BEFORE THE
WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND
TRANSPORTATION COMMISSION,

       Complainant,

          v.

PUGET SOUND ENERGY, INC.,

       Respondent.

DOCKET NOS. UE-121697
and UG-121705 (*consolidated*)

WASHINGTON UTILITIES AND
TRANSPORTATION COMMISSION,

       Complainant,

          v.

PUGET SOUND ENERGY, INC.,

       Respondent.

DOCKET NOS. UE-130137
and UG-130138 (*consolidated*)

PREFILED REBUTTAL TESTIMONY (NONCONFIDENTIAL) OF
DR. JEFFREY A. DUBIN
ON BEHALF OF PUGET SOUND ENERGY, INC.

DECEMBER 19, 2014

PUGET SOUND ENERGY, INC.

**PREFILED REBUTTAL TESTIMONY
(NONCONFIDENTIAL) OF DR. JEFFREY A. DUBIN**

CONTENTS

PUGET SOUND ENERGY, INC.

PREFILED REBUTTAL TESTIMONY (NONCONFIDENTIAL) OF
DR. JEFFREY A. DUBIN

I.   INTRODUCTION

Q.   Please state your name, business address, and occupation.

A.   My name is Jeffrey Alan Dubin. My consulting business is Jeffrey Alan Dubin

Economic Consultant, Inc. My business address is 434 Puerto Del Mar, Pacific

Palisades, California, 90272.

Q.   Have you prepared an exhibit describing your education, relevant

employment experience, and other professional qualifications?

A.   Yes, I have. It is Exhibit No. ___(JAD-2).

Q.   What are some of your duties as a consultant?

A.   I actively consult with clients on demand issues, environmental issues, market

issues, and antitrust policies. I specialize in microeconomic and micro-

econometric modeling with an emphasis on statistical and demand analysis.

Some of my current research topics include discrete-choice econometrics, energy

economics, tax compliance, sampling and survey methods, and intellectual

property valuations.

1 **Q. Do you hold any other positions?**

2 A. I am presently an Adjunct Professor of Economic, Statistics, and the Practice Area

3 at the University of Southern California.[1]

4 **Q. Please summarize the purpose of your prefiled rebuttal testimony.**

5 A. Puget Sound Energy, Inc. ("PSE") has asked that I review and comment on the

6 Prefiled Direct Testimony of Dr. Christopher A. Adolph, Exhibit No. ____(CAA-

7 1T), on behalf of the Public Counsel Unit of the Washington Attorney General's

8 Office ("Public Counsel") and the Industrial Customers of Northwest Utilities

9 ("ICNU"). Dr. Adolph was similarly asked to respond to the Prefiled Direct

10 Testimony of Dr. Michael J. Vilbert, Exhibit No. ___(MJV-1T), on behalf of PSE.

11 Dr. Vilbert provides an empirical analysis of the relationship of decoupling in

12 electric and gas utilities and the cost of capital for those utilities. Dr. Adolph

13 challenges the empirical analyses of Dr. Vilbert and the following conclusion of

14 Dr. Vilbert published in a report by The Brattle Group:[2]

15 The results of our empirical analysis of decoupling in the electric
16 industry do <u>not</u> support the hypothesis that utilities with
17 decoupling have a lower cost of capital than utilities without
18 decoupling. Our study finds that decoupling is not associated with
19 a statistically significant decrease in the estimated cost of capital.[3]

---

[1] In the course of my current assignment, I worked closely with Professor Douglas Rivers of Stanford University. Professor Rivers is a Political Scientist. He is also generally regarded as a leading statistician in his field. The opinions contained herein are nonetheless my own.

[2] Michael J. Vilbert *et al.*, "The Impact of Revenue Decoupling on the Cost of Capital for Electric Utilities: An Empirical Investigation" (Mar. 20, 2014) ("The Brattle Group Report"), a copy of which is provided by Public Counsel as Exhibit No. ___(SGH-16).

[3] Hill, Exh. No. ___(SGH-16) at page 6.

1   **Q.   Please summarize your conclusions.**

2   A.   Dr. Adolph's opinion and interpretation of the statistical results presented in the

3   Brattle study are flawed. His conclusions rest on unusual procedures, loose

4   evidentiary standards and misinterpretation of the statistical tests. His attempt to

5   turn admittedly "weak" statistical evidence[4] into preponderance of evidence in

6   favor of the opposite comes from a misinterpretation of statistical procedures and

7   statistical hypothesis testing. Contrary to Dr. Adolph's testimony, the evidence

8   presented in The Brattle Group studies and reanalyzed by Dr. Adolph is consistent

9   with decoupling having little impact or even raising the cost of capital for

10   regulated utilities.

11   **II.   DR. ADOLPH'S CONCLUSIONS REST ON UNUSUAL**
12   **PROCEDURES, LOOSE EVIDENTIARY STANDARDS, AND**
13   **MISINTERPRETATION OF THE STATISTICAL TESTS**

14   **Q.   Have you reviewed Dr. Adolph's testimony relating to The Brattle Group's**

15   **studies in this proceeding?**

16   A.   I reviewed Dr. Adolph's testimony as well as the electric study by The Brattle

17   Group, which was co-authored by Dr. Vilbert and his colleagues at The Brattle

18   Group. My focus in this rebuttal testimony is on Dr. Adolph's conclusions with

19   respect to the studies by the Brattle Group and statistical testing. I focus my

20   critique on the same case that Dr. Adolph does, the November 2014 analysis of

21   the electric industry, which Dr. Adolph suggests is Dr. Vilbert's "preferred

---

4   Adolph, Exh. No. ___(CAA-1T) at page 36, line 18.

1 analysis."[5] This was but one result presented by Dr. Vilbert, and my critique of

2 Dr. Adolph's testimony is not limited to this case.

3 **Q.** **Are you familiar with Dr. Vilbert's empirical approach that forms the basis**

4 **of his testimony in this case?**

5 A. Yes. I regularly teach econometric methods used to analyze panel data and have

6 employed such methods in my published studies. I am also familiar with

7 empirical analyses on the cost of capital. In fact, I contributed to this literature in

8 a peer-reviewed publication entitled "Regulatory Climate and the Cost of

9 Capital". [6] I am generally familiar with empirical analyses that attempt to explain

10 the cost of capital for utilities.

