PARAMETER INFERENCE AND ESTIMABILITY ANALYSIS FOR NONLINEAR MULTI-RESPONSE SYSTEMS BY MEANS OF PROFILING

by

Bahareh Golfar

A thesis submitted to the Department of Chemical Engineering

in conformity with the requirements for

the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

(September, 2014)

Abstract

Determining the inference bounds for parameter estimates in multi-response models is a challenge for research and industry since potential co-dependencies among responses add to the complexity of the problem. Statistical profiling is a valuable method in such situations that provides insight into model nonlinearity and co-dependencies among parameter estimates by studying the behavior of the likelihood function.

This thesis focuses on extending the application of profiling for parameter inference and estimability analysis to nonlinear multi-response models with unknown noise covariance terms. Profiling is explored based on Generalized Least Squares (GLS) and Determinant Criterion (DC). One issue is finding the distribution and the corresponding degrees of freedom of the profile likelihood function, which depends on the process of estimation of the noise covariance matrix in multi-response models. A method for estimating this matrix and performing profiling is proposed. This research proves that for multi-response models with m responses, p parameters and n experimental runs, the GLS-based profile likelihood function has a χ_1^2 or an $F_{1,nm-p}$ distribution for cases with known or unknown noise covariance matrices respectively. Similarly, the DC-based profile likelihood is shown to have an $F_{1,nm-p}$ distribution. The GLS-based and DC-based profiling approaches are compared theoretically and in practice, suggesting that the GLS-based method is overall more advantageous for small datasets and a combination of DC for parameter estimation and GLS for profiling generally produces reliable results when certain issues are avoided.

Furthermore, application of profiling for parameter estimability analysis is explored and a profile-based exploratory analysis is proposed which studies the GLS-based profile likelihood plots and profile traces to reveal inestimability issues. This method links parameter estimability

analysis to parameter inference and brings the two objectives of this thesis together. All the proposed methods are examined by means of several examples, the results of which validate the defined approaches.

Acknowledgements

My sincere gratitude to:

professors James McLellan and Thomas Harris for providing me with this opportunity and sharing their knowledge and wisdom; the staff of the Chemical Engineering Department at Queen's University for being such a friendly and efficient team; my family Chris, Mina, Ali and Behdad for their endless love, care and patience; and my friends and colleagues for giving me a home away from home.

I would also like to thank Professor Kim McAuley for her encouragement and suggestions that have helped strengthen this work.

The financial support of the Natural Sciences and Engineering Research Council (NSERC) in the form of Discovery Grants, MITACS, and Queen's University, is gratefully acknowledged.

This thesis is dedicated to my grandfather Hassan Mandehzadeh for his constant encouragements and interest in my achievements, from learning my first words to pursuing my dreams far away from home.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Bold lowercase characters represent vectors, and bold uppercase letters denote matrices.

DC	Determinant Criterion
GLASSO	Graphical Least Absolute Shrinkage and Selection Operator
GLS	Generalized Least Squares
HPD	Highest Posterior Density
IID	Independent Identically Distributed
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Likelihood Ratio
LSE	Least Squared Errors
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MRCE	Multivariate Regression with Covariance Estimation
MSE	Mean Squared Errors
Multi-Response	Multiple Response
MVN	Multivariate Normal
NLS	Nonlinear Least Squares
PCA	Principal Component Analysis
PSC	Profile-based Sensitivity Coefficient
PWL	Plug-in Joint Weighted LASSO
SRD	Signed Root Deviance
SRDP	Signed Root Deviance Profiling

SSE	Sum of Squared Errors
SUR	Seemingly Unrelated Regression
SVD	Single Value Decomposition
WLS	Weighted Least Squares

