NET Institute*

www.NETinst.org

Working Paper #17-15

September 2017

**What is the Value of Re-use?
Complementarities in Popular Music**

Jeremy M. Watson
Boston University

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

What is the Value of Re-use? Complementarities in Popular Music

Jeremy M. Watson
Boston University

This Draft: October 1, 2017

What is the Value of Re-use? Complementarities in Popular Music

Abstract

Digitization has drastically lowered the costs of replicating and distributing music, enabling piracy on the demand side, as well as supply side re-use and recontextualization. This paper examines cumulative creativity and re-use through the release of derivative works in the popular music industry. Combining novel data on “digital sampling” and cover songs with a new proprietary Spotify data-set tracking online music streaming, I study how the introduction of a derivative work impacts the market for the underlying good upon which it is based. With my data-set covering 11,682 artists and their daily streaming demand between 2015 and 2016, I utilize a matched-sample difference-in-differences estimator to find that, on average, re-use causes a 3% increase in demand for the treated artists. This effect is significantly mediated by prominence – with the effect neutralized for highly prominent artists, while artists of low prominence have a larger 6% boost in listening. Novel re-uses of artists that have not been subject to extensive past re-use appear to have the largest effect, resulting in an average 15% increase in online streaming. These results highlight an advertising effect of re-use, suggest that derivative works have limited *ex post* competition with upstream works, and point toward the potential benefits of permissive intellectual property rights licensing.

Keywords: Copyright, Intellectual Property, Digitization, Cumulative Innovation, Technology Strategy, Content Industries

JEL Codes: L24, O31, O34.

Funding from the NET Institute (www.netinst.org) is gratefully acknowledged. I also thank Marni Greenberg and Eliot Van Buskirk of Spotify for assistance with the data for this project. All views expressed herein are solely my own. Email: jeremymw@gmail.com

1 Introduction

Copying, re-use, and appropriation are nothing new. In 1873, Johannes Brahms composed *Variations on a Theme by Joseph Haydn*, attributing the re-used theme to the Austrian composer of the prior century. Early rock and roll musicians, especially Led Zeppelin, heavily appropriated from twelve-bar blues and the works of delta blues musicians such as Muddy Waters and Robert Johnson. As digitization decreases the costs of re-using bits and pieces of music, the incidence of re-use, appropriation, and remixing has since accelerated, particularly with user generated content (Lessig, 2008). Such re-use and recontextualization has been an accepted norm in art for centuries, yet it was only with the U.S. Copyright Act of 1976 that broad restrictions over re-use were established with controversial “derivative rights” (Voegtli, 1997; Tushnet, 2004). Hence, while digitization takes replication and distribution costs to zero (Goldfarb et al., 2015; Goldfarb and Tucker, 2017), copyright creates legal barriers and transaction costs that may impede replication, especially re-use. On the one hand, there is the view that strong intellectual property rights are needed to incentivize the creation, dissemination, and further downstream licensing of high quality works (Landes and Posner, 1989). In particular, entertainment industry professionals have voiced concern that easing restrictions over re-use “would do great damage to the value of the creator’s property” (LaPolt et al., 2014). However, absent empirical evidence, many legal scholars point to the transaction costs inherent to copyright licensing, along with the changing norms around re-use in the digital age to motivate a call for policy relief (Lessig, 2008; McLeod and DiCola, 2011; Menell, 2016). Ultimately, debate around this topic in the United States must consider copyright’s stated basis, “to promote the Progress of Science and the useful Arts.”

This paper studies creative re-use in the popular music industry to understand whether re-use harms the market for the original work, or instead acts as a complementary good that increases demand for the re-used piece of music. I draw upon two new data sources to answer this question. First, a rich data-set tracking re-use in popular music through two forms: 1) “digital samples,” wherein portions of past recordings are re-used in a new sound recordings, and 2) “cover songs,” wherein the composition and lyrics of a previously released

song are re-used in a new recording of the song. I link this data with a proprietary panel data-set from Spotify, an online music streaming platform, that tracks daily streaming, or “listens,” to each artist in my data. With the maintained assumption that the precise timing of re-use is exogenous once conditioned on observable characteristics, I employ a matched-sample difference-in-differences design, matching each re-used artist to an observably similar “control” artist based upon both time variant and time invariant features. With this data and research design, I am able to causally estimate how the release and distribution of a new song incorporating existing work affects demand for the underlying good upon which it is based, as measured through online listening.

I find that, on average, re-use by a downstream artist induces a positive impact on the upstream (re-used) artist. In an unmatched sample, the release of a derivative song increases daily streaming of the re-used artist by 2.8%. These results remain robust under a matched-sample where each re-used artist is matched to one closely related control artist. Estimates from the matched-sample design suggest that re-used artists receive a 3% boost in demand on the streaming platform once their works have been re-used. I also find that low-prominence artists receive the greatest benefit from being re-used – double the baseline estimates – with the positive effects of re-use shrinking towards zero for the most prominent artists. In a similar vein, I demonstrate that this effect is most dramatic for “novel re-use,” those of artists that have not been extensively sampled and/or covered.

This paper makes three contributions. First, I contribute to the management of IP literature, particularly research on the enforcement of intellectual property rights. Much of the work in this area has focused upon patent litigation and enforcement ([Lanjouw and Schankerman, 2001, 2004](#); [Cohen et al., 2014](#)). A more recent literature in this area has turned to copyright and piracy ([Rob and Waldfogel, 2006, 2007](#)), focusing on the effectiveness of copyright enforcement and new tools to combat unlicensed online copying, from government policy ([Danaher et al., 2014](#)), to the ineffectiveness of digital rights management (DRM) ([Zhang, 2016](#)), to online takedown procedures ([Peukert et al., 2017](#); [Reimers, 2016](#)), to the importance of communication on successful *ex post* licensing ([Luo and Mortimer, 2017](#)). While much of the

copyright literature in this area has focused on demand-side copying (e.g., digital piracy), this paper instead focuses on supply-side copying and re-use. My paper contributes to this area by informing and describing the contours where licensing can result in positive spillovers to rights-holders' past catalogs of IPR, versus when downstream re-use instead results in cannibalization of productive upstream markets.

Second, this research contributes to the growing literature on cumulative creativity. [Heald \(2009\)](#) examines the use of early 20th century musical compositions in late 20th/early 21st century films, and finds no significant effect of copyright protection on the propensity that a song is re-used. Subsequent work in [Heald \(2014\)](#), however, suggests that a composition's inclusion in the public domain dramatically increases the chance that the song appears in a high-grossing movie. A recent stream of copyright literature focuses on cumulative creativity and the innovation process, examining the effect of copyright policies on re-use of copyright-protected materials. [Nagaraj \(2017\)](#) finds that *Baseball Digest* content that has lapsed into the public domain is significantly more likely to be re-used on Wikipedia, an effect that is particularly pronounced for rich content like images, rather than text. [Watson \(2017\)](#) similarly finds that court decisions restricting digital sampling have had a negative impact on re-use in popular music, highlighting hold-up and royalty stacking mechanisms at play in licensing multiple copyrights. [Gans \(2015\)](#) considers a model where original content is remixed by a downstream creator, and compares outcomes under several proposed rights regimes. This paper follows the results in [Gans \(2015\)](#) by testing one of the key assumptions of their model – that downstream re-use harms the upstream creator's profits – finding that on average, this assumption may not hold. While empirical papers in this stream typically examine the effect of copyright protection or breadth on subsequent re-use, this paper focuses on the reverse direction by measuring the consequences of re-use on the re-used work.

Finally, the results of this paper speak directly to the policy debate regarding the balance of copyright, derivative works, and fair use. Much legal scholarship has been written about copyright and derivative works following the broad derivative work right enacted with the U.S. Copyright Act of 1976 ([Goldstein, 1982](#)). Since the early days of digitization, legal scholars

have been concerned that the rights over derivative works are overly broad, and at odds with both historical norms of appropriation in art, as well as paths offered by digital technologies (Voegtli, 1997). While many authors have suggested that the derivative work right may be a particularly onerous restriction over the first amendment and free speech (Tushnet, 1996; Rubinfeld, 2002), others have pointed towards broad derivative rights as a lever to prevent undue *ex post* competition against an artist’s original creations (Abramowicz, 2005). A more recent literature has tackled the legal challenges surrounding digital sampling itself (McLeod and DiCola, 2011), as well as the related mashup phenomena (Menell, 2016). Citing royalty stacking and transaction costs when multiple copyrights are sampled, these works have called for policy relief, such as compulsory licensing, to counteract the inherent transaction costs. My research contributes to this field by providing direct empirical evidence regarding the impact of derivative works on the market for the original, demonstrating that *ex post* competition is limited, and that derivative works may often confer a positive effect on average.