11 **Q.** **Do you agree with Dr. Adolph that expertise in the area of utility regulation,**

12 **decoupling and cost of capital determination are not required to assess the**

13 **statistical work of Dr. Vilbert?[7]**

14 A. No. I think it is very important that the statistician have a good understanding of

15 the subject area being analyzed. Dr. Adolph, by his own admission, does not have

16 expertise in "utility regulation, the cost of capital, or the policy of decoupling."[8]

---

[5] Adolph, Exh. No. ___(CAA-1T) at page 14, line 11.

[6] J. Dubin & P. Navarro, "Regulatory Climate and the Cost of Capital," in M. Crew. (ed.), *Regulatory Reform and Public Utilities* (1982); *see also* J. Dubin & P. Navarro, "The Effect of Rate Suppression on Utilities' Cost of Capital," 111 Pub. Utils. Fortnightly 18 (1983).

[7] Adolph, Exh. No. ___(CAA-1T) at page 6, lines 3-7.

[8] Adolph, Exh. No. ___(CAA-1T) at page 3, lines 5-6.

| 1 | **Q.** | **Does this have a consequence for Dr. Adolph's criticisms in your opinion?** |

2	A.	Yes. The interpretation of statistical evidence involves the evaluation of models
3		and analytic choices. Dr. Adolph's apparent reticence to engage the substance of
4		Dr. Vilbert's research is extremely problematic.

| 5 | **Q.** | **Would you explain how Dr. Adolph's lack of expertise in utility regulation** |
| 6 | | **affects his opinions?** |

7	A.	Yes. Dr. Adolph's primary argument is that Dr. Vilbert should have used
8		different statistical procedures (lower confidence levels and a one-sided
9		confidence interval instead of a hypothesis test). With these changes, he
10		interprets the evidence as supporting the opposite result to that reached by
11		Dr. Vilbert.[9]

12		Dr. Adolph, however, fails to address whether his statistical assumptions are
13		consistent with the economic analysis in Dr. Vilbert's study. Specifically,
14		Dr. Adolph's use of one-sided confidence intervals depends on the *assumption*
15		that it is impossible for decoupling to raise the cost of capital. In contrast,
16		Dr. Vilbert recognizes that decoupling may signal a period of higher rather than
17		lower risk,[10] which, according to standard financial theory, can raise the cost of
18		capital. The regulatory response in determining the appropriate level for the cost
19		of capital is complex. The reaction of a regulatory body to decoupling is not
20		certain and directionally unclear.

---

[9] *See, e.g.,* Adolph, Exh. No. ___(CAA-1T) at page 6, lines 17-20.

[10] *See, e.g.,* Vilbert, Exh. No. ___(MJV-1T) at page 5, lines 6-17.

1   In this situation it would have been reasonable to use statistical procedures that do

2   not assume that decoupling never raises the cost of capital. It is my opinion that a

3   two-sided confidence interval is appropriate in this context. Dr. Adolph

4   misinterprets Dr. Vilbert's use of a one-tailed hypothesis test as justifying the

5   assumption that increases in the cost of capital can be ignored, despite

6   Dr. Vilbert's explicit consideration of this possibility.[11]

7   **Q.   What is a one-tailed hypothesis test?**

8   A.   A hypothesis test always involves a null hypothesis, which is tested, and an

9   alternative hypothesis, which determines which evidence is counted against the

10   null hypothesis. The alternative hypothesis in a one-sided test only includes

11   alternatives on one side of the null hypothesis. The *Reference Manual on*

12   *Scientific Evidence*,[12] a manual published by the Federal Judicial Center to assist

13   judges in managing cases involving complex scientific and technical evidence,

14   describes one-tailed hypothesis tests as follows:

15   A one-tailed test would usually be applied when the expert
16   believes, perhaps on the basis of other direct evidence presented at
17   trial, that the alternative hypothesis is either positive or negative,
18   but not both. For example, an expert might use a one-tailed test in
19   a patent infringement case if he or she strongly believes that the
20   effect of the alleged infringement on the price of the infringed
21   product was either zero or negative. (The sales of the infringing
22   product competed with the sales of the infringed product, thereby
23   lowering the price.) By using a one-tailed test, the expert is in
24   effect stating that prior to looking at the data it would be very

---

[11] *See, e.g.,* Vilbert, Exh. No. ___(MJV-1T) at page 5, lines 6-17.

[12] Fed. Judicial Ctr., *Reference Manual on Scientific Evidence* (3d ed. 2011).

1   surprising if the data pointed in the direct opposite to the one
2   posited by the expert.[13]

3   In this case, the null hypothesis is that decoupling has no effect on the cost of

4   capital.

