Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

Do Songs Become More Popular After Being Sampled?

Harry Beaven

Warwick-Monash Economics Student Papers

September 2022 No: 2022-35

ISSN 2754-3129 (Online)

https://warwick.ac.uk/fac/soc/economics/research/wmesp/

Do Songs Become More Popular After Being Sampled?

By Harry Beaven*

Ever since star-studded copyright infringement cases in the early 1990s concluded that the process of music sampling did not constitute 'fair use' of intellectual property, high licensing costs have made the process prohibitively expensive. Employing streaming service data, this paper reevaluates the traditional ineligibility of the fair use doctrine by presenting empirical evidence of music sampling's effect on the popularity of sampled songs on Spotify over the period 2016-2022. It then examines for which levels of pre-sampling popularity this effect is strongest, as well as the effect of genre and the relationship between the genres of the sampled and sampling song. We find that sampled songs are added to playlists at a 20-40% higher rate for a seven week period after being repurposed within popular songs. Furthermore, original works see greater increases in the rate of playlist addition when there is more scope for sampling to act as informative advertising: when sampled songs were already well known, or had genre characteristics that imply listener familiarity (such as being repurposed in a song of the same genre), our primary findings diminished or disappeared entirely.

JEL Classifications: Z110, O340, K110

Keywords: Music Sampling, Copyright, Fair Use, Music Streaming, Spotify

---

*E-mail: harry@positive2.co.uk.

Acknowledgements: The author wishes to express his gratitude to Christian Soegaard for his supervision and feedback during this project. Any errors remain the author's own.

1 Introduction

Despite a ubiquitous role in contemporary music production, over the last four decades of popular culture no musical technique has caused greater controversy than the process of sampling. The legality of the process - using a small portion (or sample) of an earlier song as a component of a new composition - has often been drawn in to question (Schuster et al., 2019); sampling has historically elicited major copyright infringement cases, the outcomes of which, often in favour of the prosecution, restricted the creative freedoms artists had in using samples (McLeod, 2005). Detractors of sampling assert that the process does not constitute 'fair use' - a doctrine that permits liability-free use of copyrighted material in certain instances.

Despite the vast use of this ideology as evidence against sampling, this assertion has barely seen empirical study. The principal consideration in the fair use analysis is how a new work influences the market for the original (McIntyre, 2012), and as such our research discusses whether being sampled constitutes a method of informative advertising where, after streaming a sampling song, a listener is exposed to the original work and adds it to their playlist.

We present an empirical study of sampling's effect on the popularity of sampled songs on the streaming platform Spotify. To conduct this research, we catalogued a number of songs, all sampled in songs appearing in the Spotify Streaming Charts between 2016-22, and compiled relevant playlist addition data. We employ streaming service data considering that, of the small amount that exists, past research has exclusively used sales volume data, a measure unrepresentative of a global population that primarily streams their music. Using a fixed effects model our investigation finds that, to a 99% degree of statistical significance, the rate of playlist addition for sampled songs increases following the success of a song in which an excerpt was repurposed. We also extend existing discussion of the mechanism behind these findings by providing empirical evidence that this effect is most present when there is scope for sampling to act as informative advertising: we find that when sampled songs were well known prior to sampling, or had genre characteristics that implied listener familiarity (such as being a pop song or being repurposed in a song of the same genre), our findings diminished or disappeared entirely.

The paper proceeds as follows: the next section introduces the historical precedent for the legality of sampling and discusses its economic effects. This includes a review of the relevant literature and appropriate court decisions. In section three we discuss the data set and the importance of using streaming data. Section four and five present and evaluate our empirical model and results respectively, finding a statistically significant increase in the rate of playlist additions for original songs after being sampled. This section also examines the extent to which attributes of the sampling and sampled songs such as obscurity and genre influence the strength of this effect, and reflects on the limitations of our research. Section six concludes the paper, provides guidance for future research and discusses the implications of our findings not only for a reevaluation of the common ineligibility of fair use law in sampling cases, but also for music firms to allow sampling strategically to revive older releases on streaming platforms.

2 Literature Review

2.1 Background

Despite the inherence of securing adequate legal permissions in contemporary music sampling, the practice spent a great deal of time unregulated: from its origins with Jamaican DJs in the 1960s to its mainstream breakthrough in the early 1990s, securing clearance of samples before using them was uncommon (Norek, 2004). Developing in parallel with Hip-Hop culture (Smith, 2007), as sampling became more pervasive it also became infamous: sampling potentially represented an unchecked infringement of copyright law and was seen by many academics to be an uncompetitive practice in the market for recorded music, or a form of piracy (Thom, 1988). This attitude was founded largely on economic grounds: Sampling was regarded as costing artists and copyright holders millions of dollars (Houle, 1991; Flanders, 1991) and, due to the extensive use of tape machines to record a sample, was associated with the cost of taping-related displacement of music industry sales for publishers (U.S. Senate, Committee on the Judiciary, 1984). Without sufficient regulation to protect copyright holders, the anti-sampling school felt justified, especially considering many sampled artists lacked the funds to go to court (Brown, 1992).

