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

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES

What is the Effect of File Sharing on the Creation of New Music? A Critical Review of "*A Case Study of File Sharing and Music Output*"

George S. Ford, PhD*

March 6, 2014

Introduction

Earlier this year, the music industry reached a milestone—it had sent its 100 millionth piracy notice to Google.[1] This shockingly large number is but one piece of the portfolio of evidence pointing to how rampant piracy has become and how it threatens the creative arts. In the decade after the 1999 introduction of file sharing, global revenues for the music industry had been halved.[2] While recent years have shown some promise for revenue stability and recovery—driven in part by the more aggressive enforcement of copyright and the proliferation of legal alternatives[3]—piracy remains a significant problem for the music, film, and print industries. Congress is now actively reviewing copyright law, with some hoping to strengthen copyright's protections while others aim to weaken them.

As the debate heats up, an increasing amount of research effort is being devoted to the study of copyright, with significant attention directed at the effects of piracy. Intuitively, the first-order effect of piracy is to reduce the returns on investments made in the arts and sciences, and thus discourage such investments (of time and money), slowing the creation of new works.[4] This logic is embedded in the Copyright Clause of the U.S. Constitution, which empowers Congress "to promote the Progress of Science and useful Arts" by granting authors and inventors an "exclusive right" to their works.[5] There are, however, studies that challenge this traditional and constitutional view by arguing that piracy (an infringement on the exclusive right) has not hurt and, possibly even helped, professional artists and performers. Yet, this counter-intuitive claim has little, if any, sound empirical support.[6]

> ***... the empirical analysis in Dr. Lunney's paper is some of the weakest in this area, and the defects in the analysis are many and varied. Indeed, Dr. Lunney's analysis suffers from defects so severe as to render it useless for guiding public policy.***

For example, in a recent and unpublished study—*Empirical Copyright: A Case Study of File Sharing and Music Output*—Tulane University Law Professor Glynn Lunney, Jr., concludes that "file sharing has not reduced the creation of new original music."[7] The claim is based on the correlation of music sales over time to the appearance of "new artists", narrowly defined, appearing at the top of BILLBOARD's Hot 100 chart. In this PERSPECTIVE, I provide a review of Dr. Lunney's paper. Unfortunately, the empirical analysis in Dr. Lunney's paper is some

of the weakest in this area, and the defects in the analysis are many and varied. Indeed, Dr. Lunney's analysis suffers from defects so severe as to render it useless for guiding public policy. Not only does Dr. Lunney use an unsound measure of music output—a select group of hit songs chosen under different standards—but then he applies statistical techniques certain to produce meaningless results. Put plainly, his statistical analysis is inexpertly performed; the empirical model is poorly motivated, poorly designed, and improperly estimated. Moreover, contrary to his claim, his results do not support his theory. Alternative and more plausible interpretations of Professor Lunney's results suggest piracy has the expected negative consequences on the creative industries.

Professor Lunney's Model

Dr. Lunney's research question is straightforward: "*has the rise of file sharing and the parallel decline in record sales increased or decreased the creation of new music?*"[8] He proposes that the answer lies in comparing the sizes of two marginal effects (which I label ME1 and ME2 for convenience):

> ME1*: The first is whether the decline in revenue has led to fewer new artists?*
>
> ME2*: The second is whether the decline in revenue has led to increased or decreased output from existing artists?*

According to Dr. Lunney, the first marginal effect—lower revenues leads to fewer new artists (ME1)—is based on the typical incentive argument, where the resources attracted to the creative industries increase with the return on investments.[9] Dr. Lunney assumes that revenues from music sales are a good proxy for the total returns on music production, thus ignoring costs. According to ME1, piracy is expected to reduce new entry into the creative industries. (Similarly, it may be expected to lead to more exit by established artists, although this is not part of his analysis.)

The economic logic of the second effect (ME2) consists of two opposing influences. Like the first, a reduction in the returns on investments may discourage existing artists from making new records. Countering this traditional effect is the argument that the realization of high incomes from music sales make artists select leisure rather than work, thereby reducing creative output. According to Dr. Lunney, this response to high returns is akin to the backward-bending supply curve of labor economics, which holds that if wage rates get high enough, then a laborer may actually earn a higher income yet reduce the quantity of hours worked to consume more leisure.[10]

Put plainly, [Professor Lunney's] statistical analysis is inexpertly performed; the empirical model is poorly motivated, poorly designed, and improperly estimated. Moreover, contrary to his claim, his results do not support his theory. Alternative and more plausible interpretations of Professor Lunney's results suggest piracy has the expected negative consequences on the creative industries.

For his empirical test of these opposing effects, Dr. Lunney measures "music output" in multiple ways, but really just two of the measures are used to support his conclusion. First, he has a measure he calls "new artists", which is defined as artists appearing for the first time in the top fifty of BILLBOARD's Hot 100 chart, and doing so with their first release from their first album (labeled *N* here). This figure is annualized by summing such appearances from data collected in the first week of each month. The data covers the period 1985 through 2013. This measure of output is intended to test ME1.

