Bending the Curve: Modeling *The Bends* Album Sales

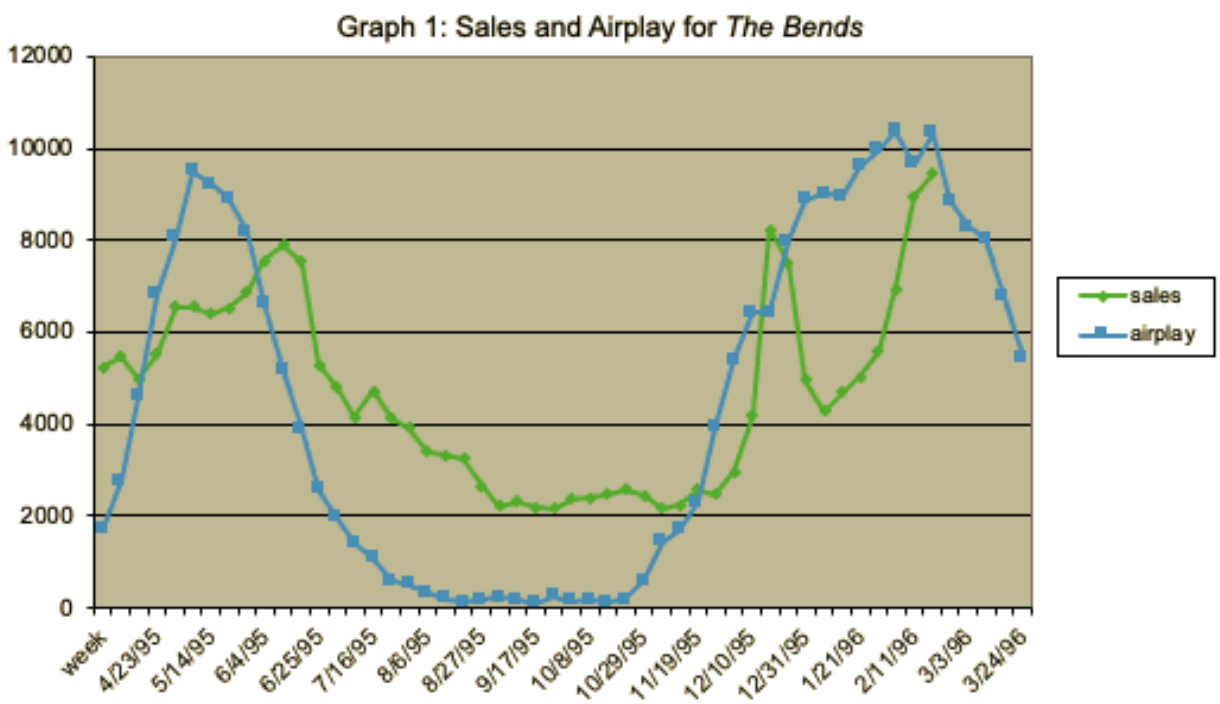
Anonymous

April 12th 2022

EXECUTIVE SUMMARY

One of the most important components of the latter half of this course has been selecting tools appropriate to model different types of data, specifically in deciding how to best encapsulate the tradeoff between heterogeneity, duration dependence and covariates in a real-life application. For album sales over the course of a year for Radiohead's 1995 album *The Bends*, the observed heterogeneity explained by covariates appears to be the main driving force behind such trends in the data; for this reason, a Burr(XII) with the covariates for radio airplay after release, radio airplay representing end-of-year popularity gain and a Christmas sales boost appeared to best model the underlying trends without closely overfitting the data. This model was chosen for parsimony, with little empirical reason to include segmentation, and story, given that the rise in airplay closely mirrors the rise in mainstream popularity Radiohead experienced at the end of 1995. The split was made in airplay to provide a distinction between the general decrease in sales in months following an album's release date and this unique gain in mainstream attention for Radiohead following. In subsequent studies, it would be interesting to fix the lag effect associated with airplay and find another covariate as a general metric for a rise in popularity following a band's breakthrough release.

1 INTRODUCTION



As an avid Radiohead fan, I feel obligated to lend some future listening suggestions to anyone unacquainted with the band before starting the analysis. The band's most critically acclaimed work is *Kid A*, which is somewhat infamous for being a huge departure in sound from their previous alternative-rock classic *OK Computer* and also being named the greatest album of the 2000s by Pitchfork. I would definitely urge you to check it out and listen straight through— it is a classic for a reason, and its dystopian-sounding, electronic-jazz-rock sound will forever be timeless, sounding like it came out yesterday when it is over 20 years old. My personal favorite is *In Rainbows*, a more traditional alternative album that is extremely beautiful, mellow and likable on first listen. If you don't have time for an album, Radiohead has amazing singles as well: "Paranoid Android," "Creep," "Karma Police," "Fake Plastic Trees" (from *The Bends*!), "Reckoner," "Idioteque"... The list goes on. If you're a fan of alternative music or good music in general, check them out!

