Abstract

BARKER, CLAYTON ADAM. The Orthogonal Interactions Model for Unreplicated Factorial Experiments. (Under the direction of Leonard Stefanski and Jason Osborne.)

Unreplicated factorial experiments arise frequently in practice because of limited resources. When fitting the standard ANOVA model to data from unreplicated experiments, the data are not sufficient to estimate the interaction terms and error variance, thus limiting the possible inferences. In experiments such as crop yield trials, investigators are interested in estimating interactions but are unable to form replications. Existing methods, such as Tukey’s one-degree-of-freedom model and the AMMI model, fit constrained interactions which allow for the error variance estimation.

We present the orthogonal interactions (OI) model for unreplicated factorial experiments. The model frees degrees of freedom for error by assuming that the main effects are orthogonal to the interactions. Through simulation we find that approximate $F$-statistics are appropriate in testing for main effects and interactions. The likelihood ratio test to compare the OI model to the ANOVA model leads to high Type I error rates, so we use a simulation corrected likelihood ratio test. Two real data sets (one with replication, one without) suggest that the OI model is appropriate for real data.

In working with the OI model and the existing models, we found a need for reliable degrees of freedom for complex statistical models. The resampling method for estimating degrees of freedom is motivated by the linear model, where there is a linear relationship between expected sums of squares and the variance of errors added to the response. Through simulation, we find that the resampling method provides reliable degrees of freedom for the OI model and for existing models as well.
The Orthogonal Interactions Model For Unreplicated Factorial Experiments

by

Clayton Barker

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Statistics

Raleigh, NC

December 1, 2006

Approved by:

Dr. Leonard Stefanski (co-chair) Dr. Jason Osborne (co-chair)

Dr. David Dickey Dr. Cavell Brownie
Dedication

To Kaye and Terry Barker
Clay Barker was born on December 15, 1978 to Kaye and Terry Barker in Tampa, Florida. After living in Tampa; Myrtle Beach, South Carolina; and Atlanta, Georgia; the Barker family settled in Chapel Hill, North Carolina in January 1987. In 1997, Clay graduated from Chapel Hill High School and began his undergraduate work at North Carolina State University as an applied mathematics major. In May 2001, Clay completed his Bachelor of Science degree in applied mathematics with a minor in statistics. Clay chose to stay at NCSU to pursue a graduate degree in statistics. He received his Master of Statistics degree in 2003 and his doctoral degree in 2006 under the direction of Dr. Leonard Stefanski and Dr. Jason Osborne. Clay received the Gertrude Cox Outstanding Ph.D. candidate award in 2004. As a graduate student, Clay had the opportunity to serve as teaching assistant for undergraduate and graduate level courses, serve as a research assistant on a project with the Environmental Protection Agency, and work as a technical student at SAS.
Acknowledgements

First of all, I must thank my advisors Dr. Leonard Stefanski and Dr. Jason Osborne for their guidance and support throughout my graduate career. They provided me with a tremendous amount of knowledge while still giving me the freedom to discover things on my own. They are excellent statisticians and people, and I am fortunate to have had the opportunity to work with them. I would also like to thank my committee members for their contributions to this document. Dr. Sastry Pantula and Dr. Bill Swallow have also provided support throughout my graduate career. I must also thank Terry Byron and Adrian Blue for all that they do for our department. We are lucky to have such excellent faculty and staff in our department.

I have endured stressful days and long nights as a graduate student, but I have not had to go about it alone because of a strong support group of friends. I am fortunate to have gone through grad school with a fine group of classmates and officemates. Besides helping me with courses and research, I have enjoyed trips to Mitch’s, NCSU athletics, trips to the gym, as well as many other activities with friends I have met in the department. I am grateful for Amanda, Darryl, Hugh, Jimmy, Joe, Justin, Kirsten, Lavanya, Matthew, Michael, Mike, Paul, Ross, Steve, and Theresa, all of whom I met in the graduate program. When times got tough in the office, we often found that a makeshift game consisting of a ball and a trashcan was sufficient to raise our spirits.

I must also thank my former roommate Brad who I met during my freshman year at NCSU. Brad has been a good friend and he has helped make my college years a lot more fun.
I would also like to thank Kimberly, who has stood by my side throughout my graduate studies. Kimberly has seen me at my best and my worst, but she has always loved me, supported me, and given me the confidence to reach my goals. I appreciate her patience and understanding through this process.

Most importantly, I must thank my family for supporting me in everything I have done. My brother Jason attended NCSU for his undergraduate and graduate studies as well, and I was very fortunate that our time at NCSU overlapped for several years. Football games, basketball games, trips to the gym, movies, concerts, and countless other activities; I could not have gotten through this process without his advice and I would not be the person I am today without his friendship. Finally, and most importantly, I would like thank my parents. They believe in me and encourage me in everything that I pursue. My parents have provided me with a tremendous amount of support, an example of how to balance family and career, and a great set of genes among other things. Without the love and support of my family, I would have never made it this far. Thank you!
Table of Contents

List of Tables ....................................................... ix

List of Figures ..................................................... xi

1 The Orthogonal Interactions Model ............................... 1
  1.1 Motivation ..................................................... 2
  1.2 Orthogonal Interactions Model Definition ..................... 4
  1.3 Existing Models ................................................. 6
  1.4 Fitting the Orthogonal Interactions Model ..................... 10
    1.4.1 Fitting the OI Model by Reparameterization ............... 11
    1.4.2 The Profile Likelihood Method for Fitting the OI Model ... 15
  1.5 Fitting the Existing Models .................................... 19
    1.5.1 Fitting the AMMI model ................................... 20
    1.5.2 Fitting the Tukey One-Degree-of-Freedom Model ........... 22
  1.6 Inference Issues ................................................. 22
    1.6.1 Inference and the Orthogonal Interactions Model .......... 23
    1.6.2 Inference and Existing Methods .......................... 32
  1.7 Recognizing Data from the Orthogonal Interactions Model ... 36
  1.8 Testing the Orthogonality Assumptions ........................ 41
    1.8.1 A Bootstrap Corrected Likelihood Ratio Test .............. 43
1.8.2 A Noise Corrected Likelihood Ratio Test ......................... 51
1.8.3 Testing the OI Model for Unreplicated Data ..................... 56
1.8.4 Determining Why the Orthogonality Assumptions are Violated . 60
1.9 Example Fits ........................................................................ 62
  1.9.1 Metal Shear Strength Data ........................................... 62
  1.9.2 Soil Data ...................................................................... 66
1.10 Simulation Results ................................................................. 74
  1.10.1 Variance Estimation ..................................................... 74
  1.10.2 Testing Simulations ..................................................... 77
1.11 Conclusions ........................................................................ 81

2 A Resampling Based Method For Estimating Degrees of Freedom 83
  2.1 Motivation ...................................................................... 84
  2.2 Resampling Method for Estimating Degrees of Freedom ........ 89
  2.3 Resampling Degrees of Freedom and the Familiar Models .......... 95
  2.4 Resampling Degrees of Freedom and the AMMI Model .......... 101
  2.5 Discussion ..................................................................... 109

Literature Cited ........................................................................ 110

Appendix .................................................................................... 113

A Technical Details ..................................................................... 113
  A.1 A Linear Algebra Identity ................................................. 113
A.2 Two Constrained Optimality Results . . . . . . . . . . . . . . . . . . . 114
# List of Tables

1.1 *Estimated Error Degrees of Freedom for a $\times b$ Data Using Rank Method.* 25
1.2 *Rank DF and Adjusted Rank DF for $6 \times 6$ OI Data.* 26
1.3 *Type I Error Rates for Bootstrap Corrected LRT of Orthogonality Assumptions for a $\times b$ Data with c Replicates (Monte Carlo SE: 6.9E-3).* 47
1.4 *Type I Error Rates for Noise Corrected LRT of Orthogonality Assumptions for a $\times b$ Data with c Replicates (Monte Carlo SE: 6.9E-3).* 52
1.5 *Full ANOVA Fit of Metal Strength Data.* 63
1.6 *Orthogonal Interactions Fit of Metal Strength Data.* 63
1.7 *Orthogonal Interactions Fit of Soil Data (Resampling DF).* 68
1.8 *AMMI(2) Fit of Soil Data (Resampling DF).* 68
1.9 *Tukey 1-DF Fit of Soil Data.* 69
1.10 *Several Different Degree of Freedom Estimates For Orthogonal Interactions Model and AMMI(2) Analysis of Soil Data.* 72
1.11 *Simulated Error Variance Estimates for Orthogonal Interactions Model ($\sigma^2 = 1$).* 75
1.12 *Simulated Error Variance Estimates for Data Generated without Main Effects or Interactions ($\sigma^2 = 1$).* 76
2.1 *Expected SS and Expected MS for Mixed Effect Model.* 87
2.2 *Summary Table of Average Sums of Squares.* 91
2.3 *Summary of Estimated DF for 2-Factor Fixed Effect ANOVA.* 97
2.4 Summary of Resampling Degrees of Freedom for Tukey’s One-DF Model
   Defined by (2.8) .......................................................... 101
2.5 Summary of Error Variance Estimation for AMMI Data ($\sigma^2 = 1$) . . . 107
2.6 Summary of Error Variance Estimation for Random Noise ($\sigma^2 = 1$) . . 108
# List of Figures

1.1 *Mean Profiles:* (a) OI Model (b) ANOVA Model  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 37

1.2 *Biplots of:* (a) Tukey Data (b) Linear-bilinear Data  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 40

1.3 *Histogram of Simulated Unadjusted LRT p-values*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 48

1.4 *Histograms of Bootstrap Adjusted p-values for Tests Based on:* (a) $\hat{p}_1$  
(b) $\hat{p}_2$  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 49

1.5 *Histogram of Unadjusted Likelihood Ratio Test Statistic and $\chi^2$ Distribution*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 50

1.6 *Histogram of Bootstrap Adjusted Likelihood Ratio Test Statistic and $\chi^2$ Distribution*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 51

1.7 *Histograms of Noise Adjusted p-values for:* (a) Adjusted LRT (b) Empirical Test  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 54

1.8 *Histogram of Noise Adjusted Likelihood Ratio Test Statistic and $\chi^2$ Distribution*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 55

1.9 *Power Curves of Test (1.27) Using Two Different Empirical Methods*  
(Monte Carlo SE: 6.9E-3)  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 58

1.10 *Power Curves of Adjusted Empirical Tests of (1.27) when Parameters do not Satisfy Orthogonality Assumptions* (Monte Carlo SE: 6.9E-3)  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 59

1.11 *Metal Strength Prediction Profiles-* (a) OI Model (b) ANOVA  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 65

1.12 *Scatterplot of ANOVA Predictions versus OI Model Predictions for Metal Strength Data*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 66

1.13 *Biplot of Metal Strength Data*  
2. . . . . . . . . . . . . . . . . . . . . . . . . . 67
The standard two-factor analysis of variance model with $A$ at $a$ levels, $B$ at $b$ levels, and $c$ replicates per factor level combination (FLC) is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

where $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i \gamma_{ij} = 0$ for all $j$, $\sum_j \gamma_{ij} = 0$ for all $i$, and $\epsilon_{ijk}$ are independent and identically distributed $N(0, \sigma^2)$ variates. If $c = 1$, the case of no replication, the data are not sufficient to estimate the interaction terms, $\gamma_{ij}$, and the error variance $\sigma^2$, thus limiting the possible inferences. In experiments such as crop yield trials, investigators are interested in main effects as well as interactions. Because of expenses or other limitations, obtaining replicates is not always an option. An alternative is to estimate a set of constrained interactions, thus providing some degrees of freedom for error variance estimation. An extreme version of this strategy assumes that all $\gamma_{ij} = 0$, thereby providing $(a - 1)(b - 1)$ degrees of freedom for estimating $\sigma^2$ in this case. Note that in this case, $(a - 1)(b - 1)$ additional constraints are imposed on the interactions. Other versions of this strategy are at the heart of Tukey’s one degree of freedom test of additivity as well as the so-called AMMI (additive main effects and multiplicative interactions) model used extensively in agricultural research. We discuss the latter two approaches in detail in Section 1.3. We propose a new set of constraints
Chapter 1. The Orthogonal Interactions Model

for the usual ANOVA model that will allow for estimation of the interaction terms while providing degrees of freedom for error in the case of no replication and compare our new model to existing models.

The remainder of this chapter is organized as follows. In Section 1.1 we provide the motivation for the orthogonal interactions model with a formal definition of the model in Section 1.2. Then in Section 1.3 we present existing models that fit constrained interactions to unreplicated two-factor models. We provide two approaches to fitting the orthogonal interactions model in Section 1.4 and details for fitting the existing models in Section 1.5. Inference is considered in Section 1.6 for the orthogonal interactions model as well as the existing models. In Section 1.7 we discuss how to recognize data from the orthogonal interactions model and in Section 1.8 we discuss testing the appropriateness of the orthogonality assumptions for both replicated and unreplicated data. We fit the orthogonal interactions model and the existing models to several real data sets in Section 1.9. In Section 1.10 we use simulations to investigate the performance of $F$-tests and to investigate the error mean square as an error variance estimator when using the orthogonal interactions model.

1.1 Motivation

We propose a set of constraints that allow the estimation of interaction terms and error variance for two-factor ANOVA data without replication. The new constraints are motivated by the two-factor random effects model. The two-factor random effects
Chapter 1. The Orthogonal Interactions Model

model with interactions is written

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, c; \quad (1.2) \]

where

\[ \alpha_i \sim iid N(0, \sigma_a^2), \quad \beta_j \sim iid N(0, \sigma_b^2), \quad \gamma_{ij} \sim iid N(0, \sigma_g^2), \quad \text{and} \quad \epsilon_{ijk} \sim iid N(0, \sigma_e^2). \]

Furthermore, it is assumed that the main effects and interactions are all mutually independent. This assumption of independence is equivalent to the assumption underlying random effect models for crossed designs, that \( \text{Cov}(\alpha_i, \gamma_{ij}) = 0 \) for all \( i \) and \( j \), which in turn implies that

\[ E(\alpha_i \gamma_{ij}) = 0 \quad \text{for all} \quad i \quad \text{and} \quad j, \]

because \( E(\alpha_i) = E(\gamma_{ij}) = 0 \). A similar argument can be made with the \( \beta_j \gamma_{ij} \) terms.

These assumptions of independence imply that the main effects vectors are orthogonal to the rows/columns of the interaction matrix in expectation. The assumption of main effects being orthogonal to interactions can be carried over to the fixed effects ANOVA model. These ideas from the random effects ANOVA model provide motivation for the orthogonal interactions (OI) model.

Comparing the assumptions of the fixed effects model (1.1) to those of the random effects model (1.2), we see a natural correspondence between effect sample moments in the fixed effects model and the population moments in the random effects model:

\[ \frac{1}{a} \sum_i \alpha_i = 0 \text{ corresponds to } E(\alpha_i) = 0, \quad (1.3) \]
Chapter 1. The Orthogonal Interactions Model

\[ \frac{1}{b} \sum_{j} \beta_j = 0 \] corresponds to \( \text{E}(\beta_j) = 0 \),

\[ \frac{1}{ab} \sum_{j} \sum_{i} \gamma_{ij} = 0 \] corresponds to \( \text{E}(\gamma_{ij}) = 0 \).

The correspondences in (1.3) are the ‘natural’ correspondences that one considers upon comparing (1.1) and (1.2). However, the random effects model has the additional population assumption that \( \text{E}(\alpha_i \gamma_{ij}) = \text{E}(\beta_j \gamma_{ij}) = 0 \). The fixed effect model finite population versions of these are:

\[ \frac{1}{b} \sum_{j=1}^{b} \beta_j \gamma_{ij} = 0, \text{ for all } i; \text{ and } \frac{1}{a} \sum_{i=1}^{a} \alpha_i \gamma_{ij} = 0, \text{ for all } j. \] (1.4)

The constraints in (1.4) have not been previously imposed on the fixed effects model. However, imposing (1.4) in the fixed effects model frees degrees of freedom for estimating \( \sigma^2 \). In the \( c = 1 \) case, imposing (1.4) allows for the estimation of the \( \gamma_{ij} \) as well as \( \sigma^2 \).

1.2 Orthogonal Interactions Model Definition

We propose a set of assumptions to be added to the fixed factor ANOVA model that allows the estimation of constrained interactions as well as the error variance in the absence of replication. The model has the same form as the two-factor ANOVA model

\[ Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}, \] (1.5)
Chapter 1. The Orthogonal Interactions Model

with the usual assumptions:

\[ \sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} \gamma_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^{b} \gamma_{ij} = 0 \text{ for all } i. \]  \hspace{1cm} (1.6)

As usual, we assume that \( \epsilon_{ij} \sim iid \sim N(0, \sigma^2) \). However, we add these orthogonality assumptions to the model:

\[ \sum_{i=1}^{a} \alpha_i \gamma_{ij} = 0, \text{ for all } j; \quad \text{and} \quad \sum_{j=1}^{b} \beta_j \gamma_{ij} = 0, \text{ for all } i. \]  \hspace{1cm} (1.7)

Using \( \alpha \) (a \( \times \) 1 vector) and \( \beta \) (b \( \times \) 1 vector) to represent the main effect parameters and \( \gamma \) (a \( \times \) b matrix) to represent the interaction parameters, the orthogonal interactions model assumptions are written

\[ \alpha^T \mathbf{1}_a = 0, \quad \beta^T \mathbf{1}_b = 0, \quad \gamma \mathbf{1}_b = \mathbf{0}, \quad \gamma^T \mathbf{1}_a = \mathbf{0}, \]  \hspace{1cm} (1.8)

\[ \alpha^T \gamma = \mathbf{0}, \text{ and } \gamma \beta = \mathbf{0}, \]  \hspace{1cm} (1.9)

where \( \mathbf{1}_k \) is a \( k \times 1 \) vector of ones. We saw in the previous section that the assumptions in (1.9) are motivated by the random effects model.

The resulting mean model

\[ \mathbb{E}(Y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}, \]

is linear, albeit with nonlinear constraints on the parameters. The model can be fit using Proc IML in SAS or similar matrix algebra programs. We use the ‘nlphqn’
optimization routine in Proc IML to obtain least squares estimates for the parameters in (1.5) using a hybrid quasi-Newton algorithm. We provide strategies for fitting the orthogonal interactions model in Section 1.4.

1.3 Existing Models

Two methods are commonly used for estimating constrained interactions and error variance for data from unreplicated two-factor experiments when factors have qualitative levels. Gauch developed the additive main effects and multiplicative interactions (AMMI) model (see [1] and [2]) which involves fitting the main effects ANOVA model to the data and then performing a principal component analysis of the residuals. Tukey’s one-degree-of-freedom model fits additive main effects and then a very restrictive multiplicative interaction term to the data. We also look at a class of linear-bilinear models that require the interactions to be a function of one of the main effects.

Tukey’s One-Degree-of-Freedom Model

Tukey [3] presents a model that adds a non-additive interaction term to the standard two factor ANOVA model. The model is written

\[ Y_{ij} = \mu + \alpha_i + \beta_j + k(\alpha_i \beta_j) + \epsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad (1.10) \]

where \( k \) is the non-additivity parameter and \( \epsilon_{ij} \sim N(0, \sigma^2) \). In (1.10), the estimates for \( \mu, \alpha_i, \) and \( \beta_j \) all come from the main effects ANOVA model. A single degree of freedom is associated with estimating \( k \), leaving \((a - 1)(b - 1) - 1\) degrees of freedom.
for error. This model fits a very specific type of interaction to the data, requiring the interaction to be a constant multiple of the product of main effects. Failure to reject the hypothesis

\[ H_0 : k = 0 \text{ vs. } H_1 : k \neq 0 \]

does not necessarily indicate that there is no interaction effect, as it may be only that the interaction, if present, is not of this form. Tukey’s model can also be fit to replicated data, requiring only a simple adjustment to (1.10).

**Linear-Bilinear Models**

van Eeuwijk et al. [4] provides a summary of linear-bilinear models in the context of unreplicated crop yield trials. Linear-bilinear models require that the interactions are a linear function of one of the main effects. A simple linear-bilinear model is written

\[
E(Y_{ij}) = \mu + \alpha_i + \beta_j + \delta_i \beta_j,
\]  
(1.11)

where \( \sum_i \alpha_i = \sum_j \beta_j = 0 \). If \( \delta_i = 0 \) for all \( i \), then model (1.11) reduces to the additive two-factor ANOVA model. The \( \delta_i \beta_j \) terms represent the interactions and are called the bilinear terms. An alternative to (1.11) would be to make the interaction a function of the \( \alpha_i : E(Y_{ij}) = \mu + \alpha_i + \beta_j + \delta_j \alpha_i \).

A more flexible linear-bilinear model was suggested by Johnson and Graybill [5]. The model is intended for unreplicated two-factor experiments and is written

\[
Y_{ij} = \mu + \tau_i + \beta_j + \lambda \alpha_i \gamma_j + \epsilon_{ij}, \quad i = 1, \ldots, a; \quad j = 1, \ldots, b;
\]  
(1.12)
Chapter 1. The Orthogonal Interactions Model

where \( \sum_i \tau_i = \sum_j \beta_j = \sum_i \alpha_i = \sum_j \gamma_j = 0 \) and \( \epsilon_{ij} \sim iid \sim N(0, \sigma^2) \). Additionally the model assumes that \( \sum_i \alpha_i^2 = \sum_j \gamma_j^2 = 1 \). In (1.12), the bilinear terms \( \lambda \alpha_i \gamma_j \) represent the interactions.

The interactions fit by the linear-bilinear models considered in this subsection are more flexible than the interactions fit by Tukey’s model. The additive main effects and multiplicative interactions (AMMI) model is a popular linear-bilinear model that features a varying number of bilinear terms. The AMMI model is used frequently in agricultural experiments and is described in the following subsection.

The AMMI Model

The AMMI model was originally developed for analyzing data from crop yield trials, where investigators grow different genotypes of a crop in a number of different environments. In this type of experiment, investigators are interested in the genotype effects and environment effects as well as the genotype-by-environment interaction. Unfortunately, investigators are not always able to form genotype/environment combination replications due to various limitations. Thus the AMMI model frequently appears in agricultural literature. Fitting the model consists of two steps: first fitting the additive main effects model to the data, and then performing a principal component analysis (PCA) of the residuals from the main effects model.

With experimental factors \( A \) and \( B \), the AMMI model is written:

\[
Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{n=1}^{N} \theta_n \tau_{in} \delta_{jn} + \epsilon_{ij}, \tag{1.13}
\]

where

\[
\sum_i \tau_i = \sum_j \beta_j = \sum_i \alpha_i = \sum_j \gamma_j = 0 \] and \( \epsilon_{ij} \sim iid \sim N(0, \sigma^2) \)
Chapter 1. The Orthogonal Interactions Model

\[ Y_{ij} \] is the response at level \( i \) and \( j \);
\[ \mu \] is the grand response mean;
\[ \alpha_i \] is the effect of the \( i^{th} \) level of \( A \);
\[ \beta_j \] is the effect of the \( j^{th} \) level of \( B \);
\[ \theta_n \] is the eigenvalue of PCA axis \( n \);
\[ \tau_{in} \] is the factor \( A \) PCA score for PCA axis \( n \);
\[ \delta_{jn} \] is the factor \( B \) PCA score for PCA axis \( n \);
\( N \) is the number of PCA axes used in the model;
\( \epsilon_{ij} \) are \( N(0, \sigma^2) \) random variates.

