BEFORE THE WASHINGTON

UTILITIES & TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION

V.

PUGET SOUND ENERGY, INC.

Dockets UE-121697 and UG-121705

AND

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION

V.

PUGET SOUND ENERGY, INC.

Dockets UE-130137 and UG-130138

DIRECT TESTIMONY OF DR. CHRISTOPHER A. ADOLPH (CAA-1T)

ON BEHALF OF

PUBLIC COUNSEL

AND

INDUSTRIAL CUSTOMERS OF NORTHWEST UTILITIES (ICNU)

DECEMBER 3, 2014

DIRECT TESTIMONY OF DR. CHRISTOPHER A. ADOLPH (CAA-1T)
DOCKETS UE-121697, UG-121705, UE-130137 & UG-130138

TABLE OF CONTENTS

FIGURES

Dr. Christopher A. Adolph's Exhibit List

I. INTRODUCTION / SUMMARY

Q: Please state your name and business address.

A: My name is Dr. Christopher A. Adolph, and my business address is Department of Political Science, Box 353530, University of Washington, Seattle, WA 98195.

Q: By whom are you employed and in what capacity?

A: I am an associate professor of political science at the University of Washington.

Q: On whose behalf are you testifying?

A: I am testifying as a private consultant on behalf of the Public Counsel Unit of the Washington Attorney General's Office (Public Counsel) and the Industrial Customers of Northwest Utilities (ICNU).

Q: What is the purpose of your testimony and provide a brief summary?

A: I was retained to provide expert statistical testimony and analysis regarding the testimony and exhibits of PSE witnesses, in particular Dr. Michael J. Vilbert's testimony regarding the statistical interpretation of the Brattle Group studies of decoupling and the cost of capital.

As I will discuss in more detail below, the preponderance of the statistical evidence in the Brattle analyses on electric utilities shows that decoupling lowers the cost of capital. This evidence is consistent across all of the Brattle models of the electric utility industry. Additionally, the Brattle studies do not provide evidence against the claim that decoupling lowers the cost of capital in the electric industry.

Q: Please describe your professional qualifications.

A: I am a tenured associate professor of political science at the University of Washington, Seattle. I am also an adjunct associate professor of statistics at the University of Washington, and a core faculty member of the Center for Statistics and the Social Sciences. I received a Ph.D. in political science from Harvard University in 2005. All of my published research involves the statistical analysis of data or the development of new statistical techniques. I teach a variety of graduate statistics courses at the University of Washington and the University of Essex Summer School in Social Science Data Analysis, including relevant courses on inference in linear models and on panel data techniques. A more detailed account of my educational background and occupational experience appears in Exhibit No. CAA-2.

Q: What is your experience as an expert witness?

In 2005, I served as an expert witness on statistical issues in *Borders v. King County*, the Washington gubernatorial election challenge regarding the disputed election of Christine Gregoire. In 2010, I was entered on court records as an expert witness in an Ohio election challenge, *Sarantou v. Lucas County*, which was subsequently dropped.

Q: What is your relevant expertise for this case?

A: I am an expert in several areas of statistics relevant to this case, including: (1) the construction, interpretation, and evaluation of linear regression models; (2) the visual presentation and interpretation of statistical models; and (3) the construction, interpretation, and evaluation of models of panel data, or models

where the data consist of multiple time series (e.g., different quarters) for different observed units (e.g., firms). I conduct and publish research and teach graduate-level courses in each of these three areas.

I am not an expert in financial accounting, utility regulation, the cost of capital, or the policy of decoupling, and will not testify on the substance of these areas. My testimony is focused only on the issue of interpreting statistical models. Because statistical models behave in the same way when applied to different subject areas, it is possible for a statistical expert to have an expert opinion on what a statistical analysis means even if that statistical expert is not deeply knowledgeable about the specific substantive issues of interest to the designer of the model. This is such a case.

Q: Do you have an opinion about the data collected by the Brattle Group to study decoupling in the electric and gas utility markets?

A: As an expert in the analysis of panel data, I recognize that the datasets constructed by Brattle are indeed panel datasets consisting of multiple observations of a set of utility companies over time. Based on my experience with panel data, I have an opinion of whether these datasets are large or small by the standards of panel data analysis, and thus whether we might reasonably expect to obtain precise results from a statistical analysis. As I am not an expert in the areas of accounting, utility regulation, cost of capital or the policy of decoupling, I do not have an opinion about whether Brattle has collected a representative sample of utilities, whether they have measured the cost of capital or degree decoupling appropriately, or

whether they have made reasonable decisions regarding which observations to exclude from these analyses.

Q: Is it your opinion that the Brattle data constitute large or small samples?

A: Panel data models, including all of the linear regressions performed by the Brattle Group for the electric and gas industries, base their inferences on a series of comparisons within the experience of each firm. After deleting missing data, the November 2014 electric model includes only 14 firms, but only 12 firms were observed to experience a change in the degree of decoupling within the set of observations modeled by the linear regression. Because these models include controls for each firm, only the 12 firms with fully observed changes in decoupling tell us anything about the effects of decoupling in the electric utility industry. Twelve studied units is a fairly small number for a panel study. Likewise, the 2014 gas model includes only 12 firms, but of these, only the 10 firms that we observed within the modeled data to experience a change in the degree of decoupling tell us anything about the effects of decoupling in the gas utility industry. Compared to other datasets studying a series of units through time, these are relatively small datasets, though not outside the range of published panel data studies. The relatively small sample creates the possibility of obtaining imprecise results even if decoupling has a real-world effect on the cost of capital. Nevertheless, despite its relatively small size, the current sample is sufficient to produce usable evidence about the effect of decoupling on the cost of capital, and does not alter my recommendations below.

Q: Do you have an opinion about the specific design and variables chosen by the Brattle Group to model the effects of decoupling on the cost of capital in the electric and gas utility markets?

A: As an expert in the analysis of panel data, I recognize that the specifications used by Brattle across the various studies of the electric and gas industries cited in Dr. Michael J. Vilbert's testimony are linear regressions on panel data. The technical specifications of these models vary in ways that are not fully explained or justified in Dr. Vilbert's testimony. A key aspect of all of Brattle's models is the use of dummy variables to control for idiosyncratic variation across firms and across time periods. This helps protect against bias that would otherwise occur due to omitted controls, but makes the models less precise than other datasets of similar size, as I note above. The Brattle models vary considerably in how they deal with changes over time in the cost of capital, with less justification. I do not have an opinion on what the best model would be for the Brattle data, and will focus my comments primarily on interpreting the findings from the range of models Dr. Vilbert has presented in his own testimony.

Q: Do you have an opinion about the statistical inferences Dr. Vilbert draws from Brattle's models of the cost of capital?

A: I have an opinion, informed by my statistical expertise and experience as a social science data analyst, on the interpretation of the estimates cited by Dr. Vilbert from the Brattle Group models of the cost of capital. In particular, I have expertise regarding the statistical meaning of the estimated relationship between the cost of capital and the degree of decoupling (the key relationship under

investigation in the Brattle studies), as well as expertise in evaluating the meaning of measures of the uncertainty in these estimates due to sampling variation, including *p*-values, significance levels, and confidence bounds. This expertise and the resulting opinions are general to statistical inference using linear regression models. It is not necessary to know anything about accounting, utility regulation, or decoupling to make statistically correct statements about these regression results.

Q: Please summarize your findings regarding the statistical claims of Dr. Vilbert.