1.1 *Key Question*

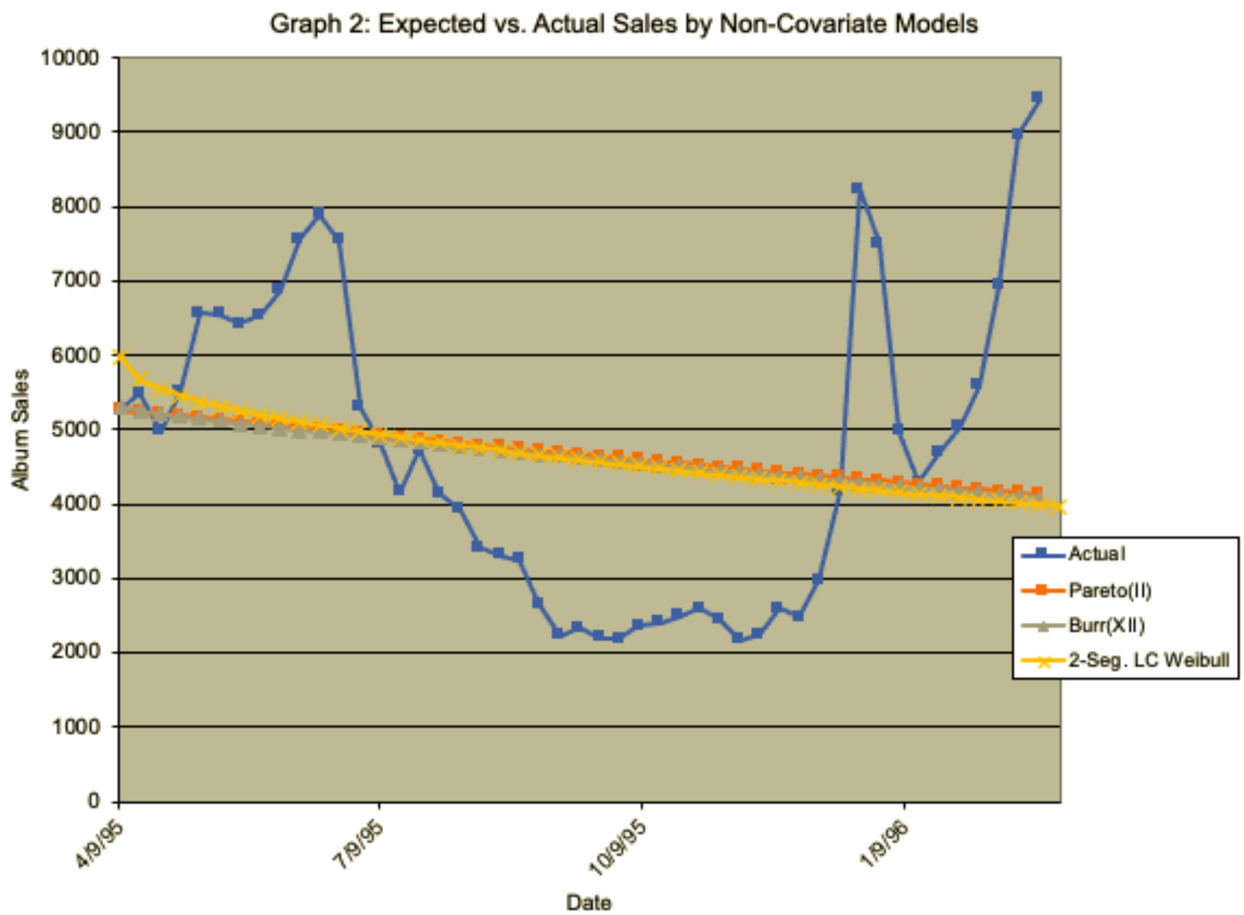
The dataset provided includes *The Bends* US album sales for the first 47 weeks after its release, along with its scaled numbers for US radio airplay in the same time frame. It is assumed each sales datapoint is an adoption for a unique customer, and the numbers associated with a given date are specifically for that day and the six days following. As such, the key question is very simple, but difficult in practice to answer: **what is the best method of modeling *The Bends* album sales data for the time period and why?**

2 ANALYSIS EXCLUDING COVARIATES

The choice of the baseline model to depict sales will be the one that best explains the unobserved heterogeneity by empirical metrics like the Bayesian Information Criterion and Mean Absolute Percent Error, as well as the most convincing story that informs the model as a whole. After hours of searching in hopes of establishing myself as a true pundit of *The Bends*, I settled on two convincing lines of reasoning to follow.

