P E R S P E C T I V E S

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES

The Lisbon Council's 2015 Intellectual Property and Economic Growth Index:

A Showcase of Methodological Blunder

George S. Ford, PhD*

June 29, 2015

Introduction

Last month, the European Commission unveiled its much-anticipated framework to create a "Digital Single Market" in Europe.[1] For one of the hallmark pillars of its Digital Single Market strategy, the Commission stated that it intends to develop a "modern, more European copyright framework."[2] To this end, the European Commission stated that it will "make legislative proposals before the end of 2015 to reduce the differences between national copyright regimes and allow for wider online access to works by users across the EU, including through further harmonisation measures" including, *inter alia*, "modernising enforcement of intellectual property rights, focusing on commercial-scale infringements (the 'follow the money' approach) as well as its cross-border applicability."[3]

Presumably intending to influence this process, the Lisbon Council, a Brussels-based think tank, released in March 2015 (with a revision in May 2015) a study by Benjamin Gibert (a Fellow of the Lisbon Council) entitled *The 2015 Intellectual Property and Economic Growth Index: Measuring the Impact of Exceptions and Limitations in Copyright on Growth, Jobs, and Prosperity*.[4] In this analysis, the Lisbon Council purports to show that weaker copyright protections are good for the economy generally and for *some* industries in particular. The Lisbon Council reaches this conclusion by combining the *IP Index Report's* own index of copyright flexibility for eight countries with numerous measures of economic activity. According to the Lisbon Council, countries "that employ a broadly 'flexible' regime of exceptions in copyright also saw higher rates of growth in value-added output throughout their economy."[5] Consequently, the Lisbon Council encourages European policymakers to increase flexibility in (i.e., weaken) copyright laws to promote economic growth.

> ***[The Lisbon Council's] IP Index Report is a showcase of methodological blunder. As I highlight below, the IP Index Report's strong claims are based on nothing more than "cherry-picking" from a set of hundreds of simple correlation coefficients computed by using no more than 8 and as few as 5 observations.***

While the Lisbon Council's conclusion is a strong one, upon inspection it becomes readily clear it cannot be supported by the evidence in its *IP Index Report*. Indeed, the *IP Index Report* is a showcase of methodological blunder. As I highlight below, the Lisbon Council's strong claims are based on nothing more than "cherry-picking" from a set of hundreds of simple correlation coefficients computed by using no more than 8 and as few as 5 observations. Out of 462 statistical tests conducted, the Lisbon Council's conclusions are based on the statistical significance of less than 5% of tests.[6] Such a small number of "successes" is readily explained by random variation (at a significance level of 5%), thereby providing reasonably strong evidence that there is *no relationship* between copyright flexibility and economic outcomes. Even worse, the economic activity variables for the eight non-randomly selected countries are measured as an average of nearly twenty-years of nominal (not inflation adjusted) time-series data that is *expressed in five different currencies*. These are fatal errors.

Given the sloppy and nonsensical statistical analysis that appears to be ends-driven, European policymakers should run, not walk, from the Lisbon Council's analysis and policy recommendations regarding copyright reform for the Digital Single Market.

And while the Lisbon Council casually asserts that for the May revision of the *IP Index Report* "correlations tests have all been re-run [] and no significant variations were found," the claim is patently false.[7] The statistical results from the revision are very much different and weaker than the original version. Given the sloppy and nonsensical statistical analysis that appears to be ends-driven, European policymakers should run, not walk, from the Lisbon Council's analysis and policy recommendations regarding copyright reform for the Digital Single Market.

IP Index Report's Methodology

Recognizing that "intangible assets have become the principal driver of growth and productivity in advanced, knowledge-based economies," the Lisbon Council senses "an urgent need to reflect on current understandings of how innovation delivers economic value."[8] Favoring the view that copyright is "restricting innovation," the *IP Index Report* aims to "measure the impact of exceptions to copyright on economic growth." To do so, the analysis must be expected to establish a causal relationship going *from* copyright law *to* economic activity.[9] Yet, the *IP Index Report* makes clear that its methods cannot be used to "attribute causality," though the *Report's* conclusion are not much tempered by this admission.[10]

While the term "econometric modeling" appears throughout the IP Index Report and there is much discussion about the value of time-series data, in fact the IP Index Report includes no econometric analysis at all and discards entirely the time-series nature of the data.

While the term "econometric modeling" appears throughout the *IP Index Report* and there is much discussion about the value of time-series data, in fact the *IP Index Report* includes no econometric analysis at all and discards entirely the time-series nature of the data. Rather, the statistics in the *IP Index Report* are merely descriptive (i.e., correlation coefficients). Specifically, the correlation coefficient is the only statistical method to be found in the *IP Index Report*. The Lisbon Council's description of a correlation coefficient as "econometric modelling" and "econometric analysis" is a fairly good signal of inexperience—a lack of skill undoubtedly confirmed upon closer inspection of the analysis.

Also, econometric analysis is used to test hypotheses derived from a formal economic model.[11] In the instant case, such a model would reveal the mechanisms by which specific types of "copyright flexibility" affect economic outcomes. Yet, nowhere in the Lisbon Council's *Report* is there an economic model that explicitly links the elements of its copyright flexibility index to the economy. In effect, the *Report* merely assumes a relationship between flexibility and the economy. In fact, the *Report's* implicit assumption is that the *only* thing that affects the economy is copyright flexibility, though, given the statistical test employed, it is correct to reverse the causality and say that the only thing that affects copyright flexibility is the economy. Such a take on economic activity is, on its face, silly.