A: Dr. Vilbert claims that although the estimates in Brattle's models indicate that decoupling substantially lowers the cost of capital, these result should either be ignored or reversed because they generally fail to reach the 0.05 level of statistical significance conventionally used in the sciences to determine whether a result found in a sample (here, a specific set of utilities) can be reliably inferred to exist in the general population. I have three main conclusions regarding Dr. Vilbert's testimony on this point.

First, the key findings in the Brattle Group's research – the statistical evidence at the heart of Dr. Vilbert's testimony – *support* the contention that decoupling is associated with lower costs of capital in the electric and gas utilities, with stronger and more consistent evidence regarding the electric utilities. Most importantly, across every model considered, the most likely effect of decoupling on the cost of capital in the electric utility industry is a substantial reduction of between 25 and 49 basis points. The only caveat offered by the Brattle Group and

Dr. Vilbert in presenting these findings is that they do not reach levels of statistical significance conventionally used in academic research. Nevertheless, as I explain later, a strong argument can be made that the fact-finder in an adjudicatory proceeding evaluating the preponderance of the evidence should consider statistical evidence below 95 percent confidence when no evidence above 95 percent confidence is available. The preponderance of empirical evidence in the Brattle analyses on electric utilities shows that decoupling lowers the cost of capital, and this evidence is consistent across a series of models and datasets. The Brattle analyses on gas utilities provide weaker and more mixed evidence, but still tend to point towards decoupling lowering the cost of capital. On balance, regardless of the conclusions of Dr. Vilbert and the Brattle Group about their research, the empirical evidence they produce should count as moderately strong evidence in favor of the conclusion that decoupling lowers the cost of capital.

Second, although it is my opinion that the Brattle Group's models provides policy makers and fact finders, such as the Commission, useful empirical evidence that decoupling lowers the cost of capital, should the Commission decide to employ in its decision the stringent standards of statistical inference used to decide whether to publish scientific research in journals, then the Commission has a choice: if the Commission accepts the Brattle Group's March 2014 models as the best available models of the electric utility industry, then it can still conclude a preponderance of statistical evidence supports the conclusion that decoupling reduces the cost of capital, as two of these models show

statistically significant evidence of a kind frequently publishable in social science journals.[1] In one case only can the Commission conclude that the Brattle studies on the electric industry fail to give evidence that decoupling reduces the cost of capital – if the Commission decides it prefers the more recent November 2014 model of the electric industry *and* insists on high levels of statistical significance. In that special case only, the Commission would not be able to draw any conclusion from the statistical evidence provided by Dr. Vilbert and the Brattle Group, and would need to look elsewhere for evidence.

Third, the Brattle studies provide no evidence at all *against* the claim that decoupling lowers the cost of capital in the electric utility industry. A scientific journal would be highly unlikely to publish the claim that the Brattle results constitute evidence that decoupling has no effect on the cost of capital. In statistical inference, the absence of evidence is not the same as evidence of absence, and results that point towards a claim but fail to reach the desired level of statistical significance are not evidence against that claim.

To better understand these three conclusions, it helps to understand what scientists mean by such technical terms as "*p*-values", "statistical significance", and "confidence bounds". Furthermore, it is important to distinguish how scientists construct standards of evidence using these concepts in order to make scientific publishing decisions, and how the decision to publish a scientific result in a journal differs from a decision in a regulatory or administrative proceeding.

[1] It is worth noting that all statistical evidence is a kind of empirical evidence.

II. UNDERSTANDING THE LEGAL IMPLICATIONS OF STATISTICAL EVIDENCE

Q: Dr. Vilbert's testimony uses linear regression to report the point estimate of the effect of decoupling on the cost of capital. What is a point estimate, and what does it tell us?

A: Linear regression – the tool used in all of the Brattle Group's cited studies on decoupling and the cost of capital – allows us to estimate the relationship between two variables while controlling for other factors. The Brattle Group uses linear regression to estimate the relationship between two things: the degree of decoupling across the units of a holding company, and the cost of capital for that holding company. To help explain this tool, I will use as an example Dr. Vilbert's preferred regression model of the electric industry, the November 2014 model of the industry using data through the second quarter of 2014, which reports a point estimate of -26 basis points and a *p*-value of 0.83. The purpose for choosing this example is not to claim it is the best model offered by Brattle – I make no claim one way or the other on which model of the electric industry is the best one – but to show that even under Dr. Vilbert's preferred model, the Brattle Group's statistical evidence supports the claim that decoupling reduces the cost of capital.

In the Brattle Group's linear regressions, a negative point estimate is evidence that decoupling reduces the cost of capital by the stated number of basis points; the more negative the point estimate, the stronger this effect. The point estimate is the most likely effect of decoupling on the cost of capital and is therefore the most important substantive piece of information provided by the regression model. In the Brattle Group's November 2014 model, the point

estimate of -26 basis points suggests that if this model is accepted as the most reliable model, the most likely effect of decoupling is to reduce the overall cost of capital by 26 basis points.

Q: Under what conditions can we trust the point estimate from a linear regression to measure the true relationship between decoupling and the cost of capital?

A: The reliability of linear regression evidence depends on four things: first, whether the data collected by the Brattle Group are representative of utility firms generally, rather than intentionally or unintentionally "cherry-picked"; second, whether the measures Brattle constructs for cost of capital and decoupling are sound; third, whether the model chosen by Brattle is an appropriate model of the relationship between cost of capital and decoupling; and fourth, whether we can trust the relationship found within the studied sample of firms to hold in general for all utility firms.

Q: What are *p*-values, and how do they measure the reliability of statistical evidence?

A: *p*-values are a tool for assessing only the fourth concern about the reliability of statistical evidence given in the previous answer: whether we can trust the result in the sample of available data to hold in the general population. They do not capture uncertainty due to errors in the collection of data, the measurement of variables, or the construction of the model. To compute a *p*-value for the estimated relationship between the cost of capital and decoupling, we first decide whether we need a two-tailed test or a one-tailed test. A two-tailed test would ask

whether the true effect of decoupling lies either far above or far below a specific value, without specifying in which direction the true effect lies (positive or negative). This is not the appropriate test in our case. Because we want to know whether the effect of decoupling is positive or negative, a one-tailed (or directional) *p*-value is appropriate.[2] This is the kind of test endorsed by Dr. Vilbert.[3]

The second step in computing a *p*-value is to choose the "null hypothesis" (often abbreviated as a "null"). It is common (but not universal) in the sciences to choose zero to be the null hypothesis for two-tailed tests. However, for a one-tailed test, the null hypothesis is not just a specific value, such as zero, but a direction, such as "zero and higher" or "zero and lower". It is thus not possible to choose a "neutral" null hypothesis for a one-tailed test. The Brattle Group chooses as its null hypothesis that decoupling *does not lower* the cost of capital (the null is that the decoupling coefficient is zero or positive), which they attempt to reject in favor of the alternative hypothesis that decoupling *does lower* the cost of capital (the alternative hypothesis is that the decoupling coefficient is negative).

If we choose a null of no negative effect, and then estimate a negative relationship between cost of capital and decoupling, as in the Brattle Group's studies, we can use a *p*-value to assess how confident we are that the negative

[2] For an explanation of the difference between one-tailed and two-tailed tests for regression coefficients, see Paul M. Kellstedt and Guy D. Whitten, 2013, *The Fundamentals of Political Science Research*, 2nd Edition, New York: Cambridge University Press.
[3] See Dr. Vilbert's testimony (Exhibit No. MJV-1TC), p. 21: "We use a so-called 'one tailed test' because the alternative hypothesis is that the effect on the cost of capital is negative."

estimate within the sample is evidence of a general negative relationship across all firms. The *p*-value tells us how likely we would be to misleadingly estimate as strong a relationship in our sample as the one we observed, even in the case where the null hypothesis is correct and decoupling has no effect or raises the cost of capital. In general, statisticians tend to trust estimates with smaller *p*-values more than estimates with large *p*-values.

