Dynamic Attraction Models for Cultural Choice

Victor Martínez-de-Albéniz* Anna Sáez-de-Tejada†

Submitted: September 8, 2014.

Abstract

In this paper, we propose a dynamic choice model for cultural preferences. Our approach is to describe an option's attractiveness as a time-dependent function, of varying degrees of complexity. We develop a methodology to easily estimate the model's parameters, even in the presence of partially missing data. We then apply the model to study the dynamics of choice of names for newborn babies and songs being played in the radio, two examples of cultural choice, for which unimodal attractiveness models work well. The model can be used to forecast the popularity over the life cycle of these products, including the times of peak popularity.

1 Introduction

Coco Chanel once said that "Fashion is not something that exists in dresses only. Fashion is in the sky, in the street, fashion has to do with ideas, the way we live, what is happening." Indeed, fashion historically referred to the production and marketing of new styles of clothing and cosmetics. But since the 1960s, many sectors experience patterns that closely resemble clothing fashion trends. Product life cycles across industries are shortening (Thompson 2012), especially in the information technology sector (Whittle 2013). This development is particularly prevalent for products with a cultural dimension. Apparel collections are indeed renewed twice a year, and recently fast fashion retailers have reduced that even further, with items staying in the assortment for a few weeks only (Caro and Martínez-de-Albéniz 2014). Artistic goods such as literature, music or movies suffer from the same fate, due to an apparent appetite for novelty. Human interests in cultural matters thus seem subject to trends, and they can be visualized with tools like Twitter Trending Topics or Google Trends. Harvey (1989) already suggested that fashion is becoming more ephemeral: "The mobilization of fashion in mass (as opposed to elite) markets provided a means to accelerate the pace of consumption not only in clothing, ornament and decoration, but also across a wide swathe

*IESE Business School, University of Navarra, Barcelona, Spain, valbeniz@iese.edu. V. Martínez-de-Albéniz's research was supported in part by the European Research Council - ref. ERC-2011-StG 283300-REACTOPS and by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economía y Competitividad, formerly Ministerio de Ciencia e Innovación) - ref. ECO2011-29536.

†UCLA Anderson School of Management, Los Angeles, California, anna.saez.de.tejada.cuenca.2019@anderson.ucla.edu

of life-styles and recreational activities (leisure and sporting habits, pop music styles, video and children's games, and the like)" (see chapter 17).

In a context of continuous changes in consumer tastes, it is a priority to understand and predict demand, at least in the short term, to take the right decisions regarding production and distribution. Indeed, this is key in the global apparel market, which was estimated at USD 1.7 trillion and employed 75 million people in 2012 (FashionUnited.com 2014), and in the creative industries, generating only in films within the United States USD 504 billion in 2011 (The Associated Press 2013). In these settings, there is consensus that predicting demand before product launch is almost impossible (Christopher et al. 2004), but knowledge of the demand over the early periods of sale provides good estimates for future demand (Fisher 2009).

In this paper, we are interested in describing how consumer tastes evolve over time. In most cases, the choice of an item does not depend solely on those, but also on characteristics of the product, such as price or quality, and constraints on consumption, such as availability, display or ease of access. These of course may affect the demand observed. We intend to isolate intrinsic variations of tastes from these extrinsic factors, by focusing on settings where consumers should choose based on taste only.

For this purpose, we develop a demand model that captures the most important features related to trends. First, demand is related to a choice probability, which allows us to integrate inter-product substitution, i.e., when the demand for one product increases, it decreases for the rest. Second, while choice modeling has a long history in economics and marketing, most of the literature works with static models, or models where the only time variation is due to changes in the offer (e.g., dynamic prices or assortments). In contrast, in our paper product attractiveness is modeled as a product- and time-dependent function. Our approach has some interesting features that are worth highlighting. It is flexible to accommodate varying degrees of complexity, ranging from simple static attractiveness useful for settings without strong trends, to more complex unimodal attractiveness more appropriate for products that experience boom and bust periods. Moreover, in contrast from most choice models, our approach makes estimation relatively easy even in the lack of complete data. Specifically, a common handicap in many empirical papers with choice models is that complex algorithms must be used to estimate the amount of unobserved demand; here unobserved data can be included in a free variable for the outside option that can be easily estimated, in part because we usually have access to the total amount of people making a choice.

After developing the model, we apply it to two settings of cultural choice, in which demand should have little exogenous influences. First, we study the choice of names for newborn babies.

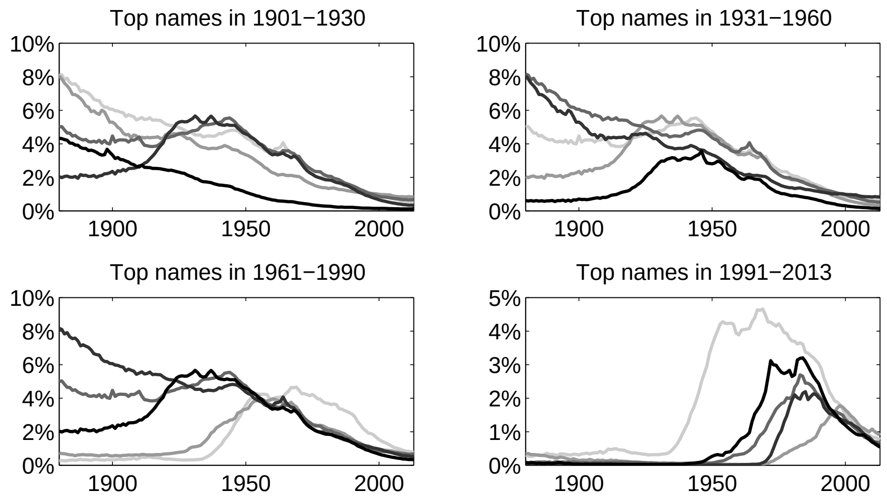

Figure 1: Evolution of male names in the US which were the top 8 in 4 different periods.

This might possibly be one the clearest situations where demand is driven by taste only, since names have infinite availability, do not carry a price tag which might create differences between them and there is no outside option for choice. Figures 1 and 2 show the evolution of name popularity of the top names in different time periods, for boys and girls respectively. Second, we study the choice of songs being played in different radio stations. In this case, song choice is constrained by the fact that new songs are continuously released, so the demand for a specific song can only exist after its introduction time. Figure 3 illustrates the popularity of songs of several artists. Note that in this analysis, we are not studying why a product is becoming popular. There might be diverse reasons for this, such as the success of a certain celebrity influencing parents naming decision (DPA 2014), or the appearance of an artist in a certain event that might boost the popularity of a certain song. But this falls outside the scope of this paper and would require a very different focus, possibly involving sociology research.

The application of our model to these two data sets reveals some intriguing insights. First, the popularity of songs and names for newborn babies presents the same qualitative features in two different time scales. Most names, even though they are available at all times, exhibit a time window in which their popularity increases and then decreases. The same is true for songs. Second, these patterns imply that our three-parameter models perform very well and provide good estimates of market shares. Third, our model can be used to predict the times of peaking of each choice quite accurately. Similarly, forecasts of future popularity can be created using our model. In this sense, we rely on the past attractiveness trajectory to predict future popularity, in the spirit of Fisher

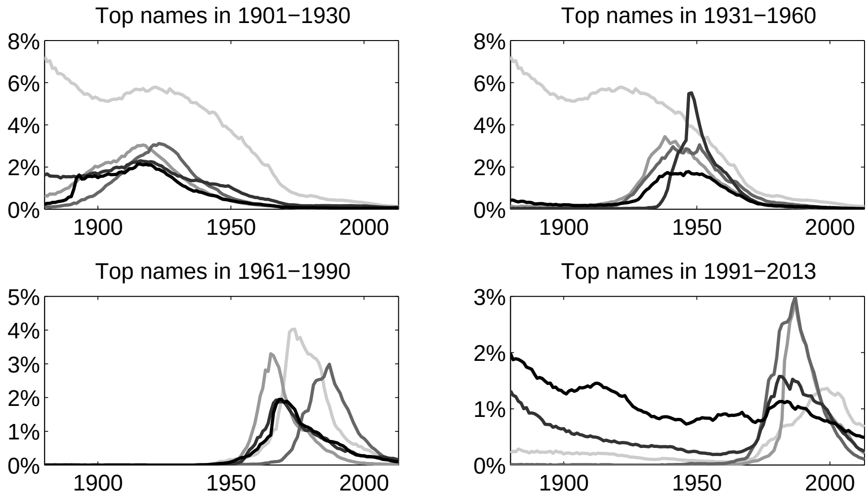

Figure 2: Evolution of female names in the US which were the top 8 in 4 different periods.

