SPECIAL TOPICS FOR RECOMBINANT INBRED INTERCROSS DATA: MODEL IDENTIFIABILITY, HYPOTHESIS TESTING AND COMPOSITIONAL METHODS

James G. Xenakis

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2019

Approved by:

Fei Zou

Joseph G. Ibrahim

Yun Li

Fernando Pardo-Manuel de Villena

William Valdar

ABSTRACT

James G. Xenakis: Special topics for recombinant inbred intercross data: model identifiability, hypothesis testing and compositional methods
(Under the direction of Fei Zou)

This dissertation addresses statistical issues in studies comprised of mice derived from the Collaborative Cross (CC) project (Churchill, 2004). Briefly, the CC is a community effort to derive novel inbred mouse strains from a genetically diverse set of eight inbred founder strains. Specifically, our interest has been in studying the effects of the parental strains on phenotypes of interest among recombinant intercrosses (RIX) derived from those strains (i.e., pairings of them). The topics explored here involve properly accounting for the relatedness of the samples created in these breeding schemes.

When polygenic effects are conceptualized as uncorrelated parental strain effects, a study design can be framed as a sparse diallel, or more generally dyadic data. In Chapter 2, we consider such designs, incorporating multiple variance components. This raises often-ignored identifiability issues, the most significant of which is the possibility of performing inference on an unidentifiable model parameter - a mistake which is actually not difficult to make in this setting. We develop an easy-to-apply condition to check for model identifiability in this setting.

In Chapter 3, our focus is on hypothesis testing in these same sparse diallels. Because variance parameters are boundary parameters, inference is considered a "non-standard" problem, with asymptotic reference distributions being (sometimes complicated) mixtures of χ^2 random variables, when they are available at all. This is further complicated by the fact that when the dependent variable is non-normal, the data structure in this complicated setting does not allow for likelihood-based methods, as integration over the random effects becomes computationally infeasible. We adapt an existing score statistic developed by Lin (1997), which is designed for models fit by penalized quasi-likelihood.

In Chapter 4, we directly model the relatedness of the strains by replacing the diallel-like random design matrices with ones derived from similarity matrices. Further, we incorporate the contributions

of the eight founder strains as fixed effects, and propose a framework incorporating the statistic employed in Chapter 3 to jointly test the fixed and random genetic effects. We show that the benefits of this approach include improved power and more meaningful interpretations of parameter estimates.

For Matt Fritsch - the best friend I will ever know,
and
for Mom and Dad.

ACKNOWLEDGEMENTS

I have truly loved my time at UNC, and consider myself fortunate to have accumulated a long list of people who have contributed so much to my experience here over the years.

First and foremost, thank you to my advisor, Dr. Fei Zou, for her constant encouragement and perspective. From North Carolina to Florida and back again, you made statistics look easy. Thanks also to Dr. Ibrahim and the National Cancer Institute for funding me throughout most of the PhD program through the Biostatistics for Research in Genomics and Cancer Training Grant (NCI grant 5T32CA106209-07). It seems like just yesterday you were encouraging me to come here.

Without the love and support of my family, I could never have finished my PhD. I couldn't have even started. Thank you, Mom and Dad - somehow I rely on you now more than ever.

Thank you to Rachel, because even 700 miles apart I don't feel closer to anyone.

Thanks to Uncle Bill, Aunt Fran, Aunt Barbi, Baris and Seval, for taking such good care of me.

And to new friends: Matt Psioda, Jonathan O'Brien, Dave Pritchard, and Sebastian Hidalgo, for getting me through the hard times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADR	Adverse Drug Reaction
ANOVA	Analysis of Variance
CC	Collaborative Cross
CISGen	Center for Integrated Systems Genomics
DNA	Deoxyribonucleic Acid
GCA	General Combining Ability
GCTA	Genome-Wide Complex Trait Analysis
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GRM	Genetic Relationship Matrix
IBD	Identity by Descent
IBS	Identity by State
LM	Linear Model
LMM	Linear Mixed Model
LRT	Likelihood Ratio Test
ML	Maximum Likelihood
OWL	One Way Loop
PQL	Penalized Quasi-Likelihood
REML	Restricted (Residual) Maximum Likelihood
rGCA	Reciprocal General Combining Ability
rSCA	Reciprocal Specific Combining Ability
SCA	Specific Combining Ability
SKAT	Sequence Kernel Association Test
SRR	Single Round Robin
PoO	Parent-of-Origin
RI	Recombinant Inbred (Strain)
RIX	Recombinant Inbred Intercross (Line)
SCZ	Schizophrenia

UNC	University of North Carolina
VCM	Vacuous Chewing Movements

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

Psychiatric disorders, including schizophrenia (SCZ), are often devastating neuropsychiatric illnesses. According to the World Health Organization, SCZ alone affects over 20 million people worldwide, with onset typically in adolescence or early adulthood (WHO, 2012). Affected subjects suffer from myriad symptoms, including distortions in cognition like hallucinations and delusions, as well as excess mortality. Antipsychotic medications like haloperidol are the current mainstays of treatment for SCZ but are associated with severe and often intolerable side effects that reduce compliance and thus also badly constrain their effectiveness. It has been estimated that 75% of patients discontinue treatment over relatively short periods of time, and that adverse drug reactions (ADR's) play a major role (Lieberman et al., 2005). The enormity of disorders like SCZ is made worse by the fact that they are so difficult to study, in part due to the inaccessibility of tissue (Weis et al., 2007; Tomita et al., 2004; Arnold et al., 2001; Smutzer et al., 1998; Harrison, 1999). While there is substantial inter-individual variation in susceptibility to ADR's and much evidence implicating genetic variation (Bakker et al., 2006; Lerer et al., 2005; Patsopoulos et al., 2005; Reynolds et al., 2005), researchers have thus far not been able to produce effective algorithms to predict them. While researchers are limited in the capacity to study the genetic underpinnings of ADR's in humans, mouse models can fill this gap. Typically, such studies involve crosses of two existing recombinant inbred (RI) strains, but such studies are limited in their ability to detect genetic effects and their possible interactions with environment that underlie complex phenotypes. For one, the limited number of existing RI strains affords insufficient power to detect the modest effect sizes that are common in such phenotypes. Further, crossing two inbred lines yields a low level of genetic variation, which is also distributed non-uniformly. These problems are addressed by the novel Collaborative Cross (CC) genetic reference panel, which serves as a superior mouse model for the more heterogeneous human population (Rogala et al., 2014; Kantor et al., 2011). The CC was derived from a genetically diverse

set of 8 founder inbred strains (Rogala et al., 2014; Kantor et al., 2011; Follenzi et al., 2004), the selection of which was designed to maximize the levels of genome-wide variation, and capitalize on the fact that no genomic regions are identical in all these founder strains. The genetic variation is randomized in the CC strains so that causal relationships can be established. Additionally, each CC RI genome is a mosaic of the eight founder strains such that the combination of alleles present in these new strains are not present in any existing mouse strain (Rogala et al., 2014). The CC is the only mammalian resource that has high and uniform genome-wide variation effectively randomized across a large, heterogeneous, and infinitely reproducible population.

The pressing need to study the genetic underpinnings of important pharmacogenomic phenotypes induced by haloperidol motivated a project at UNC, the data from which we employ in what follows. The samples comprising the study were recombinant inbred intercrosses (RIX) generated from crossing the inbred RI strains. Such crosses maintain the benefits described above (e.g., genome variation and infinite reproducibility) and confer additional advantages due to the fact that they are not inbred and are therefore more representative of the human genome. In the study, a series of behavior and pharmacogenetic phenotypes were collected from a panel of RIX treated with or without haloperidol. This behavior study can be framed as a sparse diallel, for which the mixed model has recently become an increasingly useful tool for analysis, where the parental effects are assumed to arise from a normal distribution. Although sparse diallels are known to have high power for mapping studies, the details regarding parameter estimation and testing, and effect prediction are complicated. In this context, some relevant hypotheses pertain to variances of the random effects. Such testing problems are considered "non-standard" problems in statistics, because under the null the parameters of interest lie on the boundary of the parameter space (Self and Liang, 1987). This is further complicated by the possible presence of nuisance boundary parameters. Further, because of the connected nature of the random effects in the diallel design, the response vector cannot be subdivided into independent units; this has been shown to affect the asymptotic distribution of the estimates for variance components. In addition, this makes fitting the model difficult when the response is non-normal. Further, whereas the theory is thin for linear mixed models (LMM), it is altogether lacking for generalized linear mixed models (GLMM). We formally lay out the estimability theory and testing framework.

There are several ways of modeling the parental effects. The simplest is with a sort of pedigree; pairs of mice which share a common parent are assumed to be equally related. While it was originally expected that each founder strain would contribute 12.5% of the genome of each recombinant strain, recent work has shown this not to be the case. The founder strains are not equally represented, and some founder strains are (surprisingly) not represented at all in some of the RI strains (see figure 4.18 in Chapter 4). Because of this phenomenon, it is of interest to extend the simple pedigree model to incorporate the genetic similarity of the CC lines. Such models have been used to effectively predict phenotypes based on genomic similarity (Kang et al., 2008).

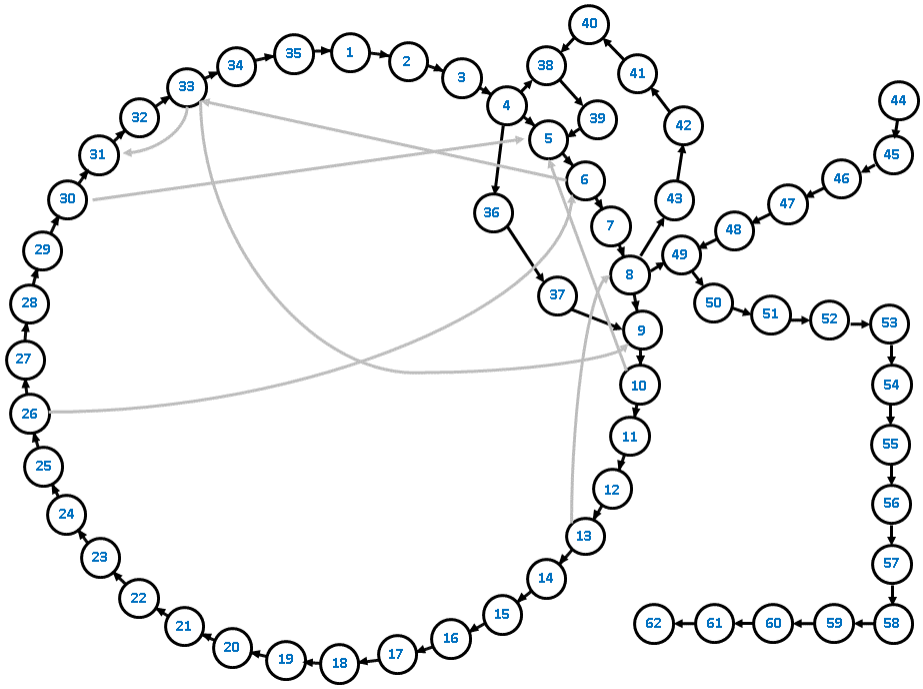
For complex polygenic traits, RI panels like the CC which are derived from a limited number of founder strains afford the opportunity to test the association between amounts of founder DNA present in each line with phenotypes of interest. This is a good starting point in the study of any trait, especially since some founder strains were specifically bred to express certain phenotypes. There is a subtle issue with testing such associations because of a constraint present in the data: the sum of the genetic contributions of the founder strains is, by definition, one. Such data are called *compositional*, and if care is not taken to address the constraint, the consequences can be uninterpretable hypotheses or spurious conclusions (Pawlowsky-Glahn and Buccianti, 2011). While methods for compositional covariates have been applied effectively in some fields (Lin et al., 2014; Li, 2015), these methods must be extended to accommodate compositions that have zero elements, which, as mentioned above, is an unexpected feature of the CC RI straines. In the future, we plan to develop models that properly handle the compositional aspect of RI data, allowing for meaningful and interpretable hypothesis testing.

1.2 Classical Diallel

We noted above that our study design can be framed as a sparse partial diallel. The data structure is shown below in Figure 1.1. Each numbered node corresponds to a unique CC inbred strain, and the arrows to recombinant crosses of those strains - that is, the samples in the dataset are composed of these crosses. The arrow points in the direction of maternal parent.

The salient feature of our study design is the large one-way “loop”, which is similar to a well-known partial diallel known as the *single round robin* (SRR). There are additional “satellite loops”

Figure 1.1: Study design



and "chains", but for now we will focus on the loop. The diallel has had a long and complicated history in genetics, beginning nearly a centrury ago (Schmidt, 1919). Contributions have come from sub-fields of applied genetics, where the experimental designs and associated analysis methods have differed based on the science relevant to the specific subfield. Although all such models can be described in a general framework, it was several decades until this was attempted. For this reason, a linear history of the diallel and all its variants would only serve to confuse. We therefore start with the general treamtent, best formalized by Hayman and Griffing (Hayman, 1954b,a, 1958, 1960; Griffing, 1956a,b) which was published more than three decades after the publication of what can be considered the first diallel experiment (Schmidt, 1919). We then circle back to provide a more linear review.

Griffing defined a diallel cross as one in which a set of p inbred lines are chosen and crosses (though not necessarily *all* crosses) are made among these lines (Griffing, 1956a,b). The structure of the data can be easily visualized in a $p \times p$ table (see, for example, Figures 1.2-1.5 below), and the p^2 possible combinations can be divided into three separate groups:

1. The p inbred parental lines themselves, corresponding to the main diagonal in the table.

2. One set of F_1 hybrids, corresponding to, say, the upper right triangle of the table, denoting offspring whose maternal parent is indexed by row label and paternal parent by column label.

3. The other set of F_1 hybrids, corresponding to the lower left triangle of the table. These are denoted *reciprocal* F_1's.

According to Griffing there are then four types of diallel experiments, depending on whether the inbred lines are to be included:

1. All p^2 combinations - that is, p inbred parents and both sets of F_1's.

2. The $\frac{p(p+1)}{2}$ crosses corresponding to the inbred parents and one set of F_1's.

3. The $p(p-1)$ combinations corresponding to all F_1's, including reciprocals, but no inbred parents.

4. The $\frac{p(p-1)}{2}$ crosses corresponding to a single group of F_1's (reciprocal and parental lines excluded).

Griffing further distinguished between the following two scenarios:

1. The parental lines are assumed to be a random sample from a population about which inferences are to be made.

2. The parental lines are deliberately chosen, and are therefore not a random sample from any population.

We note that these two scenarios were, respectively, termed Model I and Model II by Eisenhart (Eisenhart, 1947), another seminal author on the subject. This yields a total of eight scenarios to be considered (the effects in each of the four types of experiment can either be considered fixed or random). It is with these scenarios in mind that we proceed with the history of the diallel. The purpose of a diallel experiment is twofold. First, the design can be used to study the heritability of traits of interest (that is, to estimate the genetic components of yield variation among the crosses). Secondly, it can be used to estimate the actual breeding values of the parental lines included (Kempthorne and Curnow, 1961). Although Schmidt published the first diallel in 1919 (Schmidt, 1919), it can be argued that Sprague and Tatum (Sprague and Tatum, 1942) pioneered the theory of the diallel when they coined the following terms:

1. *General combining ability* (GCA): the average performance of a line in hybrid combination.

2. *Specific combining ability* (SCA): designates cases in which certain combinations do better or worse than would be expected based on the average performance of the lines involved.

We note that Sprague and Tatum's study design was of Griffing's Type 4: $\frac{p(p-1)}{2}$ F_1's with no parents, and so these terms require some clarification in order to generalize to all four study designs. We do this in the context of Griffing's model for crossing methods that incorporate reciprocal crosses (Griffing's types 1 and 3). The model for phenotypic response is:

$$y_{ijk} = g_j + g_k + s_{jk} + r_{jk} + \epsilon_{ijk} \tag{1.1}$$

Where y_{ijk} corresponds to the response of sample i with parents j and k. Here, g_j corresponds to the GCA of the jth parent, s_{jk} is the SCA effect for the cross between the jth and kth parents, and r_{jk} is the reciprocal effect involving the jth and kth parents. The variety effects for crossing methods excluding reciprocals (Griffing's type 2 and 4 above), become:

$$y_{ijk} = g_j + g_k + s_{jk} + \epsilon_{ijk} \tag{1.2}$$

1.3 Partial Diallel

Although most diallel literature focuses on the classical models described above, there is also a rich literature on partial diallel design (Kempthorne and Curnow, 1961; Greenberg et al., 2010; Viana et al., 1999; Singh and Hinkelmann, 1995, 1998, 1990; Mukerjee, 1997; Hinkelmann and Kempthorne, 1963; Gupta et al., 1995; Ghosh and Divecha, 1997; Fyfe and Gilbert, 1963; Das et al., 1998; Arya, 1983; Verhoeven et al., 2006; Clatworthy, 1955). This literature is of particular relevance to us because, as mentioned above, our study design can be framed as a partial diallel. Further, the "loop" portion of the design is actually very similar to a common partial diallel known as the *single round robin (SRR)*. For this reason, we take special care to review several papers that studied particular partial diallels known as *circulant designs*, of which the loop design is one. As mentioned above, the purpose of the diallel is twofold: to estimate genetic components of variation and to estimate/predict actual yielding capacities of crosses. Note that the number of potential crosses to

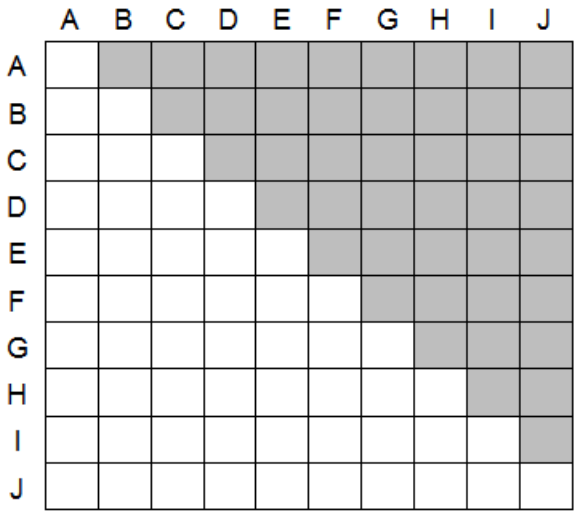
be made increases quadratically in the number of inbred parental lines, p. It is seldom feasible then to perform a complete diallel as the number of parental lines under study becomes moderate/large. For a given amount of resources (e.g., the total number of possible samples, which might be limited for economic reasons) it is of interest whether a design should involve more inbred lines with fewer crosses sampled (i.e., a more sparse design) or to construct more crosses from a smaller number of inbred lines. In general, Kempthorne and Curnow (Kempthorne and Curnow, 1961) showed that, for the estimation of variance components of interest, sparser designs can be more efficient, although there are some potential pitfalls discussed, for example, in Fyfe and Gilbert (Fyfe and Gilbert, 1963).

Kempthorne and Curnow (Kempthorne and Curnow, 1961) provide perhaps the seminal work on circulant designs, borrowing heavily from the work of Clatworthy (Clatworthy, 1955), and framing the circulant design as a *connected partially balanced incomplete block design with two associate classes and two treatments per block*. Quoting Clatworthy (with slight changes in notation), such a design is an arrangement of p treatments (parents in our context) into c blocks (crosses in our context) such that:

1. Each of the treatments occurs r times, and no treatment occurs more than once in any block.
2. Every pair among the p treatments occurs together in either λ_0 or λ_1 blocks (and are said to be ith associates if they occur together in λ_i blocks, $i = 0, 1$).
3. There exists a relationship of association between every pair of the p treatments satisfying the following conditions:
   (a) Any two treatments are either first or second associates.
   (b) Each treatment has n_1 first and n_2 second associates.
   (c) Given any two treatments that are ith associates, the number of treatments common to the jth associates of the first and the kth associates of the second is $p^i_{jk} = p^i_{kj}$ $(i, j, k = 1, 2)$ which is independent of the pair of treatments with which we start.
4. A design is said to be *connected* if for every two treatments, two blocks, or a block and a treatment, it is possible to pass from one to the other by means of a chain consisting alternately of blocks and treatments, such that every treatment of the chain occurs in each of the adjoining blocks.

Some examples of these circulant designs are shown below in Figures 1.2-1.5. These have been recurrent designs in the comparative study of diallels (Verhoeven et al., 2006). Clearly, for the loop design/SSR, $\lambda_0 = 0$ and $\lambda_1 = 1$. Most of the theoretical work on these circulant designs predates the modern mixed model, which we will study. We note that some of our results relate closely to the nonsingularity condition noted by Curnow (Curnow, 1963) and again by Ole Braaten in his excellent PhD thesis (Braaten, 1965). That is, special care should be taken in designing the circulant study to insure that the least squares matrices are not singular. This is closely related to some of our results that follow in subsequent chapters.

Figure 1.2: Half diallel without inbreds



Most of the work on fitting the diallel models (and variants thereof) discussed so far centered around the operationalizing of the basic ANOVA models first laid out by Fisher (Fisher, 1919). That is, constructing the relevant ANOVA tables and comparing observed to expected sums of squares. The shortcomings of this approach are well known and discussed further below.

