(Not) Working 9 to 5: Using Music Characteristic and Song Lyric Data to Predict Initial Claims

Jacob Scott*

May 19, 2020

Abstract

Applying a Lasso modeling strategy to music characteristic data from Spotify and lyric information from Genius, I explore whether the characteristics and sentiment of popular music can offer insight into economic conditions. I find that the features of the top songs on Spotify for a given week can predict initial claims numbers with accuracy comparable to that of professional forecasters, and can do so more than 5 days prior to the release of official numbers. This provides evidence that data on message-consumption is connected to economic conditions and can provide more frequent estimates of indicators such as initial claims, which would be valuable to policymakers, investors, and consumers alike.

JEL Codes: C55, E24, J60

*Email:jscott@colgate.edu

1 Introduction

Economic indicators such as unemployment, GDP growth, and inflation are of keen interest to policymakers, the private sector, and the public at large. Unfortunately, due to the nature of data collection, such numbers come out at intervals that can leave economic actors in the dark for weeks at a time.

Given the proliferation of real-time data, in tandem with the advancement of machine learning techniques, it may be possible to derive more frequent estimates of these indicators. Such estimates would be immensely helpful when formulating policy, when making investment decisions, and when making other important economic choices. Thus, it is worth exploring the feasibility and effectiveness of using novel methods to generate near real-time estimates of economic indicators.

In this paper, I use a unique source, music, to generate predictions of one such indicator, initial claims. As I will discuss in greater detail below, the theoretical mechanism that would make such a method work is sentiment. The music that individuals choose to consume, in theory, reveals their sentiment. Consumer sentiment, of course, is directly related to economic indicators. Thus, the sentiment revealed by music may also tell us something about economic conditions.

Using music characteristic data from Spotify and initial claims data from the Department of Labor between January 2017 and February 2020, I find that between 50 to 60 percent of out-of-sample variation in initial claims data can be explained by music characteristics alone. This is not to say that music *causes* variation in initial claims, nor is it to say music is the only way to explain the variation. It does, however, suggest that frequently released and publicly available music data can offer insights into economic indicators prior to their official release.

The rest of the paper is structured as follows. In Section II I explore the theory and literature behind musical sentiment and its connection to economic data. In Section III I explain my data. In Section IV I explain my modeling strategy. In Section V I explore the results and their implications. In Section VI I briefly conclude.

2 Theory and Literature

I am certainly not the first to use sentiment analysis in an economic context. Bollen et al. (2011) find a relationship between the sentiment of Tweets and the performance of the stock market, Gilbert and Karahalios (2010) find that the sentiment of posts on the site LiveJournal provides novel information about future stock prices, and Liu et al. (2007) find that the sentiment of blog posts can be used to predict movie ticket sales.

However, there is a drawback to using social media data for the purpose of estimating sentiment. Specifically, there is a selection effect; individuals non-randomly decide what to share, which may not be reflective of their underlying mood state. For example, people may post based on what is normal for their social circle rather than based on their own personal feelings.

Music is different. While people listen to music for a variety of reasons, including social connection and to pass the time (Greasley and Lamont, 2006), a major determinant of music choice is "mood management." Broadly speaking, mood management is the theory that individuals choose to consume messages that help them regulate their mood (Zillmann, 1988). A major category of "message" that individuals choose to consume based on their emotional state is music (Konečni, 2010; Sloboda, 1999). Particularly relevant for this study is that individuals have been shown to use music to help regulate negative emotional states (Van den Tol and Edwards, 2013; Lonsdale and North, 2011; Thayer et al., 1994). Thus, in aggregate, the overall content of music consumed should differ in times of relatively greater distress compared to quieter periods.

The direction of this effect, in terms of how music will change, is not clear *a priori*. On the one hand, there is evidence that individuals will consume happier music when they are upset, in order to improve their mood (Lonsdale and North, 2011). On the other hand, there is also evidence that individuals will use music to cathart and release their emotions (Sloboda, 1999). Regardless of the direction, there is a strong theoretical basis for assuming that general sentiment can be proxied through music choice. General sentiment, in turn, responds to economic conditions, especially changes in unemployment (Mandal et al., 2013; Stanca, 2010; Oswald, 1997; Mueller, 1966). Thus, the content of popular music should change along with economic conditions due to the mediating variable of sentiment.

There is evidence that this theory holds empirically. Pettijohn and Sacco Jr (2009) find that the lyrical content of *Billboard* songs from 1955 to 2003 correlates with the General Hard Times Measure (GHTM), a "standardized, global measure of social and economic threat." Specifically, they find that during more threatening social and economic periods, the lyrics of top songs become more meaningful, comforting, and romantic, and include more words per sentence, future references, and coverage of social issues (p. 307). Maymin (2012) finds that the beat variance of songs on the U.S. Billboard Top 100 list from 1958 to 2007 predicts market volatility and returns of the S&P 500, which he hypothesizes is due to individuals preferring simple music in more difficult times. And both Sabouni (2018) and Kaivanto and Zhang (2019) use the Billboard Hot 100 and Official Charts Company Top 100 lists to generate sentiment models that both track traditional consumer sentiment measures and that outperform traditional "buy and hold" stock investment strategies.

There are, however, drawbacks to these analyses. The most relevant omission in this context is that none of these studies use music data to predict an infrequently released economic indicator. Indeed, there have been no prior studies that have used music analysis to generate frequent estimates of a measure like initial claims.

There are also more specific drawbacks to the otherwise compelling analyses by Sabouni (2018) and Kaivanto and Zhang (2019), whose works most closely approximate what I do below. Two issues relate to their data sources themselves. First, the songs that make it onto the Billboard Hot 100 and Official Charts Company Top 100 lists are persistent. Across the 16 years covered in their data, only around 7,000 unique songs appear on the former and 9,000 appear on the latter. This persistence reduces the period-to-period variation in their sentiment measures. Second, Billboard and the Official Charts Company track the popularity of a given song using not only intentional consumption, such as streams or downloads, but also through radio plays, which reflect factors such as marketing efforts rather than the direct preferences of consumers themselves.

An additional issue relates to the aggregation method chosen by Sabouni (2018) and Kaivanto and Zhang (2019). Specifically, they take the simple unweighted monthly average of their sentiment variables. This both obscures substantial within and across month variation in these values and also ignores the relative popularity of songs. Despite by definition being more popular, the top song in a given month receives the same weight as the bottom song in their analyses, which almost certainly hides important nuances.

Finally, while the above authors have data that spans over a decade and a half, they nonetheless have relatively restricted information with which to make predictions. First, they use fewer than 20 song and lyric variables to build their models. Second, they are only able to obtain lyric information for around 70% of the songs in their data. This is certainly a respectable and useful success rate. However, those 30% of songs for which they capture no lyric information could hold important insights.

As I will discuss in greater detail below, my analysis overcomes many of these challenges. I have around 1,500 unique songs per year compared to their ˜550, popularity in my data is measured strictly by number of streams rather than both direct (streams) and indirect (radio) measures, and I aggregate using stream-weighted mean, median, and quantile values rather than just the song-weighted mean. Further, I am able to obtain more comprehensive data, with over 200 variables with which to build my model and lyric information for over 98% of songs in my data.

Regardless of the drawbacks in previous works, this body of evidence nonetheless constitutes the theoretical justification for using the characteristics of popular music to predict initial claims. Economic conditions affect the mood of individuals, individuals regulate their mood using music, and thus the content of popular music should change along with economic conditions. Below, I explain my data in more detail.

3 Data

3.1 Music Data

I obtain music characteristic data from the popular music streaming service Spotify. Each day, stretching back to 2017, Spotify has released the top 200 songs on their platform by number of streams, broken down by country. Despite non-negligible persistence in which songs appear on the list, there are nonetheless 4,558 unique songs that have broken the top 200 list since 2017. Given that there are over 270 million active monthly users on Spotify globally, and almost 70 million in North America alone, these songs are reasonably representative of popular music at any given time (Spotify, 2019). And while the Spotify user base skews younger, wealthier, and more urban (Kats, 2018), social contagion theory suggests that even if Spotify users at large are not the ones being directly effected by a given economic shock, they may nonetheless experience comparable emotional effects (Kramer et al., 2014).