Q: How do *p*-values relate to the concept of statistical significance?

A: A significance test involves the comparison of the *p*-value from a regression to an arbitrary standard. In the sciences, for example, many journals are reluctant to publish conclusions based on *p*-values greater than 0.05. Estimates with *p*-values equal to or less than 0.05 are said to be "significant at the 0.05 level". The 0.05 standard is arbitrary, and other thresholds are possible. For example, an estimate with a *p*-value less of 0.1 or less would be "significant at the 0.1 level", and an estimate with a *p*-value of 0.17 or less would be "significant at the 0.17 level".

Q: Can *p*-values tell us how much to trust results from a linear regression that uses a flawed sampling technique, uses poor measures of variables, or omits statistically important variables?

A: No. *p*-values, as well as significance tests and confidence bounds, are computed under the assumption that the chosen model adequately captures the process under study and that the data collected are meaningful, well measured, and sampled without bias from the population. If one disagrees with the assumptions of a model, or considers the data used to be suspect, there is no obvious way to correct *p*-values (or significance tests or confidence bounds) to account for these doubts;

indeed, it is often unclear whether *p*-values should go up or down in the presence of such problems.

Q: What are confidence bounds, and how do they relate to *p*-values and statistical significance?

A: If we consider many separate studies of the effect of decoupling on the cost of capital, wherein each study involves a different sample of data and reports its own estimate and 95 percent confidence bound (which corresponds to a *p*-value of 0.05), and we trust that these studies make good assumptions about the studied process and use well-measured data, then we can say the following: the true population effect of decoupling on the cost of capital lies inside the reported 95 percent confidence bound in 95 percent of all studies considered.

Confidence bounds can help us understand the consequence of believing statistical evidence at a given level of confidence (or, equivalently, with a given *p*-value). If we decide to always believe values with 95 percent confidence intervals that stop short of the null hypothesis, we can be confident that 95 percent of the time, the statistical evidence correctly identifies the null hypothesis as false. Thus our acceptance of these results will lead us to incorrectly reject the null hypothesis due to unusual samples only 5 percent of the time.

Confidence bounds can be computed for any desired level of confidence; for example, an 83 percent confidence interval (equivalently, a *p*-value of 0.17) allows us to be confident that we will be correct in rejecting the null hypothesis 83 percent of the time. Confidence intervals also allow us to see more clearly the range of null hypotheses we can reject, as any value outside the confidence bound

can be rejected at the given level of confidence.

Confidence bounds are an alternative representation of *p*-values. They rely on the same statistical concepts and assumptions as *p*-values. Explaining the meaning of a *p*-value through the equivalent confidence bound introduces no additional statistical assumptions.

Q: How do 95 percent confidence bounds and 83 percent confidence bounds help us understand the effect of decoupling on the cost of capital?

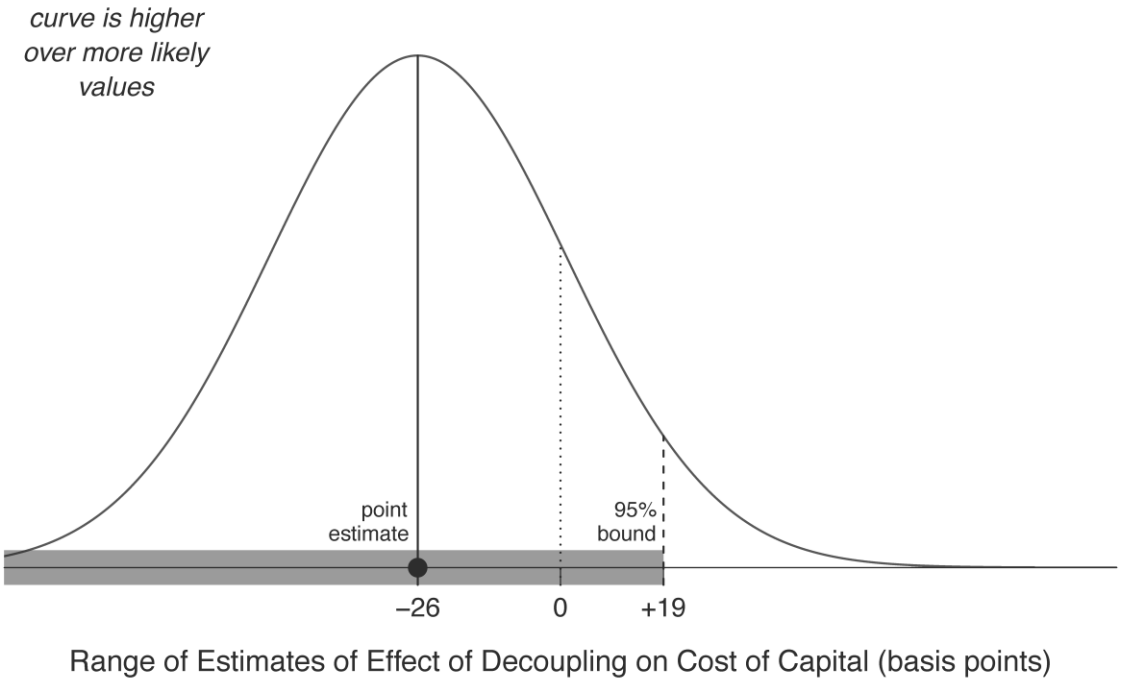
A: I will provide a somewhat technical description of what confidence bounds mean in this case, and follow it with a more intuitive explanation.

According to the Brattle Group's November 2014 analysis of the electric industry – Dr. Vilbert's preferred analysis – the effect of decoupling is estimated to lower the overall cost of capital by 26 basis points, with a one-sided *p*-value of 0.17.[4] Figure 1 represents this case graphically as a curve called the "sampling distribution", which is the probability concept underlying the computation of *p*-values. This curve is tallest over the point estimate of -26, because this is the most likely estimate of the effect of decoupling. But we are uncertain of what relationship holds between decoupling and the cost of capital in the population of all utilities. It is possible, but less likely, that the true effect of decoupling either more strongly reduces the cost of capital (as shown by the left "tail" of the curve)

[4] These are the results of Dr. Vilbert's updated analysis of decoupling in electric utilities, Exhibit No. MJV-12 and Exhibit No. MJV-13.

or that decoupling slightly raises the cost of capital (as shown by the right "tail" of the curve).

