Modeling & Optimization for Streaming Listening Demand

Kobi Abayomi

Abstract. Digital delivery of songs has radically changed the way people can enjoy music, the sort of music available for listening, and the manner by which rights holders are compensated for their contributions to songs. Subscribers can enjoy an unlimited potpourri of songs and sounds, uniquely free of incremental acquisition or switching costs. This shift reveals listening patterns governed by affinity, boredom, attention budgets, etc. Listening patterns can be driven instantaneously, dynamically, organically or programmatically (playlists, for example). Listening demand is in a new paradigm, with a commensurate change in revenue implications. These new listening phenomena deprecate past orthodoxy around content curation in which a listener made a single purchase of a song. This point-of-sale model is now insufficient: demand revenue is proportional to song affinity—e.g., by how often a song is listened to within a time interval—and can be modeled as a time dependent process. We explore modeling digital *on-demand* demand and employ a fully Bayesian probabilistic model that: (1) yields estimators for multi-level effects on song demand and (2) naturally joins with multi-stage Linear Optimization scheme to optimize the same.

In loving memory of M. Atiim Abayomi

1. "NILE, MAKE ME A HIT LIKE BOWIE'S": A SIMPLE DYNAMIC MODEL FOR STREAMING SONG LISTENING. This paper focuses on modeling demand for a "song"—generally a two to five minute musical composition—consumable via some digital delivery service or *Digital Streaming Provider* (DSP) and the strategic, macroscopic, business-useful inferences that could then be deduced from elucidation of some assumptions around the demand for those songs. We place song-level listening demand in a modeling framework—i.e., with estimable effects—that can then be managed in an optimization or demand maximization scheme.

Nile Rodgers, in an interview about his influence on popular music of the 1970s and 1980s recalled this exchange with Miles Davis:

> Miles would always ask me to make him a hit like how I did for [David] Bowie. I never took him seriously until he covered [Cyndi Lauper's] '*Time after Time.*' I listened to that track and realized he was serious, and like most artists, wanted as many people to hear him as possible. [1]

Q1

In this paper we address the *macroscopic* dynamics of song listening via a probabilistic model. We focus on the aggregate demand dynamics of a population, or subpopulation, enjoying a song as a function of time, aggregating (if not fully eliding) any individual or group-wise utility—here the *probability* of listening—into larger groupwise aggregate demand.

It is natural to use these sub-populations—or *audience segments*—as the device to collect the effects on listening as effects on demand. Music industry profession-

doi.org/10.1080/00029890.2024.2410142
MSC: 62

als, A&R[1] executives, say, may think of an audience segment as a micro-genre. This stratification is familiar in the advertising technology literature [2].

Streaming demand as a counting process. To begin, but without loss of generality, we consider a "song" a *de novo* offering: a new, or new version of a composition yielding a demand curve with a fixed point at (0,0). Time zero is just as a song is released—or, in the parlance, *dropped*. Zero observable demand at the release date.[2]

In similarity with [3], the model for volume of listening, or listener response to a listening "opportunity," is a counting process where any individual listener, i, enjoys a song with a (not-necessarily) time variant probability $P_{t,i}$. In this paper capital letters represent random processes, lower case letters refer to observed or observable values. We rely on the random processes as *U-statistics* [4] for measurability, and other, assumptions.

$$U_{t,i} = \begin{cases} 0 \; with \; probability \; P_{t,i} \\ 1 \; with \; probability \; 1 - P_{t,i}. \end{cases} \tag{1}$$

We can think of the time dependent dis-aggregated $\{P_{t,i}\}_{t=1,\ldots T}$'s as an *affinity* curve, say, for an individual listener i. An aggregation of these as *realized* utilities is given by

$$Y_t^j = \sum_{i=1}^{N_t^j} U_{t,i}, \tag{2}$$

which are the cumulative realized individual listening affinities within each *listening strata*, N_t^j, with $i \le j \le J$, yielding a song-level demand curve

$$Y_t = \sum_{j=1}^{J} Y_t^j. \tag{3}$$

Let i be an individual listener and let the listeners be divided into J (not necessarily disjoint) time-varying listening strata, $\{N_t^1, \ldots, N_t^j, \ldots, N_t^J\}$, for time $1 \le t \le T$. Then we let n_t^j denote the *size* of the jth stratum at time t, that is, $|N_t^j| = n_j^t$. Another way to state this is that the comprehension of listening stratification—as a function for measurement of the listening population—exploits subadditivity, e.g.,

$$\left| \bigcup_j N_t^j \right| \le \sum_j n_t^j. \tag{4}$$

Thus equation (4) conveys the non-disjointedness of "listening mode": strata may overlap, listeners may be compelled by (or marketed to via) multiple affinities. Segregation

[1] The A&R, Artist & Repertoire, department at a music label is typically tasked with discovering and curating music talent.

[2] On without loss of generality: in practice a song may be "pre-released"—not available for listening but in a potential listeners queue—or an available song that is in a demand trough. This need not affect the model specification here, but the audience stratification and effects within each strata might be different. The marketing strategies and listener demand response for a song that is seasonally popular, say, may as well be unique.

of listener affinity can be as idiosyncratic as the set of rules (or some of the set of rules) which increment a song as a fully "listened to" stream on a particular DSP (30 seconds of listening, say, on a particular provider), or as intuitive as merely on which DSP a song is enjoyed. These curves are models for listener preference, over time, for coherent—but not necessarily identical—patterns of listening demand or consumption unique to the "listening mode."

Typically these curves are not disaggregated by practitioners. Top line demand (number of streams, cumulative number of streams) is often illustrated; investigation of other stratifications for disjoint or overlapping segments—demographic, listening mode (playlist, first-time, library, etc.)—is uncommon. Inability to identify unique listeners can frustrate segment identification and assignment: this is one reason why segment-wise inspection is less common. Within definite intervals and for listeners who *are* identifiable (pre-saves, for example) non-disjoint segmentations are not difficult to create.