Second, after their first appearance, these new artists become established artists, for which Dr. Lunney computes the number of top fifty hits over the next ten-years (I'll label this hit count *H*). The hit count is averaged by year for all new artists in that year. This output measure is intended to test ME2.

A little explanation may be helpful. Say the band "Bad Study" appears for the first time, with the first song from their first album, at spot 30 on BILLBOARD's Hot 100 chart in January 1990. This appearance adds one unit to *N* for the year 1990. Over the next ten years, "Bad Study" has six top-fifty hits, and these hits add 6 to the computation of *H* (an average) for the year 1990.[11] To maintain the ten-year window, the data for *H* ends in 2005 (so the last year sums only over nine-years).

Dr. Lunney's empirical model can be stated (in a simplified form) as:

$$N_t = \beta_0 + \beta_1 R_t + \beta_2 F_t + \beta_3 X_t + u_t, \qquad (1)$$

$$H_t = \alpha_0 + \alpha_1 R_t + \alpha_2 F_t + \alpha_3 X_t + v_t, \qquad (2)$$

where *R* is music sales in the U.S. in all formats, *F* is a dummy variable for file sharing equal to 1.0 after year 1999, *X* is a collection of "other" factors that may determine output including two measures of national income, u_t and v_t are econometric disturbance terms, and *t* is an index of time. The data is annual, so the model is a time-series model consisting of 29 data points for Equation (1) and 21 data points for Equation (2). Based on Dr. Lunney's setup, the expectation is that *N* rises in *R* ($\beta_1 > 0$ by hypothesis ME1), but *H* may rise or fall with increases in *R* ($\alpha_1 <> 0$ by hypothesis ME2). For a given change in *R*, if *N* falls but *H* increases, then "music output," as defined here, could rise or could fall, depending on the relative sizes of *N* and *H*.

Equations (1) and (2) are linear models estimated using ordinary least squares ("OLS") regressions with the variables in their levels. (Log specifications are also estimated, but the focus of his work is on the estimates using levels data.) For *well-known* reasons (discussed later), this estimation approach is entirely inappropriate.[12] After obtaining the estimates of the β and α coefficients by OLS, the effect on "music output" is computed by Dr. Lunney as follows.

[T]he calculation of the "Hot 100 chart" had changed numerous times in significant ways over the period of Dr. Lunney's study, yet he fails to either acknowledge or account for such changes. Any observed changes in the data over time are as likely to be a consequence of the Hot 100's changing definition as it is music sales.

Dr. Lunney attributes the entire revenue decline in music sales since 1999 to file sharing ($13 billion). The estimate of β_1 is 0.000675, which when multiplied by the $13 billion reduction in sales indicates a loss of 8.77 artists per year. The estimate of α_1 is -0.0002, so a $13 billion decline in revenues increases the number of hits by the new artists over the ten-year window by 2.62 songs. Note that the (unconditional) average over the entire sample is 30.2 new artists per year which have 3.5 hits over the ten-year window. To compute the "effect" of file sharing, Dr. Lunney multiplies the 8.77 reduction in new artists by their unconditional expected hit count of 3.5 to render a reduction in hits by new artists of 30.7 songs annually. Offsetting this reduction in an increase in hit production over the "life" of the new artists (or ten years of it) by 2.62 songs. Multiplying 2.62 by 30.2 artists, Dr. Lunney concludes that hit production rises by 79.1 songs annually. The net effect of file sharing, therefore, is 48.4 hit songs

(= 79.1 – 30.7). As such, Dr. Lunney concludes "file sharing has not reduced the creation of new original music."[13]

Review of the Lunney Study

Every so often I run across a study where pretty much everything is done incorrectly; Dr. Lunney's paper is unfortunately one of these. A problem with such studies is that the volume of errors makes it difficult to give a real sense to the reader, and particularly the lay reader, of how truly defective the analysis is in a concise and not too technical way. I will try, nevertheless, to do so by focusing on a few of the more glaring defects.

To begin, the calculation of the "Hot 100 chart" had changed numerous times in significant ways over the period of Dr. Lunney's study, yet he fails to either acknowledge or account for such changes. Any observed changes in the composition of the chart over time are as likely to be a consequence of the Hot 100's changing definition as it is music sales. Next, I will demonstrate that all the statistically-significant results Dr. Lunney reports are, by his own analysis, spurious; they all disappear once the proper estimation method, given the nature of his data, is applied. I will then explain why his measure of "music output" is not a measure of output at all. Next, for argument's sake, I will show why his reported (yet invalid) results do not (and cannot) support his theory, but rather support the first-order expectation that file sharing has reduced the creation of new works. Finally, I will list a few of the many other defects in Dr. Lunney's statistical analysis.

