Title
Predicting Music Revenue: A hierarchical linear modeling approach with sensitivity analyses

Author
Whitworth, Alex

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Predicting Music Revenue: A hierarchical linear modeling approach with sensitivity analyses

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Alex Phillip Whitworth

2015

Abstract of the Thesis

Predicting Music Revenue: A hierarchical linear modeling approach with sensitivity analyses

by

Alex Phillip Whitworth

Master of Science in Statistics

University of California, Los Angeles, 2015

Professor Yingnian Wu, Chair

The music industry has undergone enormous change since the introduction of of Napster in 1999. In 1999, 100% of industry revenue was from physical sales; in 2014, United States music industry revenue was 32% physical, 37% digital downloads, 27% streaming, and 4% other minor categories. In this thesis, I present the first models in the music industry that predict monthly revenue at the album level across both revenue stream and geography within the music industry, which are based on a hierarhical linear modeling framework. In addition to the predictive models, I present several sensitivity analyses to examine interesting properties of the data. Specifically, the sensitivity analyses address the effects of data missingness, design imbalance, and the impact of outliers on the predictive results.

The thesis of Alex Phillip Whitworth is approved.

Robert Gould

Mark Stephen Handcock

Yingnian Wu, Committee Chair

University of California, Los Angeles

2015

To my parents . . .

without your constant love and support, all of my life's endeavors and educational pursuits would not have been possible.

Thank you, Mama and Dad.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to the UCLA Statistics Department, which has provided a wonderful environment for me to pursue my graduate work over the past two years. I am also sincerely grateful to the members of my thesis committee. In particular, to Professor Ying Nian Wu who provided considerable input and guidance as I pursued my research and to Rob Gould, introduced me to Concord Music Group, which led to the work that makes up this thesis. I would also like to thank Bob Valentine for both providing me with the data used in this research and his support and patience as I pursued my research.

Beyond the educational experience, I have formed many lasting friendships in the UCLA Statistics Department, which I cherish. I would like to thank James Molyneux and Katie McLaughlin for their patient help with LaTeX. I also wish to thank Bryon Aragam, Hao Ho, and Kyle Hasenstab for many helpful conversations regarding the content of my thesis.

Lastly, I would like to thank Ben Bolker and Adam Loy, the creators and maintainers of the **lme4** and **HLMdiag** R packages respectively. Their help with technical issues, quick responses to emails, and package updates have been invaluable to my research process.

CHAPTER 1

Introduction

The music industry has undergone enormous change since the 1999 introduction of Napster, which allowed consumers to share large quantities of music online anonymously. Although the Recording Industry Association of America (RIAA) successfully sued and closed Napster in 2000, other successful file sharing services such as Kazaa quickly emerged in its place. While Napster was the first successful entrant onto the digital music scene, it merely marks the beginning of a period of industry transformation. Apple launched their iTunes Music Store in 2003, becoming the first major entrant into legal downloading services and finding major success. Apple sold their billionth song via iTunes in February 2006 and their five billionth song in June 2008.[1] Apple's success made it clear that digital music was now a major source of industry revenue, making up a third of United States recorded music sales in 2008.[2] The digital deluge continued with the launch of Spotify in 2008.[3] Spotify, and other competing services, offers online music streaming services to consumers and has become a major source of industry revenue as well.[4] In 2014, music industry revenue in the United States was 32% physical sales, 37% digital downloads, 27% streaming, and 4% other minor

[1] Van Buskirk, Eliot, "iTunes Store May Capture One-Quarter of Worldwide Music by 2012." (http://archive.wired.com/entertainment/music/news/2008/04/itunes_birthday). Accessed: 2015-04-22

[2] RIAA Year-End Shipment Statistics. (http://76.74.24.142/1D212C0E-408B-F730-65A0-C0F5871C369D.pdf). Accessed: 2015-04-22

[3] "We've only just begun." (https://news.spotify.com/us/2008/10/07/weve-only-just-begun/). Accessed: 2015-04-22

[4] Friedlander, Joshua P. "News and Notes on 2014 RIAA Music Industry Shipment and Revenue Statistics." (http://riaa.com/media/D1F4E3E8-D3E0-FCEE-BB55-FD8B35BC8785.pdf). Accessed: 2015-04-21

categories.

The impact of music file sharing services has been studied closely in the empirical literature with special attention paid to the relationship beween file sharing and legal music purchases (Peitz and Waelbroeck. (2006), Waldfogel (2010), Bakos et al. (1999), Hong. (2007), Oberholzer-Gee and Strumpf (2007)). The literature has focused primarily on the the displacement of music sales by digital downloads, individuals' purchase and stealing activities, and variations in these phenomenon across geography and time. Given the recency of streaming services, the literature lacks a similar discussion of the effects of streaming on music sales or on the substitution effects of streaming on digital and physical sales. None of these studies dispute the fact that digital and streaming are now a major source of music revenue, and most industry commentators expect streaming to soon constitute the largest portion of industry revenues.[5]

This transition from exclusively physical sales to an environment in which digital and streaming sales play an increasingly important role presents two major challenges to the music industry. In a trend that started with digital downloads, individual songs have been decoupled from albums. Consumers are no longer forced to purchase a full album when they are only interested in one or two hit tracks. This has led to the first challenge facing the industry: the revenue from each individual transaction is worth increasingly less revenue to the music industry. This trend has accelerated with streaming services, where consumers are able to effectively *rent* songs each time they listen to them for a very small amount of money, paid for via either advertising or a monthly subscription fee. Digital and streaming services have also quickly eroded the importance of the prior, physical, distribution regime and replaced it with new distribution channels. These new distribution channels introduce additional barriers to industry participants who

[5]Table (A.1) in the appendix shows a detailed look at music industry revenue in the United States over time.

wish to gather detailed data about their customers and revenue composition, which is the second major challenge facing the industry.

These two trends have created a much more elaborate operating environment for the music industry, where the number of music consumers has rapidly increased, the value of each individual transation has decreased, and the data about both consumers and revenue has both multiplied and become much more difficult to obtain. It is therefore of critical interest to companies within the music industry to leverage their existing data, in order to gain more insight into both their revenue and their consumers. In this thesis, I examine the former of these, presenting the first models which predict, at the album level, monthly revenue across major revenue streams–physical, digital, and streaming–and across geographies. The rest of this thesis is composed as follows: Chapter 2 presents the modeling framework; Chapter 3 describes the data used in this study; Chapter 4 outlines the modeling approach, discusses sensitivity analyses affecting the results, and discusses the model results; and Chapter 5 concludes.

CHAPTER 2

Hierarchical Models

Hierarchical, or nested, data structures arise naturally in many data collection regimes. The most frequently cited examples are when individuals are observed over time (longitudinal studies), or when stratified samples are used, such as when sampling students within classrooms within schools. Data organized in this manner are no longer independent observations. Any statistical model used must therefore accommodate to this more general covariance structure, where observations within the same sample unit may be correlated.

Hierarchical models, also commonly called mixed models, were developed to appropriately handle such data structures. These models incorporate parameters that estimate both the overall trend–the fixed effects–and the correlated covariance structure–the random effects. These models therefore represent an explicit trade-off; practicioners have increased flexibility in modeling both random and fixed effects but accept increased complexity in the statistical modeling process.

The problems associated with parameter estimation have been extensively discussed in the literature. Practictioners typically employ either a maximum likelihood (ML) or restricted maximum likelihood (REML) approach, although Monte Carlo methods are also available.[1] In addition to estimation, many diagnostic methods have been developed as extensions of diagnostic methods for the classical linear model (Hilden-Minton (1995), Hodges (1998), Loy (2013), Ronald Christensen and Johnson. (1992), Snijders and Berkhof (2008), Zewotir and Galpin.

[1]For a thorough treatment to estimation techniques, please see Goldstein (2011), Pinheiro and Bates. (2000), and Raudenbush and Bryk (2002).

(2005)). As with classical linear models, these methods are chiefly based on residual analysis and include methods for assessing the adequacy of model assumptions and methods for influence diagnostics.

The remainder of this chapter provides a summary of hierarchical models. I first discuss an illustrative example from a Bayesian perspective and then present the general two-level hierarchical model. Finally, I include a brief discussion of residual and influence analysis for hierarchical models. In the most common case, when the estimation technique is linear, the literature refers to these models as hierarchical linear models (HLM) or multilevel models (MLM), although these terms are often used for binomial and count data as well. For consistency, I use the term HLM going forward.

2.1 Illustrative Example

In this section, I present an example of hierarchical models to fix understanding. To do so, I consider a study of educational achievement within a stratified sample; that is, where observations of student achievement are nested within classrooms. Formally, the study consists of observed student achievement y_{ij} with observations indexed by $i = 1, \ldots, n_j$ within classrooms and with classrooms indexed by $j = 1, \ldots, J$. This model has many parameters of interest, notably: (i) the parameters describing classroom achievement within each classroom, $\theta = (\theta_1, \ldots, \theta_J)$; and (ii) the hyperparameters describing overall student achievement $\phi = (\mu, \tau^2)$. To gain a full understanding of student achievement at each level of our hierarchy, I wish to estimate all of these parameters and to provide meaningful interpretations.

Before estimating θ and ϕ, some distributional assumptions must be made. Additionally, some assumptions about the exchangeability of both students and classrooms must also be made. Firstly, considering distributions, while hierarchical models exist in general form, here it is useful to consider the case where educational achievement is normally distributed. Given that most educational achievement data (SAT-V, SAT-M ACT, IQ, etc) has been shown to be normally distributed, this assumption is reasonable. Importantly, normality can be assumed for illustrative purposes here without any loss of generality. It is also important to consider assumptions regarding the exchangeability of both students and classrooms. One possibility is to view both students and classrooms as identical replications of one another. Under this assumption, all students are regarded as independent samples within a common population of students and all classrooms as independent samples within a common population of classrooms. A second possibility is to consider students as independent samples but to consider classrooms so different that the results from any one classroom provide no information about the results of the others. A third, more general possibility is to regard students within a given classroom and all classrooms as exchangeable but not necessarily either identical or completely unrelated. This third possibility represents a continuum between the first two extremes, and it is this exchangeable model that forms the basis of the analysis presented here.

To fully specificy this normal model, I assume that student achievement and classroom achievement have common variance. Defining $\bar{y}_{.j} = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$ and $\sigma_j^2 = \frac{\sigma^2}{n_j}$ as the sample mean and variance for each classroom respectively gives

$$\bar{y}_{.j}|\theta_j \sim N(\theta_j, \sigma_j^2) \tag{2.1}$$

$$\theta_j|\phi \sim N(\mu, \tau^2) \tag{2.2}$$

The joint density $P(\phi, \theta, y)$, joint posterior $P(\phi, \theta|y)$, and conditional posterior

$P(\theta|\phi, y)$ are

$$P(\phi, \theta, y) = \pi(\phi) \prod_{j=1}^{J} N(\theta_j|\phi) N(\bar{y}_{.j}|\theta_j) \tag{2.3}$$

$$P(\phi, \theta|y) \propto \pi(\phi) \prod_{j=1}^{J} N(\theta_j|\phi) N(\bar{y}_{.j}|\theta_j) \tag{2.4}$$

$$P(\theta_j|\phi, y_j) \propto N(\theta_j|\phi) \prod_{i=1}^{n_j} N(\bar{y}_{.j}|\theta_j) \tag{2.5}$$

where $\pi(\phi)$ is the prior distribution for ϕ, $(Y \perp \phi)|\theta$, and $(\theta_k \perp \theta_l)|\phi$ for all $k \neq l$. It is therefore clear that

$$\theta_j|\phi, y \sim N(\hat{\theta}_j, V_j)$$

where

$$\hat{\theta}_j = \frac{\frac{1}{\theta_j^2}\bar{y}_{.j} + \frac{1}{\tau^2}\mu}{\frac{1}{\theta_j^2} + \frac{1}{\tau^2}} \text{ and } V_j = (\frac{1}{\theta_j^2} + \frac{1}{\tau^2})^{-1}. \tag{2.6}$$

This solution is still incomplete because it depends on the unknown hyperparameters $\phi = (\mu, \tau^2)$. For the normal distribution, the marginal likelihood of the hyperparameters has a particularly simple form. The marginal posterior density can be written

$$P(\mu, \tau|y) \propto \pi(\phi) \prod_{j=1}^{J} N(\bar{y}_{.j}|\mu, \sigma_j^2 + \tau^2)$$

Further, the marginal posterior density for the hyperparameters can be factored $[P(\mu, \tau|y) = P(\mu|\tau, y)P(\tau|y)]$. I specify a uniform density on $\mu|\tau$. Therefore

$$\mu|\tau, y \sim N(\hat{\mu}, V_\mu)$$

where

$$\hat{\mu} = \frac{\sum_{j=1}^{J}(\sigma_j^2+\tau^2)^{-1}\bar{y}_{.j}}{\sum_{j=1}^{J}(\sigma_j^2+\tau^2)^{-1}} \text{ and } V_u^{-1} = \sum_{j=1}^{J}\frac{1}{\sigma_j^2+\tau^2} \tag{2.7}$$

and, additionally,

$$P(\tau|y) \propto \pi(\tau)V_\mu^{1/2}\prod_{j=1}^{J}(\sigma_j^2+\tau^2)^{-1/2}exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma_j^2+\tau^2)}\right). \tag{2.8}$$

A prior distribution on τ must still be specified to complete this illustration. The simplest specification is to use a diffuse noninformative prior density for τ, which requires an examination of the resulting posterior distribution to ensure it has a finite integral. Alternatively, if an appropriate 'best guess' and upper bound are determined for τ, a reasonable prior distribution can be constructed from the scaled inverse-χ^2 family, where the 'best guess' is matched to the mean of the scaled inverse-χ^2 and the upper bound is matched to an upper percentile.

2.1.1 A three-level Model

I next briefly discuss an extension of this illustration to three levels (higher level higherarchies are also possible) with observations y_{ijk} where school districts, for example, are indexed by $k = 1, \ldots, K$. In this stratified sampling regime, observations of student achievement are nested within classrooms, which are further nested within school districts. Here, there may be some characteristics of school district that influence student achievement, such as socioeconomic status, which are of interest. Therefore, in the three-level model, a third set of parameters is added to our list of parameters of interest: (iii) the parameters describing school district achievement, $\gamma = (\gamma_1, \ldots, \gamma_K)$. The three-level model is thus parameterized:

$$\bar{y}_{.jk}|\theta_{jk} \sim N(\theta_{jk}, \sigma^2_{jk})$$
$$\theta_{jk}|\gamma_k \sim N(\gamma_k, \alpha^2_k)$$
$$\gamma_k|\phi \sim N(\mu, \tau^2)$$

where $\alpha^2_k = \frac{\alpha}{n_k}$ represents the school district sample variance with common variance parameter α assumed and n_k the number of sampled classrooms within each school district. The total number of obserations in this three-level study would therefore be $N = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\sum_{i=1}^{n_j} y_{ijk}$. The joint density, joint posterior density, and conditional densities can be found in a similar fashion to that outlined above.

2.1.2 Extension to Covariates

I also consider an extension of this model that includes covariates at the student and classroom level.[2] In this formulation of the hierarchical model, educational achievement is considered to be a function of student-level and classroom-level attributes. It is therefore assumed that classroom-level effects are non-exchangable and depend on specific features of the classroom. In this case, multiple parameters can vary by group. For example, one might consider gender, prior test scores, and highest level of parental education at level-1 and teacher's education level, classroom prior test score average, and a factor variable describing classroom management style at level-2.