For the correlation analysis, Kendall's tau is used. Kendall's tau is a type of correlation coefficient (a non-parametric one) that analyzes the rank of data obtained in pairs.[12] By using the ranks of the data rather than actual values, Kendall's tau is not much affected by outliers in the data, since extreme values do not alter the rankings.[13] Kendall's tau is interpreted in a standard fashion—it has values between -1 and +1, with 0 implying no relationship. Whether or not the tau correlation is "statistically different from zero" can be tested. For "large" samples (> 10), which are not relevant here, the test statistic follows the standard normal distribution.[14] For smaller samples, a non-parametric approach such as the bootstrap or published tables of critical values should be employed, but the *Report* does not indicate how statistical significance is determined.[15]

As is well-known, correlation analysis is flimsy evidence at best, especially for policymaking. Correlation merely indicates whether the two variables generally move in the same direction. No causal interpretation is permitted. In fact, absent a great deal of other evidence, correlation does not indicate a relationship of any real meaning.

Many things are correlated that have no meaningful relationship. Such relationships are called "spurious correlations." For example, over time, *U.S. Spending on Science, Space and Technology* has a near perfect correlation coefficient ($r = 0.998$) with *Suicides by Hanging, Strangulation and Suffocation;* per-capita cheese consumption has a strong correlation ($r = 0.95$) with the number of people who died by becoming tangled in their bedsheets; which in turn has a strong correlation with the divorce rate in Maine ($r = 0.992$).[16] Later, using the data and methodologies of the *IP Index Report*, I show that a policy to increase gun ownership in Europe will improve economic outcomes much more than will making copyright law more flexible.

Out of 462 statistical tests conducted, the Lisbon Council's conclusions are based on the statistical significance of less than 5% of tests. Such a small number of "successes" is readily explained by random variation (at a significance level of 5%), thereby providing reasonably strong evidence that there is no relationship between copyright flexibility and economic outcomes.

Furthermore, correlation analysis does not imply any *directional* relationship between two variables. Certainly, the *IP Index Report* wishes to (and needs to) demonstrate that copyright flexibility affects economic outcomes. Yet, the two variables in a correlation analysis are "equals" so that there is no "direction" of relationship. No directional relationship is even assumed.

The shortcomings of correlation analysis are well-known, but in the case of the *IP Index Report* such concerns are a bit like complaining about a bank robber illegally parking during his crime. There is a long-list of fatal defects found in the *IP Index Report* unrelated to the choice to use the flimsiest of statistical methods.

The Lisbon Council's description of a correlation coefficient as "econometric modelling" and "econometric analysis" is a fairly good signal of inexperience—a lack of skill undoubtedly confirmed upon closer inspection of the analysis.

Data

The goal of the *IP Index Report* is to evaluate the relationship—using simple correlation—between economic activity and copyright flexibility. The *IP Index Report* (non-randomly) chooses eight countries for its analysis: the U.S., the U.K, France, Germany, Japan, the Netherlands, Spain, and Sweden. These are all advanced economies of substantially varying size. For these eight countries, data for the *Report's* analysis is obtained from two sources.

First, economic data is from the EU/World KLEMs database.[17] The *IP Index Report* lists (at Table 10) the economic output variables included in the initial round of data analysis and their definitions.[18] These data are said to include, economy-wide and by certain industries, items such as Gross Value-Added, Number of Persons Employed, Labour Compensation, Capital Compensation, Growth Rate of Value Added, Total Factor Productivity, Contribution of Hours Worked to Value Added Growth, and Contribution of TFP to Value Added Growth. However, the *Report* provides neither discussion nor results for "Capital Compensation" measures; this measure of economic activity was discarded for unspecified reasons.

The *IP Index Report* uses 1992 as the starting date through the last year provided (2009, 2010, or 2011, depending on the country). In all, the *IP Index Report* uses 77 different measures of economic activity. For each activity, the mean, median, and growth rate (average of first differences) is computed.[19] Using this data, 231 Kendall's tau correlations [$= 77 \times 3$] are computed and a statistical test of the hypothesis that the tau is equal to zero is performed (using an unspecified method).[20]

[T]here is no econometric modeling in the IP Index Report. The only statistic used is a correlation coefficient. So, what is good for econometric modeling is mostly irrelevant to the IP Index Report's approach.

The EU/World KLEMs data is chosen, according to the *IP Index Report*, because of the "availability of long, time-series data for each country," which provides for "more data points" and permits "stronger econometric modelling and [] robustness of results."[21] This concern for "long, time-series data" is entirely misplaced given the chosen methodology, and turns out to be more harmful than helpful for a number of reasons.

First, as already noted, there is no econometric modeling in the *IP Index Report*. The only statistic used is a correlation coefficient. So, what is good for econometric modeling is mostly irrelevant to the *IP Index Report's* approach.

Second, the Lisbon Council discards entirely the times-series nature of the data. Specifically, rather than use the richness of the variation in the time-series data, the *IP Index Report* instead uses averages and medians of the series, effectively turning twenty years of data into a single data point.[22] Discarding the time series nature of the data by averaging, particularly when legal changes took place during the sample period, is certainly an odd methodological choice. It's unclear what such a statistic means or what guidance it provides for policy. Imagine a case, for example, where a change in a policy in 2002 miraculously makes a previous laggard into a leader. By averaging the economic performance over an extended period of time both before and after the law change, the statistical analysis would view the country as mediocre, since the low values of data from the laggard period offset the high values of data from the leader period. Averaging the data to a single point makes absolutely no sense.

[T]he Lisbon Council discards entirely the times-series nature of the data. Specifically, rather than use the richness of the variation in the time-series data, the IP Index Report instead uses averages and medians of the series, effectively turning twenty years of data into a single data point.

Third, the data is measured in its levels and in first differences of the levels (the latter labeled a "growth rate"),[23] so the correlation analysis runs the risk of a spurious size effect. Obviously, the number of employed persons, labor compensation, value added, and so forth, will be positively correlated with the size of an economy. For example, even if the U.S. used half as much labor as Japan to create its output, then employment and labor compensation will be larger in the U.S. than in Japan. A "size" outlier, like the U.S. in this group of countries (which, by the Lisbon Council's "SFEER" Index (discussed *infra*), also has the highest level of copyright flexibility), can influence a rank correlation coefficient, conflating size and the relationship of interest. To address this problem, data is often scaled (e.g., as per-capita or as a percentage of GDP), but the *IP Index Report* does not make such an adjustment.

