Dynamics of Musical Success:
A Bayesian Machine Learning Approach

Khaled Boughanmi[1]

Asim Ansari

Abstract

People consume music through albums and playlists. Consumer reactions and consequently the success of such musical assortments depends upon the felt experience of listeners. We develop a novel multimodal machine learning model for musical success that combines data of different modalities (e.g. metadata, acoustic and textual data) to predict album success. We estimate our model on a unique dataset which we collected using different online sources. Our model integrates different types of nonparametrics. It uses a supervised hierarchical Dirichlet process to represent crowd-sourced textual tags and penalized splines to capture the nonlinear impact of acoustic features. It captures dynamics via a state-space specification. We show the predictive superiority of our model with respect to several benchmark models. Our results illuminate the dynamics of musical success over the years and uncover themes that categorize albums on several aspects of the musical experience. We then use our model for forecasting the success of new albums, playlist construction, album fusion, as well as for contextual and other types of recommendations.

Keywords: Music Industry, Success Dynamics, Experiential Design, Product Recommendations, Big Data, Probabilistic Machine Learning, Bayesian Nonparametrics, Supervised Hierarchical Dirichlet Process.

Introduction

People often consume music in listening sessions involving multiple songs played in sequence. For much of musical history these sessions have been fueled by the availability of albums and playlists. Consumer reactions to music depend upon a number of factors. Of considerable importance is the felt experience of the listener which is determined by the physical aspects of music, such as pitch, loudness, or the tempo, as well as more subjective facets such as the energy, liveness, danceability, valence, speechiness, and acousticness of songs. Moreover, consumer reactions to assortments of songs such as albums or playlists depend upon the balance (Bradlow and Rao 2000) and the distribution of such acoustic characteristics across the songs within the bundle. From the production perspective, both albums and playlists are designed to be more than just collections of songs. They are designed around themes, balance various acoustic features, and contain songs that evoke specific moods and emotions. Predicting the success of such assortments is therefore difficult as one needs to capture the subjective experience of listeners that is evoked by the acoustic features, as well as accommodate how these are distributed across the songs within the collection.

To date, albums continue to be the primary format for launching popular music. According to Nielsen, more than 70,000 albums were released by mid-year 2018 (RIAA 2018) and Vogel (2014) reports that major labels alone introduce 11,000 albums each year. Less than 10% of these are profitable, however, and fewer than a 100 sell more than half-million units. Moreover, new music is costly to produce as an album can cost between $250,000 to $400,000. Musical consumption of late has been driven by streaming providers like Spotify, Pandora and more specialized ones like Quboz who construct and curate a variety of playlists to suit different tastes, generations, and contexts. Despite the obvious importance of the music industry and the unique nature of musical consumption, relatively little work has been done in the marketing literature to model and predict musical success.

In this paper, we develop a novel multimodal machine learning framework to predict the success of musical assortments. Success is modeled in terms of the underlying acoustical

features of the songs, the musical genres that are represented within the assortment, and user-generated textual tags that capture the thematic components and consumer perceptions of the song collection. We then use our model on yearly Billboard 200 data on album success for the last 50 years. The paper has three objectives. The first is to examine how acoustic features and tag-generated themes correlate with the success of popular music in America. The second is to examine how the impact of these features have evolved over the last half century. The third is to (1) forecast the performance of new albums (and playlists), (2) recommend albums that are consistent with the musical styles associated with specific eras, (3) design contextual playlists, (4) recommend musically similar albums and (5) provide diagnostics for albums predicted to be unsuccessful.

The dependent variable in the model is a (censored) score reflecting the success of musical albums released in 1963 through 2016. This score has been used by *Billboard* magazine to rank the 200 most successful albums each year. It combines data on sales of albums, track-equivalent albums, and streaming-equivalent albums (Pietroluongo 2012) that provides information about all major musical artists, genres, and historically significant albums produced over a period of fifty-four years. We supplement the Billboard data with information about albums recorded by some of the same artists that did not appear on the Billboard 200. We collected the acoustic features using the *Spotify* web API and the tags using *Last.fm* web API . The model accommodates dynamics in the coefficients of the genre and acoustical variables to capture their heterogeneous impact over time. It flexibly captures the possible nonlinear impact of the acoustical variables using nonparametric basis splines and it uses a supervised hierarchical Dirichlet process (SHDP) to summarize the semantic content within the tags via inferred latent album themes or topics. This results in a Bayesian semiparametric machine learning model which we then use for several marketing tasks involving album recommendation and playlist construction

We analyze album data because most new music has been and continues to be launched via albums.[1] Analyzing playlists is more difficult because these are relatively recent and it is

---

[1]The aforementioned Nielsen report notes that approximately forty percent of the music consumed is released in the preceding eighteen months.

difficult to find documentation for unsuccessful playlists. In contrast, album data dates back to at least the 1960s and are available for both successful and unsuccessful albums.[2] Even if we were only interested in playlist design, we could learn a great deal by studying the musical features that distinguish between successful and unsuccessful albums. Like albums, playlists are designed, and the good ones are balanced in their musical content and consistent in their themes.

On the methodological front, our model is novel in many respects. It integrates multiple nonparametric components that allow us to flexibly model the impact of covariates of different modalities. The use of nonparametrics is important as we do not have existing knowledge of the functional forms of these effects components. Another component uses the SHDP to infer themes that provide a mixed membership representation of the albums and are useful for predicting album success. The hierarchical Dirichlet process allows for albums to share a common set of themes and automatically infers the number of those themes. While state-space dynamics (Naik 2015) and Bayesian nonparametrics (Shively et al. 2000; Ansari and Mela 2003; Wedel and Zhang 2004; Ansari and Iyengar 2006; Kim et al. 2007; Li and Ansari 2013; Rossi 2014; Dew and Ansari 2018; Bruce 2019) have been previously used in marketing studies, our model applies these tools in several novel ways. It is the first to use a supervised hierarchical Dirichlet process. It also extends the machine learning and statistical literature in its integration of state-space dynamics and different types of Bayesian nonparametric approaches to simultaneously handle different data modalities.

Our research builds on the marketing literature on the consumption and success of music. Bradlow and Fader (2001), for example, used a Bayesian approach to model the movement of songs up and down the Billboard 200 chart and Lee et al. (2003) used a Bayesian approach to forecast sales of new albums prior to their release. Elberse (2010) found that unbundling songs and selling them individually had a smaller effect on album sales when the songs were produced by a musician who had a relatively strong reputation and/or the songs' features

---

[2]Some playlists on streaming platforms have millions of followers. For example, Spotify's *Hot Country* has 5 million followers, *RapCaviar* has 10 million followers, and *Today's Top Hits* 21 million followers. Statistics indicate that 54% of online listeners use curated, sometimes personalized, playlists (DiMA 2018).

were similarly appealing. Papies and van Heerde (2016) examined the development of music marketing and the effects of technology on changes in the music industry and Papies and van Heerde (2017) studied the dynamic interplay between recorded music and live concerts and the effects of piracy, unbundling, and artist characteristics on demand elasticities. Datta et al. (2017) examined how adoption of online streaming services affected listening behavior, and Chung et al. (2009) proposed an adaptive system for music personalization. At a still broader level, our research is related to work done in other experiential contexts such as the motion picture industry (Eliashberg et al. 2000).

Numerous behavioral studies are also relevant to our research. Holbrook and Hirschman (1982) and Schmitt (2010) broadly examined experiential consumption while Bruner (1990) reviewed the effects of music on mood and emotions. Juslin and Laukka (2004) and Juslin and Västfjäll (2008) addressed mechanisms underlying emotional responses to music; Nave et al. (2018) related musical preferences to listeners' personalities; and Holbrook and Anand (1990) considered the effect of musical tempo on perceptions of activity, affective responses, and situational arousal. Most important for our research is Holbrook and Schindler (1989), which found that development of tastes for popular music follows an inverted U-shaped pattern that peaks at about twenty-four years of age.

We compared the predictive accuracy of our model to a set of 13 benchmark models. We show that our model is the best in both in-sample and out-of-sample predictive performance. More importantly, for a managerial perspective, we use a series of applications to show how our model can be used to (1) forecast the performance of new albums before they are launched, (2) recommend albums that are consistent with the musical styles associated with specific eras, (3) design contextual playlists based on user queries, (4) recommend albums that are musically similar to a given album and (5) provide diagnostics for albums that are predicted to be unsuccessful. The model also generates substantive insights relating album success with acoustics and themes. We find that the role of acoustics in shaping popular success has changed over time. We also discover a number of themes from the ser-generated tags that are useful for predicting success. Among other things, these themes can represent

variables for which we do not have explicit data (e.g., sub-genres, emotions, situations and artist characteristics).

It is important to note that while we focus on the musical context in this paper, our modeling framework can be used across a variety of consumption contexts. A distinctive aspect of our framework is its use of data of multiple modalities to capture the experiential aspects of consumption. These aspects are of prime importance in consumption contexts involving movies, podcasts, videos theatre, art, and other intangible goods. Multimodal data such as movie scripts, movie tags, video reviews, art genes can be combined with other measures such as genre and other metadata to predict the popularity and success of different items in each context.

The rest of the paper is organized as follows. The next section describes our data, and its sources and collection methods. This is followed by a section in which we develop the proposed model. Then we presents the results, discuss qualitative insights, and compare the predictions of our model to those obtained using a set of benchmark models. The penultimate section uses the model for album and playlist scoring, recommendation and design. We conclude by acknowledging the limitations of our research and identifying areas for future research.

Data Description

We used different online sources to assemble a multi-modal data set of American popular music spanning the 54 years from 1963 to 2016. The variables in our data include not only metadata such as the genres represented within the album, but also acoustical measures that characterize the listening experience, as well as textual tags that summarize the perceptions of listeners. We now describe the data and its sources, starting with our dependent variable.

Album Success: The *Billboard 200*

Billboard 200 is a ranking, released by the Billboard magazine, of the 200 best performing musical albums in a week. Since 1991, the ranking has been based on sales data obtained from Nielsen Soundscan. From 2014, these data have been augmented to include the sales of digital albums/tracks and revenue from online streaming. We scraped *The Billboard* magazine's website to obtain the weekly *Billboard 200* rankings for the 1963–2016 period.

Ideally, we would have preferred to use album sales as our dependent variable. As these are unavailable, we used, instead, the *Billboard*'s year-end score, which is obtained by summing the weekly inverse rankings of an album across all the weeks of a year in which it was on the charts (e.g., the top selling album on *Billboard 200* is assigned an inverse rank of 200 and the worst performing album an inverse rank of 1). *Billboard* exclusively used this score before switching to Nielsen Soundscan in 1991. The *Billboard* rankings and the year-end score are industry standards that have also been extensively used for research on popular music; see Bradlow and Fader (2001); Alexander (1996); Anand and Peterson (2000); Dowd (2004); Lena (2006); Lena and Pachucki (2013); Peterson and Berger (1975). A limitation of the year-end score is that it aggregates weekly rankings. Its strength, however, is that it considers both peak success and longevity of an album on the charts.

Focusing only on albums that make the charts can provide a biased picture of the determinants of musical success. We handled this selection issue by augmenting the *Billboard 200* data with albums that did not appear on the charts. To this end, we collected the full discography (a catalog of all the recordings of an artist) of all our artists and identified their albums that did not appear on the *Billboard 200*. From this set of non-charting albums, for each year, we randomly selected 200 albums that were released in the preceding three years.[3]

---

[3]This is a sufficiently long duration because a significant majority of the charting albums stay on the charts for at most one year
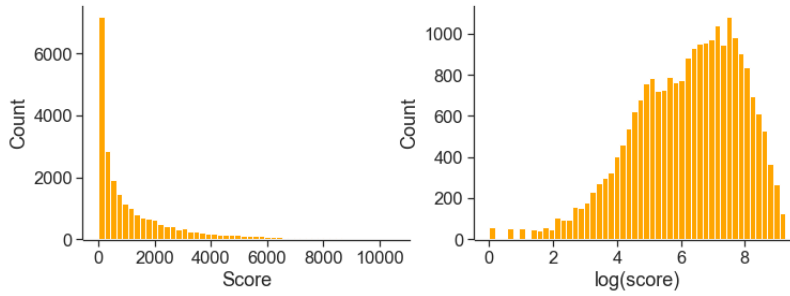
Figure 1: Distributions of the year end scores and their logarithmic transformations

Table 1: Ten albums with the all-time highest year-end scores on *Billboard 200*

Album	Artist	Label	Year	Year-end Score
Thriller	Michael Jackson	Epic	1983	10526
Jagged Little Pill	Alanis Morissette	Maverick	1996	10233
1989	Taylor Swift	Big Machine Records, LLC	2015	10181
Cracked Rear View	Hootie & The Blowfish	Atlantic Records	1995	10111
Fearless	Taylor Swift	Big Machine Records, LLC	2009	10080
Breakaway	Kelly Clarkson	RCA Records Label	2005	10044
Human Clay	Creed	The Bicycle Music Company	2000	10044
21	Adele	XL Recordings/Columbia	2012	10036
Backstreet Boys	Backstreet Boys	Jive	1998	10021
Crazysexycool	TLC	Arista/LaFace Records	1995	10016

Altogether, we obtained 34,214 observations across 22,506 unique albums that were created by 5,852 unique artists or groups. The observations include 15,233 albums that appeared, and 7,273 albums that did not appear, on *Billboard 200*. Table 1 shows the ten albums with the all-time highest year-end *Billboard 200* scores. The top album is Michael Jackson's *Thriller*, which sold an estimated 66 million copies and is the best-selling album of all time. The next two are Maverick's *Jagged Little Pill*, which sold 33 million copies; and Taylor Swift's *1989*, which has sold 10 million copies since its release in 2014.