Figure 1.3: Single round robin

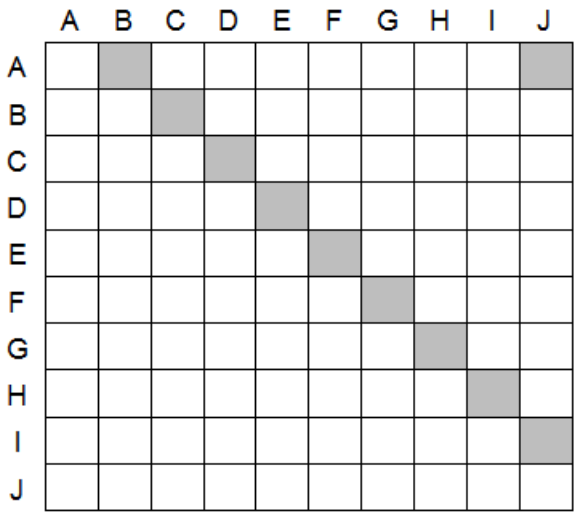


Figure 1.4: Double round robin

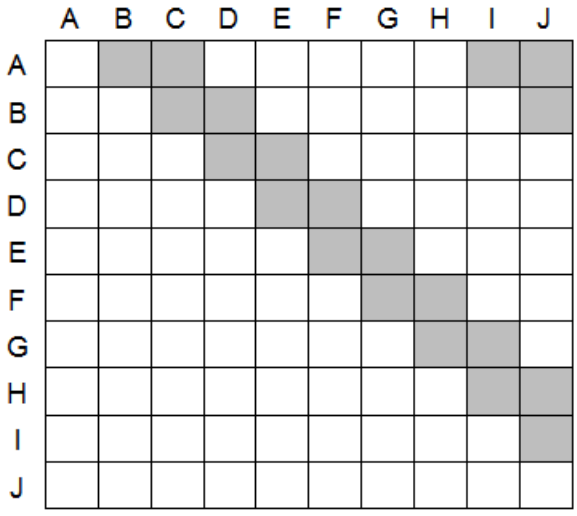
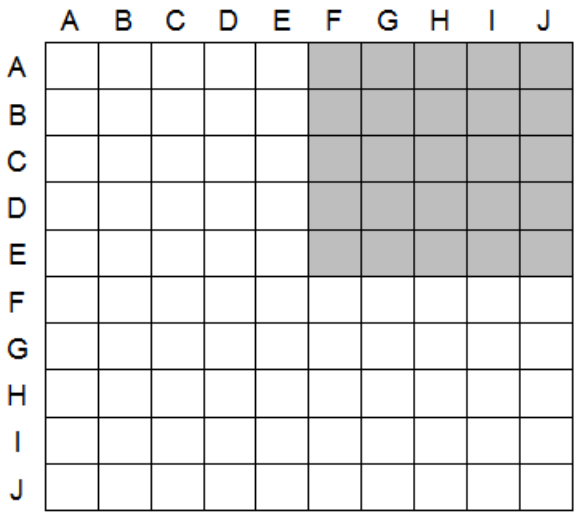


Figure 1.5: Factorial design



1.4 Modern Diallel

As the tools for analysis of the diallel became more sophisticated, so too did the ways of dissecting the genetic effects in different ways. We use the term "modern diallel" to refer to the post-ANOVA age, which we might say began with the work of Zhu and Weir (Zhu and Weir, 1996a), although they built heavily on the work of Cockerham and Weir (Cockerham and Weir, 1977), who start with a genetic experiment where the phenotypic performance is written in the most general form:

$$Y_{ijk} = \mu + G_{ij} + \epsilon_{ijk} \tag{1.3}$$

where Y_{ijk} corresponds to the response of the kth sample of a cross between a maternal strain i and paternal strain j. The authors then take into account the factorial nature of the pairwise matings by dissecting G_{ij} as follows:

$$G_{ij} = M_i + P_i + (MP)_{ij} \tag{1.4}$$

where M and P denote the maternal and paternal effects, respectively, and (MP) to the interaction thereof. Assuming these effects to be random, this would correspond to a variance components model with three mean zero variance components with variances σ^2_M, σ^2_P, and σ^2_{MP}, respectively. If the design includes reciprocals, the dissection of the genetic effect can be made even more granular, as follows:

$$G_{ij} = g_i + g_j + s_{ij} + d_i - d_j + r_{ij} \tag{1.5}$$

where g_i corresponds to the average GCA associated with strain i, and d_i to a symmetric deviation from this average associated with having a parent of strain i. Similarly, s_{ij} corresponds to the average interaction of having parental pair i, j, and r_{ij} to a symmetric deviation to account for the direction of the cross. That is:

$$g_i = \frac{(M_i + P_i)}{2} \tag{1.6}$$

$$d_i = \frac{(M_i - P_i)}{2} \tag{1.7}$$

$$s_{ij} = \frac{(MP_{ij} + MP_{ji})}{2} \tag{1.8}$$

$$r_{ij} = \frac{(MP_{ij} - MP_{ji})}{2} \tag{1.9}$$

where $s_{ij} = s_{ji}$ and $r_{ij} = -r_{ji}$. The authors further expand this into what they call a *bio model* which separates nuclear parental contribution from extranuclear maternal and paternal effects, as follows:

$$G_{ij} = n_i + n_j + t_{ij} + m_i + p_j + k_{ij} \tag{1.10}$$

where n_i is the nuclear contribution of parent i and t_{ij} the associated interaction effect, m_i and p_i are the maternal and paternal extranuclear effects, respectively, and the k_{ij} are the various interaction terms (Cockerham and Weir, 1977). This model was further expanded by Zhu and Weir to include genotype-by-environment interactions for experiments conducted in multiple environments (Zhu and Weir, 1996b). Zhu and Weir write their model in the most general form as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{u=1}^{q+1} \boldsymbol{U}_u \boldsymbol{e}_u \tag{1.11}$$

with covariance matrix:

$$\sum_{u=1}^{q+1} \sigma_u^2 \boldsymbol{U}_u \boldsymbol{U}_u^{'} \tag{1.12}$$

We certainly do not mean for this to be an exhaustive literature review of the history of and methods for the diallel; that would require an entire volume. The purpose was to arrive at the modern age of diallel analysis. Zhu and Weir (Zhu and Weir, 1996b) note some deficiencies in the ANOVA methods that had dominated diallel analysis, including the handling of unbalanced data and the inability to give separate estimates for variance components of maternal and paternal effects. They fit their model by MINQUE(1) (Rao, 1971), which has fallen out of favor in deference to restricted maximum likelihood (REML) as computing methods have improved, but we use their general framework throughout the remainder of this work. We note that some authors have pursued Bayesian methods for diallel analysis (Greenberg et al., 2010; Lenarcic et al., 2012), but as our ultimate goal is a model that can be used on the genome-wide scale, we do not pursue such approaches further.

1.5 Variance Components

Although much of the theoretical and applied work on random effects models has been driven by genetics applications, the earliest known formulation of such models dates back to astronomical work in the mid-1800's (Airy, 1861), wherein Airy studied the problem of making repeated observations

by telescope on celestial bodies on multiple nights. One interesting aspect of his formulation was that it accommodated unbalanced data (the number of observations need not be the same from night to night). This is in direct contrast with the treatment of variance component models for the following century which relied on balanced designs; the unbalanced design would not receive its due treatment until Henderson's work in the 1950's. For the better part of the twentieth century, variance component estimation was dominated by the analysis of variance (ANOVA) methods developed by R.A. Fisher, beginning with his seminal paper in the area of quantitative genetics (Fisher, 1919). Briefly, these methods rely on equating the sums of squares from an ANOVA to their expected values, yielding a system of equations which is linear in the variance components. There are two problems with this estimation approach. For one, nothing prevents the estimates of the variance components from being negative. While some suggested truncating negative estimates at zero, this immediately compromises the unbiasedness of the estimators. When fitting models with multiple variance components and the targets of interest are mean parameters it is often argued that negative variance components are perfectly acceptable, so long as the overall covariance matrix is valid (i.e., positive definite). This is not helpful when the variance components are themselves of interest, as is often the case in genetics applications. This brings us to the second problem with the ANOVA approach to variance component estimation, which is the lack of satisfactory distributional results for testing their significance. This problem was partially addressed in the 1940's by Sattherthwaite, who derived approximate sampling distributions for the variance components (Satterthwaite, 1946) and Wald (Wald, 1940, 1941), who derived confidence intervals for the ratios of variance components in the 1- and 2- way classifications, accommodating unbalanced data.

A groundbreaking work on variance component estimation with unbalanced designs came in Henderson's now famous 1953 paper (Henderson, 1953) - methodology which is still commonly used today. It is interesting from our perspective because his work bears some similarity to our own application. While we are studying how mice respond to stimuli, he was studying the milk production of dairy cows. Henderson's methods are best viewed as an effective and comprehensive generalization of the ANOVA methods. Henderson's paper consists of three methods, still very much in use today. Method I applies to models with random effects only, generalizing the sums of squares ANOVA approach to handle unbalanced data. Method II generalizes Method I, but adjusts for fixed effects in mixed models. Model III is more complicated, but was of limited use at the time due to

computational burden. A useful feature of Henderson's approach is that, in addition to imbalance, it accommodated any number and type of random effects (i.e., nested or crossed).

As useful as Henderson's methods are for estimation, being generalizations of the ANOVA method, the problems of negative estimates and the lack of a cohesive hypothesis testing framework persisted. These were not adequately addressed until the popularization of Maximum Likelihood (ML) methods and their variants in the 1960's. In contrast to ANOVA methods, which require no distributional assumptions on the random effects (they require little more than a finiteness assumption, and, in a sense, a separability of the fixed and random effects). The key paper to popularize ML estimation in random effects models was by Hartley and J.N.K. Rao (Hartley and Rao, 1967). It is cohesive in the way that Henderson's approach was, in that it accommodates any number of random effects, and lack of balance. It also cohesively solves the problem of negative variance estimates, as these are (typically) forced to be positive in the maximization routine (more on this later, because it does actually cause problems for testing). One disadvantage (though less so today compared to when the method first came out) is the computational expense of ML - it requires iteratively solving the ML equations (Miller, 1973, 1977). Searle worked on the asymptotic distribution in the general case as well (Searle, 1970). Note also that in the balanced case, the ML estimates coincide with the ANOVA estimates.

Extending ML estimation, Thompson (Thompson Jr, 1962) first introduced the idea of maximizing the portion of the likelihood which is invariant to the location parameters (fixed effects). This method is known as restricted (or residual) maximum likelihood (REML).

1.6 Hypothesis Testing

As we plan to fit models with many variance components, some of which might be null, a clarification of some important work in this area is in order. The seminal result on the distribution of the likelihood ratio test is due to Wilks (Wilks, 1938). Consider a p-dimensional vector θ, and a hypothesis that θ lies on an r-dimensional hyperplane within p-space. Under the null, the distribution of $-2\log\lambda$ is asymptotically χ^2 with $p-r$ degrees of freedom. Chernoff (Chernoff, 1954) built on Wilks' work, and his framework goes a long way to understanding the LRT from a geometric

perspective. Consider testing $H_0 : \theta \in \omega$ versus $H_A : \theta \in \tau$ Define:

$$P_\psi = \sup_{\theta \in \psi} L(X, \theta) \tag{1.13}$$

Then the likelihood ratio can be expressed as:

$$\lambda(X) = \frac{P_\omega(X)}{P_\tau(X)} \tag{1.14}$$

Chernoff first considers data following a multivariate normal distribution with mean θ and known covariance matrix Σ. By the sufficiency of the sample mean, we need only consider the case where the sample size is one, in which case:

$$P_\omega(x) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{Q_\omega(x)}{2}} \tag{1.15}$$

where $Q_\omega(x) = \inf_{\theta \in \omega}(x - \theta)' \Sigma^{-1}(x - \theta)$. The likelihood ratio test (LRT) is then:

$$-2 \log \lambda(x) = Q_\omega(x) - Q_\tau(x) \tag{1.16}$$

Note that (1.16) lends itself well to geometric interpretation: $Q_b(a)$ is the (generalized) distance from a to b. The LRT statistic is therefore easily seen as the difference between two distances. Chernoff's main contribution is then the following theorem:

Theorem 1.1. *Let Z be a random variable with a multivariate Gaussian distribution with mean θ and covariance matrix $I^{-1}(\theta_0)$, and let C_{Ω_0} and C_{Ω_1} be non-empty cones approximating Ω_0 and Ω_1 at θ_0, respectively. Then under sufficient regularity conditions, the asymptotic distribution of the likelihood ratio test statistic, $-2 \log \lambda_N$, is the same as the distribution of the likelihood ratio test of $\theta \in C_{\Omega_0}$ versus the alternative $\theta \in C_{\Omega_1}$ based on a single realization of Z when $\theta = \theta_0$.*

Self and Liang (Self and Liang, 1987) build upon the work of Chernoff in such a way so as to facilitate geometric interpretability and ease of computation, by introducing a spectral decomposition of the Fisher information: $I(\theta_0) = P \Lambda P'$. Thus, the LRT can be written as:

$$\inf_{\theta \in \tilde{C}_0} ||\tilde{Z} - \theta||^2 - \inf_{\theta \in \tilde{C}} ||\tilde{Z} - \theta||^2 \tag{1.17}$$

where $\tilde{C} = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{1/2} P' \theta \ \forall \, \theta \in C_\Omega - \theta_0\}$ and $\tilde{C}_0$ is defined analagously. This transformation allows us to work with $\tilde{Z}$, which is normal with mean zero and identity covariance; use of the Euclidean norm in equation (14) facilitates geometric interpretation. Further, this facilitates the computation of the probabilities that follow, as everything has been recast in terms of independent standard normal random variables.

We begin with a scenario that is not explicitly dealt with in Self and Liang (Self and Liang, 1987): testing interior parameters in the presence of nuisance boundary parameters. Formally, we consider a parameter vector $\theta = (\theta_1, \theta_2)$ where θ_1 corresponds to interior parameters of interest, and θ_2 to nuisance boundary parameters. Figure 1.6 illustrates the case where these are each one-dimensional, and after applying the spectral transformation above, $\Lambda^{0.5} P \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\Lambda^{0.5} P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ remain orthogonal. Note that the support of $\Lambda^{0.5} P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (plotted on the abscissa) is $\mathbb{R}$ as it is an interior parameter, and that of $\Lambda^{0.5} P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (plotted on the ordinate axis) is $\mathbb{R}^+$ as it is a boundary parameter. Applications of this scenario abound. For example, a salient feature of the linear mixed model is the asymptotic independence of the estimates for the mean and the variance parameters, so this scenario would apply to testing a mean parameter in the linear mixed model in the presence of null variance components. In this scenario, the null parameter space corresponds to the half-line $\Lambda^{0.5} P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and the alternative to the entire upper half plane. There are two regions of interest. If Z falls in the upper half-plane (region 1), under the alternative it attains the true parameter value, and thus the Euclidian distance is zero. Under the null, however, the closest point in the null parameter space is the projection onto the half-line. The LRT statistic is the difference between these two distances (visualized in red as d_1), which is simply the square of the first element of Z: a χ^2_1 random variable. If Z falls in region 2, the distance under the null corresponds to the line from the origin, and under the alternative it is the line to the projection onto the abscissa. Again, this difference (visualized in red and labeled d_2) is simply the χ^2_1 variable Z_1.

Figure 1.6: Testing a single interior parameter in the presence of a single independent nuisance boundary parameter

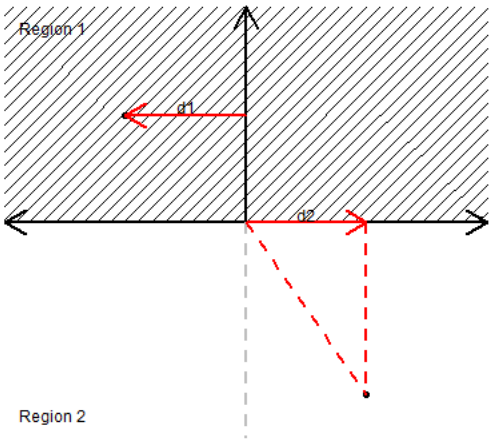
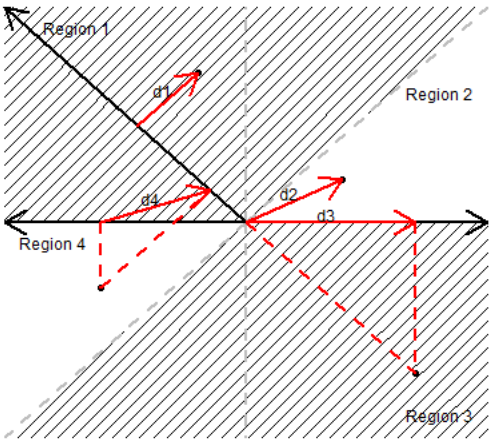


Figure 1.7: Testing a single interior parameter in the presence of a single correlated nuisance boundary parameter



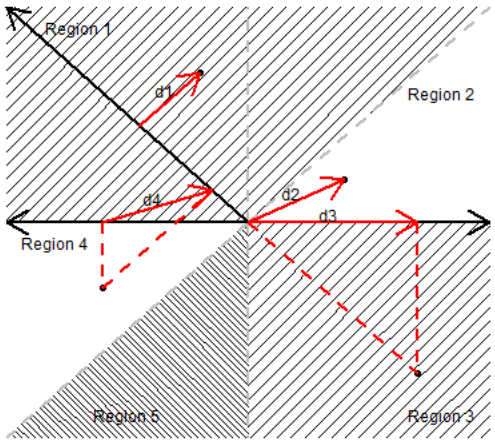
The situation is more complicated when we cannot rely on the independence of the estimates. Visually, this corresponds to the angle between $\Lambda^{0.5}P\begin{bmatrix}1\\0\end{bmatrix}$ and $\Lambda^{0.5}P\begin{bmatrix}0\\1\end{bmatrix}$ no longer being $\frac{\pi}{2}$

(Figure 1.7). The geometry is more complicated, but following the logic above, there are four relevant regions, in which the LRT statistics are labeled $d_1, \ldots, d_4$. Note that the only regions in which the distribution is tractable are regions 2 and 3, where the LRT statistic follows a χ^2_2 and a χ^2_1 distribution, repectively. The distributions in regions 1 and 4 are not necessarily easy to compute. Although this situation might seem somewhat contrived at first glance, it is actually common. Consider testing a fixed effect in the generalized linear mixed model where the covariance is "overfit", in the sense that one or more parameters are truly zero. The standard asymptotic results do not obtain here. This also corresponds precisely to the scenario where we want to use a GLMM with random effects that control for global family effects in CC mice for mapping.

Finally, we consider the case where we are interested in testing a boundary parameter in the presence of another nuisance boundary parameter. This corrresponds to case 8 in Self and Liang (Self and Liang, 1987), which is associated with their Figure 7. Both parameters being on the boundary under the null, the null parameter space is, again, the half-line $\Lambda^{0.5} P \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and the aternative is the entire upper half plane. This situation is plotted in Figure 1.8, which only differs from Figure 1.8 in that region 4 has been broken up into two separate regions. If Z falls in region 5, the closest point under both the null and the alternative hypotheses is the origin, and thus in this region, the LRT statistic is zero (this is where the ubiquitous "point mass on zero" originates).

In recent years, several authors have revisited the boundary problem from a theoretical perspective (Sinha et al., 2007; Kopylev and Sinha, 2011; Vu et al., 1997; Kopylev, 2012; Li and Cui, 2016; Han and Chang, 2009; Drton, 2009). Although much of this work of is quite general, the bulk of the cited papers have been, either explicitly or implicitly, in the context of longitudinal data analysis. See, for example, the oft-cited Stram and Lee (Stram and Lee, 1994). While this paper clearly expands on the work of Self and Liang and derives some important distributional results, it is not sufficiently general to accommodate broader classes of mixed models (explained in more detail below). Still, these two seminal papers are the canonical works when it comes to distributional results for testing the significance of variance components in mixed models. While the asymptotics governing interior parameters (e.g., mean parameters in linear regression) are fairly well-behaved, this is not the case for boundary parameters. These tests are known to be conservative in small samples for the LMM. Further, Crainiceanu and Ruppert (Crainiceanu and Ruppert, 2004) showed that the sample sizes

Figure 1.8: Testing a single boundary parameter in the presence of a single nuisance boundary parameter



need to be quite large for the asymptotics to obtain. Further, they depend on subtle assumptions which are often not met or ignored. They derived the finite sample and asymptotic null distributions of the LRT and restricted LRT using efficient simulation algorithms based on spectral decompositions for testing a single variance component. This work was extended by Greven et al to accommodate multiple nuisance random effects (Greven et al., 2008).

We have focused thus far on the LRT, but this is by no means the only testing framework (only the most common); it has been shown that the score and Wald tests follow the same asymptotic distributions as the LRT (Silvapulle, 1992; Verbeke and Molenberghs, 2003; Silvapulle and Silvapulle, 1995). Ofversten developed a method for deriving exact F tests for variance components for certain unbalanced LMM's by means of orthogonal tranformations and a resampling procedure (Ofversten, 1993).

Several Bayesian variable selection methods have also been developed for random effects. Kinney and Dunson (Kinney and Dunson, 2007) employed a reparametrization of the LMM based on a modified Cholesky decomposition followed by a stochastic search variable selection approach. Saville and Herring (Saville and Herring, 2009) compared the null and alternative models using Bayes factors. While these methods are interesting, we do not explore them further here due to

the prohibitive computational cost in genomics applications. Another interesting, but similarly computationally expensive approach is the bootstrap derived by Sinha (Sinha, 2009).

It is important to note that most methods here have focused on the LMM. Fitzmaurice and Ibrahim (Fitzmaurice et al., 2007) derived a simple permutation procedure based on randomly permuting the indices associated with a given level of a hierarchical model. Although in theory applicable to multiple (uncorrelated) random effects, the application was for testing a single variance component in the absence of nuisance variance components. Lee extended this approach to accommodate multiple correlated random effects in LMM (Lee and Braun, 2012).

Testing variance components in the GLMM is additionally complicated when the model is fit by penalized quasi-likelihood (PQL), or "pseudo-likelihood" in SAS parlance (Breslow and Clayton, 1993; Wolfinger and O'connell, 1993). Briefly, PQL is a doubly-iterative process which approximates the model by transforming the response vector, and then fitting a LMM to the transformed data (itself and iterative process). Testing requires treating the data at final iteration as normal and performing, for example, likelihood ratio tests based on LMM using the transformed data as the modified response. Lin developed a score statistic for testing one or more variance components in GLMM fit by PQL which has been employed in various other forms (Wu et al., 2011; Chen et al., 2013). We similarly employ this statistic in what follows.

1.7 Compositional Background

Data consisting of multiple components that are non-negative, subject to a sum constraint, and where the response is assumed to be related to the relative amounts of each component present, but not to the size of the sample itself are called *compositional data*. Although Karl Pearson's seminal work on spurious correlation that can result from mishandling such data was done over a century ago (Pearson, 1896), modern compositional methods derive from the more recent work of Aitchison (Aitchison and Bacon-shone, 1984) and are nicely summarized by Egozcue et al (Pawlowsky-Glahn and Buccianti, 2011). The fact that the meaningful information is relative information and not a function of the size of the sample itself is also known as *scale invariance*. To form an equivalence class, the data are typically normalized to sum to a constant k, which is formalized by the *closure*

operation. For the D-dimensional vector, $\boldsymbol{x}$, the *closure* is defined as:

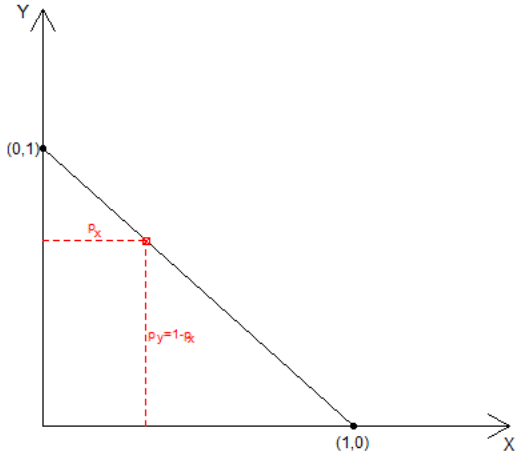
$$\mathcal{C}\boldsymbol{x} = \left(\frac{kx_1}{\sum_{i=1}^{D} x_i}, \frac{kx_2}{\sum_{i=1}^{D} x_i}, \ldots, \frac{kx_D}{\sum_{i=1}^{D} x_i} \right) \tag{1.18}$$

Often (as in the case of proportions) $k = 1$, and without loss of generality we can assume this unity constraint. The sample space for a mixture of D components is the $D - 1$ dimensional simplex:

$$\mathbb{S}^d = \{(x_1, \ldots, x_D) : x_i \geq 0 \ \ (i = 1, \ldots, D), \ \sum_{i=1}^{D} x_i = 1\} \tag{1.19}$$

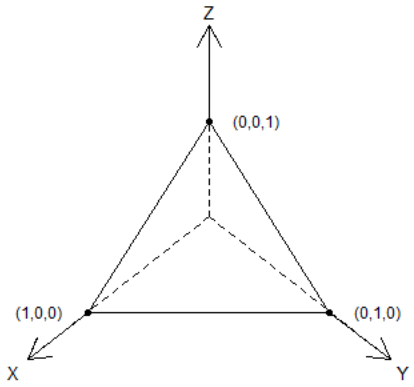
Compositional data are frequently mishandled when statistical tools designed for use on $\mathbb{R}^D$ are applied to data that reside on $\mathbb{S}^{D-1}$.