The Changing Hot 100

While shipment and sales data is likely a better (though imperfect) measure of music output, Dr. Lunney chooses to measure music output using BILLBOARD's Hot 100 chart because he claims the data on "shipment and sales across the pre- and post-file sharing eras unreliable and inconsistent."[14] But, Dr. Lunney argues, "in the pre-file sharing era and the post-file sharing era, a radio station has the same financial interest in satisfying the musical demand of its listeners."[15] This view is based on the idea that the Hot 100 is based on radio play and not sales. In fact, the Hot 100 chart is based heavily on music sales. As stated on the BILLBOARD website,

> The week's most popular current songs across all genres, ranked by radio airplay audience impressions as measured by Nielsen BDS, sales data as compiled by Nielsen SoundScan and streaming activity data from online music sources tracked by Nielsen BDS. Songs are defined as current if they are newly-released titles, or songs receiving widespread airplay and/or sales activity for the first time.[16]

BILLBOARD targets the shares of inputs as follows:

> … our Hot 100 formula targets a ratio of sales (35-45%), airplay (30-40%) and streaming (20-30%).[17]

Thus, in constructing the Hot 100 Chart, Billboard intends for radio sales to play a smaller role than music sales. In one particular week, BILLBOARD notes that,

> This week, points for the Hot 100's leader stem 50% from streaming, 43% from sales and just 7% from radio airplay.[18]

For Katy Perry's "Dark Horse," BILLBOARD notes that "99% of its points are from sales."[19] Thus, if music sales are "unreliable and inconsistent," as Dr. Lunney claims, then so is the Hot 100 chart.

Moreover, the way the Hot 100 is compiled has changed over time, and in many significant ways. For example, BILLBOARD instituted a significant change in the Hot 100 in 1998, which is about the same time file sharing became available. Prior to 1998, the Hot 100 was limited to songs that could be purchased as a single. Since "singles" declined in popularity in the

1990's, BILLBOARD changed its formula to measures "songs" rather than "singles." This change was significant one.

In 2005, BILLBOARD began including digital downloads in its formula, and given the volume of such sales, the change substantially impacted the chart. In 2007, streaming data was added to the mix. In 2013, YouTube viewing was added as an input to the chart's formula, which has a material impact on the relative popularity of certain songs (e.g., the "Harlem Shake") and which some fear will lead to chart manipulation (which is not a new phenomenon for the Hot 100).[20] Over time, there have also been material changes regarding the treatment of "remixes" and "recurrents," the latter reflecting modifications to the chart to down weight songs that have appeared on the chart for many weeks.[21]

Clearly, Dr. Lunney's reliance on the Hot 100 is in error. First, while Dr. Lunney argues that sales are a biased measure of output, the Hot 100 is highly dependent on sales. Dr. Lunney mistakenly believes that the Hot 100 is primarily determined by radio play—it is not. Second, the definition of the Hot 100 is dynamic and changing materially over time; Dr. Lunney makes no attempt to adjust the data to reflect such changes. Thus, the observed changes in the composition of the Hot 100 chart over time may merely reflect the changes in its definition, rather than a response to overall music sales as Dr. Lunney claims. For very many reasons (others discussed later), the Hot 100 is a meaningless measure of music output over time.

Improper Estimation Technique

In regression analysis, time-series data requires special attention because the past often largely determines the present. The properties of the data—how the series move about over time—determine, in part, the proper estimation method of a statistical model. A key property of time-series data is whether or not the series are "stationary." Some properties of a stationary process are that its mean and variance do not change over time and the process does not follow any trends. Or, for a series to be stationary, it must return over time to a constant mean, so it cannot move up or down over time. If a data series is non-stationary, then the typical response is to first-difference the data (that is, subtract last period's observation from the current period's observation), since the first difference of a non-stationary series is often stationary.

Clearly, Dr. Lunney's reliance on the Hot 100 is in error. First, while Dr. Lunney argues that sales are a biased measure of output, the Hot 100 is highly dependent on sales. Dr. Lunney mistakenly believes that the Hot 100 is primarily determined by radio play—it is not. Second, the definition of the Hot 100 is dynamic and changing materially over time; Dr. Lunney makes no attempt to adjust the data to reflect such changes. Thus, the observed changes in the composition of the Hot 100 chart over time may merely reflect the changes in its definition, rather than a response to overall music sales as Dr. Lunney claims.

The reason economists are concerned with stationarity is that estimating a model with non-stationary data typically renders spurious results, capturing trends and other data properties rather than revealing the genuine relationship between series.[22] Over time, things tend to grow (or shrink), and a regression of one growing series on another, regardless of whether they are truly related, will often

indicate a correlation simply due to the common trends. One example is the observed correlation of bread prices in Britain and sea levels in Venice; nominal bread prices rise along with the sea level, but one would not use this fact to argue that one variable causes the other. By using a differenced series, the correlation is computed based on the changes in the series over time, which is more informative. Using a standard estimation method like OLS in the levels (or logs) of the data when the data is non-stationary is a gross error—the estimates of the model are utterly meaningless. Walter Enders, in his classic APPLIED ECONOMETRIC TIME SERIES, states it formally as follows:

> The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order and the residual sequence contains a stochastic trend. This is the case in which the regression is spurious. The results from such spurious regressions are meaningless in that all errors are permanent. In this case, it is often recommended that the regression equation be estimated in first differences.[23]

Dr. Lunney states in his paper that "the dependent variables do not appear to be stationary."[24] It is well known that GDP data is non-stationary in the levels (yet stationary in first differences). As a result, the simple OLS regression on the levels of the series, which is what Dr. Lunney employs, is entirely inappropriate for his data. Consequently, his econometric estimates are meaningless.