Using this data, I obtain the title, artist, number of streams, and album name for the top 200 songs in the United States for each day between January 2017 and February 2020. From this I collect detailed data on song features and lyrics. I discuss each below.

3.1.1 Song Features

Through its developer API, Spotify provides detailed information about every song on its platform, as well as the artist who created it and the album it is found on. Song information includes aspects such as danceability, instrumentalness, and valence. A full description of these song features is presented in Table 1. The artist and album information is slightly less detailed. For the former I obtain number of followers and popularity, and for the latter I obtain the album type (single or full album), genre, release date, popularity, and whether it is explicit.

Feature	**Description**
Acousticness	A confidence measure of whether the track is acoustic.
Danceability	How suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
Duration	The duration of the track in milliseconds.
Energy	Perceptual measure of intensity and activity.
Instrumentalness	Predicts whether a track contains no vocals.
Key	The estimated overall key of the track.
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
Loudness	The overall loudness of a track in decibels (dB).
Mode	Major or minor.
Speechiness	Detects the presence of spoken words in a track.
Tempo	The overall estimated tempo of a track in beats per minute (BPM).
Time Signature	Estimated meter.
Valence	Musical positiveness conveyed by a track.

Table 1: Description of song features. Obtained using Spotify API (Spotify, 2020).

Despite the richness of the available song data, there is an aggregation challenge in using it to predict a weekly indicator like initial claims. Each day has 200 distinct songs, and each week has up to 1,400. Each

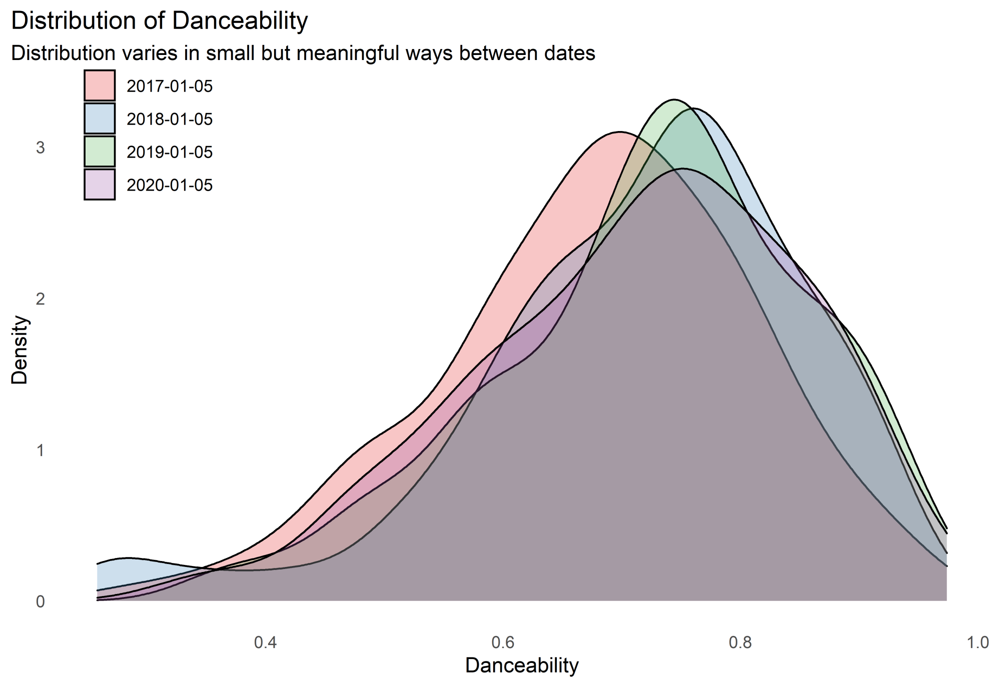

Figure 1: Distribution of danceability values for 4 different dates. Song characteristic data obtained using Spotify API (Spotify, 2020).

of these unique songs has its own combination of the above-described variables. The challenge, then, is how to aggregate into a single set of numbers the features of these 1,400 separate songs. The naive approach would be to simply take the mean, which would obscure substantial variation in the data. For example, even if some weeks have the same mean danceability, they may have substantially different distributions (Figure 1). The ideal method would be to somehow capture this full distribution for each variable for each week. Unfortunately this is not feasible. Instead, I approximate the distribution by taking the mean, median, and 0^{th}, 25^{th}, 75^{th}, and 100^{th} percentile values, all weighted by number of streams, for each variable. While necessarily imperfect, this captures at least some of the variation in numeric variables in a given week. Unfortunately, there is less that can be done for indicator variables, and I am only able to capture the stream-weighted percentage of positive values in a given week. So, for example, I can only capture the percentage of songs, weighted by stream, that are explicit in a given week.

While not the focus of this analysis, visualizing these variables over time, as done in Figures 2-5, reveals both considerable between-week variability and interesting longer term trends. For example, there is a considerable fall in the length of songs since 2017 (Figure 2) and a u-shaped relationship between time and

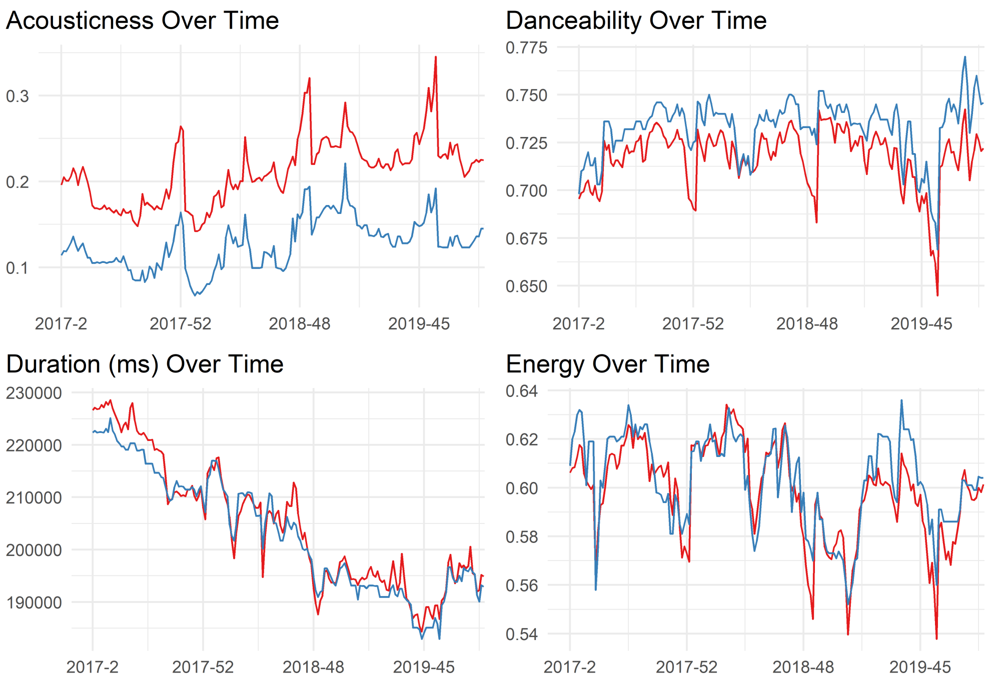

Figure 2: Values of accousticness, danceability, duration, and energy over time. Red indicates mean, blue indicates median. Song characteristic data obtained using Spotify API (Spotify, 2020).

the valence of songs (Figure 4). The week-to-week variation is a promising sign for the predictive power of a model built with this data, and the longer term trends are interesting in their own right.