Figure 1.9: One-dimensional simplex



Much of the work on compositional data applies to the case where the compositional data are themselves the response variables, but often researchers are interested in studying the effects of compositional covariates on some other response variable. This causes problems for testing hypotheses of interest. One scientific area where compositional methods have gained recent traction is the biological study of the gut microbiome (Lin et al., 2014; Li, 2015), where certain phenotypes (i.e., inflammation) are known to be related to the relative contributions of different taxa present.

Figure 1.10: Two-dimensional simplex



To geometrically illustrate the interpretability problems associated with compositional covariates, we consider some simple cases. Suppose a sample is made up of two sexes, male and female, and we wish to control for sex in our analysis. Since proportions are non-negative, Figure 1.9 corresponds to the first quadrant. The unity constraint means the sample space is the line segment connecting $(0, 1)$ and $(1, 0)$. The parameter of interest in a model that controls for sex is meaningful. For example, using females as the reference group, it would be the marginal effect of a unit-increase in the male proportion. Since there are only two components, this of course corresponds to a unit decrease in the female proportion, which is congruous with the scientific question of interest. This interepretation no longer applies when there are more than two components to a composition. Consider, for example, a mixture of three components, illustrated in Figure 1.10. This figure is viewed from within the first orthant, looking in towards the origin. In three dimensions, the unity constraint means the sample space is the triangle connecting the points $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. In fitting this model, it is tempting to simply drop one of the terms. This might even make sense if one only wishes to control for those covariates, when the target of inference is some other covariate. But when the compositional covariates are themselves the target of inference, such models are innappropriate; the parameters do not correspond to hypotheses of interest, stemming from the fact that it is impossible to alter one proportion without simultaneously altering at least one of the other proportions. More formally, a

contrast in a statistical model has a marginal interpretation. That is, in a (generalized) linear model, a monotonic function of the expected value of the response is assumed equal to a linear predictor. Thus, a parameter estimate (or a linear combination thereof) can be interpreted as the derivative of the linear predictor with respect to a variable (or a linear combination of variables) of interest. We have established that compositional data are *scale invariant*. Another way of putting this is that they are *homogeneous of degree zero*. A function $f(\cdot)$ is homogoneous of degree k if $f(\alpha \boldsymbol{x}) = \alpha^k f(\boldsymbol{x})$. Applying Euler's theorem for homogeneous functions, if $f(\cdot)$ is scale invariant and differentiable at x, then:

$$\sum_{j=1}^{n} x_j \frac{df}{dx_j}(x) = 0. \tag{1.20}$$

In words, the constraint on the domain redounds to the derivatives. Thus, the traditional interpretations for the contrasts no longer apply.

In their 1984 paper, Aitchison and Bacon-Shone (Aitchison and Bacon-shone, 1984) detailed what we now know as that *additive logratio transformation (alr)* for handling compositional covariates, which entails removing the constant-sum (that is, transforming the data from the simplex to a real space) via the following transformations:

$$z_i = \log \frac{x_i}{x_D} \quad (i = 1, \ldots, D-1). \tag{1.21}$$

Thus, denoting the expected response as $\eta(x)$, the linear model for a D-component mixture becomes:

$$\eta(x) = \beta_0 + \sum_{i=1}^{D-1} \beta_i z_i \tag{1.22}$$

which is easily shown to be equivalent to:

$$\eta(x) = \beta_0 = \sum_{i=1}^{D} \beta_i \log(x_i) \quad (\beta_1 + \cdots + \beta_D = 0). \tag{1.23}$$

While this effectively handles the unity constraint, several problems remain, one of which is interpretability. That is, all parameters are interpreted relative to the initially chosen reference component. There is also a theoretical problem here, as distances between points in the transformed space differ when different divisors are selected. One proposed solution to this was to instead employ the *centered*

logratio transformation (clr), which is defined as follows:

$$(y_1, \ldots, y_D) = (\ln \frac{x_1}{\sqrt[D]{\prod_{i=1}^{D} x_i}}, \ldots, \ln \frac{x_D}{\sqrt[D]{\prod_{i=1}^{D} x_i}}). \tag{1.24}$$

Note that with this coordinate system, the singularity is retained, along with the associated interpretability issue. The best solution for dealing with compositional covariates is to express the compositional covariates in an orthonormal basis on the simplex (with respect to the Aitchison geometry). This is achieved by the following *isometric logratio transformation (ilr)*:

$$z_i = \sqrt{\frac{D-i}{D-i+1}} \ln \frac{x_i}{\sqrt[D-i]{\prod_{j=i+1}^{D} x_j}}, \quad i = 1, \ldots, D-1. \tag{1.25}$$

Having performed this transformation, the variable z_1 is now of scientific interest. It contains all the information about the compositional part x_1 in relation to its ratios between it and all other components of the mixture. Note that the same interpretation does not pertain for z_2, as x_1 does not appear in this term. A solution is to construct D different ilr tranformations, where the role of z_1 is assumed, in turn, by each component. That is, to investigate the significance of the kth compositional part, we replace $(x_1, \ldots, x_D)$ with $(x_k, x_1, \ldots, x_{k-1}, x_{d+1}, \ldots, x_D) =: (x_1^k, \ldots, x_k^k, x_{k+1}^k, \ldots, x_D^k)$. Noting that the *ilr* transformations are all orthogonal transformations of each other, there is no need to refit the model D times, only to perform the D linear transformations and associated hypothesis tests (Hron et al., 2012).

CHAPTER 2: MODEL IDENTIFIABILITY IN LINEAR MIXED MODELS WITH MULTIPLE VARIANCE COMPONENTS WITH APPLICATION TO DIALLEL DESIGNS

2.1 Introduction

In this chapter, we seek to clear up some issues pertaining to model identifiability in the linear mixed model (LMM), when the variance components are themselves the target(s) of inference. Our particular focus is on sparse diallel designs. It is quite easy to fit such models with multiple variance components, such that the model specified is actually unidentifiable. Further, in such cases, it is not uncommon for existing statistical software to fail to return a reliable warning, which we demonstrate in simulation. There is therefore a danger of performing inference on unidentifiable model parameters. We develop a simple check for mixed model identifiability that is both easy to use, and generally applicable.

2.2 Background

Our study design (refer again to Figure 1.1) can be recast as a type of sparse diallel. This is helpful in order to reconcile our work with existing literature. The salient feature our design is the large one-way loop (*OWL*) comprised of the parental strains labeled 1 through L, where $L = 35$ in our study. We focus on this portion of the data because it is nearly identical to a well-known partial diallel known as the *single round robin (SSR)*. The diallel has had a long and complicated history in genetics, beginning nearly a century ago (Schmidt, 1919). Contributions have come from sub-fields of applied genetics, where the experimental designs and associated analysis methods have differed based on the science relevant to the specific subfield. Although all such models can be described in a general framework, it was several decades until this was attempted by Hayman and Griffing (Hayman, 1954b,a, 1958, 1960; Griffing, 1956a,b). They defined a diallel cross as one in which a set of p inbred lines are chosen and crosses (though not necessarily *all* crosses) are made among these

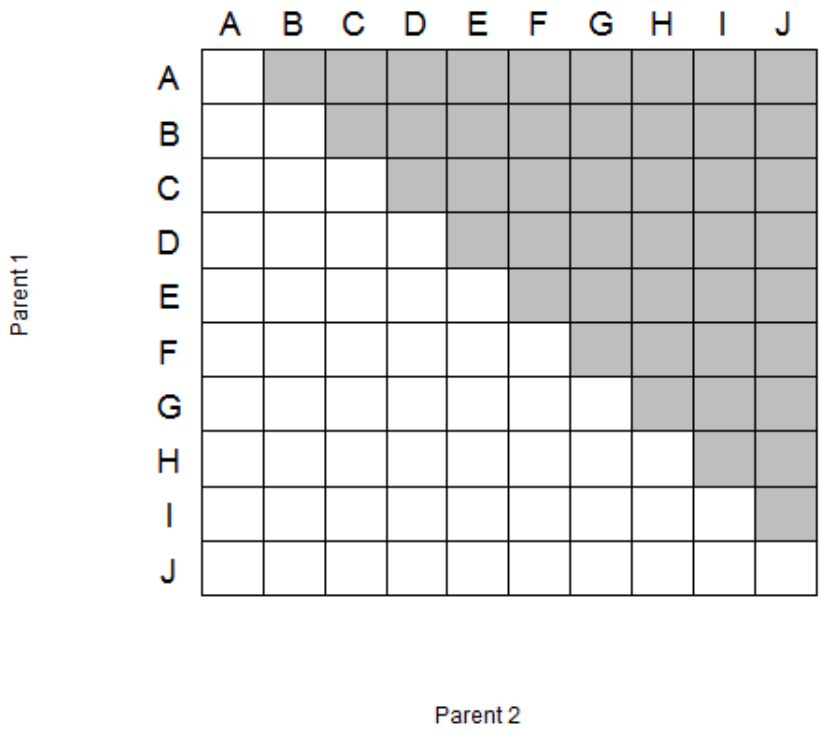


Figure 2.11: Half diallel with no selfs

lines (Griffing, 1956a,b). The structure of the data can be easily visualized in a $p \times p$ table. The p^2 possible combinations can be divided into three separate groups:

1. The p inbred parental lines themselves, corresponding to the main diagonal in the table.

2. One set of F_1 hybrids, corresponding to, say, the upper right triangle of the table, denoting offspring whose maternal parent is indexed by row label and paternal parent by column label.

3. The other set of F_1 hybrids, corresponding to the lower left triangle of the table. These are denoted *reciprocal* F_1's.

For example, Figure 2.11 below illustrates a half diallel: the $\frac{p(p-1)}{2}$ crosses corresponding to a single group of F_1's (reciprocal and parental lines excluded). This is one of the classic types of diallels detailed by Griffing, though any subset of the p^2 possible crosses (e.g., the OWL) is a partial diallel.

A difficult aspect of diallel data is that while diallels are conceptualized in terms of individual parental effects, these effects always come in pairs. Li and Locken term this type of data "dyadic" and nicely synthesize work from different fields, as dyadic data occurs in areas outside of genetics, including, for example, psychology (Li and Loken, 2002). This was presicely the feature of the data that Clatworthy stressed in the 1950's by framing the diallel as a partially balanced incomplete block design (Clatworthy, 1955), although we retain the pithy term "dyadic".

The purpose of a diallel experiment is twofold. First, the design can be used to study the heritability of traits of interest (that is, to estimate the genetic components of phenotype variation among the crosses). It can also be used to estimate the actual breeding values of the parental lines included (Kempthorne and Curnow, 1961). Although Schmidt published the first diallel in 1919 (Schmidt, 1919), it can be argued that Sprague and Tatum (Sprague and Tatum, 1942) pioneered the theory of the diallel when they coined the following terms:

1. *General combining ability* (GCA): the average performance of a line in hybrid combination.

2. *Specific combining ability* (SCA): designates cases in which certain combinations do better or worse than would be expected based on the average performance of the lines involved.

In statistical parlance, the GCA effects correspond to the main effects of parental lines, and the SCA effects to interaction terms. We note that Sprague and Tatum's study design corresponds to the half diallel shown in 2.11. The full model for this design type is:

$$Y_{(i,j),k} = \mu + g_i + g_j + s_{(i,j)} + \epsilon_{(i,j),k} \tag{2.26}$$

where g_i refers to the average performance of line i (GCA), $s_{(i,j)}$ refers to the deviation from the expected additive contributions of lines i and j when they are used in combination (SCA), $\epsilon_{(i,j),k}$ is an error term specific to the kth offspring of parents (i, j) and $Y_{(i,j),k}$ is this subject's phenotype. We use the parentheses in the above notation to stress the dyadic nature of the data. For designs that include reciprocal crosses, these effects can be further dissected. The full model becomes:

$$Y_{(i,j),k} = \mu + g_i + g_j + s_{(i,j)} + d_i - d_j + r_{(i,j)} + \epsilon_{(i,j),k} \tag{2.27}$$

Here, the GCA has been dissected into two pieces: g_i now corresponds to the average GCA associated with strain i, and d_i to a symmetric deviation from this average associated with having a parent of strain i - these are sometimes called parent-of-origin (PoO) effects. Similarly, SCA is dissected into two pieces, where $s_{(i,j)}$ corresponds to the average SCA associated with pair (i, j) and $r_{(i,j)}$ to a symmetric deviation that accounts for the direction of the cross. Note that in this formulation $g_\cdot$ and $s_{\cdot,\cdot}$ are referred to as GCA and SCA, respectively, while $d_\cdot$ and $r_{\cdot,\cdot}$ are referred to as reciprocal GCA (rGCA) and reciprocal SCA (rSCA), respectively. Although most diallel literature focuses on the classical models described above, there is also a rich literature on partial diallel designs (Kempthorne and Curnow, 1961; Greenberg et al., 2010; Viana et al., 1999; Singh and Hinkelmann, 1995, 1998, 1990; Mukerjee, 1997; Hinkelmann and Kempthorne, 1963; Gupta et al., 1995; Ghosh and Divecha, 1997; Fyfe and Gilbert, 1963; Das et al., 1998; Arya, 1983; Verhoeven et al., 2006; Clatworthy, 1955). It remains unclear which parental effects should be included when partial diallel designs are employed; no systematic work has been done in this area.

2.3 Model Identifiability

The parental effects in a partial diallel can be assumed to be either fixed or random. In either case, there are interesting and subtle identifiability issues to be discussed. An important, though in practice often ignored, facet of statistical modeling is an intitial check for model identifiability. It is of course not necessary for a model to be identifiable, only that inference be performed on identifiable parameters. However, it is good practice because we show below that it is possible to fit non-identifiable models and receive no warning from existing statistical software.

2.3.1 Fixed Effects Model Identifiability

Here, we illustrate an identifiability issue with the OWL design when a fixed effects model is employed. We consider the situation where only GCA effects are of interest and the number of RI lines is relatively small. That is, our model takes the following form:

$$Y = Gb + \epsilon \tag{2.28}$$

where Y is an $n \times 1$ vector of observed phenotypes and ϵ is an $n \times 1$ vector of random errors such that $\epsilon \sim N(0, \sigma^2_\epsilon I_{n \times n})$. Assuming a total of p parental strains, $b = (b_1, \cdots, b_p)^T$ is a $p \times 1$ vector of fixed parental strain effects, and G is named as such because it corresponds to the GCA design matrix.

In certain cases, the loop design undertaken solves an identifiability problem that exists in "unconnected" designs, like the "chain" design illustrated in Figure 2.12 below. We illustrate this identifiability problem using a simple example with four parental strains. Again, each node corresponds to a parental inbred line, and each edge to a cross of the connected strains, where the arrow indicates the direction of the cross. There is an identifiability issue with this design, which might not be immediately obvious, though it stems from the fact that the parents come in pairs, with each row of the parental design matrix having two non-zero elements. Mathematically, the parental design matrix (we present the essence matrix in Figure 2.13) is singular.

Figure 2.12: Chain design with four parents

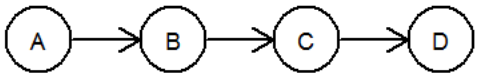


Figure 2.13: Design matrix associated with four-parent chain

A	B	C	D
1	1	0	0
0	1	1	0
0	0	1	1

Note that the sum of the odd columns (A and C) equals the sum of the even columns (B and D). Since a major goal is often to rank the additive contribution of each parental strain to a phenotype

under study, this design is unhelpful. "Closing" the chain to create a loop solves this problem in a perhaps somewhat subtle way. *For loops constructed from an even number of parental strains, the identifiability problem persists, but perhaps surprisingly, this problem is solved when the loop contains an odd number of parents.* A general design matrix for the OWL design is shown below:

$$G = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & ... & p \end{array} \\ \left[\begin{array}{ccccc} J_{n\times 1} & J_{n\times 1} & 0 & \dots & 0 \\ 0 & J_{n\times 1} & J_{n\times 1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & J_{n\times 1} & J_{n\times 1} \\ J_{n\times 1} & 0 & 0 & \dots & J_{n\times 1} \end{array}\right] \end{array}$$

For loops constructed from an even number of parental strains, this design matrix is singular. Note that the sum of the odd columns (associated with parents 1,3, etc.) equals the sum of the even columns (associated with parents 2, 4, etc.). Perhaps surprisingly, this problem is resolved only when the loop contains an odd number of parents. This is a direct consequence of the dyadic nature of the data. The first cross provides an estimate of the sum of two parental effects (say $b_1 + b_2$). In order to isolate b_1, we must also generate an estimate of $(b_1 - b_2)$. We do this by first subtracting the adjacent effect $(b_2 + b_3)$ from the original effect $(b_1 + b_2)$, yielding $(b_1 - b_3)$. To eliminate b_3, we proceed to the next cross, this time adding its effect $(b_3 + b_4)$. Propagating around the loop in this way, in order for our final estimate to be of $(b_1 - b_2)$, as needed, the last operation must be subtraction. Thus, the integers 2 through p can be paired, and p must therefore be odd. When the loop is comprised of an even number of strains, such separation is not possible, but in Appendix A we formally derive what is estimable in such designs. Specifically, we prove the following claims:

Claim 1: For a loop design with an even number of parental strains (q), we can estimate $b_j + b_{j+j'}$ $\forall j$ and for odd j' where $j, j' \in 1, ..., q$.

Claim 2: For a loop design with an even number of parental strains (q), we can estimate $b_j - b_{j+j'}$ $\forall j$ and for even j' where $j, j' \in 1, ..., q$.

Further, we derive what is estimable in a loop that incorporates reciprocal crosses (i.e., crosses are generated in both directions) when we assume the existence of only GCA and rGCA effects. In

Appendix B, we develop interesting designs that allow for the estimation of the same effects as in the reciprocal loop designs, but without incorporating any actual reciprocal crosses.

Similar conclusions to some of the above have been reached in existing studies of partial diallels (Gilbert, 1958; Fyfe and Gilbert, 1963; Kempthorne and Curnow, 1961), though in somewhat different contexts.

2.3.2 Mixed Model Identifiability

Although fixed effects models can be theoretically sound, they can be of limited use in practice for analyzing sparse diallels. For a given number of mice, it is often of interest to generate as many crosses as possible - that is, to include more parental strains, with fewer numbers of mice per strain. The number of parameters in the model therefore increases with the sample size, yielding inconsistent maximum likelihood estimates (also known as "Neyman Scott" problems). Another problem is the fact that mouse studies often suffer from missing data, for which the mixed model is better suited. We present the following theorem to establish identifiability for linear mixed models, similar to one presented in (Demidenko, 2013):

Theorem 1: Consider the following model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{k=1}^{K} \boldsymbol{Z}_k \boldsymbol{b}_k + \boldsymbol{\epsilon} \tag{2.29}$$

where the $\mathbf{b}_k$'s are random effects associated with various types of genetic effects, and the $\boldsymbol{Z}_k$'s are the associated random design matrices. Further, $\mathbf{b}_k \sim N(\mathbf{0}, \sigma_k^2 \boldsymbol{I}_{p_k})$ for $k = (1, \ldots K)$ independently of each other and also independently of $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ and $\boldsymbol{I}_{p_k}$ refer to identity matrices of dimension n and p_k, respectively, where p_k is the number of columns of $\boldsymbol{Z}_k$. We denote σ_k^{*2} and σ_k^2 to be unique values of the parameter. The model is identifiable if and only if the following n^2 equations:

$$(\sigma_\epsilon^{*2} - {\sigma_\epsilon}^2)\mathbf{I}_n + (\sigma_1^{*2} - {\sigma_1}^2)\mathbf{Z}_1\mathbf{Z}_1^{'} \ldots + ({\sigma_K^*}^2 - {\sigma_K}^2)\mathbf{Z}_K\mathbf{Z}_K^{'} = \mathbf{0} \tag{2.30}$$

imply $\sigma_\epsilon^{*2} - {\sigma_\epsilon}^2 = \sigma_1^{*2} - {\sigma_1}^2 = \ldots = \sigma_K^{*2} - {\sigma_K}^2 = 0$. *Defining* $\mathbf{A}$ *as* $[vec(\mathbf{I}_n), vec(\mathbf{Z}_1\mathbf{Z}_1'), \ldots, vec(\mathbf{Z}_K\mathbf{Z}_K')]$ *we can write the* n^2 *equations above as:* $\mathbf{A}\begin{bmatrix} \sigma_\epsilon^{*2} - {\sigma_\epsilon}^2 \\ \sigma_1^{*2} - {\sigma_1}^2 \\ \vdots \\ \sigma_K^{*2} - {\sigma_K}^2 \end{bmatrix} = \mathbf{0}$. *Thus, we need only show that the* $n^2 \times (K+1)$ *matrix* $\mathbf{A}$ *be full rank for the model to be identifiable.*

Others have noted the subtleties around model identifiability in mixed models and the need for formal checks of identifiability. Wang (Wang et al., 2013) presents a necessary and sufficient condition to establish non-identifiability, which is developed in the context of particular common covariance structures for the random effects and residual errors. While our identifiability condition pertains to a more narrow class of mixed models, it enjoys the advantage of relative simplicity and universal applicability to all such models. Specifically, our work applies to models with simple independent variance components, but with arbitrarily complex associated random design matrices. Note that, through careful re-paramaterization, this framework can accomodate a very broad range of models.

2.4 One Way Loop / Single Round Robin

We consider now our *OWL* design, which is nearly identical to the *SRR* design that has been used in the literature. For our purposes, the theory is identical. Both designs are illustrated in Figure 2.14 below, with the grey cells common to both designs; the SRR and OWL additionally include the black and shaded cells, respectively.