This model is formally described as consisting of observed educational achievement y_{ij} with student achievement indexed by $i = 1, \ldots, n_j$ within classrooms and with classrooms indexed by $j = 1, \ldots, J$ as above. In addition, there is a matrix of explanatory variables for each classroom, X_j, with fixed effects parameter vector

[2] Covariates at higher levels can also be included . Here, I revert back to the two-level model for simplicity.

β where there are q_1 coefficients in the regression model. It is also important to model variances between groups. To do so, a vector of random effects for each classroom, Z_u, is added–with u_j the vector of random effects parameters to be estimated. This model is parameterized as

$$y_{ij} \sim N(X_j(\beta_j + Z_u \otimes u_j), \sigma_y^2) \tag{2.9}$$

$$\beta_j \sim N(\mu, \Sigma_\beta) \tag{2.10}$$

$$u_j \sim N(0, \Sigma_u) \tag{2.11}$$

where $\otimes$ is the usual Kronecker product and $\Sigma_\beta = \tilde{Z}_u \Sigma_u \tilde{Z}_u$. Here $\tilde{Z}_u$ is defined as the square matrix with the elements of Z_u on the diagonal. One advantage of this model is that separate prior distributions are assigned to u_j and Σ_u, which induces a rich structure of uncertainty modeling. This model is an example of the scaled inverse-Wishart model. For a nonimformative model, the analyst can set $\Sigma_u \sim$ Inv-Wishart$_{q_1+1}(I)$ along with indepedent prior distributions on the u_j's. Or, if necessary, the analyst can add informative priors to the u_j's.

2.2 The General two-level Model

With that illustration of hierarchical models from a Bayesian perspective concluded, I now direct attention to the general two-level HLM. In the general two-level model, consider the observed response vector Y with individual responses y_{ij} where observations are indexed by $i = 1, \ldots, n_j$ within groups and with groups indexed by $j = 1, \ldots, J$. Following the notation of Goldstein (2011) and Loy (2013), the two level HLM can be specified as follows

$$Y_j = X_j\beta + Z_u u_j + \epsilon_j \tag{2.12}$$

where Y_j is the $n_j \times 1$ vector of responses within the jth group; X_j is an $n_j \times q_1$ matrix of explanatory variables for each group; β is a $q_1 \times 1$ vector of unknown fixed parameters; Z_u is the matrix of random effects; u_j is the $q_2 \times 1$ vector of random parameters; and ϵ_j is the $n_j \times 1$ vector of random errors. The random error vector ϵ_j is composed of individual and group error terms, $\epsilon_j = \epsilon_{ij} + u_j$ where ϵ_{ij} is the ith element of ϵ_j and represents the error for the ith individual of group j. u_j is the group level error which is fixed within each group. It is commonly assumed that $u_j \sim N(0, \Omega_u), \epsilon_{ij} \sim N(0, \Omega_e), Cov(u_k, \epsilon_l) = 0\ \forall\ k$ and l, and $Cov(u_k, u_l) = 0$ $\forall\ k \neq l$ where Ω_e and Ω_u are positive definite. Succinctly, in the standard model, the level-1, or individual, and the level-2, or group, errors are assumed to be independent. Further, the group errors are assumed to be independent across groups. It is often convenient for estimation to use the scaled covariance matrices where it is assumed that observations have within group homoscedastic errors $\Omega_e = \sigma_e^2 R_j$ and group variances are homoskedastic $\Omega_u = \sigma_u^2 D$.

Note than, in a special case where the scale parameters are equal $\sigma_e^2 = \sigma_u^2$, these assumptions imply $Y_j \sim N(X_j\beta, \sigma^2 V_j)$ where $V_j = R_j + Z_u D Z_u^T$. Further, defining $\xi = Z_u u_j + \epsilon_j$, the marginal model can be defined as $Y_j = X_j\beta + \xi$. If R and D are known and $\sigma_e^2 = \sigma_u^2$, then the marginal model is just a linear model with weight matrix $\sigma^2 V^{-1}$.

In the general case, $\epsilon_j = \epsilon_{ij} + u_j$ implies $V_j = V_{j(1)} + V_{j(2)}$ where $V_{j(1)} = Z_u^T \Omega_e Z_u$ and $V_{j(2)} = Z_j^T \Omega_u Z_j$. The overall covariance matrix V is therefore block-diagonal with the jth block of V, $V_j = \oplus_j \sigma_e^2 + V_{j(2)}$ where $\oplus$ is the direct-sum operator. For known β, let $Y^* = (Y - X\beta)(Y - X\beta)^T$ and $Y^{**} = vec(Y^*)$ where vec is the vector stacking operator. Therefore, $E(Y^*) = V$ and $E(Y^{**}) = Z^* u$ where Z^* is the design matrix for the random parameters.

The best least unbiased predictors (BLUPs) in the two-level model are

$$\widehat{\beta} = (\sum_{j=1}^{J} X_j^T V_j X_j)^{-1} \sum_{j=1}^{J} X_j^T V_j Y_j \tag{2.13}$$

$$\widehat{u} = (Z^{*T} V^{*-1} Z^{*})^{-1} Z^{*T} V^{*-1} Y^{**} \tag{2.14}$$

where $V^* = V \otimes V$. In practice, if V is unknown, $\widehat{V}$ is used. The iterative generalised least squares (IGLS) procedure iterates between (2.13) and (2.14). At convergence, assuming multivariate normality, the estimates are maximum likelihood (ML). The IGLS procedure produces biased estimates in general, which can be important in small samples. A simple modification (using $E(Y^*) = V - X(X^T V^{-1} X)^{-1}$) is used to obtain restricted maximum likelihood (REML) estimates, which are unbiased.

Following Hilden-Minton (1995), note that $Y - X\widehat{\beta} = V^{-1} Q Y$ where $Q = V - VX(X^T V X)^{-1} X^T V$. Keeping consitent notation, it is possible to partition Q into $J \times J$ blocks of the form

$$Q_{ij} = \begin{cases} V_i - P_{ii} & i = j \\ -P_{ij} & i \neq j \end{cases}$$

where $P_{ij} = V_j X_j (\sum_{j=1}^{J} X_j^T V_j X_j)^{-1} X_i^T V_i$. Additionally, the variances of the parameter estimates are

$$Var(\widehat{\beta}) = \sigma^2(\sum_{j=1}^{M} X_j^T V_j X_j)^{-1} \tag{2.15}$$

$$Var(\widehat{u}_j - u_j) = \sigma^2[D - DX_j^T Q_{jj} X_j D] \tag{2.16}$$

$$Cov(\widehat{\beta}, \widehat{u}_j - u_j) = (\sum_{j=1}^{J} X_j^T V_j X_j)^{-1} X_j^T V_j X_j D \tag{2.17}$$

$$Cov(\widehat{u}_k - u_k, \widehat{u}_l - u_l) = \sigma^2[DX_k^T Q_{kl} X_l D] \tag{2.18}$$

for all $k \neq l$.

2.3 Residuals

One of the key characteristics of the HLM is that it allows for uncertainty at both the individual and group levels. Considering Equation (2.12), this leads to multiple errors in the model:

- the conditional error, ϵ_{ij}
- the random effects, u_j
- the composite, or marginal, error $Z_u u_j + \epsilon_{ij}$.

Each type of error is of diagnostic interest, although none is directly observable. That is, these errors are confounded. For diagnostic purposes, the analyst would ideally like a residual term for each unobserved error that depends solely on the specific error term of interest. However, since residuals in HLM are confounded, this is impossible. This makes the analyst's life considerable more difficult. Residual terms can still be defined and analyzed; but, in contrast to the classical regression model, the presence of uncertainty at both the individual and group level leads to multiple residual terms. Defining the residuals from Equation

(2.12) and using the estimated covariance matrices $\hat{D}$ and $\hat{V}_j$, gives

$$\widehat{\epsilon}_{ij} = e_{ij} = \widehat{Y}_j - (X_j)_i\widehat{\beta} - (Z_u)_i\widehat{u}_j \tag{2.19}$$

$$\widehat{u}_j = d_j = \widehat{D}Z_u^T\widehat{V}_j^{-1}(Y_j - X_j\widehat{\beta}) \tag{2.20}$$

$$c_{ij} = \widehat{Y}_{ij} - (X_j)_i\widehat{\beta} \tag{2.21}$$

for the conditional, random, and composite residuals respectively. These are termed emperical bayes (EB) residual estimates; and, since *c, d, e* are defined in terms of BLUPs, it is clear that they are BLUPs for $Z_u u_j + \epsilon_{ij}$, u_j, and ϵ_{ij} respectively.[3]

Naturally, given that the residual terms are defined as BLUPs, the question arises–why should one care if a residual is confounded? While d and e are unbiased predictors for u_j and ϵ_{ij}, this is not true conditionally. It can be shown that $d_j|\epsilon_{ij}$ has bias $ZDZ^TQ\epsilon$, and $e|u_j$ has bias $RQZu_j$ (Hilden-Minton (1995)). This conditional dependency may lead analysts to inappropriately change the functional form of the model or to remove a few, truly concordant cases without just cause.

As with least squares, residual plots and analyses are crucial in assessing the validity of the assumptions of the HLM, and, if they have been violated, how they have been violated. To minimize the impact of confounding and following the advice of Hilden-Minton (1995) and Loy (2013), an *upward* residual analysis is preferred during the model-checking process. That is, it is preferred to first examine the level-1 residuals, and, after concluding that an appropriate model has been speficied at this level, proceed to level-2. In the case of a three-level or higher model, the residual analysis continues to proceed up the hierarchy. I use this *upward* residual procedure for model checking in this thesis.

[3]Alternatively, if the level-1 sample sizes are large enough, unconfounded level-1 residuals may be found via individual least squares regression for each group.

Another complication to consider is heterogeneity of group variance, that is to consider $\Omega_{e_j} = \sigma_j^2 R_j$ and $\Omega_{u_j} = \sigma_j^2 D_j$. Heterogeneity of group variance is related to residual analysis since, under the assumption of heterogeneity, it is often useful to work with standardized residuals. To define the standardized residuals recall that the variance matrix for the jth group V_j is $V_j = V_{j(1)} + V_{j(2)}$ where the level-1 variance is $V_{2(1)} = Z^T \Omega_e Z$ and the level-2 variance is V_2 is a block diagonal matrix with the jth block $V_{j(2)} = Z_j^T \Omega_u Z_j$. Since the residuals are defined in terms of the variance matrix, or more accurately its estimate $\hat{V}$ with jth group $\hat{V}_j$, and following Loy (2013), the following formulation for $V_{j(2)}$ can be used to standardize the level-2 residuals.

$$V_{j(2)} = \widehat{D} Z_u (\widehat{V}_j^{-1} - \widehat{V}_j^{-1} X_j (\sum_{j=1}^{J} X_i^T \widehat{V}_j^{-1} X_j)^{-1} X_j^T \widehat{V}_j^{-1}) Z_u \widehat{D} \tag{2.22}$$

Note that this formulation is only preferred for model checking. For inferential claims, the conditional variance, $Var(\widehat{u}_j - u_j)$, is preferred (Laird and Ware (1982)).

2.4 Influence Analysis

As with the classical linear model, not all observations have the same impact on parameter estimation. In the case of HLMs, this extends to higher order observations–groups–as well. Some observations, or groups, may have excessive impact. Observations, or groups, with excessive impact on parameter estimation are termed *influential*. An extensive literature exists for influence analysis in classical linear models (Cook and Weisberg, 1982; Chatterjee and Hadi, 1986; etc). More recently, influence analysis has been studied for HLMs (Loy (2013), Ronald Christensen and Johnson. (1992), Zewotir and Galpin. (2005)). In this section, I review some of the key generalization of influence measures to HLMs. For additional detail, readers are encouraged to review Loy and Hofmann (2013) and Zewotir and Galpin (2005).

One influence measure that is of interest is leverage. High leverage points are observations, or groups, that greatly influence the fitted values. In ordinary regression, the leverage values are defined to be the the diagonal elements of the hat matrix, $H = X(X^TX)^{-1}X^T$. In HLM, leverage must be defined for both fixed and random effects. Assuming a fixed covariance structure V_j, the leverage at level j is denoted $H = \partial\hat{y}_j/\partial y_j$. And the leverage of group j is the sum of the fixed and random effects H_{1j} and H_{2j} where

$$H_{1j} = X_j(X_j^T V_j^{-1} X_j)^{-1} X_j^T V_j^{-1} \tag{2.23}$$

$$H_{2j} = Z_j D Z_j^T V_j^{-1}(I - H_{1j}). \tag{2.24}$$

In practice, $\hat{V}$ and $\hat{D}$ are substituded for $\mathbf{V}$ and $\mathbf{D}$ respectively.

Another common method to assess influence is to observe the change in parameter estimates after the ith unit is deleted. Cook's distance is one common measure used to do this in ordinary least squares. Applied to HLMs, Cook's distance can be extended to measure influence on the fixed effects and random effects as follows:[4]

$$CD_i(\beta) = (\hat{\beta}_{(i)} - \hat{\beta})^T(X\hat{V}^{-1}X^T)(\hat{\beta}_{(i)} - \hat{\beta}) \tag{2.25}$$

$$CD_i(\theta) = (\hat{\theta} - \hat{\theta}_{(i)})^T \widehat{Var(\hat{\theta})}(\hat{\theta} - \hat{\theta}_{(i)}) \tag{2.26}$$

[4]I follow common convention and use the subscript (i) to denote deletion of the ith observation or group. For example, $X_{(i)}$ denotes $\mathbf{X}$ where the ith observation, or row, has been deleted.

where θ denotes the vector of variance components, that is, the vector containing σ_e^2 and the unique elements of $\mathbf{D}$. Large values of the Cook's distance indicate that an observation, or group, is influential. Note that, since the estimated covariance matrix $\mathbf{V}$ is used, there is no exact reference distribution for (2.25). The use of a bootstrap distribution is recommended Loy (2013).

Computational methods are not discussed in this thesis. Readers with a particular interest in computation methods are encouraged to read Christensen (1992) and Zewotir (2005), where they are discussed thoroughly. Pinheiro and Bates. (2000) is also an excellent reference for computational methods for the general HLM.

CHAPTER 3

Data

The data in this thesis comes from a longitudinal study conducted for a music company in Beverly Hills, CA (the Client). The study focuses on monthly music sales at the album level from January 2012 to May 2014.[1] Sales are differentiated by nine primary geographical markets (geographies) and three primary sources of revenue (revenue streams). This is therefore a three-level HLM where monthly sales are nested within albums, which are nested within geographical markets. I provide separate models for each source of revenue. That is to say that, for this study, I am particularly interested in predicting future monthly revenue at the album level across geographies and separately within each revenue stream.

The raw data is stored in the Client's database as individual revenue line-items for each sale, where each sale is uniquely identified by an associated universal product code (UPC). Importantly, UPCs may not be equated to albums. Different UPCs exist for different formats and versions of the same album.[2] For example, a specific artist's album titled "Greatest Hits" would have three distinct UPCs for versions on vinyl, on CD, and for a special, boxed-set edition, also on CD. There may be additional UPCs for different e-audio formats and distinct international versions. The database contains over 13,000 distinct UPCs, which are associated with almost 10,000 album titles and over 4,000 artists with sales over the January

[1] The data only includes catalog albums–those that were released over two years ago–and not new releases. Sales of new releases are affected by many factors outside this study, such as the popularity of an artist's last album and marketing.

[2] To simplify the data structure, albums are grouped into five main formats (CD, E Audio, Vinyl, Super Audio, and Other). Additionally, album versions are collapsed by title.