This 'golden era' of unchecked sampling was brought to an end by the landmark victory of the prosecution in the 1991 copyright infringement case Grand Upright vs Warner. This suit concluded with the court ruling that Biz Markie's unauthorised use of a sample from Gilbert O'Sullivan's 'Alone Again, Naturally', and all unauthorised sampling thereupon, constituted copyright infringement, a decision instrumental in developing the strict framework surrounding the legality of sampling in the western music industry (McLeod, 2005). This was intensified by Bridgeport v Dimension Films (2005), whose verdict that a two-second guitar sample was in violation of copyright law effectively eliminated the de minimis doctrine (that small scale use should not come under scrutiny) from sampling, and affirmed the idea that any sample, regardless of size or context, must be licensed (DiCola and McLeod, 2011). This principle has, to a large extent, made sampling prohibitively expensive to producers without access to sufficient funds (Meiselman, 2016).

2.2 Empirical Literature

The absolutism of these decisions, as well as the certainty of anti-sampling literature is discredited by the insufficient analysis that supports these claims. Of the two empirical research papers that exist on the effect that digital sampling has on subsequent music sales, both found that artists benefit from being sampled. Schuster (2014) finds that copyrighted songs sampled in Gregg Gillis' successful 2010 album 'All Day' sold 3.2% better immediately after release, significant at the 92.5% level. However, these findings are largely limited by a lack of review over heterogeneous periods and sampling artists. Furthermore, Schuster removes both sampled songs that charted close to release as well as songs released shortly before 'All Day' to eliminate sales data biased towards tracks that would see an inevitable decline over the study period but in doing so makes the relevant data set even less convincing as a random sample. This is improved upon in Schuster et al. (2019), where the authors employ a data set of songs sampled in tracks that charted on the Billboard Hot 100 from 2006 to 2015, finding that sales of sampled songs increased after being repurposed in a new work to a 99.99% level of significance. Despite the robustness of these results, the paper ignores the sales displacing effect of music streaming by using sales instead of streaming data, the use of which would provide an analysis more representative of the listening habits of the population (Aguiar and Waldfogel, 2018). This paper improves on the two that precede it in this field with the use of streaming data, discussed in the next section, extended analysis of the mechanism behind our results, discussion of how genre and popularity attributes alter findings as well as controlling for the number of active streaming service users.

2.3 Mechanism

Depictions of the mechanism by which being sampled can increase an artist's popularity are similarly hard to come by; sampling literature overwhelmingly describes the effect of hearing song excerpts in a context excluding repurposing in a new work. However, these papers still aid in the identification of an informative channel through which this mechanism acts. Peitz and Waelbroeck (2004) suggest sampling's value as a method of promotion: by manipulating a model of a simple multi-product monopoly environment, they find that hearing samples of music comprises an alternative channel of information provision in digital music markets, directing new listeners to the sampled artist. Further models by Peitz and Waelbroeck (2006) as well as Gopal et al. (2006) infer the role of this informative channel in the mechanism discussed in our research.

Evidently, there is potential for sampling to benefit the sampled artist. The pervasiveness of the argument that unauthorised sampling constitutes copyright infringement has led to millions of dollars being paid out in compensation to sampled artists every year, despite there existing minimal empirical research actually supporting the need for such a relationship to exist to these extremities.

3 Data Description

3.1 Data Set

The analysis involves cross-section of 92 Sampled Songs over 2016-2022.

3.1.1 Sampling Songs

Analysis began by identifying a set of songs that sampled earlier works. 'Hit' songs were employed considering an informative advertising effect is more likely to be identifiable if we increase the scale of public exposure. Creating a criteria for sampling songs allows us to more easily compare our findings across song attributes and reduces the need to use an unmanageably large data set to identify significant results. Therefore, we define a 'hit' song as one that charted on a global or national Spotify streaming chart during the study period. Whosampled.com was used as a resource for identifying which songs contained samples. The website is a crowdsourced database of information about sample-based music, covers and remixes. This information is amassed by over 27,000 contributors whose work is verified before being added. To this end we were able to identify a set of over 100 'hit' songs that re-used portions of earlier works.

We then collected data on the date when each of these songs reached its highest position on Spotify streaming charts to find a proxy for when the song's exposure to the public was at its highest. This date acts as a benchmark for when sampling's effect may begin to appear. By default, this date was taken as the date of highest position on the Global Streaming Chart, as this represents exposure of the sample to a larger audience than on a National Chart. If the song only charted on one National Chart, the date of peak on this chart was used. If the 'hit' song charted on several national charts, we resolved to drop songs where the dates of peak varied by over a week from our data set. For songs whose national chart peak dates did not vary to this extent, we calculated the average date of peak between all national charts, and marked this as the date of chart peak. Consequently, we were able to date the chart peak of 92 sampling songs. From this, we constructed a relative time variable where $t = n$ for n days after chart peak.