The simplest possible non-disjoint segmentation is to merely separate cumulative streaming listening curves by DSP. "DSP-wise" differences in listening affinity are well known and frequently observed. This is a first step to more particular segmentation of listening affinities: the benefit of being able to segment demand cannot be understated. Coherent audience segmentation maps the effect of either ambient or local conditions which induce a listener to listen. One way to convey this is to say that any listener, at any time, may be exposed to (and listen to) any song for any reason—in fact, multiple reasons. These can be expected to be dynamic; only a certain sort of person listens to Christmas music in July, for example.

Figure 1 shows a cartoon graphic of listening demand for a song over a 40 week interval. Curve height is the number of listeners weekly listeners on each DSP, say; the colored curve aggregates listeners within unique subscription services. The black curve is the total and the overall demand curve. This illustration would be familiar to music industry executives and/or artists: an important heuristic for modeling song performance is that *it should be clear that a song performs differentially* (over time) on different platforms. DSPs can appeal to different audiences, with possibly different listening preferences; each DSP may offer variegated subscription plans, which may appeal to listening preferences heterogenously.

The aggregate curve in Figure 1—in black at the top—conveys a slow steady growth in listening demand. The other curves, on audience (sub) segments illustrate the differential listener affinities (at least, on different DSPs). This sort of rich, differential, picture of demand is invaluable to modern content rights management.

Other counting process for streaming demand. Content rights holders typically receive intermediated information on listener demand, via the DSPs, in a way that is similar to data scientists in advertising technology. To account for this "schmutzdecke,"[3] and in place of a completely naive observed data model, modeling the extremal processes can yield inference. Let

$$Y_t^+ = \bigwedge_{N_t^1,\ldots,N_t^J} \sum_{i=1}^{n_t^j} U_{t,i} \tag{5}$$

a boundary process, on the best possible audience strata—i.e., with maximum listening affinity. And let

[3]From a past as an environmental statistician, an intermediate filtering layer.

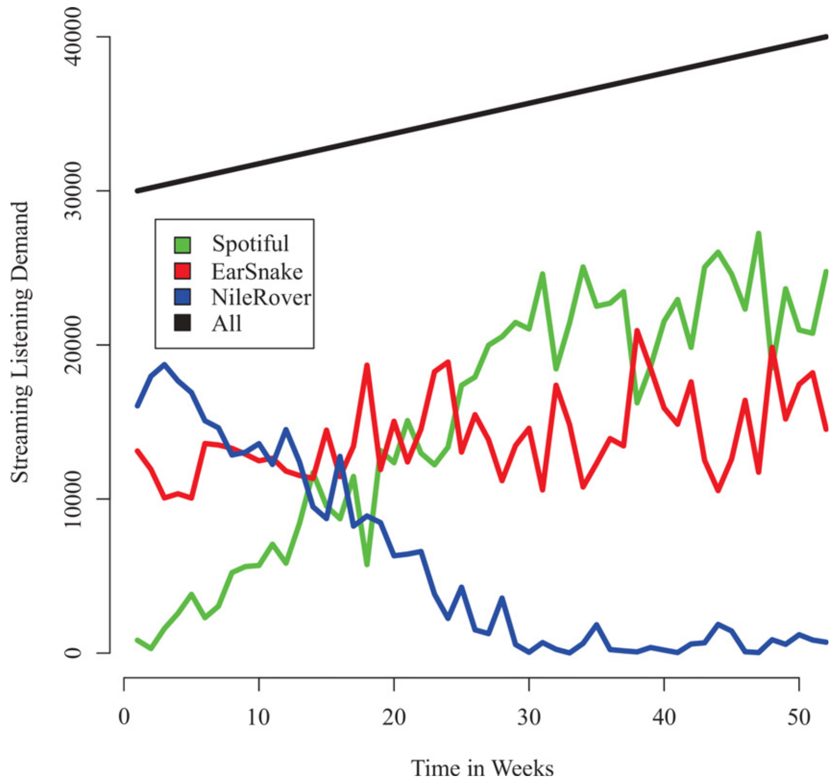


Figure 1. Illustration of song demand over time. Each curve, Y_t^j, is a illustration of demand for a particular stratum. Here the audience strata are just each DSP: the number of people listening to the song on a DSP at a coincident time within an agreed upon time. In this cartoon J = (Spotiful, EarSnake, NileRover) (three made-up DSPs). Each convey different demand patterns. One can imagine an example narrative: the demand dropped precipitously on NileRover, built slowly on Spotiful and was steady on EarSnake. These explanations are lost if only the black—total curve—is inspected. Behind each unique demand curve there is differential performance of the song over time, thus differential listening affinity, thus differential response to the song itself and marketing for the song. These strata could be playlisting behavior, or overlapping demographics—each are important to song marketing. Any listener can be in multiple strata.

$$Y_t^- = \bigvee_{N_t^1,\ldots,N_t^J} \sum_{i=1}^{n_t^j} U_{t,i} \tag{6}$$

be the lower boundary.