2012 to May 2014 timeframe. The line-item data was aggregated to revenue at the album level to create the final dataset, which has monthly sales nested within albums nested within geography.

One of the most interesting features of the data concerns the availability of data at the second and third levels of the hierarchy. Various level-2 attributes are associated with each album, the most important of which are format, as described above, artist, and release and reissue dates. Additional fields are calcuated to examine a holiday sales effect, if any, by album. Noticeably, the data are missing information on marketing expenditures and any notable exogenous events such as an artist death or other highly publicized events that might be associated with increased or decreased sales. In addition, no level-3 attributes, such as income per-capita, are specified. Quite simply, the level-2 and level-3 observed variables do not provide complete information on the level-2 and level-3 units. The potential effects of these missing variables are described later with other interesting features of the data.

3.1 Data Summary

The principal revenue streams are: physical sales, which are physical units (CDs, vinyl, etc) sold to retailers; digital sales, such as iTunes downloads; and streaming, where fractional revenue is earned whenever a user plays a song on streaming services such as Spotify and Pandora. It is worth emphasizing that, for the Client, this represents two different distribution methods. For physical sales, the Client has business to business (B2B) sales to retailers; and, for digital and streaming sales, there is no retailer–the sales are direct to consumer (DTC).

The geographies of interest for this study and their associated revenue are summarized in Table (3.1) below. As can be seen, the United States accounts for the vast majority of the Client's sales.[3] This may be due to many factors, including: the established distribution networks and business focus of the Client; the catalog composition, which is heavy on US-based artists; or cultural tastes in music and spending patterns that differ by geography. But, as noted above, variables on cultural differences and spending information are exogenous to the model, and therefore these relationships cannot be adequately examined.

Geography	Total	Physical	Digital	Streaming
Canada	4.87%	47%	49%	4%
CMG	3.83%	100%	0%	0%
iTunes Europe	1.04%	0%	100%	0%
Japan	8.45%	77%	22%	1%
Licensed Territories	4.49%	48%	33%	19%
United Kingdom	4.51%	41%	43%	16%
France	2.49%	53%	28%	19%
Germany	3.54%	73%	24%	2%
United States	66.79%	43%	48%	8%

Table 3.1: Summary of revenue by geography and revenue stream (January 2012 to April 2014). Percentage of overall revenue by geography and revenue-stream percentage within geography are provided. Revenue is for catalog albums only; new release revenue is not included.

In addition to large concentrations of revenue by geography, extreme concentrations of revenue exist when looking at revenue by either artist or album. In general, music is a business that exhibits the traits of the *economics of superstars* as outlined in Rosen (1981). Rosen notes that markets dominated by superstars, which he defines as "markets where a relatively small number of people earn an enormous share of the overall market revenue," exhibit two key features. Firstly, they exhibit a close connection between personal economic rewards and personal market share. Secondly, market size and reward tend to be highly skewed towards the most talented. As Rosen notes, from a consumer perspective, "lesser talent

[3] I note that the Client's revenue composition in the United States differs markedly from the composition of the overall industry. This may reflect differences between the Client's music catalog and the overall music industry. It may also partially reflect the focus of this study on older albums instead of new releases.

[is] often a poor substitute for greater talent." This phenomenon leads to markets where "small differences in talent become magnified in larger earnings differences, with greater magnification of the earnings-talent gradient increas[ing] sharply near the top of the [talent] scale." What Rosen is describing is markets where consumers have a strong preference for the absolute best talent available and, more importantly, where economies of scale allow individuals who possess this top talent to easily satisfy extremely broad segments of the market. In the music business, the cost of producing an additional digital copy of a song or CD is essentially 0. Music is therefore a *non-rival good*, meaning that one individual listening to a song does not exclude anyone else from listening to the same song. The *economics of superstars* phenomenon is evident in the data, which can be seen Figures (A.1), (A.2), (A.3) and (A.4) in Appendix A.1. In these figures, I have ranked both albums and artists by their aggregate log-revenue for each revenue stream individually and total revenue. I compare these rankings to the aggregate revenue associated with each rank. This results in a comparison of aggregate log-revenue across each revenue stream by both album and artist revenue-rank, which clearly shows the concentrations of revenue associated with markets dominated by superstars.

Lastly, turning to distributions, the variable on revenue in the dataset is roughly log-normally distributed but exhibits some deviation from log-normality at both level-1 and level-2, where revenue is somewhat over-dispersed. This is due to the extreme concentrations of revenue by album. The data is positively skewed at level-2, while, at level-1, the data has an extremely long, thin tail. The vast majority of monthly revenue falls into the \$0 - \$99 range; however, some albums have some months with very high revenue as can be seen in Table (A.2) in Appendix A.1. Within the physical revenue stream, note that there are two tails. That is, some albums also have some months with extreme negative revenue, where negative revenue represents refunds to retailers for unsold units. Since both digital and streaming revenue come from direct to the consumer distribution channels, there are no returns and therefore no months with negative revenue.

3.2 Interesting Features of the Data

In this section, I explore some of the interesting features of the data, particularly as they relate to HLMs. Most of these features arise from the fact that this is an observational study and not a designed experiment. The data are therefore highly imbalanced, and there is missing data, as mentioned above, on several key features of interest. I additionally discuss the possibility of cross-classification and issues of missingness that are related to both data collection and data preparation.

The first feature of note in the data is the lack of level-2 covariates. At level-2, the data are missing information on the amount spent on marketing for either specific albums or larger batches of albums such as those within a single genre or from a specific artist. These are potentially important missing variables and may have some bias on the parameter estimates. One potential mitigating effect on the importance of these missing variables is that this study is focused on older catalog albums–albums that were released at least two years ago–and many of which were released over ten years ago. Most marketing efforts are focused on new releases, which have more variable revenue and more easily influenced fans. It may be the case that there is very little to no marketing effort on these albums at all, although I do not have data to either confirm or disprove this. Additionally, older catalog albums tend to have stable fan bases. Consumers have largely decided whether they like an album or not by the time it has been out for a long time.

The other important level-2 attribute that is missing is key events related to the album, such as the death of the album's artist(s), a highly public event for the album's artist(s), or the release of a new hit from the album's artist(s). Any of these events may lead to a substantial increase in interest for the work of these artists and therefore sales of their music. The most famous example of this phenomenon was the death of Michael Jackson in 2009. In the five years since his death, Michael Jackson sold 13.2M albums, far more than in the five years prior to his death (3.9M albums).[4] Another example would be the release of a new hit album from an old artist, which leads to a dramatic increase in sales of their older work. Unfortunately, this dataset doesn't capture these data attributes. These missing variables may lead to some bias in parameter estimates.

The data is also missing level-3 covariates. At level-3, the particular covariates of interest include information capturing the consumer behavior of each geography, such as income per-capita, disposible income per-capita, information on differing musical preferences by geography, and consumer sentiment. To address this problem, I have decided to build two-level models instead of three-level models. This choice represents a trade-off, where I lose the richness of variance estimates which might provide insight into cross-geography purchasing behavior in order not to introduce a source of bias with unknown magnitude into the modeling process. But, since the missing covariates also mean that I am not able to estimate the fixed-effects parameters related to differences in consumer behavior across geography or to interpret their signifigance, this seems like a reasonable trade-off. In general, it is preferable to acknowledge shortcomings in the data, and therefore in the potential insight the data can provide, than to introduce bias into the estimates of parameters on which one does have data.

The next interesting feature of the data to consider is the hierarchical classification of monthly sales. As noted above, the data has been aggregated at the album level where albums have 1 of 5 main album-formats: CD, E Audio, Vinyl, Super Audio, and Other. One might therefore consider the data cross-classified at level-2 with level-2 variance partitioned by both album and album-format (Gold-

[4] Hughes, Jason. "Michael Jackson Has Sold More Albums Since His Death Than Over The Last 13 Years of His Life." (`http://www.thewrap.com/michael-jackson-has-sold-more-albums-since-his-death-than-the-last-13-years-of-his-life/`). Accessed: 2015-04-16

stein (1994)). But this would be a naive interpretation of the data for two reasons. Firstly, the data is highly imbalanced between formats. Looking at the United States geography alone, which is by far the largest and most significant geography for the Client, it is clear that the *vast* majority of sales are within the CD format, as is shown in Tables (3.2), (3.3), and (3.4) below.[5] Given the design imbalance, it is unlikely that sufficient sample size exists to provide accurate estimates of any cross-classified variance structure.

An additional, and more important, reason that cross-classification would be a naive interpretation is that the nature of these formats make it likely that they represent distinct consumer preferences, which should be considered distinct level-2 variance partitions. Specifically, vinyl, which is an analog music medium, was the primary music-storage medium during the early part of the twentieth century. Its sales rapidly declined after the introduction of the CD, a digital music medium, in the 1980s. By the early 1990s, vinyl had been almost completely replaced by CDs; however, vinyl has seen a resurgence as a niche market for audiophiles and music connoisseurs.[6][7] Super Audio, which is a CD with a higher bit-rate and purported better sound quality, has found a niche market for audiophiles in a similar fashion.[8]

Since cross-classified variance structures are uncorrelated by definition, this would not capture the reality where album format represents distinct consumer preferences. Instead, I consider the combination of album title and album format to be a unique specification for level-2 units. This allows for different parameter estimates for monthly sales of vinyl, super audio, CD, and other audio formats

[5]The same pattern is observed across all geographies and revenue streams.

[6]"The Death of the Vinyl LP?" (`http://mistervideo.net/the-death-of-the-vinyl-lp/`). Accessed: 2015-04-16

[7]Kornelis, Chris. "Why CDs May Actually Sound Better Than Vinyl." (`http://www.laweekly.com/music/why-cds-may-actually-sound-better-than-vinyl-5352162`). Accessed: 2015-04-16

[8]Del Colliano, Jerry. "The Symbolism of Losing Tower Records." (`http://www.avrev.com/news/1006/19.tower.shtml`). Accessed: 2015-04-16

for the same album title. It also allows for different hyperparameter estimates for monthly sales globally within a given audio format.[9] It should be expected that different audio formats have different estimates for monthly sales. Audiophiles and music connoisseurs are willing to pay the higher prices that vinyl and super audio formats command, which should be reflected in their hyperparameter estimates.

	CD	E Audio	Super Audio	Vinyl
	92.4%	0%	5.0%	2.6%
5%	1	0	0	0
25%	21	0	0	0
50%	27	0	0	0
90%	30	0	0	0
95%	30	0	6	0
99%	30	0	29	28
100%	30	0	30	30

Table 3.2: Summary of album formats for physical sales within the United States. Percentage of total observations by format is provided first. The distribution of observations within level-2 units by format is also presented.

	CD	E Audio	Super Audio	Vinyl
	91.9%	7.5%	0.5%	0.1%
5%	1	0	0	0
25%	19	0	0	0
50%	29	0	0	0
90%	29	3	0	0
95%	29	17	0	0
99%	29	29	0	0
100%	30	30	29	28

Table 3.3: Summary of album formats for digital sales within the United States. Percentage of total observations by format is provided first. The distribution of observations within level-2 units by format is also presented.

[9]In this case, the term "globally" refers to within a specific model. Specifically, by "globally", I mean within a specific geography and revenue stream.

	CD	E Audio	Super Audio	Vinyl
	93.8%	5.4%	0.5%	0.3%
5%	0	0	0	0
25%	26	0	0	0
50%	29	0	0	0
90%	29	0	0	0
95%	29	10	0	0
99%	29	29	0	0
100%	30	30	29	29

Table 3.4: Summary of album formats for streaming sales within the United States. Percentage of total observations by format is provided first. The distribution of observations within level-2 units by format is also presented.

Returning to design imbalance, it is known to effect the standard errors of parameter estimates as described in Maas and Hox (2004) and (2005). This is primarily because parameter estimates for HLMs have asymptotic properties, which translates to the requirement that sample sizes be sufficiently large for these properties to hold. The simulation studies in the literature to-date suggest that the number of level-2 units is far more important than total sample size or design imbalance between groups. In particular, Maas and Hox note that level-2 variances only appear to be underestimated when the number of groups is "substantially lower than 100" (Maas and Hox. (2004), Maas and Hox. (2005)). Since all of but wo of my models have over one hundred groups, and most have several hundred or more groups, this should not present a problem with the estimates. I note that the study is also imbalanced within formats. I investigate the effects of design imbalance more thoroughly in the simulation studies in Chapter 4.

A final feature of interest in the data concerns data missingness and is related to the motivation behind the origin of this study. A little background is thus required to properly discuss this feature. This study was motivated by the need to provide the Client with insight into the revenue data provided by the Client's distributor. In the current business environment, the Client receives monthly revenue data from their distributor but has no guarantee that the provided data is either accurate or complete. One desired piece of insight is a method for detecting revenue anomalies and errors, which is one of the key challenges facing music industry participants outlined in Chapter 1. The lack of data transparency impacts missingness in two ways. Firstly, it may be the case that there is an unknown error in the data reporting provided to the Client. This represents a potential exogenous source of missingness on the dataset. The second way in which missingness may be affected has to do with how the modeling process was designed to address the detection of revenue anomalies. Revenue is considered to be anomalous when it is either substantially below or substantially above expectations. Within the modeling process, therefore, data cleaning was done to *exclude* observations with excessively high or excessively negative monthly revenue. In general, this was done heuristically. QQ-plots were created from the monthly revenue data for each model, and any observations that were deemed to be severe outliers or otherwise deviations from log-normality were excluded. This represents a potential endogenous source of missingness on the dataset. In addition to these two sources of missingness, there may be some missingness due to the revenue reporting process. Specifically, whenever a UPC records exactly \$0 sales in a given month, the distributor may not report this \$0 monthly revenue. Such a dataset may be said to have *zero-deflation*, where there are many more true zeros than are observed.

These three possibilities for the generating mechanism for missingness have different implications. If all missingness is related to either data-reporting errors or from data cleaning, the data is Missing Completely At Random (MCAR). But, missingness due to non-reporting of \$0 revenue, or zero-deflation, is not MCAR. Data that is Not Missing At Random (NMAR) is problematic for both inferential and predictive claims. However, in the case of NMAR, the degree of missingness is also unclear: that is how many months are missing. Specifically, the possibility of data being NMAR is confounded by the possibility of its being MCAR.[10] To

[10]A fourth possibility for missingness exists. If an album was released in the last few years, it may have moved to "catalog" status within the study period. This missingness pattern is termed

investigate the impact of missingness, sensitivity analyses on the robustness of predictive results to all these potential sources of missingness are examined in Chapter 4.

Missing At Random (MAR) and is not problematic for estimation purposes.

CHAPTER 4

Model Results, Sensitivity Analyses, and Discussion

In this chapter, I describe the modeling approach and results. In addition, I present the results of sensitivity analyses examining the effects of missing data, design imbalance, and outliers at level-2. Firstly, turning to the models, I have built separate models for each of the revenue streams–physical, digital, and streaming. As discussed above, separate models were built for each geography in order to cope with the lack of covariates at level-3.

All models were built in R v3.1 (2015) with HLMs built in the **lme4** package (R Core Team (2015), Bates et al. (2014)). While **lme4** does provide shrinkage estimators for the level-2 parameters, it is not able to fit heteroskedastic or other complex covariance structures.[1] This is concerning because of the extreme concentration of revenue within each geography to the top albums as shown in Figures (A.1), (A.2), (A.3) and (A.4) in Appendix A.1. Given the software inflexibility, I have separated each geography into tiers by artist revenue such that artists are tiered within each geography roughly by the variance in their monthly album revenue.[2] While the tiering approach should adequately address the concern of heteroskedastic variance across albums, the skew in the revenue concentration does

[1]The **nlme** package in R does allow for complex covariance structures; however, **nlme** does not support downstream package development. Therefore, diagnostic methods are less flexible in **nlme**. As with any project involving a choice of modeling software, a trade-off is involved in this project. **lme4** provides the most complete suite of tools for this work.