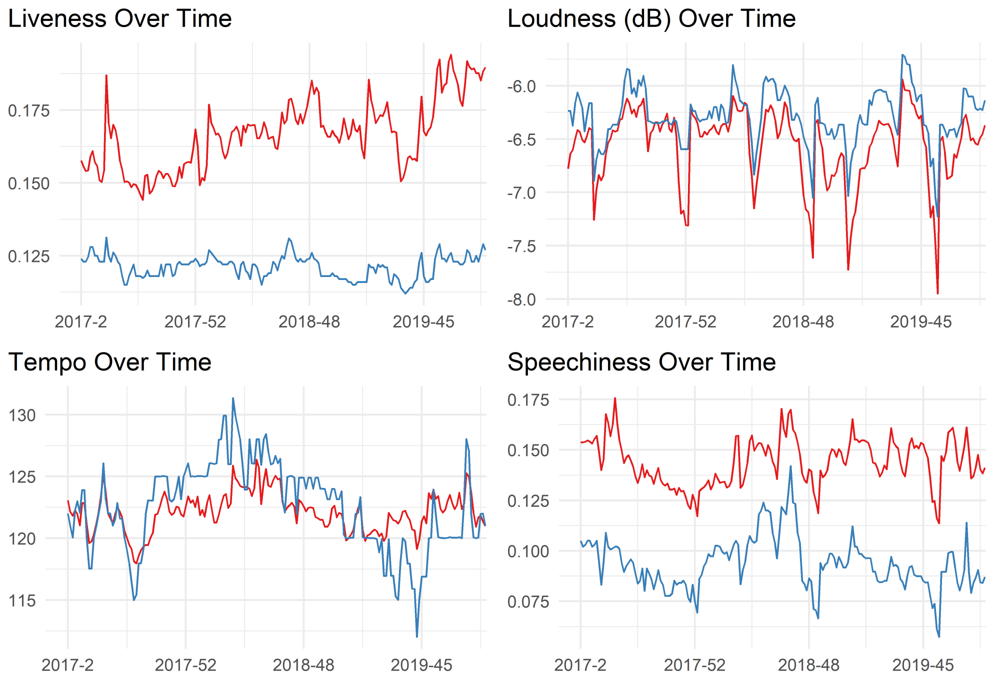

Figure 3: Values of liveness, loudness, tempo, and speechiness over time. Red indicates mean, blue indicates median. Song characteristic data obtained using Spotify API (Spotify, 2020).

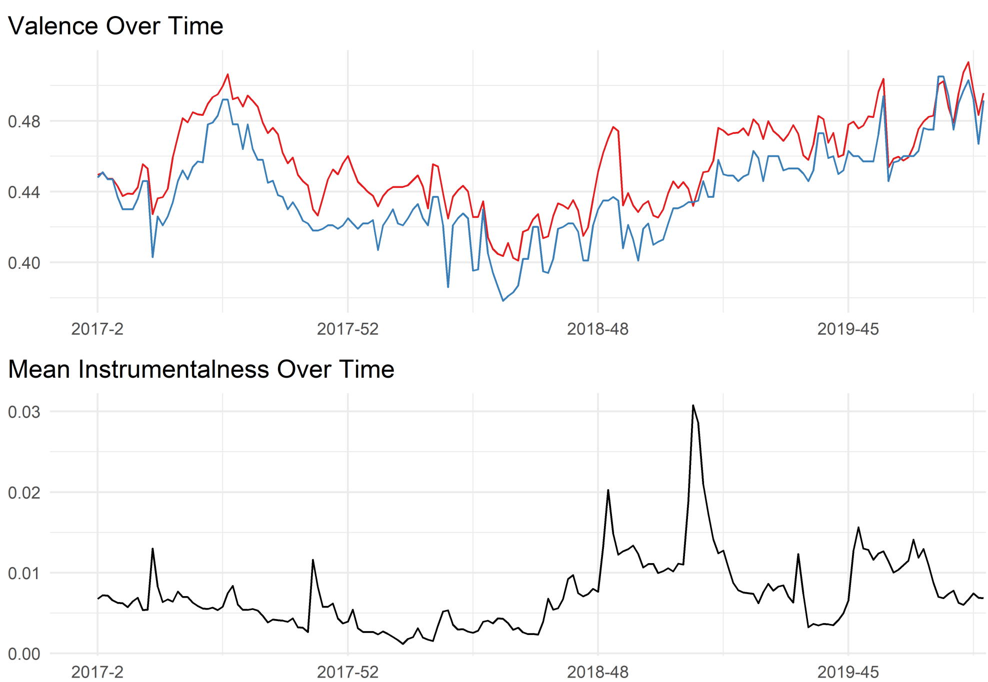

Figure 4: Values of valence and instrumentalness over time. For valence, red indicates mean and blue indicates median. Song characteristic data obtained using Spotify API (Spotify, 2020).

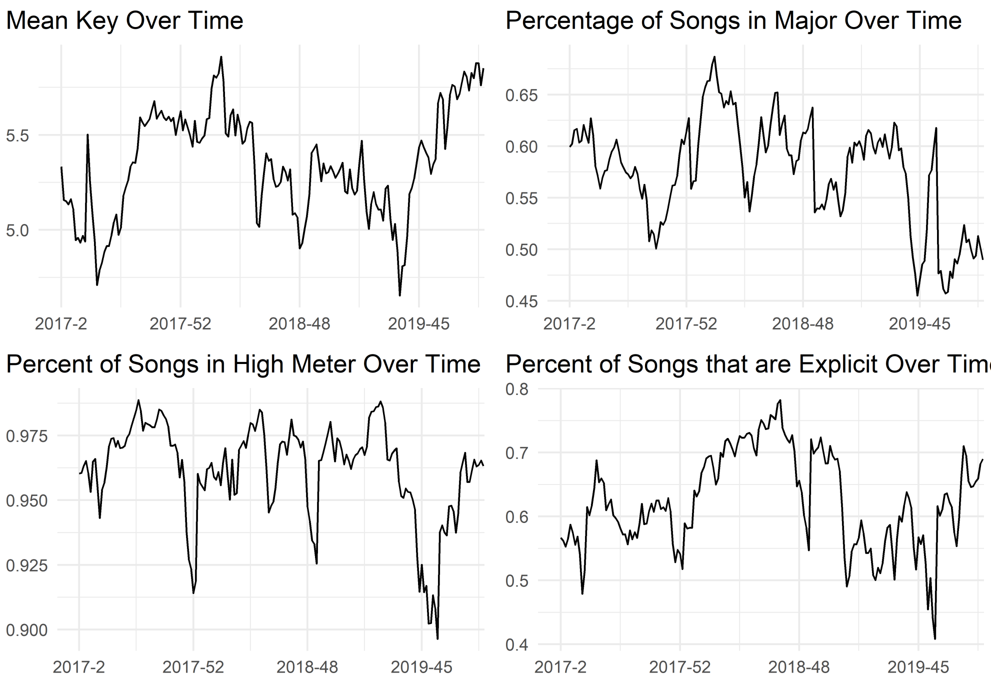

Figure 5: Song feature data over time. Clockwise: mean key, percent of songs in major, percent of songs in high meter, and percent of songs that are explicit. Song characteristic data obtained using Spotify API (Spotify, 2020).

3.1.2 Lyrics

The other type of data I use in this analysis are song lyrics. Using the API of the digital media company Genius, I obtain full lyrics for all but fewer than 75 of the 4,588 unique songs that have appeared on the top 200 list since 2017 (Genius, 2020). The songs for which I was unable to obtain lyric information represent less than 1% of the total data from 2017 to 2020. Thus, the missing lyric information is unlikely to have a meaningful effect on the overall analysis.

From this lyric data I extract sentiment information using the NRC Word-Emotion Association Lexicon, AKA EmoLex. EmoLex is a crowd-sourced lexicon that associates over 14,000 English words with eight base emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust, as well as two general sentiments: negative and positive (Mohammad and Turney, 2013). The system is illustrated for a selection of words in Table 2.