Figure 1 shows the histograms of the observed year-end scores and their log-transformed values. The year-end scores have an average value of 169 and exhibit a long tail distribution. Only 10% of the albums have scores higher than 2,887. The log-transformed scores are less skewed and have a distribution that resembles the normal. For this reason, we use the
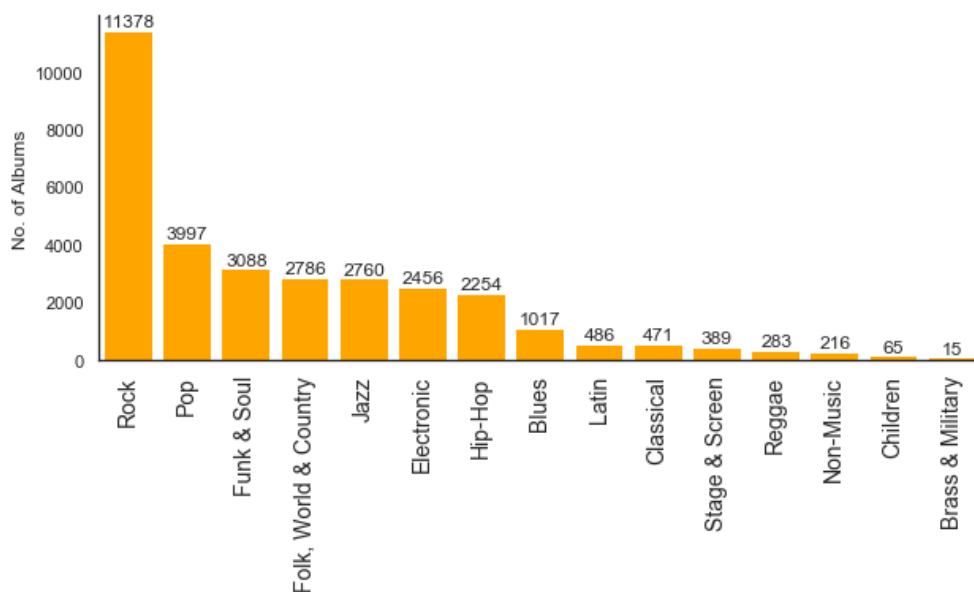
Figure 2: Distribution of genres across all observations in all years

log-transformed score as the independent variable in the following analysis.

A number of variables can be used to model album success. We consider these below.

Genres

Musical styles are conventionally described by their genre. We identify the genres of each album using the API for *Discogs*, which is a comprehensive crowd-sourced database of audio recordings. An album can belong to multiple genres if it fuses different types of music (e.g. electronic-rock combines electronic and rock music), if it results from a collaboration of artists with different backgrounds, or if its tracks have different genres. Figure 2 shows that Rock is the most common genre in our data.

The production and popularity of genres has changed over time. Jazz and Pop were dominant until Rock took over in 1965. Since then, Rock has been prevalent in at least a third of all released albums. Hip-Hop emerged in the early 70s and peaked in the 90s. Funk/Soul peaked in the 70s and held its own till the early 80s. Folk, World & Country music peaked twice, in the 60s and the 90s, and electronic music in the mid 70s. The other

genres have appeared only infrequently on the charts.

Acoustic Fingerprints

It is difficult to measure how a song impacts the listening experience. Acoustic fingerprints, which are digital summaries of a song's phonic features, are the best available measures for capturing a song's effect on a listener. Acoustic features encapsulate the creative experience on multiple dimensions, capture the underlying artistic style, and relate to the type of instruments and technologies used for producing music.

Some acoustical fingerprints capture the physical aspects of music: key, loudness, mode, tempo and time-signature. Others describe the listening experience: acousticness, danceability, energy, instrumentalness, liveness, speechiness and valence. We also consider track duration and the explicitness of lyrics as acoustic features of an album. We used the fourteen acoustic features described in the Web Appendix. We used the *Spotify* API to collect the acoustic fingerprints of the songs in each album. The fingerprints are produced using machine learning techniques by *The Echo Nest*. An album is a bundle of songs that can differ in their acoustical profiles. As in the balance model, we use the means and variances of the different acoustic measures to capture the average acoustical level and the variability of the acoustics across the songs in an album (Bradlow and Rao 2000; Farquhar and Rao 1976).[4]

User-Generated Tags

User generated tags reflect how listeners categorize and perceive different albums. *Last.fm* is an online music platform that started collecting member-generated tags describing music in 2002 (it also records the listening habits of, and recommends music to, its listeners). These tags contain a mix of factual and perceptual information about the albums. We collected

---

[4]Previous research has used acoustic features to compare and recommend songs and construct playlists (Bertin-Mahieux et al. 2008). Information retrieval professionals and researchers have also used acoustic features to develop more effective recommendation algorithms and understand diffusion and creativity in popular music (Askin and Mauskapf 2017).

Figure 3: Wordcloud of all unique tags

tags for the albums in our dataset using the *Last.fm* API.

Bag of Tags Representation   *Last.fm* data only reports the relative frequencies of the tags associated with an album. For example, it might report that 95% of the listeners who tagged an album used the word "Romantic." We used this information to construct a bag-of-words representation of size 100 for each album.[5] To prune the vocabulary and retain important tags that distinguish among albums, we used the term frequency-inverse document frequency (tf-idf) (Ullman 2011) to choose the top V tags across all albums. We deleted the tags with the 0.1% lowest tf-idf scores and retained the remaining 15,059 tags. This procedure resulted in a reasonable number of tags per album — a minimum of nine and an average of ninety-five tag applications. Less that 1% of the albums had fewer than 63 tag applications. Altogether, there were 2,145,139 tag applications across all albums and years. Figure 3 shows a word cloud of all unique tags in our data. Tags displayed with larger fonts have higher weights. Figure 4 shows the tags associated with the albums *Thriller*, *25* and *1989*. The unique tags are "male vocalists," "halloween" and "classic" for *Thriller*; "british," "blue" and "epic" for *25*; and "love at first listen," "electropop" and "synthpop"

---

[5]For example, if 100% listeners use the tag "Female Singer," 80% use the tag "Love" and 20% use the tag "Guitar," then its bag of tags includes $100/(100 + 80 + 20) = 50$ replications of "Female Singer", 40 replications of "Love" and 10 replications of "Guitar".

Figure 4: Wordcloud of tags for *Thriller* (1982) by Michael Jackson (left), *25* (2015) by Adele (middle) and *1985* (2014) by Taylor Swift (right)
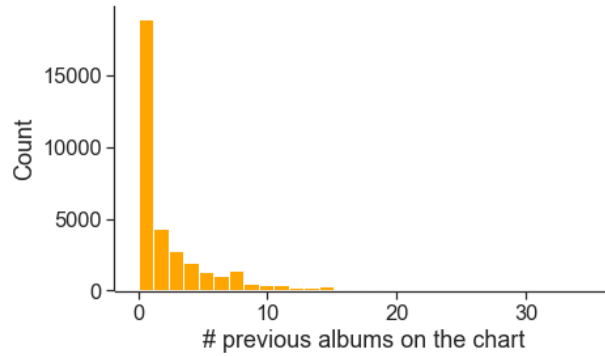

Figure 5: Histogram of previous charting albums

for *1989*. The tags "pop" and "albums I own" are common to the three albums.

Other covariates

We also collected information on the prior success of artists (superstardom) and the number of artists featured in the album.

Superstardom Artists differ in their popularity, with some being superstars. Moreover, their popularity varies over time. We measure an artist's superstardom at any point in time using the number of previous albums he or she had on the *Billboard 200*. Artists who have more previous albums on *Billboard 200* have more fans, are more visible (e.g., they appear on television shows, movies and advertisements) and are better supported by record houses. As Krueger (2005) and Giles (2007) observed, superstardom can have a spillover effect on the success of new albums.

11

Figure 5 shows the histogram of the number of previous albums an artist had on the *Billboard 200* at any point in time. The figure shows that most of the observations in our dataset correspond to artists with less than ten previous albums on the chart, with a median of one album, and 90% of the observations correspond to artists with less than seven charting albums. Barbra Streisand is the biggest superstar in our data. She is the only artist whose every album has appeared on *Billboard 200*. The latest, her thirty-fifth entry on the charts, was *Encore: Movie Partners Sing Broadway.* Five of her albums have achieved the top rank.

Number of artists We used the *Spotify* API to obtain data on the number of artists featured in each album. Most of all the albums in our data set feature one artist or group. *Symphony No.8* of Mahler is the album with the highest number artists, twelve.

Having described our data, we now present our modeling framework.

Modeling Framework

We develop a novel Bayesian modeling framework that flexibly and semiparametrically integrates the different data modalities (genres, acoustic features, and textual tags) to explain and predict album success. Let $i = 1$ to I denote albums and $t = 1$ to T the calendar years (i.e., 1963 to 2016) in our data. An album can appear on the charts for multiple years and therefore can be present several times in our data. Let album i appear in J_i years. Let $t(ij) \in \{1, \ldots T\}$ denote the calendar year associated with the jth observation for album i. Following are the variables used in the model.

- y_{ij} is the success score of album i on its jth observation, where $j \in \{1, \ldots J_i\}$. This variable is not observed for those albums that do not make it to the charts.

- $\boldsymbol{x}_i = [1, x_{i1}, \ldots, x_{iK}]$ is a $K + 1$-vector containing a constant for the intercept and the binary genre variables for album i.

- $\boldsymbol{a}_{ij}$ is a L-vector of continuous variables containing the means and variances of the

different acoustical measures, and the superstardom and number of artists covariates for album i. Of these, only the superstardom variable changes across the different observations of the album.

- $\boldsymbol{w}_i = \{w_{in}\}_{n=1}^{N_i}$ is a vector of N_i textual tags associated with album i.

Next, we describe how these variables are used in different components of our model. We begin with the dependent variable.

Censored Success Score

The *Billboard* year-end score is a yearly measure of an album's success. It is available only if an album appears on the charts. For other albums, the score is unobserved. Following Tobin (1958), we model this censored variable by assuming that the observed value of y_{ij} is a manifestation of a partially latent score y_{ij}^* such that

$$
y_{ij} = \begin{cases} y_{ij}^*, & \text{if } y_{ij}^* \geq 0, \\ \varnothing, & \text{if } y_{ij}^* < 0. \end{cases} \tag{1}
$$

We use a flexible link function to relate this partially latent variable to the different sets of covariates. In its generic form, this relationship can be written as

$$
y_{ij}^* = F_{t(ij)}\left(\boldsymbol{x}_i, \boldsymbol{a}_{ij}, \boldsymbol{w}_i\right) + \varepsilon_{ij}, \tag{2}
$$

where $F_{t(ij)}$ is a time-varying function and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is an idiosyncratic, normally distributed, error term. We now describe the link function in greater detail.

Link Function

The link function $F_{t(ij)}$ flexibly captures the impact of the different types of album features and latent variables on success via an additive specification

$$
F_{t(ij)}\left(\boldsymbol{x}_i, \boldsymbol{a}_{ij}, \boldsymbol{w}_i\right) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_{t(ij)} + f_{t(ij)}(\boldsymbol{a}_{ij}) + g_i(\boldsymbol{w}_i), \tag{3}
$$

where $\boldsymbol{\beta}_{t(ij)}$ is a vector of time-varying coefficients that captures the effects of the binary genre variables; $f_{t(ij)}(.)$ is an unknown, time-varying, multivariate function that captures the effects of the continuous variables (including the acoustics); and $g_i(.)$ is a function that quantifies the relationship between the tags and the success of an album. We separately discuss each component below.

Linear Effects of Genres

It is unlikely that the popularity of different genres remained unchanged over time. We used a state-space approach to allow for dynamics in the popularity of genres over the fifty four years for which we analyze data. We used a linear specification, $\boldsymbol{x}_j^\top \boldsymbol{\beta}_{t(ij)}$, to model the effects of the binary genre variables. Notice that the coefficients in $\boldsymbol{\beta}_{t(ij)}$ are indexed by $t(ij)$ which gives the calendar year for the jth observation of album i. All albums within a year t share the same coefficients, $\boldsymbol{\beta}_t$. We assume that $\boldsymbol{\beta}_t$ varies across the calendar years $t = 1$ to T according to a transition equation that we describe in detail in our section on dynamics.[6]

Non-linear Nonparametric Effects of Acoustics

We use the functions $f_{t(ij)}(\boldsymbol{a}_{ij})$, which are specific to calendar time, to model the dynamic effects of the acoustic fingerprints. We use an additive specification:

$$f_{t(ij)}(\boldsymbol{a}_{ij}) = \sum_{\ell=1}^{L} h_{t(ij)}^{\ell}(a_{ij\ell}),$$

where $h_{t(ij)}^{\ell}(.)$ is a smooth function for the (continuous) variable ℓ. It is not possible to *a priori* hypothesize how these covariates influence album success. Consequently, we use a nonparametric specification to model these unknown functions, using penalized splines (Wand and Ormerod 2008). We assume that for a given calendar year $t(ij)$, each component

---

[6]$t(ij)$ refers to the calendar year for the jth observation for album i, where as, t refers to the tth calendar year in our data.

$h^\ell_{t(ij)}(.)$ is a piecewise, second-order polynomial:

$$h^\ell_{t(ij)}(a_{it\ell}) = \psi^\ell_{t(ij)0} + \boldsymbol{\psi}^\ell_{t(ij)1} a_{ij\ell} + \boldsymbol{\psi}^\ell_{t(ij)2} a^2_{ij\ell} + \sum_{q=1}^{Q} \boldsymbol{\psi}^\ell_{\kappa q}(a_{ij\ell} - \kappa^\ell_q)^2_+, \qquad (4)$$

where $\{\psi_{t0}, \psi^\ell_{t1}, \psi^\ell_{t2}, \boldsymbol{\psi}^\ell_\kappa = [\boldsymbol{\psi}^\ell_{\kappa q}]^Q_{q=1}\}$ are parameters and $\kappa^\ell_1 < \cdots < \kappa^\ell_Q$ are fixed knots. We use quadratic basis splines involving the second-degree truncated basis function where $(a_{ij\ell} - \kappa)^2_+ = (a_{ij\ell} - \kappa)^2$ if $a_{it\ell} \geq \kappa$ and zero otherwise; and κ is a fixed knot in the compact support of $a_{ij\ell}$. The use of penalized splines allows us to seamlessly trade-off the flexibility and smoothness of the inferred functions. Usually, there is a tradeoff between using a larger number of knots (which increases the fidelity to the data) and the smoothness of the resulting function. The tradeoff is resolved using a roughness penalty. We use a fixed number of knots with known positions but constrain their influence by regularizing their basis coefficients using Bayesian priors.