This paper proceeds with a discussion of music licensing (Section 2), describes the data used (Section 3), explicates the research design and identification strategy (Section 4), reviews the results (Section 5), and ends with concluding remarks (Section 6).

2 Institutions of Music Licensing and Re-use

Information, a public good, is non-rival and non-excludable. In the absence of intellectual property rights (IPR), theory suggests that the threat of piracy and bootlegging may create insufficient incentives for content producers to bring new goods to market if they cannot recoup their initial sunk cost of production. Policymakers may intervene with copyright to bolster *ex ante* incentives, but must balance the length and breadth of rights granted with dynamic concerns for cumulative innovation, wherein new innovation builds upon prior work (Scotchmer, 1991; Green and Scotchmer, 1995; Murray and O’Mahony, 2007). Furthermore, copyright has significantly evolved since its origins in the Statute of Anne, and in the U.S. now incorporates peculiar statutory licensing schemes, a broad restriction over re-use through

“derivative works,” yet allows some forms of unlicensed re-use as “fair use.” In the United States, any piece of recorded music necessarily includes two separate copyrights – (1) the copyright of the sound recording, and (2) the copyright of the underlying musical composition.

Digital sampling, the practice of re-using portions of a prior sound recording in creating new works, can be traced back to the “versioning” production-style of Jamaican dub musicians that was subsequently brought to the U.S. by Jamaican-American DJ Kool Herc during the early stages of hip-hop music and culture (McLeod and DiCola, 2011). Disc jockeys at the time used vinyl turntables to isolate, repeat, and extend portions of a song, like the “break,” during a live event. As with jazz music (Phillips, 2013), geography undoubtedly shaped the development of hip hop, but it was technology that transformed hip hop from a live, event-driven medium into a commercially potent recorded art form with the advent of the digital sampler. Diffusing in the 1980s, digital samplers allowed record producers to rapidly isolate small portions of sound recordings, store them in memory, replay, and loop over these samples. Rather than having to create a new sound *de novo*, record producers could select the “right” sound from a vast catalog of past recordings and inspirations. By recontextualizing past recordings in a new song, producers can give their recordings a familiar element while creating a distinctively new work, balancing the similarity-differentiation tradeoff (Askin and Mauskopf, 2017). Prior to 1991, sampling was rampant in hip hop music, and often operated in an unauthorized, legal gray area (McLeod and DiCola, 2011; Watson, 2017). With commercial success, the copyright infringement lawsuits inevitably followed, particularly the landmark *Grand Upright Music, Ltd v. Warner Bros. Records Inc.* (1991) ruling that equated sampling with theft, precipitating a rapid decrease in the use of digital sampling. Following *Bridgeport Music, Inc. v. Dimension Films* (2005), the courts have provided a bright-line rule – unauthorized sampling of a sound recording will be regarded as infringement, regardless of whether the recordings are substantially similar.¹

¹However, clearly *de minimis* should apply on compositions, as one cannot hold a copyright over a single musical note. See the *Newton V. Diamond* 388 F.3d 1189 (9th Cir. 2004) opinion that a 3-note sample of a composition was *de minimis*. Recently, the 9th circuit disagreed with the judgement on *de minimis* in *Bridgeport Music*, see *VMG Salsoul, LLC v. Ciccone*, No. 13-57104 (9th Cir. 2016).

Although unauthorized sampling clearly still exists, major labels are wary to permit unauthorized re-use under threat of costly infringement claims, and thus new samples in popular music are cleared by negotiating a sample licensing agreement.² While some music licensing is highly standardized, sample licenses must be negotiated on a case by case basis.³ Kohn and Kohn (2010) provides guidance on the contours of these deals, and this section draws heavily from the discussion therein. The most important factors in general with sample licensing are (a) who was sampled, (b) who is doing the sampling, (c) the value of the sampled work, (d) the nature of the intended re-use, and (e) the manner in which the re-use came to the rights-holder’s attention.

Music publishers and songwriters, as the owners of composition rights, have typically been less tolerant of sampling than record labels. Hence, a composition license, known as a “mechanical license,” must be negotiated to avoid potential infringement suits.⁴ Copyright owners have a number of options to consider when they discover a downstream artist is producing, or has already distributed, music containing samples of their work. Timing is especially important due to hold-up (Williamson, 1985), as well as the typical punitive terms demanded under post-infringement licensing. At the most basic level, the copyright owner may choose to ignore the sample, especially in cases where enforcement will not justify the costs, particularly when the re-use is minimal or the downstream artist is commercially unsuccessful. Alternatively, the copyright holder may also choose to flatly refuse a mechanical license, and forego licensing revenue, perhaps in the case that the proposed re-use results in the original work being “perverted, distorted, or travestied.”⁵ In the case that the rights-holder chooses to give

²The *Illegal Art* label operated under blatant disregard for licensing of digital samples, with the associated mash-up artist Girl Talk achieving commercial success re-using hundreds of samples per album.

³Examples of standardized licensing include licensing a public performance from a performance rights organization (PRO), or authorizing a cover song via a compulsory mechanical license. Note also that the statutory licensing of music compositions precludes significant alterations or transformative use, hence does not apply to digital sampling.

⁴These licenses have historically, and continue to be known as “mechanical licenses” due to the then mechanical nature of early reproductions, such as piano rolls for player pianos, music boxes, and phonograph records. Piano rolls, for example, consisted of rolls of paper with perforations coding the notes of a musical composition. The rolls were fed through a player piano, which mechanically scanned the perforations to perform the composition.

⁵As was the case with anecdotes in the early days of hip hop, with artists objecting to their melodies being featured in expletive-filled, anti-authority focused songs.

sampling permissions, they have a number of options, (a) grant a mechanical license for a flat fee, (b) license for a royalty, (c) take a share of the new composition, along with royalties, (d) seek co-ownership in the new composition, or (e) demand full assignment of the copyright. In the most common arrangement, the rights-holder licenses the sample for a percentage below the statutory mechanical rate, while also taking a royalty share from the public performance right. However, licensor’s demands typically increase with the popularity of the re-used work, especially in cases of extensive re-use.

A master use license is additionally required if a new song incorporates any portion of an existing sound recording. Upstream sound recording rights-holders begin negotiations from a substantially more powerful position compared to the composition rights-holder, given the previous discussion on *de minimis* use. However, record labels are more lenient with sampling than songwriters and publishers, given that a record label’s own artists may engage in sampling and re-use. Again, however, the options available to a sound recording copyright holder are very similar to those of the previously discussed composition owner. When re-use is substantial and the success of the downstream work can be attributed to the sampled recording, royalty agreements are most common.⁶ Sound recording royalty rates in this context typically range from 1/2 cent to 5 cents per unit, along with an upfront royalty payment.

Digital sampling’s historical roots in hip hop music should influence the selection process in this study. Samples are more likely to be used downstream in hip hop music, although the practice has now become relatively common in electronic music and used to a degree in other genres as well. Furthermore, this norm of sampling in hip hop should affect the types of artists that are selected for re-use through samples, with hip hop music heavily drawing from funk, soul, R&B, and previous hip hop music to create the percussion-based “sound” of hip hop. Sampled songs are often not especially prominent or mainstream, with many producers priding themselves on their ability to repurpose obscure, forgotten tunes. In this empirical section of this paper, I employ a matched-sample design that controls this selection process in order to

⁶Royalty agreements may sometimes be necessary due to the recording contract between the re-used artist and their label. These recording contracts may also prevent licensing of samples without the express permission of the recording artist.

estimate the average treatment effect on the treated (ATT) artists.

In contrast with digital sample licensing, cover songs are covered under a statutory licensing doctrine that originated from antitrust concerns in the early 20th century player piano market.⁷ Congress originally set the statutory royalty rate at 2 cents per unit, which held until 1978, and currently stands at 9.10 cents per unit, or 1.75 cents per minute per unit for songs over 5 minutes in length. The Harry Fox Agency, founded in 1927 by the National Music Publishers Association, collectively manages this licensing in the U.S. market, as well as collects and distributes royalties to member songwriters and music publishers.

The practice of creating cover versions is less tied to any specific genre than the relationship between hip hop and digital sampling. It is common, however, for cover versions to be used to adapt past compositions across genres, such as Johnny Cash’s cover of Nine Inch Nail’s industrial track “Hurt.” Prominent early rock and roll artists, like The Beatles, The Rolling Stones, and Bob Dylan, are especially prone to be the subject of cover versions.