In taking the average of twenty years of the EU/World KLEMs data, the IP Index Report has averaged nominal data. This is a serious error: one U.S. Dollar in 2010 has the purchasing power of only $0.64 in 1992. The IP Index Report is, in effect, averaging apples and oranges…

In addition to these peculiar treatments of time-series data, the EU/World KLEMs data has two very important features entirely ignored in the *IP Index Report*: (a) the data is nominal (not adjusted for inflation); and (b) the data are in country-specific currencies. For the eight countries, there are *five different currencies*: the U.S.'s data are in U.S. Dollars; Japan's data are in Yen; Sweden's data are in Krona, the U.K's data are in Pounds; and the remainder in Euros. Neither of these features is addressed. The problems are apparent.

In taking the average of twenty years of the EU/World KLEMs data, the *IP Index Report* has averaged nominal data. This is a serious error: one U.S. Dollar in 2010 has the purchasing power of only $0.64 in 1992.[24] The *IP Index Report* is, in effect, averaging apples and oranges (which, as explained next, is an error made twice). Having data from multiple countries exacerbates the problem. The Japanese Yen has traded for U.S. Dollars at ratios between 83:1 to 143:1 over the years 1992-2010. Thus, the relative values of the economic data between countries are inconsistent over time in a common currency. With five different currencies represented in the data, the analysis of the data cannot escape these inconsistencies.

Let's look, for example, at Germany and Sweden for the average of labor compensation for the telecommunications industry. By averaging the nominal data, Sweden (10,939) ranks higher than Germany (10,679). If, instead, the data is adjusted for inflation, Germany (11,041) now ranks higher than Sweden (10,762).[25] This change in rank is important when using a rank statistic like Kendall's tau.

A comparison between Germany and Sweden is made even more problematic by the fact that the data for the two are expressed in different currencies. While the *IP Index Report* claims that all the data are measured in U.S. Dollars, a comparison of the *IP Index Report's* data to the source data reveals otherwise.[26] All the data remain in each country's own currency. Even the series described in the *Report* as a "growth rate"—which turns out to be almost all of the statistically significant results—is not immune to the currency problem. The *IP Index Report* computes the "growth rate" as a first-difference in values and not as a percentage growth rate. Thus, the currency problem remains intact.

While the Lisbon Council describes the *IP Index Report* as the product of a "year-long research effort,"[27] the variation in currency is immediately apparent with even a cursory review of the data. Take the variable for Labor Compensation for the Market Economy (labeled "Market Economy-lab"). For the U.S., the figure provided in the *IP Index Report*'s Table 15 is $4.6 trillion, which seems reasonable for an economy with a GDP averaging $11 Trillion between 1992 and 2010. Yet, if all the data were in U.S. dollars as the *IP Index Report* claims, then Japan's labor compensation figure is $231 trillion—an amount nearly fifty times larger than the Japanese economy ($4.7 trillion) and twenty-five times larger than the U.S. economy over the 1992-2010 period. Japan's data, in contrast to the *Report's* claim, are actually provided in Japanese Yen.

Table 1, provides the exchange rates of all the currencies in the *IP Index Report's* data relative to the U.S. Dollar in 2010.[28] The exchange rate between the Yen and the U.S. dollar in 2010 was about 82-to-1, and over the 1992-2010 period was close to about 100-to-1, a fact which explains the exceedingly high value for labor compensation. Making a currency adjustment would bring labor compensation in Japan down to about $2.7 trillion, which is a reasonable share of its $4.7 trillion in GDP.

An attentive researcher would pick up on this issue immediately; the *IP Index Report* does not. More troubling is the fact that the *IP Index Report* actually labels Japan an outlier and tosses it from the sample, but fails to give a second thought regarding the source of the problem.[29]

Table 1. Currencies of Data

Country	Currency	Exchange 2010 (to U.S. Dollar)
US	U.S. Dollar	1 : 1
UK	Pound	0.65 : 1
Germany	Euro	0.75 : 1
Sweden	Krona	6.7 : 1
Spain	Euro	0.75 : 1
Japan	Yen	82 :1
France	Euro	0.75 : 1
Netherlands	Euro	0.75 : 1

By failing to adjust the varied currencies to a single currency, Table 1 reveals that relative to the U.S. Dollar the data for the U.K., Germany, Spain, France and the Netherlands is understated and the data for Japan and Sweden is inflated. Since much of the data (and almost all the data for statistically significant results) is in levels or in first-differences of levels, the currency problem renders the correlation analysis meaningless. The differences in currencies will affect the rank of the countries, which in turn affects Kendall's tau. Apples are being compared to oranges. This mistake is so negligent and fatal it may be hard to believe it could be made. But, as my analysis or a basic review of the source data shows, it was a mistake plainly made.

While the IP Index Report claims that all the data are measured in U.S. Dollars, a comparison of the IP Index Report's data to the source data reveals otherwise. All the data remain in each country's own currency.

Using different currencies is a fatal mistake indicating a lack of experience in statistical analysis and/or profound carelessness. Failing to normalize the currencies is akin to comparing prices of gasoline in the U.S. and Europe without adjusting for the fact that prices in the U.S. are expressed in gallons but in Europe are expressed in liters. To wit, three of the eight general classes of economic indicators are impacted by the currency error (i.e., labeled "–va", "-lab", "-cap"). Of the 15 "significant" and "borderline" significant correlations listed in the *IP Index Report*, 12 of these results are impacted by the currency error. Thus, the Lisbon Council's statistical results are entirely meaningless (for many reasons).

