

# LONGITUDINAL ANALYSIS FOR BINARY AND COUNT DATA

By

Jinko Graham

B. Sc. (Mathematics) University of British Columbia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

We accept this thesis as conforming

to the required standard

---

THE UNIVERSITY OF BRITISH COLUMBIA

1991

© Jinko Graham, 1991

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of STATISTICS

The University of British Columbia  
Vancouver, Canada

Date DEC 23, 1991

## Abstract

Longitudinal data sets consist of repeated observations of an outcome over time, and a corresponding set of covariates for each of many subjects. In many fields, multivariate analysis-of-variance is commonly used to analyse longitudinal data. Such an analysis is appropriate when responses for each subject are multivariate Gaussian with a common covariance matrix for all subjects. In many cases, however, the longitudinal response cannot be transformed to satisfy these assumptions. An alternate analysis might rely on specification of an appropriate likelihood to obtain estimates of regression parameters and their standard errors. Unfortunately, the correlation structure in the data for each subject may not be well understood, making such parametric modelling difficult.

This thesis discusses two methods for the analysis of longitudinal data which require only minimal assumptions about the true correlation structure in the data for each subject to yield consistent estimates of regression parameters and their standard errors. The first method is based on the use of a “working” likelihood and extends the results of Zeger, Liang and Self (Biometrika, 1985) to the case of time-dependent covariates. The second method, first presented by Liang and Zeger (Biometrika, 1986), is based on quasi-likelihood theory. This method uses generalized estimating equations to arrive at consistent estimates of the regression coefficients and their standard errors, and can be applied to any longitudinal response with univariate marginal distributions for which the quasi-likelihood formulation is sensible. This

includes Gaussian, Poisson, binomial (bernoulli), gamma, and inverse Gaussian distributions. Both methods are extensively illustrated using the results from an experiment on hummingbird learning. These methods enable much more information to be extracted from the hummingbird data set than a more traditional analysis-of-variance, and therefore provide useful and powerful tools for researchers in this subject area.

## Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>Acknowledgement</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background to the Experiment . . . . .	1
1.2 Description of the Experiment . . . . .	3
1.3 Description of the Data . . . . .	5
1.4 Outline of the Thesis . . . . .	6
<b>2 Classical Analysis</b>	<b>9</b>
2.1 Univariate Analysis-of-Variance for Repeated Measures . . . . .	12
2.2 The Statistical Model . . . . .	13
2.3 Results of the Analysis . . . . .	14
2.4 Conclusions and Limitations . . . . .	18
<b>3 Analysis Based on a Working Likelihood Approach</b>	<b>20</b>
3.1 Logistic Regression for Binary Repeated Measures . . . . .	20

3.2	The Working Likelihood Approach . . . . .	20
3.3	Incorporating Time-dependent Covariates . . . . .	28
3.3.1	The Independence Working Model . . . . .	28
3.3.2	The First-Order Working Model . . . . .	30
3.4	Exploring the Binary Hummingbird Data . . . . .	34
3.5	Model Fitting With Sex Excluded . . . . .	37
3.5.1	The Initial Model . . . . .	37
3.5.2	An Informal Check of the Working Likelihoods . . . . .	38
3.6	Model Reductions With Sex Excluded . . . . .	40
3.6.1	Strategy for Model Reductions . . . . .	40
3.6.2	Description of the Reductions . . . . .	41
3.7	Residual Analysis for the Final Model with Sex Excluded . . . . .	47
3.7.1	Analysis of the Independence Residuals . . . . .	47
3.7.2	Analysis of the First-Order Residuals . . . . .	49
3.8	Summary of Analysis with Sex Excluded . . . . .	51
3.9	The Initial Model With Sex Included . . . . .	53
3.10	Model Reductions with Sex Included . . . . .	54
3.10.1	Strategy for Model Reductions . . . . .	54
3.10.2	Description of the Reductions for the Independence Working Likelihood . . . . .	55
3.10.3	Description of the Reductions for the First-Order Working Likelihood . . . . .	57
3.11	Summary of Analysis with Sex Included . . . . .	60

<b>4</b>	<b>Analysis Based on the GEE Approach</b>	<b>63</b>
4.1	Description of the GEE approach . . . . .	63
4.1.1	Brief Description of Quasi-Likelihood . . . . .	64
4.1.2	Generalized Estimating Equations (GEEs) for Longitudinal Data . . . . .	66
4.2	GEE Analysis of the Binary Hummingbird Data . . . . .	70
4.2.1	Description of the Model Reductions . . . . .	71
4.2.2	Fits of the Final Model Obtained by Other Working Correlations . . . . .	74
4.3	GEE Analysis of the Binomial Hummingbird Data . . . . .	76
4.3.1	Exploring the Binomial Hummingbird Data . . . . .	77
4.3.2	Model Fitting for the Binomial Hummingbird Data . . . . .	79
4.3.3	Fits of the Final Model Obtained by Other Working Correlations . . . . .	84
<b>5</b>	<b>Conclusion</b>	<b>87</b>
5.1	Discussion of Results . . . . .	87
5.2	Comparison of Methods . . . . .	89
	<b>Figures</b>	<b>92</b>
	<b>Bibliography</b>	<b>117</b>

## List of Tables

Table 2.1. Analysis of Variance for Blocks of 10 . . . . .	16
Table 2.2. Analysis of Variance for Blocks of 15 . . . . .	16
Table 3.1. Comparison of Z-scores from the Initial Model . . . . .	39
Table 3.2. Final Model Fits from the First Reduction Path . . . . .	43
Table 3.3. Final Model Fits from the Second Reduction Path . . . . .	46
Table 3.4. Estimated Autocorrelations for the Independence Residuals . . . . .	49
Table 3.5. Estimates of Transformed Lag One Autocorrelations . . . . .	50
Table 3.6. Model Reductions for the Independence Working Likelihood with Sex Included . . . . .	57
Table 3.7. Model Reductions for the First-Order Working Likelihood with Sex Included	59
Table 4.1. Reductions for the Stationary 1-Dependent Working Correlation with Sex Included . . . . .	74
Table 4.2. Robust z-Scores (Parameter Estimates) for the Binary Final Model . . . . .	75
Table 4.3. Model Reductions for the Binomial Data . . . . .	81
Table 4.4. Robust z-Score (Parameter Estimate) for the Binomial Final Model . . . . .	84



## List of Figures

Figure 1.1. Feeding Arrays for the Hummingbird Experiment . . . . .	92
Figure 2.1. Mean Treatment (Transformed) Response versus Measurement for Blocks of 15 . . . . .	93
Figure 2.2. Mean Treatment (Transformed) Response versus Measurement for Blocks of 10 . . . . .	93
Figure 2.3. Estimated Variances versus Measurement for Blocks of 15 . . . . .	94
Figure 2.4. Estimated Variances versus Measurement for Blocks of 10 . . . . .	94
Figure 3.1. Average Treatment Response in the Logit Scale for Blocks of 3 Trials . . .	95
Figure 3.2 Average Treatment Response in the Logit Scale for Blocks of 5 Trials . . .	96
Figure 3.3. Average Treatment Response in the Logit Scale for Blocks of 10 Trials . .	97
Figure 3.4. Smoothed Logit of Average Treatment Response versus Trial (Window = 10) . . . . .	98
Figure 3.5. Smoothed Logit of Average Treatment Response versus Trial (Window = 20) . . . . .	99
Figure 3.6. Smoothed Logit of Average Treatment Response versus Trial (Window = 30) . . . . .	100

Figure 3.7. Smoothed Individual Logits for Treatment 1 . . . . .	101
Figure 3.8. Smoothed Individual Logits for Treatment 2 . . . . .	102
Figure 3.9. Smoothed Individual Logits for Treatment 3 . . . . .	103
Figure 3.10. First Independence Fits without Sex . . . . .	104
Figure 3.11. Second Independence Fits without Sex . . . . .	104
Figure 3.12. Fitted Success Probabilities for Treatment 1 from the Independence Working Model . . . . .	105
Figure 3.13. Independence Pearson Residuals by Bird . . . . .	106
Figure 3.14. Boxplot of Independence Pearson Residuals by Sex . . . . .	107
Figure 3.15. Smoothed Logit of Average Response versus Trial for Treatment 1 (Window = 60) . . . . .	108
Figure 3.16. Smoothed Logit of Average Response versus Trial for Treatment 2 (Window = 60) . . . . .	108
Figure 3.17. Smoothed Logit of Average Response versus Trial for Treatment 3 (Window = 60) . . . . .	108
Figure 3.18. Binary Independence Fits for Females . . . . .	109
Figure 3.19. Binary First-Order Fits for Females . . . . .	109
Figure 3.20. Fitted Success Probabilities for Females in Treatment 1 . . . . .	110
Figure 4.1. Binary Fits for Females Obtained from the Stationary 1-Dependent Working Correlation . . . . .	111
Figure 4.2. Average Logit of Response for Treatment 1, Blocks of 10 Trials . . . . .	112

Figure 4.3. Average Logit of Response for Treatment 2, Blocks of 10 Trials . . . . .	112
Figure 4.4. Average Logit of Response for Treatment 3, Blocks of 10 Trials . . . . .	112
Figure 4.5. Smoothed Logit of Success Proportion versus Trial (Window = 30) . . . .	113
Figure 4.6. Smoothed Logit of Success Proportion versus Trial (Window = 60) . . . .	113
Figure 4.7. Smoothed Individual Logits (Binomial Data) for Treatment 1 (Window = 60) . . . . .	114
Figure 4.8. Smoothed Individual Logits (Binomial Data) for Treatment 2 (Window = 60) . . . . .	114
Figure 4.9. Smoothed Individual Logits (Binomial Data) for Treatment 3 (Window = 60) . . . . .	114
Figure 4.10. Initial Model Fits for Treatment 1 (Binomial Data) . . . . .	115
Figure 4.11. Initial Model Fits for Treatment 2 (Binomial Data) . . . . .	115
Figure 4.12. Initial Model Fits for Treatment 3 (Binomial Data) . . . . .	115
Figure 4.13. Final Fits for Females from the Stationary 1-Dependent Working Correlation . . . . .	116
Figure 4.14. Final Fits for Males from the Stationary 1-Dependent Working Correlation	116

## **Acknowledgement**

I would like to thank John Petkau for his guidance and supervision throughout the development of this thesis. I would also like to thank Nancy Heckman for her careful reading of the manuscript and subsequent helpful suggestions. Additionally, I am grateful to Gayle Brown for many instructive conversations about hummingbird learning, and, in particular, for Figure 1.1 of the text. Finally, I would like to thank the Statistical Consulting and Research Laboratory for providing the data.

## **Chapter 1**

### **Introduction**

#### **1.1 Background to the Experiment**

A desire for a more complete understanding of how organisms relate objects separated in space motivates much research on “spatial learning” in both the biological sciences and psychology. Currently, in the Department of Zoology at U.B.C., two Ph.D. students are studying spatial learning in hummingbirds. One of these students, Gayle Brown, came to the Statistical Consulting and Research Laboratory in the summer of 1990 with questions about how to analyze data she had collected in one of her experiments. The experiment was designed to answer questions about “spatial association learning” in hummingbirds. In “spatial association learning”, organisms learn to associate certain cues with reinforcement (such as food, water, or mates) separated in space from the cues. For example, Gori (1989) has shown that bees use the colours of spots on banner petals to select flowers with greater average nectar rewards.

Results from previous experiments conducted as part of the ongoing research of the investigator show that hummingbirds learn spatial associations very quickly relative to all other organisms that have been tested in the laboratory (monkeys, rats, pigeons, and children) and even perform perfectly on problems that may be unlearnable by other animals, according to

documentation in the literature (see reviews by MacKintosh, 1983; and Bowe, 1984). One explanation for the success of hummingbirds in spatial association experiments is that, in nature, hummingbirds must frequently learn floral cues in order to efficiently extract nectar hidden within flowers. The fact that many of these floral cues are visibly connected to the nectar of the flower suggests visible guides may facilitate the learning of spatial associations in hummingbirds.

To test this possibility, an experiment on the feeding behaviour of eighteen experimentally-naive adult rufous hummingbirds (10 females and 8 males) was carried out. The hummingbirds were assigned randomly by sex to one of three treatments: “No Tape” (NT or treatment 1), in which no visible guides connected cues with the feeders below them; “Partial Tape” (PT or treatment 2), in which fluorescent orange Dymo type (9 mm wide) provided a discontinuous (i.e., broken in two places) connection between each cue and its feeder; and “Full Tape” (FT or treatment 3), in which the visible guide between each cue and its feeder (fluorescent orange Dymo tape) was continuous. The NT group (treatment 1) had 4 females and 2 males, while the PT (treatment 2) and FT (treatment 3) groups each had 3 males and 3 females.

The investigator was interested in comparing learning patterns of the birds between these treatments. She expected different learning patterns on each treatment; in particular, she expected hummingbirds in the two treatment groups with visible guides between cues and feeders (2=PT and 3=FT) to become aware of the cue-feeder relationship sooner than hummingbirds in the treatment group with no visible guides (1=NT). From psychological principles of visual perception, she also expected birds in the treatment groups with discontinuous and continuous visible guides to become aware of the cue-feeder relationship at roughly the same time.

## 1.2 Description of the Experiment

On a given day, three birds (one bird from each treatment group) were tested separately in three randomly assigned rooms, 1.3 x 2.5 x 2.5 metres high. A portion of one end wall in each room contained a horizontal array of six feeders spaced 3 cm apart on a thin metal panel. Figure 1.1 displays the feeding arrays for the NT, PT, and FT groups. Feeders in each feeding array were marked by 19 mm round fluorescent orange Avery labels with central 3 mm holes. Hummingbirds hovered at the feeders and probed their bills through the holes to small food reservoirs. If the feeder was the correct or “profitable” one, a miniature solenoid valve immediately released 2  $\mu$ l of 20% sucrose solution into the reservoir. Hummingbirds were free to fly and visit feeders at all times during the experiment. Between brief foraging bouts to the feeding array and short non-foraging flights around the room, they spent most of their time on a 1.5 metre high central perch, located 1.8 metres from the feeding array and designed to place birds at eye level with the feeders.

A small red light, 4 mm in diameter, protruded slightly through the metal panel 12 cm above each feeder. These small lights served as the spatial cues. During each feeding trial one was lit to signal the feeder below it as profitable. A computer controlled the lights, dispensed the food, and recorded the time and duration of all visits to feeders and perches.

To allow the three hummingbirds to become accustomed to the feeding array used in the experiment, on the morning of the day before testing, the birds were placed in training cages identical to their home cages, except that food was available from a wall feeder similar in appearance to the feeding array. The birds quickly learned to feed from these feeders and were

moved from their training cages to the experimental rooms that afternoon. Until experimental training began the following morning, the hummingbirds fed from a standard commercial feeder marked with a fluorescent orange Avery label around the access hole. The feeder was hung in front of a central array feeder on which training would start. The other five feeders were covered.

At the beginning of the next day, the standard feeder was removed from each room, exposing a central feeder in the array to the hummingbird in each room. After the birds had fed several times from this feeder, the light above it was lit and training continued. After several more feedings, all other feeders were uncovered and the cue and profitability then reassigned to a different cue-feeder pair. This transfer procedure was repeated until birds had fed several times from each feeder when the light above it was lit. Training lasted approximately 2 hours.

Testing started between 10:30-11:00 am, immediately after training was finished. Each hummingbird was tested for 60 feeding trials on each of three successive days. Birds received initial training (as described in the preceding paragraph) on each testing day. On the first and second days, at the end of the day's 60 feeding trials, a commercial feeder was hung in front of a middle feeder in the array, the other feeders in the array were covered, and the birds remained in the experimental rooms overnight. On the third and final day, birds were returned to their home cages immediately after testing was completed.

The one profitable feeder and its cue were reassigned randomly among the six feeders each feeding trial, but the same random sequence was used for all birds. Feeding trials began 2 minutes after preceding trials ended, or as soon as preceding trials ended and the birds perched. At the beginning of each feeding trial the following events occurred simultaneously: a soft buzzer sounded for 0.5 seconds, the cue to the profitable feeder was lit, and the profitable



feeder was set to provide 2  $\mu$ l of solution on each visit. Birds could visit any sequence of feeders and obtain food from the correct feeder up to 11 times, for a total of 22  $\mu$ l each feeding trial. This quantity provided food at a rate slightly more than average *ad libitum* feeding rates.

No limit was imposed on the time for birds to respond in feeding trials. Feeding trials ended only after birds had probed the correct feeder 11 times, or after they had visited at least one feeder and returned to their perch. When a feeding trial ended, the light cue was turned off and the previously profitable feeder delivered no more food. All feeders remained exposed between trials.

### 1.3 Description of the Data

Learning and performance of the hummingbirds was to be evaluated from *first* visits in feeding trials. If the first visit was to a profitable feeder, the feeding trial was scored as a success (coded as 1); if the first visit was to any one of the five unprofitable feeders, the feeding trial was scored as a failure (coded as 0). The *binary* data set is comprised of 18 binary series of length 180, corresponding to the successes and failures of each bird on the *first* feeding foray in a trial, for 3 days of 60 consecutive feeding trials.

In addition to the information in these binary series on the first feeding foray of each trial, information on both the number of successful feeding forays and the total number of feeding forays in each trial is available. What we will refer to as the *binomial* data set consists of the total number of feeding forays (denoted by  $n$ ) and the number of *successful* feeding forays (denoted by  $y$ ) in each trial. Note that the number of successful forays in each trial is not really a binomial random variable because:

- the feeding forays within a trial are not independent, and
- even if the forays within a trial were independent, for those trials in which the bird feeds successfully the maximum number of 11 times (roughly 60% of the 3240 trials), the total number of feeding forays made during the trial (denoted by  $n$ ), rather than the number of successful feeding forays made during the trial (denoted by  $y$ ), would be the random quantity (a negative binomial random variable) because of the way the experiment has been set up.

#### 1.4 Outline of the Thesis

This thesis discusses several approaches to the analysis of binary longitudinal data, the last of which is also applicable to count (and continuous) responses.

The first approach discussed is univariate analysis-of-variance for repeated measures using the arcsine square-root variance stabilizing transformation for proportions; this was the approach taken by the investigator in her analysis of the binary hummingbird data. To use this approach, the data for each bird must be blocked over trials before the elements in the resulting response vector are transformed. For example, the response vector could consist of 3 elements, the arcsine square-roots of the proportions of successful trials on each of the 3 days in the experiment. The arcsine square-root transformation may not be entirely effective if the response is a vector of proportions because, although the variances of the proportions are stabilized, covariances remain functions of the success probabilities. Because the investigator expects success probabilities to differ across treatment groups, the covariance matrices are also expected to be different. In this context, univariate analysis-of-variance for repeated measures is of little

utility because it requires a common covariance matrix for valid statistical tests. Even if we assume a common covariance matrix is shared by all birds, conservative statistical tests are required (with this data) for checking the effects of measurement and measurement-by-treatment interaction in the analysis-of-variance. Moreover, the sex of each bird is not easily incorporated into the analysis because of the unequal number of males and females in the experiment. An alternate technique for analysis of binary longitudinal data which accounts for correlation over time within subjects and permits flexible modelling of the response would be preferable to this more classical approach.

The “working likelihood” approach of Zeger, Liang, and Self (1985), described in Chapter 3, is just such a technique. Here, a working likelihood (not necessarily the true likelihood) is used to generate estimating equations for regression parameters which model the marginal probability of success. Sections 3.1 and 3.2 outline the approach of Zeger, Liang, and Self (1985), while Section 3.3 extends to this approach to time-dependent covariates. This extension permits success probabilities on the first feeding foray to be modelled as functions of time, while taking into account correlation between measurements on each bird. Hence, comparisons of the learning trajectories of hummingbirds across treatment groups are more direct than comparisons achieved by the classical analysis. The remainder of Chapter 3 presents an analysis of the binary hummingbird data using this extension. Note that with only 18 birds in the experiment, efficient estimation of the regression parameters is an important issue. One weakness of this approach is that only one working likelihood is available for explicitly modelling correlation in the data (although, the results could presumably be extended to include working likelihoods modelling correlation in a variety of ways). Therefore, specification of a working likelihood

“close” to the true likelihood (for reasons of efficiency) may be difficult. Further, when modelling correlation explicitly, asymptotic results only hold under direct assumptions on the true correlation structure.

Chapter 4 presents the “generalized estimating equation” (GEE) approach of Zeger and Liang (1986) and Liang and Zeger (1986). The GEE approach extends the use of quasi-likelihood to longitudinal data and may be used for any multivariate response with univariate marginal distributions for which the quasi-likelihood formulation is sensible. The GEE approach is more flexible than the working likelihood approach not only because the marginal likelihoods need not be specified, but also because many possible “working correlations” are available for explicitly modelling correlation in the responses. Moreover, no assumptions about the true correlation structure are necessary for valid asymptotic results with any of these working correlations. To illustrate the flexibility of the GEE approach, both the binary and binomial hummingbird data are analyzed using different working correlation structures.

In Chapter 5, the thesis concludes with a summary of the important conclusions made resulting from the analyses and a comparison of the methods discussed.

## Chapter 2

### Classical Analysis

A standard approach to the analysis of binary longitudinal data such as the hummingbird data (for the first feeding foray) might be to calculate the proportion of correct responses over blocks of an equal number of successive time points for each subject, and then “normalize” these proportions via the arcsine square-root transformation. The resulting subject response vectors within the  $i^{th}$  treatment group might then be assumed to be approximately normally distributed with mean vector  $\mu_i$  and covariance matrix  $\Sigma_i$ . Moreover, if there is a physical basis for doing so, or if there is insufficient evidence in the data to conclude otherwise, the covariance matrices may be assumed to be equal, a sufficient assumption for the use of multivariate analysis-of-variance (MANOVA) of the response vectors.

Before using this approach, the equality of the covariance matrices should be tested. Assuming the covariance matrices of birds within each treatment group are the same, the estimated covariance matrix common to each treatment group will be non-singular only if  $q$ , the number of elements in each subject’s response vector, is fewer than one less the number of subjects in the treatment group. The hummingbird experiment has only 6 birds per treatment group, so  $q$  must be 4 or less for these estimated covariance matrices to be non-singular. The assumption of equal covariance matrices cannot be formally checked if  $q \geq 5$  because the test requires the estimated covariance matrices to be invertible. To check the equality of covariance matrices,

subject data must therefore be blocked over at least 45 trials (leading to  $q = 4$  measurements per subject). Aggregating the data of each bird over 45 trials is awkward, though, because there are 60 trials per day and day effects could be important in describing the performance of the birds.

Instead, individual data might be blocked over 60 trials (one day), allowing MANOVA with  $q = 3$  “measurements” on each bird. But with only 3 measurements (corresponding to different days) on each bird, such an analysis is likely to be too crude to capture the true nature of the temporal learning pattern on each treatment. The aggressive blocking necessary to carry out MANOVA with only 3 measurements per subject does not permit detailed inference on how success probabilities within the treatment groups change as a function of time. An alternative method of analysis able to provide higher resolution of the temporal learning patterns would be ideal.

In the MANOVA context, a step towards this ideal would be to block over fewer trials. If data were blocked over 30 trials instead of 60, there would be 6 measurements on each subject, and the previous confounding of blocks and days would be eliminated, allowing for a more detailed representation of temporal learning patterns. But the resulting MANOVA tests would not be valid unless all treatment groups shared a common covariance matrix; that is,  $\Sigma_i = \Sigma$ ,  $i = 1, 2, 3$ . This assumption of common covariance matrices across treatments is inconsistent with the investigators beliefs and contradicted by the data.<sup>1</sup> Moreover, if the

---

<sup>1</sup>The data for  $q = 3$  measurements per bird (where each measurement is the arcsine square-root of a raw success proportion) indicate that covariance matrices for the transformed success proportions are not equal across treatments. Note that if these covariance matrices are not equal when there are  $q = 3$  measurements per bird, it is unlikely that they will be equal when less aggressive blocking results in  $q > 3$  such measurements. Covariance matrices can be estimated within treatment groups when there are 3 or fewer measurements per bird, but the distribution of the test statistic for their equality is only chi-squared as the number of subjects in each treatment group (6 for this experiment) becomes large. Moreover, the test is very sensitive to non-normality

success probabilities for birds differ across treatment groups, the covariance matrices for the transformed response must differ across treatment groups as well because, even though the arcsine square-root transformation can be expected to stabilize the variances, the covariances will still be functions of the success probabilities. Although this analysis would provide a more detailed picture of temporal learning patterns, its conclusions would be somewhat tenuous because the assumption of a common covariance matrix is supported by neither the investigator's beliefs, the data, nor the nature of the variance-stabilizing transformation.

Even though assuming a common covariance matrix may not be correct, it allows blocking over as few as 15 trials. The resulting 12 measures per bird permit representations of temporal response patterns across treatments that are much more detailed than the representations obtained when the equality of covariance matrices can be tested (and we are limited to 3 or fewer measurements per bird). Another benefit of assuming a common covariance matrix is that univariate analysis-of-variance for repeated measures, a simpler method more accessible to researchers familiar with experimental design, may be used as an alternative to MANOVA.

The details of such analyses are presented in the remainder of this chapter.

---

in the data (see Morrison 1976, page 252). Hence, both the test results and the estimated covariance matrices should be examined to judge whether underlying covariance matrices appear to be approximately equal.

The estimated covariance matrices (common to birds within treatment groups) for the case of  $q = 3$  measurements per bird (where the measurements are transformed success proportions based on blocks of 60 successive trials) are,

$$.01 \begin{bmatrix} 2.5 & 3.7 & 2.7 \\ 3.7 & 5.9 & 4.2 \\ 2.7 & 4.2 & 8.1 \end{bmatrix}, \quad .01 \begin{bmatrix} 2.2 & 1.1 & -0.1 \\ 1.1 & 2.1 & 0.3 \\ -0.1 & 0.3 & 3.0 \end{bmatrix}, \quad .01 \begin{bmatrix} 5.5 & 3.7 & 1.1 \\ 3.7 & 9.8 & 7.4 \\ 1.1 & 7.4 & 11.0 \end{bmatrix},$$

for treatments 1, 2, and 3 respectively. These appear quite different, and the same conclusion is suggested by the chi-squared test; the value of the test statistic is 44.1 on 12 degrees of freedom ( $p < .001$ , based on the asymptotic chi-squared distribution). It thus seems unlikely that the covariance matrices are equal when there are  $q > 3$  measurements per subject.

## 2.1 Univariate Analysis-of-Variance for Repeated Measures

Univariate analysis-of-variance for repeated measures is strictly valid only if observations are multinormally distributed with a common covariance structure corresponding to the “type-H” pattern described by Huynh and Feldt (1970); i.e.,  $\Sigma = (\sigma_{k,l})$ , has the pattern defined by

$$\sigma_{k,l} = \alpha_k + \alpha_l + \lambda \quad k = l.$$

$$\sigma_{k,l} = \alpha_k + \alpha_l \quad k \neq l.$$

If the covariance structure of the observations does not correspond to the type-H pattern, Greenhouse and Geisser (1959) have proposed conservative tests based on adjusting the degrees of freedom for the univariate tests. Even though the univariate approach may sometimes require use of these conservative tests, it still has a strong appeal because it is easy to interpret and commonly used by those familiar with experimental design. Further, unlike multivariate analysis-of-variance, test statistics are still defined if the number of repeated measurements per subject is greater than the within treatment degrees of freedom. The binary hummingbird data may thus be blocked over even fewer than 15 (say 10, for example) trials to perform the univariate analysis, yielding representations of temporal response patterns that are more detailed than those permitted by MANOVA. However, if the binary hummingbird data is blocked over fewer than 15 trials, it will not be possible to check for the type-H pattern in the underlying covariance matrix because the estimated covariance matrix will once again be singular. Even so, an analysis using conservative univariate tests may still be carried out.

Because the investigator wishes to understand how success probabilities for birds in different treatment groups compare over time, detailed representations of temporal response patterns are



desirable. Since these are permitted by univariate analysis-of-variance for repeated measures and since this method of analysis is well understood and commonly used by researchers familiar with experimental design, this technique will be used to analyze the binary hummingbird data (on the first feeding foray), even though a key assumption for this technique (i.e., that a common covariance matrix for the transformed response vectors be shared by all subjects) is not supported by either the beliefs of the investigator, the data, or the nature of the arcsine square-root variance stabilizing transformation for binary data.

## 2.2 The Statistical Model

In order to “normalize” the hummingbird data, the proportion of correct responses for blocks of fifteen and then ten successive trials was calculated for each bird and transformed via the arcsine square-root transformation. The resulting 12- and 18-dimensional response vectors for each bird in the  $i^{th}$  treatment group were then assumed to be approximately normally distributed with mean  $\mu_i$  and common covariance matrix  $\Sigma$ .

The treatment means of these transformed proportions are shown in Figures 2.1 and 2.2. These figures suggest that the (transformed) response grows roughly linearly with time, particularly for birds in treatments 2 (PT) and 3 (FT). The lack of systematic discontinuities in these response curves at measurements corresponding to new days (measurements 7 and 13 in Figure 2.1, and measurements 5 and 9 in Figure 2.2) suggests that modelling a “day” effect is not necessary. Thus, for both blocks of 10 and blocks of 15 observations, a univariate analysis-of-variance for repeated measures was carried out under the following model:

$$Y_{ijk} = \arcsine(\sqrt{\hat{P}_{ijk}}) = \mu + \text{treatment}_i + \text{measurement}_j + (\text{treatment} * \text{measurement})_{ij} \\ + \text{subject}_{(i)k} + (\text{measurement} * \text{subject})_{(i)jk} + \epsilon_{ijk},$$

for  $i = 1, 2, 3$ ;  $j = 1, \dots, 12$  or  $18$  (for blocks of 15 and 10 observations respectively);  $k = 1, \dots, 6$ .

Here  $\hat{P}_{ijk}$  is the raw success proportion for the  $k^{\text{th}}$  subject within the  $i^{\text{th}}$  treatment at the  $j^{\text{th}}$  measurement, and  $\epsilon_{ijk}$  is the noise term for the  $k^{\text{th}}$  subject within the  $i^{\text{th}}$  treatment at the  $j^{\text{th}}$  measurement. We assume that  $(\epsilon_{ijk})_{j=1, \dots, 12 \text{ or } 18}$  ( $i = 1, 2, 3$ ;  $k = 1, \dots, 6$ ), the subject noise vectors, are independent normally distributed random vectors with mean 0 and common covariance matrix  $\Sigma$ . “Treatment” and “measurement” are fixed effect factors, and “subject” is a random effect factor nested within “treatment”.

### 2.3 Results of the Analysis

Before carrying out univariate analysis-of-variance on the binary hummingbird data (for the first feeding foray) blocked over 15 trials, the underlying covariance matrix was tested for the type-H pattern. Because the value of the appropriate chi-squared statistic (see Morrison 1976, page 251) is 259.17 with 65 degrees of freedom, the covariance matrix does not seem to have the type-H pattern. The estimated covariance matrix for the binary hummingbird data blocked over 10 trials is singular because there are 18 measurements per bird and only 18 birds in the study, so the underlying covariance matrix cannot be tested for the type-H pattern. However, it seems unlikely that the covariance matrix for blocks of 10 trials will have the type-H pattern since the covariance matrix for blocks of 15 trials does not. Univariate analysis-of-variance was

therefore carried out using the conservative tests of Greenhouse and Geisser (1959) for both blocks of 15 and blocks of 10 observations.

For the univariate analysis-of-variance tests to be strictly valid, the transformed response vectors for the individual subjects must be normally distributed with a common covariance matrix  $\Sigma$ ; that is,  $(Y_{ijk})_{j=1,\dots,12 \text{ or } 18} \sim N(\mu_i, \Sigma)$ . The investigator does not believe that a common covariance matrix is shared by all birds, but separate covariance matrices common to birds within treatment groups cannot be modelled due to the small number of birds within each treatment group. Because these estimated covariance matrices,  $\widehat{\Sigma}_i$  ( $i = 1, 2, 3$ ), are singular, a formal test for the equality of the underlying covariance matrices is not available. However, elements of the estimated covariance matrices may still be compared visually. Figures 2.3 and 2.4 plot the estimated variances of the transformed responses (the diagonal elements of  $\widehat{\Sigma}_i$ ) against measurement by treatment group for the 12- and 18-measurement analyses, respectively. Neither figure suggests an obvious pattern in the treatment variances over time, but both figures suggest that the variances for birds in treatment 3 are generally greater than those for birds in other treatments, while the variances for birds in treatment 2 are generally lower than those for birds in other treatments (the estimated variances of responses for birds in treatment 3 are anywhere from 1-11 times larger than the estimated variances for birds in treatment 2 for the 12-measurement analysis, and anywhere from 0.4-22 times larger for the 18-measurement analysis). These results suggest that the assumption of a common covariance matrix is not valid for either the 12- or the 18-measurement analysis. Hence, the distributions of test statistics formulated assuming equal covariance matrices can only be expected to roughly approximate the true distributions of the test statistics. In spite of this, the tests should still provide

useful information about the relative magnitude of the components of variability, and strong conclusions from the analyses-of-variance should be consistent with conclusions from analyses which permit unequal covariance matrices.

The analyses-of-variance are given in Tables 2.1 (18-measurement analysis) and 2.2 (12-measurement analysis). Tables 2.1 and 2.2 confirm one clear indication from Figures 2.1 and 2.2: “measurement” is an important factor. But Tables 2.1 and 2.2 give no strong evidence of “treatment” effects, either through main effects or through interactions with “measurement”. On the other hand, Figures 2.1 and 2.2 clearly indicate that the (transformed) responses for treatment 3 are generally greater than those in the other treatments, suggesting that “treatment” main effects do, in fact, exist but are not being detected because there is too much variability between subjects within treatment groups.

Table 2.1. Analysis of Variance for Blocks of 10

Source	SS	d.f.	MS	F	conservative d.f.	p
measurement	14.74	17	.87	14.20	1,15	.002
treatment	.57	2	.29	.75		.49
subject (within treatment)	5.67	15	.38			
measurement*treatment	1.26	34	.037	.62	2,15	.55
subject*measurement (within treatment)	15.33	255	.060			
total	37.57	323				

Table 2.2. Analysis of Variance for Blocks of 15

Source	SS	d.f.	MS	F	conservative d.f.	p
measurement	3.56	11	.32	8.68	1,15	.01
treatment	1.70	2	.85	1.76		.21
subject (within treatment)	7.22	15	.48			
measurement*treatment	.958	22	.044	1.17	2,15	.34
subject*measurement (within treatment)	6.15	165	.037			
total	19.59	215				

The conservative F-tests used in these univariate analyses-of-variance are based on approximate distributions of the sums of squares when the subject response vectors are multivariate

normal with a common covariance matrix  $\Sigma$ . Greenhouse and Geisser (1959) have shown that when the hypothesis of no measurement effects is true, the F-statistic for measurement has an approximate F-distribution with degrees of freedom  $(K-1)\epsilon$  and  $(K-1)(N-K)\epsilon$ , where  $K$  denotes the number of measurements per subject, and  $N$  denotes the number of subjects. The scale factor

$$\epsilon = \frac{K^2(\overline{\sigma_d} - \overline{\sigma_{..}})^2}{(K-1)(\sum_i \sum_j \sigma_{ij}^2 - 2K \sum_i \overline{\sigma_{i.}}^2 + K^2 \overline{\sigma_{..}}^2)}$$

is computed from the elements of the (common) covariance matrix  $\Sigma$ , where  $\overline{\sigma_d}$  is the mean of the  $p$  diagonal terms,  $\overline{\sigma_{..}}$  is the grand mean of all variances and covariances, and  $\overline{\sigma_{i.}}$  is the mean of the elements in the  $i^{th}$  row of  $\Sigma$ . They have also shown that when the hypothesis of no measurement-by-treatment interaction is true, the F-statistic for this interaction has an approximate F-distribution with degrees freedom  $(K-1)(p-1)\epsilon$  and  $(K-1)(N-p)\epsilon$ , where  $p$  denotes the number of treatments.