For each song, I count the number of words that fall into each emotion and sentiment category. From these song-level counts I calculate the overall percentage and the mean number of words per song that fall into each category each week. So, for example, I generate the mean number of words per song, as well as

	Trust	Fear	Negative	Sadness	Anger	Surprise	Positive	Disgust	Joy	Anticipation
Abacus	1	0	0	0	0	0	0	0	0	0
Abandon	0	1	1	1	0	0	0	0	0	0
Abandoned	0	1	1	1	1	0	0	0	0	0
Abandonment	0	1	1	1	1	1	0	0	0	0
Abba	0	0	0	0	0	0	1	0	0	0
Abbot	1	0	0	0	0	0	0	0	0	0

Table 2: Selection of words and their associated emotions and sentiments from the NRC Word-Emotion Association Lexicon, AKA EmoLex (Mohammad and Turney, 2013).

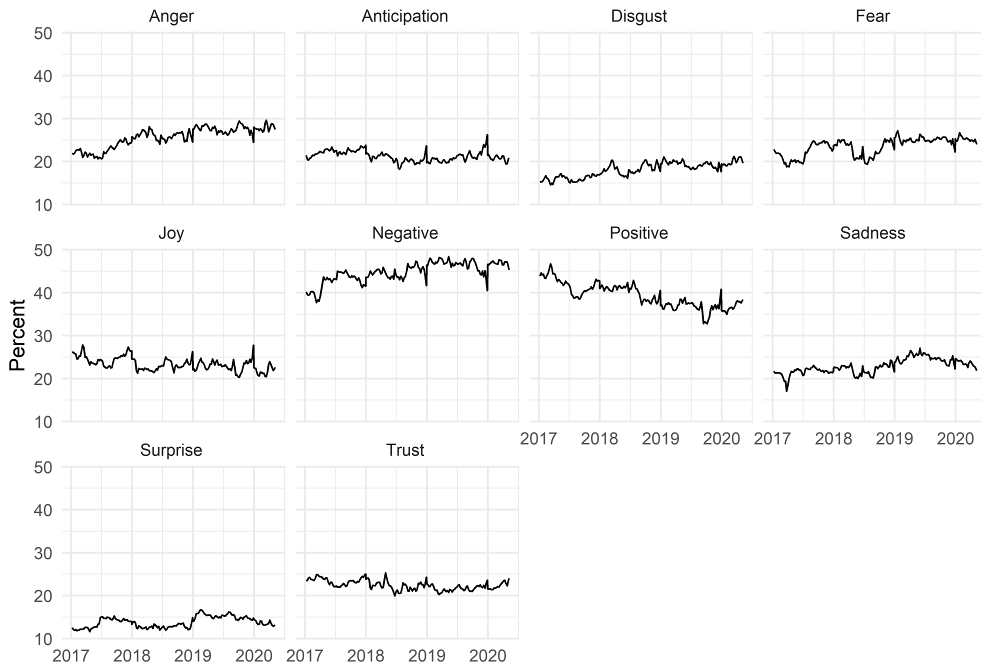

Figure 6: Stream-weighted percent of words that fall into different emotion and sentiment categories over time. Percentages do not sum to 100 since a given word can fall into multiple categories. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013).

the overall percentage of words that fall into the “anger” category, as defined by EmoLex, for a given week. Both of these values are weighted by number of streams, so that more popular songs receive more weight than less popular ones. Unfortunately, given the staggering number of unique words used in songs, even the expansive EmoLex lexicon only scratches the surface. Nonetheless, there are a sufficient number of matching words to generate meaningful numbers. For example, Figure 6 shows the change in percentage of total words that fall into each category over time. Once again, there is considerable variation between weeks, as well as some interesting longer term trends, such as a consistent decrease in the percentage of words that fall into the positive category.

In addition, as done by Kaivanto and Zhang (2019) and originally suggested by Plutchik (2001), I con-

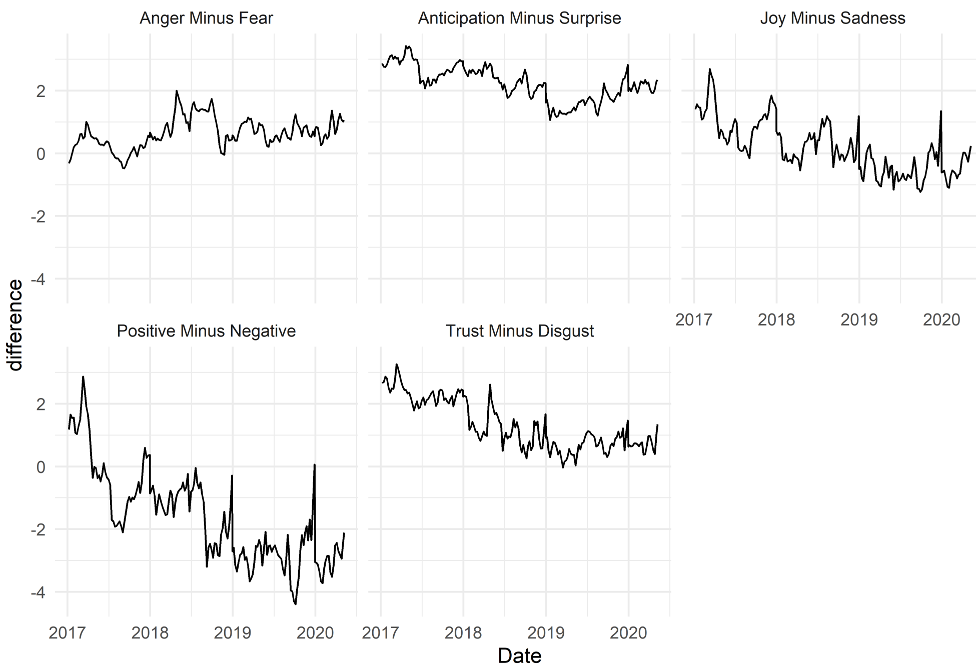

Figure 7: Difference in stream-weighted mean number of words per song in each category over time. For example, the difference between the mean number of "anger" words and the mean number of "fear" words per song over time. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013).

struct emotion category pairs, taking the difference between mean number of words that fall into each of the following pairings: joy and sadness, anger and fear, anticipation and surprise, trust and disgust, and positive and negative. A summary of these differences can be found in Figure 7. Once again, there is considerable variation between weeks, and longer term trends are apparent.

3.1.3 Misc.

In addition to the variables described above, it is reasonable to expect that the *change* in values could be predictive. For example, it may be that rather than the level of valence of songs in a given week, it is the change in valence from one week to the next that is related to changing economic conditions. Given this possibility, I also generate weekly change variables.

3.2 Initial Claims

The term "initial claims" refers to the total unemployment insurance claims filed by individuals in the United States. This metric is released on a weekly basis by the Department of Labor, and provides relatively frequent insight into the state of the labor market. However, while the official numbers cover the claims filed from Sunday to Saturday in a given week, they are not released until the following Thursday (Department of Labor, 2020). This means there is a 5 day reporting lag between the period being captured by the official numbers and the day that they are released.

Given that I have musical data only from 2017 to early 2020, the weekly nature of this indicator is useful in that it means there are sufficient observations for me to train a model. At the same time, the reporting lag between the period covered by the measure and the official release means that there is still value in generating more frequent estimates. These two features make initial claims an ideal indicator for me to demonstrate the possible usefulness of a music-based model. Thus, I use the most recently revised initial claims numbers as of February 12th 2020.

4 Modeling

The main challenge with using the data I have to predict initial claims is the so-called "curse of dimensionality." The curse of dimensionality describes a situation in which one has more features than outcomes. In my case, I have 234 variables but only around 160 weeks worth of initial claims data. Thus, to generate meaningful predictions I must employ a regularization technique. I elect to use Lasso.

Unfortunately, in the context of high dimensional data, even a Lasso regression can only generate one of many possible models. In other words, the model generated by Lasso will not be unique. Thus, rather than examining and assessing any one particular model generated by Lasso, I employ a bootstrapping strategy to obtain a broader overview. I do this by running 1000 distinct iterations of Lasso, using R's seed system to ensure that each is unique. For each of these models I extract an out-of-sample R-Squared value and coefficient information.