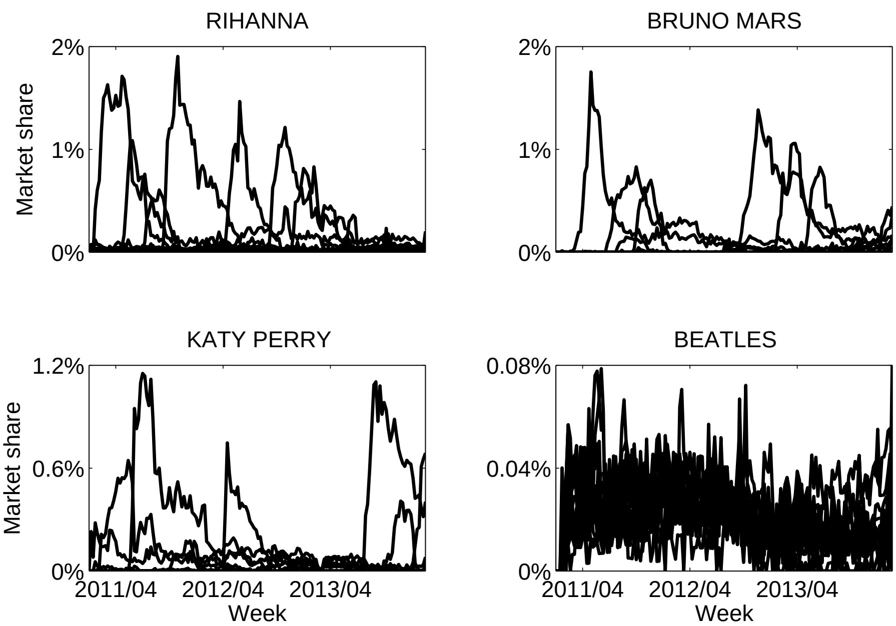

Figure 3: Evolution of the top 8 songs in the UK by Rihanna, Bruno Mars, Katy Perry and the Beatles.

(2009) where early sales are highly predictive. Lastly, deeper analysis of the data allows us to study the differences in the distribution of the demand across options, i.e., the weight of the 'head' and the 'tail' of the distribution, and its speed of change across countries and over time.

In summary, we make two main contributions. On the one hand, we develop a simple and flexible method to integrate life cycle patterns in the estimation of demand, which can be applied easily to large data sets with missing data. On the other hand, we discuss two applications that suggest that cultural choice is well captured with attractiveness functions that increase and then decrease over time.

The rest of the paper is organized as follows. In §2 we review the relevant literature. Our dynamic choice model is presented in §3 and the two applications are shown in §4. We conclude in §5.

2 Literature review

This paper is connected to several streams of literature. First, the development of a dynamic choice model that incorporates life-cycle patterns is related to the economics, marketing and operations literature. Second, our applications have been studied before and we provide an overview of the work in those areas.

Choice modelling has been explored since the 1970s. Two early works that are worth highlighting are Bell et al. (1975), who show that if market shares satisfy a number of axioms, then they must be expressed through an attraction model; and McFadden (1974), who describes the logit choice model and its statistical properties, and provides a procedure to estimate the maximum-likelihood parameters. Later on, Kök et al. (2009) provides a review of different choice models that have been used in operations management and in retail in particular, while Anderson et al. (1992) and Train (2009) are widely used textbooks. Many applications in Operations Management have used attraction models, e.g., Bernstein and Federgruen (2004), Hopp and Xu (2005, 2008), Federgruen and Yang (2009), Caro and Martínez-de Albéniz (2012) or Aksoy-Pierson et al. (2013) among others.

From a methodology perspective, one of the key challenges is to deal with missing data. Indeed, when a particular product has one period's data missing, the estimation step becomes challenging (non-convex objective functions). In the operations literature, Vulcano et al. (2010) study an airline revenue management model where customers choose according to a MNL choice model. To estimate the customer arrival rate (or equivalently the outside option), they use an expectation-maximization procedure. Vulcano et al. (2012) provide some further insights on the procedure, and in particular we discuss in §3.2 how to use the knowledge on the share of unobserved data to simplify the estimation, as in their Proposition 1. Expectation-maximization is also used in Kök and Fisher (2007), where purchase quantity is also considered. Talluri (2009) provides an alternative approach to assess the unknown market size, when there is a finite customer population. In contrast with all these papers, we have access to the total market size in a given period, but we do not observe all the individual choices. By allowing the outside option parameter to vary freely, we are able to develop a tractable model where parameters can be easily estimated even with unobserved data. As a result, our approach is well-suited for large data sets, so as to model choice probabilities of 'rare' (unpopular) items. These are the so-called *long tail* products, first described in Anderson (2004) and reviewed in Elberse (2008).

Lastly, our model focuses on characterizing the evolution of popularity over time. Such dynamics have been well studied in the movie industry. Eliashberg et al. (2006) provides an excellent review of the literature. Most of the research focuses on identifying the attributes that drive movie box office, but there are also some papers that incorporate sales dynamics. Sawhney and Eliashberg (1996) develop a simple model that predicts that the amount of customers that view a movie may be gamma distributed over time. Neelamegham and Chintagunta (1999) develop a Bayesian model and use cumulated past viewers and a trend in predicting future viewers. Elberse and Eliashberg (2003) model the interaction between US and international demand for movies, when they are first released in the US market, and find that the time lag between both releases is important. The effect of competition is incorporated in Krider and Weinberg (1998). The most relevant paper for our work is Ainslie et al. (2005), who study box office weekly sales with a choice model with gamma attractiveness where time is counted from a movie's introduction date. They use a hierarchical regression model to connect movie attributes to movie life-cycle patterns and show that this model is provides better estimates than other demand models that ignore the effect of competition across movies. Their focus is to determine the impact of movie characteristics (e.g., star actors) on movie success. Our paper uses a similar modelling approach: we also use a choice model with different shapes for product attractiveness (including gamma functions), although we do not explain it with attributes. Instead, we develop tractable estimation methods and use model parameters to understand the variations in attractiveness over time. Other relevant papers that introduce decays of attractiveness over time are Caro et al. (2014) and Çınar and Martínez-de-Albéniz (2013), who optimize product introductions subject to an exponential decay of attractiveness.

We must also acknowledge all the work related to our two applications. Regarding name popularity, work on surname popularity is not so relevant for us, since this is not usually a choice (surnames are inherited from parents). In contrast, there has been some research on first name choices, mostly in sociology. Hahn and Bentley (2003) and Bentley et al. (2004) use models from genetics, but do not model names individually: they aggregate names in frequency bins and look for probabilities of 'mutation' into new names. These papers focus on how much the name pool is diversifying over time, but do not look into popularity life cycles. Fryer and Levitt (2004) create a 'black name' index and predict adult outcomes (e.g., education, wages) as a linear regression of this index; they then relate poverty to having a name with a high 'black name' index. Yoganarasimhan (2012) uses naming decisions to test whether there are fashion cycles. Regarding song popularity, there is work in economics that proposes different approaches to analyze cultural choices. Most of the literature explains popularity from attributes. For instance, Louviere and Hensher (1983) use an attraction model to explain the popularity of different exhibitions in Australian museums, using a static model, where the drivers of attractiveness are observable characteristics such as type of exhibition, city or ticket cost. Attributes may be external variables: Crain and Tollison (2002) test the economic theories of superstardom by regressing concentration of top artists with earnings of artists, music industry variables and socio-demographic characteristics of the population. Similarly, Prieto-Rodríguez and Fernández-Blanco (2000) and Favaro and Frateschi (2007) test the influence of different socioeconomic features of the population on the consumption of popular and classical music to find differences in both groups of consumers, using a model first proposed by Lévy-Garboua and Montmarquette (1996), in which there is a learning process from past demand.

3 Modelling Frequencies over Time

3.1 A Dynamic Attraction Model

Consider a finite set of alternatives $\{1, 2, \ldots, I\}$ that consumers must take over time. The observed probability that a consumer chooses alternative i from all the alternatives is $p_i = \frac{n_i}{N}$, where n_i is the observed number of sales of this product and N is the total number of sales, $N = \sum_{i=1}^{I} n_i$. This observed probability is in fact a market share.

Static attraction models have been proposed to explain market shares: the estimated probability is written as

$$\hat{p}_i = \frac{a_i}{a_0 + \sum_{j=1}^{I} a_j},$$

where a_i is the attractiveness of each product, which is a function of a certain number of parameters and observed features, $a_i = a(\boldsymbol{\theta}_i, \boldsymbol{q}_i)$, and a_0 is the attractiveness of the outside option (or no-purchase preference). The number of parameters, namely the dimension of $\boldsymbol{\theta}_i$, is a choice driven by how complex and flexible we want our model to be. Some examples of observed features, $\boldsymbol{q}$, could be the price of the product, its quality, the reputation of the company, etc.