Additive Model Summing over all L additive nonlinear effects yields the following form for the function $f_t(.)$:

$$f_{t(ij)}(\boldsymbol{a}_{ij}) = \psi_{t(ij)0} + \sum_{\ell=1}^{L} \left[\psi^\ell_{t(ij)1} a_{ij\ell} + \psi^\ell_{t(ij)2} a^2_{ij\ell} + \sum_{q=1}^{Q} \psi^\ell_{\kappa q}(a_{ij\ell} - \kappa^\ell_q)^2_+ \right], \qquad (5)$$

where $\psi_{t(ij)0} = \sum_{\ell=1}^{L} \psi^\ell_{t(ij)0}$ is a non-identified intercept, which we fix to zero. Sun et al. (1999) showed that Equation 5 can be represented as a mixed model with additional covariate-specific hierarchical variance components $\sigma^2_{\boldsymbol{\psi}_\kappa \ell}$, which regularize the basis coefficients and therefore control the extent of the smoothing by assuming that $\psi^\ell_{\kappa q} \sim \mathcal{N}(0, \sigma^2_{\psi_\kappa \ell})$.

The polynomial coefficients $\psi^\ell_{t(ij)1}$, $\psi^\ell_{t(ij)2}$, the basis coefficients $\boldsymbol{\psi}^\ell_{\kappa q}$ and the prior variances $\sigma^2_{\boldsymbol{\psi}_\kappa \ell}$ are inferred from the data, for each continuous variable ℓ. As $\sigma^2_{\psi_{\kappa l}}$ is estimated, the implicit regularization it induces ensures that the parameters of only the important knots are given significant weights. Small values of $\sigma^2_{\boldsymbol{\psi}_\kappa \ell}$ diminish the effect of the truncated basis; larger values allow the basis coefficients to significantly deviate from zero and allow a greater

effect of the truncated basis. This type of regularization is similar to what is done in ridge regression.

State-Space Dynamics

Notice that the coefficients of the genre variables in $\boldsymbol{\beta}_{t(ij)}$, as well as the coefficients $\psi^\ell_{t(ij)1}$ and $\psi^\ell_{t(ij)2}$, in Equation 4 for each variable ℓ, vary across the calendar years, $t = 1$ to T. All albums within a given calendar year t (i.e., for albums, with $t(ij) = t$), have the same coefficients, $\boldsymbol{\beta}_t$, and ψ^ℓ_{t1} and ψ^ℓ_{t2}. In modeling the evolution of these coefficients over the calendar years, it is important to capture the fact that coefficients are likely to evolve smoothly as dramatic changes from one year to the next are unlikely. We use a linear state-space specification to model such evolution.

Collecting all the time varying coefficients into a single vector $\boldsymbol{\gamma}_t = \{\boldsymbol{\beta}_t, \{\psi^\ell_{t1}, \psi^\ell_{t2}\}^L_{l=1}\}$, we model their evolution via a second order auto-regressive AR(2) process (see Biyalogorsky and Naik (2003) for a variant of this) as follows:

$$\boldsymbol{\gamma}_t = \boldsymbol{\varphi}_1 + \boldsymbol{\varphi}_2 \circ \boldsymbol{\gamma}_{t-1} + \boldsymbol{\varphi}_3 \circ \boldsymbol{\gamma}_{t-2} + \boldsymbol{\omega}_t. \tag{6}$$

In the above equation, $\boldsymbol{x} \circ \boldsymbol{y}$ denotes the element-wise product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The residuals in $\boldsymbol{\omega}_t$ are distributed normal, $\mathcal{N}(0, \boldsymbol{\Sigma}_\gamma)$, with a diagonal covariance matrix $\boldsymbol{\Sigma}_\gamma$. Using an AR(2) transition equation allows us to capture different patterns of evolution. Depending upon the values of the $\boldsymbol{\varphi}$ coefficients, cyclical patterns of evolution are also possible.[7] The variance components $\sigma^2_{\gamma,k}$ in $\boldsymbol{\Sigma}_\gamma$ capture the volatility in the values of the coefficient over time. It is possible to include in Equation 6 additional terms to accommodate factors that drive the evolution in the coefficients (e.g., technological drivers of music). As we do not have access to these drivers in our data, we do not include these in the paper.

Having discussed the first two components of Equation 3, we now focus on the last component that incorporates the perceptual tags.

---

[7]We thank an anonymous reviewer for suggesting this stochastic process.

Nonparametric Themes

Tags can also be useful for explaining album success. Given the large number of tags in our dataset, we cannot directly include them in the model as covariates. Instead, we incorporate their effect via topics or themes that predict album success. We use a supervised hierarchical Dirichlet process (SHDP) to model the themes. This is a generalization of the commonly used Latent Dirichlet Allocation (LDA) topic model (Blei et al. 2003; Tirunillai and Tellis 2014) that summarizes documents into topics. It differs from LDA on two aspects. First, the inferred themes are obtained in a *supervised* fashion as they are informed not only by the tag collections across the albums, but also by the success score of the albums. Hence, the themes are predictive of success. Second, unlike in an LDA, the hierarchical Dirichlet process nonparamterically assumes a countably infinite number of themes and ensures that only a finite number of these themes have significant weights in any given application. In summary, the SHDP allows us to automatically infer the number of themes that correlate with album success.

Themes   A theme is a discrete probability distribution over the vocabulary of V tags. Formally, the kth theme is described by $\boldsymbol{\phi}_k^w = [\phi_{k,v}^w]_{v=1}^V$, where $\sum_{v=1}^V \phi_{k,v}^w = 1$ and $\phi_{k,v}^w$ is the probability that tag $v \in \{1, \ldots, V\}$ appears in that theme. Thus themes reside in a $V - 1$ dimensional simplex. All the tags appear in all the themes, but with different (possibly near-zero) probabilities. That is, the themes differ in the weights they place on each of the V tags. As the number of themes is a priori unknown, we consider a Bayesian nonparametric framework where we assume that there is a countably infinite number of themes $[\boldsymbol{\phi}_k^w]_{k=1}^\infty$ that are *shared* across all the albums. For each theme $\boldsymbol{\phi}_k^w$, we assign a scalar ϕ_k^y that indicates the contribution of the theme to the success of an album. Each "augmented" theme is then fully defined by the tuple $\boldsymbol{\phi}_k = (\boldsymbol{\phi}_k^w, \phi_k^y)$.

Mother Distribution of Themes   The themes that summarize the tags need to be shared across all the albums. This implies that we need a discrete distribution G_0 that generates the

countably infinite number of themes. As this distribution is unknown, we assume a Dirichlet process (DP) prior for it, i.e.,

$$G_0 | \zeta, H \sim \mathrm{DP}(\zeta, H), \tag{7}$$

where the baseline distribution $H = H^w \times H^y$ is a joint discrete-continuous probability measure on the themes and their success coefficients, and $\zeta > 0$ is a precision parameter. The baseline distribution H represents the mean of the process and sets its location, and ζ controls the dispersion of the DP realizations around H. We use a symmetric Dirichlet distribution $H^w = \mathrm{Dir}(1/V)$ for the themes and a diffuse normal distribution H^y for generating the success coefficients for the themes.

As the DP realizations are discrete distributions, G_0 places its mass over a countably infinite collection of augmented themes, (i.e., probability vectors and their success coefficients), $\boldsymbol{\phi} = [\boldsymbol{\phi}_k]_{k=1}^\infty$ and can be written as

$$G_0 = \sum_{k=1}^\infty \varsigma_k \delta_{\boldsymbol{\phi}_k}, \quad \boldsymbol{\phi}_k \sim H, \tag{8}$$

where $[\boldsymbol{\phi}_k]_{k=1}^\infty$ is an infinite sequence of augmented themes drawn from the base distribution H, $\delta_{\boldsymbol{\phi}_k}$ is a discrete random measure concentrated at atom $\boldsymbol{\phi}_k$, and $[\varsigma_k]_{k=1}^\infty$ is an infinite sequence of weights that sum to one, obtained via the stick breaking representation of a DP (Sethuraman 1994). In this way, G_0 initializes the infinite collection of themes and their success coefficients that are then shared across the albums.

The clustering properties of the DP depend upon its precision parameter. If ζ is large, the DP realizations mimic the baseline distribution H. As its value approaches zero, the DP generates distributions that are concentrated on a few mass points (i.e., augmented themes). We infer ζ from the data and therefore automatically determine the number of themes that are present in a given dataset.

Albums as a Mixture of Themes　The collection of tags for an album can be considered as a mixture of the themes $[\boldsymbol{\phi}_k]_{k=1}^\infty$. The albums differ in the proportions with which the themes are represented in them and therefore place different weights on the themes. Let

G_i denote the unknown discrete distribution, defined over the themes $[\boldsymbol{\phi}_k]_{k=1}^{\infty}$, from which album i draws its themes. To ensure that the album specific distributions G_i place their weights only on the augmented themes in the mother distribution (i.e., on the atoms of G_0) we assume a common Dirichlet process population distribution for all G_i, i.e.,

$$G_i \sim \mathrm{DP}(\alpha_0, G_0),$$

where G_0 is the baseline distribution and $\alpha_0 > 0$ is a precision parameter. G_i can be represented as an infinite mixture,

$$G_i = \sum_{k=1}^{\infty} \pi_{ik} \delta_{\boldsymbol{\phi}_k}, \tag{9}$$

where the atoms $[\boldsymbol{\phi}_k]_{k=1}^{\infty} \sim G_0$ and π_{ik} is the probability weight for atom $\boldsymbol{\phi}_k$. The G_i's are independent across albums; each G_i selects a specific subset of themes for album i.

Each tag in album i comes from its associated theme. Let $\boldsymbol{\theta}_{in} = (\boldsymbol{\theta}_{in}^w, \theta_{in}^y)$ denote the augmented theme for the n-th tag, w_{in}, of album i, where $\boldsymbol{\theta}_{in}^w$ is equal to one of the probability vectors in $[\boldsymbol{\phi}_k^w]_{k=1}^{\infty}$ and θ_{in}^y is its corresponding success score. Then

$$\boldsymbol{\theta}_{in} \mid G_i \sim G_i.$$

Given the theme, the tag w_{in} comes from the categorical distribution

$$w_{in} \mid \boldsymbol{\theta}_{in}^w \sim \mathrm{Categorical}(\boldsymbol{\theta}_{in}^w). \tag{10}$$

The set $\boldsymbol{w}_i$ of the N_i tokens within the album can contain many replications of the same tag that were assigned by different users, and two replicates of the same tags can come from different themes. The $g_i(\boldsymbol{w}_i)$ component in Equation 3 links the themes to overall album success as in

$$g_i(\boldsymbol{w}_i) = \sum_{n=1}^{N_i} \frac{\theta_{in}^y}{\mathcal{C}(\theta_{in}^w)}, \tag{11}$$

where $\mathcal{C}(\theta_{in}^w) = \sum_{n'=1}^{N_i} \mathbf{1}\{\theta_{in}^w = \theta_{in'}^w\}$ is the number of tags in the album that share the same theme as θ_{in}^w and $\mathbf{1}(.)$ is the indicator function. As the themes are linked to the overall success

19

score, the model ensures that the inferred themes are *supervised* and therefore predictive of success. Our formulation of $g_i(\boldsymbol{w}_i)$ allows us to capture the effect of the themes on musical success based solely on their presence or absence within an album. To understand how Equation 11 works, consider for example, an album with three tags $\boldsymbol{w}_i = \{w_{i1}, w_{i2}, w_{i3}\}$ and associated themes $\{\theta_{i1}^w = \phi_1^w, \theta_{i2}^w = \phi_1^w, \theta_{i3}^w = \phi_2^w\}$. This thematic profile implies that themes 1 and 2 are represented in the album. Their contribution to success is given by $g_i(\boldsymbol{w}_i) = \frac{\phi_1^y}{2} + \frac{\phi_1^y}{2} + \phi_2^y = \phi_1^y + \phi_2^y = \boldsymbol{z}_i'\boldsymbol{\phi}^y$, where $\boldsymbol{z}_i$ is an infinite dummy variable vector with its first two elements equal to one and zeros elsewhere, and the coefficients associated with this vector are given by $\boldsymbol{\phi}^y = [\phi_k^y]_{k=1}^\infty$. This is different from what is done in traditional supervised topic models, where the theme proportions within an album are used, instead of the binary variables that we employ. Our approach has the significant advantage that it yields coefficients that are readily and cleanly interpretable. Moreover, it allows us to also identify the intercept in the regression.

Figure 6 provides a schematic representation of the supervised hierarchical Dirichlet process (see Dai and Storkey 2015) for a simplified vocabulary of three tags. The figure shows both the hierarchy of distributions as well as the data generative process for the tags.

In summary, our model flexibly integrates data from multiple modalities, using a combination of static and dynamic components. The dynamics in our model are parametric and linear in scope, where as the nonlinear and nonparametric components , i.e., the basis terms and the supervised HDP components are static in nature. This results in a partially linear semi-parametric model. We now further summarize the entire modeling framework via its generative process.

Generative Process

Given the above, our model can be specified using the following generative process.