5   Both Dr. Vilbert and Dr. Adolph use a one-sided hypothesis alternative—the

6   effect of decoupling is negative. To perform the test, one computes a *p*-value

7   giving the probability of observing an estimate of the effect of decoupling *lower*

8   than that in the sample under the assumption that there is no effect. The one-sided

9   hypothesis test rejects the null hypothesis if the *p*-value is below a specified level

10   (usually 0.05, though Dr. Adolph advocates loosening this by using much higher

11   values).

12   **Q.   Could you explain how the *p*-value is calculated and what it means?**

13   A.   If the null hypothesis is assumed to be true (*i.e.,* if there is no effect of

14   decoupling), then the sampling distribution of the estimated effect would be as

15   shown in Figure 1 below. The distribution represents variation in the estimated

16   effect that would be observed if we had many different samples. In fact, we only

17   have a single sample, but statistical theory allows us to estimate the amount of

18   sampling variation.

19   The *p*-value, or observed significance level, is the probability of obtaining a more

20   extreme estimate than the sample estimate *under the assumption that there is no*

21   *effect.* If the alternative hypothesis is one-sided, then the *p*-value is the area of the

---

[13]   *See, e.g., Reference Manual on Scientific Evidence* at 321.

1 cross-hatched region in the left tail in Figure 1 below.[14] If the alternative

2 hypothesis is two-tailed, then the *p*-value is the sum of the areas of the cross-

3 hatched regions in both tails, because it is equally likely that the sample estimate

4 could have made an error in the opposite region.

5

Figure 1.

Calculation of one and two−sided P−values

Sample
Estimate   No effect

Estimated effect of decoupling (basis points)

6

---

[14] Dr. Adolph incorrectly refers to a graph centered at the point estimate of -26 (Adolph, Exh.
No. ___(CAA-1T) at page 15, Figure 1) as the "sampling distribution" (*id*. at page 14, lines 13-14).

---

1 **Q. Please explain the logic of one-sided hypothesis testing.**

2 A.   Dr. Vilbert adopted the null hypothesis that the effect of decoupling on the cost of

3   capital was larger or equal to zero. His alternative hypothesis is that decoupling

4   does lower the cost of capital. In this situation, the hypothesis is one-sided with a

5   "critical" region in the left or lower tail with extreme values of the alternative.

6   The *p*-value is the likelihood of observing the estimated effect or something lower

7   under the null hypothesis. If the *p*-value is lower than the significance level set

8   for the "critical" region, the statistician deems the result to be significant.

9 **Q. Do The Brattle Group studies provide one-tailed *p*-values?**

10 A.   Yes. The Brattle Group studies provides one-tailed *p*-values. I believe that this is

11   because directional tests are conservative and that, if the test failed in the direction

12   being tested, it would also fail to reject the more neutral hypothesis of no effect of

13   decoupling. The *Reference Manual on Scientific Evidence* describes the

14   conservative nature of a one-tailed test as follows:

15   Because using a one-tailed test produces *p*-values that are one-half
16   the size of *p*-values using a two-tailed test, the choice of a one-
17   tailed test makes it easier for the expert to reject a null hypothesis.
18   Correspondingly, the choice of a two-tailed test makes null
19   hypothesis rejection less likely. Because there is some
20   arbitrariness involved in the choice of an alternative hypothesis,
21   courts should avoid relying solely on sharply defined statistical
22   tests. Reporting the *p*-value or a confidence interval should be
23   encouraged because it conveys useful information to the court,
24   whether or not a null hypothesis is rejected.[15]

25   Courts have demonstrated a preference for two-tailed tests:

---

[15] *Reference Manual on Scientific Evidence* at 321.

1   Courts have shown a preference for two-tailed tests. *See, e.g.,*
2   *Palmer v. Shultz*, 815 F.2d 84, 95-96 (D.C. Cir. 1987) (rejecting
3   the use of one-tailed tests, the court found that because some
4   appellants were claiming overselection for certain jobs, a two-
5   tailed test was more appropriate in Title VII cases); *Moore v.*
6   *Summers*, 113 F. Supp. 2d 5, 20 (D.D.C. 2000) (reiterating the
7   preference for a two-tailed test). *See also* David H. Kaye & David
8   A. Freedman, Reference Guide on Statistics, Section IV.C.2, in
9   this manual; *Csicseri v. Bowsher*, 862 F. Supp. 547, 565 (D.D.C.
10   1994) (finding that although a one-tailed test is "not without
11   merit," a two-tailed test is preferable).[16]

12   In short, The Brattle Group's use of a one-tailed test was a conservative one. It

13   made the rejection of the null hypothesis (that decoupling does not lower the cost

14   of capital) easier. Use of the one-tailed test by The Brattle Group, however, did

15   not mean to imply that one should assume that decoupling cannot raise the cost of

16   capital.

17   **Q.**   **What is a confidence interval?**

18   A.   A confidence interval is a method for quantifying the likely size of error of a

19   statistical estimate. Instead of using a single value, we can form an interval that

20   will, with high probability, contain the parameter that we wish to estimate.

21   The effect of decoupling on the cost of capital is an example of an unknown

22   parameter. Based on data from 14 utilities, The Battle Group used regression

23   analysis to estimate the difference in the cost of capital when a utility does or does

24   not have a decoupling policy. As Dr. Adolph acknowledges, this estimate is

25   imprecise.[17] Dr. Adolph cites two introductory textbooks that recommend

---

[16] *Reference Manual on Scientific Evidence* at 321 n.49.

[17] Adolph, Exh. No. ___(CAA-1T) at page 4, lines 18-19.

1   confidence intervals be used to summarize the evidence.[18] I agree with this

2   recommendation—a confidence interval provides a range of estimates that can be

3   supported by the available data.