The first establishes a connection with the Burr(XII), and by the method of investigation, the Pareto(II) as well. Given weekly sales being readily classifiable as a continuous time dataset, these two models are the first to come to mind to create a baseline analysis. However, they also fit into the general story for how consumers vary in their likelihood of purchasing the album as more time goes by. Given the fact that depth-of-repeat models are not to be used, as what genius would buy two albums for themselves without deliberately trying to waste money, simplicity is king: individuals vary in their capacity to buy *The Bends* by a given week, with their differences in capacity captured in heterogeneity. Due to the fact that an album rollout is at mercy of time dynamics, as hype spikes upon release, slowly dying out over time, then potentially rising again by the end of the year as awards are given out, the Burr(XII) is a clear choice for a baseline model.

Despite this, all albums have their own idiosyncratic release to go along with their own unique bands. This brings in the second interpretation: a latent class Weibull that divides the buyer population into two distinct groups. Even before looking at the data, I figured segmentation would be a good place to start due to how publicly segmented Radiohead's fanbase was at the time! Coming off of their extremely successful single "Creep" (still their most successful to this day), the fanbase would be understandably divided into fans of the band, and lovers of the single who would have different capacities to buy at one time. The existence of a hard-core non-buyer population additionally seemed doubtful, largely due to how massive the population is in comparison with the total number of buyers viewed in the sample; how could Excel distinguish between these groups in such a short time period examined, as you can still buy *The Bends* to this day! Additionally, heterogeneity need not be included due to the ease of interpretability and large number of parameters without a reasonable story to support its inclusion. Still, the two-segment Weibull's story warrants its own model— the combined results of both options are shown below.



If only it were that easy. Observably, the models have such an awful fit that the word count for this section is begging me to not write about the parameter estimates, other than by noting that the two-segment story is essentially rejected with the parameter π set to 0.999999. Still, they are listed below if you're still interested.

Table 1: Non-Covariate Model Performance Data

Parameters	Pareto(II)	Burr(XII)	2-Seg. Weibull
r	1631653.435	1548.022193	
alpha	308900306.8	290477.5454	
lambda			0.005997645
c		0.995479349	0.966478277
pi			0.999999
lambda2			0.971429824
c2			0.988289939
LL	-1375364	-1375338	-1375230
BIC	2750755.445	2750755.445	2750528.64
MAPE	44.235%	43.696%	16.628%
R^2	1.213%	1.422%	2.535%

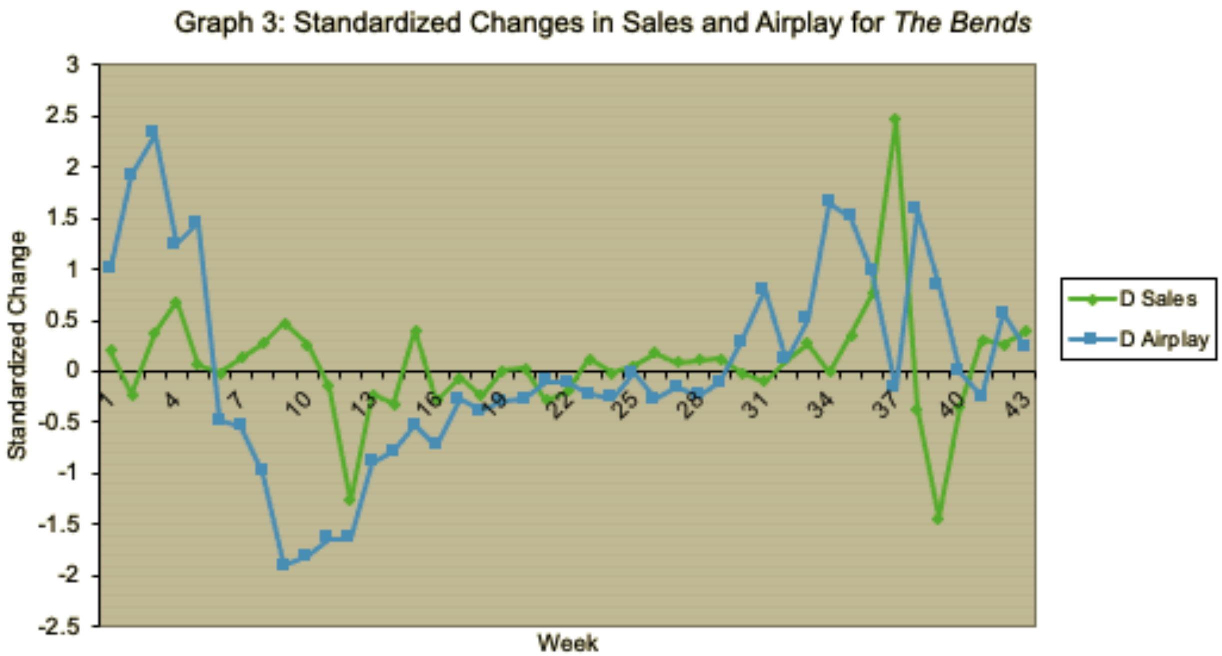
Despite this, although unobserved heterogeneity and duration dependence clearly play a role in how sales change over time, it is clear that such changes are more a direct result of observables changing simultaneously. The poor fit and reasonable explanation are convincing enough to drop these other models entirely.