One advantage to using the AMMI model is that it is easy to implement. The model can be fit using a combination of Proc GLM and Proc IML in SAS (or similar computing languages), fitting the main effects model using GLM and doing the principal component analysis in Proc IML or Proc PRINCOMP. In (1.13) \( \mu, \alpha_i, \) and \( \beta_j \) come directly from the additive main effects model, so the estimation of these terms is straightforward. Estimates for \( \theta_n, \tau_{in}, \) and \( \delta_{jn} \) are obtained from the singular value decomposition of the residual matrix from the main effects model. We provide more detail for fitting the AMMI model in Section 1.5.1. The AMMI model can be fit to replicated data as well, requiring a slight modification to (1.13).

One drawback to using this model is that the investigator must choose \( N \), the number of PCA axes in the model. Clearly each principal component axis included in the model reduces the sum of squared errors. Gauch states that if a few of the PCA axes are not sufficient for capturing a large part of the interaction sum of squares, then the interaction is very complex and it would be impossible to find a parsimonious
model \textsuperscript{[1]}. For more information about choosing the number of principal component axes to include in the model, see the work by Cornelius \textsuperscript{[6]} and the work by dos Santos Dias and Krzanowski \textsuperscript{[7]}. Many of the methods for choosing \( N \) are based on cross validation or they make use of the distribution of eigenvectors as developed by Mandel \textsuperscript{[8]}. Sometimes the dimension of the data will limit the number of PCA axes that can be included in the model. For example, \( 6 \times 4 \) data (six levels of experimental factor \( A \) and four levels of experimental factor \( B \)) are only sufficient for including one or two principal component axes in the model. Another problem with the AMMI model is that there is debate over how many degrees of freedom each PCA axis adds to the model. We will discuss assigning degrees of freedom to each source of variation in the AMMI model in Section 1.6.2. The linear-bilinear model suffers from the same problem of assigning degrees of freedom to the bilinear terms.

1.4 Fitting the Orthogonal Interactions Model

The orthogonal interactions model is simple to fit using SAS or similar statistical packages that feature built-in optimization routines. We have developed two methods for fitting the orthogonal interactions model. The first method we present uses a reparameterization to estimate main effects and interactions that satisfy the orthogonality constraints. Since this method involves optimizing over the main effects and interaction parameters, the optimization routine must minimize over \( ab - 1 \) constrained parameters. The second method finds the least squares estimates for the interaction terms as a function of the main effects, and subsequently requires the optimization routine to minimize over the \( a + b - 2 \) main effect parameters. While the first method we present
Chapter 1. The Orthogonal Interactions Model

is more appealing intuitively, the second method is more efficient computationally.

1.4.1 Fitting the OI Model by Reparameterization

Enforcing the usual sum-to-zero ANOVA constraints can be done through modifying the design matrix, $\mathbf{X}$. Unfortunately, enforcing the constraints in (1.7) is not as simple.

One way to enforce the orthogonality constraints in (1.7) is to use a reparameterization. This subsection provides a method for taking $\alpha$, $\beta$, and $\gamma$ that satisfy the usual ANOVA assumptions in (1.6), and transforming $\gamma$ so that $\alpha$, $\beta$, and $\gamma$ satisfy the orthogonality assumptions in (1.7).

Suppose that for $a \times b$ data ($a$ levels of factor $A$ and $b$ levels of factor $B$), we let $\alpha$ and $\beta$ be vectors of sizes $a \times 1$ and $b \times 1$ respectively. We assume that $\alpha$ and $\beta$ satisfy the usual sum to zero constraints, represented in matrix form as: $\alpha^T \mathbf{1}_a = \beta^T \mathbf{1}_b = 0$, where $\mathbf{1}_k$ denotes a $k \times 1$ vector of ones. Let $\gamma$ be an $a \times b$ matrix of interactions that satisfy the constraints $\gamma \mathbf{1}_b = \mathbf{0}$ and $\gamma^T \mathbf{1}_a = \mathbf{0}$, meaning that the row sums and column sums of $\gamma$ are all zero. The first step to fitting the OI model is to show that $\hat{\mu} = \overline{Y}..$

To do this, we define the error sum of squares

$$J(\mu, \alpha, \beta, \gamma) = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2.$$

Taking the derivative with respect to $\mu$

$$\frac{\partial}{\partial \mu} J(\mu, \alpha, \beta, \gamma) = -2 \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \mu - \alpha_i - \beta_j - \gamma_{ij}),$$
Chapter 1. The Orthogonal Interactions Model

and setting the partial derivative equal to zero results in

\[ 0 = - \left( \sum_{j=1}^{b} \sum_{i=1}^{a} Y_{ij} \right) + ab\mu + 0 + 0 + 0. \]

Thus

\[ \hat{\mu} = \frac{1}{ab} \sum_{j=1}^{b} \sum_{i=1}^{a} Y_{ij} = Y_{..}. \]

So just as in the usual ANOVA model, \( \hat{\mu} = Y_{..} \) for the orthogonal interactions model.

Next we provide the details for estimating the main effects and interactions.

We are using the matrix notation in (1.9) to represent the orthogonality assumptions. Now \( \gamma \) can also be represented by an \((ab) \times 1\) vector that we denote \( \gamma_v \). The relationship between \( \gamma \) and \( \gamma_v \) is written

\[
\gamma = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_a \end{bmatrix} \quad \text{and} \quad \gamma_v = \begin{bmatrix} g_1^T \\ g_2^T \\ \vdots \\ g_a^T \end{bmatrix},
\]

where \( g_i \) is the \( i^{th} \) row of \( \gamma \). Then \( \gamma_v \) can be used to express the orthogonality assumptions as

\[ \mathbf{M}_1(\alpha)\gamma_v = \mathbf{0} \quad \text{and} \quad \mathbf{M}_2(\beta)\gamma_v = \mathbf{0}, \]

where \( \mathbf{M}_1(\alpha) \) is an \( a \times (ab) \) matrix function of \( \alpha \) and \( \mathbf{M}_2(\beta) \) is a \( b \times (ab) \) matrix function.
Chapter 1. The Orthogonal Interactions Model

of β. Defining an \((a + b) \times ab\) matrix function of \(\alpha\) and \(\beta\),

\[
M(\alpha, \beta) = \begin{bmatrix}
M_1(\alpha) \\
M_2(\beta)
\end{bmatrix},
\]

the orthogonality constraints can also be represented by \(M(\alpha, \beta)\gamma_v = 0\).

If \(\gamma_0\) is an \(a \times b\) matrix of interactions that satisfies the sum-to-zero constraints and \(\alpha\) and \(\beta\) also satisfy the sum-to-zero constraints, then \(\gamma_0\) can be represented by an \((ab) \times 1\) vector denoted \(\gamma_{v,0}\) and we can perform the reparameterization

\[
\gamma_{v,1} = \{I - M(\alpha, \beta)^{\#}M(\alpha, \beta)\} \gamma_{v,0},
\]

(1.14)

where \(M(\alpha, \beta)^{\#}\) denotes the generalized inverse of the matrix \(M(\alpha, \beta)\). Then \(\gamma_{v,1}\) is reshaped into the \(a \times b\) matrix \(\gamma_1\) using the relationship

\[
\gamma_{v,1} = \begin{bmatrix}
g_1^T \\
g_2^T \\
\vdots \\
g_a^T
\end{bmatrix}
\text{ and } \gamma_1 = \begin{bmatrix}
g_1 \\
g_2 \\
\vdots \\
g_a
\end{bmatrix},
\]

where the \(g_i\) are \(1 \times b\) vectors. Now \(\alpha, \beta,\) and \(\gamma_1\) satisfy all of the orthogonal interactions model assumptions. We have shown how to take \(\alpha, \beta,\) and \(\gamma\) that satisfy the ANOVA model assumptions, and reparameterize \(\gamma\) using (1.14) so that the parameters satisfy all of the OI model assumptions.
Chapter 1. The Orthogonal Interactions Model

Using the reparameterization to obtain $\alpha$, $\beta$, and $\gamma$ that satisfy (1.8) and (1.9), the error sum of squares is written

$$J(\alpha, \beta, \gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \bar{Y}_{..} - \alpha_i - \beta_j - \gamma_{ij})^2. \quad (1.15)$$

Minimizing (1.15) yields the least squares estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ that satisfy the OI model assumptions. An optimization routine such as 'nlphqn' in SAS can be used to minimize this error sum of squares. The problem with fitting the OI model through reparameterization is that it is not efficient. For $a \times b$ data, the optimization routine is minimizing (1.15) over $ab - 1$ constrained parameters which can be computationally intense. For example, when using this method with $25 \times 25$ data, the optimization routine must minimize the error sum of squares in 624 dimensions. Optimizing over 624 parameters is computationally intense and may overwhelm some optimization routines. Although the reparameterization method is straightforward, it may be too inefficient to use with real data. In Chapter 2 we show that the resampling method uses simulation methods to estimate the degrees of freedom for the sources of variation in the OI model. Using the reparameterization method to fit the orthogonal interactions model may prevent the use of the resampling method for estimating degrees of freedom because of the computation time required to fit the model possibly thousands of times. The next subsection presents a more efficient method for fitting the orthogonal interactions model.
Chapter 1. The Orthogonal Interactions Model

1.4.2 The Profile Likelihood Method for Fitting the OI Model

In this subsection, we discuss a method for fitting the orthogonal interactions model that is more efficient than the reparameterization method discussed in the previous subsection. We show that given the main effect parameters $\alpha$ and $\beta$, the optimal solution for $\gamma$ is a simple, easily calculated function of $\alpha$ and $\beta$. This relationship allows us to minimize the error sum of squares over $a + b - 2$ parameters rather than $ab - 1$ parameters. This reduction in dimension provides a significant savings in computation time and allows the OI model to be fit to much larger data sets. Because this method involves estimating the interactions as a function of the main effects, we call this method the profile likelihood method for fitting the OI model.

Let $Y$ be an $a \times b$ matrix of data. Again we use the $a \times 1$ vector $\alpha$ and the $b \times 1$ vector $\beta$ to represent the main effect parameters. Furthermore, assume that $\alpha$ and $\beta$ satisfy the usual ANOVA sum-to-zero assumptions. We use $1_k$ to represent a $k \times 1$ vector of ones and we define

$$A = \begin{bmatrix} 1_a, \alpha \end{bmatrix}, B = \begin{bmatrix} 1_b, \beta \end{bmatrix},$$

$$P_A = A (A^T A)^{-1} A^T, \quad \text{and} \quad P_B = B (B^T B)^{-1} B^T.$$}

Here $P_A$ is an $a \times a$ projection onto $A$ and $P_B$ is a $b \times b$ projection onto $B$. Now let $r$ be an $a \times b$ matrix of residuals defined by

$$r_{ij} = Y_{ij} - \bar{Y} - \alpha_i - \beta_j. \quad (1.16)$$
Chapter 1. The Orthogonal Interactions Model

We show that given $\alpha$ and $\beta$,

$$
\hat{\gamma} = (I - P_A)r(I - P_B)
$$

(1.17)

is the profile likelihood estimate of $\gamma$ that satisfies the orthogonality constraints. In Appendix A.2, we show that $\hat{\gamma}$ satisfies the OI model assumptions stated in (1.8) and (1.9).

Next we show that $\hat{\gamma} = (I - P_A)r(I - P_B)$ minimizes the the error sum of squares given $\alpha$ and $\beta$. Let $\tilde{\gamma}$ be any $a \times b$ matrix that satisfies the orthogonality constraints. Furthermore, we define

$$
Q(\gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} (r_{ij} - \gamma_{ij})^2,
$$

where $r_{ij}$ are defined in (1.16). In Appendix A.2 we show that

$$
Q(\tilde{\gamma}) = Q(\hat{\gamma}) + \text{tr}(\Delta^T\Delta),
$$

where $\Delta = \hat{\gamma} - \tilde{\gamma}$. In Appendix A.1, we show that $\text{tr}(\Delta^T\Delta) = \sum_i \sum_j \Delta_{ij}^2$, therefore $Q(\tilde{\gamma}) \geq Q(\hat{\gamma})$ with equality only when $\Delta = 0$. That means that given $\alpha$ and $\beta$, $\hat{\gamma} = (I - P_A)r(I - P_B)$ is the profile likelihood estimate of the interaction effects. So fitting the orthogonal interactions model using the profile likelihood method is done by minimizing

$$
J(\alpha, \beta) = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{..} - \alpha_i - \beta_j - \hat{\gamma}_{ij}\right)^2,
$$

(1.18)

which is a function of only $\alpha$ and $\beta$. In (1.18), $\hat{\gamma}_{ij}$ represents the $(i, j)$ element of the matrix $\hat{\gamma}$ defined in (1.17), which is a function of $\alpha$ and $\beta$. The error sum of
Chapter 1. The Orthogonal Interactions Model

squares in (1.18) can be minimized using an optimization routine such as ‘nlphqn’ in SAS Proc IML. The error sum of squares is a function of the main effect parameters $\alpha$ and $\beta$, which means that the dimension of the optimization is only $a + b - 2$, as opposed to $ab - 1$ for the method discussed in the previous section. This reduction allows the orthogonal interactions model to be fit more efficiently. For example, using the profile likelihood method to fit the orthogonal interactions model to $25 \times 25$ data involves a minimization over 48 parameters, a substantial savings in computation over the reparameterization method.

In Appendix A.2, we show that given the interaction matrix $\gamma$, the profile likelihood estimates of $\alpha$ and $\beta$ are

$$\hat{\alpha} = (I - PU)\overline{Y}_a \text{ and } \hat{\beta} = (I - PV)\overline{Y}_b,$$

where

$$U = \begin{bmatrix} 1_a, & \gamma \end{bmatrix}, \quad P_U = U(U^TU)^gU^T,$$

$$V = \begin{bmatrix} 1_b, & \gamma^T \end{bmatrix}, \quad P_V = V(V^TV)^gV^T,$$

$$\overline{Y}_a = \begin{bmatrix} \overline{Y}_{1a} - \overline{Y}_a \\ \overline{Y}_{2a} - \overline{Y}_a \\ \vdots \\ \overline{Y}_{aa} - \overline{Y}_a \end{bmatrix}, \text{ and } \overline{Y}_b = \begin{bmatrix} \overline{Y}_{1b} - \overline{Y}_b \\ \overline{Y}_{2b} - \overline{Y}_b \\ \vdots \\ \overline{Y}_{bb} - \overline{Y}_b \end{bmatrix}.$$

While these solutions for $\hat{\alpha}$ and $\hat{\beta}$ are not useful in fitting the OI model, they are useful for proving certain results in later sections. This relationship is critical in showing that
the sums of squares as defined in Section 1.6.1 are additive, meaning that the sum of squares from each source of variation add to the total sum of squares.

**Extension to Replicated Data**

Although the primary purpose of the OI model is to estimate interactions in the absence of factor level combination replication, investigators may be interested in fitting the model to replicated data as well. If data are truly from the orthogonal interactions model, then fitting the OI model to the data rather than fitting the ANOVA model could result in more efficient error variance estimates. Extending the reparameterization method described in Section 1.4.1 to fit replicated data is straightforward but as discussed earlier, it requires extensive computation for large data sets. The profile likelihood method described in Section 1.4.2 is more attractive because of the savings in computation, but the extension to fit replicated data is not immediately obvious. We now show how to use the profile likelihood method to fit the OI model to replicated data. The replicated orthogonal interactions model is written

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \ j = 1, \ldots, b; \ k = 1, \ldots, c; \]

where \( \alpha, \beta, \) and \( \gamma \) are subject to the orthogonality and sum-to-zero constraints and \( \epsilon_{ijk} \) are independent \( N(0, \sigma^2) \) variates. Fitting this model is very similar to fitting the unreplicated orthogonal interactions model, only requiring an adjustment to the matrix \( r \) defined in (1.16). Define \( A, B, P_A, \) and \( P_B \) as in Section 1.4.2 and redefine
Chapter 1. The Orthogonal Interactions Model

the matrix $r$ by

$$r_{ij} = c^{-1} \sum_{k=1}^{c} (Y_{ijk} - \overline{Y}_{..} - \alpha_i - \beta_j).$$

Then given $\alpha$ and $\beta$, the profile likelihood estimate of $\gamma$ is

$$\hat{\gamma} = (I - P_A)r(I - P_B).$$

The orthogonal interactions model is fit to the data matrix $Y$ by minimizing the error sum of square function

$$J(\alpha, \beta) = \sum_{k=1}^{c} \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ijk} - \overline{Y}_{..} - \alpha_i - \beta_j - \hat{\gamma}_{ij})^2. \quad (1.19)$$

Minimizing (1.19) over $\alpha$ and $\beta$ ($a + b - 2$ parameters) yields estimates of main effects and interactions.

1.5 Fitting the Existing Models

In this section we discuss fitting the AMMI model as well as the Tukey one-degree-of-freedom model. Both models are simple to fit using SAS or similar computing languages. Fitting the AMMI model requires fitting a linear model and performing a principal component analysis of the residuals. Fitting the Tukey one-degree-of-freedom model requires creating a new variable from the data and fitting a linear model. Since the AMMI model is the most frequently used linear-bilinear model, we will omit the details of fitting the linear-bilinear model, instead focusing on the AMMI model.
1.5.1 Fitting the AMMI model

Suppose that investigators are interested in fitting the AMMI model to data with \( a \) levels of experimental factor \( A \) and \( b \) levels of experimental factor \( B \). The first step in fitting the AMMI model is to fit the additive main effects ANOVA model to the data:

\[
E(Y_{ij}) = \mu + \alpha_i + \beta_j \quad i = 1, \ldots, a; \quad j = 1, \ldots, b,
\]  

resulting in

\[
\hat{\mu} = \overline{Y}., \quad \hat{\alpha}_i = \overline{Y}_i. - \overline{Y}.., \quad \text{and} \quad \hat{\beta}_j = \overline{Y}.j - \overline{Y}..
\]

Here \( \overline{Y}. = (ab)^{-1} \sum_i \sum_j Y_{ij}, \overline{Y}_i. = b^{-1} \sum_j Y_{ij}, \) and \( \overline{Y}.j = a^{-1} \sum_i Y_{ij}. \) Then let \( \mathbf{e} \) be an \( a \times b \) matrix of residuals from (1.20) where

\[
e_{ij} = Y_{ij} - \hat{Y}_{ij}.
\]

Next, obtain the singular value decomposition of \( \mathbf{e} \), resulting in

\[
\mathbf{e} = \mathbf{UQV}^T \quad \text{where} \quad \mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{VV}^T = \mathbf{I}_b.
\]

Here \( \mathbf{U} \) is an \( a \times b \) matrix, \( \mathbf{Q} \) is a \( b \times b \) diagonal matrix, \( \mathbf{V} \) is a \( b \times b \) matrix, and \( \mathbf{I}_b \) is the \( b \times b \) identity matrix. This singular value decomposition can be obtained using the ‘svd’ function in SAS Proc IML or similar routines available with other computing languages.

Now let AMMI(\( N \)) represent the AMMI model that uses \( N \) principal component axes in modeling the interactions. Predictions for \( Y_{ij} \) using the AMMI(1) model are
Chapter 1. The Orthogonal Interactions Model

obtained using the formula:

\[ \hat{Y}_{ij} = \bar{Y} + \hat{\alpha}_i + \hat{\beta}_j + q_{11}u_{i1}v_{j1}, \]

where \( \hat{\mu}, \hat{\alpha}, \) and \( \hat{\beta} \) are taken from model (1.20) and \( q_{11} \) is the \((1, 1)\) element of \( Q, \) \( u_{i1} \) is the \((i, 1)\) element of \( U, \) and \( v_{j1} \) is the \((j, 1)\) element of \( V. \) Here \( q_{11}u_{i1}v_{j1} \) is the estimate of the \((i,j)\) interaction term using the AMMI(1) model. Similarly, predictions for \( \hat{Y}_{ij} \) can be obtained for the general AMMI\((N)\) model using the formula

\[ \hat{Y}_{ij} = \bar{Y} + \hat{\alpha}_i + \hat{\beta}_j + \sum_{k=1}^{N} q_{kk}u_{ik}v_{jk}, \]

where \( \hat{\mu}, \hat{\alpha}, \) and \( \hat{\beta} \) again come from the main effects model stated in (1.20). Here \( \sum_{k=1}^{N} q_{kk}u_{ik}v_{jk} = \hat{\gamma}_{ij} \) is the estimate of the \((i,j)\) interaction term when using the AMMI\((N)\) model.

After obtaining a matrix of predicted values from the AMMI\((N)\) model, \( \hat{Y}, \) sums of squares can be partitioned into sources of variation. In order to partition the sums of squares for the AMMI\((N)\) model, the AMMI(1) through AMMI\((N)\) models must all be fit to the data. Let \( \text{SSE}\{\text{AMMI}(i)\} \) represent the error sum of squares for the AMMI\((i)\) model. Then the sum of squares for the \( i^{th} \) PCA axis is defined by

\[ \text{SS} \{ \text{PCA}(i) \} = \text{SSE} \{ \text{AMMI}(i - 1) \} - \text{SSE} \{ \text{AMMI}(i) \}. \]

The AMMI(0) model is equivalent to the main effects ANOVA model, so \( \text{SSE}\{\text{AMMI}(0)\} \) comes directly from (1.20).
1.5.2 Fitting the Tukey One-Degree-of-Freedom Model

Fitting Tukey’s one-degree-of-freedom model is also straightforward. The model is written as in (1.10) and requires a new variable to be created in order to fit the model. For the two-factor ANOVA model in (1.20), we know that
\[ \hat{\alpha}_i = Y_{i.} - Y_{..}, \text{ and } \hat{\beta}_j = Y_{.j} - Y_{..}, \]

where \( Y_{..} \) is the grand mean, \( Y_{i.} = b^{-1} \sum_j Y_{ij} \), and \( Y_{.j} = a^{-1} \sum_i Y_{ij} \). Then we create the variable
\[ l_{ij} = (Y_{i.} - Y_{..})(Y_{.j} - Y_{..}) \]

to aid in fitting Tukey’s model. Using this new variable, fitting the linear model
\[ E(Y_{ij}) = \mu + \alpha_i + \beta_j + k l_{ij}, \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \]  

is equivalent to fitting Tukey’s model stated in (1.10). The model in (1.21) is simple to fit using any statistical computing language and can even be fit using Proc GLM or Proc REG in SAS. The same strategy is used when fitting Tukey’s model to replicated data. Although we use (1.21) to fit Tukey’s model, it is important to remember that Tukey’s model is a nonlinear model.

1.6 Inference Issues

Although we have not considered any linear models up to this point, we will use methods from linear models to test for main effects and interactions. Using the orthogonal
interactions model or one of the existing models, the sum of squares are partitioned to
each source of variation and $F$ statistics are formed as if we were working with linear
models. Although we do not have theory to support the use of these $F$ tests, we show
that the approximate $F$ tests yield reasonable results. Assigning degrees of freedom
to each source of variation is an important issue when using these models because we
must obtain the correct mean squares and reference $F$ distributions. In the remainder
of this section, we will discuss methods for assigning degrees of freedom to each source
of variation in the orthogonal interactions model and the existing models.