4 **Q.   What is the most common form of confidence intervals used in economics?**

5 A.   In nearly all scientific work (including the example on page 425 of the Moore and

6   McCabe textbook cited in footnote 7 of Dr. Adolph's testimony), two-sided 95%

7   confidence intervals are standard. Anything else is quite unusual.[19]

8 **Q.   What is the 95% confidence interval for the effect of decoupling on the cost**

9   **of capital in The Battle Group studies?**

10 A.   The estimated -26 basis point effect has a 95% confidence interval ranging from -

11   79 basis points to +29 basis points.[20] This is the bottom interval in Figure 2

12   below.

---

[18]   Adolph, Exh. No. ___(CAA-1T) at page 22, footnote 7 (citing Damodar N. Gujarati, *Basic Econometrics* at 134 (1995) and David S. Moore & George P. McCabe, *Introduction to the Practice of Statistics* at 425 (2006).

[19]   *See, e.g.,* Daniel L. Rubinfeld, "Reference Guide on Multiple Regression," in *Reference Manual on Scientific Evidence* at 320-21 (2011) .

[20]   The confidence interval is calculated using the estimate plus or minus 1.96 times the standard error of the estimate. The standard error can be calculated by dividing the estimate by the standard normal quartile corresponding to the one-tailed p-value. On page 9, line 14, Dr. Adolph incorrectly states that the p-value was 0.83. Elsewhere (page 12, line 14; page 13, line 20; page 14, line 13; page 16, line 15; page 20, line 21; page 23, line 19; page 31, line 15), he gives the correct p-value of 0.17.

1

Figure 2.

**Confidence Intervals for Effect
of Decoupling**
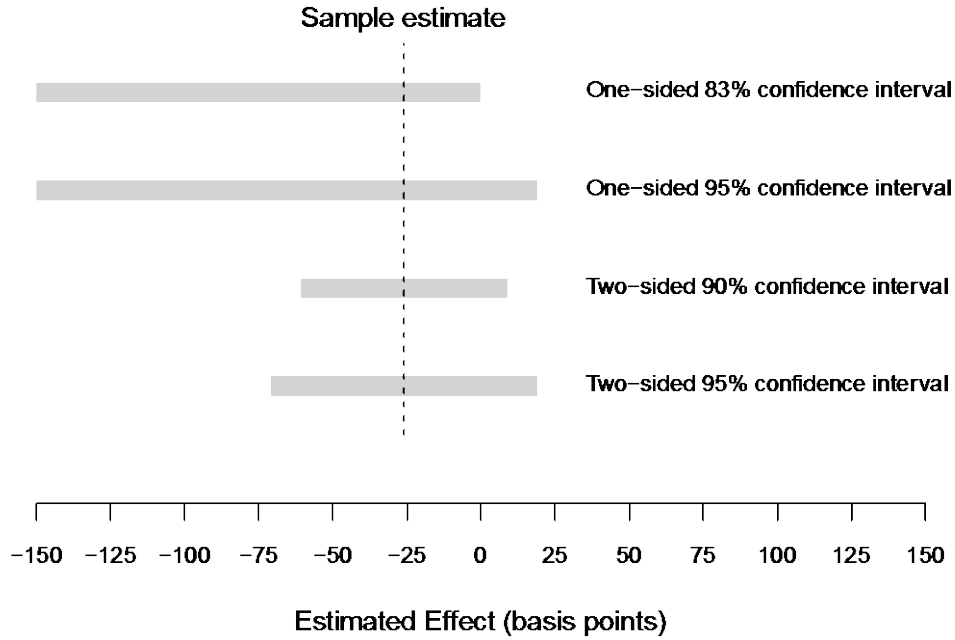
2

3 **Q.   How should the confidence interval of the effect of decoupling on the cost of**

4   **capital be interpreted?**

5 A.   The data are consistent with decoupling raising, lowering, or having no effect on

6   the cost of capital. With 95% confidence, we can rule out decreases of more than

7   79 basis points or increases of more than 27 basis points. We cannot rule that

8   there is no effect from decoupling.

9 **Q.   What exactly does "95% confidence" mean?**

10 A.   The confidence interval, as calculated by me (as well as those calculated by

11   Dr. Adolph, discussed below) should *not* be interpreted as saying "there is a 95%

1   probability that the effect of decoupling is between -79 and +29 basis points."[21]

2   The effect of decoupling is a population parameter. That unknown parameter is

3   not treated as a random variable and does not have a probability distribution.[22]

4   The confidence level refers to the reliability of the procedure used to form the

5   intervals. Randomness occurs because the confidence interval is computed using

6   a sample of data. If the same procedure is repeated on many independent

7   samples, about 95% of the computed confidence intervals will contain the

8   corresponding population parameter.

---

[21]   The Reference Manual on Scientific Evidence also recognizes that federal courts have recognized that

> it is misleading to suggest that "[a] 95% confidence interval means that there is a 95% probability that the 'true' relative risk falls within the interval" or that "the probability that the true value was . . . within two standard deviations of the mean . . . would be 95 percent." *DeLuca v, Merrell Dow Pharms., Inc.*, 791 F. Supp. 1042, 1046 (D.N.J. 1992), *aff'd*, 6 F.3d 778 (3d Cir, 1993); *SmithKline Beecham Corp. v. Apotex Corp.*, 247 F. Supp. 2d 1011, 1037 (N.D. Ill. 2003), *aff'd on other grounds*, 403 F.3d 1331 (Fed. Cir. 2005).

Reference Manual on Scientific Evidence at 247 n.92.