3 COVARIATE SELECTION

This section investigates the immediately obvious covariates to be assessed and included in the model.

3.1 *Airplay*

The only other dataset included with sales is a key component to fitting the model, based on observable trends and a readily believable story. As more radio stations play music from an album, more people become exposed to the songs, and if the songs are particularly good, they may buy the album to listen to it on their own. This simple line of reasoning is the driving force behind many trends in sales, as seen in the graph below.



It is evident that a distinct change in airplay corresponds to a change in sales on a lag of approximately 3 weeks, although the magnitudes of each respective change are unlikely to be equal. This disproves one of my original theories of an endogeneity issue between sales and airplay, where it was thought a known increase in sales could incentivize radio stations to play songs off of a given album, which would in turn gain the album more exposure overall, with a cascading effect. However, the graph debunks this theory and proves that airplay distinctly influences sales.

3.2 *Popularity*

Originally, an arbitrary measure of Radiohead's increasing popularity was to be included in the model to demonstrate how the band was steadily entering the mainstream throughout the time period examined. During this time, the band toured extensively in the U.S. with the relevant rock band R.E.M., and released distinctive music videos to accompany their singles that were consistently and widely broadcasted by MTV. MTV host Matt Pinfield also detailed the platform's adamant stance on promoting the unique music videos for each single on *The Bends* despite its lower sales numbers than other released albums at the time. Additionally, drummer Philip Selway was quoted as saying the videos helped the album "gradually seep into people's consciousness" over the course of the year. This gradual increase in the bands' popularity in the general public culminated in the album being included on many famous music critics best-of-the-year lists, which is generally correlated with the increase in sales seen in December 1995.

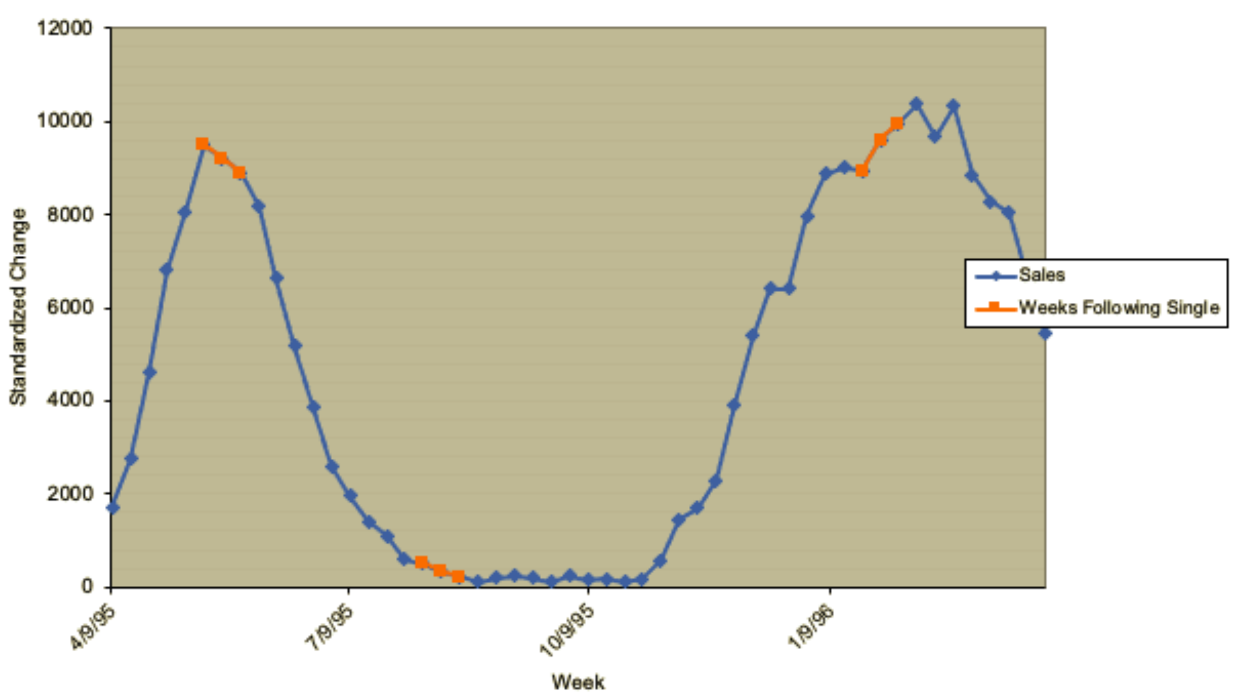
However, the following popularity increase is gradually depicted in the large increase in airplay occurring at the end of October 1995. To explain the full effect, the airplay covariate will be divided into two components: the beginning segment for acquainted listeners up to 9/17/95, and the general boost in public popularity following. Although this will not change the model's fit, the differing covariates will be interesting to interpret and directly play into the idea of increasing overall band popularity in the examined time period: in essence, the first models increase and proceeding, gradual decrease in airplay following an album's release, and the second models the unique mainstream attention Radiohead experienced at the end of 1995 as more people were exposed to the album.