Symbols and Nomenclature

B	$p \times m$ parameter matrix in SUR models
$blockdiag_n$	Operator that constructs a block diagonal matrix from n matrices
C	Deviation matrix
d	Tuning parameter
$D(\theta)$	Determinant function
$\widetilde{D}(\theta_k)$	Value of the Determinant function at the DC-optimal parameter estimates conditional on a fixed value of θ_k
e	Residual values
E{}	Expected value function
$f(x_i, \theta)$	Model function
$f(\boldsymbol{\theta}, {\sigma_\varepsilon}^2)$	Prior density function
$f(\boldsymbol{\theta}, {\sigma_\varepsilon}^2 \vert \boldsymbol{y})$	Posterior density function
$f(\boldsymbol{y} \vert \boldsymbol{\theta}, {\sigma_\varepsilon}^2)$	Probability density function
$F_{\nu_1,\nu_2,\alpha}$	Upper α quantile of the *F* distribution with ν_1 and ν_2 degrees of freedom
H	Hessian matrix
$\boldsymbol{I}_{nm}$	$nm \times nm$ identity matrix
$\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}$	Orthogonal projection operator
K_d	decay rate coefficient (hr^{-1})

K_s	Michaelis-Menten constant (g/l)
l	Number of linear dependencies among responses
m	Number of responses
m'	Number of linearly independent responses
n	Number of runs
p	Number of parameters
$\boldsymbol{P}$	Projection operator
$PL_S(\theta)$	GLS-based profile likelihood function
$PL_D(\theta)$	DC-based profile likelihood function
s^2	Least Squares or GLS-based average noise variance estimate
S	Matrix of singular values in Singular Value Decomposition
s'^2	DC-based average noise variance estimate
$S(\theta)$	Sum of Squared Errors (SSE) function
$\tilde{S}(\theta_k)$	Value of the SSE function at the least squares parameter estimates conditional on a fixed value of θ_k
$S_w(\theta)$	Weighted sum of squares function
$t_{\alpha/2,\nu_1}$	Upper $\alpha/2$ quantile of Student's *t*-distribution with ν_1 degrees of freedom
tr	Trace of the matrix
vec	Operator that stacks the elements of a matrix into a vector
U	Output rotation matrix for Singular Value Decomposition
V	Eigenvector
V	Input rotation matrix for Singular Value Decomposition
$\boldsymbol{V}(\boldsymbol{\theta})$	Sensitivity matrix

$\mathbf{W}$	Weight matrix
$W_m(\boldsymbol{\Sigma_\varepsilon}, \nu)$	m-variate Wishart distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Sigma_\varepsilon}$
x	Explanatory variable
y	Response variable
$Y_{x/s}$	yield coefficient
$z_{\alpha/2}$	Upper $\alpha/2$ quantile of the Standard Normal distribution
$\boldsymbol{Z}(\boldsymbol{\theta})$	Residual matrix
$z^*(c)$	Signed Root Deviance (SRD) function

Greek Letters

α	Significance level
β	Parameters of linear models
$\delta(\theta_k)$	Studentized parameter
ε	Disturbance term
$\eta(\theta)$	Expectation function
$\boldsymbol{\theta}$	Vector of parameters
$\hat{\theta}$	Least squared errors parameter estimate
$\tilde{\theta}_{-k}$	Least squares estimates for parameters conditional on θ_k
λ	eigenvalue
$\mu(s)$	specific growth rate function
μ_m	maximum specific growth rate (hr^{-1})
ν	Degrees of freedom
${\sigma_\varepsilon}^2$	Noise variance
$\boldsymbol{\Sigma}$	Within-run noise covariance matrix

$\boldsymbol{\Sigma}_{\theta}$	Parameter covariance matrix
$\tau(\theta)$	Profile *t* function
$\chi^2_{\nu;\alpha}$	Upper α quantile of the χ^2 distribution with ν degrees of freedom
ω_{ij}	Elements of the overall noise covariance matrix ($\boldsymbol{\Omega}$)
$\boldsymbol{\Omega}$	Overall Noise Covariance Matrix

Chapter 1

Introduction

1.1 Introduction

Chemical processes often provide datasets containing a number of explanatory variables (x) and responses (y). The main goal of process modeling is to describe the systematic relationship between the values of the explanatory variables and responses. A statistical model can be used to predict a response (or a vector of responses) for a specific set of regressor variables (Seber and Wild, 1989). Accurate models can have a significant role in design, scale up and optimization of systems (e.g., (Bates and Watts, 1988); (Seber and Wild, 1989)).

