Gold Sounds: An Empirical Analysis of Popular Music From the 1960s to Today

By: Ian Hueston Leon
Advisor: Dr. Moses Luri

Department of Economics,
The College of Wooster

Background

The recorded music industry is a rather young one, yet it has evolved into something that brings in billions of dollars per year for artists and labels alike. Recording and distribution technology have been the primary influencers of the industry, with the primary means of music consumption changing from LPs to CDs to internet streaming over the course of a little over 50 years. In this current age of streaming, it is increasingly difficult for artists to get a fair share of the profits.. Past literature has suggested that musical popularity can be predicted using musical elements such as tempo, mode, and instrumentalness. The present study aims to answer the question of if musical popularity can be predicted using a set of musical elements, and if popular elements vary across time.

Theory

Evolutions in musical production and distribution have greatly altered the way in which both those who consume and those who create music interact with the market. The inventions of the internet and the personal computer have caused the marginal cost of both production and distribution to fall near zero, thus shifting music's supply curve far out to the right, lowering the equilibrium cost and increasing the equilibrium quantity. Where a consumer needed to pay $17 for a CD 30 years ago, they can now spend $10 a month to listen to nearly anything that has ever been recorded, increasing the utility a rational consumer derives from music consumption. Label profit maximization has also been impacted by these changes. Their optimal capital and labor decisions have increased throughout time, as lower costs in distribution and consumption allow them to invest in a greater range of talent.

Data

Song popularity data was collected from Billboard.com. The Billboard Hot 100 chart is the industry standard for musical popularity and all songs that appeared on the chart from 1965 to 2021 were taken into consideration for the final dataset. Song data was collected from the Spotify.com API. The information about each song includes the name, the artist's name, the release date, and measures of various musical elements. These include measures of instrumentalness, acousticness, danceability, tempo, and the presence of explicit content among other variables. The final dataset includes 278,466 unique songs with complete observations for each one. Songs that appeared on the Hot 100 were cross referenced with the larger dataset and were recorded as having made the chart.

Methods

The empirical method of choice for this study was logistic regression. Logistic regression takes a binary response variable and creates predictions of the likelihood of the response variable either happening or not happening, in this case songs either making or not making the Hot 100 based on their era of release. Three regressions were run, one for each era of release. These eras were defined as 1965-1984 (Vinyl Era), 1985-2004 (CD Era), and 2005-2021 (Streaming Era). The regression equation is as follows:

$$logit(hot100) = \beta_0 + \beta_1 songDurationSec_i + B_2 danceability_i + B_3 energy_i + B_4 loudness_i + B_5 speechiness_i + B_6 acousticness_i + B_7 instrumentalness_i + B_8 liveness_i + B_9 tempo_i + B_{10} valence_i + B_{11} timeSignature_i + B_{12} key_i + B_{13} mode_i + B_{14} explicit_i$$

Post estimation tests were run to ensure that each of the individual estimations met the assumptions of logistic regression. Without ensuring that these assumptions are met, one cannot make valid statistical conclusions from a model.

Conclusion

The results found in my empirical analysis reflect significance in musical elements when examining the likelihood of a song being popular. The analysis also reflects significance in there being different preferences across generations regarding these musical elements and suggest that these preferences will continue to evolve moving into the future. Based on the findings of the empirical analysis, there are potential recommendations for both artists and firms who may be looking to increase their commercial success by signing artists that specialize in making music which is found to be popular in the modern era or taking existing artists and suggesting that they should follow consumer trends in what elements are popular at a given point in time. One of the elements that I expect will become increasingly negative for song popularity moving into the future will be duration, shorter songs have become increasingly popular throughout the 2020s, largely due to the influence of Tik Tok and other social media. As musical elements are significant predictors of popularity, artists and labels could attempt to forecast elements that will become popular in the near future. I believe that there is a lot of room for improvement in regard to my questions. A large group with better coding knowledge than mine could potentially scrape websites such as Wikipedia and last.fm to collect information that would be very difficult to record by hand, such as artist label affiliation, to add to their estimations.

Results

Release Era:	1965-1984	1985-2004	2005-2021
	(odds ratio)	(odds ratio)	(odds ratio)
hot100			
songDurationSec	1.000	1.001***	0.998***
	(0.000167)	(9.75e-05)	(0.000328)
danceability	1.016***	1.018***	0.992***
	(0.00195)	(0.00166)	(0.00160)
energy	1.016***	0.999	0.977***
	(0.00189)	(0.00167)	(0.00195)
loudness	0.936***	0.990	1.172***
	(0.00711)	(0.00781)	(0.0157)
speechiness	0.925***	0.974***	0.984***
	(0.00815)	(0.00325)	(0.00285)
acousticness	0.985***	0.980***	0.980***
	(0.000884)	(0.00101)	(0.00119)
instrumentalness	0.985***	0.986***	0.968***
	(0.00155)	(0.00164)	(0.00371)
liveness	0.994***	0.996***	0.992***
	(0.00147)	(0.00130)	(0.00149)
tempo	1.000	0.998**	1.001
	(0.000962)	(0.000829)	(0.000802)
valence	0.999	0.995***	0.996***
	(0.00125)	(0.00101)	(0.00104)
3.timeSignature	0.711	1.660	1.055
	(0.229)	(0.850)	(0.494)
4.timeSignature	1.009	2.317*	1.674
	(0.312)	(1.163)	(0.759)
5.timeSignature	0.449	0.844	1.457
	(0.244)	(0.569)	(0.729)
1.key	1.355***	1.578***	1.072
	(0.155)	(0.148)	(0.103)
2.key	0.936	0.829**	0.834*
	(0.0820)	(0.0769)	(0.0852)
3.key	1.092	1.742***	0.984
	(0.164)	(0.231)	(0.153)
4.key	1.081	1.160	1.239**
	(0.106)	(0.111)	(0.129)
5.key	1.090	1.419***	1.189*
	(0.105)	(0.132)	(0.124)
6.key	1.147	1.690***	1.125
	(0.150)	(0.170)	(0.121)
7.key	0.901	0.968	1.006
	(0.0795)	(0.0835)	(0.0955)
8.key	1.306**	1.327***	1.115
	(0.157)	(0.145)	(0.118)
9.key	0.893	1.034	0.937
	(0.0799)	(0.0909)	(0.0951)
10.key	1.287**	1.674***	1.279**
	(0.141)	(0.163)	(0.139)
11.key	1.151	1.282**	1.161
	(0.130)	(0.127)	(0.119)
1.mode	1.365***	1.294***	1.569***
	(0.0751)	(0.0598)	(0.0789)
1.explicit	0.350*	2.261***	2.574***
	(0.205)	(0.178)	(0.151)
Constant	0.00930***	0.00777***	0.597
	(0.00357)	(0.00428)	(0.318)
Observations	56,104	105,771	116,591

Robust seeform in parentheses
*** p<0.01, ** p<0.05, * p<0.1