[22]   In Bayesian statistics, parameters are treated as random variables, but this requires the assignment of a subjective prior probability distribution, which Dr. Adolph has not formed nor has the expertise to assess. The Reference Manual on Scientific Evidence describes Bayesian statistics as follows:

> Given the sample data, what is the probability of the null hypothesis? The question might be of direct interest to the courts, especially when translated into English; for example, the null hypothesis might be the innocence of the defendant in a criminal case. Posterior probabilities can be computed using a formula called Bayes' rule. However, the computation often depends on prior beliefs about the statistical model and its parameters; such prior beliefs almost necessarily require subjective judgment.

Reference Manual on Scientific Evidence at 241-42 (2011) (footnote omitted).

---

1   **Q.   Why do you use a 95% confidence level instead of a 90% or lower confidence**

2   **level?**

3   A.   There is nothing sacred about the 95% confidence level, but inferences rapidly

4   become unreliable as the confidence level is lowered. The *Reference Manual on*

5   *Scientific Evidence* describes the level as follows:

6   In practice, statistical analysts typically use levels of 5% and 1%.
7   The 5% level is the most common in social science, and an analyst
8   who speaks of significant results without specifying the threshold
9   probably is using this figure.[23]

10   Conventional hypothesis testing predominately uses 95% or 99% confidence and

11   rarely a lower level of 90%. Contrary to what Dr. Adolph claims, using a 95%

12   confidence interval does not represent "a heavy burden."[24] In most cases, a

13   nominal 95% confidence level overstates one's actual level of confidence in

14   sample estimates:

15   [Confidence intervals] reflect only the statistical uncertainty, and
16   thus provide a *lower bound* on the true uncertainty. The generally
17   nonquantifiable deviations of the practical assumptions from
18   reality provide an added *unknown* element of uncertainty. If there
19   were formal methods to reflect this further uncertainty
20   (occasionally there are, but often there are not), the resulting
21   interval, expressing the *total* uncertainty would clearly be longer
22   than the statistical interval alone.[25]

---

[23]   *Reference Manual on Scientific Evidence* at 251.

[24]   Adolph, Exh. No. ___(CAA-1T) at page 28, line 19, through page 29, line 1.

[25]   Gerald J. Hahn & William Q. Meeker, *Statistical Intervals: A Guide for Practitioners* at 5 (1991) (emphasis in original); *see also* Michael O. Finkelstein, *Basic Concepts in Probability and Statistics in the Law* at 86-87 (2009).

1   At several points in his testimony,[26] Dr. Adolph mentions a variety of untested

2   assumptions upon which the analyses rest and which contribute additional

3   uncertainty not reflected in the calculated confidence levels.

4   **Q.**   **Does Dr. Adolph use standard 95% confidence intervals in his analyses?**

5   A.   No. Dr. Adolph uses one-sided confidence bounds with various different

6   confidence levels below 95%. Both of these choices are unusual and

7   questionable.

8   **Q.**   **What is a one-sided confidence bound?**

9   A.   Dr. Adolph's confidence bounds have an upper limit (on the effect of decoupling)

10   but no lower limit. That is, the intervals range from minus infinity to an upper

11   limit.[27] The top two confidence intervals in Figure 2 above are one-sided

12   intervals.

13   **Q.**   **Should Dr. Adolph have used one-sided confidence bound or a standard**

14   **confidence interval?**

15   A.   One-sided confidence bounds are extremely rare in economics. Conventional

16   confidence intervals are a more appropriate way of summarizing the evidence.

17   The one-sided confidence bound calculated by Dr. Adolph includes large negative

18   effects (of -100 or more basis points) that have little support in the data, instead of

---

[26] *See, e.g.,* Adolph, Exh. No. ___(CAA-1T) at page 3, lines 20-22, at page 5, lines 12-13, and at page 10, lines 7-14.

[27] *See* Adolph, Exh. No. ___(CAA-1T) at page 15, Figure 1.

---

1   moderately positive values (between +19 and +29 basis points) which are better

2   supported by the data.

3   Dr. Adolph notes that "[c]onfidence bounds are an alternative representation of *p*-

4   values."[28] More precisely, one-sided 95% confidence bounds contain all values

5   that would not be rejected using a 0.05 significance level. In practice, however,

6   the two are quite different. A *p*-value from a one-tailed *t* or *z* test is exactly twice

7   the *p*-value from a two-tailed test of the same null hypothesis. The one-tailed test

8   almost always occurs when, as in The Brattle Group studies, the null hypothesis

9   cannot be rejected and failure to reject using the more powerful one-tailed test is a

10   sign of how weak the evidence is against the null hypothesis.

11   One-sided confidence intervals, on the other hand, are semi-infinite intervals and

12   bear no simple relationship to the finite intervals of two-sided confidence

13   intervals. The primary application of one-sided bounds is in determining a

14   tolerance level (*e.g.,* the minimum temperature at which 99% of O-rings would

15   not fail or the maximum safe dose of a drug). In the present case, the relevant

16   evidence provided by the one-sided confidence bound is that there is a good

17   chance that decoupling *raises* the cost of capital.

18   **Q.   What confidence levels does Dr. Adolph use?**

19   A.   At various points in his testimony, Dr. Adolph adopts confidence levels of 87%,[29]

20   83%,[30] and 63%.[31] He provides no justification for these confidence levels,

---

28   *See* Adolph, Exh. No. ___(CAA-1T) at page 14, lines 2-3.
29   Adolph, Exh. No. ___(CAA-1T) at page 30, line 12.

1 except that they produce negative estimates for the effect of decoupling on the

2 cost of capital.