In dynamic contexts, static attraction models usually fail to accommodate time variations. The main purpose of this paper is to develop a dynamic alternative: we define the attractiveness as a time-depending function, $a_{it} = a(\boldsymbol{\theta}_i, \boldsymbol{q}_i, t)$, where the only observed feature is the time t, and thus the market share for each $t = 1, \dots, T$ becomes

$$\hat{p}_{it} = \frac{a_{it}}{a_{0t} + \sum_{j=1}^{I} a_{jt}}. \tag{1}$$

In this work, the attractiveness functions we are considering are all in the form $a_{it} = e^{\phi(\boldsymbol{\theta}_i, t)}$, where

$$\phi(\boldsymbol{\theta}_i, t) = \sum_{d=1}^{D} \theta_i^d f_d(t)$$

is linear respect to $\boldsymbol{\theta}_i = (\theta_i^1, \dots, \theta_i^D)$. Such exponential formulation will result in a tractable estimation, especially when the objective is to maximize log-likelihood. In particular, we consider four different attractiveness functions, having each of them a different number of parameters (see Figure 4). It is worth pointing out that $f_d(t)$ can be replaced by a function of both t and i, as we do in §4.2.

	Constant	**Exponential**	**Gamma**	**Exp-Quadratic**
Dimension	$D = 1$	$D = 2$	$D = 3$	$D = 3$
Parameters	$\boldsymbol{\theta}_i = \alpha_i$	$\boldsymbol{\theta}_i = (\alpha_i, \beta_i)$	$\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i)$	$\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i)$
Attractiveness	$a_{it} = e^{\alpha_i}$	$a_{it} = e^{\alpha_i + \beta_i t}$	$a_{it} = e^{\alpha_i + \beta_i \ln(t) + \gamma_i t}$	$a_{it} = e^{\alpha_i + \beta_i t + \gamma_i t^2}$

Figure 4: Attractiveness functions

In the case of short time series, the constant or exponential formulation might be enough. For longer time series however, where there is larger variation of market shares with time (for products which show a quick increase and decrease, following a rapid fashion 'window'), we must use the exp-quadratic or the gamma function. As we will see later, these attractiveness functions have good properties respect to the estimation process.

Note that in this model there are D degrees of freedom. Specifically, for any of the parameters (say the d-th), we can shift θ_i^d by a constant κ for each $i = 1, \dots, I$ so that $\theta_i^{d'} = \theta_i^d + \kappa$. The attractiveness function thus becomes $a'_{it} = e^{\kappa f_d(t)} a_{it}$, and so the estimated probability is invariant:

$$\hat{p}'_{it} = \frac{e^{\kappa f_d(t)} a_{it}}{e^{\kappa f_d(t)} a_{0t} + \sum_{j=1}^{I} e^{\kappa f_d(t)} a_{jt}} = \hat{p}_{it}.$$

Because of these degrees of freedom, we are able without loss of generality to re-center each family of parameters without changing the goodness-of-fit of the estimation. In particular, in §4, we set the average of θ_i^d over $i = 1, \ldots, I$ to zero, for all d.

This model has the advantage of considering product characteristics, which are static and dynamic (trends may differ across products), while being relatively simple (low number of parameters to estimate) and tractable (structural properties that we develop later).

3.2 Estimation Approach

The maximum likelihood is the most common estimation approach for attraction models, e.g., Vulcano et al. (2010), Vulcano et al. (2012). Given a random variable X with density function $f(\cdot, \theta)$, the likelihood of a random sample $x_1, \ldots, x_n$ is $L(x_1, \ldots, x_n; \theta) = f(x_1, \theta) \ldots f(x_n, \theta)$. Due to the multiplicative nature of the likelihood function, it is usually more tractable to work with its natural logarithm, the log-likelihood. Maximizing it over model parameters yields the maximum-likelihood estimator $\hat{\theta}$. This estimator is well-behaved in terms of sufficiency, consistency, efficiency and bias (see Train 2009).

Estimation will be tractable when there is independence of choice over decision-makers and time. In other words, we must ensure that the choice decision for one item at a given time is not influenced by constraints due to previous choices or simultaneous choices of other items. Fortunately, this is not relevant in most of the applications we have in mind:

- When naming a baby, usually only one choice is made per year. Even when there are multiple births, these are insignificant compared to the total number of births in the population. In addition, the only real constraint is not to use the same name that was given to a previous child. But given the low probabilities given to a single name, the conditional probability of a second name is almost identical to that of a first name.

- When choosing a song to play on the radio, again the number of songs chosen by a DJ is typically very small compared to that of the entire sample, so independence can be safely assumed. Moreover, dependencies over time are also minimal due to the fact that professional DJs do renew their playlists often (especially given our time scale of weeks).

- Fashion apparel purchases also fit the independence assumption. Indeed, the number of purchases from an individual consumer is small compared to total sales, even at the store level. Moreover, previous choices usually do not constrain future ones, because assortment changes quickly compared to purchasing frequency and the amount of products is sufficiently large so, when a customer does not want to buy again the same item, the purchase probabilities of the other items is almost unchanged.

In contrast, there are some cases where independence is clearly not satisfied. For example, when there are a few platform technologies that customers need to choose prior to consumption (e.g., a videogame console or a smartphone operating system), a user will only purchase items compatible with that standard (e.g., videogames or apps), which will create a strong time dependency when the number of choices is not large. In these settings, a diffusion model will be more appropriate.

Thus, when observations are independent across decision-makers, the collection of random variables $(n_{1t}, n_{2t}, \ldots, n_{It})$ follows a multinomial distribution with parameters $(p_{1t}, \ldots, p_{It})$. When observations across time periods are also independent, then the probability of observing $(n_{it})_{i=1,\ldots,I;t=1,\ldots,T}$, letting $N_t = \sum_{i=1}^{I} n_{it}$, is

$$\prod_{t=1}^{T} \left(\frac{N_t!}{\prod_{i=1}^{I} n_{it}!} \right) \prod_{i=1}^{I} p_{it}^{n_{it}}.$$

It may be the case that there is no information about n_{it} for a specific i in period t, but the total N_t is still known. The multinomial formulation can take this into account. Let O_t (observed) be the set of i such that n_{it} is known, $U_t = \{1, \ldots, I\} \backslash O_t$ (unobserved), and

$$n_{0t} = \sum_{i \in U_t} n_{it} = N_t - \sum_{i \in O_t} n_{it}.$$

Then the probability of seeing $(n_{it})_{i \in \{0\} \bigcup O_t; t=1,\ldots,T}$ is

$$\prod_{t=1}^{T} \left(\frac{N_t!}{n_{0t}! \prod_{i \in O_t} n_{it}!} \right) \left(\prod_{i \in O_t} p_{it}^{n_{it}} \right) p_{0t}^{n_{0t}}.$$

Note that this property of the multinomial distribution is very powerful: it will allow us to work with censored information, provided that we have access to the total of occurrences N_t in every period. This is very convenient because we can include alternatives in the analysis even when we only have data for a few periods. This provides a richer picture than excluding these alternatives from the analysis. In addition, censored data usually requires making strong assumptions on the structure of the model. For instance, Vulcano et al. (2012) must estimate the strength of the outside option through an external method (total market share estimation), rather than from the individual sales data: Kök and Fisher (2007) must assume that the substitution structure of the unobserved data is the same as the observed one, with an expectation-maximization procedure. We do not need to make any specific assumption in our case. For example, in the case of names, we have access to demographic data (how many boys and girls are born each year), so even if we do not have full access to all the names in the population, we can still estimate the dynamic attraction model. This means that we do model the weight of the long tail of the data and take it into account when modeling the most popular names, although we do not have specific information for each name below a certain level of population.

Moreover, it allows us to also build dynamics into the attractiveness of an outside option, if there is one. In the case of names, there is no outside option, but when we look at choice of products in a store, it is possible that we know how many people entered the store without buying, or how many people bought products in an outside category not considered in the analysis.

Letting

$$\mathcal{C} = \sum_{t=1}^{T} \left[\ln\Big(N_t!\Big) - \ln\Big(n_{0t}!\Big) - \sum_{i \in O_t} \ln\Big(n_{it}!\Big) \right],$$

the resulting log-likelihood of a model with $p_{it} = \frac{a_{it}}{a_{0t} + \sum_{j=1}^{I} a_{jt}}$ becomes

$$\mathcal{C} + \sum_{t=1}^{T} \left[\sum_{i \in O_t} n_{it} \ln(a_{it}) + n_{0t} \ln \left(a_{0t} + \sum_{i \in U_t} a_{it} \right) - N_t \ln \left(\sum_{i=0}^{I} a_{it} \right) \right] \tag{2}$$

where $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$. The best estimates $\boldsymbol{\theta}_i$ are obtained by maximizing (2). To make the optimization tractable, it would be desirable to show that LLH is concave in $(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_I)$. However, the terms $n_{0t} \ln(a_{0t} + \sum_{i \in U_t} a_{it})$ are convex in $\boldsymbol{\theta}_i$, for $i \in \{0\} \cup U_t$, because they are logarithms of a sum of exponentials (Boyd and Vandenberghe 2004). Therefore, we redefine the sum of attractiveness of unobserved products and outside option as $c_t = a_{0t} + \sum_{i \in U_t} a_{it}$, and so $a_{0t} + \sum_{i=1}^{I} a_{it} = a_{0t} + \sum_{i \in U_t} a_{it} + \sum_{i \in O_t} a_{it} = c_t + \sum_{i \in O_t} a_{it}$. Additionally, we can write the censored attractiveness in an exponential formulation, $c_t = e^{\eta_t}$, where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_T)$ is a collection of T free parameters. With this notation, and canceling each logarithm of exponential, the log-likelihood function becomes

$$LLH(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathcal{C} + \sum_{t=1}^{T} \left[\sum_{i \in O_t} n_{it} \phi_{it} + n_{0t} \eta_t - N_t \ln \left(e^{\eta_t} + \sum_{i \in O_t} e^{\phi_{it}} \right) \right]. \tag{3}$$