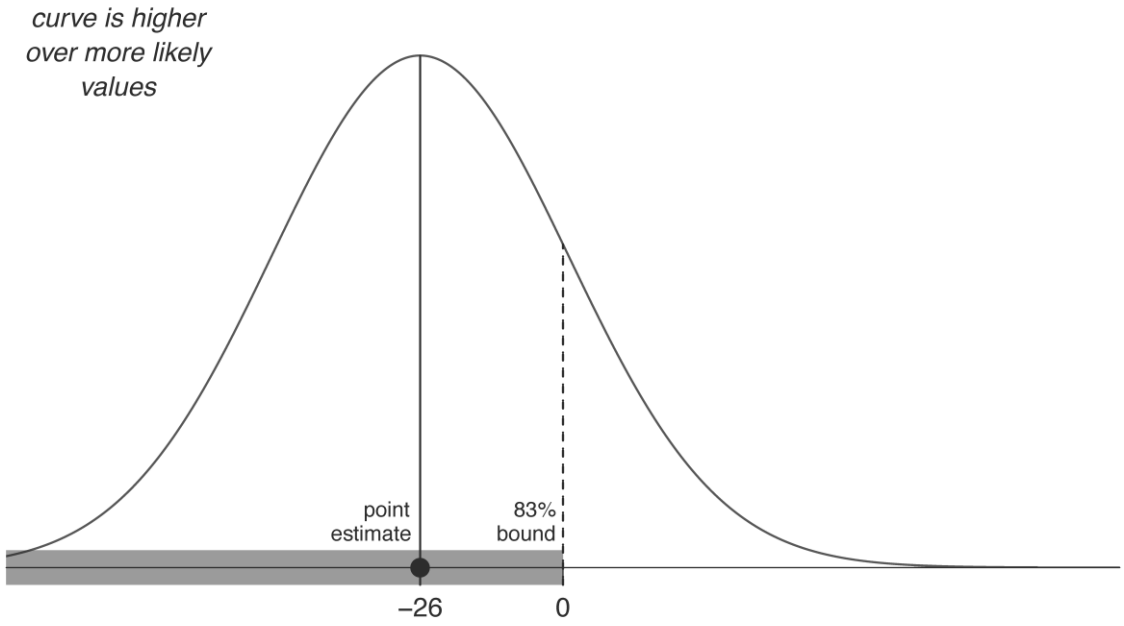
FIGURE 1: INTERPRETATION OF A 95 PERCENT CONFIDENCE BOUND USING DATA FROM BRATTLE'S NOVEMBER 2014 FULL ANALYSIS OF DECOUPLING IN ELECTRIC UTILITIES



Because the reported *p*-value is greater than 0.05, the 95 percent confidence bound in this case covers zero and extends all the way up to +19 basis points. So in this case, we can reject with 95 percent confidence a null hypothesis that claims decoupling raises the cost of capital by +19 basis points or more. But we cannot reject with 95 percent confidence the null hypothesis that decoupling raises the cost of capital modestly, somewhere between 0 and +19 basis points. If we decided Brattle's November 2014 model is the best model, and insisted on a minimum of 95 percent confidence (which I argue is unnecessary in this case), we

would not be able to decide whether decoupling lowers or raises the cost of capital from the confidence bound in Figure 1, or from the equivalent *p*-value of 0.17.

FIGURE 2: INTERPRETATION OF AN 83 PERCENT CONFIDENCE BOUND USING DATA FROM BRATTLE'S NOVEMBER 2014 FULL ANALYSIS OF DECOUPLING IN ELECTRIC UTILITIES



However, if we are willing to consider evidence that falls short of 95 percent confidence, we *can* draw useful conclusions about the effect of decoupling on the cost of capital, even in Dr. Vilbert's preferred model. Figure 2 shows results from the same regression model, but uses an 83 percent confidence interval. At the 83 percent level, we can reject the null hypothesis that decoupling raises or has no effect on the cost of capital. The upshot of this 83 percent confidence bound is the following: if, as a matter of practice, a judge of evidence took statistical significance at the 0.17 level as evidence in favor of an empirical claim, then 83

percent of the time, that conclusion would be sound, and 17 percent of the time, it would be mistaken, simply due to misleading samples.

Q: Can confidence bounds be explained in plainer language?

A: Yes. We want to know as a general matter whether decoupling leads to a lower cost of capital on average, but we do not have data on all utility companies and time periods. Instead, we have a limited sample of companies observed over a limited set of periods. In that sample, the regression coefficients estimated by the Brattle Group are negative, showing that decoupling is associated on average with overall lower costs of capital for both electric and gas utilities.

But we have a source of doubt: this limited sample of companies and periods may, just by luck of the draw, look different from the general population of utility companies and time periods. Two factors – which may or may not hold here – can make it more likely that the effect of decoupling in our sample adequately represents the general population. First, the larger the sample, the less often it will mislead. Second, if the true effect of decoupling is strong, smaller samples will more consistently detect it. Casual observation cannot tell us with certainty whether our sample size is large enough to reliably measure population effects, so we must turn to statistical inference to quantify our doubt. Statistical inference gives us tools to protect ourselves against the possibility of drawing the wrong conclusions about populations from limited samples. The confidence bound is one such tool.

In Dr. Vilbert's preferred model of the electric industry, the Brattle Group obtains a point estimate of the effect of decoupling on the cost of capital that is -

26 basis points, so that decoupling appears to lower the cost of capital in the sampled utilities. The 83 percent confidence bound on this estimate is at zero. What does this tell us about the effect of decoupling on the cost of capital for all utility firms, not just those in our sample?

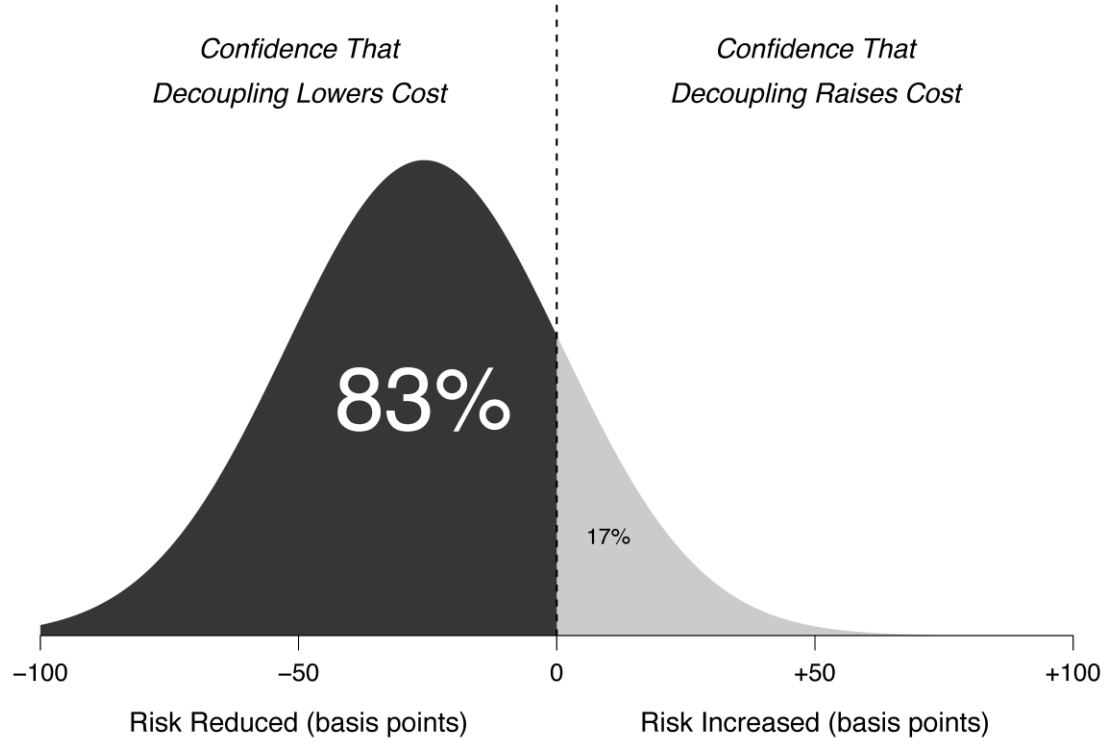
It helps to start with what confidence bounds *don't* tell us.