Content rights holders are concerned with song performance—*and the ability to characterize a song's performance*—in the presence of confounding factors: temporality, ambient head or tailwinds, DSP idiosyncrasy, bad luck, etc. There are many hard to quantify explanations for song performance. Fixing Y^+ and Y^- as the extremal demand processes, *with respect to the process model*, can yields stable comparative models for performance characteristics. The extremal processes can offer insight in cases where data (or metadata) to meaningfully stratify cumulative demand are unavailable. In Section 3 of this paper we illustrate a model (the *envelope model*) amenable to minimal or maximal listening strata (Figure 2). Q2

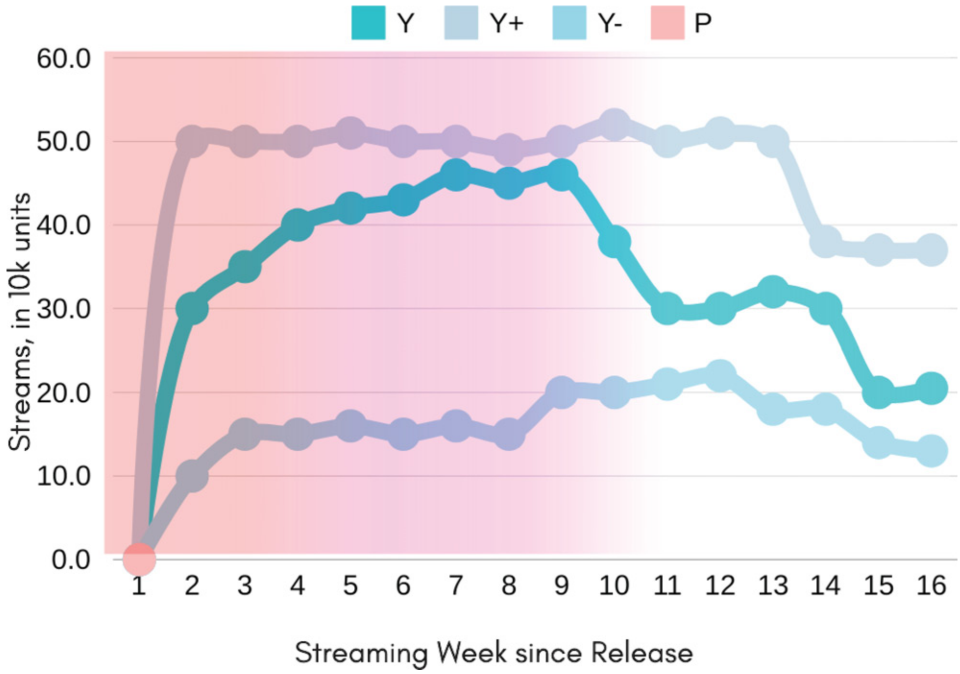


Figure 2. Illustration of processes for song demand over time. The curves—the max value process, the observed demand curve and the minimum value process—are envelopes for the expected demand over time. Here, the graph is shaded by the "temperature" of the underlying aggregate affinity process $P_{t,i}$. Affinity for the song begins to "cool" in week 8.

Model for listener affinity. To move from a counting process to a probability model for listener affinity we impose a minimal probabilistic assumption (later, this is a constraint in the optimization scheme) on the affinities, $P_{t,i}$, which naturally yields a Bernoulli distribution for the utilities, $U_{t,i}$, as:

$$P_{t,i\in j} = \boldsymbol{\theta}^j \mathbf{x}_t + \boldsymbol{\gamma}^j \mathbf{z}_t \tag{7}$$

$$U_{t,i\in j} \sim Ber(\boldsymbol{\theta}^j \mathbf{x}_t + \boldsymbol{\gamma}^j \mathbf{z}_t). \tag{8}$$

Covariates for exogenous or ambient effects on demand are collected in $\mathbf{z}_{t,j}$; those for endogenous effects (marketing, complementary media, social media, etc.) are collected in $\mathbf{x}_{t,j}$. Another modeling trick is to assume (and constrain in the optimization scheme) the C and D dimensional covariates are nonnegative such that: $\mathbf{x} \in [0, 1]^C$, $\mathbf{z} \in [0, 1]^D$. This is just to elide effects that depress song listening, i.e., we are not accounting for *dislike* of a song or sound.

Song demand via listening mode. Figure 3 is a plot and characterization of observed demand curves for 1,000 *de novo* songs, with demand curves observed in calendar year 2021, on a popular streaming service. The demand curves were classified by $k = 6$ mean centroid classification via the Python `tslearn` toolkit to illustrate similarities in types of song demand curves.

Figure 3 points to varied *modes* for listening and song demand: song demand peaks and decays with regular, differentiable characters. Modeling the incidental processes

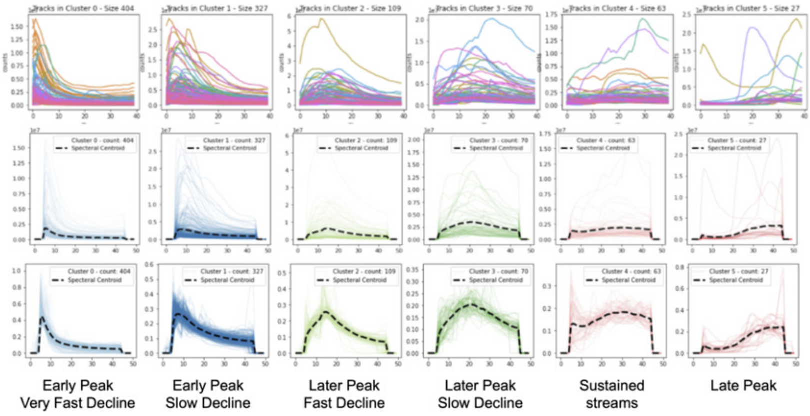


Figure 3. Illustration of *modes* of song demand, from observed song demand on a popular streaming service in calendar year 2021, via time-series clustering. Each *de novo* demand curve was translated to (0, 0), i.e., release date vs. zero number of listeners to start. Time is incremented in weeks. As processes, each curve (type) traces the number of listeners in each week. The top row are samples of songs within each (column) time-series cluster. The middle and bottom rows are lower and upper boundaries within each cluster. Captions at the bottom of each column convey an interpretation of the demand pattern for each cluster. Successful partitioning of listener types can yield empirically disjoint or differentiable curve types [20]. Each column are centroids of time series curves that we interpret as categories of types of listening demand curves. The pattern of young people listening to a song may follow an "early peak with slow decline," for example, while an older demographic may follow a "later peak with slow decline."