3 **Q. Why shouldn't lower levels of confidence than 95% be used?**

4 A. Using a 95% level of confidence (or a 0.05 significance level) is just a rule of

5 thumb, and Dr. Adolph is correct that ignoring evidence with a p-value of 0.051 is

6 arbitrary. However, Dr. Adolph proposes not to weaken conventionally accepted

7 evidentiary standards slightly but to discard them entirely. He advises the

8 Commission to reject the hypothesis that decoupling has no effect on the cost of

9 capital using a one-tailed p-value of 0.17, which corresponds to a two-tailed p-

10 value of 0.34 or a 66% two-sided confidence interval.

11 Indeed, the logical implication of Dr. Adolph's position is that the preponderance

12 of evidence standard should be based on 50% confidence levels. That result is

13 absurd as discussed below.

14 **Q. Did Dr. Adolph advocate changing the level of significance in Dr. Vilbert's**

15 **hypothesis tests because the Commission will rule based on the**

16 **preponderance of evidence standard?**

17 A. Yes. Dr. Adolph would like to lower the scientific standard of 95% significance

18 to a much lower level. He opines that "an adjudicatory proceeding evaluating the

19 preponderance of the evidence should consider statistical evidence below 95

---

[30] Adolph, Exh. No. ___(CAA-1T) at page 13, line 20.
[31] Adolph, Exh. No. ___(CAA-1T) at page 34, line 14.

1   percent confidence . . ."[32] and that "demanding a 95% percent confidence level

2   would be analogous to demanding proof beyond a reasonable doubt . . . ."[33]

3 **Q.   What does Dr. Adolph propose that the Commission accept instead of the**

4   **scientific level for statistical significance?**

5 A.   Dr. Adolph would have the Commission raise the significance level from 5% to

6   17% in this case because the legal standard is preponderance of evidence.

7   Dr. Adolph states that "[l]ower legal standards of proof, such as preponderance of

8   evidence, correspond to lower required levels of statistical significance."[34]

9   Similarly Dr. Adolph stated that "statistical evidence significant at the 0.05 level –

10   a standard I consider inappropriately high where standard of proof is

11   preponderance of evidence . . . ."[35] His proposal to the Commission is quite clear

12   but also very wrong.

13 **Q.   Do you agree with Dr. Adolph?**

14 A.   Absolutely not. Dr. Adolph does not understand the relationship between legal

15   and statistical significance. Dr. Adolph's proposal has been bandied about before.

16   It is quite instructive to review some of the commentary. For instance, Professor

17   Franklin Fisher, a distinguished econometrician, former past president of the

18   Econometric Society, and Emeritus Professor of Economics from the

19   Massachusetts Institute of Technology, explains this issue in a Columbia Law

---

[32] Adolph, Exh. No. ___(CAA-1T) at page 7, lines 2-6.

[33] Adolph, Exh. No. ___(CAA-1T) at page 24, lines 16-17.

[34] Adolph, Exh. No. ___(CAA-1T) at page 24, lines 13-14.

[35] Adolph, Exh. No. ___(CAA-1T) at page 31, lines 21-23.

1   Review to a non-statistical audience.[36] Professor Fisher uses the example of

2   testing whether a coin is weighted to show the absurdity of using a 50%

3   significance level with a one-tailed test: "The conclusion will be that the coin is

4   weighted if it comes up heads more often than tails no matter how small the

5   number of tosses."[37]

6 **Q.   Does Professor Fisher explain why 95% confidence intervals (and requiring**

7 ***p*-values of 0.05 or less to reject hypotheses) are reasonable?**

8 A.   Yes. Professor Fisher explains as follows:

9   Significance levels of five percent and one percent are generally
10   used by statisticians in testing hypotheses. That is, given a
11   significance level of five percent (or one percent for a stricter
12   researcher) it is safe to assume that the true coefficient is not zero
13   and that therefore the variable being tested has some effect on the
14   dependent variable in question. Some lawyers might question
15   whether the use of such levels imposes too severe a standard. Why
16   reject the hypothesis that a certain coefficient is zero only if the
17   probability that the results obtained are due to chance is five
18   percent or less? Where the hypothesis involved is of legal
19   importance (for example, when a nonzero coefficient would
20   indicate the presence of sex discrimination in wages), would it not
21   make more sense to use a "preponderance of the evidence"
22   standard and require only significance at fifty percent?

23   *Such an approach, however, would reflect a flawed understanding*
24   *of what significance levels really mean.* In particular, a
25   significance level of fifty percent would not correspond to a
26   "preponderance of the evidence" standard. The significance level
27   tells us only the probability of obtaining the measured coefficient
28   value *if* the true value is zero; it does *not* give the probability that
29   the coefficient's true value *is* zero, nor does subtracting the

---

[36] Franklin M. Fisher, *Multiple Regression in Legal Proceedings*, 80 Colum. L. Rev. 702 (May 1980).

[37] Franklin M. Fisher, *Statisticians, Econometricians, and Adversary Proceedings*, 81 J. Am. Stat. Ass'n 277, 280 (June 1986).

significance level from one hundred percent give the probability that the hypothesis is not true. Because, even with a large sample, it is quite possible to obtain results differing from a coefficient's true value, it is conventionally thought that there must be a very high probability that the coefficient is not zero before it can be conclusively claimed that the variable associated with the coefficient has a definite effect on the dependent variable.