1. Draw the hyper-distribution of themes and their success coefficients, $G_0 \sim \mathrm{DP}(\gamma, H)$
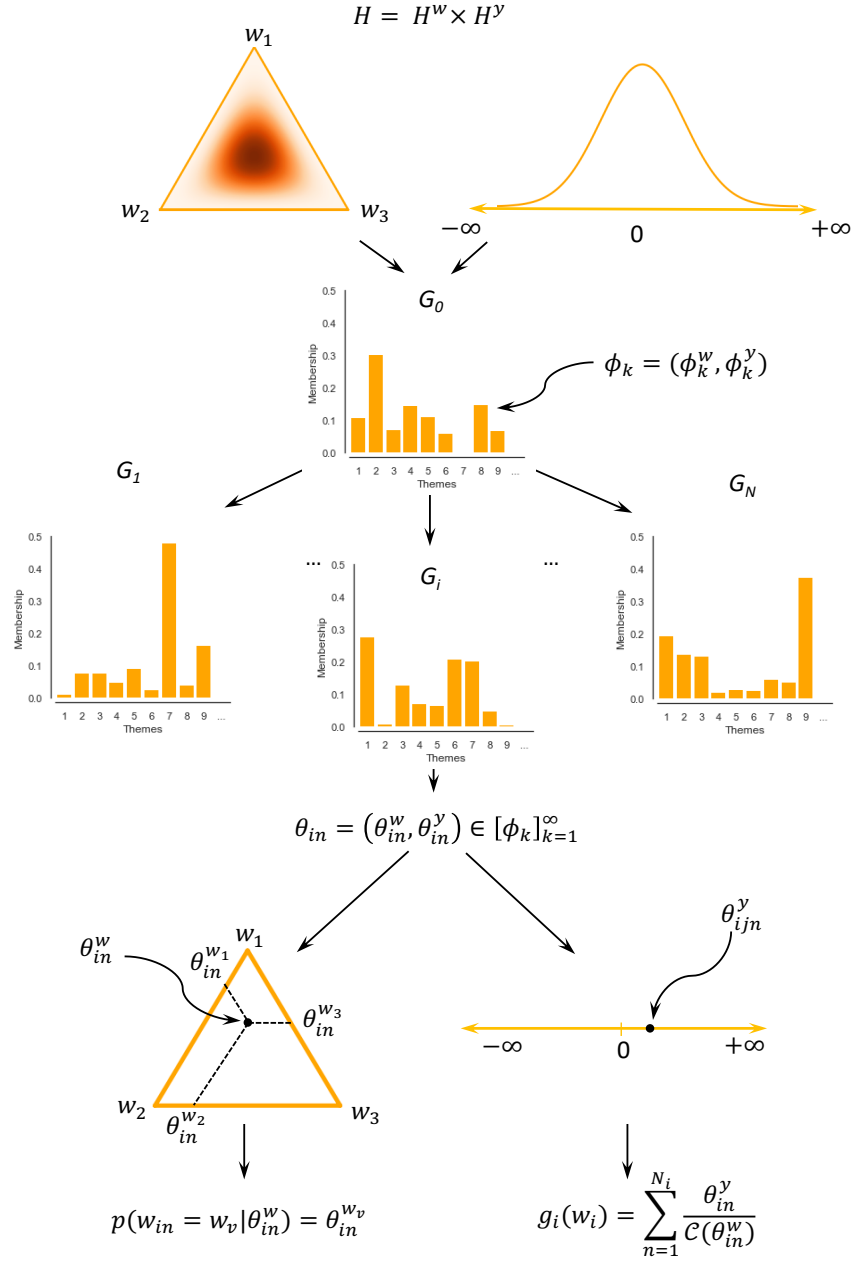
2. For each album i,

Figure 6: Illustrative representation of the SHDP with a vocabulary of three tags

(a) Draw the thematic mixture $G_i \sim \text{DP}(\alpha_0, G_0)$

(b) For each tag application n,

    i. Draw theme $\boldsymbol{\theta}_{in} = (\boldsymbol{\theta}_{in}^w, \boldsymbol{\theta}_{in}^y) \sim G_i$

    ii. Draw tag $w_{in} \sim \text{Categorical}(\boldsymbol{\theta}_{in}^w)$

3. Draw the truncated basis coefficients of the non-linear effects, $\boldsymbol{\psi}_\kappa \sim \mathcal{N}(0, \sigma_{\psi_\kappa}^2 \boldsymbol{I})$

4. For each year t, draw the time varying coefficients for the genre, acoustics and other continuous covariates,

$$\boldsymbol{\gamma}_t \sim \mathcal{N}(\boldsymbol{\varphi}_1 + \boldsymbol{\varphi}_2 \circ \boldsymbol{\gamma}_{t-1} + \boldsymbol{\varphi}_3 \circ \boldsymbol{\gamma}_{t-2}, \boldsymbol{\Sigma}_\gamma)$$

5. For each observation j of album i, in year $t(ij)$, draw score

$$y_{ij} \sim \mathcal{N}(F_t\left(\boldsymbol{x}_i, \boldsymbol{a}_{it(ij)}, \boldsymbol{w}_i\right), \sigma_\varepsilon^2)$$

Posterior Inference via MCMC methods

We use a fully Bayesian approach to infer the unknowns in our model. Let the collection of all unknowns be given by $\Delta = \{\{y_{ij}^*\}, \{\boldsymbol{\gamma}_t\}\boldsymbol{\varphi}, \boldsymbol{\psi}_\kappa, \boldsymbol{\Sigma}_\gamma, \sigma_{\boldsymbol{\psi}_\varphi}^2, \{\theta_{in}\}, \alpha_0, \zeta, \sigma_\varepsilon^2\}$. Then the joint distribution of the data and unknowns can be written as

$$p(\{y_{ij}\}, \{y_{ij}^*\}, \boldsymbol{w}_{1:I}, \Gamma) = \prod_{i=1}^{I} \prod_{j \in \mathcal{T}_i} p(y_{ij}|y_{ij}^*) p(y_{ij}^*|\boldsymbol{\gamma}_{t_{ij}}, \boldsymbol{\psi}_\kappa, \boldsymbol{w}_i, \{\theta_{in}\}, \sigma_\varepsilon^2)$$

$$\times \prod_{t=1}^{T} p(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t-1}, \boldsymbol{\gamma}_{t-2}, \boldsymbol{\varphi}, \boldsymbol{\Sigma}_\gamma) \times p(\boldsymbol{\psi}_\varphi|\sigma_{\psi_\kappa}^2) \times \prod_{i=1}^{I} \prod_{n=1}^{N_i} p(w_{in}|\boldsymbol{\theta}_{in}) p(\boldsymbol{\theta}_{in}|\alpha_0, \gamma)$$

$$\times p(\sigma_\varepsilon^2) p(\{\gamma_{-1}, \gamma_0\}) p(\boldsymbol{\varphi}) p(\boldsymbol{\Sigma}_\gamma) p(\sigma_{\psi_\kappa}^2). \tag{12}$$

Note that we integrate over the mixture distributions G_i and G_0 and perform direct inference on the albums themes $\boldsymbol{\theta}_{in}$. As the full posterior distribution $p(\Delta|\{y_{ij}\}, \boldsymbol{w}_i)$ is not available in closed form, we use MCMC methods to summarize the posterior. We use data augmentation to infer the censored score for the non-charting albums in each MCMC iteration. This, with the use of conjugate priors for the other unknowns, results in closed-from full conditionals resulting in a Gibbs sampling algorithm.[8] In other nonlinear contexts where such data

---

[8]Full conditionals are available in the web appendix.

augmentation is not possible, advanced methods as in Rubel and Naik (2017)and Bruce (2019) are needed. Integrating out the HDP priors results in a collapsed Gibbs sampling scheme for inferring the themes. Here, we rely on the algorithm based on direct assignment of the tags as in Teh et al. (2005), Heinrich (2011), and Dai and Storkey (2015). We coded the inference algorithm in `Python` and used `numba` to compile our code to processor language to ensure efficient computation in our big-data setting.

Results

We first compare our model with a set of benchmark models on a number of predictive metrics. We then highlight the qualitative insights obtained from the model.

Model estimation and comparison

We use several static, dynamic, parametric and semi-parametric models in our comparisons.

The parametric models rely on a two-stage approach that first summarizes the tags into themes and then estimates their effect on album success. We used the `gensim` package in `Python` to estimate an ordinary Latent Dirichlet Allocation (LDA) model with an assumed fixed number of themes K_ϕ. As the number of topics are not known, we used five variants involving $K_\phi = \{20, 40, 60, 80, 100\}$ themes. The LDA yields theme membership probabilities $\boldsymbol{\rho}_i$ for each album i in our data These membership vectors are then used as covariates in the parametric models. The genres, acoustics, theme memberships and other control variables are included linearly in the link function.[9] As the membership vectors sum to 1, i.e., $\sum_{k=1}^{K_\phi} \rho_{ik} = 1$, an intercept is not identifiable. We therefore fixed it to zero. For the set of static benchmark models, we assumed that the estimated coefficients do not change over time. For the set of dynamic models, we allowed the coefficients to evolve over time following the AR(2) autoregressive state-space process described by equation 6 in our model section.

---

[9] $F_{t(ij)}\left(\boldsymbol{x}_i, \boldsymbol{a}_{ij}, \boldsymbol{w}_i\right) = [x_{i1}, \ldots, x_{iK}, \boldsymbol{a}_{ij}, \boldsymbol{\rho}_i]' \boldsymbol{\beta}.$

In addition to the parametric benchmark models, we also estimated a set of three semi-parametric variants of our model. Model M_1 uses artist superstardom and number of artists on the album as predictor variables. Model M_2 adds album genre to the variables in M_1; and model M_3 adds acoustic features to the variables in M_2. We label the proposed model as M_4; it includes the variables in M_3 and the user generated tags. We included dynamics in the M_1, M_2 and M_3 using our state-space specification and we used semi-parametric functional forms for the continuous covariates.

We estimated each model using MCMC methods. The Markov chains converged within 10,000 iterations for the parametric and the semi-parametric models M_1, M_2 and M_3. We used the first 5,000 MCMC draws for burn-in and report results based on the last 5,000 MCMC. Model M_4, which includes textual data, required more iterations. We ran the chain for 50,000 iterations and retained the last 5,000 iterations after burn-in. To allow meaningful comparisons across variables, we normalized all continuous variables (e.g., acoustic features) to a scale between zero and one. The estimation time for M_4 depends upon the vocabulary size. Given the large size of our vocabulary, this model takes about two weeks to sample 50,000 MCMC draws. The most time-consuming part of the computation stems from the supervised nature of our HDP specification.

To compare these models in terms of fit and predictive ability, we randomly split the data set into a calibration sample with 30,848 observations and a holdout data set of 3,366 observations. The holdout sample contains 10% of the albums in each year. We estimated all models on the calibration data and used the estimated parameters for holdout predictions. The theme membership indicators for the holdout albums are sampled with the themes fixed to those from the calibration sample and by ignoring the information in the success scores.

Predictive performance   We evaluated the predictive performance of the models using the log-likelihood (LL) and the Pareto smoothed importance sampling leave-one-out statistic (PSIS-LOO) which is asymptotically equivalent to WAIC, but is more robust for model comparison in finite sample settings (Vehtari et al. 2017). The PSIS-LOO accounts for both

model fit and complexity via an implicit penalty for complex models. The model with the highest PSIS-LOO is considered the best when comparing a set of models. We also used the Area Under the Curve (AUC) of a model's Receiver Operating Characteristic (ROC) curve (Swets 2014) to assess the closeness of the predicted and actual probabilities of an album's appearance on a chart. Finally, we used the root mean square error (RMSE) of the observed and estimated success scores for albums that made it to the charts. Table 2 reports these statistics both for the calibration and holdout data.

Table 2: Comparison of the fit and predictive performances of the different models

	In-sample				Out-of-sample	
	LL (of log-score)	PSIS-LOO	AUC	RMSE	AUC	RMSE
# Themes	Parametric Static Models					
20	-69,547.01	-69,589.20	0.72	3.53	0.70	3.67
40	-70,090.89	-70,142.77	0.71	3.71	0.70	3.73
60	-70,300.59	-70,360.43	0.70	3.72	0.68	4.17
80	-70,772.75	-70,843.38	0.70	3.81	0.69	3.72
100	-70,826.69	-70,906.24	0.70	3.84	0.70	3.77
# Themes	Parametric Dynamic Models					
20	-67,980.66	-68,549.10	0.75	3.07	0.73	3.16
40	-67,686.73	-68,171.56	0.76	3.08	0.73	3.17
60	-67,835.68	-68,382.09	0.75	3.03	0.72	3.15
80	-68,052.16	-68,614.17	0.75	3.07	0.74	3.12
100	-67,917.17	-68,440.05	0.75	3.07	0.72	3.12
Variants	Nonparametric Model					
M_1	-71,745.73	-71,789.41	0.62	4.48	0.62	4.54
M_2	-70,955.89	-71,173.32	0.66	4.24	0.66	4.35
M_3	-67,052.16	-67,724.55	0.78	3.17	0.75	3.20
M_4	-66,633.24	-67,156.76	0.78	3.07	0.76	2.98

It is clear from the last row of the table that our proposed model, M_4 has the best in sample and out of sample performance across all these metrics. We note that the dynamic models perform better than the static models, both in-sample and out of sample. Finally,

we see that adding nonparametrics results in improvements on the PSIS-LOO statistics and in the in-sample as well as out of sample AUC and RMSE measures.

Having compared our model to other benchmarks, we now briefly discuss three sets of results. First, we examine how the popularity of certain genres has changed over time. Second, we study how the average level of the acoustic features and their variability across the songs affects album success. For brevity, we focus on a few genres and acoustics in the main body of the paper and relegate a fuller discussion to the web appendix. Finally, we assess how the perceptual and experiential aspects of music, reflected via the crowd-sourced tags, affect album success.