3 Data

3.1 Digital Sampling and Cover Song Data

A new, proprietary panel data-set tracking online music streaming was provided by Spotify, one of the largest online music streaming platforms with over 40 million paid subscribers. Spotify provided a private, disaggregated data-set measuring the daily streaming demand of

⁷In 1908, the United States Supreme Court, in *White-Smith Music Publishing Company v. Apollo Company*, ruled that piano rolls and phonograph records were not infringing copies of musical compositions due to the opinion that humans could not read these mechanical reproductions. The Æolian Company, a manufacturer of player pianos and organs, anticipated that congress was poised to supersede the controversial Supreme Court decision through recognition of mechanical reproductions as an exclusive right of the copyright holder. In anticipation, Æolian negotiated exclusive licenses with major music publishers in order to control the mechanical reproduction rights for the majority of popular music at the time, preventing competing manufacturers from acquiring their own piano roll licenses. In response, Congress included a compulsory licensing requirement in the 1909 act, mandating that anyone had the right to a mechanical reproduction license once the reproduction right of the song had been exercised at least one time.

11,682 artists between August 1, 2015 and August 29, 2016. While song-level data may have been ideal, artist-level aggregation was required as part of the data-transfer arrangement.

The streaming sample covers 11,682 artists, consisting of 1,004 treated artists – those that were re-used in the WhoSampled data described below – and 10,678 control artists. These control artists were gathered according to Spotify’s similar artist feature. For each treated artist, I collected the name and unique database identifier of each similar artist, dropping any duplicates. Upon collecting this list of artists, Spotify then provided me with daily streaming data, totaling 4,394,568 observations for 11,682 unique artists.

Data on re-use in music was collected from WhoSampled.com. WhoSampled.com is a community-driven website dedicated to sample-based music, cover songs, and remixes. Currently the largest website of this kind, WhoSampled tracks the origins of samples and cover versions used in popular music. Any of the approximately 17,000 community members of the website may upload a re-use to the database – as well, members of the community may comment on songs tracked in the database, report incorrect information, or vote on the “quality” of a particular sample, remix, or cover version. Figure 1 depicts an example of a typical sample-based song in the database, that is, a song that appropriates elements of past work. This re-use data thus unambiguously measures a form of cumulative innovation, and I am hence able to avoid many of the known pitfalls that come with inferring knowledge flows from patent citations (Alcacer and Gittelman, 2006).

Beyond just data on who and what was re-used, the WhoSampled database also tracks data on how the content was re-used, such as what elements of the prior song were captured, as well as whether the sample appears once, or throughout the new composition. Figure 2 provides an example of this raw, detailed data that I’ve collected from WhoSampled. My data from WhoSampled, in aggregate, covers 358,356 cases of re-use in popular music, including cover songs, remixes, direct samples, and interpolations.⁸ Such data was previous utilized in Watson

⁸Interpolations are a form of sampling in which the original composition is re-used, but the downstream artist creates their own, new sound recording of the original composition. Rather than licensing both the composition and the recording, the composition can be licensed and re-recorded to avoid licensing fees on a master-use license of the original sound recording – though artists may also choose to re-record the composition

(2017) to study copyright’s affect on re-use and the content of new products. Upon reducing this data-set to samples and covers that occurred over the duration of my music streaming data, this data covers 2,180 re-uses, consisting of 801 cover versions, 364 interpolations, and 1,015 samples.

Additional artist attributes and measures were collected from the Spotify database. Follow-up counts were collected for each artist, along with detailed genre information, and data on all releases by each artist (e.g., albums, singles, compilations). Spotify tracks genres at a highly disaggregated level – artists in my data-set are tagged with several genres, covering 1,186 distinct genres in total, including examples like “funk rock,” “vegan straight edge,” and “quiet storm.”⁹ I use Spotify’s release data to measure when an artist debuted – by the appearance of their first album – as well as to measure when an artist last released an album.

Due to contractual obligations, Spotify could not provide raw, uncensored streaming counts for artists at the artist-day level. Instead, Spotify provided a normalized measure of streaming demand, $Streams_{it}$, created by dividing streams for artist i in time t in the U.S. market by the average artist’s streams in the United States during the period. To aid interpretation of the regression results in Section 5 given this normalization, consider Equation 1 below.

$$\ln(Streams_{it}) = \alpha + \delta PostReuse_{it} + \varepsilon_{it} \quad (1)$$

where $Streams_{it}$ is the normalized measure of streaming, equal to the count of streams for artist i in time t , l_{it} , divided by a normalization factor, μ , in this case equal to the average artist streams during the sample period. With Equation 2 below, $\ln(\mu)$ will be absorbed by the constant term α , thus δ will be interpreted as the percentage change in l_{it} resulting from a unit-change in $PostReuse_{it}$.

to alter the sound and feel. An example of this being Dr. Dre’s re-use of David McCallum’s “The Edge” in “The Next Episode.”

⁹With vegan straight edge being a subgenre of punk rock, and quiet storm a popular form of R&B incorporating jazz and pop influences

$$\ln\left(\frac{l_{it}}{\mu}\right) = \ln(l_{it}) - \ln(\mu) = \alpha + \delta PostReuse_{it} + \varepsilon_{it} \quad (2)$$

Summary statistics for the combined data-set are shown in Table 1, with variables split between re-used artists and control group artists in Table 2. As seen in Table 1, the outcome variable of interest, $Streams_{it}$ is highly skewed – while the average is 1.00 due to the normalization, $Streams_{it}$ reaches a maximum value of 876 for the artist Drake, implying that on the observed day t , Drake received 876 times the average listening. $Followers_i$ measures the number of Spotify users that have subscribed to updates for artist i , and hence is an alternative measure of popularity on the platform – that is similarly skewed, with a maximum follower count of 11 million, compared with an average of $\sim 113,000$ followers and a minimum of 0. Approximately 9% of the sample artists were treated with re-use (average of $EverReused_i = 0.089$), a result of constructing the data-set – control artists were selected for each re-used artist from Spotify’s “similar artists” feature. It is thus important to note that I am not observing the universe of artists and songs on Spotify, but rather a sub-population that is somewhat similar even before econometric matching. Table 2 splits the sample by treatment category, between those artists that are re-used ($EverReused_i = 1$) and those that are not ($EverReused = 0$). From this table one can see that the re-used artists are much more popular on average. This is not particularly surprising given that digital platforms enable distribution of the “long tail” (Brynjolfsson et al., 2006), and thus the control group will be comprised of many low-popularity artists. I discuss methods to account for this bias in Section 4.

4 Empirical Framework and Research Design

With regards to research design and identification for this research question, it is first useful to consider the ideal experiment used for estimating demand spillovers in re-use and derivative works. In the ideal experiment, the omnipotent econometrician would, (1) randomly select a sample of treatment songs from the population of available songs, (2) create derivative works

based upon them, by appropriating sections of the treatment group songs in new compositions, and (3) distribute these new musical arrangements to the relevant market. Then, *ceteris paribus*, demand could be compared between the treatment and control group songs by observing changes in online streaming, digital downloads, and physical media purchases.

Alternatively, the same ideal analysis could be undertaken at the artist level, rather than the song level. An artist-level analysis may be necessary if spillovers exist between downstream and upstream artists, and are not confined to the reused-reusing songs. For example, consider Kanye West’s re-use of portions of Steely Dan’s 1976 hit song “Kid Charlemagne” in West’s “Champion.” If re-use confers a positive effect, it may drive more listeners to seek out “Kid Charlemagne,” but it may also drive consumers to listen to other songs and albums in Steely Dan’s discography beyond the re-used song. Thus, the econometrician would underestimate the effect of re-use if changes in listening are measured only at the song level, as re-use may spur dramatic changes in demand for the re-used artist’s expanded discography. Second, if re-use directs positive or negative spillovers to an artist’s broader discography, any re-used artist’s songs could not be used in the control group, or vice-versa, in order to maintain the treatment group restriction. These concerns justify conducting the analysis of potential spillovers at the artist-level, rather than a song-level analysis.

In this paper, I rely on observational data, and hence cannot turn to the random, controlled variation from an ideal experimental condition in order to identify the effect of interest. Rather than specifying and estimating a discrete choice demand model (e.g., [McFadden \(1973\)](#); [Berry et al. \(1995\)](#)), I measure the spillovers to re-use through a reduced-form, difference-in-differences design, based upon the specification shown in Equation 3.

$$\ln(Streams_{it}) = X'_{it}\beta + \gamma_t + \zeta_i + \delta PostReuse_{it} + \varepsilon_{it} \quad (3)$$

Where $Streams_{it}$ is a non-negative measure of an artist i ’s streams in time t , X_{it} is a vector of time-variant controls, γ_t captures time period effects, ζ_i an artist-specific fixed effect, and $PostReuse_{it} = 1$ for artists in the treatment group once they have been treated (re-used).