In practice,  $\Sigma$  is never known and estimating  $\epsilon$  from the sample covariance matrix is undesirable because more uncertainty is then introduced into the approximate analysis-of-variance. Greenhouse and Geisser have shown that  $\epsilon > 1/(K-1)$ , implying that the degrees freedom for the approximate tests cannot be less than 1 and  $(N-K)$ , and  $(K-1)$  and  $(N-K)$ , respectively. A conservative F-test results from assuming the maximum possible reduction in degrees freedom, and therefore requires larger values of the F-statistic to reject the null hypothesis (i.e., the hypothesis of no measurement effects, or the hypothesis of no treatment-by-measurement interactions) than tests which assume less than the maximum reduction in degrees freedom. These conservative F-tests are used in both the 12- and 18-measurement analyses described

above, but the conclusions from these analyses are the same as the conclusions from the analyses which estimate the scale factor  $\epsilon$  from the data ( $\hat{\epsilon}$  is .33 for the 12-measurement analysis, and .24 for the 18-measurement analysis, so the use of  $\hat{\epsilon}$  results in, respectively, degrees of freedom 3.63 and 4.08 times the conservative degrees of freedom).

## 2.4 Conclusions and Limitations

In both univariate analyses, the *conservative* test for the measurement main effects finds “measurement” to be an important factor. Figures 2.1 and 2.2 also indicate that “measurement” is an important factor. The test (not conservative) for treatment main effects in both univariate analyses-of-variance does not find “treatment” to be important, nor does the *conservative* test for the interactions of treatment with measurement find “treatment\*measurement” to be important. Thus, both analyses-of-variance indicate that treatment group has no effect either directly, or through interaction with measurement, on the (transformed) response pattern. On the other hand, plots of the fitted values (Figures 2.1 and 2.2) clearly indicate that the (transformed) responses for treatment 3 are greater than those in the other treatments, suggesting that a treatment main effect is present but cannot be detected by these methods because variability between subjects within treatments is too large. The same result would therefore be expected in the corresponding MANOVA. Some of the within-treatment variability between subjects may be explained by their sex, but modelling of a sex effect or any interactions of sex with the other factors is difficult in both univariate and multivariate analysis-of-variance because of the unequal number of males and females in treatment 1 (NT). An approach allowing modelling of sex, despite the imbalanced number of males and females across treatment groups,

might reduce the unexplained within-treatment variability, and thereby provide more sensitive tests of all effects.

Note that many other “classical” approaches to analyzing the binary hummingbird data (for the first feeding foray) have not been discussed in this section. These include parametric modelling of the “measurement” effects (for example, analysis-of-covariance, which requires that observations within each subject be independent), and analysis-of-variance (either univariate or multivariate) in the logit scale ( $\text{logit}(p) = \log(\frac{p}{1-p})$ ,  $p$  between 0 and 1) rather than in the arcsine square-root scale to correspond to logistic regression.

Although univariate and multivariate analysis-of-variance are standard techniques used to analyze data in which repeated measurements are taken for each subject, both seem to model how the data change over time in indirect ways which lack intuitive appeal. To allow changes in response over time, univariate analysis-of-variance includes “measurement” main effects and interactions in its model. Multivariate analysis-of-variance allows changes in response over time through the different elements of its treatment mean response vectors. More intuitive approaches, presented in the next sections, model treatment response patterns directly as a function of time, *while taking into account the correlation of measurements within subjects*. Treatment response patterns can then be compared visually by examining the fitted treatment response models.

## Chapter 3

### Analysis Based on a Working Likelihood Approach

#### 3.1 Logistic Regression for Binary Repeated Measures

A natural approach to the analysis of the binary hummingbird data is to extend logistic regression to the case where the binary outcome variable is observed repeatedly for each subject. Zeger, Liang, and Self (1985) present the details of one way of implementing such an approach in which, rather than the actual likelihood, a “working likelihood” based on simple assumptions about the time dependence within each subject’s data is used to generate estimating equations which lead to consistent estimates of the regression parameters, as well as standard errors for these estimates. Not only does this approach allow for correlation over time within subjects, but it allows flexible modelling of the response and is powerful enough (provided the number of subjects is large) to detect important factors and covariates, while remaining simple and easy to interpret.

#### 3.2 The Working Likelihood Approach

To establish notation, let  $Y_{i,t}$  ( $t = 1, \dots, n_i$ ) be a stationary binary time series, and let  $Z_i$  be an  $s \times 1$  vector of time-independent covariates for the  $i^{th}$  subject ( $i = 1, \dots, K$ ). For

$\pi_i \stackrel{\text{def}}{=} Pr(Y_{i,t} = 1 \mid Z_i)$ , it is assumed throughout this section that  $\text{logit } \pi_i = Z_i' \beta$ . The



primary objective is to make inference about the vector of parameters,  $\beta$ . There may be additional nuisance parameters in the working likelihood, so let  $\theta$  denote all the parameters modelled by the working likelihood. Let  $S(\theta)$  be the working likelihood score function, and let  $H(\theta)$  be the corresponding Hessian matrix. Finally, let  $\hat{\theta}$  be the estimate of  $\theta$  obtained by maximizing the *working* likelihood; in general, we have  $S(\hat{\theta}) = 0$ . Throughout, score functions and Hessian matrices will be evaluated under the working likelihood,  $E(\cdot)$  will denote the expectation under the true (unknown) likelihood and  $E_W(\cdot)$  will denote the expectation under the working likelihood.

Motivation for the results of Zeger, Liang, and Self (1985) is provided by the Taylor expansion for  $S(\theta)$  about  $\hat{\theta}$ , and the Mean Value Theorem which, together with the fact that  $S(\hat{\theta}) = 0$ , allow us to write

$$S(\theta) = S(\hat{\theta}) + H(\theta_*)(\theta - \hat{\theta}) = -H(\theta_*)(\hat{\theta} - \theta),$$

where  $\theta_*$  is between  $\theta$  and  $\hat{\theta}$ .

If the data for the different subjects is independent, then the working log-likelihood is the sum of the independent log-likelihoods for the different subjects. Thus, the score function and the Hessian are also sums of independent random quantities, and we can write,

$$\hat{\theta} - \theta = -H(\theta_*)^{-1} S(\theta) = -\left[\frac{1}{K} \sum_{i=1}^K H_i(\theta_*)\right]^{-1} \frac{1}{K} \sum_{i=1}^K S_i(\theta),$$

where  $S_i$  and  $H_i$  are the  $i^{th}$  subject's score function and Hessian, respectively.

If  $E S_i(\theta) \equiv 0$ , then provided regularity conditions for Lindeberg's Central Limit Theorem hold, under the true model,

$$\frac{1}{K} \sum_{i=1}^K S_i(\theta) = \frac{1}{K} S(\theta) \stackrel{d}{\approx} N\left(0, \frac{1}{K^2} I^*\right),$$

where  $I^*(\theta) = E S(\theta)S'(\theta)$ .

Under regularity conditions ensuring  $\hat{\theta}$  is consistent for  $\theta$ , we have

$$\frac{1}{K} \sum_{i=1}^K H_i(\theta_*) \xrightarrow{p} \lim_{K \rightarrow \infty} \frac{1}{K} E \sum_{i=1}^K H_i(\theta) = \lim_{K \rightarrow \infty} \frac{1}{K} E H(\theta).$$

Thus, by Slutsky's Theorem,

$$\hat{\theta} - \theta \stackrel{d}{\approx} N(0, [E H(\theta)]^{-1} I^*(\theta) [E H(\theta)]^{-1}),$$

where all expectations are taken under the true (unknown) likelihood, and convergence is under the true (unknown) distribution.

Note that to use this result for inference about  $\theta$ , both  $I^*(\theta)$ , the variance-covariance matrix for the score function under the true (unknown) model, and  $E H(\theta)$ , the expectation of the Hessian under the true (unknown) model, must be estimated from the data. Provided  $\hat{\theta}$  is consistent for  $\theta$ ,  $I^*(\theta) = E S(\theta)S'(\theta)$  can be consistently estimated by  $\sum_{i=1}^K S_i(\hat{\theta})S'_i(\hat{\theta})$ . The expected Hessian under the working model is identical to the expected Hessian under the true (unknown) model provided simple assumptions about the time dependence within each subject's data are valid. Therefore, if  $\hat{\theta}$  is consistent for  $\theta$ ,  $E_W H(\hat{\theta})$  is a consistent estimate of  $E H(\theta)$ . Because  $H(\theta)$  is a sum of independent random quantities, if  $\hat{\theta}$  is consistent for  $\theta$  then an alternative consistent estimate of  $E H(\theta)$  is  $H(\hat{\theta})$ , the observed Hessian evaluated at  $\hat{\theta}$ . The extension of the Zeger, Liang, and Self approach to the case of time-dependent covariates, presented in the next section, uses the observed Hessian evaluated at  $\hat{\theta}$ , rather than  $E_W H(\hat{\theta})$ , to (consistently) estimate  $E H(\theta)$ .

Zeger, Liang, and Self (1985) present two estimators of  $\beta$ . The first,  $\hat{\beta}_0$ , is obtained from the working assumption that repeated observations for a subject are independent of one another.

This estimator is consistent (as  $K \rightarrow \infty$ ) for any set of stationary binary time series for which  $\text{logit } \pi_i = Z_i' \beta$ . The second estimator,  $\hat{\beta}_1$ , is obtained from the weaker working assumption that each time series has a stationary 1-dependent correlation structure.  $\hat{\beta}_1$  is consistent for any set of stationary binary time series for which both  $\text{logit } \pi_i = Z_i' \beta$  and the first lag autocorrelation,  $\rho = \text{corr}(Y_{i,t}, Y_{i,t-1})$ , is common to all subjects.

For the independence working model with  $\text{logit } \pi_i = Z_i' \beta$ , the working log-likelihood can be written as

$$l_0(\beta) = \sum_{i=1}^K \sum_{t=1}^{n_i} [y_{i,t} Z_i' \beta + \log(1 - \pi_i)],$$

where  $\text{logit } \pi_i = Z_i' \beta$ . Let  $S(\beta)$  be the score function for the vector  $\beta$ , and let  $S_u(\beta)$  be its  $u^{\text{th}}$  element ( $u = 1, \dots, s$ ). The maximum likelihood equations under the working model are

$$0 = S_u(\beta) = \sum_{i=1}^K \sum_{t=1}^{n_i} (y_{i,t} - \pi_i) Z_{iu}, \quad u = 1, \dots, s.$$

Define  $\hat{\beta}_0$  to be the solution to these estimating equations.

Note that the expected values of the scores under the true (unknown) model are 0 provided the mean structure has been correctly specified; that is,  $E(Y_{i,t} | Z_i) = [1 + \exp(-Z_i' \beta)]^{-1}$ , or  $\text{logit } E(Y_{i,t} | Z_i) = Z_i' \beta$ . This ensures that  $\hat{\beta}_0$  is an asymptotically unbiased estimate of  $\beta$ .

The  $(u, v)^{\text{th}}$  element of the Hessian matrix is given by

$$H_{uv}(\beta) = - \sum_{i=1}^K \sum_{t=1}^{n_i} Z_{iu} Z_{iv} \pi_i (1 - \pi_i),$$

which does not depend upon the  $Y_{i,t}$ 's. Therefore, the expected Hessian is the same under both the true likelihood and the working likelihood; in other words, the working likelihood Fisher information matrix is the same as  $-E H(\beta)$  in this case. As  $K \rightarrow \infty$ , we obtain the result that

$\hat{\beta}_0$  is consistent and asymptotically normal for any set of stationary binary time series such that  $\text{logit } \pi_i = Z_i' \beta$ , as stated by Zeger, Liang, and Self (1985) in their:

**PROPOSITION 2.1.** *Let  $Y_{i,t}$  ( $t = 1, \dots, n_i < \infty$ ;  $i = 1, \dots, K$ ) be stationary binary series such that  $\text{logit } E(Y_{i,t} | Z_i) = Z_i' \beta$ . Then  $\hat{\beta}_0 - \beta$  is asymptotically multivariate Gaussian with expectation 0 and covariance matrix  $V_0(\beta) = I_0(\beta)^{-1} I_0^*(\beta) I_0(\beta)^{-1}$ , where  $I_0(\beta)$  is the working Fisher information matrix and  $I_0^*(\beta) = E S(\beta) S'(\beta)$ .*

A consistent estimate of  $I_0(\beta)$ , the Fisher information matrix under the working model, is given by

$$\hat{I}_0 = I_0(\hat{\beta}_0).$$

A consistent estimate of  $I_0^*(\beta)$ , the variance-covariance matrix of the score function, is provided by

$$\hat{I}_0^* = \sum_{i=1}^K S_i(\hat{\beta}_0) S_i'(\hat{\beta}_0),$$

where  $S_i(\cdot)$  is the score function for the  $i^{\text{th}}$  subject. Combining these results leads to a consistent estimate of  $V_0(\beta)$  given by

$$\hat{V}_0 = \hat{I}_0^{-1} \hat{I}_0^* \hat{I}_0^{-1}.$$

The estimator  $\hat{\beta}_1$  is obtained similarly, after replacing the working assumption of independence by the working assumption of a common lag one (and no other) autocorrelation. Let  $\rho = \text{corr}(Y_{i,t}, Y_{i,t-1})$  and  $\theta = (\rho, \beta)$ . If  $p_{it} \stackrel{\text{def}}{=} E(Y_{i,t} | Y_{i,t-1}, Z_i)$ , the true lag one conditional mean, the assumption of a common first-lag autocorrelation implies that

$p_{it} = \pi_i + \rho(Y_{i,t-1} - \pi_i) = E_W(Y_{i,t} | Y_{i,t-1}, Z_i)$ . The working log-likelihood can thus be written

$$l_1(\theta) = \sum_{i=1}^K \left[ y_{i,1} Z_i' \beta + \log(1 - \pi_i) + \sum_{t=2}^{n_i} [y_{i,t} \log p_{it} + (1 - y_{i,t}) \log(1 - p_{it})] \right],$$

where  $\text{logit } \pi_i = Z_i' \beta$  as before, and  $p_{it} = \pi_i + \rho(y_{i,t-1} - \pi_i)$ .

For this working model, let  $S_0(\theta)$  be the score function for  $\rho$ , and let  $S_u(\theta)$  be the score function for  $\beta_u$  ( $u = 1, \dots, s$ ). The maximum likelihood equations under the working model are

$$0 = S_0(\theta) = \frac{\partial l_1}{\partial \rho} = \sum_{i=1}^K \sum_{t=2}^{n_i} \frac{(y_{i,t} - p_{it})(y_{i,t-1} - \pi_i)}{p_{it} \bar{p}_{it}},$$

$$0 = S_u(\theta) = \frac{\partial l_1}{\partial \beta_u} = \sum_{i=1}^K \left\{ Z_{iu}(y_{i,1} - \pi_i) + (1 - \rho) \sum_{t=2}^{n_i} \frac{Z_{iu} \pi_i \bar{\pi}_i (y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \right\},$$

where  $\bar{p}_{it} = 1 - p_{it}$ , and  $\bar{\pi}_i = 1 - \pi_i$ . Define  $\hat{\theta} = (\hat{\rho}, \hat{\beta}_1)$  to be the solution to these estimating equations.

Note that the expected values of the scores under the true (unknown) model are 0 provided the lag one conditional mean structure has been correctly specified; that is,  $E(Y_{i,t} | Y_{i,t-1}, Z_i) = \pi_i + \rho(Y_{i,t-1} - \pi_i)$ . This can be seen by conditioning on  $Y_{i,t-1}$  for each  $Y_{i,t}$  in the estimating equations. Thus,  $\widehat{\beta}_1$  is an asymptotically unbiased estimate of  $\beta$ .

The elements of the Hessian matrix are given by:

$$\frac{\partial^2 l_1}{\partial \rho^2} = - \sum_{i=1}^K \sum_{t=1}^{n_i} \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{it} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \left( \frac{\partial p_{it}}{\partial \rho} \right)^2,$$

$$\frac{\partial^2 l_1}{\partial \beta_u \partial \beta_v} = \sum_{i=1}^K -\pi_i \bar{\pi}_i Z_{iu} Z_{iv}$$

$$+ \sum_{t=2}^{n_i} \left[ \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \left( \frac{\partial^2 p_{it}}{\partial \beta_u \partial \beta_v} \right) - \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{it} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \left( \frac{\partial p_{it}}{\partial \beta_u} \right) \left( \frac{\partial p_{it}}{\partial \beta_v} \right) \right],$$

and

$$\frac{\partial^2 l_1}{\partial \rho \partial \beta_u} = \sum_{i=1}^K \sum_{t=1}^{n_i} \left\{ \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \left( \frac{\partial^2 p_{it}}{\partial \rho \partial \beta_u} \right) - \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{it} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \left( \frac{\partial p_{it}}{\partial \beta_u} \right) \left( \frac{\partial p_{it}}{\partial \rho} \right) \right\},$$

where

$$\begin{aligned}\frac{\partial p_{it}}{\partial \rho} &= y_{i,t-1} - \pi_i, \\ \frac{\partial p_{it}}{\partial \beta_u} &= (1 - \rho)\pi_i\bar{\pi}_i Z_{iu}, \\ \frac{\partial^2 p_{it}}{\partial \beta_u \partial \beta_v} &= (1 - \rho)(1 - 2\pi_i)\pi_i\bar{\pi}_i Z_{iu} Z_{iv},\end{aligned}$$

and,

$$\frac{\partial^2 p_{it}}{\partial \rho \partial \beta_u} = -\pi_i\bar{\pi}_i Z_{iu}.$$

The assumption that, under the *true* likelihood, a common lag one autocorrelation is shared by all subjects, that is, that the lag one conditional mean structure has been correctly specified, yields  $E(Y_{i,t} | Y_{i,t-1}, Z_i) = \pi_i + \rho(Y_{i,t-1} - \pi_i)$ , without any further assumptions concerning the higher lag autocorrelations. Under this assumption  $E(Y_{i,t} | Y_{i,t-1}, Z_i) = E_W(Y_{i,t} | Y_{i,t-1}, Z_i)$ , and it follows that  $E H(\theta) = E_W H(\theta)$ ; in other words, under the assumption that the lag one conditional mean structure has been correctly specified, the working likelihood Fisher information matrix and  $-E H(\theta)$  are identical.

The Fisher information matrix under the first-order working model,  $I_1(\theta)$ , can be partitioned as

$$I_1 = \begin{bmatrix} I_{\rho\rho} & I_{\rho\beta} \\ I_{\beta\rho} & I_{\beta\beta} \end{bmatrix},$$

where the  $(u, v)^{th}$  element of  $I_{\beta\beta}$  is

$$\sum_{i=1}^K Z_{iu} Z_{iv} \pi_i \bar{\pi}_i \{1 + (n_i - 1)(1 - \rho) \frac{\pi_i \bar{\pi}_i (1 - 3\rho) + \rho}{(1 - \rho)^2 \pi_i \bar{\pi}_i + \rho}\},$$

the  $u^{th}$  element of  $I_{\beta\rho}$  is

$$\rho \sum_{i=1}^K (n_i - 1) Z_{iu} \pi_i \bar{\pi}_i \frac{(2\pi_i - 1)}{(1 - \rho)^2 \pi_i \bar{\pi}_i + \rho},$$

and

$$I_{\rho\rho} = \frac{(1+\rho)}{(1-\rho)} \sum_{i=1}^K (n_i - 1) \left( \frac{\pi_i \bar{\pi}_i}{(1-\rho)^2 \pi_i \bar{\pi}_i + \rho} \right).$$

As  $K \rightarrow \infty$ , we obtain the result that  $\hat{\theta}$  is consistent and asymptotically normal for any set of stationary binary time series such that  $\text{logit } E(Y_{i,t} \mid Z_i) = Z_i' \beta$  and  $\text{corr}(Y_{i,t}, Y_{i,t-1}) = \rho$ , as stated by Zeger, Liang, and Self (1985) in their:

**PROPOSITION 2.2.** *Let  $Y_{i,t}$  ( $t = 1, \dots, n_i < \infty$ ;  $i = 1, \dots, K$ ) be stationary binary series such that  $\text{logit } E(Y_{i,t} \mid Z_i) = Z_i' \beta$  and  $\text{corr}(Y_{i,t}, Y_{i,t-1}) = \rho$ . Let  $\theta = (\rho, \beta)$ . Then  $\hat{\theta} - \theta$  is asymptotically multivariate Gaussian with expectation 0 and covariance matrix  $V_1(\theta) = I_1(\theta)^{-1} I_1^*(\theta) I_1(\theta)^{-1}$ , where  $I_1(\theta)$  is the Fisher information matrix under the first-order working model and  $I_1^*(\theta) = E S(\theta) S'(\theta)$ .*

A consistent estimate of  $I_1(\theta)$ , the Fisher information matrix under the first-order working model, is given by

$$\hat{I}_1 = I_1(\hat{\theta}).$$

A consistent estimate of  $I_1^*(\theta)$ , the variance-covariance matrix of the score function, is provided by

$$\hat{I}_1^* = \sum_{i=1}^K S_i(\hat{\theta}) S_i'(\hat{\theta}),$$

where  $S_i(\cdot)$  is the score function for the  $i^{\text{th}}$  subject. Combining these results leads to a consistent estimate of  $V_1(\theta)$  given by

$$\hat{V}_1 = \hat{I}_1^{-1} \hat{I}_1^* \hat{I}_1^{-1}.$$

A technical point which might arise during the Newton-Raphson iteration to obtain  $\hat{\theta}$  is that, depending on the values of the  $\pi_i$ 's, it may be necessary to constrain  $\rho$  to a smaller region

than  $(-1, 1)$  in order to prevent the  $p_{it} = E_W(Y_{i,t} | Y_{i,t-1}, Z_i) = \pi_i + \rho(Y_{i,t-1} - \pi_i)$  ( $i = 1, \dots, K$ ) from taking on values outside the range  $[0, 1]$ .

Since the covariates,  $Z_i$ , are time independent, the success probability  $\pi_i = Pr(Y_{i,t} = 1 | Z_i)$ , does not depend on time. Zeger, Liang, and Self (1985) were concerned with short binary time series, in which there is no need for time-dependent success probabilities. However, the binary time series for the first feeding forays of the hummingbirds are long (180 time points), and the success probabilities seem to increase with time. Figures 3.1, 3.2, and 3.3 plot the transformed average response for birds within a treatment group over consecutive blocks of 3, 5, and 10 time points respectively. Note that, except for the vertical scale, Figure 3.3, which plots the average response in the logit scale, and Figure 2.2, which plots the average response in the arcsine square root scale used for the classical analysis, are identical. All the plots clearly indicate that success probabilities generally increase with time. In the next section, the results of this section are extended to the case of time-dependent covariates to allow the modelling of success probabilities which change with time.

### 3.3 Incorporating Time-dependent Covariates

#### 3.3.1 The Independence Working Model

Suppose we start with the working assumption that repeated observations for a subject are independent, but now let  $Z_{i,t}$  be an  $s \times 1$  vector of time-dependent covariates for the  $i^{th}$  subject ( $i = 1, \dots, K$ ). If  $\pi_{i,t} \stackrel{\text{def}}{=} Pr(Y_{i,t} = 1 | Z_{i,t})$ , it is assumed throughout this section that



$\text{logit } \pi_{i,t} = Z'_{i,t}\beta$ . The estimator  $\hat{\beta}_0$  is obtained from the working log-likelihood

$$l_0(\beta) = \sum_{i=1}^K \sum_{t=1}^{n_i} \left[ y_{i,t} Z'_{i,t}\beta + \log(1 - \pi_{i,t}) \right],$$

where  $\text{logit } \pi_{i,t} = Z'_{i,t}\beta$ .

Let  $S_u(\beta)$  be the score function for  $\beta_u$  ( $u = 1, \dots, s$ ) under this working likelihood. Then the maximum likelihood equations under the working model are

$$0 = S_u(\beta) = \frac{\partial l_0}{\partial \beta_u} = \sum_{i=1}^K \sum_{t=1}^{n_i} (y_{i,t} - \pi_{i,t}) Z_{i,t,u}.$$

Define  $\hat{\beta}_0$  to be the solution to these estimating equations.

Note that the expected values of the scores under the true (unknown) model are 0 provided the mean structure has been correctly specified; that is,  $\pi_{i,t} \stackrel{\text{def}}{=} E(Y_{i,t} \mid Z_{i,t}) = [1 + \exp(-Z'_{i,t}\beta)]^{-1}$ , or  $\text{logit } E(Y_{i,t} \mid Z_{i,t}) = Z'_{i,t}\beta$ . Therefore,  $\hat{\beta}_0$  is asymptotically unbiased for  $\beta$ .

The  $(u, v)^{th}$  element of  $H_0(\beta)$ , the Hessian matrix for this independence working model, is given by

$$- \sum_{i=1}^K \sum_{t=1}^{n_i} Z_{i,t,u} Z_{i,t,v} \pi_{i,t} (1 - \pi_{i,t}),$$

which does not depend upon the  $Y_{i,t}$ 's. Therefore, the expected Hessian is the same under both the true likelihood and the working likelihood; in other words, the working likelihood Fisher information matrix is the same as  $-E H_0(\beta)$  in this case. Under regularity conditions sufficient to guarantee the consistency of  $\hat{\beta}_0$ , the following generalization of Zeger, Liang, and Self's (1985) Proposition 2.1 is immediate:

**PROPOSITION 1.** *Let  $Y_{i,t}$  ( $t = 1, \dots, n_i < \infty$ ;  $i = 1, \dots, K$ ) be binary series such that*

*$\text{logit } E(Y_{i,t} \mid Z_{i,t}) = Z'_{i,t}\beta$ . Then  $\hat{\beta}_0 - \beta$  is asymptotically multivariate Gaussian with expectation*

0 and covariance matrix

$$V_0(\beta) = [E H_0(\beta)]^{-1} I_0^*(\beta) [E H_0(\beta)]^{-1},$$

where  $I_0^*(\beta) = E S(\beta)S'(\beta)$ .

A consistent estimate of  $E H_0(\beta)$ , the expectation of the Hessian under the true (unknown) model, is given by  $H_0(\hat{\beta}_0)$ . A consistent estimate of  $I_0^*(\beta)$ , the variance-covariance matrix of the score function, is provided by

$$\widehat{I}_0^* = \sum_{i=1}^K S_i(\hat{\beta}_0)S_i'(\hat{\beta}_0),$$

where  $S_i(\cdot)$  is the score function for the  $i^{th}$  subject. Combining these results leads to a consistent estimate of  $V_0(\beta)$  given by

$$\hat{V}_0 = [H_0(\hat{\beta}_0)]^{-1} \widehat{I}_0^* [H_0(\hat{\beta}_0)]^{-1}.$$

### 3.3.2 The First-Order Working Model

The first-order working model with time-dependent covariates assumes each binary series has a stationary 1-dependent correlation structure with  $\text{logit } \pi_{i,t} = Z_{i,t}'\beta$  and

$\text{corr}(Y_{i,t}, Y_{i,t-1}) = \rho$ . If  $p_{it} \stackrel{\text{def}}{=} E(Y_{i,t} \mid Y_{i,t-1}, Z_i)$ , the assumption of a common lag one autocorrelation implies that

$$p_{it} = \pi_{i,t} + \rho \sqrt{\frac{\pi_{i,t} \overline{\pi_{i,t}}}{\pi_{i,t-1} \overline{\pi_{i,t-1}}}} (y_{i,t-1} - \pi_{i,t-1}),$$

where

$$\overline{\pi_{i,t}} = (1 - \pi_{i,t}).$$

Let  $\theta = (\rho, \beta)$ . Then the working log-likelihood is

$$l_1(\theta) = \sum_{i=1}^K \left[ y_{i,1} Z'_{i,1} \beta + \log(1 - \pi_{i,1}) + \sum_{t=2}^{n_i} [y_{i,t} \log p_{it} + (1 - y_{i,t}) \log(1 - p_{it})] \right],$$

where

$$p_{it} = \pi_{i,t} + \rho \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} (y_{i,t-1} - \pi_{i,t-1}).$$

For this working model, let  $S_0(\theta)$  be the score function for  $\rho$ , and let  $S_u(\theta)$  be the score function for  $\beta_u$ . The maximum likelihood equations under the working model are

$$0 = S_0(\theta) = \frac{\partial l_1}{\partial \rho} = \sum_{i=1}^K \sum_{t=2}^{n_i} \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \frac{\partial p_{it}}{\partial \rho},$$

and

$$0 = S_u(\theta) = \frac{\partial l_1}{\partial \beta_u} = \sum_{i=1}^K \left[ (y_{i,1} - \pi_{i,1}) Z_{i,1,u} + \sum_{t=2}^{n_i} \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \frac{\partial p_{it}}{\partial \beta_u} \right],$$

where  $\bar{p}_{it} = 1 - p_{it}$ . The partial derivatives appearing in these equations are given by:

$$\begin{aligned} \frac{\partial p_{it}}{\partial \beta_u} &= \sqrt{\pi_{i,t} \bar{\pi}_{i,t}} [\sqrt{\pi_{i,t} \bar{\pi}_{i,t}} Z_{i,t,u} - \rho \sqrt{\pi_{i,t-1} \bar{\pi}_{i,t-1}} Z_{i,t-1,u}] \\ &\quad + \frac{\rho}{2} \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} [y_{i,t-1} - \pi_{i,t-1}] [(1 - 2\pi_{i,t}) Z_{i,t,u} - (1 - 2\pi_{i,t-1}) Z_{i,t-1,u}], \end{aligned}$$

and

$$\frac{\partial p_{it}}{\partial \rho} = \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} [y_{i,t-1} - \pi_{i,t-1}].$$

Define  $\hat{\theta} = (\hat{\rho}, \hat{\beta}_1)$  to be the solution to these estimating equations.

Note that the expected values of the scores under the true (unknown) model are 0 provided the lag one conditional mean structure has been correctly specified; that is,  $E(Y_{i,t} | Y_{i,t-1}, Z_{i,t}) = \pi_{i,t} + \rho \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} (Y_{i,t-1} - \pi_{i,t-1})$ . This can be seen by conditioning on  $Y_{i,t-1}$  for each  $Y_{i,t}$  in the estimating equations. Therefore,  $\hat{\theta}$  is asymptotically unbiased for  $\theta$ .

The Hessian matrix for the first-order working likelihood,  $H_1(\theta)$ , can be partitioned as

$$H_1(\theta) = \begin{bmatrix} H_{\rho\rho} & H_{\rho\beta} \\ H_{\beta\rho} & H_{\beta\beta} \end{bmatrix},$$

where the  $(u, v)^{th}$  element of  $H_{\beta\beta}$  is

$$\begin{aligned} \frac{\partial^2 l_1}{\partial \beta_u \partial \beta_v} &= - \sum_{i=1}^K \{ \pi_{i,1} \bar{\pi}_{i,1} Z_{i,1,u} Z_{i,1,v} \\ &\quad + \sum_{t=2}^{n_i} \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{i,t} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \frac{\partial p_{it}}{\partial \beta_u} \frac{\partial p_{it}}{\partial \beta_v} - \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \left( \frac{\partial^2 p_{it}}{\partial \beta_u \partial \beta_v} \right) \}, \end{aligned}$$

the  $u^{th}$  element of  $H_{\beta\rho}$  is

$$\frac{\partial^2 l_1}{\partial \beta_u \partial \rho} = - \sum_{i=1}^K \sum_{t=2}^{n_i} \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{i,t} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \frac{\partial p_{it}}{\partial \beta_u} \frac{\partial p_{it}}{\partial \rho} - \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \left( \frac{\partial^2 p_{it}}{\partial \beta_u \partial \rho} \right),$$

and

$$H_{\rho\rho} = \frac{\partial^2 l_1}{\partial \rho^2} = - \sum_{i=1}^K \sum_{t=2}^{n_i} \left[ \frac{1}{p_{it} \bar{p}_{it}} + \frac{(y_{i,t} - p_{it})}{(p_{it} \bar{p}_{it})^2} (1 - 2p_{it}) \right] \left( \frac{\partial p_{it}}{\partial \rho} \right)^2 - \frac{(y_{i,t} - p_{it})}{p_{it} \bar{p}_{it}} \frac{\partial^2 p_{it}}{\partial \rho^2}.$$

The partial derivatives appearing in these expressions are given by:

$$\begin{aligned} \frac{\partial^2 p_{it}}{\partial \beta_u \partial \beta_v} &= (1 - 2\pi_{i,t}) \pi_{i,t} \bar{\pi}_{i,t} Z_{i,t,u} Z_{i,t,v} \\ &\quad - \frac{\rho}{2} \sqrt{\pi_{i,t} \bar{\pi}_{i,t} \pi_{i,t-1} \bar{\pi}_{i,t-1}} (1 - 2\pi_{i,t}) [Z_{i,t,u} Z_{i,t-1,v} + Z_{i,t-1,u} Z_{i,t,v}] \\ &\quad + \frac{\rho}{2} \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} [y_{i,t-1} - \pi_{i,t-1}] a_{it}, \end{aligned}$$

where

$$\begin{aligned} a_{it} &= .5[(1 - 2\pi_{i,t})Z_{i,t,u} - (1 - 2\pi_{i,t-1})Z_{i,t-1,u}][(1 - 2\pi_{i,t})Z_{i,t,v} - (1 - 2\pi_{i,t-1})Z_{i,t-1,v}] \\ &\quad - 2[\pi_{i,t} \bar{\pi}_{i,t} Z_{i,t,u} Z_{i,t,v} - \pi_{i,t-1} \bar{\pi}_{i,t-1} Z_{i,t-1,u} Z_{i,t-1,v}], \\ \frac{\partial^2 p_{it}}{\partial \beta_u \partial \rho} &= .5 \sqrt{\frac{\pi_{i,t} \bar{\pi}_{i,t}}{\pi_{i,t-1} \bar{\pi}_{i,t-1}}} [y_{i,t-1} - \pi_{i,t-1}] [(1 - 2\pi_{i,t})Z_{i,t,u} - (1 - 2\pi_{i,t-1})Z_{i,t-1,u}] \\ &\quad - \sqrt{\pi_{i,t} \bar{\pi}_{i,t} \pi_{i,t-1} \bar{\pi}_{i,t-1}} Z_{i,t-1,u}, \end{aligned}$$

and

$$\frac{\partial^2 p_{it}}{\partial \rho^2} = 0.$$

Under regularity conditions sufficient to guarantee the consistency of  $\hat{\theta}$ , the following generalization of Zeger, Liang, and Self's (1985) Proposition 2.2 is immediate:

**PROPOSITION 2.** *Let  $Y_{i,t}$  ( $t = 1, \dots, n_i < \infty$ ;  $i = 1, \dots, K$ ) be binary series such that  $\text{logit } E(Y_{i,t} \mid Z_{i,t}) = Z'_{i,t}\beta$  and  $\text{corr}(Y_{i,t}, Y_{i,t-1}) = \rho$ . Let  $\theta = (\rho, \beta)$ . Then  $\hat{\theta} - \theta$  is asymptotically multivariate Gaussian with expectation 0 and covariance matrix*

$$V_1(\theta) = [E H_1(\theta)]^{-1} I_1^*(\theta) [E H_1(\theta)]^{-1},$$

where  $I_1^*(\theta) = E S(\theta)S'(\theta)$ .