Significantly, Dr. Lunney makes a casual reference to what happens when he does estimate the model in first differences, where he states "all of the correlations become statistically insignificant."[25] So, when his model is estimated using a technique more suited to his data, Dr. Lunney finds nothing. This result is a severe indictment against his study.

Dr. Lunney makes two peculiar statements about the non-stationarity problem that suggest a lack of experience in statistical analysis. First, he states he transforms the data using the natural log function to "account for [non-stationarity]," but a knowledgeable statistician knows that such a transformation does not account for non-stationarity.[26] Differencing the data is the solution. Second, Dr. Lunney claims that the failure of the first-differenced model "tends to confirm our interpretation of the results."[27] Yet again, an experienced statistician knows that this claim has zero support from econometric theory.

Dr. Lunney states in his paper that "the dependent variables do not appear to be stationary." It is well known that GDP data is non-stationary in the levels (yet stationary in first differences). As a result, the simple OLS regression on the levels of the series, which is what Dr. Lunney employs, is entirely inappropriate for his data. Consequently, his econometric estimates are meaningless.

The risk of using simple least squares regression with non-stationary data can be illustrated with an admittedly ridiculous example. Say, in an effort to encourage strong copyright protection, someone argues that higher music sales encourage employment in U.S. manufacturing. If this were so, then I would think politicians would be very interested in reducing piracy. If I estimate a model like Dr. Lunney's, testing whether music sales determine manufacturing employment (years 1985-2012), then the estimated regression equation looks like this,

$$E = 11.7 + 0.24R - 0.46Y + u \qquad (2)$$

where E is the natural log of manufacturing employment, R is the natural log of music sales, Y is the natural log of real Gross Domestic Product ("GDP"), and u is the econometric disturbance term. Both R and Y are statistically-significant determinants of E, with both coefficients having t-statistics greater than 10 (in absolute value). The regression says that an increase in music sales is associated with an increase in manufacturing employment; increases in GDP, however, are associated with reduced manufacturing employment. Neither result makes sense; both are simply reflecting trends in the data.

Dr. Lunney does not estimate his model properly; if he did, his own analysis would imply that the key statistical results of his paper disappear.

As with Dr. Lunney's data, none of the variables used in this model are stationary in their levels, but all are stationary in first differences. Re-estimating the model in first differences renders the regression equation,

$$\Delta E = -0.05 + 0.05\Delta R + 1.27\Delta Y + u \qquad (2)$$

where the coefficient on ΔR is not statistically significant (one can't reject the hypothesis that the true relationship between music sales and manufacturing employment is zero), and the positive sign on GDP (Y) is now more in keeping with expectations (higher economic output is associated with higher employment) and still statistically significant. This exercise demonstrates that, when using time-series data, the estimation technique must account for the properties of the data. Spurious results are quite common with time-series data. Dr. Lunney does not estimate his model properly; if he did, his own analysis would imply that the key statistical results of his paper disappear.

Another important consideration when using times series data is whether or not the series are cointegrated.[28] Dr. Lunney fails to even mention cointegration and provides no tests of it. Without any information from such tests, I cannot comment on the implications for his dataset. Thus, I merely mention the fact that the failure to evaluate the presence of cointegration is one more defect of Dr. Lunney's analysis.

When Output is not Output

The stated purpose of Dr. Lunney's paper is to test whether "file sharing and the parallel decline in record sales increased or decreased the creation of new music?"[29] The first question to ask, then, is whether the appearance of "new artists," narrowly defined, in the top fifty spots of BILLBOARD's Hot 100 is a meaningful measure of new music output? Plainly not, no more than the player of the week is a measure of the resources devoted to a football league. Scarcely any "new music" appears in the BILLBOARD's Hot 100 chart, and many artists have no desire to serve the "pop music" market,[30] where songs like "What Does the Fox Say?" measure artistic genius.[31]

Scarcely any "new music" appears in the Billboard's Hot 100 chart, and many artists have no desire to serve the "pop music" market, where songs like "What Does the Fox Say?" measure artistic genius.