3.1.2 Sampled Songs

From our set of 92 sampling songs, we identified 92 corresponding songs that were each sampled by one of the songs in this set. To control for heterogeneity in song mood, these songs are reasonably evenly distributed across mode and key. We resolved to strive for an even distribution instead of one more representative of the mood and key characteristics of the current charts, as these trends are likely to change and could limit the relevance of our research to future discussion.

As a measure of a song's popularity on streaming services, we employ the weekly rate of playlist addition on Spotify. We chose playlist additions instead of the daily number of streams a song receives primarily due to easier accessibility. This measure arguably also proves a better indicator of the longevity of popularity increases; a song being added to a playlist is more likely to indicate repeated listens in the long term than a single stream. We believe the assumption that a more popular song would be added to playlists at a faster rate is a reasonable one, as playlists represent one of the primary forms of interaction on streaming platforms (Hagen, 2015). Daily data measuring the historical number of playlists a song was included in on Spotify was retrieved from Spotify's application programming interface (API) and compiled in to one data set. Each date of observation was then paired, by song, to our constructed relative time variable. Playlist addition rate was constructed by taking the 7-day difference of the number of playlists a song was included in on a given day. Following Schuster et al. (2019) as well as allowing for a larger reference period, we dropped observations that took place over 11 weeks before and 10 weeks after the sampling song's chart peak.

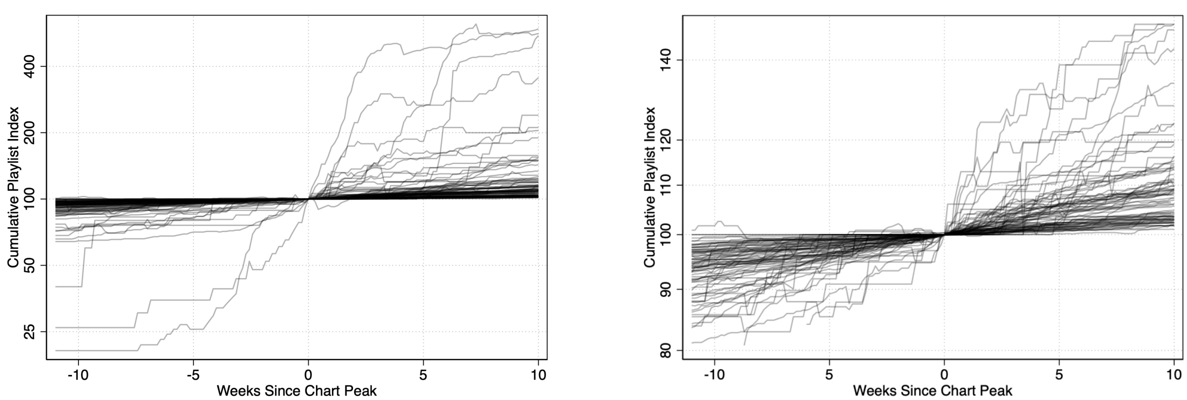
Following a visual analysis, we identified semi-regular anomalies caused by measurement error in Spotify's API where the number of playlists a song was included in would drop significantly, often by over 90%, and then immediately return to the previous observation. To locate these anomalies, we employed a criteria where if an observation was seen to cause two equal spikes in opposite directions on a visual inspection of the 7-day difference variable, it was dropped from the data set as this likely represented an isolated error in measurement and not a decision of a mass of users to remove and then restore a song from their playlist.

Whosampled.com was also used to retrieve genre information for all sampling and sampled songs. Although our genre categories are broad, this is necessary for ease of analysis in 5.2 when we split our data in to subsets.

3.1.3 Streaming Data

The need for analysis of sampling's effects using streaming data was noted in Schuster et al. (2019). Physical music sales volume has declined consistently since 2011, whereas streaming income continues to consolidate its position as the largest contributor to recorded music revenues since 2015. Despite this, literature that exists on this topic exclusively uses music sales volume data as a measure of popularity: this is not representative of consumers' current listening habits and revenue sources for rights holders. For this reason, we cannot understate the importance of using streaming data in our analysis.

3.2 Preliminary Analysis



(A) Full Dataset

(B) 14 Fastest Growing Songs Removed

FIGURE 1: Indexed Playlist Count Pre- and Post-Sampling

We constructed the 'Cumulative Playlist Index' (CPI), an index of the number of playlists a song was in over the 21 week period, with $CPI = 100$ on the date of chart peak, to make visual analysis simpler. This analysis allows us to better understand the distribution of sampled song responses to the success of the sampling song within our data set. Figure 1A makes evident that experiences after being sampled vary dramatically between songs; although some songs are included in 3 times as many playlists only 5 weeks after the sampling song peaks in chart performance, roughly 85% of songs find themselves in around 0-50% more playlists (Figure 1B). We discuss in 5.2 some potential determinants of this variation. Still, it is clear that after sampling, a considerable amount of songs are added to playlists at a faster rate than before.