[2]Tiers were created based on natural breaks in aggregated artist revenue within each geography from January 2013 - April 2014.

still leave the possibility of outlier level-2 units, particularly at the highest artist tier. Simulation results which examine the robustness of parameter estimates to outliers are discussed in this Chapter.

Generally speaking, the intraclass correlation (ICC), a measure of proportional partition of variance at level-2 (or level-3), is very high across all revenue streams, geographies, and artist tiers. This result is consistent with other empirical studies of longitudinal data. Given the sparsity of level-2 and level-3 covariates, I used a heuristic approach for model selection. Specifically, model selection was done by starting with a full model and applying backward selection to prune predictors such that the overall model fit was significantly improved at the $\alpha = 0.05$ level. After model fitting, diagnostic tests involving residual analysis and influence analysis were performed. In particular, the **HLMdiag** package was instrumental for many of these calculations (Loy and Hofmann (2014)). Any level-1 or level-2 units that were highly influential were removed. In the remainder of this chapter, I present the predictive results and discuss the results of the sensitivity analyses.

4.1 Sensitivity to Zero-Deflation

In Chapter 3, I discussed three possibilities for the missing data generating mechanism: errors in the data reporting process, exclusion of revenue anomalies during the modeling process, and zero-deflation. In this section, I investigate the possibility of zero-deflation, which would indicate the data is NMAR; and, examine the impact of missingness on predictive accuracy.

To do this, I conducted a sensitivity analysis where I randomly imputed a percentage (15%, 30%, 50%, and 70%) of missing data for each album and format type to zero. I then refit the HLMs on each of these augmented datasets and examined out-of-sample prediction accuracy. Prediction accuracy was compared for both the augmented data and the non-augmented data. Since the literature does not have a consensus on a single best measure of predictive accuracy, I have assessed prediction accuracy with a variety of metrics. Specifically, I use:

1. **Root Mean Squared Error (RMSE):** $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
2. **Mean Absolute Error (MnAE):** $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
3. **Weighted Mean Absolute Error (WMnAE):** $\frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i|y_i - \hat{y}_i|$ where the weights are the log of the absolute predicted values where a small constant has been added to upweight very small predicted values ($w_i = log(|\hat{y}_i| + 2)$)
4. **Maximum Absolute Error (MxAE):** Rather than being concerned with the average deviation, this metric is concerned with the maximum error in the test observations. $max_{i \in n}|y_i - \hat{y}_i|$
5. **Classification Rate:** Here, I calculate the 95% prediction interval $y_{i,PI} = \hat{y}_i \pm z_{.975}SE(\hat{y}_i)$ and classify observations as either within the prediction interval or outside it. The classification rate is then $\frac{1}{n}\mathbb{1}_{(y_i \in y_{i,PI})}$ where $\mathbb{1}$ is the indicator function.

The results of this analysis are shown below in Tables (4.1), (4.2), and (4.3). For brevity, only the United States models are shown. Results were similar across all geographies and revenue streams. These tests make it clear that the baseline models, without zero imputation, perform best across a variety of prediction metrics. Based on these results, I am confident that the data does not violate Missing At Random (MAR) assumptions.

Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 1						
Baseline	2,471	105.87	302.49	426.04	16,765.89	96.24%
Imp-15%	2,471	221.32	356.20	433.67	16,859.64	96.33%
Imp-30%	2,471	213.28	349.86	431.17	16,864.96	96.12%
Imp-50%	2,471	206.90	345.30	432.10	16,871.84	96.12%
Imp-70%	2,471	198.53	339.49	430.88	16,876.55	95.92%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 2						
Baseline	2,276	21.31	158.90	209.27	28,060.11	97.34%
Imp-15%	2,276	24.42	158.12	207.66	28,045.88	97.12%
Imp-30%	2,276	23.73	158.30	207.60	28,038.10	97.16%
Imp-50%	2,276	24.51	158.48	207.71	28,039.29	97.16%
Imp-70%	2,276	26.32	158.21	206.96	28,023.96	97.03%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 3						
Baseline	5,912	13.51	74.04	87.07	4,148.52	95.98%
Imp-15%	5,912	16.10	74.20	86.94	4,138.30	95.71%
Imp-30%	5,912	16.76	74.45	87.11	4,133.27	95.52%
Imp-50%	5,912	17.64	74.50	86.99	4,128.04	95.48%
Imp-70%	5,912	18.68	74.58	87.04	4,120.40	95.25%

Table 4.1: Summary of prediction accuracy for sensitivity analysis for zero-deflation by artist tier. Results for the United States geography and physical revenue stream shown.

Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 1						
Baseline	2,421	85.02	141.97	212.02	8,188.17	96.65%
Imp-15%	2,421	127.07	161.10	251.52	8,189.14	96.15%
Imp-30%	2,421	138.36	166.01	263.36	8,189.43	95.79%
Imp-50%	2,421	143.23	167.16	267.17	8,189.67	95.53%
Imp-70%	2,421	146.75	166.58	269.20	8,189.92	95.29%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 2						
Baseline	2,502	48.44	67.46	73.94	5,889.59	89.13%
Imp-15%	2,502	53.83	64.71	79.49	5,896.57	99.00%
Imp-30%	2,502	54.76	65.58	80.81	5,897.10	99.167%
Imp-50%	2,502	56.99	66.50	82.14	5,898.27	99.08%
Imp-70%	2,502	56.99	66.50	82.14	5,898.27	99.08%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 3						
Baseline	6,239	11.11	20.39	21.48	2,246.73	95.70%
Imp-15%	6,239	14.90	20.43	24.26	2,231.95	99.42%
Imp-30%	6,239	15.70	20.71	24.49	2,233.78	99.36%
Imp-50%	6,239	16.38	20.98	24.78	2,234.82	99.37%
Imp-70%	6,239	17.85	21.72	25.81	2,236.51	99.36%

Table 4.2: Summary of prediction accuracy for sensitivity analysis for zero-deflation by artist tier. Results for the United States geography and digital revenue stream shown.

Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 1						
Baseline	2,167	170.54	216.98	399.19	10,968.02	96.49%
Imp-15%	2,167	202.24	225.47	411.61	11,541.01	91.62%
Imp-30%	2,167	207.34	235.12	415.69	11,621.24	91.45%
Imp-50%	2,167	214.58	229.59	418.41	11,780.75	91.08%
Imp-70%	2,167	216.29	230.24	419.01	11,818.69	90.91%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 2						
Baseline	2,272	75.39	95.05	110.44	5,916.90	86.13%
Imp-15%	2,272	87.14	96.45	114.31	5,918.67	89.13%
Imp-30%	2,272	88.79	97.04	115.10	5,919.33	88.81%
Imp-50%	2,272	91.09	97.10	116.66	5,920.21	88.45%
Imp-70%	2,272	93.83	99.18	118.49	5,920.84	87.49%
Imputed Percentage	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate
Tier 3						
Baseline	6,696	27.02	32.54	37.46	2,243.20	95.96%
Imp-15%	6,696	29.58	32.97	36.77	2,246.29	92.27%
Imp-30%	6,696	29.84	32.99	36.80	2,246.25	91.97%
Imp-50%	6,696	30.03	33.11	37.04	2,246.24	91.63%
Imp-70%	6,696	30.57	33.17	37.08	2,246.25	90.96%

Table 4.3: Summary of prediction accuracy for sensitivity analysis for zero-deflation by tier. Results for the United States geography and streaming revenue stream shown.

4.2 Sensitivity to Design Imbalance

The maximum likelihood estimates commonly used in HLM analysis (ML or REML) are asymptotic. This roughly translates into the assumption that sample sizes must be sufficiently large for properties of assymptotic consistency and normality to hold. An obvious question is what constitutes an acceptable lower limit for sample sizes. Simulation studies in the literature have shown that the number of observations per level-2 unit is far more important than the total sample size. In addition, it has been shown that the number of level-2 units is very important for efficient estimates. In general, one hundred level-2 units are suggested as a minimum, although Kreft and De Leeuw (1998) note that the smallest acceptable number is 30. In simulation studies, Maas and Hox (2005) show that "[level-2] variances are estimated too small when the number of [units] is substantially lower than 100." With 30 units, they show that standard errors are about 15% too small and, with 50 units, about 9% too small. Maas and Hox (2005) also investigate the impact of level-2 units being non-normally distributed about the hyperparameter mean, finding that non-normality has little-to-no effect on fixed effects but does affect the variance components. They note that while the variance estimates are unbiased, their standard errors are not unbiased in all simulated conditions.

While many simulation studies in the literature have investigated the issue of sufficient sample size for both level-1 observations and level-2 units, I have not found simulation studies investigating the effect of design imbalance on the bias and efficiency of estimators. The effects of design imbalance are particularly important for this study, where it is substantial.[3] In this section, I present a simulation study to examine the effect of design imbalance.

[3] See Tables (3.2), (3.3), and (3.4) for the design imbalance of the United States geography.

To do so, I use a simplistic two-level model with a single explanatory variable at level-1 and no explanatory variables at level-2, which is designed to mimic simplified conditions of the models. Random intercepts are included for the level-2 units, defined by the unique combination of album title and album format. Unique hyperparameter fixed effects for each of the four different types of album formats and for the single explanatory variable are included. I also include a random slope for the single explanatory variable, a binary variable modeling the increase in album revenue during holiday months. Ten percent of level-2 units are simulated to have a valid significant holiday effect, while the other 90% have a small random effect. The model is provided by Equations (4.1), (4.2), (4.3), and (4.4) below.

$$\beta_j \sim N(\mu, \Sigma_\beta) \tag{4.1}$$

$$y_{ij} \sim N(X_j(\beta_j + Z_u \otimes u_j), \sigma_y^2) \tag{4.2}$$

$$\mu = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \mu_{03} \\ \mu_{04} \\ \mu_{05} \end{bmatrix} \sim N \left(\phi = \begin{bmatrix} \phi_{01} \\ \phi_{02} \\ \phi_{03} \\ \phi_{04} \\ \phi_{05} \end{bmatrix}, \begin{bmatrix} \tau_{01} & 0 & 0 & 0 & 0 \\ 0 & \tau_{02} & 0 & 0 & 0 \\ 0 & 0 & \tau_{03} & 0 & 0 \\ 0 & 0 & 0 & \tau_{04} & 0 \\ 0 & 0 & 0 & 0 & \tau_{05} \end{bmatrix} \right) \tag{4.3}$$

$$u_j \sim N(0, \Sigma_u) \tag{4.4}$$

Here X_j is the matrix of explanatory variables for the jth group, β_j is the vector of level-2 parameters for each album format and increased album revenue during holiday months. The hyperparameter vector μ describes overall monthly album revenue with the associated variances Σ_β. The hyperparameter for increased holiday revenue during holiday months μ_{05} is a mixture distribution where 10% of level-2 units have significant increases in revenue during holiday months and 90% of level-2 units do not. In all cases, the hyperparameter for the mean is fixed: $\phi = [4.4, 3.2, 1.5, 4, 0.5]$ where 10% of level-2 units have $\phi_{05} = 4.4$. The diagonal elements of Σ_β are $[1, 1, 1, 1, 0.5]$. Z_u is the design matrix for random effects and the random effects u_j are normally distributed about zero.

There are several conditions, listed below, that I am interested in. I use 1,000 datasets for each test condition.

1. I first create a balanced case where there are 100 groups for each album format and each group has 30 observations. This case is used as a baseline.

2. I vary group size for a single album format, looking at cases where one album-format has group sizes of 10, 15, and 20.

3. I vary group size for two album-formats, looking at group sizes of 10 and 20.

4. I investigate the case where group size varies within a single album-format. Within this format, level-2 units have group sizes of 10, 15, 20, and 30.

5. I investigate the case where group number is imbalanced by album-format. In this case, two album-formats have 100 level-2 groups, while the other two have 10, 30, 50, and 70 in different cases.

6. I investigate the case where both group size and group number vary. Here, two album-formats have group numbers of 30 and 50 and group sizes of 20 and 15. The other album-formats have group number of 100 and group size of 30. This presents four cases.[4]

[4]The results with a group number of 50 are similar to those with a group number of 30. The results with a group number of 50 are not shown for brevity.

7. I investigate the case where I vary group size both within a single format and between formats, and, simultaneously, vary group number between formats.

8. Lastly, I examine results where group number is small with varying group size. I look at the case where group number is either 40 or 80, which is either balanced between formats or imbalanced. In conjunction with the varying group numbers, I use group sizes of 10, 20, and 30.

To indicate the accuracy of parameter estimates, the percentage bias is used. For a given parameter θ, the percentage bias is defined $\frac{\hat{\theta}-\theta}{\theta} \times 100\%$. Since I am primarily interested in the level-2 parameter estimates for predictive purposes, I calculate the 2.5th, 50th, and 97.5th percentiles of percentage bias at level-2 for each simulation and then average across all simulations in a given simulated condition. Level-3 percentage bias is also calculated for each dataset and the 2.5th, 50th, and 97.5th percentiles of percentage bias in a given simulated condition is reported. To give an estimate of estimation efficiency at level-2, I use Mean Squared Error (MSE), where MSE($\hat{\theta}$) is defined MSE($\hat{\theta}$) $= Var(\hat{\theta}) + (\hat{\theta} - \theta)^2$. As with percentage bias, I report the average 2.5th, 50th, and 97.5th percentiles of level-2 MSE in each simulated condition.

The results of these simulations are presented in Appendix A.2. Looking at the results of the baseline case (Table A.3), it is clear that the largest range for percentage bias comes from the Other format, which has a hyperparameter mean ϕ_{03} that is the most different from the other hyperparameter means. While albums with the Other format are the least common in the true data, this does suggest that estimation accuracy suffers when estimating heterogeneous variance structures, which are not currently estimated in **lme4**. The MSE for all the hyperparameter means in the baseline case are near 1, although underestimated by 1%. This indicates that there is very litle bias and accurate estimation of the true variance parameter at baseline. I also note that the MSE for the holiday fixed effect is substantially higher than the true values, indicating that the mixture distribution of parameters is poorly estimated. HLMs do not have estimation techniques for parameters with a mixture distribution.

Looking at the results for cases where there is design imbalance between the group sizes of the audio formats (cases 2 and 3 above) and where there is design imbalance within a single audio format (case 4), there is little change in estimation efficiency at a group size of 20, although a group size of either 10 or 15 does increase the range of percentage bias in level-2 units. Specifically, while the 50th percentile level-2 unit estimate is still unbiased, the range of percentage bias roughly doubles as group size moves from 30 observations to 10 observations. The smaller group size also appears to impact variance estimates, which are underestimated by 2% (group size of 20) or 3% (group size of 10). This same pattern is observed regardless of whether group size is imbalanced in a single audio-format or with two audio-formats. Consistent with the literature, group size does not appear to have an effect on the level-3 parameter estimates. The number of level-2 units is more important than the total sample size or sample size per group for level-3 estimates.

Moving to the results of cases where there is design imbalance in group number (case 5), a different pattern emerges. While the lower group number has little affect on the level-2 percentage bias or MSE, it does substantially increase the range of level-3 bias. Once again, the median estimate is relatively unbiased; but, when group size is 10, instead of 100, the range of level-3 bias roughly triples. The change in the range of level-3 bias narrows as group number increases. At 70 groups, the difference is minor.