For a set of n experimental runs, a statistical single-response model can be written as (Bates and Watts, 1988):

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{1-1}$$

where $\boldsymbol{y}$ represents the $n \times 1$ vector of observations of the response at the run conditions, $\boldsymbol{\theta}$ is the $p \times 1$ parameter vector and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of disturbances at the run conditions. $\boldsymbol{\eta}(\boldsymbol{\theta})$ is the expectation mapping and contains the model functions evaluated at the run conditions as follows:

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = [f(\boldsymbol{x}_i, \boldsymbol{\theta})];\ (i = 1,2, \dots, n) \tag{1-2}$$

in which $\boldsymbol{x}_i$ are the values of the regressor variables for run i.

Depending on the number of the observed response variables, the model can be categorized as a single-response or a multiple-response (multi-response) model. Examples of multi-response cases include observing different types of variables (such as

product concentration and yield) over the course of experimental runs, or measuring a functional quality at a number of different index values such as a particle size distribution. The time response of a single-response model can also be treated in this manner. With advanced analytical tools which make simultaneous measurements possible, multi-response models are becoming more and more common (Box et al., 1973). These systems have the advantage of gathering information from several sources to draw conclusions about the process. Various methods have been used for parameter estimation and inference in multi-response systems which are discussed in detail in Chapter 2. There are still a few challenges in this regard, including determining and accounting for the co-dependencies between different response variables.

Statistical models can be categorized as linear or nonlinear with respect to their parameters or their explanatory variables. In the present work, the term "nonlinear model" refers to models in which the parameters appear nonlinearly. A model is nonlinear when at least one of the derivatives of the expectation function with respect to one of the parameters depends on at least one of the parameters (Bates and Watts, 1988). The expectation surface for a linear model is a plane of infinite extent, whereas for the nonlinear case it has the form of a p-dimensional curved surface of possibly infinite or finite extent (with p being the number of parameters of the model). In chemical engineering, nonlinear models are often encountered in reaction rate expressions, phase equilibria, adsorption, cell growth kinetics, heat transfer coefficients, and specific heat expressions. Material and energy balances also give rise to product, or bilinear terms (e.g., flow * temperature * specific heat, or flow * concentration).

Nonlinearity might arise from the natural curvature of the expectation surface (i.e., intrinsic curvature) or can be associated with the mapping from the parameter space onto the expectation surface (i.e., parameter-effects curvature) (Bates and Watts, 1980). In the latter case, the nonlinear model can often be transformed into a linear one by re-parameterization, but intrinsic nonlinearity should be accounted for by using appropriate modeling techniques. Nonlinear regression problems arise frequently in steady-state and dynamic models of single- or multiple-response types. One of the main challenges when the model depends nonlinearly on its parameters is finding reliable parameter inference intervals. The uncertainties in the estimated parameters are expressed by inference regions that contain true parameter values with specific probability levels. These regions can be built by focusing on one parameter at a time (marginal inference regions) or by considering the joint distribution of the parameter estimates (joint inference regions) (Bates and Watts, 1988).

While it is possible to obtain unique parameter estimates and exact likelihood bounds for linear models, it is more complicated to do so for nonlinear models (Watts, 1994). Parameter inference for nonlinear models is often based on some kind of sequential linearization method. There are several methods to measure the accuracy of the linear approximations such as curvature measurements discussed by Bates and Watts (1988). Other techniques for parameter inference include likelihood contour plots, Bayesian and empirical techniques such as Monte Carlo and Bootstrapping methods (Bates and Watts, 1988); (Seber and Wild, 1989); (Fraser, 1993); (Jitjareonchai et al., 2006). These methods can be good alternatives to linearization if the numerical steps involved can be carried out reliably and efficiently from a computational point of view. However, the inference

methods for nonlinear models are still not fully reliable and there is ongoing research in this field.

Because of the inadequacies of linearization approximation methods for parameter inference, Bates and Watts (1988) introduced the method of statistical profiling, suggesting that it would provide more accurate parameter inference intervals. They initially defined profiling in terms of the Sum of Squared Errors (SSE) function but later on other researchers, notably Chen and Jennrich (1996), defined a more general form of profiling in terms of the likelihood function. Plots of the profiling function versus the profiling parameter (profile plots) and pairwise plots of the conditional parameter values versus the profiling parameter (profile traces) can be used to form parameter likelihood intervals and regions. Furthermore, these plots indicate the presence of nonlinear co-dependencies among the parameter estimates that could otherwise be difficult to detect, and they also indicate the impact of nonlinearity on the parameter estimation problem. This method and some of its applications are explained in detail in Chapter 2. Profiling has been pursued by several researchers for single-response models. However, except for a few papers (e.g., (Soo and Bates, 1996); (Raue et al., 2009)), profiling for multi-response models has not been sufficiently explored.