U_t through to the extremal process curves, Y_t^+, Y_t^- lets the model be flexible for the available data granularity.

The importance of audience segmentation. One can think of an audience segment as a listening group which responds similarly to listening stimuli (at a particular time); within each segment we model the utilities as i.i.d.—random but identically distributed. The $\{N_t^j\}_{j=1,\dots,J}$ are non-disjoint because individual listeners may occupy more than one utility for listening (at a particular time) a particular song.[4] The ability to segregate demand as unique audience segments and model differences in effects is important.

Contemporary work on streaming demand [3, 5] elides listener level utility with aggregation, perhaps as user level data are hard to come by. The audience segmentation device in this paper joins varied hierarchical level listening demand data with listener level utility models [6, 7]. This resonates with the both the spirit of [8] and the similarities in theoretical process models they derive and both they and we observe in data.

With a natural probability model in hand for listening demand curves we can now address the estimation problem directly. A music marketer should be, mainly, interested in the estimation of $\boldsymbol{\theta}^j$, that is, the effect of marketing on music demand. Including data which allows the $\boldsymbol{\gamma}^j$ to be estimated allows a marketer to control for ambient effects, competitive releases, etc.

2. "JUST BECAUSE A RECORD HAS A GROOVE, DON'T MAKE IT IN THE GROOVE": COVARIATE MODELS FOR PROCESSES & FORECASTING.[5]

Within any coherent audience segment N_t^j, the estimator for segment-wise affinity can be accessed via a logistic model,

[4] N.B. that the time index for streaming demand modeling can be coarse, where each increment is one week.
[5] From the lyrics to "Sir Duke" [9].

$$\hat{\mathbb{P}}(U_{i,t} = 1) = logit^{-1}\{\boldsymbol{\theta}^j \mathbf{x}_{t,i} + \boldsymbol{\gamma}^j \mathbf{z}_{t,i}\} = \hat{P}_{t,i}. \tag{9}$$

As usual this is a well known model for a binary process: here where a song listen is realized. Straightaway the estimators for effects of business levers ($\boldsymbol{\theta}$) via observed data covariates ($\mathbf{x}$) or ambient effects ($\boldsymbol{\gamma}$) via ($\mathbf{z}$) can be modeled using individual, user level data where available. Where these data aren't available—for example Apple Music's API does not offer granular, user level data—one can use segment-wise counts and covariates and then can appeal to the extremal counting process models.[6] For example, for observed data demand curve y_t, for audience segment j, the distribution of the size of the audience strata is

$$\mathbb{P}(N_t^j = n_t) = \binom{n_t - 1}{y_t - 1} P_{t,i}^{y_t}(1 - P_{t,i})^{n_t - y_t}. \tag{10}$$

The Negative Binomial distribution relates the demand curves' observed value, y_t to the size of the listening strata N_t *in terms of the covariates* as P_t is covariate dependent. More straightforwardly, Poisson or Negative Binomial regression can specify the effects of the covariates on the demand curves.

Fully Bayesian workflow for streaming demand. Here it is important to invoke a modeling perquisite: translating the songs to a time-demand interval beginning at $(0, 0)$. This condition is met if data for release dates and listening demand beginning from release are available. This condition is not always necessary, nor it is necessarily sufficient. Consider a model forecasting demand behavior for a song in *deep catalog*: a song that was released many years ago. We illustrated in Figures 1–3 the *growth-decay* character of listening demand for *de novo* songs; these demand patterns may exist within several alternate or similar periodic behaviors.

For example, when an audience segment of young listeners discover Stevie Wonder: the mode of growth and decay of listening can be similar, for this strata, to a new release. A forecaster who wants to consider aggregate future demand for a re-release of Stevie's *Jesus Children of America*, say, can't rely fully on only the dynamics of *de novo* songs by comparable artists or even Stevie Wonder himself but *within strata* the assumption is tenable and *across stratum* models are fit on the convolution.

This is a fully Bayesian setup for collecting, training, estimating, and updating the model(s) for streaming demand (see [**10**]). This setup co-ordinates demand response, covariate information, and metadata in a framework that is useful for monitoring and gauging song performance in-the-moment and yields a full-distributional tableau for a subsequent optimization scheme as well.

In what follows, we elucidate two versions of Bayesian models which capture listener stratum and artist level effects—accounting for differences in utility, say, among the effects on listeners who enjoy only the unique rhythm guitar, organ and synthesizer on the Ohio Players' single *Ecstasy* and those who have an ear for it in the rest of the album.

We can think of both models as "forcing" models: the first—the 'Null Model'—in the sense that the effects of covariates on listening affinity "force" audience segment-wise demand. The second—the "Envelope Model"—conveys these same effects, via

[6] For example, a straightforward extremal segment is first time listener.

the probability model, but mediated by structural equations for the growth and decay of listening demand for any song.

Null model These effects in this model are time-invariant and the model itself only accounts for time dependent effects via the value of the predictor processes. This model does not intermediate the covariate effects on listener affinity within audience segments. One can imagine an affinity process where covariate effects "row" and/or "decay" differentially across non-disjoint strata, yielding overall growth/decay curve modes similar to those in Figure 3. But there is only one Bob Marley.[7] In practice, where an audience segment is identifiable—say via high resolution user information, or for songs & artists where listening affinities persist—we recommend using the Null model for effect estimation and optimization.