Dynamics of Genre Popularity

We illustrate how our model captures the dynamics of genre popularity by focusing on Rock and Pop, the two main genres in our data set. The results for the other genres are presented in the web appendix. Figure 7 shows how the appeal of Rock and Pop genres has changed over time. The vertical axis in each plot shows the coefficient for each genre.
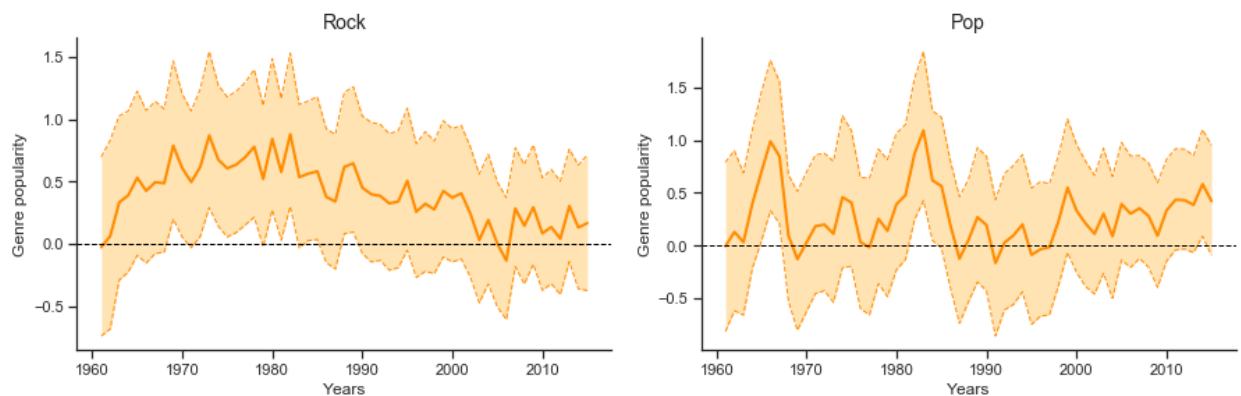
Figure 7: Patterns in the popularity of Rock and Pop over time

Rock: Rock music is the most produced genre in our dataset. The "British invasion" of American popular music coincides with the peak of rock music in the 60s. Classic albums of the period include *Revolver*, *The Beatles* and *Sgt. Pepper's Lonely Hearts Club Band* by the

Beatles. Albums like Pink Floyd's *The Dark Side of the Moon* drove Rock music to another peak in the 70s. Since then, the popularity of Rock music has declined while Pop and Hip Hop have risen; and Rock music itself has fragmented into specialized forms. The recent resurgence since 2010s has been led by albums like *Blurryface* by Twenty One Pilots.

Pop: Pop music has had significant periods of popularity. The 60s peak in the figure corresponds with the release of *More of the Monkeys* by the Monkeys and Herb Alpert's *Whipped Cream & Other Delights*. New styles of Pop emerged in the 80s, the most notable being Michael Jackson's *Thriller* and Prince's *1999*. Recently, Pop has again shown a positive trend in popularity with the launch of albums like Taylor Swift's *1989*, Adele's *21*, and Ed Sheeran's ×.

Acoustic Balance and Album Success

We illustrate how the success of an album is related to the average levels of the acoustic features and their variances across its songs. As noted, our acoustic fingerprints contain "technical" and "experiential" features. Technical features refer to physical characteristics of the music itself and experiential features relate to how music is experienced by the listener. Recall that, we modeled album success to be nonlinearly related with the acoustic measures in any given year. We further allowed the nature of this nonlinear relationship to vary across the years. Our results indicate that such flexibility is needed, as many of the acoustics indeed show dynamic nonlinear effects. We now discuss the results for Tempo and Valence, and relegate the detals for the remaining acoustic fingerprints to the web appendix.

Tempo: Tempo characterizes the speed of music. Figure 8 shows the nonlinear effects for mean and the standard deviation of tempo on the success of albums. The sub-figures show that on average, albums with fast tracks were more likely to succeed for most of the years. Success is also associated with a moderate level of variation in tempo across an album's songs. As listening to a succession of fast tempo songs can become tiresome, mixing these with slow tempo songs can be more relaxing. Moreover, this provides a more balanced and variegated listening experience.
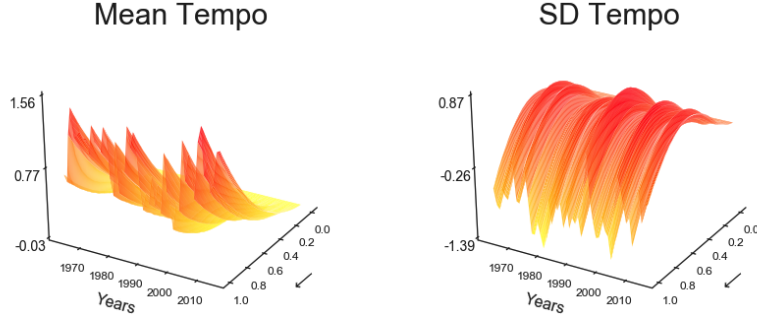
Figure 8: Estimated nonlinear effects for mean and the standard deviation of tempo
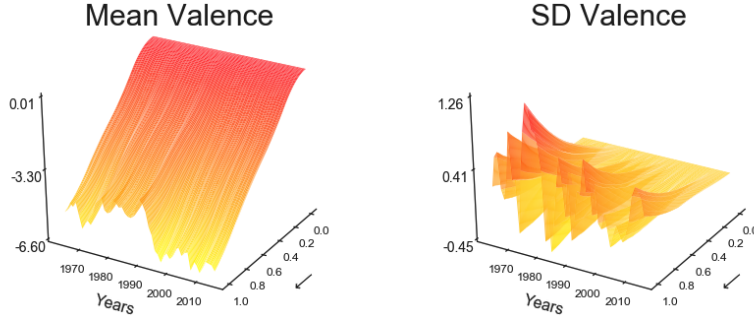

Figure 9: Estimated nonlinear effects for mean and the standard deviation of valence

Valence: Figure 9 shows that, albums with negative (sad or angry) valence are more likely to succeed. Moreover the nature of this relationship is the same across the years. This is consistent with the findings by Koelsch et al. (2006) that sad music can be more enjoyable because it has a greater aesthetic appeal and activates other positive emotions (Scherer 2004; Zentner et al. 2008). Our results also suggest that a more successful album balances the emotional content across its songs — that is, it does not have all happy or sad songs. This "demand" for variety increased in the mid 1970s then lowered in the subsequent years.

Thematic Insights

The supervised HDP component of our model yielded a total of 60 themes. These are the atoms of the G_0 distribution with substantial weights. Figure 10 shows the traceplot for the number of themes discovered across the MCMC iterations. To stabilize the computations, the number of themes is fixed to 50 for the first 5,000 MCMC draws. The histogram of the
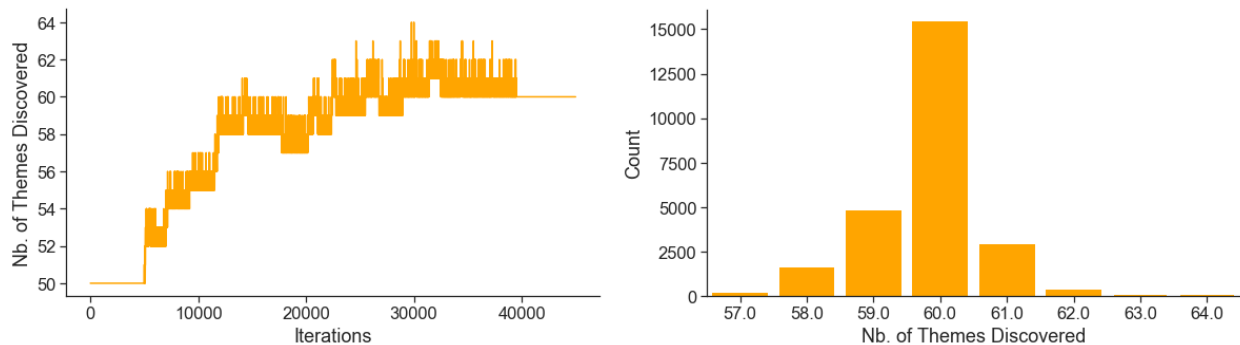
Figure 10: Number of Themes Discovered by the Supervised HDP.

number of themes is based on the last 25,000 iterations. After 35,000 iterations, the number of themes stabilized between 58 and 63 and the discovery of themes was frozen. We report results based the last 5,000 iterations, for which the number of themes were fixed to the modal value of 60.

Theme Distribution and Prevalence Table 3 shows the top tags (i.e., those with the highest probability of occurring within the theme) associated with the most prevalent themes $(p(\phi_k) \geq 0.01)$. The second column of the table shows the success coefficient associated with the theme ϕ^y, where as the third column shows the probability with which the theme appears across all the tag applications. These 42 themes account for 92% of all themes represented within the albums. The themes yield meaningful categories: specific types of music, artist gender, musical eras, and different groupings relating to music consumption. For example, themes 7, 15, 20 and 34 refer to sub-genres (e.g., country pop, classic rock, rap, and christian). Theme 54 is about female vocalists singing love songs. Themes 11, 22, 31, 47 and 49 are about the music from particular decades (e.g., 1960, 1970, 1980, 1990 and 2000s). Theme 15 is about the British influence, theme 21 is about Christmas and holidays, and theme 38 is about smooth jazz.

The themes themselves are ordered in descending order of ϕ^y, their success coefficient such that the theme that impacts success the most appear at the top of table 3. For example, the presence of theme 7, which is about modern country music, improves by 0.28 the success
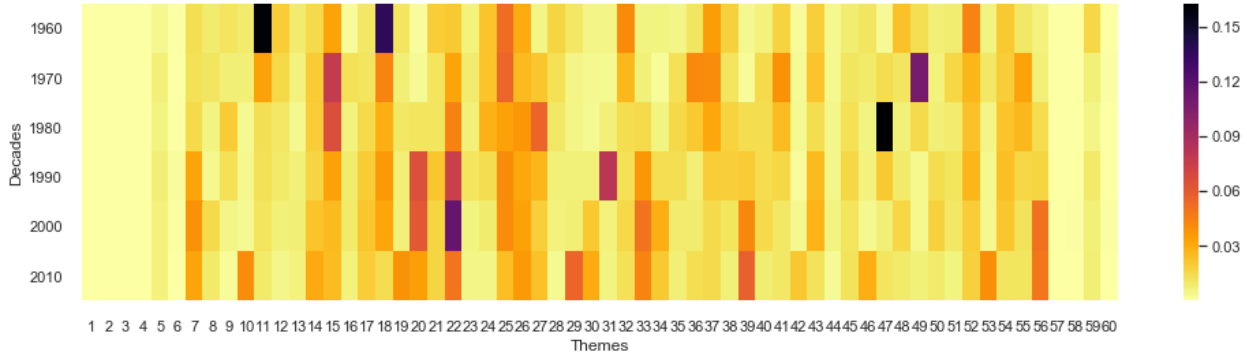
29

Figure 11: Theme proportions across decades

score of an album. However, theme 56, which is about indie rock, negatively impacts the success score by -0.21.

Figure 11 shows the average of the theme proportions of all the albums within each decade. A darker cell means that the corresponding theme is present in higher proportions among the albums in a decade. This can be verified by looking at the themes in Table 3 that focus on particular decades. These themes have a higher density in the corresponding row in Figure 11. For example, theme 11, which has "60s", "1966" and "50s" as the top tags in Table 3, peaks exactly during the 1960s and this is true for each of the decade related themes — theme 22 (2000s), theme 31 (1990s) , theme 47 (1980s), theme 49 (1970s) and theme 11 (1960s).

The themes that are associated with different styles and listening experiences in Table 3 also appear in the relevant years in Figure 11. For example, rap and hip-hop (Theme 20) was more prevalent in the 1990s and 2000s. Classic rock, hard rock and garage rock (Theme 15) appeared more in the 1970s and 1980s. This was driven by the British invasion (Beatles, Rolling Stones) in the 60s. Hard rock became mainstream with the success of bands like Iron Maiden, Saxon, and Def Leppard. Jazz, swing jazz and saxophone (Theme 18) appear more frequently in the 60s albums, which is consistent with the success of vocal jazz and swing bands in the 50s.