Indeed, the top fifty of the Hot 100 list is the narrowest of slivers of music output. Take, for example, the first week of December 2013. In that week, there were 50 top fifty songs, yet there were about 850 reported album releases by the recording industry, representing perhaps about 8,000 songs.[32] Dr. Lunney's definition of "new music" implies there are only 105 new songs per year.[33] In 2011, data indicates that

there were 76,875 album releases, which probably represents about a million individual tracks (and this number includes only those tracks that are officially measured).[34]

Furthermore, the top fifty list cannot, by definition, measure output because there is always fifty songs in it. It is a count that neither goes up nor down. And, by its nature as a ranking, there will always be a top fifty, regardless of how much music output is being produced or the quality of such music.

Dr. Lunney's measure of "music output" is really a measure of the "artist composition" of an infinitesimal subset of new music. Recognizing this, his results (which again are not legitimate) prescribe a very different conclusion than the one Dr. Lunney provides. Suppose we ignore all the econometric issues and take his regression results and data definitions as valid. At least two alternative meanings of his reported results are possible.

To begin, the first-order effect of file sharing and piracy is to reduce the return on investments in creative works, including music. This reduction in returns is expected to reduce entry into the music industry by new artists. As entry falls and competition from new entry wanes, it becomes easier for established acts to have success (measured, say, by reaching the Hot 100 list). Consequently, the expectation of a rise in file sharing is that the number of new artists will fall (as Dr. Lunney concludes) and the success rate of established artists will rise (as Dr. Lunney concludes) as competition softens. Dr. Lunney's findings do not imply a backward bending supply curve, but rather are entirely consistent with the expectations of the standard theory of copyright: *Piracy reduces the creation of new works.*

Second, recall that *H* is defined as how often a particular group of "new artists" are able to hit a top fifty spot in the Hot 100 chart after their first hit. There is, however, another group of untracked artists that are also trying to hit the same top-fifty list. This group of artists, while successful, does not meet Dr. Lunney's strict criteria for a "new artist." These "untracked artists" write songs and compete with the "new artists" group for hits.

Suppose music sales rise and the efforts and entry of the "untracked artists" increase by more than the tracked "new artists." The increased competition would result in a reduction in how often the "new artists" group was able to hit the top fifty spots. Hence a negative correlation between the variable *H* and music sales would be generated, yet the narrow sliver of artists being tracking in *H* still had a positive response in effort, production, and entry to music sales (the supply curve slopes upward); the "untracked artists" simply had a greater response to sales than the tracked group. In this scenario, all supply curves slope upward and do not bend backward, yet the expectations match Dr. Lunney's empirical results.

Dr. Lunney's findings do not imply a backward bending supply curve, but rather are entirely consistent with the expectations of the standard theory of copyright: Piracy reduces the creation of new works.

In neither of these alternative interpretations of Dr. Lunney's results is there a justification or benefit to society of allowing the piracy of music through file sharing sites, thereby destroying record sales. To me, these alternatives are much more plausible than Dr. Lunney's claim that artists produce less output when they get paid more for that output. In any case, his results need not confirm his conceptual framework, and statistically his results are entirely meaningless.

A Condensed List of a Long List of Errors

As noted above, the errors in Dr. Lunney's statistical analysis are too numerous even to mention them all. Instead, I've listed some of the problems below with brief explanations.

Sample Size. Dr. Lunney's conclusions are based on only 21 annual observations for the *H* equation and 29 annual observations for the *N* equation. Obviously, this is not a lot of data. To reach the conclusion that "file sharing has not reduced the creation of new original music" with so little data, and encourage the U.S. Congress to make policy on such findings, is questionable.[35] While Dr. Lunney recognizes that he has "a very small data set,"[36] this fact fails to temper his conclusions. Moreover, in both cases, but particularly in the case of 21 observations, the reported t-statistics from the statistical package used are dubious measures of statistical significance. Dr. Lunney's sample is a "small sample," and procedures should be used which address that fact. (Of course, all this presumes his estimation method is appropriate, which it plainly is not.)

A Backward Bend. As noted above, the core of Dr. Lunney's conceptual framework is the concept of a backward bending supply curve, where the response of labor supplied to an *increase* in the wage rate may be positive or negative depending on the level of the wage rate. Thus, the relationship between wages and effort is non-linear and non-monotonic. Dr. Lunney's model is a linear one (in the parameters) and, as such, permits *only* a positive or negative relationship to exist, not both depending on the "wage" rate (or revenues in his model). Consequently, his empirical model cannot capture the very relationship he proposes exists. If you are looking for a backward-bending supply curve, then you better estimate a curve that can bend backward.

Income or Music Sales. The amount of effort put forth by an artist is a function of his or her entire compensation, and music recording sales are only one of many sources of income for artists. Some artists—including Taylor Swift, Pink, and Queen Latifah—make income as CoverGirl models.[37] Other artists have clothing lines, or appear as judges on televised talent shows. Most of these endorsements and side-jobs are driven by their popularity in the music industry, so the continuation of such income requires continued effort in musical recordings. Also, some artists make a large share of their income touring. U2's 360° tour grossed nearly $750 million dollars.[38] By including only music sales in the regression, Dr. Lunney's model is mis-specified. Whether we view *R* as a mis-measurement of income or view this "other income" as an omitted variable separate from *R*, Dr. Lunney's estimates are theoretically biased.