It is generally intuitive that general popularity would be correlated with the amount of times a band's songs are played in a given time period. Additionally, it is difficult to find any public data sources that could be interpreted as a similar measure, such as changes in Billboard standings outside of album sales, or the yet-to-be-invented weekly streaming numbers. In fact, the inclusion of an arbitrary curve to model Radiohead's increasing popularity from 1995 into 1996 may violate the endogeneity principle, as sales are attributed to increases in popularity just as popularity is attributed to increases in album sales.

3.3 *Single Releases*

Radiohead released five singles for The Bends, two of which were released prior to the album, with the others separated by 3 and 6 months respectively. The changes in sales corresponding to the week of a single release, along with two weeks following that specific date are highlighted in the graph below. The two week period was arbitrarily chosen to allow the reader to observably note any trends in sales changes following the single release date.

Graph 4: Changes in Sales Following *The Bends* Single Releases



I initially found it unlikely that the single releases would lead to a change in album sales for two reasons:

empirically, the effect does not appear significant due to the lack of a consistent sales pattern in the graph above, and by story, the band's popularity in this period was not specifically due to one popular single on *The Bends*— "High and Dry" peaked the highest on the US Billboard charts at a low 78, despite it being such a beautiful song! The last single, "Street Spirit" peaked at 5 on the UK charts and effectively proved the band to be a non-one-hit wonder following "Creep." Its release corresponds with a spike in sales, but the timing amidst an already steady incline makes its effect dubious at best. This makes sense, as a lack of US Billboard recognition would translate to a lack of change in US sales data, especially when coupled with how their popularity steadily increased due to gradual mainstream acceptance of their complete product, not one part of it.

For lack of a better reason to do otherwise, the singles were not coded into the model. Had we modeled the sales of *Pablo Honey*, their debut album with a chart-smashing lead single, I definitely would have included it.