It has been claimed (Verhoeven et al., 2006) that for the *SSR*, model (2.26) cannot be fit as the GCA and SCA are not estimable. This is, however, only true in the fixed effects setting (the same logic applies to the *OWL*). We present the design matrices for the *OWL* design below. In deference to tradition, we label the GCA and SCA random design matrices G and S, respectively. We adopt the notation $G_{(r)}$ and $S_{(r)}$ to refer to the reciprocal GCA and SCA effects, respectively. For a balanced OWL design with p parental strains and n observations per cross:

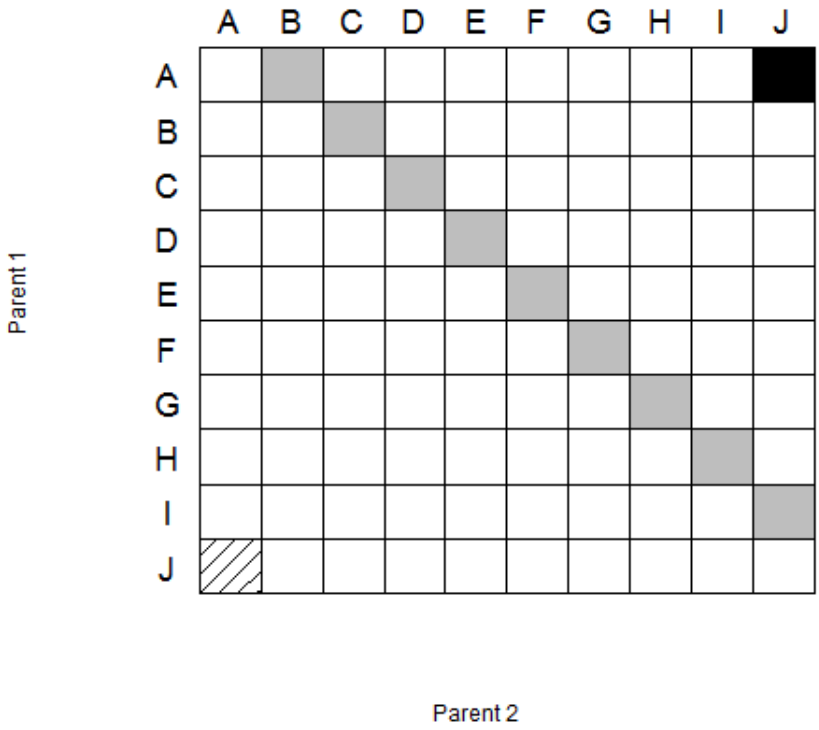


Figure 2.14: SRR/OWL design

$$G = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & ... & p \end{matrix} \\ \begin{bmatrix} J_{n\times 1} & J_{n\times 1} & 0 & \dots & 0 \\ 0 & J_{n\times 1} & J_{n\times 1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & J_{n\times 1} & J_{n\times 1} \\ J_{n\times 1} & 0 & 0 & 0 & J_{n\times 1} \end{bmatrix} \end{array}$$

$$G_{(r)} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & ... & p \end{matrix} \\ \begin{bmatrix} J_{n\times 1} & -1 * J_{n\times 1} & 0 & \dots & 0 \\ 0 & J_{n\times 1} & -1 * J_{n\times 1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & J_{n\times 1} & -1 * J_{n\times 1} \\ -1 * J_{n\times 1} & 0 & 0 & 0 & J_{n\times 1} \end{bmatrix} \end{array}$$

Note the reciprocal GCA matrix is created from the GCA matrix, changing the sign of each entry associated with a maternal parent to negative one. The SCA design matrix is $S = I_p \otimes J_{n\times 1}$ where $\otimes$ denotes the Kronecker product. The design matrix associated with the error term is I_{np}. By using G, S, and I_{np} as Z_1, Z_2, and Z_3 in the identifiability condition above, it is easily shown that $\boldsymbol{A}$ is full rank and thus the model is identifiable. Some straightforward but tedious calculations yield:

$$A = \begin{bmatrix} 6n^2p & 2n^2p & 2np \\ 2n^2p & n^2p & np \\ 2np & np & np \end{bmatrix}$$

which is clearly full rank. If, however, we add reciprocal GCA effects to the fitted model, such calculations yield:

$$A = \begin{bmatrix} 6n^2p & 2n^2p & 2n^2p & 2np \\ 2n^2p & 6n^2p & 2n^2p & 2np \\ 2n^2p & 2n^2p & n^2p & np \\ 2np & 2np & np & np \end{bmatrix}$$

which is of rank 3. For example, four times the third column equals the sum of the first two columns, and thus the model is no longer identifiable.

2.5 Simulation Study

Our preferred software for fitting mixed models to dyadic data is SAS, as PROC MIXED allows for easy specification of random design matrices with multiple nonzero elements per row (we have included sample code in the Appendix C. We note that this is not easily accomplished in the common mixed model software packages in R, as the piping syntax used to specify random effects precludes multiple nonzero elements per row. Of possible interest, one could take advantage of a link between ridge regression and random effects to effectively "trick" R into fitting a random effects model with (a single) arbitrarily complicated random design matrix. This link is detailed in Appendix D.

Statistically, there is no problem with fitting an unidentifiable model. A potential pitfall, however, arises when the researcher makes inference from such a model. If the researcher has not checked model identifiability prior to model fitting, there is no guarantee that the software will produce output that indicates the unidentifiability. Wang (Wang et al., 2013) documented this phenomenon, noting that "software may or may not indicate a problem with model identifiability". In his simulation study, he checks for warnings of non-identifiability through non-convergence of the model as well as zero or extremely large standard errors for the variance components. We add to this research with our own experience with the *OWL* sparse diallel design.

We simulated data from the *OWL* design with GCA and SCA effects (an identifiable model) under four scenarios:

1. $\sigma^2_G = \sigma^2_S = \sigma^2_\epsilon = 1$
2. $\sigma^2_G = \sigma^2_S = 1,\ \sigma^2_\epsilon = 2$
3. $\sigma^2_G = 0.5,\ \sigma^2_S = 1,\ \sigma^2_\epsilon = 2$
4. $\sigma^2_G = 1,\ \sigma^2_S = 0.5,\ \sigma^2_\epsilon = 2$

For each scenario, we simulated 10,000 datasets, and fit a non-identifiable model with GCA, rGCA, and SCA effects. Rather than investigate the standard errors for the variance components,

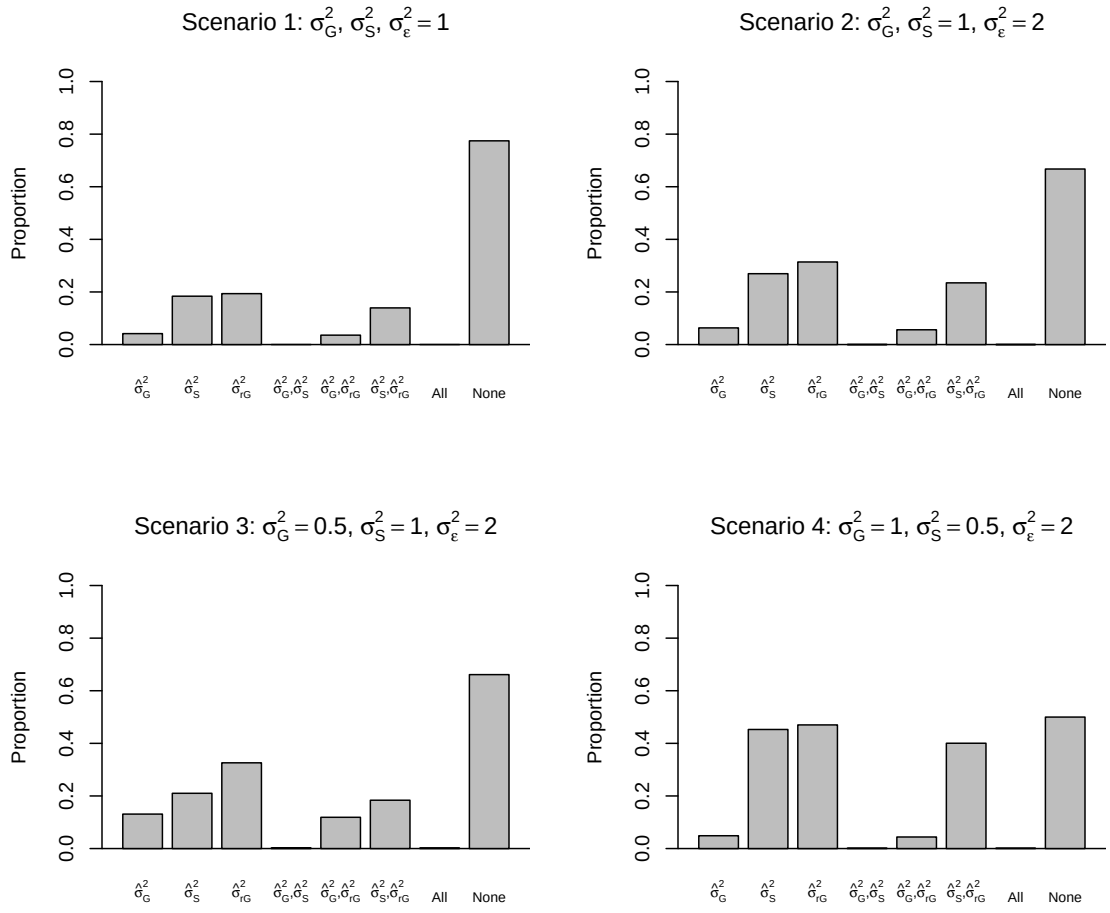


Figure 2.15: Proportion of simulations wherein variance components are estimated at zero

we note that PROC MIXED generates two notes in the log which provide evidence of model unidentifiability. Whenever all variance components are estimated in the interior of the parameter space (that is, none are estimated at zero), SAS returns a note in the log indicating the final Hessian is not positive definite. In our simulations, $(\sigma^2_G, \sigma^2_{rG}, \sigma^2_S) \in \mathbb{R}^3_+$ 20.5, 33.4, 34.6, and 50.4 percent of the time in scenarios 1-4, respectively. The distributions of the zero estimates are illustrated in Figure 2.15.

When at least one variance parameter is estimated at zero, SAS simply returns a note that the estimated covariance matrix for the random effects (G in SAS parlance - not to be confused with our G above) is not positive definite. This is a potentially misleading note, as it does not distinguish between an unidentifiable model and an identifiable model for which the data support one or more of

the variance components truly being zero. A potential pitfall is to proceed by simply eliminating the particular variance component(s) that happened to be estimated at zero in the unidentifiable model.

2.6 Discussion

In this chapter, we have cleared up some important and often overlooked issues concerning model identifiability which were called to our attention through our experience with data from a sparse diallel design. Through our simulations we showed that model non-identifiability may not be obviously revealed based on model results alone. When one or more variance components are estimated on the boundary, the note that SAS returns can falsely convince the researcher to simply drop the zero terms from the model in order to yield a positive definite random effect covariance matrix. This demonstrates the need to check model identifiability prior to model fitting in the setting of linear mixed models with multiple variance components. We have provided a simple model check that can be used for this purpose. Although our application was in complex mouse studies, in particular sparse diallels, the identifiability condition we established applies to all mixed models with multiple variance components.

CHAPTER 3: VARIANCE COMPONENT TESTING IN SPARSE DIALLEL DESIGNS

3.1 Introduction

In Chapter 2, we developed an identifiability condition to be used in study designs with complex correlation structures, and demonstrated the importance of checking for model identifiability and the pitfalls associated with not doing so. Having established identifiability, an important next step might be to test the significance of one or more of the variance components in the model. In this chapter, we demonstrate that the standard tools are not designed for hypothesis testing in our complicated setting. As a remedy, we extend an existing score statistic and demonstrate its value.

3.2 Model Parametrization

Ideally, a well-designed study relates a scientific question of interest to a parameter in a statistical model; this allows for a cohesive hypothesis testing framework. In some cases, accessible statistical tools do not permit researchers to construct such models. In these scenarios, *ad hoc* or purely descriptive approaches are often taken to reconcile estimates produced by a model with the original scientific goal. Consider, for example, the parametrization employed in a recent study of the effect of perinatal nutrition on behavioral phenotypes in adult CC RIX mice (Shorter et al., 2018). A salient result of this study was that perinatal nutrition acts on these phenotypes in a parent-of-origin (PoO) dependent manner. To study this phenomenon, twenty unique RI strains were paired and crosses in both directions within these pairings were created (see Table 3.1). That is, each strain was used both maternally (as dam) and paternally (as sire). Dams were exposed to various diets and the RIX

offspring of these crosses were phenotyped. The following mixed model was fit to the data:

$$
\begin{aligned}
Y_{ijk} &= \beta_0 + \beta_j * I_{\{Dam=j\}} + b_{1j} + b_{2k} + \epsilon_{ijk} && (3.31)\\
b_{1j} &\sim N(0, \sigma_1^2) && (3.32)\\
b_{2k} &\sim N(0, \sigma_2^2) && (3.33)\\
\epsilon_{ijk} &\sim N(0, \sigma_\epsilon^2) && (3.34)
\end{aligned}
$$

where Y_{ijk} is the phenotype of the ith mouse of dam of CC strain j and sire of strain k. The random effects b_{1j} and b_{2k} correspond to the dam and sire effects, respectively, which are typically assumed to be independent, and also independent of the error term, ϵ_{ijk}. In this parametrization, β_j corresponds to a reciprocal effect. That is, a RIX offspring of dam j and sire k receives the additive "dose" $b_{1j} + b_{2k}$, while a RIX from dam k and sire j receives, in addition to this, a dose β_j to account for the fact that the direction of the cross has been reversed. Note that in this parametrization, the ten reciprocal effects to be estimated are fixed terms, in contrast with the background parental effects, which are modeled as random. Although there are some technical advantages to such specification (described below), there are several reasons it might be undesirable. We present the study design described above in table form below, noting that one pairing was dropped from the study because it yielded no viable offspring.

It should be obvious that, as we want to estimate strain and reciprocal effects for all pairings, an identifiability issue exists if we were to employ a purely fixed-effects model. That is, we want to estimate three model terms per pairing, but have only two crosses from which to estimate them. Modeling the parental strains as random solves this identifiability problem; the levels of the parental effects would no longer be model parameters. Rather, the variances of these levels would be model parameters (one parameter per random term). This study also illustrates another important issue that the random effects approach resolves known as "Neyman Scott problems". Suppose momentarily that we were not interested in the reciprocal effects and there were therefore no identifiability issue. Because many strains contribute low/moderate sample sizes, a fixed-effects approach would not yield consistent strain estimates. Modeling these effects as random solves this problem, shrinking their estimates toward the strain grand mean. Further, if we were convinced of the desirability of shrinking

Table 3.1: Study design from Schoenrock et al.

RIX	Mice Phenotyped (n)			
	Diet 1	Diet 2	Diet 3	Diet 4
CC001xCC011	19	13	11	20
CC011xCC001	12	14	17	11
CC041xCC051	12	24	11	20
CC051xCC041	13	12	16	14
CC004xCC017	12	6	12	14
CC017xCC004	9	2	15	13
CC023xCC047	10	9	4	8
CC047xCC023	6	4	3	6
CC006xCC026	6	5	10	9
CC026xCC006	6	5	11	14
CC003xCC014	4	3	13	9
CC014xCC003	1	4	8	6
CC035xCC062	1	2	5	5
CC062xCC035	4	7	4	6
CC032xCC042	7	19	21	19
CC042xCC032	6	9	9	6
CC005xCC040	14	24	–	–
CC040xCC005	16	15	–	–

the strain effects, we might analagously wish to shrink the reciprocal effects in a model that incorates them.

Perhaps even more importantly, the random design relates the scientific question of interest (Does this study provide evidence of reciprocal effects?) to an actual model parameter - that is, a variance component. Modeling the reciprocal terms as fixed, we are constrained to asking whether or not a PoO effect exists *for a particular cross*. We would then have to rely on some *ad hoc* approaches like concluding reciprocal effects exist when they are observed in multiple crosses, for example. Note that, given the modest sample sizes employed in such studies, estimates would be subject to substantial type I error inflation, which would need to be addressed through some corrective approach.

3.3 Problems Associated with Connectivity

The aforementioned connectivity causes several problems, both technical and theoretical. Two related problems simply involve the ability to input the relevant random design matrices into common statistical software. Most software designed for diallel anlaysis can only accomodate a limited

number of parental strains and is not sufficiently general to incorporate the arbitrary sparse diallel designs that arise in studies of RIX mice (see, for example, the arbitrary satellite tails and connections in Figure 1.1). One simple workaround is to abandon specialized software and simply employ general mixed model software. However, much existing software is not flexible enough to accomodate the idiosyncratic structure in the parametrization above, which requires multiple nonzero elements per row of each design matrix. For example, the piping syntax employed for the specification of mixed models in R presupposes that the column labels for each random design matrix can be specified in a single column. To our knowledge, the only out-of-the-box software permitting this level of flexibility are SAS's *PROC MIXED* (for LMM) and *PROC GLIMMIX* and *NLMIXED* (for GLMM). Sample code to fit models employing our specification above in *PROC GLIMMIX* is included Appendix C.

Another major problem caused by connectivity is computational. Typically, estimation of fixed effects proceeds by integrating the random effects out of the likelihood, as follows:

$$\int_{\boldsymbol{b}} \prod_{i=1}^{n} p(\boldsymbol{y}|\boldsymbol{b})p(\boldsymbol{b}) \tag{3.35}$$

where $p(\cdot)$ denotes a distribution function. Note that the dimension of the integration is the dimension of b. For computational advantage, it is often desirable to exchange the order of multiplication and integration, and write the integrated likelihood as follows:

$$\prod_{i=1}^{n} \int_{\boldsymbol{b_i}} p(y_i|\boldsymbol{b_i})p(\boldsymbol{b_i}) \tag{3.36}$$

When such exchange is possible, a single high dimensional integration is replaced with n integrations of much lower order. For example, in a classical longitudinal model including a subject-specific random intercept and slope, each integral above would be two-dimensional. Note that 3.36 is actually the standard starting point for most conceptualizations of the integrated likelihood for mixed models, though this is not sufficiently general. In many settings, the likelihood cannot be broken up into n independent units. In our setting (admittedly, an extreme scenario) the likelihood cannot be written as a product at all. In fact, in designs that are completely connected, it is more accurate to say that the entire data vector is a single (computational) subject.

Writing the likelihood as a product provides great computational advantage in the GLMM. The reason is that for non-normal response, the integrations do not exist in closed form, and require approximation, typically via a quadrature rule. However, the computational cost of using quadrature increases nonlinearly with the number of random effects integrated over. For data layouts like ours that must be processed as in 3.35 above, the integration quickly becomes intractable.

The major appeal of integration via quadrature is the fact that, by increasing the number of quadrature points, the integral approximation can be made arbitrarily precise. But, as intimated above, these approximations really only obtain for models with at most a few random effects (i.e., longitudinal models). Instead, the model can be fit by penalized-quasi likelihood (*PQL*). Briefly, *PQL* is a doubly-iterative process which, rather than approximate the above integral, approximates the model by transforming the response vector at each outer iteration, and then fitting a LMM to the transformed data (itself an iterative process). Recall that our interest is in testing variance components. This is a notoriously difficult hypothesis testing problem because these variance components are boundary parameters - that is, under the null hypothesis, they lie on the boundary of the parameter space. The likelihood ratio test (*LRT*) is by far the most common tool for hypothesis testing in this setting, which Self and Liang, termed "non-standard". In a frequently cited paper, they derived the asymptotic distribution of the *LRT* in certain special cases involving boundary parameters (Self and Liang, 1987). For example, for a model with a single random effect, the asymptotic null distribution of the *LRT* for testing the significance of the associated variance component is known to be a 50:50 mixture of χ^2_1 and χ^2_0 (a point mass on zero) random variables. Crainiceanu and Ruppert (Crainiceanu and Ruppert, 2004) point out that this distribution, while commonly used in practice, is highly dependent upon certain subtle assumptions, and that even when the assumptions are met, the distribution can still perform poorly. One such assumption is that the response vector can be partitioned into independent units under both the null and alternative hypotheses - an assumption that is typically not met in studies involved RIX mice. The situation is further complicated by the fact that *PQL* does not actually approximate the liklihood. It rather approximates the GLMM itself with a LMM at each outer iteration using a set of transformed residuals. Still, the hypothesis testing framework that is standard in statistical software for testing the significance of these variance components is to base inference on an *LRT* treating the data at final iteration as being normally distributed, and using that transformed data to fit the reduced model.

We present a cohesive strategy based on a score statistic developed by Lin (Lin, 1997) that applies to hypothesis testing for models fit by *PQL* (and therefore applies to both LMM and GLMM), where the hypotheses to be tested involve variance parameters. Of note, our approach applies to scenarios where one wishes to test multiple parameters simultaneously, and when nuisance variance parameters are present.

3.4 Score Test

We frame our model as:

$$g\{E[Y|\boldsymbol{b}]\} = X\beta + \sum_{k=1}^{K} Z_k b_k \tag{3.37}$$

where $g\{\cdot\}$ is a link function that links the expected value of the response (conditional on the random effects) to the linear predictor. Having specified our model thusly, we build on the work of Lin (Lin, 1997) to develop a framework for testing one or more variance components, possibly in the presence of one or more nuisance variance components, for models fit by *PQL*. We note that, having specified our model, an important (though sometimes neglected) step is to check for model identifiability. In Chapter 2 above, we have provided a simple means of checking for model identifiability in the LMM, but note that this process is more challenging for the GLMM due to the dependence of the variance on the mean. Although Lin's main result is a "global" score test (that is, a simultaneous test of all variance parameters in a model), she also presents a statistic for testing the jth random effect, while treating the other random effects as nuisance. The statistic is as follows:

$$U_{\sigma_j^2} = \frac{1}{2}[(Y^* - X\beta)^T V_{-j}^{-1} Z_j Z_j^T V_{-j}^{-1}(Y^* - X\beta) - tr(Z_j^T V_{-j}^{-1} Z_j)] \tag{3.38}$$

where Y^* is the pseudo-data at final iteration, and V_{-j} is the variance of the linear predictor with the jth term removed:

$$V_j = (W^{b_{-j}})^{-1} + \sum_{k \neq j} \sigma_k^2 Z_k Z_k^T \tag{3.39}$$

where W is a diagonal matrix with elements:

$$w_i = [V(\mu_i)\{g'(\mu_i)\}^2]^{-1} \tag{3.40}$$

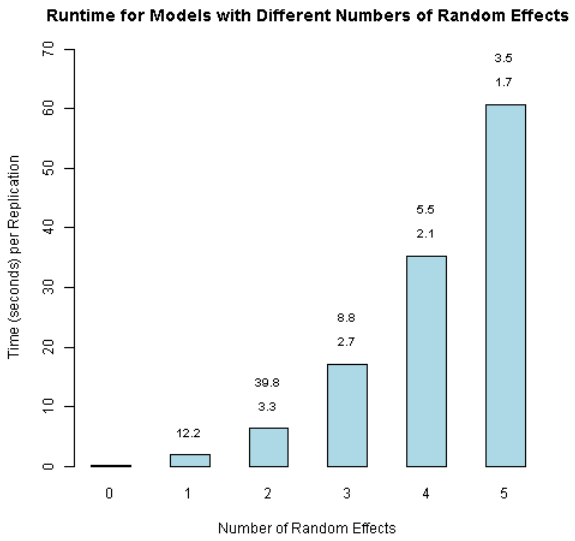


Figure 3.16: Run time for mixed models with increasing number of variance components

and W^{b-j} corresponds to W having been computed with the jth random effect removed. The efficient information for the jth random effect, denoted $\tilde{\mathscr{I}}_{jj}$, is:

$$\tilde{\mathscr{I}}_{jj} = E[\{\mathscr{U}_{\theta_j}(\hat{\alpha}, \hat{\theta}_{-j})\}^2] \tag{3.41}$$

which can then be approximated with $\tilde{\mathscr{I}}^*_{jj}$, where:

$$\tilde{\mathscr{I}}^*_{jj} = \mathscr{I}^*_{\theta_j\theta_j} - (\mathscr{I}^*_{\theta_{-j}\theta_j})^T(\mathscr{I}^*_{\theta_{-j}\theta_{-j}})^{-1}\mathscr{I}^*_{\theta_{-j}\theta_j} \tag{3.42}$$

where we define:

$$\mathscr{I}^*_{\theta_k\theta_{k'}} = \frac{1}{2}tr(Z_{k'}^T V_j^{-1} Z_k Z_k^T V_j^{-1} Z_{k'}). \tag{3.43}$$

A very attractive feature of this score test is how well it dovetails with SAS's *PROC GLIMMIX*, as most of the elements required to compute it are readily exported by the procedure (with the exception of the data vector at final iteration). Another attractive feature is its scalability. That is, as it is readily seen that the computational burden of fitting a GLMM increases nonlinearly with the number of random effects (see Figure 3.16), the benefit of not having to fit the model under the alternative can be large, especially when a testing procedure is applied on the genome-wide scale.