4 Methodology

To correct for omitted variables that may have affected playlist additions for individual songs and considering methods used by Schuster et al. (2019), we employ a fixed effects model. Our main estimating equation takes the following form:

$$log(additionrate_{it}) = \delta_t + o_i + \alpha + \sum_{k=1}^{19} \beta_k \sum_{n=-9}^{9} nweekspostpeak_{it} + \epsilon_{it}$$

where our dependent variable is a logarithmic transformation of our constructed 7-day difference variable. To capture the effect of sampling week-by-week we constructed 19 weekly dummies, which take a value of 1 if an observation is taken from n weeks after the sampling song peaked in the charts for $Z \cap \{n : -9 \geq n \geq 9\}$ respectively (9 weeks before chart peak to 9 weeks after). We exclude dummies for the weeks 10 and 11 prior to chart peak to avoid perfect colinearity. As a result, our estimates should be interpreted as relative to this two week period. Therefore, with $k = n + 10$, $100 \times \beta_k$ represents the % change in a song's weekly playlist addition rate for week n after the chart peak of the song it was sampled in compared to the period 10-11 weeks before this chart peak.

To control for factors that may have increased playlist additions for all songs, we also provide a second specification which includes a constructed variable measuring the change in quarterly users on Spotify. This allows us to control for increases in playlist addition for sampled songs that are caused by there simply being more active users on Spotify. Both specifications control for song (o_i) and day (δ_t) fixed effects.

To test if the variance of the error term is dependent on the explanatory or control variables, we perform a modified Wald test for groupwise heteroscedasticity (Baum, 2001). We reject the null hypothesis of homoscedasticity and so use robust standard errors in our regression to obtain more accurate t-ratios for significance testing.

5 Empirical Results and Analysis

5.1 Primary Results

Our primary results are presented in Table 1. Column 1 confirms that there is strong evidence of original works becoming more popular on streaming services after being sampled. We identify an 8 week period beginning the week of the sampling song's chart peak where sampled songs are added to playlists at a rate significantly higher than in the reference period. This effect is strongest 1 week after the sampling song peaks, or one week after public exposure to the sample was at its highest, where the playlist addition rate was 40.3% higher than in the reference period. All eight of these estimates are significant at the 90% level, six at the 95% level and three at the 99% level (these also being the weeks with the three largest increases: Weeks 1, 2 and 6). We identify no significant increases in playlist addition rate before the sampling song reaches its peak chart position. Our estimates almost quadruple those of Schuster et al. (2019) in magnitude, implying the effect of sampling on the market for the original work is even more pronounced on streaming services than in physical music sales.

Particularly surprising is the result that, though playlist addition rate (and significance) does decrease following its apex until week 4, we identify a second peak 6 weeks after chart peak of an addition rate 30.9% higher than the reference period, significant at the 95% level. It is unlikely that this is explained by a second peak in public exposure to the song; Asai (2009) demonstrates how song sales' geometric decline is likely to shift even further towards the earlier stage of the sales cycle as more people use the internet to listen to music. Although this undulation may disappear with a larger data set, this could indicate that the mechanism

Table 1: Playlist Addition Rate of Sampled Song Around Chart Peak of Sampling Song

	Log of Playlist Addition Rate			Log of Playlist Addition Rate	
	(1)	(2)		(1)	(2)
9 Weeks Prepeak	0.015	0.006	2 Weeks Postpeak	0.337^{***}	0.310^{**}
	(0.051)	(0.052)		(0.120)	(0.119)
8 Weeks Prepeak	0.062	0.053	3 Weeks Postpeak	0.284^{**}	0.260^{**}
	(0.062)	(0.062)		(0.110)	(0.111)
7 Weeks Prepeak	0.089	0.078	4 Weeks Postpeak	0.153^{*}	0.124
	(0.060)	(0.061)		(0.091)	(0.094)
6 Weeks Prepeak	-0.012	-0.025	5 Weeks Postpeak	0.247^{**}	0.216^{**}
	(0.074)	(0.075)		(0.096)	(0.099)
5 Weeks Prepeak	0.038	0.028	6 Weeks Postpeak	0.309^{***}	0.277^{**}
	(0.076)	(0.076)		(0.106)	(0.109)
4 Weeks Prepeak	0.028	0.017	7 Weeks Postpeak	0.213^{**}	0.180^{*}
	(0.066)	(0.067)		(0.104)	(0.106)
3 Weeks Prepeak	0.102	0.086	8 Weeks Postpeak	0.181	0.146
	(0.074)	(0.076)		(0.117)	(0.122)
2 Weeks Prepeak	0.042	0.028	9 Weeks Postpeak	0.133	0.102
	(0.088)	(0.089)		(0.105)	(0.109)
1 Week Prepeak	-0.022	-0.044	New Spotify Users		0.014^{**}
	(0.094)	(0.097)			(0.006)
Week of Chart Peak	0.212^{*}	0.184^{*}	Constant	1.373^{***}	1.173^{***}
	(0.107)	(0.108)		(0.055)	(0.082)
1 Week Postpeak	0.401^{***}	0.372^{***}			
	(0.106)	(0.105)			
Observations				13447	13445
Groups				92	92
Day and Song Fixed Effects				Yes	Yes
Change in Spotify Users				No	Yes

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

depicted in these results works through two channels: one acting around the time the sample reaches its highest level of public exposure, and the other five or six weeks after. This defies the idea expressed in the literature that sampling works through one informative channel.