A consistent estimate of  $E H_1(\theta)$ , the expectation of the Hessian under the true (unknown) model, is given by  $H_1(\hat{\theta})$ . A consistent estimate of  $I_1^*(\theta)$ , the variance-covariance matrix of the score function, is given by

$$\widehat{I}_1^* = \sum_{i=1}^K S_i(\hat{\theta})S'_i(\hat{\theta}),$$

where  $S_i(\cdot)$  is the score function for the  $i^{\text{th}}$  subject. Combining these results leads to a consistent estimate of  $V_1(\theta)$  given by

$$\widehat{V}_1 = [H_1(\hat{\theta})]^{-1} \widehat{I}_1^* [H_1(\hat{\theta})]^{-1}.$$

Because this extension to the Zeger, Liang, and Self approach incorporates time dependent covariates, it provides the necessary tools for the analysis of the binary hummingbird data. The extension presented in this section permits the success probabilities to be modelled as a function of time, while taking into account the correlation between measurements on each

bird, and thereby allows comparisons of temporal response patterns across treatments which are more direct than the comparisons achieved by either univariate analysis-of-variance for repeated measures or MANOVA.

### 3.4 Exploring the Binary Hummingbird Data

Prior to applying the extension of the methods of Zeger, Liang, and Self (1985) to the analysis of the binary hummingbird data, a series of plots was examined to determine appropriate parametric forms to be incorporated into modelling of the temporal response patterns.

To check if a day effect was apparent, the average treatment responses on the logit scale ( $\log(\frac{\bar{y}+c}{1-\bar{y}+c})$ ,  $c = .5$ ) for successive blocks of 3, 5, and 10 trials were plotted (see Figures 3.1-3.3). Because there were no systematic discontinuities at those trials corresponding to new days (trials 61 and 121), there seemed to be no need to model a day effect. To get an overall picture of how treatment responses changed over time, the averages over the 6 birds within each treatment were calculated separately at each time point, and then the logits of the results were smoothed using robust locally weighted linear least squares (“lowess”).

Robust locally weighted linear regression is a method for smoothing a scatterplot,  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , in which the fitted value at  $x_k$  is the value of a linear fit to the data using weighted least squares, where the weight for  $(x_i, y_i)$  is large if  $x_i$  is close to  $x_k$  and small if it is not. “Lowess” centres a window at  $x_k$  which includes the  $r$  nearest neighbours of  $x_k$  on either side. The  $2r + 1$  points in this window are weighted symmetrically about  $x_k$ . If  $x_k$  is near the endpoints, however, “lowess” cannot centre the window there. Points close to the endpoint, including  $x_k$ , get larger weights than they would if the window was centred at  $x_k$ . If the data

has extreme values near the endpoints, the resulting fit for  $x_k$  may not be consistent with fits for points at which the window can be centred. Consequently, “lowess” fits near the endpoints may have peculiar features.

Figures 3.4-3.6 show the results of this “lowess” smoothing with windows of 10, 20, and 30 points. In these figures, the fits near the endpoints do not seem peculiar because logits of treatment average responses (over 6 birds) have been smoothed. This averaging reduces the number of extreme values in the data and protects against extreme values near the endpoints. Figure 3.6 seems to best display the basic features of the change in transformed responses over time. The average responses for birds in treatments 2 (PT) and 3 (FT) appear to grow roughly linearly with time in the logit scale, with the responses for birds in treatment 3 (FT) being higher, in general, than the responses for birds in treatments 1 (NT) and 2 (PT). This is contrary to the expectations of the investigator, who used psychological principles of visual perception to predict that birds in the treatment groups with discontinuous (PT) and continuous (FT) visible guides should have similar learning trajectories. For birds in treatment 1, improvement in the average response appears to be either linear or quadratic. The responses for birds in treatment 1 seem to be higher, in general, than the responses for birds in treatment 2, but not as high as those for birds in treatment 3. Of all the birds, those in treatment 3 seem to have the highest responses throughout the entire experiment.

To study individual-to-individual variation within each treatment, the logits of individual results ( $\log(\frac{y+c}{1-y+c})$ ,  $c = .5$ ) at each time point were also smoothed by “lowess”, using a window of 30 consecutive time points. Figures 3.7-3.9 display the results of this smoothing. Step-like patterns associated with sudden awareness of the relationship between light cue and feeder

were expected by the experimenter, but are not present in these plots. In all the figures, the peculiarities of “lowess” fits near the endpoints occur because individual logits, rather than logits of treatment averages are smoothed. The individual logits are more likely to have extreme values at the endpoints, leading to fits which may not accurately describe the patterns in the data of each individual. In Figure 3.7, note that two females in treatment 1 share the same “lowess” fits given by the horizontal solid line. This is because both birds have binary series which are virtually all zeroes; in other words, almost all of their *first* feeding forays were unsuccessful. In addition, a third female has a binary series which is predominantly zero up to the 110<sup>th</sup> trial. In the interior of the data sets, the plots do not show any obvious change in response variability over time for any of the treatments. Moreover, at any given time, the variability of responses for birds in treatment 1 (NT) appears to be larger than the variability of responses for birds in treatments 3 (FT). At any given time, the variability of responses for birds in treatment 2 (PT) appears to be the smallest of all three treatments.

These plots of the “lowess”-smoothed binary responses for each bird do not suggest an obvious difference between the responses of females and males in treatment 2, but in treatments 1 and 3, the male responses seem to be higher, on average, than the female responses. In fact, when designing the experiment, the investigator did not anticipate a difference in the success probabilities for the first feeding foray of male and female hummingbirds. However, since sex may be incorporated into the model with little extra effort, two analyses of the binary hummingbird data are presented in the following sections; the first analysis does not include sex as a covariate, while the second does. The first analysis corresponds to the classical analysis (univariate analysis-of-variance for repeated measures) discussed in the previous section, where



sex was ignored. Note that the unequal number of males and females on one of the treatments in the experiment would complicate including sex as a covariate in the classical analysis.

For the binary hummingbird data on the first feeding foray, both the analysis excluding sex and the analysis including sex will be repeated under the *independence* and the *first-order* working model. The *independence* working model assumes that observations within individuals, as well as between individuals, are independent. The *first-order* working model, assumes that the observations for all individuals share a common first lag autocorrelation.

### 3.5 Model Fitting With Sex Excluded

#### 3.5.1 The Initial Model

For the analysis of the binary data which excludes sex as a covariate, the exploratory plots suggest that the transformed probability of success on the first feeding foray may be modelled, initially, as

$$\text{logit } \pi_{i,t} = \mu_i + \beta_{1,i}t + \beta_{2,i}t^2 \quad i = 1, \dots, 3, \quad T = 1, \dots, 180, \quad t = \frac{T - \bar{T}}{s.d.(T)}$$

where  $\mu_i$  ( $i = 1, 2, 3$ ) may be interpreted as the main effect of the  $i^{th}$  treatment. Since birds in different treatment groups are not restricted to have the same linear and quadratic coefficients for the pattern over time, interactions between time and treatment group are permitted by this model. Note that the time covariate has been standardized to orthogonalize the highly collinear linear and quadratic time covariates.

### 3.5.2 An Informal Check of the Working Likelihoods

When calculating the standard errors of the parameter estimates, we must take into account that the working likelihood may not be the same as the true likelihood. Section 3.3 provides consistent estimators of the covariance matrix for the parameter estimates from which “robust” standard error estimates can be obtained. Because these “robust” standard error estimates account for the fact that the working likelihood may be incorrectly specified, we expect them to be *different* from the “naive” standard error estimates calculated assuming that the working likelihood correctly specifies the true likelihood. If the working likelihood is correctly specified, the robust standard error estimates should be the same as the naive standard error estimates as the number of subjects becomes large. Thus, an informal method for checking how well the working likelihood approximates the true likelihood (when the number of subjects is sufficiently large) might compare the naive and robust standard error estimates or the corresponding naive and robust z-scores. Differences between robust and naive standard error estimates could result from two possibilities: (i) although the working likelihood adequately described the true likelihood, the number of subjects was not large enough for accurate estimation of the true standard errors; or, (ii) the working likelihood was not adequate for the data.

Table 3.1 presents the naive and robust z-scores resulting from the initial model fits for both the independence and first-order working likelihoods. Notice that the naive z-scores for the independence fit are consistently larger, and typically substantially so, than the corresponding robust z-scores. If we assume that the robust z-scores reflect the associations between the outcome variable and the covariates accurately, we might conclude that the naive z-scores for the independence working likelihood tend to overstate the association, a well-known result

when the observations on a subject are positively correlated. More efficient parameter and standard error estimates might then be expected from the first-order working likelihood which models some correlation. Although substantial differences still exist for many parameters, the agreement between the naive and robust z-scores appears slightly improved for the first-order working likelihood. This suggests the first-order working likelihood may better reflect the actual likelihood than the independence working likelihood; a gain in efficiency of estimation would then result from using the first-order working likelihood instead of the independence working likelihood. This conclusion is somewhat tenuous since the assumption that the robust z-scores reflect the associations between the outcome variables and the covariates accurately may not be reasonable with only 18 birds in the experiment. Qualitatively, the conclusions obtained from the two working likelihoods are identical: modelling of the treatment main effects and the linear terms seems worthwhile, while modelling of the quadratic terms for treatments 2 and 3 may be unnecessary.

Table 3.1. Comparison of Z-scores from the Initial Model

parameter	Independence			First-Order		
	estimate	robust z-score	naive z-score	estimate	robust z-score	naive z-score
$\mu_1$	-.31	-.97	-3.35	-.32	-1.00	-2.95
$\mu_2$	-.71	-3.41	-7.23	-.68	-3.68	-6.32
$\mu_3$	.16	.42	1.72	.17	.42	1.54
$\beta_{1,1}$	.27	2.31	4.14	.26	2.37	3.51
$\beta_{1,2}$	.50	3.03	7.27	.48	3.31	6.42
$\beta_{1,3}$	.60	2.70	9.13	.62	2.54	7.92
$\beta_{2,1}$	-.21	-2.37	-2.82	-.20	-2.23	-2.41
$\beta_{2,2}$	.04	.37	.47	.03	.35	.41
$\beta_{2,3}$	-.09	-.68	-1.29	-.09	-.66	-1.02

### 3.6 Model Reductions With Sex Excluded

#### 3.6.1 Strategy for Model Reductions

The primary aim of this analysis is to understand the differences in hummingbird learning across treatment groups. Because birds in different treatments may learn in different ways, the initial model for the success probabilities allows all possible interactions of time and treatment. The purpose of subsequent model reductions is to obtain a parsimonious final model which reliably addresses the investigator's questions about the way hummingbird learning differs across treatments. To focus the analysis on treatment differences, the reduction procedure should eliminate unnecessary terms for negligible interactions between treatment and time first, leaving the examination of treatment main effects for the final steps. The elimination of interaction terms reduces the complexity of the model to be used for interpretation; in particular, a final model with no interactions between treatment and time permits differences in learning patterns for birds in different treatment groups to be easily described by treatment main effects only.

One such reduction procedure for the binary hummingbird data (with sex excluded as a covariate) is outlined in the following steps:

1. Fit the initial model.
2. Check the quadratic terms.
  - Can they be set equal?
  - If so, can this common term be eliminated?
3. Check the linear terms.

- Can they be set equal?
- If so, can this common term be eliminated?

4. Examine the treatment main effects.

This procedure is not meant to be followed rigidly. Instead, it is meant as a rough guideline for model reductions, to be used in conjunction with existing scientific theory and exploratory plots of the data.

### 3.6.2 Description of the Reductions

Using the strategy outlined above, two plausible model reduction paths arise for both the independence and first-order working likelihoods. In this section, the two models are compared, and for each working likelihood, the model most consistent with patterns in the data suggested by the exploratory plots is selected to represent the experimental results.

To examine how the results of the analysis depend on the choice of the working likelihood, these final models are then compared across working likelihoods. Final models with similar parameter and standard error estimates for different working likelihoods suggest that the efficiency of the two working likelihoods is about the same. Note that with only 18 birds in the experiment, the number of subjects may not be sufficiently large to ensure the adequacy of the asymptotic multivariate normal approximation to the distribution of the parameter estimates. The Wald statistics used to test the model reductions are based on this approximation.

The first model reduction path is described by the following steps:

1. Check the quadratic terms.

- Can they be set equal? Yes; the p-values for this reduction are .17 and .21 for the independence and first-order working likelihoods, respectively.
- Can the common quadratic term be eliminated? Yes; the p-values for this reduction are .17 and .19 for the independence and first-order working likelihoods, respectively.  
(P-values are calculated using the reduced model from the previous step.)

2. Check the linear terms.

- Can they be set equal? Yes; the p-values for this reduction are .29 and .26 for the independence and first-order working likelihoods, respectively.
- Can this common linear term be eliminated? No; the p-value for this reduction is less than .001 using both the independence and the first-order working likelihoods.

3. Examine the treatment main effects.

- Can they be set equal? Not entirely clear; the p-value for this reduction is .14 using both the independence and first-order working likelihoods.

This first plausible reduction path leads to models with coefficients and standard error estimates given in Table 3.2. Note that the predicted success probabilities resulting from the final models are virtually identical. Figure 3.10 displays the final model fit for the independence working likelihood and suggests the success probabilities for birds in treatments 1 and 2 are very similar. Since  $\text{corr}(\widehat{\mu}_1, \widehat{\mu}_2) \approx 0$ , the standard errors in Table 3.2 indicate that the model could be further reduced by taking main effects for treatments 1 and 2 to be the same. In fact, the p-values for the reduction are .69 and .71, with the common fitted values for  $\mu_1$  and  $\mu_2$  being

$-.60$  (s.e.=.18) and  $-.58$  (s.e.=.17), for the independence and first-order working likelihoods, respectively.

Table 3.2. Final Model Fits from the First Reduction Path

Parameters	Independence		First-Order	
	estimate	s.e.	estimate	s.e.
$\rho$	-	-	.17	.03
$\mu_1$	-.53	.31	-.52	.30
$\mu_2$	-.67	.17	-.65	.15
$\mu_3$	.07	.33	.08	.34
$\beta_{1,1} = \beta_{1,2} = \beta_{1,3}$	.45	.11	.45	.10
$\beta_{2,1} = \beta_{2,2} = \beta_{2,3}$	-	-	-	-

The estimated lag one autocorrelation and standard error provided by the first-order working likelihood suggest the presence of (positive) correlation in the data of each bird. On the other hand, the first-order and independence working likelihoods result in virtually the same estimates of the regression parameters and their standard errors. Both final models indicate the transformed response depends only linearly on time, with a common slope for all three treatment groups; no terms for interaction of treatment and time are included. This lack of interaction in the fitted model allows for a simple interpretation of treatment differences through examination of treatment main effects only. For example, Figure 3.10 suggests the success probabilities of birds in treatment 1 are about the same as the success probabilities of birds in treatment 2, while birds in treatment 3 have higher success probabilities than birds in either treatment 1 or treatment 2 (about .67 units greater in the logit scale at any given time).

These final models describe the success probabilities of birds in treatment 1 by a linear trajectory. This seems to disagree with the exploratory lowess plot of the transformed outcomes in Figure 3.6, which suggests the success probabilities of birds in treatment 1 could be better

described by a quadratic trajectory. Figure 3.6 also suggests birds in treatment 2 start out with lower success probabilities than birds in treatment 1, but develop higher success probabilities by the end of the experiment.

A possible explanation for the discrepancy between the lowess plot and the first final model emerges upon examining the path of model reductions leading to the first final model. Although the p-values for setting the quadratic terms for all three treatments equal are .17 and .21 for the independence and first-order working likelihoods respectively, and the p-values for setting this common quadratic term to be equal to zero are .17 and .19 for the independence and first-order working likelihoods, respectively, the p-values for proceeding directly from the full model to a model with no quadratic terms are a marginal .12 and .14 for the independence and first-order working likelihoods, respectively. This suggests the final model may be neglecting some curvature present in the data. In fact, z-scores for the quadratic term in treatment 1 ( -2.37 and -2.23 for the independence and first-order working likelihoods, respectively; see Table 3.1) lend support to what is suggested by Figure 3.6: much of the curvature present in the data comes from the learning trajectory of birds in treatment 1.

Table 3.3 displays the coefficient and standard error estimates resulting from an alternate reduction path which allows curvature in the fitted learning trajectory of birds in treatment 1. This second reduction path is described by the following steps:

1. Check the quadratic terms.

- Can they be eliminated? No; the p-values for this reduction are a marginal .12 and .14 for the independence and first-order working likelihoods, respectively.



2. Check the linear terms.

- Can they be set equal? Yes; the p-values for this reduction are .32 and .27 for the independence and first-order working likelihoods, respectively.
- Can the common linear term be eliminated? No; the p-values for this reduction are less than .001 for both the independence and first-order working likelihoods.

3. Examine the treatment main effects.

- Different quadratic terms for different treatment groups are still present in the model, so checking whether the treatment main effects are equal corresponds to checking whether the success probabilities for all three treatment groups are equal at  $t = 0$ , or at trial 90.5. The p-values for this reduction are .12 and .13 for the independence and first-order working likelihoods, respectively, indicating that this reduction may not be justified. Note that even if this reduction was justified, it may be of little practical interest because it does not simplify interpretation of the final model.

4. For interpretability of the final model...

- Eliminate the quadratic terms for treatments 2 and 3. The p-values for this reduction are .62 and .74 for the independence and first-order working likelihoods, respectively. These p-values justify what is suggested by the lowess plot: the success probabilities of birds in treatments 2 and 3 can be adequately described by a linear trajectory.

Table 3.3. Final Model Fits from the Second Reduction Path

Parameters	Independence		First-Order	
	estimate	s.e.	estimate	s.e.
$\rho$	-	-	.17	.03
$\mu_1$	-.31	.32	-.52	.30
$\mu_2$	-.67	.17	-.65	.15
$\mu_3$	.07	.33	.08	.34
$\beta_{1,1} = \beta_{1,1} = \beta_{3,1}$	.46	.11	.45	.10
$\beta_{2,1} = \beta_{2,2} = \beta_{2,3}$	-.23	.10	-.22	.10

As was the case for the first reduction path, the estimated first lag autocorrelation and standard error from the first-order working likelihood suggest the presence of (positive) correlation in the data of each bird, but the predicted success probabilities resulting from the independence and first-order fits are virtually identical. Figure 3.11 plots the final model for the second reduction path using the independence working likelihood. Comparison of Figures 3.10 and 3.11 suggests the only noticeable difference between the final models for the two reduction paths is the fit for the learning trajectory of birds in treatment 1. Because the final model from the second reduction path allows curvature in the fitted learning trajectory of treatment 1 and the final model from the first reduction path does not, the former appears to be more consistent with patterns displayed by the exploratory lowess plots, and may thus be considered the more reliable summary of the experimental results.

Note that the final fits for birds in treatment 1 resulting from the two different reduction paths do not differ greatly in the probability scale except at the very beginning and the very end of the experiment; see Figure 3.12. The largest discrepancies in the fitted success probabilities occur at the end of the experiment, with the maximum difference in fitted probabilities on the last trial. Here the linear model fits a success probability of .56 (with a standard error of about

.09) and the quadratic model fits a probability of .45 (with a standard error of about .10).

### 3.7 Residual Analysis for the Final Model with Sex Excluded

Analysis is based on the Pearson residuals (given by  $(y_{it} - \widehat{\pi}_{it}) / \sqrt{\widehat{\pi}_{it}(1 - \widehat{\pi}_{it})}$ ) from both the independence and first-order fits. To determine whether the independence working assumption is reasonable (recall that the independence working assumption isn't necessary for asymptotically valid results), the independence residuals are examined to see if they are approximately uncorrelated with mean 0 and variance 1. After analyzing the independence residuals, the first-order residuals are examined to determine whether the assumption of a common lag one autocorrelation for all birds (*necessary* for asymptotically valid results when using the first-order working likelihood) is reasonable.

#### 3.7.1 Analysis of the Independence Residuals

Under the independence working assumption, the 180 Pearson residuals for each bird should be approximately uncorrelated, with mean 0 and variance 1. Figure 3.13 displays scatterplots of the Pearson residuals for each bird and suggests the distributions of residuals for all birds have mean 0 and variance 1, with those for birds in treatments 1 and 2 being positively skewed, and those for birds in treatment 3 being more symmetric. This is explained by the relative number of 0's and 1's in each treatment. Only 38% of treatment 1 responses (407/1080) and 35% of treatment 2 responses (374/1080) are 1's, whereas 52% of the responses in treatment 3 are 1's (558/1080). The final model therefore fits the "1" responses in treatment 3 as well as the "0" responses, but for treatments 1 and 2, fits "0" responses better than "1" responses since the

majority of responses in these treatments is 0. Note that the gaps around zero in Figure 3.13 arise because the responses are either 0 or 1, and the denominator of the Pearson residuals ( $\sqrt{\widehat{\pi}_{it}(1 - \widehat{\pi}_{it})}$ ) is essentially constant.

To check if Pearson residuals from the independence final model were approximately uncorrelated within birds, estimates of the first, second, and third lag Pearson residual autocorrelations were calculated for each bird (see Table 3.4). Under the independence working assumption, the standard error of each of these autocorrelation estimates is about  $1/\sqrt{180} = .075$ . From Table 3.4, we see that 4 out of 18, or 2/9 of the first lag autocorrelations appear significantly different from zero at level  $\alpha = .05$  (birds 5, 6, 14, and 15 have  $\hat{\rho}_1$ 's greater than .15). Because birds are independent, we would expect this to happen for only 1 out of every 20 estimated first lag autocorrelations. Therefore, the assumption that observations within birds are independent seems questionable for these data. It is interesting to note that when the three autocorrelations are compared across birds, birds 14 and 15 (both in treatment 3) appear to have higher first, second, and third lag autocorrelations than the others.

Checking the adequacy of the final model is limited by the fact that the only explanatory variables modelled are treatment group and time. The patterns of Pearson residuals across treatment groups can be seen in Figure 3.13 and have already been explained by the relative number of 0's and 1's in each treatment group. Note that the Pearson residuals for the  $i^{th}$  treatment group can be characterized by two functions of time - one (essentially the negative fit, or  $-\widehat{\pi}_{it}$ ) for those responses which are 0; the other (essentially the negative fit plus one, or  $1 - \widehat{\pi}_{it}$ ) for those responses which are 1. Thus, plotting residuals against time is not a useful check of model adequacy. Another variable, excluded as a covariate in the model because its

role in hummingbird learning was thought to be minimal, is the sex of each bird. Figure 3.14, a boxplot of the Pearson residuals by sex, suggests sex should be included in the model. The residual distribution for males appears to be positively skewed, whereas for females it appears to be negatively skewed. The skewness of the distributions is a consequence of the females having a smaller proportion (607/1800 or 34%) of successful first feeding forays (response=1) than the males (732/1440 or 51%).

Table 3.4. Estimated Autocorrelations for the Independence Residuals

	Bird (Treatment 1)					
	1	2	3	4	5	6
$\hat{\rho}_1$	.07	-.02	.000	-.06	.16	.16
$\hat{\rho}_2$	.11	-.09	.09	-.05	.02	.02
$\hat{\rho}_3$	.08	-.09	.03	-.04	.10	.08

	Bird (Treatment 2)					
	7	8	9	10	11	12
$\hat{\rho}_1$	-.005	.13	-.09	-.07	.10	.02
$\hat{\rho}_2$	-.05	.07	.13	-.09	.03	.04
$\hat{\rho}_3$	-.009	.10	-.07	.10	-.06	.002

	Bird (Treatment 3)					
	13	14	15	16	17	18
$\hat{\rho}_1$	.01	.24	.34	-.07	.02	.10
$\hat{\rho}_2$	.14	.16	.28	.03	.15	.07
$\hat{\rho}_3$	.02	.28	.37	-.006	-.01	.12

### 3.7.2 Analysis of the First-Order Residuals

For the first-order working model, let  $r_i$  denote the estimate of the  $i^{th}$  subject's lag 1 autocorrelation obtained by regressing the Pearson residuals for that subject on their predecessors. The variance of  $r_i$ , approximated by  $1/n_i(1 - \rho_i^2)$ , where  $n_i$  and  $\rho_i$  are, respectively, the number

of time points and the lag one autocorrelation for the  $i^{th}$  subject, can be stabilized by taking the inverse hyperbolic sine of  $r_i$ . If the number of time-points for each subject is large,  $\sinh^{-1}(r_i)$  is approximately normally distributed with mean  $\sinh^{-1}(\rho_i)$  and variance  $1/n_i$ . The standard error of all the transformed individual first-lag autocorrelation estimates is thus approximately  $1/\sqrt{180} = .075$ , allowing a check on the assumption of a common first lag autocorrelation for all subjects. Testing the values in Table 3.5 for homogeneity (average = .16, sample variance = .021), leads to a chi-squared statistic of 65.2 on 17 degrees of freedom; the data thus provide quite convincing evidence that a common lag one autocorrelation is not shared by all subjects.

Table 3.5. Estimated Lag One Autocorrelations for the First-Order Residuals

		Bird (Treatment 1)					
		1	2	3	4	5	6
$r_i$		.16	.04	.27	.04	.33	.19
$\sinh^{-1}(r_i)$		.16	.04	.27	.04	.33	.19
		Bird (Treatment 2)					
		7	8	9	10	11	12
$r_i$		-.005	.13	.06	-.07	.17	.03
$\sinh^{-1}(r_i)$		-.005	.13	.06	-.07	.17	.03
		Bird (Treatment 3)					
		13	14	15	16	17	18
$r_i$		.04	.27	.47	.15	.45	.15
$\sinh^{-1}(r_i)$		.04	.27	.45	.15	.44	.15

The assumption of a common lag one autocorrelation is critical to deriving the asymptotic results for the first-order working model, so the fact that this assumption appears to be violated casts doubt on the validity of such an analysis. However, except for birds 3, 5, 14, 15 and 17, the estimated first-lag autocorrelations are not particularly large. The reason for compelling

evidence against homogeneity is that the standard errors of the estimates are quite small due to the large number of responses ( $n_i = 180$ ) for each bird. Although the assumption of a common lag one autocorrelation does not appear to hold for the binary responses of the 18 birds in this experiment, the conclusions drawn from the analysis based on the first-order working likelihood are virtually identical to those from the analysis based on the independence working likelihood. Therefore, the conclusions drawn from the first-order analysis appear to be reasonable.

The Pearson residuals obtained from the first-order working likelihood are virtually the same as those obtained by the independence working likelihood because the final models are almost identical. The plot of the first-order residuals against sex has the same features as the plot of the independence residuals displayed in Figure 3.14. Hence, analysis of both types of Pearson residuals indicates sex is an important covariate.

### 3.8 Summary of Analysis with Sex Excluded

The models obtained by the independence and first-order working likelihoods are virtually identical. For both working likelihoods, the final model corresponding to the second reduction path has been selected to represent the experimental results. This final model indicates an increase in success probabilities over successive trials for treatments 2 (PT) and 3 (FT), and an increase, peaking at about trial 110 (near the end of the second day), followed by a decline in success probability for treatment 1 (NT).

The final model uses a quadratic trend in the logit scale to describe the pattern of improvement in treatment 1, and a linear trend to describe the patterns of improvement in treatments 2 and 3. Its main features are summarized as follows:

- (i) birds in treatment 3 (FT) have greater fitted success probabilities than birds in treatments 1 (NT) and 2 (PT) throughout the entire experiment;
- (ii) the fitted rate of improvement for birds in treatment 2 (PT) is the same as the fitted rate of improvement for birds in treatment 3 (FT); and
- (iii) although birds in treatment 2 (PT) have lower fitted success probabilities than birds in treatment 1 (NT) for the intermediate trials of the experiment (trials 30-120), sometime during the course of the third day of the experiment (at about the 150<sup>th</sup> trial), they develop greater fitted success probabilities than birds in treatment 1.

Note that different working assumptions lead to virtually identical fits. This is because both approaches give similar parameter and standard error estimates, suggesting the difference in efficiency between the two approaches is minimal. The estimated lag one auto-correlation and standard error provided by the first-order working likelihood seem to indicate the presence of positive correlation in the data (see Tables 3.2 and 3.3), suggesting the first-order parameter estimates and standard errors may be more efficient. On the other hand, the first-order Pearson residuals suggest a common lag one autocorrelation may not be shared by all subjects. Because this assumption is critical in deriving the asymptotic results for the first-order parameter estimates and standard errors, the independence model fit (which does not require any assumptions about the time-dependence within each subject's responses) may be considered the more reliable summary of the experimental results.

Both the exploratory plots and the residual analysis suggest sex may be an important covariate worth including in the model. In the next section, the binary hummingbird data is



analyzed with sex included as a covariate.

### 3.9 The Initial Model With Sex Included

The importance of sex as a covariate is clearly indicated by Figures 3.15-3.17 which plot the lowess smoothed logits of the average response by sex for each treatment group. On average, males seem to have a higher success probability than females in all three treatments throughout the entire experiment. For treatments 2 (PT) and 3 (FT), both males and females appear to have success probabilities which increase roughly linearly with time, whereas for treatment 1 (NT), the male and female success probabilities appear to increase over the first two days, but to decrease over at least part of the third day. For all three treatment groups, it appears a model incorporating an additive (on the logit scale) effect for sex, with males having higher transformed success probabilities than females, may be adequate to explain the patterns in the data. Furthermore, sex-by-treatment interaction may also be present, since the figures suggest that a different additive effect for sex in each treatment group may be necessary.

These exploratory plots suggest the transformed probability of success on the first feeding foray may be modelled, initially, as

$$\text{logit } \pi_{i,t,s} = \mu_i + \text{sex}_{i,s} + \beta_{1,i,s}t + \beta_{2,i,s}t^2 \quad i = 1, \dots, 3, \quad s = 1, 2, \quad T = 1, \dots, 180, \quad t = \frac{T - \bar{T}}{s.d.(T)}$$

where  $\mu_i$  may be interpreted as the main effect of the  $i^{\text{th}}$  treatment group,  $\text{sex}_{i,1}$  is the female-to-male differential sex effect for the  $i^{\text{th}}$  treatment group, and  $\text{sex}_{i,2} \stackrel{\text{set}}{=} 0$  (for males in the  $i^{\text{th}}$  treatment group). The female-to-male differential sex effect is permitted to vary across treatments so treatment-by-sex interaction is incorporated. Similarly, the three-way interaction of sex, treatment and time is incorporated because different linear and quadratic coefficients

for each sex and each treatment group are permitted. Again, the time covariate has been standardized to orthogonalize the highly collinear linear and quadratic time covariates. As in Sections 3.5-3.8, the analysis is repeated under both the *independence* and the *first-order* working model.

### 3.10 Model Reductions with Sex Included

#### 3.10.1 Strategy for Model Reductions

The reduction strategy for the initial model which includes sex as a covariate is similar to its counterpart with sex ignored. Because the investigator does not anticipate complicated sex-by-time interactions, the reduction strategy checks for the presence of such interactions before it checks for the interaction of treatment group and time. In addition, quadratic terms common to males and females within a treatment are treated differently from the corresponding linear terms because Figure 3.15 suggests males and females in the treatment 1 have success probabilities (for the first feeding foray) which may be described by a quadratic trend in the logit scale. The general reduction strategy is outlined by the following steps:

1. Fit the initial model.
2. Check the quadratic terms.
  - Can the coefficients for males and females be set equal within all treatment groups?
  - If so, can they be eliminated?
3. Check the linear terms.
  - Can the coefficients for males and females be set equal within all treatment groups?

- If so, can these coefficients be set equal for all treatment groups?
  - If so, can this common term be eliminated?
4. Check the differential sex effects.
- Can the coefficients be set equal for all treatment groups?
  - If so, can this common term be eliminated?
5. Examine the treatment main effects.

As in Section 3.6, this procedure is meant as a rough guideline for model reductions, to be used in conjunction with existing scientific theory and exploratory plots of the data.

### 3.10.2 Description of the Reductions for the Independence Working Likelihood

The model reduction path for the independence working likelihood including sex as a covariate is described by the following steps:

1. Check the quadratic terms.
  - Can the coefficients for males and females be set equal within all treatment groups?  
Yes; the p-value for this reduction is .80.
  - Can the quadratic coefficients be eliminated? No; the p-value for the simultaneous elimination of all three quadratic coefficients is a marginal .12. Leave the quadratic terms for each treatment in the model for the time being.
2. Check the linear terms.

- Can the linear coefficients for males and females be set equal within each treatment group? Yes; the p-value for this reduction is .91.
- Can the resulting linear coefficients be set equal for all treatment groups? Yes; the p-value for this reduction is .38.
- Can this common linear term be eliminated? No; the p-value for this reduction is less than .001.

3. Check the differential sex effects.

- Can they be set equal for all treatment groups? Yes; the p-value for this reduction is .40.
- Can this additive sex effect be eliminated? No; the p-value for this reduction is .007.

4. Examine the treatment main effects.

- Different quadratic terms for different treatment groups are still present in the model, so checking whether the treatment main effects are equal corresponds to checking whether the success probabilities for all three treatment groups are equal at  $t = 0$ , or at trial 90.5. The p-value for this reduction is .10, indicating this reduction may not be justified. As explained in Section 3.6, a reduction like this is of little practical interest because it does not simplify interpretation of the final model.

5. For interpretability of the final model...

- Eliminate the quadratic terms for treatments 2 and 3. The p-value for this reduction is .74, justifying what is suggested by the lowess plots in Figures 3.16 and 3.17: the

success probabilities of birds in treatments 2 and 3 can be adequately described by a linear trajectory.

Note that quadratic terms for all three treatment groups are permitted to remain in the model until the final step, leading to a final model which is consistent with the patterns observed in the lowess plots. The reduction path for the independence working likelihood is summarized in Table 3.6.

Table 3.6. Model Reductions for the Independence Working Likelihood with Sex Included

Parameters	Model					
	1	2	3	4	5	6
$\mu_1$	.36(.03)	.36(.17)	.36(.17)	.39(.17)	.18(.27)	.18(.27)
$\mu_2$	-.59(.31)	-.53(.26)	-.53(.24)	-.53(.24)	-.36(.26)	-.32(.20)
$\mu_3$	.53(.75)	.60(.71)	.60(.69)	.58(.67)	.54(.48)	.45(.41)
$sex_{1,1}$	-1.03(.42)	-1.02(.42)	-1.03(.43)	-1.06(.44)	-.75(.28)	-.75(.28)
$sex_{2,1}$	-.25(.40)	-.37(.30)	-.36(.30)	-.36(.31)	set= $sex_{1,1}$	set= $sex_{1,1}$
$sex_{3,1}$	-.72(.76)	-.86(.66)	-.86(.64)	-.83(.60)	set= $sex_{1,1}$	set= $sex_{1,1}$
$\beta_{1,1,1}$	.21(.09)	.21(.09)	.29(.14)	.48(.11)	.47(.11)	.48(.11)
$\beta_{1,2,1}$	.55(.21)	.54(.21)	.50(.16)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,1}$	.54(.38)	.54(.38)	.63(.24)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,1,2}$	.41(.31)	.41(.31)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,2,2}$	.46(.23)	.46(.24)	set= $\beta_{1,2,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,2}$	.74(.27)	.73(.27)	set= $\beta_{1,3,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{2,1,1}$	-.21(.08)	-.21(.09)	-.22(.08)	-.24(.09)	-.24(.10)	-.24(.10)
$\beta_{2,2,1}$	-.04(.05)	.03(.10)	.04(.10)	.04(.10)	.04(.11)	-
$\beta_{2,3,1}$	-.16(.15)	-.09(.12)	-.10(.15)	-.10(.14)	-.09(.14)	-
$\beta_{2,1,2}$	-.22(.21)	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$
$\beta_{2,2,2}$	.10(.18)	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	-
$\beta_{2,3,2}$	-.007(.18)	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	-
Working Deviance	4029.67	4031.50	4035.64	4048.42	4062.55	4064.58
Pearson $X^2$	3235.89	3233.99	3236.09	3252.95	3252.44	3252.62
degrees freedom	3222	3225	3228	3230	3232	3234

### 3.10.3 Description of the Reductions for the First-Order Working Likelihood

The model reduction path for the independence working likelihood including sex as a covariate is described by the following steps:

## 1. Check the quadratic terms.

- Can the coefficients for males and females be set equal within all treatment groups?