The confidence bound does *not* mean that there is an 83 percent chance that the effect of decoupling is negative (zero or less). In statistics, there is no objective way to calculate the probability that an effect holds in the population based on a single sample.

The confidence bound does *not* mean that in 83 percent of future studies, the effect of decoupling will be found to lie within the current confidence bound, although this is a common misinterpretation.

What this confidence bound means is this: the truth lies inside the reported 83 percent bound in 83 percent of all studies, but we don't know which 83 studies out of every 100 studies are correct, and which 17 studies out of every 100 studies are misleading. If we decide to believe in a negative effect only when the 83 percent confidence bound is zero or less, then misleading samples would only lead us to incorrectly conclude there was a negative effect 17 percent of the time, while we would correctly identify negative effects 83 percent of the time.

FIGURE 3: CONFIDENCE FOR AND AGAINST A COST REDUCTION FROM DECOUPLING USING DATA FROM BRATTLE'S NOVEMBER 2014 FULL ANALYSIS OF DECOUPLING IN ELECTRIC UTILITIES



This state of affairs is shown in Figure 3, which presents the sampling distribution for the cost of capital in Dr. Vilbert's preferred model of decoupling in the electric industry. Clearly, the majority of the mass of this distribution lies to the left, over the range of negative effects of decoupling on the cost of capital. This reflects that fact that if we treated evidence like that provided by this sample as sufficient to prove that decoupling lowers the cost of capital, we would be correct about the general population 83 percent of the time in similar cases (provided, of course, our model was appropriately designed and our data appropriately measured). But if we insisted this confidence interval actually provided evidence that the cost of capital stayed the same or rose due to decoupling, we would be

right in only 17 percent of similar cases. In the other 83 percent of cases, we would have gotten the evidence backwards.

Q: If an 83 percent confidence bound provides useful information to decision makers, why do scientific journals tend to require higher levels of confidence before publishing studies, and why do they often require 95 percent confidence (or equivalently, significance at the 0.05 level)?

A: Space in scientific journals is limited, in part because the attention of the scientific community is limited. Journals tend to only publish research based on samples sufficiently large as to allow very high levels of confidence. If the best available study on topic *X* employs a sample that is not sufficiently large to produce a statistically significant result, many journals would urge the study's authors to collect more data (in order to reach 95 percent confidence) before submitting the paper for publication. Typically, journals do not need to publish a paper on topic *X* in any given time frame; instead, most journal editors would prefer to wait until studies of topic *X* meet the conventional standards of scientific publication. This is because science is mainly concerned with the production of knowledge, rather than making timely judgments. Requiring high confidence (or setting a tough significance test) is one way in which many scientific journals seek to encourage the production of highly reliable knowledge.

A journal that declines to publish a paper with an 83 percent confidence bound or a *p*-value of 0.17, is not deciding in favor of the null hypothesis: the journal editor is merely deciding to wait until sufficient data are available to devote scarce journal attention to the topic. The journal is making no judgment at

all about the conclusions of the research, merely about whether there is yet sufficient confidence to publish a scientific paper. The journal has the luxury of waiting years to publish a particular paper, because academic debates are rarely urgent. If academic journals were deciding policy, rather than publishing papers, they might have a different standard urging consideration of the most reliable currently available evidence, but journals are not policy makers.

Scientific journals tend to require low *p*-values for publication of statistical results. Some, but by no means all, journals also require *p*-values to fall below a particular value. The choice of a specific threshold *p*-value is arbitrary. There is nothing special about any particular significance threshold -- estimates with a *p*-value of 0.051 are essentially as trustworthy as estimates with *p*-values of 0.049. Instead, the selection of a significance threshold simplifies decisions about what statistical evidence to accept. [5]

Q: Do statisticians and scientists consider evidence that does not reach 95 percent confidence?

A: Quite a few scientists – myself included – think that arbitrary significance thresholds are too limiting, and that it is always valuable to consider a variety of confidence levels before drawing a conclusion from a sample.

Support for this view can be found in mainstream introductory statistics

[5] One introductory statistics text notes that "[a]pplied statisticians increasingly prefer *p*-values to classical testing because classical tests involve setting α [the significance level] arbitrarily (usually at 5 percent). Rather than introduce such an arbitrary element, it is often preferable just to quote the *p*-value, leaving the readers to pass their own judgment on H_0 [the null hypothesis]." Thomas H. Wonnacott and Ronald J. Wonnacott. 1990. Introductory Statistics. Fifth Edition. New York: Wiley, p. 302.

textbooks. In his popular econometrics text, Damodar Gujarati advises that:

> It is better to give up fixing α [the significance level] arbitrarily at some level and simply choose the *p* value of the test statistic. It is preferable to leave it to the reader to decide whether to reject the null hypothesis at the given *p* value. If in an application the *p* value happens to be, say, 0.145, or 14.5 percent, and if the reader want[s] to reject the null hypothesis at this (exact) level of significance, so be it. Nothing is wrong with taking a chance of being wrong 14.5 percent of the time if you reject the null hypothesis.[6]

Gujarati also advises presenting confidence bounds rather than significance tests to make it easier for readers to select their own desired level of confidence; in their introductory statistics text, David S. Moore and George P. McCabe agree.[7]

Q: Is the choice of a null hypothesis neutral, or does this choice affect whether an estimate is statistically significant?

A: The choice of the null hypothesis is important and far from neutral, especially for directional hypotheses such as whether decoupling has a positive or negative effect on the cost of capital. The Brattle Group chooses its null hypothesis to capture the possibility that the effect of decoupling on the cost of capital is *not* negative. In their updated study of electric utilities, they find that this null hypothesis can only be rejected with 83 percent confidence, and conclude that this is insufficient evidence to conclude decoupling has a negative effect under the 95 percent standard (or equivalently, at the 0.05 level of significance).

But suppose we choose a different null hypothesis: suppose our new null hypothesis is that the effect of decoupling was negative (that is, that it reduced the

[6] Damodar N. Gujarati, 1995, *Basic Econometrics*, 3rd Edition, New York: McGraw Hill, p. 133.
[7] Damodar N. Gujarati, 1995, *Basic Econometrics*, 3rd Edition, New York: McGraw Hill, p. 134; David S. Moore and George P. McCabe, 2006, *Introduction to the Practice of Statistics*, 5th Edition, New York: W. H. Freeman and Co, p. 425.

cost of capital). This is the mirror image of Brattle's null hypothesis. Choosing a different null radically changes our conclusion at the 0.05 significance level: now we find that we cannot reject the possibility that the effect of decoupling is negative. The *p*-value for this test is 0.83, so we can "reject" the null of a negative effect only at the 17 percent level. (To see this, note that in Figure 3, only 17 percent of the area of the sampling distribution lies on the positive side of zero.) This is a standard of confidence that essentially no scientist would find adequate to draw a conclusion in favor of a zero or positive effect of decoupling on the cost of capital.

Because rejecting the null is intentionally difficult to do under conventional scientific standards, it is possible to stack the deck in favor of your preferred conclusion by making that preferred conclusion the null hypothesis. Then instead of being a conservative scientific tool, the machinery of statistical significance will be biased towards your preferred conclusion. In science, this is avoided in two ways: first, a scientific paper is not allowed to impose as the null hypothesis the very claim being considered. Second, and even more important, failure to reject the null in a scientific paper is not treated by scientists as evidence that the null is correct. If one adopts the 95 percent standard, we can draw no conclusions in either direction from a *p*-value of 0.17 or an 83 percent confidence interval. If we accept a lower standard of evidence, such as 83 percent confidence, then the evidence from the sample of data points towards the conclusion that decoupling lowers the cost of capital for electric utilities.