This does not mean that only results significant at the five percent level should be presented or considered. Less significant results may be suggestive, even if not probability, and suggestive evidence is certainly worth something.[38]

Relaxing the confidence level (or raising the significance level used to test hypotheses) fails to address the fundamental problem with weak evidence. If a confidence interval provides considerable support for two opposing positions, it has little evidentiary value. Dr. Adolph's proposal to this Commission to raise the significance level reflects a "flawed understanding".

Q.   Do you agree with Professor Fisher?

A.   Absolutely.

Q.   What standards are adopted by courts that the Commission might review regarding the relationship of preponderance of evidence and statistical significance levels?

A.   According to Professor Michelle Mello, courts have generally required that statistical evidence of discrimination meet the 95% criterion for statistical significance in order to be deemed to have satisfied the plaintiffs burden of proof under the preponderance standard:

---

[38] *Multiple Regression in Legal Proceedings*, 80 Colum. L. Rev. at 717-18 (emphasis added).

1	The relationship between statistical significance and satisfaction of
2	a legal burden of proof has been addressed most directly by courts
3	and commentators in the context of disparate-impact
4	discrimination cases. There, courts have generally required that
5	statistical evidence of discrimination meet the 95% criterion for
6	statistical significance in order to be deemed to have satisfied the
7	plaintiffs burden of proof under the preponderance standard.[39]

8   In other words, the 5% significance level used in hypothesis should apply, even

9   when the preponderance of evidence of standard is followed by the court.

10   **Q.**   **Did Professor Mello discuss why application of the preponderance of**

11   **evidence standard is incorrect?**

12   A.   Yes. Professor Mello refers to this as "a misguided conflation of legal

13   adjudicators' confidence levels and the concept of statistical significance."[40] She

14   explains that "[t]he confusion [by some courts] is perhaps wrought by

15   statisticians' unfortunate tendency to refer to *p*-values as 'confidence levels.'"[41]

16   **Q.**   **Did Dr. Adolph refer to *p*-values as confidence levels?**

17   A.   Yes. For instance, Dr. Adolph's Figure 3 labels the darkly shaded area as the

18   "Confidence that Decoupling Lowers Cost".[42] Recall that Professor Fisher

19   explained as follows:

20	The significance level tells us only the probability of obtaining the
21	measured coefficient value if the true value is zero; it does not give

---

[39]   Michelle M. Mello, *Using Statistical Evidence to Prove the Malpractice Standard of Care: Bridging Legal, Clinical, and Statistical Thinking*, 37 Wake Forest L. Rev. 821, 841 (2002) (citing Marcel C. Garaud, *Comment, Legal Standards and Statistical Proof in Title VII Litigation: In Search of a Coherent Disparate Impact Model*, 139 U. Pa. L. Rev. 455, 468 (1990)).

[40]   37 Wake Forest L. Rev. at 839.

[41]   37 Wake Forest L. Rev. at 840.

[42]   Adolph, Exh .No. ___(CAA-1T) at page 19, Figure 3.

---

1 | the probability that the coefficient's true value is zero, nor does
2 | subtracting the significance level from one hundred percent give
3 | the probability that the hypothesis is not true.[43]

4 | Similarly, Professor Mello explains as follows:

5 | But it is also not appropriate to conclude that because the p-value
6 | is below the required level, the weight of the evidence shows that
7 | the proposition is false. A p-value of 0.49 does not mean that 51%
8 | of the evidence points to the falsity of the proposition.[44]

9 | Consequently a p-value of 17% does not mean that 83% of the evidence points to

10 | the falsity of the proposition (decoupling has a positive or no effect). In short,

11 | Dr. Adolph cannot conclude that there is 83% (100% - 17%) confidence that

12 | decoupling has a negative effect.

13 | **Q.** **Are there recommendations to help courts not fall into this trap?**

14 | A. To avoid courts being misled about statistical evidence, Professor David Kaye has

15 | suggested as follows:

16 | When a confidence interval is used in court, . . . it should not be
17 | denominated a "confidence" interval because the confidence
18 | coefficient does not equal the subjective confidence that one
19 | should have in the truth of a relevant proposition.[45]

20 | Dr. Adolph's labeling in his figures is misleading and can only cause confusion.

21 | **Q.** **Has this issue been discussed elsewhere?**

22 | A. Yes. Professor David Kaye has written extensively on this issue. Professor Kaye

23 | notes that *p*-values are the probability of observing the data given the null

---

[43] *Multiple Regression in Legal Proceedings*, 80 Colum. L. Rev. at 717-18 (emphasis added).

[44] 37 Wake Forest L. Rev. at 840.

[45] David H. Kaye, *Is Proof of Statistical Significance Relevant?,* 61 Wash. L. Rev. 1333, 1349 n.78 (1986).

hypothesis (not the converse). The *Reference Manual on Scientific Evidence*

describes this relationship as follows

> Because *p* is calculated by assuming that the null hypothesis is
> correct, *p* does not give the chance that the null is true. The *p*-
> value merely gives the chance of getting evidence against the null
> hypothesis as strong as or stronger than the evidence at hand.
> Chance affects the data, not the hypothesis.[46]

In other words, *p*-values give the likelihood of the observed parameter estimate or

results more extreme (the data) given that the null hypothesis of no effect is

actually true. Hence, *p*-values give the conditional probability P(data | H0)

(in the notation of Professor Kaye). A result is deemed significant when the *p*-

value is smaller than the significance level because the chance of getting such an

extreme outcome is predetermined to be small.[47]

Q.   **Do significance *p*-values translate to expressions of certitude?**

A.   No. As Professor Kaye explains: "The probability of the alternative is *not*

generally equal one minus the significance probability."[48] As discussed above,

Dr. Adolph makes this error when he labels his charts showing that there is a 83%

significance of obtaining a negative change in ROE from decoupling. Professor

Kaye, Mello and Fisher all are saying exactly the same thing. Dr. Adolph has it

wrong.