Thus δ captures the potential positive/negative spillovers that result from being re-used.

Additionally, I combine the differences-in-differences estimator above with a matched-samples approach to estimate the Average Treatment Effect on the Treated (ATT).¹⁰ Identification in this setting thus rests on the key assumption of unconfoundedness of selection into treatment, conditioned on observable characteristics. In my empirical context, I observe a set of songs that have been re-used (as both digital samples and cover songs), the characteristics of both upstream/downstream songs, the characteristics of both upstream/downstream artists, and a time-series of demand for the artists that have been re-used. In addition, I also observe a pool of artists that have not been re-used during the study time period, characteristics of these artists, and a time series of demand for each of these “untreated” artists. Dealing with treatment group selection in the data generating process is the primary empirical challenge in this setting. It is reasonable to be concerned about the proposition that artist’s repertoires are not randomly selected when derivative works are created, and that this selection process may be correlated with unobserved characteristics of the artist’s appeal. If this is the case, the unobserved error term in a regression framework correlates with the treatment condition, and parameter estimates will hence be biased. I consider two types of selection problems, (1) time invariant, and (2) time varying.

Some artists may have a higher baseline propensity of being re-used than other artists. Heterogeneity in prominence between artists exists over time, and this prominence may be correlated with the chance that the artist’s work is used in a new derivative work. Some artists also have a history of consistent re-use, such as the “Amen break,” a 6-second section of “Amen, Brother” by The Winstons, a sample that has been re-used over 2,700 times since it was released in 1969.¹¹ This time invariant selection can be remedied with panel data, as the within estimator – in the form of an artist-level fixed effect – can control for unobserved-to-the-econometrician, time invariant characteristics of each artist in the sample.

It is more difficult to deal with potential time-variant selection into treatment, but I argue

¹⁰See e.g., [Azoulay et al. \(2010\)](#) for examples of prior use of this approach

¹¹<http://www.whosampled.com/The-Winstons/Amen,-Brother/> (Accessed 9/1/2017)

that the precise timing of re-use is plausibly exogenous, and rely on this timing assumption for identification. However, one may still be worried that career dynamics of an artist influence the propensity that an artist’s works are re-used. Potential dynamic sources of treatment selection could come from examples such as (a) new artists, rising in popularity, are sampled by downstream artists due their rising profile, (b) promotional activities (e.g., increased radio airplay) of an upstream artist remind downstream artists of these works, influencing the chance that the promoted works are incorporated in downstream innovations, or (c) short lived fads arise in particular types of sounds and melodies, which are extensively re-used during the fad. All of these examples could result in trend heterogeneity between treatment and control group, leading to violation of the difference-in-differences central assumption of parallel trends between treatment and control groups. In light of these concerns, and in order to bolster the timing exogeneity assumption, I create a matched sample, matching each treatment-group artist to a closely matched control-group artist based upon both time invariant and time variant characteristics of the observed artists. Treatment artists are matched to control group artists in my sample based upon (1) artist genre, (2) artist vintage, and (c) artist popularity trends before treatment.

To ensure balance between treatment and control groups, I implement matching between treatment and control artists based upon coarsened exact matching (CEM) of pre-treatment variables (Iacus et al., 2012). I match artists based upon vintage, genre, and pre-treatment popularity. In the matching procedure, I coarsen vintage into the decade that an artist’s first album was released (e.g., 1960s, 1970s, etc.), disaggregate genre into main genre (e.g., garage rock $\rightarrow$ rock), and pre-treatment popularity into percentile rank “buckets.” For each treatment artist, i , I then select a control-group artist, j , that exactly matches artist i on the coarsened variables. If many control artists exactly match artist i , the tie is broken by selecting one control from the potential controls at random. Treated artists are dropped from the matched-sample in cases where there is no exact match between the treated artist and potential control group artists.

Three variables were coarsened for use in the matching procedure - genre, vintage, and

pre-treatment streaming. Each artist in the data-set is assigned, on average, 5.17 genres of the total 1,186 observed genres. Exactly matching on this highly detailed genre information thus clearly presents a dimensionality problem, as it is unlikely that any given artist will exactly match the ~ 5 genres of a focal artist. In order to tractably match on genre, I hand-code each of the 1,186 genres into one of 18 broader genres, based upon AllMusic’s genre classification.¹² The broad genres include “rock,” “r&b,” “jazz,” “blues,” “folk,” “avant garde,” and so on. With this coding in hand, I then count how many times an artist is tagged with each of the broad genres, and select one genre for each artist based upon the greatest number of tags. An artist’s “vintage” is coarsened according to the date of their first album debut. For each artist in the data, I select their earliest album release – and coarsen this to the decade level. Thus an artists that debuted in 1983 has a coarsened vintage of “1980s.” Five measures of pre-treatment popularity were coarsened for the matching procedure. For each artist, rolling averages of $\log(\text{streams}_{it})$ for 5, 10, 25, 50, and 100 days before day t for artist i were calculated. The percentile rank of these measures were calculated for each day, and then coarsened into 27 “buckets,” with successively tighter break points approaching the top of the distribution.

5 Results

5.1 Baseline Estimates

First, using an unmatched sample, I consider the average effect of being re-used on the original, underlying artist. Table 3 displays the baseline results in which the dependent variable is $\log(\text{streams}_{it})$, a log-transformed measure of streaming for artist i on day t , and the main variable of interest, PostReuse_{it} , = 1 for re-used artists after they have been re-used. Following the previous discussion in Section 4, we can interpret the coefficient on PostReuse_{it} as the percent change in raw stream-counts when artist i is re-used. Standard errors are clustered at the artist level.

¹²<http://www.allmusic.com/genres>

Column (1) includes no control, and implies an enormous benefit to being re-used, but does not account for treatment-selection issues, or confounding secular time trends in listening on the platform. Not controlling for these confounders would imply a 243% boost in streaming resulting from downstream re-use. Additionally, it is well established that the time lag between the release date of music and the current date has a significant impact on its popularity (Waldfoegel, 2012), thus in Column (2) I control for the time since artist i last released an album by including $\log(\text{sincealbum}_{it})$, measuring the log-transformed days since an artist’s last album release. As expected, an artist’s popularity dissipates as the time since album release grows and associated promotional activity dies out. However, even after controlling for this distance in time, the re-use parameter remains extremely high, implying a 239% increase in popularity on the platform as a result of downstream re-use. I add an individual artist fixed effect for the re-used artists in Column (3) in order to control for time invariant selection effects in my estimation. With the addition of this artist-level effect, the parameter on PostReuse_{it} significantly shrinks in magnitude, from $\sim 240\%$ down to an estimated 19.8% positive increase in listening – suggesting that an artist’s popularity has a large role in the propensity that they are re-used downstream in cover songs and sample-based music. I alternatively control for time trends in Column (4), by adding time fixed effects, in the form of a day-of-week fixed effect to capture cyclical patterns in streaming throughout the week, and a month-of-year fixed effect to capture secular time trends on the platform. These potentially confounding time trends appear to have little effect on the results, as the introduction of the time fixed effects in Column (4) yields results consistent with the model lacking controls in Column (1). Finally, I include all fixed effects and control for time since last release in Column (5). This fully saturated model implies that re-use, on average, results in a 2.9% increase in streaming for the re-used artist, a result that is significant at the 5% level

To explore cover version vs digital sample heterogeneity, I restrict the treatment group to artists that were subject to re-use through cover songs, dropping those artists that were re-used through downstream digital sampling. The results of this analysis are shown in Table 4. Column (1) includes no controls, Column (2) adds a control for time since an artist’s last album release, with Columns (3) and (4) alternatively controlling for artist and time fixed effects

respectively. These results in Columns (1)-(4) of Table 4 are relatively consistent with the analogous columns in Table 3, that is, showing a positive spillover as a result of a downstream cover version release. However, the positive effect of re-use disappears for cover songs once both artist and time fixed effects are included in the model. Additionally, the coefficient mean estimates appears negative, translating to a 1.8% drop in listening post-reuse, though this result is not significant at typical levels. Thus if there is any negative effect from the release of cover songs, it is below the noise threshold in my sample. These results suggest that cover songs, wherein the entire composition and lyrics of a past song are re-used in a new recording, do not have a significant impact on the consumption of the re-used artist.