However, there is no interpretation under either standard that favors the conclusion that decoupling has no effect on the cost of capital.

Q: How should a fact-finder seeking to establish the preponderance of evidence interpret *p*-values and confidence bounds?

A: As discussed above, scientists with the luxury of waiting for more data might refrain from publishing a conclusion in a scientific journal until such time as sufficient evidence arises to meet scientific convention. Only when there is a great reluctance to make a mistake in a particular direction – such as in criminal cases – are the needs of courts and scientists similar. The analogy is this: just as criminal courts require prosecutors to meet a high burden of proof ("beyond reasonable doubt") in order to protect defendants from wrongful punishment, so do scientific journals impose a special burden on scientists to publish new findings. Lower legal standards of proof, such as a preponderance of evidence, correspond to lower required levels of statistical significance. It is my opinion that if a fact-finder trusts the data and model which produced a statistical estimate, then demanding a 95 percent confidence level would be analogous to demanding proof beyond a reasonable doubt, placing a much heavier burden of proof on whichever side is asked to reject the null hypothesis.

III. UNDERSTANDING THE BRATTLE GROUP'S ELECTRIC UTILITY STATISTICAL RESULTS

Q: What are Dr. Vilbert's statistical claims about the relationship between decoupling and the cost of capital for electric utilities?

A: Dr. Vilbert cites results from six different models estimating the relationship between decoupling and the cost of capital in the electric industry. In each case,

he notes that the cost of capital is estimated to fall as the degree of decoupling rises.[8] But he argues that because each of these estimates fails to meet the conventional scientific threshold of statistical significance at the 0.05 level, these results do not support the claim that decoupling lowers the cost of capital.

Q: What are the six models of the cost of capital for electric utilities cited by Dr. Vilbert?

A: The six models cited by Dr. Vilbert come from two studies of the electric utility industry conducted by the Brattle Group. The most recent study, from November 2014, presents two models. The first of these recent models, which I label "Brattle Nov-2014 Full", estimates a linear regression of the cost of capital as a function of an index of decoupling, while controlling for the period and for each combination of a firm and an "epoch". Standard errors in this model are adjusted to account for clustering by firm, and the data analyzed includes fourteen holding companies over the period 2005Q1 to 2014Q2. The second model, which I label "Brattle Nov-2014 Limited", is the same in all respects, except that it considers a narrower range of data, from 2005Q1 to 2012Q4.[9]

Brattle's earlier study from March 2014 presents four models. These

[8] Regarding the most recent and complete model, Dr. Vilbert's testimony (page 28) indicates Brattle found a negative coefficient: "What are the results of the electric industry decoupling study? ... The statistical results do not reject the neutral hypothesis that there is no impact on the cost of capital from adoption of decoupling. The results are shown in Exhibit No. MJV-12. The coefficient of the decoupling index variable is -26 bps." Dr. Vilbert reports a similar result for a second Brattle model on page 28 of his testimony: "These results are similar to the results for the full period to 2014. The coefficient on the decoupling index is -25 bps." Finally, regarding the baseline case among four models from the earlier Brattle study, Dr. Vilbert again notes a negative point estimate of decoupling's effect on the cost of capital (page 29): "The p-value in the original study was 0.14 and the coefficient was -41 bps."

[9] These two models are discussed in Michael J. Vilbert's testimony. The "Brattle Nov-2014 Full" estimates appear in Exhibit No. MJV-12, and the "Brattle Nov-2014 Limited" estimates appear in Exhibit No. MJV-13. The studies are referred to in this testimony as November studies because of the date these results were provided, although it appears they were conducted in October 2014.

models differ from the November 2014 models in a number of respects, including the number of quarters of data used, the method of computing the variables, and the use of firm dummies (which control for the many ways in which firms differ from each other) rather than firm-epoch dummies (which also control for unmeasured changes over time in the character of firms).[10] The four March 2014 models differ from each other in a further respect: whether they assume the cost of capital is immediately affected by changes in decoupling (as assumed in Brattle's later November 2014 models), or whether the effect of decoupling is assumed to occur at a lead of one, two, or three quarters. There are thus four different models in the March 2014 report: "Brattle Mar-2014 Immediate", "Brattle Mar-2014 Lead-1", "Brattle Mar-2014 Lead-2", and "Brattle Mar-2014 Lead-3".[11]

Q: Were you able to replicate these models from Brattle's data?

A: As the first step in producing the figures presented in this report, I was able to replicate the November 2014 models using data provided by Dr. Vilbert through the discovery process. I used the statistical package R (version 3.0.2) to conduct my replication and obtained model coefficients and *p*-values almost exactly the same as those reported in Brattle's own study. I did not have the opportunity to

[10] Specifically, according to the PSE's response to Public Counsel/ICNU Data Request No. 4: "The differences between the March 2014 electric study presented here and the new October 2014 electric study presented in PSE's Response to Public Counsel/ICNU Data Request No. 003 include: not having 6 more quarterly observations on all variables, using single-stage DCF, not having fixed-variable rates as decoupling, and that the company indicator variable is not defined for the epochs." As noted in note 9 above, the study referred to in the Response to Public Counsel/ICNU Data Request No. 4 as "the new October 2014 electric study" is referred to in this testimony as the November 2014 study

[11] These four models are presented in the Brattle Group's March 2014 study "The Impact of Revenue Decoupling on the Cost of Capital for Electric Utilities: An Empirical Investigation", Michael J. Vilbert et al, 2014, which is included in the record as Exhibit No. SGH-16.

replicate the March 2014 models, because it was unclear how to construct the variables and cases for analysis from the Excel workbook provided. However, I do not have any reason to doubt that the March 2014 models were correctly estimated by the Brattle Group.

Q: For models you did not replicate, can you interpret the meaning of *p*-values and confidence bounds?

A: Yes. Brattle's March 2014 working paper provides sufficient information to interpret the confidence bounds and *p*-values of the estimated relationship between the cost of capital and decoupling.

Q: The main result from the "Brattle Nov-2014 Full" model is that decoupling lowers the cost of capital by 26 basis points, with a *p*-value of 0.17. What does this mean?

A: This result – Dr. Vilbert's preferred model – was used as the running example in Part II of this testimony.[12] As stated there, we can conclude with 83 percent confidence that decoupling lowers the cost of capital for electric utilities. If we decide to believe in a negative effect when the 83 percent confidence bound is zero or less, then misleading samples would only lead us to incorrectly conclude there was a negative effect 17 percent of the time.

Q: In the "Brattle Nov-2014 Full" model, the coefficient for decoupling was negative, but not statistically significant at conventional levels. Is this

[12] Dr. Vilbert appears to prefer the two models of the electric industry he produces in his November 2014 update of the Brattle Group's March 2014 working paper. These models differ only in whether data after 2012 Q4 is included in the analysis, and the two models yield extremely similar results. To simplify exposition, the model used as an example in this section is the November 2014 model using data through

evidence in favor of Dr. Vilbert's null hypothesis that decoupling does not lower the cost of capital?