Another major consideration in modeling is whether a unique set of parameters can be estimated based on the model configuration and experimental dataset (McLean and McAuley, 2012). Some analyses such as parameter identifiability and estimability are widely used for this matter. This thesis focuses on parameter estimability analysis. An important step in fitting a mathematical model to a set of experimental data is assessing the estimability of the parameters in the model and finding a set of estimable parameters

(Vajda et al., 1989). Parameter estimability depends on the structure and parameterization of the model, and the experimental design used to gather the data. Estimability analysis is not only necessary to ensure the quality of the estimation results but also helps with the experimental design. If the analysis is done on the planned data and the estimable parameters are identified, experiments can be designed specifically to estimate those parameters more precisely and additional experiments can be designed to improve the estimability of the remaining parameters. The issue of parameter inestimability is particularly noticeable in biological processes (e.g., pharmacokinetics and systems biology models), and chemical process models (e.g., polymer reactor and combustion models) that deal with a large number of parameters and also face some limitations around measurements and data collection (Dochain et al., 1995); (Wu et al., 2008). More information about estimability analysis and the current methods for performing it is provided in Chapter 5. As explained in that chapter, estimability analysis techniques have not been completely successful yet and more research is required in this field.

1.2 Motivation and Objectives

As mentioned in the previous section, there are still quite a few challenges and problems regarding the statistical model-building procedure for multi-response models. There is ongoing research towards presenting new techniques as well as improving the current methods. The present work has focused mainly on statistical model-building based on profiling for multi-response models by proposing some modifications and extensions to the profiling method.

The existing literature on profiling (e.g., (Bates and Watts, 1988); (Kang and Bates, 1990)) suggests that statistical profiling is a very advantageous technique for parameter inference with graphical results that are relatively easy to interpret. The majority of the results for profiling have been obtained for single-response models and this field is relatively well established (e.g., (Bates and Watts, 1988); (Chen and Jennrich, 1996)). However, only a few papers have been published regarding profiling for multi-response models which arise quite frequently in parameter estimation problems (Soo and Bates, 1996); (Raue et al., 2009). On the other hand, parameter likelihood intervals for multi-response systems have been explored by a number of researchers but not in a profiling context (e.g., (Zellner, 1962); (Bard, 1974); (Bates and Watts, 1985); (Kang and Bates, 1990); (Oxby et al., 2003)). This constitutes the main motivation of this thesis in which the primary objective is to find appropriate ways to extend the application of profiling to nonlinear multi-response models. Realizing the potential of the profiling method and considering the improvements in the available software are motivations to investigate more effective ways of building profile plots and extending their application.

One particular area in which profiling seems to be appealing is the analysis of large-scale systems consisting of several response variables such as those encountered in systems biology (see for example, (Raue et al., 2009)). Chemical process models in general can be of high complexity due to the number of differential equations, complicated dependencies among the state and response variables, and nonlinearity of model terms. Current methods for parameter inference in complicated models are not very reliable or efficient. Addressing these issues is one of the objectives of this research in which I seek to provide a more comprehensive set of diagnostics for assessing and building models. A profile-

based technique for determining the parameter likelihood intervals can be useful for many practical applications for such systems.

Profiling for multi-response systems can be done based on the common parameter estimation approaches for multi-response models namely, the Generalized Least Squares (GLS) and the Determinant Criterion (DC) methods. Soo and Bates (1996) have introduced DC-based profiling but the GLS-based profiling has not been sufficiently explored. Raue et al. (2009) have done some research on GLS-based profiling focusing on models with known noise covariance matrix with a diagonal structure in the systems biology field. However, since in practice the noise variance and covariance values are often unknown (e.g., (Seber and Wild, 1989); (Kang and Bates, 1990)), in this research I have focused on GLS-based profiling for systems with unknown noise covariance matrix.