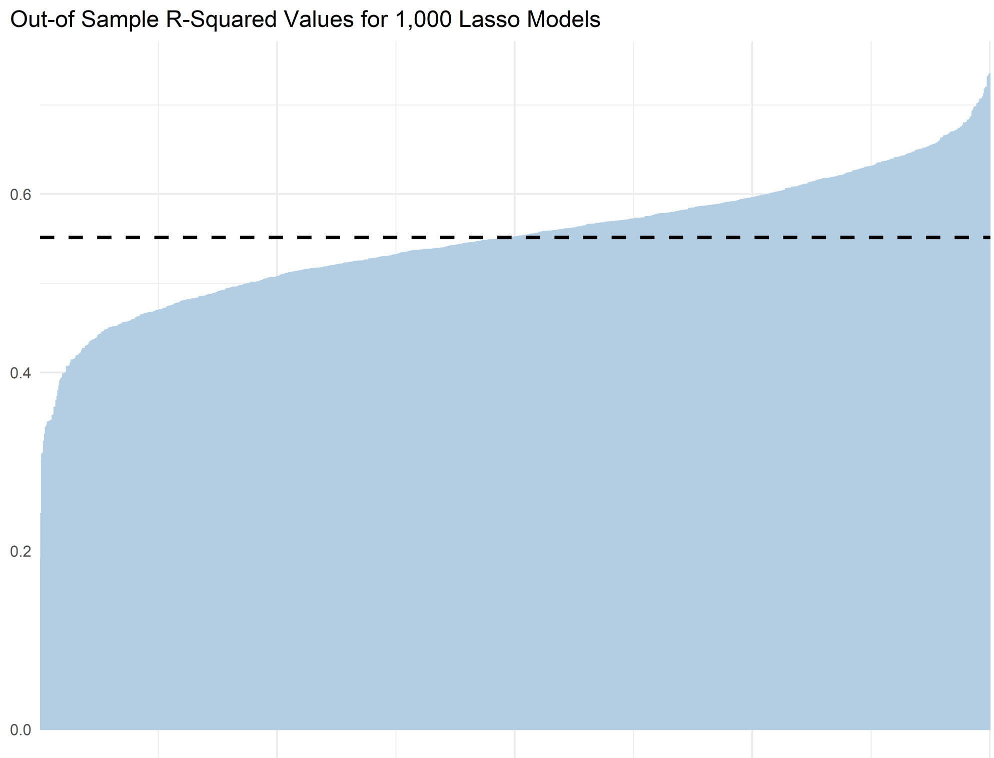

Figure 8: Out-of-sample R-Squared values for 1000 models that predict initial claims using music data. All models generated using Lasso. Arranged in ascending order by R-Squared values. Dashed line represents median R-Squared value. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

5 Results

5.1 R-Squared and Predictive Power

The median out-of-sample R-Squared value of these models is 0.55 ($\sigma = 0.072$), and the full distribution can be seen in Figure 8. As is readily apparent, most of the 1000 R-squared values fall within the 0.50 to 0.60 range, though there are some extremes on both the high and low end. This suggests that, on average, a model using my music data can explain around 55% of the variation in initial claims numbers in a sample that it was not trained on.

This predictive power can be seen in Figure 9. Not only do models generated with this data oftentimes successfully predict whether initial claims will increase or decrease from one week to the next, their point estimates are often quite accurate. This provides compelling evidence that there is in fact a connection between economic conditions (as measured by initial claims) and the characteristics of popular music, and

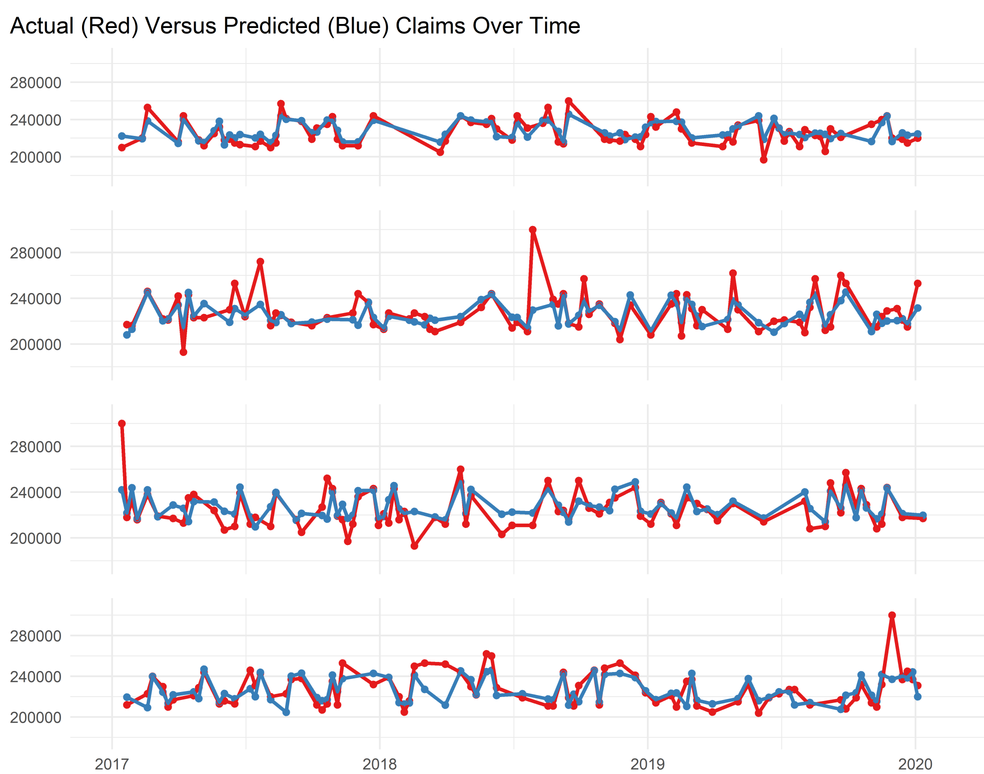

Figure 9: Predicted versus actual initial claims values for four random Lasso models. Red line represents actual initial claims. Blue line represents predicted initial claims. Predictions generated using out-of-sample test data. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