In light of the selection effect apparent from controlling for re-used artist-level fixed effects, I further explore this in Table 5, where I introduce the dependent variable $EverReused_i$, which = 1 if artist i is re-used in the data-set, and 0 otherwise. I collapse the panel to a cross section, by averaging each artists stream measure, $streams_{it}$ over all days, t , to yield $avg(streams_{it})$. Table 5 thus explores, with linear probability models, how artist popularity, genre, and vintage affect the propensity with which artists are re-used downstream. Column (1) includes just the streaming measure of popularity, $log(avg.streams_{it})$, as across specifications this has a small, but positive and significant effect on the propensity of re-use. Similar results are observed when I include another correlated measure of artist popularity, the log-transformed count of an artists followers, $log(Followers_i)$, in Column (2). Again, popularity appears to increase the chance that an artist is subject to re-use across specifications. In Column (3) I explore the effects of artist vintage, by including a set of dummies for the decade of artist debut, with 2010 as the omitted category, with Column (5) including genre fixed effects. Interestingly, across Columns (3)-(5), we see that artists from the 1960s and 1970s are much more likely to be re-used, with results from Column (5) implying that artists from the 1960s are 7.4% more likely to be re-used, and artists from the 1970s 8.3% more likely to be re-used than artists from the 2010s. Regression results also imply that genre also plays a significant role in re-use. Columns (6) and (7) explore this by comparing the R&B genre category against all other genres – estimates that imply R&B artists are 8.5%-9.7% more likely to be re-used than other genres of music.

I then contrast the magnitude of the treatment effect, versus the “selection effect” in Table 6. The dependent variable is the log-transformed streaming variable. Columns (1)-(2) include time fixed effects, while Column (3) adds vintage fixed effects, and Column (4) additionally includes genre fixed effects. The variables of interest are $PostReuse_{it}$ as before, which equals 1 for re-used artists post-reuse, along with $EverReused_i$ which = 1 for artists that are part of the treatment group – those that are re-used at any point during the sample period. The results across columns are fairly consistent, there appears to be a large, positive bias in streaming for treatment-group artists, that is, treatment group artists have a significantly higher baseline of listening as compared with artists in the control-group, and this effect is much greater in magnitude than the treatment effect, assuming exogeneity. These results are not unexpected given the procedure for constructing the unmatched-sample. Digital marketplaces and distribution platforms are known to enable the “long tail” (Brynjolfsson et al., 2006) – and a large fraction of the $\sim 10,000$ control group artists are relatively obscure acts that receive minuscule daily demand. In order to ensure comparability and balance between the re-used artists and the control artists, I then turn to a matched-sample estimator in the next section.

5.1.1 Matched Sample

A matched-sample, implemented as described in Section 4, was used to match each treatment artist to one closely-related control group artist. In this way, I can ensure that the treatment group and control group are not observably different *ex ante*, and better control for time-variant selection into treatment by matching popularity trends between treatment and control artists. However, we trade bias for variance, as the matched-sample is reduced to 652 treatment-control pairs, compared with the approximately 11,000 artists in the unmatched sample.

Figure 3 graphs my estimates of the pre and post re-use trajectory of artist listening using the matched-sample. The 95 percent confidence interval of these estimates is overlaid, based on artist-clustered robust standard errors. The model used for this graph includes a full set of time and artist fixed effects, along with controls for time since last album. In this model, the treated artist’s streaming rate remains relatively level compared with the control group

from 50 days prior to the day of re-use, and the error bands show that I cannot reject the null hypothesis of no difference in the pre re-use trends. Thus, the re-used artists and control artists have similar trends in listening up to the point of re-use, once conditioned on observable characteristics. Furthermore, the point estimates in the graph show that re-used artists gain an immediate boost in listening upon re-use, with this effect most precise from five to nine days after re-use, with the error bands from two weeks after exposure to re-use indicating the noise in my estimates. Part of the variance in my estimates is attributable to the high-frequency (day-level) of observation, but this level of disaggregation is also useful for picking up the apparent immediate effect. This graph must also be interpreted in context – the baseline estimate of re-use’s effect is rather small at 3%.

The results of the baseline regression on the matched-sample are shown in Table 7, with these results representing Average Treatment Effects on the Treated (ATT). Column (1) regressed $\log(\text{streams}_{it})$ on PostReuse_{it} without no controls, and shows a much more modest effect compared with the same regression on the unmatched sample in Table 3, implying an average 9% boost in streaming from re-use compared to the larger 243% estimate in the unmatched sample. Controlling for time since last album in Column (2) does not significantly change the significance, but the point estimate does increase to an implied positive effect of 18%. However, addition of an upstream artist fixed effect in Column (3) yields statistically significant results that are comparable in magnitude to the same regression on the unmatched sample, with an estimated positive effect of 19.5% increase in listening on the matched sample, compared with estimates of 19.8% on the unmatched sample. Time fixed effects are included in Column (4), with the fully saturated model in Column (5) suggesting that re-use results in an average increase in listening of 3% for re-used artists. It is important to note that this estimate is quite consistent with results from the unmatched sample, where the estimated effect was 2.8% in the fully saturated model. This consistency between unmatched/matched samples implies that time-variant selection effects are only of minor concern – most of the selection appears driven by time invariant heterogeneity in artist popularity.

5.2 Artist Prominence

With some robust evidence that re-use leads to a positive spillover of an approximate 3% boost in plays, I then explore how this effect is mediated by the prominence of an artist. Measuring artist prominence as a function of their followers on Spotify, I create a measure $prominent_i$, which = 1 when artist i is in the top 50% of the followers distribution, and 0 otherwise. Interacting this variable with the previously used treatment effect on the matched sample produces the results shown in Table 8. Columns (1) and (2) of Table 8 omit the artist fixed effect, and we see that these high-prominence artists receive much greater listening on average, corresponding to an over 400% increase in listening compared to artists with fewer followers. Conversely, once I introduce an artist fixed effect in Columns (3)-(5), the negative interaction between our post-treatment re-use variable and $prominent_i$ points to the moderating impact of prominence on re-use. Column (3) includes just the main treatment variable, $PostReuse_{it}$, along with the interaction effect $PostReuse_{it} \times Prominent_i$, without controlling for time trends, while Column (4) controls for time since last album. Columns (3) and (4) are rather consistent in inference, with a 22-23% estimated increase in listening, paired with a 4-6% negative moderating effect for high-prominence artists. In the fully saturated model, Column (5), I find that on average, the artists of low prominence receive a 6% boost in plays from being re-used, but that this effect is nearly entirely nullified for highly prominent artists, as the interaction of these effects, $PostReuse_{it} \times prominent_i$, is estimated at negative 5.8%. These results point to an advertising effect at play, wherein low-prominence artists benefit from the exposure of being re-used, but with the most prominent artists deriving no benefit from the exposure, having sufficiently saturated the market.

Figure 4 separately graphs my estimates of the pre and post re-use trajectory for artists of high and low prominence. The sample is split by prominence. The left plot, in blue, displays the effect of re-use for low prominence artists, while the right plot, in red, for high prominence artists. This figure is based upon the matched-sample, and 95 percent confidence intervals are overlaid based on artist-clustered robust standard errors. The model in both graphs controls for time since last album release, along with a full set of time and artist fixed effects. Rather

than the high-frequency day-level estimates used in Figure 3, the effects are estimated at the week level to diminish variance.

On the left plot of Figure 4, I estimate the listening trajectory for low prominence artists before and after re-use. The artist’s streaming rate remains close to zero relative to the controls from 3 weeks prior up to the week of re-use, and the error bands overlapping with zero-effect show that I cannot reject the null hypothesis of no difference in the pre re-use trends. Thus, once conditioned on observables, the low prominence re-used and control artists display similar trends in listening up to the point of re-use. However, upon re-use, the artists immediately gain an increase in listening during the first week post re-use, and this effect remains statistically significant for four weeks after the re-use date.

These results contrast with the right-hand plot of Figure 4, which estimates the listening trajectory for high prominence artists. Again, I find no difference in pre re-use trends between the high prominence re-used artists and the control artists. After re-use, however, the high prominence artists receive only a modest boost in listening, based upon the point estimates, with the error bars implying that I cannot reject the null hypothesis of no effect. The point estimates then trend toward zero-effect starting at two weeks after re-use, suggesting that high prominence artists receive little, if any, positive effect from downstream re-use.