One of the challenges of GLS-based profiling for models with unknown noise covariance matrix is the process of estimating this matrix and incorporating it into the profile likelihood function. In GLS-based parameter estimation it is a common practice to make assumptions about this matrix in order to simplify the estimation problem. However, these assumptions can impact the parameter estimation outcomes and consequently lead to unreliable profiling results. Some common assumptions in this regard and their influence on the parameter estimates have been discussed in this thesis. Furthermore, I have investigated several approaches for estimating the noise covariance matrix during parameter estimation and profiling stages and made a suggestion about the most reliable approach.

Another issue in using GLS-based profiling for parameter inference is finding the distribution of the GLS-based profile likelihood function. The distribution of this function

and the corresponding degrees of freedom are studied in this research and the outcomes and conclusions are presented in Chapter 3.

DC-based profiling was proposed by Soo and Bates (1996) and discussed in Chapter 4 of this thesis. The suggested distribution and degrees of freedom for the DC-based likelihood function were not clear to us and that prompted us to investigate this method and present an argument in this regard in Chapter 4. In this argument it has been explained why the degrees of freedom used by Soo and Bates (1996) for the DC-based profiling might not represent the characteristics of the profile likelihood function for multi-response models in general. In addition, different degrees of freedom are proposed that are believed to be more applicable.

Another focus of the present research is the parameter estimability analysis. Unfortunately even though determination of parameter estimability is an advantageous and necessary step in modeling, it is often ignored (Yao et al., 2003). This motivated us to explore the potential of applying profiling to this analysis. Profiling can be useful for parameter estimability analysis due to its ability to form and study parameter inference intervals. Raue et al. have investigated this application for multi-response systems with known noise covariance matrices (Raue et al., 2009); (Raue et al., 2011). However, there are still issues regarding the distribution of the profile likelihood function and the cut-off values for this function to determine parameter estimability for models with unknown noise covariance matrix. Thus, proposing a profile-based parameter estimability analysis for nonlinear multi-response systems with unknown noise covariance matrix is another objective of this research and has been explored in Chapter 5.

In summary, the objectives of this thesis are as follows:

- Extending profiling applications to parameter inference in nonlinear multi-response systems using the GLS method and focusing on the cases with unknown noise covariance terms.
- Exploring the DC-based profiling and its application for parameter inference in nonlinear multi-response models.
- Improving the application of profiling for model analysis and interpretation, specifically parameter estimability analysis.

1.3 Scope and Organization of Thesis

In order to meet the objectives and the expected contributions, this thesis is organized as follows:

Chapter 1 (this chapter) contains a summary of statistical model building, briefly describing different types of models that are encountered in chemical engineering. It mentions some areas that are in need of further improvements and how this has motivated us to carry out this research.

Chapter 2 provides a detailed background on nonlinear regression including current estimation and inference methods for single- and multi-response models including the profiling method. The chapter also includes an extensive description of the profiling method, profile plots and sketches and their applications. Profiling for multi-response models is reviewed and some areas which need further research are pointed out. Profile likelihood functions based on both GLS and DC methods are introduced.

Chapter 3 is devoted to GLS-based profiling, focusing mostly on profiling with unknown noise covariance matrix. The importance of the estimation of the noise covariance matrix

is discussed and an approach for this estimation process during parameter estimation and profiling is suggested. The distribution of the GLS-based profile likelihood function and the corresponding degrees of freedom are explored and a GLS-based profiling approach for parameter inference for nonlinear multi-response models is proposed. The proposed approach is applied to some case studies and the outcomes are summarized and discussed.

Chapter 4 contains the DC-based profiling method for parameter inference. A discussion regarding the distribution of the DC-based profile likelihood function and the degrees of freedom is presented. The application of this method is demonstrated using several case studies and eventually the two methods of GLS-based profiling and DC-based profiling are compared, revealing their advantages and disadvantages.

In Chapter 5 the parameter estimability analysis and its significance are explained and the current methods for this analysis are reviewed. A profile-based exploratory parameter estimability method is proposed and applied to a case study. The application of profiling for resolving some inestimability issues is described.

Lastly, Chapter 6 presents the final conclusions and contributions. Some recommendations for future work on the matters discussed in this research are also summarized in this chapter.