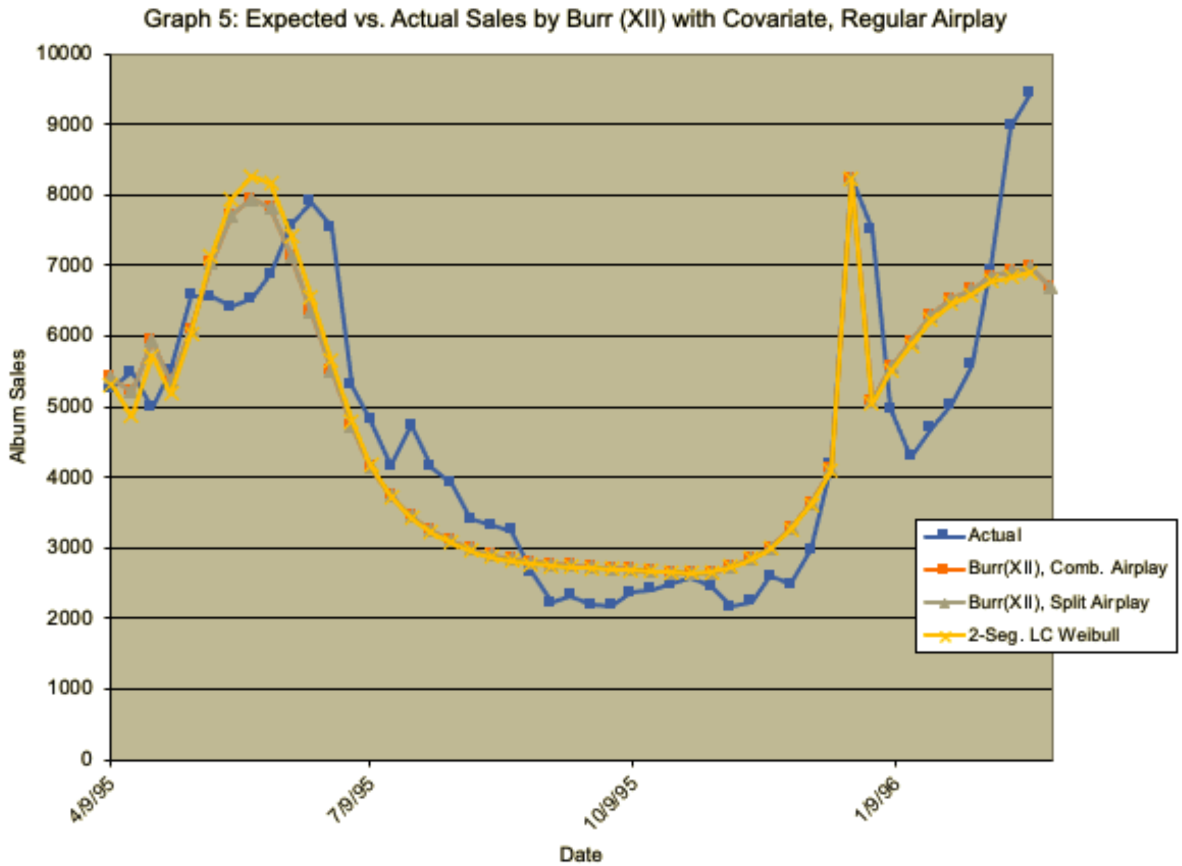
3.4 *Christmas*

In order to account for the seemingly random-without-context peak on the week before Christmas 1995, a dummy variable will be coded on 12/24/95 to account for the spike in sales. Given the massive increase without any explanation in the bands' history at this moment in time, it is reasonable to assume that some people went to the record store (or maybe a 1995 Urban Outfitters) and bought *The Bends* for one of their loved ones. Additionally, the fact that many critics included the album on their best of the year lists, it seems like it would be a great present for music listeners at the time and a justifiable reason to include a Christmas covariate.

4 ANALYSIS INCLUDING COVARIATES

Thus, the covariates for radio airplay after release, radio airplay representing end-of-year popularity gain and a Christmas dummy variable were included in the model and assessed with Burr(XII) and 2-Segment Latent Class Weibull models. Airplay was combined into one variable for the Weibull on the basis of sparsity and parsimony, as well as its lack of distinction between the two covariates on first analysis. A Burr(XII) with combined airplay was also analyzed to compare summary statistics with those of the split airplay model.

4.1 Analysis



Just as before, the models display marginal differences from observation, so the summary statistics must be examined to draw distinctions between each model.

Table 2: Covariate Model Performance Data

Parameters	Burr(XII)	Burr(XII) Split	2-Seg. Weibull
r	914.2332352	240.9403529	
alpha	221015.8763	52629.70274	
lambda			0.003962031
c	0.928509384	0.90376232	0.936647013
Bairplay	0.000107687		0.000109871
Bchristmas	0.597520058	0.584805848	-0.044042142
Boldplay		0.000100307	
Bnewplay		0.000112849	
pi			0.99548456
lambda2			0.15768923
c2			0.0499740
B2airplay			-0.000104
B2christmas			168.832236
LL	-1358554.043	-1358468.361	-1358538.944
BIC	2717177.163	2717019.614	2717202.227
MAPE	16.628%	16.423%	16.520%
R^2	76.268%	77.082%	76.509%

Of all the models analyzed, the Burr(XII) with the split airplay covariates is the best by every statistic: it has the highest log-likelihood, the lowest BIC, the lowest MAPE and the highest R-squared. As another confidence check, a Likelihood Ratio Test was performed for the Burr(XII) models, depicted in the table below.