5.3 Novel Re-uses

Finally, I also explore how this effect differs for artists that have not traditionally been the target of digital sampling and cover versions. Some artists have been particularly subject to re-use, such as James Brown, The Winstons, and George Clinton’s related Parliament and Funkadelic acts.¹³ Instead, I look at artists that have yet to receive attention through re-use, by restricting the sample to artists that have previously been re-used one or fewer times. Of course, this restricts the sample size substantially, and the resulting sample consists of only 73 matched treatment-control pairs. While the statistical inference in this setting is limited, the

¹³Particularly James Brown’s “Funky Drummer” drum break and The Winstons’ “Amen Break”

results may still be provocative.

The results of this analysis are shown in Table 9. Column (1) includes no controls, Column (2) introduces an artist fixed effect, Column (3) only time fixed effects, Column (4) includes both artist and time fixed effects, and Column (5) presents the fully saturated model including controls for time since last album. While the results are somewhat limited due to the small sample size and limited statistical power, we observe consistent trends and inference between these results in Table 9 with the previous results in Table 7, albeit with greater magnitude in point estimates. Without controls, Column (1) suggests an insignificant effect, similar in magnitude to the same regression in Table 7. However, once I include an artist fixed effect in Column (2), the estimate jumps to a significant 30% increase in listening as a result of re-use. The fully saturated model in Column (5) suggests that re-use leads to a 15% boost in plays. While that result is significant at only the 10% level, this effect is quite provocative given that it is 5 times as large as the positive spillover estimated for the more general matched-sample. If these results hold, it would suggest that less-targeted artists receive a particularly large benefit from re-use. Thus encouraging downstream re-use of both emerging acts and out-of-spotlight musicians through permissive licensing may be an effective promotional tool.

6 Concluding Remarks

With this research, I have found no evidence that derivative works compete with the original works that they are based upon. Instead, the release and online distribution of a derivative work induces a positive effect on demand for the underlying good. This effect is particularly pronounced for artists that have been re-used for the first time, as well as for low-prominence artists. Derivative works thus appear to have, on average, a net positive advertising effect for the work re-used. Consequently, these results contribute to the debate around the appropriate balance of rights to derivative works, aims to adapt fair use to the digital age, as well as providing quantitative evidence to content industry IPR strategy. In particular, these results speak to the potentially unexpected benefits of a permissive licensing strategy – that easing

licensing terms of music publishers' and record labels' catalogs may actually drum up demand for these under-utilized works. This research additionally may inform licensing negotiations, as I've thus far seen no evidence that digital sampling and cover version licensing deals internalize this positive effect in pricing and licensing terms. That is, when an eminent artist intends to re-use a piece of music, pricing should incorporate the chance that this downstream re-use will actually boost the market for the original.

From a policy perspective, the finding of a small, positive effect from downstream re-use is particularly useful for designing the appropriate balance of copyright in a cumulative setting. In fact, finding any non-negative effect from downstream re-use would still be policy relevant as this would imply that re-use causes no economic harm to the original artist. If re-use does not harm the market for the original, easing restrictions over re-use should not affect the incentives for artists to create new works. In the United States, the copyright system is rooted in consideration for the incentives to create new creative works, with the copyright clause of the U.S. Constitution granting congress the ability "to promote the Progress of Science and useful Arts by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." The U.S. constitution does not, however, justify copyright by considering the profits of the creator or the moral right of the artist. This focus on economic incentives in the U.S. system contrasts with other copyright systems, particularly those in Europe, which emphasize the moral rights of the original artist. Many objections to easing re-use restrictions (e.g., through compulsory licensing) focus on the desire for artists to control the context in which their work is re-used to avoid undesired distortions. However, at least in the United States, such objections have little basis.

Evidence of these potential positive spillovers is particularly relevant to rapidly changing content industries. Particularly in the music industry, publishers and record labels face new threats from digitization. Additionally, traditional music industry strategies for promoting back catalog works, such as the n -th anniversary re-issue box set, are becoming less relevant due to digitization and the successive unbundling of content. Despite this, labels now possess a repertoire covering decades of copyrights on once-popular music, and these rights are extremely

long lived – nearly all sound recordings will be protected by U.S. copyright until 2067. While labels have many avenues for promoting currently active artists – through new albums and singles, collaborations with established acts, and live tours – methods are more limited for promoting dated, now-forgotten artists. Thus, this research provides a suggestive avenue for monetizing the rights portfolio of out-of-spotlight artists, and the industry should consider supporting mechanisms to ease the transaction costs of re-use – through collective licensing and/or compulsory licensing schemes.¹⁴

Finally, these results are also useful for adapting fair use to the digital environment. Rather uniquely, the copyright policy of the United States includes a fair use doctrine that permits legal, unauthorized use of copyright-protected material in limited contexts.¹⁵ Fair use is determined in a case by case basis using a four factor test, with the fourth factor considering “the effect of the use upon the potential market for or value of the copyrighted work.” I have demonstrated in this paper a context in which re-use, on average, has a positive effect on the potential market of the upstream work, speaking directly to the fourth factor of consideration under fair use.¹⁶ As a balance to copyright’s restrictions of first amendment free speech, the fair use doctrine is still quite flawed.¹⁷ Due to the necessarily vague statute, any given fair-use case is subject to the opinions of a judge and jury, such that determining whether a use qualifies as a fair use *ex ante* is extremely difficult, if not impossible, even under guidance from a legal professional. In the cases that a use will be judged to fall under fair use, a successful defense of this in court is inevitably very costly. And yet, digital age policies like ‘notice-and-takedown’ procedures further threaten fair use with their ability to be abused to invoke censorship. This paper thus provides provocative results that contribute to the debate of how to best re-design fair use.

¹⁴See e.g., [Menell \(2016\)](#) detailing how compulsory licensing could be expanded for this context

¹⁵Israel, Poland, and South Korea current have small fair use provisions. The European Union may also be moving towards fair use style exemptions.

¹⁶Although it is the case that the fourth factor, the effect of the use upon the potential market, includes consideration for the effect of the use upon the market for authorized derivative works. These results do not refute the relevance of this factor, but are intended to aid discussion and debate of the matter.

¹⁷See e.g., [Tushnet \(2004\)](#); [Carroll \(2007\)](#) for a thorough analysis of this point.

References

- Abramowicz, M. (2005). A theory of copyright's derivative right and related doctrines. Minnesota Law Review 90, 317.
- Alcacer, J. and M. Gittelman (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. The Review of Economics and Statistics 88(4), 774–779.
- Askin, N. and M. Mauskopf (2017). What makes popular culture popular? product features and optimal differentiation in music. American Sociological Review 82(5), 910–944.
- Azoulay, P., J. S. Graff Zivin, and J. Wang (2010). Superstar extinction. The Quarterly Journal of Economics 125(2), 549–589.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. Econometrica 63(4), 841–890.
- Brynjolfsson, E., Y. J. Hu, and M. D. Smith (2006). From niches to riches: Anatomy of the long tail. MIT Sloan Management Review 47(4), 67.
- Carroll, M. W. (2007). Fixing fair use. North Carolina Law Review 85(4), 1087.
- Cohen, L., U. Gurun, and S. D. Kominers (2014). Patent trolls: Evidence from targeted firms. Technical report, National Bureau of Economic Research.
- Danaher, B., M. D. Smith, R. Telang, and S. Chen (2014). The effect of graduated response anti-piracy laws on music sales: evidence from an event study in france. The Journal of Industrial Economics 62(3), 541–553.
- Gans, J. S. (2015). Remix rights and negotiations over the use of copy-protected works. International Journal of Industrial Organization 41, 76–83.
- Goldfarb, A., S. M. Greenstein, and C. E. Tucker (2015). Economic Analysis of the Digital Economy. University of Chicago Press.

- Goldfarb, A. and C. Tucker (2017, August). Digital economics. Working Paper 23684, National Bureau of Economic Research.
- Goldstein, P. (1982). Derivative rights and derivative works in copyright. Journal of the Copyright Society of the USA 30, 209.
- Green, J. R. and S. Scotchmer (1995). On the division of profit in sequential innovation. The RAND Journal of Economics 26(1), 20–33.
- Greenstein, S., J. Lerner, and S. Stern (2013). Digitization, innovation, and copyright: What is the agenda? Strategic Organization 11(1), 110–121.
- Heald, P. J. (2009). Does the song remain the same? An empirical study of bestselling musical compositions (1913-1932) and their use in cinema (1968-2007). Case Western Reserve Law Review 60(1), 1–47.
- Heald, P. J. (2014). How copyright keeps works disappeared. Journal of Empirical Legal Studies 11(4), 829–866.
- Iacus, S. M., G. King, and G. Porro (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis 20(1), 1–24.
- Kohn, A. and B. Kohn (2010). Kohn on Music Licensing (4 ed.). Aspen Publishers.
- Landes, W. M. and R. A. Posner (1989). An economic analysis of copyright law. The Journal of Legal Studies 18(2), 325–363.
- Lanjouw, J. O. and M. Schankerman (2001). Characteristics of patent litigation: A window on competition. RAND Journal of Economics 32(1), 129–151.
- Lanjouw, J. O. and M. Schankerman (2004). Protecting intellectual property rights: Are small firms handicapped? The Journal of Law and Economics 47(1), 45–74.
- LaPolt, D., J. Rosenthal, and J. Meller (2014). A response to professor menell: A remix compulsory license is not justified. Columbia Journal of Law & the Arts 38, 365.