If you are looking for a backward-bending supply curve, then you better estimate a curve that can bend backward.

Counting Artists. The dependent variable *N* is a count of new artists. Count data has special distribution properties and thus requires special attention. If the counts are small, then an estimation method like Poisson or Negative Binomial Regression is recommended.[39] In some cases, particularly if the count is large, a natural log transformation is suitable to the task. In either case, Dr. Lunney estimates the count in its level form, which is not proper technique and prone to render poor estimates.

Outside the Range. When using regression coefficients to make predictions, theory implies that the most accurate interpretation of computed changes are obtained when small changes are computed from the mean of the data. Dr. Lunney evaluates very large changes far from the mean of the data. The mean of music sales over Dr. Lunney's sample period is about $13 billion, with a minimum of about $7

billion and a maximum of $20 billion. Thus, his computation of changes using a $13 billion figure is a change equal to the full range of the data, and if evaluated from the mean, will lie well outside the range of the data—an econometric no-no. Marginal effects from such a large change will have very large confidence intervals, so it is unclear whether the change in "output" Dr. Lunney computes is statistically different from zero. I doubt it, but I can't say it definitively without the data.

Hit Predetermination, Sales Disconnect. The *H* variable from Equation (2) is computed as the sum of hits realized by a "new artists" over a ten-year horizon. So, if the band "The Bunglers" had a first-hit in 2000 (becoming a "new artist") and hits in each year thereafter (to 2013), the *H* variable rise by 13 in year 2000. Of course, sales in 2000 say nothing about the incentive to produce hits in 2001, 2002, …, 2013. In effect, Dr. Lunney has assumed that the incentive to work is determined by market conditions at the time of an artist's first hit. This assumption is senseless. There are many other problems related to the definition of *H*, including (but not limited to) the arbitrary choice of ten-years and the arbitrary horizon of nine-years for the last group of new artists.

Related or Unrelated. Dr. Lunney's model assumes that the flow of new artists and hit production by them are related to music sales (and other factors), and presumably *H* and *N* are generated from a related process. Yet, the two equations are estimated independently. An experienced statistician would likely have estimated the model using a technique known as *Seemingly Unrelated Regression* (though an experienced statistician would not have estimated his model in the first place) in which the equations are estimated jointly.

A File Sharing Dummy. Dr. Lunney includes in his model a dummy variable equal to one after 1999 to account for the emergence of file sharing. However, file sharing is not an on-off proposition. File sharing started small and has risen substantially since its appearance. A dummy variable is not the right approach to measure file sharing.

The purpose of copyright is to promote the creation of new works. It is not, as Dr. Lunney appears to believe, to promote exclusively the creation of "new artists" and their Hot 100 hit count—especially when "new artist" is defined in such a narrow way. Copyright aims to encourage creative output by established artists, new artists, and even artists that will never have a song appear in the Hot 100 chart. Once such material is created, the mix of commercial success between new and established artists is then determined by the market. Commercial success is not the point of copyright—creation is.

Lagging Away a Problem. In many econometric models, what is being determined (the dependent variable) is arguably also determining the explanatory variables. If two variables are jointly determined, then they are said to be endogenous. Dr. Lunney expresses some concern about the endogeneity of the composition of the top fifty Hot list, and he claims the following: "to avoid endogeneity, I used one-year lagged values for record sales, gross domestic product, and gross national income per capita."[40] Using lagged values is commonly used in econometrics to avoid endogeneity, but the lagging approach is not valid if the series are non-stationary. A non-stationary series is "persistent," so that this

year's value is pretty much the same as last year's value. Thus, if this year's data is endogenous, the odds are so is last year's data. Lagging is no solution.

Miscalculated Marginal Effect. Above I provided description of how Dr. Lunney computed the effect on "new music" using his econometric results. His computation is wrong, and for many reasons, of which I'll mention two. First, such changes are best computed from the mean and the proper comparison is the change in both series from one revenue figure to another. Second, while he claims to seek the effect of "file sharing *and* the parallel decline in record sales," he ignores his estimated coefficients on the direct effect of file sharing.

For example, the sample means of *N* and *H* are 30.2 and 3.5, indicating an "annual" hit production of 105.7 songs. With a $13 billion dollar change (which, from the sample, essentially takes revenues to zero), the number of artists shrinks by 8.77 units and the number of hits rises by 2.62 units. The new "annual" hit count is 21.4 artists multiplied by 6.1 songs, or 127.3 songs. Thus, the change in hits is not 48 songs, but 21.6 songs (less than half the number calculated by Dr. Lunney).

But that's not all. Dr. Lunney states that he wishes to evaluate the effect on new music from "the rise in file sharing and the parallel decline in record sales." As shown in Equations (1) and (2), his models include a dummy variable for file sharing. These coefficients are relevant to the answer he claims to seek. Including the coefficients from these dummy variables in the calculation, the change is hit count is now only 7.3 songs (105.7 versus 113). This is a very small change in hits (and likely not statistically significant) for a $13 billion change in music sales.