SFEER Index

In light of the absence of an index of copyright flexibility across countries, the *IP Index Report* sets out to create its own index of copyright flexibility, which is labeled the *Scope and Flexibility of Exceptions to Exclusive Rights Index* ("SFEER Index"). Looking at the copyright regimes for eight advanced economies (listed in Tables 1-3), this index attempts to quantify flexibility by accounting for variations in fair use rules, private copying, exceptions related to education, parody and criticism, and other factors.[30] The construction of the SFEER Index is based on the author's interpretation of copyright laws. No attention is paid to legal decisions regarding the application of such laws, and the *IP Index Report* includes the explicit assumption of uniform (and presumably effective) enforcement.[31]

One problem with creating such an index is that laws change. According to the *IP Index Report*, "[o]nly statutes which were in force for over half of the time period were included. Since the economic data surveyed ranged from 1993 to 2010, the cut-off date for legislation was set as 2002."[32] The Lisbon Council's reasoning is as follows:

> … these changes are omitted from the SFEER Index because it would be misleading to evaluate the relationship between historical economic performance and copyright law over a given timeframe based on statutes that were only enacted late in the period in question. Only statutes which were in force for over half of the time period were included. [] All exceptions and limitations to copyright that were introduced after 2002 are not included in the scores.[33]

This focus on the timing of legal changes reveals another fatal flaw in the analysis. If flexibility is important, as the *IP Index Report* presumes, then an early legal change will affect much of the time series data, whereas a later legal change will affect less of the time series data. Of course, by averaging twenty years of data, it is impossible to test for the effects of "changes" in the laws or the economic outcomes.

Moreover, explicit in the construction of the SFEER Index is the assumption that changes in copyright laws will not have much influence on outcomes in the eight-year window following the change. At the risk of getting off task, I think it is worth mentioning that in another of Benjamin Gibert's studies—*The Economic Value of Fair Use in Copyright*—the effects of a legal change are assumed to occur in the same year (and the subsequent six years) as that change.[34] Obviously, Mr. Gibert's view on the matter of the timing of economic responses is flexible and inconsistent.

Considering the temporal nature of laws, the variation in enforcement, the relevance of judicial interpretation, and the general sloppiness of the *IP Index Report's* analysis, it is difficult to assign much credibility to the SFEER Index. Nevertheless, digging into the details of its construction is beyond the scope of this PERSPECTIVE.

Table 3. SFEER Index

Rank	Country	SFEER Score (May 2015) [Rank]	SFEER Score (March 2015) [Rank]
1	US	8.13 [1]	8.13 [1]
2	UK	7.19 [2]	7.81 [2]
3	Germany	5.94 [3/4]	7.50 [3]
4	Sweden	5.94 [3/4]	7.19 [4]
5	Spain	5.63 [5]	6.88 [5]
6	Japan	5.31 [6/7]	6.25 [6/7]
7	Netherlands	5.31 [7/7]	5.94 [8]
8	France	4.38 [8]	6.25 [6/7]

Table 3 lists the SFEER Index for the eight countries; values for the initial version and revision are provided. According to the *IP Index Report*, the maximum value for the SFEER Index is 10.0 and the minimum value is 1. Higher values of the index indicate a more flexible regime.[35] The most flexible regime is in the United States—which is also, by far, the largest economy in the sample—with a SFEER Index of 8.13 and the least flexible is the Netherlands at 4.38. Note also that for 5 of the 8 countries in the sample the range of the SFEER Index is only 5.31 to 5.94. In effect, the copyright regimes of these countries are almost identical (based on the index). Without much variability, statistical evidence of relationships is very difficult to obtain, and correlations will be unduly influenced by a few observations (as demonstrated later).

The *Report's* revision in May 2015 is principally related to changes in the SFEER Index.[36] Table 1 shows that the SFEER Index values do change quite a bit, eliminating one tie and introducing two new ties. There are also rank changes for France and the Netherlands. These ties (accounting for four of the eight countries!) and changes in ranks are very important when using a rank statistic like Kendall's tau (especially in small samples), and we should expect these changes to materially impact the results. They do. Naturally, most of my discussion will focus on the latest *IP Index Report*. However, given the significant differences in the results between the two releases, I will provide some discussion of the first release of the *Report* when relevant.[37]

Using different currencies is a fatal mistake indicating a lack of experience in statistical analysis and/or profound carelessness. Failing to normalize the currencies is akin to comparing prices of gasoline in the U.S. and Europe without adjusting for the fact that prices in the U.S. are expressed in gallons but in Europe are expressed in liters. *** Thus, the Lisbon Council's statistical results are entirely meaningless…

Statistical Analysis

Let's turn now to the statistics. For eight countries, there is a single-valued SFEER Index measuring copyright flexibility. Paired with this data are 231 single data points computed from 77 measures of economic activity. These measures include the average of levels, the median of levels, and the average of first differences over about twenty-years of data for each of the 77 economic activity variables. As already discussed, the source data was in nominal terms and expressed in five different currencies. No adjustments were made for these features of the data. Consequently, the statistical tests are meaningless, but I will proceed as if they are not for the purpose of illustrating a few methodological principles.

Using data for all the countries, 231 Kendall tau correlation coefficients are computed under the null hypothesis that the correlation is equal to zero. The correlations are based on at most 8 and as few as 5 observations. These very small samples strain the credibility of the research. Small samples render weak statistical tests, and are also prone to spurious correlations. One research paper notes, "[s]purious results are virtually certain with small [samples], a large number of explanatory variables, and an intense search for statistical significance."[38] Similarly, as the book STATISTICS FOR EVIDENCE-BASED PRACTICE AND EVALUATION states, "[s]tudies with very small samples might find strong correlations due to sampling error."[39] The other problems with the data—the use of levels, non-normalization, nominal data, and currency inconsistencies—ensures spurious results.

