

# GENERAL BLENDING MODELS FOR MIXTURE EXPERIMENTS: DESIGN AND ANALYSIS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER  
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

January 2014

By  
Liam John Brown  
Mathematics

# Contents

<b>Declaration</b>	<b>11</b>
<b>Copyright</b>	<b>12</b>
<b>Acknowledgements</b>	<b>13</b>
<b>1 Introduction</b>	<b>14</b>
<b>2 Methodology</b>	<b>17</b>
2.1 Scheffé polynomials . . . . .	21
2.1.1 Simplex lattice designs and canonical polynomial models .	22
2.1.2 Simplex centroid designs and Scheffé $q$ -tenery models . . .	26
2.1.3 What can be described using Scheffé's polynomials? . . . .	27
2.1.4 Comments upon Scheffé's ideas . . . . .	28
2.1.5 Concluding remarks on Scheffé polynomials . . . . .	29
2.2 Becker's models . . . . .	30
2.2.1 Linear and quadratic blending . . . . .	30
2.2.2 Becker: New models . . . . .	32
2.2.3 Further linear blending models . . . . .	34
2.2.4 Summary . . . . .	35
2.3 Inverse and logarithmic terms . . . . .	35
2.4 Reparameterisations . . . . .	37
2.5 Generalised Q-fractions mixture model . . . . .	43
2.6 Summary . . . . .	44
<b>3 The GBMM</b>	<b>46</b>
3.1 Regressor defining parameters . . . . .	46
3.2 The Dirichlet distribution . . . . .	49

3.3	Interpreting the terms of joint effect . . . . .	50
3.4	General blending models for mixture experiments . . . . .	55
<b>4</b>	<b>Fitting the GBMM</b>	<b>57</b>
4.1	Model estimation . . . . .	57
4.1.1	Maximum likelihood estimation . . . . .	58
4.1.2	Least squares regression . . . . .	59
4.2	Partially linear models . . . . .	60
4.3	Stepwise regression and model selection . . . . .	62
4.4	Fitting the GBMM: an introduction . . . . .	63
4.5	Adding a term . . . . .	65
4.6	Reestimating the regressor defining parameters of each term . . .	67
4.7	Reestimating the regressor defining parameters $\mathbf{t}$ . . . . .	68
4.8	Fitting the GBMM: a graphical representation . . . . .	69
4.8.1	Adding a term . . . . .	71
4.8.2	Reestimating the regressor defining parameters of each term	72
4.8.3	Reestimating the regressor defining parameters $\mathbf{t}$ . . . . .	73
4.9	How to fit the GBMM: discussion . . . . .	73
4.9.1	Ternary terms of joint effect . . . . .	73
4.9.2	How many times shall the terms be reestimated? . . . . .	74
4.9.3	Should $\mathbf{t}$ be reestimated separately? . . . . .	74
4.9.4	What set of initial values shall be used for $r_{ij}$ , $r_{ji}$ and $t_{ij}$ ? .	74
4.9.5	How many general binary terms of joint effect can be con- sidered for each pair of components? . . . . .	75
4.9.6	$AIC_c$ and the number of parameters . . . . .	76
4.9.7	Stopping criterion . . . . .	76
4.10	Summary . . . . .	77
<b>5</b>	<b>Fitting the GBMM: Examples</b>	<b>78</b>
5.1	Example F1 . . . . .	80
5.2	Example F2 . . . . .	80
5.3	Example F3 . . . . .	81
5.4	Example F4 . . . . .	81
5.5	Example F5 . . . . .	82
5.6	Example F6 . . . . .	83
5.7	Example F7 . . . . .	83

5.8	Example F8 . . . . .	84
5.9	Comparison of GBMMs . . . . .	85
5.10	Comparison to other models . . . . .	86
5.11	Conclusions . . . . .	87
<b>6</b>	<b>Design of Experiments</b>	<b>90</b>
6.1	Continuous and exact designs and their optimality . . . . .	90
6.2	General Equivalence Theorem . . . . .	92
6.3	Experimental design for nonlinear models . . . . .	93
6.4	The Fedorov exchange . . . . .	94
6.5	D-optimum design for the Scheffé polynomials . . . . .	96
<b>7</b>	<b>Designs for the GBMM</b>	<b>98</b>
7.1	Example D1 . . . . .	100
7.2	Example D2 . . . . .	101
7.3	Example D3 . . . . .	107
7.4	Example D4 . . . . .	108
7.5	Example D5 . . . . .	111
7.6	Conclusions . . . . .	113
<b>8</b>	<b>Examples</b>	<b>116</b>
8.1	What can the GBMM describe? . . . . .	116
8.1.1	Two component example . . . . .	117
8.1.2	Three component example . . . . .	119
8.1.3	Discussion . . . . .	122
8.2	Using a design for the GBMM: an example from industry . . . . .	122
8.3	Fitting where there are more components: a second example from industry . . . . .	124
<b>9</b>	<b>Discussion</b>	<b>128</b>
9.1	What has been achieved? . . . . .	128
9.2	What is still to be achieved? . . . . .	130
9.3	Summary . . . . .	132
<b>A</b>	<b>Designs for the GBMM</b>	<b>133</b>

Word Count: 42,652

# List of Tables

5.1	Statistics comparing GBMM found, with different fitting procedure, for chick diet data . . . . .	85
5.2	Statistics comparing GBMM to a reduced Scheffé cubic polynomial, for chick diet data . . . . .	87
7.1	Statistics for locally D-optimum 15 point design 7.1a for GQBMM with priors equal to implicit values in the Scheffé quadratic polynomial . . . . .	101
7.2	Statistics for locally D-optimum 19 point design 7.1b for GSCBMM with priors equal to implicit values in Scheffé cubic polynomial . . . . .	101
7.3	Statistics for locally D-optimum designs for GQBMM with priors equal to implicit values in the Scheffé quadratic polynomial . . . . .	103
7.4	Statistics for contrasting locally D-optimum 15 point designs for GQBMM, chosen with different priors . . . . .	110
7.5	Characteristics for contrasting locally D-optimum 19 point designs for GQBMM, chosen with different priors . . . . .	110
7.6	Characteristics for contrasting D-optimum 15 point designs for GQBMM . . . . .	113
8.1	Statistics comparing GBMM to Scheffé polynomials for simulated two component example . . . . .	118
8.2	Statistics comparing GBMM to Scheffé polynomials with additional inverse terms for simulated two component example . . . . .	119
8.3	Statistics comparing GBMM to Scheffé polynomials for simulated three component example . . . . .	119
8.4	Parameter estimates for the linear terms of the GBMM fit to 8 component FM data . . . . .	126

8.5	Parameter estimates for the terms of the joint effect of the GBMM fit to 8 component FM data . . . . .	126
8.6	Comparison of models fit to 8 component FM data . . . . .	127
A.1	15 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial . . . . .	133
A.2	19 point design for GSCBMM identified with priors equal to values dictated by the Scheffé cubic polynomial . . . . .	134
A.3	18 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial . . . . .	135
A.4	21 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial . . . . .	136
A.5	19 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial . . . . .	137
A.6	15 point design for GQBMM identified with priors equal to values dictated by the Becker H2 model . . . . .	137
A.7	15 point design for GQBMM identified with priors equal to values dictated by Becker H2 model and the Scheffé quadratic polynomial	138
A.8	19 point design for GQBMM identified with priors equal to values dictated by the Becker H2 model . . . . .	138
A.9	19 point design for GQBMM identified with priors equal to values dictated by Becker H2 model and the Scheffé quadratic polynomial	139

# List of Figures

2.1	Illustration of the difference in Quenouillé and Scheffé's understanding of component blending . . . . .	31
3.1	Blending effects for $s_{ij} = 0.2, 0.5, 1, 2$ and 5. . . . .	48
3.2	Effect of binary term of joint effect for $g_{ij} = 20$ and $g_{ij} = 0.2$ and $h_{ij} = 0.75$ . . . . .	52
3.3	Plots of the response described by the general ternary term of joint effect for different values of $g_{ijk}$ , $h_{ijk}$ and $g_{jki}$ . . . . .	54
5.1	Response surfaces for GBMM for chick diet data . . . . .	86
5.2	Predicted response surface for the chick diet data . . . . .	88
6.1	Designs for the Scheffé polynomials . . . . .	97
7.1	Design plots for saturated locally D-optimum designs, for GQBMM and GSCBMM, with point priors equal to implicit values in the Scheffé polynomials . . . . .	100
7.2	Prediction variance plots for saturated locally D-optimum designs, for GQBMM and GSCBMM, with point priors equal to implicit values in Scheffé polynomials . . . . .	102
7.3	Design plots for non-saturated locally D-optimum designs, for GQBMM, with point priors equal to implicit values in the Scheffé quadratic polynomial . . . . .	104
7.4	Prediction variance plots for non-saturated locally D-optimum designs, for GQBMM, with point priors equal to implicit values in the Scheffé quadratic polynomial . . . . .	105
7.5	Design plot and prediction variance plot for 15 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model . . . . .	108



7.6	Design plot and prediction variance plot for 15 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model and Scheffé quadratic polynomial .	108
7.7	Contrasting 15 point designs for GQBMM, chosen with different priors . . . . .	109
7.8	Design plot and prediction variance plot for 19 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model . . . . .	111
7.9	Design plot and prediction variance plot for 19 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model and Scheffé quadratic polynomial .	111
7.10	Plots of prediction variance for designs where priors are defined through Becker H2 Model and Scheffé quadratic polynomial . . .	112
8.1	Fitted response surface for GQBMM v. Underlying response surface	118
8.2	Fitted response surface for alternative models v. underlying response surface for simulated two component example . . . . .	120
8.3	Plots relating to simulated 3 component example . . . . .	121
8.4	Designs for brake compaction data . . . . .	123
8.5	Predicted response surface for models fitted to brake compaction data . . . . .	125

# GENERAL BLENDING MODELS FOR MIXTURE EXPERIMENTS: DESIGN AND ANALYSIS

Liam John Brown

A thesis submitted to the University of Manchester  
for the degree of Doctor of Philosophy, January 2014

It is felt the position of the Scheffé polynomials as the primary, or sometimes sole recourse for practitioners of mixture experiments leads to a lack of enquiry regarding the type of blending behaviour that is used to describe the response and that this could be detrimental to achieving experimental objectives. Consequently, a new class of models and new experimental designs are proposed allowing a more thorough exploration of the experimental region with respect to different blending behaviours, especially those not associated with established models for mixtures, in particular the Scheffé polynomials.

The proposed General Blending Models for Mixtures (GBMM) are a powerful tool allowing a broad range of blending behaviour to be described. These include those of the Scheffé polynomials (and its reparameterisations) and Becker's models. The potential benefits to be gained from their application include greater model parsimony and increased interpretability. Through this class of models it is possible for a practitioner to reject the assumptions inherent in choosing to model with the Scheffé polynomials and instead adopt a more open approach, flexible to many different types of behaviour. These models are presented alongside a fitting procedure, implementing a stepwise regression approach to the estimation of partially linear models with multiple nonlinear terms.

The new class of models has been used to develop designs which allow the response surface to be explored fully with respect to the range of blending behaviours the GBMM may describe. These designs may additionally be targeted at exploring deviation from the behaviour described by the established models. As such, these designs may be thought to possess an enhanced optimality with respect to these models. They both possess good properties with respect to optimality criterion, but are also designed to be robust against model uncertainty.

# Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

# Copyright

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.manchester.ac.uk/library/aboutus/regulations>) and in The University’s policy on presentation of Theses

# Acknowledgements

I would like to thank Federal Mogul for their support. I would particularly like to thank Alastair Bissett for his part in initiating the project and the support he has given throughout. I would like to thank Alex Donev for his assistance and advice over the past 3 years and not only with the project; I have learnt a lot. I will be continually grateful to Maria for everything.

# Chapter 1

## Introduction

In mixture experiments products are made by blending two or more ingredients. These ingredients are known as components. Examples of such experiments can be found in numerous different contexts. For example:

1. Petrol blends are blends of organic compounds formed from the fractional distillation of petroleum.
2. Pringles are blends of wheat starch, flours (potato, corn, rice), vegetable oils, salt and seasoning.
3. Marketers choose how their budget is to be split between different advertising media.
4. Brake pad formulations are blends of metal powders with fine powder abrasives and graphite.
5. Glass include blends of oxides such as silicon dioxide and sodium oxide.

In each case, the practitioner is interested in one or more properties of the mixture: in Example 1, they could be interested in the octane rating of the performance of the blend; in Example 2, the optimum balance of ‘hardness’ and fat content in the crisps (Kang, Roshan Joseph, and Brenneman (2011)); in Example 3, the allocation of resources in order to maximise the impact of advertising (van Vuuren and Goos (2012)); in Example 4, flow characteristics of the formulated blend (Brown, Bissett, and Donev (2012)), effectiveness of the brakes under intensive repeated use, or noisiness of the brakes when braking at high speed and in Example 5, any number of properties dependent on the purpose of the glass,

where Piepel and Redgate (1997) have looked at viscosity, chemical durability and crystallinity for use in nuclear waste disposal.

It is assumed there is a functional relationship between the mixture composition, that is the proportions in which the mixture components are blended, and the property of interest. Practitioners attempt to understand this relationship often with the aim of identifying the ‘best’ blend, with respect to the investigated property, or to gain a better general understanding through exploring the roles played by the individual components in this relationship.

The characterising feature of mixture experiments is that the values taken by the mixture components represent proportionate amounts in the mixture rather than unconstrained amounts. Thus, the property under investigation depends on the proportions of the mixture components, but not on the amount of the mixture. Hence, the component proportions may not change independently of each other. It is this which necessitates a distinct statistical strategy for mixture experiments. Cornell (2002) provides the best summary of work on mixture experiments up to the point of its publication.

A comprehensive methodology was first proposed by Scheffé (1958, 1963). This solution was effective and elegant in its treatments of the problem presented and while there have been subsequent developments of these ideas, with manifold objectives, Scheffé’s ideas endure as the primary recourse for practitioners of mixture experiments. Scheffé expressed the functional relationship between the investigated property and mixture components as a statistical model. However, the manner in which this (and subsequent statistical models) characterise this relationship is limited. This thesis will discuss why this is the case, hence explaining the motivation behind the General Blending Models for Mixtures (GBMM), which are the principal novelty of this work.

While one example does demonstrate their application to a constrained experiment, this new class of models are presented in the context of unconstrained experiments only. Additionally, there is no attempt to address experiments including process variables. However, future work can be expected to address both these areas.

The thesis shall be split into four parts: models, model fitting, experimental design and examples.

Chapters 2 and 3 introduce the GBMM. Chapter 2 will establish the context for the new class of models. It shall present the ideas of Scheffé and in

light of these, those of his successors. This will highlight those deficiencies in the preexisting statistical models which it is felt our new class of models addresses. Subsequently, in Chapter 3, the new class of models will be introduced, with a comprehensive discussion of the advantages it presents in relation to these other models. Chapters 4 and 5 address model fitting. The structure of the newly proposed models is unusual. Therefore, they require an unusual method of model fitting. This is discussed in Chapter 4. Chapter 5 then uses a simple example to both illustrate this process and begin to demonstrate the advantages the GBMM has over established models. Chapters 6 and 7 address experimental design. Experimental design is the study of how to collect data in order to achieve the statistical and functional goals of the experiment. Typically, the goals of experimental design relate directly to some feature of the proposed model and the GBMM presents an interesting challenge in this respect. Chapter 6 establishes some background on experimental design, while Chapter 7 proposes both several designs for the GBMM and a new paradigm for developing experimental designs for mixture experiments; one with far greater allowance for model uncertainty than has previously been considered. Finally, Chapter 8 shall give further examples in order to evidence ideas from earlier chapters. We will conclude with some discussion of the work's progress, how far it represents a completed methodology and how it may be developed in the future. R code (R Development Core Team (2010)) allowing the reproduction of the results and figures in the thesis can be obtained from the author, or his supervisor, Dr. Alex Donev.



# Chapter 2

## Methodology

In statistics, response surface methodology explores the relationship between explanatory variables and one or more response variables. A statistical model is a description of a relationship between these variables in the form of mathematical equations. It may describe how one dependent, response variable,  $y$ , is influenced by other independent, explanatory variables  $\mathbf{x}$ , where  $\mathbf{x}$  is a vector of variables  $x_i$ . Such a situation may be formally expressed

$$y_j = f(\mathbf{x}) + \epsilon_j, \quad (2.1)$$

where  $\epsilon_j$  accounts for the influence of random variation upon the  $j^{th}$  observation of the response. The model is identified with the intention of providing the best fit to these observations.

There may be different objectives in identifying such models. They can allow the practitioner to identify the optimum combination of variables in order to obtain a desired response, identify variables with a significant influence upon the response or examine how each variable influences the response. They may also be used for prediction across the entire experimental region, that is across all values of interest of the explanatory variables. This is the principal purpose of the class of models proposed in this thesis. However, as shall be seen, inference may also be made regarding the effects of the variables.

In a mixture experiment, these  $q$  independent variables are known as mixture components and the observed response is dependent only on their relative proportions, rather than their absolute amounts. Taking a simple example of a juice drink formed by the mixing of grapefruit juice and pineapple juice, it can be seen

the values taken by the mixture components are subject to certain characteristic constraints. Taking  $x_1$  to be grapefruit juice and  $x_2$  to be pineapple juice, how tasty the drink is found to be will depend on whether it is half pineapple juice and half grapefruit juice, where  $x_1 = \frac{1}{2}$  and  $x_2 = \frac{1}{2}$ , two thirds pineapple juice and one third grapefruit juice, where  $x_1 = \frac{2}{3}$  and  $x_2 = \frac{1}{3}$ , or all pineapple juice, where  $x_1 = 1$  and  $x_2 = 0$ . It may be desired to have more juice, but this will not change opinion of its flavour.

Hence, in a mixture experiment,

$$\sum_{i=1}^q x_i = 1 \quad (2.2)$$

and

$$0 \leq x_i \leq 1, \quad (2.3)$$

where these constraints define the distinctive geometry of mixture experiments; the unconstrained response surface is described over a  $q - 1$  dimensional simplex representing all possible mixtures of the  $q$  components.

The components may also be subject to additional upper or lower bound constraints such as

$$0 < L_i \leq x_i \leq U_i < 1, \quad (2.4)$$

where  $L_i$  and  $U_i$  are the lower and upper bound constraints, respectively. Moreover, they may be subject to multicomponent linear constraints of the form

$$a < A_1x_1 + \dots + A_qx_q < b, \quad (2.5)$$

where  $a$  and  $b$  are the lower and upper bounds, respectively, and  $A_i$  are scalar constants. In addition it is possible to have nonlinear multicomponent constraints, but these are less common and rarely addressed in the theoretical literature (Atkinson, Donev, and Tobias (2007)). Subjected to constraints, the composition space is the constrained experimental region within the simplex.

The simplest forms of statistical model are the polynomial models, where  $f(\mathbf{x})$  in Equation 2.1 is a polynomial function, that is a function of the variables  $\mathbf{x}$  and estimable parameters  $\beta$ , formed only from addition, subtraction, multiplication and non-negative integer exponents. Examples are the linear, quadratic and cubic

polynomial models:

$$E[y] = \beta_0 + \sum_{i=1} \beta_i x_i, \quad (2.6)$$

$$E[y] = \beta_0 + \sum_{i=1} \beta_i x_i + \sum_{i=1} \beta_{ii} x_i^2 + \sum_{i \neq j} \beta_{ij} x_i x_j \quad (2.7)$$

and

$$\begin{aligned} E[y] = & \beta_0 + \sum_{i=1} \beta_i x_i + \sum_{i=1} \beta_{ii} x_i^2 + \sum_{i \neq j} \beta_{ij} x_i x_j \\ & + \sum_{i=1} \beta_{iii} x_i^3 + \sum_{i \neq j} \beta_{iij} x_i^2 x_j + \sum_{i \neq j \neq k} \beta_{ijk} x_i x_j x_k, \end{aligned} \quad (2.8)$$

where  $\beta_0$  represents a mean effect,  $\beta_i$  main effects,  $\beta_{ii}$  and  $\beta_{iii}$  curvature effects and other terms interaction effects, where two or more variables work together to influence the response. However, in the context of mixture experiments, the Characteristic constraints 2.2 and 2.3 render resolution of the response into mean effects, main effects and interaction effects meaningless. Therefore, ordinary polynomial models may not be applied to mixture experiments.

To see why, take the linear polynomial model and introduce an arbitrary constant  $\gamma \neq 0$ :

$$\begin{aligned} E[y] &= \beta_0 + \sum_{i=1} \beta_i x_i \\ &= \beta_0 + \gamma - \gamma + \sum_{i=1} \beta_i x_i \\ &= \beta_0 + \gamma - \gamma(1) + \sum_{i=1} \beta_i x_i \\ &= \beta_0 + \gamma - \gamma \left( \sum_{i=1}^q x_i \right) + \sum_{i=1} \beta_i x_i \quad \text{as } \sum_{i=1}^q x_i = 1 \\ &= (\beta_0 + \gamma) + \sum_{i=1} (\beta_i + \gamma) x_i \\ &= \beta'_0 + \sum_{i=1} \beta'_i x_i. \end{aligned}$$

Thus it is shown that, for mixture experiments, the linear polynomial model may

not be uniquely determined as

$$\beta_0 + \sum_{i=1} \beta_i x_i = \beta'_0 + \sum_{i=1} \beta'_i x_i, \quad (2.9)$$

where

$$\beta_i \neq \beta'_i \quad 0 \leq i \leq q. \quad (2.10)$$

It is similarly possible to demonstrate that higher order polynomial models, such as the quadratic and cubic models, may not be uniquely determined for mixture experiments. Consequently, it was necessary to develop alternative statistical models for mixture components.

Before 1958, little was done to achieve this and it was not until Scheffé (1958, 1963) that a comprehensive methodological approach was presented. Prior to this ordinary polynomial models in mathematically independent variables (MIV) had been used and these proved to be used occasionally by later practitioners. Clar- ingbold (1955), Draper and Lawrence (1965a,b), Thompson and Myers (1968), and Becker (1970) consider cases where the MIV are linear combinations of the component proportions, while Hackler, Kriegel, and Hader (1956) and Kenwor- thy (1963) take the MIV to be ratios of the component proportions. Both these strategies resolve the linear dependence of the component proportions, but in a manner which may complicate the understanding of their influence upon the re- sponse. By contrast, Scheffé's canonical polynomial models allowed the response to be represented directly as a function of the component proportions. Conse- quently, his work proved the impetus for mixture experiments as a distinct area of research and the models he developed remain the recourse for many practitioners today.

This said, they are not without criticism, nor have later researchers failed to indicate dissatisfaction with Scheffé's models through the proposal of alternative parameterisations or extensions. The linear blending of Becker's models, the reparameterisations of Darroch and Waller (1985), Piepel and Cornell (1994), Piepel, Szychowski, and Loeppky (2001) and Draper and Pukelsheim (1998), and the inverse terms of Draper and St John (1977a,b) and the log terms of Chen, Zhu, and Hu (1985) have all been proposed with varied claims for their benefits; some describe behaviour which the Scheffé polynomials cannot, while others promise more parsimonious models, that is models with fewer estimable parameters. Each

of these alternative models shall be discussed as it is felt there presence reveals the motivation for an alternative to the Scheffé polynomials.

To date, there has been little work done on nonlinear models for mixture experiments. Linear models, such as the polynomial models, may be expressed

$$E[y] = f(\mathbf{x}, \beta) = \beta f_i(\mathbf{x}), \quad (2.11)$$

that is as a weighted sum of functions of  $\mathbf{x}$ . Nonlinear models differ in the presence of the nonlinear parameters  $\theta$ , such that

$$E[y] = f(\mathbf{x}, \theta). \quad (2.12)$$

In particular, partially linear models may be expressed

$$E[y] = \beta f_i(\mathbf{x}, \theta). \quad (2.13)$$

The class of models proposed in this thesis are partially linear. They can be considered novel for being broadly applicable nonlinear models, well seated within the canon of mixture experiments literature. The only case of other nonlinear models proposed for mixture experiments are given by Focke, Ackermann, and Coetzer (2012), who discuss the generalised Q-fraction mixture models, with respect to a limited applications. These models do not appear broadly applicable. However they are certainly worthy of discussion, particularly in relation to how models for mixture experiments will progress.

The following chapter looks at all of these proposed models for mixture experiments. This will inform both the discussion of how the GBMM was developed, but also the analysis of its merits and how far it goes to providing a comprehensive modelling strategy for mixture experiments. It is not a comprehensive discussion of all literature relating to the modelling of mixture experiments, but it is thorough in providing a background to the GBMM.

## 2.1 Scheffé polynomials

Priority in the treatment of designs for mixture experiments may be claimed by Quenouillé (1953). However, the work of Scheffé (1958, 1963) proved the greater impetus for increased interest in the analysis of mixture experiments. Prompted

by the ideas of Claringbold (1955), Scheffé introduced easily interpretable canonical polynomial models in the mixture components. For the reasons given above, it was not possible to apply ordinary polynomial models to the mixture components and previously the response in mixture experiments had been modelled through standard polynomials in mathematically independent variables (MIV), these being either linear transformations or ratios of the component values. However, as Scheffé's polynomials and other mixture models became better known the use of ordinary polynomials in MIV declined.

In 1958, Scheffé introduced the simplex-lattice designs alongside his first canonical polynomial models for mixture experiments. Subsequently, in 1963, he proposed the simplex-centroid designs and corresponding alternative models, to which extensions were later given by Gorman and Hinman (1962) and Lambrakis (1969). This section looks at Scheffé's models, introducing them in a manner that makes note of how they are relevant to the new ideas to be discussed later. Key to this are some criticisms levelled at Scheffé's work at the time of its inception. Of these, those of Quenouillé are particularly important.

Having explained some of the significant early ideas, this section should provide a satisfactory introduction to the methods used in experiments with mixtures. However, it is important to stress that while Scheffé's work precedes almost all of what we are to discuss it endures, often without refinement, as the methodological recourse of practitioners today.

### 2.1.1 Simplex lattice designs and canonical polynomial models

In mixture experiments, if it is desired to model the response in a  $q$ -component mixture experiment, an intuitively appealing design is that where the design points are distributed evenly across the experimental region. Thus motivated, Scheffé proposed the simplex lattice designs. For these the values taken by the component proportions are the  $m + 1$  equally spaced values  $x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$ , such that the  $\{q, m\}$  lattice designs are all

$$\binom{m+q+1}{m} = \frac{(m+q+1)!}{m!(q+1)!} \quad (2.14)$$

possible mixtures where the component proportions take these values.

The  $\{q, m\}$  ordinary polynomial model

$$E[y] = \alpha_0 + \sum_{i=1}^q \alpha_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \alpha_{ij} x_i x_j + \dots, \quad (2.15)$$

of interactions up to degree  $m$ , where the  $\alpha$  values are the model coefficients, are meaningful in the context of mixture experiments only in light of the Characteristic constraint 2.3. Scheffé instead proposed symmetric canonical polynomial models formed through suitable substitution into Model 2.15 of identities dictated by the Characteristic constraint 2.3. This resulted in polynomials of  $\binom{m+q+1}{m}$  coefficients, which could be uniquely determined and simply interpreted, with respect to the mixture components. Additionally, they possess a simple, intuitive relationship to  $\{q, m\}$  simplex lattice designs, which, even at a time without computers, allowed them to be easily estimated.

In this way, the  $\{q, 1\}$  linear Scheffé polynomial,

$$E[y] = \sum_{i=1}^q \beta_i x_i, \quad (2.16)$$

is obtained through replacement of  $\alpha_0$ , in the ordinary  $\{q, 1\}$  polynomial, with  $\alpha_0 (\sum_{i=1}^q x_i)$ . Hence,

$$E[y] = \alpha_0 + \sum_{i=1}^q \alpha_i x_i \quad (2.17)$$

$$= \alpha_0 \left( \sum_{i=1}^q x_i \right) + \sum_{i=1}^q \alpha_i x_i \quad (2.18)$$

$$= \sum_{i=1}^q (\alpha_0 + \alpha_i) x_i \quad (2.19)$$

$$= \sum_{i=1}^q \beta_i x_i. \quad (2.20)$$

The quadratic Scheffé polynomial is found where, in addition, each  $x_i^2$  in the ordinary  $\{q, 2\}$  polynomial

$$E[y] = \alpha_0 + \sum_{i=1}^q \alpha_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \alpha_{ij} x_i x_j + \dots, \quad (2.21)$$

is replaced by

$$x_i^2 = x_i \left( 1 - \sum_{j \neq i} x_j \right) = x_i - \sum_{j \neq i} x_i x_j. \quad (2.22)$$

The quadratic, cubic and quartic Scheffé polynomials are then

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j}^q \beta_{ij} x_i x_j, \quad (2.23)$$

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j}^q \beta_{ij} x_i x_j + \sum_{i \neq j}^q \beta_{ij} x_i x_j (x_i - x_j) \quad (2.24)$$

$$+ \sum_{i \neq j \neq k} \beta_{ijk} x_i x_j x_k \quad (2.25)$$

and

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j}^q \beta_{ij} x_i x_j + \sum_{i \neq j}^q \beta_{ij} x_i x_j (x_i - x_j) \quad (2.26)$$

$$+ \sum_{i \neq j}^q \beta_{ij} x_i x_j (x_i - x_j)^2 + \sum_{i \neq j \neq k} \beta_{ijk} x_i^2 x_j x_k \quad (2.27)$$

$$+ \sum_{i \neq j \neq k \neq l} \beta_{ijkl} x_i x_j x_k x_l, \quad (2.28)$$

respectively. The quartic model, not given by Scheffé, was proposed by Gorman and Hinman in their critique of Scheffé's method. They cited a lack of complexity in the response surface which could be described by Scheffé's polynomials and therefore introduced another more complex model. This lack of complexity also motivates the ideas presented in this thesis.

A general form for a polynomial of degree  $m$  was suggested by Lambrakis (1969)

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{n=2}^m \sum_{i \neq j}^q \beta_{ij} x_i x_j (x_i - x_j)^{n-2} + \sum_{n=3}^m \sum_{1 \leq i_1 < \dots < i_n \leq q} x_{i_1}^{s_1} \dots x_{i_n}^{s_n}, \quad (2.29)$$

where  $\mathbf{s}$  is defined subject to the constraint  $3 \leq n \leq m$ . Through adding the



cubic terms  $+\sum_{i \neq j \neq k} \beta_{ijk} x_i x_j x_k$  to the Model 2.23, the special cubic polynomial,

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j}^q \beta_{ij} x_i x_j + \sum_{i \neq j \neq k} \beta_{ijk} x_i x_j x_k, \quad (2.30)$$

may also be obtained. This extension, suggested by Scheffé, was a precursor to the models he would later propose in 1963.

The Scheffé polynomials require different interpretations of the coefficients compared to those applied to Model 2.15. Let  $\eta_i$  denote the response where  $x_i = 1$  and  $x_j = 0, i \neq j$ , and suppose only pure component effects are observed. The main component effect

$$\beta_i = \eta_i. \quad (2.31)$$

Therefore, more generally, the main effects  $\beta_i$  can be interpreted as giving the response for a pure mixture of the  $i^{th}$  component.

In light of this, the other coefficients are best described with respect to the deviation from the linear blending described by these pure component effects. These deviations are here said to occur through joint effects between the components. This is in contrast to describing them as interaction terms, as found in the ordinary polynomial models, which would indicate the effect of a variable is dependent on the value of another.

Hence, in the Scheffé polynomials,  $\beta_{ij}$  are the quadratic coefficients of binary joint effect between the  $i^{th}$  and  $j^{th}$  components. Where the cubic model is fitted,  $\gamma_{ij}$  are the cubic coefficients of binary joint effect and the  $\beta_{ijk}$  cubic coefficients of ternary joint effect.  $\delta_{ij}$  and  $\gamma_{ijk}$  in the quartic model are the quartic coefficients of binary and ternary joint effect, respectively.

The characteristic constraint of mixture experiments dictates that mixture components are non-orthogonal. This is the reason for the contrasting interpretation of interaction terms in ordinary polynomials and terms of joint effect in the Scheffé polynomials. For this reason, independence of the effect of variables is a more unusual property in mixture components; removal of terms of joint effect in a particular component does not imply independence of the effect of that component in a mixture polynomial.

### 2.1.2 Simplex centroid designs and Scheffé $q$ -tenery models

A potential objection to the simplex lattice designs was that they were intended to be used to predict the response for mixtures of  $q, q-1, q-2, \dots$  components, but fit a model using the observed responses of mixtures of up to  $m$  components. Thus the  $\{3, 2\}$  simplex is used to fit a quadratic model, which may be used to predict the response for mixtures of three components, but is fit using the responses for mixtures of up to two components. In reality this is unlikely to be the case and data is likely to be observed for values outside of these designs. However, when Scheffé's polynomials were first proposed, it was judged to be a failing that needed to be addressed and hence inspired a new model form.

The simplex centroid design overcomes this problem by taking the response at those mixtures where up to  $q$  components take equal proportions in the mixture. Thus it includes pure mixtures, those where two components  $x_i = x_j = 1/2$ , three components  $x_i = x_j = x_k = 1/3$  and so forth. Using these designs, Scheffé chose to fit a model which again had an equal number of coefficients as points in the design:

$$\begin{aligned}
 E[y] = & \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j}^q \beta_{ij} x_i x_j + \sum_{i \neq j \neq k} \beta_{ijk} x_i x_j x_k \\
 & + \sum_{i \neq j \neq k \neq l} \beta_{ijkl} x_i x_j x_k x_l + \dots
 \end{aligned} \tag{2.32}$$

Here the  $\beta_i$  are the main effects,  $\beta_{ij}$  the binary joint effects between the  $i^{th}$  and  $j^{th}$  components,  $\beta_{ijk}$  the ternary joint effects between the  $i^{th}$ ,  $j^{th}$  and  $k^{th}$  components and so forth. Thus, in this model, each coefficient is associated with a particular sub-simplex within the factor space and has no effect upon mixtures for which any of its associated components equal zero. Hence, for example,  $\beta_{ij}$  has no effect upon the response of any mixture where  $x_i = 0$  or  $x_j = 0$ .

With hindsight these  $q$ -tenery models can be said to be redundant beyond those special cases which are included among the Scheffé canonical polynomials; there is almost no record of their application beyond those of the quadratic and special cubic model, which could be viewed as reduced cases of the earlier proposed Models 2.23 and 2.30 (where some terms have been eliminated).

### 2.1.3 What can be described using Scheffé's polynomials?

An important motivating factor for the ideas given in this thesis is the limitations on what types of surfaces and particularly, what joint effects can be described by the Scheffé polynomials. As has been noted, this was also a consideration for Gorman and Hinman, prompting them to propose a higher order model than given by Scheffé.

To consider the particular joint effects described by the canonical polynomial models and  $q$ -tenery model we can first consider the two component case. For a binary mixture described by a quadratic model the response surface may have a maximum or a minimum, but not both, and no point of inflection. Any departure from linearity, as created by the quadratic terms of binary joint effect, is symmetric with maximum departure occurring at the 50:50 mixture. More generally, this will be the case for the second order  $q$ -tenery model or the quadratic canonical polynomial model and the restriction upon the maximum departure from linearity in binary mixtures endures for the cubic and quartic models, for which it may lie between 21% and 79%  $x_1$  and 15% and 85%  $x_1$ , where the cubic and quartic terms of binary joint effect, respectively, have offered increased flexibility.

These two cases are relevant to binary and ternary joint effects. The  $q$ -tenery model offers terms describing joint effects in 4 or more components. For  $q$  component cases the restrictions upon the  $q$ -tenery model can be described thus: in Scheffé's terminology, the 'pull-up' effect (possibly negative) of each term of joint effect is symmetric in the components involved. This is seen in the particular case of the quadratic terms of binary synergism  $\beta_{ij}x_i x_j$ , which cause a symmetric effect about the 50:50 mixture. This concept extends to more than 2 components. Thus the departure from linearity effected by each term is greatest in the centre of the sub-simplex to which that term applies. Therefore, for example, in a ternary mixture experiment where the quadratic Scheffé polynomial,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \sum_{i \neq j}^3 \beta_{ij} x_i x_j, \quad (2.33)$$

is fitted, adding the  $\beta_{ijk}x_i x_j x_k$  term, assuming all other coefficients remain the same, results in a symmetric change in the response surface, about the centre of three dimensional simplex, where  $x_i = x_j = x_k = \frac{1}{3}$ , and similar statements can be made when adding the terms of the  $q$ -tenery models in greater than 3

components. For this reason, these terms can be viewed as severely limited in what they can describe. This will prove important as new ideas are proposed in later chapters.

### 2.1.4 Comments upon Scheffé's ideas

Finally, we explore the comments levelled at Scheffé in the discussion given in the appendix of his 1963 paper. Nadler, Cox and Plackett each made noteworthy contributions relevant to the development of later ideas. However, the most significant comments, with respect to the ideas presented here, were made by Quenouillé.

Quenouillé's work on mixtures predated Scheffé's, but lacked its appeal. His contention was that Scheffé's models were the poorer for not allowing certain situations he would like to describe. Let A, B and C be the vertices of the 2-dimensional simplex, attributed to the pure mixtures  $x_1 = 1$ ,  $x_2 = 1$  and  $x_3 = 1$ , respectively, as shown in Figure 2.1a on page 31. Supposing D to be any point along the edge between A and B, that is the edge representing binary mixtures of  $x_1$  and  $x_2$ , then the reduced Scheffé polynomial,

$$E[y] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2, \quad (2.34)$$

fails to describe linear blending in  $x_3$ , that is a lack of curvature in the response surface between C to D. For Quenouillé, the curvature that this model instead describes is the result of a joint effect between the component  $x_3$  and supercomponent  $x_1 + x_2$ , which cannot be attributed to the joint effect between  $x_3$ , and  $x_1$  and  $x_2$ , respectively.

In response, Scheffé reiterated that he considered that traditional understanding of interaction to be unsuitable for mixture experiments, and suggested that a model of the form

$$E[y] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} \frac{x_1 x_2}{x_1 + x_2}, \quad (2.35)$$

would represent what Quenouillé described. This model would be the basis of Becker's work on linear blending, which shall be introduced in the next section. Furthermore, its use of a supercomponent, a sum of two or more components, gives an example of how Scheffé suggests other types of curvature in the blending

of components may be represented. These ideas will be significant as we explore the alternative mixture polynomials to those proposed by Scheffé and moreover, will feed directly into the new ideas to be proposed later.

Also important, in this respect, is the contribution of Plackett, who identifies that the Scheffé polynomials, instead of predicting the response, take their respective designs and use the model in effect as a smoothing formula. Each model, when fit to its allocated design, will pass through the mean of the observations at each of the design points and therefore additional points are required in order that the models can be checked for goodness of fit. This means that the models may appear to be artificially well fit, where the Scheffé polynomial may in fact be the wrong choice of model. More generally, this statement identifies that the choice of the Scheffé polynomials, by a practitioner, contains inherent assumptions about the type of response surface which is expected and regarding how the individual components influence the response.

Quenouillé highlights the choice of the correct form of polynomial is therefore important and hence, models and designs are required which can explore more varied behaviour than that investigated by the Scheffé's polynomials. Plackett in turn compounds Quenouillé's point by suggesting that non-polynomial methods of modelling mixtures could be beneficial, foreshadowing the ideas proposed here.

### 2.1.5 Concluding remarks on Scheffé polynomials

The Scheffé polynomials present an effective recourse for practitioners of mixture experiments. There is lack of flexibility in the type of behaviour which may be described, particularly by the lower order models and the  $q$ -tenery models, but in many cases this will not sufficiently hinder the practitioner in obtaining apparently effective statistical analysis. That said, Gorman et al and Quenouillé correctly identify deficiencies in what they may describe, and the statements of Plackett and Quenouillé create space to explore different model forms describing alternative or more varied behaviour. The following section presents the alternatives developed by Becker, who was building directly on the ideas of Scheffé in response to Quenouillé.

## 2.2 Becker's models

As was discussed in the previous section, Quenouillé (1959) criticised Scheffé for adopting a polynomial smoothing method, which assumed a specific form of polynomial and collected data accordingly. Scheffé's models failed to describe the linear blending in the component proportions which Quenouillé found intuitively sensible. He suggested some alternative model should be devised, but failed to propose any such solution. Instead Scheffé made a suggestion which adapted his own canonical polynomial forms. This suggestion would go on to be developed by Becker (1968, 1978), who would make proposals which addressed not only linear blending, but also inactivity in the effect of the components. This section shall explain further the nature of Quenouillé's criticism and examine how Becker's ideas developed to address them.

### 2.2.1 Linear and quadratic blending

The reduced 3 component Scheffé polynomial,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} x_1 x_2, \quad (2.36)$$

can be said to describe quadratic curvature in the blending of the components,  $x_1$  and  $x_2$ , with the remainder of the mixture. This shall be referred to as quadratic blending. This is in contrast to the linear blending Quenouillé desired Scheffé's models to describe.

To identify what this means, look at Figure 2.1a, where A, B and C are the vertices of the 3-dimensional simplex attributed to the pure mixtures  $x_1 = 1$ ,  $x_2 = 1$  and  $x_3 = 1$ , respectively. For arbitrary D along the edge between A and B, the components  $x_1$  and  $x_2$  remain in fixed relative proportion to each other along the line between C and D, that is

$$\frac{x_1}{x_1 + x_2} = \alpha$$

and

$$\frac{x_2}{x_1 + x_2} = 1 - \alpha.$$

for  $0 \leq \alpha \leq 1$ .

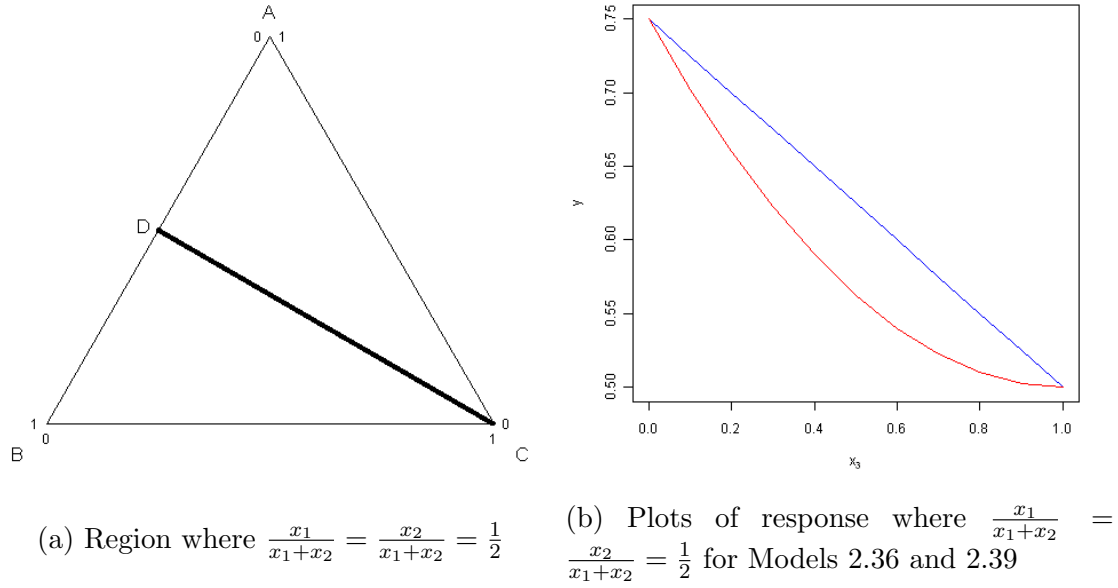


Figure 2.1: Illustration of the difference in Quenouillé and Scheffé's understanding of component blending

In the case illustrated

$$\frac{x_1}{x_1+x_2} = \frac{x_2}{x_1+x_2} = \frac{1}{2}. \quad (2.37)$$

The response, in this case, as described by Model 2.36, is given as

$$E[y] = \frac{\beta_1 + \beta_2}{2} + \left( \beta_3 - \frac{\beta_1 + \beta_2}{2} \right) x_3 + \frac{\beta_{12}}{4} (1 - x_3)^2, \quad (2.38)$$

and is shown by the red line in Figure 2.1b for  $\beta_1 = \beta_2 = \beta_{12} = 1$ . In contrast, the response as described by the model

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1 x_2}{x_1 + x_2}, \quad (2.39)$$

is given by

$$E[y] = \frac{\beta_1 + \beta_2}{2} + \left( \beta_3 - \frac{\beta_1 + \beta_2}{2} \right) x_3 + \frac{\beta_{12}}{4} (1 - x_3) \quad (2.40)$$

and is shown in blue in Figure 2.1b.

More generally, for any ray where  $x_1$  and  $x_2$  remain in fixed relative proportion to each other Model 2.36 will describe a quadratic response, with respect to  $x_3$ , while Model 2.39 will describe a linear response, with respect to  $x_3$ . This is to say Model 2.39 provides the kind of linear blending Quenouillé felt was not described by Scheffé's model. In this case, it may be said that Model 2.39 allows component  $x_3$  to blend linearly with the remainder of the mixture, that is without curvature in the response as the component value  $x_3$  changes.

### 2.2.2 Becker: New models

Developing on Scheffé's proposal, Becker (1978) identified that a more general model describing linear blending in a component  $x_q$  would take the form

$$E[y] = \beta_q x_q + (1 - x_q) f(x_1 / (1 - x_q), \dots, x_{q-1} / (1 - x_q)), \quad (2.41)$$

where the interpretations of the terms in  $f$  are invariant under the introduction of the linear blending component, that is where  $f$  is homogenous of degree one. A function  $f$  is homogenous of degree one where

$$f(t\mathbf{x}) = tf(\mathbf{x}). \quad (2.42)$$

To allow easier fitting Becker suggests that  $f$  may be expressed

$$f(\mathbf{x}) = \beta_i h_i(x_i) + \beta_{ij} h_{ij}(x_i, x_j) + \dots + \beta_{ij\dots q-1} h_{ij\dots q-1}(x_i, x_j, \dots, x_{q-1}), \quad (2.43)$$

where each term  $h$  is homogenous of degree one. Becker describes three such terms which may be employed:

$$h_{ij\dots k}(x_i, x_j, \dots, x_k) = \min \left( \frac{x_i}{1 - x_q}, \frac{x_j}{1 - x_q}, \dots, \frac{x_k}{1 - x_q} \right), \quad (2.44)$$

$$h_{ij\dots k}(x_i, x_j, \dots, x_k) = \frac{x_i x_j \dots x_k}{1 - x_q} \quad (2.45)$$

and

$$h_{ij\dots k}(x_i, x_j, \dots, x_k) = \sqrt{\frac{x_i x_j \dots x_k}{1 - x_q}}. \quad (2.46)$$

where each of these three types of terms can be used to construct the H1, H2 and H3 models, respectively. For a three component mixture experiment, where  $x_3$  is



said to blend linearly, these models are

$$E[y] = \beta_3 x_3 + (1 - x_3) \left( \beta_1 \frac{x_1}{1 - x_3} + \beta_2 \frac{x_2}{1 - x_3} + \beta_{12} \min \left( \frac{x_1}{1 - x_3}, \frac{x_2}{1 - x_3} \right) \right), \quad (2.47)$$

$$E[y] = \beta_3 x_3 + (1 - x_3) \left( \beta_1 \frac{x_1}{1 - x_3} + \beta_2 \frac{x_2}{1 - x_3} + \beta_{12} \frac{x_1 x_2}{1 - x_3} \right) \quad (2.48)$$

and

$$E[y] = \beta_3 x_3 + (1 - x_3) \left( \beta_1 \frac{x_1}{1 - x_3} + \beta_2 \frac{x_2}{1 - x_3} + \beta_{12} \sqrt{\frac{x_1 x_2}{(1 - x_3)^2}} \right), \quad (2.49)$$

respectively. Here, the H2 model is that proposed by Scheffé in response to Quenouillé as it may be expressed

$$E[y] = \sum_i^3 \beta_i x_i + \beta_{12} \frac{x_1 x_2}{1 - x_3}. \quad (2.50)$$

Similarly, the H1 and H3 models may be expressed as

$$E[y] = \sum_i^3 \beta_i x_i + \beta_{12} \min(x_1, x_2) \quad (2.51)$$

and

$$E[y] = \sum_i^3 \beta_i x_i + \beta_{12} \sqrt{x_1 x_2}, \quad (2.52)$$

respectively. The terms proposed by Becker may of course be used in other models which do not describe strict linear blending in any one component but do provide the best fit to the collected data. Each term permits an alternative description in comparison to a particular term in the Scheffé polynomials and may prove a beneficial addition to a model. For example, it may be that a model which combines both Becker and Scheffé terms, such as

$$E[y] = \sum_i^3 \beta_i x_i + \beta_{12} \frac{x_1 x_2}{1 - x_3} + \beta_{23} x_2 x_3, \quad (2.53)$$

provides the best overall fit to a set of data.

More generally, Becker's work suggests three models as alternatives to the

Scheffé polynomials. These are the full H1, H2 and H3 models, respectively given by

$$E[y] = \sum_i \beta_i x_i + \sum_{i \neq j} \beta_{ij} \min(x_i, x_j) + \sum_{i \neq j \neq k} \beta_{ij} \min(x_i, x_j, x_k) + \cdots, \quad (2.54)$$

$$E[y] = \sum_i \beta_i x_i + \sum_{i \neq j} \beta_{ij} \frac{x_i x_j}{x_i + x_j} + \sum_{i \neq j \neq k} \beta_{ij} \frac{x_i x_j x_k}{(x_i + x_j + x_k)^2} + \cdots \quad (2.55)$$

and

$$E[y] = \sum_i \beta_i x_i + \sum_{i \neq j} \beta_{ij} \sqrt{x_1 x_2} + \sum_{i \neq j \neq k} \beta_{ij} \sqrt[3]{x_1 x_2 x_3} + \cdots, \quad (2.56)$$

where each term is homogenous of degree one. Hence, reduced forms of these models are capable of describing linear blending in any one component. However, it is not necessary to apply these models in this way and it could just be that application of the terms of these models, instead of those found in the Scheffé polynomials, provides the best possible description of the response surface or a more parsimonious description. In this context, it is better to think of the linear blending described by a term  $\frac{x_i x_j}{x_i + x_j}$  to be between the pair of components  $x_i$  and  $x_j$  and the remainder of the mixture.

Choosing between the binary term of joint effect,  $x_1 x_2$ , of the Scheffé polynomial and the binary term of joint effect,  $\frac{x_1 x_2}{x_1 + x_2}$ , of Becker's H2 model would go some way to testing the inherent assumptions of the 'polynomial smoothing method' that Plackett identified Scheffé's method to represent. It also leads to the question of which further alternatives could be proposed to the models of Scheffé.

### 2.2.3 Further linear blending models

In a subsequent paper, Becker (1978) extended his ideas for the treatment of linear blending components. This was done through proposing the more general term

$$h_{ijq-1}(x_i, x_j, \dots, x_{q-1}) = \prod_{i=1}^{q-1} \left( \frac{x_i}{1 - x_q} \right)^{s_i}, \quad (2.57)$$

which includes all possible terms of the type included in the H2 and H3 models, but also provides asymmetry in the joint effect between the components included in the term, thus addressing one of the significant shortcomings of the Scheffé

polynomials. Very ‘unusual’ behaviour could be most readily described through Scheffé type polynomials by the extension proposed by Gorman and Hinman - the quartic polynomial. Yet this is a relatively complex model to describe what potentially could be described by far fewer terms of the type suggested by Becker.

However, although Becker’s terms allow particular behaviour to be described in a manner which was not possible with the Scheffé polynomials, they may not necessarily be applied in a straight forward manner. This is because the estimation of a large number of nonlinear parameters is likely to be problematic. It is probably for this reason that Becker does not apply these models and consequently his 1978 paper has seen little attention although, as shall be seen, his work is very closely related to the class of models which are proposed here.

### 2.2.4 Summary

Becker’s models give the potential to describe linear blending in one or more components in the manner which Quenouillé considered useful. More importantly, the full H1, H2 and H3 models represent an alternative to the models of Scheffé and the work of Becker’s 1978 paper indicates a path towards a far more flexible modelling strategy. Both these points are influential in the work which is presented here.

However, while Becker’s proposed alternatives to the models of Scheffé are certainly the most important with respect to our work, they are not the only alternatives and, as much as the Scheffé polynomials possess certain inherent assumptions about the influence of the components upon the response, Becker’s models possess similar assumptions. Therefore, there are other behaviours which neither sets of models can represent, some of which were considered important to other practitioners.

## 2.3 Inverse and logarithmic terms

A phenomenon that may occur in mixture experiments is extreme change in the response at the boundary of the experimental region, for example, where  $x_i \rightarrow 0$ . In general, neither the Scheffé polynomial models nor the Becker models can handle this behaviour. To deal with these cases the model

$$f(\mathbf{x}) \rightarrow \inf, \quad (2.58)$$

as  $x_i \rightarrow 0$ , is needed. Ratio terms which, before Scheffé proposed his canonical polynomial models, were commonly applied in mixture experiments as MIV in ordinary polynomial models, describe such types of behaviour. For example,  $x_1/(x_2 + x_3)$  and  $x_1/x_3$  describe extreme change as  $x_2 + x_3 \rightarrow 0$  and  $x_3 \rightarrow 0$ , respectively. However, their effect is dependent on the value of  $x_1$ . Draper and St John (1977a,b) propose instead

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j} \beta_{ij} x_i x_j + \sum_{i=1}^q \beta_{-i} x_i^{-1}, \quad (2.59)$$

that is, the Scheffé polynomial models with additional inverse terms of the individual mixture components. These preserved a model in the mixture components rather than MIV, while allowing extreme behaviour through the new inverse terms  $\beta_{-i} x_i^{-1}$ .

Although in this case  $x_i = 0$  is the active boundary of interest, to allow fitting, the experimental region should not actually be considered to include the boundary itself. Hence there exists the implicit constraint,

$$x_i > 0, \quad (2.60)$$

imposed upon  $x_i$  in the inverse term. To achieve this the practitioner may add a small value  $c_i$ . Thus the inverse terms take the form

$$\beta_{-i} \frac{1}{x_i + c_i}. \quad (2.61)$$

Draper proposed the inverse terms be fitted in addition to the terms of the Scheffé polynomials. Of course, they may equally be added to the Becker models. Akay (2007) discussed the comparison of Becker models and inverse polynomial models of the type described, but did not consider inclusion of inverse terms and Becker terms in the same model.

An alternative to the inverse terms are logarithmic terms. These were first suggested by Chen et al. (1985). The expectation of a practitioner wishing to apply inverse terms is that the extreme behaviour will be relevant to values of  $x_i$  approximately less than 0.05, in line with the suggestions of Draper and St John; Draper and St John. This means the change in the effect is perceived to be less steep than that described by a logarithmic term. Hence, the models proposed by

Chen, such as

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j} \beta_{ij} x_i x_j + \sum_{i=1}^q \beta_{-i} \log(x_i), \quad (2.62)$$

the Scheffé polynomials with additional logarithmic terms of the individual components, consider the extreme behaviour to be relevant for lower values of  $x_i$  than the inverse terms.

As a result of the appeal of the polynomial models, inverse terms and particularly logarithmic terms have not been considered much beyond the papers already mentioned. This may be attributed to the limited nature of the option they offer - it is a manner of handling more extreme behaviour, but in a very specific way. While their existence suggests dissatisfaction with the limitations of the behaviour described in models already discussed, they do not allow a very broad range of extreme behaviour to be explored; they do not describe extreme behaviour where values other than  $x_i$  approach zero.

A better solution might be to introduce a term or terms which not only allowed extreme change to be described as  $x_i \rightarrow 0$ , but also at other extremes of the experimental region, that is, say, at the two component edges where  $x_i + x_j \rightarrow 0$ , or the three component sides where  $x_i + x_j + x_k \rightarrow 0$ .

## 2.4 Reparameterisations

The models of Becker present terms describing contrasting joint effects in comparison to those of the Scheffé polynomial. Therefore, these models present the possibility of describing the response surface in a different and potentially more effective manner, regardless of whether one or more components are described as blending linearly. The logarithmic and inverse terms also allow the response surface to be described in a contrasting manner to the Scheffé polynomials.

A different manner in which to improve upon the Scheffé polynomials is to suggest reparameterisations. These will describe the same behaviour, but allow more parsimonious representation of the response surface through the use of equivalent terms. For simplicity, this section will focus on the reparameterised quadratic Scheffé polynomial, but there are equivalent reparameterisations for the higher order models.

To introduce such reparameterisations consider a three component experiment

in which the response is described by the model:

$$\begin{aligned}
 E[y] &= 3x_1 + 4x_2 + 5x_3 + 15x_1x_2 + 15x_1x_3, \\
 &= 3x_1 + 4x_2 + 5x_3 + 15x_1(x_2 + x_3), \\
 &= 3x_1 + 4x_2 + 5x_3 + 15x_1(1 - x_1), \\
 &= 3x_1 + 4x_2 + 5x_3 + 15x_1(1 - x_1), \\
 &= 3x_1 + 4x_2 + 5x_3 + 15x_1 - 15x_1^2, \\
 &= 18x_1 + 4x_2 + 5x_3 - 15x_1^2.
 \end{aligned}$$

Hence, it is seen that having considered the square term  $x_1^2$ , a more parsimonious description is given of the response - there are fewer terms. This is achieved due to the equivalence of component effects through the Characteristic constraint 2.2. The differences between methods of reparameterising the Scheffé polynomials depend upon the manner in which the identities due to the Characteristic constraint 2.2 are used.

### Slack variable models

A quadratic slack variable model, for a  $q$ -component mixture experiment, describes the response as a function of  $q - 1$  components. One component  $x_i$  is denoted slack and is removed from the model through the identity  $x_i = 1 - \sum_{j \neq i} x_j$ . Its linear effect is represented by an intercept and its joint effects absorbed into square terms and the terms of binary joint effect of the other components. Thus, in a ternary experiment, through substitution of  $x_3$  in the quadratic canonical Scheffé polynomial,

$$E[y] = \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3, \quad (2.63)$$

the slack variable model,

$$E[y] = \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \alpha_{11}x_1^2 + \alpha_{22}x_2^2 + \alpha_{12}x_1x_2, \quad (2.64)$$

may be obtained, using the identity  $x_3 = 1 - x_1 - x_2$ . Here

$$\begin{aligned}\alpha_0 &= \beta_3, \\ \alpha_1 &= \beta_1 - \beta_3 + \beta_{13}, \\ \alpha_2 &= \beta_2 - \beta_3 + \beta_{23}, \\ \alpha_{12} &= \beta_{12} - \beta_{13} - \beta_{23}, \\ \alpha_{11} &= -\beta_{13}\end{aligned}$$

and

$$\alpha_{22} = -\beta_{23}.$$

These two full-model representations describe the same predicted response, but the square terms of the slack variable model provide a different representation of curvature in the response surface in comparison to that of the canonical polynomial model. This is discussed by Piepel and Cornell (1994).

Previously, Marquardt and Snee (1974) and Snee and Rayner (1982) had identified the complexity of comparing the interpretation of the parameters in the slack variable and canonical polynomial models as potentially prohibitive. However, the interpretation of the slack variable model itself is straightforward and not inherently more complex than that of the canonical polynomial model. Ease of understanding should not prohibit any practitioner from applying the slack variable model rather than the canonical polynomial model. More importantly, the work of Cornell (2000) and Khuri (2005) indicates that, potentially, these models could prove to be more parsimonious when fitting in a reduced form.

However, the use of the slack variable models does present difficulties. Each distinct choice of  $x_i$  as the slack variable represents a different parameterisation of the Scheffé canonical polynomial, which itself also represents an alternative parameterisation. Thus the canonical polynomial model and each of the  $q$  slack variable models, fitted with a different component taken as slack, will have to be examined in order to find the most effective subset of square terms,  $x_i^2$  and terms of joint effect,  $x_i x_j$ . This could be considered an unwanted complication. However, Piepel, Szychowski, and Loepky (2001) and Piepel and Landmesser (2009) present a solution.

### Partial quadratic mixture models

Piepel recognises that the slack variable model can be used advantageously to provide a more parsimonious model. However, he also raises valid concerns that through denoting one variable slack, a practitioner, particularly one inexperienced in mixture experiments, may draw incorrect conclusions about the effects of mixture components. This could occur where the practitioner does not appreciate the relationship between the estimated coefficients of the canonical polynomial and slack variable models. Cornell (2002) is also suspicious of recommending a method which encourages practitioners to incorrectly consider one component to be a non-entity. Therefore, Piepel proposes an alternative: the Partial Quadratic Mixture (PQM) model. This is found by adding either square terms or terms of binary joint effect, in any components, to the Scheffé linear mixture model

$$E[y] = \sum_{i=1}^q \beta_i x_i. \quad (2.65)$$

Thus, the full PQM model is expressed

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^q \beta_{ii} x_i^2, \quad (2.66)$$

where  $q$  joint effects or square terms are removed through an appropriate constraint in order to avoid over-parameterisation. Both the quadratic canonical polynomial and each of the  $q$  quadratic slack variable models are equivalent to special forms of the PQM model. This is also true of the quadratic additive model,

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} x_i^2, \quad (2.67)$$

inferred by Darroch and Waller (1985) and discussed further by Chan, Guan, and Zhang (1998) and Chan, Meng, Jiang, and Guan (1998). This model is gained by the same substitutions by which the slack variable models were obtained above and provides similar advantages, without treating a variable as slack and hence artificially diminishing its importance. Higher order forms, with no terms of joint effect, can also be found for additive, slack variable and PQM models.

In addition, the PQM models also include other models, which are not equivalent to slack variable, quadratic additive or canonical polynomial models. Thus,



the best reduced PQM model is of at least equal fit to the best reduced forms of each of these three alternatives and for this reason its application would be expected to be most effective in finding the most parsimonious description of the response surface, provided the correct reduced form was identified.

Piepel demonstrated the superiority of the PQM model through a three component example, which looked at the tint strength of house paint and was used by Cornell (2002) and Khuri (2005) to support the viability of the slack variable approach. Starting from the linear model, Piepel implements a stepwise regression procedure for the quadratic terms of the canonical polynomial, slack variable and PQM models. This use of stepwise regression where there are more possible terms than can actually be fit will be referred to later in reference to the fitting procedure for the GBMM.

The resultant reduced PQM model provides a better fit than the reduced canonical polynomial model or reduced slack variable models. It is also not equivalent to an additive quadratic model. With some caveats, it is demonstrated to be the most useful of these type of parameterisations, where linear terms are preserved in the model. However, one further reparameterisation provides an alternative to what are conventionally described as linear terms.

### Kronecker polynomial models

The Kronecker polynomial models (K-models), introduced by Draper and Pukelsheim (1998), are based on the Kronecker algebra of vectors and matrices. The mixture components may be expressed as a vector,

$$\mathbf{x} = (x_1, \dots, x_q). \quad (2.68)$$

The Kronecker square,

$$\mathbf{x} \otimes \mathbf{x} = (x_1^2, x_1x_2, \dots, x_1x_q, x_2x_1, x_2^2, \dots, x_{q-1}x_q, x_q^2), \quad (2.69)$$

is the vector of the order 2 cross product terms  $x_i x_j$ , including the square terms  $x_i^2$ . The cross product terms of the  $i^{th}$  and  $j^{th}$  components appear twice as  $x_i x_j$  and  $x_j x_i$ . Similarly the Kronecker cube,

$$\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} = (x_1^3, x_1^2x_2, \dots, x_{q-1}x_q^2, x_q^3), \quad (2.70)$$

is the vector of the order 3 crossproduct terms  $x_i x_j x_k$ , including the cube terms  $x_i^3$ , where the third order terms are repeated six or three times dependent upon the subscripts  $iiij$  or  $ijk$ .

The first and second order K-models are expressed

$$E[y] = \mathbf{x}' \theta = \sum_{i=1}^q \theta_i x_i \quad (2.71)$$

and

$$E[y] = (\mathbf{x} \otimes \mathbf{x})' \theta = \sum_{i=1}^q \theta_i x_i^2 + \sum_{i=1}^q \sum_{j=1}^q \theta_{ij} x_i x_j. \quad (2.72)$$

Since the terms  $x_i x_j$  and  $x_j x_i$ , of the second order model, are identical, it is assumed that  $\theta_{ij} = \theta_{ji}$ . The third order model is given similarly.

Draper considers it incorrect to state that the second order K-model loses linear terms or the third order K-model loses linear or quadratic terms. The full second order K-model features the same number of terms as the full second order Scheffé polynomial and describes the same surface and this is similarly true of a comparison of the third order models. Moreover, the square terms  $x_i^2$  of the second order K-model represent the individual components in the manner of the  $x_i$  terms of the Scheffé polynomials, while the  $x_i x_j$  terms describe a joint effect of the  $i^{th}$  and  $j^{th}$  components. Similarly, terms within the third order K-model may be associated with each of the individual components and the joint effects of any combination of two or three components.

However, the reduced models are not equivalent. Tests based upon the coefficient estimates of the terms of joint effect in each model do not return the same results. The parameter estimates in each case represent different types of behaviour and for this reason it could be felt that Draper is incorrect to state that linear terms are preserved in K-models. This is because a linear term typically has a particular effect upon the response, not preserved by the  $x_i^2$  terms of the second order K-model or the  $x_i^3$  terms of the third order K-model. Similarly, the representation of two component synergy in the third order K-model does not preserve the quadratic curvature of the  $x_i x_j$  terms. In short, interpretation is complicated by engaging all terms as representative of quadratic or cubic curvature, respectively.

This is not to say the interpretation is particular complex. However, it contrasts with both the common understanding of linear, quadratic and cubic curvature and the interpretation of other models used for the analysis of mixture experiments. While it may not be considered prohibitive to their application, it could be expected to be deleterious to their appeal. This said, if it is acknowledged that the linear effects are not important, and instead it is chosen to explore more flexible single component effects, the Kronecker models represent an important starting point.

## 2.5 Generalised Q-fractions mixture model

Prior to recent years the closest attempt to producing nonlinear models for mixture experiments were those of Becker. In 1978, he suggested a model including nonlinear parameters, but without application nor discussion of their estimation. More recently, the work of Focke, Sandrock, and Kok (2007) and Focke et al. (2012) has led to the development of a nonlinear model for mixture experiments, based upon chemical mixing rules commonly applied in thermodynamics. Their final model is

$$E[y] = \left( \frac{\sum_{i=1}^q a_i x_i \beta_i^r}{\left( \sum_{j=1}^q a_j^s x_j \right)^{\frac{1}{s}}} \right)^{\frac{1}{r}}, \quad (2.73)$$

where  $\beta_i$  represents a known value, associated with a measurable physical property, and  $a_i$ ,  $r$  and  $s$  are estimable parameters.  $\beta_i$  represent the same property as they would in the Scheffé polynomial, but here they are fixed at the observed pure component effect, rather than being estimable, thus not being influenced by observations in the rest of the response surface.

This model can instead be represented as

$$E[y] = \left( \sum \beta_i^r Q_i \right)^{\frac{1}{r}}, \quad (2.74)$$

where

$$Q_i = \frac{a_i x_i}{\left( \sum_{j=1}^q a_j^s x_j \right)^{\frac{1}{s}}}. \quad (2.75)$$

Hence, these are linear blending models upon functions of the mixture components and the Scheffé linear model is a special case where  $r = s = a_i = 1$  (and  $\beta_i$  are

estimable).

As a nonlinear model it is highly adaptable in a manner which is intriguing. It can describe a broad range of behaviour, including asymmetry in the joint effects of the components, which is not possible for the Scheffé polynomials. However, the structure of the model means it can have a large number of nonlinear parameters as the number of components increases. This would particularly be the case where  $\beta_i$  cannot be fixed at some theoretically logical value. For this reason, estimation has not yet been made in more than 3 components (to our knowledge). This opens a space for similarly flexible, but more broadly applicable models.

## 2.6 Summary

Scheffé's models represented a significant progression in the modelling of mixture experiments. However, this has not prevented others attempting to develop upon his models and present wholesale alternatives. The stated objective for several of these developments were to explore blending behaviour for which Scheffé's models did not allow (in the case of Becker's models, inverse and logarithmic terms and the Generalised Q-fraction model). The view of Plackett, in his initial assessment of the Scheffé polynomials, that they represented a polynomial smoothing method, seems consistently problematic; the assumptions in Scheffé's models regarding the blending of the components, appear too rigid.

In short, it would appear desirable to possess a model which could adapt to describe a wider range of blending effects. The Generalised Q-fraction model is one such attempt. However, this model includes nonlinear parameters which could not be easily handled in cases of more than a small number of components. Moreover, it is not well seated within the established methodology for mixture experiments and cannot be shown to represent a theoretical advance upon the firmly established methods, i.e. the Scheffé polynomials, Becker models and PQM models.

In developing the new class of models proposed in this thesis, it was considered desirable to develop a model which included as special cases the models of Scheffé and Becker, as well as a continuous range of other possibilities. Thus, it would be capable of describing a broad range of blending behaviours, including that given by the models of Scheffé and Becker, as well as similar behaviour to that desired by Chen and Draper when proposing the logarithmic and inverse terms.

In the manner of the Generalised Q-fraction models, such a class of models could be expected to possess a large number of nonlinear parameters, particularly in cases where there are a large number of components. As such, the new class of models is represented here alongside a detailed discussion of their fitting, in order that the reader understands how to handle the potentially large number of parameters.

## Chapter 3

# General blending models for mixture experiments

The linear in parameters models discussed thus far will describe well situations where their terms accommodate the specific joint effects of the mixture components. However, the situations where this will be the case are limited and there are joint effects outside the scope of what they may describe. It is felt this could lead them to perform poorly when these effects are required and they may not adequately represent the response or do so in a manner detrimental to model parsimony. This section proposes a general class of models, utilising flexible regressors, which allow for the description of responses of mixtures whose components have a wide range of different effects. The discussion will first look at joint effects of two components, and then the presented ideas will be extended to three components. The joint effects of more than three components are rarely considered when modelling mixture experiments using existing methodology and therefore are not considered here, although such terms could be developed similarly to the two and three component cases shown.

### 3.1 Regressor defining parameters

In order to establish the link between the existing methodology we start by describing an idea for combining and generalising standard quadratic models. It is from this that the entire class of models has developed.

The models

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ij} x_i x_j \quad (3.1)$$

and

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ij} \frac{x_i x_j}{x_i + x_j}, \quad (3.2)$$

where  $1 \leq i, j \leq q$ ,  $i \neq j$ , characterise the response surface in contrasting ways with respect to the joint effect of  $x_i$  and  $x_j$ . While Model 3.2 allows an additive blending effect through Becker's H2 model, Model 3.1 utilises the quadratic blending effect of the Scheffé polynomial. This contrast is reflected in the form of the regressors for  $x_i$  and  $x_j$ .

Where more than one pair of mixture components demonstrate joint effects, the best model fit may be achieved where the joint effect of the term in Model 3.1 is used for one pair of components and that of the regressor in Model 3.2 for another (Johnson and Zabik, 1981). However, the choice of regressors in this context may be far more extensive and can be defined by the introduction of *regressor defining parameters*.

Firstly, a generalised binary blending effect is defined by introducing the parameter  $s_{ij}$  in the model

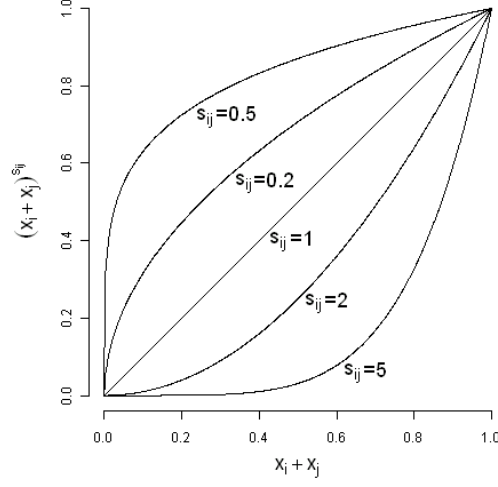
$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right) \left( \frac{x_j}{x_i + x_j} \right) (x_i + x_j)^{s_{ij}}. \quad (3.3)$$

The blending effects corresponding to five different values of  $s_{ij}$  ( $s_{ij} = 0.2, 0.5, 1, 2, 5$ ) are shown in Figure 3.1, including those for Models 3.1 and 3.2, for  $s_{ij} = 2$  and  $s_{ij} = 1$ , respectively.

Further flexibility can be added with the introduction of the regressor defining parameters  $r_{ij}$  and  $r_{ji}$  to the model, which gives

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{r_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{r_{ji}} (x_i + x_j)^{s_{ij}}, \quad (3.4)$$

where, if  $s_{ij} = 1$ , this term of joint effect is the same as in Model 2.57 proposed (but not applied) by Becker in 1978. Thus we define a flexible regressor for the

Figure 3.1: Blending effects for  $s_{ij} = 0.2, 0.5, 1, 2$  and  $5$ .

joint effect of  $x_i$  and  $x_j$ , governed by the regressor defining parameters  $s_{ij}$ ,  $r_{ij}$ ,  $r_{ji}$  and the corresponding model parameters  $\beta_{ij}$ .

This idea of regressor defining parameters can be extended to introduce a flexible regressor describing general ternary joint effects in the model

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ijk} \left( \frac{x_i}{x_i + x_j + x_k} \right)^{r_{ijk}} \left( \frac{x_j}{x_i + x_j + x_k} \right)^{r_{jki}} \times \left( \frac{x_k}{x_i + x_j + x_k} \right)^{r_{kij}} (x_i + x_j + x_k)^{s_{ijk}}. \quad (3.5)$$

Here, the joint effect of the components  $x_i$ ,  $x_j$  and  $x_k$  is governed by  $s_{ijk}$ ,  $r_{ijk}$ ,  $r_{jki}$ ,  $r_{kij}$  and the corresponding  $\beta_{ijk}$ . In particular,  $s_{ijk}$  governs the blending effect between  $x_i + x_j + x_k$  and the remainder of the mixture, in an analogous manner to  $s_{ij}$ , for  $x_i + x_j$  above. Thus, contrasting effects may be seen along any ray where  $x_i$ ,  $x_j$  and  $x_k$  remain in fixed relative proportions. The flexible regressors, for the binary and ternary cases, shall be referred to as the general term of binary and ternary joint effect, respectively.

Model 3.4 may alternatively be expressed

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{g_{ij} h_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{g_{ji} (1-h_{ij})} (x_i + x_j)^{s_{ij}}, \quad (3.6)$$

where  $g_{ij} h_{ij} = r_{ij}$  and  $g_{ij} (1 - h_{ij}) = r_{ji}$ , so that  $g_{ij} = r_{ij} + r_{ji}$  and  $h_{ij} = r_{ij}/g_{ij}$ .



Similarly, Model 3.5 may alternatively be expressed

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{ijk} x_i x_j x_k \left( \frac{x_i}{x_i + x_j + x_k} \right)^{g_{ijk} h_{ijk}} \left( \frac{x_j}{x_i + x_j + x_k} \right)^{g_{ijk} h_{jki}} \times \left( \frac{x_k}{x_i + x_j + x_k} \right)^{g_{ijk} (1 - h_{ijk} - h_{jki})} (x_i + x_j + x_k)^{s_{ijk}}, \quad (3.7)$$

where  $g_{ijk} = r_{ijk} + r_{jki} + r_{kij}$ ,  $h_{ijk} = r_{ijk}/g_{ijk}$ ,  $h_{jki} = r_{jki}/g_{ijk}$ ,  $g_{ijk} h_{ijk} = r_{ijk}$ ,  $g_{ijk} h_{jki} = r_{jki}$  and  $g_{ijk} (1 - h_{ijk} - h_{jki}) = r_{kij}$ . These reparameterisations are important in understanding how the regressor defining parameters govern the joint effects these terms describe.

## 3.2 The Dirichlet distribution

To assist this, it is useful to look at the Dirichlet distribution (Huang (2005)). The Dirichlet distribution is a probability distribution in  $K \geq 2$  variables,  $x_i$ , with probability density function

$$f(\mathbf{x}, \alpha_i) = \frac{1}{B(\alpha_1, \dots, \alpha_k)} \prod_{i=1}^K x_i^{\alpha_i - 1}, \quad (3.8)$$

where  $\sum x_i = 1$  and  $0 \leq x_i \leq 1$ ,  $\alpha_1, \dots, \alpha_k$  are parameters to be estimated and  $B(\alpha_1, \dots, \alpha_k)$  is the Beta function,

$$B(\alpha_1, \dots, \alpha_k) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}. \quad (3.9)$$

The constraints upon  $x_i$  mean the probability density function of the Dirichlet distribution is supported on the  $K - 1$  dimensional simplex - this support is analogous to the unconstrained experimental region of a mixture experiment.

It can be shown the binary term of joint effect can be expressed as a product of a constant,  $\beta'_{ij}$ , and two functions, each with the same form as the probability

density function of the Dirichlet distribution for  $K = 2$ :

$$\begin{aligned} & \beta_{ij} \times \left( \frac{x_i}{x_i + x_j} \right)^{g_{ij}h_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{g_{ji}(1-h_{ij})} \times (x_i + x_j)^{s_{ij}} (1 - x_i - x_j)^0 = \\ & \beta'_{ij} \times \frac{1}{B(\gamma_1^{ij}, \gamma_2^{ij})} \left( \frac{x_i}{x_i + x_j} \right)^{(\gamma_1^{ij}-1)} \left( \frac{x_j}{x_i + x_j} \right)^{(\gamma_2^{ij}-1)} \\ & \times \frac{1}{B(\delta_1^{ij}, \delta_2^{ij})} (x_i + x_j)^{\delta_1^{ij}-1} (1 - x_i - x_j)^{\delta_2^{ij}-1} \end{aligned} \quad (3.10)$$

where  $g_{ij} = \sum_i \gamma_i - 1$ ,  $h_{ij} = \frac{\gamma_i - 1}{g_{ij}} = \frac{\gamma_i - 1}{\sum_i \gamma_i - 1}$ ,  $s_{ij} = \delta_1 - 1$  and  $\beta_{ij} = \frac{\beta'_{ij}}{B(\delta_1^{ij}, \delta_2^{ij})B(\gamma_1^{ij}, \gamma_2^{ij})}$ . The ternary term of joint effect can similarly be written as a product of a constant, one function of the form of the probability density function of the Dirichlet distribution for  $K = 2$  and one function of the form of the probability density function of the Dirichlet distribution for  $K = 3$ .

It is possible to interpret a Dirichlet distribution through the mean,

$$M = \left( \frac{\alpha_1}{S}, \dots, \frac{\alpha_K}{S} \right), \quad (3.11)$$

and the concentration,

$$S = \sum_i \alpha_i, \quad (3.12)$$

where  $M$  describes the point within the simplex where the distribution is at its maximum and  $S$  describes the concentration about that point. By looking at the two functions embedded within the general binary term of joint effect, it is also possible to interpret this through mean and concentration effects: the mean and concentration effects with respect to  $x_i$  and  $x_j$ , and the mean and concentration effects with respect to  $x_i + x_j$ .

### 3.3 Interpreting the terms of joint effect

From Equation 3.10 it is evident that the binary term of joint effect can be separated into two functions and these functions can be interpreted in terms of mean and concentration effects, in the manner of the Dirichlet distribution. One of these functions is governed by the regressor defining parameters  $h_{ij}$  and  $g_{ij}$ , and the other by the regressor defining parameter  $s_{ij}$ .

In the first of these functions,

$$\left(\frac{x_i}{x_i + x_j}\right)^{g_{ij}h_{ij}} \left(\frac{x_j}{x_i + x_j}\right)^{g_{ji}(1-h_{ij})} = \left(\frac{x_i}{x_i + x_j}\right)^{(\gamma_1^{ij}-1)} \left(\frac{x_j}{x_i + x_j}\right)^{(\gamma_2^{ij}-1)}, \quad (3.13)$$

$g_{ij}$  defines the concentration effect of the term with respect to  $x_i$  and  $x_j$ , since  $S^{ij} = g_{ij} + 2$ , where  $S_{ij}$  is the concentration parameter as defined for the Dirichlet distribution above. Therefore, since

$$\begin{aligned} h_{ij} &= \frac{\gamma_1^{ij} - 1}{S^{ij} - 2} \\ \frac{h_{ij}}{S^{ij}} &= \frac{\gamma_1^{ij} - 1}{S^{ij}(S^{ij} - 2)} \\ (S^{ij} - 2) \frac{h_{ij}}{S^{ij}} &= \frac{\gamma_1^{ij} - 1}{S^{ij}} \\ (S^{ij} - 2) \frac{h_{ij}}{S^{ij}} + \frac{1}{S^{ij}} &= \frac{\gamma_1^{ij}}{S^{ij}}, \end{aligned}$$

the parameter  $h_{ij}$  can be said to govern the mean effect of the term with respect to  $x_i$  and  $x_j$ ,

$$M^{ij} = \left(\frac{\gamma_1^{ij}}{S^{ij}}, \frac{\gamma_2^{ij}}{S^{ij}}\right) = \left(\frac{(S^{ij} - 2)h_{ij}}{S^{ij}} + \frac{1}{S^{ij}}, \frac{(S^{ij} - 2)(1 - h_{ij})}{S^{ij}} + \frac{1}{S^{ij}}\right). \quad (3.14)$$

In particular, it can be shown that  $h_{ij} = 0.5$  indicates a symmetric effect.

To illustrate the way the joint effect of  $x_i$  and  $x_j$  changes with these regressor defining parameters, the effects for  $g_{ij} = 20$  and  $g_{ij} = 0.2$  (that is concentration effects  $S^{ij} = 22$  and  $S^{ij} = 2.2$ , respectively) are shown in Figure 3.2 for  $h_{ij} = 0.75$ , that is mean effect  $M^{ij} = (0.727, 0.273)$ .

In the second function contained within the binary term of joint effect,

$$(x_i + x_j)^{s_{ij}} (1 - x_i - x_j)^0 = (x_i + x_j)^{\delta_1^{ij}-1} (1 - x_i - x_j)^{\delta_2^{ij}-1}, \quad (3.15)$$

where  $\delta_2^{ij} = 1$  is seen to be fixed and  $\delta_1^{ij} = s_{ij} - 1$ . The influence of  $s_{ij}$  has already been seen in Figure 3.1, and is understood to govern the blending between  $x_i + x_j$  and the remainder of the mixture. The concentration effect of the binary term of joint effect with respect to  $x_i + x_j$ , is  $S_s^{ij} = s_{ij}$  and the mean effect,  $M_s^{ij} = \left(\frac{s_{ij}-1}{s_{ij}}, \frac{1}{s_{ij}}\right)$ . Hence, it is seen  $s_{ij}$  governs both mean and concentration effects with respect to  $x_i + x_j$ .

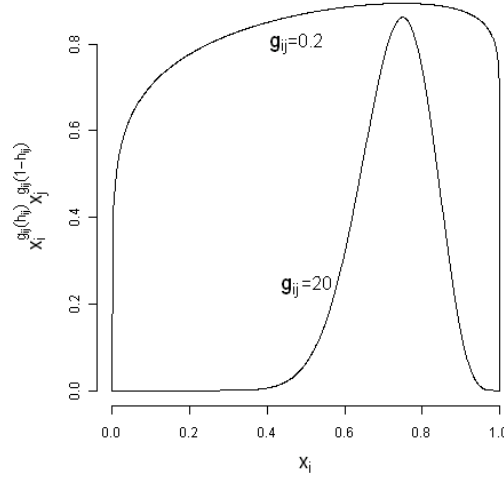


Figure 3.2: Effect of binary term of joint effect for  $g_{ij} = 20$  and  $g_{ij} = 0.2$  and  $h_{ij} = 0.75$

At this point, it seems useful to reiterate the terminology for the terms of binary joint effect:

- $M^{ij}$  - the mean effect of the binary term of joint effect, with respect to  $x_i$  and  $x_j$ ;
- $S^{ij}$  - the concentration effect of the binary term of joint effect, with respect to  $x_i$  and  $x_j$ ;
- $M_s^{ij}$  - the mean effect of the binary term of joint effect, with respect to  $x_i + x_j$ ;
- $S_s^{ij}$  - the concentration effect of the binary term of joint effect, with respect to  $x_i + x_j$ ;
- $h_{ij}$  - the regressor defining parameter governing  $M^{ij}$ ;
- $g_{ij}$  - the regressor defining parameter governing  $S^{ij}$ ;
- $s_{ij}$  - the regressor defining parameter governing  $M_s^{ij}$  and  $S_s^{ij}$ ;
- $r_{ij}$ ,  $r_{ji}$ ,  $s_{ij}$  - regressor defining parameters combining mean and concentration effects, used for estimation.

The interpretation of the ternary term of joint effect can similarly be separated into mean and concentration effects with respect to  $x_i$ ,  $x_j$  and  $x_k$ , governed by  $h_{ijk}$ ,  $h_{jki}$  and  $g_{ijk}$ , and mean and concentration effects with respect to  $x_i + x_j + x_k$ , governed by  $s_{ijk}$ . As mentioned, the influence of  $s_{ijk}$  is similar to that of  $s_{ij}$ . The effect of different  $h_{ijk}$ ,  $h_{jki}$  and  $g_{ijk}$ , are illustrated in Figure 3.3.

In the Dirichlet distribution there are restrictions placed upon  $\alpha$ . For the binary terms of joint effect these are most usefully reexpressed in terms of  $r_{ij}$ ,  $r_{ji}$  and  $s_{ij}$ , as these are the parameters estimated when fitting the GBMM:

$$r_{ji}, r_{ij}, s_{ij} \geq -1. \quad (3.16)$$

For now, estimation has only occurred for

$$r_{ji}, r_{ij}, s_{ij} > 0, \quad (3.17)$$

in order to preserve the interpretation of the linear parameters as in established models, but this is something which could be looked at again. The parameters  $r_{ijk}$ ,  $r_{jki}$ ,  $r_{kij}$ , and  $s_{ijk}$  would similarly be constrained for the general ternary term of joint effect,

$$r_{ijk}, r_{jki}, r_{kij}, s_{ijk} \geq -1, \quad (3.18)$$

but have been restricted to values greater than 0, for the purpose of estimation.

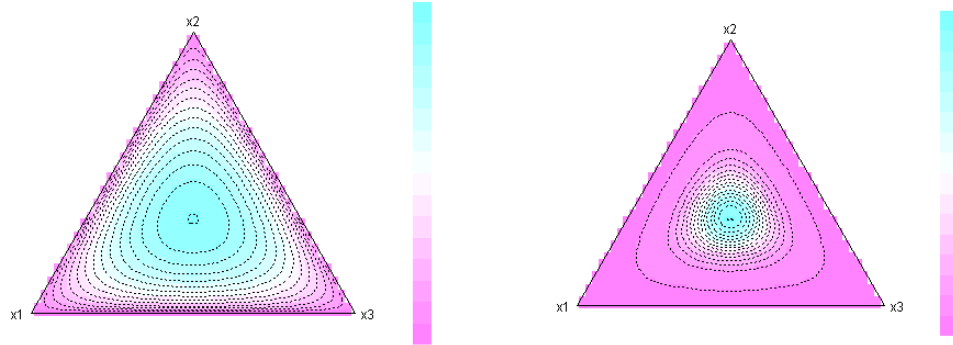
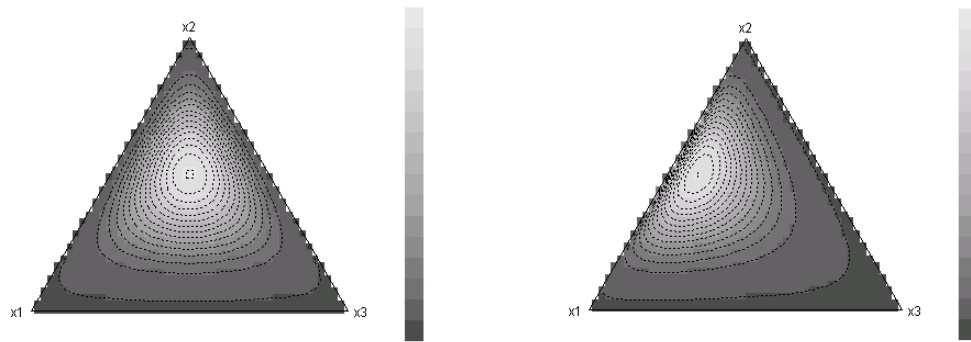
(a)  $g_{ijk} = 20, h_{ijk} = h_{jki} = \frac{1}{3}$ (b)  $g_{ijk} = 2, h_{ijk} = h_{jki} = \frac{1}{3}$ (c)  $g_{ijk} = 7, h_{ijk} = \frac{1}{4}, h_{jki} = \frac{1}{4}$ (d)  $g_{ijk} = 7, h_{ijk} = \frac{1}{10}, h_{jki} = \frac{4}{10}$ 

Figure 3.3: Plots of the response described by the general ternary term of joint effect for different values of  $g_{ijk}$ ,  $h_{ijk}$  and  $g_{jki}$

### 3.4 General blending models for mixture experiments

Using the terms of binary and ternary joint effect, the new class of General Blending Models for Mixtures (GBMM) is proposed of the form

$$\begin{aligned}
 E[y] = & \sum_{i=1}^q \beta_i x_i \\
 & + \sum_{i \neq j} \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{r_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{r_{ji}} (x_i + x_j)^{s_{ij}} \\
 & + \sum_{i \neq j \neq k} \beta_{ijk} \left( \frac{x_i}{x_i + x_j + x_k} \right)^{r_{ijk}} \left( \frac{x_j}{x_i + x_j + x_k} \right)^{r_{jki}} \times \\
 & \quad \left( \frac{x_k}{x_i + x_j + x_k} \right)^{r_{kij}} (x_i + x_j + x_k)^{s_{ijk}}. \tag{3.19}
 \end{aligned}$$

It is possible to include more than one term of each type in the model, that is for each pair or triple of components, but for simplicity this is not considered. Model 3.19 can have many terms, but usually only a small number of such terms will be necessary to represent the studied relationship.

As discussed earlier, the GBMM can also be reparameterised as

$$\begin{aligned}
 E[y] = & \sum_{i=1}^q \beta_i x_i \\
 & + \sum_{i \neq j} \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{g_{ij} h_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{g_{ji} (1-h_{ij})} (x_i + x_j)^{s_{ij}} \\
 & + \sum_{i \neq j \neq k} \beta_{ijk} \left( \frac{x_i}{x_i + x_j + x_k} \right)^{g_{ijk} h_{ijk}} \left( \frac{x_j}{x_i + x_j + x_k} \right)^{g_{jki} h_{jki}} \times \\
 & \quad \left( \frac{x_k}{x_i + x_j + x_k} \right)^{g_{kij} (1-h_{ijk}-h_{jki})} (x_i + x_j + x_k)^{s_{ijk}}. \tag{3.20}
 \end{aligned}$$

Additionally, they may be expressed

$$\begin{aligned}
 E[y] = & \sum_{i=1}^q \beta_i x_i + \sum_{i \neq j} \beta_{ij} \frac{x_i^{r_{ij}} x_j^{r_{ji}}}{(x_i + x_j)^{t_{ij}}} \\
 & + \sum_{i \neq j \neq k} \beta_{ijk} \frac{x_i^{r_{ijk}} x_j^{r_{jki}} x_k^{r_{kij}}}{(x_i + x_j + x_k)^{t_{ijk}}}, \tag{3.21}
 \end{aligned}$$

where  $t_{ijk} = r_{ijk} + r_{jki} + r_{kij}$ . This form is useful for estimation and shall be used in the following chapter.

Models 3.19, 3.20 and 3.21 may be used to establish a broad range of joint effects. In fact, most models presented in the literature are special cases of our class of models. For example, the quadratic crossproduct terms in the Scheffé polynomial or the PQM model occur when  $h_{ij} = 0.5$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$ . The squared terms in the PQM model occur when  $h_{ij} = 0$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$ . The quadratic terms of Becker's H2 and H3 models occur when  $h_{ij} = 0.5$ ,  $s_{ij} = 1$  and  $g_{ij} = 2$  or 1, respectively. Furthermore, the ternary term of the special cubic model occurs when  $h_{ijk} = h_{jki} = \frac{1}{3}$ ,  $g_{ijk} = 3$  and  $s_{ijk} = 3$  in the general ternary term of joint effect. Thus, the GBMM allows us to consider commonly used terms, as well as new terms, in a context of notable flexibility.

It is possible that a practitioner may choose not to include the terms of ternary joint effect in a GBMM. Therefore, cases where they aren't included shall be referred to as the General Quadratic Blending Mixture Model (GQBMM), acknowledging the relationship to the Scheffé polynomials. Cases where they are included shall be referred to as the General Special Cubic Blending Mixture Model (GSCBMM).

It should be noted that while the possible interpretation can be made of the particular binary and ternary terms of joint effect, it is not foreseen that interpretation will necessarily be useful. It is expected that more than one such term will be included in a model and this will blur the interpretation of the individual terms. However, this is not to say that interpretation will be impossible and evidently there is a clear framework for doing so, which could be useful as methodologies for the GBMM are developed. In fact, the possibility of interpretation will be evident in some of the examples of the following chapters.



## Chapter 4

# Model estimation and model selection: a fitting procedure for the GBMM

In fitting the GBMM, it is not expected to fit a full model, i.e. one with at least one term of joint effect between every pair or triplet of components. It is necessary to select only a subset of such terms, with the objective of achieving model parsimony. Parsimony, in the context of statistical models, provides the best fitting model with the fewest number of parameters. Hence, the simplest model is chosen judged to possess a sufficiently good description of the observed responses, that is, sufficient goodness of fit.

The difficulty of fitting a reduced GBMM is that, in order to ascertain such a subset, it is also necessary to estimate the values of the regressor defining parameters for each term of joint effect within the subset. Therefore, the challenge of fitting the GBMM, is to achieve both model selection and nonlinear model estimation simultaneously. In order to describe the process chosen, certain concepts regarding both model estimation and model selection are discussed at the beginning of this chapter. Subsequently, the details of the fitting procedure for the GBMM are presented.

### 4.1 Model estimation

In order to fully inform the fitting procedure proposed, it is necessary to look at two methods of model estimation: maximum likelihood and least squares.

Maximum likelihood is addressed briefly, in the context of linear models, while least squares is additionally explored in the context of nonlinear models.

### 4.1.1 Maximum likelihood estimation

A linear in parameters model may be expressed

$$E[\mathbf{y}] = \mathbf{F}\beta, \quad (4.1)$$

where  $\beta$  is the vector of linear parameters,  $\mathbf{F}$  is the extended design matrix whose rows  $\mathbf{F}_i(\mathbf{x})$  give the values of each of the terms of the model at the values of  $\mathbf{x}$ , for which  $\mathbf{y}$ , the vector of responses, were observed. Thus for the two component quadratic Scheffé polynomial,

$$E[y] = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2, \quad (4.2)$$

if observations are made at the pure mixtures,  $\mathbf{x} = (1, 0)$  and  $\mathbf{x} = (0, 1)$ , and the 50-50 mixture  $\mathbf{x} = (\frac{1}{2}, \frac{1}{2})$ ,

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

Under the assumptions of the linear model, the vector of observations are normally distributed such that

$$\mathbf{y} \sim N(\mathbf{F}\beta, \sigma_\epsilon^2 \mathbf{I}), \quad (4.3)$$

where  $\mathbf{I}$  is the identity matrix.

The joint probability density of these observations is

$$p(\mathbf{y}) = \frac{1}{(\sigma_\epsilon \sqrt{2\pi})^n} \exp \frac{(\mathbf{y} - \mathbf{F}\beta)' (\mathbf{y} - \mathbf{F}\beta)}{2\sigma_\epsilon^2}, \quad (4.4)$$

of which the log likelihood is

$$\log L(\beta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{(2\sigma^2)} (\mathbf{y} - \mathbf{F}\beta)' (\mathbf{y} - \mathbf{F}\beta), \quad (4.5)$$

where  $n$  is the number of observations taken in the experiment. The maximum likelihood estimators of the model parameters are found by maximising the log likelihood.

In the context of the fitting procedure presented for the GBMM, maximum likelihood is important with respect to a measure of model parsimony. Of the several statistics which can measure model parsimony, here the corrected AIC

$$AIC_c = -2L + \frac{2n(k+1)}{n-k-1}, \quad (4.6)$$

is used, which is a function of  $L$ , the maximised value for the likelihood function of the estimated model. This can be viewed as a measure of goodness of fit with an additional penalisation for the number of parameters, where  $k$  is the number of parameters.

### 4.1.2 Least squares regression

More significant to the fitting of GBMM is least squares regression. In order to estimate  $\beta$ , both in the case of least squares and maximum likelihood estimation, the functional

$$\|\mathbf{y} - \mathbf{F}\beta\|_2^2, \quad (4.7)$$

can be minimised, where  $\|\cdot\|_2^2$  is the Euclidean distance. Therefore, a measure of the goodness of fit

$$R^2 = 1 - \frac{\|\mathbf{y} - \mathbf{F}\beta\|_2^2}{\|\mathbf{y} - \bar{\mathbf{y}}\|_2^2}. \quad (4.8)$$

The estimates of  $\beta$ , defined by maximising Equation 4.7, are expressed

$$\hat{\beta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y}, \quad (4.9)$$

the variance of which are given as

$$\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1}. \quad (4.10)$$

The confidence ellipsoid of the  $p$  parameters can be defined in  $p$ -dimensional space as

$$S(\beta) - S(\hat{\beta}) = ps^2 F_{p,v,\alpha}, \quad (4.11)$$

where  $s^2$  is an estimate of  $\sigma^2$ ,  $p$  is the number of parameters and  $F_{p,v,\alpha}$  is the critical value for the desired false-rejection probability,  $\alpha$ , of an F-distribution with  $p$  and  $v = n - p$  degrees of freedom. A confidence ellipsoid is used to suggest

the accuracy of the estimates  $\hat{\beta}$  in relation to their true value. The volume of this ellipsoid is inversely proportionate to the square root of the determinant of the information matrix,

$$|\mathbf{F}^T \mathbf{F}|. \quad (4.12)$$

In experimental design it is often viewed as desirable to reduce the size of this ellipsoid in some respect. As such many criteria for assessing experimental design relate to  $|\mathbf{F}^T \mathbf{F}|$ . In particular, experimental designs which maximise the determinant of the information matrix are known as D-optimum. This will be of interest when experimental design for the GBMM is discussed later.

Given estimates  $\hat{\beta}$ , the predicted value of the response for any  $\mathbf{x}$  is now given as

$$\hat{y} = \hat{\beta}^T f^T(\mathbf{x}), \quad (4.13)$$

such that

$$\text{var}\{\hat{y}(\mathbf{x})\} = \sigma^2 f^T(\mathbf{x}) (\mathbf{F}^T \mathbf{F})^{-1} f(\mathbf{x}). \quad (4.14)$$

The prediction variance is then given as

$$N \frac{\text{var}\{\hat{y}(\mathbf{x})\}}{\sigma^2} = N f^T(\mathbf{x}) (\mathbf{F}^T \mathbf{F})^{-1} f(\mathbf{x}). \quad (4.15)$$

This is of interest with respect to several other criteria relating to experimental design. In particular a G-optimum design minimises the maximum generalised prediction variance across the design space. Again this topic shall be returned to later.

## 4.2 Partially linear models

As discussed, the GBMM is a particular type of nonlinear model, that is a partially linear model. In a partially linear model, the  $i^{th}$  row of the extended design matrix  $\mathbf{F}$  is expressed

$$\mathbf{F}_i^T(\mathbf{x}, \alpha), \quad (4.16)$$

where  $\alpha$  is a vector of unknown nonlinear parameters to be estimated. A feature of such models and all nonlinear models, initiating estimation requires prior values for the nonlinear parameters  $\alpha$ . Therefore, some prior information is required regarding  $\alpha$  to initiate estimation. Additionally, the choice of experimental design for partially linear models is also dependent on  $\alpha$ . The use of prior information

for  $\alpha$  both in estimation of the GBMM and choice of experimental design are important issues to be addressed.

Estimation of the model parameters of partially linear models can be achieved through nonlinear optimisation of the functional

$$\|\mathbf{y} - \mathbf{F}(\alpha) \beta\|_2^2, \quad (4.17)$$

where  $\mathbf{F}(\alpha)$  is the matrix of the nonlinear functions  $\mathbf{F}_i(\mathbf{x}, \alpha)$ , now dependent on  $\alpha$ . However, because of the partially linear structure, the variable projection method of Golub and Pereyra (1973, 2003) permits the parameters be estimated instead through optimisation of the functional

$$\|(\mathbf{I} - \mathbf{F}(\alpha)\mathbf{F}^+(\alpha)) \mathbf{y}\|_2^2, \quad (4.18)$$

where  $\mathbf{F}^+$  is the Moore-Penrose generalized inverse of  $\mathbf{F}$ . This functional is an expression of the nonlinear parameters only, although this is not to say that it does not estimate the linear parameters, as this still occurs implicitly. In this manner, estimation is achieved through a lower-dimensional although more complex optimisation problem, which nevertheless achieves the estimation of both the linear and nonlinear parameters simultaneously. This can prove advantageous for reasons relevant to the proposed estimation procedure for the GBMM.

These reasons are twofold. Firstly, optimisation may be quicker than it would be with more widely used methods of nonlinear optimisation. This is provided the reduction in dimensionality sufficiently compensates the increased complexity. Secondly, it requires only starting values of the nonlinear parameters (and not the linear parameters) to begin optimisation. In other methods of nonlinear optimisation, prior information regarding the nonlinear parameters would be used to gain the corresponding initial values for the linear parameters. A search for optimal values can then be initiated from these starting values, in the case of our examples using a Gauss-Newton algorithm. By removing the step of identifying the initial values of the linear parameters, time can be saved in the overall estimation time.

The variable projection method allows that the starting values for the linear parameters need not be identified. This will be seen to be particularly advantageous in the proposed fitting procedure of the GBMM, as a large proportion of the parameters can be excluded from the optimisation problem. Hence, the

potentially large numbers of parameters in the GBMM is overcome, reducing the dimensionality of the problem. This said, even excluding the linear parameters, the number of parameters in the GBMM increases rapidly as the number of regressors in the model increases. Therefore a creative application of the existing statistical software is still required.

### 4.3 Stepwise regression and model selection

This creative application is achieved using ideas from stepwise regression (Efroymson (1965)). Stepwise regression is a commonly applied method of achieving model parsimony and can be broken down into forward, backward and stepwise selection, of which the former is relevant to the fitting procedure proposed for the GBMM. All three present a method for the selection of a subset of terms, from a linear model, in order to achieve a parsimonious reduced model. Forward and stepwise selection are of particular interest where it is infeasible to fit the full model, which is most likely to be the case when fitting the GBMM to data from a mixture experiment in a large number of components.

It is also important where all the terms of a model may not be fit simultaneously because of confounding effects. This is the case with the PQM models,

$$E[y] = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \sum_{j=1}^q \beta_{ij} x_i x_j + \sum_{i=1}^q \beta_{ii} x_i^2, \quad (4.19)$$

where it is only possible to fit a model with  $\frac{q(q-1)}{2}$  of the full  $\frac{q(q-1)}{2} + q$  terms. This is because, for example,

$$x_i^2 = x_i \left( 1 - \sum_{j \neq i} x_j \right). \quad (4.20)$$

As stepwise regression was recommended by Piepel as a method for selecting a PQM, it is used here to illustrate both the concepts of forward and stepwise selection.

Forward selection involves beginning with a reduced model and subsequently adding terms in order to improve model parsimony. Piepel suggests beginning

with the linear Scheffé polynomial,

$$E[y] = \sum_{i=1}^q \beta_i x_i. \quad (4.21)$$

Each of the models formed from adding one of the  $\frac{q(q-1)}{2} + q$  additional square or joint effect terms to the model are then found and the  $AIC_c$  calculated. That model with the lowest  $AIC_c$  is then chosen as the second model. The process then continues by adding one of the remaining other  $\frac{q(q-1)}{2} + q - 1$  terms to this second model. Hence forward selection is an iterative process where one term is added at each iteration. When no additional terms can be added without damaging model parsimony, according to  $AIC_c$ , say, the final model has been reached.

Stepwise selection differs from forward selection in that it allows for terms to be removed from the model. At any stage in forward selection, model parsimony may be improved by removing a term added earlier. This will be because terms added later in the fitting procedure have rendered this term ‘less important’ in contributing to model fit. Hence, any iteration of stepwise selection involves consideration of all models formed by removing a term from the model, in addition to all models formed by adding a term to the model.  $AIC_c$  is then used similarly to choose the most parsimonious model from which to begin the next iteration.

Methods of stepwise regression are generally criticised for causing stochastic errors. This means that, at any stage, the process may take a step away from selecting the most parsimonious final model. This mistake can then be compounded at later stages, resulting in a solution very distant from the most parsimonious possibility. This concept shall be discussed with respect to the GBMM. More generally, in order to avoid problems of stochastic errors, best subset selection or elastic net methods are applied in selection of parsimonious reduced models. However, for the GBMM these options were not possible.

## 4.4 Fitting the GBMM: an introduction

Fitting a parsimonious GBMM is made relatively complex because model uncertainty exists not only in the terms to be included in the model, but also the form of these terms. This is made particularly complicated by the large number of regressor defining parameters which will have to be chosen in any GBMM of more than a handful of terms. To address this forward selection can be used,

adding the terms of joint effect sequentially. This allows the consideration of only those regressor defining parameters of one general term of joint effect at a time, provided the regressor defining parameters of general terms of joint effect added at earlier steps remain fixed.

However, by fixing the regressor defining parameters of terms chosen at earlier steps of the forward selection procedure, the GBMM may be viewed as particularly susceptible to stochastic errors. The estimation procedure proposed introduces an additional step to combat this problem. Upon adding a new term of joint effect, through forward selection, it is proposed to re-estimate the terms of joint effect already included in the model. Hence, the regressor defining parameters of a term do not remain fixed once the term has entered the model. Not only does this improve upon the approach which could have been undertaken, it continues to take advantage of the flexibility of the GBMM.

At each step of the forward selection procedure, a large number of models will be estimated in order to choose both the best term of joint effect to be added to the model and the values of its regressor defining parameters. Each of these models presents a nonlinear optimisation problem. This requires the starting values be given for the model parameters, in order to initiate the search for their optimum values. However, the strength of the GBMM is its approach to model uncertainty, which is to say the form of the model terms are unknown prior to fitting and hence the best starting values are unknown. To tackle this problem the variable projection method of Golub and Pereyra is applied, greatly reducing the required number of starting values to only those for the regressor defining parameters of the term of joint effect being estimated at that point. Then, rather than one set of starting values, a list of possible starting values has then been used and the same model estimated multiple times. It is hoped this approach maximises the advantages taken from the flexibility of the GBMM.

The rest of this chapter begins with a discussion of how the process of estimating multiple models, using multiple starting values, is implemented. Hence, Section 4.5 can be viewed as an in depth discussion of how each term is added to the model. Section 4.6 elaborates on this, introducing the idea of reestimating terms of joint effect added to the model at earlier steps of the forward selection process. Finally, in Section 4.7, the additional separate reestimation of the regressor defining parameters  $t_{ij}$  and  $t_{ijk}$  is discussed.

Section 4.8 then gives the full iterative process of model estimation graphically,



for the simplest case, that of the GQBMM. This permits discussion of each of the steps of this process and how they could be applied differently. We will also identify how the process changes when using the GSCBMM. Finally, the following chapter will look at several simple examples. The purpose of these is twofold. Firstly, to inform a discussion of how the changes to the model fitting procedure, introduced in Section 4.8, can change the consequent model and secondly, to begin to illustrate how the GBMM can be applied advantageously in comparison to the models for mixture experiments presented in Chapter 2.

## 4.5 Adding a term

The final form of any GBMM can mean the regressor defining parameters of a very large number of terms need be identified in selecting the final model. Estimation of all of these parameters simultaneously would present a very high-dimensional nonlinear optimisation problem. Particularly challenging is the choice of which general terms of joint effect to include in the model, given the unknown values of their regressor defining parameters. As discussed above, it is proposed to overcome this challenge using forward selection. This section discusses the process of selecting each term to be added to a fitted GBMM.

To do this we look at the fitting of a GQBMM, to a three component mixture experiment. In such a model it is necessary to select a reduced form of the model

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \sum_{i=1}^3 \sum_{j<i} \beta_{ij} \frac{x_i^{r_{ij}} x_j^{r_{ji}}}{(x_i + x_j)^{t_{ij}}}. \quad (4.22)$$

Using forward selection, first a model with one general binary term of joint effect is selected. This requires the three models

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{r_{12}} x_2^{r_{21}}}{(x_1 + x_2)^{t_{12}}}, \quad (4.23)$$

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{13} \frac{x_1^{r_{13}} x_3^{r_{31}}}{(x_1 + x_3)^{t_{13}}} \quad (4.24)$$

and

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{23} \frac{x_2^{r_{23}} x_3^{r_{32}}}{(x_2 + x_3)^{t_{23}}}, \quad (4.25)$$

be fitted and compared using some measure of goodness of fit.

Each of these models is nonlinear in the regressor defining parameters  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$  and hence require the implementation of methods of nonlinear optimisation for their estimation. To do this the variable projection method is applied. Hence, for each pair of components  $x_i$  and  $x_j$ , the estimation occurs through optimisation of the functional

$$\|\mathbf{I} - \mathbf{F}(r_{ij}, r_{ji}, t_{ij}) \mathbf{F}^+(r_{ij}, r_{ji}, t_{ij}) \mathbf{y}\|_2^2, \quad (4.26)$$

rather than

$$\|\mathbf{y} - \mathbf{F}(r_{ij}, r_{ji}, s_{ij}) \beta\|_2^2, \quad (4.27)$$

where  $\beta$  is the vector of linear coefficients,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_{ij}$ , and  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$  are equivalent to  $\alpha$  in Equation 4.17. In this manner, a nonlinear optimisation problem in seven dimensions, including the linear coefficients, is replaced by a somewhat more complex optimisation problem in only three dimensions.

It should be noted that, when considering one general term of binary joint effect in a GBMM, the optimisation problem inherent in estimation will always reduce to three dimensions, regardless of the number of components. Similarly, for a general term of ternary joint effect, the optimisation problem can be expressed in four dimensions, those of the respective regressor defining parameters. Where there are a large number of components, this then has the potential to massively reduce the dimensionality of the optimisation problem required in estimating each model.

For each pair of components  $x_i$  and  $x_j$ , the Functional 4.26 is optimised using numerous sets of initial values for  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$ . This is because, using only one set of starting values may only identify a local optimum for the values of these parameters or indeed no optimum may be found; using a list a different sets of initial values gives the greatest opportunity of finding the global optimum values.

Having estimated Models 4.23, 4.24 and 4.25, using the numerous sets of starting values, the best fitted of each of these models is then identified and each of these three models is then compared to each other and one model selected as 'best'. This model is expressed

$$E[y] = \sum_{i=1}^3 \hat{\beta}_i x_i + \hat{\beta}_{i_1 j_1} \frac{x_{i_1}^{\hat{r}_{i_1 j_1}} x_{j_1}^{\hat{r}_{j_1 i_1}}}{(x_{i_1} + x_{j_1})^{\hat{t}_{i_1 j_1}}}, \quad (4.28)$$

where  $i_1$  and  $j_1$  identify the components present in the chosen general term of joint effect and  $\hat{r}_{i_1j_1}$ ,  $\hat{r}_{j_1i_1}$ ,  $\hat{t}_{i_1j_1}$ ,  $\hat{\beta}_i$  and  $\hat{\beta}_{i_1j_1}$  indicate the estimated values of the parameters. Without loss of generality, it is supposed  $i_1 = 1$  and  $j_1 = 2$

This model found, a second term can now be added. This requires the two models,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{\hat{r}_{12}} x_2^{\hat{r}_{21}}}{(x_1 + x_2)^{\hat{t}_{12}}} + \beta_{13} \frac{x_1^{r_{13}} x_3^{r_{31}}}{(x_1 + x_3)^{t_{13}}} \quad (4.29)$$

and

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{\hat{r}_{12}} x_2^{\hat{r}_{21}}}{(x_1 + x_2)^{\hat{t}_{12}}} + \beta_{23} \frac{x_1^{r_{23}} x_3^{r_{32}}}{(x_2 + x_3)^{t_{23}}}, \quad (4.30)$$

to be estimated and compared. Here,  $\hat{r}_{12}$ ,  $\hat{r}_{21}$  and  $\hat{t}_{12}$  indicating the values of these regressor defining parameters estimated at the previous step, will be fixed during this current round of estimation at the values in Model 4.28.

Hence, fitting Models 4.29 and 4.30 requires the estimation of the regressor defining parameters of the new term only. Therefore, using the variable projection method, estimation is once again possible through optimisation of the functional

$$\|\mathbf{I} - \mathbf{F}(r_{ij}, r_{ji}, t_{ij}) \mathbf{F}^+(\mathbf{x}, r_{ij}, r_{ji}, t_{ij}) \mathbf{y}\|_2^2, \quad (4.31)$$

where now  $\beta = (\beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{ij})$ , and the  $i^{th}$  and  $j^{th}$  components are now the pairs of components for which there is not already a general term of binary joint effect present in the model, in this case  $x_2$  and  $x_3$ , and  $x_1$  and  $x_3$ . As with the fitting of Models 4.23, 4.24 and 4.25 this is a problem in 3 dimensions. Thus, it is seen that by fixing the regressor defining parameters of those terms fitted at earlier steps, the problem of defining the regressor defining parameters of a large number of general terms of joint effect can be overcome.

## 4.6 Reestimating the regressor defining parameters of each term

Having chosen either Models 4.29 or 4.30, once again through estimation for multiple sets of initial values, it may be advantageous to adjust the regressor defining parameters of the term added at the first iteration. This can be done by now fixing the regressor defining parameters estimated for the second term and reestimating those for the first. Supposing Model 4.29 to have been selected over

Model 4.30, then the model,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{r_{12}} x_2^{r_{21}}}{(x_1 + x_2)^{t_{12}}} + \beta_{13} \frac{x_1^{\hat{r}_{13}} x_3^{\hat{r}_{31}}}{(x_1 + x_3)^{\hat{t}_{13}}}, \quad (4.32)$$

is fitted, in the same manner as above, thus finding new values for  $r_{12}$ ,  $r_{21}$  and  $t_{12}$ . Here,  $\hat{r}_{13}$ ,  $\hat{r}_{31}$  and  $\hat{t}_{13}$  appear with hats to indicate they are fixed at the values found for them at the most recent, previous step of the estimation procedure. Throughout this chapter hats upon model parameters indicate they are fixed at such values and therefore, are not estimated in fitting the presented model. The model

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{\hat{r}_{12}} x_2^{\hat{r}_{21}}}{(x_1 + x_2)^{\hat{t}_{12}}} + \beta_{13} \frac{x_1^{r_{13}} x_3^{r_{31}}}{(x_1 + x_3)^{t_{13}}} \quad (4.33)$$

can then be refitted to identify new values for  $r_{13}$ ,  $r_{31}$  and  $t_{13}$  in light of the new (fixed) values  $\hat{r}_{12}$ ,  $\hat{r}_{21}$  and  $\hat{t}_{12}$ .

## 4.7 Reestimating the regressor defining parameters $\mathbf{t}$

In addition to sequential reestimation of the regressor defining parameters of each of the terms in a GBMM, a practitioner may also choose to subsequently reestimate the parameters  $\mathbf{t}$ , made up of all parameters  $t_{ij}$  currently in the model. This can equivalently be viewed as separate reestimation of the vector of parameters  $\mathbf{s}$ , made up of all parameters  $s_{ij}$ . This group of parameters could be considered an interesting distinct subset for such treatment, as their values can not be discussed with respect to a particular two or three component sub-simplex. Hence, continuing the example above, it is proposed to take one more step, where the model,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \beta_{12} \frac{x_1^{\hat{r}_{12}} x_2^{\hat{r}_{21}}}{(x_1 + x_2)^{t_{12}}} + \beta_{13} \frac{x_1^{\hat{r}_{13}} x_3^{\hat{r}_{31}}}{(x_1 + x_3)^{t_{13}}}, \quad (4.34)$$

is estimated with fixed values  $\hat{r}_{ij}$  and  $\hat{r}_{ji}$ , in order to reestimate new values for  $\mathbf{t}$ .

## 4.8 Fitting the GBMM: a graphical representation

The previous section presented a method through which to handle estimation of the regressor defining parameters of multiple general terms of joint effect. This now allows us to present the estimation procedure in full. As the discussion above suggests, this is an iterative procedure, where each iteration has two or three stages. At each iteration a new term is first chosen for the model, and then the regressor defining parameters for each term of joint effect in the model are reconsidered in light of this new term. Finally, the practitioner may then choose to reestimate all values  $\mathbf{t}$  again, simultaneously, where this is the vector of all values  $t_{ij}$  already fit to the model. These three stages are presented separately in Sections 4.8.1, 4.8.2 and 4.8.3, respectively.

The following notation is used in the description below:

- the vector of all parameters  $\beta_i$  and  $\beta_{ij}$  included in the model found at the  $j^{th}$  iteration of the fitting procedure -

$$\beta^{(j)}; \quad (4.35)$$

- unfitted linear model -

$$f_0(\mathbf{x}, \beta^{(0)}); \quad (4.36)$$

- fitted linear model -

$$f_0(\mathbf{x}, \hat{\beta}^{(0)}); \quad (4.37)$$

- the list of unestimated models to be fit in order to identify the new term added at the  $k^{th}$  iteration -

$$f_k(\mathbf{x}, \beta^{(\mathbf{k})}); \quad (4.38)$$

- the estimations of these models found to have the best, respective, goodness

of fit -

$$f_k \left( \mathbf{x}, \hat{\beta}^{(\mathbf{k})} \right); \quad (4.39)$$

- the model decided upon at the  $k-1^{th}$  iteration, with fixed regressor defining parameters for those terms added to the model up to this point -

$$F_{k-1} \left( \mathbf{x}, \beta^{(\mathbf{k}-1)} \right); \quad (4.40)$$

- the model to be estimated in reconsidering the regressor defining parameters of the terms added at the  $j^{th}$  iteration during the  $k^{th}$  iteration -

$$f_k^j \left( \mathbf{x}, \beta^{(\mathbf{k})} \right); \quad (4.41)$$

- the estimation of this model providing the greatest improvement in goodness of fit -

$$F_k^j \left( \mathbf{x}, \hat{\beta}^{(\mathbf{k})} \right); \quad (4.42)$$

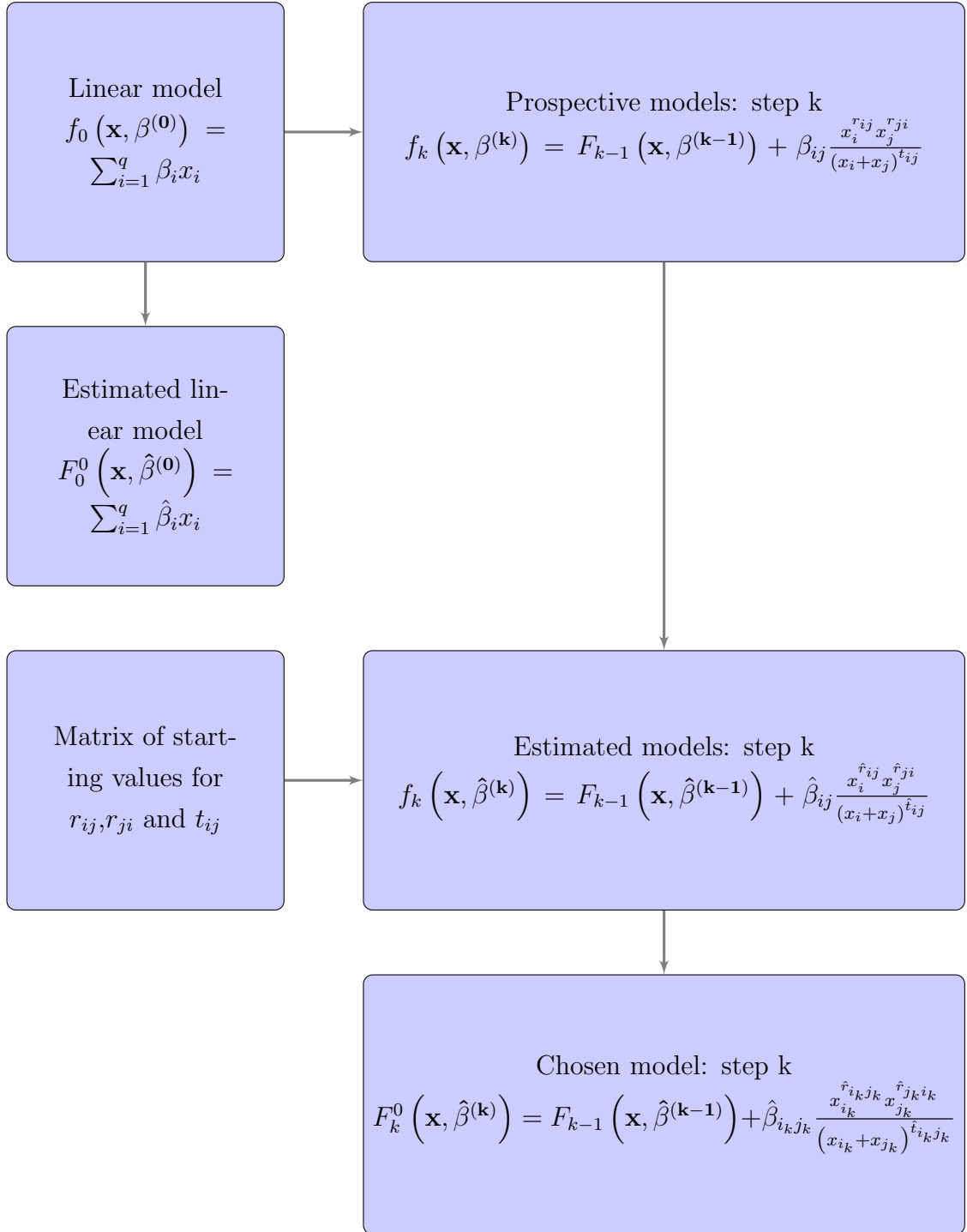
- the current model of the  $k^{th}$  iteration with the term added at the  $j^{th}$  iteration -

$$F_k^{-j} \left( \mathbf{x}, \hat{\beta}^{(\mathbf{k})} \right); \quad (4.43)$$

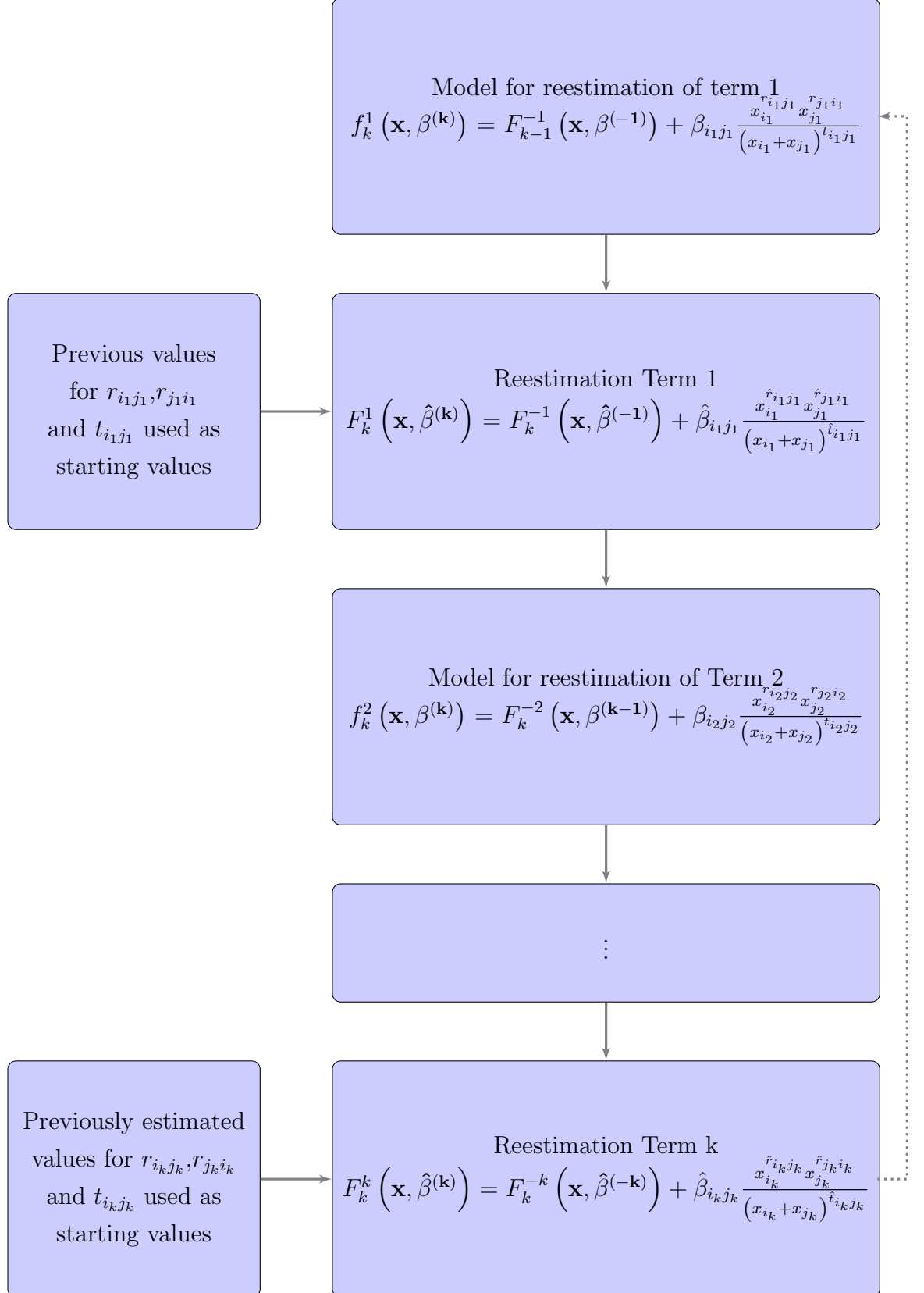
- the model to be estimated with fixed values of  $r_{ij}$  and  $r_{ji}$  of the terms added to that point, but estimable  $t_{ij}$  -

$$f_k^{\mathbf{t}} \left( \mathbf{x}, \beta^{(\mathbf{k})} \right). \quad (4.44)$$

## 4.8.1 Adding a term

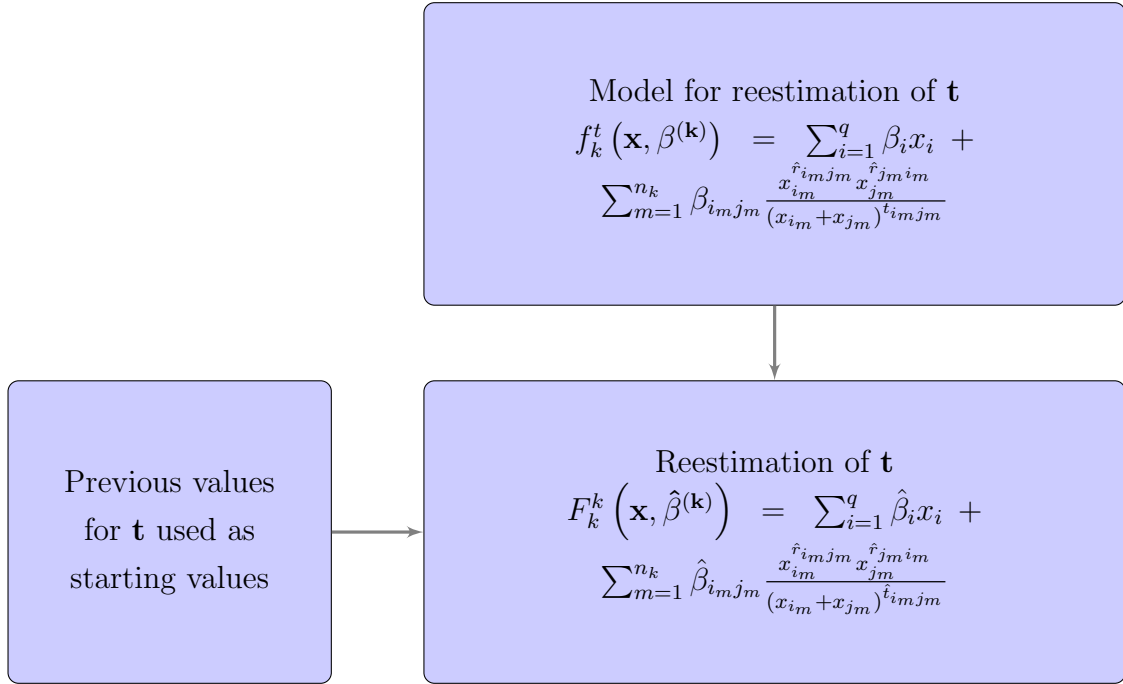


### 4.8.2 Reestimating the regressor defining parameters of each term





### 4.8.3 Reestimating the regressor defining parameters $\mathbf{t}$



## 4.9 How to fit the GBMM: discussion

Thus far the discussion of the fitting procedure for the GBMM has looked exclusively at the GQBMM. Moreover, certain details have been left aside in order to maintain the clarity of presentation of the full procedure. These are now addressed. Some of them ascertain to variations in how the fitting procedure is applied and some of these in turn have been explored in the examples of the following section. There it is hoped to discuss these details of how the fitting procedure shall be implemented, although not necessarily answer this definitively.

### 4.9.1 Ternary terms of joint effect

The additional inclusion of ternary terms of joint effect is simple. At stage 1, each possible ternary term of joint effect not already included in the model is considered for inclusion; at stage 2 all ternary terms included in the model thus far are reestimated in the same manner as the binary terms of joint effect, and at stage 3 the parameters  $t_{ijk}$  in the model are reestimated along side the parameters

$t_{ij}$ . How different the final model may be, as a result of fitting a GSCBMM rather than a GQBMM, will be explored in the next chapter.

#### 4.9.2 How many times shall the terms be reestimated?

In stage 2 of the estimation procedure, presented in section 4.8.2, a dotted line indicates that, having reestimated the regressor defining parameters of each of the terms once, the process of reestimation of terms in the model can occur repeatedly. This is aimed at achieving convergence before progressing to adding another term. This can be governed by continuing until the level of model improvement is below a set threshold or through setting the number times to undertake the process of reestimation.

The latter option is explored in the examples of the following chapter. However, it is also questionable how worthwhile it is to make small adjustments in the regressor defining parameters at this stage, when they may be changed a lot once another term is added at the next iteration. To investigate the effectiveness of such reestimations, examples are given, in the next chapter, where 0, 1 and 10 reestimations are made. In this way it is also investigated whether Stage 2 and Stage 3 should occur at all.

#### 4.9.3 Should $t$ be reestimated separately?

Similarly, the next chapter will also look at whether estimation should occur without Stage 3. Stage 3 could be considered unnecessary in light of the amount of reestimation that may have occurred by that point. Moreover, where a large number of general terms of joint effect have already been added to the model, Stage 3 will demand a high-dimensional nonlinear optimisation problem for which convergence may not easily occur. This is to say, Stage 3 may not be an option. The next chapter will consider the difference inclusion of Stage 3 in the fitting procedure can make to the final model.

#### 4.9.4 What set of initial values shall be used for $r_{ij}$ , $r_{ji}$ and $t_{ij}$ ?

While it has been suggested that in Stage 1, when adding a new term to the model, a list of different sets of initial values should be used, no attempt has yet

been made to define how these values should be chosen. Moreover, no discussion has been made of the initial values used in Stage 2 and Stage 3.

At this point, it should be acknowledged that fitting of a model for multiple sets of starting values is time consuming. Therefore, reasoned ways of reducing the number of sets of initial values would be useful in order to reduce the implementation time of the fitting procedure. Often many sets of starting values will result in the same locally optimum values for  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$  and hence, it will not be necessary to explore any more than one of these. This said, the number of sets of starting values should not be so limited as to be detrimental to the final model. The next chapter looks specifically at how using respectively larger and smaller lists of sets of initial values in Stage 1 can change the final model.

Regarding the initial values used for reestimation in Stages 2 and 3, at this point all parameters have been previously estimated, and it is felt these estimated values should not be discarded completely. Therefore, it is suggested to use as the initial values for any regressor defining parameters to be reestimated, their current estimated values.

#### 4.9.5 How many general binary terms of joint effect can be considered for each pair of components?

Where forward regression is implemented upon the Scheffé quadratic polynomial, which can occur in the same manner as discussed for the PQM, only one term of joint effect is considered between the components  $x_i$  and  $x_j$ . This is because there is only one such term between  $x_i$  and  $x_j$ , that is,  $\beta_{ij}x_ix_j$ . In the case of the general binary term of joint effect, it is possible to have a model where two regressors  $\beta_{ij} \frac{x_i^{r_{ij}} x_j^{r_{ji}}}{(x_i + x_j)^{t_{ij}}}$  are used, with different values for the regressor defining parameters  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$ . So, once one of these terms is added to the model, should another be considered? In examples given here this has not been done. Thus, there may only be one blending term of each type between  $x_i$  and  $x_j$ , or equivalently  $x_i$ ,  $x_j$  and  $x_k$ , for the general cubic term of joint effect. However, it is possible that considering more than one term would allow a worthy extension of the types of GBMM which could be fit.

### 4.9.6 $AIC_c$ and the number of parameters

The  $AIC_c$  has been used here, as a measure of model parsimony, in the examples of subsequent chapters. Only the linear parameters have been counted when calculating the penalisation. The logic for this is thus: given there are existing models which can be considered special cases of the GBMM, and fitting these models and comparing them would not take into account that there exist implicit values of the regressor defining parameters in these models, it would be remiss to count such parameters in the GBMM, when what was proposed was a systematic manner of exploring a greater number of special cases, i.e. model estimation should not be penalised for now additionally performing the process of model selection. For example, say the two models

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{12} x_1 x_2 \quad (4.45)$$

and

$$E[y] = \sum_{i=1}^q \beta_i x_i + \beta_{12} \frac{x_1 x_2}{x_1 + x_2}, \quad (4.46)$$

are estimated and compared. The term of joint effect of the first model has implicit value  $s_{12} = 2$  (or equivalently,  $t_{12} = 0$ ) and the term of joint effect of the second model has implicit value  $s_{12} = 1$  ( $t_{12} = 2$ ), but these parameters would not be counted in a penalised measure of model parsimony. Similarly, 100 models could be fit separately and compared, each with  $s_{12}$  equal to 100 different values between 0 and 1, each providing slightly different fit. Once again the implicit presence of the regressor defining parameters would not be considered. However, should nonlinear regression be used to choose  $t_{12}$ , even when restricted between 0 and 1, the model fit would be penalised for the presence of  $t_{12}$ . I have considered this illogical, so it is proposed not to count regressor defining parameters when penalising the fit of the GBMM.

### 4.9.7 Stopping criterion

The final model of each full iteration of the fitting procedure is  $F_k^k$ . Where Stage 2 and Stage 3 are not applied,  $F_k^0$  becomes  $F_k^k$ . In the typical manner of forward selection this model is then compared, using some criteria, to the model found at the previous iteration of the fitting procedure. In the examples given here that

criterion is the  $AIC_c$ . Should the  $AIC_c$  be reduced from the previous iteration of the fitting procedure, another iteration is undertaken.

## 4.10 Summary

As the previous section discussed, there are several ways to vary the fitting procedure proposed for the GBMM. The examples that shall be given in the following chapter illustrate some of these variations. In particular, these examples will look at the benefits from including general ternary terms of joint effect, increasing the number of sets of initial values for  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$  used in Stage 1, increasing the number of reestimations in Stage 2 and the use of Stages 2 and 3 at all.

## Chapter 5

# Application of fitting procedure for the GBMM

This chapter shall give examples illustrating the fitting procedure for the GBMM. Multiple GBMM will be fitted to data presented by Cornell (2002, p.543-546), where the effects of the proportions of protein ( $x_1$ ), fat ( $x_2$ ) and carbohydrate ( $x_3$ ), in a force-fed diet, upon the fat gain ( $y$ ) of chicks were studied. This shall be done using several of the variations described in the previous chapter. There were 30 observations and the ranges taken by the components were

$$0.05 \leq x_1 \leq 0.40,$$

$$0.06 \leq x_2 \leq 0.86$$

and

$$0.02 \leq x_3 \leq 0.89.$$

Eight examples shall be given, where each example may contain more than one model. The purpose of each example is to progress the discussion around the fitting procedure with respect to a specific feature. The first four examples include only GQBMM.

- Example F1 fits a GQBMM without Stage 2 or Stage 3 of the fitting procedure; this example will help examine how the fitting procedure benefits from reestimation of regressor defining parameters.

- Example F2 fits two GQBMM without Stage 3, exploring specifically reestimation of regressor defining parameters in Stage 2; this example will look at how multiple reestimations of the regressor defining parameters, that is multiple implementations of Stage 2, can change the final estimated model.
- Example F3 fits a GQBMM without Stage 3, but with a fixed number of passes of Stage 2; compared to Example F2, the model will be fit using a different sized set of initial values for the regressor defining parameters in Stage 1. Hence this example looks at how this can impact upon the final model. The sets of initial values are described here with reference to parameters  $t_{ij}$  and  $t_{ijk}$ , but could instead be described in respect to  $s_{ij}$  and  $s_{ijk}$ .
- Example F4 fits one GQBMM with the full fitting procedure; the purpose of this example is to look at how the inclusion of Stage 3 in the fitting procedure can change the final model.

The subsequent four examples, Examples F5-F8, have the same objectives as Examples F1-F4, respectively. However, in these examples GSCBMM are fit where GQBMM were fit before. This will allow a discussion of the potential benefits of inclusion of general ternary terms of joint effect to be initiated. In all examples the model presented is that chosen by using the  $AIC_c$  as a stopping criterion as described in the previous section. All models are presented with their parameters given to 4 significant figures alongside their corresponding  $R^2$  and  $AIC_c$  values.

R code giving the different versions of the fitting procedure is available from the author. The fitting procedure applied in Examples F1-F3, is given in Fitting Procedure 1.txt, as applied in Example F4, is given in Fitting Procedure 2.txt, as applied in Examples F5-F7, is given in Fitting Procedure 3.txt and as applied in Example F8, is given in Fitting Procedure 4.txt.

Subsequent to the presentation of these examples, the response surface shall be given for two models from Examples F1-4 and two models from Examples F5-8. This is done alongside a reduced Scheffé cubic model. This will begin to establish the potentially advantageous use of the GBMM over established models, while also continuing to progress the discussion on the benefits of the ternary terms of joint effect.

## 5.1 Example F1

In this example, a GQBMM is fitted, using the set of initial values found from all combinations of  $(0, 3, 6, 9, 12)$  for  $r_{ij}$  and  $r_{ji}$  and  $(-12, -9, -6, -3, 3, 6, 9, 12, 15, 18, 21, 24)$  for  $t_{ij}$ . There is no attempt at reestimation of terms estimated at earlier iterations of the fitting procedure, that is no implementation of either Stage 2 or Stage 3 of the fitting procedure.

The resultant model,

$$E[y] = -18.34x_1 + 61.73x_2 + 42.26x_3 - 73.01 \frac{x_2^{0.8186} x_3^{0.5777}}{(x_2 + x_3)^{-21.67}} + 1.172 \times 10^8 \frac{x_1^{18.91} x_3^{2.280 \times 10^{-8}}}{(x_1 + x_3)^{0.5426}}, \quad (5.1)$$

contains two terms of joint effect, those between  $x_1$  and  $x_3$ , and  $x_2$  and  $x_3$ . It has  $R^2$  value, 0.9230, and  $AIC_c$ , 144.5738.

## 5.2 Example F2

In this example, two GQBMMs are fit using the same list of initial values for  $r_{ij}$ ,  $r_{ji}$  and  $t_{ij}$  outlined in Example F1. However, for both models, reestimation of the terms added at previous iterations of the fitting procedure has now occurred. This has been undertaken once and 10 times, respectively, upon each iteration. Alternatively, it could be said there have been one and ten passes of Stage 2, respectively, upon each addition of a new term to the model.

The resultant models are the same, although subtly different from that found without reestimation of terms added at previous iterations of the fitting procedure. Given that the new model (found with both 1 and 10 reestimations),

$$E[y] = -18.34x_1 + 61.72x_2 + 42.27x_3 - 73.17 \frac{x_2^{0.8192} x_3^{0.5810}}{(x_2 + x_3)^{-21.64}} + 1.172 \times 10^8 \frac{x_1^{18.91} x_3^{9.143 \times 10^{-8}}}{(x_1 + x_3)^{0.5458}}, \quad (5.2)$$

has parameter estimates very similar to those for the model found in the previous example, they can be thought to give essentially the same representation of the response surface.  $R^2 = 0.9230$  and  $AIC_c = 144.5725$ .



### 5.3 Example F3

This example is very similar to Example F2. The fitting procedure is undertaken with 10 passes of reestimation upon the inclusion of each additional term. However, in contrast to Example F2, the list of initial values for  $r_{ij}$ , for  $r_{ji}$  and for  $t_{ij}$  is now larger, taking all combinations of the sequence from 0 to 24, at intervals of 3, for  $r_{ij}$  and  $r_{ji}$  and the sequence from -24 to 48, at intervals of 3, for  $t_{ij}$ .

The difference between the consequent model and those presented in the previous examples is, once again, negligible. The model,

$$E[y] = -18.34x_1 + 61.72x_2 + 42.27x_3 - 73.21 \frac{x_2^{0.8196} x_3^{0.5813}}{(x_2 + x_3)^{-21.64}} + 1.172 \times 10^8 \frac{x_1^{18.91} x_3^{2.9069 \times 10^{-7}}}{(x_1 + x_3)^{0.5427}}, \quad (5.3)$$

has  $R^2 = 0.9230$  and  $AIC_c = 144.5737$ . Interestingly, this indicates a model slightly less well fit than that given in the previous example, a manifestation of the stochastic nature of forward selection. This indicates that, while expanding the list of initial values and increasing the number of times reestimation of terms is undertaken is theoretically more thorough, it may lead to deviations away from the path towards the most parsimonious model. Admittedly, in this case that deviation is very small.

### 5.4 Example F4

This section looks at the difference the inclusion of Stage 3 can make to the estimation procedure. Hence, in addition to the reestimation of each term in turn,  $\mathbf{t}$  is also subsequently reestimated. As shall be seen, the result can be a different choice of model. The number of times previously added terms are reestimated, that is the number of passes of Stage 2, is now fixed at 10 and the smaller list of initial values, introduced in Example F1, has been used. The effect of reestimating  $\mathbf{t}$  is to arrive at a three term model which wasn't gained using

the fitting procedures of the previous examples. This model,

$$E[y] = -17.19x_1 + 63.74x_2 + 42.02x_3 - 46.69 \frac{x_2^{0.6430} x_3^{0.3658}}{(x_2 + x_3)^{-16.93}} \\ + 6.402 \times 10^7 \frac{x_1^{18.65} x_3^{0.0254}}{(x_1 + x_3)^{1.833}} - 1.048 \times 10^8 \frac{x_1^{9.491} x_2^{14.62}}{(x_1 + x_2)^{17.52}}, \quad (5.4)$$

has  $R^2 = 0.9363$  and  $AIC_c = 142.0389$ . Evidently, the reestimation of  $\mathbf{t}$  can be used advantageously.

The following section progresses to explore GSCBMM. In fitting such models, more time is required in order to consider the ternary terms of joint effect. As such, fitting can be more time consuming. In order to capture this, each of the models in Examples F5-8 possess analogous examples in Examples F1-4 and comparison is made of the time taken to fit these models, on a computer with a dual core 3GHz processor and 3.5GB RAM.

## 5.5 Example F5

As with Example F1, here there is no attempt at reestimation of terms fitted to the model at earlier iterations and the list of initial values for the binary terms of joint effect is all combinations of  $(0, 3, 6, 9, 12)$  for  $r_{ij}$  and  $r_{ji}$  and  $(-12, -9, -6, -3, 0, 3, 6, 9, 12, 15, 18, 21, 24)$  for  $t_{ij}$ . The set of initial values for the ternary terms of joint effect is all combinations of  $(0, 6, 12)$  for  $r_{ijk}$ ,  $r_{jki}$  and  $r_{kij}$  and  $(-12, -6, 0, 6, 12, 18, 24, 30, 36)$  for  $t_{ijk}$ .

However, in spite of the introduction of possible ternary terms of joint effect, the model identified is no different from that found in Example F1. Obviously, this is possible given that the GQBMM is a special case of the GSCBMM. In addition to the models being the same, the difference in the time taken for the model to be fit is small: the model found in Example F1 took 20 minutes to fit and that found in this example took 21 minutes to fit. However, as shall be seen in later examples, this very small difference in estimation time is anomalous among these examples.

## 5.6 Example F6

Reestimation of terms is now added to the fitting of the GSCBMM. This has been done once and 10 times, as with Example F2 above, resulting in two models which can be ostensibly thought of as the same model. However, unlike the model found in Example F5, this model includes a term of ternary joint effect. The model,

$$\begin{aligned}
 E[y] = & -26.76x_1 + 67.99x_2 + 47.23x_3 - 39.42 \frac{x_2^{0.4443}x_3^{0.2524}}{(x_2 + x_3)^{-10.94}} \\
 & + 3.553 \times 10^7 \frac{x_1^{17.63}x_3^{1.614}}{(x_1 + x_3)^{1.325}} - 1.023 \times 10^8 \frac{x_1^{9.391}x_2^{15.04}}{(x_1 + x_2)^{18.37}} \\
 & + x_1^{15.89}x_2^{5.825}x_3^{66.81}, \tag{5.5}
 \end{aligned}$$

has  $R^2 = 0.9486$  and  $AIC_c = 137.5857$ . This evidences that the inclusion of a ternary term can improve the fit of a model. Moreover, as shall be shown later, this can result in a contrasting description of the response surface.

Given that the models found, when reestimating terms once or 10 times, were almost the same, this leads to a question of whether this is worthwhile? Obviously the latter is more time consuming. In this example, reestimating once allowed the model to be fit in 52 minutes - 5 minutes faster than when reestimating 10 times. Yet, this resulted in the same model. Practitioners may wish to consider whether the gains that can be made by multiple reestimations are worthwhile, while further research should establish just how substantial such gains can be.

At this point it is also worth noting that the additional consideration of ternary terms meant that the equivalent fitting procedure for the GQBMM, used in Example F2, took only 20 minutes (in both cases). In this situation, this extra time appears to have been worthwhile, but in other contexts, with a larger number of components, more data and with more terms being fitted, a practitioner may veer towards only exploring GQBMM in order to save time.

## 5.7 Example F7

Fixing the number of times for reestimation at 10, but using the larger list of initial values, of all combinations of (0, 3, 6, 9, 12, 15, 18, 21, 24) for  $r_{ijk}$ ,  $r_{jki}$  and

$r_{kij}$  and at all values at intervals of 6, from -24 to 72, for  $t_{ijk}$ , the model,

$$\begin{aligned} E[y] = & -15.78x_1 + 58.51x_2 + 42.62x_3 - 59.08 \frac{x_2^{0.5921}x_3^{0.6735}}{(x_2 + x_3)^{-23.32}} \\ & + 1.197e + 08 \frac{x_1^{19.24}x_3^{0.0441}}{(x_1 + x_3)^{1.253}} + 1.473e + 37x_1^{17.32}x_2^{104.1}x_3^{8.141}, \end{aligned} \quad (5.6)$$

is fitted. It is interesting that in introducing a larger set of initial values a model is reached which is less well fit than that found with the smaller set of initial values:  $R^2 = 0.9363$  and  $AIC_c = 140.8038$ . Here the stochastic nature of forward selection deviates from the path to that model found in Example F6.

## 5.8 Example F8

In this final example, as was done with respect to the GQBMM in Example F4, separate reestimation of  $\mathbf{t}$  is introduced, that is Stage 3 is introduced to the fitting procedure. Reestimation of terms occurs 10 times, that is Stage 2 is implemented 10 times upon each addition of a new term to the model, and the smaller list of initial values, used for Examples F5 and F6, is applied.

The model,

$$\begin{aligned} E[y] = & -24.07x_1 + 68.53x_2 + 44.58x_3 - 38.60 \frac{x_2^{0.5119}x_3^{0.2615}}{(x_2 + x_3)^{-11.74}} \\ & + 1.013e + 08 \frac{x_1^{18.73}x_3^{0.0004581}}{(x_1 + x_3)^{0.8778}} - 1.059e + 08 \frac{x_1^{8.729}x_2^{16.74}}{(x_1 + x_2)^{20.17}} \\ & - 3.135e + 18x_1^{4.779}x_2^{26.37}x_3^{13.66}, \end{aligned} \quad (5.7)$$

has  $R^2 = 0.9464$  and  $AIC_c = 140.2696$ . This means it has a worse fit than the model in Example F6, in spite of the more convoluted fitting procedure. Once again, this is clearly an effect of the stochastic nature of the model building procedure; here, as with Example F7, through attempting to optimise improvements in model fit at each step, the model identified in Example F6 is not obtained. Consequently, the models describe different response surfaces. This is explored in the following section.

	Model 5.2	Model 5.4	Model 5.5	Model 5.7
$R^2$	0.9230	0.9363	0.9486	0.9464
$AIC_c$	144.5725	142.0389	137.5857	140.2696

Table 5.1: Statistics comparing GBMM found, with different fitting procedure, for chick diet data

## 5.9 Comparison of GBMMs

Following on from the work of the first half of this chapter, this section compares four of the models found thus far. These are Models 5.2 and 5.5, the most parsimonious GQBMM and GSCBMM fitted without separate reestimation of  $\mathbf{t}$ , and Models 5.4 and 5.7, the GQBMM and GSCBMM found in Examples F4 and F8, where Stage 3 of the fitting procedure was implemented and separate reestimation of  $\mathbf{t}$  did occur.

The goodness of fit statistics for these models are given in Table 5.1 and the plots of their predicted response surfaces are given in Figure 5.1. Evidently, there are advantages to the inclusion of ternary terms, which have allowed a better fit both with and without reestimation of  $\mathbf{t}$ . Moreover, they have identified shapes in the response surface which are not evident in those described by the GQBMM. A manner through which to test the veracity of these descriptions of the response surface does not exist in this specific context, but it is evident the ternary terms can be a worthy use of the extra time required to fit them.

There is some contrast in the response surface representation given by the two GQBMM, Models 5.2 and 5.4. There are two obvious features, the contours arriving from either side of the experimental region, while the plot for Model 5.4 includes an additional set of contours, arriving from the bottom right. Each of these three features is evident in the plots for Models 5.5 and 5.7, where each has different additional features, attributable to their respective ternary terms of joint effect. For Model 5.5 this is the slight additional curvature at the top of the experimental region, where for Model 5.7 it is the far more obvious feature.

These contrasting surfaces for the two GSCBMM could be viewed as problematic. However, there is only one term of ternary joint effect in each model, since the fitting procedure was deliberately limited to fitting only one term of joint effect between each pair or triplet of components. More than one term of any one type could enter a final model and it may be that this would allow both these features to be picked up in a final description of the response surface, if so

desired.

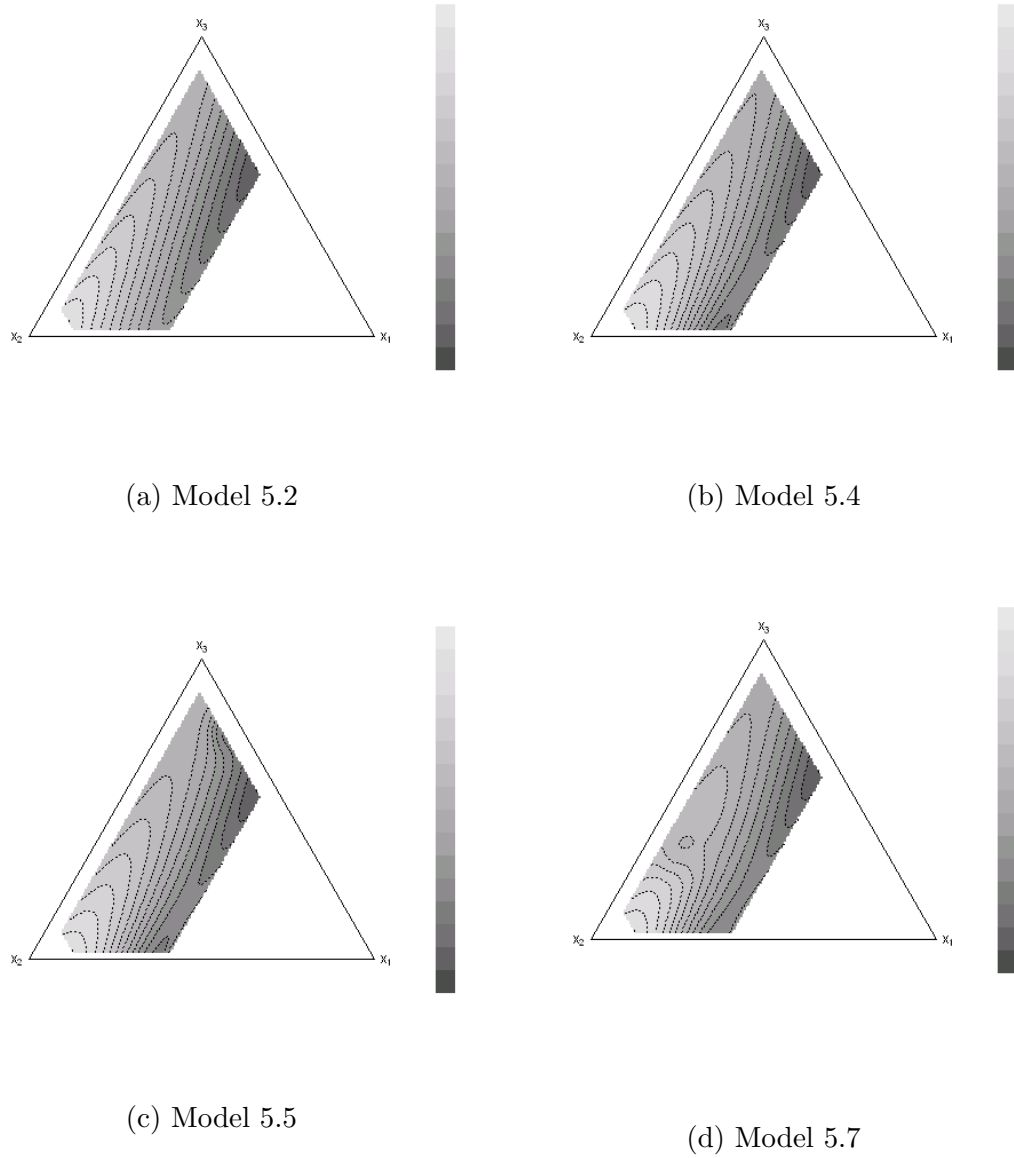


Figure 5.1: Response surfaces for GBMM for chick diet data

## 5.10 Comparison to other models

Having compared the different models identified with variations of the fitting procedure, the advantage of using the GBMM is now assessed. The best fitting

	Linear	Reduced Scheffé cubic	Model 5.5
$R^2$	0.8661	0.9250	0.9486
$AIC_c$	155.5686	150.3390	137.5857
No of Terms	3	7	7

Table 5.2: Statistics comparing GBMM to a reduced Scheffé cubic polynomial, for chick diet data

model identified above (Model 5.5) is compared to the model,

$$\begin{aligned}
 E[y] = & 561.43x_1 + 52.12x_2 + 33.54x_3 - 1061.54x_1(x_3 + x_2) - 33.71x_2x_3 \\
 & - 638.44x_1(x_2(x_1 - x_2) + x_3(x_1 - x_3)) + 1448.40x_1x_2x_3.
 \end{aligned} \tag{5.8}$$

This was found through reducing the full Scheffé cubic polynomial through combining and removing terms. This seemed necessary for an effective comparison as no reduced form of the Scheffé cubic polynomial, without such manipulation, proved to have a lower  $AIC_c$  than the linear model.

The statistics of fit for these two models, along with those for the linear model, are given in Table 5.2, from which it is evident the GBMM gives the better fit. There is a degree of agreement between the two models about the major features in the response surface. However, Model 5.5 gives additional features in the bottom right and the top left of the plot, as can be seen in Figure 5.2. This is significant as the Scheffé cubic polynomial, the most complex model commonly applied to mixture experiments, has not described these additional features; there is scope with the GBMM to describe things that commonly applied models, in all their reparameterisations, do not describe. However, in order to benefit from this it will prove important to insure against overfitting.

## 5.11 Conclusions

The purpose of the previous two chapters has been to introduce the fitting procedure for the GBMM and discuss some of its variations. Most significantly, it is evident that the fitting procedure allows the GBMM to be used effectively. This initial example suggests the description of the response surface given by GBMM possesses possible advantages over those of Scheffé polynomials: it can provide better fit and it can describe features which other models cannot. This shall be evidenced further in the later examples.

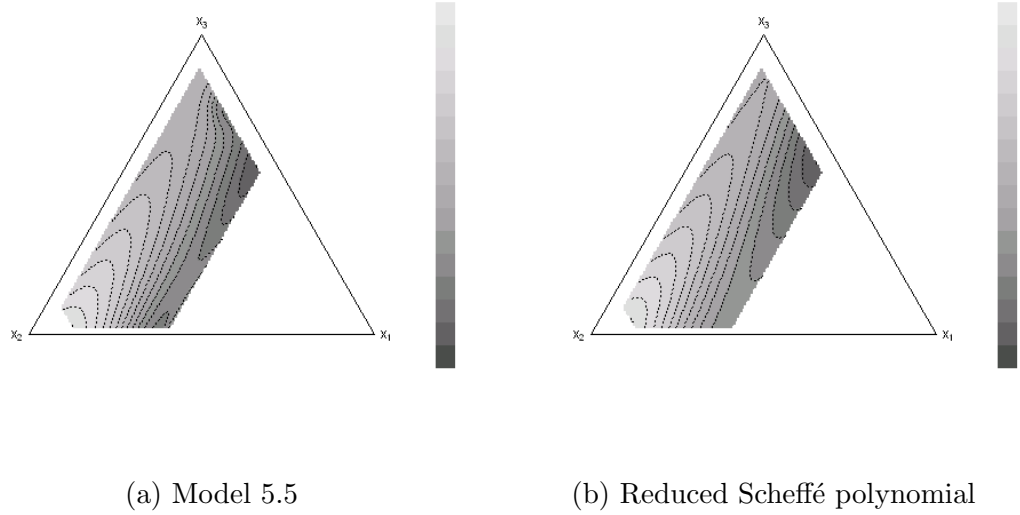


Figure 5.2: Predicted response surface for the chick diet data

It is possible that the fitting procedure itself is a novelty. It was necessary, required in order to fit the relatively complex models being proposed. Therefore, it cannot be said to be well informed by the canon on nonlinear model fitting literature. Moreover, it would be generous to describe it as elegant. However, the exploitation of the partially linear structure in order to build models of multiple nonlinear terms is evidently effective and at least can be viewed as an adroit solution to the fitting problem posed.

The example discussed in this chapter demonstrates that, depending on the variation of the fitting procedure, different models may be reached for the same data. As seen, this does not necessarily mean substantially different conclusions and is really to be expected given each variation upon the fitting procedure can be viewed as a separate procedure entirely. Different results were sought in order to begin a discussion of which variation of the fitting procedure could be considered best. However, it does lead to the question of which model to ultimately select.

Certainly, the differences in the final model are attributable to the stochastic nature of all forward selection procedures and were expected, but it means that concrete conclusions on how best to undertake the model fitting procedure are difficult. Moreover, the quality of the estimation procedure when applied to other data will depend upon the size and the quality of the data, the complexity of the problem and the level of variability.



Based on current experience, the best suggestion appears to be to undertake several variations of the fitting procedure as has been done here. Using general ternary terms of joint effect appears effective, but when implemented as described may still lead away from the most parsimonious modelling solution. This said, they potentially allow behaviour to be described which otherwise cannot. For this reason if it is felt they should always be included, time permitting. Such concrete suggestions, regarding the other variations are less comfortably made.

The reestimation of  $\mathbf{t}$  has been demonstrated to result in improved model fit in some cases, but not in others. Intuitively, the idea of reestimating some parameters simultaneously across terms of joint effect appeals and the additional reestimation will always create improvements within each iteration of the fitting procedure, if not overall. This can similarly be said of additional reestimation of each term, although it appears, in the cases above, the gains to be made through reestimation diminished across each consecutive reestimation. It is likely time gains could be made by limiting this aspect of the fitting procedure, although the more terms, the greater the gains are likely to be from multiple reestimations.

Finally, larger sets of initial values did not appear worthwhile above, and it could be concluded that time gains can be made using smaller sets. However, it is difficult to make conclusions on how to choose these based on such a simple study.

# Chapter 6

## Design of Experiments

The next two chapters shall look at the initial work on developing experimental designs for the GBMM. By dictating the appropriate collection of data, optimal experimental design theory helps provide answers with the most efficient use of experimental resources. Such theory provides algorithms for the selection of experimental designs and criteria by which to assess them. Both shall be presented in this chapter in addition to ideas on experimental designs for non-linear models. Designs for the GBMM are then presented in the following chapter.

The issue of optimal experimental designs for other mixture models will also be addressed briefly. However, it is reiterated that the designs here are dictated by the newly proposed model. In taking an approach motivated by a model not yet considered in other work and one with greater flexibility than other models, we feel a novel method is presented of exploring the influence of component blending on the response and how this may deviate from that described by the established models. For a thorough summary of work on experimental design for mixture experiments the reader may refer to Chan (2000), in addition to Cornell (2002).

### 6.1 Continuous and exact designs and their optimality

The design region of an experiment is all values  $\mathbf{x}$  for which responses  $\mathbf{y}$  may be observed. This may or may not coincide with the region of interest of the experiment. The support points of a design are the distinct values  $\mathbf{x}$  at which responses  $\mathbf{y}$  are observed. An experimental design defines the choice of support

points.

An experimental design in which the distribution of trials over a design region  $\chi$  that is specified by a measure,  $\xi$ , is called continuous. Supposing there to be three sets of values  $\mathbf{x}$  at which observations are made, a measure  $\xi$ , where there are  $n$  distinct support points of the design, is expressed

$$\xi = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\ w_1 & w_2 & w_3 \end{Bmatrix},$$

where observations are made at the factor values  $\mathbf{x}_i$ , with weighting  $w_i$ . Since  $\xi$  is a measure  $\sum_i w_i = 1$ .

In practice all designs are exact, such that they contain a specific number of design points  $N$ . In this case the measure

$$\xi_N = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\ r_1 & r_2 & r_3 \end{Bmatrix},$$

where  $r_i$  are the integer number of observations at  $\mathbf{x}_i$  and  $\sum_i r_i = N$ . Often good exact designs can be found approximating the continuous design  $\xi$ .

There can be many motivations for choosing a particular experimental design. In the theory of continuous designs the objective is to minimise a measure of imprecision  $\Psi(M(\xi))$ . In particular, D-optimal designs minimise

$$\Psi(M(\xi)) = -\log |M(\xi)|, \quad (6.1)$$

or equivalently maximise the the determinant of the information matrix,

$$|F^T F| = |M(\xi_i)|. \quad (6.2)$$

D-optimality achieves the minimum volume of the confidence ellipsoid of the parameters. The length of the axes of this ellipsoid are proportionate to the squares of the eigenvalues,  $\lambda_i$ , of the information matrix. Therefore, a D-optimal design also minimises the product of the eigenvalues,

$$\min \prod_i \lambda_i. \quad (6.3)$$

There exist other types of design optimality related to the eigenvalues of the

information matrix (see Atkinson, Donev, and Tobias (2007)), although these are not relevant to the discussion of the following chapter.

Two other design criteria which shall be referred to later are G-optimality and V-optimality. A G-optimal design minimises the maximum of the generalised prediction variance over the design region i.e.

$$\min \max d\{\mathbf{x}, \xi^*\}, \quad (6.4)$$

where

$$d\{\mathbf{x}, \xi^*\} = f^T(\mathbf{x}) M^{-1}(\xi^*) f(\mathbf{x}), \quad (6.5)$$

while a V-optimal design seeks to minimize the average prediction variance over the design points, that is,

$$\min \int d\{\mathbf{x}, \xi^*\} d\mathbf{x}. \quad (6.6)$$

## 6.2 General Equivalence Theorem

In general, for continuous designs, a design satisfying one type of optimality will also satisfy another. Hence, a D-optimal design will also be G-optimal. This is because of the General Equivalence Theorem (GET). This states the equivalence of the following three conditions for  $\xi$ :

- The design  $\xi$  minimises  $\Psi(M(\xi))$ ;
- The design  $\xi$  maximises the minimum of  $\psi(M(\xi))$ , the derivative of  $\Psi(M(\xi))$ ;
- The maximum of  $\psi(M(\xi))$ , over the design region, is equal to 0. This maximum occurs at the points of support of the design.

For a design with  $p$  support points, this also means that the maximum of  $d\{\mathbf{x}, \xi^*\} = p$  at the support points of the design.

While true for continuous designs, it is not necessarily so that an exact design, for a particular  $N$ , will satisfy the GET. This means that a design of size  $N_1$  may be D-optimal and another of size  $N_2$  G-optimal. Identifying such designs requires the implementation of computer searches. The manner in which this has been undertaken for the GBMM will be described once optimal design theory for nonlinear models has been discussed.

### 6.3 Experimental design for nonlinear models

Thus far experimental design has been discussed in the context of linear in parameters models. Therefore, the choice of design has not depended on the values of the parameters of the model. This section addresses the contrasting case of nonlinear models. In this case, the optimum design depends upon the values taken by the *nonlinear* parameters. There will be particular reference to partially linear models.

For simplicity, consider the case of a model with a single nonlinear parameter  $\alpha$ . This parameter is unknown and therefore requires assumptions to be made about its value. These assumptions may take the form of a point prior, where it is assumed that

$$\alpha = \alpha_0, \quad (6.7)$$

allowing locally optimal designs. Alternatively, a prior distribution may be assumed, where  $\alpha$  is assumed to take one of a range of values each with probability,  $p(\alpha)$ . We shall address both cases here and use each to propose designs for the GBMM.

Suppose

$$E[y] = \eta(\beta, \alpha), \quad (6.8)$$

where  $\beta$  are linear coefficients. As the  $\beta$  enter the model linearly, their values do not effect the optimum design. However, the design is influenced by the value of  $\alpha$ . As such, prior information on  $\alpha$  is required to identify an optimal design. Initially,  $\alpha$  takes the point prior  $\alpha_0$ .

In such a case a Taylor expansion creates the linearised model

$$\begin{aligned} E[y] = \eta(\beta, \alpha) &\approx \eta(\beta, \alpha_0) + (\alpha - \alpha_0) \frac{\partial \eta}{\partial \alpha} \Big|_{\alpha=\alpha_0} \\ &= \eta(\beta, \alpha_0) + (\alpha - \alpha_0) f(\mathbf{x}, \alpha_0). \end{aligned} \quad (6.9)$$

From this it is possible to construct an optimal design. The extension of this to models with  $p$  nonlinear parameters is straightforward, where now the linearised equation is given

$$\eta(\beta, \alpha) + \gamma f^T(\mathbf{x}, \alpha), \quad (6.10)$$

where

$$\gamma = (\alpha^1 - \alpha_0^1, \dots, \alpha^p - \alpha_0^p)^T \quad (6.11)$$

and

$$f^T(\mathbf{x}, \alpha) = \left( \frac{\partial \eta}{\partial \alpha^1} \Big|_{\alpha=\alpha_0}, \dots, \frac{\partial \eta}{\partial \alpha^p} \Big|_{\alpha=\alpha_0} \right). \quad (6.12)$$

Taking point priors, as the prior information for  $\alpha$ , will be an effective strategy where the choice of prior is well informed by subject matter knowledge. However, in general, as the purpose of the design is to allow the estimation of the model parameters, dependence on  $\alpha$  is unfortunate. It is possible to address this problem using sequential designs, identifying an initial design from which more information on  $\alpha$  can be obtained, in light of which more points can be added to the design. This process can be undertaken multiple times, as far as the experimental resources allow.

Alternatively, prior distributions  $p(\theta)$  can be chosen for the model parameters rather than point priors. While D-optimum designs, in the presence of point priors  $\alpha_0$ , maximise  $\log |M(\xi, \alpha_0)|$ , where a prior distribution is available for  $\alpha$ , D-optimum designs can be shown to maximise

$$E_\theta \log |M(\xi, \theta)| = \int_\theta \log |M(\xi, \theta)| p(\theta) d\theta. \quad (6.13)$$

There then exists an analogous equivalence theorem for the variances,

$$E_\theta d(\mathbf{x}, \xi, \theta) = \int_\theta d(\mathbf{x}, \xi, \theta) p(\theta) d\theta, \quad (6.14)$$

meaning a relationship between G- and D- optimality continues to apply.

## 6.4 The Fedorov exchange

As stated above, an exact D-optimum design,  $\xi_i$ , maximises the determinant of the information matrix. This design is identified from points across the design region,  $\chi$ . While in simple cases analytic solutions may be possible, usually identifying the optimal design is achieved using numerical methods. This is particularly the case in irregular shaped, constrained regions common in mixture experiments. The Fedorov exchange is one such method.

In the Fedorov Exchange an initial design  $\xi_0$ , of size  $N$ , is subjected to iterative improvement. This is done either by adding a point  $x_l$  to the design, from a list of alternative candidate points, to achieve a design of size  $N + 1$ , removing a point  $x_k$ , to achieve a design of size  $N - 1$  or exchanging a point  $x_k$  currently in the design, for a point  $x_l$  in the list of candidate points. The change providing the most improvement in the determinant of the information matrix is implemented. This should identify the locally optimal design with respect to the chosen criteria.

The value of the determinant of the information matrix of design  $\xi_{i+1}$ , when a point  $x_l$  is added to design  $\xi_i$  is

$$|M(\xi_{i+1})| = (1 + d(x_l, \xi_i)) |M(\xi_i)|, \quad (6.15)$$

when a point  $x_k$  is removed from design  $\xi_i$  is

$$|M(\xi_{i+1})| = (1 - d(x_k, \xi_i)) |M(\xi_i)| \quad (6.16)$$

and when a point  $x_k$  is replaced by a point  $x_l$ ,

$$|M(\xi_{i+1})| = [(1 - d(x_k, \xi_i))(1 + d(x_l, \xi_i)) + d^2(x_k, x_l, \xi_i)] |M(\xi_i)|, \quad (6.17)$$

where

$$d(x_l, x_k, \xi_i) = f_l^T M^{-1}(\xi_i) f_k. \quad (6.18)$$

Typically, it is difficult to obtain an optimal design of size  $N$  from one of size  $N + 1$  or  $N - 1$ . For this reason, experimental designs are usually identified for a given  $N$ . This has been done for all designs given in the following chapter, although there is opportunity to compare designs for different  $N$ .

Even simplifying the search for a optimal design by fixing  $N$ , it is still possible not to achieve a globally optimal design. In order to increase the chance of this happening the choice of the initial design,  $\xi_0$ , is important. For this reason, in the cases of the examples given, random sampling has been applied to obtain multiple initial designs, for each of which the determinant of the information matrix has been calculated, in order to choose the one design currently closest to maximising this value.

In order to decrease the length of time taken to find a design in the examples of the next chapter, the initial designs are calculated for a coarse grid of values, that is the support points are chosen from a relatively small set (mesh) of possible

mixtures. The Fedorov exchange is then applied to find the D-optimal design of support points chosen from this mesh. Improvements to the design are then made by making small changes, swapping the chosen design points with others on an increasingly fine mesh. Implicit in this approach is the expectation that the final design will be obtained by only shifting a small amount from the design identified from the original coarse grid.

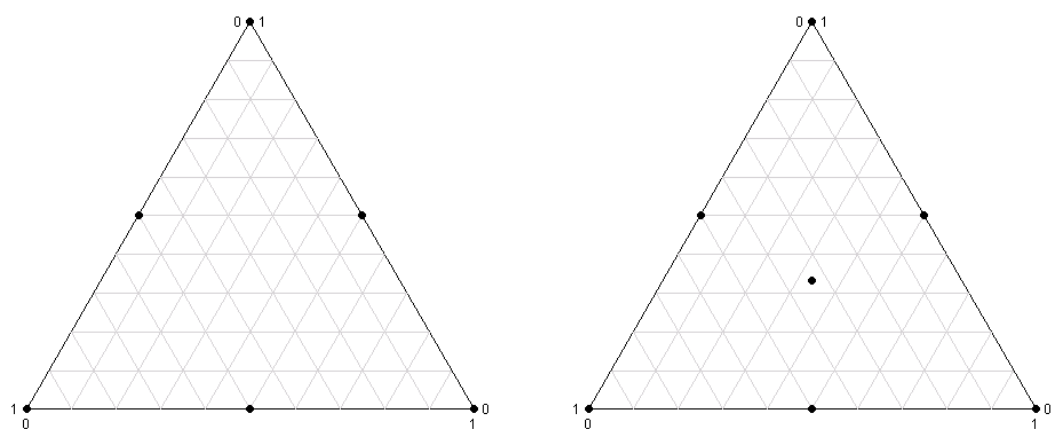
## 6.5 D-optimum design for the Scheffé polynomials

Before progressing to discuss some designs identified for fitting the GBMM, it will be useful to look at design in relation to the Scheffé polynomials. This will allow some comparison to be made later. Scheffé chose his designs to satisfy a particular motivation, the equal spread of data across the design region,  $\chi$ . Hence, the  $\{q, m\}$  simplex lattice designs place support points where the components take the  $m$  equally spaced values  $\{1, \frac{1}{m}, \dots, \frac{m-1}{m}, 1\}$ .

Scheffé's failure to consider optimality criteria in proposing his designs was of particular significance to Kiefer (1959, 1961). Kiefer explored D-optimal designs for the models of Scheffé's 1958 paper. He found the  $\{q, 2\}$  simplex lattice design in Figure 6.1a and the design in Figure 6.1b, proposed for the special cubic model, to be D-optimal for their respective models. These designs shall be useful to refer to in relation to the designs proposed in the next chapter.

As has been discussed, the quadratic and special cubic Scheffé polynomials, are special cases of the GBMM and the approach taken to building experimental designs for the GBMM has been to exploit this relationship. Therefore, what is presented in the next chapter represents a new method of building designs for mixture experiments in general, not just when applying the GBMM.



(a)  $q, 2$  simplex lattice design

(b) Scheffé special cubic design

Figure 6.1: Designs for the Scheffé polynomials

# Chapter 7

## Designs for the GBMM

The following section proposes several designs for fitting the GBMM. All examples look at experiments where there are 3 components and each example contains more than one design with the objective of advancing the discussion with respect to a particular feature. In line with the theory of optimal designs for nonlinear models, creating designs for the GBMM requires assumptions to be made about the likely values of the regressor defining parameters in each term; prior information is required.

In order to build a broadly applicable logic for the choice of prior values, the relationship between the GBMM and those models already commonly applied has been exploited. As mentioned previously, both the H2 and H3 Becker models and Scheffé quadratic and special cubic models are special cases of the GBMM. For example, the quadratic term of the Scheffé polynomials is the term of binary joint effect of the GBMM, where  $h_{ij} = 0.5$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$ . When this is noted, applying these models is in effect applying the GBMM with strict assumptions fixing the values of the regressor defining parameters.

Where subject matter knowledge does not dictate the priors in some other manner, it is proposed to use these inherent values to define the prior information for the regressor defining parameters. For example, in Example D1 below, a design for the GQBMM is constructed where the point priors taken are  $h_{ij} = 0.5$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$ , for all  $i$  and  $j$ . The consequent D-optimal designs can be viewed to possess an enhanced D-optimality for the 3 component quadratic Scheffé polynomial. It is assumed that the terms fitted to the GQBMM are likely to be of the form of the Scheffé quadratic terms, and the design best serves fitting that model. However, the flexibility of the GBMM allows alternatives also to

be fit, other than the Scheffé polynomial, where different values for the regressor defining parameters would allow better fit. Thus, this enhanced D-optimal design can also be viewed to introduce robustness against model uncertainty into the choice of a D-optimal design for the Scheffé quadratic polynomial.

The examples of this section are separated to address particular issues. There are five sections each containing more than one design. Example D1 contrasts the D-optimal saturated designs for the GQBMM and GSCBMM respectively. Saturated designs have the same number of design points as parameters to be estimated. Example D2, gives several non-saturated D-optimal designs for the GQBMM with the same priors as in Example D1. Thus there can be comparison of how the number of observations changes the design. In each of these first two examples, point priors are taken, defined by their implicit values in the Scheffé polynomials. In Examples D3 and D4, the priors are dictated by the implicit values of the regressor defining parameters in both the Becker H2 model and the Scheffé quadratic polynomial. In these two examples, these two sets of priors can best be viewed as two sets of point priors. Example D3 looks at saturated designs, while Example D4 looks at non-saturated designs. Example D5 uses the implicit values of the regressor defining parameters in the Scheffé and Becker models as the end points for a uniform prior distribution for the parameters. This is a move away from use of point priors to prior distributions, thus meaning Example D5 allows some exploration of the use of Bayesian D-optimal designs, in the context of GBMM.

Information on the designs is provided through plots of the designs and of the generalised prediction variance. All designs are some sort of D-optimal design and therefore the determinant of the information matrix is provided for each of them. Additionally, statistics relating to G- and V- optimality are also given: the maximum generalised prediction variance value and the mean of the generalised prediction variance across the entire simplex, respectively. In addition, the median value is also given. The designs are given in full in Appendix A.

## 7.1 Example D1

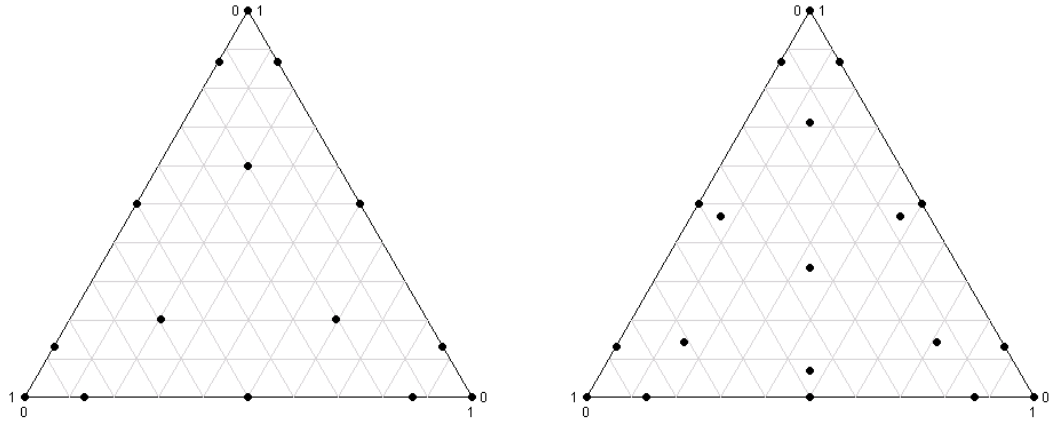
In this set of examples, two designs are given. Firstly, for the GQBMM,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \sum_{i \neq j}^3 \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{g_{ij} h_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{g_{ij}(1-h_{ij})} (x_i + x_j)^{s_{ij}}, \quad (7.1)$$

where the point priors  $h_{ij} = .5$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$  are the implicit values for these parameters in the Scheffé quadratic polynomial and secondly, for the GSCBMM,

$$E[y] = \sum_{i=1}^3 \beta_i x_i + \sum_{i \neq j}^3 \beta_{ij} \left( \frac{x_i}{x_i + x_j} \right)^{g_{ij} h_{ij}} \left( \frac{x_j}{x_i + x_j} \right)^{g_{ij}(1-h_{ij})} (x_i + x_j)^{s_{ij}} \\ + \beta_{123} x_1^{g_{123} h_{123}} x_2^{g_{123} h_{231}} x_3^{g_{123}(1-h_{123}-h_{231})}, \quad (7.2)$$

where the point priors for the additional parameters  $h_{123} = \frac{1}{3}$ ,  $h_{231} = \frac{1}{3}$  and  $g_{123} = 3$ , are the implicit values for these parameters in the Scheffé special cubic polynomial.



(a) 15 point design for GQBMM

(b) 19 point design for GSCBMM

Figure 7.1: Design plots for saturated locally D-optimum designs, for GQBMM and GSCBMM, with point priors equal to implicit values in the Scheffé polynomials

The two D-optimal designs, for the GQBMM and GSCBMM, are plotted in Figures 7.1a and 7.1b. The table of statistics relevant to D-optimality, G-optimality and V-optimality are given in Tables 7.1 and 7.2, respectively, while

the associated plots of the generalised prediction variance are given in Figures 7.2a and 7.2b.

$ F^T F $			
4.985e-22			
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	13.7640	13.5780	16.5150

Table 7.1: Statistics for locally D-optimum 15 point design 7.1a for GQBMM with priors equal to implicit values in the Scheffé quadratic polynomial

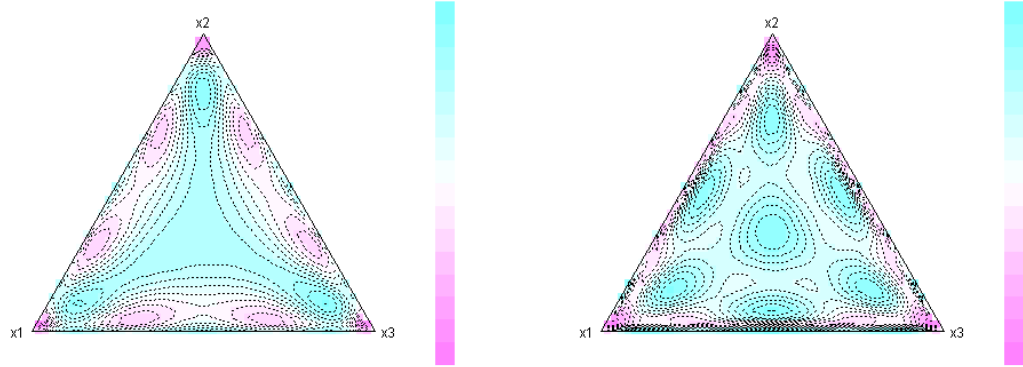
$ F^T F $			
1e-38			
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	16.0113	15.9106	18.9943

Table 7.2: Statistics for locally D-optimum 19 point design 7.1b for GSCBMM with priors equal to implicit values in Scheffé cubic polynomial

Allowing for a degree of imprecision regarding the position of the design points, we see that the design for the GSCBMM satisfies the GET, while that for the GQBMM does not. That a saturated design satisfies the GET in one case but not the other is intriguing, particularly given the saturated D-optimal designs for the Scheffé quadratic and special cubic models themselves satisfy the GET. The points of the D-optimal designs of the Scheffé quadratic and special cubic models are respectively subsets of the points of the designs given here for the GQBMM and GSCBMM.

## 7.2 Example D2

The previous section gave the opportunity to look at saturated designs for the 3 component GQBMM and GSCBMM, where point priors were taken at the values for the special case Scheffé quadratic and special cubic polynomials. The focus is now narrowed to only the GQBMM, using the same point priors, but now looking at D-optimal designs of 18, 19, 21, 24 design points. Those for 18, 21 and



(a) 15 point design 7.1a for GQBMM      (b) 19 point design 7.1b for GSCBMM

Figure 7.2: Prediction variance plots for saturated locally D-optimum designs, for GQBMM and GSCBMM, with point priors equal to implicit values in Scheffé polynomials

24 points, that is, for  $p + 3$ ,  $p + 6$  and  $p + 9$  points, respectively, permit common conclusions to be drawn from adding a point for each binary term of joint effect, that is adding three points at a time. The 19 point design with  $p + 4$  points gives a somewhat contrasting design.

The two saturated D-optimal 18 and 21 point designs are plotted in Figures 7.3a and 7.3b. The 24 point design is the same as the 21 point design except with repeated observations at the pure mixtures.

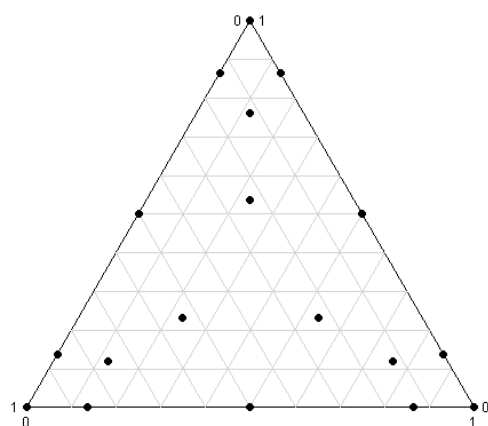
The statistics relevant to D-optimality, G-optimality and V-optimality are given in Table 7.3. This also includes those for the 15 point design for the GQBMM, given in the previous set of examples. The plots of the prediction variance for each design are given in Figure 7.4.

In the designs where points are added in multiples of 3, the design points are allocated symmetrically. In this context, this seems intuitively sensible; the terms of joint effect, for each pair of components, are subject to the same assumptions. However, it raises the question of whether, and if so how, this changes as the terms are subjected to different assumptions, that is, when point priors are not taken at values dictated by the Scheffé quadratic polynomial and when different priors are taken for different terms of joint effect.

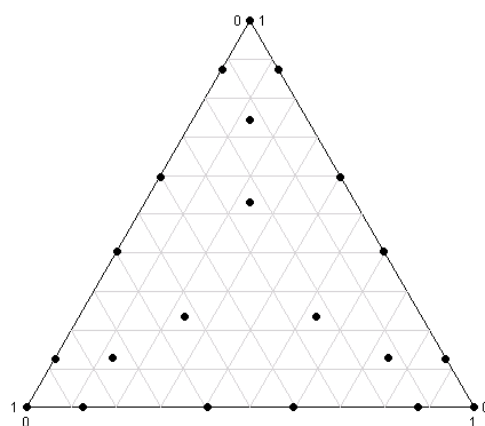
The 19 point design is shown in Figure 7.3c. This design is not symmetric

Saturated 15 point design			
$ F^T F $			
4.9850e-22			
$d(\mathbf{x}, \xi)$			
Median      Mean      Max			
13.7640      13.5780      16.5150			
18 point design			
$ F^T F $			
5.0771e-21			
$d(\mathbf{x}, \xi)$			
Median      Mean      Max			
11.8728      12.0042      17.4168			
19 point design			
$ F^T F $			
1.0140e-20			
$d(\mathbf{x}, \xi)$			
Median      Mean      Max			
11.9529      12.1163      18.3863			
21 point design			
$ F^T F $			
4.0437e-20			
$d(\mathbf{x}, \xi)$			
Median      Mean      Max			
12.3123      12.1842      17.5077			
24 point design			
$ F^T F $			
1.3393e-19			
$d(\mathbf{x}, \xi)$			
Median      Mean      Max			
13.9176      13.7256      20.1480			

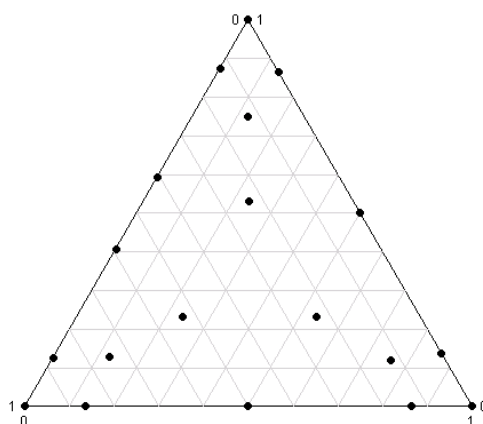
Table 7.3: Statistics for locally D-optimum designs for GQBMM with priors equal to implicit values in the Scheffé quadratic polynomial



(a) 18 point design



(b) 21 point design



(c) 19 point design

Figure 7.3: Design plots for non-saturated locally D-optimum designs, for GQBMM, with point priors equal to implicit values in the Scheffé quadratic polynomial

about the centre of the simplex which, in this particular situation, can be attributed to not having added points in multiples of 3. The design is a subset of the 21 point design and can clearly be viewed as a compromise between the 18 point design and the 21 point design.

Increasing the number of points in the design appears to be advantageous with respect to D-optimality, but disadvantageous with respect to the other two



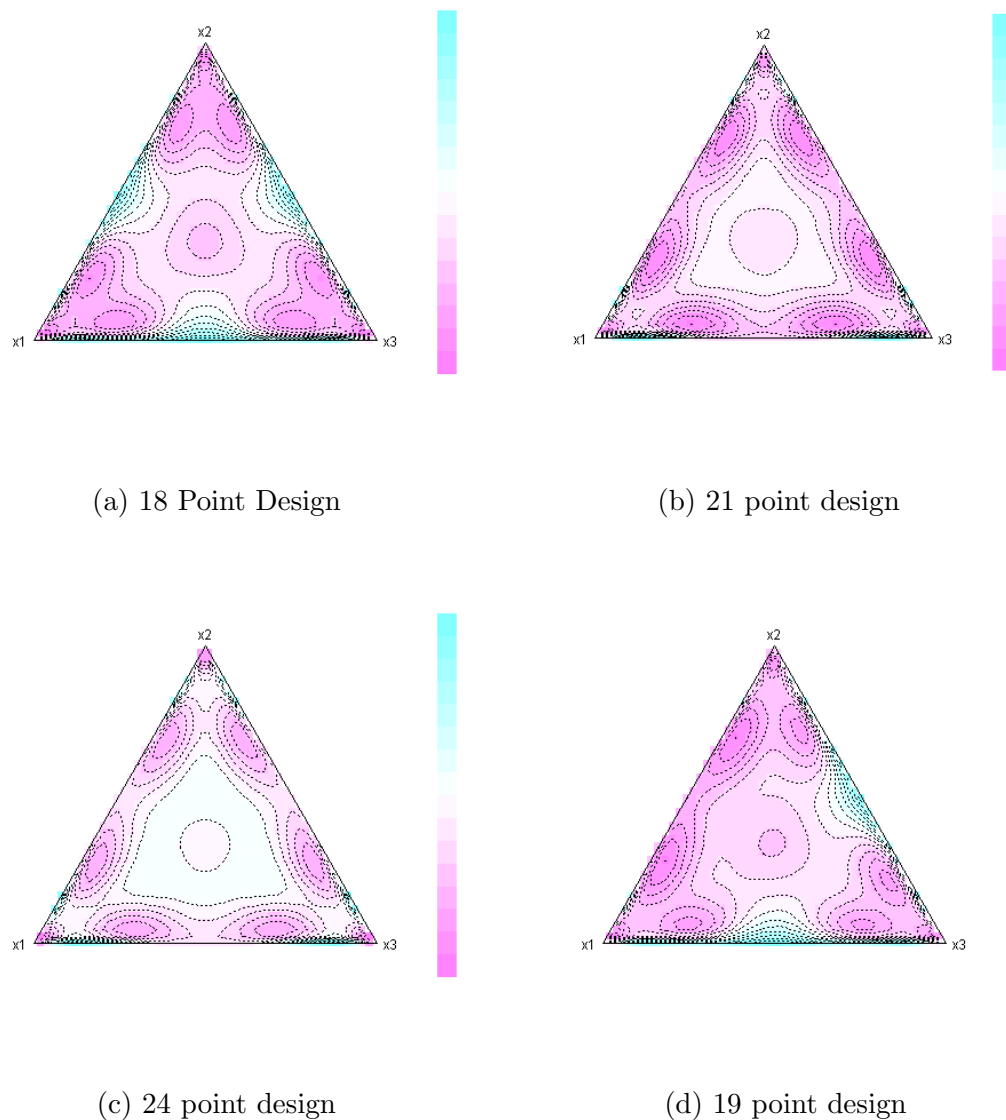


Figure 7.4: Prediction variance plots for non-saturated locally D-optimum designs, for GQBMM, with point priors equal to implicit values in the Scheffé quadratic polynomial

criteria. Adding 3 points appears to provide improvement with respect to V-optimality, but in general, adding points increases the average prediction variance over the design region. The asymmetry of the design with 4 extra points appears particularly influential with respect to G-optimality. However, this does not appear to be true with V-optimality, where it can be assumed the increase in the maximum prediction variance in one part of the design region occurs in

conjunction with a reduction in the prediction variance in another. Given the saturated design for the GSCBMM, found in the previous section, satisfied the GET, it is interesting that adding more points appears to improve the design for the GQBMM with respect to D-optimality. More needs to be done to look at designs for mixture experiments satisfying the GET.

### 7.3 Example D3

In the previous section, the designs given for the GQBMM differed from that given in the first section in the number of design points. In this section, the focus is instead on how changing the priors taken for the regressor defining parameters effects the design. Therefore, the designs are saturated, as for the design for the GQBMM in Example D1.

For simplicity, in this case the priors now take two sets of values, each of which is of particular interest. The first as before, is those taken by the regressor defining parameters in the special case of the Scheffé quadratic polynomial, while the second is the values taken by the regressor defining parameters in the Becker H2 model, where  $s_{ij} = 1$  for all  $i, j$ , rather than for the Scheffé quadratic polynomial, where  $s_{ij} = 2$ .

The 15 point design for the GQBMM, where the priors are dictated only by the values taken by the regressor defining parameters in the Becker H2 model, is plotted in Figure 7.5a and a plot of the prediction variance is given in Figure 7.5b. In contrast to the 15 point design for the GQBMM, given in Example D1, this design is not symmetric, nor does it place as many points on the edge of the design space. These contrasts are interesting regarding the following design, which should attempt to compromise between the 15 point design, where the point priors are dictated by the Scheffé quadratic polynomial and the 15 point design, where the point priors are dictated by the Becker H2 model.

The 15 point design where both sets of values are taken as priors is given in Figure 7.6a and a plot of the prediction variance for this design is given in Figure 7.6b. The design appears similar to the design found where the point priors are dictated by the Scheffé quadratic polynomial, although there is a slight shift in the position of the full mixtures. To illustrate this, these designs are given side by side in Figure 7.7. In the previous section the designs tended towards symmetry where possible. The same can be thought to have happened here.

Looking at the statistics for the two designs introduced in this section, given in Table 7.4, it is seen the design where the priors are defined by the Becker H2 model is less efficient with respect to G-optimality. Once again the asymmetry of the design increases the maximum prediction variance in a particular region of the response surface. The compromise design appears far better in this respect.

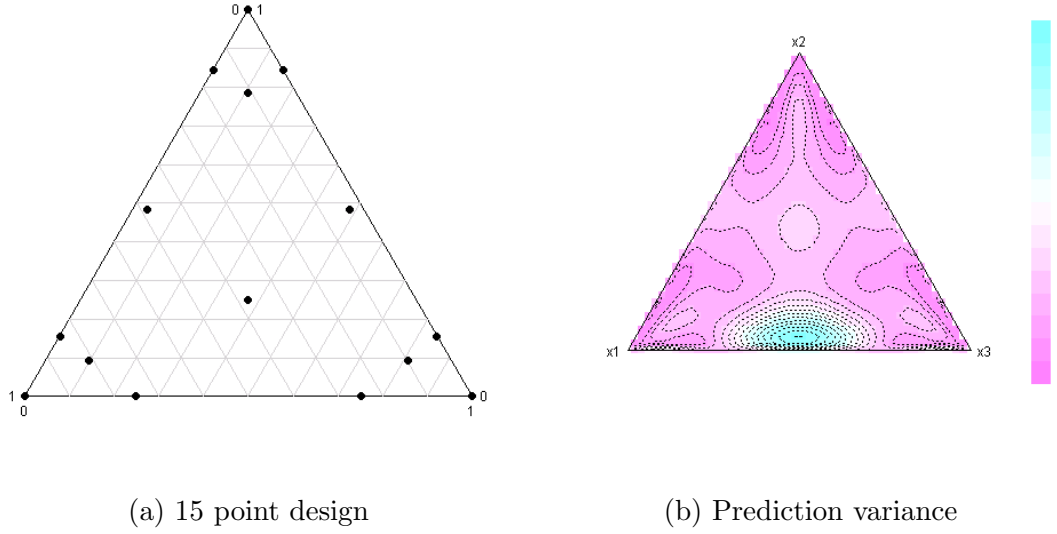


Figure 7.5: Design plot and prediction variance plot for 15 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model

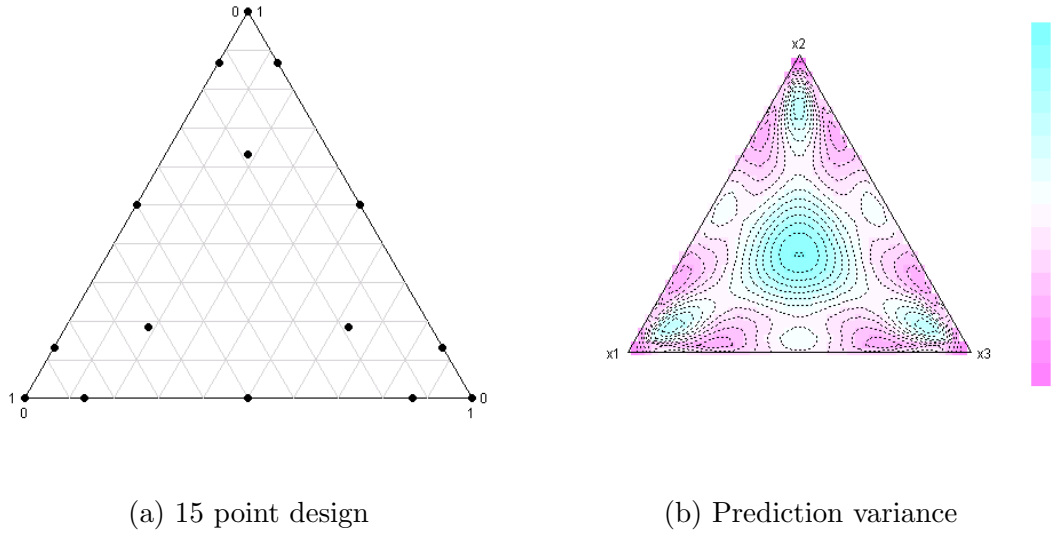
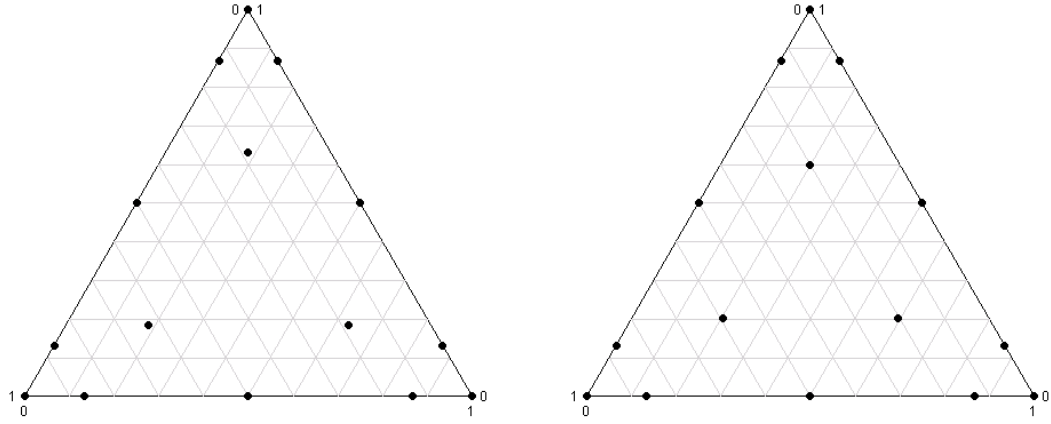


Figure 7.6: Design plot and prediction variance plot for 15 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model and Scheffé quadratic polynomial

## 7.4 Example D4

This pair of designs is very similar to those given in Example D3. However, here the designs are non-saturated: the designs contain 19 points, i.e.  $p + 4$ . The plot



(a) Design chosen with point priors equal to implicit values in the Scheffé quadratic polynomial (b) Design chosen with point priors equal to implicit values in the Becker H2 Model and Scheffé quadratic polynomial

Figure 7.7: Contrasting 15 point designs for GQBMM, chosen with different priors

of the 19 point D-optimum design, where the priors are determined solely by the Becker model, is given in Figure 7.8a and the plot of its prediction variance is given in Figure 7.8b. The equivalent pair of figures where the priors are the two sets of values dictated by the Becker H2 model and Scheffé quadratic polynomial are given in Figures 7.9a and 7.9b.

The most interesting aspect of these two designs is how the former contrasts with all the designs proposed for the GQBMM with priors dictated by the Scheffé quadratic polynomial, but then how the latter design does not, including points close to the centre of the design region in the manner of the designs for the Scheffé priors and then adding an additional point near the edge, in the manner of the design found for Becker priors. This suggests that with a greater number of points the designs will develop to provide effective estimation of models describing different effects. The GBMM, used in combination with these designs, could be particularly robust to model uncertainty.

Becker H2 priors			
$ F^T F $			
	2.6506e-18		
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	14.0565	14.7405	33.4800
Becker H2 and Scheffé quadratic priors			
$ F^T F $			
	2.2797e-18		
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	14.8125	14.9295	20.4300

Table 7.4: Statistics for contrasting locally D-optimum 15 point designs for GQBMM, chosen with different priors

Becker H2 priors			
$ F^T F $			
	1.1864e-16		
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	11.9168	12.1961	16.7333
Becker H2 and Scheffé quadratic priors			
$ F^T F $			
	8.4155e-17		
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	11.8769	12.1163	17.7555

Table 7.5: Characteristics for contrasting locally D-optimum 19 point designs for GQBMM, chosen with different priors

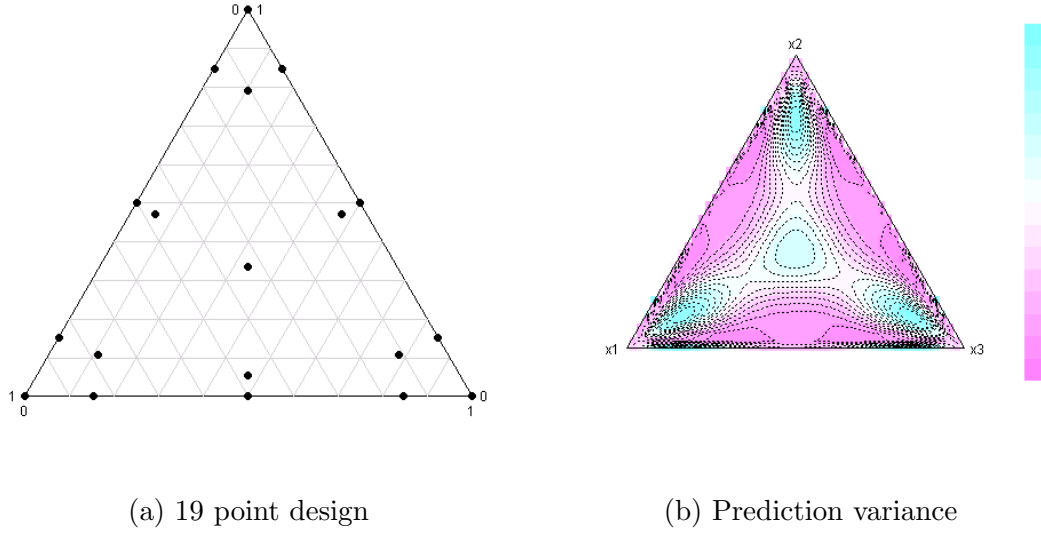


Figure 7.8: Design plot and prediction variance plot for 19 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model

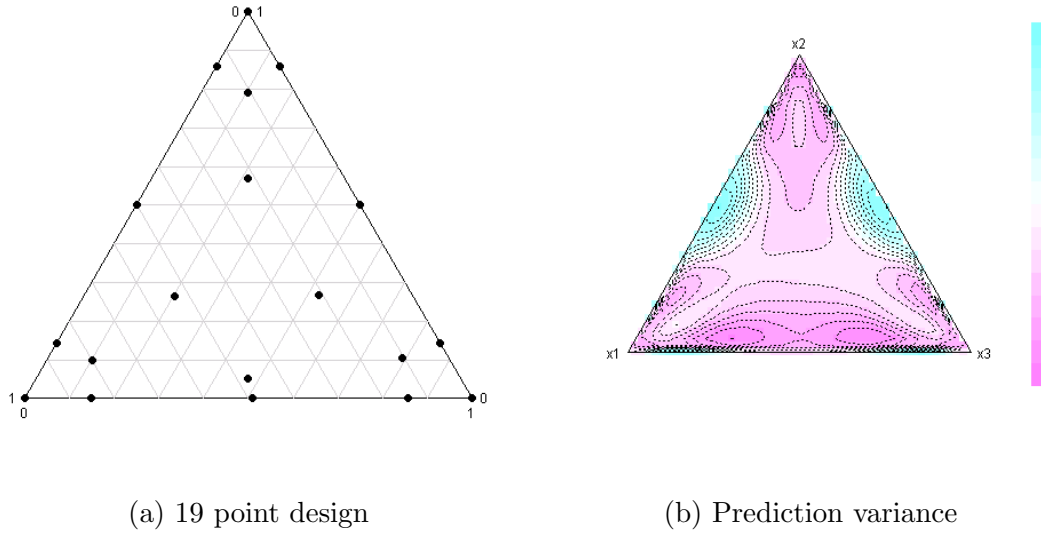


Figure 7.9: Design plot and prediction variance plot for 19 point locally D-optimum design, for GQBMM, with point priors equal to implicit values in the Becker H2 Model and Scheffé quadratic polynomial

## 7.5 Example D5

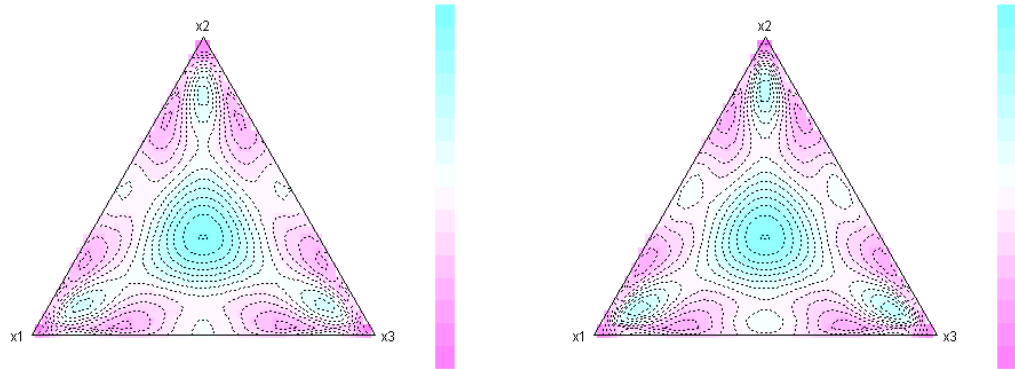
In this final set of designs, a continuous prior distribution is assumed for all  $s_{ij}$ . However, the  $s_{ij}$  have also been assumed to remain equal for all  $i$  and  $j$ . Allowing

for this constraint, the prior distribution,

$$p(s_{ij}) = \begin{cases} 1 & \text{if } 1 \leq s_{ij} \leq 2 \\ 0 & \text{otherwise} \end{cases}, \quad (7.3)$$

is a uniform distribution between the values  $s_{ij} = 1$  and  $s_{ij} = 2$ , that is between the values of  $s_{ij}$  defined by the Becker H2 model and Scheffé quadratic polynomial. To achieve the Bayesian D-optimal design, the mean value of the criteria is calculated over 11 equally spaced values across the distribution.

With respect to the degree of accuracy applied here the design is the same as that found for the average saturated design between the Becker H2 and Scheffé quadratic polynomials found in Example D3. Looking at the plots of the prediction variance for these two models there seems subtle differences, suggesting that maybe larger differences would be found when calculations are made with greater accuracy. The characteristics in Table 7.6 also suggest only subtle differences,



(a) Bayesian D-optimal design  $1 \leq s_{ij} \leq 2$       (b) Locally D-optimal design  $s_{ij} = 1, 2$

Figure 7.10: Plots of prediction variance for designs where priors are defined through Becker H2 Model and Scheffé quadratic polynomial

the larger value for the mean of the determinant of the information matrix attributable to calculating more values close to  $s_{ij} = 1$ . The design is less effective with respect to D-optimality when  $s_{ij}$  is close to 1. This is unsurprising given how different this design is in comparison to the first design given in Example D3.



Locally D-optimal design $s_{ij} = 1, 2$			
$ F^T F $			
2.2797e-18			
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	14.8125	14.9295	20.4300
Bayesian D-optimal design $1 \leq s_{ij} \leq 2$			
$ F^T F $			
4.0724e-19			
$d(\mathbf{x}, \xi)$			
	Median	Mean	Max
	14.4660	14.6550	20.6250

Table 7.6: Characteristics for contrasting D-optimum 15 point designs for GQBMM

## 7.6 Conclusions

The set of designs presented might all be considered for use when constructing an experiment to which it is intended to fit a GBMM. Collectively, they also suggest an idea of how to go about building further designs for the GBMM dependent on prior knowledge of the values of the regressor defining parameters. Generally this may come from subject matter knowledge allowing an understanding of how the joint effects of the components impact upon the response or from a previous study. However, alternatively the prior assumptions regarding the values of the regressor defining parameters may be based upon the inherent assumptions which have always been present when practitioners have fitted the more established models to mixture data. In this way, the designs presented can be viewed as the starting point for a new, more robust method of analysis to be introduced.

As was mentioned in Example D1, the designs for the GBMM where the point priors of the regressor defining parameters are equal to the implicit values of these parameters in the Scheffé quadratic and special cubic polynomials, includes the D-optimal designs for these models as a subset of the design points. Moreover, the designs for the GBMM also include additional points chosen in a manner which favours good estimation of all parameters of the Scheffé quadratic and special cubic models. Hence the designs for the GBMM are well chosen to fit these models.

Similar designs, but with more design points, were presented in Example D2. For designs of up to 21 points, rather than 15, no repetition of points were included. New points were added to the 15 point design of Example D1, further improving the coverage of the design space in a manner that lends itself to fitting the Scheffé quadratic polynomial.

In these cases, the selection of the prior values of the regressor defining parameters explicitly suggests that the final models are likely to take the form of the Scheffé polynomials. However, because the GBMM is a partially linear model, there is an implicit acceptance of model uncertainty. The designs for the GBMM address both the likely model forms, underlying the choice of design, as well as additional model uncertainty. Hence it is possible to construct a model building strategy which uses these designs to explore first, how well the most likely model forms fit the data and secondly, what alternative model could be used should these models prove insufficiently effectual.

One particular manner through which this issue of addressing model uncertainty, through the design, has been looked at before, is through the choice between the Scheffé quadratic polynomial and Becker's models. This is alluded to in a broad discussion by Becker (1978) and more specifically by Cornell and Gorman (1978), although without complete resolution of the problem. Here, by introduction of the regressor defining parameters  $s_{ij}$ , it is possible to propose designs, driven by ideas of design optimality, specifically capable of choosing between these models. In fact, the designs of Examples D3-D5 address the choice between the 3 component quadratic Scheffé polynomial and the Becker H2 model. However, this is done with the additional consideration that  $s_{ij}$  may be estimated such as to choose a term which is found in neither of these models. This is a novelty. Additional work of optimal designs for Becker's models has been done by Liu and Neudecker (1997).

In this way the designs for the GBMM can be chosen to possess an enhanced optimality for the more traditional mixture models, such as the Scheffé polynomials and Becker's models. These designs are robust against the uncertainty which the GBMM reveals about these model forms, will identify if these models are inadequate, allow for a choice between the established model forms and permit other model forms to be fit if necessary. In this manner, the designs here can be viewed to have application beyond just the fitting of the GBMM, and present an approach to undertaking mixture experiments far more open to the wealth of

joint effects the GBMM allow us to explore. In the following chapter, an example is presented which begins to explore how such designs could be applied as part of a multi-stage sequential experiment.

# Chapter 8

## Examples

This section provides applications of GBMM, both to simulated data and data from real industrial experiments. Across three sections, four examples are presented, where each section intends to address a particular issue. Section 8.1 provides examples of behaviour which may be easily described by a GBMM, but which other, more established models can find difficult to describe. This is done using two examples of simulated data, in two and three components, respectively. Section 8.2 presents an example, in three components, from an industrial experiment, designed and then executed by Federal Mogul to serve joint aims with this project. This applies one of the designs presented in the previous chapter with the intention of fitting the Scheffé quadratic polynomial. The purpose of this example is to examine how designs found for the GBMM, can be used to reveal whether, and if so where, established models inadequately describe the response surface. This leads to new ideas for building sequential designs for mixture experiments. Section 8.3 presents a final example, again from an industrial experiment. In this case, the example is in eight components and the intention is to demonstrate that GBMM can be fit to cases where there are more than three components and that the required model is of high complexity. For each example it is specified how the fitting procedure was applied, with reference to the variations discussed in Chapters 4 and 5.

### 8.1 What can the GBMM describe?

This section contains two examples. Both are for simulated data, in two and three components, respectively. The intention is to demonstrate that there are

types of behaviour which the GBMM can easily describe, but which other models cannot, or at least, cannot without a large number of model terms. For each model fit,  $AIC_c$  and  $R^2$  values are provided, in addition to plots comparing the predicted response surface to the underlying behaviour.

### 8.1.1 Two component example

This example uses a data set of 11 points, spaced uniformly across a 1 dimensional simplex. Hence they occur at mixtures where  $x_1 = (0, 0.1, 0.2, \dots, 0.9, 1)$  and  $x_2 = 1 - x_1$ . At these points a response is simulated for

$$y = 5x_1 + 2.5x_2 + 500000x_1^6x_2^{10}, \quad (8.1)$$

which in the terminology introduced in Chapter 3, describes an asymmetric joint effect, more concentrated than that of the quadratic terms of the Scheffé quadratic polynomial. Random variation is added such that

$$\epsilon_i \sim N\left(0, \frac{\bar{y}}{10}\right). \quad (8.2)$$

To this data set 7 models are fitted: the two component quadratic, cubic and quartic Scheffé polynomials,

$$E[y] = \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2, \quad (8.3)$$

$$E[y] = \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{12}x_1x_2(x_1 - x_2) \quad (8.4)$$

and

$$E[y] = \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{12}x_1x_2(x_1 - x_2) + \beta_{12}x_1x_2(x_1 - x_2)^2; \quad (8.5)$$

the two component quadratic, cubic and quartic models with additional inverse terms

$$\beta_{-1} \frac{1}{x_1 + 0.1} \quad (8.6)$$

and

$$\beta_{-2} \frac{1}{x_2 + 0.1} \quad (8.7)$$

and a GQBMM.

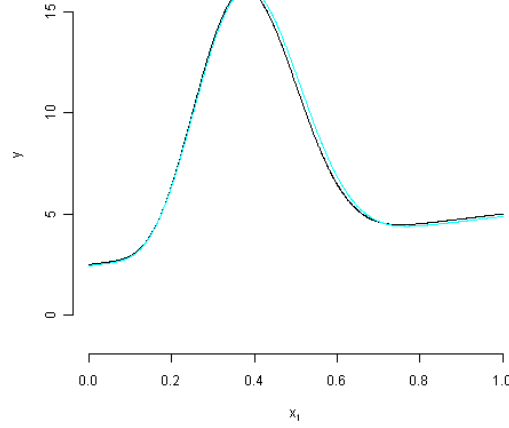


Figure 8.1: Fitted response surface for GQBMM v. Underlying response surface

Figure 8.2 plots the predicted response for each of the first six models (excluding the GQBMM) over the underlying response surface, while Figure 8.1 does the same for the GQBMM. From Tables 8.1 and 8.2 it can be seen that, other than the GQBMM, only the Scheffé cubic and quartic models with additional inverse terms provide an effective fit to the data. However, the plots for these models, Figures 8.2e and 8.2f, reveal that this relatively good fit is achieved by describing a simple underlying response surface with one which is relatively complex. This is reflected in the statistics for model parsimony, for which these models compare poorly in comparison to the fitted GQBMM,

$$E[y] = 4.87x_1 + 2.44x_2 + 293414.34x_1^{5.753}x_2^{9.353}, \quad (8.8)$$

which matches the simulated response surface very closely. Here the GBMM was fitted without either Stage 2 or Stage 3 of the fitting procedure as there is only one possible pair of components for which to define a term of joint effect.

	Scheffé quadratic	Scheffé cubic	Scheffé quartic	GQBMM
$R^2$	0.5363	0.6979	0.8011	0.9814
$AIC_c$	66.67	67.20	69.94	31.30
No of Terms	3	4	5	3

Table 8.1: Statistics comparing GBMM to Scheffé polynomials for simulated two component example

	Scheffé quadratic	Scheffé cubic	Scheffé quartic	GQBMM
$R^2$	0.7300	0.9016	0.9039	0.9814
$AIC_c$	73.30	73.20	91.27	31.30
No of Terms	5	6	7	3

Table 8.2: Statistics comparing GBMM to Scheffé polynomials with additional inverse terms for simulated two component example

### 8.1.2 Three component example

In contrast to the previous example, here the data have been simulated for a three component experiment. The design is as indicated in Figure 8.3d and the underlying response surface is described by the model

$$E[y] = 3x_1 + 4x_2 + 5x_3 + 500000 \frac{x_1^6 x_2^{10}}{(x_1 + x_2)^{12}} + 100x_1^{0.5} x_2^{0.5} x_3^{2.5}, \quad (8.9)$$

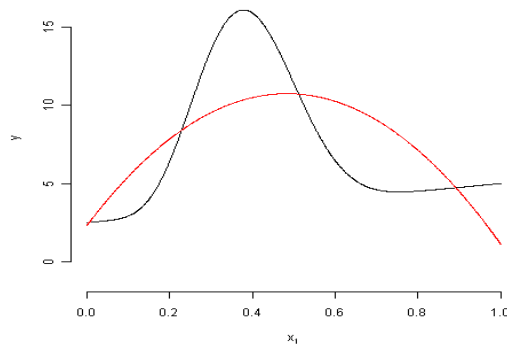
to which has been added random variation, as in the two component example. Therefore, this model includes the same effect as investigated in the two component example and an additional three component joint effect; again asymmetric and somewhat more concentrated than the cubic term of ternary joint effect found in the cubic Scheffé polynomial. To this has been fit a GSCBMM, using all three stages of the fitting procedure, and full quadratic, cubic and quartic Scheffé polynomials. The  $R^2$  and  $AIC_c$  values for these four models are given in Table 8.3. Also given, in Figures 8.3a, 8.3b and 8.3c, respectively, are the plots of the un-

	Scheffé quadratic	Scheffé cubic	Scheffé quartic	GSCBMM
$R^2$	0.4275951	0.7068914	0.9491928	0.9882141
$AIC_c$	93.54475	101.2201	94.16494	4.269948
No of Terms	6	10	16	5

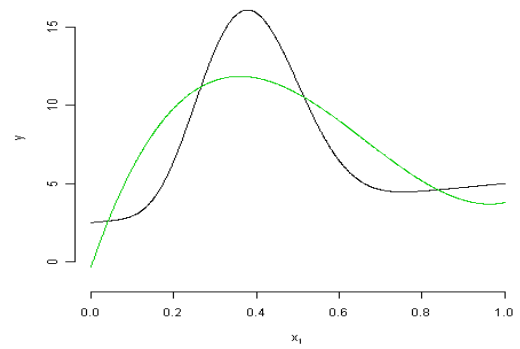
Table 8.3: Statistics comparing GBMM to Scheffé polynomials for simulated three component example

derlying response surface, as described by Model 8.9 (without additional random variation), the response surface described by the fitted GSCBMM,

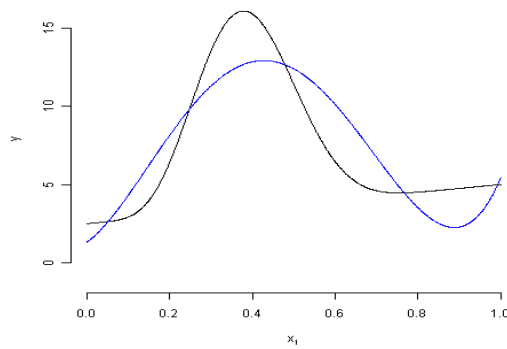
$$E[y] = 3.021x_1 + 4.002x_2 - 0.1172x_3 - 1.275e + 07 \frac{x_1^{10.23} x_2^{14.78}}{(x_1 + x_2)^{19.65}} + 94.09x_1^{0.3299} x_2^{0.5712} x_3^{2.5}, \quad (8.10)$$



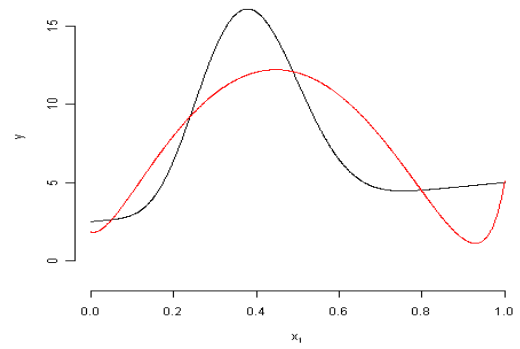
(a) Scheffé quadratic polynomial



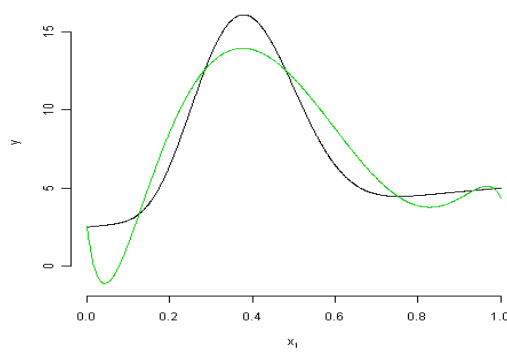
(b) Scheffé cubic polynomial



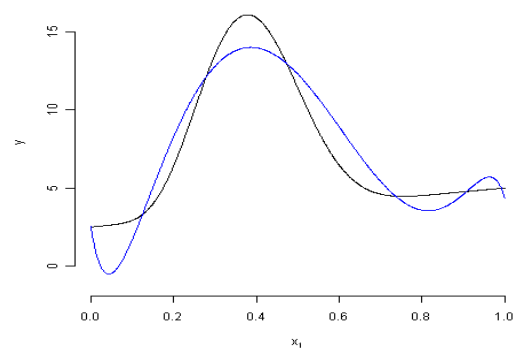
(c) Scheffé quartic polynomial



(d) Scheffé quadratic polynomial with inverse terms



(e) Scheffé cubic polynomial with inverse terms

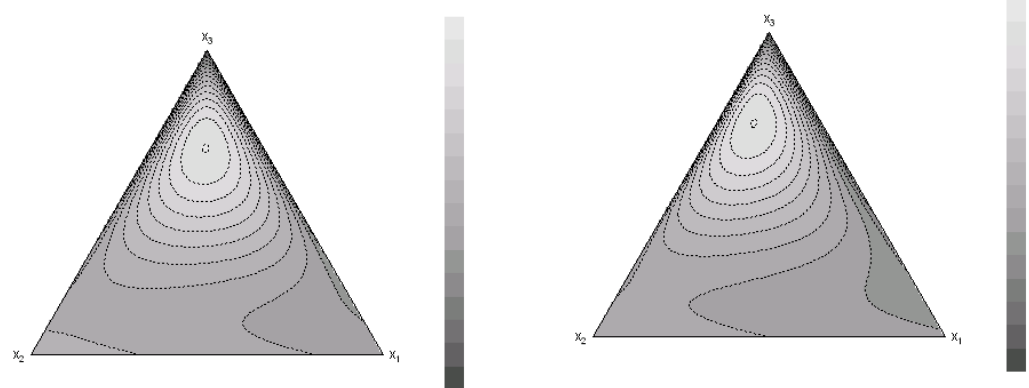


(f) Scheffé quartic polynomial with inverse terms

Figure 8.2: Fitted response surface for alternative models v. underlying response surface for simulated two component example

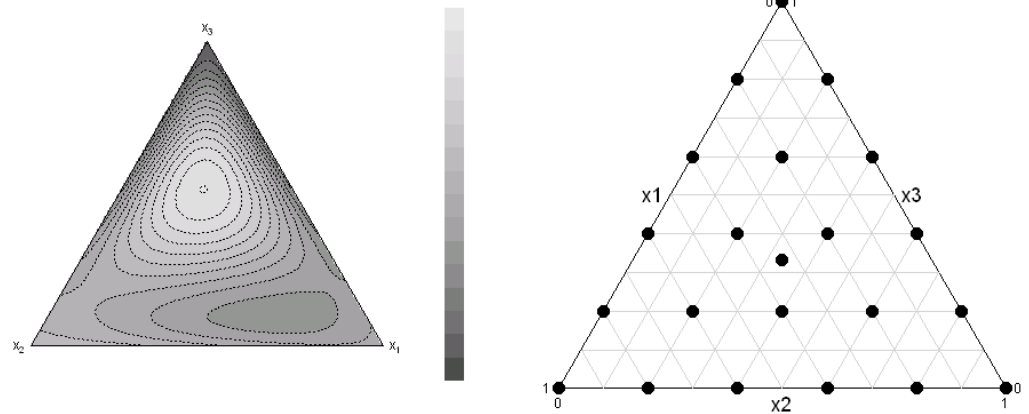


and the fitted quartic Scheffé polynomial. These demonstrate the fitted GSCBMM matches the underlying response relatively strongly in comparison to the quartic Scheffé polynomial, the only alternative considered which can be thought to provide a reasonable fit.



(a) Underlying surface

(b) Response surface predicted by fitted GSCBMM



(c) Response surface predicted by fitted Scheffé quartic polynomial

(d) Design for simulated 3 component example

Figure 8.3: Plots relating to simulated 3 component example

### 8.1.3 Discussion

There is a bias toward the GBMM in these examples; as the underlying response surface is simulated using a GBMM it is likely that a GBMM will prove to describe the surface well. However, the focus here is on the inability of the established models to describe such surfaces. Even when using models with a large number of varied terms, none of the models given above provide effective representation of the underlying response. The problem is particularly felt with the quadratic and cubic Scheffé polynomials, which are also the only models which have been frequently applied to mixture experiments elsewhere. In almost all other cases, mixture experiments are represented through the behaviour described by the terms of the quadratic and cubic Scheffé polynomials. Yet these terms appear inadequately equipped to describe surfaces which we would consider reasonable to expect in mixture experiments.

## 8.2 Using a design for the GBMM: an example from industry

The purpose of this example is somewhat different to those given in the previous section. Here the focus is upon the relationship between the GQBMM and the Scheffé quadratic polynomial and how this may be used to propose new ways of undertaking mixture experiments. As previously discussed, the Scheffé quadratic polynomial can be viewed as a special case of the GQBMM, where  $h_{ij} = 0.5$ ,  $g_{ij} = 2$  and  $s_{ij} = 2$ . For a three component experiment, assuming these values as point priors, Design 7.1a is produced. This design can be viewed as the D-optimal design of the Scheffé quadratic polynomial with 9 additional check points, chosen in a manner which will specifically challenge the inherent assumptions which exist when choosing to model a mixture experiment with the Scheffé quadratic polynomial.

The data comes from an industrial mixture experiment, that took place at Federal Mogul, on formulations for motorcycle disc brake pads. These products are blends of metal powders with fine powder abrasives and graphite. The mixes are compacted under high pressure at ambient temperature and then sintered at high temperatures in an inert atmosphere to form individual disc brake pads. This experiment explores the compaction properties of the mixture, during the

manufacturing process.

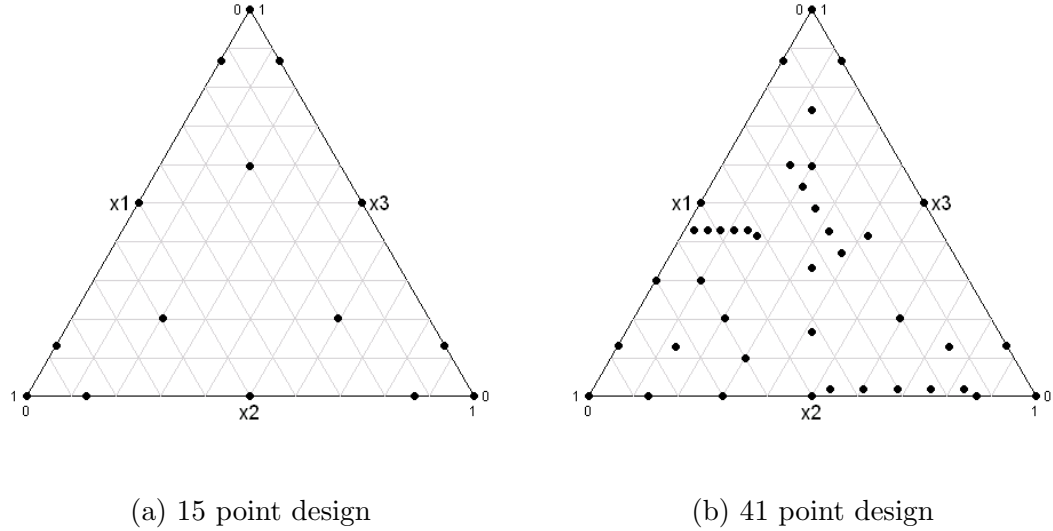


Figure 8.4: Designs for brake compaction data

The data consists of 41 observations, of which 15 were those of Design 7.1a. The full design is shown in Figure 8.4b, alongside the plot of the 15 point design in Figure 8.4a, originally seen in the previous chapter. Initially, the full Scheffé quadratic polynomial,

$$E[y] = 50.58x_1 + 26.41x_2 + 15.55x_3 + 25.19x_1x_2 + 25.05x_1x_3 - 20.31x_2x_3, \quad (8.11)$$

and GQBMM,

$$\begin{aligned} E[y] = & 49.12x_1 + 26.31x_2 + 16.61x_3 + 1801.43 \frac{x_1^{5.281}x_3^{2.474}}{(x_1 + x_3)^{7.754}} \\ & - 36.65 \frac{x_2^{1.439}x_3^{1.142}}{(x_2 + x_3)^{2.582}} + 19.58 \frac{x_1^{0.8515}x_2^{0.9150}}{(x_1 + x_2)^{-2.424}}, \end{aligned} \quad (8.12)$$

were fit to the latter 15 points, returning  $R^2$  values of 0.9671 and 0.9972, at these points, and 0.9386 and 0.9508 for observations across the entire data set. There is little difference in the fit of these models. However, looking at the plots of the response surfaces they describe, in Figures 8.5a and 8.5b respectively, there are evidently implications from choosing between these models; the fitted GQBMM identifies curvature in the response surface where there is no such behaviour described by the Scheffé quadratic polynomial. A logical experimental strategy

would be to, subsequently, take observations which particularly focus on whether it is correct to expect such behaviour.

Without reference to the regressor defining parameters this could be done by spreading the additional points randomly or uniformly over the region of interest, whether that is the whole of the initial experimental region or some part of it. This is a strategy which could also be taken with higher order models indicating curvature where the Scheffé quadratic polynomial did not. However, just as the regressor defining parameters were used to define the initial design, their newly estimated values can also be used to define the subsequent additional design points. In this context, the GBMM may be viewed as a tool through which a more systematic exploration of the experimental region can take place, whether or not a GBMM should, ultimately, be used for prediction. Hence, if necessary, experimental resources can be focussed upon those regions of interest where there is expected to be curvature which deviates from what would be described by the established models.

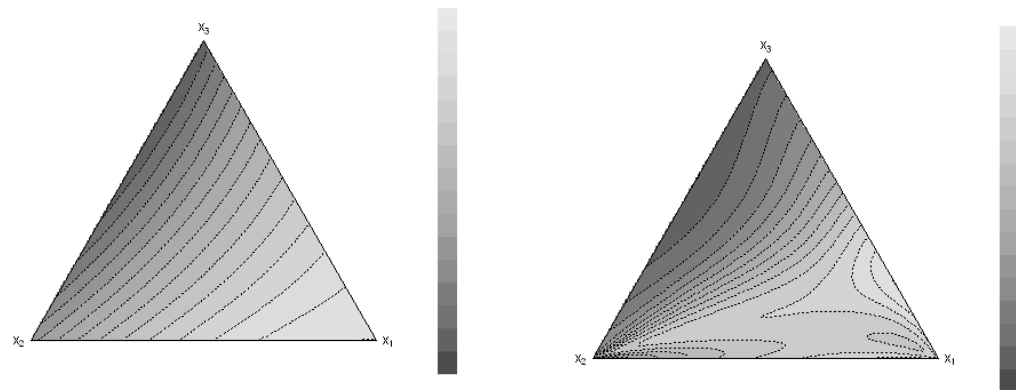
The final GBMM, fitted to the full 41 points is

$$E[y] = 49.12x_1 + 26.31x_2 + 16.61x_3 + 1801.43 \frac{x_1^{4.259} x_3^{2.574}}{(x_1 + x_3)^{6.44}} - 36.65 \frac{x_2^{5.233} x_3^{11.03}}{(x_2 + x_3)^{11.61}} + 19.58 \frac{x_1^{1.597} x_2^{2.687}}{(x_1 + x_2)^{0.6414}}. \quad (8.13)$$

This possesses an  $R^2$  value 0.9791 and the response surface it describes can be seen in Figure 8.5c. Assuming this later GQBMM to describe something closer to the true response surface, given the greater amount of data, it is evident that the initial GQBMM did identify a region in which it would have been informative to collect more data. Hence, in this case, the proposed design building methodology would have been worthwhile as a method of systematically allocating the additional experimental resources, although the curvature does not appear to be as dramatic as first indicated.

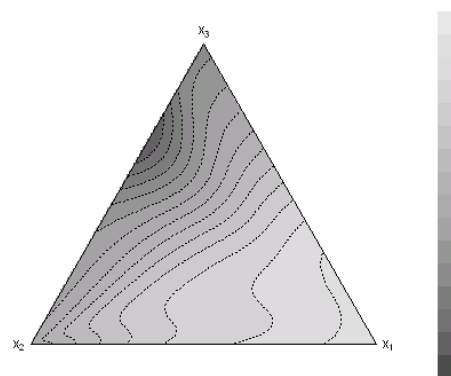
### 8.3 Fitting where there are more components: a second example from industry

Thus far, all examples presented have been in three components or less. However, the strength of the proposed fitting procedure for the GBMM is that it can be used



(a) Response surface predicted by fitted Scheffé quadratic polynomial to 15 data points

(b) Response surface predicted by fitted GQBMM to 15 data points



(c) Response surface predicted by fitted GQBMM to 41 data points

Figure 8.5: Predicted response surface for models fitted to brake compaction data

where there are far more components, in spite of the model selection requiring the values of much greater numbers of regressor defining parameters to be identified. This final example gives a model in 8 components and 9 terms.

The data, consisting of 101 observations, once again comes from an experiment on formulations for motorcycle disc brake pads. The aim of this experiment was to examine how the individual components behave to determine the overall flow of the mixture, quantified through the bulk flow energy response. The flow characteristics for each pure component and all blends were measured using a powder rheometer.

The GBMM was obtained with 9 terms describing binary joint effects. No terms of ternary joint effect were included in the model, although the full fitting procedure, looking for ternary terms, with both Stages 2 and 3, was applied. The estimates of the model parameters for the individual components are given in Table 8.4, while those for the terms of joint effect are given in Table 8.5.

Component	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
$\hat{\beta}_i$	976.1	-398.8	3478	6563	4677	2944	7819	7380
s.d. $(\hat{\beta}_i)$	316.9	367.8	361.7	390.2	400.6	324.9	346.0	412.5

Table 8.4: Parameter estimates for the linear terms of the GBMM fit to 8 component FM data

Components		$\hat{\beta}_{ij}$	s.d. $(\hat{\beta}_{ij})$	$r_{ij}$	$r_{ji}$	$s_{ij}$
$x_2$	$x_4$	-8.382e+10	1.397e+10	7.581	17.46	22.19
$x_3$	$x_5$	-2.445e+07	6.032e+06	5.945	7.370	-1.225
$x_2$	$x_7$	-5.741e+10	1.414e+10	0.0434	24.68	16.93
$x_4$	$x_6$	-4.534e+05	9.254e+04	5.749	1.614	-39.73
$x_2$	$x_8$	-1.055e+10	2.399e+09	9.914	12.21	21.04
$x_1$	$x_7$	7.057e+05	1.737e+05	9.313	1.541e-05	6.829
$x_5$	$x_9$	-5.039e+10	1.409e+10	15.27	9.278	-16.86
$x_3$	$x_8$	4.440e+10	1.123e+10	20.02	7.855	0.2484
$x_4$	$x_8$	2.974e+03	987.7	0.8299	0.3282	0.0914

Table 8.5: Parameter estimates for the terms of the joint effect of the GBMM fit to 8 component FM data

The GBMM was compared to the best reduced Scheffé quadratic, cubic and special cubic models and the PQM model, found using stepwise regression. The  $R^2$  and  $AIC_c$  values for these models are given in Table 8.6. The GBMM is best

Model	$AIC_c$	$R^2$	$p$
GBMM	803.0	0.9632	9
Scheffé quadratic	841.1	0.9285	10
PQM	836.1	0.9285	9
Scheffé special cubic	817.3	0.9758	15
Scheffé cubic	810.1	0.9895	19

Table 8.6: Comparison of models fit to 8 component FM data

with respect to the  $AIC_c$  criterion and its  $R^2$  value is only slightly lower than those for the Scheffé special cubic and cubic models, but the latter have 6 and 10 more parameters, respectively. In this case, the GBMM provides a better fit with fewer terms.

# Chapter 9

## Discussion

This final chapter shall be used to summarise the ideas discussed. It shall be separated into two parts. The first shall discuss what it is felt has been achieved by the work presented, making claims to novel contribution. The second shall look at how the work presented should be developed in the future.

### 9.1 What has been achieved?

The PhD, of which this work is the culmination, was funded by Federal Mogul (FM) Friction Materials. FM, among other activities, manufacture brake discs for many of the world's leading automobile manufacturers. As discussed in the examples, brake disc manufacture involves the formulation of mixtures.

When initiating the project, the objective was to explore the experimental region of mixture experiments more thoroughly, with respect to different types of blending behaviour. It was felt that the position of the Scheffé polynomials as the primary, or sometimes sole, recourse for practitioners led to a lack of enquiry regarding the type of blending behaviour that was used to describe a response. This could be detrimental to achieving the experimental objectives. More specifically, it was felt that there did not exist a method for building experimental designs which would systematically allow blending behaviour to be explored which was not associated generally with the established models and more specifically, with the Scheffé polynomials. As such, while the initial remit of the project was vague, were there to have been two stated aims, they would have been these:

1. Develop a model capable of describing a broad range of blending behaviour.



2. Develop experimental designs specifically directed towards exploring blending behaviour other than that given by the Scheffé polynomials.

The GBMM is a powerful tool allowing a very broad range of blending behaviours to be described. These include those of the Scheffé polynomials (and its reparameterisations) and Becker's models. There are potential benefits to be gained from the GBMM through fitting more parsimonious models and possibly, although it has not been explored here, interpreting the terms of the GBMM with respect to the mean and concentration effects described in Chapter 3. Through this class of models it's possible for a practitioner to reject the assumptions inherent in choosing to model with the Scheffé polynomials and instead adopt a more open approach, flexible to many different types of behaviour. It is foreseen that the approach presented here will prove only one manner in which this class of models can be used; a practitioner, acutely aware of the manner in which the components in their experiment influence the response, now has a tool more capable of both reflecting the subject knowledge they have been able to build up and of adapting to new knowledge they may acquire.

However, the flexibility of the GBMM is a challenge. Should a practitioner wish to exploit the full range of possible behaviours, a powerful method for model fitting is required. The method proposed here may be considered crude, essentially being a brute force approach. However, as a response to the challenge presented, it is effective; it astutely exploits the partially linear structure using the variable projection method and can at least be viewed as an effective response to the challenge presented. The novelty of taking the obvious step of applying stepwise methods to partially linear models is doubtful. However, as an answer to how to channel the flexibility of the GBMM, it is a useful and adroit response. Moreover, for other cases where partially linear models exist, in a large number of nonlinear terms, this method could be applied with little redevelopment; theoretically, it is a fitting procedure with application beyond the GBMM.

What novelty can be attributed to the designs presented here is drawn from developing designs for the new class of models. However, in taking this approach it has been possible to make proposals which allow designs specifically capable of exploring deviation from the behaviour described by the established models.

More generally, it has been shown the flexibility of the models can be used to drive the design process systematically, in a manner which allows the response surface to be explored fully, with respect to the blending behaviours common in

most models for mixture experiments. This is in contrast to cruder approaches to providing effective coverage of the design space, such as uniform designs (Wang and Fang (1996)), which would similarly allow the GBMM to be used. The result is that new designs can be presented for established models, which can be said to be D-optimal, but which also allow for uncertainty regarding the form of the regressors, something that has not previously been considered.

With respect to the objectives above, it is felt that the project has been successful. In particular, it is felt the new class of models, the GBMM, are a novel contribution. It is already the case that these models can be advantageously applied. However, with the development of holistic experimental approaches, with these models at their centre, it is believed the benefits from their application can be substantial.

## 9.2 What is still to be achieved?

In order that the full benefits be gained from application of the GBMM, it will be necessary to undertake further research. Some of what follows should be considered imperative. However, other points are only suggestions. The primacy of developing a complete and universally applicable methodology, for the design and analysis of mixture experiments using the GBMM, should be the driving factor in the direction of all future research.

One important challenge is to look at the possibility of overfitting. In Examples F5-F8, the selected general ternary terms of joint effect were inconsistent, depending upon the fitting procedure applied. However, the general binary terms of joint effect were always the same or very similar; these were also consistently fitted before the ternary terms. It is possible that these differences in the fitted models, using different fitting procedures, reveal that the features described by the ternary terms were unnecessary. Certainly, looking at the plots, their influence is more subtle.

While the flexibility of the GBMM is obviously a boon, theoretically, the range of available terms increases the chance of overfitting. It is beneficial to explore blending effects which other models cannot. However, often, where other models cannot describe a more subtle feature, the GBMM has a solution. In this manner, features could be erroneously identified as important to representing the response, rather than being a reflection of variability in the experiment. There

would be overfitting and consequently, one of the most important challenges of future work on the GBMM, is to establish a method for avoiding such situations from happening.

More generally, how the fitting procedure is applied needs clarification. Several variations were presented above, but with no conclusions regarding which was better. Further work could be done in order to establish one method. Furthermore, more effective programming could result in a more efficient program, reducing the relatively long amount of time taken to fit GBMM.

In addition to these two tasks relating to the fitting procedure, there are also several other areas which definitely need addressing. While Examples F1-F8 look at an experiment across a constrained region of the full simplex, there is yet to be a theoretical examination of application of the GBMM in such cases. Moreover, no work has yet looked at GBMM for mixture - process variable (MPV) experiments. Both constrained experimental regions and MPV experiments are common and it is believed that, in such situations, the GBMM could be advantageously applied. This is particularly so with constrained mixture experiments, where there is commonly a problem with collinearity. The ability of the GBMM to present a more parsimonious model would go some way to addressing this problem in some cases.

There are several directions for further work on experimental designs for the GBMM. Certainly, designs could be developed and applied in the sequential manner discussed in Example 3. Moreover, it remains to expand upon the designs presented here, looking at those for more components. Particularly with the saturated designs, it is suspected that there will be simple rules, relating to the values of the regressor defining parameters, defining the allocation of the design points. Finally, it is important to look at constrained mixture experiments, where the allocation of design points for optimal designs can become more complex.

Regarding development to the GBMM itself, the path is less clear. There could be benefits from interpreting the terms of the GBMM. For example, Model 5.3, of Example F3, described two distinct features in the response surface, through two distinct terms. A precise understanding of these two effects would be gained through the regressor defining parameters of these terms. However, it is unclear how generally such an approach could be applied, particularly where there are a larger number of components and a larger number of terms.

One other possible way to develop the new class of models would be to introduce more terms. In Chapter 3, it was demonstrated that the general terms of joint effect could be shown to be the product of two functions, each taking a similar form to the probability density function of the Dirichlet distribution. It may prove interesting to explore other known functions upon the simplex, in order to similarly construct new terms describing types of behaviour the terms of the GBMM cannot describe.

### 9.3 Summary

The manner in which the GBMM is applied can still be improved; it is evident that clarification is needed regarding the fitting procedure and comprehensive ideas on experimental design are still in development. Moreover, in order to maximise its utility, several areas are still to be looked at, including MPV experiments. However, it seems undeniable that the GBMM has the potential to prove a useful tool in the analysis of mixture experiments.

# Appendix A

## Designs for the GBMM

$x_1$	$x_2$	$x_3$
0	0.132	0.868
0.132	0	0.868
0.5	0	0.5
0.500	0	0.5
0	0.5	0.5
0	0.868	0.132
0.868	0.132	0
0.868	0	0.132
0.203	0.594	0.203
0.200	0.200	0.600
0.594	0.203	0.203
0.132	0.868	0
1	0	0
0	1	0
0	0	1

Table A.1: 15 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial

$x_1$	$x_2$	$x_3$
0.5	0.5	0
0	0.5	0.5
0.5	0	0.5
0.144	0.712	0.144
0.144	0.144	0.712
0	0.132	0.868
0.468	0.064	0.468
0.868	0	0.132
0.132	0	0.868
0.132	0.868	0
0	0.868	0.132
0.712	0.144	0.144
0.064	0.468	0.468
0.332	0.332	0.336
0.468	0.464	0.068
0.868	0.132	0
1	0	0
0	1	0
0	0	1

Table A.2: 19 point design for GSCBMM identified with priors equal to values dictated by the Scheffé cubic polynomial

$x_1$	$x_2$	$x_3$
0.120	0.760	0.120
0.136	0.864	0
0	0.5	0.5
0.12	0.12	0.76
0	0.864	0.136
0.136	0	0.864
0.5	0.5	0
0.232	0.232	0.536
0.864	0	0.136
0.864	0.136	0
0.232	0.536	0.232
0.760	0.120	0.120
0.536	0.232	0.232
0	0.136	0.864
0.5	0	0.5
1	0	0
0	1	0
0	0	1

Table A.3: 18 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial

$x_1$	$x_2$	$x_3$
0.404	0.596	0
0	0.125	0.875
0.128	0.744	0.128
0	0.404	0.596
0	0.875	0.125
0.875	0.125	0
0.125	0	0.875
0.596	0.404	0
0.404	0	0.596
0.128	0.128	0.744
0.125	0.875	0
0	0.596	0.404
0.744	0.128	0.128
0.532	0.234	0.234
0.596	0	0.404
0.234	0.532	0.234
0.875	0	0.125
0.234	0.234	0.532
1	0	0
0	1	0
0	0	1

Table A.4: 21 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial



$x_1$	$x_2$	$x_3$
0.125	0	0.875
0.864	0.136	0
0.532	0.236	0.232
0.59375	0.000	0.40625
0	0.136	0.864
0.406	0	0.594
0.232	0.236	0.532
0	0.864	0.136
0.136	0.864	0
0.232	0.536	0.232
0.748	0.124	0.128
0.5	0.5	0
0	0.5	0.5
0.12	0.76	0.12
0.875	0	0.125
0.128	0.124	0.748
1	0	0
0	1	0
0	0	1

Table A.5: 19 point design for GQBMM identified with priors equal to values dictated by the Scheffé quadratic polynomial

$x_1$	$x_2$	$x_3$
0.844	0	0.156
0.484	0.0312	0.484
0	0.752	0.248
0.156	0	0.844
0.156	0.844	0
0	0.248	0.752
0.784	0.108	0.108
0.844	0.156	0
0.484	0.484	0.0312
0.092	0.812	0.096
0.092	0.096	0.812
0.250	0.375	0.375
1	0	0
0	1	0
0	0	1

Table A.6: 15 point design for GQBMM identified with priors equal to values dictated by the Becker H2 model

$x_1$	$x_2$	$x_3$
0	0.132	0.868
0.868	0	0.132
0.184	0.632	0.184
0.5	0	0.5
0.184	0.184	0.632
0.5	0.5	0
0.132	0	0.868
0	0.5	0.5
0.632	0.184	0.184
0.868	0.132	0
0.132	0.868	0
0	0.868	0.132
1	0	0
0	1	0
0	0	1

Table A.7: 15 point design for GQBMM identified with priors equal to values dictated by Becker H2 model and the Scheffé quadratic polynomial

$x_1$	$x_2$	$x_3$
0.472	0.472	0.056
0	0.152	0.848
0.5	0.5	0
0.056	0.472	0.472
0.848	0.152	0
0.784	0.108	0.108
0.104	0.104	0.792
0	0.848	0.152
0.152	0	0.848
0.5	0	0.5
0.152	0.848	0
0.332	0.332	0.336
0.472	0.056	0.472
0.848	0	0.152
0.108	0.784	0.108
0	0.5	0.5
1	0	0
0	1	0
0	0	1

Table A.8: 19 point design for GQBMM identified with priors equal to values dictated by the Becker H2 model

$x_1$	$x_2$	$x_3$
0.144	0.856	0
0.5	0.5	0
0.052	0.476	0.472
0.265625	0.531250	0.203125
0	0.852	0.148
0	0.492	0.508
0.144	0	0.856
0.268	0.208	0.524
0.5	0	0.5
0.568	0.216	0.216
0.1	0.8	0.1
0.104	0.104	0.792
0.792	0.104	0.104
0.859	0.141	0
0	0.144	0.856
0.86	0	0.14
1	0	0
0	1	0
0	0	1

Table A.9: 19 point design for GQBMM identified with priors equal to values dictated by Becker H2 model and the Scheffé quadratic polynomial

# Bibliography

- Akay, K. U. (2007). A note on model selection in mixture experiments. *Journal of Mathematics and Statistics* 3(3), 93.
- Atkinson, A. C., A. N. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*. New York: Oxford University Press.
- Becker, N. G. (1968). Models for the response of a mixture. *Journal of the Royal Statistical Society, Series B* 30, 349–358.
- Becker, N. G. (1970). Mixture designs for a model linear in the proportions. *Biometrika* 57, 329–338.
- Becker, N. G. (1978). Models and designs for experiments with mixtures. *Australian Journal of Statistics* 20, 195–208.
- Brown, L., A. Bissett, and A. N. Donev (2012). A general model for mixture experiments. *Technometrics*.
- Chan, L.-Y. (2000). Optimal designs for experiments with mixtures: a survey. *Communications in statistics-theory and methods* 29(9-10), 2281–2312.
- Chan, L. Y., Y. N. Guan, and C. Q. Zhang (1998). A-optimal designs for an additive quadratic mixture model. *Statistica Sinica* 8, 979–990.
- Chan, L. Y., J. H. Meng, Y. C. Jiang, and Y. N. Guan (1998). D-optimal axial designs for quadratic and cubic additive mixture models. *Australian and New Zealand Journal of Statistics* 40, 359–371.
- Chen, W. S., W. Y. Zhu, and C. J. Hu (1985). A new mixture model with three or four components and their quasi D-optimality. Technical report, Northern Illinois University.

- Claringbold, P. J. (1955). Use of the simplex design in the study of joint action of related hormones. *Biometrics* 11, 174–185.
- Cornell, J. and J. Gorman (1978). On the detection of an additive blending component in multicomponent mixtures. *Biometrics*, 251–263.
- Cornell, J. A. (2000). Fitting a slack-variable model to mixture data: Some questions raised. *Journal of Quality Technology* 32, 133–147.
- Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data (3rd edition)*. New York: Wiley.
- Darroch, J. N. and J. Waller (1985). Additivity and interaction in three-component experiments with mixtures. *Biometrika* 72, 153–163.
- Draper, N. R. and W. Lawrence (1965a). Mixture designs for four factors. *Journal of the Royal Statistical Society, Series B* 27, 473–478.
- Draper, N. R. and W. Lawrence (1965b). Mixture designs for three factors. *Journal of the Royal Statistical Society, Series B* 27, 450–465.
- Draper, N. R. and F. Pukelsheim (1998). Mixture models based on homogeneous polynomials. *Journal of Statistical Planning and Inference* 71(1), 303–311.
- Draper, N. R. and R. C. St John (1977a). Designs in three and four components for mixtures models with inverse terms. *Technometrics* 19, 117–130.
- Draper, N. R. and R. C. St John (1977b). A mixtures model with inverse terms. *Technometrics* 19, 37–46.
- Efroymson, M. A. (1965). Multiple regression analysis. In A. Ralston and H. S. Wilf (Eds.), *Mathematical Methods for Digital Computers*. New York: Wiley.
- Focke, W. W., M. H. Ackermann, and R. L. J. Coetzer (2012). Mixing rules based on power means and generalized Q-fractions. *Journal of Chemical Engineering of Japan* 45, 1–8.
- Focke, W. W., C. Sandrock, and S. Kok (2007). Weighted-power-mean mixture model: Application to multicomponent liquid viscosity. *Industrial Engineering Chemistry Research* 46, 4660–4666.

- Golub, G. H. and V. Pereyra (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis* 10, 413–432.
- Golub, G. H. and V. Pereyra (2003). Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems* 19, R1–R26.
- Gorman, J. W. and J. E. Hinman (1962). Simplex lattice designs for multicomponent systems. *Technometrics* 4, 463–487.
- Hackler, W. C., W. W. Kriegel, and R. J. Hader (1956). Effect of raw-material ratios on absorption of whiteware compositions. *Journal of the American Ceramic Society* 39, 20–25.
- Huang, J. (2005). Maximum likelihood estimation of dirichlet distribution parameters. *Distribution. CMU Technique Report*, 1–9.
- Johnson, T. M. and M. E. Zabik (1981). Response surface methodology for analysis of protein interactions in angel food cakes. *Journal of Food Science* 46(4), 1226–1230.
- Kang, L., V. Roshan Joseph, and W. A. Brenneman (2011). Design and modeling strategies for mixture-of-mixtures experiments. *Technometrics* 53(2), 125–136.
- Kenworthy, O. O. (1963). Fractional experiments with mixtures using ratios. *Industrial Quality Control* 19, 24–26.
- Khuri, A. I. (2005). Slack-variable models versus Scheffé’s mixture models. *Journal of Applied Statistics* 32, 887–908.
- Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B* 21, 272–319.
- Kiefer, J. (1961). Optimum designs in regression problems, II. *The Annals of Mathematical Statistics* 32, 298–325.
- Lambrakis, D. (1969). Experiments with mixtures: an alternative to the simplex-lattice design. *Journal of the Royal Statistical Society. Series B* 31, 234–245.

- Liu, S. and H. Neudecker (1997). Experiments with mixtures: optimal allocations for Beckers models. *Metrika* 45(1), 53–66.
- Marquardt, D. W. and R. D. Snee (1974). Test statistics for mixture models. *Technometrics* 16(4), 533–537.
- Piepel, G. and T. Redgate (1997). Mixture experiment techniques for reducing the number of components applied for modeling waste glass sodium release. *Journal of the American Ceramic Society* 80(12), 3038–3044.
- Piepel, G. F. and J. A. Cornell (1994). Mixture experiment approaches: examples, discussion, and recommendations. *Journal of Quality Technology* 26(3), 177–196.
- Piepel, G. F. and S. M. Landmesser (2009). Mixture experiment alternatives to the slack variable approach. *Quality Engineering* 21, 262–276.
- Piepel, G. F., J. M. Szychowski, and J. L. Loepky (2001). Augmenting Scheffé linear mixture models with squared and/or crossproduct terms. *Journal of Quality Technology* 34(PNNL-SA-33717).
- Quenouillé, M. H. (1953). *The Design and Analysis of Experiments*. New York: Hafner.
- Quenouillé, M. H. (1959). Experiments with mixtures. *Journal of the Royal Statistical Society, Series B* 21, 201–202.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Scheffé, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society, Series B* 20, 344–360.
- Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society, Series B* 25, 235–263.
- Snee, R. D. and A. A. Rayner (1982). Assessing the accuracy of mixture model regression calculations. *Journal of Quality Technology* 14, 67–79.

- Thompson, W. and R. Myers (1968). Response surface designs for experiments with mixtures. *Technometrics* 10, 739–756.
- van Vuuren, W. and P. P. Goos (2012). Mixture-amount models for media mix optimization. *ERASMUS*.
- Wang, Y. and K. Fang (1996). Uniform design of experiments with mixtures. *Science in China, Series A* 39, 264–275.