Cases 6 and 7 look at design imbalances where both group size and group number are imbalanced. A similar pattern to the prior simulations emerges. As group number decreases from 100, the range of percentage bias at level-3 increases, although the median estimate is still unbiased. As group size decreases from 30, the range of percentage bias at level-2 increases, although the median estimate is still unbiased. In all cases, the variance estimates are slightly underestimated. A similar pattern is clear in case 8, when there are few total groups across all formats. The decrease in group number leads to an increasing range of potential bias at level-2 and the decrease in group size leads to an increasing range of potential bias at level-3.

The overall conclusion from these simulations is clear. A large sample size at both level-2 and level-3 is preferred. As group number decreases from 100, the range of percentage bias at level-3 increases. And, as group size decreases from 30, the range of percentage bias at level-2 increases. In all cases, the variance estimates are slightly underestimated, and the median bias in level-2 and level-3 parameter means is negligible. These results increase my confidence in the robustness of the predictive results for this study, where I have well over 100 groups in all but 2 models and where the majority of groups are of size 30.

4.3 Sensitivity to Outliers

In this section, I present a simulation study examining the robustness of parameter estimates to level-2 outliers. One possible problem with using the **lme4** package in R is that it does not offer full flexibility for estimating complex covariance structures of the level-2 (or higher) parameters. Despite the artist tiering approach taken to reduce the possibility of heteroskedastic variance of level-2 units, the possibility of outlying level-2 units still exists. These outlying units may introduce bias into the parameter estimates. As Maas and Hox (2004) note, when assumptions of normality and large samples are not met, parameter estimates have been shown to have a small downward bias, and level-2 variance components estimates may be underestimated.

As above, I use a simplistic two-level model with a single explanatory variable at level-1 and no explanatory variables at level-2, conforming to Equations (4.1), (4.2), (4.3), and (4.4) reprinted here.

$$
\begin{aligned}
\beta_j &\sim N(\mu, \Sigma_\beta) \\
y_{ij} &\sim N(X_j(\beta_j + Z_u \otimes u_j), \sigma_y^2) \\
\mu = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \mu_{03} \\ \mu_{04} \\ \mu_{05} \end{bmatrix} &\sim N\left(\phi = \begin{bmatrix} \phi_{01} \\ \phi_{02} \\ \phi_{03} \\ \phi_{04} \\ \phi_{05} \end{bmatrix}, \begin{bmatrix} \tau_{01} & 0 & 0 & 0 & 0 \\ 0 & \tau_{02} & 0 & 0 & 0 \\ 0 & 0 & \tau_{03} & 0 & 0 \\ 0 & 0 & 0 & \tau_{04} & 0 \\ 0 & 0 & 0 & 0 & \tau_{05} \end{bmatrix} \right) \\
u_j &\sim N(0, \Sigma_u)
\end{aligned}
$$

This formulation is similar to that above. In this formulation, however, each of the first 4 elements of the hyperparameter ϕ is a mixture distribution where some proportion of level-2 units comes from an outlying mean $\phi_{k,b}$ s.t. $\phi_{k,b} = 3\phi_{k,a}$, $k = 1, 2, 3, 4$. In all cases, the hyperparameter for the mean is again fixed: $\phi_{.a} = [4.4, 3.2, 1.5, 4, 0.5]$ where 10% of level-2 units have $\phi_{05} = 4.4$, and outlying level-2 units have means 3 times greater than $\phi_{.a}$. The diagonal elements of Σ_β are $[1, 1, 1, 1, 0.5]$. I use 1,000 datasets in each test condition and use percentage-relative-bias and MSE to evaluate the simulations, as I did previously.

I am once again interested in several test cases. Along with a baseline case, I test cases where: (i) all album formats have the same percentage of level-2 units as outliers (5%, 10%, 15%, and 20% outliers); (ii) where only a single format has a percentage of level-2 units as outliers (5%, 10%, and 20%); and (iii) two cases where all album formats have some various percentages of outliers. In one case, these are 5%, 10%, 5%, and 15%; and, in another case 1%, 10%, 5%, and 1%. The results of these simulations are presented in Appendix A.3.

Looking at the results, increasing the percentage of outliers from 0% up to 20% has very little effect on the level-2 percentage bias. This is encouraging as it suggests that the predictive results will be robust to any outlying level-2 units. Somewhat surprisingly, outliers seem to have a substantial effect on the efficiency of the estimates for holiday effects. This may indicate that, in the presence of outliers, the holiday effect is confounded with the effects of outliers. More troubling, however, the MSE and level-3 percentage biases are greatly increased by the presence of outliers. Some of this increase is to be expected. After all, if $\phi_{0k}, k = 1, 2, 3, 4$ is truly a mixture distribution but is being estimated as a single parameter, it should be expected that the hyperparameter estimate is upwardly biased. The degree of upward bias can be roughly calculated. Excluding superscripts for convenience, it is expected that

$$\phi_a < \hat{\phi} < \phi_b \text{ such that}$$
$$\hat{\phi} \approx A\phi_a + 3B\phi_b$$

where $A \in \{.80, .85, .90, .95, .99\}$, $B \in \{.01, .05, .10, .15, .20\}$ and $A + B = 1$. This implies that for every 1% increase in the percentage of level-2 outliers, a 2%

increase in level-3 percentage bias is roughly expected.[5] Looking at Appendix A.3, this matches the observed pattern of level-3 percentage bias at the $50th$ percentile across all cases where outliers are present.

Taken together, these results suggest that the presence of outliers at level-2 should not have a substantial effect on the main predictive results, which are based on the parameter estimates at level-2. But the presence of outliers will tend to introduce substantial upward bias to the hyperparameter estimates. Since it is these hyperparameters which are used to make predictions for out-of-sample level-2 units–that is, unique combinations of album title and audio format which were not included in the training dataset–there is some concern about the predictive accuracy for these new units. In practice, it may be best to simply identify these new units and monitor their predicted and actual future revenue more closely.

4.4 Predictive Results

Now satisfied that the data, and predictive results, are robust to missing data, design imbalance, and outlying level-2 units, I proceed to evaluating the models' predictive performance. To assess predictive performance, I use out-of-sample monthly revenue observations from May 2014 to August 2014. The five metrics described above for evaluating data missingness (see section (4.1)) are used here to produce a consensus evaluation of prediction accuracy.[6] I again use multiple metrics since a consensus "best" predictive metric does not exist in the literature.

As can be seen, the models do a fairly good job of prediction with a RMSE typically less than \$125 per month, MnAE typically less than \$160 per month,

[5]This relationship is dependent on the fact that $\phi_b = 3\phi_a$. If the relationship between ϕ_a and ϕ_b were to change, this relationship would also change.

[6]Several of the models from smaller non-US geographies have very few out-of-sample observations for the test period. In these cases, I use a random sample of 20% of in-sample observations.

and accurate classification rate of roughly 96%. Models that have noticably higher predictive error are the physical revenue models for: tier-1 artists in Japan, CCR, tier-1 artists in the UK, and tier-1 artists in the Licensed Territories (LT). In each of these cases, the revenue is highly concentrated by album. That is, these models exhibit both level-2 heterogeneity and level-2 outliers.

	N	RMSE	MnAE	WMnAE	MxAE	Classification Rate	Sampled
US_physical_tier1	2471	105.87	302.49	426.04	16765.89	96.24	FALSE
US_physical_tier2	2276	21.31	158.90	209.27	28060.11	97.34	FALSE
US_physical_tier3	5912	13.51	74.04	87.07	4148.52	95.98	FALSE
US_digital_tier1	2421	85.02	141.97	212.02	8188.17	96.65	FALSE
US_digital_tier2	2502	48.44	67.46	73.94	5889.59	89.13	FALSE
US_digital_tier3	6239	11.11	20.39	21.48	2246.73	95.70	FALSE
US_streaming_tier1	2421	170.54	216.98	399.19	10968.02	96.49	FALSE
US_streaming_tier2	2502	75.39	95.05	110.44	5916.90	86.13	FALSE
US_streaming_tier3	6239	27.02	32.54	37.46	2243.20	95.96	FALSE
CCR_physical	192	381.49	877.65	1187.80	29601.59	94.27	FALSE
CCR_digital	253	214.56	406.75	581.72	9118.12	89.72	FALSE
CCR_streaming	169	115.67	329.02	658.22	10150.28	93.49	FALSE
CAN_physical_tier1	106	56.29	448.64	506.40	5648.13	90.57	FALSE
CAN_physical_tier2	612	7.96	42.23	52.55	1172.42	94.77	FALSE
CAN_physical_tier3	799	12.76	86.57	129.07	5200.92	99.75	FALSE
CAN_digital_tier1	365	116.61	128.86	171.25	9116.40	88.77	FALSE
CAN_digital_tier2	664	12.29	18.63	34.80	1076.81	95.03	FALSE
CAN_digital_tier3	1098	8.73	11.17	14.83	744.95	95.90	FALSE
CAN_streaming_tier1	329	15.83	19.96	26.68	325.96	92.40	FALSE
CAN_streaming_tier2	2710	1.01	1.31	2.18	124.35	96.90	FALSE
JAP_physical_tier1	398	1568.74	1633.62	1912.34	34978.79	80.49	TRUE
JAP_physical_tier2	508	16.59	119.01	128.58	469.93	92.59	TRUE
JAP_physical_tier3	642	211.78	443.98	445.51	10400.82	95.45	TRUE
JAP_digital_tier1	123	120.08	146.92	189.33	1884.44	97.56	TRUE
JAP_digital_tier2	319	24.14	33.63	52.64	1232.18	98.75	TRUE
JAP_digital_tier3	884	14.40	20.22	25.95	973.69	98.19	TRUE
JAP_streaming_tier1	88	6.56	8.38	11.24	100.68	98.86	TRUE
JAP_streaming_tier2	277	1.04	1.43	2.46	37.76	99.64	TRUE
JAP_streaming_tier3	681	1.29	1.71	2.26	45.23	98.97	TRUE
CMG_physical_tier1	194	143.44	393.20	459.98	5146.25	92.78	FALSE
CMG_physical_tier2	314	52.14	144.06	158.93	3082.61	96.18	FALSE
CMG_physical_tier3	362	62.60	149.99	172.27	2807.12	93.65	FALSE
FRA_physical1	243	265.02	279.90	284.45	6147.20	93.85	TRUE
FRA_physical2	360	24.24	67.34	73.05	598.02	97.01	TRUE
FRA_digital	287	21.92	31.00	55.89	1012.73	100.00	TRUE
GER_physical1	530	28.48	89.76	116.84	1607.23	96.60	TRUE
GER_physical2	515	16.02	53.30	66.29	736.52	97.48	TRUE
GER_digital	678	9.36	15.19	23.16	280.66	98.97	TRUE
GER_streaming	749	0.47	0.75	0.94	9.38	100.00	TRUE
LT_physical1	51	369.04	1031.12	1052.35	5923.75	90.20	TRUE
LT_physical2	643	18.36	86.26	99.28	1401.57	90.51	TRUE
LT_digital	997	29.35	37.12	58.25	927.17	98.50	TRUE
LT_streaming	812	24.07	31.88	47.02	665.51	99.14	TRUE
iEUR_digital	1132	0.19	10.83	14.78	151.74	99.20	FALSE
iEUR_streaming	1632	0.18	0.21	0.29	15.12	99.02	FALSE
UK_physical1	396	639.26	725.39	770.65	3210.37	57.14	TRUE
UK_physical2	350	78.19	160.22	167.93	774.97	94.44	TRUE
UK_digital	1882	10.95	22.85	37.72	2057.66	99.57	TRUE
UK_streaming	426	5.47	5.53	18.54	159.26	89.20	FALSE

Table 4.4: Summary of predictive results for all models. The first column, "N", indicates the number of observations used in the evaluation. The final column, "Sampled", indicates whether the 20% random sample of in-sample observations were used.

CHAPTER 5

Concluding Remarks

In this thesis I note two key challenges facing the music industry as its revenue model continues to transition from that of purely physical sales to an environment with multiple revenue channels dominated by digital mediums. These challenges require successful music industry participants to gain a deeper understanding of their data, particularly the data about their customers and revenue. I presented what I believe are the first models predicting monthly revenue at the album level across both geographies and revenue streams within the music industry. Finally, I discussed the impact of several interesting aspects of the data and how they impact the robustness of predictive results, specifically focusing on data missingness, design imbalance, and extreme concentrations of revenue by albums.

This works suggests several areas of additional research. Firstly, these models could be improved by collecting, and including, currently missing information on level-2 and level-3 attributes of the data such as the amount spent on marketing per album and geography, significant events by each album's artists, consumer music preferences by geography, and consumer income and spending patterns by geography. These variables would allow modeling of some of the major sources of uncertainty surrounding the models presented in this thesis. They would also allow analysts to gain a deeper understanding of the mechanics underlying music industry revenue, particularly how it varies across geographies for a particular album or artist.

Secondly, the modeling framework presented in this thesis focuses on a hierarchical linear modeling (HLM) approach. Alternative approaches, such as Dirichlet Process Mixture Models, could be evaluated and their predictive performance compared to the HLM results. Different modeling approaches have different underlying assumptions. It may be the case that a different set of assumptions, such as the non-parametric assumptions underlying Dirichlet Process Mixture Models, would produce a better fitting set of models.

A third avenue of further research is methodological and specific to the **lme4** software package in R used in this thesis. Specfically, future versions of the **lme4** package should focus on providing flexible estimation techniques for complex covariance structures including level-2 heterogeneity $\Omega_{e_j} = \sigma_j^2 R_j$ and $\Omega_{u_j} = \sigma_j^2 D_j$ and correlation of residual terms. For instance, the **nlme** package includes estimation for AR1, ARMA, CAR1, and several spatial correlation structures of HLMs. Work that advances the **lme4** package and provides estimation for the covariance structures available in **nlme** would prove fruitful to future analysts.

APPENDIX A

Appendix

A.1 Appendix A: Data Summary

Year	Total	Physical	Digital	Streaming	Other
1990	$13,659	100%	0%	0%	0%
2000	$19,692	100%	0%	0%	0%
2005	$14,897	91.1%	4.1%	1.4%	3.4%
2008	$9,651	65.7%	19.5%	3.7%	11.1%
2010	$7,615	52.2%	32.1%	6.6%	9.1%
2014	$6,972	32.6%	37%	26.8%	3.7%

Table A.1: Summary of music industry revenue in the United States at retail value for selected years and by revenue stream. Revenue is reported in millions of dollars and is inflation adjusted to chained 2013 dollars. Source: RIAA.