Null Model

$$
\begin{aligned}
y_t^{j[a]} &\sim NegBin(e^{\boldsymbol{\theta}^{j[a]}\mathbf{x}_{t,i}+\boldsymbol{\gamma}^{j[a]}\mathbf{z}_{t,i}}, \omega^{j[a]}) \\
\theta &\sim Normal(\boldsymbol{\mu}_a^x, \boldsymbol{\Sigma}_a^x) \\
\gamma &\sim TruncNormal(\boldsymbol{\mu}_a^z, \boldsymbol{\Sigma}_a^z) \\
\boldsymbol{\Sigma}_a^x &\sim LkjCorr(\eta_a^x) \\
\boldsymbol{\Sigma}_a^y &\sim LkjCorr(\eta_a^z) \\
\eta_a^x &\sim \chi^2(\tau^x) \\
\eta_a^y &\sim \chi^2(\tau^z) \\
\omega_j &\sim \Gamma(\alpha_a, \beta_a); \{\alpha_a, \beta_a\}_{a\in A}\ constants.
\end{aligned}
\tag{11}
$$

ADSR/forced model The illustrations in Figure 4a and 4b picture a forcing, or phase shift model, that we find useful to convey covariate effects through while simultaneously capturing common growth-decay song demand phenomena.
Forced (envelope) Model

$$
\begin{aligned}
y_t^{j[a]} &\sim NegBin(\mathbb{E}(Y_t), \omega^{j[a]}) \\
\mathbb{E}(Y_t) &= \alpha_r^{j[a]} + \beta_r^{j[a]} \cdot t \\
\theta^a &\sim Normal(\boldsymbol{\mu}_a^x, \boldsymbol{\Sigma}_a^x) \\
\gamma^a &\sim TruncNormal(\boldsymbol{\mu}_a^z, \boldsymbol{\Sigma}_a^z) \\
\boldsymbol{\Sigma}_a^x &\sim LkjCorr(\eta_a^x) \\
\boldsymbol{\Sigma}_a^z &\sim LkjCorr(\eta_a^z) \\
\eta_a^x &\sim \chi^2(u^x) \\
\eta_a^y &\sim \chi^2(u^z) \\
\omega_j &\sim \Gamma(\alpha_a, \beta_a); \\
&\quad \{\alpha_a, \beta_a\}_{a\in A}\ constant
\end{aligned}
\qquad
\begin{aligned}
\alpha_r^{j[a]} &= \begin{cases} |\alpha| \geq 0\ , & r \leq \tau_A \\ |\alpha| \approx 0\ , & \tau_A \leq r \leq \tau_S \\ |\alpha| \leq 0\ , & \tau_S \leq r \leq \tau_D \\ |\alpha| \approx 0\ , & \tau_D \leq r \leq \tau_R \end{cases} \\
\beta_r^{j[a]} &= \begin{cases} |\beta| \geq 0\ , & r \leq \tau_A \\ |\beta| \approx 0\ , & \tau_A \leq r \leq \tau_S \\ |\beta| \leq 0\ , & \tau_S \leq r \leq \tau_D \\ |\beta| \approx 0\ , & \tau_D \leq r \leq \tau_R \end{cases} \\
\tau_A &\sim \frac{1}{T-2} \\
\tau_{D,S,R} &\sim \frac{1}{T-2}\sum_{t=2}^{T}\frac{1}{T-t}.
\end{aligned}
\tag{12}
$$

[7] Or Michael Jackson, or Jan Hammer, or KraftWerk.

The ADSR model is a Bayesian Hierarchical Model for "always on" prediction of streaming demand with change points and phase shift forcing. Listener strata are indexed $\{1, \ldots, J\}$ as before. Vector valued estimators for endogenous and exogenous predictors enter the first level of the hierarchy *via the linear equations in equation (12)*. These are now the *main effects* per each subspace of this model. Contrast the null model: the main effects are simply the covariate coefficients.

This phase shift model has four phases: A attack or growth; D decay; S sustain; R release. The change points for each phase can be estimated simultaneously or before the remainder of the posterior for y_t (here the prior is Restricted Uniform—see [**11**]). Figure 4 is an illustration of the ADSR model.

This model fixes growth-decay conditions on segment-wise counting processes. In this version of the model the main effects estimators, for the utility forcings, are estimated as projection on subspaces of a phase transition model and in this way mediated or attenuated depending upon the phase of the process. One reason for *not* treating this as a fully Gaussian Process with a Latent Variable [**12**] is that the generating processes here are only Gaussian in a large numbers regime. Starting from first principles here yields distributional inference even for less popular songs and artists, i.e., that stretch the Gaussian assumption on the feature space [**13**,**14**]. In practice we found this model to be useful for prediction of segments that we could not discriminate across with metadata: in particular the extremal demand processes.

3. "WHIP IT": FULLY OPTIMIZING LISTENING DEMAND. "Whip It," a single by the new wave group DEVO on their 1980 album *Freedom of Choice* is memorable for its synthetic C-G-D chord chorus punctuated by a 5-4 synthesizer suffix that is immediately and famously recognizable [**15**]. When James Ambrose Johnson, Jr.—the legendary Rick James—was working at the end of 1980 on "Super Freak" with Alonzo Miller he was aware of and a fan of DEVO's single. In *Bitchin': The Sound & Fury of Rick James* James was looking to imbue his track—already a likely hit with an unforgettable walking blues-ish bassline and doubled piano chords throughout—with a sound that, as he put it, would get him a new audience:

> James insisted on that 5-4 DEVO sound. He wouldn't relent. He said he knew it would get him a new wave audience, a white audience. What the kids were listening to. He fought me and I put it in. He was right. [**16**]

It is important to note that any song, while a work of art, can be regarded via its constitution from parts, each having a possible effect on listening behavior. Producers and dj's are aware of these differential affinities when they mix songs live or in studio, when they search for a hook or break that has just the right sound for the audience they intend the song for. Inasmuch as the models written here can encode, measure, and be optimized for marketing levers, so can sound information encoded on partitions of a song be measured against a dictionary of segmented audience response. Optimization for listening affinity can address the tuning of a song as well.[8]

Recall that $|N|$ is the total audience available for a song; fix it constant for each time t over the period $\{1, \ldots, T\}$; T is usually quite large, each t often a week. Recall that the $\{N_t^j\}_{0 \leq j \leq J}$ form a *non-disjoint covering* for N; individual listeners i may be in more than one audience segment (at a time) N_t^j. The audience segment covering permits a differential response to marketing strategies $\mathbf{x_t}$, say, and ambient events $\mathbf{z_t}$ that affect listening affinity—within each equal time interval t—via effects $\boldsymbol{\theta}^j$ and

[8]This happens often *post hoc*, for example when a song is sped up, slowed down, remixed or the well known conversions to Musak.