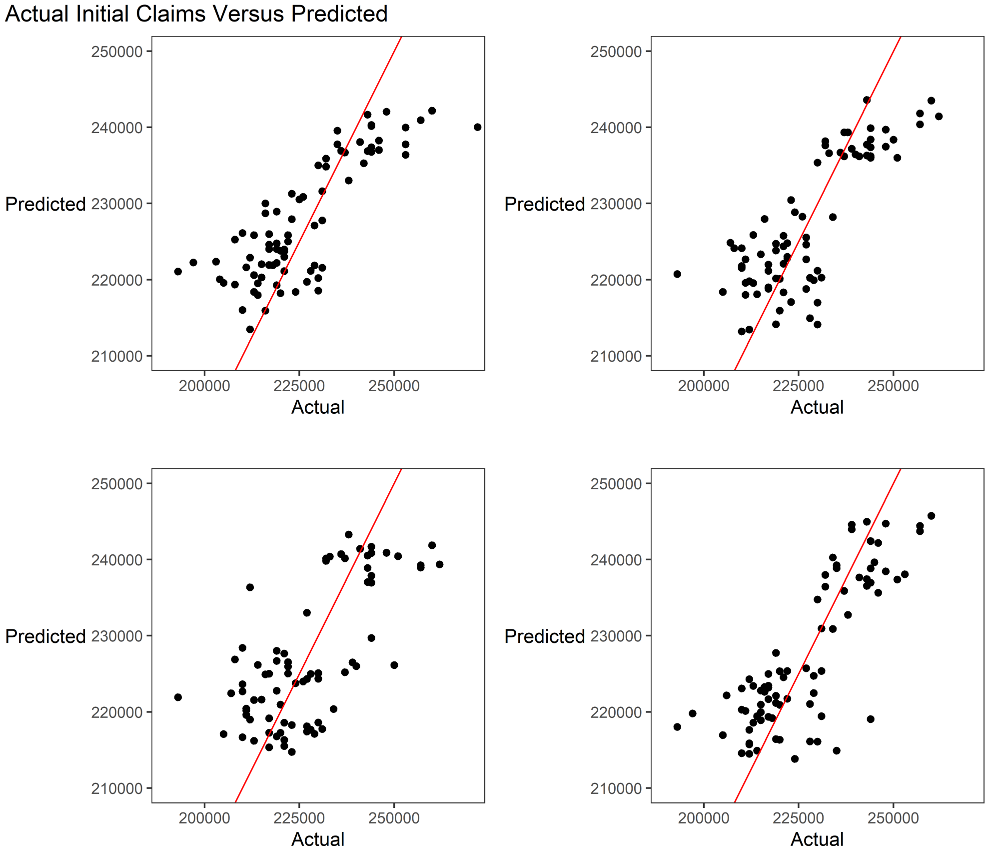

Figure 10: Predicted versus actual initial claims values for four random Lasso models. Red line indicates where predicted claims equal actual claims. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

that this connection is strong enough to generate meaningful predictions.

Despite this impressive predictiveness, the models are of course imperfect, and appear to be somewhat conservative. They generally underestimate initial claims when actual claims are relatively high and overestimate them when they are relatively low (Figure 10). In stable times such as the period between 2017 and early 2020 this poses little challenge. However, in more turbulent times characterized by drastic changes in initial claims, the models may become less accurate. This can be explored empirically as time goes on and more data becomes available.

5.2 Variable Importance

Along with capturing the R-Squared values and predictiveness of the model, I also measure the importance of each predictor in two ways. First, I use a bootstrapping method by running 1000 models and counting the number of appearances for each variable. While 183 of the 243 variables in my data appear in at least one model, Figure 10 demonstrates that there are a handful of variables that appear far more often than others. Indeed, over 95% of the variables appear in fewer than half of the models.

I also use a more conventional measure of variable importance. Using the *Caret* package in R, I assess variable importance by fitting a linear model to each feature and looking at the the absolute value of the t-statistic (Kuhn, 2020). While the exponential nature is less steep using this method, Figure 11 nonetheless shows that certain variables are far more useful in predicting initial claims than others.

To determine which of these variables are driving the models, I look at the top 10 in terms of number of appearances and variable importance, respectively. The results can be seen in Figure 12.

For the bootstrapping method, 6 out the top 10 most common variables used in the 1000 models are related to lyrical sentiment (and the 11th and 12th most common, joy mean and anticipation mean, are also sentiment-related). Similarly, 5 of the top 10 are related to lyrical sentiment when using the repeated linear model method of measuring importance, including the 3 most important. This suggests that the sentiment of popular songs is indeed correlated with economic conditions (as measured by initial claims).

Of the four other variables that appear in the top 10 for the bootstrapping method, danceability and the percent of songs in the major key seem to make sense. Individuals may be less inclined to listen to "dance" music and more inclined to listen to songs in the minor key (generally sadder) in economically difficult times. On the other hand, it is not immediately obvious why mean liveness and album popularity would be related to economic conditions.

For the 5 non-sentiment features in the top 10 using repeated linear models, those related to number of streams and duration of songs could be connected to economic conditions insofar as individuals listen not only to different types of music, but different amounts of music during more challenging times. Again, it is less obvious what connection there would be between economic conditions and mean liveness and album popularity.

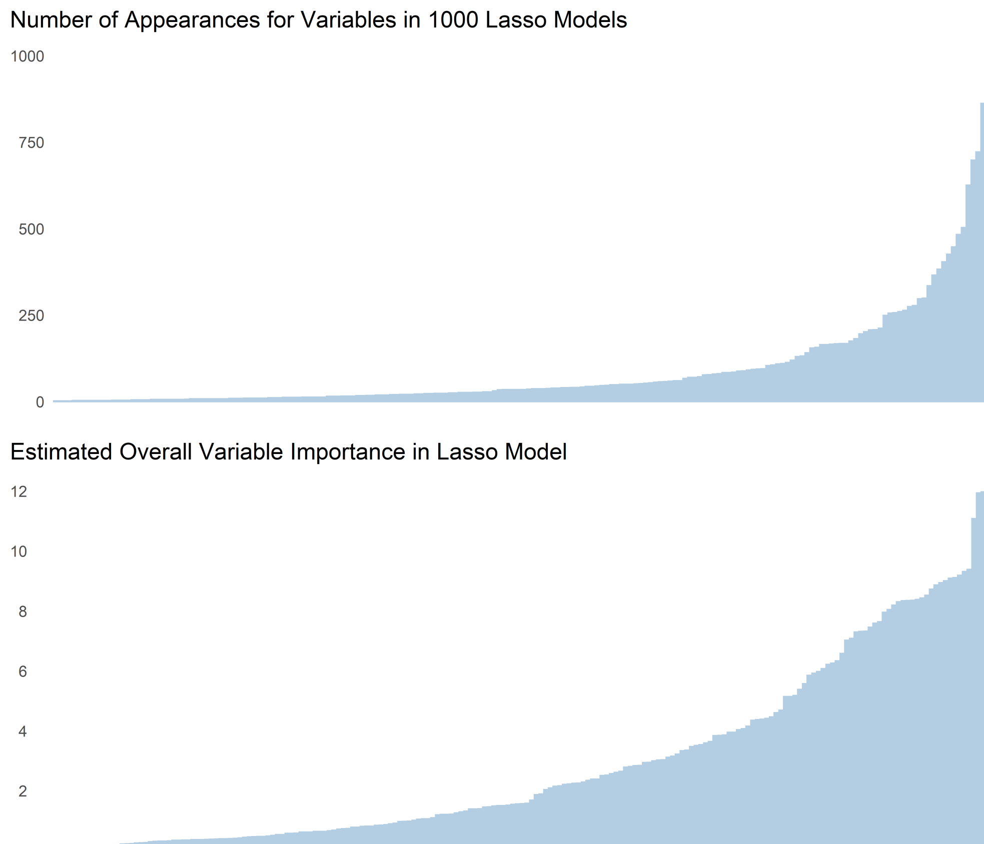

Figure 11: Variable importance estimated via bootstrapping (top) and repeated linear models (bottom) for 243 predictors. 1000 runs of Lasso aggregated and appearances for a given predictor summed for bootstrap method. Linear model fit on each predictor and absolute value of the t-statistic used for repeated linear models method (Kuhn, 2020). Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