Yes; the p-value for this reduction is .74.

- Can the quadratic coefficients be eliminated? Yes; the p-value for this reduction is .20.

## 2. Check the linear terms.

- Can the linear coefficients for males and females be set equal within each treatment group? Yes; the p-value for this reduction is .88.

- Can the resulting linear coefficients be set equal for all treatment groups? Yes; the p-value for this reduction is .34.

- Can this common linear term be eliminated? No; the p-value for this reduction is less than .001.

## 3. Check the differential sex effects.

- Can they be set equal for all treatment groups? Yes; the p-value for this reduction is .33.

- Can this additive sex effect be eliminated? No; the p-value for this reduction is .02.

## 4. Examine the treatment main effects.

- Can they be set equal? Not entirely clear; the p-value for this reduction is a marginal .10.

The reduction path for the first-order working likelihood is summarized in Table 3.7.

Since  $\text{corr}(\widehat{\mu}_1, \widehat{\mu}_2) \approx 0$ , the standard errors for the final model in Table 3.7 indicate that taking the main effects for treatments 1 and 2 to be equal would be a permissible further reduction of the model. In fact, the p-value for this reduction is .35, with the fitted value common to  $\mu_1$  and  $\mu_2$  being  $-.19$  (s.e. = .21). This reduction could be carried out for interpretability of the final model but is not included in this analysis.

Table 3.7. Model Reductions for the First-Order Working Likelihood with Sex Included

Parameters	Model					
	1	2	3	4	5	6
$\rho$	.13(.03)	.13(.03)	.13(.03)	.13(.03)	.13(.03)	.14(.03)
$\mu_1$	.36(.04)	.35(.20)	.16(.29)	.15(.25)	.17(.26)	-.04(.27)
$\mu_2$	-.58(.30)	-.53(.24)	-.49(.19)	-.50(.18)	-.50(.17)	-.30(.20)
$\mu_3$	.59(.87)	.65(.81)	.56(.76)	.54(.69)	.51(.64)	.46(.42)
$\text{sex}_{1,1}$	-1.03(.40)	-1.02(.42)	-1.02(.42)	-1.02(.41)	-1.06(.42)	-.73(.28)
$\text{sex}_{2,1}$	-.23(.37)	-.34(.27)	-.34(.27)	-.33(.27)	-.33(.27)	set= $\text{sex}_{1,1}$
$\text{sex}_{3,1}$	-.79(.88)	-.89(.75)	-.89(.76)	-.87(.69)	-.84(.64)	set= $\text{sex}_{1,1}$
$\beta_{1,1,1}$	.20(.08)	.20(.08)	.19(.07)	.27(.13)	.47(.11)	.46(.11)
$\beta_{1,2,1}$	.52(.20)	.51(.20)	.52(.20)	.49(.15)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,1}$	.54(.38)	.53(.38)	.53(.38)	.64(.26)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,1,2}$	.42(.35)	.42(.34)	.42(.36)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,2,2}$	.46(.22)	.46(.22)	.47(.23)	set= $\beta_{1,2,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,2}$	.77(.33)	.76(.33)	.77(.37)	set= $\beta_{1,3,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{2,1,1}$	-.19(.09)	-.20(.10)	-	-	-	-
$\beta_{2,2,1}$	-.03(.05)	.03(.10)	-	-	-	-
$\beta_{2,3,1}$	-.15(.15)	-.10(.13)	-	-	-	-
$\beta_{2,1,2}$	-.21(.24)	set= $\beta_{2,1,1}$	-	-	-	-
$\beta_{2,2,2}$	.09(.18)	set= $\beta_{2,2,1}$	-	-	-	-
$\beta_{2,3,2}$	-.04(.19)	set= $\beta_{2,3,1}$	-	-	-	-
Working Deviance	3980.30	3981.21	3988.20	3993.00	4005.18	4017.33
Pearson $X^2$	3230.33	3230.80	3234.58	3229.04	3234.18	3228.03
degrees freedom	3221	3224	3227	3230	3232	3234

As in the analysis where sex was not included as a covariate, the estimated lag one autocorrelation and standard error provided by the first-order working likelihood suggest the data for each bird are positively correlated. As before, the fits resulting from the independence and

first-order working likelihoods are similar. When sex is included in the analysis, the only difference in the fits is the learning trajectory for birds in treatment 1. The independence working likelihood models a quadratic learning trajectory for birds in treatment 1, whereas the first-order working likelihood models a linear learning trajectory. The fitted learning trajectories resulting from the independence and first-order working likelihoods are virtually identical for both treatments 2 and 3. Figures 3.18 and 3.19 display the fitted female learning trajectories in the logit scale for the independence and first-order working likelihoods respectively; because the effect of sex has been modelled as additive, the male trajectories are parallel to those for females, with the fitted (transformed) response for males being greater than that for females by an additive constant (-.75 and -.73 for the independence and first-order working likelihood, respectively).

### 3.11 Summary of Analysis with Sex Included

Except for the fitted learning trajectory for birds in the first treatment group, the final models obtained from the independence and first-order working likelihoods are virtually identical. Both final models describe the effect of sex as additive only; interactions of sex with time or with treatment group are not present in either model. Both models describe the success probabilities of males as being about .74 (s.e.= .28) units higher in the logit scale than those for females throughout the entire experiment; sex is clearly an important covariate for these binary response data. As in the previous models where sex was not included as a covariate, the transformed success probability for birds in treatments 2 (PT) and 3 (FT) appears to improve linearly at the same rate in both final models, with the transformed success probabilities for birds in



treatment 3 being about .77 (s.e.= .35) units higher than those for birds in treatment 2, for both the working likelihoods, throughout the entire experiment.

For the learning trajectories of birds in the first treatment group, different working likelihoods lead to different final models. The final model resulting from the independence working likelihood describes the learning trajectory for birds in treatment 1 (NT) by a quadratic trend in the logit scale for both males and females, whereas the final model resulting from the first-order working likelihood describes the same trajectory by a linear trend. For birds in treatment 1, the independence final model suggests an increase, peaking at about trial 110 (near the end of the second day), followed by a decline in success probability in the logit scale. The first-order final model, on the other hand, suggests a steady increase throughout the entire experiment (see Figures 3.18 and 3.19).

Figure 3.20 displays the final fits resulting from the two working likelihoods for female birds in treatment 1. Although these fits do not differ greatly in the probability scale, they may be interpreted quite differently. For example, the independence fit suggests success probabilities for birds in treatment 2 are higher than those for birds in treatment 1 at the end of the experiment (see Figure 3.18). On the other hand, the first-order fit suggests the success probabilities for birds in treatment 1 are higher than those for birds in treatment 2 (see Figure 3.19). Collecting more data on the first treatment group in the form of additional subjects or further trials might resolve this conflict in interpretation, since the presence or absence of a quadratic effect would presumably become more clear. In this analysis, the final model resulting from the independence working likelihood is viewed as the more reliable summary of the experimental results because the independence fits are more consistent with patterns in the exploratory plots (see Figures

3.15-3.18) than the first-order fits.

The results of this analysis indicate that sex is important in describing the probabilities of success on the first feeding foray for the hummingbirds in this experiment. One possible explanation for the importance of sex is the higher body-weight to wing-span ratio of males. Because males must work harder to obtain their food, they be more likely to succeed on the first feeding foray than females. Since sex is important in this analysis, it seems that any complete analysis of the hummingbird data (binary or binomial) should include sex as a covariate.

In the next chapter both the binary and binomial hummingbird data are analyzed with sex *included* as a covariate, using the generalized estimating equation (GEE) approach of Liang and Zeger (1986) and Zeger and Liang (1986). The GEE approach extends quasi-likelihood theory (see McCullagh and Nelder, 1989) to arrive at estimating equations for the regression parameters, and is applicable to any longitudinal response with univariate marginal distributions for which the quasi-likelihood formulation is sensible. Unlike the working likelihood approach described in this section, the GEE approach does not require assumptions about the true correlation structure when explicitly modelling correlation in the responses to improve the efficiency of estimation. In these cases, the GEE approach permits many alternatives to the independence “working correlation” structure, unlike the working likelihood approach where only one alternative is available.

## Chapter 4

### Analysis Based on the GEE Approach

This chapter presents the generalized estimating equation (GEE) approach described in Liang and Zeger (1986) and Zeger and Liang (1986). This unified approach to the analysis of longitudinal data can be applied to any repeated measures data with univariate marginal distributions for which the quasi-likelihood formulation is sensible. The GEE approach is first used to analyze the binary hummingbird data. Then, to illustrate an application to non-binary longitudinal data, it is used to analyze the binomial hummingbird data.<sup>1</sup>

#### 4.1 Description of the GEE approach

Let  $Y_{it}$ ,  $i = 1, \dots, K$  and  $t = 1, \dots, n_i$ , be the outcome variable for the  $i^{th}$  subject at the  $t^{th}$  measurement, and  $x_{it}$  be the corresponding  $p \times 1$  vector of covariates. Suppose that we are interested in how the outcome variable  $Y_{it}$  depends on the covariates  $x_{it}$ . If there were only one time of observation for each subject, a generalized linear model (see McCullagh and Nelder, 1989) could be used to model a variety of continuous and discrete outcome variables. When dealing with longitudinal data, however, the situation is complicated by the fact that there may

---

<sup>1</sup>All the GEE analyses reported here were carried out using a SAS macro received from K.Y. Liang, and written by M. Rezaul Karim. The SAS macro permits modelling of the following “working” correlation matrices: independence, stationary M-dependent, non-stationary M-dependent, AR-M, exchangeable, unspecified, and user-specified. Only working correlations which are *common* to all subjects may be specified. In addition, if observation times are unequally spaced or there is missing data, only the independence, exchangeable, or user-specified working correlations should be used.

be correlation among the repeated observations for a given subject. This correlation must be taken into account to obtain a valid analysis.

Modelling the correlation directly is often difficult for non-Gaussian longitudinal data because there are few natural models for the joint distribution of the repeated observations. Liang and Zeger (1986) and Zeger and Liang (1986) propose a methodology based on an extension of the quasi-likelihood approach described by Wedderburn (1974) and McCullagh (1983). They model the marginal distribution of the univariate outcome at each time point by specifying a known function of the mean as a linear function of the covariates, and by assuming the (marginal) variance is a known function of the mean. Rather than specify a form for the joint distribution of the outcome vector for each subject, they merely specify a “working” correlation matrix for this outcome vector. This partial specification of the joint distribution gives rise to generalized estimating equations (GEEs) which, under weak assumptions, provide consistent estimates of the regression coefficients and their asymptotic covariance matrix.

#### 4.1.1 Brief Description of Quasi-Likelihood

Since the GEE approach is an extension of the quasi-likelihood approach to modelling independent univariate outcomes, a brief description of the pertinent aspects of quasi-likelihood theory is provided to motivate the GEE approach and establish notation.

Quasi-likelihood is a unified methodology for the regression analysis of (independent) discrete or continuous outcomes first proposed by Wedderburn (1974) and examined further by McCullagh (1983). A desirable property of quasi-likelihood analysis is its flexibility. Unlike likelihood analysis, the actual form of the outcome distribution does not have to be specified;

only the relationships between the outcome mean and the covariates and between the mean and the variance are required.

To fix notation, let  $Y_i$  be the univariate response for the  $i^{th}$  individual, and let  $x_i$  be the corresponding  $p$ -dimensional vector of covariates for that individual. Define  $\mu_i$  and  $v_i$  to be the expectation and variance of  $Y_i$ . The expectation is assumed to be a known function of a linear combination of the covariates, i.e.,

$$\mu_i = h(x_i' \beta),$$

where  $\beta$  is a  $px1$  vector of regression parameters. In addition, except for an unknown scale parameter  $\phi$ , the variance is assumed to be a known function of the expectation, i.e.,

$$v_i = g(\mu_i)/\phi.$$

Inference for the regression parameters  $\beta$ , rather than for the nuisance parameter  $\phi$ , is the primary objective.

The quasi-likelihood estimator for  $\beta$  is obtained by solving a score-like system of equations where the equation for the  $l^{th}$  regression parameter,  $\beta_l$ , is given by

$$S_l(\beta) = \sum_{i=1}^K \frac{\partial \mu_i}{\partial \beta_l} v_i^{-1} (y_i - \mu_i) = 0, \quad l = 1, \dots, p. \quad (4.1)$$

These are in fact score equations for  $\beta$  when the distribution of the outcome is in the exponential family. The solution to these quasi-score equations is the quasi-likelihood estimator of the regression parameters, and can be obtained by the method of iteratively re-weighted least squares (for example, by the use of GLIM). McCullagh (1983) has shown that, under mild regularity conditions, the quasi-likelihood estimator of the regression parameters is asymptotically Gaussian.

#### 4.1.2 Generalized Estimating Equations (GEEs) for Longitudinal Data

To extend the quasi-likelihood approach to longitudinal data, Zeger and Liang (1986) adopt the structure described above for each of the univariate outcomes  $Y_{it}$ . Their extension involves the specification of a “working” correlation matrix for  $Y_i$ , the response vector of the  $i^{th}$  subject, and the incorporation of this correlation into the estimating equations for  $\beta$ . The observation times and the working correlation matrices are allowed to differ across subjects in the GEE approach.

Let  $R_i(\alpha)$  be the  $n_i \times n_i$  “working” correlation matrix for the  $i^{th}$  subject. Zeger and Liang (1986) call  $R_i(\alpha)$  the working correlation matrix because it is *not necessarily* specified correctly. The collection of working correlation matrices is assumed to be fully specified by  $\alpha$ , an  $s \times 1$  vector of unknown parameters. The working covariance matrix for  $Y_i$  is then

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} / \phi,$$

where  $A_i$  is an  $n_i \times n_i$  diagonal matrix with  $g(\mu_{it})$  as the  $t^{th}$  diagonal element.

Liang and Zeger (1986) extend the earlier quasi-likelihood equations in (4.1) and propose generalized estimating equations,

$$\sum_{i=1}^K S_i(\beta, \alpha) = \sum_{i=1}^K D_i' V_i^{-1} (y_i - \mu_i),$$

where  $S_i(\beta, \alpha)$  denotes the  $p$ -dimensional quasi-score function for the  $i^{th}$  subject,  $\mu_i = E(Y_i) = (\mu_{i1}, \dots, \mu_{in_i})' = (h(x_{i1}'\beta), \dots, h(x_{in_i}'\beta))'$ , and  $D_i = \partial \mu_i / \partial \beta$ . Note that these equations reduce to (4.1) in the case  $n_i=1$  for all  $i$ . More generally, the quasi-score function for the  $i^{th}$  subject,  $S_i(\beta, \alpha) = D_i' V_i^{-1} (y_i - \mu_i)$ , is equivalent to the estimating function suggested by Wedderburn (1974), except that here the  $V_i$ 's are functions of  $\alpha$  as well as  $\beta$ . Because  $D_i' V_i^{-1}$  does not

depend on the  $y_i$ 's, as long as  $E(Y_i - \mu_i) = 0$  (i.e., as long as the mean is correctly specified), the average of the quasi-score functions will converge to zero, ensuring that the equations have consistent roots.

Since the nuisance parameters  $\alpha$  (a parameter defined only for the working correlation) and  $\phi$  are unknown and the primary goal is inference for the regression parameters  $\beta$ , an obvious way to proceed is to replace  $\alpha$  in the expression for  $V_i$  by a consistent estimator,  $\tilde{\alpha}(y_1, \dots, y_K, \beta, \phi)$ , replace  $\phi$  in  $\tilde{\alpha}$  by a consistent estimator  $\tilde{\phi}(y_1, \dots, y_K, \beta)$ , and solve the altered estimating equations for  $\beta$ . Liang and Zeger (1986) adopt this approach and define  $\tilde{\beta}_G$ , the GEE estimate of  $\beta$ , as the solution of

$$\sum_{i=1}^K S_i\{\beta, \tilde{\alpha}(y_1, \dots, y_K, \beta, \tilde{\phi}(y_1, \dots, y_K, \beta))\} = 0.$$

Provided  $\tilde{\alpha}$  is consistent given  $(\beta, \phi)$ ,  $\tilde{\phi}$  is consistent given  $\beta$ , and mild regularity conditions hold, Liang and Zeger (1986, Theorem 2, page 16) show that as the number of subjects,  $K$ , becomes large,  $\tilde{\beta}_G$  consistently estimates  $\beta$ . They also show that under these same conditions,  $\sqrt{K}(\tilde{\beta}_G - \beta)$  is asymptotically multivariate Gaussian with zero mean and a covariance matrix given by

$$\begin{aligned} V_G &= \lim_{K \rightarrow \infty} K \left( \sum_{i=1}^K D_i' V_i^{-1} D_i \right)^{-1} \left[ \sum_{i=1}^K D_i' V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right] \left( \sum_{i=1}^K D_i' V_i^{-1} D_i \right)^{-1} \\ &= \lim_{K \rightarrow \infty} K V^{-1} V_* V^{-1}, \end{aligned} \quad (4.2)$$

where  $\text{cov}(Y_i)$  is the actual (unknown) covariance matrix for the response vector of the  $i^{\text{th}}$  individual. Although in general one might expect  $V_G$ , the asymptotic covariance matrix of  $\tilde{\beta}_G$ , to depend on which estimators are used for  $\alpha$  and  $\phi$ , the above result indicates that it does not depend upon how  $\alpha$  and  $\phi$  are estimated, as long as they are *consistently* estimated. The

covariance matrix  $V_G$  can be consistently estimated by replacing  $cov(Y_i)$  by  $(y_i - \tilde{\mu}_i)(y_i - \tilde{\mu}_i)'$ , where  $\tilde{\mu}_i = (h(x'_{i1}\tilde{\beta}), \dots, h(x'_{in_i}\tilde{\beta}))'$ , and  $\alpha$  and  $\beta$  by consistent estimates in (4.2).

The estimate of the regression coefficient,  $\tilde{\beta}_G$ , is obtained by iterating between solving the generalized estimating equations for  $\beta$  and consistently estimating  $\alpha$  and  $\phi$ . These two steps are repeated until convergence of all the estimates. To consistently estimate  $\alpha$  and  $\phi$ , Liang and Zeger (1986) recommend the use of the standardized residuals based on the updated estimate of  $\beta$ ,  $r_{it} = (y_{it} - \tilde{\mu}_{it}) ([\tilde{V}_i^{-1}]_{tt})^{-1/2}$ . Provided the fourth moments of the  $Y_{it}$ 's are finite, at a given iteration,  $\phi$  can be consistently estimated by the longitudinal analog of the Pearson  $X^2$  statistic

$$\phi^{-1} = \sum_{i=1}^K \frac{r_{it}^2}{N - p},$$

where  $N = \sum_{i=1}^K n_i$ . As in many quasi-likelihood problems (where the mean-variance relationship is specified by a scale parameter times a function of the mean), direct estimation of the parameter  $\phi$  may not be necessary for estimation of  $\beta$  and  $V_G$ . In particular, if the working correlation matrix is common to all subjects, i.e.,  $R_i(\alpha) = R(\alpha)$ , and the elements of this common working correlation matrix are all multiples of the parameter  $\alpha$ , an estimate of  $\phi$  need not be calculated. (Because  $\phi$  appears as a multiplicative constant in the formula for  $\tilde{\alpha}$ , it cancels in the calculation of  $V_i = A_i^{1/2} R(\alpha) A_i^{1/2}$ , so is not required for estimation of  $\beta$  and  $V_G$ .) This is the case for many choices of the working correlation matrix which are of practical interest. To estimate  $\alpha$ , we need to borrow strength across subjects. The specific estimator depends upon the choice of the  $R_i(\alpha)$ , but when  $R_i(\alpha) = R(\alpha)$  for  $i = 1, \dots, K$ , the general idea is to estimate



$\alpha$  by a simple function of

$$\tilde{R}_{uv} = \sum_{i=1}^K \frac{r_{iu}r_{iv}}{N-p} .$$

In practice, the true correlation structure of the response is rarely known. At best, we may only have vague ideas about the actual correlation structure present in the data. Nevertheless, we may wish to use this knowledge in order to obtain more efficient estimates of the regression parameters. The GEE approach allows us to incorporate this knowledge into the estimation process by specifying working correlation matrices we expect might be close to the true correlation structure. The GEE approach has a strong appeal because it does not require the working correlations to be correctly specified for asymptotically valid results.

For binary longitudinal data, the GEE approach is more flexible than the working likelihood approach of Section 3 because working correlations other than independence and stationary 1-dependent may be used, permitting increased efficiency of estimation in many different situations. Further, when a working correlation other than the identity matrix is specified in the GEE approach, observations belonging to different subjects need not share the same correlation structure for asymptotically valid results. This is not the case in the working likelihood approach. (Recall that for valid asymptotic results using the first-order working likelihood, all subjects are required to share a common first-order autocorrelation.)

However, a trade-off between efficiency and validity exists in the GEE approach, just as in the working likelihood approach. When specifying the working correlation  $R_i(\alpha)$  as something other than the independence working correlation in the GEE approach, the parameter  $\alpha$  must be consistently estimated from the data in order to obtain asymptotically valid parameter and standard error estimates. Because  $\alpha$  (a parameter defined only for the working correlation) must

be consistently estimated, it should describe some aspect of the true correlation structure, at least indirectly. An analogous situation arises when applying the working likelihood approach developed for binary data. When using the first-order working likelihood, the stationary 1-dependent working correlation is consistently estimated from the data provided all subjects have a common lag one autocorrelation.

One final point regarding the GEE approach is that it may be used for any multivariate outcomes with (univariate) marginal distributions for which the quasi-likelihood formulation is sensible. A wide variety of univariate distributions such as Gaussian, Poisson, binomial (binary), gamma, and inverse Gaussian, are included in this family. Liang and Zeger (1986) suggest the approach may even be extended to include multivariate multinomial and ordinal data by using formulations described by McCullagh and Nelder (1989) for these types of variables and choosing an independence working correlation.

## 4.2 GEE Analysis of the Binary Hummingbird Data

For the GEE analysis of the binary hummingbird data, the logit link with the binomial mean-variance relationship (up to an unknown scale factor  $\phi$ ) is used. If the independence working correlation is specified using this choice of link and mean-variance relationship, the scale parameter  $\phi$  does not enter into the estimation of  $\beta$  and  $V_G$ . Therefore, the GEE parameter and standard error estimates using the independence working correlation will be identical to those obtained in the earlier analysis based on the independence working likelihood. In contrast, fits for the first-order working likelihood may be different from the GEE fits based on a stationary

1-dependent working correlation (the quasi-likelihood analog of the first-order working likelihood) because the working likelihood models the correlation directly, whereas quasi-likelihood treats the correlation as a nuisance parameter, obtaining a method-of-moments estimator after first estimating the regression parameters at each iteration of a process which continues until convergence of all the parameter estimates.

In order to compare the fits obtained by the working likelihood approach and the GEE approach, in particular, the fits obtained by the first-order working likelihood and its quasi-likelihood analog, the binary hummingbird data was analyzed using the GEE approach with a stationary 1-dependent working correlation. Sex was included as a covariate in the GEE analysis because of the clear indications of its importance in the working likelihood analyses of Chapter 3. To examine the robustness of parameter and standard error estimates across working correlations, the final model obtained from the stationary 1-dependent working correlation was re-fit using several other working correlations.

#### 4.2.1 Description of the Model Reductions

For the GEE analysis including sex as a covariate, the same general reduction procedure outlined in Section 3.10.1 was applied to the initial model for the success probabilities on the first feeding foray described in Section 3.9. The model reduction path for the stationary 1-dependent working correlation is described by the following steps:

1. Check the quadratic terms.

- Can the coefficients for males and females be taken to be equal within all treatment groups? Yes; the p-value for this reduction is .81.

- Can the quadratic coefficients be eliminated? No; the p-value for the simultaneous elimination of all three quadratic coefficients is a marginal .12. Leave the quadratic terms for each treatment in the model for the time being.

2. Check the linear terms.

- Can the linear coefficients for males and females be set equal within each treatment group? Yes; the p-value for this reduction is .91.
- Can the linear coefficients for all treatment groups be set equal? Yes; the p-value for this reduction is .21. (Note that the fitted linear coefficients of Model 3 in Table 4.1 suggest different slopes might be identified if more birds were included in the experiment.)
- Can this common linear term be eliminated? No; the p-value for this reduction is less than .001.

3. Check the differential sex effects

- Can they be set equal across treatment groups? Yes; the p-value for this reduction is .48.
- Can this additive sex effect be eliminated? No; the p-value for this reduction is .006.

4. Examine the treatment main effects.

- Different quadratic terms for different treatment groups are still present in the model, so checking whether the treatment main effects can be set equal corresponds to checking whether the success probabilities for all three treatment groups can be set

to be equal at  $t = 0$  or at trial 90.5. The p-value for this reduction is .02, so it is not justified by the data.

5. For interpretability of the final model...

- Eliminate the quadratic terms for treatments 2 and 3. The p-value for this reduction is .79, justifying what is suggested by the lowess plots in Figures 3.16 and 3.17; the success probabilities of birds in treatments 2 and 3 can be adequately described by a linear trajectory.

Interestingly, the reduction path for this GEE analysis with a stationary 1-dependent working correlation matches that for the independence working likelihood (described in Section 3.10.2), rather than that for the first-order working likelihood (described in Section 3.10.3); the path leads to a final model with a quadratic term for the trajectory of birds in treatment 1.

Figure 4.1 plots the fitted final model for females (for the males, the curves within each treatment are simply shifted .77 units higher on the logit scale), while Table 4.1 summarizes the model reductions. Although the path of reductions is the same, the final model fits obtained from the independence working likelihood and the stationary one-dependent working correlation are slightly different. In particular, the fitted linear coefficient obtained here (.39, with standard error .10), is somewhat less than that obtained from the independence working likelihood (.48, with standard error .11; see Table 3.6), although these estimates are still reasonably close considering the standard errors. The estimated female-to-male differential sex effect (-.77, with standard error .28) is very similar to that obtained from the independence working likelihood (-.75, with standard error .28).

Table 4.1. Reductions for the Stationary 1-Dependent Working Correlation with Sex Included

Parameters	Model					
	1	2	3	4	5	6
$\mu_1$	.36(.03)	.36(.17)	.35(.16)	.37(.16)	.19(.26)	.19(.26)
$\mu_2$	-.59(.31)	-.52(.24)	-.50(.24)	-.53(.24)	-.35(.26)	-.34(.20)
$\mu_3$	.53(.75)	.60(.71)	.60(.70)	.57(.66)	.55(.48)	.46(.41)
$sex_{1,1}$	-1.03(.42)	-1.02(.42)	-1.02(.41)	-1.05(.42)	-.77(.28)	-.77(.28)
$sex_{2,1}$	-.25(.40)	-.37(.30)	-.39(.30)	-.44(.30)	set= $sex_{1,1}$	set= $sex_{1,1}$
$sex_{3,1}$	-.72(.76)	-.86(.66)	-.86(.64)	-.81(.59)	set= $sex_{1,1}$	set= $sex_{1,1}$
$\beta_{1,1,1}$	.22(.09)	.22(.09)	.18(.11)	.39(.10)	.39(.10)	.39(.10)
$\beta_{1,2,1}$	.55(.22)	.54(.21)	.42(.17)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,1}$	.55(.38)	.54(.38)	.63(.24)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,1,2}$	.41(.31)	.41(.31)	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,2,2}$	.46(.23)	.46(.24)	set= $\beta_{1,2,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{1,3,2}$	.74(.27)	.73(.27)	set= $\beta_{1,3,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$	set= $\beta_{1,1,1}$
$\beta_{2,1,1}$	-.21(.08)	-.21(.09)	-.21(.09)	-.23(.09)	-.22(.10)	-.22(.10)
$\beta_{2,2,1}$	-.04(.06)	.03(.10)	.03(.10)	.01(.11)	.01(.11)	-
$\beta_{2,3,1}$	-.16(.15)	-.09(.12)	-.10(.15)	-.10(.14)	-.09(.14)	-
$\beta_{2,1,2}$	-.22(.21)	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$	set= $\beta_{2,1,1}$
$\beta_{2,2,2}$	.10(.18)	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	set= $\beta_{2,2,1}$	-
$\beta_{2,3,2}$	-.008(.17)	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	set= $\beta_{2,3,1}$	-

Estimates of the lag one (working) autocorrelation at each stage of the model reductions are not included in Table 4.1. These are .12, .12, .13, .13, .13, and .13, for Models 1 through 6, respectively, essentially the same as the fitted lag one autocorrelations for the first-order working likelihood which were roughly .13 (s.e.  $\approx$  .03) for all model fits.

#### 4.2.2 Fits of the Final Model Obtained by Other Working Correlations

The final model obtained using the stationary 1-dependent working correlation was re-fit using a variety of working correlation structures; the results are summarized in Table 4.2. Qualitatively, the conclusions obtained with each working correlation are the same; all terms in the final model appear to be important. In general, the robust z-scores do not reflect as much association between the covariates and the response as the corresponding naive z-scores (not presented), which are up to 4 times larger. Results across working correlation structures are

remarkably similar, although the exchangeable working correlation fits a larger male-to-female sex differential and a larger (negative) main effect for treatment 2 (as well as larger standard errors for both these estimates) than the other working correlations. This suggests parameter and standard error estimates may be reasonably accurate, even with only 18 subjects.

Table 4.2. Robust z-Scores (Parameter Estimates) for the Binary Final Model\*

Working Corr.	$\mu_1$	$\mu_2$	$\mu_3$	$sex_{1,2}^{(1)}$	$\beta_{1,1,1}^{(2)}$	$\beta_{2,1,1}^{(3)}$
Independent	.68(.18)	-1.54(-.32)	1.09(.45)	-2.64(-.75)	4.31(.48)	-2.47(-.24)
Stat. 1-dep.	.73(.19)	-1.71(-.34)	1.12(.46)	-2.76(-.77)	3.89(.39)	-2.27(-.22)
Stat. 2-dep.	.73(.19)	-1.71(-.34)	1.12(.46)	-2.75(-.77)	3.89(.39)	-2.26(-.22)
Stat. 5-dep.	.76(.20)	-1.74(-.34)	1.13(.46)	-2.76(-.77)	3.98(.40)	-2.34(-.23)
Stat. 8-dep.	.75(.20)	-1.69(-.33)	1.14(.45)	-2.75(-.76)	4.03(.40)	-2.40(-.23)
AR-1	.73(.19)	-1.71(-.34)	1.12(.46)	-2.76(-.77)	3.89(.39)	-2.27(-.22)
AR-2	.73(.19)	-1.71(-.34)	1.12(.46)	-2.75(-.77)	3.90(.40)	-2.26(-.23)
AR-5	.76(.20)	-1.74(-.34)	1.13(.46)	-2.76(-.77)	4.01(.40)	-2.38(-.24)
Exchangeable	.55(.16)	-1.83(-.54)	1.14(.50)	-2.69(-.95)	3.51(.40)	-2.28(-.23)

\* Entries in table are z-score (parameter estimate).

1.  $sex_{1,1} = sex_{1,2} = sex_{1,3} = sex$ , the common differential sex effect.
2.  $\beta_{1,1,1} = \beta_{1,1,2} = \beta_{1,2,1} = \beta_{1,2,2} = \beta_{1,3,1} = \beta_{1,3,2}$ , the common linear coefficient.
3.  $\beta_{2,1,1} = \beta_{2,1,2}$ , the common quadratic coefficient for birds in treatment 1.

Note that among the working correlations for which results are presented in Table 4.2, the exchangeable structure is scientifically the least plausible candidate for the true structure because it models the correlations for responses distant in time to be the same as those for responses adjacent in time. Oddly enough, of all the working correlations presented, the naive and robust z-scores for the exchangeable structure are the closest together. Although this does not necessarily imply the exchangeable structure best approximates the true correlation structure, it is interesting to note that when the working correlation structure is specified to be stationary 8-dependent, the estimated lag one through eight working correlations are all

about .12 (slightly larger than .08, the fitted exchangeable working correlation), suggesting correlations for responses relatively far apart in time are not too different from correlations for responses adjacent in time.

### 4.3 GEE Analysis of the Binomial Hummingbird Data

To illustrate the versatility of the GEE approach, we turn now to an analysis of the “binomial” data. Recall that the maximum number of successful feeding forays on any given trial in the experiment was 11, while the number of unsuccessful feeding attempts was unrestricted. Because the total number of feeding forays in a trial is random, the “binomial” response, i.e., the number of successful feeding forays on each trial, cannot have a binomial distribution. Nevertheless, we assume the mean-variance relationship for the number of successful feeding forays in a trial is that of a binomial random variable and interpret the fitted success probability for a particular trial as a rough average of the success probabilities for all the (possibly dependent) forays within that trial, rather than as a success probability for a binomial experiment (i.e., the trial) in which the total number of (independent) forays is non-random.

The binomial data is not as meaningful to the investigator as the binary data because hummingbird researchers believe the first feeding forays are the most accurate measure of a bird’s expectations about feeder profitability. Many of the birds which are unsuccessful on the first foray have not yet associated the light cue with the feeder. For these birds, any forays subsequent to the initial unsuccessful attempt will reflect their strategy for finding food after failing the first time. This strategy varies with each individual; so, on a particular trial, a bird’s binomial response (which summarizes information on all of its forays during that trial)



is expected to convey less information about the bird's feeding expectations than its binary response at the first feeding foray. Even though analysis of the binomial hummingbird data may not be as scientifically relevant as the earlier analysis of the binary data, it is worthwhile as an example of the GEE approach applied to non-binary data.

### 4.3.1 Exploring the Binomial Hummingbird Data

A series of plots was used to determine appropriate parametric forms to be incorporated into modelling of the temporal response patterns.

To check if a day effect was apparent, the logits of the individual responses ( $\text{logit}(y) = \log(\frac{y+c}{n-y+c})$ ,  $c = .5$ , where  $y$  is the number of *successful* forays out of  $n$  total forays in the trial for a particular bird) were averaged over all birds of the same sex within each treatment group for successive blocks of 10 trials and then plotted (see Figures 4.2-4.4). Because there were no systematic discontinuities at those trials corresponding to new days (trials 61 and 121), there seemed to be no need to model a day effect. To get a better overall picture of the binomial data, the logits of responses within each treatment-sex group were smoothed using robust locally weighted linear least squares ("lowess") with windows of 30 and 60 points.