Lin proves that, under the following regularity conditions, the (normalized) score statistic follows a $N(0,1)$ distribution:

1. *Condition 1*: Each of the design matrices Z_j has only zeros and ones, with exactly one 1 in each row and at least one 1 in each column.

2. *Condition 2*: As n, $q_j \to \infty$ $(j = 1, \ldots, m)$, the number of observations at any level of any random factor is bounded by a constant E for all n. It follows that $q_j = O(n)$ $(j = 1, \ldots, m)$.

3. *Condition 3*: The log-quasilikelihood of α has the usual asymptotic properties, including consistency of $\hat{\alpha}_0$ and the linear expansion (White, 1982)

$$n^{\frac{1}{2}}(\hat{\alpha}_0 - \alpha_0) = 1 \tag{3.44}$$

4. *Condition 4*: There exists a positive definite matrix I^0, which has a partition (see Lin Appendix 2), such that:

$$\lim_{n, q_j \to \infty} n^{-1} I = I^0 \tag{3.45}$$

   Note that this assumption is reasonable if Conditions 1 and 2 hold.

5. *Condition 5*: The sequences $\{w_i \delta_i^{-1}(y_i - \mu_i)\}$, $\{x_i\}$, $\{w_{oi}\}$ are uniformly bounded for all $i, \ldots, n$.

6. *Condition 6*: For any given $m \times 1$ constant vector λ_2, let

$$\Omega_{\lambda_2} = \sum_{j=1}^{m} \frac{1}{2} \lambda_{2j} \Delta^{-1} W Z_j Z_j^T W \Delta^{-1} \tag{3.46}$$

Then $y^*_{\lambda_2} = \Omega_{\lambda_2}(y - \mu) = (y^*_{\lambda_{2,1}}, \ldots, y^*_{\lambda_{2,n}})$ forms an M-dependent sequence for some constant M.

We note that (6) constitutes the key condition; it essentially says that the matrix Ω_{λ_2} is sufficiently sparse, which is what is then used to establish the asymptotic normality of the statistic. This assumption is more or less reasonable given conditions (1) and (2). However, these assumptions are clearly not met in our model. In fact, the untenability of condition (1) was precisely our original motivation. That is, our random effect design matrices have two nonzero elements per row, which are 1 or -1. Further, condition 6 is not met, because the loop design, for example, is clearly not

M-dependent. In adapting her proof, the only change required is that the diagonal elements of each $Z_j Z_j'$ are no longer 1, they are 2, and so the $\frac{1}{2}$ in $\Omega_{\lambda_{2j}}$ above can be removed. Similarly, where the following appears:

$$\Gamma_{\lambda_2,i} = \frac{1}{2}(\sum_{j=1}^{m} \lambda_{2j}) w_{oi} \tag{3.47}$$

the $\frac{1}{2}$ can be removed. Lastly, Lin argues that, properly normalized:

$$\lambda^T U = \sum_{i=1}^{n} U_{\lambda,i} = \sum_{i=1}^{n} \{\lambda_1^T U_{\alpha,i} + (y_i - \mu_i) y^*_{\lambda_2,i} - \Gamma_{\lambda_2,i}\} \tag{3.48}$$

is asymptotically normal since $\{U_{\lambda,i}\}$ is an M-dependent sequence. Although this is not the case for our data, we can heuristically argue that we have approximate M-dependence so long as the connections in the design are few relative to the overall size of the loop, and thus her "Proposition 1" can be extended (Lin, 1997).

3.5 Employing the Quadratic Form Portion of the Score Statistic

Rather than formalizing the heuristic logic above, we instead pursue an approach taken by others - for example, by Wu et al in his *Sequence Kernel Association Test (SKAT)* (Wu et al., 2011) - where we employ only the first portion of the statistic above (i.e., the quadratic form). That is, we take as our statistic:

$$Q = (Y^* - X\hat{\beta})' \hat{V}_{-j}^{-1} Z_j Z_j' \hat{V}_{-j}^{-1} (Y^* - X\hat{\beta}) \tag{3.49}$$

In contrast with *SKAT*, Y^* is the pseudo-data at final (outer) iteration (or the actual data if a LMM is fit, in which case there is only a single "outer" iteration). For convenience, we omit this asterisk in what follows. That is, in contrast to above, Y actually refers to the pseudo-data in what follows. Plugging in the weighted least squares estimate for β:

$$\begin{aligned}
Q &= \{Y - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y\}'\hat{V}^{-1}Z_j Z_j'\hat{V}^{-1}\{Y - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y\} \\
&= Y'\{I - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}'\hat{V}^{-1}Z_j Z_j'\hat{V}^{-1}\{I - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}Y \\
&= Y'\{\hat{V}^{-1} - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}' Z_j Z_j'\{\hat{V}^{-1} - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}Y \\
&= Y'P'KPY
\end{aligned}$$

for appropriately defined K and P. To derive the distribution,we start by deriving the distribution of $Y - X\hat{\beta}$. It is normal with expectation zero and variance derived as follows:

$$\begin{aligned}
\hat{Var}(Y - X\hat{\beta}) &= \{I - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}\hat{V}\{I - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}' \\
&= \{I - X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\}\hat{V}\{I - \hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1}X'\} \\
&= \{\hat{V} - X(X'\hat{V}^{-1}X)^{-1}X'\}\{I - \hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1}X'\} \\
&= V - X(X'\hat{V}^{-1}X)^{-1}X' \\
&= P
\end{aligned}$$

We therefore can rewrite the statistic as:

$$\begin{aligned}
Q &= (Y - X\hat{\beta})'P^{-0.5}P^{0.5}\hat{V}^{-1}Z_jZ_j'\hat{V}^{-1}P^{0.5}P^{-0.5}(Y - X\hat{\beta}) && (3.50) \\
&= \tilde{Y}'P^{0.5}\hat{V}^{-1}Z_jZ_j'\hat{V}^{-1}P^{0.5}\tilde{Y} && (3.51)
\end{aligned}$$

where $\tilde{Y} \sim N(0, I)$. Thus, we have expressed the statistic as a quadratic form of a multivariate normal random variable with identity covariance matrix. The distribution can therefore be shown to be a mixture of χ^2_1 distributions, weighted by the eigenvalues of the interior term. For simplicity, we note that:

$$\begin{aligned}
\lambda(P^{0.5}\hat{V}^{-1}Z_jZ_j'\hat{V}^{-1}P^{0.5}) &= \lambda(Z_j'\hat{V}^{-1}P^{0.5}P^{0.5}\hat{V}^{-1}Z_j) \\
&= \lambda(Z_j'\hat{V}^{-1}P\hat{V}^{-1}Z_j)
\end{aligned}$$

where $\lambda(\cdot)$ denotes the spectral decomposition; we therefore need only compute the eigenvalues of $Z_j'PZ_j$. This distribution can be computed analytically, using the results of Davies (Davies, 1980b,a).

3.6 Testing Multiple Variance Components Using Quadratic Forms

In what follows, we will refer to models that include up to five variance components: a batch effect, a parental strain effect, a parental strain-by-treatment interaction effect, a sex-by-strain effect,

and a sex-by-strain-by-treatment interaction effect. The associated random design matrices are constructed as follows:

1. Each row of the random design matrix associated with the batch effect has a single nonzero element, which is a one in the column corresponding to the batch to which the sample belongs.

2. Each row of the random design matrix associated with the parental strain effect has two nonzero elements, which are ones in the columns corresponding to that sample's parental strains.

3. Each row of the random design matrix associated with the strain-by-treatment effect is constructed from the parental strain matrix by switching the ones to negative ones for mice that are treated with placebo.

4. Each row of the random design matrix associated with the strain-by-sex effect is constructed from the parental strain matrix by switching the ones to negative ones for female mice.

5. Each row of the random design matrix associated with the strain-by-sex-by-treatment effect is constructed from the strain-by-sex matrix by switching the ones to negative ones for female mice.

In such scenarios, we wish to test the significance of multiple variance components simultaneously. Specifically, we might wish to test for treatment effects or sex effects, each of which would require testing two terms at once. For example, a test of the treatment effect corresponds to the following hypothesis:

$$H_0 : \sigma^2_{strain*tmt} = \sigma^2_{strain*sex*tmt} = 0$$

vs.

$$H_A : \text{as least one departure from null}$$

As in 3.49 above, a test of each individual variance component is achieved with quadratic forms that are identical except for the Z matrices. That is, if we let Z_j correspond to the strain-by-treatment random design matrix and $Z_{j'}$ correspond to the strain-by-treatment-by-sex random design matrix, we propose employing the following statistic:

$$Q = Q_j + Q_{j'} = \tilde{Y}' P^{0.5} \hat{V}^{-1} (Z_j Z_j' + Z_{j'} Z_{j'}') \hat{V}^{-1} P^{0.5} \tilde{Y} \tag{3.52}$$

Since this itself is a quadratic form, we can compute the distribution analytically, as above, rejecting for large Q.

3.7 Simulation Comparators

In the simulations that follow, we compare our analytic χ^2 approach described above (referred to as χ^2_{mix}) with the simple LRT approach, which is the default in SAS. Here, an LRT is performed using the data at final PQL iteration, which is treated as Gaussian. In addition, for completeness, we include Lin's actual asymptotic χ^2 statistic, as well as a possible adjustment incorporating permutation. That is, we use the simple permutation strategy laid out by Fitzmaurice et al (Fitzmaurice et al., 2007), where instead of using the asymptotic distribution, we derive the reference distribution by repeatedly permuting the column labels of the random design matrix/matrices of interest.

3.8 Results

3.8.1 Simulation Results

We test our method in a comprehensive simulation study, the aim of which was to investigate the performance of the proposed method under a range of study designs, distributional assumptions, and parameter configurations. To investigate the effect of study design, we restrict ourselves to different partial diallel designs: the half diallel, the double round robin (DRR), and two versions of the single round robin (SRR). In all cases, the total sample size was maintained more or less constant. We used a half diallel (with no selfs) composed of six parents, yielding 15 filled cells; employing 12 mice per cell yields a sample size of 180. Our DRR was designed to nearly match the half diallel's sample size; with eight parents and 12 mice per cell, it has a sample size of 192. Our first SRR was composed of 16 parental strains and 12 mice per cell (the sample size was maintained at 192). Lastly, in our second version of the SRR, we increased the number of parental strains to 24, and to maintain the sample size at 192 we decreased the number of mice per cell to eight.

Each of the results tables corresponds to a unique parameter configuration. Specifically, we vary which parameters are being tested (the strain effect, strain-by-treatment interaction, sex-by-strain effect, and sex-by-strain-by-treatment interaction). We explore testing a single parameter of interest or two parameters simultaneousy. When testing two parameters simultaneously, we investigate

scenarios where both parameters are in the interior (i.e., the alternative) of the parameter space, and where only one parameter is the interior and the other is on the boundary (i.e., it is set to the null value of zero). We also explore the effect of varying the number of nuisance parameters in the null model, and whether these parameters lie on the boundary or interior of the parameter space.

For each scenario (or, equivalently, correlation structure) we simulated data from binary, Poisson, and Gaussian distributions. In each case, we compared the type I errors of the proposed procedures. We also provided selected simulation results in order to compare the methods in terms of power. Note that for these power simulations, the alternative parameter values for each distribution were picked more or less arbitrarily in order to make a reasonable comparison *within distributions*. Comparing power between distributions is, of course, essentially meaningless. In Table 3.2 below, we present results for our simplest (and in some sense our most tractable) scenario - one in which we have a single nuisance parameter that lies in the interior of its parameter space, and we are testing a single parameter. We compare results based on the *LRT* statistic given by SAS to our analytic χ^2 mixture statistic. An important point to note (and one that remains relevant in all of what follows) is that the p-values used to gauge significance were not actually computed by SAS. The reason for this is that the null distribution of the *LRT* is dependent upon whether nuisance parameters are on the boundary or interior of the parameter space. For certain configurations, the null distribution is tractable - see, for example (Self and Liang, 1987) or (Shapiro, 1988). Of course, outside of simulation studies, whether or not a nuisance parameter lies on the boundary is itself estimated. SAS computes the reference distribution and associated p-value when it is mathematically available. However, when the reference distribution is unavailable, SAS simply returns the statistic, without a corresponding p-value. Of course, for our simulations that include no nuisance boundary parameters, we know the appropriate reference distribution; for the situation in Table 3.2, it is a 50:50 mixture of χ^2_0 (a point mass on zero) and χ^2_1 random variables. The p-values used to gauge significance were computed accordingly.

As shown, the analytic mixture approach achieves type I error rates much closer to the nominal 0.05 level, and as a result a modest improvement in power. This advantage is consistent across study designs and distributions, although the effect atrophies as the diallel becomes increasingly sparse. As in Table 3.2, the simulations illustrated in Table 3.3 were conducted with a single (large) interior nuisance variance parameter (a batch effect). Here, however, we jointly test the significance of two

variance components simultaneously - the strain effect and the strain-by-treatment interaction. To compute the power in this scenario, we set both tested variance components to interior values. As shown, the results are similar to those in Table 3.2 where we test a single effect; the type I error is much closer to the nominal rate in most scenarios. Here, however, the improvement in power is even more pronounced. In Table 3.4, we explore scenarios where we test two parameters, one of which is in the interior of the parameter space, and the other of which is truly on the boundary (null), in the presence of a single interior nuisance parameter. In Table 3.5, we investigate the robustness of our proposed approach when testing a single parameter of interest when there are two nuisance parameters, one of which lies on the boundary and the other remains in the interior. Note that the commonly used distributions for these two scenarios (a 50:50 mixture of χ^2_0 and χ^2_1 and a 25:50:25 mixture of χ^2_0, χ^2_1 and χ^2_2) are technically incorrect. We include these simulations as a means of investigating the consequences of this common error. As shown, the analytic mixture approach seems more robust to this misspecification of the null reference distribution, as demonstrated by the improvements in type I error rates and power. In the simulations in which we test two parameters of interest, we note that while the analytic mixture approach maintains the corrective effect on the type I error rates, it does not enjoy an analagous improvement in power. In fact, the analytic χ^2 mixture approach does almost uniformly worse than the simple LRT. We explore a possible remedy to this deficiency in analyses below, by incorporating a simple permutation strategy analagous to that laid out by Fitzmaurice et al (Fitzmaurice et al., 2007). That is, we derive the reference distribution by repeatedly permuting the column labels of the random design matrix/matrices of interest. For completeness, we also include results using Lin's asymptotic statistic as well (on which we based the permutation). The results are shown below in Table 3.6.

3.8.2 Real Data Analysis

We present the results of our method when applied to real data from a recent study of RIX mice collected by UNC's Center for Integrated Systems Genomics (CISGen). The study design has been discussed above. Although there are numerous satellite tails and connections, it most closely resembles the SRR, in that the salient feature of the study is a large one-way loop (OWL), a design which is nearly identical to the SRR. The study was intended to include 12 mice per cross, although

Table 3.2: Simulation results for testing a single interior parameter in the presence of a single interior nuisance parameter

Study Design	Distribution	Type I Error		Power	
		LRT	χ^2_{mix}	LRT	χ^2_{mix}
HD*	Binary†	0.0278	0.0430	0.1265	0.1655
	Poisson††	0.0327	0.0489	0.1861	0.2344
	Gaussian‡	0.0339	0.0469	0.1355	0.1677
DRR**	Binary	0.0300	0.0443	0.1381	0.1720
	Poisson	0.0342	0.0501	0.2110	0.2536
	Gaussian	0.0362	0.0460	0.1522	0.1786
SRR1***	Binary	0.0313	0.0423	0.1264	0.1537
	Poisson	0.0360	0.0489	0.2026	0.2410
	Gaussian	0.0370	0.0436	0.1446	0.1610
SRR2****	Binary	0.0306	0.0404	0.1050	0.1251
	Poisson	0.0395	0.0499	0.1804	0.2095
	Gaussian	0.0398	0.0424	0.1275	0.1306

*Half Diallel with 6 parental strains, 12 mice per cross
**Double Round Robin with 8 parental strains, 12 mice per cross
***Single Round Robin with 16 parental strains, 12 mice per cross
****Single Round Robin with 24 parental strains, 8 mice per cross
†$\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.05)$
†† $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.005)$
‡ $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.01)$

there were some deviations. One purpose of this study was to determine the genetic basis of adverse reactions to haloperidol. This is a very important research question because these adverse reactions are the main drivers of noncompliance with treatment, which is quite high with such antipsychotics.

Tremors, tongue movements, and vacuous chewing movements (VCM) are three such side effects. As exemplified in Figure 3.8.2, the minority of mice experienced tremors and tongue movements, but those that did were highly prone and experienced a large number. Because of these highly skewed distributions, these two endpoints were dichotomized to model their presence/absence. VCM were modeled with a negative binomial distribution. The results for the tongue movement endpoint are illustrated in Table 3.7. The first four columns correspond to the fixed-effect estimates, and the next five to the variance component estimates from the model. Each row represents a separate run of the model, with the variance components indicated by dashes removed and tested. Variance components marked as NA were excluded from a particular model run entirely (that is, they were not considered part of the alternative model). The p-values from the analytic χ^2 mixture distribution are shown in the rightmost column. Treatment has a strong marginal effect; subjects treated with haloperidol are much more likely to experience tongue movements. The second row illustrates that there are

Table 3.3: Simulation results for testing two interior parameters in the presence of a single interior nuisance parameter

Study Design	Distribution	Type I Error		Power	
		LRT	χ^2_{mix}	LRT	χ^2_{mix}
HD*	Binary†	0.0278	0.0430	0.1265	0.1655
	Poisson††	0.0327	0.0489	0.1861	0.2344
	Gaussian‡	0.0339	0.0469	0.1355	0.1677
DRR**	Binary	0.0300	0.0443	0.1381	0.1720
	Poisson	0.0342	0.0501	0.2110	0.2536
	Gaussian	0.0362	0.0460	0.1522	0.1786
SRR1***	Binary	0.0313	0.0423	0.1264	0.1537
	Poisson	0.0360	0.0489	0.2026	0.2410
	Gaussian	0.0370	0.0436	0.1446	0.1610
SRR2****	Binary	0.0306	0.0404	0.1050	0.1251
	Poisson	0.0395	0.0499	0.1804	0.2095
	Gaussian	0.0398	0.0424	0.1275	0.1306

*Half Diallel with 6 parental strains, 12 mice per cross
**Double Round Robin with 8 parental strains, 12 mice per cross
***Single Round Robin with 16 parental strains, 12 mice per cross
****Single Round Robin with 24 parental strains, 8 mice per cross
† $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.05), \sigma^2_{strain*tmt} = 0.0(0.05)$,
†† $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.005), \sigma^2_{strain*tmt} = 0.0(0.005)$
‡ $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.01), \sigma^2_{strain*tmt} = 0.0(0.01)$

strong additive strain effects, and the third row provides marginal evidence for a strain-by-treatment interaction. The fourth row jointly tests the sex effects; there is no evidence for their existence. The last row likewise jointly tests the treatment effects; there is marginal evidence that treatment affects the strains differently.

Table 3.8 illustrates the results for the tremor endpoint. Here, there is evidence only for a strong additive strain effect. Lastly, Table 3.9 illustrates the results for VCM; there is evidence of a strong strain effect, and marginal evidence that treatment affects strains differently.

Table 3.4: Simulation results for testing two parameters (one interior and one boundary) in the presence of a single interior nuisance parameter

Study Design	Distribution	Type I Error		Power	
		LRT	χ^2_{mix}	LRT	χ^2_{mix}
HD*	Binary†	0.0265	0.0453	0.1773	0.1532
	Poisson††	0.0327	0.0499	0.5550	0.4940
	Gaussian‡	0.0341	0.0405	0.2329	0.2051
DRR**	Binary	0.0242	0.0414	0.1903	0.1906
	Poisson	0.0337	0.0521	0.6064	0.5564
	Gaussian	0.0319	0.0442	0.1960	0.1592
SRR1***	Binary	0.0257	0.0375	0.1841	0.1879
	Poisson	0.0355	0.0490	0.6364	0.6104
	Gaussian	0.0376	0.0390	0.2108	0.1862
SRR2****	Binary	0.0271	0.0339	0.1427	0.1375
	Poisson	0.0363	0.0497	0.6199	0.5870
	Gaussian	0.0428	0.0366	0.1859	0.1432

*Half Diallel with 6 parental strains, 12 mice per cross
**Double Round Robin with 8 parental strains, 12 mice per cross
***Single Round Robin with 16 parental strains, 12 mice per cross
****Single Round Robin with 24 parental strains, 8 mice per cross
†$\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.1), \sigma^2_{strain*tmt} = 0.0(0.0)$
†† $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.01), \sigma^2_{strain*tmt} = 0.0(0.0)$
‡ $\sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.02), \sigma^2_{strain*tmt} = 0.0(0.0)$

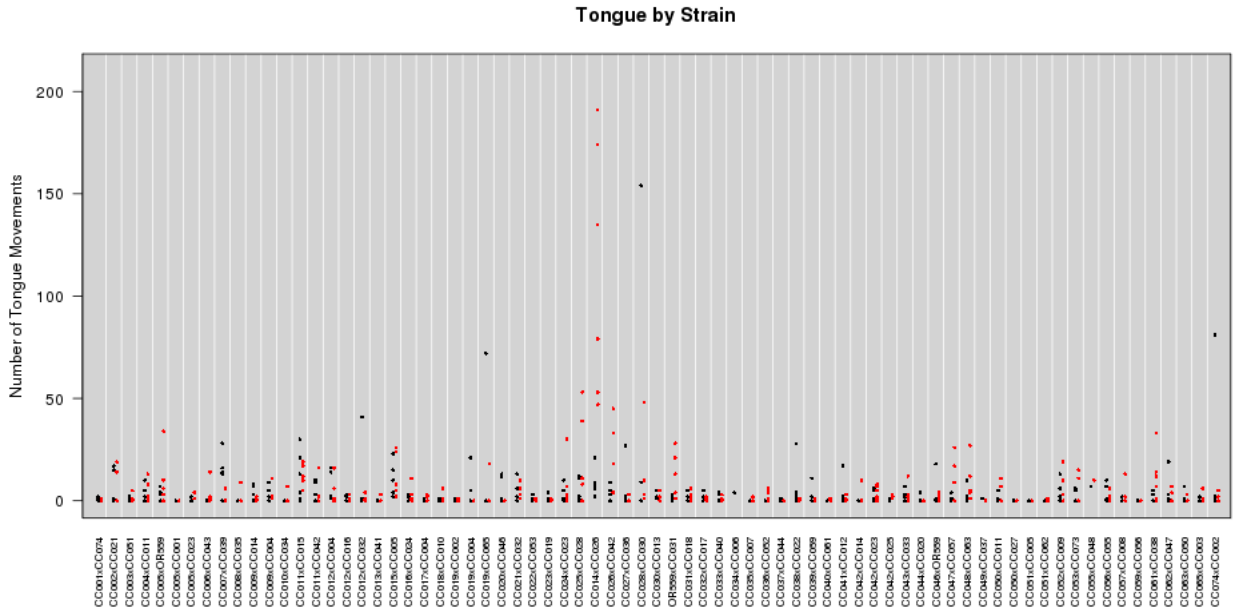


Figure 3.17: Number of tongue movements by RIX line

Table 3.5: Simulation results for testing a single interior parameter in the presence of an interior nuisance parameter and a boundary nuisance parameter

Study Design	Distribution	Type I Error		Power	
		LRT	χ^2_{mix}	LRT	χ^2_{mix}
HD*	Binary†	0.0278	0.0430	0.1265	0.1655
	Poisson††	0.0327	0.0489	0.1861	0.2344
	Gaussian‡	0.0339	0.0469	0.1355	0.1677
DRR**	Binary	0.0300	0.0443	0.1381	0.1720
	Poisson	0.0342	0.0501	0.2110	0.2536
	Gaussian	0.0362	0.0460	0.1522	0.1786
SRR1***	Binary	0.0313	0.0423	0.1264	0.1537
	Poisson	0.0360	0.0489	0.2026	0.2410
	Gaussian	0.0370	0.0436	0.1446	0.1610
SRR2****	Binary	0.0306	0.0404	0.1050	0.1251
	Poisson	0.0395	0.0499	0.1804	0.2095
	Gaussian	0.0398	0.0424	0.1275	0.1306

*Half Diallel with 6 parental strains, 12 mice per cross
**Double Round Robin with 8 parental strains, 12 mice per cross
***Single Round Robin with 16 parental strains, 12 mice per cross
****Single Round Robin with 24 parental strains, 8 mice per cross
$\dagger \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.0), \sigma^2_{strain*tmt} = 0.0(0.1)$
$\dagger\dagger\ \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.0), \sigma^2_{strain*tmt} = 0.0(0.01)$
$\ddagger\ \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.0), \sigma^2_{strain*tmt} = 0.0(0.02)$

Table 3.6: Simulation results for testing two parameters (one interior and one boundary) in the presence of a single interior nuisance parameter, incorporating permutation

Design	Dist.	Type I Error				Power			
		LRT	$\chi^2_{mix.}$	$\chi^2_{asymp.}$	$\chi^2_{perm.}$	LRT	$\chi^2_{mix.}$	$\chi^2_{asymp.}$	$\chi^2_{perm.}$
HD*	Binary†	0.0255	0.0435	0.0348	0.0561	0.1776	0.1578	0.1087	0.1504
	Poisson††	0.0329	0.0499	0.0540	0.0564	0.2941	0.2582	0.2494	0.2339
DRR**	Binary	0.0257	0.0397	0.0344	0.0524	0.1958	0.1977	0.1616	0.1991
	Poisson	0.0339	0.0539	0.0528	0.0604	0.3327	0.3178	0.3031	0.2886
SRR1***	Binary	0.0263	0.0402	0.0361	0.0546	0.1841	0.1853	0.1790	0.2130
	Poisson	0.0362	0.0506	0.0480	0.0559	0.3316	0.3345	0.3141	0.2966
SRR2****	Binary	0.0254	0.0335	0.0323	0.0531	0.1419	0.1386	0.1350	0.1678
	Poisson	0.0369	0.0492	0.0470	0.0550	0.2940	0.3007	0.2652	0.2411

*Half Diallel with 6 parental strains, 12 mice per cross
**Double Round Robin with 8 parental strains, 12 mice per cross
***Single Round Robin with 16 parental strains, 12 mice per cross
****Single Round Robin with 24 parental strains, 8 mice per cross
$\dagger \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.1), \sigma^2_{strain*tmt} = 0.0(0.0)$
$\dagger\dagger\ \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.01), \sigma^2_{strain*tmt} = 0.0(0.0)$
$\ddagger\ \sigma^2_{batch} = 0.1, \sigma^2_{strain} = 0.0(0.02), \sigma^2_{strain*tmt} = 0.0(0.0)$

Table 3.7: Real data analysis of number of tongue movements

Fixed Effects				Variance Components*					
Intercept	Tmt	Sex	Sex*tmt	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	p-value
0.1416	-0.1997	0.1046	-0.0280	0.1345	0.4526	0.0309	NA	NA	NA
0.1365	-0.1741	0.0745	-0.0268	0.4202	–	0.0130	NA	NA	<0.0001
0.1425	-0.2049	0.1047	-0.0279	0.1333	0.4518	–	–	–	0.1277
0.1416	-0.1997	0.1046	-0.0280	0.1345	0.4526	0.0309	–	–	0.4449
0.1425	-0.2049	0.1047	-0.0279	0.1333	0.4518	–	0.0000	–	0.0951

* $\sigma_1^2 = \sigma_{batch}^2, \sigma_2^2 = \sigma_{strain}^2, \sigma_3^2 = \sigma_{strain*tmt}^2, \sigma_4^2 = \sigma_{strain*sex}^2, \sigma_5^2 = \sigma_{strain*sex*tmt}^2$

Table 3.8: Real data analysis of number of tremors

Fixed Effects				Variance Components*					
Intercept	Tmt	Sex	Sex*tmt	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	p-value
0.0391	-0.3734	-0.0447	-0.0433	0.0000	1.5101	0.0000	NA	NA	NA
0.0210	-0.2804	-0.0049	-0.0431	0.5445	–	0.0000	NA	NA	<0.0001
0.0391	-0.3734	-0.0447	-0.0433	0.0000	1.5101	–	NA	NA	0.9690
0.0391	-0.3734	-0.0447	-0.0433	0.0000	1.5101	0.0000	–	–	0.4853
0.0391	-0.3734	-0.0447	-0.0433	0.0000	1.5101	–	0.0000	–	0.6957

* $\sigma_1^2 = \sigma_{batch}^2, \sigma_2^2 = \sigma_{strain}^2, \sigma_3^2 = \sigma_{strain*tmt}^2, \sigma_4^2 = \sigma_{strain*sex}^2, \sigma_5^2 = \sigma_{strain*sex*tmt}^2$

Table 3.9: Real data analysis of number of vacuous chewing movements (VCM)

Fixed Effects				Variance Components*					
Intercept	Tmt	Sex	Sex*tmt	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	p-value
4.5790	0.1806	0.0959	-0.0143	0.0211	0.0504	0.0013	NA	NA	NA
4.5827	0.1805	0.1075	-0.0169	0.0597		0.0001	NA	NA	<0.0001
4.5805	0.1802	0.0958	-0.0144	0.0208	0.0499	–	NA	NA	0.1833
4.5790	0.1806	0.0960	-0.0143	0.0211	0.0504	0.0013	–	–	0.2240
4.5805	0.1802	0.0958	-0.0144	0.0208	0.0499	–	0.0000	–	0.0910

* $\sigma_1^2 = \sigma_{batch}^2, \sigma_2^2 = \sigma_{strain}^2, \sigma_3^2 = \sigma_{strain*tmt}^2, \sigma_4^2 = \sigma_{strain*sex}^2, \sigma_5^2 = \sigma_{strain*sex*tmt}^2$

3.9 Software for Testing Variance Components

The ability to compute the distribution of a quadratic form of *iid* normal random variables analytically can be attributed to Davies (Davies, 1980a), whose work has been implemented in the *R* package *CompQuadForm* (Lafaye De Micheaux, 2013). This code is in fact the engine behind the popular *SKAT* and was also used by Pirinen as the basis for variance component testing in his LMM software *MMM*, which was coded in *C* (Pirinen et al., 2013). We have adapted Pirinen's *C* code (itself an adaptation of the source code from *CompQuadForm*) and combined it with original SAS IML code to create our software for variance component testing in GLMM which can be used with *PROC GLIMMIX*. The code will be freely available under the GNU public license and can easily be adapted for use with any software capable of fitting GLMM by *PQL*.

3.10 Discussion

This chapter was motivated by the existence of study designs that arise in, for example, mouse genetics that feature a certain type of connectivity of the random effects. At the extreme, the response vector cannot be subdivided into any independent units, which causes myriad problems for model fitting and hypothesis testing. For example, the piping syntax in R presupposes that, for any variance component, the column headings can be written in a single column. For connected data, such specification is not adequate to capture the required complex correlation structures, nor does it allow for a comprehensive hypothesis testing framework in many scenarios of interest. Although the required models can be specified in SAS, when the response variable is non-normal the same connectivity precludes standard ML estimation routines due to the intractably high dimension of the integration required to average over the random effects. This connectivity also creates problems for hypothesis testing of variance components, for which the standard theory relies on the ability to subdivide the response vector into independent units under both the null and alternative hypotheses.

The approach we take is to fit the models by penalized quasi-likelihood (for example, in SAS's *PROC GLIMMIX*). While the models can be fit with this approach, it generates another problem - that is, the classic *LRT* is no longer appropriate because *PQL* relies on approximating the GLMM with a LMM, rather than maximizing the original likelihood. We propose adapting a test statistic developed

by Lin (Lin, 1997) that is appropriate for very general models like ours. Specifically, we adopt an approach similar to a well-known adaptation of Lin's statistic where we employ only the quadratic form portion of the statistic as in *SKAT* (Wu et al., 2011) and compute the distribution analytically using the method of Davies (Davies, 1980a). We showed that in certain scenarios (i.e., those where we test a single variance component in the presence of nuisance variance components) this approach dominates the LRT as implemented in standard software. To overcome the poor performance of our method in certain scenarios, we argued that Lin's regularity conditions can be relaxed in order to prove the asymptotic normality of the complete score statistic in our setting. We also demonstrated that a simple permutation scheme could be used, though with mixed results.

Several additional points are worthy of discussion here. For one, we spent some time above discussing the seminal results of Self and Liang (Self and Liang, 1987), and how the reference distributions depend upon whether the nuisance parameters included lie on the boundary or interior of the parameter space. In our work, no attempt was made to address this issue. Indeed, computing the reference distribution is easy only in some simple scenarios. We did show that in some scenariois (i.e., when testing a single variance component) the method was robust to whether or not nuisance parameters were on the boundary or in the interior.

We also note that there might seem to be a type of inconsistency with the reference distributions for the *LRT* and the asymptotic score statistics used in our comparisons. Specifically, for the *LRT*, we employ the common mixture distributions, but for the score statistic we use the classic χ^2 distributions. This inconsistency was pointed out by Verbeke and Molenbergh (Verbeke and Molenberghs, 2003) who frame this as an error on the part of Lin. We, however, disagree with this assessment. The mixture distribution arises from constrained optimization in the variance component setting. That is, variance components that would be estimated as negative if the maximization were unconstrained are typically set to zero. Some have suggested using unconstrained optimization, but that potentially results in uninterpretable parameter estimates. This is not a problem when the variance and covariance parameters are regarded as nuisance, but in situations like ours they are themselves of primary interest. It makes little sense to constrain a parameter that is not actually in the fit model (as in a score statistic). A significant result (regardless of the hypothetical sign of the tested parameter) is evidence of model misspecification - when an estimate for the parameter is required, that term would of course be included in the model, which would be fit with constraints for

the purpose of interpretability. This problem is completely obviated when we use only the quadratic form as in our main results, because we are not employing an asymptotically pivotal statistic, but rather a pseudo-statistic (aka, a plug-in statistic) using the estimates from the model.

CHAPTER 4: IMPLICATIONS OF GENERAL SIMILARITY MATRICES AND COMPOSITIONAL CONSIDERATIONS FOR HYPOTHESIS TESTING

4.1 Introduction

In Chapters 1 and 2, our interest was strictly in establishing identifiability and testing the significance of variance components in very particular types of models where we modeled the parental strain effects as random. Technically, this amounted to random design matrices that were more complicated than those seen in standard longitudinal or clustered models, but albeit with some structure and sparsity: in every row, there were two nonzero elements. We established a formal identifiability check and reviewed the limitations of the standard asymptotic reference distributions for the LRT when testing boundary parameters (they are often mixtures of χ^2 distributions).

In this chapter, we consider two modifications to such models. First, we allow for more general random design matrices that flexibly accomodate the relatedness in the parental strains. Secondly, we bring the proportions of the genome attributable to each of the eight founder strains into the model (henceforward referred to as *founder compositions*) as fixed effects. In such models, entertaining hypotheses about overall genetic effects requires jointly testing fixed and random terms. At first glance, this would seem to be a trivial problem, the relevant asymptotic theory having been worked out already by Self and Liang (Self and Liang, 1987) and described in detail in our literature review. However, there are subtle issues at play which, if not properly considered, can lead the researcher down potentially disastrous paths.

4.2 Background

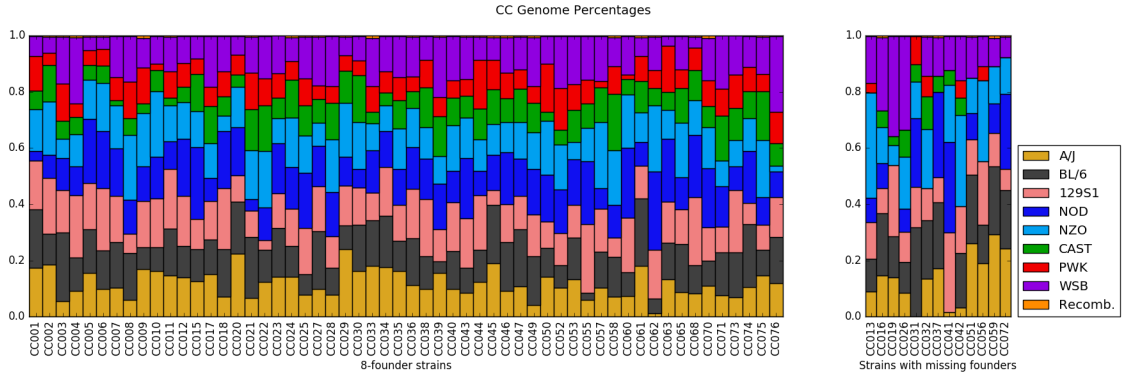
Resources such as the Collaborative Cross (CC) present unique opportunities and challenges for incorporating genetic similarity into statistical models for phenotypes of interest. For example, the CC project recombines the genomes of eight founder strains in a novel way for each newly created recombinant inbred (RI) strain by means of a funnel breeding scheme. Although the RI strains

are entirely new, they remain mosaics of the eight founder strains: A/J, C57BL/6J, 129S1/SvImJ, NOD/SHiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/Eij. Further, we will make use of the fact that the autosomal DNA of these founders represent three mouse subspecies, as follows:

1. Mus Domesticus: includes A/J, C57BL/6J, 129S1/SvImJ, NOD/SHiLtJ, NZO/HiLtJ, and WSB/Eij
2. Mus Musculus: PWK/PhJ
3. Mus Castaneus: CAST/EiJ

In the models discussed in Chapters 2 and 3 above, no genetic information was employed other than the implicit assumption that all parental RI lines are equally distant from each other in terms of genetic relatedness. This is true only in expectation. That is, although each founder strain contributes, on average, 12.5% to the genome of each recombinant strain, there is substantial variability in these contributions. Not only are the founder strains not equally represented, but some of the founder strains are (surprisingly) not represented at all in some of the RI strains. This is illustrated in Figure 4.18. This phenomenon necessitates an extension to the models discussed thus far to incorporate the

Figure 4.18: Founder contributions to CC RI strains



genetic similarity of the CC strains.

Another important question is whether phenotypes of interest in RIX samples are associated with the amount of genetic material from particular founders. For example, the New Zealand Obese (NZO) strain was bred to be overweight to serve as a model for the study of type II diabetes. A significant association between the proportion of NZO and body weight in study samples would provide evidence of the highly polygenic nature of body weight. In proposing this model that

incorporates the relative proportions of the founder strains, we introduce a sum constraint (these proportions always sum to one). To address this constraint properly, and thus allow for the testing of meaningful hypotheses of interest, we incorporate some ideas from the field of compositional data analysis.

A somewhat comprehensive review of compositional data analysis was included in Chapter 1 and is not repeated here. However, it is worth stressing again that, without accounting for the linear constraint, model parameters do not correspond to hypotheses of interest, stemming from the fact that it is impossible to alter one proportion without simultaneously altering at least one of the other proportions. More formally, a contrast in a statistical model has a marginal interpretation. That is, in a (generalized) linear model, a monotonic function of the expected value of the response is assumed to be equal to the linear predictor. Thus, a parameter estimate (or a linear combination thereof) can be interpreted as the derivative of the linear predictor with respect to a variable (or a linear combination of variables) of interest. We have established that compositional data is *scale invariant*. Another way of putting this is that it is *homogeneous of degree zero*. A function $f(\cdot)$ is homogoneous of degree k if $f(\alpha \boldsymbol{x}) = \alpha^k f(\boldsymbol{x})$. Applying Euler's theorem for homogeneous functions, if $f(\cdot)$ is scale invariant and differentiable at x, then:

$$\sum_{j=1}^{n} x_j \frac{df}{dx_j}(x) = 0. \tag{4.53}$$

In words, the constraint on the domain redounds to the derivatives. Thus, the traditional interpretations for the contrasts no longer apply. Most solutions for dealing with compositional data involve a series of log transformations which remove the constant-sum constraint by moving the data from the simplex to real space (Aitchison and Bacon-shone, 1984). The most promising solution is to express the compositional covariates in an orthonormal basis on the simplex (with respect to the Aitchison geometry). This is achieved by the following *isometric logratio transformation (ilr)*:

$$z_i = \sqrt{\frac{D-i}{D-i+1}} \ln \frac{x_i}{\sqrt[D-i]{\prod_{j=i+1}^{D} x_j}}, \quad i = 1, \ldots, D-1 \tag{4.54}$$

Having performed this transformation, the variable z_1 is now of scientific interest. It contains all the information about the compositional part x_1 in relation to the ratios between it and all other components of the mixture. Note that the same interpretation does not apply to z_2, as x_1

does not appear in this term. The solution is to construct D different ilr transformations, where the role of z_1 is assumed, in turn, by each component. That is, to investigate the significance of the kth compositional part, we replace $(x_1, \ldots, x_D)$ with $(x_k, x_1, \ldots, x_{k-1}, x_{d+1}, \ldots, x_D) =: (x_1^k, \ldots, x_k^k, x_{k+1}^k, \ldots, x_D^k)$. Noting that the *ilr* transformations are all orthogonal transformations of each other, there is no need to refit the model D times, only to perform the D linear transformations and associated hypothesis tests.

As previously mentioned, much of our knowledge on inference involving variance components has been developed in the context of longitudinal data analyis. There, it is common knowledge that, when testing the significance of variance components (for example, when performing model selection on the covariance structure) we hold the fixed effects constant and fit the models by restricted maximum likelihood (REML). This corrects the well-known downward bias in estimating the covariance parameters by maximum likelihood. Furthermore, this yields valid restricted maximum likelihood tests. When testing the fixed effects, however, models are fit (and likelihood ratio tests performed) by ML, as the distributional results are not valid when comparing models with different mean structures. It would therefore seem that, for joint tests of fixed and random effects, models should be fit using ML. However, here we demonstrate an example where ML tends to yield inconsistent parameter estimates. This compromises the validity of inference, and necessitates an alternative approach to the LRT. Since score tests do not require fitting the alternative model, they have the potential to solve this problem. Our method involves fitting a simple linear model with all the fixed effects and combining the p-values from score tests for fixed effects with the p-values from the SKAT score statistic (Wu et al., 2011) for the variance component using Fisher's method. In what follows, we refer to our method as simply the "composite score" method.

4.3 Method

We begin with a model of the following form:

$$Y = X_0\beta_0 + \sum_{m=1}^{n_m} x_m\alpha_m + Z\gamma + \epsilon \tag{4.55}$$

where Y is an $n \times 1$ vector of (continuous) phenotypes and X_0 is a matrix that includes an intercept and nuisance fixed covariates, for which β_0 are the associated (nuisance) fixed effects. Each x_m is

an $n \times 1$ column, the elements of which correspond to a proportion. If the effects of the founder proportions are of interest, then $n_m = 8$ and $x_m[i]$ corresponds to the proportion of founder m in the genome of subject i. If the contributions of subspecies DNA are of interest, then $n_m = 3$ (musculus, castaneus, and domesticus) and $p_m[i]$ would then correspond to the proportion of subspecies m in sample i. The associated α_m are the effects of founder (or subspecies) m. Applying an *ilr* transformation (4.54) to $X = (x_1, \ldots, x_m)$ we can then fit the following model:

$$Y = X_0\beta_0 + \sum_{m=1}^{n_m-1} z_m\alpha_m^* + Z\gamma + \epsilon \tag{4.56}$$

where the variable z_1 (and only z_1) is now of scientific interest. It contains all the information about the compositional part x_1 in relation to the ratios between it and all other components of the mixture. Note that the same interpretation does not pertain for z_2, as x_1 does not appear in this term. As mentioned above, we can construct D different ilr tranformations, and in each case interpret α_1^*, which can be interpreted as the effect of a unit increase in that component, while maintaining the log ratios of the other components constant.

If Z in models (4.55) and (4.56) above were an actual genotype matrix, it would be of extremely high-dimension (tens or even hundreds of thousands). Instead, we start by writing the variance of a LMM more generally:

$$Var(Y) = K\sigma_b^2 + I\sigma_\epsilon^2 \tag{4.57}$$

where K is interpreted as a "genetic relationship matrix" (GRM). The measure of similarity can be, for example, as in the popular Genome-Wide Complex Trait Analysis (GCTA) (Yang et al., 2011), the GRM where:

$$K_{jk} = \sum_{i=1}^{N} \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)} \tag{4.58}$$

where x_{ij} is the number of copies of the reference allele (i indexes SNP and j indexes individual), and p_i is the frequency of the reference allele. Alternatively, the kinship coefficient matrix (i.e., the expected proportion of the genome shared IBD) can be used. After computing the similarity matrix of one's choice, an eigendecomposition can be performed, and rather than (4.56) above, the following model is instead fit:

$$Y = X_0\beta_0 + \sum_{m=1}^{n_m-1} z_m\alpha_m^* + K^{\frac{1}{2}}\gamma + \epsilon \tag{4.59}$$

As described in GCTA, the model is typically fit by REML. However, in our setting, we are interested in testing the following hypotheses:

$$H_0 \quad : \quad \alpha_1^* = \alpha_2^* = \cdots = \alpha_{n_m-1}^* = 0, \;\; \sigma_b^2 = 0, \;\; vs. \tag{4.60}$$

$$H_A \quad : \quad at\ least\ one\ departure\ from\ null \tag{4.61}$$

As our hypotheses involves both fixed and random effects, a restricted LRT (rLRT) would not be valid (in the sense that the well-known asymptotic reference distribution would be incorrect). However, we can demonstrate an even more fundamental problem, stemming from the fact that GCTA was designed for human data. Although it theoretically applies to our mouse data as well, the samples are far more closely related than human samples. As a result, as the samples become more closely related, K looks more and more like a compound symmetric matrix. Because the entire dataset can be thought of as a single "subject", this becomes like estimating a compound symmetric matrix with a sample of size one - an impossibility. Approaching this scenario causes problems for both ML and REML. Specifically, the ML estimates for the variance component are enriched with zeros and the REML estimates are highly biased. For intuition, we present in Appendix E an identifiable model for which the ML estimate of the variance component can be shown mathematically to be inconsistent and *always* estimated at zero (and for which the REML estimate is highly biased upwards).

In a very recent paper, Su et al (Su et al., 2018) used a composite score statistic to jointly test both fixed and random effects, which employed the SKAT statistic for the random portion (Wu et al., 2011). In this vein, we start with the asymptotic score as follows:

$$U = \begin{bmatrix} U_{\beta_t} \\ U_{\sigma_b^2} \end{bmatrix} = \begin{bmatrix} (Y - X_0\hat{\beta}_0)'\hat{V}_0^{-1}X_t \\ \frac{1}{2}(Y - X_0\hat{\beta}_0)'\hat{V}_0^{-1}K\hat{V}_0^{-1}(Y - X_0\hat{\beta}_0) - tr(K^{\frac{1}{2}}\hat{V}_0^{-1}K^{\frac{1}{2}}) \end{bmatrix} \tag{4.62}$$

where $\hat{\beta}_0 = (X_0^{'}X_0)^{-1}X_0^{'}Y$ and $\hat{V}_0 = \hat{\sigma}_\epsilon^2 I_{n\times n}$. Further,

$$\mathscr{I}_{\boldsymbol{\beta}} = X'\hat{V}_0^{-1}X \tag{4.63}$$

$$\mathscr{I}_{\sigma_b^2} = \frac{1}{2}tr(K^{\frac{1}{2}}\hat{V}_0^{-1}K\hat{V}_0^{-1}K^{\frac{1}{2}}) \tag{4.64}$$

The asymptotic score statistic is $U'\mathscr{I}^{-1}U$ for appropriately defined $\mathscr{I}$. As we will show below, it is useful to employ the quadratic form portion as an alternative score statistic for the variance component (SKAT):

$$U^*_{\sigma_b^2} = (Y - X_0\hat{\beta}_0)'K(Y - X_0\hat{\beta}_0) \tag{4.65}$$

for which the p-value can be computed analytically as in Chapter 3, using the method of Davies (Davies, 1980b,a). This p-value can then be combined with p-value associated with the fixed effects using Fisher's method. We note that Su's approach involved fitting a modified null where all fixed effects were left unconstrained to compute $U^*_{\sigma_b^2}$; we do not employ that approach here. Instead, we use the same (restricted) null to compute both the score statistics.

4.4 Results

4.4.1 Simulation Results

Beyond the implications of testing fixed and random effects jointly, modeling the compositional effects correctly really only has implications for interpretability. Thus, in the interest of simplicity, the following simulations include only a single random term (and null fixed effects, one of which is an intercept). That is, we consider the following model:

$$Y = \beta_0 * J_{n\times 1} + \beta_1 X_1 + K^{\frac{1}{2}}b + \epsilon \tag{4.66}$$

$$b \sim N(0, \sigma_b^2) \tag{4.67}$$

$$\epsilon \sim N(0, \sigma_\epsilon^2). \tag{4.68}$$

In all cases, the sample size was set to 200, and β_0 and σ_ϵ^2 were both set to 1. We manually constructed the similarity matrix, K, to investigate the following four scenarios:

1. Scenario 1: Off diagonal elements of K are sampled from a truncated normal distribution with mean 0.125, standard deviation of 0.05, a minimum of zero, and a maximum of 0.3. This scenario was meant to mimic the IBD matrix observed in our RI data.

2. Scenario 2: Off diagonal elements of K are sampled from a truncated normal distribution with mean 0.725, standard deviation of 0.01, a minimum of 0.7, and a maximum of 0.75. This scenario was meant to mimic the IBS matrix observed in our RI data.

3. Scenario 3: Off diagonal elements of K are sampled from a truncated normal distribution with mean 0.7, standard deviation of 0.025, a minimum of 0.65, and a maximum of 0.75.

4. Scenario 4: Off diagonal elements of K are sampled from a truncated normal distribution with mean 0.7, standard deviation of 0.05, a minimum of 0.6, and a maximum of 0.8.

In addition, we included a "reference scenario" involving a sparse design matrix. In this scenario $K^{\frac{1}{2}}$ was simply set to a random design matrix associated with a random intercept model (20 levels, 10 subjects per level). In all other scenarios, the similarity matrices were forced to be symmetric and positive definite. In each scenario, we incremented the values of $\beta_1 \in \{0.0, 0.1, 0.5, 1.0\}$ and $\sigma_b^2 \in \{0.0, 0.2\}$ and computed the type I error and power over 2500 simulations.

The results for the reference scenario are shown in Table 4.10.These results are included to demonstrate two features, one of which was discussed in the previous chapter. For this sparse random design matrix, all methods perform as expected. The LRT based on maximum likelihood and the asymptotic score tests are nearly identical, with the composite score test conferring no benefit above the asymptotic test. Note also the deficienty of the REML-based LRT. It has extremely conservative type I error, which translated to markedly reduced power. For all other scenarios, the REML-based ML was removed.

For scenarios 1-4, the performance of the different approaches diverge wildly. The results for scenario 1 (meant to mimic using the IBD matrix for 200 unique RI samples) are shown in Table 4.11. As shown, the LRT performance is much better than the score. Scenario 2, the results of which are shown in 4.12 is meant to mimic use of the IBS matrix. In this setting, no method performs well. There is simply no power to detect the variance component. The power jumps discretely when β_1 is incremented, but within each set of simulations ($\beta_0 = 0.0,\ \beta_0 = 0.20$), the power does not increase

Table 4.10: Reference scenario results

β	σ_b^2	LRT_{ML}	LRT_{REML}	$Score$	$Score_{composite}$
0.0	0.0	0.042	0.009	0.038	0.046
0.0	0.1	0.514	0.312	0.539	0.561
0.0	0.5	0.996	0.992	0.997	0.997
0.0	1.0	1.000	1.000	1.000	1.000
0.2	0.0	0.736	0.470	0.690	0.697
0.2	0.1	0.884	0.745	0.867	0.903
0.2	0.5	1.000	0.999	1.000	1.000
0.2	1.0	1.000	1.000	1.000	1.000

Simulation based on OWL design with 10 mice per cross (n=200), $\sigma_\epsilon^2 = 1$

Table 4.11: Scenario 1 results

β	σ_b^2	LRT_{ML}	$Score$	$Score_{composite}$
0.0	0.0	0.037	0.030	0.032
0.0	0.1	0.036	0.021	0.031
0.0	0.5	0.366	0.075	0.160
0.0	1.0	0.835	0.183	0.451
0.2	0.0	0.700	0.649	0.678
0.2	0.1	0.728	0.713	0.712
0.2	0.5	0.762	0.761	0.756
0.2	1.0	0.939	0.786	0.840

Simulation based on 200 unique samples, $\sigma_\epsilon^2 = 1$

with σ_b^2. The tightness of the correlation structure is causing all the methods to fail. Scenarios 3 and 4, shown in Tables 4.13 and 4.14, respectively, are of key interest. These scenarios decrease the similarity relative to scenario 3, and increase the spread in the elements of K. As shown, for these simulations, the composite score has the ability to resolve the lack of power, detecting the variance component when the asymptotic tests fail.

4.4.2 Real Data Analysis

Table 4.15 illustrates the results of an analysis of an actual dataset comprised of 840 RIX mice which were included in a project with UNC's Center for Integrated Systems Genomics (CISGen). The phenotype was body weight, which was log transformed before analysis. The samples in this study represent 72 RIX crosses, made from 63 unique parental strains (most parental strains were used multiple times). As shown, there is a highly significant polygenic genetic effect. The most significant compositional component was NZO, which was expected as this strain was in fact bred to

Table 4.12: Scenario 2 results

β	σ_b^2	LRT_{ML}	$Score$	$Score_{composite}$
0.0	0.0	0.024	0.020	0.018
0.0	0.1	0.022	0.023	0.024
0.0	0.5	0.026	0.032	0.028
0.0	1.0	0.016	0.046	0.020
0.2	0.0	0.714	0.689	0.686
0.2	0.1	0.699	0.679	0.672
0.2	0.5	0.644	0.720	0.656
0.2	1.0	0.581	0.719	0.615

Simulation based on 200 unique samples, $\sigma_\epsilon^2 = 1$

Table 4.13: Scenario 3 results

β	σ_b^2	LRT_{ML}	$Score$	$Score_{composite}$
0.0	0.0	0.025	0.019	0.028
0.0	0.1	0.023	0.022	0.031
0.0	0.5	0.020	0.036	0.086
0.0	1.0	0.021	0.065	0.201
0.2	0.0	0.710	0.684	0.692
0.2	0.1	0.678	0.686	0.676
0.2	0.5	0.626	0.707	0.720
0.2	1.0	0.617	0.763	0.792

Simulation based on 200 unique samples, $\sigma_\epsilon^2 = 1$

Table 4.14: Scenario 4 results

β	σ_b^2	LRT_{ML}	$Score$	$Score_{composite}$
0.0	0.0	0.023	0.020	0.034
0.0	0.1	0.027	0.027	0.061
0.0	0.5	0.029	0.037	0.218
0.0	1.0	0.016	0.058	0.535
0.2	0.0	0.715	0.678	0.691
0.2	0.1	0.685	0.675	0.699
0.2	0.5	0.641	0.722	0.800
0.2	1.0	0.609	0.732	0.890

Simulation based on 200 unique samples, $\sigma_\epsilon^2 = 1$

Table 4.15: Real compositional data analysis

Effect	Estimate	p-value
Overall Genetic Effect		<0.0001
Intercept	3.224	<0.0001
Sex (female)	-0.250	<0.0001
129	-0.017	0.4371
A	-0.002	0.9053
C57BL	-0.052	0.0133
CAST	-0.049	0.0010
NOD	-0.004	0.746
NZO	0.115	<0.0001
PWK	0.027	0.1048
WSB	-0.018	0.3928

be obese. Increases in body weight associated with increasing contributions of NZO DNA are most offset by contributions of C57BL and CAST, which are negatively assocated with body weight.

4.4.3 Discussion

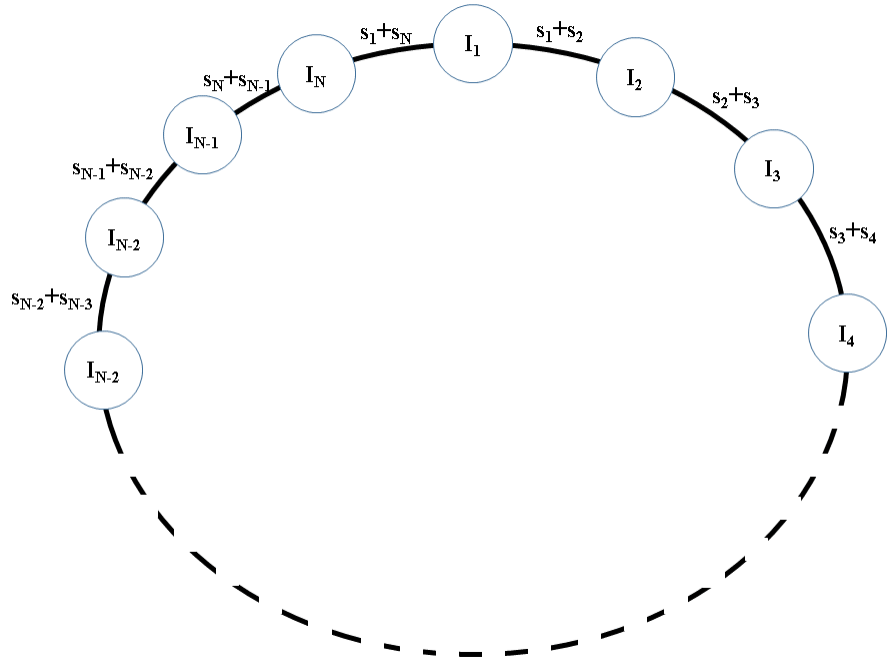
In this chapter, we have demonstrated a deficiency with the standard (ML) LRT hypothesis testing framework when testing fixed and random effects jointly. While in ideal scenarios, the LRT performs as expected, it also has the potential to fail catastrophically. The composite score approach has the ability to resolve this deficiency in certain settings. The relative performance of these methods is a function of the particular similarity matrix employed. Because this is known a priori, in future work we hope to develop formal conditions to guide the inferential approach.

We note that it is also possible for a REML-based LRT to resolve the problem. In fact, REML estimation resolves the problem of inconsistent estimation that we demonstrated in the defective scenario in Appendix E. However, an adjustment would have to be made because the distributional results that apply to the ML-based LRT do not apply to the REML-based LRT. We explored a bootstrap based on an extension of the work of Sinha (Sinha, 2009), but we did not include the results here because it did not control type I error. This also remains a topic for future research.

APPENDIX A: COMMON LOOP DESIGNS

In the figure below, each node represents an inbred stain, and each of the edges connecting these nodes corresponds to a cross. Mice of these crosses are the units of study. Note that the crosses are not directional because we here assume an additive model. As shown in the figure, for each cross, we can obviously estimate the additive contribution of the two parental strains (labeled $s_i + s_j$ in the figure). Suppose we wish to obtain strain-specific estimates; that is, for example, we wish

Figure 7.19: Additive loop design



to obtain separate estimates for s_1 and s_2. Using the terminology of Sprague and Tatum (1942) we are estimating the generalized combining ability (GCA) of the parental strains. We can show that this is only possible if the loop is composed of an odd number of parental strains. This is a direct consequence of the fact that the parents come in pairs - in order to subtract out the effect of one parent, we introduce the effect of another. This is more easily seen arithmetically. We have an estimate for $s_1 + s_2$, so we therefore must obtain an estimate of $s_1 - s_2$, so that then taking half the sum of these estimators yields an estimate of s_1. We do this starting with the estimate of $s_1 + s_N$ and traversing the loop in the counter-clockwise direction. That is, since we want to eliminate s_N, we

must subtract our estimate of $s_N + s_{N-1}$, arriving at an estimate of $s_1 - s_{N-1}$. To eliminate s_{N-1} we must add the estimate of $s_{N-1} + s_{N-2}$. We propagate around the loop until we are left with an estimate involving only s_1 and s_2. Writing out the arithmetic:

$$\begin{aligned} s_1 \quad & + \quad s_N \\ & - \quad (s_N + s_{N-1}) \\ & + \quad (s_{N-1} + s_{N-2}) \\ & - \quad (s_{N-2} + s_{N-3}) \\ & \quad \vdots \\ & + \quad (s_4 + s_3) \\ & - \quad (s_3 + s_2) \end{aligned}$$

In order for our estimate to be one of $s_1 - s_2$, as needed, the last operation must be subtraction, as above. Thus, integers 2 through N can be paired, and N must therefore be odd. We can therefore make the following claims:

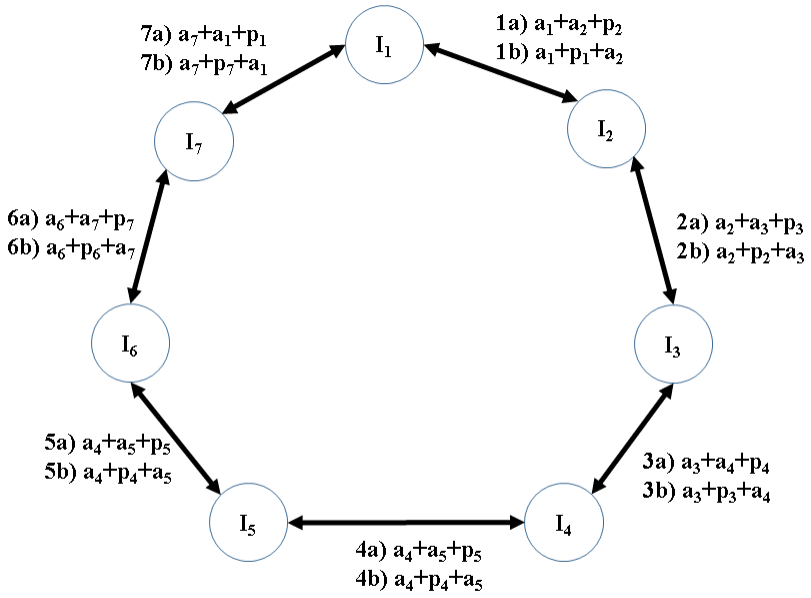
Claim 1: For a loop design with an even number of parental strains (N), we can estimate $a_i + a_{i+j}$ $\forall i$ and for odd j where $i, j \in 1, ..., N$.

Claim 2: For a loop design with an even number of parental strains (N), we can estimate $a_i - a_{i+j}$ $\forall i$ and for even j where $i, j \in 1, ..., N$.

Reciprocal Loop

In the figure below, we show the design for the reciprocal loop; we have crosses in both directions. Note that I have now labeled the additive effect with an $a_.$, instead of $s_.$, as above. Further, $p_.$ corresponds to a parent-of-origin (PoO) effect. Utilizing both directions of a cross, we can estimate:

Figure 7.20: Reciprocal cross loop design



$$(1a) a_1 + a_2 + p_1$$

$$(1b) a_2 + a_1 + p_2$$

Since we can do this for each pair of reciprocal crosses, we can take differences and estimate all the differences $p_i - p_j$.

Next, consider traversing the loop in the clockwise direction. We can estimate the following:

$$(i)a_1 + a_2 + p_2$$

$$(ii)a_2 + a_3 + p_3$$

$$(iii)a_3 + a_4 + p_4$$

$$(iv)a_4 + a_5 + p_5$$

$$(v)a_5 + a_6 + p_6$$

$$(vi)a_6 + a_7 + p_7$$

$$(vii)a_7 + a_1 + p_1$$

Note that:

$$(i) + (ii) + (iii) + (iv) - (v) + (vi) - (vii) = 2a_1 + p_2 - p_3 + p_4 - p_5 + p_6 - p_7 + p_1$$

Pairing relevant PoO terms and subtracting their estimates, and continuing this logic to other strains,we can estimate:

$$2a) : 2a_1 + p_1$$

$$2b) : 2a_2 + p_2$$

$$2c) : 2a_3 + p_3$$

$$2d) : 2a_4 + p_4$$

$$2e) : 2a_5 + p_5$$

$$2f) : 2a_6 + p_6$$

$$2g) : 2a_7 + p_7$$

We note that, if the above design incorporated the seven inbred strains as well, it would add nothing of value in terms of estimability if the true model has only additive and PoO effects. This is seen because if we have two inbred strains and both associated reciprocal crosses, that would add no new estimable parameters. The inbred strains would allow for estimating additional dominance effects.

Now, traversing the loop in the opposite direction we have the following readily estimable:

$$(i)a_1 + a_2 + p_1$$

$$(ii)a_2 + a_3 + p_2$$

$$(iii)a_3 + a_4 + p_3$$

$$(iv)a_4 + a_5 + p_4$$

$$(v)a_5 + a_6 + p_5$$

$$(vi)a_6 + a_7 + p_6$$

$$(vii)a_7 + a_1 + p_7$$

Note that:

$$(i) + (ii) + (iii) + (iv) - (v) + (vi) - (vii) = 2a_1 + p_1 - p_2 + p_3 - p_4 + p_5 - p_6 + p_7$$

Pairing the relevant PoO terms and substracting their estimates, we can estimate:

$$3a) : 2a_1 + p_7$$

$$3b) : 2a_2 + p_1$$

$$3c) : 2a_3 + p_2$$

$$3d) : 2a_4 + p_3$$

$$3e) : 2a_5 + p_4$$

$$3f) : 2a_6 + p_5$$

$$3g) : 2a_7 + p_6$$

Now, subtracting 2a) from 3a), 2b) from 3b), etc. we can estimate:

$$2(a_1 - a_2) + (p_1 - p_2)$$

$$2(a_2 - a_3) + (p_2 - p_3)$$

$$2(a_3 - a_4) + (p_3 - p_4)$$

$$2(a_4 - a_5) + (p_4 - p_5)$$

$$2(a_5 - a_6) + (p_5 - p_6)$$

$$2(a_6 - a_7) + (p_6 - p_7)$$

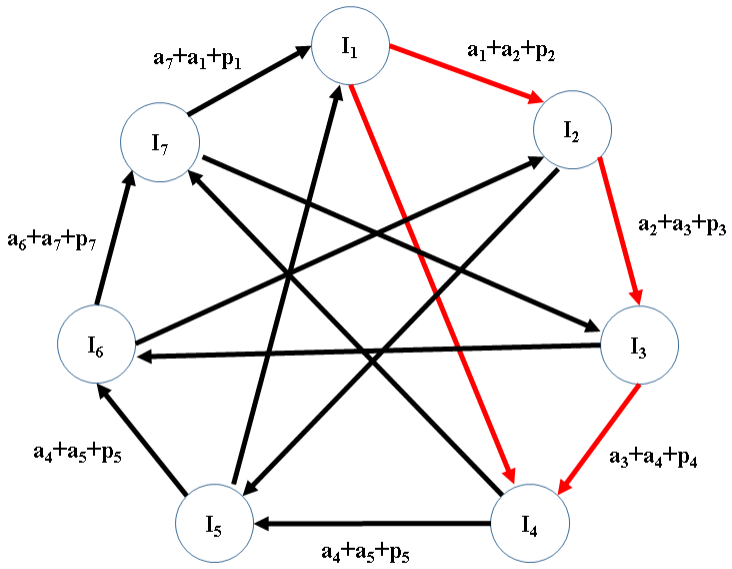
$$2(a_7 - a_1) + (p_7 - p_1)$$

Since we can estimate all pairwise differences $p_i - p_j$, we can also estimate all $a_i - a_j$. Note that it does not matter whether we have an odd or even number of parental strains in the loop.

APPENDIX B: AN ALTERNATIVE LOOP DESIGN

If the true model has only additive and PoO, we can actually estimate the same parameters as in the reciprocal loop design without employing any reciprocal crosses. Consider the design illustrated in the figure below: We have maintained the original uni-directional outer loop, but introduced many

Figure 8.21: An alternative loop design



inner loops. Note the loop highlighted in red, and the fact that it is not uni-directional. That is, two strains use I_4 as the maternal strain. We can estimate:

$$(i) : a_1 + a_2 + p_2$$

$$(i) : a_2 + a_3 + p_3$$

$$(i) : a_3 + a_4 + p_4$$

$$(i) : a_1 + a_4 + p_4$$

Subtracting (iv) from (i) we get:

$$(i) - (iv) = a_1 - a_2 + p_2 - a_1 - a_4 - p_4 = (a_2 - a_4) - (p - 2 - p_4)$$

Similarly, subtracting (iii) from (ii) yields:

$$(ii) - (iii) = a_2 + a_3 + p_3 - a_3 - a_4 - p_4 = (a_2 - a_4) + (p_3 - p_4)$$

And subtracting these, we can estimate $p_2 - p_3$. Continuing this logic, we can estimate all $p_i - p_j$.

Above, we showed we could traverse the outer loop in one direction to estimate:

$$2(a_1 - a_2) + (p_1 - p_2)$$

$$2(a_2 - a_3) + (p_2 - p_3)$$

$$2(a_3 - a_4) + (p_3 - p_4)$$

$$2(a_4 - a_5) + (p_4 - p_5)$$

$$2(a_5 - a_6) + (p_5 - p_6)$$

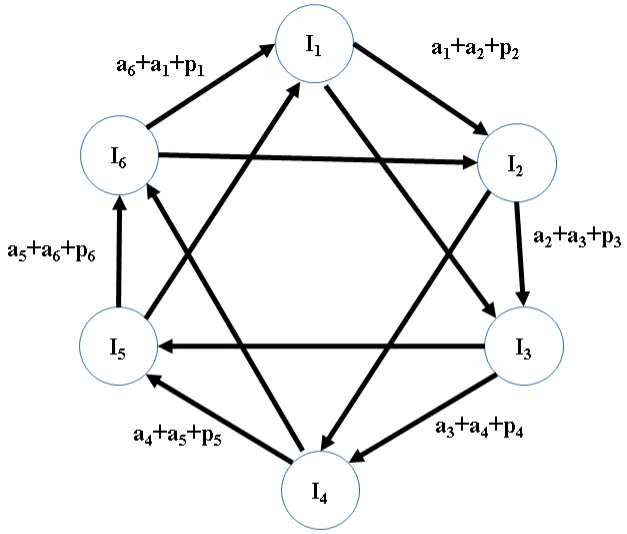
$$2(a_6 - a_7) + (p_6 - p_7)$$

$$2(a_7 - a_1) + (p_7 - p_1)$$

Thus, we can estimate all $p_i - p_j$ and $a_i - a_j$, as in the design where we had reciprocal crosses.

It can similarly shown that for an even number of parental strains, we achieve the same by using the following design shown in Figure 8.22. Note that this logic does not apply if there are dominance effects; in that case, the reciprocal design is preferable.

Figure 8.22: An alternative loop design with an even number of parental strains



APPENDIX C: FITTING MODELS IN SAS

```
proc mixed data=dataset;
    model dependent_variable = &fixed_effects. / solution;
    random batch1-batch&numbatch. / type=TOEP(1);
    random addint1-addint&numparent. / type=TOEP(1);
    random addtmt1-addtmt&numparent. / type=TOEP(1);
    random addsex1-addsex&numparent. / type=TOEP(1);
    random addsextmt1-addsextmt&numparent. / type=TOEP(1);
run;
```

Here, *TOEP(1)* refers to a 1-banded Toeplitz structure, *fixed_effects* to a list of fixed effects, and *numbatch* and *numparent* to the number of unique batches and parental strain, respectively.

APPENDIX D: LINK BETWEEN RIDGE REGRESSION AND MIXED MODELS

Consider the simplest mixed model with a single random term:

$$\begin{aligned} Y &= X\beta + Z\gamma + \epsilon \\ \gamma &\sim N(0, \sigma^2_\gamma) \\ \epsilon &\sim N(0, \sigma^2_\epsilon) \end{aligned}$$

Further, consider the ridge regression estimates for γ which we denote $\hat{\gamma}_{RR} = (Z'Z + \lambda I)^{-1}Z'y$. We ignore X for now, and start by applying the Sherman-Morrison-Woodbury formula (Woodbury, 1950):

$$\begin{aligned} \hat{\gamma}_{RR} &= \{(\lambda I)^{-1} - [(\lambda I)^{-1}Z'(I + Z(\lambda I)^{-1}Z')^{-1}Z(\lambda I)^{-1}]\}Z'y \\ &= \frac{1}{\lambda}\{I - [Z'(\lambda I + ZZ')^{-1}Z]\}Z'\lambda \\ &= \frac{1}{\lambda}\{Z' - Z'(\lambda I + ZZ')^{-1}ZZ'\}y \\ &= \frac{1}{\lambda}Z'[I - (Z'Z + \lambda I)^{-1}ZZ']y \\ &= \frac{1}{\lambda}Z'(Z'Z + \lambda I)^{-1}[(ZZ' + \lambda I) - ZZ']y \\ &= Z'(ZZ' + \lambda I)^{-1}y \end{aligned}$$

Next, consider the prediction of the BLUPs. To predict the BLUPs, we start by specifying the joint distribution of y and γ, and then use the formula for conditional expectation to obtain the estimates:

$$\begin{bmatrix} \mathbf{y} \\ \gamma \end{bmatrix} \sim N\Big(\begin{bmatrix} X\beta \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_\gamma ZZ' + \sigma^2_\epsilon I & \sigma^2_\gamma Z \\ \sigma^2_\gamma Z' & \sigma^2_\gamma I \end{bmatrix}\Big) \tag{6.69}$$

Now, we derive the BLUPs:

$$\begin{aligned} E[\gamma|y] &= \sigma^2_\gamma X'(\sigma^2_\gamma ZZ' + \sigma_\epsilon I)^{-1}y \\ &= Z'\Big(ZZ' + \frac{\sigma^2_\epsilon}{\sigma^2_\gamma}I\Big)^{-1}y \end{aligned}$$

So, we see that the BLUPs can be thought of as the estimates from ridge regression with $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\gamma^2}$. We know from ridge regression that there is a unique solution for all $\lambda > 0$.

APPENDIX E: INCONSISTENT ML ESTIMATES OF VARIANCE COMPONENTS

We consider the most extreme example of "confounding" between the fixed and the random effects, such that the fixed effects design matrix exactly equals the random effects design matrix. Consider a simple clustered model where there are n independent clusters with m subjects per cluster. We can specify the model as follows:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{z}_i b_i + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{Y}_i$ is an $m \times 1$ column vector, β is an $m \times 1$ parameter vector, and $\boldsymbol{X}_i$ is an $m \times m$ matrix of zeros with ones in column i. Further, $\boldsymbol{z}_i = \boldsymbol{J}_{m\times 1}$. At first glance, this model might seem unidentifiable, but it is in fact identifiable. The fixed effects matrix, $\boldsymbol{X}$, is of full rank, and so the fixed effects are identifiable. The covariance parameters are also identifiable in the sense detailed in Chapter 2, where we need not have bothered with the fixed effects (other than to implicitly assume the fixed effects matrix was full rank). Having established that the fixed effects and covariance terms are, when viewed separatedly, identifiable, the only way to have an identifiability issue would involve an interplay between the fixed and the random terms. But this cannot be the case - in order for that to occur, we would be saying that two normal distribution (μ_1, Σ_1) and (μ_2, Σ_2) have identical distributions, which is impossible. We define the following:

$$\begin{aligned}
\Sigma_i &= \sigma_e^2 I + Z_i G Z_i' \\
\Sigma_i &= \sigma_e^2 D_i \\
D_i &= I + Z_i G^* Z_i'
\end{aligned}$$

We will use the following result:

$$|A + XX'| = |A|(I + X'A^{-1}X)$$

We write the loglikelihood as:

$$l = -\frac{1}{2}nm\log(\sigma_e^2) - \frac{1}{2}n\log|\boldsymbol{D}_i| - \frac{1}{2\sigma_e^2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)$$

Conditional on σ_e^2 and $\boldsymbol{D}_i$, we estimate $\boldsymbol{\beta}$ by GLS:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

Substituting $\hat{\boldsymbol{\beta}}$ into the loglikelihood, we find it is proportional to the following:

$$l \propto -nm\log(\sigma_e^2) - n\log|\boldsymbol{D}_i| - \frac{1}{\sigma_e^2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)$$

Using (2) above and differentiating with respect to σ_e^2 we can solve for the MLE:

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)}{nm}$$

and substituting these into the loglikelihood, the relevant portion (that involving $\boldsymbol{D}_i$ is):

$$l(\boldsymbol{D}_i) = -n\log|\boldsymbol{D}_i| - \hat{\sigma}^{-2}\frac{\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)}{nm}$$

Now, differentiating with respect to $\boldsymbol{G}^*$:

$$\frac{dl(\boldsymbol{D}_i)}{d\boldsymbol{G}^*} = -n\hat{\sigma}^{-2}\boldsymbol{Z}_i'\boldsymbol{D}_i^{-1}\boldsymbol{Z}_i - \hat{\sigma}^{-4}\sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}\boldsymbol{Z}_i$$

Now,

$$\frac{dl(\boldsymbol{D}_i)}{d\boldsymbol{G}^*} < 0$$

$$\iff \quad -n\hat{\sigma}^{-2}\boldsymbol{Z}_i'\boldsymbol{D}_i^{-1}\boldsymbol{Z}_i < \hat{\sigma}^{-4}\sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'\boldsymbol{D}_i^{-1}\boldsymbol{Z}_i$$

But since, in our setting $\boldsymbol{Z} = \boldsymbol{X}$, the right hand side of the inequality is always zero (since $(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)$ is in the nullity of $\boldsymbol{D}_i^{-1}\boldsymbol{Z}_i$), making the inequality always true. Therefore, $\hat{\boldsymbol{G}}_i^*$ will always be set to zero.

BIBLIOGRAPHY

Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Macmillan&Company.

Aitchison, J. and Bacon-shone, J. (1984). Log contrast models for experiments with mixtures. *Biometrika*, 71(2):323–330.

Arnold, S. E., Han, L.-Y., Moberg, P. J., Turetsky, B. I., Gur, R. E., Trojanowski, J. Q., and Hahn, C.-G. (2001). Dysregulation of olfactory receptor neuron lineage in schizophrenia. *Archives of general psychiatry*, 58(9):829–835.

Arya, A. (1983). Circulant plans for partial diallel crosses. *Biometrics*, pages 43–52.

Bakker, P. R., van Harten, P. N., and van Os, J. (2006). Antipsychotic-induced tardive dyskinesia and the ser9gly polymorphism in the drd3 gene: a meta analysis. *Schizophrenia research*, 83(2):185–192.

Braaten, M. O. (1965). The union of partial diallel mating designs and incomplete block environmental designs. Technical report, North Carolina State University. Dept. of Statistics.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Chen, H., Meigs, J. B., and Dupuis, J. (2013). Sequence kernel association test for quantitative traits in family samples. *Genetic epidemiology*, 37(2):196–204.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, pages 573–578.

Clatworthy, W. H. (1955). Partially balanced incomplete block designs with two associate classes and two treatments per block. *J. Res. Nat. Bur. Standards*, 54:177–190.

Cockerham, C. C. and Weir, B. (1977). Quadratic analyses of reciprocal crosses. *Biometrics*, pages 187–203.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):165–185.

Curnow, R. (1963). Sampling the diallel cross. *Biometrics*, 19(2):287–306.

Das, A., Dean, A. M., and Gupta, S. (1998). On optimality of some partial diallel cross designs. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 511–524.

Davies, R. B. (1980a). Algorithm as 155: The distribution of a linear combination of χ 2 random variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3):323–333.

Davies, R. B. (1980b). The distribution of a linear combination of x2 random variables. *Applied Statistics*, 29(3):323–333.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons.

Drton, M. (2009). Likelihood ratio tests and singularities. *The Annals of Statistics*, pages 979–1012.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, 3(1):1–21.

Fisher, R. A. (1919). Xv.the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(02):399–433.

Fitzmaurice, G. M., Lipsitz, S. R., and Ibrahim, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics*, 63(3):942–946.

Follenzi, A., Battaglia, M., Lombardo, A., Annoni, A., Roncarolo, M. G., and Naldini, L. (2004). Targeting lentiviral vector expression to hepatocytes limits transgene-specific immune response and establishes long-term expression of human antihemophilic factor ix in mice. *Blood*, 103(10):3700–3709.

Fyfe, J. and Gilbert, N. (1963). Partial diallel crosses. *Biometrics*, 19(2):278–286.

Ghosh, D. and Divecha, J. (1997). Two associate class partially balanced incomplete block designs and partial diallel crosses. *Biometrika*, 84(1):245–248.

Gilbert, N. (1958). Diallel cross in plant breeding. *Heredity*, 12(3):477–492.

Greenberg, A. J., Hackett, S. R., Harshman, L. G., and Clark, A. G. (2010). A hierarchical bayesian model for a novel sparse partial diallel crossing design. *Genetics*, 185(1):361–373.

Greven, S., Crainiceanu, C. M., Küchenhoff, H., and Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17(4):870–891.

Griffing, B. (1956a). Concept of general and specific combining ability in relation to diallel crossing systems. *Australian journal of biological sciences*, 9(4):463–493.

Griffing, B. (1956b). A generalized treatment of the use of diallel crosses in quantitative inheritance. *Heredity*, 10(1):31–50.

Gupta, S., Das, A., and Kageyama, S. (1995). Single replicate orthogonal block designs for circulant partial diallel crosses. *Communications in Statistics-Theory and Methods*, 24(10):2601–2607.

Han, S. S. and Chang, J. T. (2009). Reconsidering the asymptotic null distribution of likelihood ratio tests for genetic linkage in multivariate variance components models under complete pleiotropy. *Biostatistics*, page kxp054.

Harrison, P. J. (1999). The neuropathology of schizophrenia. *Brain*, 122(4):593–624.

Hartley, H. O. and Rao, J. N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54(1-2):93–108.

Hayman, B. (1954a). The analysis of variance of diallel tables. *Biometrics*, 10(2):235–244.

Hayman, B. (1954b). The theory and analysis of diallel crosses. *Genetics*, 39(6):789.

Hayman, B. (1958). The theory and analysis of diallel crosses. ii. *Genetics*, 43(1):63.

Hayman, B. (1960). The theory and analysis of diallel crosses. iii. *Genetics*, 45(2):155.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9(2):226–252.

Hinkelmann, K. and Kempthorne, O. (1963). Two classes of group divisible partial diallel crosses. *Biometrika*, 50(3/4):281–291.

Hron, K., Filzmoser, P., and Thompson, K. (2012). Linear regression with compositional explanatory variables. *Journal of Applied Statistics*, 39(5):1115–1128.

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723.

Kantor, B., Bayer, M., Ma, H., Samulski, J., Li, C., McCown, T., and Kafri, T. (2011). Notable reduction in illegitimate integration mediated by a ppt-deleted, nonintegrating lentiviral vector. *Molecular Therapy*, 19(3):547–556.

Kempthorne, O. and Curnow, R. (1961). The partial diallel cross. *Biometrics*, 17(2):229–250.

Kinney, S. K. and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics*, 63(3):690–698.

Kopylev, L. (2012). Constrained parameters in applications: Review of issues and approaches. *ISRN Biomathematics*, 2012.

Kopylev, L. and Sinha, B. (2011). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary. *Sankhya B*, 73(1):20.

Lafaye De Micheaux, P. C. (2013). Distribution function of quadratic forms in normal variables. *R package version*, 1(1).

Lee, O. E. and Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, 68(2):486–493.

Lenarcic, A. B., Svenson, K. L., Churchill, G. A., and Valdar, W. (2012). A general bayesian approach to analyzing diallel crosses of inbred strains. *Genetics*, 190(2):413–435.

Lerer, B., Segman, R. H., Tan, E.-C., Basile, V. S., Cavallaro, R., Aschauer, H. N., Strous, R., Chong, S.-A., Heresco-Levy, U., Verga, M., et al. (2005). Combined analysis of 635 patients confirms an age-related association of the serotonin 2a receptor gene with tardive dyskinesia and specificity for the non-orofacial subtype. *International Journal of Neuropsychopharmacology*, 8(3):411–425.

Li, G. and Cui, Y. (2016). Assessing statistical significance in variance components linkage analysis: A theoretical justification. *Journal of Statistical Planning and Inference*, 178:70–83.

Li, H. (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annual Review of Statistics and Its Application*, 2:73–94.

Li, H. and Loken, E. (2002). A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statistica Sinica*, pages 519–535.

Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., Keefe, R. S., Davis, S. M., Davis, C. E., Lebowitz, B. D., et al. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, 353(12):1209–1223.

Lin, W., Shi, P., Feng, R., Li, H., et al. (2014). Variable selection in regression with compositional covariates. *Biometrika*, 101(4):785–797.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika*, pages 309–326.

Miller, J. J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical report, DTIC Document.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, pages 746–762.

Mukerjee, R. (1997). Optimal partial diallel crosses. *Biometrika*, pages 939–948.

Ofversten, J. (1993). Exact tests for variance components in unbalanced mixed linear models. *Biometrics*, pages 45–57.

Patsopoulos, N. A., Ntzani, E. E., Zintzaras, E., and Ioannidis, J. P. (2005). Cyp2d6 polymorphisms and the risk of tardive dyskinesia in schizophrenia: a meta-analysis. *Pharmacogenetics and genomics*, 15(3):151–158.

Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional data analysis: Theory and applications*. John Wiley & Sons.

Pearson, K. (1896). Mathematical contributions to the theory of evolution.–on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the royal society of london*, 60(359-367):489–498.

Pirinen, M., Donnelly, P., Spencer, C. C., et al. (2013). Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *The Annals of Applied Statistics*, 7(1):369–390.

Rao, C. R. (1971). Estimation of variance and covariance componentsminque theory. *Journal of multivariate analysis*, 1(3):257–275.

Reynolds, G. P., Templeman, L. A., and Zhang, Z. J. (2005). The role of 5-ht2c receptor polymorphisms in the pharmacogenetics of antipsychotic drug treatment. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 29(6):1021–1028.

Rogala, A. R., Morgan, A. P., Christensen, A. M., Gooch, T. J., Bell, T. A., Miller, D. R., Godfrey, V. L., and de Villena, F. P.-M. (2014). The collaborative cross as a resource for modeling human disease: Cc011/unc, a new mouse model for spontaneous colitis. *Mammalian Genome*, 25(3-4):95–108.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6):110–114.

Saville, B. R. and Herring, A. H. (2009). Testing random effects in the linear mixed model using approximate bayes factors. *Biometrics*, 65(2):369–376.

Schmidt, J. (1919). La valeur de lindividu atitre de générateur appréciée suivant la méthode du croisement diallele. *CR Trav. Lab. Carlsberg*, 14(6):1–33.

Searle, S. R. (1970). Large sample variances of maximum likelihood estimators of variance components using unbalanced data. *Biometrics*, pages 505–524.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review/Revue Internationale de Statistique*, pages 49–62.

Shorter, J. R., Maurizio, P. L., Bell, T. A., Shaw, G. D., Miller, D. R., Gooch, T. J., Spence, J. S., McMillan, L., Valdar, W., and de Villena, F. P.-M. (2018). A diallel of the mouse collaborative cross founders reveals strong strain-specific maternal effects on litter size. *bioRxiv*, page 458877.

Silvapulle, M. J. (1992). Robust wald-type tests of one-sided hypotheses in the linear model. *Journal of the American Statistical Association*, 87(417):156–161.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association*, 90(429):342–349.

Singh, M. and Hinkelmann, K. (1990). On generation of efficient partial diallel crosses plans. *Biometrical journal*, 32(2):177–187.

Singh, M. and Hinkelmann, K. (1995). Partial diallel crosses in incomplete blocks. *Biometrics*, pages 1302–1314.

Singh, M. and Hinkelmann, K. (1998). Analysis of partial diallel crosses in incomplete blocks. *Biometrical journal*, 40(2):165–181.

Sinha, B., Kopylev, L., and Fox, J. (2007). Some new aspects of dose-response multistage models with applications. In *Environmental and Ecological Statistics (accepted for publication). Platinum Jubilee Conference of ISI, World Scientific Publishing, Singapore. Earlier version is available as UMBC technical report: http://www. math. umbc. edu/˜ kogan/technical_papers/2007/Sinha_Kopylev_Fox. pdf*. Citeseer.

Sinha, S. K. (2009). Bootstrap tests for variance components in generalized linear mixed models. *Canadian Journal of Statistics*, 37(2):219–234.

Smutzer, G., Lee, V. M.-Y., Trojanowski, J. Q., and Arnold, S. E. (1998). Human olfactory mucosa in schizophrenia. *Annals of Otology, Rhinology & Laryngology*, 107(4):349–355.

Sprague, G. F. and Tatum, L. A. (1942). General vs. specific combining ability in single crosses of corn. *Agronomy Journal*, 34(10):923–932.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, pages 1171–1177.

Su, Y.-R., Di, C., Bien, S., Huang, L., Dong, X., Abecasis, G., Berndt, S., Bezieau, S., Brenner, H., Caan, B., et al. (2018). A mixed-effects model for powerful association tests in integrative functional genomics. *The American Journal of Human Genetics*, 102(5):904–919.

Thompson Jr, W. (1962). The problem of negative estimates of variance components. *The Annals of Mathematical Statistics*, pages 273–289.

Tomita, H., Vawter, M. P., Walsh, D. M., Evans, S. J., Choudary, P. V., Li, J., Overman, K. M., Atz, M. E., Myers, R. M., Jones, E. G., et al. (2004). Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. *Biological psychiatry*, 55(4):346–352.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, 59(2):254–262.

Verhoeven, K., Jannink, J., and McIntyre, L. (2006). Using mating designs to uncover qtl and the genetic architecture of complex traits. *Heredity*, 96(2):139–149.

Viana, J. M. S., Cruz, C. D., and Cardoso, A. A. (1999). Theory and analysis of partial diallel crosses. *Genetics and Molecular Biology*, 22(4):591–599.

Vu, H., Zhou, S., et al. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *The Annals of Statistics*, 25(2):897–916.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. *The Annals of Mathematical Statistics*, 11(1):96–100.

Wald, A. (1941). On the analysis of variance in case of multiple classifications with unequal class frequencies. *The Annals of Mathematical Statistics*, 12(3):346–350.

Wang, W. et al. (2013). Identifiability of linear mixed effects models. *Electronic Journal of Statistics*, 7:244–263.

Weis, S., Llenos, I., Dulay, J., Elashoff, M., Martinez-Murillo, F., and Miller, C. (2007). Quality control for microarray analysis of human brain samples: the impact of postmortem factors, rna characteristics, and histopathology. *Journal of neuroscience methods*, 165(2):198–209.

WHO (2012). Schizophrenia fact sheet. 2010. *Available at https://www.who.int/news-room/fact-sheets. Accessed January 2019*, 16.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

Wolfinger, R. and O'connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation*, 48(3-4):233–243.

Woodbury, M. A. (1950). Inverting modified matrices. *Memorandum report*, 42(106):336.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89(1):82–93.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82.

Zhu, J. and Weir, B. S. (1996a). Diallel analysis for sex-linked and maternal effects. *Theoretical and Applied Genetics*, 92(1):1–9.

Zhu, J. and Weir, B. S. (1996b). Mixed model approaches for diallel analysis based on a bio-model. *Genetical research*, 68(03):233–240.