1.6.1 Inference and the Orthogonal Interactions Model

A key issue in using the orthogonal interactions model is assigning degrees of freedom
to each source of variation. In the standard two-factor ANOVA without replication,
including interactions in the model contributes $(a - 1) \times (b - 1)$ degrees of freedom to
the model and results in zero degrees of freedom left for error. Because of the nonlinear
constraints on $\alpha$, $\beta$, and $\gamma$, the interactions may no longer contribute $(a - 1) \times (b - 1)$
degrees of freedom to the model. Since the orthogonality assumptions can also be
viewed as constraints on the main effects as well, it may no longer be appropriate
to assign $a - 1$ and $b - 1$ degrees of freedom to the main effects. We use ideas from
linear models theory to find the error degrees of freedom for the orthogonal interactions
model. In this subsection we will discuss estimating the degrees of freedom for each
source of variation in the OI model.

Linear models theory tells us that when testing the linear hypothesis $H_0 : \mathbf{g}(\theta) = \mathbf{0}$,
where $g(\theta)$ is a linear function of $\theta$, the degree of freedom for the test is

$$df = \text{rank} \left( \frac{\partial g(\theta)}{\partial \theta^T} \Big|_{\theta = \theta^*} \right) = \text{rank} \{ g(\theta) \}, \quad \text{where } \theta^* \text{ satisfies } g(\theta^*) = 0.$$ 

Here, if $\theta$ is a $p \times 1$ vector and $g(\theta)$ is an $r \times 1$ vector function of $\theta$, we know that $df = r$ only when none of the functions in $g(\cdot)$ are redundant. Using the same idea, we can find an estimator for the error degrees of freedom for the orthogonal interactions model.

Consider data from the orthogonal interactions model as stated in (1.5). We use the matrix representation of the main effects: $\alpha$ and $\beta$ ($a \times 1$ and $b \times 1$ vectors respectively) as well as the interactions: $\gamma$ ($a \times b$ matrix), and we use $1_k$ to denote a $k \times 1$ vector of ones. As stated earlier, the OI model assumptions can be stated in matrix form using (1.8) and (1.9). The orthogonality constraints in (1.9) can be represented as a matrix function of $\alpha$, $\beta$, and $\gamma$:

$$g(\alpha, \beta, \gamma) = \begin{bmatrix} \alpha^T \gamma \\ \gamma \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$ 

Writing the orthogonality assumptions in this functional form resembles testing a function of parameters in the usual linear model. We define

$$\frac{\partial}{\partial(\alpha, \beta, \gamma)} g(\alpha, \beta, \gamma) = \dot{g}(\alpha, \beta, \gamma).$$
Then the orthogonality assumptions add

$$\text{rank}\{\mathbf{g}(\alpha^*, \beta^*, \gamma^*)\} = \hat{d}f_E$$

degrees of freedom for error where $\alpha^*$, $\beta^*$, and $\gamma^*$ satisfy the orthogonality assumptions. Using numerical differentiation and simulation methods, this idea can be used to find the error degrees of freedom for the general case of $a \times b$ data. Using $a - 1$ and $b - 1$ degrees of freedom for the main effects and if $\hat{d}f_E$ represents the estimated error degrees of freedom, then the degrees of freedom assigned to interaction is estimated by $\hat{d}f_{AB} = ab - a - b - \hat{d}f_E + 1$. We call this method for assigning degrees of freedom the rank method. Table 1.1 gives the error degrees of freedom computed using the rank method for several values of $a$ and $b$. The rank method typically yields $\hat{d}f_E = a + b - 3$, so when we reference the rank method in following sections, we will be referring to the use of $\hat{d}f_E = a + b - 3$. 

<table>
<thead>
<tr>
<th>$a$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>
The rank method described in this subsection provides degrees of freedom for error by subtracting degrees of freedom from the traditional \((a - 1) \times (b - 1)\) degrees of freedom for interaction. Because the orthogonality assumptions put constraints on the main effects and interactions, it may be inappropriate to provide degrees of freedom for error merely by subtracting degrees of freedom from interaction. A variation of the rank method involves rescaling the degrees of freedom for the main effects and interactions so that the orthogonality constraints impact them both. The degrees of freedom for error are found in the same way as above, but then define a constant \(r\) and let

\[
\hat{df}_A = r(a - 1), \quad \hat{df}_B = r(b - 1), \quad \text{and} \quad \hat{df}_{AB} = r(a - 1)(b - 1).
\]

The constant \(r\) is found by forcing the estimated degrees of freedom to sum to \(ab - 1\), resulting in

\[
r = \frac{1}{ab - 1 - \hat{df}_E}.
\]

Table 1.2 shows the estimated degrees for \(6 \times 6\) OI data using the rank method as well as the adjusted rank method (adjustment term \(r = .743\)). Using the adjusted rank degrees of freedom results in higher estimated degrees of freedom for interaction and lower estimated degrees of freedom for main effects.
Chapter 1. *The Orthogonal Interactions Model*

A resampling-based method can also be used to estimate the degrees of freedom for each source of variation in the orthogonal interactions model. The method involves adding noise to the data in varying magnitudes and observing the change in the average sums of squares. We present this method in Chapter 2. The resampling method offers more reliable estimates for the degrees of freedom, but it requires more computation than the rank method described in this section.

**Partitioning Sums of Squares**

After estimating the degrees of freedom for each source of variation, the next step in using the orthogonal interactions model is partitioning the sums of squares. We borrow from linear models theory again, using the following definitions of sums of squares

\[
SST = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \bar{Y}_{..})^2, \quad (1.22)
\]

\[
SS(A) = b \sum_{i=1}^{a} \hat{\alpha}_i^2, \quad SS(B) = a \sum_{j=1}^{b} \hat{\beta}_j^2,
\]

\[
SS(AB) = \sum_{j=1}^{b} \sum_{i=1}^{a} \hat{\gamma}_{ij}^2, \quad \text{and} \quad SSE = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \bar{Y}_{..} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij})^2.
\]

When fitting the OI model to replicated data, the sums of squares in (1.22) are modified by summing over the replicates. Using this partitioning of the sum of squares, it can be shown that the sums of squares are additive, meaning that the total sum of squares is equal to the sum of the sum of squares for the other sources of variation in the model. We now focus on showing that the sums of squares are additive by manipulating the
Chapter 1. The Orthogonal Interactions Model

total sum of squares:

\[ \text{SST} = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \bar{Y}_{..})^2 \]
\[ = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \bar{Y}_{..} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_{ij} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij})^2 \]
\[ = \sum_{j=1}^{b} \sum_{i=1}^{a} (\hat{e}_{ij} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij})^2 \]
\[ = \sum_{j=1}^{b} \sum_{i=1}^{a} \left\{ \hat{e}_{ij}^2 + 2\hat{e}_{ij}(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}) + (\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij})^2 \right\} \]
\[ = \text{SSE} + \text{SS(A)} + \text{SS(B)} + \text{SS(AB)} \]
\[ + 2 \sum_{j=1}^{b} \sum_{i=1}^{a} \left\{ \hat{e}_{ij}(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}) + 2\hat{\alpha}_i\hat{\beta}_j + 2\hat{\alpha}_i\hat{\gamma}_{ij} + 2\hat{\beta}_j\hat{\gamma}_{ij} \right\} . \]

From here, we know that:

\[ \sum_{j=1}^{b} \sum_{i=1}^{a} \hat{\alpha}_i\hat{\beta}_j = \sum_{j=1}^{b} \hat{\beta}_j \left( \sum_{i=1}^{a} \hat{\alpha}_i \right) = 0, \]
\[ \sum_{j=1}^{b} \sum_{i=1}^{a} \hat{\alpha}_i\hat{\gamma}_{ij} = \sum_{i=1}^{a} \hat{\alpha}_i \left( \sum_{j=1}^{b} \hat{\gamma}_{ij} \right) = 0, \]
and

\[ \sum_{j=1}^{b} \sum_{i=1}^{a} \hat{\beta}_j\hat{\gamma}_{ij} = \sum_{j=1}^{b} \hat{\beta}_j \left( \sum_{i=1}^{a} \hat{\gamma}_{ij} \right) = 0. \]
Chapter 1. The Orthogonal Interactions Model

Now the total sum of squares can be written:

$$\text{SST} = \text{SSE} + \text{SS}(A) + \text{SS}(B) + \text{SS}(AB) + 2 \sum_{j=1}^{b} \sum_{i=1}^{a} \left\{ \hat{e}_{ij}(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}) \right\}. $$

Showing that $\sum_j \sum_i \hat{e}_{ij}(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}) = 0$ would be sufficient to prove that the sums of squares are additive. We will do this by showing that

$$\sum_j \sum_i \hat{e}_{ij}\hat{\alpha}_i = \sum_j \sum_i \hat{e}_{ij}\hat{\beta}_j = \sum_j \sum_i \hat{e}_{ij}\hat{\gamma}_{ij} = 0. $$

Using the results in Appendix A.1,

$$\sum_{j=1}^{b} \sum_{i=1}^{a} \hat{e}_{ij}\hat{\alpha}_i = \text{tr}(\hat{e}^T\hat{\alpha}1_b^T)$$

$$= \text{tr}\left\{ (Y - \bar{Y}\cdot1_a1_b^T - \hat{\alpha}_i1_b^T - 1_a\hat{\beta}^T - \hat{\gamma})^T\hat{\alpha}1_b^T \right\}$$

$$= \text{tr}\left\{ (Y^T - \bar{Y}\cdot1_b1_a^T - 1_b\hat{\alpha}^T - \hat{\beta}1_a^T - \hat{\gamma}^T)^T\hat{\alpha}1_b^T \right\}$$

$$= \text{tr}(Y^T\hat{\alpha}1_b^T - \bar{Y}\cdot1_b1_a^T\hat{\alpha}1_b^T - 1_b\hat{\alpha}^T\hat{\alpha}1_b^T - \hat{\beta}1_a^T1_b^T - \hat{\gamma}^T\hat{\alpha}1_b^T)$$

$$= \text{tr}(Y^T\hat{\alpha}1_b^T - 1_b\hat{\alpha}^T\hat{\alpha}1_b^T)$$

$$= \text{tr}\left\{ Y^T(I - P_U)\bar{Y}\cdot1_b^T - 1_b\bar{Y}\cdot1_a^T(I - P_U)\bar{Y}\cdot1_b^T \right\}$$

$$= \text{tr}\left\{ 1_b^TY^T(I - P_U)\bar{Y}\cdot1_a^T - 1_b^TY^T(I - P_U)\bar{Y}\cdot1_a^T \right\}$$

$$= \text{tr}\left\{ 1_b^TY^T(I - P_U)\bar{Y}\cdot1_a^T - b\bar{Y}\cdot1_a^T(I - P_U)\bar{Y}\cdot1_a^T \right\}. $$
Chapter 1. The Orthogonal Interactions Model

Remember that \( \bar{Y}_a = b^{-1}Y_b - \bar{Y}_a 1_a \), which means that \( Y_b 1_a = b \left( \bar{Y}_a^T + \bar{Y}_a 1_a \right) \).

\[
\sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\alpha}_i = \text{tr} \left\{ b \left( \bar{Y}_a^T + \bar{Y}_a 1_a \right) (I - P_U) \bar{Y}_a - b \bar{Y}_a^T (I - P_U) \bar{Y}_a \right\} = \text{tr} \left\{ b \bar{Y}_a (I - P_U) \bar{Y}_a - b \bar{Y}_a^T (I - P_U) \bar{Y}_a \right\}, (P_U \text{ projects onto } 1_a) = 0.
\]

We use a similar argument to show that \( \sum_j \sum_i \hat{e}_{ij} \hat{\beta}_j = 0 \). Using results from Appendix A.1,

\[
\sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\beta}_j = \text{tr} \left( \hat{e}^T 1_a \hat{\beta}^T \right) = \text{tr} \left\{ (Y - \bar{Y}_a 1_a)^T 1_b - \hat{\alpha}_i 1_b^T - 1_a \hat{\beta}^T - \hat{\gamma}^T 1_a \hat{\beta}^T \right\} = \text{tr} \left( Y^T 1_a \hat{\beta}^T - a \bar{Y}_b \hat{\beta}^T 1_b + a \bar{Y}_a \hat{\beta}^T 1_a \hat{\beta}^T - \hat{\gamma}^T 1_a \hat{\beta}^T \right) = \text{tr} \left( Y^T 1_a \hat{\beta}^T - \hat{\beta}^T 1_a \hat{\beta}^T \right) = \text{tr} \left\{ Y^T 1_a \bar{Y}_b^T (I - P_V) - a (I - P_V) \bar{Y}_b \bar{Y}_b^T (I - P_V) \right\} = \text{tr} \left( \bar{Y}_b^T (I - P_V) Y^T 1_a - a \bar{Y}_b^T (I - P_V) \bar{Y}_b \right)\right\}.
\]

Remember that \( \bar{Y}_b = a^{-1}Y^T 1_a - \bar{Y}_a 1_b \) and that \( Y^T 1_a = a (\bar{Y}_b + \bar{Y}_b 1_b) \). Then,

\[
\sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\beta}_j = \text{tr} \left\{ a \bar{Y}_b^T (I - P_V) (\bar{Y}_b + \bar{Y}_a 1_b) - a \bar{Y}_b^T (I - P_V) \bar{Y}_b \right\} = \text{tr} \left\{ a \bar{Y}_b^T (I - P_V) \bar{Y}_b - a \bar{Y}_b^T (I - P_V) \bar{Y}_b \right\}, (P_V \text{ projects onto } 1_b) = 0.
\]
Chapter 1. The Orthogonal Interactions Model

The last step in showing that the sums of squares are additive is to show that \( \sum_j \sum_i \hat{e}_{ij} \hat{\gamma}_{ij} = 0 \). Again, using results from Appendix A.1,

\[
\sum_{j=1}^{b} \sum_{i=1}^{a} \hat{e}_{ij} \hat{\gamma}_{ij} = \text{tr} ( \hat{\gamma}^T \hat{\gamma} )
\]

\[
= \text{tr} \left\{ \left( Y - \overline{Y} \cdot 1_a 1_b^T - \hat{\alpha}_1 1_b^T - 1_a \hat{\beta}^T - \hat{\gamma} \right)^T \hat{\gamma} \right\}
\]

\[
= \text{tr} \left\{ \left( Y^T - \overline{Y} \cdot 1_b 1_a^T - 1_b \hat{\alpha}^T - \hat{\beta} 1_a^T - \hat{\gamma}^T \right) \hat{\gamma} \right\}
\]

\[
= \text{tr}(Y^T \hat{\gamma} - \overline{Y} \cdot 1_b \underbrace{1_a^T - 1_b \hat{\alpha}^T - \hat{\beta} 1_a^T - \hat{\gamma}^T}_{0} \underbrace{0}_{0} - \overbrace{0} \underbrace{0}_{0} \underbrace{0}_{0})
\]

\[
= \text{tr}(Y^T \hat{\gamma} - \hat{\gamma}^T \hat{\gamma})
\]

\[
= \text{tr}\left\{ Y^T (I - P_A) r (I - P_B) - (I - P_B) r^T (I - P_A) r (I - P_B) \right\}
\]

\[
= \text{tr}\left\{ Y^T (I - P_A) r (I - P_B) - (I - P_B) r^T (I - P_A) r \right\}.
\]

Now recall that \( r = Y - \overline{Y} \cdot 1_a 1_b^T - \hat{\alpha} 1_b^T - 1_a \hat{\beta}^T \). Then

\[
\sum_{j=1}^{b} \sum_{i=1}^{a} \hat{e}_{ij} \hat{\gamma}_{ij} = \text{tr}\{ Y^T (I - P_A) r (I - P_B) \}
\]

\[
- (Y - \overline{Y} \cdot 1_b 1_a^T - 1_b \hat{\alpha}^T - \hat{\beta} 1_a^T)^T (I - P_A) r (I - P_B) \}
\]

\[
= \text{tr}\left\{ Y^T (I - P_A) r (I - P_B) \right\} - \text{tr}(\kappa).
\]
Chapter 1.  The Orthogonal Interactions Model

Now focusing on \(\text{tr}(\kappa)\),

\[
\text{tr}(\kappa) = \text{tr}\left\{ (Y - \bar{Y}.1_a1_b^T - \hat{\alpha}_b1_b^T - 1_a\hat{\beta}_a^T)(I - P_A)r(I - P_B) \right\}
= \text{tr}\left\{ (Y^T - \bar{Y}.1_b1_a^T - 1_b\hat{\alpha}_a^T - \hat{\beta}_a1_a^T)(I - P_A)r(I - P_B) \right\}
= \text{tr}\{Y^T(I - P_A)r(I - P_B) - (Y.1_b1_a^T - 1_b\hat{\alpha}_a^T - \hat{\beta}_a1_a^T)(I - P_A)r(I - P_B)\}
= 0 \text{ because } P_A \text{ projects onto } 1_a \text{ and } \hat{\alpha}.
\]

So then:

\[
\sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\gamma}_{ij} = \text{tr}\{Y^T(I - P_A)r(I - P_B)\} - \text{tr}\{Y^T(I - P_A)r(I - P_B)\}
= 0.
\]

Now we have shown that \(\sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\alpha}_i = \sum_{j=1}^b \sum_{i=1}^a \hat{e}_{ij} \hat{\beta}_j = \sum_{i=1}^a \sum_{j=1}^b \hat{e}_{ij} \hat{\gamma}_{ij} = 0\),
which is sufficient to prove that

\[
\text{SST} = \text{SS}(A) + \text{SS}(B) + \text{SS}(AB) + \text{SSE},
\]

where the sums of squares are defined in (1.22).

1.6.2 Inference and Existing Methods

In order to perform inference on the AMMI model, degrees of freedom must be assigned
to each source of variation in the model. Several methods exist for assigning degrees
of freedom in the AMMI model. Since the main effects part of the model is additive,
Chapter 1. The Orthogonal Interactions Model

the degrees of freedom for the main effects are the same as in the additive ANOVA model. Gauch [2] provides a survey of the different methods of estimating the degrees of freedom for the interaction terms in the AMMI model. For $a \times b$ data, the most simple way of assigning degrees of freedom to this model is to assign $a + b - 1 - 2m$ degrees of freedom to the $m^{th}$ PCA axis included in the model. This method of assigning degrees of freedom is called Gollub’s Rule. Cornelius [6] suggests that the Gollub degrees of freedom lead to high Type I error rates in tests for main effects and interactions.

Other methods for assigning degrees of freedom to the AMMI model involve simulation. The pure noise method estimates degrees of freedom for each PCA axis for general $a \times b$ data, independent of the data of interest. For each of the $r$ simulation replicates, an $a \times b$ matrix $e$ is created where $e_{ij} \sim N(0, \sigma^2)$. If we are interested in estimating degrees of freedom for the AMMI($m$) model, we must fit the main effects ANOVA model to $e$, as well as the AMMI(1) through AMMI($m$) models. For each simulated data matrix $e$, sums of squares are calculated for each PCA axis using the following equations:

$$SS(PCA 1) = SSE(\text{Main Effects ANOVA}) - SSE\{\text{AMMI}(1)\}$$

$$SS(PCA 2) = SSE\{\text{AMMI}(1)\} - SSE\{\text{AMMI}(2)\}$$

$$\vdots$$

$$SS(PCA m) = SSE\{\text{AMMI}(m)\} - SSE\{\text{AMMI}(m-1)\}.$$ 

Let $SS_j(PCA i)$ represent the sum of squares attributed to the $i^{th}$ principal component axis for the $j^{th}$ simulated data set. For each simulated data set, the degrees of freedom
Chapter 1. The Orthogonal Interactions Model

For each principal component axis are estimated using the following equations:

\[
\hat{df}_j(\text{PCA 1}) = \frac{SS_j(\text{PCA 1})}{\sigma^2}
\]
\[
\hat{df}_j(\text{PCA 2}) = \frac{SS_j(\text{PCA 2})}{\sigma^2}
\]
\[
\vdots
\]
\[
\hat{df}_j(\text{PCA } m) = \frac{SS_j(\text{PCA } m)}{\sigma^2}.
\]

Averaging over simulation replicates yields the following estimate for the degrees of freedom associated with the \(i^{th}\) principal component axis:

\[
\hat{df}(\text{PCA } i) = r^{-1} \sum_{j=1}^{r} \hat{df}_j(\text{PCA } i).
\]

The estimated degrees of freedom for error are then obtained by subtraction. For simplicity, \(\sigma^2 = 1\) is a natural choice for the variance of the generated noise. The pure noise method assumes that there is no signal in the data, which may limit interest in this method because it does not make use of the data of interest. Another complication with this method is that it could yield a negative estimate for the error degrees of freedom. If the degrees of freedom for each PCA axis is overestimated, then the estimated error degrees of freedom may be negative. The method does not offer an obvious correction if the estimated error degrees of freedom is negative, but a negative estimate may suggest that fewer PCA axes should be included in the model.

The \textit{pattern plus noise} method is similar to the pure noise method, but it makes use of the data of interest. To use this method, first fit the main effects ANOVA model to the data and estimate the error variance, \(\hat{\sigma}_1^2 = \text{MSE}\). Then each simulated data set
Chapter 1. *The Orthogonal Interactions Model*

is generated by adding $N(0, \sigma_2^2)$ noise to the response variable. The AMMI(1) through AMMI($m$) models are fit to each simulated data set. Partitioning the sum of squares for each PCA axis in the same way as the pure noise method, the degrees of freedom for the $i^{th}$ PCA axis for the $j^{th}$ simulation replicate is estimated by:

$$\hat{df}_j(\text{PCA } i) = \frac{SS_j(\text{PCA } i)}{\hat{\sigma}_1^2 + \sigma_2^2}.$$  

Averaging over the simulation replicates yields the degrees of freedom estimate

$$\hat{df}(\text{PCA } i) = r^{-1} \sum_{j=1}^r \hat{df}_j(\text{PCA } i)$$

for the $i^{th}$ principal component axis in the model. Just as with the pure noise method, the estimated error degrees of freedom are obtained by subtraction, which could potentially yield a negative estimate.

Unlike the AMMI model, assigning degrees of freedom to Tukey’s model is straightforward. For $a \times b$ data, experimental factors $A$ and $B$ are assigned $a - 1$ and $b - 1$ degrees of freedom respectively. As the name suggests, the multiplicative interaction parameter for the model is assigned a single degree of freedom.

After fitting the AMMI model or the Tukey one-degree-of-freedom model and estimating degrees of freedom, inference on these models is simple. As in linear models, $F$ ratios are formed to test for main effects and interaction using the error mean square as the denominator. For example, testing the significance of the $i^{th}$ principal component
axis in the AMMI model can be done using

\[ F = \frac{\text{MS}(\text{PCA } i)}{\text{MSE}} , \]

and comparing \( F \) to the critical value of the appropriate reference distribution.

### 1.7 Recognizing Data from the Orthogonal Interactions Model

When working with data from a two-factor experiment, investigators often look at mean profiles to visualize interactions. When fitting the ANOVA model for example, parallel mean profiles suggest that interactions should not be included in the model. We investigated several simulated data sets to see if mean profiles can be used to determine whether or not the orthogonal interactions model is appropriate for a given data set. After considering these simulated data sets, we believe that although mean profiles do provide evidence of interactions, they provide little evidence in determining if interactions are of the form of the OI model.