Table 3: Description of top themes with $p(\phi_k^w) \geq 0.01$

Theme k	ϕ_k^y	$p(\phi_k^w)$	Top Tags
7	0.28	0.027	country, modern country, london symph, 2004, george strait taggradio, country pop, contemporary country, television soundtrack
8	0.24	0.010	2010, philly soul, best of 2010, b s orchestra, 1968, my gang 10, 10s, e harris, mehta, r buchanan
10	0.22	0.011	2015, best of 2015, 1972, sjass, best albums of the 2010s, rb, 2010s, j smith, 10s, george duke
11	0.21	0.025	60s, oldies, psychedelic, bernstein, psychedelic rock, 1967, got it on 8 track, 50s, garage rock, 1966
14	0.10	0.022	electronic, disco, dance, electronica, synthpop, house, ambient, electro, chillout, rock and roll
15	0.08	0.041	rock, classic rock, album rock, british, hard rock, rock n roll, arena rock, american, guitar, blues rock
17	0.06	0.018	punk, punk rock, americana, alt-country, bluegrass, soft-rock, ska, covers, cover, bbk
18	0.06	0.037	jazz, swing, jazz vocal, jazz piano, saxophone, trumpet, modernjazz, vocalistas femeninas, vocal jazz, bebop
19	0.05	0.015	2012, gospel, basically bass, best of 2012, aor, melodic rock, new orleans, 10s, windham hill records, c m rae
20	0.03	0.037	rap, hip-hop, hip hop, gangsta rap, west coast rap, my collection great 150 albumz of rap, east coast rap, best album, hiphop, dirty south
21	0.03	0.016	christmas, southern rock, country christmas, xmas, holiday, christmas music listening, christmas music, weihnachten, wants, jazz organ
22	0.03	0.066	albums i own, favorite albums, 00s, favourite albums, 2006, 2005, 2000s, 2003, rock, favorites
24	0.02	0.015	comedy, weallgetold, latin jazz, stand-up, emusic, jazztrumpet, california, rotfl, spoken word, funny
25	0.02	0.039	soul, rnb, r&b, motown, neo-soul, rhythm and blues, smooth, new jack swing, sexy, urban
26	0.02	0.033	pop, pop rock, male vocalists, female vocalist, dance, soft rock, american, adult contemporary, vocal, dance-pop
27	0.01	0.026	hard rock, heavy metal, metal, glam metal, hair metal, glam rock, musicalmessradio, beach music, rock n roll, speed metal
29	-0.01	0.018	2014, soundtrack, best of 2014, playlist, 10s, soundtracks, broadway, musical, 2014
30	-0.01	0.015	post-hardcore, pop punk, emo, hardcore, my albums, metalcore, screamo, melodic hardcore, emocore, registret
31	-0.01	0.019	90s, 1993, 1996, 1999, 1991, 1994, 1992, 1995, 1990, 1997
32	-0.01	0.018	folk, folk rock, killforpeace, canadian, country rock, awesome, leadbelly and co, not available on last-fm radio yet, bob dylan, janis ian
33	-0.01	0.033	alternative rock, killforpeace, alternative, grunge, post-grunge, power pop, alternative metal, pop rock, favorite, hard rock
34	-0.02	0.016	christian, christian rock, contemporary christian, worship, gospel, piano, praise & worship, praise and worship, praise, kids
35	-0.02	0.011	vinyl i own, new age, favourite albums, instrumental, spoken word, lopez 88, yanni, croatian pop, jazz-soul-funk, piano
36	-0.03	0.016	progressive rock, 1970, progressive, space rock, art rock, symphonic rock, 1987, regine-disc, psychedelic rock, psychedelic
37	-0.03	0.023	laptop, i have this album, usa, 1001 albums you must hear before you die, male singer songwriter, american musician, female singer songwriter
38	-0.03	0.013	smooth jazz, jazz guitar, saxophone, smooth, 1960's, cash, country taggradio, 2000, instrumental, tammy wynette taggradio
39	-0.03	0.03	metal, metalcore, thrash metal, progressive metal, alternative metal, nu metal, death metal, hardcore, deathcore, groove metal
40	-0.04	0.013	live, reggae, roots reggae, debut, m haggard, chill, 2001, dancehall, dub, jamaica
41	-0.04	0.018	funk, classic country, latin grammy nominated, jazz-funk, nelson, classic, 1998, donald byrd, wishlist, c adderley
43	-0.05	0.021	singer-songwriter, folk, acoustic, mellow, sundaymix, folk rock, my private work station, records i own, 1965, the supremes
45	-0.06	0.013	experimental, industrial, industrial metal, industrial rock, gammarec, avant-garde, 1982, post-rock, freepurple, 1975
46	-0.06	0.011	2011, best of 2011, 1974, honky tonk, c atkins, 10s, top stuff, wb recording, temptations
47	-0.07	0.034	80s, new wave, post-punk, 1985, 1986, 1984, 1980s, synthpop, 1983, 1988
48	-0.07	0.013	easy listening, 2009, top cd, best of 2009, albums i own on vinyl, desert island discs, 2014 albums
49	-0.08	0.022	70s, 1973, 1971, 1977, 1978, 1970s, 1974, 1972, 1975, lp
50	-0.11	0.012	latin, latin pop, reggaeton, 1988, salsa, latin grammy nominated, spanish, crooners, trap, vocal jazz
51	-0.12	0.010	irish, celtic, allboutguitar, brc blues band karlsruhe, allbout guitar lessons - blues workshops karlsruhe
52	-0.14	0.024	blues, blues rock, instrumental, grammy nominated, classical, opera, guitar, piano, blues-rock, bluezzz
53	-0.16	0.013	2013, vinyl, best of 2013, selbst, 10s, doo wop, liblistened, o liebert, jfuzz, gl4m
54	-0.17	0.020	female vocalists, happy music for happy people, rock n roll, love songs, 1983, female vocalist, female, bette midler, sweet
55	-0.20	0.018	jazz fusion, fusion, guitar, guitar virtuoso, jazz rock, instrumental, bass, instrumental rock, chick corea, corea
56	-0.21	0.031	indie, indie rock, 2008, alternative, indie pop, 2007, rock, british, 00s, alternative rock

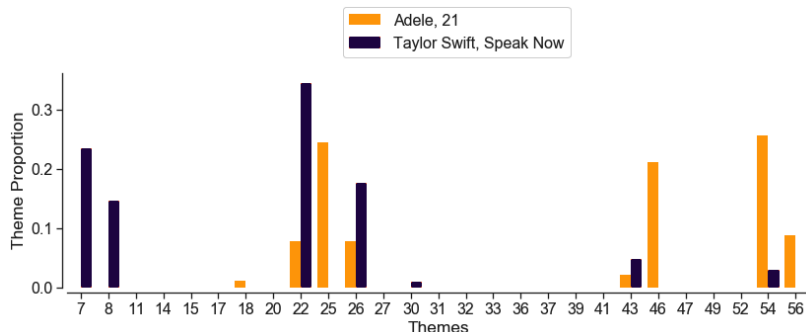* These 42 themes account for 92% of all the themes of the entire tags.

.

Figure 12: Theme proportions for *21* by Adele and *Speak Now* by Taylor Swift (themes with weights less than 0.001 are not displayed in the graph)

Themes Proportions for One Album   Recall that an album is a mixture of themes, and therefore, the bag-of-tags representation of an album draws from different themes. Figure 12 shows the theme proportions for *21* by Adele and *Speak Now* by Taylor Swift. The theme proportions provide striking descriptions of these studio albums. The two albums draw upon themes 22 and 26, but with different weights. Theme 22 is strongly associated with successful music of the 2000s, and theme 26 is associated with pop rock and adult contemporary. The two albums differ on a number of themes. For example, *Speak Now* is strongly associated with theme 7, which is about modern country music. Themes 25, 46, 54 and 56 are strongly associated with Adele's *21*. Theme 25 refers to Motown which is one of the influences of Adele's previous album *19* and exhibits soul music inflections that are very present in *21*. Theme 54 is about female vocalists and love songs (the album displays a range of emotions — bitterness and moving forward — after a broken relationship). Theme 56 refers to indie rock and pop of the 2010s.

Having described the qualitative insights that can be generated from our results, we now look at how our model can be used for marketing decisions of different types.

Album & Playlist Scoring, Recommendation and Design

We illustrate the use our model for (1) forecasting the performance of new albums, (2) Album diagnostics and tuning, (3) recommending albums that are consistent with the musical styles

associated with specific areas, (4) designing contextual playlists, and (5) recommending musically similar albums.

Forecasting the Performance of New Albums   Consider a new album that is yet to be released. The proposed nonparametric model can be used to predict its latent score y^* and the probability $p(y^* \geq 0) = 1 - \Phi(-F(x, a, w)/\sigma)$ that it will appear on the charts (Φ denotes the cdf for the unit Normal distribution). To evaluate the model's performance for this purpose, we re-estimated the proposed model and two competing models using data up to 2015 and used them to predict the charting probabilities for the 751 albums (551 charted, 200 uncharted) in our dataset that were released in 2016. We used the state equation 6 to project forward the time-varying components of our model (i.e., the linear effects and the linear basis of the splines). We used the themes that are estimated on the training data and assigned each tags of the test albums to one of these themes in a semi-supervised fashion (no new themes were considered).

Table 4 reports the average predicted probability of an album appearing on the charts using the proposed nonparametric model (M_4) and the two competing parametric models, the latter using the best PSIS-LOO value, to forecast the performance of the 2016 albums. It shows these probabilities separately for the albums that appeared and did not appear on

Table 4: Charting probabilities predicted by the three models

			Best Parametric	
Statistic	Album Type	Nonparametric (M_4)	Dynamic	Static
$p(y^* \geq 0)$	Charting	0.86	0.84	0.85
	Non-charting	0.66	0.73	0.71
AUC	All albums	0.80	0.76	0.75

the charts. The nonparametric model predicted a marginally higher charting probability (0.86) for the charting albums than the two competing models (0.84 and 0.85), and also a lower charting probability (0.66) for the non-charting albums than the competing models (0.73 and 0.71). The AUC (area under the curve) measure suggests that the nonparametric

model performed better (0.80) than the two parametric models (0.76 and 0.75) in classifying the albums into charting and non-charting categories.
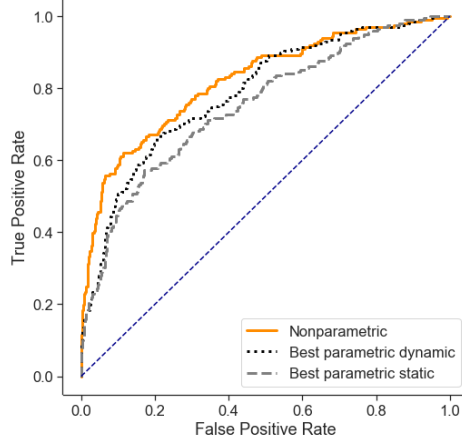

Figure 13: Receiver operating characteristic (ROC) curves of the three models

Figure 13 exhibits the Receiver Operating Characteristic (ROC) curves of the three models. A model with a higher curve has greater predictive power. The figure shows that our model has the best predictive power among the three models.

We can use different probability thresholds to predict if an album appears on the charts. A threshold value of ϑ means that an album is predicted to appear on the charts if a model predicts its charting probability to be at least ϑ. A threshold of $\vartheta = 1$ means that no album is predicted to appear on the charts. Correspondingly, the hit rate is zero for the 2016 albums that appeared on the charts and one for albums that did not appear on the charts. As the threshold value decreases, the hit rate increases for the charting albums and decreases for the non-charting albums. The lowest threshold displayed is $\vartheta = 0.5$, for which all three models have a hit rate of almost one for charting albums and low hit rates for non-charting albums. The nonparametric model performs the best: given a hit rate for a charting album, it predicts the highest hit rate for a non-charting album across the three models. The threshold $\vartheta = 551/75 = 0.734$, the proportion of successful 2016 albums in our data set, corresponds to the probability with which we would predict an album to be successful without any other information. The corresponding hit rates obtained by (1) the
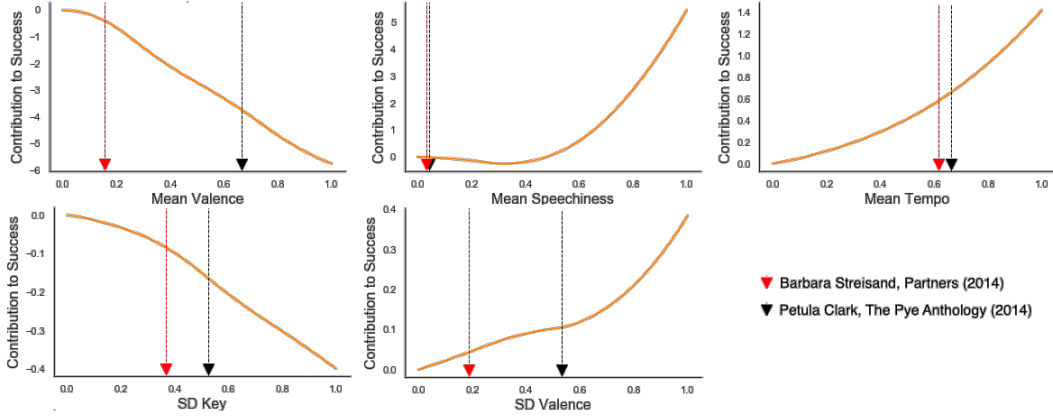
Figure 14: Acoustic diagnostic of Barbara Streisand's Partners (2014) and Petula Clark *The Pye Anthology* (2014) for musical success for 2014

nonparametric model are 0.94 for charting albums and 0.55 for non-charting albums, (2) the dynamic parametric model are 0.95 for charting albums and 0.35 for non-charting albums, and (3) the static parametric model are 0.90 for charting albums and 0.45 for non-charting albums. The highest overall hit rates for the three models are 0.84 ($\vartheta = 0.734$) for the nonparametric model, 0.80 ($\vartheta = 0.761$) for the best parametric dynamic model, and 0.79 ($\vartheta = 0.696$) for the best prametric static model.

Album Diagnostic & Tuning Our model can be used for album diagnostic, tuning and balancing. For an illustration, we consider two albums released in 2014, *Partners* by Barbara Streisand and *The Pye Anthology* by Petula Clark. The nonparametric model correctly predicted that *Partners*, which has an estimated (actual) log-score of 7.82 (7.60), would appear on the Billboard 200 and that *The Pye Anthology*, which had an estimate score of -0.34, would not.

Figure 14 shows the values of five acoustic components for the two albums and their contributions to album success[10]. Both albums have low scores on mean speechiness (absence of spoken words in their songs) and high scores on average tempo. *The Pye Anthology* has more emotional variety, a higher mix of happy and sad songs, and features songs with higher

---

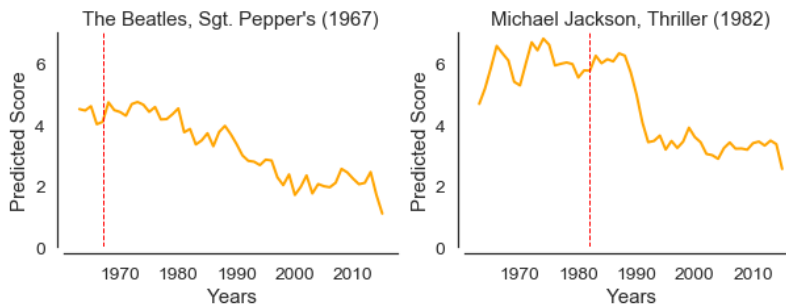[10]The remainder of the acoustic features are provided in the web appendix

Figure 15: Scores across different years for the albums *Sgt. Pepper's Lonely Hearts Club Band* by The Beatles and *Thriller* by Michael Jackson

variety of keys. *Partners* has lower average valence (on average, sadder songs). This results in a gap of about 20 points on the orginal year-end scale between the two albums. Our model suggests that Clark's album would have a higher probability of appearing on the charts if it had a lower average valence.

Scoring Albums and Playlists for Different Generations of Listeners As noted, Holbrook and Schindler (1989) found that people's preferences for popular music are formed in early adulthood. We can use the proposed model to predict the score for an album or playlist in different years and recommend it to listeners whose musical tastes match the periods in which it has high scores.

To illustrate, consider the albums *Sgt. Pepper's Lonely Hearts Club Band* by The Beatles ans *Thriller* by Michael Jackson. Figure 15 shows the predicted scores for each album in years before and after their release. Our model suggests that *Sgt. Pepper's Lonely Hearts Club Band* is most appealing to people whose musical tastes were formed in the sixties, seventies and eighties. Its score begins to decline after 1980; we would not recommend it to the present generation of listeners. Similarly, we would recommend *Thriller* to listeners whose taste gravitates to the music of the seventies, eighties and nineties, but not to listeners who grew up after 1990.

