	94.90	(4.53)	1.43	3.16	1	1
98	103.07	(2.55)	1.61	1.58	1	105.62	2.55	1.61	1.58	1	1
99	102.60	(1.64)	1.07	1.53	1	104.24	1.64	1.07	1.53	1	1
100	99.55	1.66	0.84	1.96	1	97.90	(1.66)	0.84	1.96	1	1
					8					8	8