Those opposed to copyright, on whatever grounds, surely bear a substantial burden in making their case. The general principle that payment to producers encourages production, so far as I know, is not under attack in any venue except copyright. Claims of special and unique circumstances, though, require extraordinary evidence—a burden that Professor Lunney has totally failed to meet.

Missing the Point. The purpose of copyright is to promote the creation of *new works*. It is not, as Dr. Lunney appears to believe, to promote exclusively the creation of "new artists" and their Hot 100 hit count—especially when "new artist" is defined in such a narrow way. Copyright aims to encourage creative output by established artists, new artists, and even artists that will never have a song appear in the Hot 100 chart. Once such material is created, the mix of commercial success between new and established artists is then determined by the market. Commercial success is not the point of copyright—creation is. As stated by the artist Todd Rundgren,

> [E]verybody loves music, but I think many people don't realize that musicians, aside from making a living and doing what they do, provide a service to the rest of humanity in some sense. People get comfort from music.[41]

By focusing alone on a very narrow definition of commercial success rather than the creation of *new works of art*, Dr. Lunney's analysis entirely misses the point.

Conclusion

As the copyright review debate heats up, so will the amount of academic research on the effects of copyright, piracy, fair use, and other relevant concepts. Given the importance of intellectual property, it is vital that any research used to formulate policy be subjected to close scrutiny. In this PERSPECTIVE, I have reviewed a recent paper on the effects of piracy on the music industry by Professor Glynn Lunney which purports to show that "file sharing has not reduced the creation of new original music." As I have demonstrated here, not only is this conclusion unsupported, but Professor Lunney's analysis suffers from defects so severe as to render it useless for guiding public policy.

One can go farther: copyright is established in the U.S. Constitution on the theory that one obtains more goods and services when one pays for them. Those opposed to copyright, on whatever grounds, surely bear a substantial burden in making their case. The general principle that payment to producers encourages production, so far as I know, is not under attack in any venue *except* copyright. Claims of special and unique circumstances, though, require extraordinary evidence—a burden that Professor Lunney has totally failed to meet.

NOTES:

[*] **Dr. George S. Ford** *is the 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.*

[1] C. Sherman, *Google's 100 Million Notices*, THE HILL (January 13, 2014) (available at: http://thehill.com/blogs/ballot-box/195107-googles-100-million-notices); *Understanding the Role of Search in Online Piracy*, Millward Brown Digital, Prepared for the MPAA (2013) (available at: http://www.mpaa.org/resources/38bc8dba-fe31-4a93-a867-97955ab8a357.pdf).

[2] Recording Industry Association of America (RIAA), U.S. Recording Industry Shipment Statistics (available at http://www.riaa.com/keystatistics.php?content_selector=research-shipment-database-overview).

[3] V. Luckerson, *Revenue Up, Piracy Down: Has the Music Industry Finally Turned a Corner?*, TIME.COM (February 28, 2013) (available at: http://business.time.com/2013/02/28/revenue-up-piracy-down-has-the-music-industry-finally-turned-a-corner); N. Perry, *Popular File-Sharing Website Megaupload Shut Down*, USA TODAY (December 2, 2012) (available at: http://usatoday30.usatoday.com/tech/news/story/2012-01-19/megaupload-feds-shutdown/52678528/1).

[4] T.R. Beard, G.S. Ford, L.V. Stern and M.L. Stern, *Theft and Welfare in General Equilibrium: A Theoretical Note,* 2 THEORETICAL ECONOMIC LETTERS 470 (2012) (available at: http://www.phoenix-center.org/papers/TheftandWelfareEconLetters.pdf).

[5] *See* Article I, Section 8, Clause 8 of the United States Constitution.

[6] *See, e.g.,* S.J. Liebowitz, *How Reliable is the Oberholzer-Gee and Strumpf Paper on File-Sharing?* Unpublished Manuscript (September 1, 2007) (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1014399); G.S. Ford, PHOENIX CENTER POLICY PERSPECTIVE No. 13-05, *Piracy and Movie Revenues: A Critical Review of "A Tale of the Long Tail"* (September 16, 2013) (available at: http://phoenix-center.org/perspectives/Perspective13-05Final.pdf).

[7] G.S. Lunney, Jr., *Empirical Copyright: A Case Study of File Sharing and Music Output*, TULANE UNIVERSITY SCHOOL OF LAW PUBLIC LAW AND LEGAL THEORY WORKING PAPER SERIES NO. 14-2 (January 2014) (available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2372630) at p. 4.

[8] *Id.* at p. 14.

[9] *See*, *e.g.*, W. Landes and R. Posner, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW (2003), at Ch. 3; Beard *et al.*, *supra* n. 4.