Figures 4.5 and 4.6 show the results of this "lowess" smoothing. Comparing across treatment groups, the success probabilities of birds in treatment 3 (FT) are generally higher than those of birds in the other treatments, suggesting the presence of treatment main effects. Success probabilities of males in treatment 1 (NT) appear to be almost as high as those of males in treatment 3, and roughly the same as those of females in treatment 3. In contrast, success probabilities of females in treatment 1 appear to be lower than those of females in treatment

3 and closer to those of females in treatment 2 (PT). In general, males appear to have higher probabilities of success than females, especially for birds in treatment 1. This suggests the presence of sex-by-treatment interaction in this data. The success probabilities of all birds appear to increase roughly linearly over time, with the slope for females in treatment 1 being somewhat less than the slopes for the other birds. This suggests a three-way interaction of sex, treatment group and time may be present in this data.

To study individual-to-individual variation within each treatment, the logits of individual results at each time point were smoothed using lowess, with a window of 60 consecutive time points; see Figures 4.7-4.9. Peculiarities of these “lowess” fits near the endpoints occur because only the individual logits, rather than logits for all the individuals of a certain sex within a treatment, have been smoothed. The individual logits are more likely to have clusters of extreme values with few moderate values to counteract their effect at the endpoints, leading to fits for an individual which may not accurately describe the patterns in the individual’s data near the endpoints. In the interior of the the data sets, the plots do not show any obvious change in response variability over time for any of the treatments. The plots also suggest that, at any given time, the variability of responses in treatments 1 (NT) and 2 (PT) is roughly the same, but that responses in treatment 3 (FT) are more variable at than responses in treatments 1 and 2.

### 4.3.2 Model Fitting for the Binomial Hummingbird Data

These exploratory plots suggest the transformed probability of success may be modelled, initially, as

$$\text{logit } \pi_{i,t,s} = \mu_i + \text{sex}_{i,s} + \beta_{1,i,s}t + \beta_{2,i,s}t^2 \quad i = 1, \dots, 3, \quad s = 1, 2, \quad T = 1, \dots, 180, \quad t = \frac{T - \bar{T}}{\text{s.d.}(T)}$$

where  $\mu_i$  may be interpreted as the main effect of the  $i^{\text{th}}$  treatment group,  $\text{sex}_{i,1}$  is the female-to-male differential sex effect for the  $i^{\text{th}}$  treatment group, and  $\text{sex}_{i,2} \stackrel{\text{set}}{=} 0$  (for males in the  $i^{\text{th}}$  treatment group). As in the binary analysis, a quadratic learning trajectory (in the logit scale) has been modelled to allow for some departure from linearity in the responses. In addition, the time covariate has been standardized to orthogonalize the highly collinear linear and quadratic time covariates. The female-to-male differential sex effect is permitted to vary across treatments to allow for treatment-by-sex interaction. Similarly, the three-way interaction of sex, treatment and time is incorporated because different linear and quadratic coefficients for each sex and each treatment group are permitted.

Because the stationary 1-dependent working correlation models *some* correlation in the responses, it is expected to provide more efficient estimates than the independence working correlation. Model reductions have therefore been carried out using the *stationary 1-dependent* working correlation and are summarized in Table 4.3. The path of these reductions is described by the following steps:

1. Check the quadratic terms.

- Can the coefficients for males and females be taken to be equal within all treatment groups? Yes; the p-value for this reduction is .65. Set these coefficients to be equal.

- Can the quadratic coefficients in the resulting model be eliminated? Yes; the p-value for the simultaneous elimination of all three quadratic coefficients is .58. Remove the quadratic terms from the model.

2. Check the linear terms in the resulting model.

- Can the linear coefficients for males and females be set equal within each treatment group? No; the p-value for this reduction is less than .001. From the coefficients and standard errors of Model 3 in Table 4.3, it appears that essentially all the evidence against this reduction comes from the responses for birds in treatment 1.
- Can the linear coefficients for all treatment groups be set equal within sexes? Yes; the p-value for this reduction is .25. To simplify the model, set two common linear slopes: one for females and one for males.
- Can the resulting linear terms for females or males be eliminated? For the females, yes; the p-value for the reduction is .67, so remove this term from the model. For the males, no; the p-value for this reduction is less than .001.

3. Check the differential sex effects.

- Can they be taken to be the same for all treatment groups? Yes; the p-value for this reduction is .47.
- Can this the common sex effect be eliminated? No; the p-value for this reduction is .05.

4. Examine the treatment main effects.

- Can they be taken to be equal? No; the p-value for this reduction is .02.

For interpretability of the final model, a further reduction might set the main effects for treatments 1 and 2 to be equal, as suggested by standard errors of these estimates under Model 6 (see Table 4.3). The p-value for this reduction is .63, indicating the reduction is permissible. The resulting model (not included in Table 4.3) leads to a simpler representation of the learning trajectories because, within each sex, birds in treatments 1 and 2 are then modelled as having the same learning trajectory in the logit scale.

Table 4.3. Model Reductions for the Binomial Data

Parameter	Model					
	1	2	3	4	5	6
$\mu_1$	1.08(.15)	1.08(.08)	1.07(.01)	1.08(.02)	1.08(.02)	1.12(.18)
$\mu_2$	1.08(.14)	1.12(.12)	1.12(.12)	1.11(.13)	1.12(.13)	1.24(.17)
$\mu_3$	2.22(.71)	2.36(.66)	2.27(.63)	2.24(.57)	2.25(.57)	1.91(.30)
$sex_{1,1}$	-.35(.49)	-.35(.36)	-.35(.36)	-.38(.38)	-.37(.37)	-.43(.22)
$sex_{2,1}$	-.12(.20)	-.20(.23)	-.20(.23)	-.19(.23)	-.18(.23)	set= $sex_{1,1}$
$sex_{3,1}$	-.76(.71)	-.95(.63)	-.96(.64)	-.93(.57)	-.93(.57)	set= $sex_{1,1}$
$\beta_{1,1,1}$	-.06(.09)	-.06(.09)	-.06(.09)	.06(.09)	-	-
$\beta_{1,2,1}$	.23(.09)	.23(.09)	.23(.10)	set= $\beta_{1,1,1}$	-	-
$\beta_{1,3,1}$	.10(.24)	.10(.25)	.10(.26)	set= $\beta_{1,1,1}$	-	-
$\beta_{1,1,2}$	.31(.02)	.31(.02)	.31(.01)	.37(.09)	.37(.09)	.37(.09)
$\beta_{1,2,2}$	.39(.18)	.39(.18)	.39(.18)	set= $\beta_{1,1,2}$	set= $\beta_{1,1,2}$	set= $\beta_{1,1,2}$
$\beta_{1,3,2}$	.48(.30)	.45(.30)	.47(.33)	set= $\beta_{1,1,2}$	set= $\beta_{1,1,2}$	set= $\beta_{1,1,2}$
$\beta_{2,1,1}$	-.01(.11)	-.01(.09)	-	-	-	-
$\beta_{2,2,1}$	-.04(.08)	-.01(.08)	-	-	-	-
$\beta_{2,3,1}$	-.14(.04)	-.09(.07)	-	-	-	-
$\beta_{2,1,2}$	-.01(.17)	set= $\beta_{2,1,1}$	-	-	-	-
$\beta_{2,2,2}$	.04(.15)	set= $\beta_{2,2,1}$	-	-	-	-
$\beta_{2,3,2}$	.05(.16)	set= $\beta_{2,3,1}$	-	-	-	-

The fits for the initial model with the stationary 1-dependent working correlation given in Table 4.3 are plotted in Figures 4.10-4.12; similar fits are obtained with both the independence and the AR-1 working correlations. These fits do not reflect some features of the lowess plots in Figures 4.5 and 4.6. For example, the lowess plots suggest the male-to-female differential sex effect should be largest for birds in treatment 1 (NT), yet the GEE fits for the initial model

suggest this effect is largest for birds in treatment 3 (FT). In addition, the lowess plots suggest success probabilities of treatment 1 males and treatment 3 females and males are higher than those of birds in treatment 2 (PT) on the logit scale, but in the fits for the initial model, only the success probabilities (on the logit scale) of treatment 3 males are clearly higher than those of birds in treatment 2. Finally, the lowess plots suggest success probabilities for females in treatment 1 increase slightly over time, whereas the fits for the initial model indicate they decrease slightly over time. The initial fit for females in treatment 1 is probably overly influenced by the female with a large dip in success probabilities near trial 140; see Figure 4.7. Discrepancies such as these may occur because, unlike the lowess fits for intermediate trials, the GEE fits at intermediate trials may be influenced by extreme data at the beginning and end of the experiment. Runs of extreme data are more likely to occur for a particular individual at the beginning or end of the experiment and may have high leverage because the number of subjects in the experiment (18) is very *small* relative to the number of trials (180). Such discrepancies between lowess and initial model fits may disappear once data on more birds are collected under these conditions.

A notable peculiarity of the GEE results in Table 4.3 is the large drop (from .08 to .01) in the standard error of the estimated effect for treatment 1 compared to those for the other treatments (from .12 to .12, and from .66 to .63, for treatments 2 and 3, respectively) when the quadratic terms are removed from Model 2. The corresponding naive standard error estimate stays relatively constant for this model reduction, dropping from .11 to .09. Another jump (from .02 to .18) in the standard error of the estimated effect for treatment 1 occurs when the differential sex effects for the treatment groups are set equal (Model 5 to Model 6 in the table).

Again, the corresponding naive standard error estimate remains relatively constant (dropping from .09 to .07). These results suggest the robust standard error estimate for the main effect of treatment 1 may not accurately reflect the true standard error; perhaps there are too few birds (18) in the experiment for the asymptotic approximation to be adequate. If the asymptotic approximation to the covariance matrix of the parameter estimates is inaccurate, the validity of all the model reductions (and hence the final model) is questionable. Somewhat reassuringly, the values of standard error estimates for the other parameters are fairly constant.

Figures 4.13 and 4.14 display the final model fits for the binomial hummingbird data obtained with the stationary 1-dependent working correlation. Model reductions are intended to lead to a parsimonious final model which describes essential features of learning trajectories for hummingbirds in general, rather than those for the eighteen hummingbirds in this experiment. Hence, some features evident in the lowess plots are absent in the final model fits, presumably because of the large variability of the responses. For example, the variability of female success probabilities, primarily in treatments 1 and 3 (see Figures 4.7-4.9), permits the fit of a common linear slope for females in Model 4 of Table 4.3. This feature of the fits is not consistent with the lowess plots, which suggest success probabilities for females in treatments 2 and 3 increase more quickly over time on the logit scale than those for females in treatment 1. When averaging all the female curves, those for the four females in treatment 1 dominate, leading to a negligible fitted slope which is eliminated in Model 5 of Table 4.3. Because both a common (negligible) female and a common male slope are modelled, a common differential sex effect is permissible in Model 6. This feature of the fits is also inconsistent with the lowess plots, which suggest a higher male-to-female differential sex effect in treatment 1. Specifying a common differential

sex effect affects the placement of the fitted curve for males in treatment 1 relative to the fitted curves for males in other treatments. For this reason, the fitted curve for males in treatment 1 is lower than the curve suggested by the lowess plots.

In order to identify which of these patterns more accurately reflects the true response probabilities of hummingbirds under these experimental conditions, more birds need to be included in the experiment. However, one feature of the binomial data suggested by the lowess plots is confirmed by the GEE analysis: the (binomial) success probabilities for birds in treatment 3 are higher, in general, than those for birds in treatments 1 and 2.

#### 4.3.3 Fits of the Final Model Obtained by Other Working Correlations

To examine the robustness of parameter and standard error estimates across working correlations, the final model was re-fit using a variety of working correlation structures; results are summarized in Table 4.4.

Table 4.4. Robust z-Score (Parameter Estimate) for the Binomial Final Model\*

Working Corr.	$\mu_1$	$\mu_2$	$\mu_3$	$sex_{1,1}^{(1)}$	$\beta_{1,1,2}^{(2)}$
Independent	6.03(1.08)	6.48(1.13)	5.76(1.81)	-1.81(-.40)	3.95(.36)
Stat. 1-dep.	6.23(1.12)	7.33(1.24)	6.32(1.91)	-1.94(-.43)	4.12(.37)
Stat. 2-dep.	6.44(1.18)	8.56(1.43)	7.10(2.04)	-2.13(-.47)	4.06(.37)
Stat. 3-dep.	5.82(1.31)	9.86(1.75)	7.39(2.23)	-2.16(-.54)	3.41(.38)
AR-1	6.27(1.12)	7.49(1.26)	6.42(1.92)	-1.97(-.43)	4.13(.37)
AR-2	6.37(1.22)	9.23(1.57)	7.46(2.15)	-2.19(-.50)	3.98(.38)

\* Entries in table are z-score (parameter estimate).

1.  $sex_{1,1} = sex_{1,2} = sex_{1,3} = sex$ , the common differential sex effect.

2.  $\beta_{1,1,2} = \beta_{1,2,2} = \beta_{1,3,2}$ , the common linear coefficient for males.



The robust z-scores for different working correlations are reasonably close, leading to qualitatively similar conclusions: all terms in the final model are important. For the working correlations used in the analysis of this binomial hummingbird data, the robust z-scores do not reflect as much association between the covariates and the response as the naive z-scores (not presented), which are up to 5 times larger. When naive and robust z-scores are compared across the working correlations used in this analysis, the differences between naive and robust z-scores for the stationary 3-dependent working correlation are the smallest. This, together with the fact that the stationary 3-dependent correlation structure is scientifically plausible, might suggest it approximates the true correlation better than the other working correlations presented in the table. For the stationary 3-dependent working correlation, estimates for the first, second and third lag autocorrelations are .24, .26, and .26, respectively, suggesting non-negligible correlation may indeed be present in the data.

Note that estimates of all the effects, and, in particular, the treatment main effects, increase as more correlation structure is incorporated into the working assumptions. In fact, when solving the generalized estimating equations iteratively using the stationary 5-dependent, stationary 8-dependent, AR-3, and exchangeable working correlations (not included in the table), successive estimates of the treatment main effects become very large and do not converge. This is in contrast to the binary data on the first feeding foray analyzed in the previous section, where convergence was achieved with all of these working correlations. For the binomial hummingbird data, lack of convergence with more detailed working correlations may be related to the large number of trials in the experiment (180) relative to the small number of subjects (18).

This same feature of the data set (180 trials on only 18 subjects) may explain discrepancies between the lowess fits and the GEE fits for the full model in the binomial analysis: because the number of trials (180) is large relative to the number of subjects (18), the GEE fits at intermediate trials may be overly influenced by a run of extreme data for a particular individual at the beginning or end of the experiment. This problem does not arise in the binary analysis because there are no extreme responses for binary data (binary data takes on only two values). Hence, the GEE and lowess fits for the binary data are reasonably consistent. The binomial analysis seems to illustrate that when a long series of observations is made on relatively few subjects, the GEE approach should be used with caution since it relies on combining strength across subjects to

- (i) obtain accurate estimates of the regression parameters; and
- (ii) compensate for making no strong assumptions about the correlation structure of each subject's response.

## Chapter 5

### Conclusion

#### 5.1 Discussion of Results

Exploratory plots of the binary responses for each bird on the first feeding foray (Figures 3.7-3.9) do not reveal the step-like pattern associated with sudden awareness of the spatial association between light cue and feeder expected by the experimenter. These exploratory plots suggest a more gradual increase in individual success probabilities over time, but with considerable fluctuation throughout the experiment.

Evaluating the performance of birds across treatment groups by the three methods described in this thesis, both the working likelihood and GEE analysis indicate that, as expected, birds in treatment 3 (FT) have the highest success probabilities throughout the entire experiment. The univariate ANOVA, however, does not detect an important difference in success probabilities across treatment groups.

Unlike the ANOVA, both the working likelihood and GEE approaches permit modelling of the response directly as a function of time. When the learning trajectories of birds are modelled by a quadratic trend in the logit scale using these approaches, birds in treatments 2 (PT) and 3 (FT) appear to improve linearly at the same rate. However, success probabilities for birds in treatment 2 appear to start out (and remain) lower than those for birds in treatment 3

throughout the entire experiment. Unlike the linear trajectories (in the logit scale) for birds in treatments 2 and 3, the trajectory for birds in treatment 1 (NT) appears to be quadratic, with success probabilities at the beginning of the experiment closer to those for birds in treatment 3 than those for birds in treatment 1. However, at the end of the experiment, success probabilities of birds in treatment 1 appear to drop closer to and perhaps below those of birds in treatment 2. Contrary to the expectations of the investigator, birds in treatment 2 (PT) appear to have lower success probabilities than the other birds except, perhaps, at the end of the experiment, where their success probabilities may be greater than or equal to those for birds in treatment 1. One possible explanation proposed by the investigator is that the orange Dymo tape for the partially taped feeding array creates three *horizontal* bands which may interfere with the intended *vertical* pattern to be perceived by the birds in treatment 2 (see Figure 1.1).

Contrary to the accepted wisdom among researchers in hummingbird learning, both the working likelihood and GEE analyses reveal that sex is important in describing the probabilities of success on the first feeding foray for the hummingbirds in this experiment. Since the effect of sex appears to be additive for the binary data, the fitted models indicate the odds of success for males are more than twice the odds of success for females. (For the fit resulting from the independence working likelihood, the odds for males are about 2.1 the odds for females, with a standard error of roughly 0.6.) One possible explanation for the importance of sex is the higher body-weight to wing-span ratio of males. Because males must work harder to obtain their food, they may be more likely to succeed than females. Note that in any ANOVA of the results in this experiment, sex would be difficult to incorporate because of the imbalance in the number of males and females in treatment 1 (NT).

Although not as scientifically relevant as the analysis of the binary data for the first feeding foray, the analysis of the binomial data for all the feeding forays in a trial confirms that birds in treatment 3 (FT) have higher success probabilities, in general, than birds in the other treatment groups. Again, sex appears to be an important covariate, although its effect on the binomial response (where sex appears to interact with time) seems to be more complicated than its effect on the binary response (where the effect of sex is simply additive).

Because both the working likelihood and GEE approaches model success probabilities as a function of time while taking into account the (unknown) correlation present in the data, they enable much more information to be extracted from the hummingbird data set than a more traditional analysis-of-variance. These methodologies therefore provide useful and powerful tools for researchers in this subject area.

## 5.2 Comparison of Methods

Longitudinal data consists of repeated measurements over time for each of many subjects. Hence, the response of each subject may be considered a random vector with possibly correlated elements corresponding to measurements at each time of observation. If the response vector is normally distributed, classical methods permit modelling of the joint distribution as a function of the covariates. For example, when sex is not included as a covariate in the analysis of the binary hummingbird data, MANOVA permits testing of parallelism and equality of treatment means, provided the number of subjects is greater than the number of time points. MANOVA may be carried out when the number of subjects is less than the number of time points if data are blocked over time points, but this leads to a loss of information. Modelling of responses which

are not normally distributed is more difficult. Transformations which stabilize variance for univariate responses do not necessarily stabilize covariance matrices for multivariate responses. Hence, equality of covariance matrices, a necessary assumption for valid F-tests in analysis-of-variance, may not hold.

Instead of attempting to transform non-normal data into forms more suitable for analysis-of-variance, a more natural approach might be to model the marginal expectations of the responses as functions of the covariates while taking into account the (unknown) correlation present in the data of each subject. For example, with binary longitudinal data, logistic regression might be extended in the manner described by Zeger, Liang and Self (1985). Rather than the “actual” likelihood, this approach uses a “working” likelihood based on simple assumptions about the correlation within each subject’s data to generate estimating equations for the regression parameters. Under mild assumptions about the true correlation structure, these estimating equations lead to consistent estimates of the regression parameters and their standard errors.

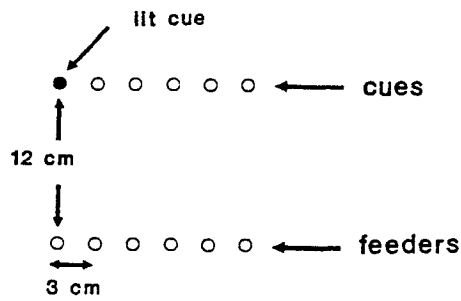
The approach described by Zeger, Liang, and Self (1985) was developed for short binary series. Because detailed models for the change in response over time are unnecessary in this context, the authors considered time-independent covariates only. In the case of the hummingbird learning experiments, where birds are expected to improve over a long series of measurements, the probability of success should be modelled as a function of time. In this thesis, the results of Zeger, Liang, and Self (1985) have been extended to allow this. The resulting “working likelihood” approach may be applied to any binary series with time-dependent covariates. One weakness of this approach is that when correlation is explicitly modelled (in order to increase the efficiency of estimation), asymptotic results only hold under certain assumptions on the

true correlation structure. Other than the independence working likelihood, only the first-order Markov working likelihood, which models stationary one-dependent correlation in the responses, has been considered in this thesis.

Unlike the working likelihood approach, the “generalized estimating equation” or GEE approach of Liang and Zeger (1986) and Zeger and Liang (1986) requires no strong assumptions about the true correlation structure when modelling correlation explicitly. The GEE approach extends quasi-likelihood theory to the case where the response is observed repeatedly over time for each subject, and uses a “working” correlation for these responses to arrive at estimating equations for the regression parameters. Hence, the GEE approach is applicable to any longitudinal response with univariate marginal distributions for which the quasi-likelihood formulation is sensible. This includes binary, binomial, count, and continuous longitudinal responses. The GEE approach thus provides a unified methodology for the analysis of longitudinal data.

Future work investigating the extension of the GEE approach to the case of ordinal longitudinal data, and to the case of data with censored or missing observations would be valuable. For instance, the approach could be used to analyze the results of an ongoing clinical trial, in which U.B.C. is one of the participating centres, concerning the effects of beta-interferon on the progression of multiple sclerosis. In this clinical trial, several outcome variables are observed on each patient repeatedly over time, including the so-called “Kurtzke score”, an ordinal response variable measuring the severity of disease.

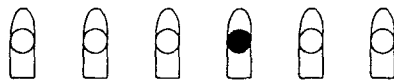
Figure 1.1. Feeding Arrays for the Hummingbird Experiment



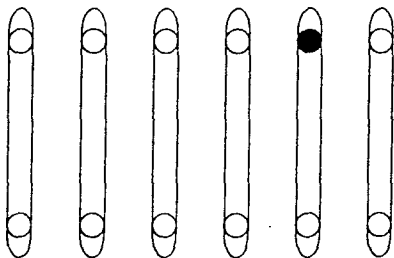
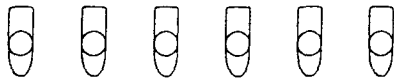
Key



No Tape Treatment



Partial Tape Treatment



Full Tape Treatment



Figure 2.1. Mean Treatment (Transformed) Response versus Measurement for Blocks of 15

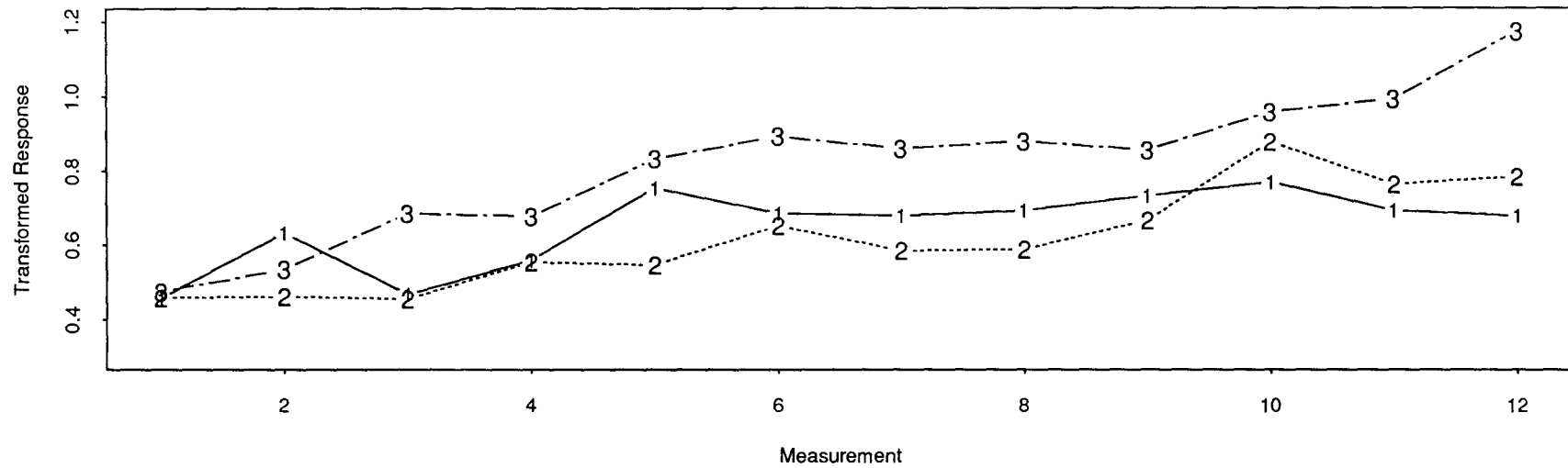


Figure 2.2. Mean Treatment (Transformed) Response versus Measurement for Blocks of 10

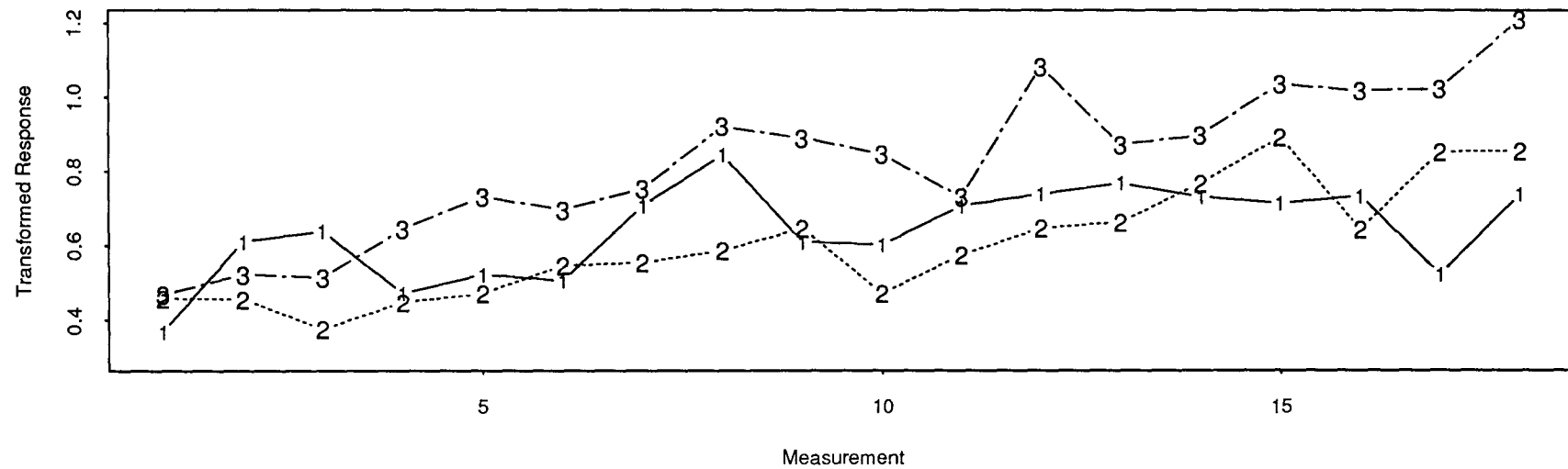


Figure 2.3. Estimated Variances versus Measurement for Blocks of 15

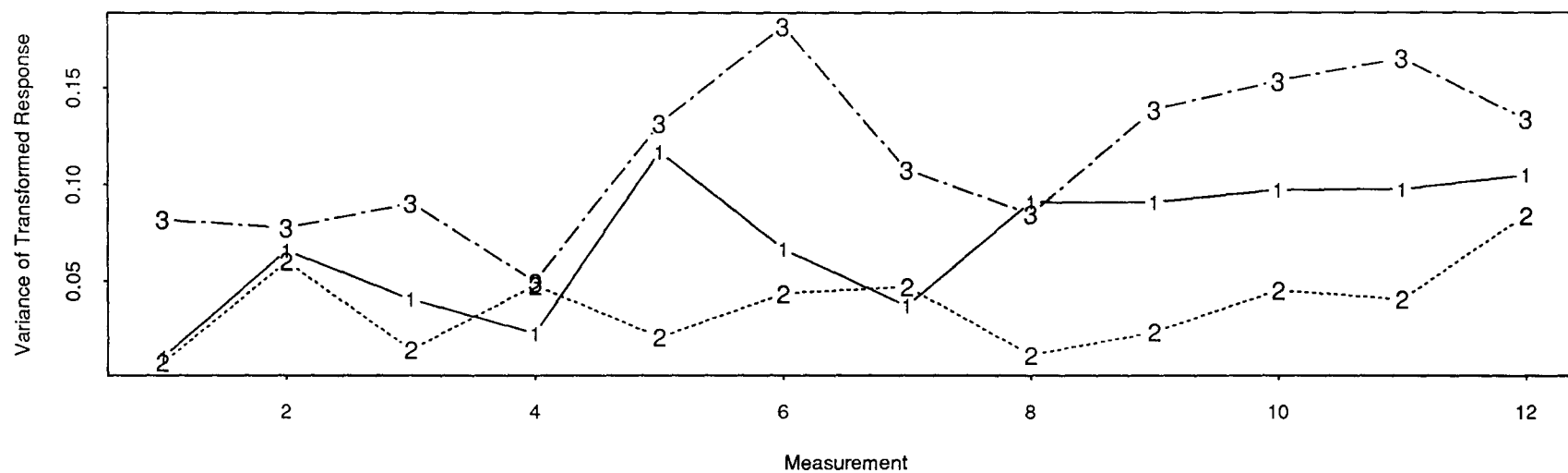


Figure 2.4. Estimated Variances versus Measurement for Blocks of 10

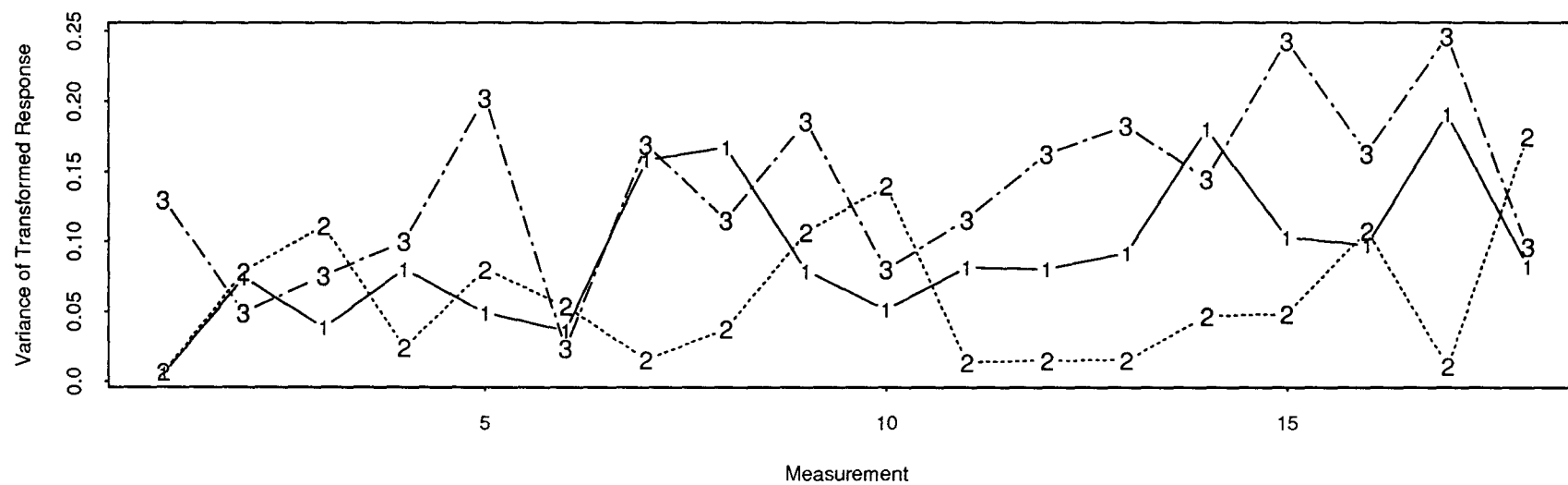


Figure 3.1. Average Treatment Response in the Logit Scale for Blocks of 3 Trials

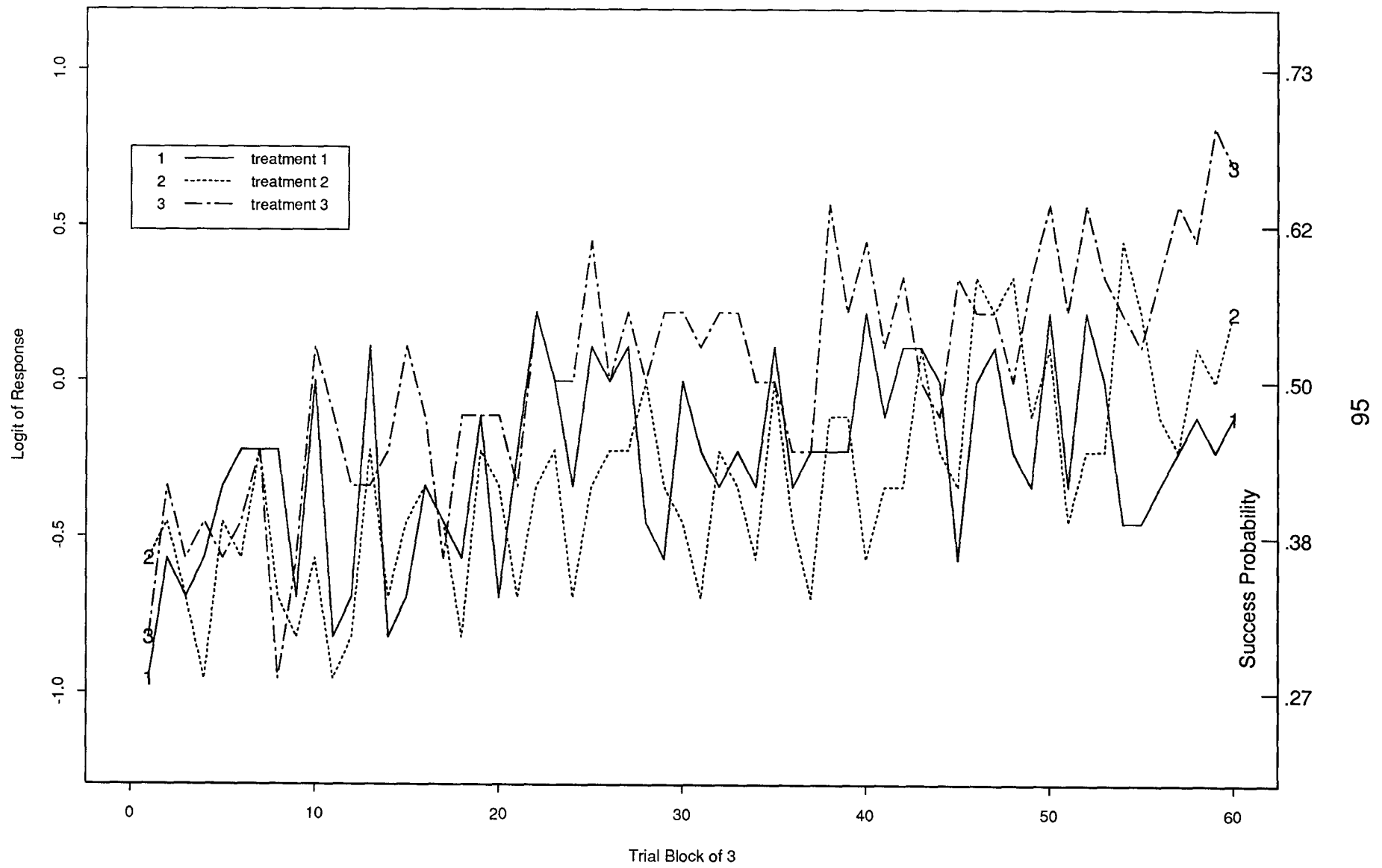


Figure 3.2. Average Treatment Response in the Logit Scale for Blocks of 5 Trials

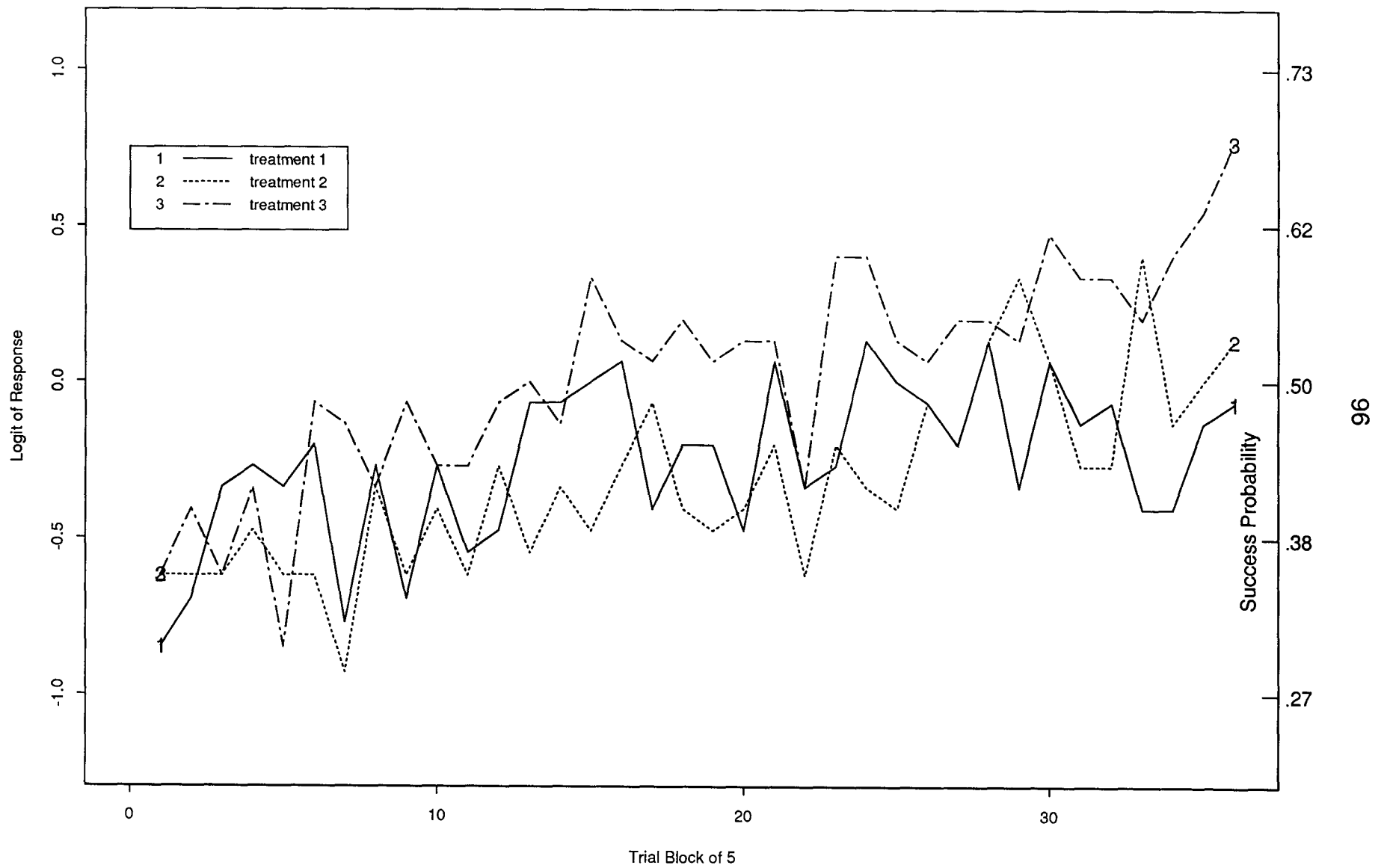


Figure 3.3. Average Treatment Response in the Logit Scale for Blocks of 10 Trials