Indeed, the fair use classification put forward by Fagundes (2013) illuminates four channels through which repurposing can affect the market for an original work. Alongside 'recognition' and 'reincarnation', both information provision channels, Fagundes describes two alternate characterisations of the mechanism: 'affirmation', where repeated use of a copyrighted work creates the perception that this work is valuable and 'innovation', the potential for novel uses of copyrighted works to supplement enjoyment of the original work. Schuster et al. (2019) find no statistically significant difference in the increase in sales between samples that had been repeatedly sampled and those that had not, suggesting the second peak is unlikely to be driven by the former classification. It is then arguable that the second peak we identify may comprise an innovation effect: as listeners form an emotional connection with a song containing a sample, they may listen to the song where the sample originated as an extension of that experience. Future research regarding user behaviour on streaming platforms is needed to examine this theory.

Excluding the estimate for four weeks post-peak, these results remain significant when we control for changes in the number of active Spotify users in column 2. This helps reassure us that our results are not being driven by industry-wide trends. Regardless of econometric specification, the central result that, after being sampled, songs are added to playlists at a rate up to 40% higher than before is robust.

5.2 Extensions

We next extend our research to empirically examine how our findings vary across different levels of obscurity and genres. This allows us to further investigate the mechanism behind our primary findings. Considering our fixed effects model is incompatible with time-invariant variables, we instead split our data set in to subsets and compare results.

5.2.1 Obscurity Effects

To categorize song obscurity before being sampled, we segment the data into tertiles by the number of playlists a sampled song was included in 11 weeks before $t = 0$ and attribute each song a rank of 1-3 dependent on which tertile it was included in, then running our primary

Table 2: Playlist Addition Rate Across Pre-Sampling Popularity Subsets

	Log of Playlist Addition Rate				Log of Playlist Addition Rate		
	(1)	(2)	(3)		(1)	(2)	(3)
	Low	Medium	High		Low	Medium	High
9 Weeks Prepeak	-0.011	0.012	0.041	1 Week Postpeak	0.698***	0.309**	0.198
	(0.044)	(0.115)	(0.094)		(0.184)	(0.148)	(0.215)
8 Weeks Prepeak	0.094	0.012	0.080	2 Weeks Postpeak	0.620***	0.326**	0.048
	(0.070)	(0.133)	(0.110)		(0.181)	(0.159)	(0.275)
7 Weeks Prepeak	0.047	0.186	0.024	3 Weeks Postpeak	0.441***	0.310**	0.092
	(0.058)	(0.120)	(0.121)		(0.150)	(0.149)	(0.266)
6 Weeks Prepeak	0.124	-0.071	-0.088	4 Weeks Postpeak	0.433***	0.0288	0.002
	(0.089)	(0.150)	(0.138)		(0.126)	(0.136)	(0.201)
5 Weeks Prepeak	0.086	-0.026	0.057	5 Weeks Postpeak	0.548***	0.135	0.061
	(0.066)	(0.143)	(0.169)		(0.149)	(0.110)	(0.218)
4 Weeks Prepeak	0.132	0.020	-0.070	6 Weeks Postpeak	0.688***	0.155	0.089
	(0.091)	(0.122)	(0.126)		(0.192)	(0.130)	(0.212)
3 Weeks Prepeak	0.228*	0.078	-0.0003	7 Weeks Postpeak	0.532***	0.148	-0.045
	(0.127)	(0.119)	(0.137)		(0.161)	(0.154)	(0.213)
2 Weeks Prepeak	0.210	-0.077	-0.001	8 Weeks Postpeak	0.522***	0.181	-0.167
	(0.141)	(0.131)	(0.184)		(0.149)	(0.140)	(0.286)
1 Week Prepeak	0.192	-0.202	-0.047	9 Weeks Postpeak	0.323**	0.246*	-0.185
	(0.132)	(0.153)	(0.199)		(0.127)	(0.135)	(0.255)
Week of Chart Peak	0.448***	0.009	0.190	Constant	0.138*	1.127***	2.903***
	(0.151)	(0.154)	(0.241)		(0.072)	(0.078)	(0.125)
Observations					4425	4694	4328
Groups					30	32	30
Day and Song Fixed Effects					Yes	Yes	Yes

Standard errors in parentheses

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

specification on each group. Pre-sampling popularity can be thought of as a proxy for the potential for an informative effect of sampling: we intuitively expect that the more obscure a song is before sampling, the larger the base of consumers who can be informed of that song's existence. Table 2 demonstrates that the strength of the mechanism increases with the potential for informative advertising. Starting from Week 0, we identify a 10 week period of addition rate increase of up to 69.8% for songs in the lower tertile (significant at 99% for the first 9 weeks). For the medium tertile, we identify a three week period of an increase around the 30% range significant at the 95% level. We do not identify any significant results for the tertile of songs that were the most popular prior to sampling. This is indicative that the mechanism is driven first and foremost by an informative channel: there is significantly reduced scope to benefit from being sampled if the work in question is already well known.

