

This paper attempts to replicate some of the results from an influential paper by Felix Oberholzer-Gee and Koleman Strumpf (henceforth OS), which examines the effect of file sharing on legal music sales. While the main estimation in the original paper is based on proprietary data, the paper also includes four quasi-experiments which are designed to support the main findings, and these quasi-experiments are based on data which are more readily available. These experiments are replicated here.

I appreciate what this replication paper tries to do. The author shows that the supporting quasi-experimental evidence in OS has some flaws. In particular, I like the expansions which include a control for unemployment and (more importantly) a new measure of genre-specific piracy intensity.

It is difficult to evaluate a replication paper which only replicates the (admittedly not very strong) supporting evidence given in the original paper, but I won't blame the author for that. Instead, I will take a closer look at the replications at hand. Some of these replications are more convincing than others.

1. Temporal variation in vacations:

The underlying assumption of the first quasi-experiment is that piracy in the US would decrease as US college students leave campus during the summer. This replication paper examines the foundation of this (admittedly strong) assumption. The number of file sharers per month varies over time, with a decrease in piracy in the first summer but not in the following summers. The regressions seem to be done correctly, although the data on file sharers is less than ideal for this. How do we know that these file sharers shared as many files as they did when they had high-speed internet?

2. Spatial variation in exposure to European file sharers:

Due to time zone differences, OS assumes that there is more piracy on the East Coast (which is close to Europe), compared to the West Coast. I agree with the author that regardless of the findings, the identification strategy is weak, and I wouldn't place much weight on this experiment.

The author shows percent changes in sales between different regions over time (Table 2). These changes show a slightly different story than OS. The author could do more for this particular experiment. Perhaps OS used region specific linear time trends, consistent with a difference-in-differences estimation. In that case I can see how the analysis would show no significant difference between the regions.

3. Regressing total sales on total file sharers

OS assumes that on the industry level, interest in music would not have changed were it not for the emergence of file sharing. Disregarding the problems with this assumption, the direct replication here finds different results than the original paper.

The author continues with a "broader" replication, in which he adds a control for the economic conditions to the regression. Here, I am unsure whether unemployment is the right variable to control for, but I could believe this. I might make sense to try further control variables.

The second to last row in Table 3 displays changes in sales over 3 years. Why not leave it at the monthly level, or at least at an annual level? Similarly, the last row in this table shows what percentage of the decline is due to piracy. If the absolute decline was small, then 20% (or 33%) of

something small is small as well. The author could explain these magnitudes better. The currently table dramatizes the effect.

4. Variation in download intensity across genres

This experiment has the most promise within the OS paper, and it thus deserves more focus. The underlying assumption is that some genres are more affected by piracy than others. This paper lays out the issues with this “quasi-experiment” clearly. It is difficult to know what OS did, based on the paragraph in their paper, especially concerning the different genres and the measure of piracy intensity.

I do appreciate the extensions in this replication – in fact, I also find this replication more convincing than the others. I agree that a per-sale measure of piracy is better than a per-title measure, and dropping the year 1999 is sensible as well. I find this replication convincing.

Regarding the economic impact, again, 12% of something small (16.6%) gives us approximately a 2-percent decrease in sales at the mean. Using the “OS corrected” measure of piracy, the estimated decrease due to piracy is only 1.3%, and not statistically significant.

5. Data replication

It is true that OS are vague in their claims of rising sales (and the definition of “major markets”), and a reference for the claims would have been helpful. That being said, I do not see much value in the data replication, as the author just used one possible definition of markets. Perhaps OS talked about unit sales, either of songs or albums, rather than revenue. Or they talked about MSAs, rather than international markets.

To summarize, some replications are more informative than others. Overall, the results in the replications are different from those reported by OS, but not necessarily significantly so. This paper concludes that “[n]one of the economic conclusions from OS’s quasi-experiments hold up under replication.” This assessment seems unnecessarily harsh.

Other comments:

For the genre specific analysis, is the dependent variable measured as sales changes in absolute terms, or in percentage terms? More information on the regression equation would be helpful even if it is the same as in OS.

On page 12, the quote suggests a typo in OS. I don’t see that typo.

Footnote 16 refers to Liebowitz (2008). This paper is missing in the references.