$\boldsymbol{\gamma}^j$. Conversationally, the audience segment covering $\{N_t^j\}_{0\le j\le J}$ conveys the *audience segment-wise reason* at a particular time for listening: one time during exercise, another time in an algorithmic playlist of new songs, another time to prepare for sleeping.

Any budget for listening—from the perspective of the listener—is a function of the utility curves' $\{U_{t,j}\}_{0\le j\le J}$ response to marketing or ambient impulses $\mathbf{x}_t, \mathbf{z}_t$—*i.e., the magnitude of the coefficients* ϕ *and* ψ—and models incremental listening as membership in a different audience segment (e.g., listeners' ability to listen for a different reason).[9] The impacts of endogenous & exogenous forcings are conveyed via the individual listening utilities, i.e., *realized probabilities*. The final piece to consider is what the equations for process maximization, for either model, are.

Null model. Consider the maximization of listening under the null model, where the sole dynamic is listener affinity. From equation (7) the user level utility curves are a function of endogenous and exogenous dynamics via effects, respectively, $\mathbf{x}_{t,j}, \mathbf{z}_{t,j}; \boldsymbol{\theta}^j; \boldsymbol{\gamma}^j$—i.e., spend per marketing channel, impulse per social channel, demand per marketing spend, and demand per social channel.

Let the endogenous budget B (the amount of money the rights holder has to spend through T) for a song be:

$$B = \sum_t B_t = \sum_t \mathbf{1}^T \mathbf{x}_t \tag{13}$$

with $\mathbf{1}$ a vector of ones the same length as $\mathbf{x}$. This is just to say that the rights holder has a finite & necessarily and wholly exhaustible budget for endogenous forcing.

Maximization of Null Model.

$$\begin{aligned}
\max \mathbb{E}U_{t,i\in j} = \max P_{t,i\in j} = \max_{\mathbf{x}_t} \boldsymbol{\theta}^j\mathbf{x}_t + \boldsymbol{\gamma}^j\mathbf{z}_t& \\
s.t.& \\
\boldsymbol{\theta}^j\mathbf{x}_t + \boldsymbol{\gamma}^j\mathbf{z}_t &\le 1 \\
\boldsymbol{\theta}^j\mathbf{x}_t + \boldsymbol{\gamma}^j\mathbf{z}_t &\ge 0 \\
\mathbf{1}^T\mathbf{x}_t &\le B_t \\
\mathbf{1}^T\mathbf{z}_t &\le S \\
\mathbf{x}_t &\ge \mathbf{0} \\
\mathbf{z}_t &\ge \mathbf{0}.
\end{aligned} \tag{14}$$

Above is the maximization scheme for the Null model. Maximization of the expected utility for any listener, audience-segment-group-wise is equivalent to maximizing the probability of listening within a segment. The probability term must remain a probability; the budget across channels at a time t is constrained by the total budget available at t. Assume that marketing spend and social buzz can only increment.

A program for the maximization of expected utility for a listener within a particular segment j at time window t is in equation (14). The maximal input for the path, as a function of time, is derived from the Lagrangian for the optimization scheme in (14):

[9]This is an important distinction between the song and utility of listening to it at a particular time, for a particular reason. From the perspective of the listener this a model for listening choices; from the perspective of the inventory holder (song creator or curator) it is a model for song demand.

$$\mathbf{x}^*_{t,i\in j} = \begin{cases} B_t[\boldsymbol{\theta}^j]^{-1} \ \ where \ \ 0 < B_t \leq (\mathbf{1} - \boldsymbol{\gamma}^j \mathbf{z})[\boldsymbol{\theta}^j]^{-1} \\ (1 - \boldsymbol{\gamma}^j \mathbf{z}^t)[\boldsymbol{\theta}^j]^{-1} \ \ where \ \ B_t > (1 - \boldsymbol{\gamma}^j \mathbf{z})[\boldsymbol{\theta}^j]^{-1} \end{cases} \tag{15}$$

where $[\cdot]^{-1}$ is a vector pseudo-inverse. This is to take the maximum of either the scaled available budget B_t, or the scaled residue beyond the endogenous effects $\mathbf{z}$; each "scaled" by the relative effect of endogenous—or business-wise levers—on the song utility, *within each audience segment*. In practice, the budget can be reallocated across audience segments—and it should be—to follow the (estimated) effect for greatest gain in audience magnitude.

ADSR model. Remember, the forced model imposes a pattern, or a template of, over-arching listening affinity (or song uptake). Refer again to Figure 3. The use of the forcing model is to exploit the regular patterns in aggregate song demand with a model that reduces the inference burden while increasing the explanatory power. Here, we use the ADSR or *envelope model*—common to the sound engineering literature as a model for the intensity of a sound over time [**17**], and a well-known *generative* tool for modifying a sound. Statistically this model is a special case of a *phase transition model* (see [**18**])—characterized by discontinuities between the phases at the transitions. Referring to the elucidation in equation (12) this model is fit in two steps:

I **Fit the change points.** The four phases of the ADSR model yield 3 change—or discontinuity—points. These can be fit *a priori*, prior to the fully Bayesian estimation of the remainder of the model parameters, or either *a priori* or jointly via the restricted uniform distributional specification in [**19**].