It is also worth noting that we see the same significant undulation in magnitude of estimates in the first two groups, albeit over different time periods. In the low group we observe peaks in weeks 1 and 6, with a trough in week 4; in the medium, peaks appear in the second and ninth week. This further evidences the idea that the mechanism does not act solely through an informative effect: there exists at least one other channel that acts several weeks after exposure has declined. Again, further research is warranted in this regard, as our data set does not suffice to empirically explain this.

5.2.2 Genre Effects

Considering that genre changes the way in which sampling is used (Boone, 2013) as well as impacting listener demography (Roy and Dowd, 2010), we expect the strength of our results to vary across genres. We assembled a data set that codes the genre of all 184 sampling and sampled songs and test for each genre subset (separately for sampled and sampling songs), whether playlist addition rate increases significantly after sampling. Considering the amount of subsets and for simplicity of interpretation, this regression only uses three dummies for time periods: one for the entire 9-week period before chart peak, one for the first 5 weeks after peak, and another for 5 weeks after that. We condense the prepeak period in to one dummy variable considering we failed to identify any significant results at all over this period in Table 1.

Considering Table 3 at the 95% significance level, we see original works sampled in popular Hip-Hop, Rap or R&B songs were likelier to become more popular after sampling than works sampled in hits of other genres. However we fail to identify any significant increases when Hip-Hop, Rap or R&B songs are the original work. This indicates that there is little scope for songs of these genres to benefit from being sampled, but considerable scope for songs of *any* genre to benefit if sampled in a hit Hip-Hop, Rap or R&B song. As sampling *of* Hip-Hop songs

Table 3: Playlist Addition Rate by Genre of Sampled and Sampling Song

	Log of Playlist Addition Rate							
	Genre of Sampled Song					Genre of Sampling (Hit) Song		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Electronic/ Dance	Jazz/ Blues	Pop/ Rock	Hip-Hop/ Rap/RnB	Soul/Funk/ Disco	Electronic/ Dance	Pop/ Rock	Hip-Hop/ Rap/RnB
Prepeak Period	0.195	0.333	-0.105	0.047	-0.004	0.383*	-0.139	0.036
	(0.165)	(0.182)	(0.108)	(0.096)	(0.091)	(0.193)	(0.115)	(0.058)
Weeks 0 to 5	0.380**	0.676***	-0.130	0.223	0.462***	0.604	0.086	0.277***
	(0.154)	(0.164)	(0.219)	(0.139)	(0.162)	(0.346)	(0.171)	(0.099)
Weeks 5 to 9	0.484*	0.428**	-0.006	0.029	0.331	0.592*	-0.054	0.200*
	(0.235)	(0.159)	(0.241)	(0.154)	(0.210)	(0.255)	(0.277)	(0.112)
Constant	0.889***	0.353**	2.109***	1.765***	0.925***	0.831***	1.938***	1.315***
	(0.129)	(0.131)	(0.130)	(0.096)	(0.105)	(0.210)	(0.116)	(0.063)
Observations	1325	1179	2906	3933	4104	1179	2173	10095
Groups	9	8	20	27	28	8	15	69

Standard errors in parentheses

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

is commonly not transformative, often taking the form of simply re-using a beat (Berry, 2017), we might expect the homogeneity of these two products to discourage listening to the song from which the beat originated. However, sampling *in* Hip-Hop/Rap/R&B songs tends to be transformative, pulling from many genres (Schloss, 2014), and so this proposed 'beat recycling' effect is less apparent in column 8.

Furthermore, we find no evidence that sampling increased the popularity of sampled songs when either the sampled or sampling song was considered pop or rock music. This is unsurprising considering, within our data set, over 50% of songs sampled in Pop/Rock hits were in the highest tertile of prior popularity; this indicates Pop/Rock songs tend to sample already popular songs. As such, the power of the informative channel is limited.

We also note the strength of the effect for sampled songs of largely antiquated genres comprised of older releases. We identify significant increases of up to 67.6% for sampled songs categorized under Jazz/Blues and 46.2% for Soul/Funk/Disco: these categories do not describe any 'hit' *sampling* songs included in this study. This supports Schuster et al.'s (2019) finding