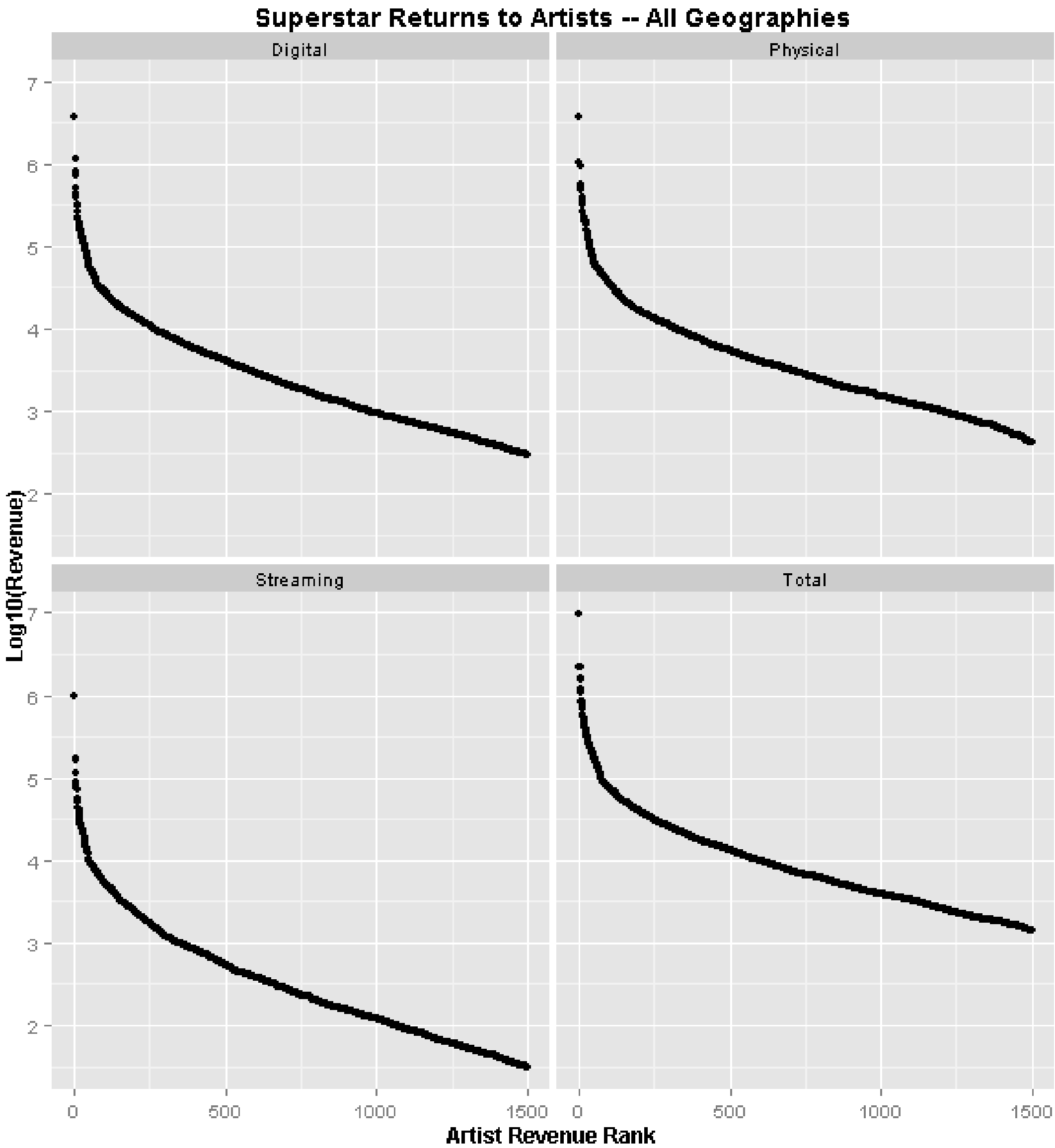

Figure A.1: Total Log10 revenue across all geographies by revenue stream and artist by artist rank. Panels, clockwise from top left: (1) Digital Revenue, (2) Physical Revenue, (3) Total Revenue, (4) Streaming Revenue.

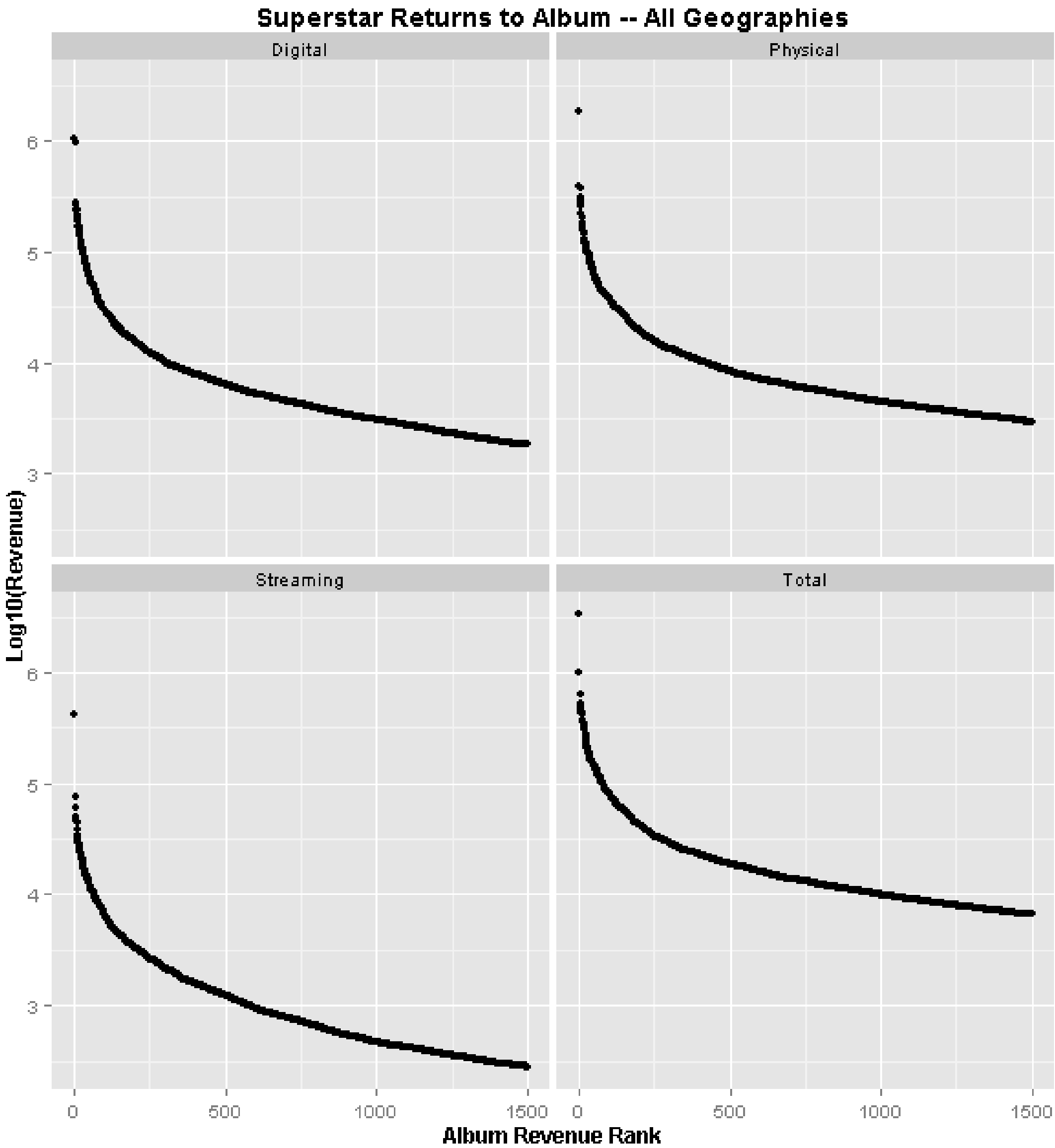

Figure A.2: Total Log10 revenue across all geographies by revenue stream and album by album rank. Panels, clockwise from top left: (1) Digital Revenue, (2) Physical Revenue, (3) Total Revenue, (4) Streaming Revenue.

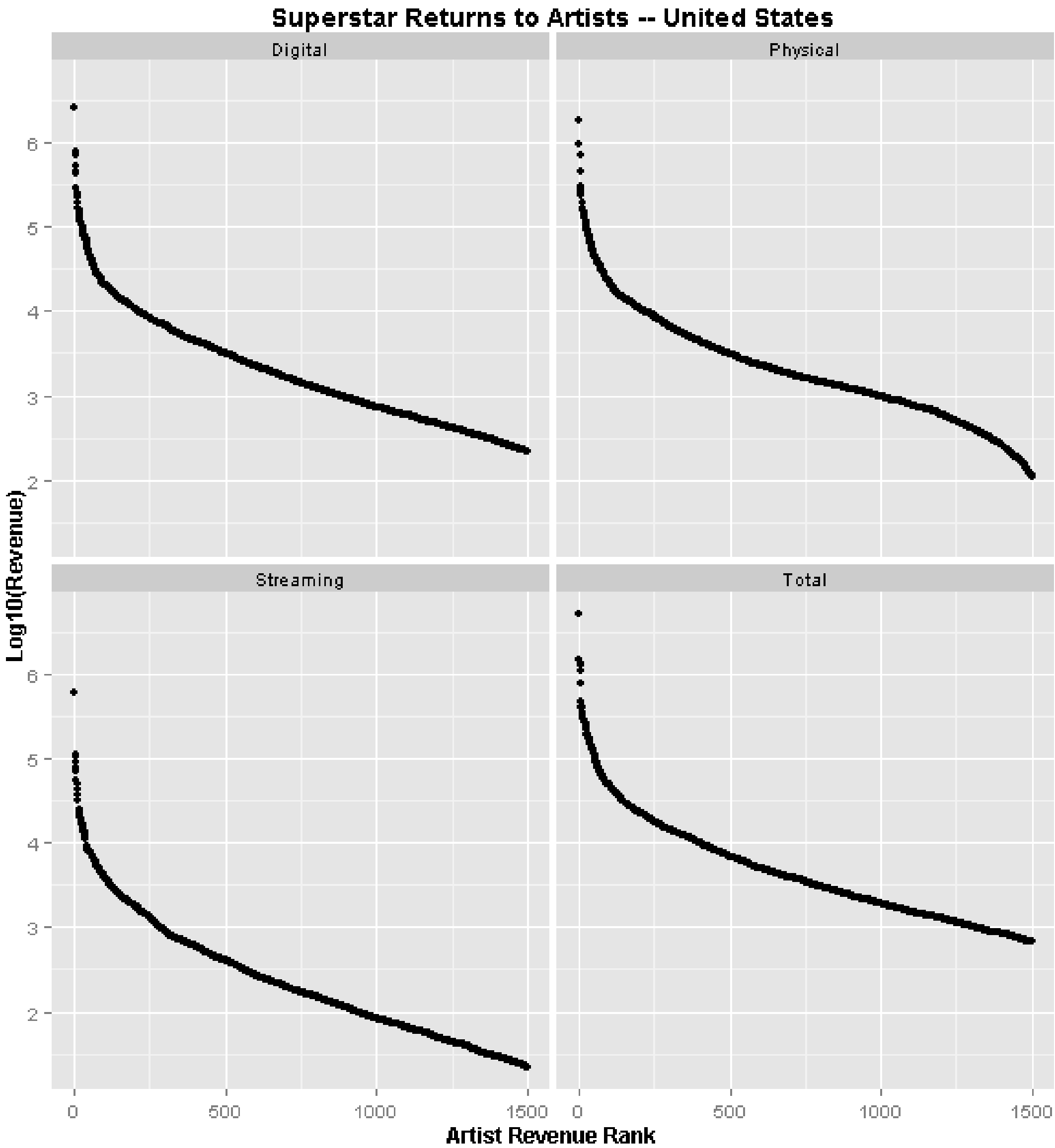

Figure A.3: Total Log10 revenue in USA geography by revenue stream and artist by artist rank. Panels, clockwise from top left: (1) Digital Revenue, (2) Physical Revenue, (3) Total Revenue, (4) Streaming Revenue.

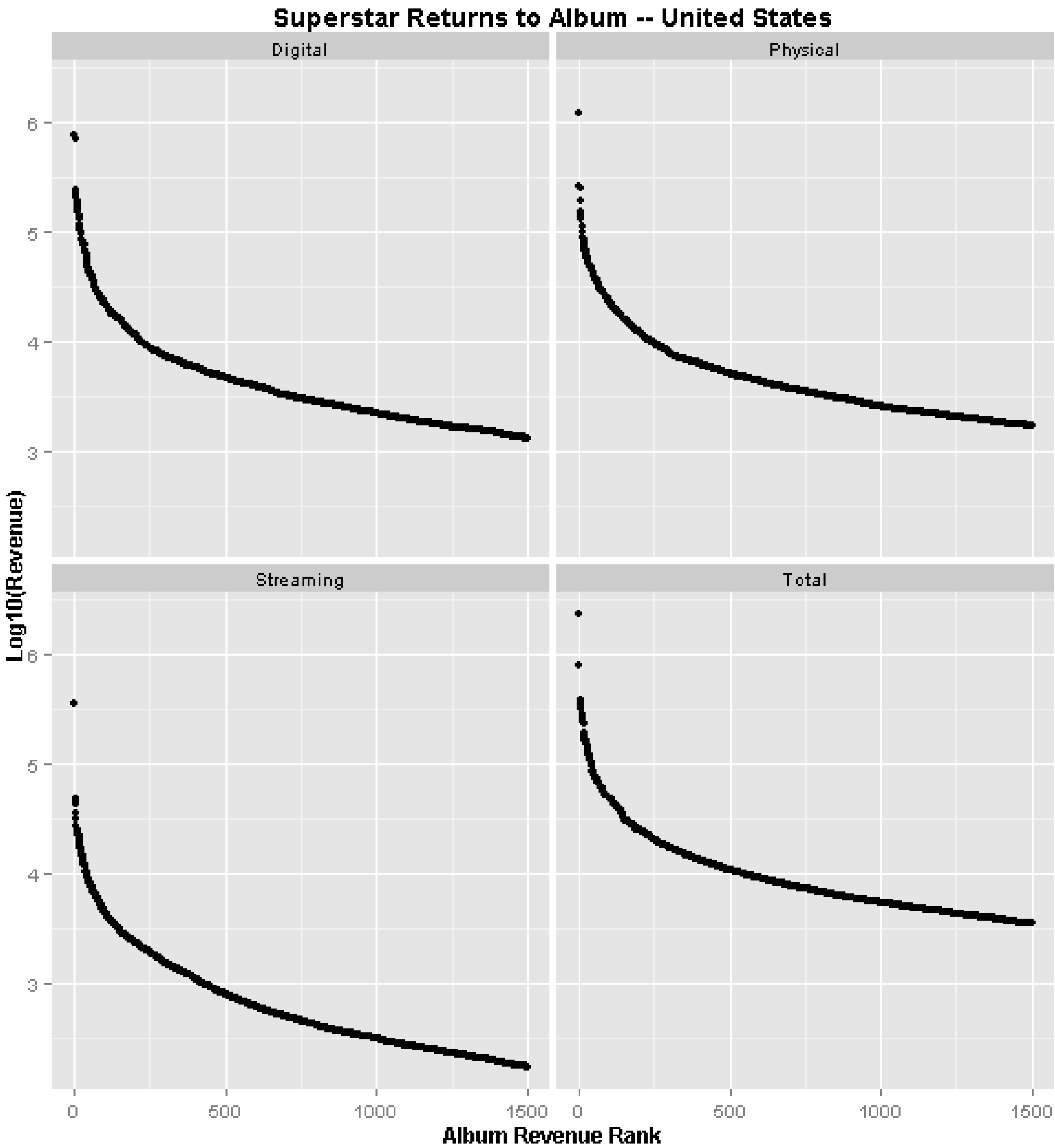

Figure A.4: Total Log10 revenue in USA geography by revenue stream and album by album rank. Panels, clockwise from top left: (1) Digital Revenue, (2) Physical Revenue, (3) Total Revenue, (4) Streaming Revenue.