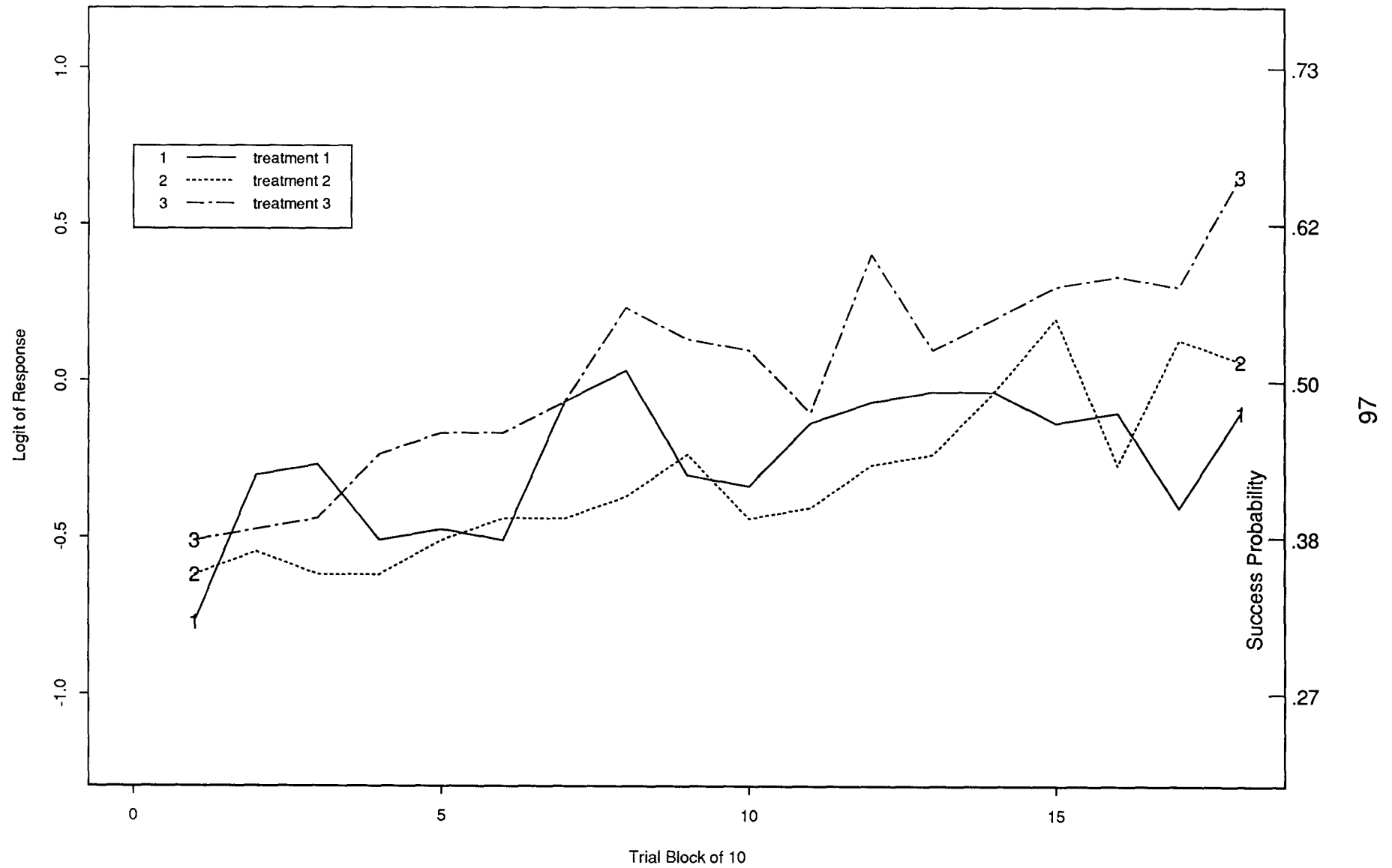


Figure 3.4. Smoothed Logit of Average Treatment Response versus Trial (Window=10)

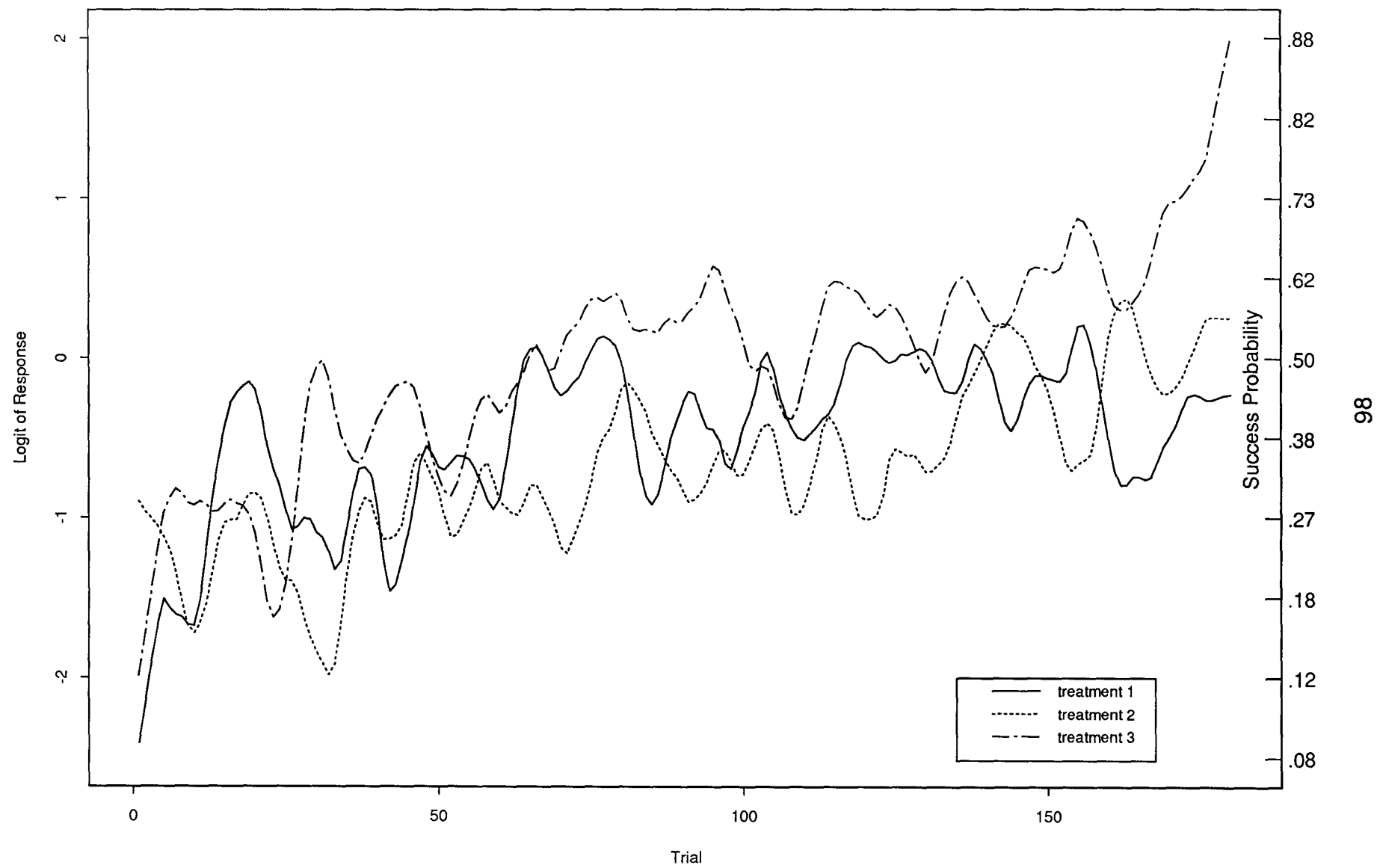


Figure 3.5. Smoothed Logit of Average Treatment Response versus Trial (Window = 20 Trials)

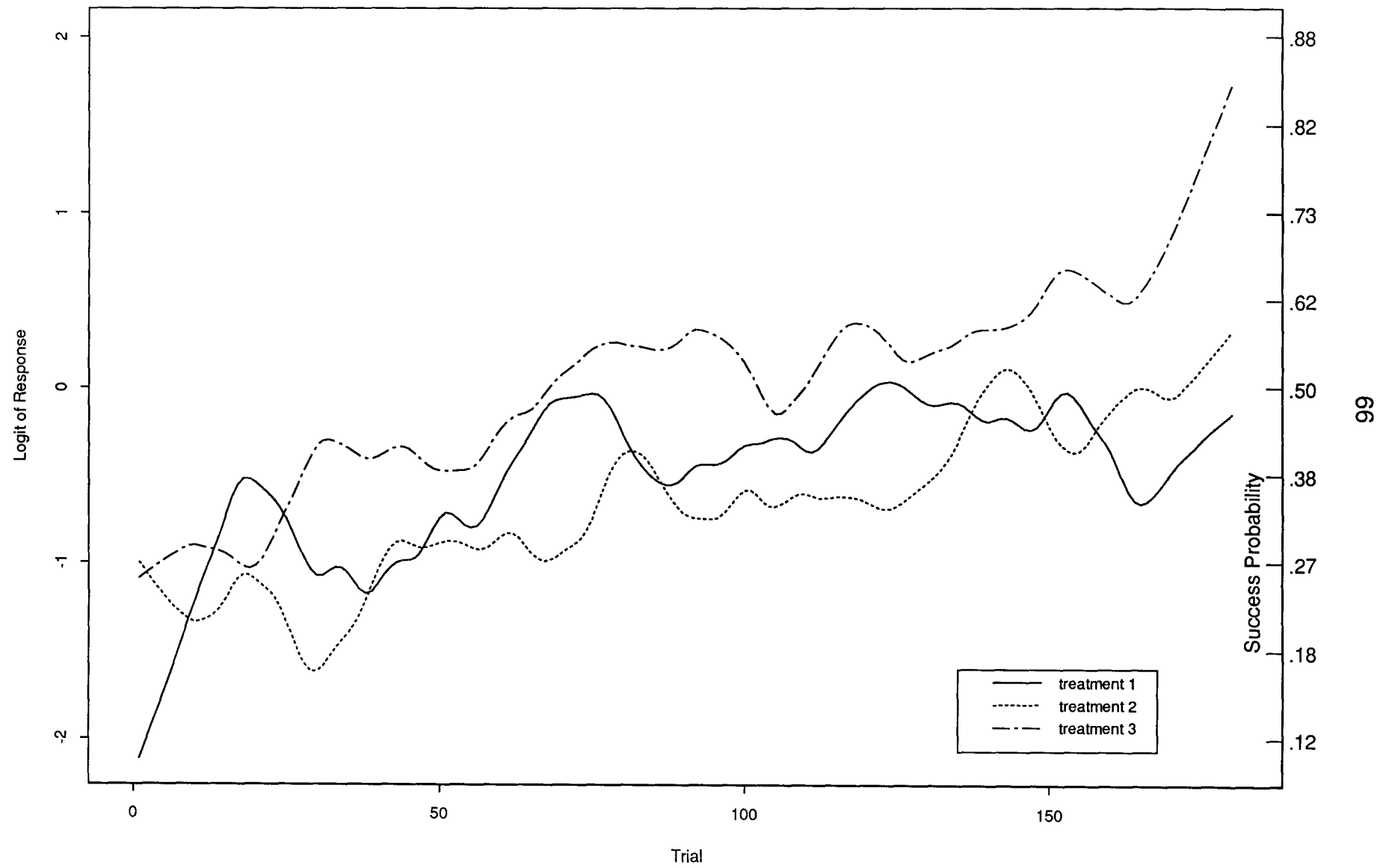


Figure 3.6. Smoothed Logit of Average Treatment Response versus Trial (Window = 30)

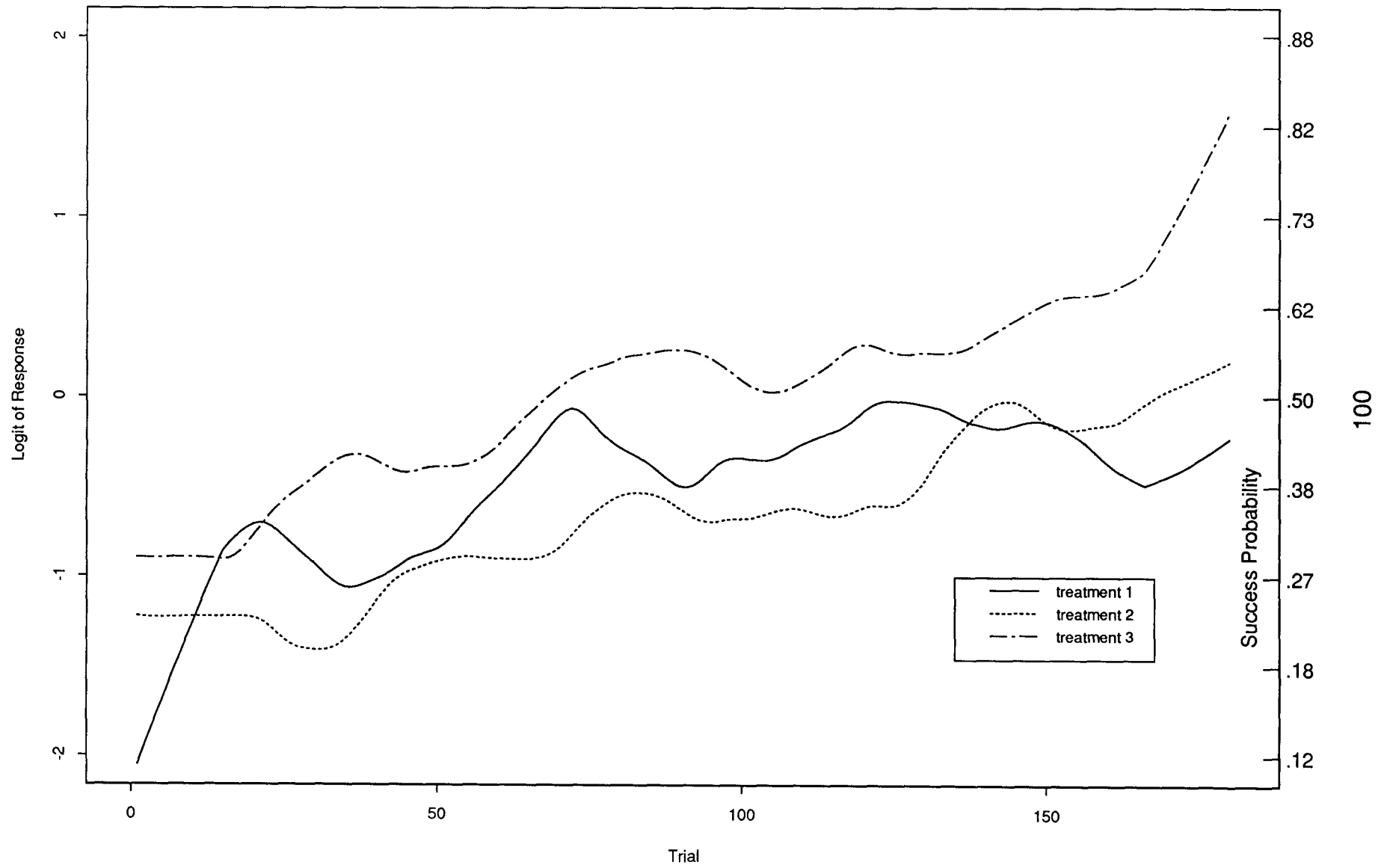




Figure 3.7. Smoothed Individual Logits for Treatment 1

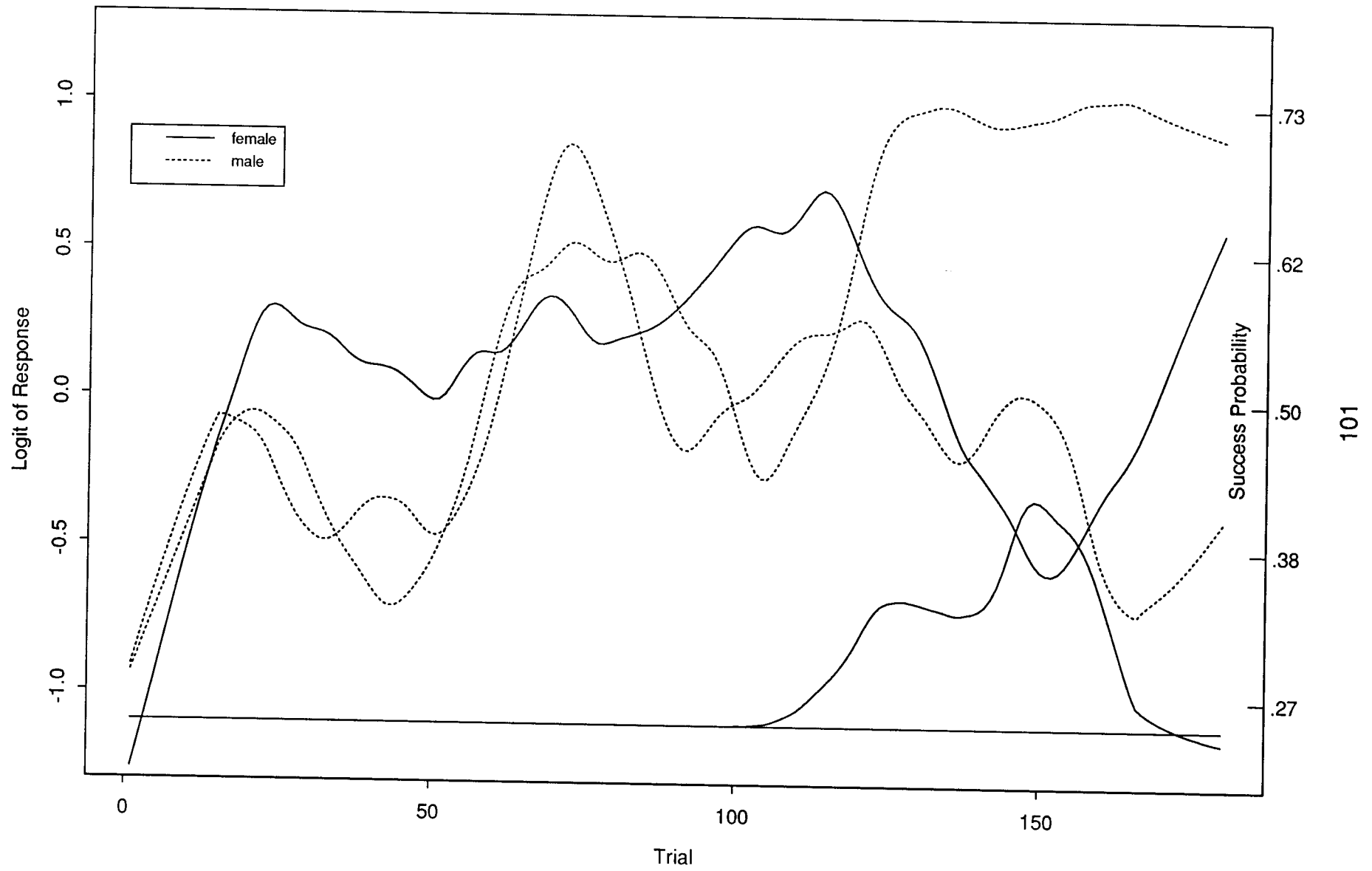


Figure 3.8. Smoothed Individual Logits for Treatment 2

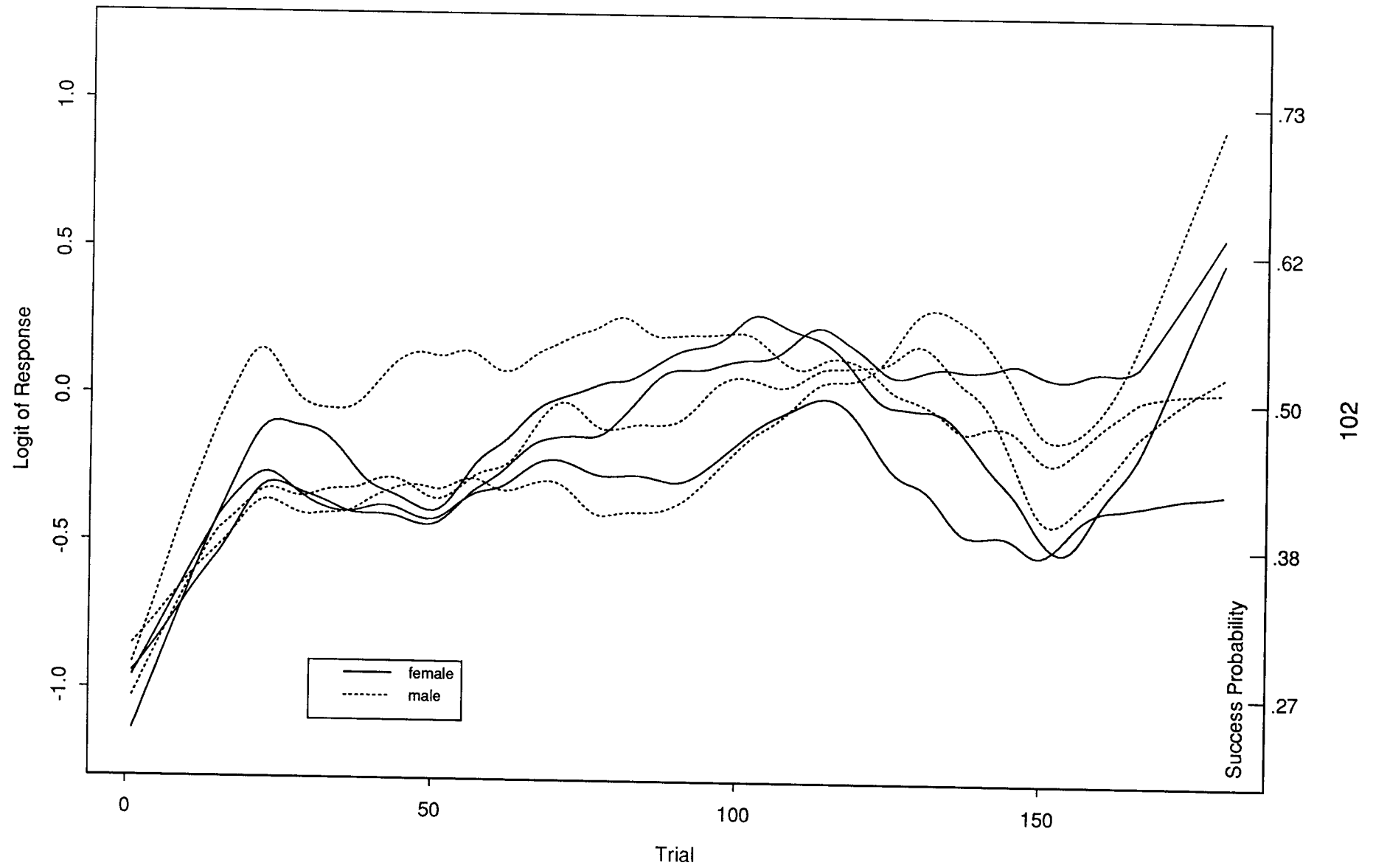


Figure 3.9. Smoothed Individual Logits for Treatment 3

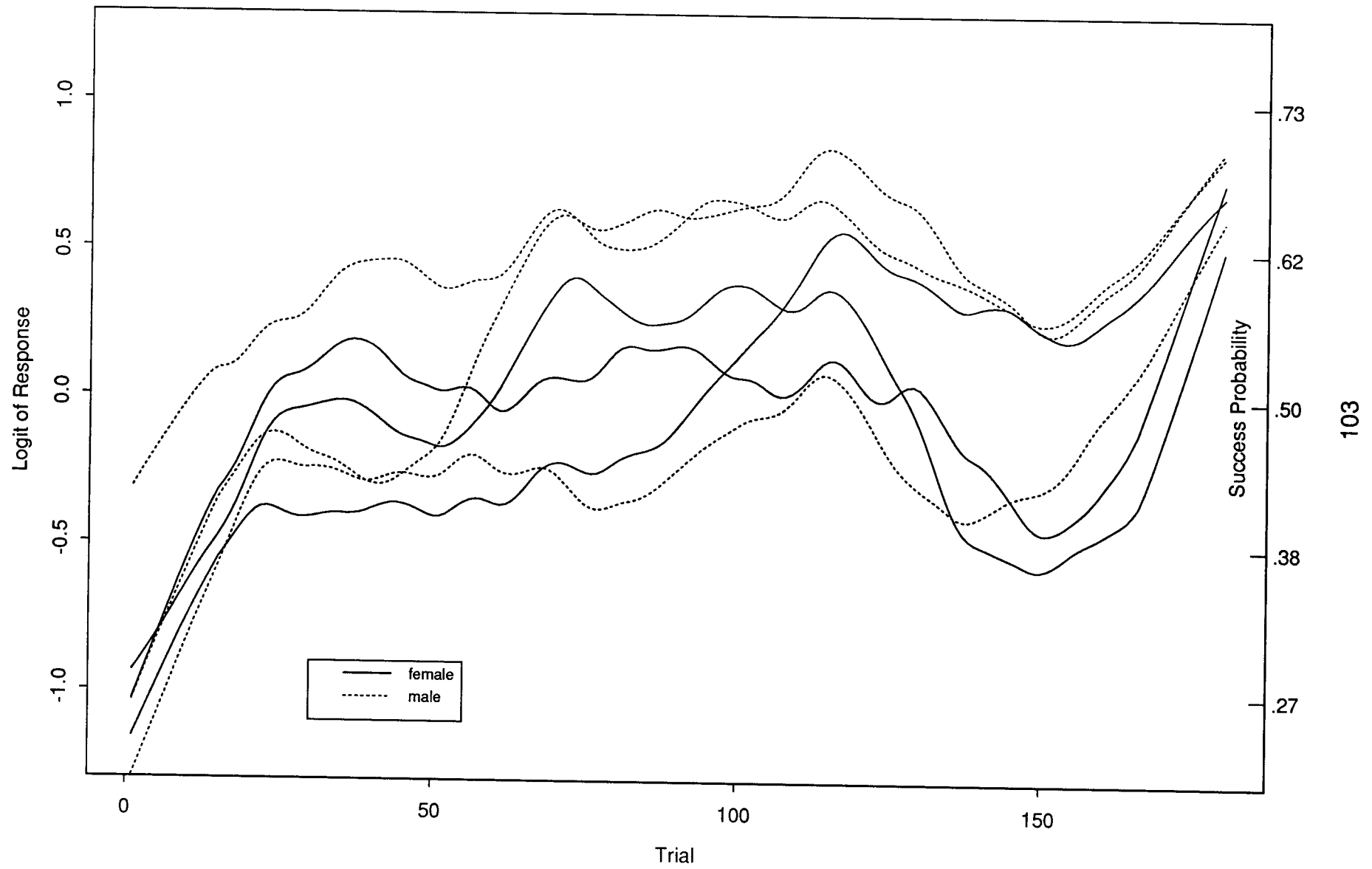


Figure 3.10. First Independence Fits without Sex

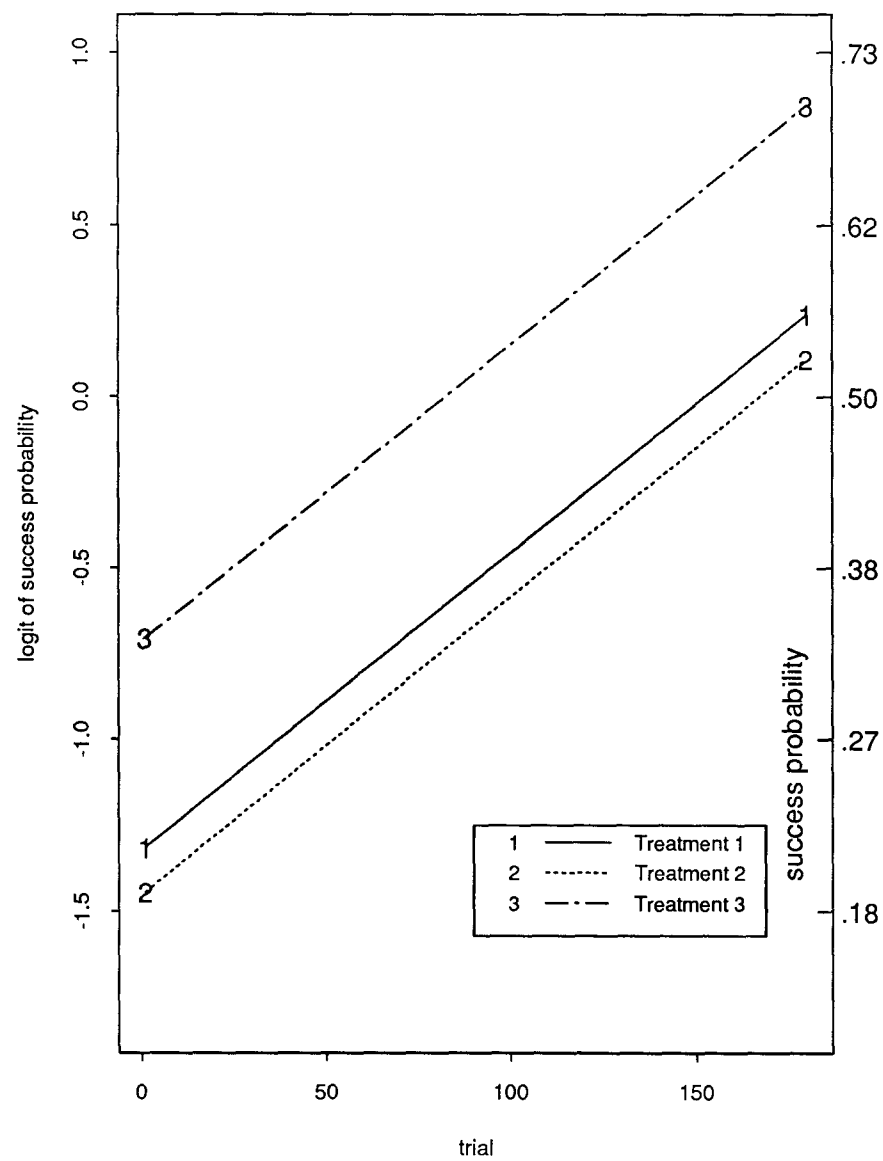


Figure 3.11. Second Independence Fits without Sex

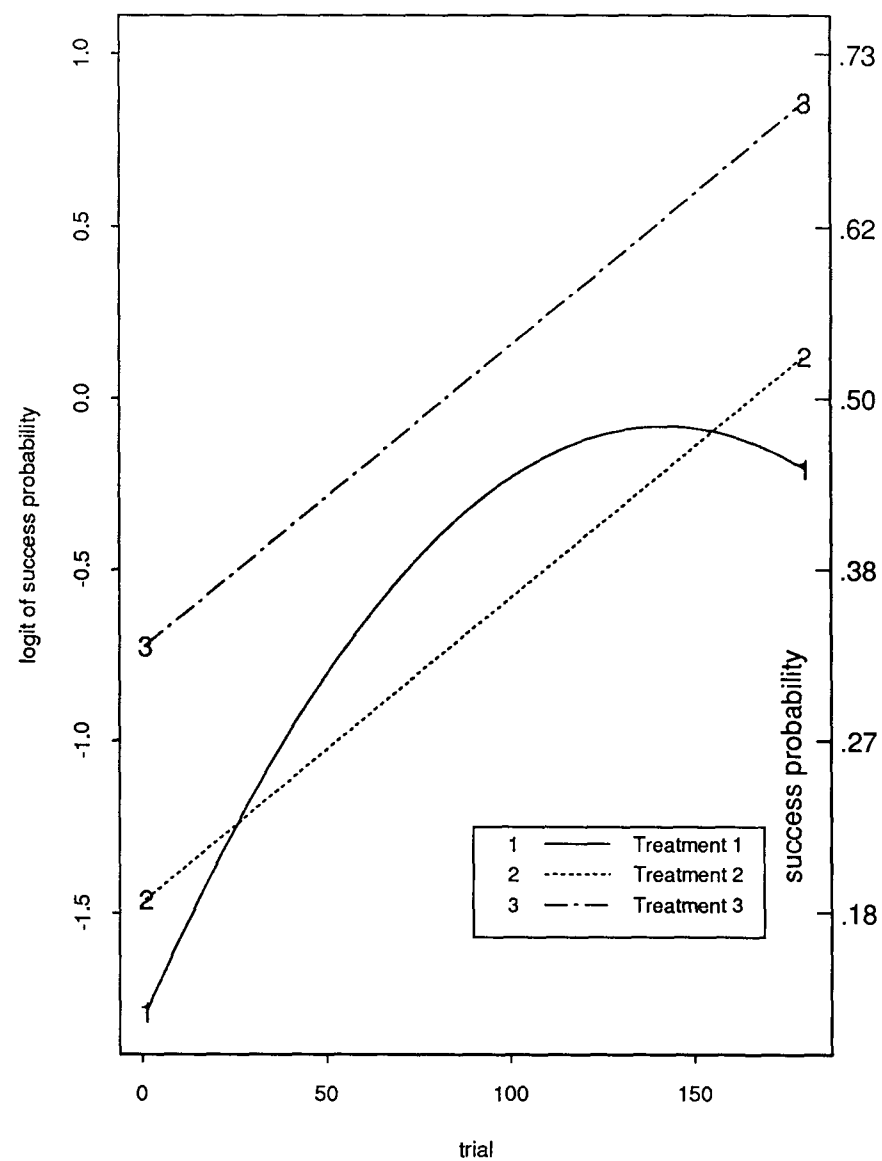


Figure 3.12. Fitted Success Probabilities for Treatment 1 from the Independence Working Model