Chapter 2

Background and Literature Review

2.1 Introduction

In this chapter, the problem of parameter estimation and inference in statistical models is reviewed for single-response models, with techniques such as Least Squares, Maximum Likelihood and Bayesian methods, and for multi-response models by means of the Generalized Least Squares (GLS) and the Determinant Criterion (DC) approaches. The issue of nonlinearity and its impact on the estimation problems are discussed. Furthermore, the statistical profiling method is introduced, explaining its applications, and focusing in particular on its application in parameter inference. For multi-response models, profile functions are defined based on the GLS as well as the DC approaches. Since the goal of this thesis is to extend the application of profiling for parameter inference in nonlinear multi-response models, a literature review of the current state-of-the-art research in this field is provided and the outstanding issues are explained.

2.2 Parameter Estimation and Inference in Single-Response Models

There are various methods for parameter estimation in statistical models so that based on the dataset, model structure, noise structure, purpose of modeling, and available software the user can choose the most relevant technique. Least Squared Errors (LSE), Maximum Likelihood Estimation (MLE) and Bayesian analysis are some of the most common estimation techniques (Bates and Watts, 1988); (Seber and Wild, 1989).

A single-response model with p parameters and n experimental runs can be written as:

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{2-1}$$

in which $\boldsymbol{y}$ represents the $n \times 1$ vector of observations of the response at the n run conditions, $\boldsymbol{\theta}$ is the $p \times 1$ parameter vector, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of disturbances at the run conditions. $\boldsymbol{\eta}(\boldsymbol{\theta})$ is the vector of expectation function evaluated at the run conditions and defined by:

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = f(\boldsymbol{x}_i, \boldsymbol{\theta});\ (i = 1,2, \dots, n) \tag{2-2}$$

with $\boldsymbol{x}_i$ being the regressor variables for run i. Using the LSE method, and assuming that the noise terms are *IID Normal* (Independent Identically Distributed) with known variance ${\sigma_\varepsilon}^2$ (i.e., $\varepsilon \sim$IID $N(0, {\sigma_\varepsilon}^2)$), parameter estimates are obtained from minimizing the Sum of Squared Errors (SSE) function defined as:

$$S(\boldsymbol{\theta}) = \big(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\big) \tag{2-3}$$

Therefore, the least squares parameter estimates are obtained from:

$$\widehat{\boldsymbol{\theta}}_{LS} = \text{argmin}_{\boldsymbol{\theta}}\, S(\boldsymbol{\theta}) \tag{2-4}$$

If the model is nonlinear in the parameters, the Nonlinear Least Squares (NLS) method can be used in which a nonlinear iterative technique (such as Gauss-Newton, Newton-Raphson, Levenberg-Marquardt Compromise, or Nelder-Mead approach) is employed in order to find the optimal set of parameter estimates. NLS has many advantages such as generating parameter estimates that are asymptotically unbiased, consistent and efficient, but there are some concerns about this method. For example, NLS needs reasonable initial guesses for parameters to assure convergence to correct answers. Long computing times might be required, especially if iterations are carried out on poor guesses (Thie et al., 1997); (Campbell, 2007). Moreover, misleading initial guesses might cause convergence to local minima. In fact, even with the application of more sophisticated

optimization techniques, NLS has a risk of having several local minima in the sum of squared residuals and not converging to the least of them (e.g., (Marquardt, 1963); (Thie et al., 1997); (Campbell, 2007)). The possibility of multiple solutions (multiple sums of squares minima) undoubtedly represents one of the greatest problems in nonlinear estimation and can be very severe sometimes (Polic et al., 2004). This problem is expected to be particularly pronounced in models with a larger number of exponential terms, such as the case with pharmacokinetics or chemical reaction network models. The number of minima increases very rapidly as the number of exponential terms increases (Pedersen, 1977). Another concern arises from the fact that the dominant computation method in NLS is the Gauss-Newton method which works best on quadratic functions due to the way it is defined (Seber and Wild, 1989). Thus, when the function is not entirely quadratic in the neighborhood of $\widehat{\boldsymbol{\theta}}$, the results might be approximate.

Analysts who use NLS might encounter some difficulties when there are strong co-dependencies between the parameter estimates. One result of such co-dependencies is that confidence ellipsoids (ellipsoids corresponding to the contours of constant residual sum of squares) are greatly elongated. In practical terms, this means that one may be able to change the parameter estimates substantially without changing the residual sum of squares, or in other words, very different sets of estimates of the parameters give about equally good fits of the model to the data (Westlake, 1973). This implies that the precision of the estimates is poor. In spite of these issues, NLS is still widely used particularly by estimation software. Advanced techniques are being implemented in order to develop more reliable solution methods, for example shooting methods and profile-

based methods for dynamic systems and ODE models (Bock, 1983); (Poyton et al., 2006); (Varziri et al., 2008).