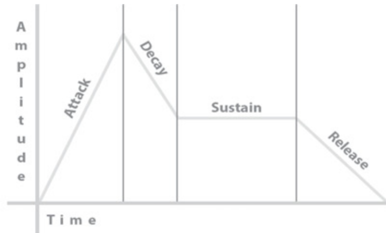
II **Fit the partite models.** Each phase of the ADSR model is essentially linear: the parameters to be fit are the slopes and intercepts for each linear part. The effects between the endogenous and exogenous covariates, the distributional hyperparameters for dependency between, and precision of those effects—each of those are parameters to fit within each phase.

The model is designed to capture dynamics for *de novo* songs—songs new to an audience of listeners,[10] yet is flexible to serve for songs with varied observed release times and listener exposure.

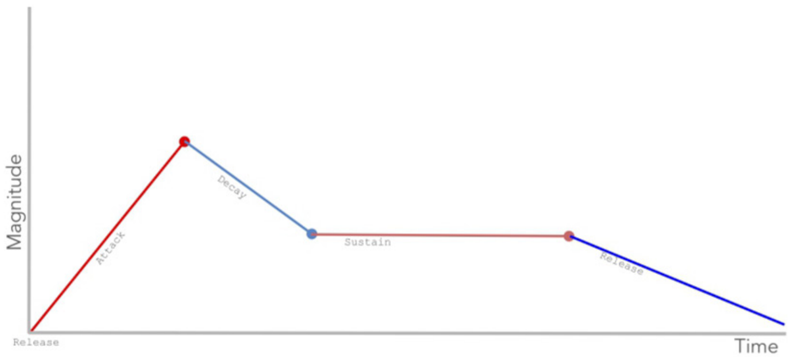
In the forcing model the endogenous and exogenous effects are estimated jointly with the partite linear model parameters. This is simply to say that the model flexibly estimates the effect on listener affinity within audience segment and subject to the growth/decay phase of the song, given the ADSR model.

The equations in (12) now specify a Bayesian hierarchy similar to the unforced model but with estimators for effects θ, γ that are constant within phase. This simplifies the maximization scheme. For example, in phase $[I]$ the maximum expectation is at time t_A, within this phase the estimating equations for effect are $\alpha = 0$ & $\beta = \frac{\mu_{t_A}}{t_A}$. The mean value function in this phase, μ_{t_A} is defined as in the unforced model.

[10]To borrow jargon from advertising technology, the *in-flight* period for an advertisement is the length of time an advert is placed within media for *impressions*.



(a) ADSR model for individual sound



(b) ADSR model for aggregate listening demand

Figure 4. Comparative illustrations of processes for song demand over time. In Figure (a) the model is illustrated as typically used in a Digital Audio Workstations (DAW). In Figure (b) the model is applied to the "*in-flight*" for a *de novo* song from release date. This is a special case of a *phase-transition model* [18]; the discontinuities here (at the nodes with enlarged circles) are where we fit partite models for each phase.

Maximization of Forcing Model, *at phase extrema.*

$$
\begin{aligned}
&[I] \quad \mathbb{E}(Y_t) = \frac{\mu_{t_A}}{t_A} \cdot t \\
&[II] \quad \mathbb{E}(Y_t) = \frac{\mu_{t_A} t_S - \mu_{t_S} t_A}{t_S - t_A} + \frac{\mu_{t_S} - \mu_{t_A}}{t_S - t_A} \cdot t \\
&[III] \quad \mathbb{E}(Y_t) = \frac{\mu_{t_S} t_D - \mu_{t_D} t_S}{t_D - t_S} + \frac{\mu_{t_D} - \mu_{t_S}}{t_D - t_S} \cdot t \\
&[IV] \quad \mathbb{E}(Y_t) = \frac{\mu_{t_D} t_R}{t_R - t_D} - \frac{\mu_{t_D}}{t_R - t_D} \cdot t
\end{aligned}
\tag{16}
$$

Above is the maximization scheme for the ADSR model. Maximization of the expected utility for any listener, audience-segment-group-wise is equivalent to maximizing the probability of listening within a segment, *which is equivalent to maximizing each of these equations at their rightmost point*. As the mean value function for each phase has a constant first derivative, the maximal path **x** is constant within phase. The budget across channels at a time t is constrained by the total budget available at t. Again we assume that marketing spend and social buzz, etc., can only increment positively (Figure 5).

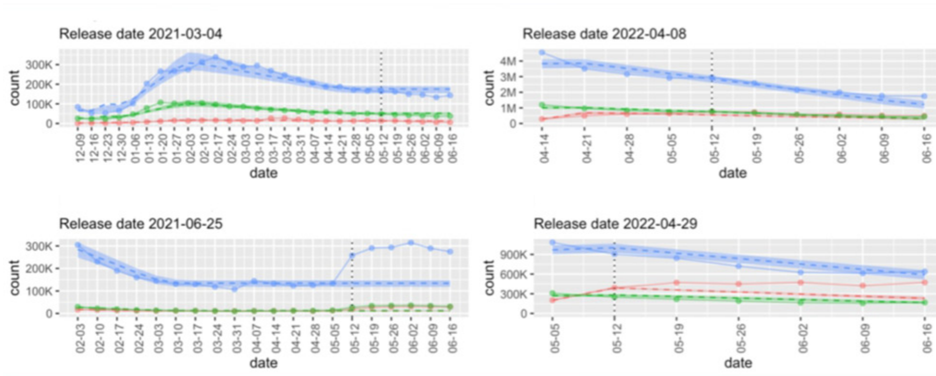


Figure 5. Illustration of demand curves estimation for audience segmented songs. Audience segments at each song are illustrated in colored line; shaded regions are predicted posterior intervals. Vertical dotted lines are optimal intervention times.

4. COMMENTS AND RECOMMENDATIONS. Either of these models should "fit" nicely within current rights holder management schemes. Either model can be dynamically instantiated—in particular the phased/forcing model—with a simple LP. The forcing model needs only (linear) estimators for the mean value function at the change of phase after the change points themselves are estimated. Knowledge of these estimators—especially for this model—make a straightforward optimal path for listening maximization.