Theorem 1. **[McFadden (1974)]** *$LLH(\boldsymbol{\theta}, \boldsymbol{\eta})$ is jointly concave in $(\boldsymbol{\theta}, \boldsymbol{\eta})$.*

As a result, there is a unique maximum of the problem, which can be solved using simple algorithms, such as gradient methods. Moreover, we can use first-order conditions to further simplify the estimation procedure. Since there are T different free parameters η_t, we have the following T decoupled first-order conditions:

$$\frac{\partial LLH}{\partial \eta_t} = n_{0t} - \frac{N_t e^{\eta_t}}{e^{\eta_t} + \sum_{i \in O_t} e^{\phi_{it}}} = 0 \quad \forall t = 1 \ldots, T.$$

These conditions provide an analytic expression $\boldsymbol{\eta}(\boldsymbol{\theta})$:

$$\eta_t = \ln \left(\frac{n_{0t}}{N_t - n_{0t}} \right) + \ln \left(\sum_{i \in O_t} e^{\phi_{it}} \right). \tag{4}$$

Redefining the constant term as

$$\mathcal{C}' = \mathcal{C} + \sum_{t=1}^{T} \left[n_{0t} \ln \left(\frac{n_{0t}}{N_t - n_{0t}} \right) - N_t \ln \left(\frac{N_t}{N_t - n_{0t}} \right) \right],$$

we obtain the following expression for the log-likelihood which only includes the attractiveness of each product in each observed period:

$$\mathcal{C}' + \sum_{t=1}^{T} \left[\sum_{i \in O_t} n_{it}\phi_{it} - (N_t - n_{0t}) \ln \left(\sum_{i \in O_t} e^{\phi_{it}} \right) \right]. \tag{5}$$

This final expression has the advantage of remaining jointly concave and including only the observable data. We can thus simplify the optimization from a numerical standpoint. For instance, in our names application, LLH in (3) involves a sum of about 500,000 terms, while (5) involves a sum of 130,000 terms. This reduction is even stronger when the variety of choices is larger.

Finally, the expression in (5) yields interesting insights when all the items under consideration are always observable, i.e., $O_t = O$, and we consider the constant model, i.e., when $\phi_{it} = \alpha_i$. First-order conditions imply for $i \in O$:

$$\sum_{t=1}^{T} \left[n_{it} - (N_t - n_{0t}) \left(\frac{e^{\alpha_i}}{\sum_{j \in O} e^{\alpha_j}} \right) \right] = 0$$

and thus, letting $N_i = \sum_{t=1}^{T} n_{it}$, $N_{obs} = \sum_{t=1}^{T} N_t - n_{0t}$, $\left(\frac{e^{\alpha_i}}{\sum_{j \in O} e^{\alpha_j}} \right) = \hat{p}_{it} = \frac{N_i}{N_{obs}}$.

However, in general first-order conditions for different α_i will be coupled and it will not be possible to obtain closed-form expressions. But some insight can still be extracted from first-order conditions. Specifically, the first-order conditions with respect to θ_i^d is:

$$\sum_{t | i \in O_t} f^d(t) \Big[(N_t - n_{0t}) \hat{p}_{it} \Big] = \sum_{t | i \in O_t} f^d(t) n_{it}.$$

In other words, the parameters for i are such that the weighted average of choice probability times number of observable occurrences is equal to the weighted average of occurrences of i, with the weights being equal to $f^d(t)$.

Maximizing LLH in (5) is an unconstrained concave maximization problem. Thus, we can use a standard method to find the optimal set of parameters $\boldsymbol{\theta}^*$. We use Matlab's built-in function *fminunc*, with the trust region algorithm. Once we have obtained $\boldsymbol{\theta}^*$, we can find $\boldsymbol{\eta}^*$ through Equation (4) and obtain the final estimate for this set of parameters.

3.3 Other Attractiveness Models and Approaches

Using an exponential form of attractiveness has good properties as we have seen. We have nevertheless explored other families of functions with other advantages. Specifically, we considered polynomials, in the form $a_{it} = \sum_{d=1}^{D} \theta_i^d t^{d-1}$. Unfortunately, despite their simplicity (they are linear in the parameters $\boldsymbol{\theta}$), they make the log-likelihood objective non-concave in the parameters: the term $-N_t \ln\left(\sum_{i=1}^{I} a_{it}\right)$ is convex in $\boldsymbol{\theta}$. As a consequence, we cannot guarantee that there is a unique maximum of the log-likelihood. Furthermore, we must require all the attractiveness to be positive in all periods (which is always true for exponential functions). With polynomials, the only way to ensure positivity is to impose constraints, which increases the complexity of the maximization problem significantly. For instance, if $a_{it} = \alpha_i + \beta_i t$, we need two constraints for each $i = 1, \ldots, I$, i.e., $2I$ constraints. Depending on the size of the data (the number of products to consider in the estimation), the amount of constraints can be very large.

In addition, one could consider minimizing the quadratic error instead of maximizing the likelihood. This method works very well for linear regression and other simple approximation functions. However, in our case, since the estimated probability in (1) is a quotient of functions, the quadratic error is non-convex, which again makes this approach not tractable.

Finally, we have also tried other approaches, in particular autoregressive models (see for instance Elberse and Eliashberg 2003), but these have some drawbacks. Most importantly, they usually assume an intrinsically stationary process, and time variation is due to changes in product characteristics, e.g., number of screens in movie box office. This is clearly not appropriate in our setting. In such models, each point of a time series is estimated as a linear combination of the previous s points, where s is the order of the model. With our data, the autocorrelation of 300 names in our database with different levels of popularity was computed, and all of them had significant autocorrelation for large lags, mostly $s \geq 20$. Thus, we would need to estimate a large amount of parameters for each product, which is not feasible. In contrast, we obtain good estimates with $D \leq 3$ with our proposed model.

4 Applications

4.1 Names for Newborn Babies

4.1.1 Data Description

We collected yearly data about the newborn names frequencies from four different countries or regions: the United States (1880-2013, 7,695 names), the United Kingdom (1996-2011, 28,806 names), Sweden (1998-2012, 1,575 names) and Catalonia in Spain (1997-2011, 2,612 names). These countries provided well-organized data. In contrast, many others had to be discarded due to inconsistency in the reporting, lack of centralized reports or high fees. The data provides the number of boys or girls with a given name in a given year. For statistical privacy purposes, names with few occurrences were not reported, e.g., in the UK, names with two occurrences or less were removed, while in Sweden, names with nine or less.

The main problem in some countries is the short time window. In most cases, the information is stored digitally since the late 1990s. This allows us to study the dynamics over a very short period, of less than 15 years. This is the case for the databases corresponding to the UK, Sweden and Catalonia.

The country with the longest, most consistent data was the US. Their database about newborn names, starting from 1880, is publicly available on the Social Security website. However, they only offer the 2,000 most popular names of each year (1,000 for boys and 1,000 for girls). They also provide the yearly number of births for each sex. Due to this longer time window, we chose to validate our model on US data. In total, there is data for $3,534$ boys' and $4,161$ girls' names, over a period of 134 years.

4.1.2 Results

The estimation of the parameters was made separately for boys and girls, because the sex of the baby is given and not a choice to be modeled. We focused on 100 names for each sex. Given the number of births for that sex in each year, N_t, the unobserved part of the data was easily computed as $n_{0t} = N_t - \sum_{i \in O_t} n_{it}$. A sample of the results is shown in Figure 5. Only the three-parameter models, i.e., Gamma and Exp-Quadratic, have been included in the figure.

To provide a comprehensive comparison of the models, we compare the log-likelihood of each model to the one obtained when there is no model (i.e., $\theta_i^d = 0$, $\forall i = 1, \ldots, I$ and $\forall d = 1, \ldots, D$) and the one where the model estimates exactly the probability (i.e., $\hat{p}_{it} = p_{it}$, $\forall i = 1, \ldots, I$ and $\forall t = 1, \ldots, T$).

To compare the goodness-of-fit of each model, we propose the following index, similar to Train (2009) pp. 68-69,

$$LRI = \frac{LLH(\boldsymbol{\theta}, \boldsymbol{\eta}) - LLH(\mathbf{0})}{LLH(\hat{\boldsymbol{p}} = \boldsymbol{p}) - LLH(\mathbf{0})}, \tag{6}$$

where $LLH(\mathbf{0})$ is the log-likelihood obtained with no model (i.e., when all parameters have zero value), and $LLH(\hat{\boldsymbol{p}} = \boldsymbol{p})$ is the log-likelihood with as many parameters as data points so that the model estimates exactly the probability for each name and time period. Thus, index LRI takes values between 0 and 1.