- Lessig, L. (2008). Remix: Making art and commerce thrive in the hybrid economy. Penguin Press.
- Luo, H. and J. H. Mortimer (2016, May). Copyright infringement in the market for digital images. American Economic Review 106(5), 140–45.
- Luo, H. and J. H. Mortimer (2017). Copyright enforcement: Evidence from two field experiments. Journal of Economics & Management Strategy 26(2), 499–528.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers of Econometrics. New York: Academic Press.
- McLeod, K. and P. DiCola (2011). Creative license: The law and culture of digital sampling. Duke University Press.
- Menell, P. S. (2016). Adapting copyright for the mashup generation. University of Pennsylvania Law Review 164(2), 441.
- Murray, F. and S. O’Mahony (2007). Exploring the foundations of cumulative innovation: Implications for organization science. Organization Science 18(6), 1006–1021.
- Nagaraj, A. (2017). Does copyright affect reuse? Evidence from google books and wikipedia. Management Science. Articles in Advance.
- Peukert, C., J. Claussen, and T. Kretschmer (2017). Piracy and box office movie revenues: Evidence from megaupload. International Journal of Industrial Organization 52, 188 – 215.
- Phillips, D. J. (2013). Shaping Jazz Cities, Labels, and the Global Emergence of an Art Form. Princeton, NJ: Princeton University Press.
- Reimers, I. (2016). Can private copyright protection be effective? Evidence from book publishing. The Journal of Law and Economics 59(2), 411–440.
- Rob, R. and J. Waldfogel (2006). Piracy on the high C’s: Music downloading, sales displacement, and social welfare in a sample of college students. The Journal of Law and Economics 49(1), 29–62.

- Rob, R. and J. Waldfogel (2007). Piracy on the silver screen. The Journal of Industrial Economics 55(3), 379–395.
- Rubinfeld, J. (2002). The freedom of imagination: Copyright’s constitutionality. The Yale Law Journal 112(1), 1–60.
- Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. The Journal of Economic Perspectives 5(1), 29–41.
- Tushnet, R. (1996). Legal fictions: Copyright, fan fiction, and a new common law. Loyola of Los Angeles Entertainment Law Journal 17, 651.
- Tushnet, R. (2004). Copy this essay: How fair use doctrine harms free speech and how copying serves it. The Yale Law Journal 114(3), 535.
- Voegtli, N. A. (1997, December). Rethinking derivative rights. Brooklyn Law Review 63(4), 1213–1269.
- Waldfogel, J. (2012). Copyright protection, technological change, and the quality of new products: Evidence from recorded music since napster. The Journal of Law and Economics 55(4), 715–740.
- Watson, J. (2017). Copyright and the production of hip hop music. Working Paper.
- Williamson, O. E. (1985). The Economic Institutions of Capitalism. New York, NY: The Free Press.
- Zhang, L. (2016). Intellectual property strategy and the long tail: Evidence from the recorded music industry. Management Science. Articles in Advance.

Tables and Figures

Figure 1: WhoSampled: Song-level Information

Car Thief
by [Beastie Boys](#)
Paul's Boutique
Capitol 1989
Producers: [Beastie Boys](#), [The Dust Brothers](#)

Buy this Track▼

Contains samples of 5 songs

Rien Ne Va Plus by Funk Factory (1975)	Multiple Elements Soul / Funk / Disco
Hurdy Gurdy Man by Donovan (1968)	Drums Rock / Pop
I'll Bet You by The Jackson 5 (1970)	Hook / Riff Soul / Funk / Disco
Drop the Bomb by Trouble Funk (1982)	Multiple Elements Soul / Funk / Disco

The WhoSampled.com community provides data on the samples incorporated, and re-used to create each song tracked in database. This screen capture displays the raw data covering the samples used in Beastie Boy's 1989 song "Car Thief"

Figure 2: WhoSampled: Detailed Sample Information

Direct Sample of Multiple Elements

12 Votes

Beastie Boys
Car Thief
Paul's Boutique
Capitol 1989

Sample appears at **0:21** (and throughout)

Download this Track

Buy on Vinyl/CD

iTunes

amazon

ebay

amazon

Producers: [Beastie Boys](#), [The Dust Brothers](#)
Main genre: [Hip-Hop](#) / [Rap](#) / [R&B](#)

Report wrong information or missing video

Funk Factory
Rien Ne Va Plus
Funk Factory
Atco 1975

Sample appears at **0:13**

Download this Track

Buy on Vinyl/CD

iTunes

amazon

ebay

amazon

Producer: [Michal Urbaniak](#)
Tags: [French](#)
Main genre: [Soul](#) / [Funk](#) / [Disco](#)

Contributed by
[Lihi](#) (1 submission)

The WhoSampled.com community provides detailed re-use data for each song in their database. For the same downstream song shown in Figure 1, details about each re-use are also collected, such as the timing and extent of re-use, and the elements of the upstream song that were re-used. This screen capture displays the raw data covering Beastie Boy’s 1989 sample of Funk Factory’s 1975 song “Rien Ne Va Plus”

Figure 3: Pre and Post Re-use Effects

Model: $\log(\text{streams}_{it}) = \sum_{j=-50}^{50} \mathbb{1}\{\text{t days after re-use}\} \times \delta_t \times \text{PostReuse}_{it} + \beta \times \log(\text{since album}_{it}) + \text{artist}_i + \text{day}_d + \text{month}_m + \varepsilon_{it}$

Note: 95% confidence intervals displayed, based on artist clustered standard errors

Figure 4: Pre and Post Re-use Effects: Comparison of Artist Types

Model: $\log(\text{streams}_{it}) = \sum_{j=-3}^4 \mathbb{1}\{t \in \text{week } j \text{ after re-use}\} \times \delta_j \times \text{PostReuse}_{it} + \beta \times \log(\text{since album}_{it}) + \text{artist}_i + \text{day}_d + \text{month}_m + \varepsilon_{it}$

Note: 95% confidence intervals displayed, based on artist clustered standard errors

Table 1: Summary Statistics

	Mean	S.D.	Min	Max
$\log(\text{Streams})_{it}$	-3.59	3.13	-10.03	6.76
Streams_{it}	1.00	6.26	0	876
PostReuse_{it}	0.067	0.25	0	1
EverReused_i	0.088	0.28	0	1
Followers_i	113,652	476,558	0	1.14e+07
Popularity_i	39.33	20.02	0	100
Since_album_{it}	1,603	2,599	1	19,800

Table 2: Summary Statistics - Continued

	$\text{EverReused}_i = 0$	$\text{EverReused}_i = 1$	Total
$\log(\text{Streams}_{it})$	-3.80 (3.04)	-1.43 (3.17)	-3.59 (3.13)
Streams_{it}	0.62 (2.90)	4.89 (18.43)	1.00 (6.26)
PostReuse_{it}	0 (0)	0.75 (0.431)	0.067 (0.249)
Followers_i	74,007 (254,737)	520,108 (1,307,252)	113,652 (476,558)
Popularity_i	37.96 (19.40)	53.37 (20.92)	39.33 (20.02)
Since_album_{it}	1,602 (2,594)	1608 (2,650)	1,603 (2,599)

Note: This table displays sample means, with standard deviations in parentheses, split by re-used/control group category. The first column displays statistics ($\text{EverReused}_i = 0$) for the control group artists, those artists that were not covered or sampled. The second column for those artists that were re-used through new cover songs or sample-based music ($\text{EverReused}_i = 1$). Pooled statistics are displayed in the third column.