We also used our model to score a playlist recommended by a website, themortonreport.com, that specializes in culture and entertainment. A weekly column on the website,
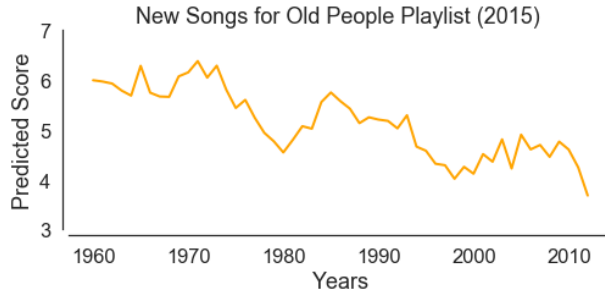
Figure 16: Scores across different years for the *New Music for Old People* playlist

entitled "New Music for Old People," recommends music to "*fill the gap for those of us who were satiated musically in the '60s and then searched desperately as we aged for music we could relate to and get the same buzz from nowadaze.*"[11] Table 5 shows the list of songs recommended on April 17, 2015 column. All songs, other than Billy Swan's *Don't be cruel*, were released in the 2000s.

Table 5: *New Music for Old People* playlist

Song Name	Artist/Group	Release Year
Run	Delta Rae	2015
Honey	The London Souls	2015
I Can See for Miles	Vanilla Fudge	2015
This Is a Man s Man s Man s World	Gov't Mule	2015
Glimmer (Orchestral)	Neil Young	2014
Don't Wait	Mapei	2014
Nice and Slow	Max Frost	2013
Long Haul	Michael McDonald and Robben Ford	2013
Tonight Insomnia	Eye to Eye	2005
Don't Be Cruel	Billy Swan	1973

We collected and used the information on the acoustic features to score the ten songs and the playlist for all 54 years. Figure 16 shows that the estimated score for the playlist indeed does peak in the 1960s. It decreases with time, reaching a low in 1980, after which it rises again for a brief period until 1985, before decreasing again.

[11]http://www.themortonreport.com/entertainment/music/new-music-for-old-people-the-london-souls-neil-young-govt-mule-delta-rae-and-more/

37

Contextual Recommendations Tags contain information on the context in which a song or album is heard and the mood it evokes. We can use this information to recommend albums and playlists for particular contexts. Contextual playlists have become increasingly popular on platforms like Spotify (see Joven 2018). A context may be an activity (e.g., running, romantic dinner) or a time-related event (e.g., new year, birthday). A user queries a recommendation system by specifying tags that describe a context, and may also indicate the importance of a tag (which is converted to a weight; otherwise, each tag is weighted equally). The response is a matching playlist.

We convert each query to a bag of 100 tags, duplicated in proportion to their weights. Then we extract the corresponding thematic representations by sampling the thematic allocations of each tag in the contextual query (without updating the thematic profiles of the training albums). We retrieve the thematic membership of the query after 100 iteration.

Let ρ_{ik} denote the weight of theme k for album i, and ρ_{qk} the weight of theme k for query q. Let S denote a target set of albums. We use the Hellinger distance (Nikulin 2001)

$$ H(i,q) = \left(\frac{1}{2} \sum_{k=1}^{K} (\sqrt{\rho_{ik}} - \sqrt{\rho_{qk}})^2 \right)^{\frac{1}{2}} $$

to measure the closeness of album $i \in S$ to query q, and recommend albums with the smallest $H(i,q)$ values in S. We illustrate the method for two contexts.

The first context considers a listener looking for a playlist for a romantic evening. Suppose the user selects the tags "romantic," "love," "smooth" and "vocalist" and assigns them equal weights. The second context considers an individual who wants music recommendations for a dance party and selects "disco," "dance," "fun" and "party" as four equally-weighted tags. Table 6 shows the recommendations obtained from our model for each context. The first context implies three main themes. The first theme has a weight of 0.57 and relates to romantic and energetic pub rock; the second theme has a weight of 0.24 and relates to love music of the late 70s; and the third theme has a weight of 0.19 and relates to smooth old jazz. The three albums that are thematically closest to this context are *Dimensions* by The

Box Tops, *Back Home* by Caedmon's Call and *La Semana* by Ottmar Liebert. The second

Table 6: Contextual Recommendations

Context 1: Romantic, Love, smooth, vocalist	
Themes (top 5 tags)	**Weight**
fully streamable albums, romantic, pub rock, fun, energetic	0.57
1979, love, love it, american idol, now [..] radio	0.24
smooth jazz, jazz guitar, saxophone, smooth, 1960s	0.19
Closest Three Albums	
The Box Tops, *Dimensions*; Caedmon's Call, *Back Home*; Ottmar Liebert, *La Semana*	

Context 2: Fun, Party, Disco, Dance	
Themes (top 5 tags)	**Weight**
electronic, disco, dance, electronica, synthpop	0.55
fully streamable albums, romantic, pub rock, fun, energetic	0.45
Closest Three Albums	
The Call, *Into the Woods*; Tangerine Dream, *Summer in Nagasaki*; Black Box, *Dreamland*	

context relates to two main themes: an electronic and disco music theme with a weight of 0.55, and a romantic, fun and energetic music theme with a weight of 0.45. The three albums that are thematically closest to this context are *Into the Woods* by The Call, *Summer in Nagasaki* by Tangerine Dream and *Dreamland* by Black Box. Observe that the two contexts use different tags, yet our model yields a common theme. But the thematic memberships differ for the two contexts, which results in the recommendation of different albums.

"*Because you liked...*" Type Recommendations Consider a listener who likes a particular album. We can use the proposed model to recommend other albums that are acoustically and/or thematically similar.

We measure the distance between two albums (or two playlists) as the average of their acoustic and thematic distances. The acoustic distance is measured by the Euclidean distance between acoustic profiles, which contain information on the means and the standard deviations of the fingerprints of the songs on an album. The thematic distance is measured by the Hellinger distance $H(i, i')$, which compares the probability distributions over the themes

Table 7: Top five thematically and acoustically similar albums to *Thriller* (1982)

Acoustic Similarity		
Album	Artist	Distance
I Feel For You (1984)	Chaka Khan	0.13
Aliens Ate My Buick (1988)	Thomas Dolby	0.15
Dream Street (1984)	Janet Jackson	0.16
Heroes of the Harvest (2001)	Arrested Development	0.16
Raise! (1981)	Earth, Wind & Fire	0.16
Thematic Similarity		
Album	Artist	Distance
Bad (1987)	Michael Jackson	0.18
Just Gets Better With Time (1987)	The Whispers	0.27
Control (1986)	Janet Jackson	0.30
Rhythm Nation 1814 (1989)	Janet Jackson	0.32
Love Is The Place (1982)	Curtis Mayfield	0.33
Thematic & Acoustic Similarity		
Album	Artist	Distance
Bad (1987)	Michael Jackson	0.21
Just Gets Better With Time (1987)	The Whispers	0.24
Dream Street (1984)	Janet Jackson	0.26
Triumph (1980)	The Jacksons	0.26
Richard Marx (1987)	Richard Marx	0.27

appearing in a pair of albums, i and i'.[12]

Table 7 shows the top-five albums that are acoustically and/or thematically similar to Michael Jackson's *Thriller*. These lists have face validity: most of these albums were released in 80s and reflect the style of music most popular at that time; and listening to these albums makes it apparent that these suggestions are indeed meaningful. The acoustically-similar albums are different from the thematically-similar albums. One thematically-similar album, *Bad*, was also recorded by Michael Jackson; and *Control* and *Rhythm Nation 1984*, were recorded by Janet Jackson, who has a similar musical style.[13] The other two thematically-

---

[12]We normalize the acoustic distance so that it ranges between 0 and 1, to be consistent with the range of the Hellinger distance, before computing the average.

[13]Michael and Janet Jackson collaborated on several songs. Their single *Scream* (1995) received a Grammy award for best Music Video.

similar albums are *Just Gets Better With Time* by the Whispers and *Love Is The Place* by Curtis Mayfield. The top three thematically-similar albums appear on the list of albums that are both thematically and acoustically similar to *Thriller*.

Playlist Compilation   Our model can be used to compile playlists. To illustrate, suppose our task was to construct a playlist of five songs selected from Adele's albums *19*, *21* and *25*. We could choose her five greatest hits. But our model suggests choosing a different mix of songs tailored to match the preferences of people whose musical tastes better match one or another time period.

Table 8: Adele's top five Billboard songs and the five songs generated by the model for listeners from different eras

Billboard's Top 5 Songs		Model Generated Compilations			
–		1973		2016	
Album	Song	Album	Song	Album	Song
21	Rolling In The Deep	19	Cold Shoulder	21	Take It All
19	Chasing Pavements	25	Million Years Ago	21	I'll Be Waiting
21	Someone Like You	21	Rolling in the Deep	21	One And Only
25	Hello	21	Don't You Remember	21	Lovesong
19	Hometown Glory	21	One And Only	21	Someone Like You

We used the acoustic features of the songs to score all possible playlists consisting of five songs selected from the two albums. We scored the playlists separately for 1973 and 2016. Table 8 shows the highest scoring playlists for each year, and Adele's top five hits according to Billboard magazine critics[14]. The playlists have one song each from Adele's five greatest hits (Rolling in the Deep and Someone Like You), and have only one common song (One And Only). Figure 17 compares the 1973 and 2016 playlists in terms of the means and standard deviations of the acoustic fingerprints. The 1973 playlist has higher mode and more acoustic songs, and the 2016 playlist has greater variation in the liveness of the songs.

We can also use the proposed model to create playlists by merging albums of different

---

[14]https://www.billboard.com/articles/columns/pop/8448730/adele-songs-best-ranked-critics-picks
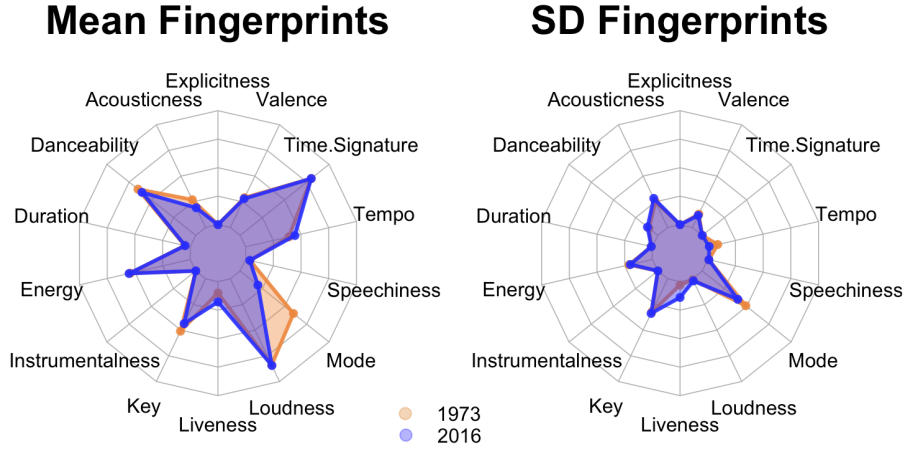
Mean Fingerprints   SD Fingerprints

Figure 17: Means and standard deviations of acoustic fingerprints for the 1973 and 2016 playlists

Table 9: *Thriller* and *1999*'s top five songs merger generated by the model for listeners from different eras

2000		2016	
Album	Song	Album	Song
Thriller	Baby Be Mine	Thriller	Baby Be Mine
Thriller	The Lady in My Life	Thriller	Billie Jean
1999	1999	Thriller	The Lady in My Life
1999	Little Red Corvette	1999	1999
1991	D.M.S.R.	1999	Little Red Corvette

artists. To illustrate, consider selecting ten tracks from *Thriller* (by Michael Jackson) and *1999* (by Prince) for group of listeners with tastes corresponding to different eras. One group has a taste for the music of 2000 and the other a taste for the music of 2016. Table 9 shows the list of five songs with the highest success scores for each group. The 2000 playlist contains two songs from *Thriller* and three songs from *1999*. The 2016 playlist contains three songs from *Thriller* and two songs from *1999*. The two playlists have four common songs: *Baby Be Mine*, *The Lady in My Life*, *1999*, and *Little Red Corvette*.

Conclusion

We developed a novel multimodal machine learning framework to predict the success of musical assortments. Success is modeled in terms of acoustical features that determine the felt experience of consumers, genres of the songs within the assortment, and themes that summarize the semantic content of user-generated tags. We estimated the model on *Billboard 200* data on album success for the past five decades. Our model uses different types of Bayesian nonparametric components including penalized splines for flexibly capturing the effect of the acoustics and a novel supervised HDP component to infer themes that are predictive of success. The model also includes dynamics that capture the evolving impact of the genres and acoustics over the years.

Our model provides several insights into the evolution of American popular music. Some of our findings relate to broad trends in the popularity and decline of different genres and the emergence of new forms of music. For instance, Rock has remained the most successful form of music in America, although its popularity has steadily declined; and pop music is almost as successful as hip hop, which took off in the early 80s and appears to have peaked in the 2000s. It also uncovers how different acoustics have been instrumental in album success over the years.

The supervised hierarchical Dirichlet process component of our model uncovered a number of themes. These themes, that are probability distributions over the tags, represent how listeners perceive and experience the albums. Themes emphasize different types of tags and can be characterized as focusing on sub-genres (e.g., soft rock, new soul, jazz vocal), consumption contexts (e.g., holiday, worship, party), emotions (e.g., love, anger), intensity (e.g., relaxing, kill for peace, must hear), nostalgia (e.g., oldies, good times), mood (e.g., smooth, energetic, ambient), quality of listening experience (e.g., virtuoso, deep lyrics, easy listening), type of experience (e.g., seen live, own on vinyl) and other aspects of music. Together they offer a vivid summary of how listeners experience and perceive music.