To ensure the robustness of our conclusions, we can also consider other statistical measures,

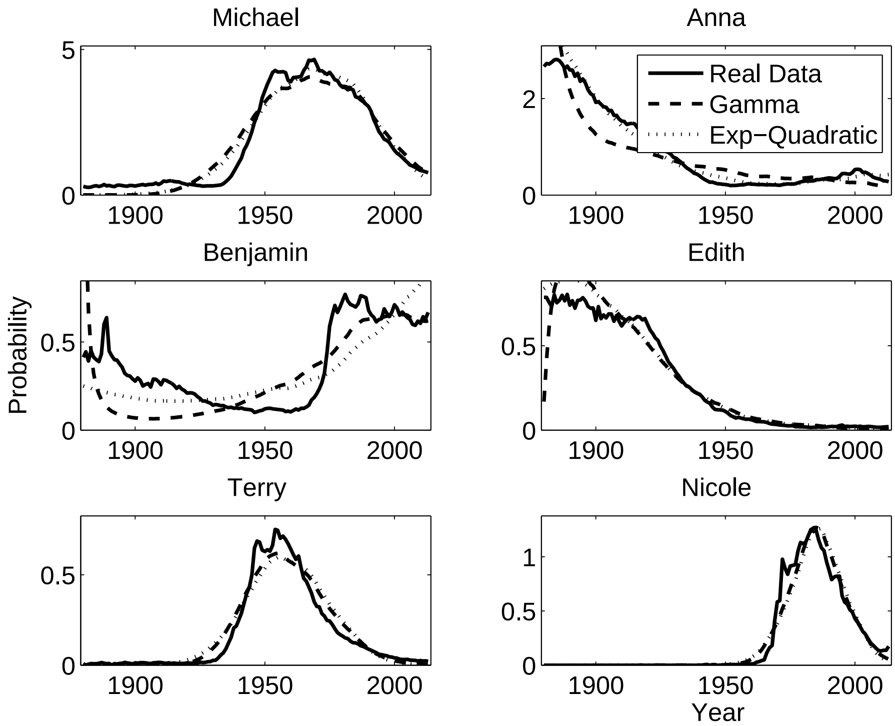

Figure 5: Comparison of frequency and estimate for boys (left) and girls (right), ranked 5 (top), 50 (middle) and 95 (bottom) in popularity over the sample.

such as the average relative error, computed as as

$$ARE = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{1+|O_t|}\sum_{i\in\{0\}\cup O_t}\left|\frac{\hat{p}_{it}-p_{it}}{p_{it}}\right|\right]. \tag{7}$$

The obtained likelihood index and relative error for each sex and model are shown in Table 1.

		Constant	**Exponential**	**Gamma**	**Exp-Quadratic**
Boys	LRI	0.572	0.862	0.934	0.947
	ARE	$4.227 \cdot 10^{-2}$	$1.162 \cdot 10^{-2}$	$3.160 \cdot 10^{-3}$	$3.285 \cdot 10^{-3}$
Girls	LRI	0.336	0.636	0.856	0.886
	ARE	$4.711 \cdot 10^{-2}$	$1.664 \cdot 10^{-2}$	$4.542 \cdot 10^{-3}$	$4.256 \cdot 10^{-3}$

Table 1: Comparison of different models for the top 100 names.

These results show that more complex models perform better than simple ones. In addition, we find that the improvement is quite significant: adding a second parameter increases LRI by 0.3, and a third parameter by 0.07-0.08 (boys) and 0.22-0.25 (girls). The potential improvement for more parameters becomes very small after three, as the maximum improvement is 0.06 (boys) and 0.14 (girls).

We observe that the models with one or two parameters per name (Constant and Exponential) generally provide a significantly worse fit than the models with three parameters. The reason for this is mainly that names over 134 years tend to exhibit extended periods of increase and decrease, which are well captured through a unimodal attractiveness, but not with a constant or monotonic one. In other words, to capture the fact that a name is 'fashionable', it is necessary to include at least two balancing effects, one dominating in the periods of increase (e.g., $\beta_i \ln(t)$ in the Gamma model with $\beta_i > 0$), the other in the periods of decrease (e.g., $\gamma_i t$ in the Gamma model with $\gamma_i < 0$).

Finally, it is worth highlighting the difference between sexes. Fitting the popularity of boys names seems to be easier and simpler models are already very accurate; in contrast, the dynamics of girls names popularity are more difficult to characterize, and three-parameter models for girls do as well as two-parameter models for boys. One reason for this seems to be that, for girls, there is much more variation over time: the number of names appearing in the top 1,000 of a given year, over 134 years, is 4,161, 20% higher for girls compared to boys, and as a result names move from less to more popular or vice-versa faster. The same is true for the 100 names considered in our estimation: the percentage of newborn boys whose names fall out of the top 100 names is, on average, 41% (and 73% counting only from year 2000), while for girls is 59% (and 89% from year 2000).

4.1.3 Discussion

The comparisons provided so far focused on model performance to fit the data. There are several other aspects that require discussion. First, we can use our model outside the sample time range, for forecasting purposes. Second, our data provides valuable information regarding popularity even when there is scarce data about obscure names, so it provides insights regarding the long tail. Moreover, combining long tail insights with forecasting of the future could yield interesting predictions regarding the variation in concentration of popularity over time. Finally, it is worth examining whether the findings from US data are generalizable to other countries and cultures.

Forecasts of popularity. To develop a forecast, we studied the Gamma and Exp-Quadratic models for the top 100 names, over the period 1951-2013. For this, we had to provide an estimate of the outside option parameter a_{0t} for $t = 1951, \ldots, 2013$, which we obtained by fitting a simple regression (specifically, our estimate $a_{0t} = e^{\phi_{0t}}$ where $\alpha_0, \beta_0, \gamma_0$ minimized $\sum_{t=1951}^{2013} \Big(\phi_{0t} - \ln(a_{0t})\Big)^2$, i.e., the best fit with the actual data). Note that this estimate only affects the absolute frequency of names, not their relative ratios in the sample. The choice of time window for the forecast was due to the change of behavior of the long tail we can observe starting in the 1950s. With this estimate of a_{0t}, and knowing $\boldsymbol{\theta}_i$ for all the names, we calculated the estimated probability p_{it} for $t = 2014, \ldots, 2030$.

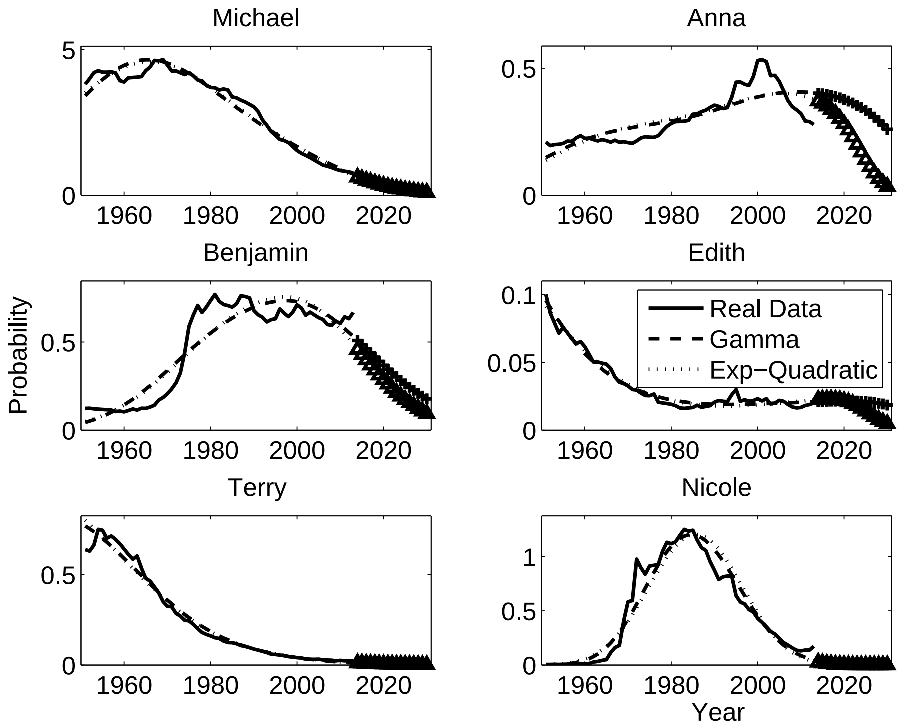

Figure 6: Forecast until 2030 of names for boys (left) and girls (right), ranked 5 (top), 50 (middle) and 95 (bottom) in current popularity over the sample.

Figure 6 provides a measurement of the quality of the forecast for these two metrics. With this forecasted probability, our model would predict that in 2030, the 5 most popular names in the US (out of those observed until today) will be Jack, Henry, Samuel, William and Alexander for boys, if we use the Gamma formulation, and Jack, Henry, Samuel, William and Anthony, if we consider the Exp-Quadratic. For girls, the most popular names will be Ella, Emma, Lillian, Grace and Evelyn, according to the Gamma model, and Nellie, Gertrude, Myrtle, Ella and Bessie, according to the Exp-Quadratic. It thus seems that more traditional names will be popular again for girls, but not so much for boys. Of course, forecasting individual names is not very reliable. In contrast, a safer forecast would be to evaluate the percentage of the top 100 names in the future. We discuss the evolution of the top names next.

Name life cycles. The estimates of our model allow us to directly describe the pattern of popularity of a name. We can contrast the predictions of the model with the actual life cycle of the name. Specifically, for each name we can calculate t_i^m as the time between maximum popularity

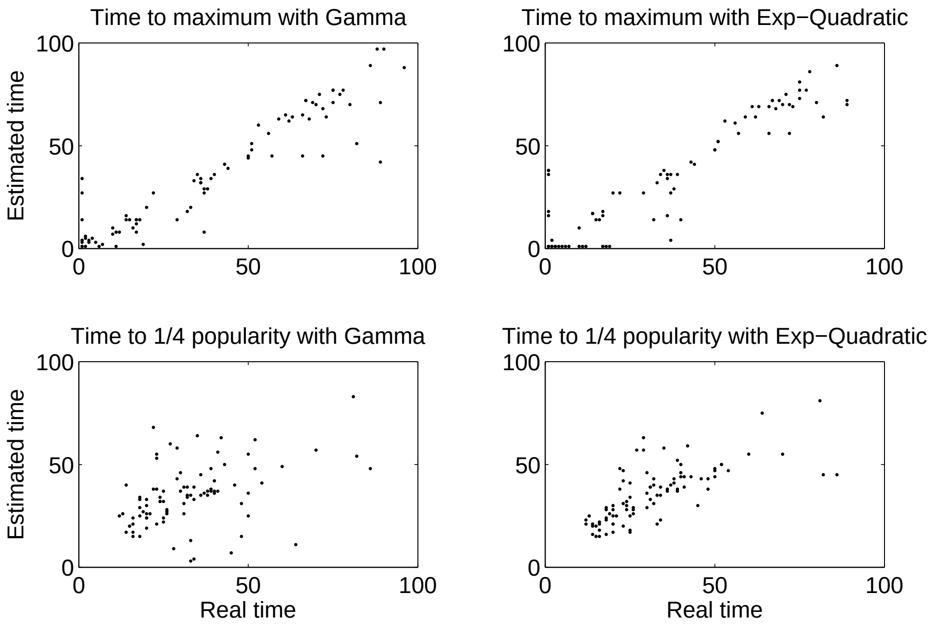

Figure 7: Comparison of t_i^m and $\hat{t}_i^m$ (top) and $t_i^{1/4}$ and $\hat{t}_i^{1/4}$ (bottom), for female names. Time is measured in years.

(its peak) and the beginning of our window (1880). We can also calculate the ω-life time t_i^ω, i.e., the time elapsed between the peak and the time in which a name's popularity reaches ω of the peak. We focus on $\omega = 1/4$ here. While these parameters can be calculated directly from the data, they can also be retrieved from our model. Indeed, the maximum popularity in our Gamma or Exp-Quadratic specifications can be easily calculated numerically. Figure 7 compares t_i^m and $\hat{t}_i^m$ (top) and $t_i^{1/4}$ and $\hat{t}_i^{1/4}$ (bottom). We observe that our model correctly approximates the peak time and the decay pattern of each name.

Popularity over time and across countries. To further understand the distribution of popularity across names, we can evaluate the importance of the 'head' vs. the 'tail' of the distribution. Although complex measures can be used (e.g., the Gini coefficient), one simple indicator is to calculate the popularity of the top 100 names over time. The top of Figure 8 compares popularity for each year's top 100 names in the US. We observe that indeed the most popular names are becoming less popular over time, which indicates that the 'tail' in names is growing while the 'head' is narrowing. The same conclusions apply for the top 10 or top 50 names, not only the top 100. This insight suggests that the long tail phenomenon discussed in Anderson (2004) is observed in name choices too. In other words, families are moving away from mainstream names, so that their babies

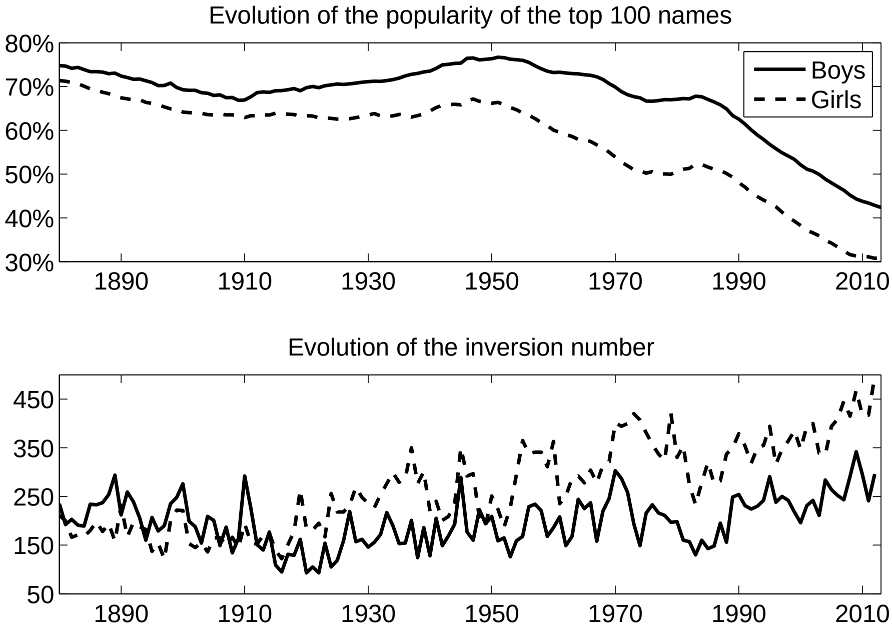

Figure 8: Evolution of the popularity of top 100 names (top) and inversion number (bottom) in the US over time.

have a more unique and distinctive identity. This contrasts with studies of other cultural products such as Tan and Netessine (2009), which find that both the head and the tail are getting stronger, while intermediate products are becoming weaker. The differences across sexes are worth noting: as discussed earlier, girls exhibit more variety than boys.

Furthermore, to understand better the variations in popularity, we can look at measures of change. For this purpose, we use the concept of inversion number from mathematics (Stanley 2011). Specifically, we compare the rankings between different years. In year t_1, the ranking of popularity if $i_1, \ldots, i_{100}$; in year t_2, denote j_k the popularity rank of name i_k. For each pair $i_k < i_l$, if $j_k > j_l$ then we call this pair an inversion. The inversion number is defined to be the number of inversions. Intuitively, an inversion number of zero denotes that there has been no change in the ranking, while an inversion number of $\frac{100 \cdot 99}{2} = 4950$ denotes that all the pairs have been inverted, i.e., the ranking has been completely reversed. The bottom of Figure 8 shows the average inversion numbers between t and $t + 1$ in the US. The figure suggests that average inversion is much larger for girls compared to boys, which again confirms that there is not only larger tails but also faster variation in girls. In addition, the speed of change has remained relatively stable over time, although for girls there has been some slight increase since the 1920s.

Finally, Table 2 compares average inversion number and popularity of the top 100 names across countries. We find that the US names have a larger tail, perhaps due to its larger size and more varied cultures (this is why it is called the 'melting pot'). The UK has a large tail as well compared to Sweden and Catalonia, for the same reasons. The differences across sexes are prevalent in all countries but largest in the US and the UK. Interestingly, the differences in inversion numbers are actually small, which suggests that the speed of change is similar in the four countries but the variety levels are different.

		US	**UK**	**Sweden**	**Catalonia**
Inversion number	Boys	251	295	331	347
	Girls	391	390	391	415
Popularity of top 100	Boys	48%	63%	71%	69%
	Girls	35%	52%	66%	67%

Table 2: Inversion number (top) and popularity of the top 100 names (bottom), for different countries. Average over 1990-2013.

4.2 Songs Played in the Radio

4.2.1 Data Description

In order to apply our model to another cultural choice process, we collected data about which songs are played on commercial radio. In this case, the choice makers we consider are not the actual radio listeners, but each station's DJs, who choose which song to play in every moment.

The information was obtained from the company BMAT. This firm has developed the technology (Vericast) to monitor the songs being played in each moment on any radio station in any country. This is done by identifying the musical 'fingerprint' of each song and matching it with their large database of songs. We obtained reliable data from March 2011 to March 2014, from stations in the UK, the US, Germany, and an aggregation of continental Europe. To choose among databases, the continental Europe database was not studied in detail due to important variations across countries. We focused on the UK database because it was the largest one, both in number of different tracks and number of weeks monitored, so we chose this one to validate our model. In total, there was information about $60,747$ different songs over a period of 160 weeks.