Table 4: Playlist Addition Rate for Same-Genre and Cross-Genre Relationships

	Log of Playlist Addition Rate			Log of Playlist Addition Rate	
	Same-Genre	Cross-Genre		Same-Genre	Cross-Genre
9 Weeks Prepeak	0.130	-0.052	1 Week Postpeak	0.334**	0.440***
	(0.104)	(0.053)		(0.164)	(0.140)
8 Weeks Prepeak	0.157	0.006	2 Weeks Postpeak	0.139	0.451***
	(0.118)	(0.070)		(0.154)	(0.166)
7 Weeks Prepeak	0.117	0.072	3 Weeks Postpeak	0.247	0.306**
	(0.101)	(0.075)		(0.167)	(0.146)
6 Weeks Prepeak	-0.017	-0.009	4 Weeks Postpeak	0.112	0.177
	(0.126)	(0.093)		(0.135)	(0.122)
5 Weeks Prepeak	0.087	0.009	5 Weeks Postpeak	0.213	0.266**
	(0.143)	(0.088)		(0.150)	(0.125)
4 Weeks Prepeak	0.072	0.002	6 Weeks Postpeak	0.204	0.370**
	(0.125)	(0.075)		(0.132)	(0.150)
3 Weeks Prepeak	0.168	0.065	7 Weeks Postpeak	0.036	0.318**
	(0.124)	(0.091)		(0.118)	(0.149)
2 Weeks Prepeak	0.034	0.046	8 Weeks Postpeak	0.109	0.223
	(0.173)	(0.098)		(0.170)	(0.158)
1 Week Prepeak	-0.030	-0.018	9 Weeks Postpeak	0.131	0.134
	(0.174)	(0.111)		(0.171)	(0.134)
Week of Chart Peak	0.170	0.236*	Constant	1.658***	1.208***
	(0.173)	(0.138)		(0.085)	(0.072)
Observations				4939	8508
Groups				34	58
Day and Song Fixed Effects				Yes	Yes

Standard errors in parentheses

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

that sampling may increase popularity of a song most when songs are relatively older at the time of sampling. Table 2's evidence for the existence of the informative channel also explains this: intuitively, songs are more likely to become more obscure the older they become and as such there is larger potential for sampling to comprise informative advertising.

To further analyse the discrepancy of our genre subset results between sampling and sampled songs, Table 4 splits our data set in to two subsets depending on whether or not the sampling-sampled pair is homogeneous in genre. Playlist addition rate increases are strongest when the songs differ in genre; for this subset we identify two three-week periods, separated by week 4, where songs were added to playlists at an average rate of around 35% (significant at the 95% level). In the same-genre subset, we only identify one week of a 33.4% increase in playlist addition rate. Our mechanism can therefore be seen to also depend on how *transformative* a sample's use is.

5.3 Limitations

However, the present results are subject to some limitations. Firstly, because our research focused on using streaming data, our study period was constrained to a shorter length and thus we involve less songs than in Schuster et al. (2019). Although this is unlikely to have a large impact on our primary findings, to be confident in our subset-based analyses in 5.2 would require a much larger pool of data. This is particularly true in our genre-based analysis, where some regressions employed as few as 8 songs and as such are unlikely to be representative of the entire genre.

The scope of our data also limits the relevance of our findings. Considering we only involve songs sampled in *popular* songs, it is uncertain whether sampling increases popularity when the exposure of the sample to the public is much more limited.

Finally although our fixed effects model controls for individual song effects, our 'Change in Spotify Users' variable likely does not suffice to control for all industry-wide effects. This variable lacks precision, being quarterly data compared to the daily data we employ as our dependent variable. Our research could therefore be improved by a more nuanced set of control variables, perhaps measuring our results against total playlist additions on the date of observation.

6 Conclusion

This paper aimed to discuss whether original works became more popular after being sampled by hit songs. In this regard we have two main empirical findings. First, sampled songs are added to playlists at a rate about 20-40% faster for a 7-week period after being sampled. These results are statistically significant and robust. Second, original works see greater increases in the rate of playlist addition when there is more scope for sampling to act as informative advertising: when sampled songs were already well known, or had genre characteristics that imply listener familiarity (such as a same-genre relationship), our primary findings diminished or disappeared entirely.

These findings support a judicial determination that in many cases sampling does constitute fair use. This is antithetical to the popular belief that unauthorised sampling represents copyright infringement, that being sampled without permission is to the detriment of the original artist. Our results instead suggest that being sampled in a popular song does not steal revenue from an artist, but instead exposes new audiences to their music and creates engagement on streaming platforms. Publishers could theoretically use this to their advantage by clearing sample usage to popular artists, whereby sampling could introduce the copyrighted work to a large audience. However, this is subject to some caveats. Our results also evidence that the more familiar an audience is with a work, the less likely it is that being sampled will increase listener engagement. In this sense, sampling is most likely to constitute fair use through a beneficial effect on the original market in cases where the sample is either sufficiently obscure or used transformatively.

Examining whether our findings scale with the success of the hit song represents a priority for future research. Although our results suggest there are opportunities for music publishers to resurrect obscure back catalog works by clearing their use as samples, these results only consider songs sampled in *popular* songs. As such, we cannot state confidently whether it would be beneficial to publishers to clear samples to smaller artists who are unlikely to have as broad a reach as those in our data set.