A set of main effect and interaction parameters \((\alpha, \beta, \text{ and } \gamma_1)\) were created randomly to satisfy the assumptions of an \(8 \times 4\) ANOVA model. Then using the reparameterization described in Section 1.4.1, \(\gamma_1\) was reparameterized to create \(\gamma_2\) so that \(\alpha, \beta, \text{ and } \gamma_2\) satisfy the OI model assumptions. Figure 1.1 displays population mean profiles for the two different sets of parameters. The means in Figure 1.1(a) come from the orthogonal interactions model, while the means in Figure 1.1(b) come from a two-factor ANOVA model with interactions. The profiles in Figure 1.1 are similar, suggesting that a mean
profile from the OI model can be similar to a mean profile from the ANOVA model. The similarity in the two profiles suggests that mean profiles are not a useful tool in determining if the orthogonal interactions model is appropriate for a particular data set. However, the profiles also suggest that the OI model is capable of modeling ‘real’ data well because predictions from the OI model can be very similar to predictions from the two-factor ANOVA model with interactions.

When considering the mean profiles in Figure 1.1, it is important to consider how the means were generated. Because $\alpha$, $\beta$, and $\gamma_1$ were generated using independent random normal variates, they essentially satisfy a random effects model. Since the random effects model provides the motivation for the orthogonal interactions model, the similarity between the population mean profiles in Figure 1.1 should not be unexpected. If the main effects and interactions are not generated independently, then the
Chapter 1. The Orthogonal Interactions Model

mean profile for the ANOVA model may look much different than the profile for the orthogonal interactions model. However, in Section 1.9 we present real data where the prediction profile from the OI model is very similar to the prediction profile from the ANOVA with interactions.

Using Biplots to Aid in Model Selection

Biplots are used frequently in determining what type of interactions are appropriate for a particular data set. See Bradu and Gabriel [9] or Crossa, Cornelius, and Yan [10] for more information on using the biplot for studying interaction. For unreplicated two-factor data with $a$ levels of factor $A$ and $b$ levels of factor $B$, let $r$ be the $a \times b$ matrix of residuals from the main effects ANOVA model. Then obtain the singular value decomposition of $r$ as described in Section 1.5.1, yielding $r = UQV^T$. The matrices $U$ and $V$ can be represented as

$$U = \begin{bmatrix} u_1, & u_2, & \ldots & u_b \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_1, & v_2, & \ldots & v_b \end{bmatrix},$$

where $u_i$ are $a \times 1$ eigenvectors and $v_i$ are $b \times 1$ eigenvectors. Let $Y_1$ denote the $a \times 1$ vector of row means of $Y$ and let $Y_2$ denote the $b \times 1$ vector of column means of $Y$. In other words,

$$\overline{Y}_{1,i} = b^{-1} \sum_j Y_{ij} \quad \text{and} \quad \overline{Y}_{2,j} = a^{-1} \sum_i Y_{ij}.$$

The biplot is obtained by plotting $u_1$ versus $\overline{Y}_1$ and $v_1$ versus $\overline{Y}_2$ on the same set of axes.

The biplot is a useful tool in determining if the interactions are of the Tukey one-
Chapter 1. The Orthogonal Interactions Model

degree-of-freedom form or if they are of the form of the linear-bilinear model. If all of the points in the biplot lay on a single line, then Tukey’s model is appropriate for the data. If the points associated with the row effects are linear (or likewise for the column effects), then the linear-bilinear model is appropriate for the data.

Figure 1.2 shows example biplots for data from Tukey’s model as well as data from a linear-bilinear model. Tukey’s model is clearly appropriate for the data in Figure 1.2(a) because all of the points lay on a single line. The biplot in Figure 1.2(b) clearly comes from a linear-bilinear model because the row effects (represented by circles) are linear while the column effects (represented by triangles) are not linear.

Similar to the linear-bilinear model, the AMMI model is appropriate for data if the column effects lay on multiple lines and the row effects also lay on multiple lines in the biplot. If the number of levels of each experimental factor is small, then the biplot may not have enough points to aid in deciding on what model is appropriate. For three levels of each experimental factor, the biplot would only consist of six points, which would not be sufficient for drawing any conclusions about the nature of the interactions.

After inspecting many biplots of data generated from the orthogonal interactions model, we do not find a feature unique to biplots of data from the OI model. This is due to the fact that interactions from the OI model are not multiplicative in nature as they are in the Tukey, linear-bilinear, and AMMI models. Because the biplot uses the singular value decomposition of the residuals, it is effective in detecting multiplicative interactions. So looking at a biplot does not immediately allow investigators to determine whether or not the OI model is appropriate for data. However, if the biplot does not clearly suggest that Tukey’s model, the linear-bilinear model, or the AMMI model
Figure 1.2: Biplots of: (a) Tukey Data (b) Linear-bilinear Data
are appropriate for the data, then the OI model is a possible choice.

1.8 Testing the Orthogonality Assumptions

For replicated data, fitting the orthogonal interactions model instead of the full ANOVA model may result in a more efficient estimate of the error variance. Although the orthogonal interactions model is not a linear model, it is in fact nested within the full two-factor ANOVA model with interactions. The model is nested within the full ANOVA model in the sense that when both models are fit to data, the error sum of squares for the OI model will always be greater than or equal to the error sum of squares for the full ANOVA model. If the OI and ANOVA models are written \( E(Y) = X\theta_1 \) and \( E(Y) = X\theta_2 \) respectively, let \( \Theta_1 \) represent all possible solutions for \( \theta_1 \) and let \( \Theta_2 \) represent all possible solutions for \( \theta_2 \). We know that \( \Theta_1 \subset \Theta_2 \) because the OI model is a special case of the full ANOVA model. Since \( \Theta_1 \subset \Theta_2 \), we know that the error sum of squares for the OI model is greater than or equal to the error sum of squares for the ANOVA model, and therefore the OI model is nested within the ANOVA model. Because of this nesting, the likelihood ratio test is the natural choice for testing the orthogonality assumptions for replicated data. The remainder of this section discusses using the likelihood ratio test for testing the appropriateness of the orthogonality assumptions.

We are interested in testing the hypothesis

\[
H_0 : \alpha^T \gamma = 0, \gamma \beta = 0 \quad \text{vs.} \quad H_1 : \alpha^T \gamma \neq 0 \text{ or } \gamma \beta \neq 0,
\]
Chapter 1. The Orthogonal Interactions Model

which is equivalent to

\[ H_0 : \text{OI model sufficient} \quad \text{vs.} \quad H_1 : \text{Full ANOVA necessary.} \quad (1.23) \]

The hypothesis written in (1.23) is a full-versus-reduced model comparison where the full model does not require the model parameters to satisfy the orthogonality constraints. If \( l_0 \) represents the log-likelihood of the reduced model and \( l_1 \) represents the log-likelihood of the full model, then

\[ W = 2(l_1 - l_0) \quad (1.24) \]

is the usual likelihood ratio test statistic. The test statistic in (1.24) is compared to the \( \chi^2_d \) distribution where \( d \) is equal to the difference in the number of parameters in the two models (see Section 1.6.1 for estimating degrees of freedom for the OI model). If the rank degrees of freedom are used for the OI model, then the likelihood ratio test is compared to a \( \chi^2_{a+b-3} \) distribution where \( a \) and \( b \) are the levels of experimental factors \( A \) and \( B \) respectively.

The problem with the likelihood ratio test statistic defined in (1.24) is that the statistic converges to the \( \chi^2 \) distribution as \( c \), the number of replicates per factor-level combination, tends to infinity. For small \( c \), the Type I error rate can be very high as we show later. Real data typically do not feature a large enough number of replicates to avoid this problem of inflated Type I error rate, so the test statistic must be adjusted to correct this problem. Bartlett ([11] and [12]) and Lawley [13] provide the basis for the Bartlett correction, a parameter that can be multiplied by the likelihood ratio
Chapter 1. The Orthogonal Interactions Model

test statistic in order to improve the quality of the $\chi^2$ approximation for small sample sizes (small $c$). More recent work by Barndorff-Nielsen and Cox [14] and Cordeiro [15] provide more justification for the Bartlett correction and extend the correction to nonlinear models. Simulations by Cordeiro [15] show that the Bartlett correction can reduce the Type I error rate from 30% to a more acceptable 8% in nonlinear models. Since the OI model is nonlinear, we expect that the Bartlett correction can also reduce the Type I error rate for the likelihood ratio test of (1.23). We do not provide a general form for the Bartlett correction to the test statistic in (1.24) though it might be possible with extensive algebraic manipulations. In the following subsections we provide simulation-based corrections to the likelihood ratio test in the spirit of the Bartlett correction.

1.8.1 A Bootstrap Corrected Likelihood Ratio Test

Suppose that investigators are interested in comparing two models by testing the hypothesis

$$H_0 : \text{Model I sufficient} \quad \text{vs.} \quad H_1 : \text{Model II necessary},$$

where Model I is nested within Model II and there are $d$ more parameters in Model II than in Model I. Then the likelihood ratio statistic $W$ as defined in (1.24) converges in distribution to $\chi^2_d$ under the null hypothesis. But for small sample sizes, $E(W \mid H_0) \neq d$ in general. The idea behind the Bartlett correction is to find a parameter $k$ such that $E(kW \mid H_0) = d$, even for small sample sizes. Finding a general form for the estimator of $k$ is troublesome and algebraically intense. In this subsection we consider two different bootstrap Bartlett corrections to the likelihood ratio test of the appropriateness of the
Chapter 1. The Orthogonal Interactions Model

orthogonal interactions model for replicated data. Rocke [16] was the first to suggest using bootstrap methods to correct the likelihood ratio test.

We are interested in testing the appropriateness of the orthogonality assumptions by testing the hypothesis stated in (1.23). In the same spirit as the Bartlett correction, we seek to find a constant $k$ such that $E(kW \mid H_0) = d$, where $d$ is the difference in the number of parameters between the OI and ANOVA models. After fitting the OI and ANOVA models to the data vector $Y$, let $\hat{Y}_0$ represent the vector of predicted values from the OI model. Then let

$$\tilde{Y}_j = \hat{Y}_0 + e_j \quad j = 1, ..., r; \quad (1.25)$$

be vectors of the same dimension as $Y$. In (1.25), $e_j$ is a bootstrap sample of residuals from the OI model and $r$ is the number of replicates chosen for the simulation. As usual, the bootstrap sample is obtained from a simple random sample with replacement from the vector $e_0 = Y - \hat{Y}_0$. So each of the simulated data sets, $\tilde{Y}_j$, is generated according to the null model (the OI model). For each of the generated data sets, the reduced and full models are fit to the data and the log-likelihoods are calculated. Then a likelihood ratio test statistic is formed for each replicate

$$W_j = 2(l_{1,j} - l_{0,j}),$$

where $l_{1,j}$ and $l_{0,j}$ are the log-likelihoods of the ANOVA and OI models respectively for the $j^{th}$ simulated data set. Then the distribution of $W$ under the null hypothesis should be very similar to the distribution of the $W_j$. Under the null hypothesis, the
Chapter 1. The Orthogonal Interactions Model

sample mean of the $W_j$ is an appropriate estimator of the mean of $W$. Then using

$$\hat{k} = \bar{W}^{-1}d, \quad \text{where} \quad \bar{W} = r^{-1}\sum_j W_j,$$

the expected value of $\hat{k}W$ should be approximately $d$ under the null hypothesis ($H_0$: OI model is sufficient). Comparing $W^* = \hat{k}W$ to the reference $\chi^2_d$ distribution yields $\hat{p}_1$ as the p-value for the likelihood ratio test stated in (1.23). We would reject the null hypothesis in (1.23) when $\hat{p}_1$ is less than the nominal testing level.

The second bootstrap method for testing (1.23) uses the empirical distribution of the likelihood ratio test statistic under the null model. The sample distribution of the $W_j$ is a natural estimator of the distribution of the likelihood ratio test statistic under the null model (OI model). So an empirical p-value can be obtained using

$$\hat{p}_2 = r^{-1}\sum_{j=1}^r I(W_j \geq W),$$

where $I(W_j \geq W) = 1$ when $W_j \geq W$, and zero otherwise. The test rejects when $\hat{p}_2$ is less than or equal to the nominal testing level.

So using bootstrap methodology, we can compute p-values from an adjusted likelihood ratio test statistic or we can obtain p-values from the empirical distribution of the likelihood ratio test statistic under the null hypothesis. These simulation methods are not complicated to implement, they merely require choosing bootstrap samples and fitting the full and reduced models. For simple models, the number of simulation replicates $r$ can be chosen to be arbitrarily large because computation time is not an issue. However for more complicated models such as the orthogonal interactions
Chapter 1. The Orthogonal Interactions Model

model, computation times can restrict the magnitude of $r$. These bootstrap methods are attractive because they do not require an analytic computation of the Bartlett correction. The rest of this subsection looks at a simulation that evaluates the performance of tests conducted using $\hat{p}_1$ and $\hat{p}_2$, as well as the unadjusted likelihood ratio test of the hypothesis stated in (1.23).

One thousand data sets were generated from the orthogonal interactions model with two factors ($a$ and $b$ levels of each factor respectively) and $c$ replicates at each factor level combination ($a \times b \times c$ OI data). For each simulated data set, we tested the hypothesis in (1.23) using an unadjusted likelihood ratio test statistic as well as the bootstrap tests using $\hat{p}_1$ and $\hat{p}_2$. For the sake of simplicity, we use the rank method to estimate degrees of freedom in the OI model so that the difference in number of parameters in the OI and ANOVA models is $d = a + b - 3$. Using the resampling method to assign degrees of freedom is straightforward, but would require significantly more computation. The hypothesis was tested at the $\alpha = .05$ level and for each generated data set, a simulation with $r = 100$ replicates was performed to obtain $\hat{p}_1$ and $\hat{p}_2$.

The simulation results are summarized in Table 1.3, where the columns headed by ‘LRT’ represent the rejection rates for the unadjusted likelihood ratio test, the columns headed by ‘$\hat{p}_1$’ represent the rejection rates for tests conducted using $\hat{p}_1$, and the columns headed by ‘$\hat{p}_2$’ represent the rejection rates for tests conducted using $\hat{p}_2$. Clearly the bootstrap tests come closer to the nominal testing level than the unadjusted likelihood ratio test. The table also suggests that the unadjusted likelihood ratio test statistic is converging to the $\chi^2_{a+b-3}$ distribution as $c$ tends to infinity. The high rates of Type I errors for the unadjusted likelihood test (up to 57%) are not acceptable, while
Table 1.3: Type I Error Rates for Bootstrap Corrected LRT of Orthogonality Assumptions for $a \times b$ Data with $c$ Replicates (Monte Carlo SE: 6.9E-3)

<table>
<thead>
<tr>
<th>$a \times b$</th>
<th>$c=2$ LRT</th>
<th>$\hat{p}_1$</th>
<th>$\hat{p}_2$</th>
<th>$c=5$ LRT</th>
<th>$\hat{p}_1$</th>
<th>$\hat{p}_2$</th>
<th>$c=10$ LRT</th>
<th>$\hat{p}_1$</th>
<th>$\hat{p}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 $\times$ 4</td>
<td>.286 .056 .053</td>
<td></td>
<td></td>
<td>.081 .054 .062</td>
<td></td>
<td></td>
<td>.073 .051 .060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 $\times$ 6</td>
<td>.377 .050 .041</td>
<td></td>
<td></td>
<td>.116 .053 .046</td>
<td></td>
<td></td>
<td>.085 .050 .063</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 $\times$ 8</td>
<td>.412 .047 .046</td>
<td></td>
<td></td>
<td>.127 .050 .050</td>
<td></td>
<td></td>
<td>.082 .051 .065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 $\times$ 10</td>
<td>.572 .046 .039</td>
<td></td>
<td></td>
<td>.145 .050 .042</td>
<td></td>
<td></td>
<td>.076 .041 .048</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Type I error rates for the bootstrap tests are more reasonable (between 3.9% and 6.5%). Tests based on $\hat{p}_1$ come closer to the nominal testing level than the tests based on $\hat{p}_2$.

The improved performance by using one of the bootstrap tests can also be seen in the histogram of the p-values for each of the tests. When the null hypothesis is true, the p-values should be uniformly distributed between zero and one. Figure 1.3 shows the histogram of simulated p-values for the unadjusted likelihood ratio test statistic in the case of $4 \times 4 \times 2$ OI data. Histograms of the p-values associated with the bootstrap adjusted tests are in Figure 1.4 and come closer to uniformity than the p-values associated with the unadjusted likelihood ratio test.

Under the null hypothesis, we would like for the likelihood ratio test statistic to have approximately a $\chi^2_{a+b-3}$ distribution. Figure 1.5 shows the histogram of values of the simulated unadjusted likelihood ratio test statistic values as well as the $\chi^2_5$ reference distribution for the case of $4 \times 4 \times 2$ OI data. Clearly the distribution of the unadjusted test statistic has a heavy tail which leads to the high rate of Type I
Chapter 1. The Orthogonal Interactions Model

Figure 1.3: Histogram of Simulated Unadjusted LRT p-values

errors. In Figure 1.6 we see that the adjusted likelihood ratio test statistic comes much closer to the $\chi^2_5$ distribution. These two figures show that under the null hypothesis of orthogonal interactions, the adjusted likelihood ratio test statistic is more appropriate because the distribution is closer to the $\chi^2$ reference distribution. A goodness of fit test was performed on the adjusted likelihood ratio test statistic values in Figure 1.6 using the tests readily available with SAS Proc UNIVARIATE (Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling). None of the goodness of fit tests reject the null hypothesis that the $\chi^2_5$ distribution is appropriate for the adjusted test statistic ($p \geq .15$ for all three tests). Meanwhile, all three goodness of fit tests reject for the unadjusted values in Figure 1.5 ($p < .0001$ for all three tests).
Figure 1.4: Histograms of Bootstrap Adjusted p-values for Tests Based on: (a) $\hat{p}_1$  (b) $\hat{p}_2$
Figure 1.5: *Histogram of Unadjusted Likelihood Ratio Test Statistic and $\chi^2$ Distribution*

We conducted a similar simulation to investigate the power of the bootstrap adjusted test to detect data from the ANOVA model. One thousand $4 \times 4 \times 2$ data sets were generated from the ANOVA model with interactions, where the orthogonality assumptions are not necessarily true. Again, a simulation with $r = 100$ replicates was used to perform the bootstrap tests for each replicate of data. The main effect and interaction parameters were generated such that $\sum_i \alpha_i^2 = 2$, $\sum_j \beta_j^2 = 2$, and $\sum_i \sum_j \gamma_{ij}^2 = 4$, but the orthogonality constraints were not enforced. Here the rejection rate for the unadjusted likelihood ratio test was 96.3% while the rejection rates for the
tests based on $\hat{p}_1$ and $\hat{p}_2$ are 85.4% and 85.0% respectively. So we have found that while the bootstrap adjusted likelihood ratio tests do not lead to high Type I error rates, the tests still exhibit power to detect data from the ANOVA model.

## 1.8.2 A Noise Corrected Likelihood Ratio Test

In Section 1.8.1, we added bootstrap errors to predictions from the orthogonal interactions model in an attempt to correct the likelihood ratio test in (1.23). Instead of
Chapter 1. The Orthogonal Interactions Model

Table 1.4: Type I Error Rates for Noise Corrected LRT of Orthogonality Assumptions for $a \times b$ Data with c Replicates (Monte Carlo SE: 6.9E-3)

<table>
<thead>
<tr>
<th>$a \times b$</th>
<th>c=2</th>
<th>c=5</th>
<th>c=10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LRT</td>
<td>$\tilde{p}_1$</td>
<td>$\tilde{p}_2$</td>
</tr>
<tr>
<td>4 $\times$ 4</td>
<td>.286</td>
<td>.061</td>
<td>.069</td>
</tr>
<tr>
<td>6 $\times$ 6</td>
<td>.377</td>
<td>.049</td>
<td>.056</td>
</tr>
<tr>
<td>8 $\times$ 8</td>
<td>.412</td>
<td>.040</td>
<td>.047</td>
</tr>
<tr>
<td>10 $\times$ 10</td>
<td>.572</td>
<td>.033</td>
<td>.049</td>
</tr>
</tbody>
</table>

using a bootstrap sample of errors, random noise could be added to the predictions instead. In other words, if $\hat{Y}_0$ represents the vector of predictions from the OI model, let

$$Y_j = \hat{Y}_0 + \hat{\sigma} z_j,$$

(1.26)

where $z_j$ is a vector of standard normal variates and $\hat{\sigma}^2$ is the maximum likelihood estimate of error variance from the OI model: $\hat{\sigma}^2 = (abc)^{-1} (Y - \hat{Y}_0)^T (Y - \hat{Y}_0)$. Using the same methods as in Section 1.8.1, p-values can be obtained from the adjusted likelihood ratio test statistic or the empirical distribution of the $W_j$ values. We denote the p-values from the adjusted likelihood ratio test statistic $\tilde{p}_1$ and we use $\tilde{p}_2$ to represent the p-values based on the empirical distribution of the adjusted test statistic.

A simulation similar to the one conducted in Section 1.8.1 (one thousand simulated data sets and $r = 100$ replicates to perform tests) was used to assess the performance of tests based on $\tilde{p}_1$ and $\tilde{p}_2$. The rejection rates are summarized in Table 1.4 where columns headed by ‘LRT’ represent the rejection rates of the unadjusted likelihood ratio test, columns headed by ‘$\tilde{p}_1$’ represent the rejection rates of tests conducted using
Chapter 1. The Orthogonal Interactions Model

\( \tilde{p}_1 \), and columns headed by ‘\( \tilde{p}_2 \)’ represent the rejection rates of tests conducted using \( \tilde{p}_2 \). Again, the rejection rates for the adjusted tests come closer to the nominal testing level than the rejection rates of the unadjusted test.

Histograms of the simulated p-values provide further insight into whether or not the noise adjusted tests are appropriate. Figure 1.7 shows histograms of the p-values associated with the noise adjusted test and empirical noise adjusted test. Once again the p-values associated with the adjusted and empirical tests come closer to having a uniform distribution than the p-values associated with the unadjusted test (which were shown in the previous subsection). Figure 1.8 shows the histogram of the noise adjusted test statistic values with the reference \( \chi^2 \) distribution (data had four levels of each experimental factor). Clearly the histogram comes much closer to the \( \chi^2 \) reference distribution that the histogram of the unadjusted test statistic (see Figure 1.5). Goodness of fit tests suggest that the \( \chi^2 \) distribution is appropriate for the noise adjusted test statistic (\( p = .25 \)).

Another simulation was conducted to evaluate the power of the noise adjusted tests to detect data from the ANOVA model. One thousand data sets were generated in the same fashion as the power simulation conducted in the previous section and \( r = 100 \) simulation replicates were used to conduct the noise adjusted likelihood ratio tests. The rejection rates for the test based on \( \tilde{p}_1 \) and \( \tilde{p}_2 \) are 81.6% and 81.8% respectively. So clearly the two methods exhibit power to reject the null hypothesis when the ANOVA model is necessary.

As we have seen in this subsection and the previous subsection, the unadjusted likelihood ratio test is not appropriate for comparing the orthogonal interactions model to the full ANOVA with interactions for replicated data. In order for the unadjusted
Figure 1.7: Histograms of Noise Adjusted p-values for: (a) Adjusted LRT (b) Empirical Test
test to perform well, \( c \) (the number of replicates) must be unrealistically large. We have presented two simulation based methods that provide four different corrected likelihood ratio tests. The four tests perform comparably for data with normal errors, but if the errors are (suspected to be) non-normal, the bootstrap based corrections should perform better. The computation required for the bootstrap adjusted tests is the same as the computation required for the noise adjusted tests, so the investigator is free to choose any of the four tests. All four methods offer reasonable power to reject the null hypothesis when the orthogonality assumptions do not hold. The tests based
on the empirical p-values do not require the degrees of freedom to be estimated for each source of variation, making these tests more attractive (in some situations because estimating degrees of freedom for the OI model may require extra work).

1.8.3 Testing the OI Model for Unreplicated Data

In the previous subsections, we considered testing the orthogonal interactions model against the full ANOVA model with interactions for replicated data. For unreplicated data, the analogous test would be

\[ H_0 : \text{Main effects ANOVA sufficient vs. } H_1 : \text{OI model necessary.} \]  

\[(1.27)\]

Testing the hypothesis stated in (1.27) is similar to testing the hypothesis stated in (1.23), except the OI model is no longer the null model. A slight modification to the bootstrap adjusted and noise adjusted testing strategies can be used to test (1.27). We focus on the empirical testing methods because they do not require degree of freedom estimation for the OI model.

If \( l_0 \) denotes the log-likelihood of the main effects model and \( l_1 \) denotes the log-likelihood for the OI model, the unadjusted likelihood ratio test statistic \( W \) is calculated as in (1.24). In order to obtain the corrected tests, simulated data sets are generated according to

\[ \tilde{Y}_j = \hat{Y}_0 + e_j \]

where \( \hat{Y}_0 \) is the vector of predicted values from the main effects model and \( e_j \) is a vector of errors. The error vector \( e_j \) can be generated using bootstrap samples from residuals.
Chapter 1. The Orthogonal Interactions Model

from the main effects model (similar to Section 1.8.1). Vectors of normal variates can also be used for the \( e_j \) (similar to Section 1.8.2), using the error mean square from the ANOVA model for \( \hat{\sigma}^2 \) in (1.26).

A simulation was used to evaluate the performance of the empirical methods for testing the hypothesis stated in (1.27). Data were generated from the \( 5 \times 5 \) OI model with standard normal errors:

\[
E(Y_{ij}) = \mu + \alpha_i + \beta_j + \kappa \gamma_{ij} \quad i = 1, \ldots, 5; \quad j = 1, \ldots, 5
\]

where \( \alpha_i \) and \( \beta_j \) were randomly generated such that \( \sum_i \alpha_i^2 = \sum_j \beta_j^2 = 2 \). The \( \gamma_{ij} \) were randomly generated so that \( \sum_i \sum_j \gamma_{ij}^2 = 1 \), \( \alpha^T \gamma = 0 \), and \( \gamma \beta = 0 \), using \( \kappa \) to control the magnitude of the interactions. When \( \kappa = 0 \), the data generated come from a main effects model and when \( \kappa > 0 \) the data come from the orthogonal interactions model. One thousand data sets were generated in this fashion for several values of \( \kappa \). For each data set, the hypothesis stated in (1.27) was tested at the .05 level using \( r = 100 \) simulation replicates. For \( \kappa = 0 \), the rejection rates were 6% and 6.5% for the bootstrap method and added noise method respectively. So the Type I error rates are close to the nominal testing level. Figure 1.9 shows the rejection rates for each of the two testing procedures for varying values of \( \kappa \). Clearly the two testing methods offer similar power to reject the null hypothesis. Similar simulations were performed for different levels of factor A and factor B to investigate the Type I error rate. Results were similar to the results presented in Table 1.3 and Table 1.4 where the Type I error rates were always close to the nominal testing level.

It may also be interesting to see how these tests perform when there are additive
errors in the model that do not satisfy the orthogonality constraints. A simulation was conducted to see how the corrected likelihood ratio tests of hypothesis (1.27) fare when the data feature interactions that are not orthogonal to the main effects. One thousand data sets were generated according to ANOVA model with additive interactions

\[ Y_{ij} = \mu + \alpha_i + \beta_j + \kappa_{ij} + \epsilon_{ij} \quad i = 1, \ldots, 5 \quad j = 1, \ldots, 5; \]
where $\epsilon_{ij}$ are independent standard normal variates. For each simulated data set, the parameters were generated randomly such that $\sum_i \alpha_i^2 = \sum_j \beta_j^2 = 2$ and $\sum_i \sum_j \gamma_{ij}^2 = 1$, but the orthogonality constraints are not enforced. Empirical tests were performed using $r = 100$ simulation replicates. The results of the simulation can be seen in Figure 1.10. The two testing methods offer similar power to reject the null hypothesis. The results of this simulation are interesting because we see that adjusted tests of (1.27)
Chapter 1. The Orthogonal Interactions Model

have some power to detect interactions, even when the parameters do not necessarily satisfy the orthogonality assumptions.

1.8.4 Determining Why the Orthogonality Assumptions are Violated

After fitting the full ANOVA model to replicated data, \( \mathbf{R} = \hat{\gamma} \hat{\beta} \) is an \( a \times 1 \) vector and \( \mathbf{S} = \hat{\alpha}^T \hat{\gamma} \) is a \( 1 \times b \) vector. Testing (1.23) is equivalent to testing \( H_0 : E(\mathbf{R}) = \mathbf{0}, E(\mathbf{S}) = \mathbf{0} \). In this section we focus on the mean and variance of the elements of \( \mathbf{R} \) and \( \mathbf{S} \) under the OI model in order to determine which elements of \( \mathbf{R} \) and \( \mathbf{S} \) may cause the test of (1.23) to reject.

Let \( R_i \) denote the \( i^{th} \) element of the vector \( \mathbf{R} \). Then \( R_i = \hat{\gamma}_i \hat{\beta} \) where \( \hat{\gamma}_i \) represents the \( i^{th} \) row of \( \hat{\gamma} \). Under the null hypothesis of orthogonal interactions, \( E(R_i) = 0 \) for all \( i \). The variance of \( R_i \) is

\[
\text{var}(R_i) = E \left( \hat{\gamma}_i \hat{\beta} \hat{\beta}^T \hat{\gamma}_i^T \right) - \underbrace{E(\hat{\gamma}_i \hat{\beta})}_0 \underbrace{E(\hat{\gamma}_i \hat{\beta})^T}_0
\]

\[
= E \left\{ E \left( \hat{\gamma}_i \hat{\beta} \hat{\beta}^T \hat{\gamma}_i^T \mid \hat{\gamma}_i \right) \right\}
\]

\[
= E \left\{ \hat{\gamma}_i E(\hat{\beta} \hat{\beta}^T) \hat{\gamma}_i^T \right\}
\]

\[
= E \left( \hat{\gamma}_i \hat{\beta} \hat{\beta}^T \hat{\gamma}_i^T \right) + E \left\{ \hat{\gamma}_i \text{cov}(\hat{\beta}) \hat{\gamma}_i^T \right\}
\]

\[
= E \left\{ \text{tr} \left( \hat{\gamma}_i \hat{\beta} \hat{\beta}^T \hat{\gamma}_i^T \right) \right\} + E \left[ \text{tr} \left\{ \hat{\gamma}_i \text{cov}(\hat{\beta}) \hat{\gamma}_i^T \right\} \right]
\]

\[
= \text{tr} \left\{ E (\hat{\gamma}_i^T \hat{\gamma}_i) \hat{\beta} \hat{\beta}^T \right\} + \text{tr} \left\{ E (\hat{\gamma}_i^T \hat{\gamma}_i) \text{cov}(\hat{\beta}) \right\}
\]

\[
= \text{tr} \left\{ \text{cov} \left( \hat{\gamma}_i + \hat{\gamma}_i^T \hat{\gamma}_i \right) \right\} \left\{ \hat{\beta} \hat{\beta}^T + \text{cov}(\hat{\beta}) \right\}.
\]
Chapter 1. The Orthogonal Interactions Model

So under the null hypothesis of $\gamma \beta = 0$, $E(R_i) = 0$ with variance as defined above for all $i$.

With $S_i$ denoting the $i^{th}$ element of $S$, we find the variance of $S_i$ under the null hypothesis $\alpha^T \gamma = 0$. Here $S_j = \hat{\alpha}^T \hat{\gamma}_j$, where $\hat{\gamma}_j$ is the $j^{th}$ column of $\hat{\gamma}$. Then the variance of $S_j$ is written

$$
\text{var}(S_i) = E(\hat{\alpha}^T \hat{\gamma}_j \hat{\gamma}_j^T \hat{\alpha}) - E(\hat{\alpha}^T \hat{\gamma}_j) E(\hat{\gamma}_j^T \hat{\alpha}) = E \{ E(\hat{\alpha}^T \hat{\gamma}_j \hat{\gamma}_j^T \hat{\alpha} \mid \hat{\alpha}) \} = E \{ \hat{\alpha}^T E(\hat{\gamma}_j \hat{\gamma}_j^T) \hat{\alpha} \} = E \left[ \text{trace} \left\{ \hat{\alpha}^T \text{cov}(\hat{\gamma}_j) \hat{\alpha} + \hat{\alpha}^T \gamma_j \gamma_j^T \right\} \right] = \text{tr} \left( E \left[ \hat{\alpha} \hat{\alpha}^T \left\{ \text{cov}(\hat{\gamma}_j) + \gamma_j \gamma_j^T \right\} \right] \right) = \text{tr} \left( E(\hat{\alpha} \hat{\alpha}^T) \left\{ \text{cov}(\hat{\gamma}_j) + \gamma_j \gamma_j^T \right\} \right) = \text{tr} \left[ \left\{ \text{cov}(\hat{\alpha}) + \alpha \alpha^T \right\} \left\{ \text{cov}(\hat{\gamma}_j) + \gamma_j \gamma_j^T \right\} \right].
$$

So under the null hypothesis of $\alpha^T \gamma = 0$, $E(S_j) = 0$ with variance as defined above.

In this subsection we have derived expectations and variances of the elements $R = \hat{\gamma} \hat{\beta}$ and $S = \hat{\alpha}^T \hat{\gamma}$ under the null hypothesis of orthogonal interactions for replicated data, where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are all obtained from the ANOVA model with interactions. If one of the likelihood ratio tests rejects the null hypothesis of orthogonal interactions, looking at the elements of $R$ and $S$ with the associated variance estimates can assist in determining which element caused the null hypothesis to reject. If $S_i$ is small compared to $\text{var}(S_i)$, then we would conclude that the $i^{th}$ element of the $S$ contributes to the violation of the orthogonality assumptions and likewise for the elements of $R$. 
Chapter 1.  The Orthogonal Interactions Model

1.9 Example Fits

In this subsection we fit the orthogonal interactions model to two real data sets to see if the orthogonal interactions model is appropriate for real data. We consider replicated data (metal shear strength data) as well as unreplicated data (soil data).

1.9.1 Metal Shear Strength Data

We consider the joint shear strength data investigated by Tomlinson and Cooper [17]. An experiment was conducted to determine how antimony affects the shear strength of tin-lead solder joints. Tin-lead solder specimens were made using one of four possible cooling methods and one of four levels of antimony (0%, 3%, 5%, and 10%) was added to the composition. Each weight/cooling method combination was replicated three times. The shear strength was recorded for each specimen. This is a $4 \times 4 \times 3$ experiment (four levels of each experimental factor and three replicates of each factor level combination).

First we fit the full ANOVA model with interactions to the data. The orthogonal interactions model is also fit to the data, using the rank method as well as the resampling method to assign degrees of freedom to the model. When using the resampling method with the OI model, we chose (see Section 2.2) the vector of added error variances to be

$$\lambda = \left[ 0, \ 1.31, \ 2.62, \ 3.92 \right]^T$$

with five hundred simulation replicates to estimate the degrees of freedom. In Chapter 2 we provide more details for using the resampling method, as well as how $\lambda$ and the number of simulation replicates were chosen. The fit of the ANOVA and OI models are summarized in Table 1.5 and Table 1.6 respectively. The two models provide similar
Table 1.5: *Full ANOVA Fit of Metal Strength Data.*

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>3</td>
<td>104.2</td>
<td>34.7</td>
<td>20.1</td>
<td>1.6E-7</td>
</tr>
<tr>
<td>Method</td>
<td>3</td>
<td>28.6</td>
<td>9.5</td>
<td>5.5</td>
<td>3.6E-3</td>
</tr>
<tr>
<td>W × M</td>
<td>9</td>
<td>25.1</td>
<td>2.8</td>
<td>1.6</td>
<td>.15</td>
</tr>
<tr>
<td>Error</td>
<td>32</td>
<td>55.2</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.6: *Orthogonal Interactions Fit of Metal Strength Data.*

<table>
<thead>
<tr>
<th>Source</th>
<th>Resampling DF</th>
<th>Rank DF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF</td>
<td>SS</td>
</tr>
<tr>
<td>Weight</td>
<td>3.1</td>
<td>103.9</td>
</tr>
<tr>
<td>Method</td>
<td>2.7</td>
<td>28.4</td>
</tr>
<tr>
<td>W × M</td>
<td>5.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Error</td>
<td>35.4</td>
<td>62.8</td>
</tr>
</tbody>
</table>

Conclusions about the data. Both models suggest that there are strong weight and cooling method effects, while the interaction effect is questionable \( (p \approx .15 \text{ for both models}).\)

Although the orthogonal interactions model is not a linear model, in Section 1.8 we saw that it is nested within the ANOVA model with interactions. Thinking of the ANOVA model as the full model and the OI model as the reduced model, a full versus reduced \( F \) test is a possible choice for comparing the ANOVA model to the OI model. Although the orthogonal interactions model is nonlinear, our simulation evidence suggests that the full-versus-reduced tests will provide reasonable results. The
approximate $F$ statistic for comparing the two models (using the resampling degrees of freedom) is
\[ F = \frac{62.8 - 55.2}{35.4 - 32} \frac{55.2}{32} = 1.3 \approx F_{3.4,32}. \]

Clearly this statistic would not reject at a reasonable testing level, so we would conclude that the OI model is sufficient for these data. Using the rank degrees of freedom instead of the resampling degrees of freedom leads to the full versus reduced test statistic $F = .88$ which would not be significant at any testing level. Although we believe that the full versus reduced $F$ test is approximate, statistics formed using both methods of estimating degrees of freedom suggest that the orthogonal interactions model is sufficient for these data.

We can also use the likelihood ratio test from Section 1.8 to test the appropriateness of the OI model. The unadjusted likelihood ratio test statistic of hypothesis (1.23) is $W = 6.17$ with $p = .29$. The simulation based methods from Section 1.8 with $r = 1000$ simulation replicates were also used to test the hypothesis. Using the bootstrap or noise adjusted likelihood ratio test statistics yields p-values of .50 and .48 respectively. The bootstrap and noise adjusted empirical p-values are .48 and .47 respectively. Clearly the (unadjusted and adjusted) likelihood based tests suggest that the OI model is sufficient for modeling the data.

The orthogonal interactions model and the ANOVA model can also be compared by looking at the predicted values from the two models. Figure 1.11 shows prediction profiles for the orthogonal interactions model as well as the ANOVA model. The profiles in (a) and (b) are similar, suggesting that predictions from the orthogonal interactions model are nearly identical to the predictions from the ANOVA model. Figure 1.12
Chapter 1. The Orthogonal Interactions Model

Figure 1.11: Metal Strength Prediction Profiles- (a) OI Model (b) ANOVA

is a scatterplot of the predictions from the OI model versus the predictions from the ANOVA model with interactions. The reference line in Figure 1.12 has slope equal to one and clearly the points lay close to the reference line. A biplot of the data (see Figure 1.13) does not suggest that the interactions are of the Tukey or linear-bilinear form, leaving the full ANOVA model or OI model as the appropriate methods for modeling the data. Because there are only four levels of each experimental factor, the biplot may not contain enough points to provide information about which model is most appropriate for the data.

This data set is interesting because it is an example of real replicated data where the orthogonal interactions model is appropriate. Full-versus-reduced $F$ tests as well as likelihood ratio tests suggest that the OI model is sufficient. We also saw that the two methods for estimating degrees of freedom for the orthogonal interactions model
1.9.2 Soil Data

Figueruelo et al. [18] outline an experiment involving seven soil samples from the Canary Islands. Each sample was divided into five subsamples and a different level of phosphate was added to each of the subsamples and exchangeable calcium was measured on each subsample. The investigators were interested in the main effects of
sample and phosphate as well as the interaction. These data are an example of a $7 \times 5$ experiment where factor level combination replication is not possible due to limited resources.

Fitting the main effects ANOVA model to the data yields an SSE of 3.09 with 24 degrees of freedom for error. The orthogonal interactions model as well as the AMMI(2) and Tukey one-degree-of-freedom models were fit to the data. When using the resampling method to estimate degrees of freedom for the OI and AMMI(2) models, we used

$$\lambda = \begin{bmatrix} 0, & .086, & .172, & .258 \end{bmatrix}^T$$

Figure 1.13: Biplot of Metal Strength Data
Chapter 1.  The Orthogonal Interactions Model

Table 1.7: Orthogonal Interactions Fit of Soil Data (Resampling DF)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>4.7</td>
<td>84.9</td>
<td>18.3</td>
<td>256.5</td>
<td>3.2E-8</td>
</tr>
<tr>
<td>Phosphate</td>
<td>3.2</td>
<td>1.4</td>
<td>.45</td>
<td>6.3</td>
<td>.02</td>
</tr>
<tr>
<td>S × P</td>
<td>18.6</td>
<td>2.6</td>
<td>.1</td>
<td>1.9</td>
<td>.18</td>
</tr>
<tr>
<td>Error</td>
<td>7.6</td>
<td>.54</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.8: AMMI(2) Fit of Soil Data (Resampling DF)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>6</td>
<td>84.9</td>
<td>14.2</td>
<td>406.6</td>
<td>2.3E-6</td>
</tr>
<tr>
<td>Phosphate</td>
<td>4</td>
<td>1.4</td>
<td>.36</td>
<td>10.2</td>
<td>.01</td>
</tr>
<tr>
<td>PCA1</td>
<td>11.7</td>
<td>2.4</td>
<td>.20</td>
<td>5.8</td>
<td>.04</td>
</tr>
<tr>
<td>PCA2</td>
<td>7.5</td>
<td>.57</td>
<td>.08</td>
<td>2.2</td>
<td>.21</td>
</tr>
<tr>
<td>Error</td>
<td>4.8</td>
<td>.17</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and five hundred simulation replicates. We chose to fit the AMMI(2) model to the data because the data were only sufficient for fitting the AMMI(1) and AMMI(2) models. The orthogonal interactions fit is summarized in Table 1.7, the AMMI(2) fit is summarized in Table 1.8, and the Tukey 1-DF fit is summarized in Table 1.9. The parameter $k$ in Table 1.9 is the non-additivity parameter described in Section 1.3. Table 1.7 and Table 1.8 both use the resampling method for estimating degrees of freedom. All of the models suggest a strong sample effect ($p < .0001$ for each model) and a significant phosphate effect, rejecting the null hypothesis of no phosphate effect.
Chapter 1. The Orthogonal Interactions Model

Table 1.9: Tukey 1-DF Fit of Soil Data.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>6</td>
<td>84.9</td>
<td>14.2</td>
<td>115.9</td>
<td>6E-16</td>
</tr>
<tr>
<td>Phosphate</td>
<td>4</td>
<td>1.4</td>
<td>.36</td>
<td>2.9</td>
<td>.04</td>
</tr>
<tr>
<td>k</td>
<td>1</td>
<td>.28</td>
<td>.28</td>
<td>2.3</td>
<td>.14</td>
</tr>
<tr>
<td>Error</td>
<td>23</td>
<td>2.8</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

at the $\alpha = .05$ level. Neither the orthogonal interactions model nor Tukey’s model suggest that there is a significant interaction between sample and phosphate. But the AMMI(2) model suggests a significant interaction because the first principal component axis is significant at the $\alpha = .05$ level. So the different models yield slightly different conclusions about the data. A biplot of the data is available in Figure 1.14. The biplot suggests that Tukey’s model is not appropriate for the data, but the AMMI and OI models can not be ruled out by looking at Figure 1.14.

Empirical p-values for the test of hypothesis (1.27) were obtained using $r = 1000$ simulation replicates. The p-values were .11 and .08 for the bootstrap method and added noise method respectively. We used the empirical methods for testing (1.27) because they do not require us to estimate the degrees of freedom. Neither of the empirical tests would reject at the .05 level, but they do offer fairly strong evidence that the main effects ANOVA model is not adequate for these data and hence the orthogonal interactions model is more appropriate. The difference between the OI and ANOVA models can also be seen in the prediction profiles from each model in Figure 1.15. The profiles in Figure 1.15(c) are not parallel, suggesting that there may
be an interaction between sample and phosphate level. The OI and ANOVA models clearly provide different predictions because of the difference in the two profiles.

As noted earlier, we could have used different methods of assigning degrees of freedom to the sources of variation in the OI model and the AMMI(2) model. These alternate degree of freedom estimates are summarized in Table 1.10. The column denoted ‘PPN’ in Table 1.10 represents the estimated degrees of freedom using the pattern plus noise method. For the simulation based methods, we used five hundred simulation replicates to estimate the degrees of freedom. Using the different degrees of freedom estimates in Table 1.10 can lead to different conclusions about the data. For example, if the rank degrees of freedom were used with the orthogonal interactions

Figure 1.14: Biplot of Soil Data
Chapter 1. The Orthogonal Interactions Model

Figure 1.15: Soil Data Profiles- (a) OI Predictions (b) ANOVA Predictions (c) Raw Data
Chapter 1. The Orthogonal Interactions Model

Table 1.10: Several Different Degree of Freedom Estimates For Orthogonal Interactions Model and AMMI(2) Analysis of Soil Data.

<table>
<thead>
<tr>
<th>Source</th>
<th>OI Model</th>
<th>AMMI(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank</td>
<td>Resampling Gollub Pure Noise PPN Resampling</td>
</tr>
<tr>
<td>Sample</td>
<td>6</td>
<td>4.7</td>
</tr>
<tr>
<td>Phosphate</td>
<td>4</td>
<td>3.2</td>
</tr>
<tr>
<td>Interaction</td>
<td>15</td>
<td>18.6</td>
</tr>
<tr>
<td>PCA 1</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.7</td>
</tr>
<tr>
<td>PCA 2</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.5</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.8</td>
</tr>
</tbody>
</table>

One way to compare the fit of the orthogonal interactions model to the fit of the AMMI(2) model is to plot the predicted values from each model. Figure 1.16 is a scatterplot of predicted values from the AMMI(2) model versus the predicted values from the OI model, with a reference line with slope equal to one through the origin.
Fitting a simple linear regression to the points in Figure 1.16 suggests a strong linear relationship \( R^2 = .93 \) where the slope is not significantly different from one. The strong linear relationship between the predictions from the two models suggests that as far as prediction is concerned, the OI model and AMMI(2) model are very similar. Although the two models provide similar predictions, the orthogonal interactions model may be more attractive because it does not require the user to choose the number of PCA axes included in the model.
1.10 Simulation Results

Error variance estimation and testing power are two important features of any statistical model. In this section, we use simulation methods to investigate the performance of the error mean square as an error variance estimator for the orthogonal interactions model. We also investigate performance of $F$-statistics for tests of main effects and interactions in the OI model.

1.10.1 Variance Estimation

For linear models, the error mean square is used to estimate the error variance. Since the orthogonal interactions model borrows heavily from linear models theory, we use the error mean square to estimate error variance with the OI model. Data were generated from the $4 \times 4$ orthogonal interactions model (1.5) with error variance $\sigma^2 = 1$ and

$$\sum_{i=1}^{4} \alpha_i^2 = 2, \sum_{j=1}^{4} \beta_j^2 = 2, \text{ and } \sum_{i=1}^{4} \sum_{j=1}^{4} \gamma_{ij}^2 = 4.$$  

Parameters were generated in this fashion to ensure that the data feature main effects and interactions. We used the rank method and the resampling method to estimate degrees of freedom in each model in order to compare the performance of the two methods. Five hundred data sets were generated in this fashion and error variance estimates are summarized in Table 1.11. We use Stein’s loss function [19] to evaluate the performance of the error variance estimators. Stein’s loss is defined

$$SL(\hat{\sigma}^2) = E \left\{ \frac{\hat{\sigma}^2}{\sigma^2} - 1 - \log \left( \frac{\hat{\sigma}^2}{\sigma^2} \right) \right\}, \quad (1.28)$$

74
and is popular for evaluating variance estimators because it penalizes underestimation and overestimation equally. The column headed by ‘10% TM(\(\hat{\sigma}^2\))’ represents the 10% trimmed mean of the error variance estimates. Neither method of estimating the error variance yields unbiased results, in fact neither method provides good estimates of the error variance. The relative bias of the estimates using rank degrees of freedom is lower than that of the estimates obtained using the resampling degrees of freedom. Stein’s loss function also favors using the rank degrees of freedom when estimating the error variance.

Another simulation was conducted where the 4 \(\times\) 4 orthogonal interactions model was fit to standard normal noise. Five hundred data sets of this form were created and the error variance was estimated using both methods for obtaining the error degrees of freedom. Table 1.12 summarizes the results of this simulation. Error variance estimates obtained using the resampling degrees of freedom are nearly unbiased. Here the error variance estimates obtained using the resampling degrees of freedom have a lower relative bias than the estimates obtained using the rank degrees of freedom. Stein’s loss function also favors the estimates obtained using the resampling degrees of freedom. We see that if there is no signal in the data, the rank method does not provide reliable degrees of freedom estimates. Because the mean squared error using
Chapter 1. The Orthogonal Interactions Model

Table 1.12: Simulated Error Variance Estimates for Data Generated without Main Effects or Interactions ($\sigma^2 = 1$)

<table>
<thead>
<tr>
<th>DF Method</th>
<th>Median($\hat{\sigma}^2$)</th>
<th>Mean($\hat{\sigma}^2$)</th>
<th>10% TM($\hat{\sigma}^2$)</th>
<th>Std. Dev.($\hat{\sigma}^2$)</th>
<th>SL($\hat{\sigma}^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>.39</td>
<td>.46</td>
<td>.42</td>
<td>.29</td>
<td>.45</td>
</tr>
<tr>
<td>Resampling</td>
<td>.78</td>
<td>1.09</td>
<td>.91</td>
<td>1.2</td>
<td>.34</td>
</tr>
</tbody>
</table>

rank degrees of freedom is underestimating the error variance, the rank method must be overestimating the degrees of freedom for error.

The rank method performs well when there is signal in the data, but it performs poorly in the absence of signal. This suggests that for data from the orthogonal interactions model, the degrees of freedom for each source of variation depend on the data. Because the resampling method makes use of the data, we suggest using the resampling method when unsure if there is a signal. The idea that the degrees of freedom for each source depend on the data makes sense intuitively. The orthogonal interactions model provides degrees of freedom for interaction by adding nonlinear constraints to the usual ANOVA model. If the true main effect vectors ($\alpha$ and $\beta$) are zero, then $\alpha^T \gamma = 0$ and $\gamma \beta = 0$ for any matrix $\gamma$, making the orthogonality assumptions unrestrictive. This lack of restriction is reflected in the degrees of freedom. Conversely if $\alpha$ and $\beta$ are both nonzero vectors, then the orthogonality assumptions are restrictive.

The variance estimation simulations in this subsection suggest that the error mean square is not a satisfactory estimator of the error variance for the orthogonal interactions model. We also found evidence that the degrees of freedom in the model depend on the data. Because the rank degrees of freedom perform so poorly in the absence of
signal, we suggest using the resampling degrees of freedom when estimating error variance with the orthogonal interactions model. However if the investigator knows that the data exhibits main effects from prior knowledge, then the rank method degrees of freedom provide satisfactory error variance estimates. For some applications such as crop yield trials, investigators know that strong main effects exist, so using the rank method to obtain degrees of freedom would be sufficient. Although the error mean square does not appear to be a good estimator of $\sigma^2$, we will see in the next subsection that $F$-statistics to test for main effects and interactions still exhibit favorable qualities.

1.10.2 Testing Simulations

We conducted several simulation studies to look at using $F$-statistics to test for main effects and interactions in the orthogonal interactions model. Data were generated according to the $4 \times 4$ OI model

$$Y_{ij} = \mu + \kappa_1(\alpha_i + \beta_j) + \kappa_2\gamma_{ij} + \epsilon_{ij}, \quad (1.29)$$

where $\epsilon_{ij}$ are standard normal variates and the main effects and interactions are generated such that

$$\sum_i \alpha_i^2 = \sum_j \beta_j^2 = \sum_i \sum_j \gamma_{ij}^2 = 1.$$  

The magnitude of the main effects and interactions in (1.29) are controlled by varying $\kappa_1$ and $\kappa_2$ respectively.
Chapter 1. The Orthogonal Interactions Model

The first simulation we conducted looked at testing the hypothesis

$$H_0: \alpha_i = 0 \text{ for all } i \text{ vs. } H_1: \text{Not all } \alpha_i \text{ are zero,}$$

using the test statistic

$$F = \frac{\text{MS}(A)}{\text{MSE}}.$$ 

Again we are using these $F$-statistics despite the fact that the OI model is nonlinear. Data were generated according to (1.29) with $\kappa_2 = 4$ for varying levels of $\kappa_1$. Five hundred data sets were generated at each level of $\kappa_1$ and we conducted the $F$-tests at the .05 level using the rank method and resampling method for estimating degrees of freedom. The resampling method used 250 replicates to estimate degrees of freedom. The rejection rates can be seen in Figure 1.17. The Type I error rate for tests formed using rank degrees of freedom and tests formed using resampling degrees of freedom are 3.8% and 1.0% respectively. So when interactions are present, tests formed using both degrees of freedom methods are conservative, but tests formed using the rank degrees of freedom come closer to the nominal level.

A similar simulation was conducted where we looked at testing the hypothesis

$$H_0: \gamma_{ij} = 0 \text{ for all } i, j \text{ vs. } H_1: \text{Not all } \gamma_{ij} \text{ are zero,}$$

using the test statistic

$$F = \frac{\text{MS}(AB)}{\text{MSE}}.$$ 

Data were generated according to (1.29) with $\kappa_1 = 2$ for varying levels of $\kappa_2$. Five hundred data sets were generated at each level of $\kappa_2$ and tests were performed at the
Chapter 1. The Orthogonal Interactions Model

Figure 1.17: Power of Main Effect F-tests Formed Using Both Degree of Freedom Methods (Monte Carlo SE: 9.7E-3)

.05 level using both methods of estimating degrees of freedom. The rejection rates are available in Figure 1.18. The Type I error rates for the tests formed using rank degrees of freedom and tests formed using resampling degrees of freedom are 11.0% and 3.8% respectively. Although the tests formed using the resampling degrees of freedom are slightly conservative, they are more appealing because the tests formed using the rank degrees of freedom exhibit a high Type I error rate.

We also performed a simulation where five hundred data sets were generated according to (1.29) with $\kappa_1 = \kappa_2 = 0$ (standard normal noise). The Type I error rates for tests formed using the rank degrees of freedom are 12.8% and 35% for main effects and interactions respectively. For tests formed using the resampling degrees of freedom, the Type I error rates are 2.2% and 3.2% for main effects and interactions respectively.
Figure 1.18: *Power of Interaction F-tests Formed Using Both Degree of Freedom Methods (Monte Carlo SE: 9.7E-3)*

Clearly the tests formed using the resampling degrees of freedom are slightly conservative, but the rank degrees of freedom are not acceptable in this situation because of the high Type I error rate.

The simulations conducted in this subsection suggest that the orthogonal interactions model has power to detect main effects and interactions when using the same $F$-tests that are used for linear models. Tests formed using the rank degrees of freedom exhibit high Type I error rates when testing for interactions and when fitting the OI model to random noise. Tests formed using the resampling degrees of freedom are slightly conservative for both tests of main effects and tests of interaction. Conservative tests are typically more attractive than tests with high Type I error rates, so we suggest using the resampling degrees of freedom to test for main effects and interactions in the
Chapter 1. The Orthogonal Interactions Model

orthogonal interactions model.

1.11 Conclusions

For data from two-factor experiments without replication, restrictions must be placed on the interactions in order to allow for the estimation of the main effects, interactions, and error variance. The most severe set of restrictions would be to require that all of the interactions are zero. We presented a set of nonlinear constraints to add to the standard two-factor analysis of variance model that do not require the interactions to be of the multiplicative form of Tukey’s model or the AMMI model. We call the resulting model the orthogonal interactions model. The nonlinear constraints free degrees of freedom for error by reducing the degrees of freedom attributed to other sources of variation.

We found that the orthogonal interactions model has several favorable qualities. Although the error mean square is not a good estimator of the error variance, approximate $F$-tests exhibit power to detect main effects and interactions. We also found simulation-based corrections to the likelihood ratio test that allow us to test the appropriateness of the orthogonality assumptions for replicated data. The likelihood ratio test comparing the OI model to the main effects ANOVA model can also be corrected using simulation based methods. In Section 1.9 we found real data for which the OI model is appropriate. Although we do not believe that graphical methods such as the biplot are useful in determining whether or not the OI model is appropriate for a given data set, if the biplot can rule out Tukey’s model or the AMMI model, then the OI model would be a logical choice for modeling the data.
Chapter 1. The Orthogonal Interactions Model

The orthogonal interactions model is a useful alternative to Tukey’s one-degree-of-freedom model and the AMMI model. The OI model does not require interactions to have a very restrictive multiplicative form like Tukey’s model, and it does not require the investigator to choose the number of axes used to model the interactions like the AMMI model. The OI model can be fit efficiently using the profile likelihood method described in Section 1.4.2. Degrees of freedom can be estimated using two different methods, but we suggest using the resampling method discussed in Chapter 2.

In Section 1.8 we found that an empirical corrected adjusted likelihood ratio test was appropriate for comparing the orthogonal interactions model to the ANOVA model. Because the empirical test is sufficient, we must determine whether or not estimating degrees of freedom for the orthogonal interactions model is an important issue. Although we did not discuss using information criteria (such as Akaike’s information criteria) in conjunction with the OI model, information criteria may be useful in assessing the fit of the OI model. Degrees of freedom are necessary in using information criteria, so we believe that estimating degrees of freedom for the OI model is important even though it is unnecessary for the empirical tests.
Chapter 2

A Resampling Based Method For

Estimating Degrees of Freedom

Consider the usual two-factor ANOVA model

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, c; \]  \hspace{1cm} (2.1)

subject to the constraints

\[ \sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0, \]

\[ \sum_{i=1}^{a} \gamma_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j=1}^{b} \gamma_{ij} = 0 \text{ for all } i. \]

Additionally we assume that \( \epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2) \). For this simple linear model, we have theory for assigning degrees of freedom to each source of variation. If a set of non-linear constraints, such as the orthogonality constraints in model (1.5), are added to the ANOVA model, assigning degrees of freedom to each source of variation is no longer a straightforward task. In this chapter we suggest a method for estimating degrees of freedom that is based on adding noise of varying magnitudes to the response variable and observing the behavior of the average sums of squares. The resampling method is motivated by the linear model, but we will see that it performs well with more complex
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

nonlinear models such as the orthogonal interactions model and the AMMI model.

2.1 Motivation

In this section we look at how the linear fixed effects model as well as the linear mixed effects model provide the motivation for the resampling method of estimating degrees of freedom. In these two models, there is a linear relationship between expected sum of squares and the variance of noise added to the data. While the relationships hold exactly for these simple models where we know the correct degrees of freedom, we believe that the relationship will also be approximately true for more complicated models where degrees of freedom calculations are difficult.

Fixed Effects ANOVA Motivation

The fixed effects ANOVA model provides the motivation for the resampling method of estimating degrees of freedom. Consider the usual linear model

\[ Y = X\beta + \epsilon, \]  

(2.2)

where \( Y \) is an \( n \times 1 \) vector, \( X \) is an \( n \times p \) design matrix of rank \( p \), and \( \beta \) is a \( p \times 1 \) vector. Furthermore we assume that \( \epsilon \) is an \( n \times 1 \) vector of errors that is distributed multivariate normal with mean zero and covariance \( \sigma^2 I \).

The sums of squares in a standard analysis of variance decomposition all have the form \( Y^T P Y \) where \( P \) is a projection matrix. We know from linear models theory that the degrees of freedom associated with \( Y^T P Y \) is \( \text{rank}(P) = \text{tr}(P) \). The expectation of
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

the sum of squares associated with $P$ is

$$E(Y^T P Y) = \theta^2 + \sigma^2 \text{rank}(P),$$

where $\theta^2 = E(Y^T P E(Y))$ is a non-centrality parameter. Now let $U = Y + \sqrt{\lambda}Z$, where $Z$ is an $n \times 1$ vector of standard normal variates. Then model (2.2) can be fit to the augmented data: $E(U) = X\beta$. We will show that the degrees of freedom can be obtained by differentiating $E(U^T P U \mid Y)$ with respect to $\lambda$. The sum of squares associated with projection matrix $P$ for the augmented response is written $U^T P U = (Y + \sqrt{\lambda}Z)^T P (Y + \sqrt{\lambda}Z)$ with conditional expectation

$$E(U^T P U \mid Y) = E \left\{ (Y + \sqrt{\lambda}Z)^T P (Y + \sqrt{\lambda}Z) \mid Y \right\}$$

$$= Y^T P Y + 2\sqrt{\lambda} \left\{ Y^T P E(Z) \right\} + \lambda E(Z^T P Z)$$

$$= Y^T P Y + \lambda \text{rank}(P).$$

Taking the derivative of $E(U^T P U \mid Y)$ with respect to $\lambda$,

$$\frac{\partial}{\partial \lambda} E(U^T P U \mid Y) = \text{rank}(P)$$

$$= df_P,$$

where $df_P$ represents the degrees of freedom associated with $P$. So when $N(0, \lambda)$ noise is added to the response vector, the expected sum of squares associated with projection matrix $P$ increases by $\lambda \text{rank}(P)$. For linear models, we know that there are $\text{rank}(P)$ degrees of freedom associated with $P$. So the expected sum of squares associated with
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

\( \mathbf{P} \) is linearly related to \( \lambda \) (variance of noise added to \( Y \)) with slope equal to the degrees of freedom associated with \( \mathbf{P} \). The degrees of freedom for \( \mathbf{P} \) can be estimated by the slope of the simple linear regression that fits average sum of squares as a function of added error variance.

We have shown that adding noise to the response variable and observing the rate of change of each sum of squares provides information about the degrees of freedom for each source of variation in the linear model. Later we use this relationship to estimate the degrees of freedom for each source of variation in the model. In the next subsection we see that the same linear relationship between expected sums of squares and added error variance is true for the linear mixed-effects model.

Balanced Mixed Models Motivation

In the previous subsection we found that for a fixed effects linear model, the expected sum of squares has a linear relationship with the variance of the added errors. We will show that the same relationship is true for the balanced two-factor mixed effects model. Consider data from the mixed effects model where factor \( A \) is a fixed effect and factor \( B \) is a random effect:

\[
Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, c;
\]  

(2.3)

where \( \beta_j \) are independent \( N(0, \sigma_B^2) \) variates, \( (\alpha\beta)_{ij} \) are independent \( N(0, \sigma_{AB}^2) \) variates, and \( \epsilon_{ijk} \) are \( N(0, \sigma^2) \) variates. Furthermore we assume that the \( \beta_j \), \( (\alpha\beta)_{ij} \), and \( \epsilon_{ijk} \) are all mutually independent. The expected sums of squares and mean squares for this model are summarized in Table 2.1.
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Table 2.1: Expected SS and Expected MS for Mixed Effect Model

<table>
<thead>
<tr>
<th>Source</th>
<th>E(SS)</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$bc \sum_{i=1}^{a} \alpha_i^2 + c(a - 1)\sigma_{AB}^2 + (a - 1)\sigma^2$</td>
<td>$\frac{bc}{a-1} \sum_{i=1}^{a} \alpha_i^2 + c\sigma_{AB}^2 + \sigma^2$</td>
</tr>
<tr>
<td>B</td>
<td>$ac(b - 1)\sigma_B^2 + c(b - 1)\sigma_{AB}^2 + (b - 1)\sigma^2$</td>
<td>$ac\sigma_B^2 + c\sigma_{AB}^2 + \sigma^2$</td>
</tr>
<tr>
<td>A×B</td>
<td>$c(a - 1)(b - 1)\sigma_{AB}^2 + (a - 1)(b - 1)\sigma^2$</td>
<td>$c\sigma_{AB}^2 + \sigma^2$</td>
</tr>
<tr>
<td>Error</td>
<td>$ab(c - 1)\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

Now let $U = Y + \sqrt{\lambda}Z$ where $\lambda$ is a known constant and $Z$ is an $n \times 1$ vector of independent standard normal variates. When model (2.3) is fit to $U$, the augmented response, we will show that the conditional expected sums of squares have a linear relationship with the sums of squares from fitting model (2.3) to $Y$. Using $SS_U(A)$ to represent the sum of squares for factor $A$ when $U$ is the response variable,

$$E\{SS_U(A) \mid Y\} = E\left\{bc \sum_{i} (U_{i.} - \bar{U}_{..})^2 \mid Y\right\}$$

$$= E\left[bc \sum_{i} \left\{\left(\bar{Y}_{i.} + \sqrt{\lambda}Z_{i.}\right) - \left(\bar{Y}_{..} + \sqrt{\lambda}Z_{..}\right)\right\}^2 \mid Y\right]$$

$$= f_1(Y) + (a - 1)\lambda,$$

where $f_1(Y)$ is a function of the response $Y$. Then

$$\frac{\partial}{\partial \lambda} E\{SS_U(A) \mid Y\} = a - 1.$$
A similar argument can be used to establish that

\[ E \{ SS_U(B) \mid Y \} = f_2(Y) + (b - 1)\lambda, \]

and hence

\[ \frac{\partial}{\partial \lambda} E \{ SS_U(B) \mid Y \} = b - 1. \]

The same relationship holds for the error sum of squares

\[
E \{ SS_U(E) \mid Y \} = E \left\{ \sum_{i,j,k} (U_{ijk} - \bar{U}_{ij})^2 \mid Y \right\} \\
= E \left\{ \sum_{i,j,k} \left\{ (Y_{ijk} + \sqrt{\lambda}Z_{ijk}) - (Y_{ij.} - \sqrt{\lambda}Z_{ij.}) \right\}^2 \mid Y \right\} \\
= f_3(Y) + ab(c - 1)\lambda,
\]

and hence

\[ \frac{\partial}{\partial \lambda} E \{ SS_U(E) \mid Y \} = ab(c - 1). \]

Since the relationship between conditional expectation and added error variance holds for the sums of squares for \( A, B, \) and error, we know that the relationship must also hold for the interaction sum of squares.

In the previous section we found that for fixed-effects linear models, if noise is added to the response variable, observing the rate of increase in each sum of squares will yield information about the degrees of freedom for each source of variation. So we know that for linear models in general, the conditional expected sum of squares is a linear function of the noise added to the response variable. In fact, we can take as the
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

definition of degree of freedom

\[ df = \frac{\partial}{\partial \lambda} \mathbb{E} \{ SS_U(j) \mid Y \} \]  

(2.4)

for the \( j^{th} \) source of variation in the model.

2.2 Resampling Method for Estimating Degrees of Freedom

We propose a new method for estimating degrees of freedom for each source of variation in complex statistical models. As stated in the previous section, the method involves adding noise of varying magnitudes to the response variable and observing the change in sums of squares when the model is fit to the augmented data. Using simulation methods, extra errors with different levels of variance \( \lambda \) are added to the original response and average sums of squares are computed for each level of \( \lambda \). For each source of variation, a simple linear regression is used to model average sum of squares as a function of \( \lambda \). The slope of this linear regression is the estimated degrees of freedom.

The remainder of this section provides details for using this method to estimate degrees of freedom for any model where sums of squares are partitioned.

The first step to using this method is to fit the model of interest to the response variable \( Y \) and partition the sum of squares, calling these sums of squares \( SS_1(Y) \), \( SS_2(Y) \), and so on. Next, an initial estimate of the error variance must be obtained. This information may come from prior knowledge or though fitting a reduced, or more familiar model, to the data. For example, when working with the orthogonal interac-
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

tions model as stated in (1.5), the mean squared error from the main effects ANOVA model can be used as an initial estimate of error variance. This initial error variance estimate, $\hat{\sigma}_0^2$, is used to determine the magnitude of the noise that will be added to the data.

After obtaining $\hat{\sigma}_0^2$, let $\lambda$ be an $l \times 1$ vector that represents the different variances of the errors to be added to the data. We let $\lambda_1 = 0$, $\lambda_l = 2\hat{\sigma}_0^2$, and the points $\lambda_2$ through $\lambda_{l-1}$ be equally spaced values between zero and $2\hat{\sigma}_0^2$. For example, if $\lambda$ is a $5 \times 1$ vector, then the elements of $\lambda$ are written

$$\lambda = \left[ 0, \frac{5}{2}\hat{\sigma}_0^2, \hat{\sigma}_0^2, 1.5\hat{\sigma}_0^2, 2\hat{\sigma}_0^2 \right]^T.$$  

The maximum magnitude of the added error variance is chosen to be $2\hat{\sigma}_0^2$ because adding this magnitude of noise to the data is essentially equivalent to reducing the amount of information in the data by a third.

After obtaining $\lambda$, $B$ vectors of $N(0, \lambda_2)$ variates are generated to be added to the response vector $Y$. The augmented response vector is denoted $U_b$, where $U_b = Y + Z_b$ and $Z_b$ is a vector of $N(0, \lambda_2)$ variates. For each $U_b$, the model of interest is fit to the augmented response and the sums of squares are partitioned. Once the model has been fit to all of the $U_b$, the average sum of squares are calculated for each source of variation. For augmented data with added error variance $\lambda_2$, the average sum of squares associated with the $j^{th}$ source of variation is written

$$\overline{SS}_{j(\lambda_2)} = B^{-1} \sum_{b=1}^{B} SS_{j,b}(\lambda_2),$$
where $SS_{j,b}(\lambda_2)$ represents the sum of squares associated with the $j^{th}$ source of variation in the $b^{th}$ simulated data set with added error variance $\lambda_2$. The process is then repeated using $\lambda_3$ up to $\lambda_l$ as the extra error variance. After the average sum of squares have been obtained for each level of added error variance, the results can be summarized in a manner such as Table 2.2.

With the help of Table 2.2, we use a simple linear regression to model average sum of squares as a function of $\lambda$ for each source of variation:

$$E\{SS_1(\lambda)\} = \tau_1 + \lambda \theta_1$$

$$E\{SS_2(\lambda)\} = \tau_2 + \lambda \theta_2$$

$$\vdots$$

$$E\{SS_E(\lambda)\} = \tau_E + \lambda \theta_E.$$
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

$\theta_E$ is the natural estimator for the degrees of freedom for error and $\theta_k$ is the natural estimator for the degrees of freedom for the $k^{th}$ source of variation in the model. Since the investigator typically knows that the estimated degrees of freedom should sum to a specific value $d$, constraining the regressions in (2.5) so that

$$\theta_1 + \theta_2 + \ldots + \theta_E = d$$

is not complicated. To obtain the constrained estimates, replace $\theta_1$ by $d - (\theta_2 + \ldots + \theta_E)$ and simultaneously fit the models in (2.5). Using the resampling method for estimating degrees of freedom can be seen graphically in Figure 2.1. The figure shows the average error sum of squares as a function of $\lambda$, the variance of the noise added to the data. The slope of the simple linear regression line in Figure 2.1 is the estimated error degrees of freedom.

When using the resampling method to estimate the degrees of freedom, the investigator must choose the number of simulation replicates, $B$, as well as $l$, the length of the $\lambda$ vector. Choosing $B$ and $l$ typically depends on the complexity of the model being considered. Using the resampling method involves fitting the model to $B(l - 1) + 1$ data sets. For a complex model that is computationally intense, choosing $B$ and $l$ to be large may result in tremendous computing times. For simple models or if computation time is not an issue, using large $B$ and $l$ will not be a problem. Through numerous simulation studies (see Section 2.3), we have found that $l = 4$ and $B = 100$ are sufficient to produce satisfactory results for linear models as well as Tukey’s model. For more complex models, increasing $B$ is necessary to obtain satisfactory estimates.

Occasionally, the resampling method will yield a negative estimate for one or more
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Figure 2.1: *Average Error Sum of Squares as a Function of Added Error Variance*
degree of freedom. This happens when by chance, the average sum of squares drops when a small amount of noise is added to the data. For obvious reasons, negative degrees of freedom are not acceptable. This problem can be solved by increasing $B$ or by increasing the magnitude of the errors added to the response (modifying the vector $\lambda$). For example, if using

$$\lambda_{(i)} = \left[ 0, \frac{2}{3}\hat{\sigma}^2_0, \frac{4}{3}\hat{\sigma}^2_0, 2\hat{\sigma}^2_0 \right]^T$$

yields a negative estimate for one or more sources of variation, then using

$$\lambda_{(ii)} = \left[ 0, \frac{2}{3}\hat{\sigma}^2_0, \frac{4}{3}\hat{\sigma}^2_0, 2\hat{\sigma}^2_0, \frac{8}{3}\hat{\sigma}^2_0, \frac{40}{3}\hat{\sigma}^2_0, 4\hat{\sigma}^2_0 \right]^T$$

may yield more appropriate estimates. If adding $N(0,.5\hat{\sigma}^2_0)$ noise to the data actually lowers one or more of the average sums of squares, adding noise up to the magnitude of $4\hat{\sigma}^2_0$ will dampen the effect of the initial drop in average sum of squares. Another option would be to use

$$\lambda_{(iii)} = \left[ 0, \frac{4}{3}\hat{\sigma}^2_0, \frac{10}{3}\hat{\sigma}^2_0, 4\hat{\sigma}^2_0 \right]^T$$

if computation time prohibits the use of $\lambda_{(ii)}$. Any modification to $\lambda$ that lessens the effect of the initial drop in average sum(s) of squares will solve the problem of the negative degree of freedom estimate(s). Other methods for estimating degrees of freedom (such as the pattern plus noise method described in Section 1.6.2) can produce negative estimates as well, but do not provide an obvious solution to the problem. So even though the resampling method can produce negative degree of freedom estimates,
the method does offer an intuitive solution to the problem.

2.3 Resampling Degrees of Freedom and the Familiar Models

In Section 2.1 we found that the linear relationship between sums of squares and added error variance is exact for all linear ANOVA models. Since we know the correct degrees of freedom for each source of variation in these two situations, we can use simulation techniques to evaluate the performance of the resampling method when analyzing these types of data. Additionally, the Tukey one-degree-of-freedom model is a slightly more complicated nonlinear model where we know the correct degrees of freedom for each source of variation. In this section, we use simulation to evaluate a particular combination of $B$ and $\lambda$ for linear models and to see how the resampling method performs with data from Tukey’s model.

Resampling Degrees of Freedom and the Two-Factor Fixed ANOVA

A simulation was conducted to evaluate the performance of the resampling degrees of freedom for the standard two-factor ANOVA with replication. We know (by theory in Section 2.1) that as $B \to \infty$, the resampling method gives us the true degrees of freedom. The simulations in this subsection allow us to evaluate the performance of the resampling method using a particular choice of $B$ and $\lambda$ for a simple linear model. Data were generated according to the fixed effect ANOVA model stated in (2.1). The
data consist of five levels of each experimental factor \((a = b = 5)\) and \(c = 2\) replicates per factor level combination. Here the \(\alpha, \beta, \) and \(\gamma\) parameters were generated using standard normal variates under the usual sum to zero assumptions and we chose \(\mu = 0\). The error terms were generated as standard normal variates as well. One thousand simulated data sets were created in this fashion. Although assigning degrees of freedom to each source of variation is trivial for this model, we use the resampling method to see how it performs.

First, the main effects ANOVA model is fit to each data set in order to obtain \(\hat{\sigma}_0^2\), the preliminary estimate of the error variance. The main effects ANOVA model would be the obvious choice for the reduced model if we did not know the correct degrees of freedom for interaction. For each data set, we use a \(4 \times 1\) \(\lambda\) vector. Equally spacing the elements of \(\lambda\) yields

\[
\lambda = \left[ 0, \frac{2}{3}\hat{\sigma}_0^2, \frac{4}{3}\hat{\sigma}_0^2, 2\hat{\sigma}_0^2 \right]^T
\]

as the set of extra error variances. At each level of \(\lambda\) (except zero), \(B = 100\) simulation replicates were used to calculate the average \(\text{SS}(A), \text{SS}(B), \text{SS}(AB),\) and SSE values. Then fitting the linear model

\[
E\{\text{SS}_A(\lambda_i)\} = \tau + \lambda_i\theta_A \quad i = 1, \ldots, 4;
\]

yields \(\hat{\theta}_A\) as the estimate of the degrees of freedom for experimental factor \(A\). Degrees of freedom for the other sources of variation are estimated in a similar manner. For
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Table 2.3: Summary of Estimated DF for 2-Factor Fixed Effect ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>True DF</th>
<th>Mean($\hat{df}$)</th>
<th>Std. Dev.($\hat{df}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>4</td>
<td>4.01</td>
<td>.82</td>
</tr>
<tr>
<td>$B$</td>
<td>4</td>
<td>3.98</td>
<td>.85</td>
</tr>
<tr>
<td>$A \times B$</td>
<td>16</td>
<td>15.99</td>
<td>.67</td>
</tr>
<tr>
<td>Error</td>
<td>25</td>
<td>25.01</td>
<td>.82</td>
</tr>
</tbody>
</table>

each simulated data set, constraining the linear regressions so that

$$\hat{\theta}_A + \hat{\theta}_B + \hat{\theta}_{AB} + \hat{\theta}_E = 49$$

guarantees that the estimated degrees of freedom sum to the correct value. The simulation results are summarized in Table 2.3. As the table shows, the average degree of freedom estimates are all close to the true values. Note that although the degrees of freedom are constrained to sum to $n - 1 = 49$ for each replicated data set, the averages presented in Table 2.3 do not sum to 49 merely because of rounding. So when using the resampling method for a simple linear model, this simulation suggests that $B = 100$ and $\lambda$ as defined in (2.6) are adequate to yield unbiased estimates.

We conducted a similar simulation study to investigate the power of tests conducted using the resampling degrees of freedom for crossed two-factor experiments with fixed effects. For each of the 1000 simulated data sets, $F$-tests of main effects and interactions were calculated using the resampling degrees of freedom as well as the true degrees of freedom.
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Data were generated from the ANOVA model

\[ Y_{ijk} = \mu + \kappa_1(\alpha_i + \beta_j) + \kappa_2 \gamma_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, 5; \ j = 1, \ldots, 5; \ k = 1, 2; \]  

(2.7)

where \( \alpha_i, \beta_j, \) and \( \gamma_{ij} \) are all randomly generated and the \( \epsilon_{ijk} \) are independent standard normal variates. The main effects and interactions were generated such that

\[ \sum_i \alpha_i^2 = 1, \ \sum_j \beta_j^2 = 1, \ \text{and} \ \sum_i \sum_j \gamma_{ij}^2 = 1 \]

so that \( \kappa_1 \) and \( \kappa_2 \) can be used to control the magnitude of the effects. Because we already found evidence that they are adequate for linear models, we chose \( B = 100 \) replicates and \( \lambda \) as defined in (2.6) to estimate the degrees of freedom for each simulated data set. Figure 2.2 shows the rejection rates for \( F \)-tests for one of the main effects formed using the resampling degrees of freedom as well as the true degrees of freedom. The magnitudes of the interactions were held constant using \( \kappa_2 = 4 \) and the magnitudes of the main effects were controlled by varying the value of \( \kappa_1 \). As the figure shows, the two curves are nearly identical which means that tests formed using the resampling degrees of freedom have the same power to detect main effects as tests formed using the true degrees of freedom.

Another simulation was used to investigate the rejection rates of \( F \)-tests for interaction formed using the resampling degrees of freedom as well as the true degrees of freedom. Data were generated by (2.7) with \( \kappa_1 = 2 \), while varying the value of \( \kappa_2 \) to control the magnitude of the interactions. The resampling method used \( B = 100 \) replicates and \( \lambda \) as defined in (2.6). Figure 2.3 shows that tests formed using the re-
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Figure 2.2: Power of Main Effect F-tests using True DF and Resampling DF (Monte Carlo SE: 6.9E-3)

Figure 2.3: Power of Interaction F-tests using True DF and Resampling DF (Monte Carlo SE: 6.9E-3)
sampling degrees of freedom and tests formed using the true degrees of freedom have identical power to detect interactions.

Through simulation, we have found that the resampling method provides reliable degrees of freedom for the two-factor fixed effects ANOVA model when $B = 100$ and $\lambda$ is defined in (2.6). The resampling estimates are unbiased and lead to $F$-tests with the same properties as tests formed using the true degrees of freedom.

**Resampling Degrees of Freedom and Tukey’s One-DF Model**

Tukey’s one-degree-of-freedom model is a simple non-linear model where it is accepted that the interactions contribute a single degree of freedom to the model. Estimating the non-additivity parameter $k$ from (1.10) requires a single degree of freedom, hence the name of the model. A simulation was performed to determine whether or not the resampling method assigns a single degree of freedom to the non-additivity parameter. One thousand data sets were generated from Tukey’s model with $k = .5$

\[
E(Y_{ij}) = \mu + \alpha_i + \beta_j + .5\alpha_i\beta_j \quad i = 1, ..., 5; \quad j = 1, ..., 5; \quad (2.8)
\]

where $\alpha_i$ and $\beta_j$ were generated such that $\sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1$. The error terms are independent standard normal variates. The main effects model was used to obtain $\hat{\sigma}_0^2$ because we want to check the degrees of freedom associated with the non-additivity parameter. For each simulated data set, the resampling method was used with $B = 100$ replicates to estimate the degrees of freedom with $\lambda$ defined in (2.6). We chose these values of $B$ and $\lambda$ because they appear to be adequate for linear models. The results of the simulation are summarized in Table 2.4. The table shows that the resampling
method produces unbiased degree of freedom estimates for Tukey’s model when using this particular choice of $B$ and $\lambda$. Because the resampling method performed well for this nonlinear model, we have reason to believe that the resampling method will perform well with more complex nonlinear models, such as the OI model and the AMMI model.

### 2.4 Resampling Degrees of Freedom and the AMMI Model

As discussed in previous sections, the AMMI model is a complex nonlinear model where we do not have statistical theory to assign degrees of freedom to the principal component axes included in the model. Assigning degrees of freedom for this model is not a trivial task and several techniques have been suggested for approaching this problem. Gollub’s Rule, the pure noise method, and the pattern plus noise methods are all described in Section 1.6.2. Because they are simple to calculate, the Gollub
degrees of freedom are used frequently in practice. We conducted a simulation study
to compare the performance of $F$-tests constructed using the resampling degrees of
freedom to the performance of tests constructed using Gollub’s rule, the pure noise
method, and the pattern plus noise method for assigning degrees of freedom to the
AMMI model.

AMMI Model Testing Simulations

Data were generated from an AMMI(2) model as stated in (1.13) with five levels of
each experimental factor and no replication ($c = 1$). The model is written

$$E(Y_{ij}) = \mu + \kappa_1(\alpha_i + \beta_j) + \kappa_2\gamma_{ij}, \quad (2.9)$$

where interactions are of the AMMI(2) form stated in (1.13)

$$\gamma_{ij} = \sum_{n=1}^{2} \theta_n \tau_{in} \delta_{jn}.$$ 

For each simulated data set, the main effect parameters were randomly generated
subject to the usual sum to zero constraints ($\sum_i \alpha_i = \sum_j \beta_j = 0$), and subject to
the constraints: $\sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1$. Generating the main effect parameters in this
fashion allows us to use $\kappa_1$ to control the magnitude of the main effects. The interaction
parameters were also randomly generated to satisfy the sum to zero constraints and
the constraint $\sum_i \sum_j \gamma_{ij}^2 = 1$, so that $\kappa_2$ can be used to control the magnitude of the
interactions. Two simulations were performed, one to investigate power to identify
main effects and another to investigate the power to identify interaction effects.
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

One thousand simulated data sets were generated from (2.9) with $\kappa_1 = 4$ and varying values of $\kappa_2$ to investigate the power to reject the hypothesis

$$H_0 : \gamma_{ij} = 0 \quad \text{for all } i,j.$$

The hypothesis was tested using

$$F = \frac{\text{MS(PCA 1)}}{\text{MSE}} \quad \text{(2.10)}$$

at the .05 level using the various methods for estimating the degrees of freedom for error and the degrees of freedom for the first principal component axis. The simulation based methods use $B=200$ simulation replicates to estimate the degrees of freedom and the resampling method uses $\lambda$ as defined in (2.6). We chose this definition of $\lambda$ because it was adequate for linear models as well as Tukey’s model and the number of simulation replicates was increased to $B = 200$ to account for the fact that the AMMI model is more complex than Tukey’s model (where $B = 100$ was adequate).

Figure 2.4 shows the rejection rates of $F$-tests of size $\alpha = .05$ formed using the various degrees of freedom methods. Clearly tests formed using the Gollub degrees of freedom offer the greatest power to detect an interaction effect, but the Type I error rate is unacceptable (31%). Cornelius also found that using Gollub degrees of freedom can lead to high Type I error rates [6]. Tests formed using the resampling degrees of freedom are conservative, but the Type I error rate is more appropriate (only 1.5%). Tests formed using the pure noise degrees of freedom are more conservative than the tests formed using the resampling degrees of freedom. Figure 2.4 does not
show the curve associated with the pattern plus noise method because these tests are more conservative than the tests formed using the pure noise degrees of freedom. So according to Figure 2.4, tests for interaction formed using the resampling degrees of freedom are the most appropriate because although they are slightly conservative, they provide the best power while still achieving the nominal Type I error rate.

A similar simulation was conducted where $\kappa_2 = 8$ was fixed and $\kappa_1$ was varied to investigate the power to detect main effects by testing the hypothesis

$$H_0 : \alpha_i = 0 \quad \text{for all } i$$

at the .05 level. The results of the simulation are provided in Figure 2.5. Again, tests formed using the Gollub degrees of freedom offer the best power to detect main effects.
However unlike the tests for interaction, the Type I error rate for the Gollub test is more acceptable, only 6.5%. The tests formed using the resampling degrees of freedom are more conservative than the tests formed using Gollub degrees of freedom, but the resampling degrees of freedom offer the greatest power among the simulation based degrees of freedom methods. The Type I error rate for tests formed using the resampling degrees of freedom is .5%, which is very low. Once again the curve associated with the pattern plus noise method is omitted because it is the most conservative.

The simulations performed in this subsection suggest that when using $F$-tests to perform inference on the AMMI model, the resampling method should be used to estimate degrees of freedom in the model. Although the tests are conservative (Type I error rate less than 2% when testing at the .05 level), the power to detect main effects
and interactions is better than the other simulation based methods, and the Type I error rates are better than the Gollub tests. The low Type I error rates for tests formed using resampling degrees of freedom could be due to the fact that the statistic in (2.10) does not have an $F$ distribution exactly. When fitting the AMMI model, investigators are typically most interested in testing for interaction effects, so the high Type I error rates associated with Gollub tests are a concern. Similar simulations were conducted fitting the AMMI model to random noise and when the Gollub degrees of freedom are used to conduct tests at the .05 level, the Type I error rates for main effects can be as high as 30% and the Type I error rate for testing interactions can be as high as 53%. Clearly the Gollub degrees of freedom are not reliable and the resampling degrees of freedom should be used instead.

**AMMI Model and Error Variance Estimation**

Another simulation was conducted to evaluate how the different methods of assigning degrees of freedom perform in terms of estimating the error variance. One thousand data sets were generated according to (2.9) with $\kappa_1 = 4$ and $\kappa_2 = 8$ with standard normal errors ($\sigma^2 = 1$). For each simulated data set, the error mean square (using each degree of freedom method) was used to estimate the error variance. For the simulation based methods, $B=200$ replicates were used to estimate the degrees of freedom. The results of the simulation are summarized in Table 2.5. The column labeled ‘10% TM($\hat{\sigma}^2$)’ represents the 10% trimmed mean of the error variance estimates and the column labeled ‘SL($\hat{\sigma}^2$)’ represents Stein’s loss function defined in (1.28). No matter which method for estimating degrees of freedom is used, the error mean square
Table 2.5: Summary of Error Variance Estimation for AMMI Data ($\sigma^2 = 1$)

<table>
<thead>
<tr>
<th>DF Method</th>
<th>Median($\hat{\sigma}^2$)</th>
<th>Mean($\hat{\sigma}^2$)</th>
<th>10% TM($\hat{\sigma}^2$)</th>
<th>Std. Dev.($\hat{\sigma}^2$)</th>
<th>SL($\hat{\sigma}^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gollub</td>
<td>.58</td>
<td>.74</td>
<td>.66</td>
<td>.53</td>
<td>.29</td>
</tr>
<tr>
<td>Pure Noise</td>
<td>1.53</td>
<td>1.89</td>
<td>1.69</td>
<td>1.36</td>
<td>.51</td>
</tr>
<tr>
<td>Pattern+Noise</td>
<td>3.12</td>
<td>3.38</td>
<td>3.21</td>
<td>1.88</td>
<td>1.32</td>
</tr>
<tr>
<td>Resampling</td>
<td>1.21</td>
<td>1.67</td>
<td>1.41</td>
<td>1.46</td>
<td>.46</td>
</tr>
</tbody>
</table>

does not do a good job of estimating the error variance. Gollub’s method leads to underestimating the true error variance, but the estimates have the lowest relative bias and they are favored by Stein’s loss function as well. Of the simulation based methods, the resampling method produces the best estimates of error variance in terms of relative bias and Stein’s loss function.

A similar simulation was conducted for estimating the error variance when fitting the AMMI model to random noise ($\kappa_1 = \kappa_2 = 0$). For each simulated data set, the mean square error was used to estimate the error variance ($\sigma^2 = 1$). The results are summarized in Table 2.6. The results suggest that the pattern plus noise and pure noise methods offer the best estimates when there is no signal in the data. It should not be surprising that the pure noise method performs well in this situation because the method is motivated by analyzing pure noise. The median and trimmed mean error variance estimate using the resampling degrees of freedom are close to the true value, suggesting that the resampling method comes close to providing unbiased estimates but several large outliers have a heavy influence on the mean value. The Gollub degrees of freedom provide the worst estimates in this case in terms of relative bias and Stein’s
Chapter 2. A Resampling Based Method For Estimating Degrees of Freedom

Table 2.6: Summary of Error Variance Estimation for Random Noise ($\sigma^2 = 1$)

<table>
<thead>
<tr>
<th>DF Method</th>
<th>Median($\hat{\sigma}^2$)</th>
<th>Mean($\hat{\sigma}^2$)</th>
<th>10% TM($\hat{\sigma}^2$)</th>
<th>Std. Dev.($\hat{\sigma}^2$)</th>
<th>SL($\hat{\sigma}^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gollub</td>
<td>.32</td>
<td>.39</td>
<td>.35</td>
<td>.27</td>
<td>.58</td>
</tr>
<tr>
<td>Pure Noise</td>
<td>.85</td>
<td>1.02</td>
<td>.91</td>
<td>.73</td>
<td>.26</td>
</tr>
<tr>
<td>Pattern+Noise</td>
<td>.84</td>
<td>.98</td>
<td>.91</td>
<td>.61</td>
<td>.21</td>
</tr>
<tr>
<td>Resampling</td>
<td>.82</td>
<td>1.31</td>
<td>.98</td>
<td>2.56</td>
<td>.54</td>
</tr>
</tbody>
</table>

The simulation results summarized in Table 2.5 and Table 2.6 are disappointing because using the resampling degrees of freedom to calculate the error mean square does not provide unbiased error variance estimates. We found that these error variance estimates do not improve by increasing $B$ or by using a different $\lambda$ vector. We do not think that this is due to a shortcoming of the resampling method, instead suggesting that the error mean square is not a good estimator of error variance for the AMMI model. But investigators are typically more interested in testing, and the $F$-tests conducted using the resampling degrees of freedom exhibit favorable qualities (although they are conservative).

In this section we compared the resampling degrees of freedom to the existing degrees of freedom methods for the AMMI model. The resampling degrees of freedom are competitive with the other methods of estimating degrees of freedom in terms of testing and error variance estimation. Tests formed using the resampling degrees of freedom exhibit power to detect main effects and interactions without the complications of high Type I error rates. We also found that for the AMMI model, the error mean
square is not a good estimator of the error variance.

2.5 Discussion

In this chapter we have presented a new method for estimating degrees of freedom for complex models. The resampling method is motivated by the linear model, where adding noise to the response variable results in a linear increase in the expected sums of squares. The method is simple to implement, as it only involves adding errors of varying magnitude to the response variable and observing the rate of change of the average sums of squares. Through simulation we found that the resampling method provides unbiased degree of freedom estimates for the two factor fixed-effects model as well as Tukey’s one-degree-of-freedom model when $B$, the number of simulation replicates, is relatively small. For the more complex AMMI model where we do not have a generally accepted method for estimating degrees of freedom, the resampling method performs better than other methods of estimating degrees of freedom in terms of forming test statistics and estimating error variance. Although $F$-tests conducted using the resampling degrees of freedom for the AMMI model are conservative, they avoid the high Type I error rates which are a problem with the Gollub degrees of freedom. Furthermore, we saw in the first chapter that the resampling method provides reliable degrees of freedom for the orthogonal interactions model as well. The resampling method is attractive because it can be used with virtually any statistical model. Although the resampling method is computationally intense, using the method with a single data set is rarely a problem.
Literature Cited


Literature Cited


Appendix
Here we provide proofs of several results used in the thesis.

A.1 A Linear Algebra Identity

Proposition 1. If \( X \) is an \( a \times b \) matrix with elements \( x_{ij} \), then

\[
\sum_i \sum_j x_{ij}^2 = \text{tr}(XX^T) = \text{tr}(X^TX).
\]

Proof. Let

\[
X = \begin{bmatrix} C_1, & C_2, & \ldots, & C_b \end{bmatrix} \quad \text{where} \quad C_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{aj} \end{bmatrix}.
\]

Then

\[
X^TX = \begin{bmatrix} C_1^T \\ C_2^T \\ \vdots \\ C_b^T \end{bmatrix} \begin{bmatrix} C_1, & C_2, & \ldots, & C_b \end{bmatrix} = \begin{bmatrix} C_1^TC_1 & C_1^TC_2 & \ldots & C_1^TC_b \\ C_2^TC_1 & C_2^TC_2 & \ldots & C_2^TC_b \\ \vdots & \vdots & \ddots & \vdots \\ C_b^TC_1 & C_b^TC_2 & \ldots & C_b^TC_b \end{bmatrix}.
\]

The trace of \( X^TX \) is

\[
\text{tr}(X^TX) = \sum_{j=1}^b C_j^TC_j \quad \text{where} \quad C_j^TC_j = \sum_{i=1}^a x_{ij}^2.
\]

Then

\[
\text{tr}(X^TX) = \sum_{j=1}^b \sum_{i=1}^a x_{ij}^2.
\]
Appendix A. Technical Details

Now let $A = X^T$, then $\text{tr} (X^T X)$ can be written:

$$\text{tr} (X^T X) = \sum_{j=1}^{b} \sum_{i=1}^{a} A_{ij} x_{ji} = \sum_{j=1}^{b} \sum_{i=1}^{a} x_{ji} A_{ij} = \text{tr} (X A) = \text{tr} (X X^T).$$

So

$$\text{tr} (X X^T) = \text{tr} (X^T X) = \sum_{j=1}^{b} \sum_{i=1}^{a} x_{ij}^2.$$

A.2 Two Constrained Optimality Results

**Proposition 1.** Given an $a \times b$ matrix $Y$, and vectors $\alpha$ and $\beta$ of size $a \times 1$ and $b \times 1$ respectively, define $A = \begin{bmatrix} 1_a, \alpha \end{bmatrix}$ an $a \times 2$ matrix and $B = \begin{bmatrix} 1_b, \beta \end{bmatrix}$ a $b \times 2$ matrix, where $1_k$ denotes a $k \times 1$ vector of ones. Also define

$$P_A = A(A^T A)^{-1} A^T \quad \text{and} \quad P_B = B(B^T B)^{-1} B^T,$$

the projection matrices corresponding to $A$ and $B$ respectively. Then

$$\hat{\gamma} = (I - P_A) Y (I - P_B)$$

minimizes

$$Q(\gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \gamma_{ij})^2$$

subject to the constraints

$$\gamma^T \alpha = 0, \quad \gamma^T 1_a = 0, \quad \gamma \beta = 0, \quad \text{and} \quad \gamma 1_b = 0. \quad (A.1)$$

**Proof.** We first establish that $\hat{\gamma}$ satisfies all of the stated constraints. Consider $\hat{\gamma}^T 1_a = 0$. Using properties of $P_A$, $P_B$, $I - P_A$, and $I - P_B$, it follows that

$$\hat{\gamma}^T 1_a = (I - P_B) Y^T (I - P_A) 1_a = 0,$$
Appendix A. Technical Details

because \((\mathbf{I} - \mathbf{P}_A) \mathbf{1}_a = \mathbf{1}_a - \mathbf{1}_a = \mathbf{0}\). Similarly,

\[
\hat{\gamma}_b = (\mathbf{I} - \mathbf{P}_A) \mathbf{Y} (\mathbf{I} - \mathbf{P}_B) \mathbf{1}_b = 0.
\]

Thus we have established that the columns and rows of \(\hat{\gamma}\) both sum to zero. Next we show that \(\hat{\gamma}^T \alpha = \mathbf{0}\)

\[
\hat{\gamma}^T \alpha = (\mathbf{I} - \mathbf{P}_B) \mathbf{Y}^T (\mathbf{I} - \mathbf{P}_A) \alpha = 0,
\]

using the fact that \(\mathbf{P}_A \alpha = \alpha\). Similarly,

\[
\hat{\gamma}^T \beta = (\mathbf{I} - \mathbf{P}_A) \mathbf{Y} (\mathbf{I} - \mathbf{P}_B) \beta = 0.
\]

So \(\hat{\gamma} = (\mathbf{I} - \mathbf{P}_A) \mathbf{Y} (\mathbf{I} - \mathbf{P}_B)\) satisfies the constraints stated in (A.1). Next we show that \(\hat{\gamma}\) minimizes \(Q(\gamma)\).

Let \(\tilde{\gamma}_\Delta = \hat{\gamma} + \Delta\). We will show that if \(\tilde{\gamma}_\Delta\) satisfies the constraints in (A.1), then it follows that \(\Delta^T \alpha = \mathbf{0}\), \(\Delta \beta = \mathbf{0}\), \(\Delta^T \mathbf{1}_a = \mathbf{0}\), and \(\Delta \mathbf{1}_b = \mathbf{0}\). First, note that \(\tilde{\gamma}_\Delta^T \alpha = \mathbf{0}\) implies

\[
\tilde{\gamma}_\Delta^T \alpha = \mathbf{0} = (\hat{\gamma} + \Delta)^T \alpha = \hat{\gamma}^T \alpha + \Delta^T \alpha = \Delta^T \alpha,
\]

and thus \(\Delta^T \alpha = \mathbf{0}\). Next, note that \(\tilde{\gamma}_\Delta^T \mathbf{1}_a = \mathbf{0}\):

\[
\tilde{\gamma}_\Delta^T \mathbf{1}_a = \mathbf{0} = (\hat{\gamma} + \Delta)^T \mathbf{1}_a = \hat{\gamma}^T \mathbf{1}_a + \Delta^T \mathbf{1}_a = \Delta^T \mathbf{1}_a.
\]
Appendix A. Technical Details

and thus \( \Delta^T 1_a = 0 \). Similarly

\[
\tilde{\gamma}_\Delta \beta = 0 = (\hat{\gamma} + \Delta) \beta = \begin{pmatrix} \hat{\gamma} \beta + \Delta \beta \end{pmatrix}_0 = \Delta \beta
\]

implies that \( \Delta \beta = 0 \). And finally

\[
\tilde{\gamma}_\Delta 1_b = 0 = (\hat{\gamma} + \Delta) 1_b = \begin{pmatrix} \hat{\gamma} 1_b + \Delta 1_b \end{pmatrix}_0 = \Delta 1_b
\]

implies \( \Delta 1_b = 0 \). We now show that \( Q(\tilde{\gamma}_\Delta) = Q(\hat{\gamma}) + \text{tr} (\Delta^T \Delta) \). The first step is to write \( \tilde{\gamma}_\Delta \) in terms of \( \hat{\gamma} \) and \( \Delta \),

\[
Q(\tilde{\gamma}_\Delta) = \text{tr} \left\{ (Y - \tilde{\gamma}_\Delta)^T (Y - \tilde{\gamma}_\Delta) \right\}
\]

\[
= \text{tr} \left\{ (Y - \hat{\gamma} - \Delta)^T (Y - \hat{\gamma} - \Delta) \right\}
\]

\[
= \text{tr} \left\{ (Y - \hat{\gamma})^T (Y - \hat{\gamma}) + \Delta^T \Delta - (Y - \hat{\gamma})^T \Delta - \Delta^T (Y - \hat{\gamma}) \right\}.
\]

For any square matrix \( M \), \( \text{tr} (M) = \text{tr} (M^T) \). So then \( \text{tr} \left\{ (Y - \hat{\gamma})^T \Delta \right\} = \text{tr} \left\{ \Delta^T (Y - \hat{\gamma}) \right\} \) and \( Q(\tilde{\gamma}_\Delta) \) can be written

\[
Q(\tilde{\gamma}_\Delta) = Q(\hat{\gamma}) + \text{tr} (\Delta^T \Delta) - 2 \text{tr} \left\{ (Y - \hat{\gamma})^T \Delta \right\}.
\]

Now showing that \( \text{tr} \left\{ (Y - \hat{\gamma})^T \Delta \right\} = 0 \) will be sufficient to show that \( Q(\tilde{\gamma}_\Delta) = Q(\hat{\gamma}) + \text{tr} (\Delta^T \Delta) \). To this end, consider that

\[
\text{tr} \left\{ (Y - \hat{\gamma})^T \Delta \right\} = \text{tr} \left\{ (Y - (I - P_A)Y(I - P_B))^T \Delta \right\}
\]

\[
= \text{tr} \left\{ (Y - (Y - YP_B - P_A Y + P_A Y P_B))^T \Delta \right\}
\]

\[
= \text{tr} \left\{ (YP_B + P_A Y - P_A Y P_B)^T \Delta \right\}
\]

\[
= \text{tr} (P_B Y^T \Delta) + \text{tr} (Y^T P_A \Delta) - \text{tr} (P_B Y^T P_A \Delta).
\]
Appendix A. Technical Details

Furthermore,

\[
P_A \Delta = A (A^T A)^{-1} A^T \Delta
\]

\[
= A (A^T A)^{-1} \begin{bmatrix} a^T \\ 1^T_a \end{bmatrix} \Delta
\]

\[
= A (A^T A)^{-1} \begin{bmatrix} a^T \Delta \\ 1^T_a \Delta \end{bmatrix}
\]

\[
= 0,
\]

from which it follows that \( \text{tr} (Y^T P_A \Delta) = \text{tr} (P_B Y^T P_A \Delta) = 0 \); and also that

\[
\text{tr} \left\{ (Y - \hat{\gamma})^T \Delta \right\} = \text{tr} (P_B Y^T \Delta) + \text{tr} \left( Y^T P_A \Delta \right) - \underbrace{\text{tr} (P_B Y^T P_A \Delta)}_{0}
\]

\[
= \text{tr} (P_B Y^T \Delta)
\]

\[
= \text{tr} (\Delta P_B Y^T)
\]

\[
= \text{tr} \left\{ \Delta B (B^T B)^{-1} B^T Y^T \right\}
\]

\[
= \text{tr} \left\{ \begin{bmatrix} \Delta 1_b, & \Delta \beta \end{bmatrix} (B^T B)^{-1} B^T Y^T \right\}
\]

\[
= 0.
\]

So \( -2 \text{tr} \{(Y - \hat{\gamma})^T \Delta\} = 0 \), implying that

\[
Q(\hat{\gamma}_\Delta) = Q(\hat{\gamma}) + \text{tr} (\Delta^T \Delta).
\]

But from Appendix A.1, \( \text{tr} (\Delta^T \Delta) = \sum_i \sum_j \Delta^2_{ij} \geq 0 \). Thus \( Q(\hat{\gamma}_\Delta) \geq Q(\hat{\gamma}) \) with equality only when \( \Delta \) is a matrix of zeros. So we have shown that for \( \hat{\gamma}_\Delta \) satisfying the constraints in (A.1), that \( Q(\hat{\gamma}_\Delta) \geq Q(\hat{\gamma}) \). Therefore \( \hat{\gamma} = (I - P_A) Y (I - P_B) \) minimizes \( Q(\gamma) \) subject to the constraints in (A.1). \( \square \)
Proposition 2. Given $a \times b$ matrices $Y$ and $\gamma$, define $U = \begin{bmatrix} 1_a & \gamma \end{bmatrix}$, $V = \begin{bmatrix} 1_b & \gamma^T \end{bmatrix}$.

$P_U = U(U^TU)^{-}U^T$, $P_V = V(V^TV)^{-}V^T$,

$\bar{Y}_a = \begin{bmatrix} Y_{11} - \bar{Y}_- \\ Y_{21} - \bar{Y}_- \\ \vdots \\ Y_{a1} - \bar{Y}_- \end{bmatrix}$, and $\bar{Y}_b = \begin{bmatrix} Y_{1b} - \bar{Y}_- \\ Y_{2b} - \bar{Y}_- \\ \vdots \\ Y_{b1} - \bar{Y}_- \end{bmatrix}$,

where $1_k$ represents a $k \times 1$ vector of ones and $X^g$ denotes the generalized inverse of the matrix $X$.

Then

$\hat{\alpha} = (I - P_U)\bar{Y}_a$ and $\hat{\beta} = (I - P_V)\bar{Y}_b$

minimize

$J(\alpha, \beta) = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \alpha_i - \beta_j)^2$,

subject to the constraints

$\alpha^T 1_a = 0$, $\beta^T 1_b = 0$, $\gamma^T \alpha = 0$, and $\gamma \beta = 0$. (A.2)

Proof. First we show that $\hat{\alpha}$ and $\hat{\beta}$ satisfy the relevant constraints. In showing that $\hat{\alpha}$ and $\hat{\beta}$ satisfy these constraints, we make frequent use of the fact that $P_U$ projects onto $\gamma$ and $1_a$, as well as the fact that $P_V$ projects onto $\gamma^T$ and $1_b$. Using the properties of these projections, it follows that

$\hat{\alpha}^T 1_a = \left\{ (I - P_U)\bar{Y}_a \right\}^T 1_a$

$= \bar{Y}_a^T (I - P_U) 1_a$

$= \bar{Y}_a^T (1_a - P_U 1_a)$

$= 0$. 

118
Similarly,
\[
\hat{\beta}^T 1_b = \{(I - P_V)Y_b\}^T 1_b
\]
\[
= Y_b^T (I - P_V) 1_b
\]
\[
= Y_b^T (1_b - P_V 1_b) =_{1_b}
\]
\[
= 0.
\]
So \(\hat{\alpha}\) and \(\hat{\beta}\) both sum to zero as required. Next we show that \(\hat{\alpha}^T \gamma = 0\)
\[
\hat{\alpha}^T \gamma = \{(I - P_U)Y_a\}^T \gamma
\]
\[
= Y_a^T (I - P_U) \gamma
\]
\[
= Y_a^T (\gamma - P_U \gamma) =_{\gamma}
\]
\[
= 0.
\]
Similarly,
\[
\gamma \hat{\beta} = \gamma (I - P_V) Y_b
\]
\[
= (\gamma - \gamma P_V) Y_b
\]
\[
= \{\gamma - (P_V \gamma^T)^T\} Y_b =_{\gamma^T}
\]
\[
= (\gamma - \gamma) Y_b
\]
\[
= 0.
\]
We have established that \(\hat{\alpha}\) and \(\hat{\beta}\) satisfy the constraints in (A.2). Next we prove the optimality of \(\hat{\alpha}\) and \(\hat{\beta}\) by showing that for any other \(\tilde{\alpha}\) and \(\tilde{\beta}\) that satisfy the constraints in (A.2), \(J(\hat{\alpha}, \hat{\beta}) \leq J(\tilde{\alpha}, \tilde{\beta})\).
Assuming that \(\tilde{\alpha}^T 1_a = 0\), it follows that
\[
\tilde{\alpha}^T 1_a = 0 = (\hat{\alpha} + \Delta_1)^T 1_a
\]
\[
= \hat{\alpha}^T 1_a + \Delta_1^T 1_a
\]
\[
= \Delta_1^T 1_a,
\]
119
Appendix A. Technical Details

and thus $\Delta_1^T 1_a = 0$. Similarly, $\hat{\alpha}^T \gamma = 0$ implies
\[
\hat{\alpha}^T \gamma = 0 = (\hat{\alpha} + \Delta_1)^T \gamma = \hat{\alpha}^T \gamma + \Delta_1^T \gamma = \Delta_1^T \gamma.
\]

Next if $\tilde{\beta}^T 1_b = 0$, then
\[
\tilde{\beta}^T 1_b = 0 = (\hat{\beta} + \Delta_2)^T 1_b = \tilde{\beta}^T 1_b + \Delta_2^T 1_b = \Delta_2^T 1_b.
\]

Similarly, if $\gamma \tilde{\beta} = 0$, then
\[
\gamma \tilde{\beta} = 0 = \gamma (\hat{\beta} + \Delta_2) = \gamma \hat{\beta} + \gamma \Delta_2 = \gamma \Delta_2.
\]

Now we show that $J(\hat{\alpha}, \hat{\beta}) \leq J(\tilde{\alpha}, \tilde{\beta})$. The first step is finding an expression for $J(\hat{\alpha}, \hat{\beta})$.

\[
J(\hat{\alpha}, \hat{\beta}) = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_j)^2 = \text{tr} \{(Y - \hat{\alpha} 1_b^T - 1_a \hat{\beta}^T)(Y - \hat{\alpha} 1_b^T - 1_a \hat{\beta}^T)\} = \text{tr} \{Y^T(Y - \hat{\alpha} 1_b^T - 1_a \hat{\beta}^T) - 1_b \hat{\alpha}^T Y + \hat{\alpha} \hat{\beta}^T + 1_b \hat{\alpha} \hat{\beta}^T - \hat{\beta} 1_a^T Y\} + \text{tr} \{\hat{\beta} 1_b^T \hat{\alpha} - \hat{\beta} 1_a^T Y + a \hat{\beta} \hat{\beta}^T\}.
\]
Appendix A. Technical Details

Focusing on $J(\tilde{\alpha}, \tilde{\beta})$,

$$J(\tilde{\alpha}, \tilde{\beta}) = \sum_{j=1}^{b} \sum_{i=1}^{a} (Y_{ij} - \tilde{\alpha}_i - \tilde{\beta}_j)^2$$

$$= \text{tr}\{ (Y - \tilde{\alpha}1_b^T - 1_a\tilde{\beta}^T)(Y - \tilde{\alpha}1_b^T - 1_a\tilde{\beta}^T) \}$$

$$= \text{tr}\{ Y^T(Y - \tilde{\alpha}1_b^T - 1_a\tilde{\beta}^T) - 1_b\tilde{\alpha}^TY + b\tilde{\beta}^TY + 1_b\tilde{\alpha}\tilde{\beta}1_b^T \}$$

$$= \text{tr}\{ Y^T(Y - (\tilde{\alpha} + \Delta_1)1_b^T - 1_a(\tilde{\beta} + \Delta_2)^T) - 1_b(\tilde{\alpha} + \Delta_1)^TY + b(\tilde{\alpha} + \Delta_1)^T(\tilde{\alpha} + \Delta_1) - (\tilde{\beta} + \Delta_2)1_a^TY + a(\tilde{\beta} + \Delta_2)^T(\tilde{\beta} + \Delta_2) \}$$

$$= J(\tilde{\alpha}, \tilde{\beta}) + b\Delta_1^T\Delta_1^T + a\Delta_2^T\Delta_2$$

$$+ \text{tr}\{ 2(b\tilde{\alpha}\Delta_1 - Y^T\Delta_11_b^T) + 2(a\tilde{\beta}\Delta_2 - Y^T1_a\Delta_2^T) \}. \quad (I)$$

$$+ \text{tr}\{ 2(b\tilde{\alpha}\Delta_1 - Y^T1_a\Delta_2^T) \} \quad (II)$$

Now we focus on $\text{tr}\{ 2(b\tilde{\alpha}\Delta_1 - Y^T\Delta_11_b^T) + 2(a\tilde{\beta}\Delta_2^T - Y^T1_a\Delta_2^T) \}$ by breaking the sum into two pieces. Let $\text{tr}(I) = \text{tr}\{ 2(b\tilde{\alpha}\Delta_1 - Y^T\Delta_11_b^T) \}$ and $\text{tr}(II) = \text{tr}\{ 2(a\tilde{\beta}\Delta_2^T - Y^T1_a\Delta_2^T) \}$. Here we utilize the fact that $\tilde{Y}_a = b^{-1}Y1_b - \tilde{\gamma}_a1_a$ and $\tilde{Y}_b = a^{-1}Y^T1_a - \tilde{\gamma}_b1_b$. First consider $\text{tr}(I)$:

$$\text{tr}(I) = 2\text{tr}\left( b\tilde{\alpha}\Delta_1 - Y^T\Delta_11_b^T \right)$$

$$= 2\text{tr}\left\{ b\tilde{Y}_a^T(I - P_U)\Delta_1 - 1_b^TY^T\Delta_1 \right\}$$

$$= 2\text{tr}\left\{ b\tilde{Y}_a^T\Delta_1 - b\tilde{Y}_a^TP_U\Delta_1 - b(\tilde{Y}_a^T + \tilde{\gamma}_a1_a^T)\Delta_1 \right\}$$

$$= 2\text{tr}\left\{ b\tilde{Y}_a^T\Delta_1 - b\tilde{Y}_a^TU^T(U^TU)^g \begin{bmatrix} 1_a^T \\ \tilde{\gamma}_T^T \end{bmatrix} \Delta_1 \right\}$$

$$= 2\text{tr}\left\{ - b\tilde{Y}_a^TU^T(U^TU)^g \begin{bmatrix} 1_a^T \Delta_1 \\ \tilde{\gamma}_T^T \Delta_1 \end{bmatrix} \right\}$$

$$= 0.$$
Appendix A. Technical Details

A similar argument shows that

\[
\text{tr}(\text{II}) = 2\text{tr}\left( a\hat{\beta}\Delta_2^T - Y^T 1_a\Delta_2^T \right) \\
= 2\text{tr}\{a(I - P_V)\bar{Y}_b\Delta_2^T - a(\bar{Y}_b + \bar{Y}_1\Delta_2^T)\} \\
= 2\text{tr}(-a\Delta_2^TP_V\bar{Y}_b) \\
= 2\text{tr}\{-a\Delta_2^T V(V^T V)^g\bar{Y}_b\} \\
= 2\text{tr}\{-a\left[ \Delta_2^T 1_b, \Delta_2^T \gamma^T \right](V^T V)^g\bar{Y}_b\} \\
= 0.
\]

So \(\text{tr}(\text{I}) = \text{tr}(\text{II}) = 0\). Therefore,

\[
J(\hat{\alpha}, \hat{\beta}) = J(\hat{\alpha}, \hat{\beta}) + b\Delta_1^T\Delta_1 + a\Delta_2^T\Delta_2,
\]

from which it follows that \(J(\hat{\alpha}, \hat{\beta}) \geq J(\hat{\alpha}, \hat{\beta})\) with equality only when \(\Delta_1 = 0\) and \(\Delta_2 = 0\). Therefore given \(\gamma, \hat{\alpha} = (I - P_U)\bar{Y}_a\) and \(\hat{\beta} = (I - P_V)\bar{Y}_b\) minimize \(J(\alpha, \beta)\) where \(\alpha, \beta,\) and \(\gamma\) satisfy the constraints in (A.2). \(\square\)