Time scales for marketing in aural media are discrete. Typically song performance is evaluated from week-to-week; advertising & social campaigns can be adjusted weekly. Optimization schemes work well on a portfolio of assets. Use of either version of these models on a suite of assets is preferable. It is conceivable that estimators for marketing or ambient effects on listening affinity trade or switch magnitude and sign across time periods, e.g., Halloween music, Christmas music.

An innovation shared by both the null and forcing models is to simply be willing to segregate the sources of (listening) demand and keep track of the marketing actions within each segment to yield usable time-aware effect estimators. Zooming out: audience segmentation for listening demand is key, perhaps even more for sound media demand than visual. The differential effects of marketing & exposure to a sound once it is observed are not difficult to measure. This paper is an argument for the importance of— and illustrates how to model and optimize over—these song specific effects *differentially across* different listener preferences (in different ways at different times, etc.).

ACKNOWLEDGMENT. The author thanks his colleagues—the many hardworking data, optimization, sound & machine learning scientists at the Georgia Institute of Technology (GT), Warner Music Group (WMG) & Warner Media (WM). At WM: Vivek Vasudeva, Xinxin Zhu, Emanuel Hof & Haile Owusu. At WMG: Daniel Lee, Julien DeMori, Guannan Zhao & Kaidy Guo. Additional thanks to Brian Magerko & Jason Freeman—both of GT. As well, Yifeng Yu of the Music Information Program at GT (via his advisor Alexander Lerch) whose vital collaboration on similar work will appear in a sequel. Lastly, the author dedicates this paper to his father, M. Atiim Abayomi, whom the world lost October 2023. Atiim shared his love of all things musical—from Sylvester to Schubert—openly. He inspired this work. He is sorely missed.

DISCLOSURE STATEMENT. No potential conflict of interest was reported by the author. Q3

REFERENCES

[1] Diggin' In The Crates With Nile Rodgers. Season 4, Episode 8. Available from: https://www.youtube.com/watch?v=Uy70I5KbHno

[2] Carbajal J, Williams P, Popescu A, Chaar W. Turner Blazes a trail for audience targeting on television with operations research and advanced analytics. INFORMS J Appl Anal. 2019;49(1):64–68.

[3] Ivaldi M, Nicolle A, Verboven F, Zhang J. Displacement and complementarity in the recorded music industry: evidence from France. J Cult Econ. 2023;48:43–94.

[4] Hoeffding W. Probability inequalities for sums of bounded random variables [Ph.D. thesis]. University of North Carolina Department of Statistics; May 1962.

[5] Ordanini A, Nunes J, Nanni A. The featuring phenomenon in music: how combining artists of different genres increases a song's popularity. Mark Lett. 2018;29:485–499.

[6] Tam K, vanTilburg W, Chan C, Igou E, Lau H. The boredom feedback model. Personal Social Psychol Rev. 2021;25:3.

[7] Wojtowicz Z, Chater N, Loewenstein GF. Boredom and flow: an opportunity cost theory of motivational attention. 2019 March 13. Available at SSRN: https://ssrn.com/abstract=3339123 or http://dx.doi.org/10.2139/ssrn.3339123

[8] Candia C, Jara-Figueroa C, Rodriguez-Sickert C, Barabási A, Hidalgo C. The universal decay of collective memory and attention. Nat Human Behav. 2019;3:82–91.

[9] Morris S. *Sir Duke*. Tamla recordings; 1977.

[10] Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian data analysis. 3rd ed. Boca Raton (FL): CRC Press; 2014.

[11] Koop G, Potter S. Prior elicitation in multiple change-point models. Int Econ Rev. 2009;30(3):751–772.

[12] Li P, Chen S. A review on Gaussian latent process models. CAAI Trans Inf Technol. 2016;1:366–376.

[13] Li B, Song J. Nonlinear sufficient dimension reduction for functional data. Ann Stat. 2017;45(3):1059–1095.

[14] Cook C. An introduction to envelopes: Dimension reduction for efficient estimation in multivariate statistics. Hoboken (NJ): Wiley; 2018.

[15] Huey S. Whip it review; 2024. Available from: https://www.allmusic.com/song/whip-it-mt0008523888

[16] Banker D, James T, Malhotra V, Bittenbender P, Jenkins S. Bitchin': the sound and fury of Rick James; 2021.

[17] Puckett M. The theory and technique of electronic music. Hackensack (NJ): World Scientific Publishing Company; 2007.

[18] Gomez H, Bures M, Moure A. A review on computational modelling of phase-transition problems. Philos Trans R Soc A. 2019;377:20180203.

[19] Polunchenko A, Tartovsky A. State of the art in sequential change point detection. Methodol Comput Appl Probab. 2012;14:649–684.

[20] Petitjean F, Ketterlin A, Gancarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit. 2011;44(3):678–693.

[21] Barr DR, Sherrill ET. Mean and variance of truncated normal distributions. Amer Stat. 1999;53(4):357–361.

[22] Lewandoski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. J Multivariate Anal. 2009;100(9):1989–2001.

KOBI ABAYOMI received his Ph.D. in Probability & Statistics from Columbia University. He has held post-doctoral positions at Duke University, the Statistical & Applied Mathematical Sciences Institute (SAMSI) and Stanford University before joining the faculty of Industrial & Systems Engineering at the Georgia Institute of Technology. Dr. Abayomi continued academic appointments at Binghamton University & the University of Cuenca. He has led data science teams at Dun & Bradstreet, Barnes & Noble Education, Warner Media and most recently Warner Music Group in what he has coined *Data Science for Digital Media*. Dr. Abayomi currently holds an appointment at Seton Hall University and is the Head of Science for Gumbel Demand Acceleration.

Gumbel Demand Acceleration, Carlsbad, CA 92011
kobi@gatech.edu