In assessing statistical significance, the *IP Index Report* uses a significance level of 5%.[40] Given the small samples, it is improper to appeal to the normal distribution for tests of statistical significance, but the *IP Index Report* does not say how the confidence intervals on the test statistics were determined. As such, my discussion of statistically significant results will rely primarily on the critical values for small samples reported in numerous papers and textbooks.[41]

Forming Expectations for a Large Number of Tests

Prior to discussing the results in the *Report*, it is worthwhile to consider what results might be expected from conducting 231 hypothesis tests at the 5% significance level. When a researcher says "the result is statistically significant at the 5% level," then what he or she means is that a statistically-significant result (i.e., a rejected null hypothesis) would only be observed due to chance about 5% of the time. Thus, at the 5% significance level, the researcher can say that he or she is 95% confident the result is real and not due to random variation. We can never be absolutely certain in a statistical result. Statistical tests are subject to Type I (a false positive) and Type II (a false negative) errors.

[T]he IP Index Report is junk science and should be ignored. But, if it can be said to demonstrate anything, then it is that there is no meaningful statistical relationship between copyright flexibility and economic outcomes.

At the 5% significance level, the Type I error rate is also 5%. That is, if a researcher conducted 100 statistical tests and judged significance at the 5% level, then the null hypothesis of about 5 of those tests will be improperly rejected due simply to random variation. In fact, 95% of the time we can expect to see a number of false positives between 1 and 10 for 100 tests.[42] Given the 231 tests conducted in the *IP Index Report*, the number of "statistically significant results" that are, in fact, false positives will (under typical conditions) lie between 6 and 18 tests.[43]

A simple simulation confirms this result. In this simulation, I generate two variables each with 8 observations and with two ties on one of the variables (to match the *IP Index Report*). These two variables are randomly created and thus have no relationship to each other than that possibly arising from random variation.[44] Using these variables, a Kendall's tau correlation coefficient is computed and a test of the null hypothesis of "zero correlation" is performed at the 5% alpha level.[45] This process is repeated 231 times and then the number of statistically-significant results is counted. This process is then repeated 1,000 times to generate a distribution of rejection rates for the 231 tests.

Keep in mind that these data are entirely fabricated using a random number generator and, by design, the two variables have no relationship other than that which appears randomly. Nevertheless, this simulation produces an average of 13 rejections of the null hypothesis.[46] As expected, the average number of statistically-significant correlations equals a little over 5% of tests.[47] Yet, each "round" of the 231 simulated tests does not produce exactly a 5% rejection rate; the rejection rate has a distribution. The number of rejections falls between 6 and 20 statistically significant results 95% of the time (i.e., the 95% confidence interval). Thus, out of 231 Kendall tau tests performed on data known to be independent, we can expect to observe as many as 20 statistically-significant correlation coefficients that can be attributable solely to random variation (i.e., statistical noise).

The IP Index Report's Results

Given the carelessness of the data collection, the apparent lack of experience in statistical analysis, the failure to indicate how statistical significance is determined, and the failure to report the results for the median measures, I replicated the *IP Index Report's* data and computed all the relevant statistics. There was a high coincidence between the replicated and reported correlation coefficients. For my replication analysis, statistical significance (at the 5% level) was based on the proper critical values for small samples.

Using the results from this replication, the *IP Index Report's* method produces 8 statistically significant results for the 231 tests on the full sample.[48] At the 5% significance level, however, we expect there to be 12 statistically significant results due solely to random variation. As such, the Lisbon Council's findings are no different than what would be observed if the SFEER Index had no relationship at all to the economic data. The few statistically significant results are just noise.

Rather than fess-up to the large differences across the two versions of the Report and conduct a serious rewrite of the IP Index Report in light of the changed results, the Lisbon Council concludes that "no significant variations were found" in the results and "the overall conclusions of the paper do not change." The statements are obviously disingenuous; there are significant differences in the results. *** To claim that the statistical tests are unchanged is a falsehood and the failure to alter the conclusions is a serious indictment reeking of an ends-driven analysis.

For the numerous reasons discussed above, the *IP Index Report* is junk science and should be ignored. But, if it can be said to demonstrate anything, then it is that there is *no meaningful statistical relationship* between copyright flexibility and economic outcomes. Indeed, the statistical results of the *IP Index Report* are not inconsistent with random variation and, based on these results, it would not be possible to reject the hypothesis of "no effect" of flexibility on economic outcomes.[49]

Tossing out Japan

Apparently unhappy with the weak results for many of the industries expected to benefit from copyright flexibility, the *IP Index Report* offers two explanations: either the (a) output in these industries is not correlated with the [SFEER Index]; or (b) correlation results are being skewed by an outlier in the data.[50] In light the prior discussion, guess who the outlier is? Japan, of course. The *IP Index Report* notes, "Japan was a consistent outlier along certain output variables—such as value-added and labour compensation."[51] Since both "value added" and "labour compensation" are measured in disparate currencies and in levels, the "outlier" status is no surprise. By expressing Japan's data in Yen (trading at about 100:1 to the U.S. Dollar), Japan's data indeed sticks out like a sore thumb.

Addressing outliers is a common practice in statistical analysis, but tossing out data is justified only when there is an expected coding error or some other explanation for the outlier status. Before tossing out outliers, a concerted effort is required to determine why there is an outlier. Rather than making an effort to figure out why Japan was so different—which should have certainly led the author to the currency problem—Japan is tossed out of the sample simply because its data didn't fit the *IP Index Report's* preferred theory. This is known as "cherry-picking," an improper practice discussed in more detail later.

Furthermore, one reason Kendall's tau correlation is used is that it is mostly immune from the influence of outliers in the raw data because it is a rank statistic. As noted in the book STATISTICS IS EASY!, "If outliers are a problem, then a rank transformation might work better."[52] So, in addition to exposing a clear case of cherry picking, the outlier approach suggests that the *Report's* chosen statistic is not very well understood by its author. The use of a rank statistic given a high incidence of ties in such a small sample likewise indicates a lack of understanding of the rank statistic.