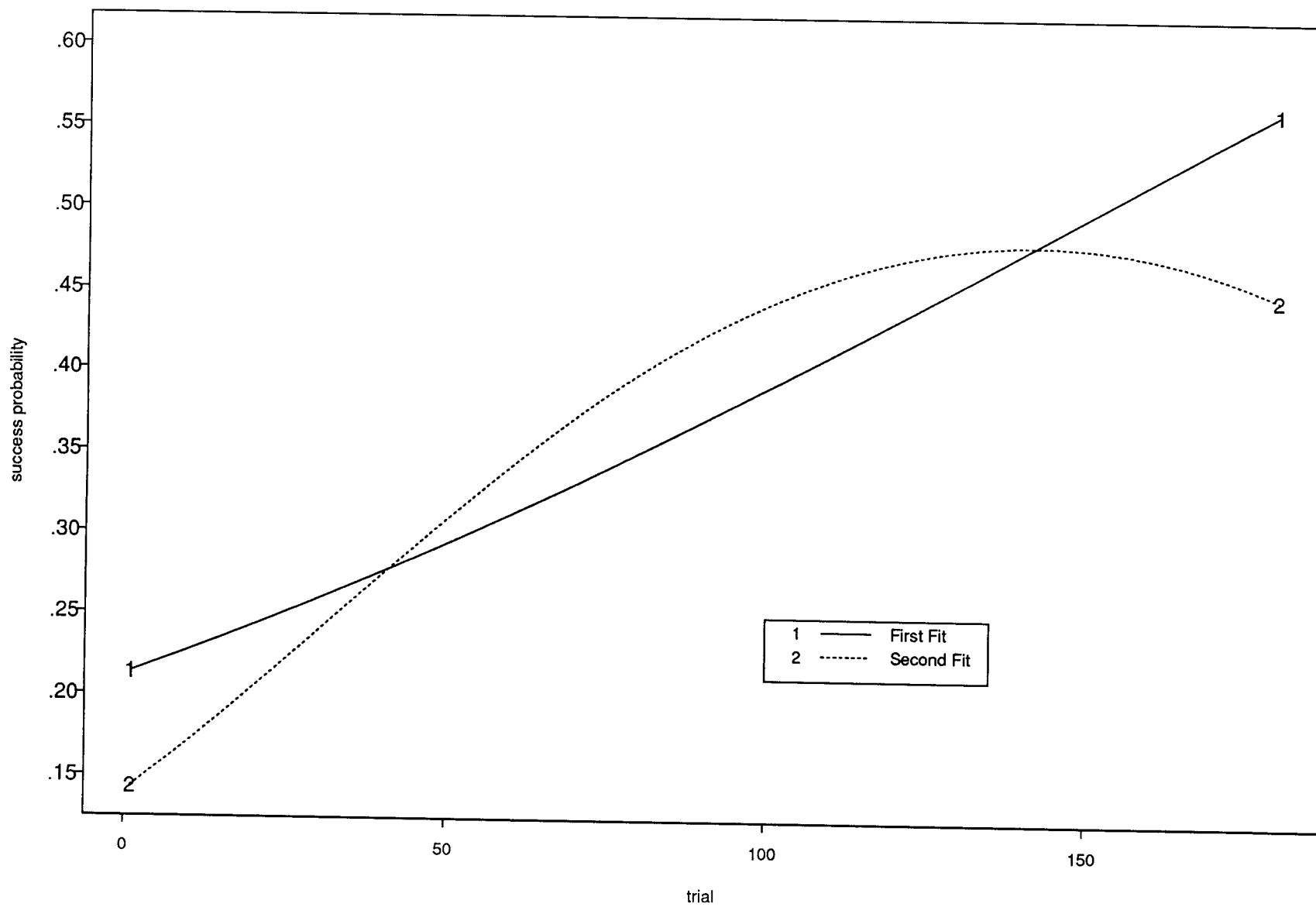


Figure 3.13. Independence Pearson Residuals by Bird

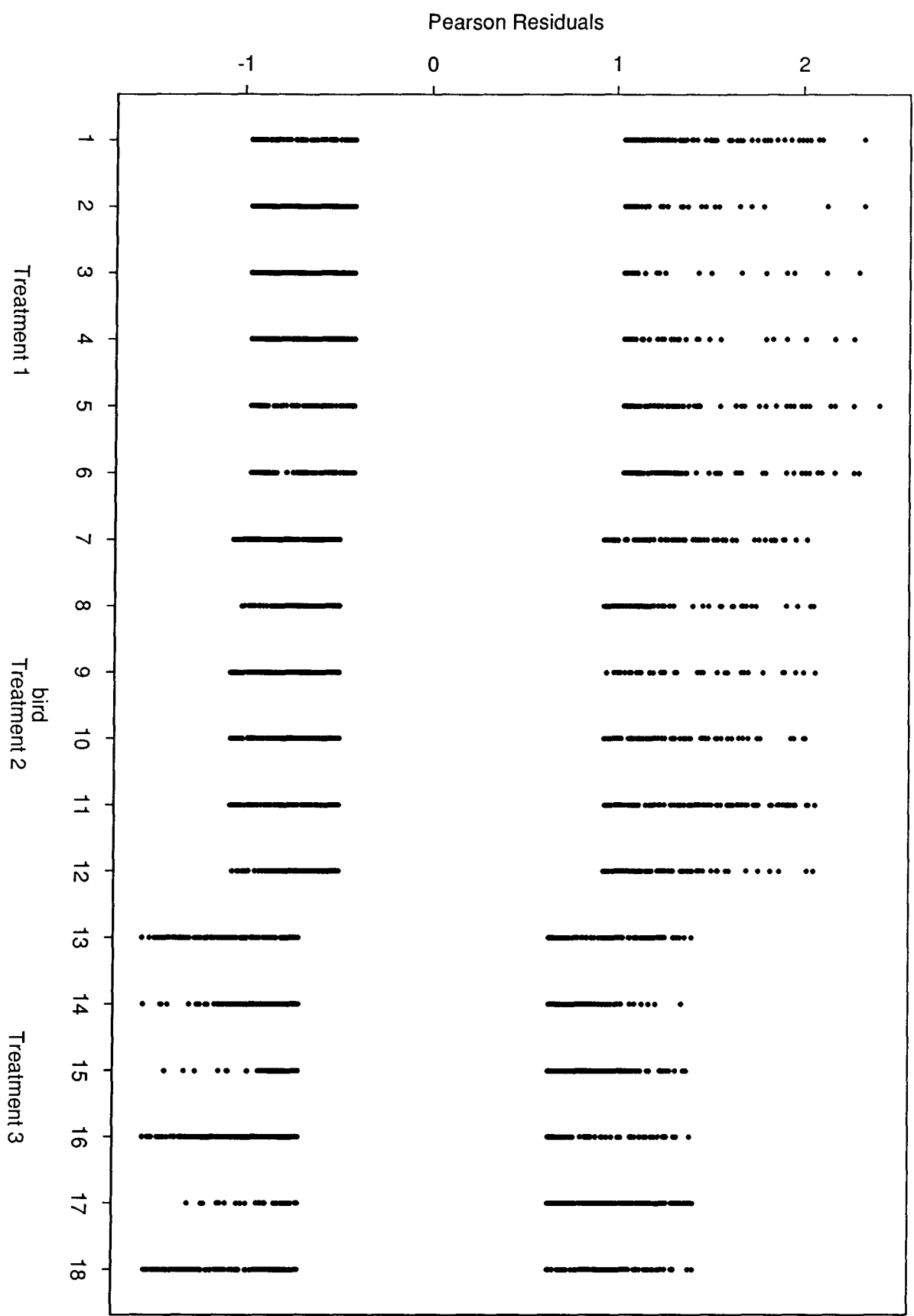


Figure 3.14. Boxplot of Independence Pearson Residuals by Sex

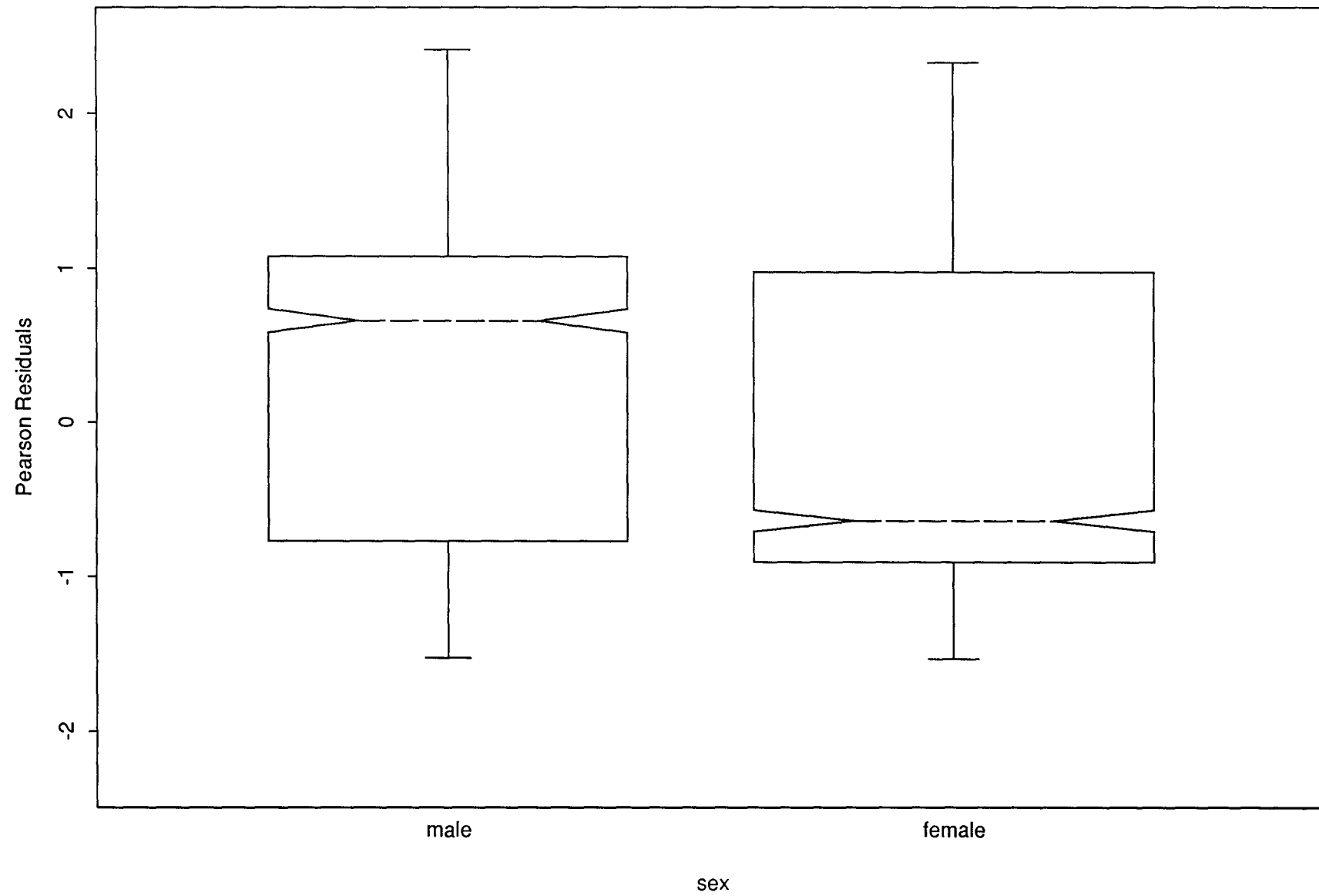


Figure 3.15. Smoothed Logit of Average Response vs. Trial for Treatment 1 (Window = 60)

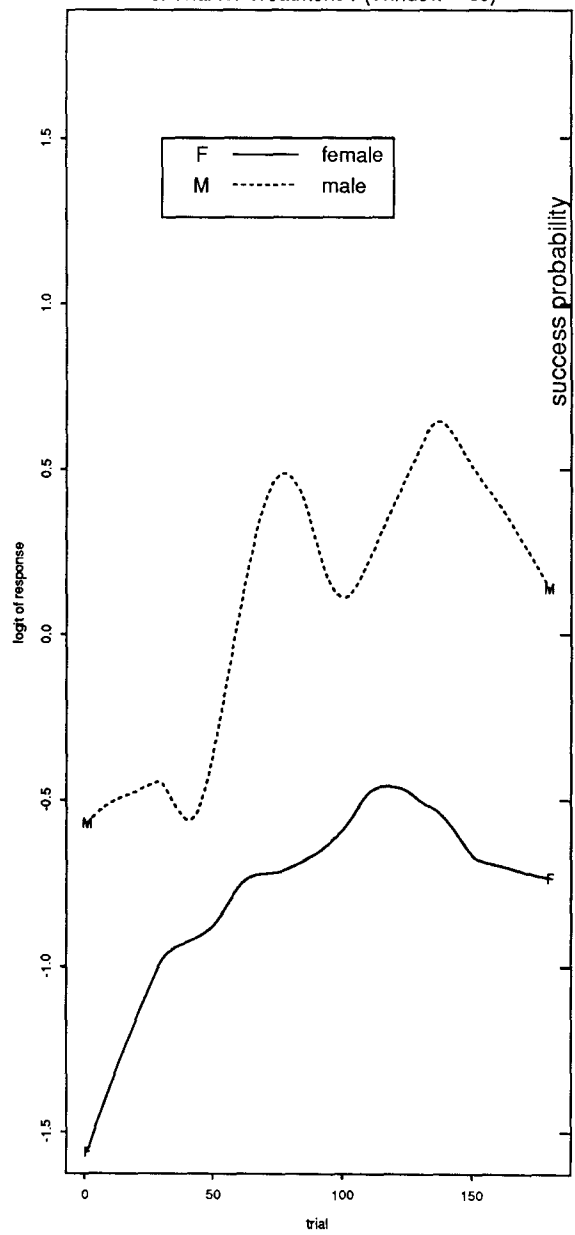


Figure 3.16. Smoothed Logit of Average Response vs. Trial for Treatment 2 (Window = 60)

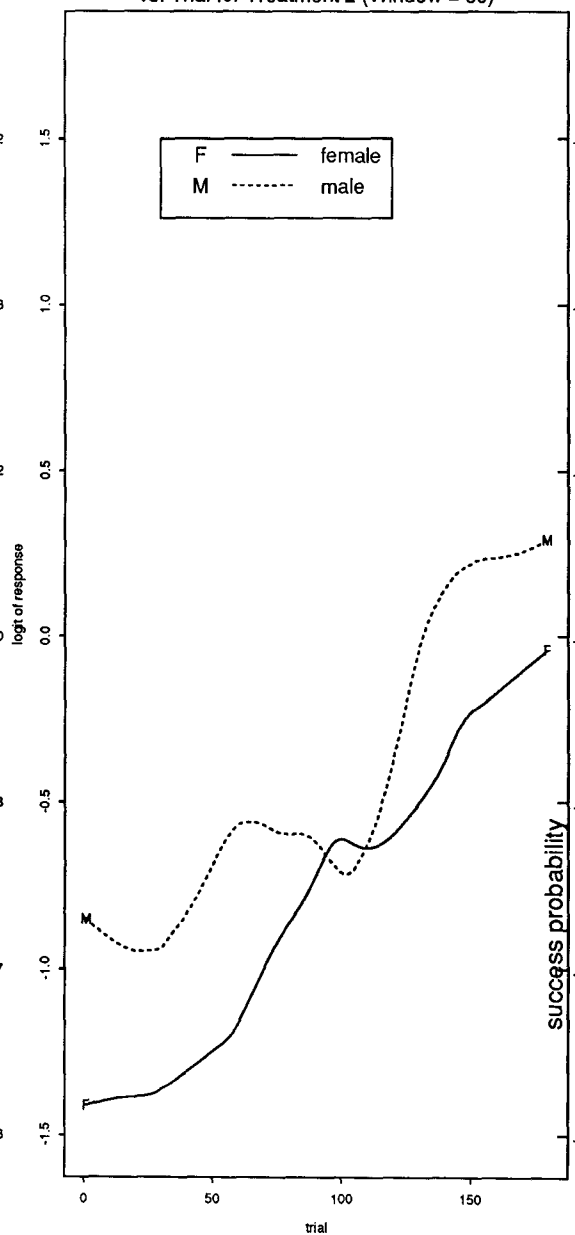


Figure 3.17. Smoothed Logit of Average Response vs. Trial for Treatment 3 (Window = 60)

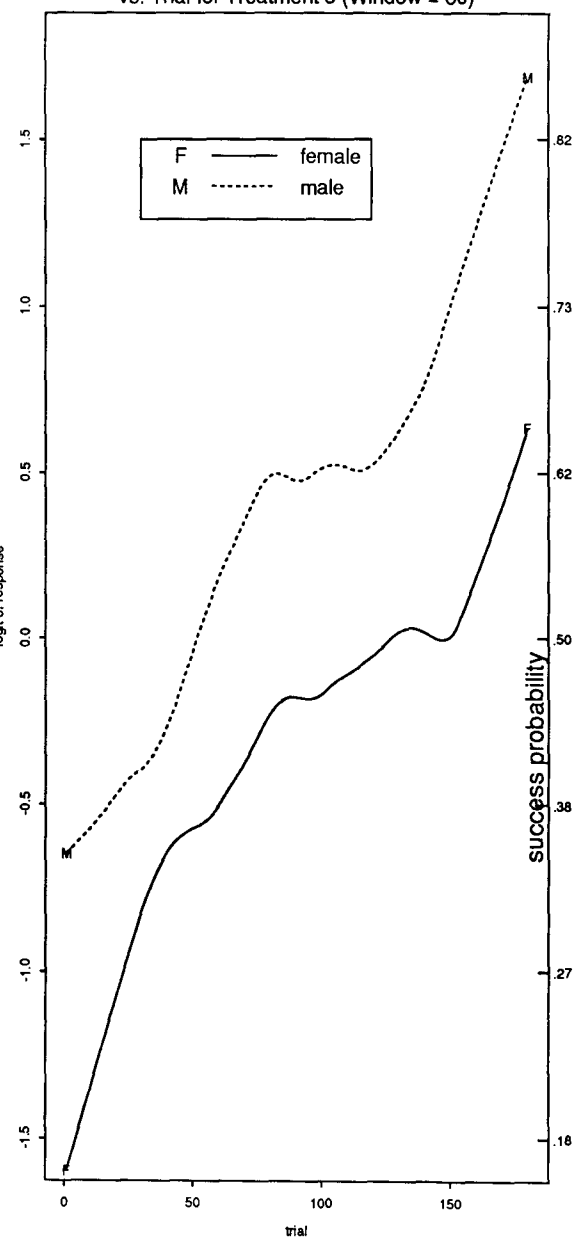




Figure 3.18. Binary Independence Fits  
for Females

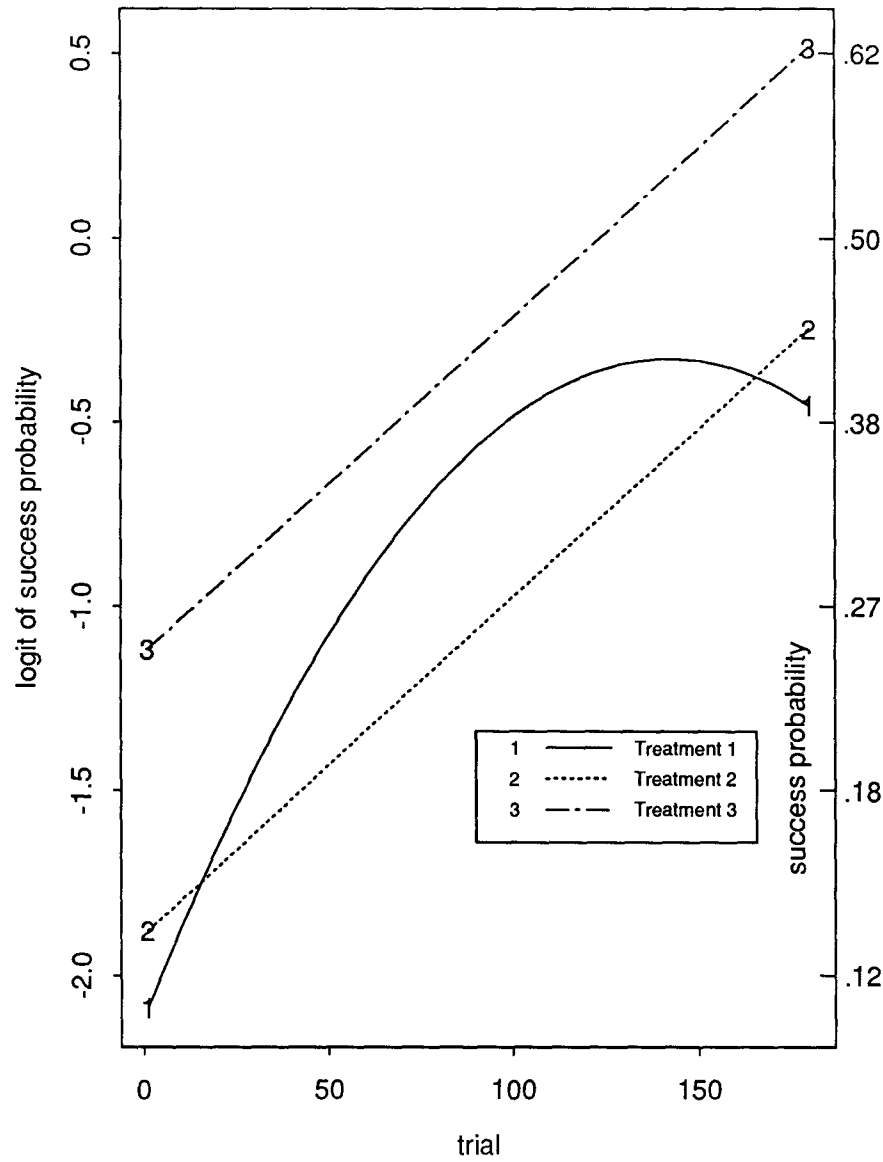


Figure 3.19. Binary First-Order Fits  
for Females

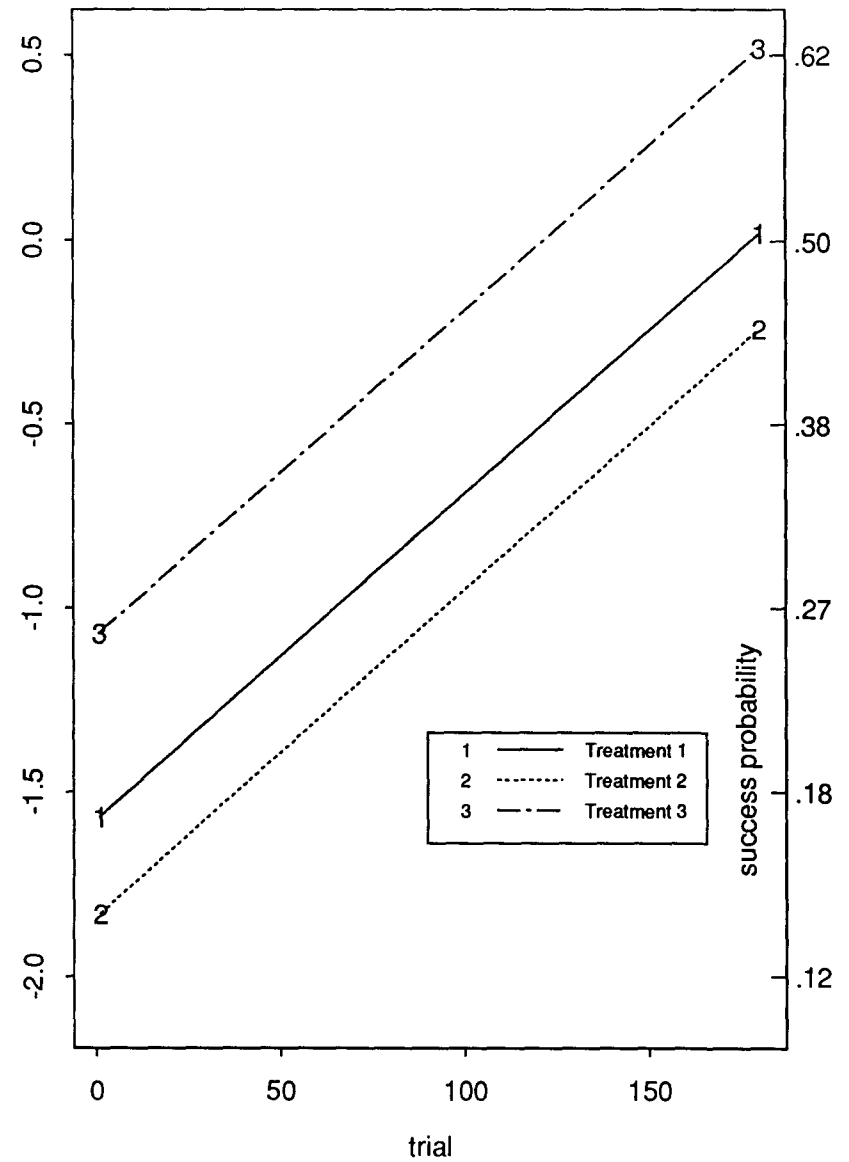


Figure 3.20. Fitted Success Probabilities for Females in Treatment 1

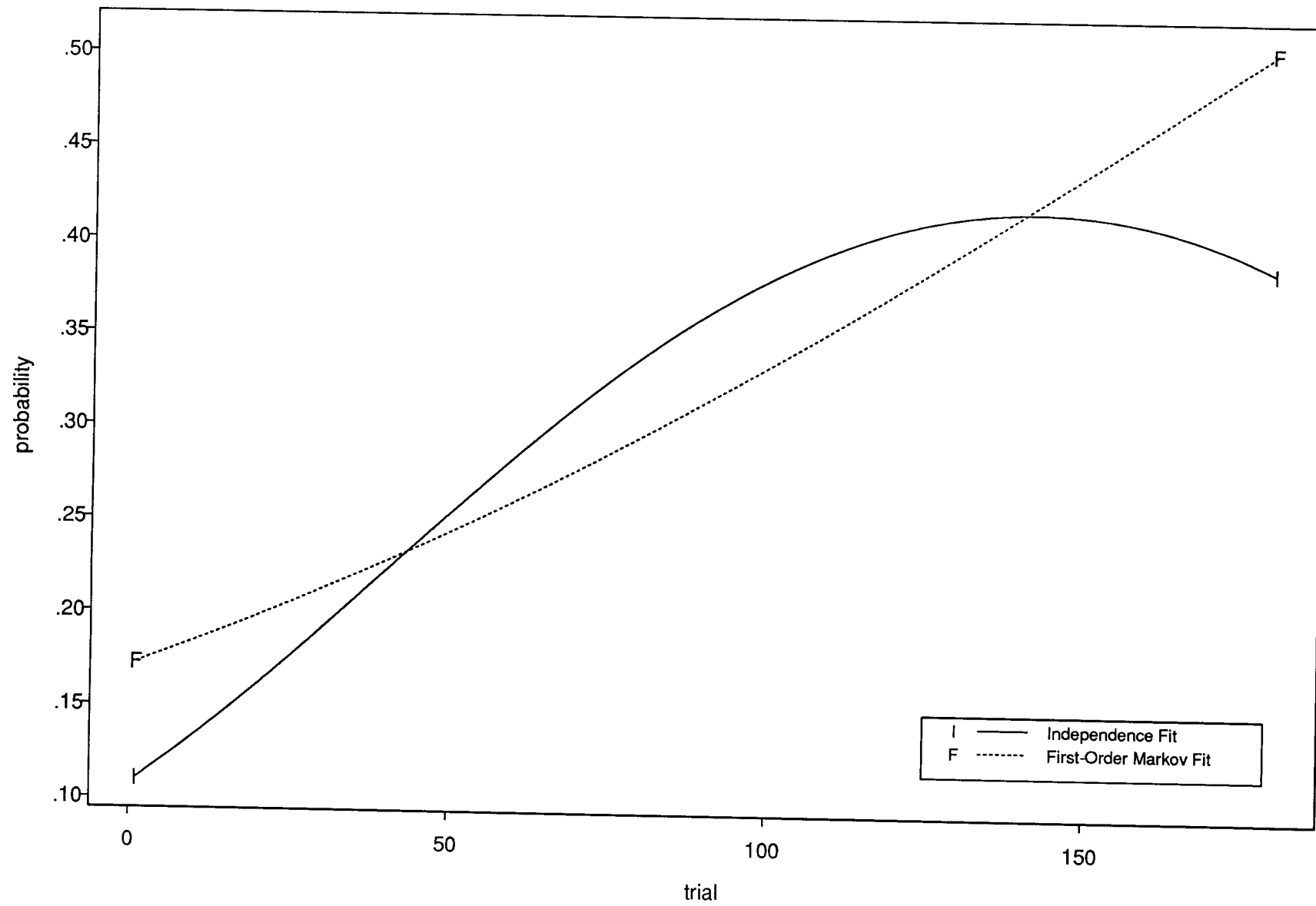


Figure 4.1. Binary Fits for Females Obtained from the Stationary 1-Dependent Working Correlation

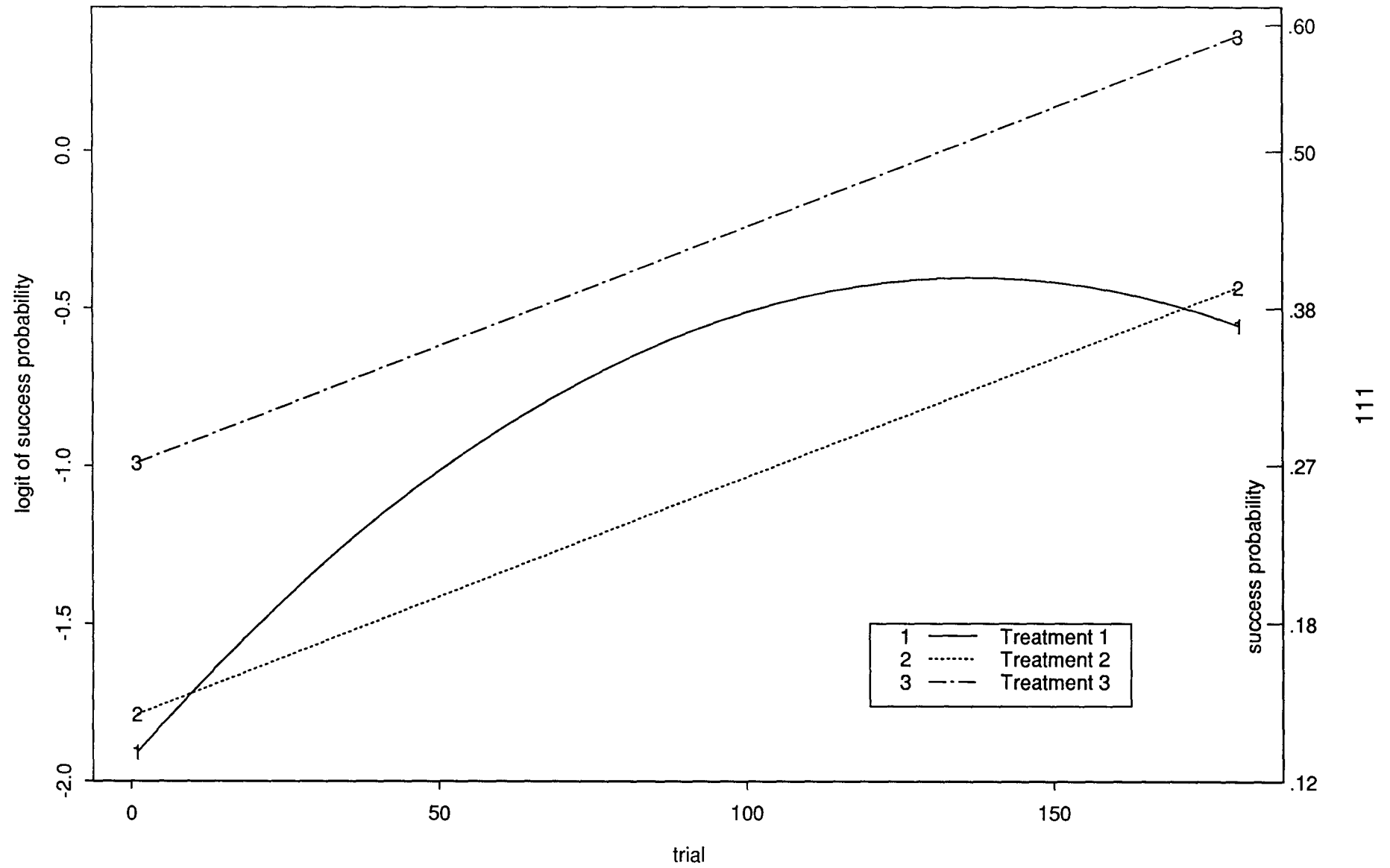


Figure 4.2. Average Logit of Response by Sex  
for Treatment 1, Blocks of 10 Trials