The database included information about artists and names of each song, day and time of the day being played, and radio station of that play, but for simplicity we aggregated these plays by week across all stations. Note that we observed all the songs that were played on each station every week, so there were no unobserveds: when we observe zero plays/week, it means that the song was not played at all. Although the data was very clean, before applying our model we had to pre-process it to remove special characters, correct different spellings of the same artist, etc.

4.2.2 Estimation

In this case, all the data is fully observable. However, a song i does not appear in the database until it is released. We call R_t the set of songs that have been released in or before week t. Once a song is released, i.e., $i \in R_t$, then $i \in R_{t'}$ for $t' \geq t$. Let t_i be the release week for the song i. We center the attractiveness with respect to this introduction time: specifically, $a_{it} = a(\boldsymbol{\theta}_i, \tilde{t}_i)$, where $\tilde{t}_i = t - t_i + 1$. Thus, at the release week t_i, we have that $\tilde{t}_i = 1$ (and hence $\ln(\tilde{t}_i) < \infty$). The estimated probability thus becomes

$$\hat{p}_{it} = \begin{cases} \frac{a_{it}}{\sum_{j \in R_t} a_{jt}} & \text{if } t \geq t_i \\ 0 & \text{otherwise} \end{cases}$$

If we are only considering a subset of all songs, we can consider all the others as an outside option. We let the attractiveness of this outside option (the aggregate of other songs) be a free parameter as before. As a result, the log-likelihood function to maximize is that of Equation (3) with the sums are taken over i within R_t instead of O_t. For the estimation, we considered the top $I = 500$ songs that were released in the 3-year time window. The rest of songs were considered as an outside option.

4.2.3 Results

Figure 9 shows the result of the estimation for different songs with different degrees of popularity. We find that visually the dynamics for songs are quite similar to those of names (cf. Figure 5), with the difference that songs have a 'birth' date, which did not occur with names. Furthermore, the shapes of increase and decrease are also steeper for songs, due to the different time scale. Finally, it is worth noting that the increase period is very short for songs, while the decay tends to take longer. This suggests that songs generally behave as blockbusters, as opposed to sleepers (Ainslie et al. 2005).

Table 3 shows the goodness-of-fit of our model to the songs' data. As with Table 1, the quality of the model is quite satisfactory for the Gamma and Exp-Quadratic models, and adding more dimensions $f^d(t)$ does not improve fit significantly. One major difference with the analysis of names is that the attractiveness function that works better for songs is Gamma, whilst for names it was Exp-Quadratic.

4.2.4 Discussion

Song life cycles. Our data allows us to directly describe the popularity pattern of a song. This pattern is typically increasing quickly and decreasing more slowly. As in §4.1.3, for each song we

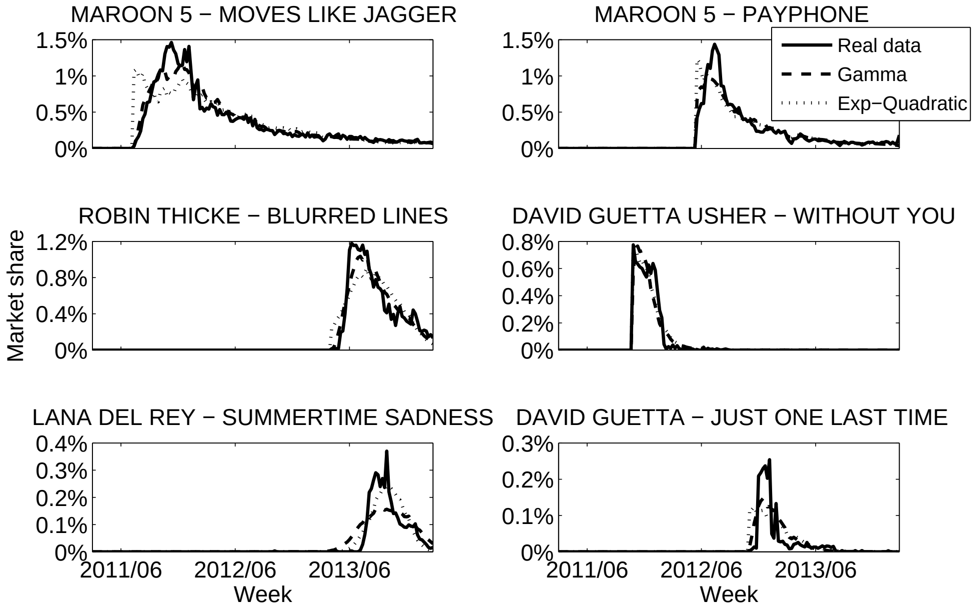

Figure 9: Comparison of frequency and estimate for songs ranked 1, 15, 25, 150, 250, 500 in popularity over the sample.

	Constant	**Exponential**	**Gamma**	**Exp-Quadratic**
LRI	0.561	0.752	0.833	0.809
ARE	$1.08 \cdot 10^{-2}$	$6.22 \cdot 10^{-3}$	$4.39 \cdot 10^{-3}$	$4.77 \cdot 10^{-3}$

Table 3: Comparison of different models for the top 500 songs.

can calculate t_i^m as the time between maximum popularity (its peak) and introduction, and t_i^ω, i.e., the time elapsed between the peak and the time in which a song's popularity reaches ω of the peak. We again focus on $\omega = 1/4$. The typical pattern from the data is the following: the median t_i^m is 10 weeks, while the median $t_i^{1/4}$ is 8 weeks. These are median values and we find significant variation across songs.

Figure 10 compares the values of $\hat{t}_i^m$ and t_i^m, and $\hat{t}_i^{1/4}$ and $t_i^{1/4}$ respectively for the top 500 songs for which t_i occurs within our sample. Again we see that our model is able to correctly approximate the life cycle of songs. We observe that for 14% of the songs our model exactly predicts the time of peaking, and that for 47% of the songs the prediction is within two weeks of the actual peak. The predictions are less accurate for the decay pattern: our model estimates the values of the $t_i^{1/4}$ exactly in 9% of the cases and with at most 5 weeks difference in 27% of the cases only.

Furthermore, as in §4.1.3 we could use our model to forecast song popularity evolution. Specifically, for songs that have already been released, and early popularity data is available, we can

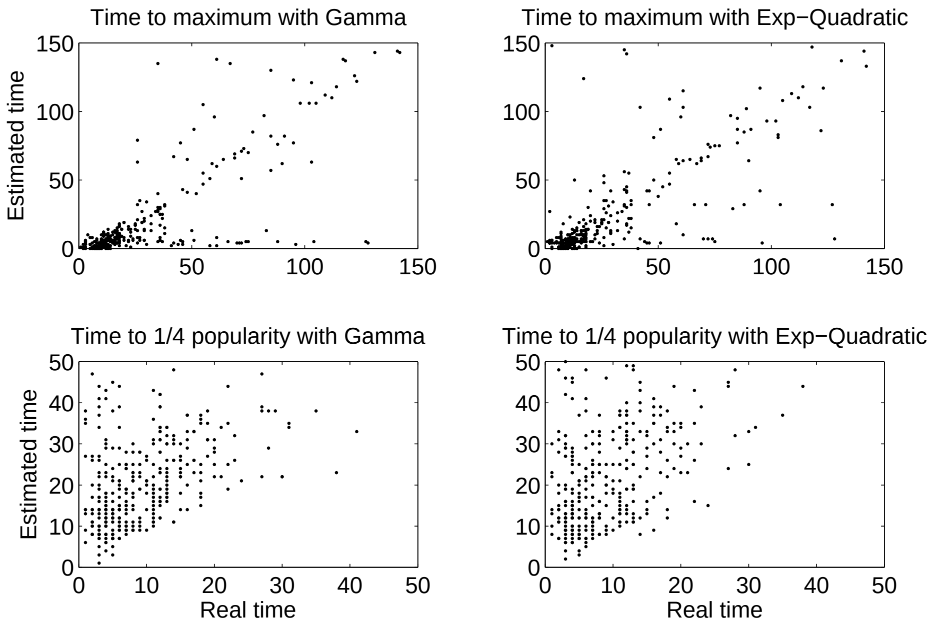

Figure 10: Comparison of t_i^m and $\hat{t}_i^m$ (top) and $t_i^{1/4}$ and $\hat{t}_i^{1/4}$ (bottom), for songs. Time is measured in weeks.

estimate the model parameters for the song and create a forecast of attractiveness and future popularity. In contrast, for new songs where no data is available, our model could still be applied by estimating $\alpha_i, \beta_i, \gamma_i$ on song characteristics such as artist, time of introduction, etc. These values would also provide estimates of peak popularity, t_i^m and t_i^ω for new songs.

Popularity across countries. Similarly to the discussion in §4.1.3, we are interested in the variations of popularity. These are shown in Table 4. First, for songs we do not observe any particular variation of 'head' popularity over time. But the differences across countries are strong. As with names, anglo-saxon countries (UK and US) have the thinnest head, as the top 100 songs only take 28-29% of the total number songs played, while in Germany it reaches 34%. The speed of change is also more pronounced in the UK and the US, with inversion numbers of 861 and 754 respectively, while it is only 524 in Germany. Given that songs are not country-specific (they are released globally), this suggests that in the German market top songs capture more attention and for a longer duration.

	UK	US	Germany
Inversion number	861	754	524
Popularity of top 100	28%	29%	34%

Table 4: Average inversion number (top) and popularity of the top 100 songs (bottom), for stations from different countries.

5 Conclusion

In this paper, we developed a choice model that integrates time variation in attractiveness of each item. This type of model is well-suited for the analysis of cultural choices, which typically exhibit intrinsic variations in tastes over time. We build an estimation methodology that can cope with unobserved data and large data inputs. Our methods provide an efficient alternative to other dynamic models such as those that contain auto-regressive features.

We tested our methodology in two distinct areas. First, we analyzed the popularity of names over time, through several data sets, some of them spanning over 100 years. We found that three-parameter models perform well, which implies that names may exhibit periods of increasing popularity followed by decay, but typically do not see more complex cycles such as increase, decrease and increase again. Our methods allow us to forecast future popularity and times of peaking and decay. They also shed light on the evolution of 'head' and 'tail' of the name distribution and can be used to compare evolutions across countries. Second, we studied the popularity of songs in the radio, through a 3-year data set with weekly market shares of songs in different countries. We again found that three-parameter models are sufficient to fit the typical life-cycle pattern of introduction, steep increase and slower but progressive decay. Our model, again, can be used to forecast song popularity evolution.

While this paper provides the methodological basis to study cultural choices over time, it uncovers several directions for future work. The methodology can of course be expanded to incorporate other explanatory variables, such as sociological and demographic data (as in most economic studies of cultural choice). In particular, application of our approach into 'product' settings would require the inclusion of drivers of choice that arise from the distribution channel, such as price, features (e.g., quality level) or availability (e.g., inventory level). One such application can be found in Boada and Martínez-de-Albéniz (2014). Furthermore, our two applications have uncovered unexpected similarities. Namely, there seems to be an inherent boom-and-bust pattern in both settings, which means that our three-parameter models perform quite well. Interestingly, the time scales may be very different (years in names, weeks in songs) and the choice process itself may involve different mental schemes (names seem to require reflection and rational assessments, songs have an element of discovery and impulse) but the outcomes can be described with the same model. More is needed to understand why this is so. Such investigation probably requires theory building across multiple disciplines such as psychology, sociology and network research.

References

Ainslie, A., X. Drèze, and F. Zufryden. 2005. Modeling movie life cycles and market share. *Marketing Science* 24 (3): 508–517.

Aksoy-Pierson, M., G. Allon, and A. Federgruen. 2013. Price competition under mixed multinomial logit demand functions. *Management Science* 59 (8): 1817–1835.

Anderson, C. 2004. The long tail. *Wired magazine* 12 (10): 170–177.

Anderson, S. P., A. De Palma, and J. F. Thisse. 1992. *Discrete choice theory of product differentiation.* MIT Press.

Bell, D. E., R. L. Keeney, and J. D. C. Little. 1975. A Market Share Theorem. *Journal of Marketing Research* 12 (2): 136–141.

Bentley, R. A., M. W. Hahn, and S. J. Shennan. 2004. Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271 (1547): 1443–1450.

Bernstein, F., and A. Federgruen. 2004. A general equilibrium model for industries with price and service competition. *Operations Research* 52 (6): 868–886.

Boada, P., and V. Martínez-de-Albéniz. 2014. Estimating and Optimizing the Impact of Inventory on Consumer Choices in a Fashion Retail Setting. Working paper, IESE Business School.

Boyd, S. P., and L. Vandenberghe. 2004. *Convex optimization.* Cambridge university press.

Caro, F., and V. Martínez-de Albéniz. 2012. Product and price competition with satiation effects. *Management Science* 58 (7): 1357–1373.

Caro, F., and V. Martínez-de-Albéniz. 2014. How Fast Fashion Works: Can It Work for You, Too? *IESE Insight Review* 21:58–65.

Caro, F., V. Martínez-de-Albéniz, and P. Rusmevichientong. 2014. The Assortment Packing Problem: Multiperiod Assortment Planning for Short-Lived Products. *Management Science* Forthcoming:N/A.

Christopher, M., R. Lowson, and H. Peck. 2004. Creating Agile Supply Chains in the Fashion Industry. *International Journal of Retail and Distribution Management* 32 (8): 367–376.

Çınar, E., and V. Martínez-de-Albéniz. 2013. A Closed-loop Approach to Dynamic Assortment Planning. Working paper, IESE Business School.

Crain, W. M., and R. D. Tollison. 2002. Consumer choice and the popular music industry: A test of the superstar theory. *Empirica* 29 (1): 1–9.

DPA 2014. El nombre Neymar causa furor en Perú. *Marca.com* 26 July:online.

Elberse, A. 2008. Should you invest in the long tail? *Harvard business review* 86 (7/8): 88.

Elberse, A., and J. Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science* 22 (3): 329–354.

Eliashberg, J., A. Elberse, and M. A. Leenders. 2006. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science* 25 (6): 638–661.

FashionUnited.com 2014. Global fashion industry statistics - International apparel. online. Accessed on July 30, 2014.

Favaro, D., and C. Frateschi. 2007. A discrete choice model of consumption of cultural goods: the case of music. *Journal of Cultural Economics* 31 (3): 205–234.

Federgruen, A., and N. Yang. 2009. Optimal supply diversification under general supply risks. *Operations Research* 57 (6): 1451–1468.

Fisher, M. 2009. OR FORUM–Rocket Science Retailing: The 2006 Philip McCord Morse Lecture. *Operations Research* 57 (3): 527–540.

Fryer, R. G., and S. D. Levitt. 2004. The causes and consequences of distinctively black names. *The Quarterly Journal of Economics* 119 (3): 767–805.

Hahn, M. W., and R. A. Bentley. 2003. Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270 (Suppl 1): S120–S123.

Harvey, D. 1989. *The condition of postmodernity*, Volume 14. Blackwell Oxford.

Hopp, W. J., and X. Xu. 2005. Product line selection and pricing with modularity in design. *Manufacturing & Service Operations Management* 7 (3): 172–187.

Hopp, W. J., and X. Xu. 2008. A static approximation for dynamic demand substitution with applications in a competitive market. *Operations Research* 56 (3): 630–645.

Kök, A. G., and M. Fisher. 2007. Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application. *Operations Research* 55:1001–1021.

Kök, A. G., M. L. Fisher, and R. Vaidyanathan. 2009. Assortment planning: Review of literature and industry practice. In *Retail Supply Chain Management*, 99–153. Springer.

Krider, R. E., and C. B. Weinberg. 1998, February. Competitive dynamics and the introduction of new products: The motion picture timing game. *Journal of Marketing Research* 35 (1): 1–15.

Lévy-Garboua, L., and C. Montmarquette. 1996. A microeconometric study of theatre demand. *Journal of cultural economics* 20 (1): 25–50.

Louviere, J. J., and D. A. Hensher. 1983. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research* 10 (3): 348–361.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. Academic Press, New York.

Neelamegham, R., and P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science* 18 (2): 115–136.

Prieto-Rodríguez, J., and V. Fernández-Blanco. 2000. Are popular and classical music listeners the same people? *Journal of Cultural Economics* 24 (2): 147–164.

Sawhney, M. S., and J. Eliashberg. 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science* 15 (2): 113–131.

Stanley, R. P. 2011. *Enumerative combinatorics*, Volume 1. Cambridge university press.

Talluri, K. 2009. A finite-population revenue management model and a risk-ratio procedure for the joint estimation of population size and parameters. Working paper, UPF.

Tan, T. F., and S. Netessine. 2009. Is Tom Cruise threatened? Using Netflix Prize data to examine the long tail of electronic commerce. Working paper, Wharton, University of Pennsylvania.

The Associated Press 2013. Hollywood, Creative Industries Add $504 Billion to U.S. GDP. *The Hollywood Reporter* December 5:online.

Thompson, D. 2012. The 100-Year March of Technology in 1 Graph. *The Atlantic* April 7:online.

Train, K. E. 2009. *Discrete choice methods with simulation*. Cambridge university press.

Vulcano, G., G. van Ryzin, and W. Chaar. 2010. Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. *Manufacturing & Service Operations Management* 12 (3): 371–392.

Vulcano, G., G. Van Ryzin, and R. Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* 60 (2): 313–334.

Whittle, D. 2013, November 12. Going live with a mobile app: Part 1 Planning a mobile app.

Yoganarasimhan, H. 2012. Identifying the Presence and Cause of Fashion Cycles in the Choice of Given Names. Working paper, University of California Davis.