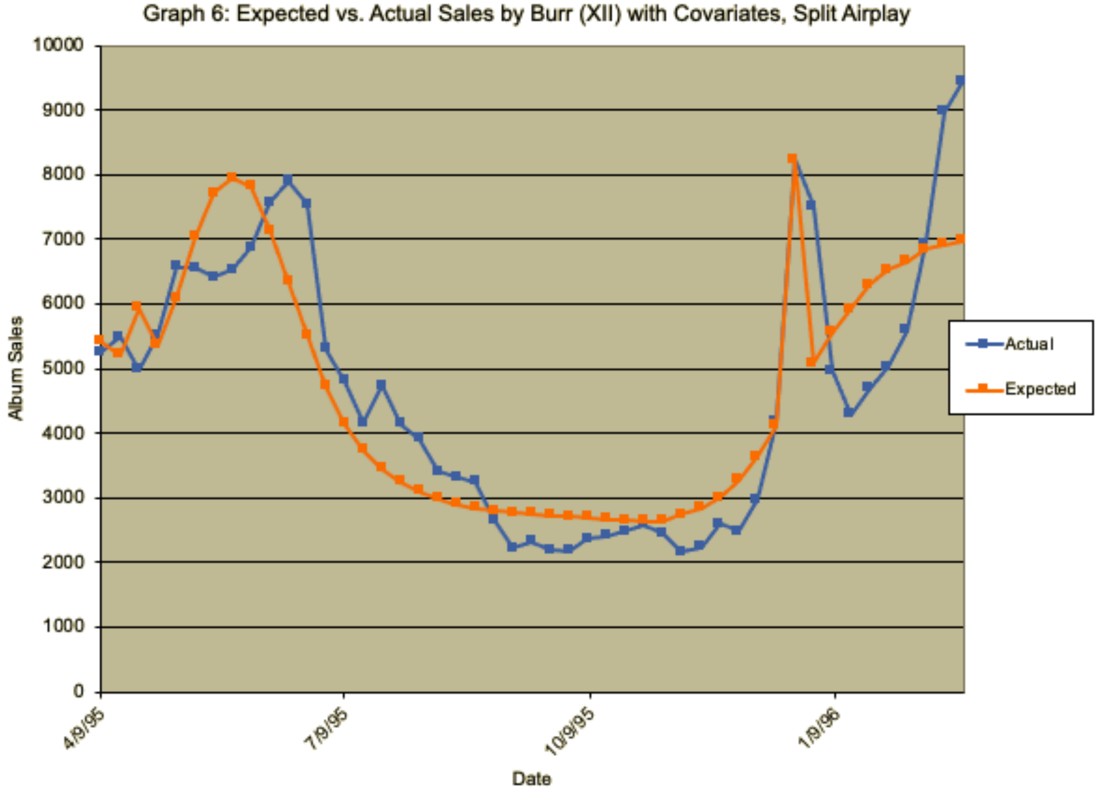
Table 3: Likelihood Ratio Test

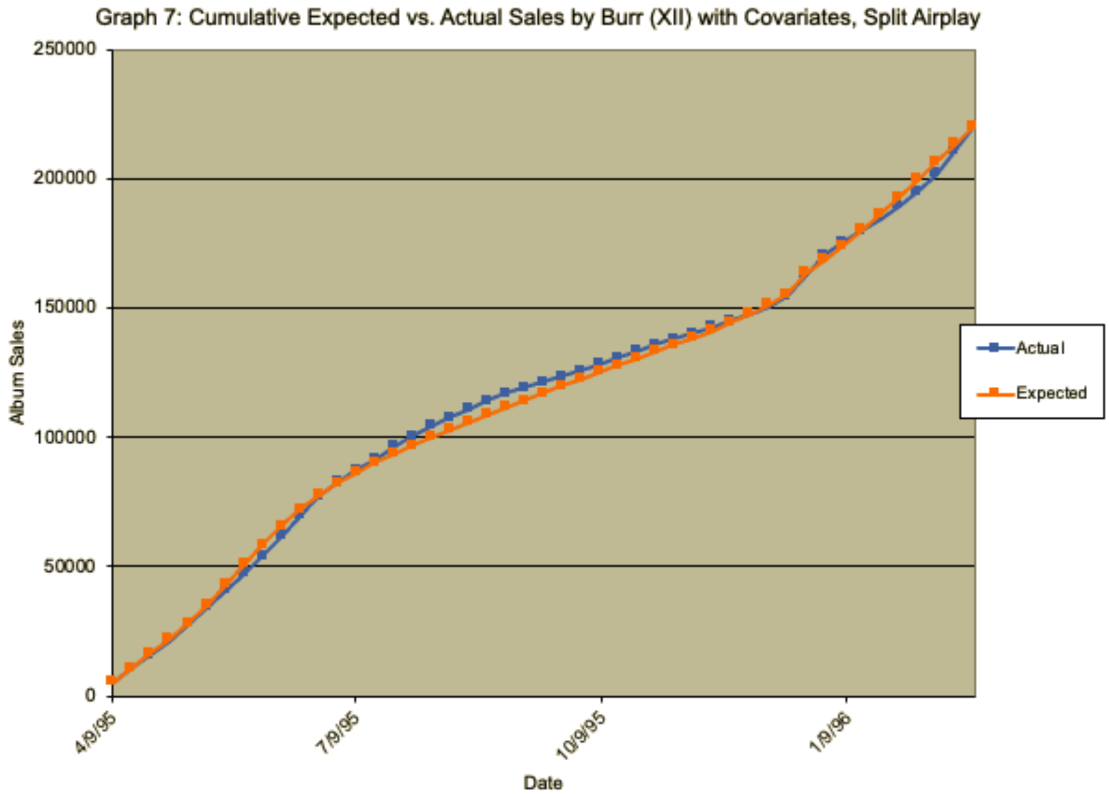
LRT	171.278
Degrees of Freedom	1
p-Value	3.89085E-39