[10] Since revenues do not measure wages and Dr. Lunney is not evaluating hours worked, the nexus between the labor supply curve and his analysis is hazy. In support of the ME2, Lunney points to the 19th century Italian composer Giuseppe Verdi as an example of such a response to copyright. Copyright was introduced during Verdi's productive life, and his output prior to copyright was greater than that after, and apparently some of his correspondence suggests that the higher income derived after copyright permitted him to reduce his output. However, it is also true that some of Verdi's later works are considered by some to be his best, so the effect may reflect a quality-for-quantity tradeoff more than pure quantity effect. The Verdi example is an interesting one, but it is easily countered with numerous examples of wealthy artists that continue to produce output and tour aggressively. Taylor Swift, for example, earned $55 million in 2013 yet is a highly productive musician, releasing five albums in six years and touring regularly. If $55 million isn't enough to make a musician reduce effort and consume more leisure, then the backward-bending supply curve seems a dubious proposition. Of course, such anecdotes are insufficient to make any general claim about the shape of the supply curve of music industry labor—a very good empirical model is required to reveal such an effect. On Taylor Swift, *see The Celebrity 100*, FORBES (as of June 2013) (available at: http://www.forbes.com/profile/taylor-swift).

[11] I suppose the addition to *H* could be 5 songs; it is not clear from Dr. Lunney's paper exactly how the mean is computed.

[12] *See, e.g.,* C. W. J. Granger and P. Newbold, "*Spurious Regressions in Econometrics*, 2 JOURNAL OF ECONOMETRICS 111–120 (1974).

[13] Lunney, *supra* n. 7 at p. 4.

[14] *Id.* at p. 15.

[15] *Id.*

NOTES CONTINUED:

16 http://www.billboard.com/charts/hot-100.

17 http://www.billboard.com/articles/columns/chart-beat/5740625/ask-billboard-how-does-the-hot-100-work.

18 *Id*.

19 *Id*.

20 J. Dasilva, *Billboard New Hot 100 Chart Formula May Lead to More Chart Manipulation*, EXAMINER.COM (February 21, 2013) (available at: http://www.examiner.com/article/billboard-new-hot-100-chart-formula-may-lead-to-more-chart-manipulation).

21 http://en.wikipedia.org/wiki/Billboard_Hot_100.

22 Granger and Newbold, *supra* n. 12.

23 W. Enders, APPLIED ECONOMETRIC TIME SERIES (2nd) (2004).

24 Lunney, *supra* n. 7 at p. 26. Formal tests of stationarity are not provided.

25 *Id.* at p. 29.

26 I confirmed this fact by conducting unit root tests on industry revenues and GDP.

27 Lunney, *supra* n. 7 at p. 29.

28 Granger and Newbold, *supra* n. 12; Enders, *supra* n. 23. In some cases, cointegrated yet non-stationary series may be evaluated in their levels and still produce meaningful results.

29 Lunney, *supra* n. 7 at p. 14.

30 *See, e.g.*, http://www.guitarcenter.com/JoJo-Electro-No-Top-40-T-Shirt-H87493-i2628223.gc.

31 W. Gruger, *Ylvis' Viral Hit 'The Fox' Debuts on Hot 100*, BILLBOARD (September 11, 2013) (available at: http://www.billboard.com/articles/news/5687397/ylvis-viral-hit-the-fox-debuts-on-hot-100).

32 http://www.allmusic.com/newreleases/all.

33 Based on 30.2 new artists with an average of 3.5 hits each.

34 *The Nielsen Company & Billboard's 2011 Music Industry Report*, BUSINESSWIRE (January 5, 2012) (available at: http://www.businesswire.com/news/home/20120105005547/en/Nielsen-Company-Billboard%E2%80%99s-2011-Music-Industry-Report#.Uv0LbfldXNm); D. Taylor, *Do Most Music CDs Have 12 Tracks*, ASKDAVETAYLOR.COM (July 14, 2006) (available at: http://www.askdavetaylor.com/do_most_music_cds_have_12_tracks).

35 G. S. Lunney, Jr., *Statement of Glynn S. Lunney, Jr., at The Scope of Copyright*, U.S. House of Representatives, 113th Congress, Committee on the Judiciary, Subcommittee on Courts, Intellectual Property, and the Internet (January 14, 2014) (available at: http://docs.house.gov/meetings/JU/JU03/20140114/101642/HHRG-113-JU03-Wstate-LunneyG-20140114.pdf).

36 Lunney, *supra* n. 7 at p. 23.

37 A. Finlayson, *Pink is the New Face of CoverGirl*, US MAGAZINE (August 6, 2012) (available at: http://www.usmagazine.com/celebrity-beauty/news/pink-is-the-new-face-of-covergirl-201268).

38 R. Waddell, *U2's '360' tTour Gross: $736,137,344*, BILLBOARD (July 29, 2011) (available at: http://www.billboard.com/biz/articles/news/touring/1176894/u2s-360-tour-gross-736137344).

39 J. S. Long, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES (1997).

40 Lunney, *supra* n. 7 at p. 22.

41 http://ultimateclassicrock.com/todd-rundgren-interview-2014.