Canada	**-\$10,000 - -\$1,001**	**-\$1000 - <\$0**	**\$0 - \$99**	**\$100 - \$999**	**\$1,000 - \$9,999**	**\$10,000+**
Digital	0%	0%	97.9%	1.9%	0.2%	0%
Physical	0.1%	23.9%	70.5%	4.9%	0.6%	0%
Streaming	0%	0%	99.9%	.1%	0%	0%
CMG	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	99.1%	0.7%	0.2%	0.%
Physical	0.2%	2.4%	61.3%	33.1%	2.9%	0.1%
Streaming	0%	0%	99.9%	0.1%	0%%	0%
iTunes Europe	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	97.7%	2.2%	0.1%	0%
Streaming	0%	0%	100%	0%	0%	0%
Japan	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	94.6%	4.7%	0.6%	0%
Physical	0%	27.9%	57.1%	11.1%	3.7%	0.2%
Streaming	0%	0%	99.7%	0.3%	0%	0%
Licensed Territories	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	93.8%	5.5%	0.7%	0.1%
Physical	0.1%	10.8%	75.9%	11.4%	1.6%	0.1%
Streaming	0%	0%	92.8%	6.8%	0.4%	0%
United Kingdom	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	94.5%	4.9%	0.6%	0%
Physical	0.1%	8.1%	79.4%	11.3%	1.%	0.1%
Streaming	0%	0%	99.5%	0.5%	0%	0%
France	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	96.9%	2.8%	0.3%	0%
Physical	0.1%	19.%	67.9%	11.4%	1.6%	0%
Streaming	0%	0%	99.%	0.9%	0.1%	0%
Germany	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	96.7%	3%	0.3%	0%
Physical	0%	13.1%	68%	17.4%	1.4%	0.1%
Streaming	0%	0%	99.9%	0.1%	0%	0%
United States	-\$10,000 - -\$1,001	-\$1000 - <\$0	\$0 - \$99	\$100 - \$999	\$1,000 - \$9,999	\$10,000+
Digital	0%	0%	84.7%	13.8%	1.4%	0.1%
Physical	0.5%	16.4%	60.2%	20.6%	2.3%	0.1%
Streaming	0%	0%	97.1%	2.7%	0.2%	0%

Table A.2: Summary of monthly revenue by geography, revenue stream, and bin (January 2012 to April 2014). Percentage of observations with recorded monthly revenue in each revenue bin is provided across each geography and revenue stream.

A.2 Appendix B: Simulation Results - Design Imbalance

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.2%	0.1%	6.3%	0.98	0.99	1.08
Level-2 (E Audio)	-8.7%	0.0%	8.7%	0.98	0.99	1.09
Level-2 (Other)	-18.6%	0.2%	18.1%	0.98	0.99	1.08
Level-2 (SACD)	-6.8%	0.1%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.23	2.28	2.93
Level-3 (CD)	-4.3%	0.1%	4.6%			
Level-3 (E Audio)	-6.5%	0.0%	6.7%			
Level-3 (Other)	-12.4%	0.3%	12.6%			
Level-3 (SACD)	-4.8%	0.1%	4.9%			

Table A.3: Bias and efficiency statistics from the baseline case.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-7.5%	0.1%	7.6%	0.97	0.99	1.12
Level-2 (E Audio)	-8.6%	0.0%	8.7%	0.98	1.00	1.09
Level-2 (Other)	-18.6%	-0.2%	18.1%	0.98	0.99	1.08
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.23	2.29	2.94
Level-3 (CD)	-4.5%	0.2%	5.0%			
Level-3 (E Audio)	-6.1%	0.0%	6.4%			
Level-3 (Other)	-13.9%	0.3%	12.8%			
Level-3 (SACD)	-4.9%	0.2%	4.7%			

Table A.4: Bias and efficiency statistics from the case where the CD album format has 20 observations per level-2 group. All other album formats have 30 observations per level-2 group.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.8%	0.1%	9.1%	0.96	0.98	1.17
Level-2 (E Audio)	-8.7%	0.0%	8.6%	0.98	0.99	1.09
Level-2 (Other)	-18.5%	-0.2%	18.2%	0.98	0.99	1.08
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.22	2.28	2.95
Level-3 (CD)	-4.9%	0.1%	4.6%			
Level-3 (E Audio)	-5.9%	0.1%	5.8%			
Level-3 (Other)	-13.5%	-0.2%	12.0%			
Level-3 (SACD)	-4.9%	0.1%	5.0%			

Table A.5: Bias and efficiency statistics from the case where the CD album format has 15 observations per level-2 group. All other album formats have 30 observations per level-2 group.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-10.0%	0.1%	10.3%	0.95	0.97	1.22
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.98	0.99	1.09
Level-2 (Other)	-18.5%	-0.2%	18.1%	0.98	0.99	1.08
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.22	2.28	2.95
Level-3 (CD)	-4.4%	0.2%	4.7%			
Level-3 (E Audio)	-6.0%	0.1%	6.1%			
Level-3 (Other)	-12.8%	-0.2%	14.5%			
Level-3 (SACD)	-4.7%	0.1%	4.7%			

Table A.6: Bias and efficiency statistics from the case where the CD album format has 10 observations per level-2 group. All other album formats have 30 observations per level-2 group.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-7.6%	0.1%	7.7%	0.97	0.98	1.12
Level-2 (E Audio)	-10.4%	0.0%	10.3%	0.97	0.98	1.12
Level-2 (Other)	-18.5%	-0.2%	18.1%	0.98	0.99	1.08
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.21	2.28	3.01
Level-3 (CD)	-4.4%	0.1%	4.8%			
Level-3 (E Audio)	-5.9%	0.2%	6.3%			
Level-3 (Other)	-12.8%	-0.2%	14.7%			
Level-3 (SACD)	-4.7%	0.1%	4.6%			

Table A.7: Bias and efficiency statistics from the case where the CD and E Audio album formats have 20 observations per level-2 group. All other album formats have 30 observations per level-2 group.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-10.0%	0.1%	10.4%	0.94	0.97	1.21
Level-2 (E Audio)	-13.9%	0.0%	13.9%	0.94	0.97	1.21
Level-2 (Other)	-18.5%	-0.2%	18.1%	0.98	0.99	1.08
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.98	0.99	1.08
Holiday Fixed Effect				2.21	2.30	3.07
Level-3 (CD)	-4.4%	0.3%	4.7%			
Level-3 (E Audio)	-6.2%	0.2%	6.0%			
Level-3 (Other)	-13.7%	-0.4%	13.1%			
Level-3 (SACD)	-5.0%	0.0%	4.5%			

Table A.8: Bias and efficiency statistics from the case where the CD and E Audio album formats have 10 observations per level-2 group. All other album formats have 30 observations per level-2 group.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.6%	0.0%	8.6%	0.97	0.98	1.16
Level-2 (E Audio)	-13.9%	-0.2%	13.9%	0.94	0.97	1.21
Level-2 (Other)	-18.6%	-0.2%	18.2%	0.97	0.98	1.08
Level-2 (SACD)	-6.9%	0.0%	6.9%	0.98	0.99	1.08
Holiday Fixed Effect						
Level-3 (CD)	-4.4%	-0.1%	4.5%			
Level-3 (E Audio)	-6.5%	0.0%	6.0%			
Level-3 (Other)	-13.8%	0.0%	12.7%			
Level-3 (SACD)	-4.9%	0.0%	4.8%			

Table A.9: Bias and efficiency statistics where group size is imbalanced between within a single format. In this case, the CD format has 25 level-2 groups with 10, 15, 20, and 30 observations each. All other formats have 30 observations in all level-2 groups.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.2%	0.0%	6.3%	0.98	0.99	1.08
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.98	0.99	1.08
Level-2 (Other)	-14.8%	-0.2%	14.4%	0.99	1.00	1.06
Level-2 (SACD)	-5.2%	0.0%	5.3%	0.98	0.99	1.06
Holiday Fixed Effect				1.55	1.60	2.15
Level-3 (CD)	-4.5%	0.3%	4.8%			
Level-3 (E Audio)	-5.7%	0.0%	5.8%			
Level-3 (Other)	-41.2%	-1.0%	39.7%			
Level-3 (SACD)	-15.3%	-0.6%	15.4%			

Table A.10: Bias and efficiency statistics where group number is imbalanced between album formats. In this case, CD and E Audio have 100 groups while Other and Super Audio have 10 groups. All formats have 30 observations in all level-2 groups.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.2%	0.0%	6.4%	0.98	0.98	1.07
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.98	0.99	1.08
Level-2 (Other)	-17.3%	-0.2%	16.9%	0.96	0.97	1.06
Level-2 (SACD)	-6.4%	0.0%	6.4%	0.98	0.99	1.07
Holiday Fixed Effect				1.82	1.87	2.42
Level-3 (CD)	-4.4%	0.0%	4.5%			
Level-3 (E Audio)	-6.2%	0.0%	6.0%			
Level-3 (Other)	-24.1%	-0.5%	24.3%			
Level-3 (SACD)	-9.1%	0.1%	9.2%			

Table A.11: Bias and efficiency statistics where group number is imbalanced between album formats. In this case, CD and E Audio have 100 groups while Other and Super Audio have 30 groups. All formats have 30 observations in all level-2 groups.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.3%	0.0%	6.4%	0.98	0.99	1.08
Level-2 (E Audio)	-8.6%	0.0%	8.7%	0.98	0.98	1.08
Level-2 (Other)	-17.9%	-0.2%	17.4%	0.98	0.99	1.07
Level-2 (SACD)	-6.6%	0.0%	6.6%	0.98	0.99	1.07
Holiday Fixed Effect				1.99	2.04	2.60
Level-3 (CD)	-4.4%	0.1%	4.5%			
Level-3 (E Audio)	-6.3%	-0.1%	6.2%			
Level-3 (Other)	-19.7%	0.1%	19.6%			
Level-3 (SACD)	-6.9%	0.1%	7.0%			

Table A.12: Bias and efficiency statistics where group number is imbalanced between album formats. In this case, CD and E Audio have 100 groups while Other and Super Audio have 50 groups. All formats have 30 observations in all level-2 groups.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.2%	0.0%	6.3%	0.98	0.99	1.08
Level-2 (E Audio)	-8.6%	0.1%	8.6%	0.98	0.99	1.09
Level-2 (Other)	-18.4%	-0.2%	18.4%	0.99	1.00	1.08
Level-2 (SACD)	-6.8%	0.0%	6.9%	0.98	0.99	1.08
Holiday Fixed Effect				2.12	2.17	2.72
Level-3 (CD)	-4.3%	0.2%	4.9%			
Level-3 (E Audio)	-6.1%	-0.1%	6.0%			
Level-3 (Other)	-16.1%	-0.1%	15.3%			
Level-3 (SACD)	-5.7%	0.2%	6.1%			

Table A.13: Bias and efficiency statistics where group number is imbalanced between album formats. In this case, CD and E Audio have 100 groups while Other and Super Audio have 70 groups. All formats have 30 observations in all level-2 groups.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.3%	0.0%	6.4%	0.97	0.98	1.07
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.97	0.98	1.07
Level-2 (Other)	-21.0%	-0.3%	20.2%	0.95	0.96	1.08
Level-2 (SACD)	-7.7%	0.0%	7.8%	0.96	0.98	1.10
Holiday Fixed Effect				1.81	1.86	2.49
Level-3 (CD)	-4.3%	0.2%	4.9%			
Level-3 (E Audio)	-6.1%	-0.1%	6.0%			
Level-3 (Other)	-16.1%	-0.1%	15.3%			
Level-3 (SACD)	-5.7%	0.2%	6.1%			

Table A.14: Bias and efficiency statistics where group number and group size are imbalanced between album formats. In this case, CD and E Audio have 100 groups of size 30 while Other and Super Audio have 30 groups of size 20.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.3%	0.0%	6.4%	0.98	0.99	1.08
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.99	1.00	1.09
Level-2 (Other)	-24.3%	-0.3%	23.8%	0.97	1.09	1.15
Level-2 (SACD)	-9.0%	0.0%	9.2%	0.95	0.97	1.14
Holiday Fixed Effect				1.81	1.87	2.51
Level-3 (CD)	-4.3%	0.0%	4.2%			
Level-3 (E Audio)	-6.0%	0.1%	6.4%			
Level-3 (Other)	-24.6%	0.1%	24.9%			
Level-3 (SACD)	-9.7%	0.0%	9.9%			

Table A.15: Bias and efficiency statistics where group number and group size are imbalanced between album formats. In this case, CD and E Audio have 100 groups of size 30 while Other and Super Audio have 30 groups of size 15.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.4%	0.0%	8.6%	0.96	0.97	1.15
Level-2 (E Audio)	-8.7%	0.0%	8.7%	0.97	0.98	1.07
Level-2 (Other)	-21.4%	-0.2%	20.7%	0.96	0.97	1.11
Level-2 (SACD)	-6.4%	0.0%	6.4%	0.98	0.99	1.8
Holiday Fixed Effect						
Level-3 (CD)	-4.6%	0.1%	4.5%			
Level-3 (E Audio)	-6.1%	0.0%	6.1%			
Level-3 (Other)	-23.1%	0.0%	24.4%			
Level-3 (SACD)	-8.7%	0.3%	9.3%			

Table A.16: Bias and efficiency statistics where group number and group size are imbalanced between album formats. In this case, CD and E Audio have 100 groups Other and Super Audio have 30 groups. CD has 25 groups of size 10, 15, 20, 30 each; E Audio has 100 groups of size 30; Other has 15 groups of 15 and 30 each; and SACD has 30 groups of size 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.5%	0.1%	8.6%	0.97	0.98	1.16
Level-2 (E Audio)	-8.6%	0.0%	8.6%	0.98	0.99	1.08
Level-2 (Other)	-22.4%	-0.3%	21.4%	0.96	0.98	1.11
Level-2 (SACD)	-6.6%	0.0%	6.8%	0.98	0.99	1.08
Holiday Fixed Effect						
Level-3 (CD)	-4.3%	0.0%	4.2%			
Level-3 (E Audio)	-6.0%	0.1%	6.4%			
Level-3 (Other)	-24.6%	0.1%	24.9%			
Level-3 (SACD)	-9.7%	0.0%	9.9%			

Table A.17: Bias and efficiency statistics where group number and group size are imbalanced between album formats. In this case, CD and E Audio have 100 groups Other and Super Audio have 50 groups. CD has 25 groups of size 10, 15, 20, 30 each; E Audio has 100 groups of size 30; Other has 25 groups of 15 and 30 each; and SACD has 50 groups of size 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.4%	0.1%	8.6%	0.97	0.99	1.16
Level-2 (E Audio)	-11.6%	0.0%	11.6%	0.97	0.98	1.16
Level-2 (Other)	-22.6%	-0.3%	21.6%	0.97	0.98	1.13
Level-2 (SACD)	-6.7%	0.1%	6.7%	0.98	0.99	1.08
Holiday Fixed Effect						
Level-3 (CD)	-4.5%	0.0%	4.7%			
Level-3 (E Audio)	-6.0%	0.1%	6.5%			
Level-3 (Other)	-18.4%	-0.4%	19.8%			
Level-3 (SACD)	-7.1%	0.2%	7.0%			

Table A.18: Bias and efficiency statistics where group number and group size are imbalanced between album formats. In this case, CD and E Audio have 100 groups Other and Super Audio have 50 groups. CD and E Audio have 25 groups of size 10, 15, 20, 30 each; Other has 25 groups of 15 and 30 each; and SACD has 50 groups of size 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-4.7%	0.0%	4.8%	0.99	1.00	1.06
Level-2 (E Audio)	-6.5%	0.0%	6.5%	0.99	1.00	1.06
Level-2 (Other)	-14.6%	-0.1%	14.4%	0.95	0.97	1.03
Level-2 (SACD)	-5.2%	0.0%	5.3%	0.98	1.00	1.06
Holiday Fixed Effect				2.24	2.29	2.77
Level-3 (CD)	-14.2%	0.1%	14.7%			
Level-3 (E Audio)	-21.2%	0.2%	20.0%			
Level-3 (Other)	-41.2%	0.7%	43.2%			
Level-3 (SACD)	-14.9%	0.0%	16.6%			