Significantly, Japan's alleged outlier status was very useful in the initial release of the *IP Index Report*. Significance of the tests rose from 5 tests in the full sample to nearly 40 significant correlations in the reduced sample.[53] In the revised May 2015 version of the study, however, dumping Japan was entirely unhelpful. In the revision, there were 12 reported significant correlations at the 5% level for the full sample (based, apparently, on improper significance testing), but without Japan in the group that number fell to only 7 tests. All 7 of these results are "-lab" or "-va" variables and thus tainted with the currency problem.

My replication of the analysis indicates 13 rejections of the 231 tests when Japan is excluded (2 of which are from the unreported median measures), which is 4 more than found in the full sample. This number of statistically significant results is again not inconsistent with random variation. Again, based on the evidence from the *IP Index Report* (either as reported or replicated), *it is not possible to reject the hypothesis that copyright flexibility has no effect on economic outcomes*.

Guilty of Cherry-Picking

In his book MORE DAMNED LIES AND STATISTICS: HOW NUMBERS CONFUSE PUBLIC ISSUES, Joel Best defines cherry-picking as "selecting statistics that support a particular thesis and drawing attention to those numbers, while ignoring other figures that might lead to a different conclusion."[54] As one prominent scientist observed, cherry-picking is the "hallmark of poor science or pseudo-science."[55] The *IP Index Report* cherry-picks both statistics (almost all of which are meaningless anyway) and the data.

First, the *IP Index Report's* conclusions are based generously on 22 statistically significant results out of 462 tests.[56] Thus, the bulk of the evidence (95% of tests) weighs against the hypothesis that copyright flexibility improves economic outcomes. As discussed above, these few significant results are entirely consistent with random variation. Interpreting results in this way is an excellent example of "selecting statistics that support a particular thesis and drawing attention to those numbers, while ignoring other figures that might lead to a different conclusion." That is, the *IP Index Report* engages in the most blatant form of cherry-picking.

Second, based solely on the fact that Japan's data did not fit into the analysis the way the author preferred, the country was removed from the sample and the correlation coefficients recomputed. No attempt is made to explain why Japan is an outlier, or why there may be errors in Japan's data. (Recall, it was Yen related.) Quite simply, Japan is excluded only because its data doesn't fit the theory. Removing Japan from the analysis simply because its data doesn't fit the theory is another case of conspicuous cherry-picking.

Third, without explanation, the *IP Index Report* provides detailed results only for the mean and growth rate variables, excluding those from the median measures. The *Report* states explicitly that the tests were conducted and in the *initial* release of the *Report* some statistically significant results are reported for the median measures. Yet, there is no discussion of the median variables in the revision.

From the replication, there are scarcely any statistically significant correlations from the median data. For the full sample, there is only one statistically significant result (and 2 when Japan is excluded). Perhaps the median measures were excluded because of their relatively poor performance. Again, it appears that the statistical results have been cherry picked.

Revision Misdirection

There are two versions of the *IP Index Report*—the initial release and the revision. Given the changes in the SFEER Index that motivated the revision, a change in the statistical results is expected. Yet, the revised *IP Index Report's* claims that "correlations tests have all been re-run [] and no significant variations were found."[57] This claim is a transparent misrepresentation of the truth.

Quite simply, Japan is excluded only because its data doesn't fit the theory. Removing Japan from the analysis simply because its data doesn't fit the theory is another case of conspicuous cherry-picking.

In the first release, tossing Japan was helpful, and the *IP Index Report* concluded, "this new test demonstrated a much stronger correlation between the SFEER Index and economic output variables in a number of industries."[58] This statement did not appear in the May 2015 revision, of course, since the actual number of statistically significant correlations fell (as indicated in the *Report*). Obviously, going from 40 reported rejections in the first release to only the 7 reported in the revised *Report* is a "significant variation[]" in the findings.[59]

Given the lack of any help from excluding Japan (with the *Report* claiming a drop from 8 to 7 significant results), it's a mystery why the *IP Index Report* even bothers to make the exclusion. Perhaps the best explanation is that the initial release of the *IP Index Report* had made a strong case for Japan being an outlier (which increased the number of statistically significant results), so backtracking wasn't feasible. Rather than fess-up to the large differences across the two versions of the *Report* and conduct a serious rewrite of the *IP Index Report* in light of the changed results, the Lisbon Council concludes that "no significant variations were found" in the results and "the overall conclusions of the paper do not change."[60] The statements are obviously disingenuous; there are significant differences in the results. If it is true that the conclusions do not change, then it must be the case that the conclusions have nothing at all to do with the statistical tests. (The evidence presented here suggests that is the case.) To claim that the statistical tests are unchanged is a falsehood and the failure to alter the conclusions is a serious indictment reeking of an ends-driven analysis.

[W]hile the Report claims that there were no changes in the statistical results between the initial release in March and the revision May of 2015, there are sizeable differences. For the Lisbon Council to claim otherwise is disingenuous.

Spurious Correlation

Even if the number of rejections exceeded that expected from random variation, the small samples and a-theoretic statistical approach risks spurious correlation. To demonstrate this possibility, I added to the data a variable measuring the number of guns-per-capita in each country.[61] My guess is that the number of guns and economic outcomes are not related. Nevertheless, out of 231 tests, there are 25 correlations that are statistically different from zero.[62] This rejection rate is about twice that reported in the *IP Index Report* for the SFEER Index and the economic variables, and outside the 95% confidence interval of Type I error (which is 18 statistically significant correlations). By the logic of the *IP Index Report*, the European Union might be wiser to push gun ownership rather than copyright flexibility.