The results display that the differences between the models are indeed significant, despite the small observable difference from the incremental graph.

Most importantly, the Burr(XII) with the split airplay covariates provides the best story given the boost in Radiohead's mainstream popularity at the end of 1995. The notion of two segments is largely rejected by the model, with a π greater than 0.995 coupled with the lack of convincing argument to include two segments existing throughout the whole time period. If anything, it is more likely that a new group was introduced to the band at the end of 1995, which is similarly represented with the split, popularity component of airplay used in the final model. The graphs below display the isolated Burr(XII) with split covariates model against the actual sales data reported.

Graph 6: Expected vs. Actual Sales by Burr (XII) with Covariates, Split Airplay





The large value for r makes sense, given that the heterogeneity is not explained by differences in people's buying-at-time propensities, but the state of the covariates at a given week. The high value for the dummy variable at Christmas additionally has intuitive meaning, as the model captures the main outlier increase from the holiday. Although the popularity half of airplay has a marginal increase in magnitude of the first half, their shared positive relationship and a difference reflecting the band's entrance into the mainstream are verified through the small change.

4.2 *Final Projections*

Based on this model, the projections for the last five weeks of sales are **6690, 6242, 5903, 5289 and 4765** respectively.

5 IMPLICATIONS AND IMPROVEMENTS

In order to better assess the model's robustness and usefulness, a holdout period of 15% was enacted with a split made at the 42th entity, the week of 12/24/95. This split was made to allow for the effect of the newfound popularity covariate to settle in, resulting in a MAPE of 20.092%. Although it is at a clear disadvantage to the full model, the less than 5% increase in MAPE supports that the model is relatively robust. However, that does not mean that it is entirely infallible, as there are clearly adjustments to be made in the future.

If some data could be collected on the somewhat exponential increase in the mainstream popularity bands over time, it would be a massively important covariate in this investigation. In particular, it could help capture the increase in sales at the very end of the model that my final model missed out on. For example, the band was nominated for multiple awards at the Brit Awards in February 1996 for *The Bends*, which would have gained them more popularity, coinciding with more sales, which may cause fans to spread word to others, increasing sales further. Given more time, this may have been captured by examining sales data on the breakthrough album for rock or alternative bands with a similar present day fanbase as Radiohead.

Additionally, if the lag of approximately three weeks of sales on airplay could be altered, a potentially more indicative variable of airplay may have been created.

Overall, this project displays the difficulty of applying probability models to real-life situations where the data is simply not perfect, as people may not follow the same buying patterns as represented by a typical mixture model. The true work lies in finding this representative data and incorporating it, with only one way to truly practice: by diving into the details and doing it yourself. I'm just happy I got to learn on a dataset for one of my favorite bands!

6 BIBLIOGRAPHY

https://books.google.com/books?id=0q9bTJp2Bi4Cq=My+Iron+Lung+Live+at+the+Astoriapg=PT203
https://pitchfork.com/features/ok-computer-at-20/10036-this-is-what-you-get-an-oral-history-of-radioheads-karma-police-video/
https://www.stereogum.com/1784465/qa-radioheads-philip-selway-remembers-the-bends/franchises/interview/