---

[46] *Reference Manual on Scientific Evidence* at 250.

[47] *See, e.g.,* David H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 Law & Contemp. Probs. 13 (1983).

[48] 61 Wash. L. Rev. at 1347 (italics in original).

1 **Q.   Why does the preponderance of evidence standard not affect the significance**

2 **level as Professors Kaye, Mello, and Fisher have written in peer-reviewed**

3 **published articles?**

4 A.   Courts must weigh all the evidence before them and subjectively determine the

5   likelihood of the hypothesis—e.g., who should be the prevailing party in a civil

6   matter or whether someone has committed the crime. The Commission must

7   consider the likelihood that the decoupling effect is non-zero based on all the

8   evidence.

9   The null hypothesis, H0, is that decoupling has no effect on the cost of capital.

10   The alternative hypothesis, H1, is that decoupling raises or lowers the cost of

11   capital (two-sided) or lowers the cost of capital (one-sided). The Commission

12   must assess the probabilities P(hypothesis | data). Hypothesis testing only gives

13   us the probability P(data | hypothesis).

14   The preponderance of evidence standard reflects the decision rule that

15   P(H0 | data) is larger or smaller than the probability of the alternative

16   P(H1 | data). This is completely different than what hypothesis testing sets out to

17   calculate—the *p*-value equal to P(data | H0). These probabilities are not the same,

18   and one cannot infer one from the other absent prior information.[49] Indeed, the

19   *Reference Manual on Scientific Evidence* warns against this transposition fallacy:

---

[49] *See, e.g.,* Morris H. DeGroot, *Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio*, 68 Journal of the American Statistical Association 966 (1973).

1 We call this a converse probability because it is of the form
2 P(H0 | data) rather than P(data | H0); an equivalent phrase,
3 "inverse probability," also is used. Treating P(data | H0) as if it
4 were the converse probability P(H0 | data) is the transposition
5 fallacy. For example, most U.S. senators are men, but few men are
6 senators. Consequently, there is a high probability that an
7 individual who is a senator is a man, but the probability that an
8 individual who is a man is a senator is practically zero. . . . The
9 frequentist p-value, P(data | H0), is generally not a good
10 approximation to the Bayesian P(H0 | data); the latter includes
11 considerations of power and base rates.[50]

12 **Q. Does Dr. Adolph interpret *p*-values correctly?**

13 A. No. Dr. Adolph states as follows with respect to *p*-values:

14 Suppose the Commission saw a series of 100 cases over time with
15 evidence of this kind – results which are significant at the 0.17
16 level. If the Commission took the side with the preponderance of
17 evidence in each of those cases [*i.e.,* rejected the null hypothesis of
18 no effect when $p < 0.17$], then it would be right on the facts in 83
19 cases, and wrong due to sampling error in 17 cases.[51]

20 This reflects a fundamental misunderstanding of *p*-values. The *p*-value is

21 computed *conditional* upon the null hypothesis being true and does not, in

22 general, give the probability (or expected proportion) of correct decisions when

23 the null hypothesis does not hold.

24 For example, suppose we have 100 coins of unknown provenance and wish to test

25 which ones are fair. We flip each coin twice and, if it comes up heads both times,

26 we declare it to be biased. The null hypothesis is that each coin is fair or,

27 equivalently, that its probability of landing heads on each toss is one half. If the

28 null hypothesis is correct for a given coin, then there is probability $1/4 = 0.25$ of

---

[50] *Reference Manual on Scientific Evidence* at 258 n.119.

[51] Adolph, Exh. No. ___(CAA-1T) at page 31, lines 12-20.

1 obtaining two heads in two tosses. In how many tosses will this decision rule—of

2 declaring every coin biased that lands heads twice in two tosses—lead to the right

3 decision? That is, how many mistakes would we expect to make using a *p*-value

4 of 0.25?

5 The answer, unfortunately, is that not enough information has been provided to

6 answer the question. Suppose, for instance, that half the coins are fair (they land

7 heads with probability 0.5) and half are biased (say, they land heads with

8 probability 0.6). Of the 50 fair coins, on average, 12.5 will be rejected as biased.

9 Of the 50 biased coins, there is probability $0.6 \times 0.6 = 0.36$ of getting two heads in

10 two tosses. So, in expectation, 18 of the 50 would be detected as biased.

11 Together, the wrong decision would be made about 44.5% of the time. The errors

12 would go in both directions (sometimes biased coins would be declared fair and

13 sometimes fair coins would be declared biased). 32 of the 44.5 coins (71.9%) that

14 the test identifies as biased would actually be biased—somewhat less than the

15 75% confidence level that Dr. Adolph's reasoning would suggest.

16 The correct interpretation of the 0.25 significance level is that *if the null*

17 *hypothesis is true*, then the test will reject the null incorrectly 25% of the time.

18 **Q.** **In conclusion, does Dr. Adolph present persuasive evidence that decoupling**

19 **lowers the average cost of capital for public utilities?**

20 A. The evidence described by Dr. Adolph that decoupling lowers the average cost of

21 capital is very weak. The data presented are entirely consistent with decoupling

22 having no impact on the cost of capital or even raising it. Dr. Adolph's

1 conclusions rest on unusual procedures, loose evidentiary standards, and

2 misinterpretation of the statistical tests.

III. CONCLUSION

4 **Q.   Does this conclude your prefiled rebuttal testimony?**

5 A.   Yes, it does.