More evidence on spurious correlation is available. As mentioned already, the U.S. has the highest SFEER Index and also is the largest economy in the study. Since much of the data is measured in levels (i.e., compensation to employees, value added, etc.), the U.S.'s relatively large size could influence the correlation coefficients.

To evaluate the influence of a size effect with respect to the U.S., the 231 correlation tests are conducted without the U.S. in the sample. Of these tests, only 5 statistically-significant correlations are found. There appears to be a size effect driving spurious correlation. Still, even with that size effect, the results of the *IP Index Report* are not inconsistent with random variation. On all accounts, the *Report* is an exceedingly weak and incompetently conducted study.

Conclusion

The Lisbon Council's *IP Index Report* shows a near total disregard for the most basic of scientific methods. These problems with the study covered in this PERSPECTIVE represent only a sample of the many errors, inconsistencies, and misleading statements contained in the *Report*, all of which are individually fatal to the Lisbon Council's credibility.

The *IP Index Report's* conclusion that "[c]ountries that employ a broadly 'flexible' regime of exceptions in copyright also saw higher rates of growth in value-added output throughout their economy" has zero support.[63] In fact, given that the results can be fully explained by random variation, the results presented in the *IP Index Report* are better and more honestly interpreted as evidence of no relationship between copyright flexibility and economic outcomes.

Furthermore, while the *Report* claims that there were no changes in the statistical results between the initial release in March and the revision in May of 2015, there are sizeable differences. For the Lisbon Council to claim otherwise is disingenuous. Moreover, to state that the "overall conclusions of the paper do not change" across the two versions seems to imply that the statistical results had no bearing on the conclusions.

With the start of its Digital Single Market initiative, the European Union is about to undertake one of the most complex legal overhauls of its regulations perhaps since the formation of the Union itself. Given the stakes at hand, therefore, the junk science proffered by the Lisbon Council with regards to copyright reform should be summarily ignored.

NOTES:

* **Dr. George Ford** ***is Chief Economist of the Phoenix Center for Advanced Legal and Economic Public Policy Studies. The views expressed in this*** **PERSPECTIVE** ***do not represent the views of the Phoenix Center or its staff. Dr. Ford has testified as an expert before the copyright royalty authorities in the U.S. and in Canada. Dr. Ford may be contacted at*** ford@phoenix-center.org.

1 *See* FINAL COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS: *A Digital Single Market Strategy for Europe* (6 May 2015) (available at: http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf).

2 *Id.* at ¶ 2.4.

3 *Id.*

4 B. Gibert, *The 2015 Intellectual Property and Economic Growth Index: Measuring the Impact of Exceptions and Limitations in Copyright on Growth, Jobs, and Prosperity*, Innovation Economics Center, The Lisbon Council (2015) (hereinafter the "*IP Index Report*") (available at: http://www.lisboncouncil.net/publication/publication/122-the-2015-intellectual-property-and-economic-growth-index.html).

5 *Id.* at p. 3.

6 There are 19 statistically significant results, but 4 are excluded for undecipherable reasons. *Id.* at p. 26.

7 *Id.* at p. 1.

8 *Id.* at p. 2.

9 *Id.* at pp. 2, 23.

10 *Id.* at p. 23.

11 *See*, *e.g.*, J. Wooldridge, INTRODUCTORY ECONOMETRICS (Cengage 2003) at p. 2 ("How does one go about structuring an empirical economic analysis? It may seem obvious, but it is worth emphasizing that the first step in any empirical analysis is the careful formulation of the question of interest."); D. Gujarati, BASIC ECONOMETRICS (McGraw Hill 1995) at p. 3 ("Broadly speaking, traditional econometric methodology proceeds along the following lines: (1) Statement of theory or hypothesis; (2) Specification of the mathematical model of the theory; (3) Specification of the econometric model of the theory; (4) Obtaining the data; (5) Estimation of the parameters of the econometric model; (6) Hypothesis testing; (7) Forecasting or prediction; and (8) Using the model for control or policy purposes.").

12 H. Abdi, *Kendall Rank Correlation*, ENCYCLOPEDIA OF MEASUREMENT AND STATISTICS (N.J. Salkind, Ed.) (2007) at pp. 508-510 (available at https://www.utdallas.edu/~herve/Abdi-KendallCorrelation2007-pretty.pdf); R. Sommer and B. Sommer, PRACTICAL GUIDE TO BEHAVIORAL RESEARCH: TOOLS AND TECHNIQUES (Oxford University Press 2001). The *IP Index Report* does not indicate whether the "a", "b", or "c" version of Kendall's tau is used, but replication suggests it is Kendall's tau-b.

13 *Id.*

14 J.D. Long and N. Cliff, *Confidence intervals for Kendall's Tau*, 50 BRITISH JOURNAL OF MATHEMATICAL AND STATISTICAL PSYCHOLOGY 31-41 (1997); G. Kanji, 100 STATISTICAL TESTS (2006) at p. 110.

15 Abdi, *supra n*. 12.

16 A number of interesting examples are provided at: http://www.tylervigen.com/spurious-correlations.

17 http://www.euklems.net.

18 *IP Index Report, supra* n. 4 at p. 24 and Table 10.

19 *Id.* ("The time series data for each country was collapsed into three summary statistics for each industry group and variable (mean, median, and average annual growth rate). Each one was tested for significant correlation with the SFEER Index score using the Kendall tau rank correlation methodology."). There were some omissions for individual countries due to a lack of data.

NOTES CONTINUED:

20 Without explanation, the *Report* includes detailed tables of results for only the mean and growth rate measures (excluding the median measures). The initial release of the *IP Index Report* did, however, report some significant correlations of the median data (see *IP Index Report* (first version), at Table 17). So clearly, the correlation coefficients using the median data were computed but, oddly, not reported. (My own explanation for this omission is provided *infra*.)

21 *IP Index Report, supra* n. 4 at p. 17.