Future research should also examine further if other fair use classifications can be applied to sampling. This paper stresses the importance of sampling's role as informative advertising, yet the bipeakedness of our results suggests that there may exist a second channel through which sampling promotes the sampled work. Again, these findings are limited by the size and scope of our data set. Our research could be expanded by examining songs from a wider study period to involve more sampled songs and considering a greater variation in the success of the sampling song. This should also include more precise controls for industry-wide effects, possibly combined with gathering information on user behaviour on streaming platforms.

Bibliography

Aguiar, L. and Waldfogel, J. (2018). As streaming reaches flood stage, does it stimulate or depress music sales? *International Journal of Industrial Organization*, 57:278–307.

Asai, S. (2009). Sales patterns of hit music in japan. *Journal of Media Economics*, 22(2):81–101.

Baum, C. (2001). Xttest3: Stata module to compute modified wald statistic for groupwise heteroskedasticity.

Berry, P. A. (2017). 24 rappers rapping over other artist's beats recently. *XXL Magazine*.

Boone, C. (2013). Mashing: Toward a typology of recycled music. *Music theory online*, 19(3).

Brown, J. H. (1992). They don't make music the way they used to: The legal implications of sampling in contemporary music comment. *Wisconsin Law Review*, 1992(6):1941–1992.

DiCola, P. C. and McLeod, K. (2011). Creative license: The law and culture of digital sampling.

Fagundes, D. (2013). Market harm, market help, and fair use. *Stan. Tech. L. Rev.*, 17:359.

Flanders, B. L. (1991). Barbarians at the gate. *American Libraries*, 22(7):668–669.

Gopal, R., Bhattacharjee, S., and Sanders, G. (2006). Do artists benefit from online music sharing? *The Journal of Business*, 79(3):1503–1533.

Hagen, A. N. (2015). The playlist experience: Personal playlists in music streaming services. *Popular Music and Society*, 38(5):625–645.

Houle, J. R. (1991). Digital audio sampling, copyright law and the american music industry: Piracy or just a bad rap. *Loyola Law Review*, 37(4):879–902.

McIntyre, S. (2012). Private rights and public wrongs: Fair use as a remedy for private censorship. *Gonz. L. Rev.*, 48:61.

McLeod, K. (2005). *Freedom of expression®: Overzealous copyright bozos and other enemies of creativity*. Doubleday.

Meiselman, J. (2016). Sampled or stolen? Untangling the knotty world of hip-hop copyright. *FACT Magazine*.

Norek, J. (2004). You can't sing without the bling: The toll of excessive sample license fees on creativity in hip-hop music and the need for a compulsory sound recording sample license system. *UCLA Ent. L. Rev.*, 11:83.

Peitz, M. and Waelbroeck, P. (2004). File-sharing, sampling, and music distribution. *International University in Germany Working Paper*, (26).

Peitz, M. and Waelbroeck, P. (2006). Why the music industry may gain from free downloading — the role of sampling. *International Journal of Industrial Organization*, 24(5):907–913.

Roy, W. G. and Dowd, T. J. (2010). What is sociological about music? *Annual Review of Sociology*, 36:183–203.

Schloss, J. G. (2014). *Making beats: The art of sample-based hip-hop*. Wesleyan University Press.

Schuster, M., Mitchell, D., and Brown, K. (2019). Sampling increases music sales: An empirical copyright study. *American Business Law Journal*, 56(1):177 – 229.

Schuster, W. M. (2014). Fair use, girl talk, and digital sampling: An empirical study of music sampling's effect on the market for copyrighted works. *Okla. L. Rev.*, 67:443.

Smith, A. L. (2007). Other people's property: Hip-hop's inherent clashes with property laws and its ascendance as global counter culture. *Va. Sports & Ent. LJ*, 7:59.

Thom, J. C. (1988). Digital sampling: Old-fashioned piracy dressed up in sleek new technology comment. *Loyola Entertainment Law Journal*, 8(2):297–336.

U.S. Senate, Committee on the Judiciary (1984). Video and audio home taping: hearing before the subcommittee on patents, copyrights, and trademarks of the committee on the judiciary, united states senate, ninety-eight congress, first session on s. 31 ... and s. 175 ... october 25, 1983. v, 525 p.

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

"*We are very happy to introduce the Warwick Monash Economics Student Papers (WM-ESP). The Department of Economics of the University of Warwick and the Economics Department at Monash University are very proud of their long history of collaboration with international partner universities, and the Monash Warwick Alliance reflects the belief in both Universities that the future will rely on strong links between peer Universities, reflected in faculty, student, and research linkages. This paper series reflects the first step in allowing our Undergraduate, Honours, and Masters students to learn from and interact with peers within the Alliance.*"

Ben Lockwood (Head of the Department of Economics, University of Warwick) and Michael Ward (Head of the Department of Economics, Monash University)

Recommended citation: Beaven, H. (2022) Do Songs Become More Popular After Being Sampled? *Warwick Monash Economics Student* Papers 2022/35.

[1] Warwick Economics would like to thank Lory Barile, Gianna Boero, and Caroline Elliott for their contributions towards the selection process.