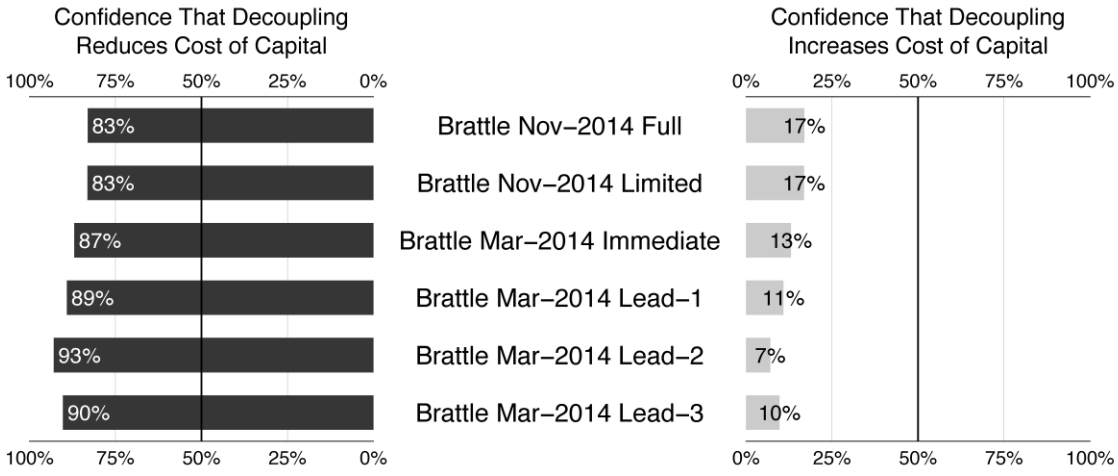
A: No. Statistical significance is meant to be a difficult test for a sample to meet and is purposefully designed to minimize the chance of falsely rejecting the null hypothesis (here, the Brattle Group null hypothesis that decoupling does not reduce the cost of capital), even at the risk of failing to conclude from a sample that decoupling lowers costs when that is the true effect of decoupling in the population. This creates an asymmetric burden of proof. In scientific publishing, this asymmetry places a heavy burden of proof on scientists seeking to publish new claims, at the expense of ignoring scientific studies with less that highly significant evidence against the null hypothesis. As a result, under the strict scientific use of statistical significance as a standard of evidence, it is well known that failure to reject the null hypothesis cannot be taken as evidence that the null hypothesis is true.

Dr. Vilbert's report reveals confusion regarding this basic principle. On page 17 of his testimony,[13] Dr. Vilbert argues that "[i]f decoupling substantially reduced the cost of capital, the coefficient on the decoupling-index variable would be negative and statistically significant." This is not necessarily the case.[14] Because statistical significance is a heavy burden, it is quite possible to obtain a non-significant or near significant negative coefficient in a sample when the

2014 Q2 (what is referred to later in this testimony as "Brattle Nov-2014 Full", but the conclusions in this section would not substantively change were the example instead the model using data through 2012 Q4.
[13] Direct Testimony of Dr. Michael J. Vilbert, Exhibit No. MJV-1T.

population value is in fact negative. The conclusion to be drawn from Dr. Vilbert's report is only that it is not possible to conclude with 95 percent confidence that decoupling reduces the cost of capital. But, as discussed above, one could still conclude that decoupling reduces the cost of capital if one were willing to accept a lower confidence level.

FIGURE 4: CONFIDENCE REGARDING THE EFFECT OF DECOUPLING ON THE COST OF CAPITAL ACROSS MODELS OF THE ELECTRIC INDUSTRY



Q: How do the results from the other five models of the cost of capital for electric utilities compare to the "Brattle Nov-2014 Full" model?

A: Figure 4 collects the *p*-values for each of Brattle's six models of decoupling for the electric industry; note that these are the same quantities as the percentages shown in Figure 3, so this figure summarizes our confidence that decoupling lowers the cost of capital (on the left) and our confidence that decoupling raises the cost of capital (on the right) for each of the six electric utility models.

[14] The converse of Dr. Vilbert's statement is true under certain conditions: if the coefficient on the decoupling index is negative and significant, then (if we believe the model is correctly specified and the data correctly measured) we could take this result as evidence that decoupling reduced the cost of capital.

Recall that these models differ in the range of data analyzed and in the treatment of delays in the effect of decoupling on the cost of capital, as well as in several other respects. Yet across all of these models, the estimated effect of decoupling is *always negative*. The confidence in this effect, shown at the left side of Figure 4, varies from 83 percent (in Brattle's most recent analysis) to 93 percent (in Brattle's earlier analysis).

The earlier models from Brattle's March 2014 report yield notably stronger evidence that decoupling reduces the cost of capital in the electric utilities industry. In these four models, the range of point estimates suggests the most likely effect of decoupling is to lower the cost of capital by between 41 and 49 basis points, a highly consistent set of estimates. The range of confidence is between 87 percent and 93 percent, with two models reaching statistical significance at the 0.1 level. It is worth noting that many social science journals are willing to publish evidence at the 90 percent confidence level (or 0.1 significance level). Thus the results from the last two models (specifically, those models from the Brattle Group March 2014 report controlling for a two or three quarter lead in the effect of decoupling) would be publishable in many social science journals as statistically significant evidence that decoupling reduces the cost of capital.

Q: How should the Commission interpret the statistical evidence of decoupling's effect on the cost of capital in the electric utility industry?

A: In his testimony, Stephen G. Hill argues that these early models from Brattle's March 2014 analysis use more reliable data and assumptions. If the Commission

finds these models are the best available for the electric utility industry, we can conclude there is relatively strong statistical certainty of a large substantive reduction in the cost of capital in the electric industry under decoupling. These four models estimate a reduction in the cost of capital on the order of -41 to -49 basis points, and have confidence in the neighborhood of 90 percent. From a statistical point of view, assuming we trust the March 2014 Brattle Group model, it is my opinion that a preponderance of the statistical evidence supports the claim that decoupling lowers the cost of capital in the electric utility industry.

If the Commission decides that the more recent November 2014 analyses presented in Dr. Vilbert's direct testimony in this proceeding are more reliable, it is still the case that a preponderance of the statistical evidence favors the claim that decoupling reduces the cost of capital in the electric industry. Suppose the Commission saw a series of 100 cases over time with evidence of this kind – results which are significant at the 0.17 level. If the Commission took the side with the preponderance of evidence in each of those cases, then it would be right on the facts in 83 cases, and wrong due to sampling error in 17 cases. From a statistical point of view, assuming we trust the Brattle Group's models and variables, it is my opinion that a preponderance of the statistical evidence supports the claim that decoupling lowers the cost of capital in the electric utility industry.

But even if the Commission decides to trust only statistical evidence significant at the 0.05 level – a standard I consider inappropriately high where the standard of proof is the preponderance of evidence – there is no reason to treat

Brattle's results on the electric industry as evidence *against* the claim that decoupling lowers the cost of capital. Given the consistency and relatively high *p*-values of many of these results, it would be unusual for a scientific publication to describe the range of results produced by Brattle as supporting that conclusion. As the right side of Figure 4 makes clear, our confidence that decoupling does not have a negative effect on the cost of capital is extremely low – less than 20 percent in every model.

IV. UNDERSTANDING THE BRATTLE GROUP'S GAS UTILITY STATISTICAL RESULTS

Q: What are Dr. Vilbert's statistical claims about the relationship between decoupling and the cost of capital for gas utilities?