22 *Id.* at p. 24 ("The time series data for each country was collapsed into three summary statistics for each industry group and variable (mean, median, and average annual growth rate). Each one was tested for significant correlation with the SFEER Index score using the Kendall tau rank correlation methodology.").

23 A growth rate is properly expressed in percentage terms [say, $(x_t - x_{t-1})/x_{t-1}$]. In the *IP Index Report*, however, the "growth rate" is simply an average of the first differences of the data [the average of the $(x_t - x_{t-1})$ for about twenty years].

24 http://data.bls.gov/cgi-bin/cpicalc.pl.

25 The price index for gross value added (provided in the EU/World KLEMs data) is used to convert the nominal to real values.

26 *IP Index Report, supra* n. 4 at Table 10, and Tables 15-19.

27 *Id.* at ft. 1.

28 www.oanda.com/currency/historical-rates.

29 Japan isn't the only victim of the careless data handling; the currency problem infects all the data. Five different currencies are represented in the data. For Sweden, labor compensation is listed as what would be $884 billion if the data were actually in U.S. Dollars, but the Swedish economy is only about $400 billion in U.S. Dollars over the relevant period. As with the Yen, it takes many Kronor to make a dollar, thereby substantially inflating the Swedish data.

30 *IP Index Report, supra* n. 4 at p. 12.

31 *Id.* at p. 16.

32 *Id.* at p. 5.

33 *Id.*

34 R. Ghafele & B. Gibert, *The Economic Value of Fair Use in Copyright Law: Counterfactual Impact Analysis of Fair Use Policy on Private Copying Technology and Copyright Markets in Singapore*, Unpublished Manuscript (2012) (available at: https://ideas.repec.org/p/pra/mprapa/41664.html) at p. 20 (Pre-intervention period = 1999 – 2004; Intervention date: 01/01/2005; Post-intervention period = 2005 – 2010).

35 *IP Index Report, supra* n. 4 at p. 18.

36 *Id.* at p. 1.

37 The Lisbon Council has removed the earlier version from its webpage. However, I can provide a copy of the original *IP Index Report* upon request.

38 D.R. Anderson, K.P. Burnham, W.R. Gould and S. Cherry, *Concerns about Finding Effects That are Actually Spurious*, 29 BIOMETRICS 311-316 (2001). Large samples may also lead to spurious results by reducing the variance of estimators. *See, e.g.*, N. Salkind, ENCYCLOPEDIA OF MEASUREMENT AND STATISTICS (SAGE Publications 2006) at 938. Small samples also reduce the power of statistical tests, leading to a large Type II error.

39 A. Rubin, STATISTICS FOR EVIDENCE-BASED PRACTICE AND EVALUATION (Cengage Learning 2012) at p. 214.

40 *IP Index Report, supra* n. 4 at p. 25. "Significant" results are measured at the 1% level and "Borderline" results at the 5% level.

41 Abdi, *supra* n. 12.

NOTES CONTINUED:

42 These values are based on the binomial distribution with a 95% confidence interval.

43 A large number of rejections may be obtained by the careful selection of data. For example, if economic variables are defined in levels, then the size of the countries in a sample may produce numerous statistically-significant correlations.

44 Each variable is drawn randomly from a total of 10,000 random numbers from the normal distribution.

45 Abdi, *supra* n. 12, at Table 2 (the threshold correlation value of ±0.5714 to indicate significance).

46 Given the small samples, the bootstrapped confidence interval is used.

47 The actual number of rejections has a distribution (i.e., it will vary simulation by simulation).

48 In Table 12, the *Report* shows 12 statistically significant results (from 154 tests, due to the exclusion of the median measures) at the 5% level, of which 4 are discarded for undecipherable reasons. (*See IP Index Report, supra* n. 4 at p. 25: "All of the historical data plots for the highlighted cases were checked for each country in order to determine [whether the summary statistic used was suitable.] If a single country plot (out of seven) did not fit the summary statistic, the result was rejected.") Of the 12 results, 8 are tainted by the currency problem.

49 This statement is based on a statistical test using the binomial distribution with 231 trials, a 5% probability of failure, and a 95% confidence level. The simulation presented in this PERSPECTIVE could also be used for the statistical test.

50 *IP Index Report, supra* n. 4 at p. 25.

51 *Id.* at p. 25.

52 D. Shasha and M. Wilson, STATISTICS IS EASY! (Morgan and Claypool 2010) at p. 12; *see also* Sommer and Sommer, *supra* n. 12.

53 *IP Index Report* (initial release) at Tables 14, 17.

54 J. Best, MORE DAMNED LIES AND STATISTICS: HOW NUMBERS CONFUSE PUBLIC ISSUES (University of California Press 2004) at p. 157.

55 R. Somerville, *Devious deception in displaying data: Cherry picking*, *Science or Not* (March 3, 2012) (available at: http://scienceornot.net/2012/04/03/devious-deception-in-displaying-data-cherry-picking).

56 That's 9 for the 261 tests using the full sample and 13 for the 231 tests conducted without Japan.

57 *IP Index Report, supra* n. 4 at p. 1.

58 *IP Index Report* (March 2015 version) at p. 25.

59 Replication indicates 11 rejections (not 7) for the mean and growth rate variables (154 tests).

60 *IP Index Report, supra* n. 4 at p. 1.

61 http://www.nationmaster.com/country-info/stats/Crime/Gun-ownership/Guns-per-100-residents/2007; http://en.wikipedia.org/wiki/Number_of_guns_per_capita_by_country.

62 The full sample has 13 statistically-significant correlations while the Japan-excluded sample has 35 statistically-significant correlations.

63 *IP Index Report, supra* n. 4 at p. 3.

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES
5335 Wisconsin Avenue, NW, Suite 440
Washington, D.C. 20015
Tel: (+1) (202) 274-0235 • Fax: (+1) (202) 318-4909
www.phoenix-center.org

NOTES CONTINUED: