HOW SMARTPHONE APPS CAN HELP PREDICTING MUSIC SALES

Nikolaeva, Irina, Technical University Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany, nikolaeva@emarkets.tu-darmstadt.de

Hinz, Oliver, Technical University Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany, hinz@emarkets.tu-darmstadt.de

Abstract

Predicting music sales is of particular interest for sales managers (e.g. for pricing), inventory management (for CD sales) and server balancing (for music download). In the past years, research therefore proposed several models for music sales prediction. These models have, however, some shortcomings which we want to overcome with a new approach. We suggest using a novel data set that is a byproduct of smartphone apps that help users to identify music. Shazam is probably the most popular of these music identification services for smartphones. This study examines the relationship between Shazam charts and song sales using data from the UK over a period from September 2010 to May 2011. Using seemingly unrelated regression we identify that Shazam charts precede sales charts by about two weeks and serve as good predictors for sales charts. Further, we find that the rock, pop, and hip hop genre and artist's popularity positively affect song sales.

Keywords: User-Generated Content, Prediction, Social Tagging, Shazam.

1 Introduction

The modern consumer faces the problem to choose the best product or service from a vast amount of new releases. Moreover, in the area of entertainment products the quality cannot be evaluated before consumption (Arrow, 1962). Although offline and online stores provide audio and reading tests, the search and choice of the best product remains time consuming. With the purpose to deal with the "product overload" rankings are well established in entertainment area, e.g. top seller books, album and single charts. Consumers use rankings as a compass when making buying decisions (Hinz and Eckert, 2010). Online and offline sellers anticipate consumer's behaviour and provide the top seller products well positioned places in their stores. If a product achieves the top 10 or top 20 charts it experiences carry-over effects (Clement et al., 2010) and benefits from a positive feedback loop. Further, product rankings serve as a performance feedback for firms. They have implications for immediate revenue generation as well as for future releases, concert attendances, and product endorsements. Sales managers can adapt their short-term pricing and inventory decisions (Moe and Fader, 2002; Lee et al., 2003; Hann et al., 2011) and music download stores can better balance server load. Further, sales managers can boost products which are on the verge of entering top 10 or 20 charts with the purpose to benefit from carry-over effects (Clement et al., 2010).

Former research proposed several models for the prediction of future song's and album's sales and their survival in charts, e.g. Lee et al. (2003), Hann et al. (2011), Strobl and Tucker (2000). Previous studies used advanced purchase orders (Moe and Fader, 2002), MySpace plays (Chen and Chellappa, 2009), blog post volume (Dhar and Chang, 2009; Dewan and Ramprasad, 2009), or artist's past performance (Lee et al., 2003) as predictors for future sales. However, the availability of such kind of data is not assured. Hence, we propose an alternative song's ranking prediction model which makes use of steadily available data that is moreover easy to analyse.

Shazam is an application (colloquially called "app") which allows consumers to identify the artist and song name of a music track. A user calls up the Shazam service on his mobile device and samples up to 15 seconds of music being played. The sample is sent to a server where it is compared to the songs in the database. After successfully matching the track title, album and artist's name are sent back to the user. The song is saved on the user's tag list, so that the user can buy it later (Shazam, 2010). Shazam aggregates individual tags to daily charts, which are freely available on the provider's website. There exist other music identification services e.g. SoundHound, Tunatic. Shazam data offers in comparison to other services following advantages. First, Shazam is the oldest and the most popular music identification service. Google Trends shows a steadily growing search rate for Shazam since its release in 2008. Second, no other service provides an access to the music tagged by the users. Third, Shazam has about 150 Million users across all platforms and is freely available. Hence, we expect that songs tagged using Shazam represent preferences for its large user base and may be used to predict future sales.

We believe that Shazam provides valuable information about the latent consumer preferences. While "shazaming" a favoured song, app users act spontaneously and independently of each other. On the aggregated level such consumers' behaviour data become "wisdom of crowds" and offer a strong explanatory power of a new song's success. This data will help researchers as well as companies to better understand the consumers' tastes in music and will also offer the opportunity for new intermediaries to broker such information.

Before adopting innovations or buying new products, the prospective buyer needs to get aware of the innovation (Rogers, 1995). This stage of the diffusion process is typically not observable. The newly available tagging behaviour of smartphone users might overcome this missing information and serve as a very good proxy for the awareness stage in the diffusion process. If this is the case, this information might also serve as a success indicator for the subsequent adoption stage.

Figure 1 shows an example of two songs: This figure depicts a typical lifecycle in the Shazam charts and the actual sales charts. This example illustrates that Shazam indeed precedes actual sales in some cases. Therefore, we develop a prediction model for songs sales incorporating the information from the Shazam charts. We collected the Top-100 Single and Shazam charts from the UK over the period of 40 weeks. Then we investigated the relationship between Shazam and sales charts using seemingly unrelated regression while controlling for song's and artist's characteristics as well as for promotional effort.

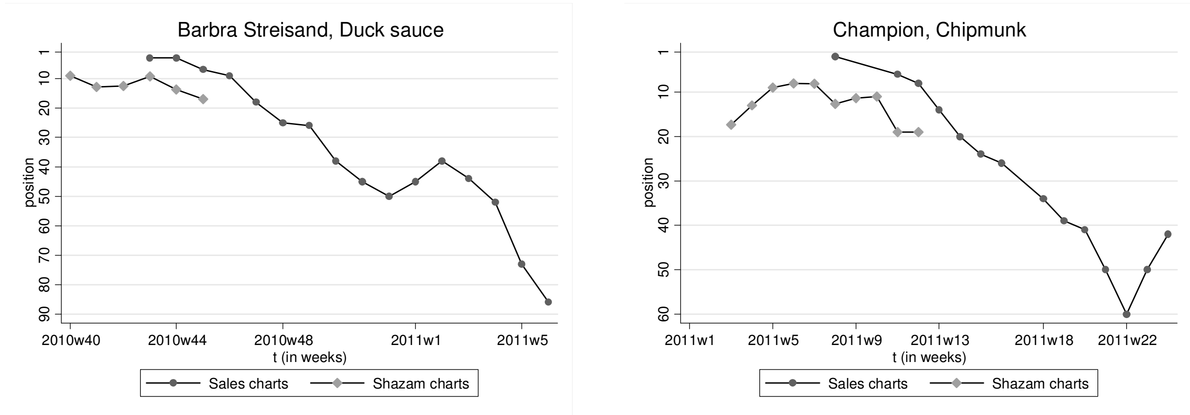

Figure 1. Diffusion pattern of new released songs.

The remainder of the paper is structured as follows. In Section 2 we discuss former research and outline the shortcoming of existing prediction models. In Section 3 we present our model, describe our data set and provide estimation results. Finally, we discuss managerial implications and limitations of our study and give suggestions for further research.

2 Previous Research on Music Sales Prediction

Previous research proposed several models for predicting music sales. These models differ in timing when the forecasts are made (pre-release vs. after-release), in used predictors, and in applied forecasting methods. The following summarises and discusses previous research.

Two studies considering pre- and post-release forecasts are widely cited: Moe and Fader (2002) and Lee et al. (2000). Both studies use a Hierarchical Bayes Model, which consists of two stages: First, these models build prior estimates and then update them when new sales data are available. The main difference between these studies exists in building priors. The study of Moe and Fader (2002) uses advance purchase orders, whereas Lee et al. (2003) use artist's past performance data like the total number of released albums or the number of gold and platinum albums.

These approaches provide good models for forecasting, but they have some shortcomings. First, the model of Moe and Fader (2002) can only be applied by sellers who offer opportunities for advanced orders. With the emergence of online stores and growing number of music downloads consumers are less likely to make advanced orders any more. Even when advanced orders are available, they are likely to exist only for already established artists. Thus, the model cannot account for newcomers, whose albums or songs may be associated with higher risks by consumers. Similarly, the study of Lee et al. (2003) cannot account for newcomers for whom there is no historical data, so that the building of prior forecasts is not possible. Further, it is unclear whether past sales in general are an appropriate and a reliable source of information: 3 platinum albums do not prevent an artist from future failures.

Hence, the success of prediction models is highly dependent on the data availability. Therefore, it is important to have alternative sets of predictors. Nowadays, with the emergence of Web 2.0 technologies and ubiquitous computing marketers and forecasting experts have access to a novel kind of data. User-generated content becomes an important source of information about consumers' activities, opinions and preferences (Hinz et al., 2011, Decker and Trusov, 2010; Dhar and Chang, 2009; Godes and Mayzlin, 2004).

The following five studies use such data sets to predict music sales. Chen and Chellappa (2009) investigate the influence of MySpace plays on the song and album sales. They find that MySpace plays can significantly predict the song's sales. Bischoff et al. (2009) propose a learning classifier method to predict whether the song will be a hit or not. As predictors they use social tags on last.fm, a music website. Dewan and Ramprapasad (2009) and Dhar and Chang (2009) propose models for estimating album sales using the music blog buzz. Hann et al. (2011) use demand and supply data from peer-to-peer networks to make pre-release forecasts for music sales. Our study ties in with studies which use novel kind of data and proposes a new model for predicting music sales using charts of tagged songs on the website of the music identification software Shazam. Whereas the approaches based on MySpace, music blog, and peer-to-peer networks data include cumbersome processes of data gathering, the advantage of Shazam data is their availability in already structured and ready-to-analyze form. Moreover, it is good for music managers to have alternative data sources to predict future music sales.

3 Modelling and Estimation

3.1 Model Development and Variables

This model aims to predict a song's position in the sales charts in week t $poscharts_t$ using Shazam data. According to the theory of Rogers (1995) the diffusion process consists of 5 stages: knowledge (awareness), persuasion, decision (adoption), implementation and confirmation. For our study two of these stages are of particular interest: awareness and adoption. Shazam and sales charts serve as a feedback of song's popularity. We suppose that Shazam charts can be used as proxy for the awareness stage while sales charts represent the actual adoption. Whereas users call up the Shazam service spontaneously, buying a song is a rational act and needs time. Therefore, we expect that Shazam charts precede sales charts and may serve as predictor for future sales. Figure 2 shows a song's typical lifecycle in Shazam and sales charts. To better explain the relationship between the two charts and for easier interpretation of the results we used a reversed scale for song's position in sales and Shazam charts: position one is the worst position and 100 in sales and 20 in Shazam charts are the best positions respectively. On the first glance sales charts seem to be a parallel shift of Shazam charts. Hence, we expect that the song's current position in Shazam charts $posshazam_t$ may predict its current position in sales charts and include it into our estimation equation.

Considering points A and B (which have about the same positions in Shazam charts) one can see that the corresponding positions in sales charts are different. We suppose that the song's relative position in the lifecycle accommodates this difference. Point A lies on the ascending part of the song's lifecycle in Shazam charts, point B on the descending branch. According to the preceding nature of Shazam charts, we expect that if a song is ascending in the Shazam charts, it is ascending in the sales charts (i.e. it receives relatively worse positions due to the initial stage of the adoption process). If a song is descending in Shazam charts, it either holds its position (saturation phase) or is already descending in sales charts. To capture this relationship we include the song's change in position from the previous week as well as its lags, i.e. the position changes in the past in Shazam charts $\sum_{\tau=0}^{T} \alpha_\tau (posshazam_{i,t-\tau} - posshazam_{i,t-\tau-1})$ in our estimation equation. The corresponding hypothesis is:

H1: A s*ong's current position and lagged position differences in Shazam charts affect its position in sales charts.*

Because Shazam and sales charts belong to two stages of the song's adoption process, they might be influenced by a common set of other factors. Hence, we introduce a second equation which includes song's ranking position in Shazam charts $posshazam_t$ on the week *t* on the left-hand side. Shazam charts serve as predictor for song's sales charts position and shares partially the influence of a set of exogenous factors with sales charts. Figure 3 depicts our final model which we successively discuss in the following.

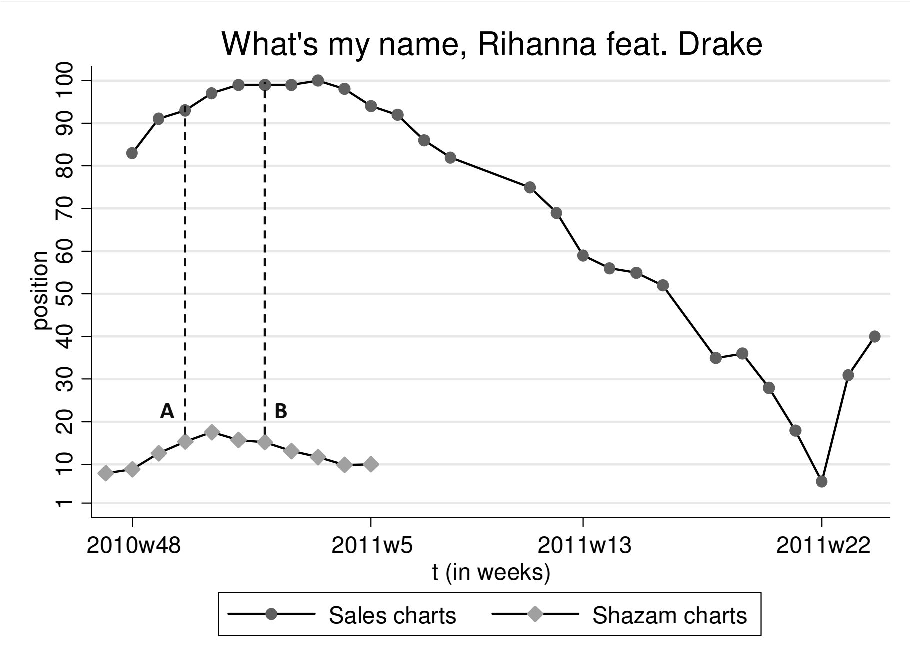

Figure 2. Relationship between song's positions in Shazam and sales charts.

The more a song is promoted, the more aware consumers are of it, the higher the subsequent adoption (sales). Because record companies treat these data as very confidential, promotional effort is usually measured by proxies: Lee et al. (2003), Montgomery and Moe (2002) and Chen and Chelappa (2009) use airplay to measure promotional efforts; Hann et al. (2011) use the number of comments on YouTube before release; Dhar and Chang (2009) and Dewan and Ramprapasad (2009) use the label as proxy for promotional efforts (major labels have more budget to promote songs or albums). We expect that airplay charts capture best the dynamical nature of promotional efforts and include song's current position in airplay charts $posairplay$ as well as lagged position differences $\sum_{\tau=0}^{T} \beta_\tau (posairplay_{i,t-\tau} - posairplay_{i,t-\tau-1})$ into the equation for song's position in Shazam. While previous models investigated promotional effects on sales directly, our model can make an interesting distinction: We propose that promotional effort raises consumers' awareness of new songs, and higher awareness drive the subsequent sales. Shazam charts can be seen as consumers' response to promotional effort. Thus, the hypothesis H2 proposes:

H2: *A song's current position and lagged position differences in the airplay charts have impact on song's position in Shazam charts.*

Further, we assume that both diffusion stages are affected by common factors such as artist and song characteristics as well as time effects. We expect that Shazam charts and music sales follow some time-dependent patterns. In the linear case, the sales and consumers' interest in a new song are higher right after the release of the song. With elapsing time the consumers already know the song and there is no need to "shazam" the song, so that the song loses positions in the Shazam ranking. Similarly sales go down after some time. We measure time effects by including elapsed weeks since song's release $weeksrelease_{it}$ and allow for different lifecycle forms by including a quadratic and cubic term for the time since release.

H3: *Song's position in Shazam* (H3a) *and sales charts* (H3b) *is affected by time effects.*

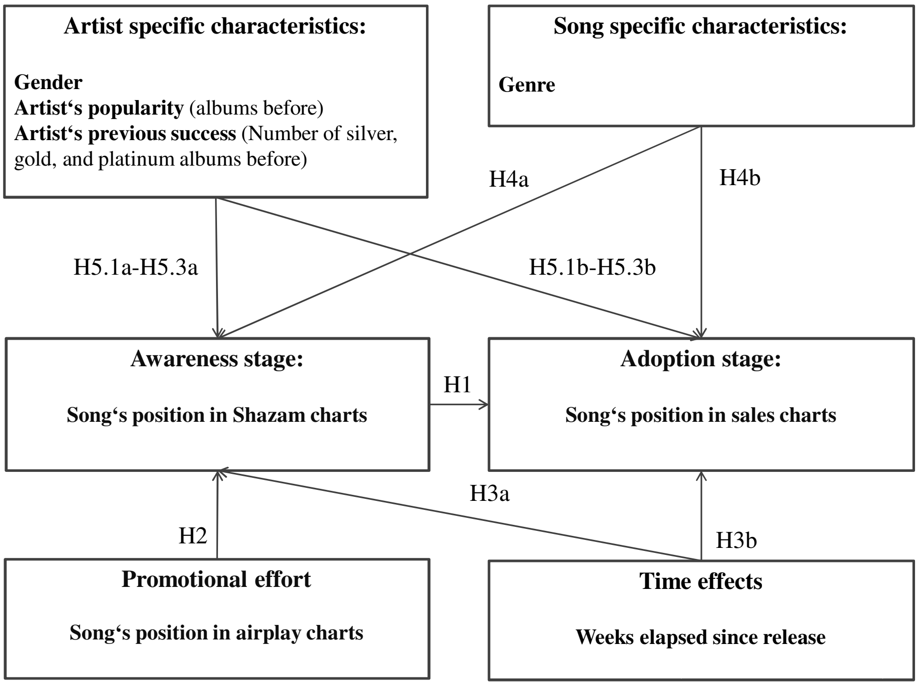

Figure 3. Relationship between Shazam and sales charts.

Further, music genre is expected to influence awareness and sales patterns (Lee et al., 2003; Hann et al., 2011). We include a set of dummy variables $\sum_{j=1}^{J} \gamma_j \, genre_{ij}$ into both equations. $genre_{ij}$ equals 1, if song *i* is of genre *j* and equals 0 otherwise.

H4: A *song's position in Shazam* (H4a) *and sales* (H4b) *charts is affected by the genre the song belongs to.*

The next set of factors which might influence Shazam and sales charts is artist's characteristics (Lee et al., 2003; Moe and Fader, 2002; Hann et al., 2011). These variables are 1) gender of the artist, 2) artist's popularity 3) artist's previous success. We distinguish between solo female, solo male and band and include a set of dummy variables $\sum_{g=1}^{G} \delta_g \, gender_{ig}$ into both equations, where $gender_{\,ig}$ equals 1 if the artist of the song *i* is of gender *g*, and 0 otherwise. Because we cannot observe artist's popularity and previous success directly, we use proxies. We follow Lee et al. (2003), Hann et al. (2011), Moe and Fader (2002) and measure artist's popularity by the number of previous albums $albumsbefore_i$. We suppose the effect of this factor to be different for both charts. We expect that newcomers are more often "shazamed" than well-known artists, whereas well-known artists may profit from their established fan base when it comes to sales. We distinguish between artist's popularity and artist's previous success. An artist may be well-known, but might not have hits yet. Thus, we include number of silver $UKsilver_i$, gold $UKgold_i$, and platinum albums $UKplatinum_i$ with purpose to control for previous success. The next set of hypotheses looks as follows:

H5.1: A s*ong's position in Shazam* (H5.1a) *and sales* (H5.1b) *charts is affected by artists's gender.*

H5.2: A s*ong's position in Shazam* (H5.2a) *and sales* (H5.2b) *charts is positively affected by artists's popularity.*

H5.3: A s*ong's position in Shazam* (H5.3a) *and sales* (H5.3b) *charts is positively affected by artists's previous success.*

Our final estimation equations look as follows:

$$1)\quad poscharts_{it} = \sum_{\tau=0}^{T} \alpha_\tau (posshazam_{i,t-\tau} - posshazam_{i,t-\tau-1}) + \omega_{1c}\, weeksrelease_{i,t} + \omega_{2c} weeksrelease_{i,t}^2 + \omega_{3c} weeksrelease_{i,t}^3 + \sum_{j=1}^{J} \gamma_{jc}\, genre_{i,j} + \sum_{g=1}^{G} \delta_{gc}\, gender_{i,g} + \theta_{1c} albumsbefore_i + \theta_{2c} UKsilver_i + \theta_{3c} UKgold_i + \theta_{4c} UKplatinum_i + \varepsilon_{i,c}$$

$$2)\quad posshazam_{it} = \sum_{\tau=0}^{T} \beta_\tau (posairplay_{i,t-\tau} - posairplay_{i,t-\tau-1}) + \omega_{1s}\, weeksrelease_{i,t} + \omega_{2s} weeksrelease_{i,t}^2 + \omega_{3s} weeksrelease_{i,t}^3 + \sum_{j=1}^{J} \gamma_{js}\, genre_{i,j} + \sum_{g=1}^{G} \delta_{gs}\, gender_{i,g} + \theta_{1s} albumsbefore_i + \theta_{2s} UKsilver_i + \theta_{3s} UKgold_i + \theta_{4s} UKplatinum_i + \varepsilon_{is}$$

The subscripts s and c of the coefficients refer to the first and the second equation respectively.

3.2 Data

The data analysed in this study consist of three panel data sets, namely sales charts, Shazam charts and airplay charts for the UK for a period of 40 weeks from September 2010 to June 2011. The *Top-100 Single* sales charts are provided on a weekly basis on http://www.theofficialcharts.com and comprise weekly offline and online sales (Official Charts Company (2009)). For detailed information concerning charts rules see (Official Charts Company, 2009). We collected daily Shazam charts data, which comprises 20 positions, from the homepage http://www.Shazam.com.

Further, we used airplay data to control for promotional effects. Airplay data was friendly provided by the broadcasting monitoring service provider "radiomonitor" (Official web site radiomonitor.de). We thank "radiomonitor" for supporting our research.

We gathered additional information about songs and artists from discogs.de, Wikipedia, and from artist's personal web pages. We classified observed songs into six genres: rock, pop, hip hop, jazz, folk, and alternative music (all songs not categorized to the first six genres).

Our data comprises 11,600 observations for 811 singles that appeared at least once in sales, Shazam, or airplay charts. Shazam charts data are on a daily basis in contrast to sales and airplay charts. Therefore, we aggregated the daily Shazam charts positions to weekly Shazam charts ranking by calculating the average of the particular week. Unfortunately, we could not find all information for all songs. Thus, we removed 118 songs due to incomplete information. Further, most songs have short durations in all three charts, so that many songs cannot be used in our estimation. We excluded songs with a duration of less than four weeks in Shazam charts and less than three weeks in sales and airplay charts. Further, we exclude our variables for jazz and folk genre due to collinearity issues. In a final estimation we used 195 observations for 61 songs.

3.3 Estimation Results

We estimated our set of equations simultaneously using seemingly unrelated regression (Zellner, 1962). Further, we expect that song's positions in all charts are interdependent within the week. Therefore, we clustered our observations for the particular week and used robust standard errors.

Table 1 provides our final estimation results. The song's current position in Shazam charts as well as its first and lagged first difference predicts the current ranking position of the song in sales charts. The coefficients amount to 1.960, -0.480, and -0.469, respectively. The first coefficient of 1.960 has two implications. First, when we observe a song in both charts, its position in the sales charts is c.p. about twice its position in the Shazam charts (level effect). Second, changes in the current position in Shazam charts have a moderate effect in the sales charts: If the song's current position in Shazam charts increases by one position, its position in sales charts increases by about two.

Dependent Variable: ***posshazam***				**Dependent Variable:** ***poscharts***			
Ind. variable	Coeff.	St.error	p-value	Ind.variable	Coeff.	St.error	p-value
Dummy pop	-2.144	2.022	0.289	Dummy pop	7.406	3.925	0.059*
Dummy rock	-3.668	2.091	0.079*	Dummy rock	8.630	4.342	0.047**
Dummy hiphop	-2.037	2.108	0.334	Dummy hiphop	6.558	3.932	0.095*
Dummy female	0.498	0.916	0.586	Dummy female	2.242	1.744	0.199
Dummy band	1.060	1.117	0.342	Dummy band	-1.238	2.031	0.542
Albums before	-0.622	0.314	0.048**	Albums before	1.131	0.478	0.018**
Number of silver albums	-0.420	0.889	0.637	Number of silver albums	-0.088	1.482	0.953
Number of gold albums	1.516	1.149	0.187	Number of gold albums	-4.387	1.850	0.018**
Number of platinum albums	0.291	0.147	0.048**	Number of platinum albums	-0.233	0.215	0.278
Weeks since release	0.762	0.211	0.000***	Weeks since release	-0.560	0.147	0.000***
Weeks since release (quadr.)	-0.035	0.010	0.001***	Weeks since release (quadr.)	0.005	0.002	0.018**
Weeks since release (cub.)	0.001	0.000	0.000***	Shazam position at time *t*	1.960	0.334	0.000***
Airplay position at time *t*	0.098	0.027	0.000***	Difference between Shazam positions at time *t* and *t-1*	-0.480	0.196	0.014***
Difference between airplay positions at time *t* and *t-1*	0.064	0.038	0.095*	Difference between Shazam positions at time *t-1* and *t-2*	-0.469	0.209	0.025**
Difference between airplay positions at time *t-1* and *t-2*	0.164	0.048	0.001***	Constant term	67.441	4.725	0.000***
Constant term	4.470	1.982	0.024**	*p<0.1, **p<0.05, ***p<0.01			

Table 1. Estimation results of seemingly unrelated regression.

The latter two coefficients imply that week-to-week changes in the Shazam charts in the last two weeks have an additional significant negative effect on the sales charts position. That is, a song which has improved its position in the Shazam charts in the past two weeks, i.e. it is on the ascending part of the Shazam lifecycle, will have a negative lagged and non-lagged first difference and thus a worse current sales charts position. Likewise, when a song is on the descending part of the Shazam lifecycle and its ranking worsened in the last two weeks, the differences will be positive and the current sales charts position will be better. Thus our estimation results support our conjecture that Shazam charts precede sales charts by about two weeks. Further, our model accounts for all other cases in life cycle patterns, e. g. when the first difference is positive and the lagged difference is negative, we observe slight changes in sales ranking position.

Next, our estimation results show that the promotional effort affects significantly song's current position in Shazam charts. The coefficients for the first and the lagged differences are positive. This indicates that airplay and Shazam charts proceed parallel. If promotional activities for the song increase in previous weeks, the song's likelihood to be "shazamed" increases. Promotion activity for a song might trigger subsequent promotion activities and further airplay (e.g. on TV or DJs might play the song more often in clubs).

Our estimation moreover shows that gender has no effect on the songs' sales and on the chance of being tagged with Shazam. The song's genre does not affect its likelihood, except for rock category, of being "shazamed" but explains its position in sales charts. Rock songs are "shazamed" about 4 ranking points less in contrast to alternative genre (constant term). In sales charts rock, pop and hip hop music obtain about 9, 7, and 7 ranking points, respectively, more than alternative genre. Assuming that Shazam charts are a valid sample of the general population, this result might indicate that the rock, pop and hip hop genres are favoured in the population.

Next, we consider the effect of the artist's popularity and previous success. Our results indicate that the more albums an artist has, the less his new released song is "shazamed". Well-known artists are recognised by their voices, so that the users do not need to "shazam" their songs. In sales charts the number of previous albums accounts for a positive effect. Well-known artists may thus benefit from an established fan base. We account for previous success using number of silver, gold, and platinum albums. The songs of platinum album holders are more often tagged using Shazam. A similar effect appears in sales charts; for each previously released album an artist obtains about one ranking point more for the newly released song. Although the effect is not especially large, it may nevertheless indicate that artists who possess platinum albums have more capabilities to meet consumers' preferences and release new hit songs. A newly released song obtains about 4 ranking positions less for each additional gold album. This finding is ambiguous: It is more difficult for newcomers (no previous albums yet) to reach high positions in the rankings while it is also difficult for well-established successful artists with prior bestseller albums to repeat their success.

Further, we controlled for time-dependent patterns in Shazam and sales charts. All time variables are significant. The linear and the quadratic value of weeks since release represent best the time-dependent trend in sales charts, while the Shazam charts follow a more sophisticated lifecycle which is reflected by a significant cubic time trend variable.

The constant terms stand for alternative music interpreted by male artists and amount to 67.441 (sales charts) and 4.47 (Shazam charts) ranking positions respectively.

4 Discussion

Former studies propose pre- and post-release forecasting models, using different data as predictors and different estimation methods. The success or applicability of the proposed methods strongly depends on the availability of relevant data. In this study, we focus on the data created by users of the music identification service Shazam in the UK through tagging of their favoured music.

We find that Shazam charts precede sales charts by two weeks and serve as good predictors of sales charts. Further, we find that the rock, pop, and hip hop genre, artist's number of previous albums positively correlate with sales charts. Therefore, our model can be used as an alternative prediction model and offers the following advantages:

First, Shazam charts are freely available and offer thereby the first proxy for prospective buyers' awareness. The data can be used by a wide variety of stakeholders. Whereas e.g. advanced purchase orders are available only for retailers, Shazam data can be used by event managers and concert planners, retailers and server farm admins who cannot access advanced orders.

Second, Shazam tags reflect consumer preferences that are not distorted by price effects and thus offer information for sophisticated pricing strategies. Prices for promising singles that do well on Shazam could be offered more expensive to extract more consumer surplus. Or singles that might reach sales charts could be offered more cheaply in order to reach certain cut-off positions (e.g. position 10 or 20) to realize carry-over effects.

Third, Shazam data offer a means to reveal promising newcomers. While many other prediction models build upon the past success of artists and thus cannot predict the success of newcomers, our approach allows predicting the short-term sales of newcomers. This would allow labels to renew the contract or event manager to arrange concerts. Artists should also keep an eye on their performance on the Shazam charts in order to delay contract renewals to anticipate success and get better offers. These data offers interesting information for the negotiation process.

Fourth, in contrast to prediction models that analyse other Web2.0 data e.g. MySpace plays or blogs, Shazam tags data are already provided in a structured, ready-to-use form and can simply be used to extend existing models.

One major shortcoming of our model is that it can only be used to forecast the success of songs that stay several weeks in the charts. Songs staying in the charts less than 4 weeks are not reflected in our model since we had to discard this information. This shortcoming certainly opens venues for further research and new modelling approaches.

References

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. Nber.

Bischoff, K., Firan, C., Georgescu, M., Nejdl, W. and Paiu R. (2009). Social knowledge-driven music hit prediction. Advanced Data Mining and Applications, 43-54.

Chen, C. and Chellappa, R. (2009). Music Sales and Online User Activity at Social Networking Sites. Working Paper, Goizueta Business School, Emory University.

Clement, M., Hille, A., Lucke, B., Schmidt-Stölting, C. and Sambeth, F. (2010). Der Einfluss von Rankings auf den Absatz-Eine empirische Analyse der Wirkung von Bestsellerlisten und Rangpositionen auf den Erfolg von Büchern. Zeitschrift für betriebswirtschaftliche Forschung

Decker, R. and Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing, 27**,** 293-307.

Dewan, S. and Ramprasad, J. (2009). Chicken and Egg? Interplay Between Music Blog Buzz and Album Sales. PACIS 2009 Proceedings**,** 87.

Dhar, V. and Chang, E. (2009). Does chatter matter? The impact of user-generated content on music sales. Journal of Interactive Marketing, 23**,** 300-307.

Godes, D. and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. Marketing Science**,** 545-560.

Hann, I. H., Oh, J. H. and James, G. (2011). Forecasting the Sales of Music Albums: A Functional Data Analysis of Demand and Supply Side P2P Data. Working paper.

Hinz, O. and Eckert, J. (2010). The Impact of Search and Recommendation Systems on Sales in Electronic Commerce. Business & Information Systems Engineering, 52**,** 67-77.

Hinz, O., Skiera, B., Barrot, C. and Becker, J. (2011). Seeding Strategies for Viral Marketing: An Empirical Comparison. Journal of Marketing, 75 (6)**,** 55-71.

Lee, J., Boatwright, P. and Kamakura, W. A. (2003). A Bayesian model for prelaunch sales forecasting of recorded music. Management Science**,** 179-196.

Moe, W. W. and Fader, P. S. (2002). Fast-Track Article: Using Advance Purchase Orders to Forecast New Product Sales. Marketing Science**,** 347-364.

Montgomery, A. L. and Moe, W. W. (2002). Should Music Labels Pay for Radio Airplay? Investigating the Relationship Between Album Sales and Radio Airplay. Working paper.

Official Charts Company (2009). Official Charts Company: Rules For Chart Eligibility. Singles. http://c0903002.cdn.cloudfiles.rackspacecloud.com/25-official-uk-singles-chart-rules-august-2009.pdf

Rogers, E. M. (1995). Diffusion of innovations, Free Pr.

Shazam (2010). http://www.shazam.com

Strobl, E. A. and Tucker, C. (2000). The Dynamics of Chart Success in the U.K. Pre-Recorded Popular Music Industry. Journal of Cultural Economics, 24, 113-134.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association**,** 348-368.