A: Dr. Vilbert cites results from five different linear regression models estimating the relationship between decoupling and the cost of capital in the gas utility industry.[15] Regarding the most recent analysis of the gas utility industry, conducted by the Brattle Group in 2014, Dr. Vilbert notes that "[t]he coefficient of the decoupling index variable is -8.7 bps"[16]. He argues that because this estimate fails to meet the conventional scientific threshold of statistical significance at the 0.05 level, this result does not support the claim that decoupling lowers the cost of capital. Regarding a set of four earlier linear regression models from a 2012 Brattle Group analysis, Dr. Vilbert argues that "[t]he qualitative conclusion is the same: no statistical evidence to disprove the

[15] Dr. Vilbert also cites an early Brattle Group t-test analysis that I have not replicated or studied. In the analysis of panel data, a simple t-test does not provide sufficient controls for unmeasured differences across firm and periods. Dr. Vilbert cites no t-test from the more recent Brattle study of the gas industry.
[16] Direct Testimony of Dr. Michael J. Vilbert, Exhibit No. MJV-1T, p. 20.

neutral hypothesis of no impact". [17]

Q: What are the five models of the cost of capital for gas utilities cited by Dr. Vilbert?

A: The five models cited by Dr. Vilbert come from two studies of the gas utility industry conducted by the Brattle Group. From the most recent study, conducted in 2014, Dr. Vilbert cites one model, which I label "Brattle 2014".[18] This model consists of a linear regression of the cost of capital as a function of an index of decoupling, while controlling for the period and firm. Standard errors in this model are adjusted to account for clustering by firm, and the data analyzed includes twelve holding companies over the period October 2005 to May 2012.

Brattle's earlier study from 2012 presents four linear regression models. According to PSE's response to Public Counsel/ICNU Data Request No. 3, these models differ from the 2014 models in a number of respects, including the use of Newey-West standard errors in place of clustered standard errors. The four 2012 models differ from each other in two ways: whether they measure decoupling using a continuous index or a simpler but potentially less informative binary indicator, and whether the model controls for the past levels of the cost of capital (that is, whether the model includes a "lag"). Because the Brattle Group considered each combination of these choices, there are four models: "Brattle 2012, Indicator", "Brattle 2012, Index", "Brattle 2012, Lag & Indicator", and

[17] Direct Testimony of Dr. Michael J. Vilbert, Exhibit No. MJV-1T, p. 22.
[18] The "Brattle 2014" estimates are presented in Exhibit No. MJV-10.

"Brattle 2012, Lag & Index".[19]

Q: Were you able to replicate these models from Brattle's data?

A: Yes. As the first step in producing the figures presented in this report, I was able to closely replicate the "Brattle 2014" using data provided by Dr. Vilbert through the discovery process. I used the statistical package R (version 3.0.2) to conduct my replication and obtained model coefficients and *p*-values almost exactly the same as those reported in Brattle's own study. I was also able to replicate the models from Brattle's 2012 study.

Q: The main result from the "Brattle 2014" model is that decoupling lowers the cost of capital by 8.7 basis points, with a *p*-value of 0.37. What does this mean?

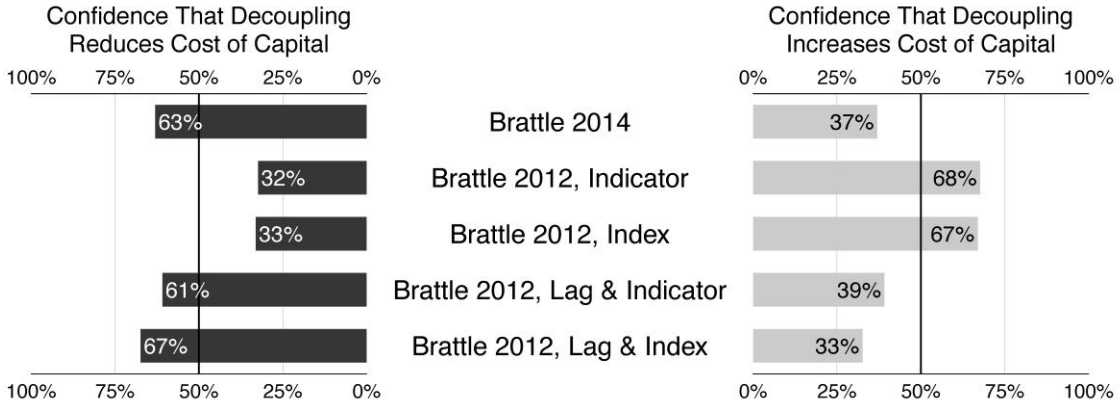
A: Following the logic of confidence bounds, we can conclude with 63 percent confidence that decoupling lowers the cost of capital for electric utilities. If we decide to believe in a negative effect when the 63 percent confidence bound is zero or less, then misleading samples would only lead us to incorrectly conclude there was a negative effect 37 percent of the time. Compared to the evidence for the electric industry, we have less confidence that decoupling lowered the cost of capital based on the gas utility sample.

Q: In the "Brattle 2014" model, the coefficient for decoupling was negative, but not statistically significant at conventional levels. Is this evidence in favor of Vilbert's null hypothesis that decoupling does not lower the cost of capital?

[19] The structure of these four models is described in PSE's response to Public Counsel/ICNU Data Request No. 2.

A: No. As noted above in the discussion of the electric industry results, under the strict scientific use of statistical significance as a standard of evidence, it is well known that failure to reject the null hypothesis cannot be taken as evidence that the null hypothesis is true. This is the case even for the weaker levels of confidence found in the gas utility models.

FIGURE 5: CONFIDENCE REGARDING THE EFFECT OF DECOUPLING ON THE COST OF CAPITAL ACROSS MODELS OF THE GAS INDUSTRY



Q: How do the results from the other four models of the cost of capital for gas utilities compare to the "Brattle Nov-2014 Full" model?

A: Figure 5 collects the *p*-values for each of Brattle's five linear regression models of decoupling for the gas industry. This figure summarizes our confidence that decoupling lowers the cost of capital (on the left) and our confidence that decoupling raises the cost of capital (on the right) for each of the five gas utility regression models.

The evidence from the four earlier models of the gas utility industry is mixed. The two models that do not control for past levels of the cost of capital actually produce point estimates in favor of a positive effect of decoupling on the

cost of capital. The two models controlling for the cost of capital produce point estimates in favor of negative effect of decoupling on the cost of capital, in line with the more recent results from the 2014 Brattle Group study. There was little difference between models that used an index to measure decoupling and those that used a binary indicator.

Two points should be emphasized regarding the gas utility results. First, the two models from 2012 that find evidence that decoupling raises the cost of capital are similar in structure to the 2014 model which, using more data, found weak evidence in the opposite direction. Second, across all five models, the available levels of confidence are low.

Q: How should the Commission interpret the statistical evidence on decoupling's effect on the cost of capital in the gas utility industry?

A: There are two options for interpreting these models, depending on the level of confidence the Commission feels is needed to rely on statistical evidence. First, if the Commission is comfortable accepting the best available statistical evidence, given that such that evidence will tend to be reliable in 63 cases out of 100, and misleading in 37 cases out 100, then the Commission can conclude based on the most comprehensive Brattle Group analysis that there is evidence, albeit weak, in favor of the claim that decoupling lowers the cost of capital in the gas industry. It is my opinion that in the absence of any additional relevant statistical or non-statistical evidence, this weak support for a reduction in the cost of capital constitutes a preponderance of the evidence on the effect of decoupling.

Second, if the Commission is uncomfortable relying on a 63 percent

confidence bound, then by the logic of significance testing, the Commission's only option is to ignore the statistical evidence of decoupling's effect on the gas industry in favor of other kinds of evidence. As noted above, failure to reject a null hypothesis is not evidence in favor of the null hypothesis.

Q: Does this conclude your testimony, Dr. Adolph?

A: Yes, it does.