Table 3: Baseline Regression: Non-matched sampled

	(1)	(2)	(3)	(4)	(5)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	2.4266*** (0.104)	2.3947*** (0.099)	0.1983*** (0.014)	2.4209*** (0.105)	0.02876** (0.014)
$\log(since\ album_{it})$		-0.4410*** (0.015)			-0.05862*** (0.002)
Constant	-3.7552*** (0.029)	-0.7040*** (0.098)			
Artist FE	No	No	Yes	No	Yes
Day FE	No	No	No	Yes	Yes
Month FE	No	No	No	Yes	Yes
Observations	4327084	3928784	4327059	4327084	3928774

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. Columns (1) and (2) include no fixed effects, with Column (2) controlling for the number of days between time t and the last album released by artist i . Column (3) adds an artist fixed effect, Column (4) includes time fixed effects in the form of day-of-week and month-of-year fixed effects. Column (5) includes both artist and time fixed effects, as well as controlling for time since last album

Table 4: Baseline Regression: Cover Songs Only

	(1)	(2)	(3)	(4)	(5)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	3.6635*** (0.167)	3.4615*** (0.160)	0.1434*** (0.021)	3.6504*** (0.168)	-0.01836 (0.021)
$\log(since\ album_{it})$		-0.4496*** (0.015)			-0.05848*** (0.003)
Constant	-3.7752*** (0.029)	-0.6702*** (0.101)			
Artist FE	No	No	Yes	No	Yes
Day FE	No	No	No	Yes	Yes
Month FE	No	No	No	Yes	Yes
Observations	4069076	3686099	4069051	4069076	3686089

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. The treatment group is restricted to only those artists that were re-used through cover versions, not digital samples. Columns (1) and (2) include no fixed effects, with Column (2) controlling for the number of days between time t and the last album released by artist i . Column (3) adds an artist fixed effect, Column (4) includes time fixed effects in the form of day-of-week and month-of-year fixed effects. Column (5) includes both artist and time fixed effects, as well as controlling for time since last album

Table 5: Treatment Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>EverReused_i</i>	<i>EverReused_i</i>	<i>EverReused_i</i>	<i>EverReused_i</i>	<i>EverReused_i</i>	<i>EverReused_i</i>	<i>EverReused_i</i>
$\log(\text{avg. Streams}_{it})$	0.01830*** (0.001)	0.01554*** (0.002)	0.01821*** (0.001)	0.01485*** (0.002)	0.01085*** (0.002)	0.02316*** (0.001)	0.02318*** (0.001)
$\log(\text{Followers}_i)$		0.003661** (0.002)		0.004539** (0.002)	0.01896*** (0.002)		
vintage=1960			0.02852** (0.011)	0.02502** (0.011)	0.07427*** (0.013)		0.04518*** (0.012)
vintage=1970			0.04449*** (0.011)	0.04017*** (0.011)	0.08337*** (0.012)		0.05853*** (0.012)
vintage=1980			0.01001 (0.010)	0.005395 (0.010)	0.04661*** (0.012)		0.02743** (0.011)
vintage=1990			-0.01129 (0.008)	-0.01601** (0.008)	0.01854** (0.009)		0.009970 (0.009)
vintage=2000			-0.02794*** (0.006)	-0.03230*** (0.007)	-0.01167 (0.008)		-0.01472* (0.008)
R&B						0.09704*** (0.014)	0.08589*** (0.013)
Constant	0.1565*** (0.005)	0.1146*** (0.020)	0.1580*** (0.007)	0.1095*** (0.021)	-0.03247 (0.041)	0.1562*** (0.006)	0.1452*** (0.008)
Genre FE	No	No	No	No	Yes	Yes	Yes
Observations	11671	11589	11671	11589	9473	9474	9474

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays coefficients from linear probability model regressions in which the dependent variable is a binary variable, EverReused_i , which = 1 if an artist is the target of re-use in my data-set, and 0 otherwise. Column (1) includes only $\log(\text{avg.streams}_{it})$ as a regressor, while Column (2) adds $\log(\text{Followers}_i)$. Columns (3) and (4) include an artist's vintage, their decade of first appearance, as explanatory variables, with the 2010's group omitted as the base factor. Column (5) additionally controls for genre fixed effects. Column (6) and (7) explore the effect of genre by comparing the R&B genre against all other genres.

Table 6: Non-matched sampled: Treatment vs. Selection

	(1)	(2)	(3)	(4)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	0.3062** (0.126)	0.3413*** (0.117)	0.3216*** (0.117)	0.2379** (0.104)
$EverReused_{it}$	2.1459*** (0.142)	2.0745*** (0.133)	2.1039*** (0.134)	2.0878*** (0.121)
$\log(since\ album_{it})$		-0.4409*** (0.015)	-0.4500*** (0.015)	-0.4138*** (0.015)
Day FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Vintage FE	No	No	Yes	Yes
Genre FE	No	No	No	Yes
Observations	4327084	3928784	3928784	3504823

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. The variable $PostReuse_{it} = 1$ for artists that have been re-used. The variable $EverReused_i = 1$ for artists that are re-used at any point during the sample period. Columns (1) and (2) include time fixed effects in the form of day-of-week and month-of-year fixed effects. Column (3) controls for time effects, as well as vintage effects – the year of an artist’s debut. Column (4) controls for time, vintage, and genre.

Table 7: Baseline Regression: Matched Sampled

	(1)	(2)	(3)	(4)	(5)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	0.09511 (0.140)	0.1770 (0.132)	0.1948*** (0.014)	0.02535* (0.015)	0.03046** (0.014)
$\log(since\ album_{it})$		-0.3636*** (0.041)			-0.05991*** (0.005)
Constant	-1.7424*** (0.096)	0.7557*** (0.271)			
Artist FE	No	No	Yes	Yes	Yes
Day FE	No	No	No	Yes	Yes
Month FE	No	No	No	Yes	Yes
Observations	506394	474540	506394	506394	474540

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. The sample is restricted to a matched-sample, wherein each re-used artist is matched to one closely related control-group artist. This procedure reduces the sample to 652 such treatment-control pairs. Columns (1) and (2) include no fixed effects, with Column (2) controlling for the number of days between time t and the last album released by artist i . Column (3) adds an artist fixed effect, Column (4) adds time fixed effects in the form of day-of-week and month-of-year fixed effects. Column (5) includes both artist and time fixed effects, as well as controlling for time since last album

Table 8: Heterogeneity: Prominence

	(1)	(2)	(3)	(4)	(5)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	0.08263 (0.150)	0.03121 (0.156)	0.2168*** (0.022)	0.2322*** (0.023)	0.06103*** (0.023)
$Prominent_i$	4.4447*** (0.125)	4.4456*** (0.125)			
$PostReuse_{it} \times Prominent_i$	-0.1504 (0.187)	-0.1491 (0.187)	-0.04368 (0.028)	-0.06194** (0.028)	-0.05841** (0.028)
$\log(since\ album_{it})$				-0.05031*** (0.005)	-0.06004*** (0.005)
Constant	-3.9373*** (0.102)				
Artist FE	No	No	Yes	Yes	Yes
Day FE	No	Yes	No	No	Yes
Month FE	No	Yes	No	No	Yes
Observations	505955	505955	505955	474540	474540

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. The sample is restricted to a matched-sample of 652 treatment-control pairs. The variable $Prominent_i = 1$ when artist i is of “high prominence,” falling into the top 50% of followers on Spotify. Columns (1) contains no controls, while Column (2) adds time fixed effects. Column (3) adds an artist fixed effect, while dropping the main effect of $prominence_i$. Column (4) adds a control for time since last album. Column (5) includes a full set of controls for both artist and time fixed effects, as well as controlling for time since last album

Table 9: Heterogeneity: Novel re-use

	(1)	(2)	(3)	(4)	(5)
	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$	$\log(streams_{it})$
$PostReuse_{it}$	0.1316 (0.434)	0.3219*** (0.076)	0.05993 (0.467)	0.1168 (0.084)	0.1509* (0.083)
$\log(since\ album_{it})$					-0.1145*** (0.024)
Constant	-3.9542*** (0.287)				
Artist FE	No	Yes	No	Yes	Yes
Day FE	No	No	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes
Observations	54817	54817	54817	54817	42744

Robust standard errors, clustered at the artist level, in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table displays regression coefficients from regressions in which the dependent variable is the log-transformed streaming measure for artist i in time t . Linear regression is used for estimation. The sample is restricted to a matched-sample, wherein each re-used artist is matched to one closely related control-group artist. The sample is further restricted to those artists that have previously been re-used one or fewer times. This procedure reduces the sample to 73 such treatment-control pairs. Column (1) contains no controls. Column (2) includes an artist fixed effect, while Column (3) contains only time fixed effects. Column (4) controls for both artist and time fixed effects, while Column (5) fully controls for artist, time, and time since last album.