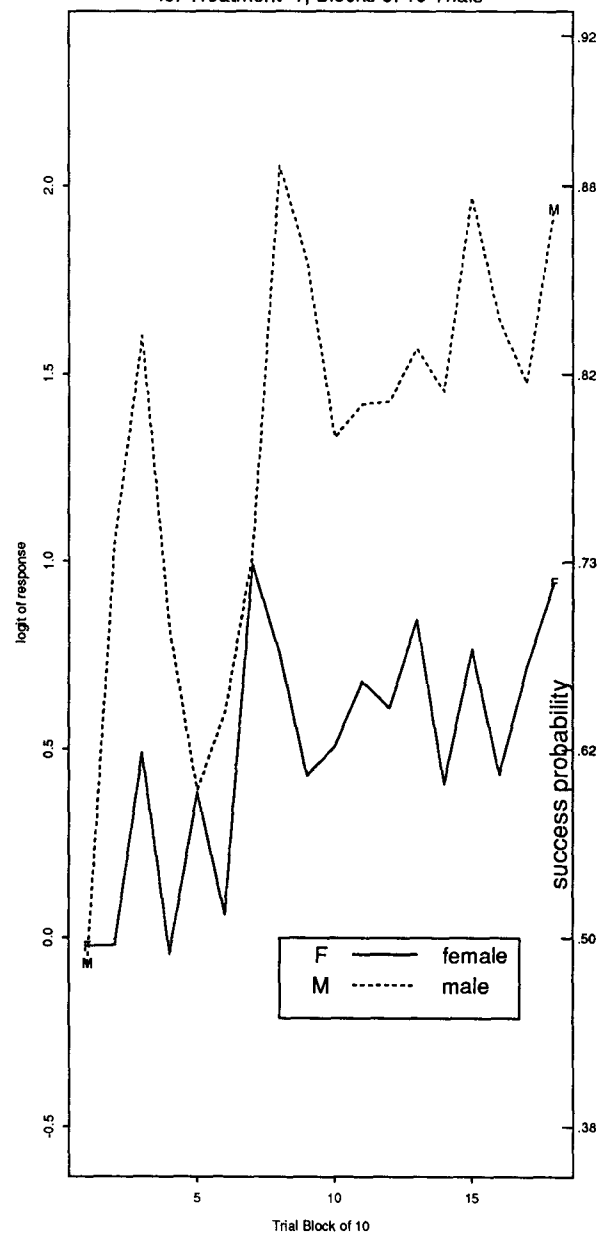


Figure 4.3. Average Logit of Response by Sex  
for Treatment 2, Blocks of 10 Trials

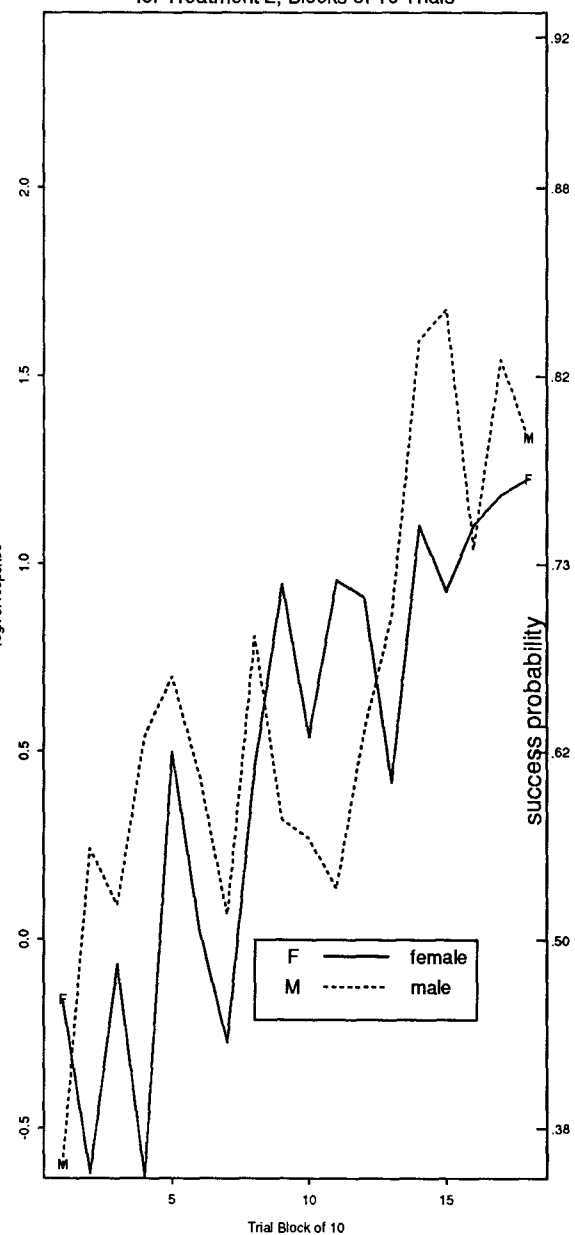


Figure 4.4. Average Logit of Response by Sex  
for Treatment 3, Blocks of 10 Trials

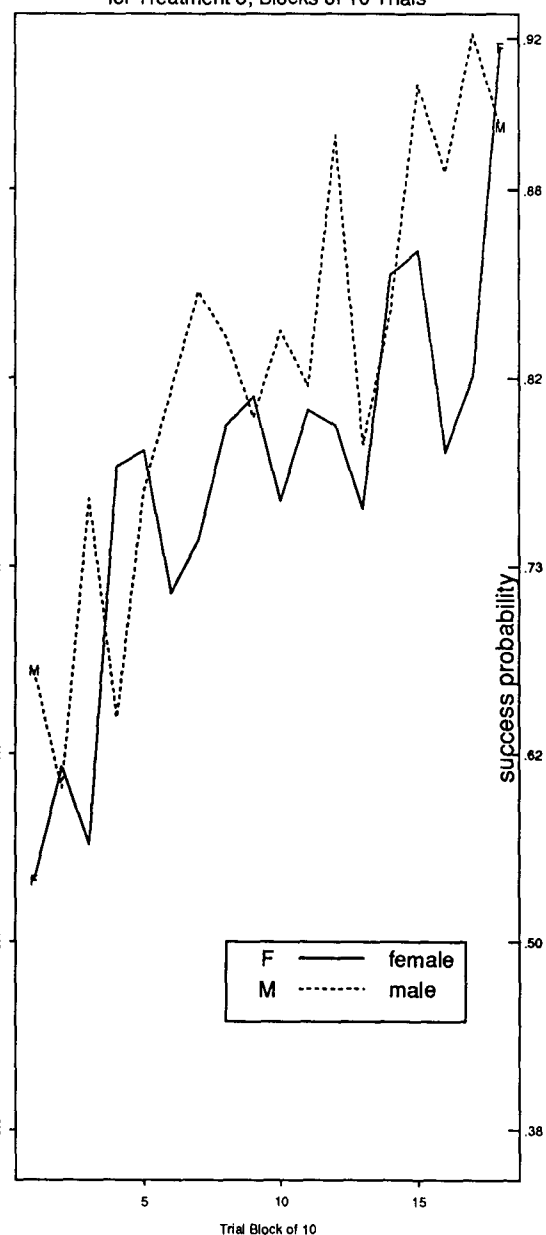


Figure 4.5. Smoothed Logit of Success Proportion versus Trial (Window = 30)

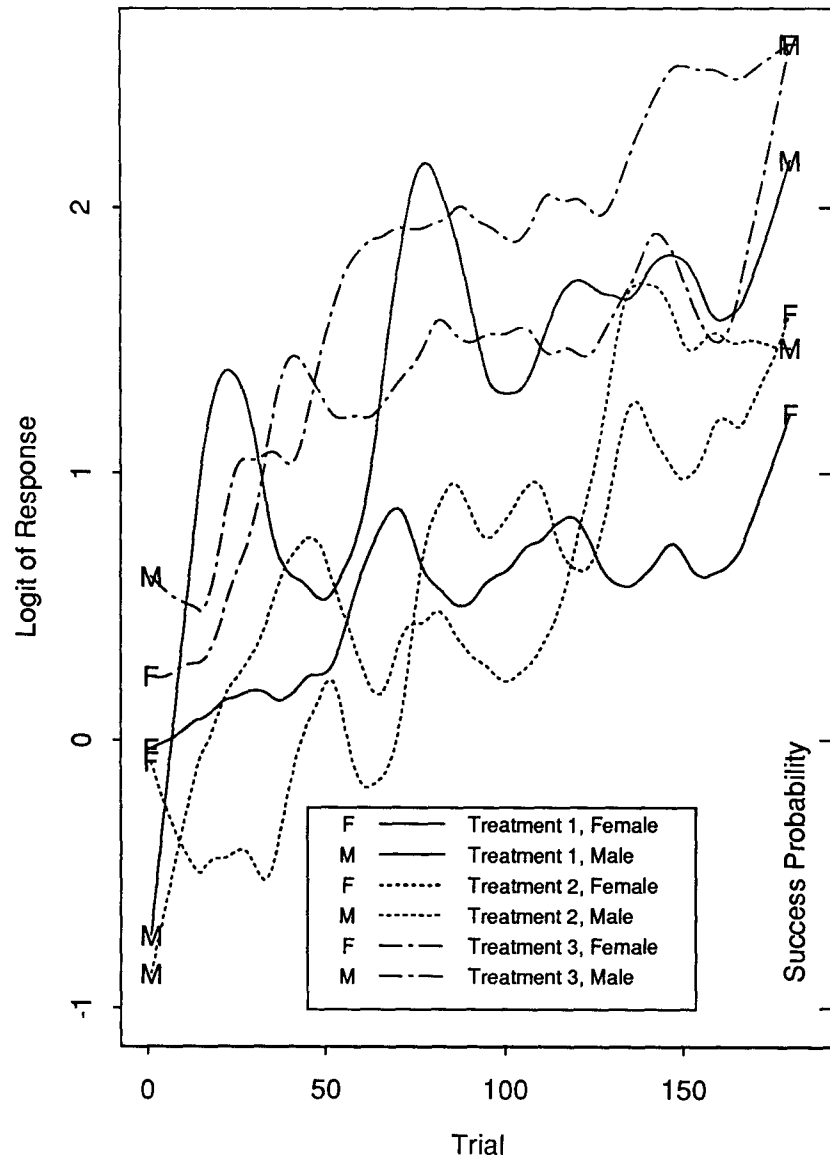


Figure 4.6. Smoothed Logit of Success Proportion versus Trial (Window = 60)

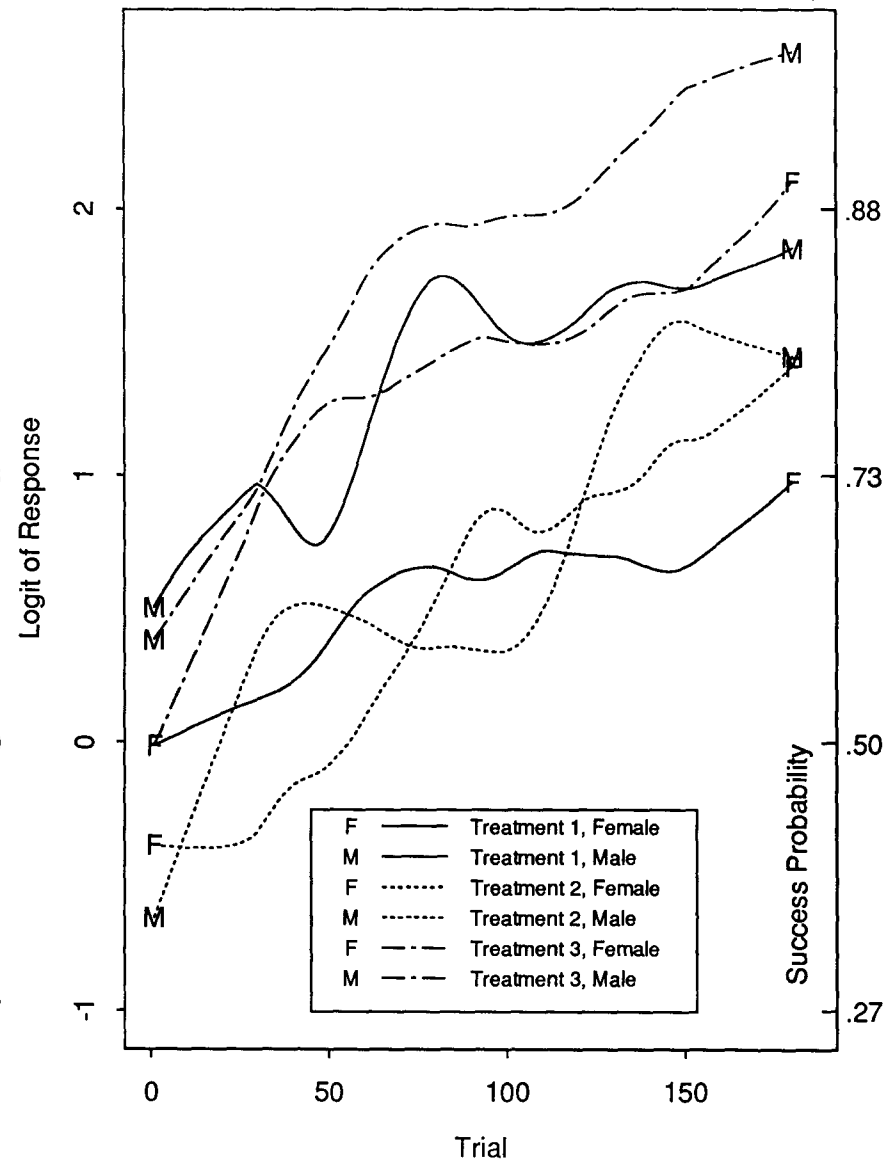


Figure 4.7. Smoothed Individual Logits (Binomial Data)  
for Treatment 1 (Window = 60)

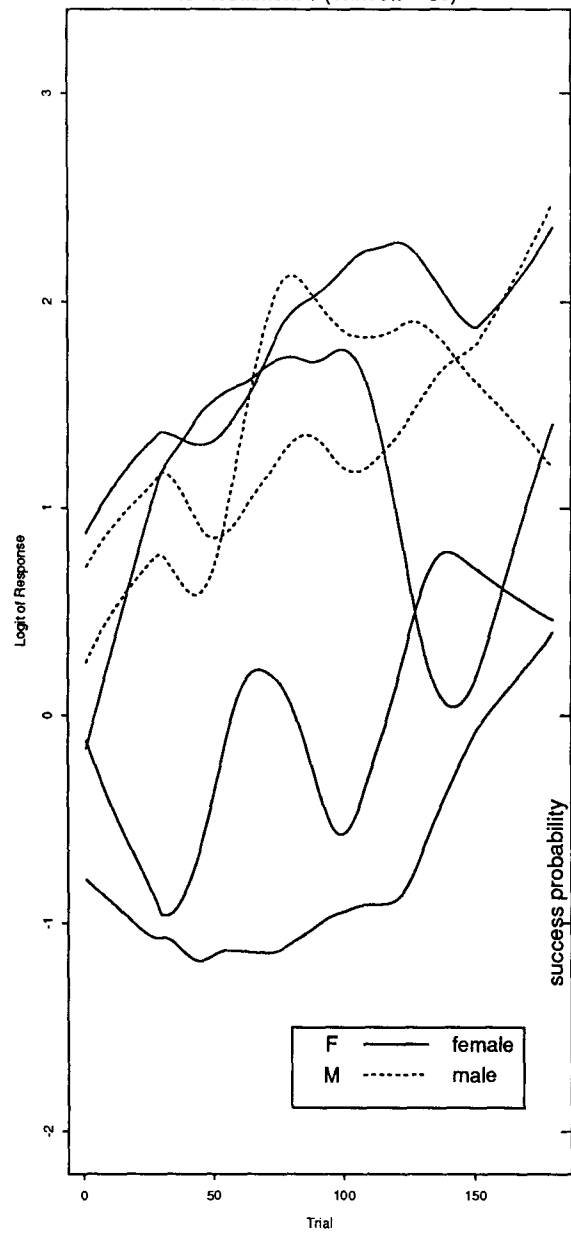


Figure 4.8. Smoothed Individual Logits (Binomial Data)  
for Treatment 2 (Window = 60)

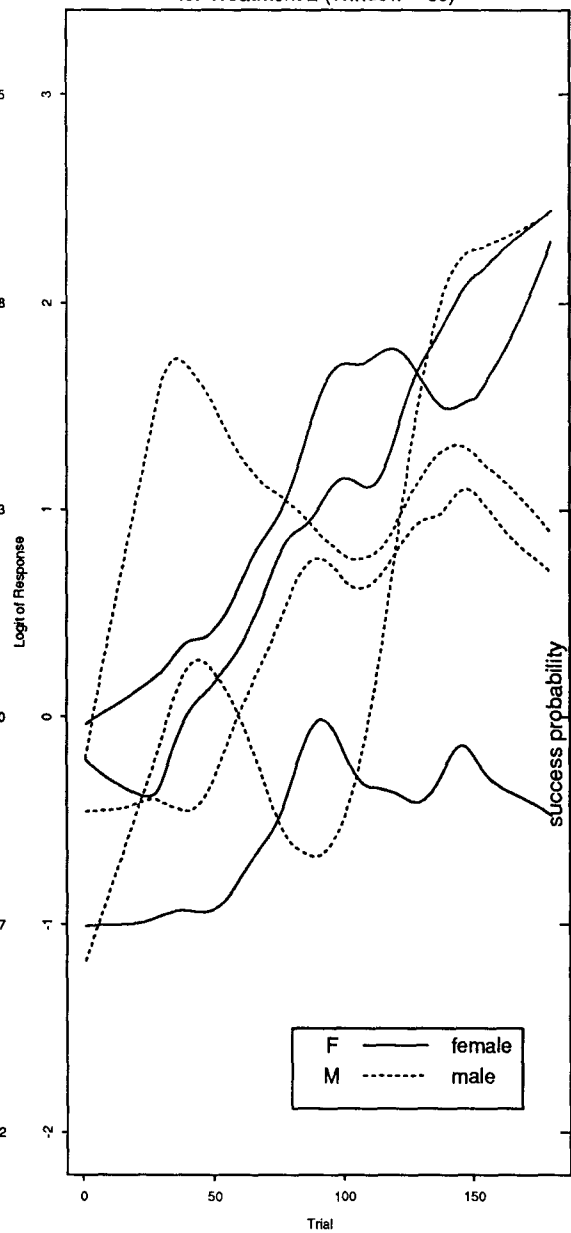


Figure 4.9. Smoothed Individual Logits (Binomial Data)  
for Treatment 3 (Window = 60)

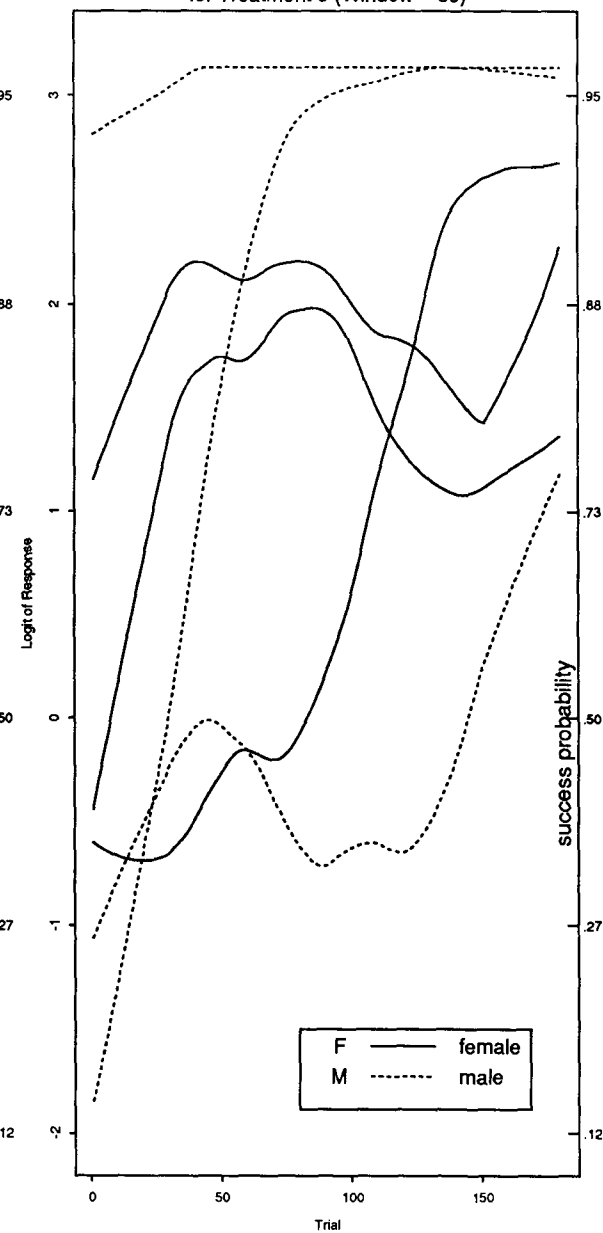


Figure 4.10. Initial Model Fits for Treatment 1  
(Binomial Data)

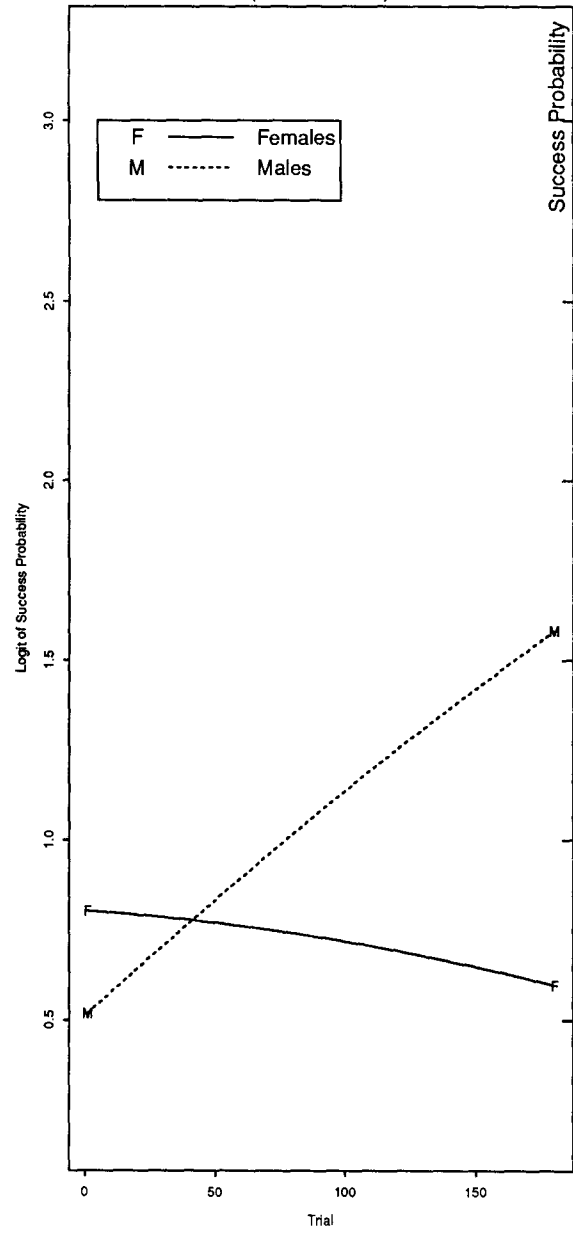


Figure 4.11. Initial Model Fits for Treatment 2  
(Binomial Data)

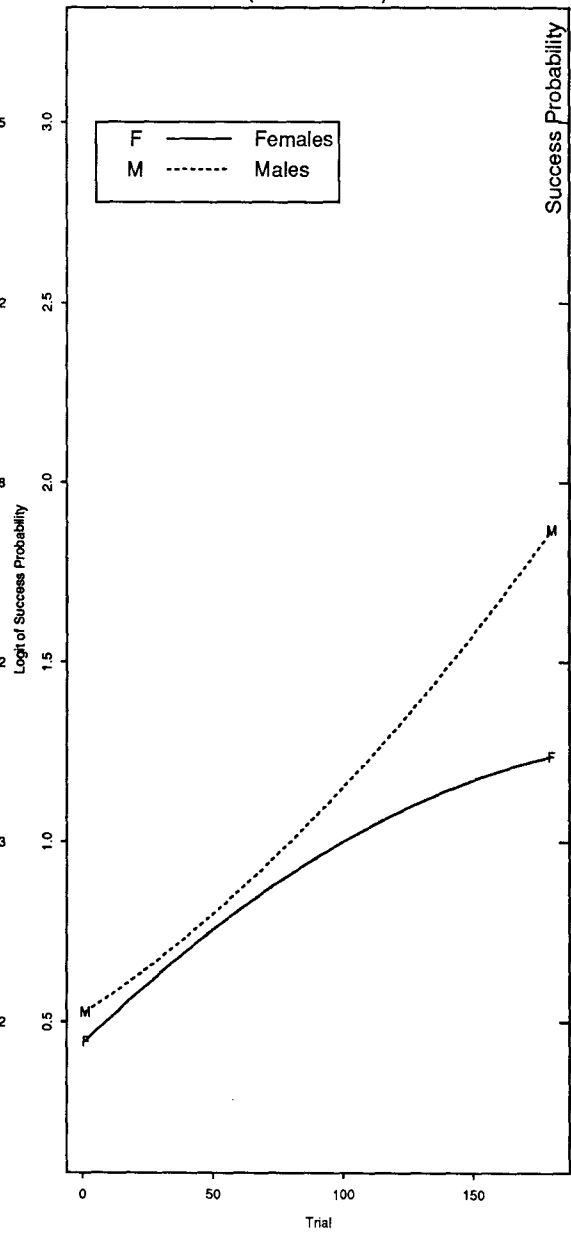


Figure 4.12. Initial Model Fits for Treatment 3  
(Binomial Data)

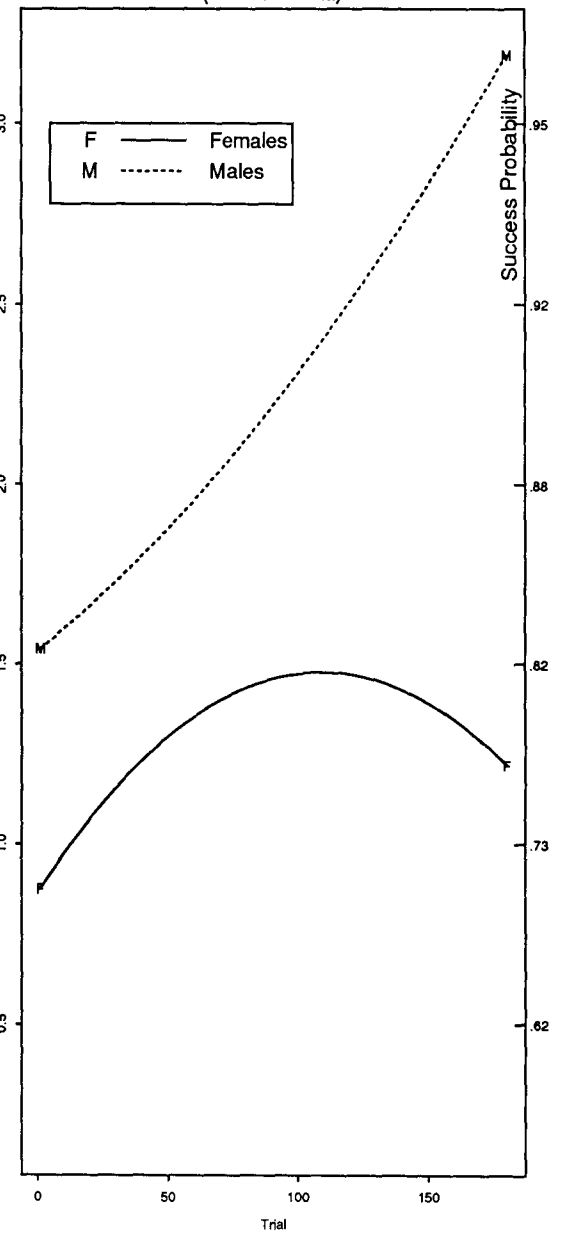


Figure 4.13. Final Model Fits for Females from the Stationary 1-Dependent Working Correlation

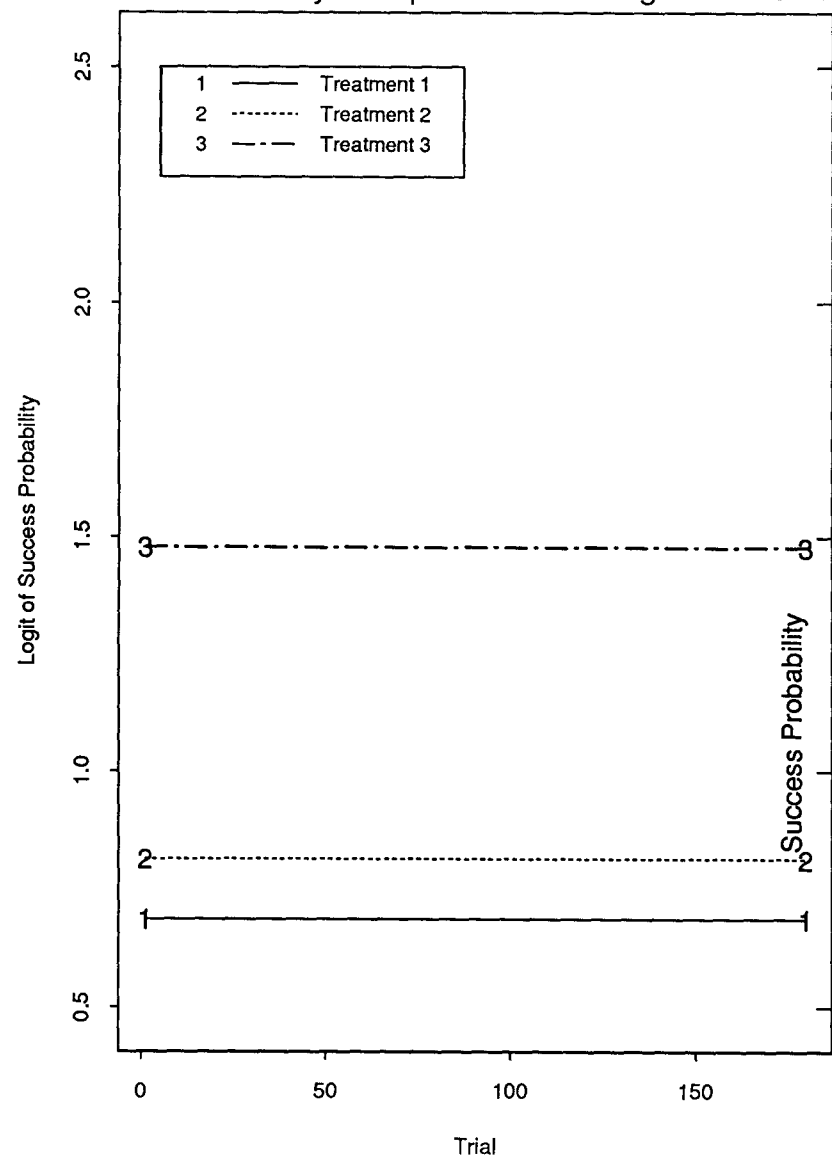
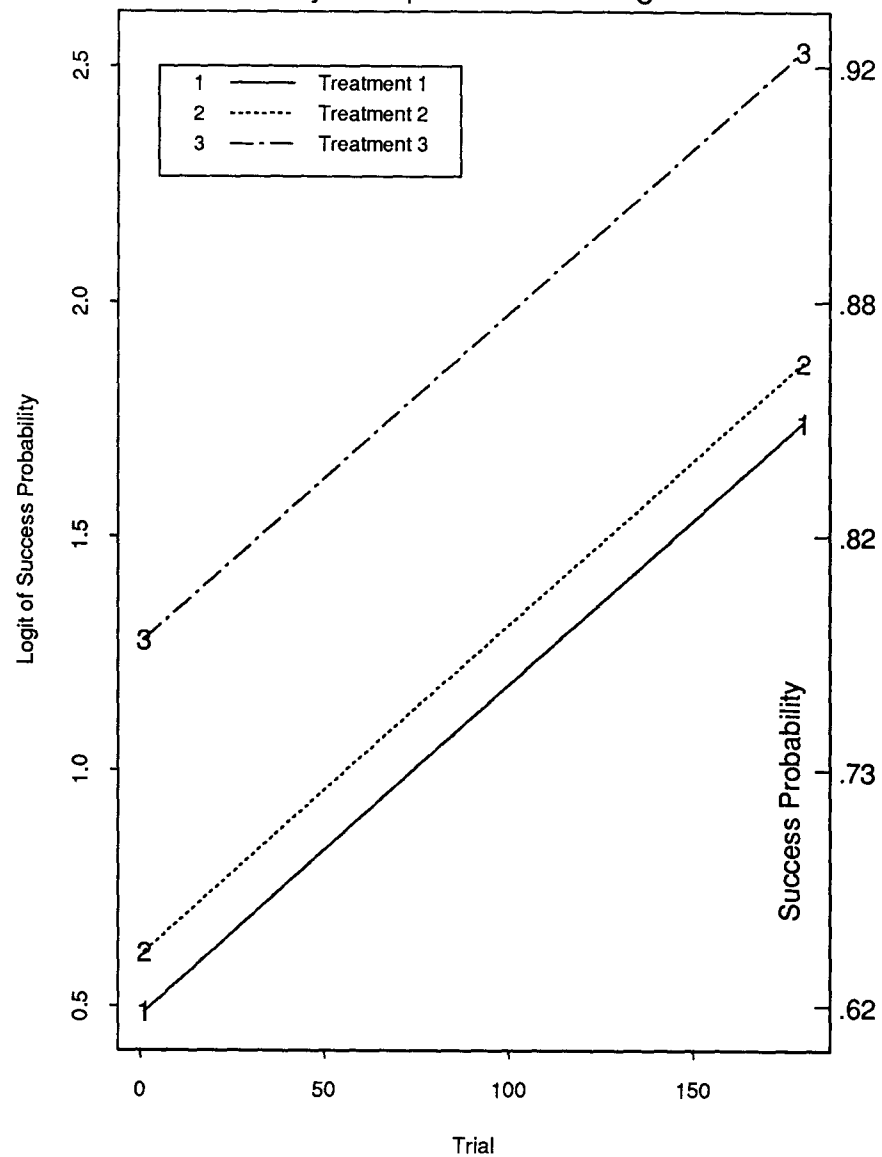


Figure 4.14. Final Model Fits for Males from the Stationary 1-Dependent Working Correlation





## Bibliography

- [1] Bowe, C.A. (1984). Spatial relations in animal learning and behavior. *Psychological Record*, *34* 181-209.
- [2] Gori, D.F. (1989). Floral color change in *Lupinus argenteus* (Fabaceae): why should plants advertise the location of unrewarding flowers to pollinators? *Evolution*, *43* 870-881.
- [3] Greenhouse, S.W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24* 95-112.
- [4] Huynh, H., and Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions, *Journal of the American Statistical Association*, *65* 1582-1589.
- [5] Liang, Kung-Yee, and Zeger Scott L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73* 13-22.
- [6] Mackintosh, N.J. (1983). *Conditioning and Associative Learning*. Oxford: Clarendon Press.
- [7] McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics*, *11* 59-67.
- [8] McCullagh, P., and Nelder, J.A., (1989). *Generalized Linear Models (2nd edition)*. Chapman and Hall, London.
- [9] Morrison, Donald F. (1976). *Multivariate Statistical Methods*. McGraw-Hill, New York.
- [10] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, *61* 439-447.
- [11] Zeger, Scott L., and Liang, Kung-Yee (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42* 121-130.
- [12] Zeger, Scott L., Liang, Kung-Yee, and Self, Steven G. (1985). The analysis of binary longitudinal data with time-independent covariates, *Biometrika*, *72* 31-78.