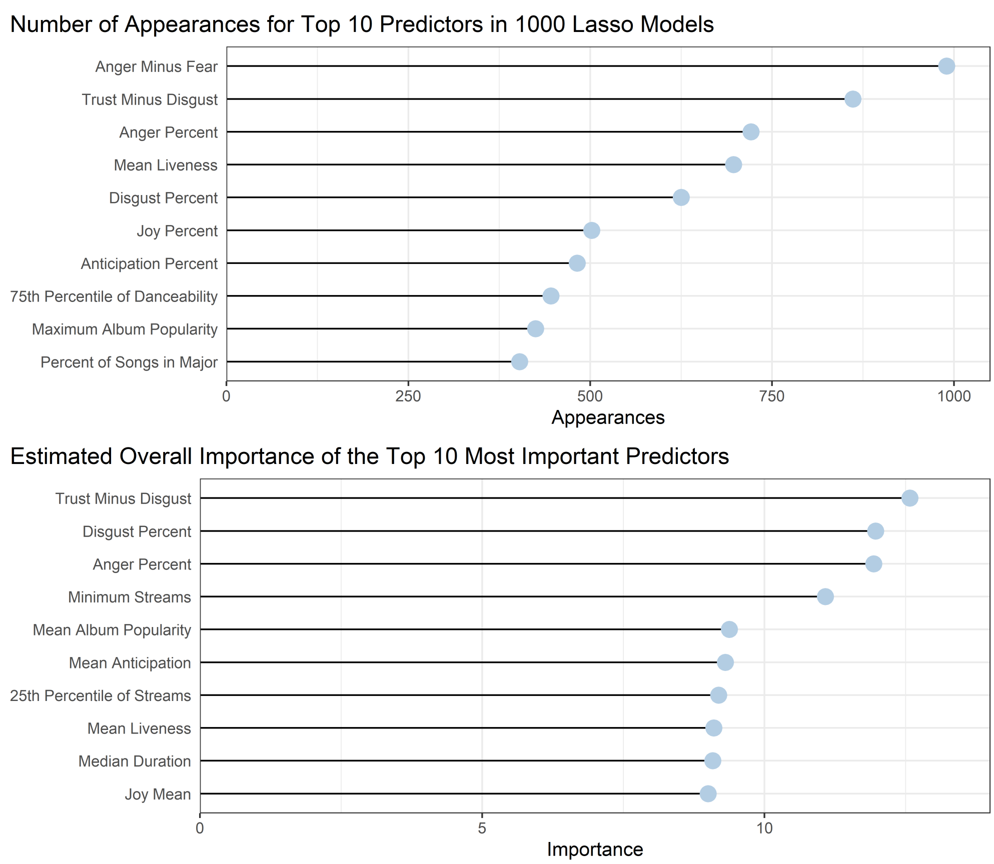

Figure 12: Variable importance estimated via bootstrapping (top) and repeated linear models (bottom) for top 10 predictors. 1000 runs of Lasso aggregated and appearances for a given predictor summed for bootstrap method. Linear model fit on each predictor and absolute value of the t-statistic used for repeated linear models method (Kuhn, 2020). Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

5.3 Coefficients

Beyond which variables are important, I also investigate the influence they actually have on the predictions generated by my model. To do this I take the median standardized coefficient across all 1000 models. Thus, the coefficient can be interpreted as, on average, the predicted change in initial claims given a one standard deviation change in a predictor. I do this both without weights, and weighting by R-Squared values to give the coefficients in more effective models more weight (Table 3).

Variable	Coefficient	Weighted Coefficient	Unweighted σ
Trust Minus Disgust	2566	2608	727
Joy Percent	898	843	876
Anticipation Percent	655	632	752
Percent of Songs in Major	-539	-498	717
75th Percentile of Danceability	-656	-639	712
Maximum Album Popularity	-772	-729	936
Mean Liveness	-954	-923	957
Anger Percent	-1215	-1223	1083
Disgust Percent	-1998	-2015	1516
Anger Minus Fear	-2926	-2989	1176

Table 3: Top 10 most common variables in 1000 Lasso models, sorted by coefficient size. Coefficients are the median coefficient across all of the 1000 models a given predictor appeared in. Weighted coefficients are weighted by the R-Squared value of the model a given observation was in, giving more weight to more effective models. All variables are standardized so that the coefficients can be interpreted as the predicted change in initial claims given a one standard deviation change in a given variable. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

While the coefficients appear sensible individually, together they appear to tell somewhat conflicting stories. For example, the coefficient on *Trust Minus Disgust* suggests that lyrics become more trusting when initial claims are higher, whereas the coefficient on *Anger Minus Fear* suggests they become more fearful. The coefficient on *Joy Percent* suggests that songs become more joyful, whereas the coefficients on *Danceability* and *Percent of Songs in Major* suggests they become less danceable and are more often in the minor key (generally associated with sadder songs).

This apparent incongruity disappears, however, when one considers that individuals use music both to change *and* to enhance their mood (Lonsdale and North, 2011; Sloboda, 1999). Given this duality, it should come as no surprise that the coefficients suggest that songs become more negative in some dimensions and more positive in others.

5.4 Pseudo Real-Time Predictions

The findings above are doubtlessly of theoretical interest. They provide evidence that economic conditions do indeed affect individuals' music choice, and that the connection is sufficiently strong to make meaningful predictions possible. However, the initial motivation behind building this model was practical: I wanted to see whether a music-based model could generate more frequent yet accurate estimates of an economic indicator like initial claims.

In one sense I have already done this. My models can take the music data for a given week as an input on Saturday and output the estimated initial claims for that period. This prediction would be available 5 days before official numbers are released, given the 5 day reporting lag described above.

However, my models can also generate even more frequent predictions by taking daily data as their input. For example, my models can be given data on a Wednesday and generate a prediction for what initial claims will be on the coming Saturday. These two methods demonstrate the "higher-frequency" feature of my model. The other feature necessary to make the model practical is accuracy.

To assess the accuracy of my model, I compare the weekly predictions and actual values for each of my 1000 models. Overall, I find that 95% of my predictions are within 1.3 standard deviations of the actual initial claims value, and 90% are within 1 standard deviation. Further, the models do not systematically over or underestimate actual claims, with the mean error close to zero (112). I do the same using the daily estimates, and find similar results: 95% of predictions are within 1.2 standard deviations and 90% are within 1.

In isolation it is difficult to gauge whether this is a reasonable level of accuracy. To put these findings in context, I benchmark the accuracy against the preliminary numbers released by the Department of Labor, consensus expectations, and Moody's Analytics estimates.

The preliminary estimates are the numbers that the DOL actually releases each given Thursday (ALFRED, 2020). These are estimates and are subject to revision over time. Comparing these preliminary estimates to the most recent revisions as of February 2020, I find that the median absolute error is 2000, compared to an out-of-sample median absolute error of 5940 for my models. Given that the DOL uses official data to generate preliminary numbers while my model simply uses music data, it comes as no surprise that the preliminary estimates are more accurate. But while the predictions from my model are somewhat less accurate, they can be generated up to 5 days before the preliminary estimates are released (and even earlier if using daily music data as the input, as explained above).

A fairer comparison is between my models and expert expectations, given that both involve predicting initial claims without official numbers. To make this comparison I collect the Bloomberg consensus expectations [1] and the Moody's Analytics predictions for each week in my data (Moody's Analytics, 2020; Bloomberg, 2020). These predictions have a median absolute error of 5250 and 6000 respectively. The out-of-sample absolute error of 5940 for my models falls between these numbers, suggesting that predictions generated using my music-based models are approximately as good at predicting initial claims as more traditional measures, and can do so with relative ease and on a daily basis.

5.5 Limitations

The accuracy of my models is certainly impressive and somewhat surprising. However, there are two limitations that raise concerns about the validity of the predictions and that could hinder their effectiveness in the future.

First, as I have discussed above, the models are quite conservative. They generally underestimate relatively high values and overestimate relatively low ones. This is especially apparent when comparing the daily estimates (which should, in theory, be noisier) to the actual weekly initial claims values (Figure 13).

This does not hinder the models in the timeframe used in this analysis. However, the timeframe, January 2017 to February 2020, is unusual in its level of stability. Actual initial claims do not fall below 192,000 or rise above 300,000 in the period, and the standard deviation is around 16,000 claims. This is historically unusual, as the full initial claims data spanning from January 1967 to February 2020 has a standard deviation of almost 90,000 (73,000 even excluding recessions). Thus, the models as they currently stand would likely perform relatively poorly in less stable periods.

A similar but more serious problem is that of trends over time. Both initial claims and some of my predictors have general trends across the sample. The reason this is problematic can be illustrated with an

[1] Consensus expectations refer to the median estimate of 50 economists surveyed by Bloomberg. Accessed via Bloomberg Terminal.