We demonstrated the predictive superiority of our model by comparing its performance to

that of 13 competing specifications. Together these benchmark models span static, dynamic and nonparametric alternatives. We also showed how the time-varying parameters in our model can be used for forecasting the success of new albums before they are launched.

In a series of applications we demonstrated the useful of our model for a variety of marketing tasks. These include 1) recommending albums that are consistent with the musical styles associated with specific eras or generations of listeners, 2) fusing different albums to create playlists for specific tastes, 3) designing contextual playlists based on queries that users may make to a recommendation engine, 4) recommending albums that are similar to a given album, 5) forecasting the performance of new albums, before they are released, and 6) diagnosing album performance and tuning albums based on the acoustical features that may promote or inhibit success.

Methodologically, we developed a novel nonparametric framework that captures the information contained in data of different modalities. Our use of a supervised hierarchical Dirichlet process for modeling text is new to the marketing literature, and our integration of different styles of nonparametrics in a single machine learning model is also a contribution to the statistics and computer science.

Although constructed specifically for music, our model can also be used in other experiential contexts where multi-modal data are available. These include the consumption of intangible goods such as movies, videos, podcasts, theatre and other contexts, where the felt experience of the user is important to capture so as to predict the success of different items. In each of these contexts, different types of textual data is also available, such as movie scripts, description of scenes, user-generated tags and reviews, that can be leveraged via variants of our model to predict success.

While we obtained a number of substantive insights and showcased several marketing applications of our model, it is also important to acknowledge some limitations of our research. Our results describe how popular music has evolved over time, but the use of observational data means that we cannot assert causality about the effects of the covariates. While we have extensively validated our models on holdout data of different types, we have not tested

our recommendations in a real-life context. Finally, although our model can be used for constructing playlists for listeners with different types of preferences, we cannot recommend individualized playlists as we do not have access to individual-level data. However, it is relatively straightforward to capture such sources of individual heterogeneity if such data were to be available. We modeled the themes and the acoustics additively in our framework. It is possible that they may have an interactive impact. We leave such an investigation to future research. Finally, the transition dynamics within our state-space framework can be modeled in terms of observed covariates that drive the evolution of parameters. Unfortunately, we did not have access to such data, as so have not modeled such effects. We look forward to working on these extensions and we also hope that researchers will use our modeling framework in other experiential contexts.

References

Alexander, Peter J. (1996), "Entropy and popular culture: product diversity in the popular music recording industry," *American Sociological Review*, 61 (1), 171–174.

Anand, Narasimhan and Richard A. Peterson (2000), "When market information constitutes fields: Sensemaking of markets in the commercial music industry," *Organization Science*, 11 (3), 270–284.

Ansari, Asim and Carl F. Mela (2003), "E-customization," *Journal of Marketing Research*, 40 (2), 131–145.

—— and Raghuram Iyengar (2006), "Semiparametric thurstonian models for recurrent choices: A Bayesian analysis," *Psychometrika*, 71 (4), 631.

Askin, Noah and Michael Mauskapf (2017), "What makes popular culture popular? product features and optimal differentiation in music," *American Sociological Review*, 82 (5), 910–944.

Bertin-Mahieux, Thierry, Douglas Eck, Francois Maillet and Paul Lamere (2008), "Autotagger: A model for predicting social tags from acoustic features on large music databases," *Journal of New Music Research*, 37 (2), 115–135.

Biyalogorsky, Eyal and Naik, P.(2003), "Clicks and mortar: the effect of on-line activities on off-line sales," *Marketing Letters*, 14 (1), 21–32.

Bradlow, Eric T. and Peter S. Fader (2001), "A Bayesian lifetime model for the "Hot 100" Billboard songs," *Journal of the American Statistical Association*, 96 454), 368–381.

Bradlow, Eric T. and Vithala R. Rao (2000), "A hierarchical Bayes model for assortment choice," *Journal of Marketing Research*, 37(2), 259–268.

Bruce, Norris I (2019), "Bayesian Nonparametric Dynamic Methods: Applications to Linear and Nonlinear Advertising Models," *Journal of Marketing Research*, 56 (2), 211–229.

Blei, David M, Ng, Andrew Y and Jordan, Michael I (2003), "Latent dirichlet allocation," *Journal of machine Learning research*, 37(2), 993–1022.

Bruce, Norris I. (2019), "Bayesian nonparametric dynamic methods: Applications to linear and nonlinear advertising models," *Journal of Marketing Research*, 1–19.

Bruner, Gordon C. (1990), "Music, mood, and marketing," *Journal of Marketing*, 54 (4), 94–104.

BuzzAngle (2018), "Buzzangle music 2017 U.S. report," https://www.buzzanglemusic.com /wp-content/uploads/buzzangle-music-2017-us-report.pdf.

Chung, Tuck Siong, Roland T. Rust and Michel Wedel (2009), "My mobile music: An adaptive personalization system for digital audio players," *Marketing Science*, 28 (1), 52–68.

Dai, Andrew M., and Amos J. Storkey (2015), "The supervised hierarchical Dirichlet process," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 243–255.

Datta, Hannes, George Knox and Bart J. Bronnenberg (2017), "Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery," *Marketing Science*, 37 (1), 5–21.

Dew, Ryan and Asim Ansari (2018), "Bayesian nonparametric customer base analysis with model-based visualizations," *Marketing Science*, 37 (2), 216–235.

DiMA (2018), "Dima -digital media association, annual music report," https://dima.org/wp-content/uploads/2018/04/dima-streaming-forward-report.pdf

Dowd, Timothy J. (2004), "Concentration and diversity revisited: Production logics and the us mainstream recording market, 1940–1990," *Social Forces*, 82 (4), 1411–1455.

Eliashberg, Jehoshua, Jedid-Jah Jonker, Mohanbir S. Sawhney, and Berend Wierenga, "MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures," *Marketing Science*, 19 (3), 226–243.

Elberse, Anita (2010), "Bye-bye bundles: The unbundling of music in digital channels," *Journal of Marketing*, 74 (3), 107–123.

Farquhar, Peter H and Vithala R. Rao (1976), "A balance model for evaluating subsets of multiattributed items," *Management Science*, 22 (5), 528–539.

Giles, David E. (2007), "Increasing returns to information in the us popular music industry," *Applied Economics Letters*, 14(5), 327–331.

Heinrich, Gregor (2011), "Infinite LDA implementing the HDP with minimum code complexity," *Technical note, Feb*, Citeseer 170.

Holbrook, Morris B. and Elizabeth C. Hirschman (1982), "The experiential aspects of consumption: Consumer fantasies, feelings, and fun," *Journal of Consumer Research* 9 (2), 132–140.

Holbrook, Morris B. and Punam Anand (1990) , "Effects of tempo and situational arousal on the listener's perceptual and affective responses to music," *Psychology of Music* 18 (2), 150–162.

—— and Robert M. Schindler (1989), "Some exploratory findings on the development of musical tastes," *Journal of Consumer Research*, 16 (1), 119–124.

IFPI (2018), "Global music report 2018, annual state of the industry," https://www.ifpi.org/news/ifpi-global-music-report-2018.

Iwanaga, Makoto (1995), "Relationship between heart rate and preference for tempo of music," *Perceptual and Motor Skills*, 81 (2), 435–440.

Jovev, Jason (2018), "Spotify: The Rise of the Contextual Playlist," *Medium*, March 13, https://blog.chartmetric.io/spotify-the-rise-of-the-contextual-playlist-c6f2c26900f4.

Juslin, Patrik and N. Daniel Västfjäll (2008), "All emotions are not created equal: Reaching beyond the traditional disputes," *Behavioral and Brain Sciences*, 31 (5), 600–621.

Juslin, Patrik N. and Petri Laukka (2004), "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of New Music Research*, 33 (3), 217–238.

Kim, Jin Gyo, Ulrich Menzefricke, and Fred M. Feinberg (2004), "Assessing heterogeneity in discrete choice models using a Dirichlet process prior." *Review of Marketing Science*, 2 (1).

——, Ulrich Menzefricke and Fred M. Feinberg (2007), "Capturing flexible heterogeneous utility curves: A Bayesian spline approach," *Management Science*, 53 (2), 340–354.

Koelsch, Stefan, Thomas Fritz, D. Yves v. Cramon, Karsten Müller and Angela D Friederici (2006), "Investigating emotion with music: an FMRI study," *Human brain mapping*, 27 (3), 239–250.

Krueger, Alan B. (2005), "The economics of real superstars: The market for rock concerts in the material world," *Journal of Labor Economics*, 23 (1), 1–30.

Lafata, Alexia (2014), "Is this your song? The science behind what determines your taste in music," https://www.elitedaily.com/life/culture/determines-music-taste/641213.

Lee, Jonathan, Peter Boatwright and Wagner A. Kamakura (2003), "A Bayesian model for prelaunch sales forecasting of recorded music," *Management Science*, 49(2), 179–196.

Lena, Jennifer C. (2006), "Social context and musical content of rap music, 1979–1995," *Social Forces*, 85(1), 479–495.

Lena, Jennifer C. Mark C. Pachucki (2013), "The sincerest form of flattery: Innovation, repetition, and status in an art movement," *Poetics*, 41 (3), 236–264.

Li, Yang and Asim Ansari (2013), "A Bayesian semiparametric approach for endogeneity and heterogeneity in choice models," *Management Science*, 60 (5), 1161–1179.

Naik, Prasad A. (2015), "Marketing dynamics: A primer on estimation and control," *Foundations and Trends in Marketing*, 9 (3), 175–266.

Nave, Gideon, Juri Minxha, David M. Greenberg, Michal Kosinski, David Stillwell and Jason Rentfrow (2018), "Musical preferences predict personality: evidence from active listening and facebook likes," *Psychological Science*, 29 (7), 1145–1158.

Nikulin, M. S. (2001), "Hellinger distance," *Encyclopedia of Mathematics*, Springer.

Ocasio, William, Mauskapf, Michael and Steele, Christopher W. (2016), "History, society, and institutions: The role of collective memory in the emergence and evolution of societal logics," *Academy of Management Review*, 41 (4), 676–699.

Papies, Dominik and Harald J. van Heerde (2017), "The dynamic interplay between recorded music and live concerts: The role of piracy, unbundling, and artist characteristics," *Journal of Marketing*, 81 (4), 67–87.

Peterson, Richard A. and David G. Berger (1975), "Cycles in symbol production: The case of popular music," *American Sociological Review*, 40 (2), 158–173.

Pietroluongo, Silvio (2012), "How we chart the year," https://www.billboard.com/articles/news/1481472/the-best-of-2012-the-year-in-music.

RIAA (2018), "Mid-year 2018 RIAA music revenues report," https://www.riaa.com/ wp-content/uploads/2018/09/RIAA-Mid-Year-2018-Revenue-Report-News-Notes.pdf.

Rossi, Peter (2014), *Bayesian Non- and Semi-parametric Methods and Applications*, Princeton University Press.

Rossman, Gabriel (2012), *Climbing the charts: What radio airplay tells us about the diffusion of innovation*, Princeton University Press.

Rubel, Olivier and Naik, P. (2017), "Robust dynamic estimation," *Marketing Science*, 36 (3), 453–467.

Ruppert, David and Wand, M. P. and Carroll, R. J. (2003), *Bayesian Semiparametric Regression*, Cambridge University Press.

Scherer, Klaus R. (2004), "Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them?" *Journal of New Music Research*, 33 (3), 239–251.

Schmitt, Bernd H (2010), *Customer experience management: A revolutionary approach to connecting with your customers*, John Wiley & Sons.

Schonfeld, Zach (2019), "Albums are getting shorter than ever — and the Grammys are catching up," https://uproxx.com/music/grammys-albums-shorter-than-ever-nominees/

Serrà, Joan, Àlvaro Corral, Marià Boguñá, Martín Haro, and Josep Ll Arcos (2012), "Measuring the evolution of contemporary western popular music," *Scientific Reports*, 2, 521.

Sethuraman, Jayaram (1994), "A constructive definition of Dirichlet priors," *Statistica Sinica*, (4) 2, 639–650.

Shively, Thomas S., Greg M. Allenby and Robert Kohn (2000), "A nonparametric approach to identifying latent relationships in hierarchical models," *Marketing Science*, 19 (2), 149–162.

Smith, Adam N., Peter E. Rossi and Greg M. Allenby (2018) , "Inference for product competition and separable demand," *Marketing Science*, Forthcoming.

Starr, Larry and Christopher Alan Waterman (2018), *American popular music from minstrelsy to MP3*, Oxford University Press, USA.

Sun, Dongchu, Robert K. Tsutakawa and Paul L. Speckman (1999), "Posterior distribution of hierarchical models using CAR(1) distributions", *Biometrika*, 86 (2), 341–350.

Swets, John A. (2014), *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*, Psychology Press.

Teh, Yee W., Michael I. Jordan, Matthew J. Beal and David M. Blei (2005), "Sharing clusters among related groups: Hierarchical Dirichlet processes," *Advances in Neural Information Processing Systems*, 1385–1392.

Tirunillai, Seshadri, and Gerard J. Tellis (2014), "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation," *Journal of Marketing Research*, 51 (4) 463–479.

Tobin, James (1958), "Estimation of relationships for limited dependent variables," *Econometrica*, 26 (1), 24–36.

Ullman, Jeffrey David (2011), *Mining of massive datasets*, Cambridge University Press.

Vehtari, Aki and Gelman, Andrew and Gabry, Jonah (2017), "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and computing*, 27 (5), 1413–1432.

Vogel, Harold L. (2014), *Entertainment industry economics: A guide for financial analysis*, Cambridge University Press.

Wand, MP and JT Ormerod (2008), "On semiparametric regression with O'Sullivan penalized splines," *Australian & New Zealand Journal of Statistics*, 50 (2), 179–198.

Wedel, Michel and Jie Zhang (2004), "Analyzing brand competition across subcategories," *Journal of Marketing Research*, 41 (4), 448–456.

Zentner, Marcel, Didier Grandjean and Klaus R. Scherer (2008), "Emotions evoked by the sound of music: characterization, classification, and measurement," *Emotion*, 8 (4), 494.