Table A.19: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 10 groups with group size of 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.7%	0.0%	5.8%	0.96	0.98	1.07
Level-2 (E Audio)	-7.9%	0.1%	8.2%	0.97	0.99	1.08
Level-2 (Other)	-17.8%	-0.3%	17.1%	0.96	0.97	1.06
Level-2 (SACD)	-6.3%	0.1%	6.4%	0.98	1.00	1.09
Holiday Fixed Effect				2.24	2.32	3.08
Level-3 (CD)	-13.5%	0.2%	14.0%			
Level-3 (E Audio)	-20.8%	-0.3%	19.7%			
Level-3 (Other)	-41.2%	-0.3%	39.4%			
Level-3 (SACD)	-15.6%	-0.5%	14.9%			

Table A.20: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 10 groups with group size of 20.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-7.6%	0.1%	8.0%	0.95	0.98	1.15
Level-2 (E Audio)	-10.6%	0.1%	10.9%	0.95	0.98	1.15
Level-2 (Other)	-24.1%	-0.4%	22.8%	0.97	1.00	1.17
Level-2 (SACD)	-8.6%	0.0%	8.7%	0.93	0.96	1.13
Holiday Fixed Effect				2.24	2.41	4.25
Level-3 (CD)	-14.8%	0.3%	15.5%			
Level-3 (E Audio)	-22.0%	0.5%	20.2%			
Level-3 (Other)	-43.7%	-0.7%	41.8%			
Level-3 (SACD)	-15.8%	0.1%	15.6%			

Table A.21: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 10 groups with group size of 10.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.4%	0.1%	5.6%	0.98	0.99	1.07
Level-2 (E Audio)	-7.6%	0.0%	7.6%	0.97	0.98	1.06
Level-2 (Other)	-16.8%	-0.3%	16.1%	0.98	0.99	1.07
Level-2 (SACD)	-6.0%	0.1%	6.2%	0.98	0.99	1.07
Holiday Fixed Effect				2.24	2.29	2.81
Level-3 (CD)	-9.2%	0.1%	10.1%			
Level-3 (E Audio)	-13.8%	-0.2%	14.8%			
Level-3 (Other)	-31.6%	0.2%	28.4%			
Level-3 (SACD)	-11.3%	-0.1%	11.5%			

Table A.22: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 20 groups with group size of 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.6%	0.0%	6.8%	0.96	0.97	1.09
Level-2 (E Audio)	-9.1%	0.0%	9.2%	0.97	0.98	1.10
Level-2 (Other)	-20.0%	-0.4%	19.5%	0.96	0.98	1.09
Level-2 (SACD)	-7.3%	0.1%	7.4%	0.96	0.98	1.09
Holiday Fixed Effect				2.23	2.31	3.12
Level-3 (CD)	-10.8%	0.0%	9.8%			
Level-3 (E Audio)	-13.6%	0.1%	14.1%			
Level-3 (Other)	-29.5%	-0.1%	32.4%			
Level-3 (SACD)	-11.3%	-0.1%	10.5%			

Table A.23: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 20 groups with group size of 20.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-8.8%	0.0%	9.1%	0.95	0.97	1.18
Level-2 (E Audio)	-12.4%	0.0%	12.4%	0.95	0.97	1.18
Level-2 (Other)	-27.3%	-0.4%	26.1%	0.94	0.97	1.17
Level-2 (SACD)	-9.8%	0.0%	10.0%	0.94	0.97	1.18
Holiday Fixed Effect				2.21	2.37	4.34
Level-3 (CD)	-10.6%	0.2%	10.3%			
Level-3 (E Audio)	-13.8%	0.2%	14.2%			
Level-3 (Other)	-31.4%	0.8%	32.0%			
Level-3 (SACD)	-11.1%	0.2%	11.4%			

Table A.24: Bias and efficiency statistics with small group number and balanced group size. In this case, all album formats have 20 groups with group size of 10.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-4.7%	0.1%	4.8%	0.97	0.98	1.04
Level-2 (E Audio)	-6.6%	0.0%	6.6%	0.98	0.99	1.06
Level-2 (Other)	-17.1%	-0.2%	16.8%	0.98	0.99	1.08
Level-2 (SACD)	-6.4%	0.1%	6.5%	0.98	0.99	1.07
Holiday Fixed Effect				2.46	2.52	3.03
Level-3 (CD)	-14.7%	0.2%	13.3%			
Level-3 (E Audio)	-20.3%	0.4%	19.2%			
Level-3 (Other)	-23.3%	0.2%	24.4%			
Level-3 (SACD)	-8.5%	0.3%	8.9%			

Table A.25: Bias and efficiency statistics with small group number (80), which are imbalanced between format and balanced group size. In this case, CD and E Audio have 10 groups with group size of 30. Other and SACD have 30 groups with group size of 30.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.6%	0.0%	5.7%	0.98	0.99	1.08
Level-2 (E Audio)	-7.9%	0.0%	8.1%	0.95	0.97	1.06
Level-2 (Other)	-20.3%	-0.5%	19.6%	0.96	0.98	1.09
Level-2 (SACD)	-7.2%	0.1%	7.5%	0.95	0.97	1.08
Holiday Fixed Effect				2.40	2.48	3.27
Level-3 (CD)	-14.3%	0.2%	14.0%			
Level-3 (E Audio)	-18.9%	0.3%	20.4%			
Level-3 (Other)	-30.6%	-0.4%	30.5%			
Level-3 (SACD)	-11.2%	-0.1%	10.1%			

Table A.26: Bias and efficiency statistics with small group number (60) and group number (20). Groups size is imbalanced between format and group sizes are balanced. In this case, CD and E Audio have 10 groups with group size of 20. Other and SACD have 20 groups with group size of 20.

A.3 Appendix C: Simulation Results - Outliers

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.7%	0.0%	5.9%	4.54	4.55	4.64
Level-2 (E Audio)	-7.9%	0.0%	7.9%	2.91	2.93	3.02
Level-2 (Other)	-17.1%	0.2%	16.7%	1.43	1.44	1.53
Level-2 (SACD)	-6.4%	0.0%	6.4%	3.94	3.95	4.05
Holiday Fixed Effect				4.70	4.75	5.30
Level-3 (CD)	-4.3%	0.1%	4.6%			
Level-3 (E Audio)	1.0%	9.6%	20.6%			
Level-3 (Other)	0.4%	9.7%	20.4%			
Level-3 (SACD)	0.2%	9.4%	20.2%			

Table A.27: Bias and efficiency statistics from the case where all album formats have 5% level-2 units as outliers.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.3%	0.0%	5.4%	7.86	7.87	7.96
Level-2 (E Audio)	-7.3%	0.0%	7.3%	4.65	4.66	4.75
Level-2 (Other)	-15.7%	0.1%	15.3%	1.81	1.82	1.92
Level-2 (SACD)	-5.8%	0.0%	5.9%	6.65	6.67	6.76
Holiday Fixed Effect				7.03	7.08	7.63
Level-3 (CD)	7.4%	19.8%	32.3%			
Level-3 (E Audio)	7.6%	19.9%	34.5%			
Level-3 (Other)	2.4%	19.7%	38.1%			
Level-3 (SACD)	7.0%	19.4%	33.2%			

Table A.28: Bias and efficiency statistics from the case where all album formats have 10% level-2 units as outliers.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-4.8%	0.0%	5.0%	10.81	10.82	10.91
Level-2 (E Audio)	-6.7%	0.0%	6.7%	6.18	6.19	6.29
Level-2 (Other)	-14.6%	0.1%	14.2%	2.15	2.16	2.25
Level-2 (SACD)	-5.4%	0.0%	5.5%	9.10	9.11	9.21
Holiday Fixed Effect				9.16	9.21	9.76
Level-3 (CD)	15.7%	29.8%	44.8%			
Level-3 (E Audio)	15.5%	29.9%	45.2%			
Level-3 (Other)	9.7%	29.2%	49.2%			
Level-3 (SACD)	14.1%	29.9%	44.7%			

Table A.29: Bias and efficiency statistics from the case where all album formats have 15% level-2 units as outliers.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-4.5%	0.0%	4.6%	13.30	13.31	13.41
Level-2 (E Audio)	-6.2%	0.0%	6.3%	7.52	7.53	7.62
Level-2 (Other)	-13.5%	0.1%	13.2%	2.45	2.46	2.55
Level-2 (SACD)	-5.0%	0.1%	5.1%	11.23	11.24	11.33
Holiday Fixed Effect				11.06	11.11	11.66
Level-3 (CD)	24.6%	39.7%	57.3%			
Level-3 (E Audio)	23.3%	40.0%	57.4%			
Level-3 (Other)	18.2%	39.5%	61.2%			
Level-3 (SACD)	24.1%	39.8%	56.0%			

Table A.30: Bias and efficiency statistics from the case where all album formats have 20% level-2 units as outliers.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.8%	0.0%	5.8%	4.57	4.58	4.68
Level-2 (E Audio)	-8.6%	0.0%	8.7%	0.99	1.00	1.10
Level-2 (Other)	-18.5%	0.2%	18.2%	1.00	1.01	1.10
Level-2 (SACD)	-6.9%	0.0%	7.0%	0.99	1.00	1.10
Holiday Fixed Effect				3.42	3.47	4.01
Level-3 (CD)	1.3%	9.7%	20.6%			
Level-3 (E Audio)	-6.3%	0.1%	6.3%			
Level-3 (Other)	-14.4%	0.0%	14.4%			
Level-3 (SACD)	-4.6%	0.0%	4.7%			

Table A.31: Bias and efficiency statistics from the case where the CD album formats has 5% level-2 units as outliers. All other album formats have no outlying level-2 units.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.3%	0.0%	5.3%	7.89	7.90	7.99
Level-2 (E Audio)	-8.7%	0.0%	8.7%	1.00	1.01	1.11
Level-2 (Other)	-18.6%	-0.2%	18.2%	1.01	1.02	1.11
Level-2 (SACD)	-6.9%	0.0%	7.0%	1.00	1.01	1.10
Holiday Fixed Effect				4.61	4.67	5.21
Level-3 (CD)	8.5%	19.7%	33.9%			
Level-3 (E Audio)	-6.3%	0.1%	6.4%			
Level-3 (Other)	-14.3%	0.0%	13.7%			
Level-3 (SACD)	-4.6%	0.0%	4.8%			

Table A.32: Bias and efficiency statistics from the case where the CD album formats has 10% level-2 units as outliers. All other album formats have no outlying level-2 units.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-4.5%	0.0%	4.6%	13.18	13.20	13.29
Level-2 (E Audio)	-8.7%	0.0%	8.7%	1.01	1.02	1.11
Level-2 (Other)	-18.6%	-0.1%	18.2%	1.01	1.02	1.11
Level-2 (SACD)	-6.9%	0.0%	7.0%	1.01	1.02	1.11
Holiday Fixed Effect				6.85	6.91	7.45
Level-3 (CD)	23.8%	39.0%	56.7%			
Level-3 (E Audio)	-6.3%	0.0%	6.4%			
Level-3 (Other)	-14.2%	0.0%	13.7%			
Level-3 (SACD)	-4.6%	0.0%	4.8%			

Table A.33: Bias and efficiency statistics from the case where the CD album formats has 20% level-2 units as outliers. All other album formats have no outlying level-2 units.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-5.7%	0.0%	5.9%	4.56	4.57	4.67
Level-2 (E Audio)	-7.3%	0.0%	7.3%	4.65	4.66	4.76
Level-2 (Other)	-17.1%	-0.1%	16.7%	1.43	1.44	1.54
Level-2 (SACD)	-5.4%	0.0%	5.4%	9.08	9.09	9.18
Holiday Fixed Effect				6.84	6.90	7.45
Level-3 (CD)	1.0%	9.6%	20.6%			
Level-3 (E Audio)	7.4%	19.9%	34.5%			
Level-3 (Other)	-5.7%	9.6%	26.6%			
Level-3 (SACD)	14.1%	29.9%	44.6%			

Table A.34: Bias and efficiency statistics from the case where album formats have 5%, 10%, 5%, and 15% level-2 units as outliers respectively.

Parameter	Percentage Bias			MSE		
	(Q 2.5%)	(Q 50%)	(Q 97.5%)	(Q 2.5%)	(Q 50%)	(Q 97.5%)
Level-2 (CD)	-6.1%	0.0%	6.3%	1.73	1.74	1.84
Level-2 (E Audio)	-7.3%	0.0%	7.3%	4.61	4.62	4.71
Level-2 (Other)	-17.1%	-0.2%	16.7%	1.42	1.43	1.53
Level-2 (SACD)	-6.8%	0.0%	6.9%	1.62	1.63	1.73
Holiday Fixed Effect				3.56	3.62	4.16
Level-3 (CD)	-3.3%	1.8%	8.6%			
Level-3 (E Audio)	7.5%	19.9%	34.5%			
Level-3 (Other)	-5.7%	9.7%	26.6%			
Level-3 (SACD)	-3.6%	1.9%	8.6%			

Table A.35: Bias and efficiency statistics from the case where album formats have 1%, 10%, 5%, and 1% level-2 units as outliers respectively.

Bibliography

Bakos, Y., Brynjolfsson, E., and Lichtman, D. (1999). Shared information goods. *Journal of Law and Economics*, 42(1):117–156.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

Gelman, A. et al. (2014). *Bayesian Data Analysis*. Chapman & Hall (CRC), Boca Raton FL, 3th edition.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods and Research*, 22(3):364–375.

Goldstein, H. (2011). *Multilevel Statistical Models*. John Wiley & Sons, Chichester NY, 4th edition.

Hilden-Minton, J. (1995). *Multilevel Diagnostics for Mixed and Hierarchical Linear Models*. PhD thesis, University of California: Los Angeles.

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):497–536.

Hong., S.-H. (2007). The recent growth of the internet and changes in household-level demand for entertainment. *Information Economics and Policy*, 19(3-4):304–318.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biomatrics*, pages 963–974.

Loy, A. and Hofmann, H. (2014). HLMdiag: A suite of diagnostics for hierarchical linear models in R. *Journal of Statistical Software*, 56(5):1–28.

Loy, A. M. M. (2013). *Diagnostics for mixed/hierarchical linear models*. PhD thesis, Iowa State University. Paper 13277.

Maas, C. J. and Hox., J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis*, 46(3):427–440.

Maas, C. J. and Hox., J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3):86–92.

Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy*, 115(1):1–42.

Peitz, M. and Waelbroeck., P. (2006). Why the music industry may gain from free downloadingâĂŤthe role of sampling. *International Journal of Industrial Organization*, 24(5):907–913.

Pinheiro, J. C. and Bates., D. M. (2000). *Mixed-effects models in S and S-PLUS.* Springer Science & Business Media.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. Vol. 1.* Sage, Thousand Oaks CA, 2 edition.

Ronald Christensen, L. M. P. and Johnson., W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, 34(1):38–45.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, pages 845–858.

Snijders, T. A. and Berkhof, J. (2008). *Diagnostic checks for multilevel models*, chapter 3th, pages 141–175. Springer-Verlag, New York.

Waldfogel, J. (2010). Music file sharing and sales displacement in the itunes era. *Information Economics and Policy*, 22(4):306–314.

Zewotir, T. and Galpin., J. S. (2005). Influence diagnostics for linear mixed models. *Journal of data science*, 3:153–177.