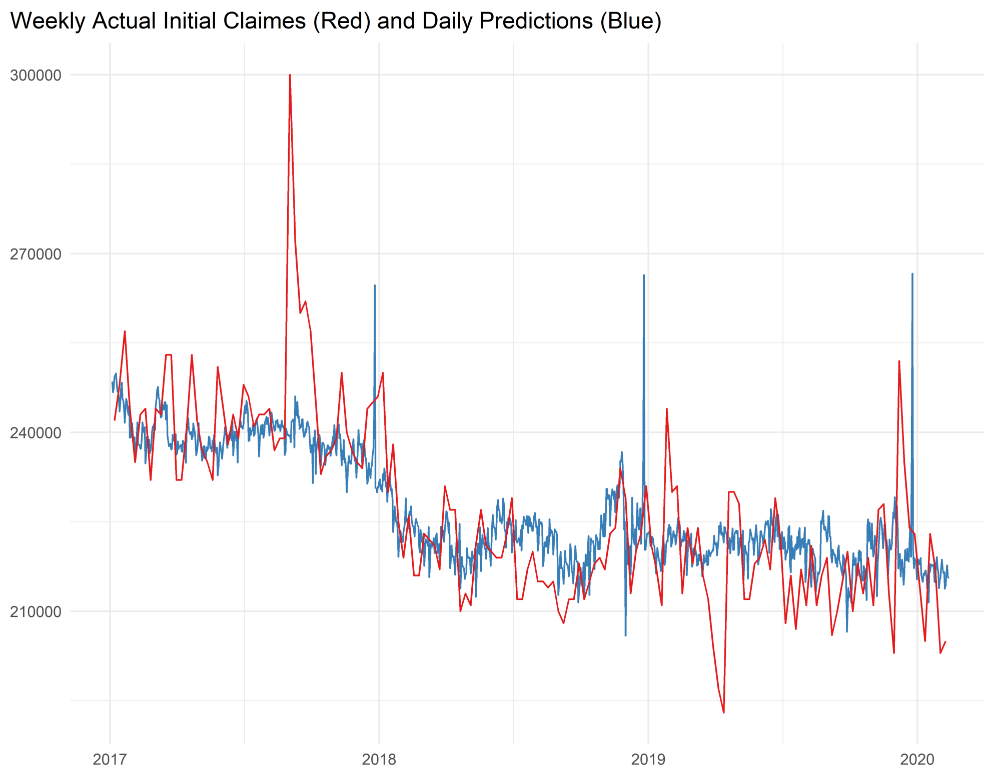

Figure 13: Actual weekly initial claims versus daily predicted values for median Lasso model (by R-Squared Value). Red line represents actual weekly initial claims. Blue line represents predicted initial claims on a daily basis. Song data obtained using Spotify API (Spotify, 2020). Lyric data obtained using Genius API (Genius, 2020). Emotion and sentiment categories obtained using NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). Initial claims data obtained from Department of Labor (Department of Labor, 2020).

example: given that the duration of songs and the number of initial claims both decline fairly consistently in this time frame, the model could improve its accuracy simply by estimating that claims will be lower when the value of song duration is lower [2]. While this happens to be true in this data, it is almost certainly spurious, and would likely reduce the accuracy of the model for predictions outside of this sample. Future models of this sort should account for such time-based issues.

There is a caveat however. These limitations apply only to the specific models generated here. Unless all of the predictiveness is driven by the relative stability of initial claims in this data, by potential spurious trend-based correlations, or by a mix of both, the strategy used in this analysis could still prove fruitful. For example, models trained on data with more variation in initial claims may overcome the challenge of overly conservative estimates, and models trained on a longer span of data, in which there are no general trends spanning the whole series, may similarly overcome spurious trend-based correlations.

6 Conclusion

The goal of this analysis was to assess whether music data could be used to create a model that generates frequent and reasonably accurate estimates of an economic indicator. Using music data from Spotify and lyric data from Genius, I demonstrate that one can indeed build a model that predicts initial claims with accuracy comparable to that of professional forecasters, at least over the period of January 2017 to February 2020. I also find that sentiment-based variables, such as the percent of words classified as angry, are important in driving the predictiveness of these models. This suggests that general sentiment acts as a mediating variable that connects economic conditions and music choice.

Despite the predictiveness of my models, caution must be exercised. The period covered by the sample is historically unusual in its level of stability (in terms of initial claims) and the short timeframe means that spurious correlations between time-trending variables could be driving some of the results. As time goes on, more data can be collected, and these limitations can be assessed empirically. Regardless of whether these limitations prove present, this analysis should encourage other researchers to explore using novel real-time data (especially message-consumption data, like music) to generate more frequent estimates of economic

[2] The repeated linear models method of assessing variable importance does suggest that duration is an important feature in my models, heightening this concern.

indicators.

References

ALFRED (2020). Vintage initial claims. https://alfred.stlouisfed.org/series?seid=ICSA.

Bloomberg (2020). Initial claims historical price table. Accessed via Bloomberg Terminal.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Department of Labor (2020). Unemployment insurance weekly claims news release. https://www.dol.gov/ui/data.pdf.

Genius (2020). Genius API documentation. *Genius*. https://docs.genius.com/.

Gilbert, E. and Karahalios, K. (2010). Widespread worry and the stock market. In *Fourth International AAAI Conference on Weblogs and Social Media*.

Greasley, A. E. and Lamont, A. M. (2006). Music preference in adulthood: Why do we like the music we do. In *Proceedings of the 9th international conference on music perception and cognition*, pages 960–966. Citeseer.

Kaivanto, K. and Zhang, P. (2019). Popular music, sentiment, and noise trading. Working paper, Lancaster University, Department of Economics.

Kats, R. (2018). Who is using Spotify? *eMarketer*. https://www.emarketer.com/content/who-is-using-spotify.

Konečni, V. J. (2010). The influence of affect on music choice. In Juslin, P. N. and Sloboda, J. A., editors, *Handbook of music and emotion: Theory,research, applications*. Oxford University Press.

Kramer, A. D., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790.

Kuhn, M. (2020). *caret: Classification and Regression Training*. R package version 6.0-86.

Liu, Y., Huang, X., An, A., and Yu, X. (2007). Arsa: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 607–614.

Lonsdale, A. J. and North, A. C. (2011). Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology*, 102(1):108–134.

Mandal, A., McCollum, J., et al. (2013). Consumer confidence and the unemployment rate in new york state: A panel study. *New York Economic Review*, 44(1):3–19.

Maymin, P. (2012). Music and the market: Song and stock volatility. *The North American Journal of Economics and Finance*, 23(1):70–85.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Moody's Analytics (2020). United States Jobless Claims. *Economic View Real Time*. https://www.economy.com/economicview/indicator/usa_claims.

Mueller, E. (1966). The impact of unemployment on consumer confidence. *Public Opinion Quarterly*, 30(1):19–32.

NBER (2020). NBER based recession indicators for the United States from the peak through the trough. https://fred.stlouisfed.org/series/USRECDM.

Oswald, A. J. (1997). Happiness and economic performance. *The economic journal*, 107(445):1815–1831.

Pettijohn, T. F. and Sacco Jr, D. F. (2009). The language of lyrics: An analysis of popular billboard songs across conditions of social and economic threat. *Journal of Language and Social Psychology*, 28(3):297–311.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Sabouni, H. (2018). *The Rhythm of Markets*. PhD thesis.

Sloboda, J. A. (1999). Everyday uses of music listening: A preliminary study. *Music, mind and science*, pages 354–369.

Spotify (2019). 2019 annual report. *Spotify Investors*. https://investors.spotify.com/financials/default.aspx.

Spotify (2020). Get audio features. *Spotify for Developers*. https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/.

Stanca, L. (2010). The geography of economics and happiness: Spatial patterns in the effects of economic conditions on well-being. *Social Indicators Research*, 99(1):115–133.

Thayer, R. E., Newman, J. R., and McClain, T. M. (1994). Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension. *Journal of personality and social psychology*, 67(5):910.

Van den Tol, A. J. and Edwards, J. (2013). Exploring a rationale for choosing to listen to sad music when feeling sad. *Psychology of Music*, 41(4):440–465.

Zillmann, D. (1988). Mood management through communication choices. *American Behavioral Scientist*, 31(3):327–340.