Another commonly used parameter estimation method is the Maximum Likelihood method in which parameter estimates are obtained from maximizing the likelihood function described by (Seber and Wild, 1989):

$$l(\boldsymbol{\theta}, {\sigma_\varepsilon}^2|\boldsymbol{y}) = \frac{1}{(2\pi{\sigma_\varepsilon}^2)^{n/2}} \exp\left(\frac{-1}{2{\sigma_\varepsilon}^2}\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)\right) \quad (2\text{-}5)$$

Again in this equation ${\sigma_\varepsilon}^2$ represents the known noise variance. The logarithm of the above equation directly follows as:

$$L(\boldsymbol{\theta}, {\sigma_\varepsilon}^2|\boldsymbol{y}) \propto -\frac{n}{2}\ln({\sigma_\varepsilon}^2) - \frac{1}{2{\sigma_\varepsilon}^2}\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big) \quad (2\text{-}6)$$

Subsequently, the log-likelihood function can be defined as:

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\ln({\sigma_\varepsilon}^2) - \frac{1}{2{\sigma_\varepsilon}^2}\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big) \quad (2\text{-}7)$$

and the maximum likelihood parameter estimates can be obtained from:

$$\widehat{\boldsymbol{\theta}}_{MLE} = \text{argmin}_{\boldsymbol{\theta}}\, L(\boldsymbol{\theta}) \quad (2\text{-}8)$$

Since for models with *IID Normal* noise $S(\boldsymbol{\theta}) = \big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y}-\boldsymbol{\eta}(\boldsymbol{\theta})\big)$ is the least squares function, it can be seen in equation 2-7 that maximization of $L(\boldsymbol{\theta})$ happens simultaneously with minimization of $S(\boldsymbol{\theta})$ and therefore both LSE and MLE methods yield the same estimation results.

In single-response models, the noise is usually assumed to be *IID Normal*. However, if that is not the case and the noise is heteroscedastic, the single-response problem can be viewed somehow similar to a multi-response model in a sense that the noise variance cannot simply be represented by ${\sigma_\varepsilon}^2$, but rather there will be a covariance matrix ($\boldsymbol{\Sigma}$) reflecting the noise variances associated with different run conditions on the main

diagonal. If in addition the noise is not independent, $\boldsymbol{\Sigma}$ will contain the noise covariances as the off-diagonal elements. Consequently, the noise terms have a *Multivariate Normal (MVN)* distribution that can be described by $\boldsymbol{\varepsilon} \sim \mathrm{MVN}(0, \boldsymbol{\Sigma})$. Thus, the log-likelihood function for single-response models with heteroscedastic noise takes the following form:

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\big(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\big)\boldsymbol{\Sigma^{-1}}\big(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\big) \qquad (2\text{-}9)$$

This pertains to the general case where the noise terms from different runs are neither independent nor identically distributed. This equation is particularly interesting since it shows the importance of the noise variance terms. However, if the noise terms are assumed to be independent between runs, $\boldsymbol{\Sigma}$ will be a diagonal matrix with elements of ${\sigma_{\varepsilon_i}}^2$, and the log-likelihood function of equation 2-9 becomes:

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\sum_i \ln({\sigma_{\varepsilon_i}}^2) - \frac{1}{2}\sum_{i=1}^{n} \frac{\big(y_i - \eta_i(\boldsymbol{\theta})\big)^{\mathbf{2}}}{{\sigma_{\varepsilon_i}}^2} \qquad (2\text{-}10)$$

Solving the log-likelihood function of equation 2-10 is like solving a Weighted Least Squares (WLS) problem with the inverse of noise variances as the weighting factors. However, for this to be feasible, these weighting factors should be known.

If the noise terms are unknown and correlated (noise covariance between different runs), estimating the noise covariance matrix becomes much more complicated and replicate runs are required. Since this is analogous to the situation often encountered in multi-response estimations, these issues will be discussed further in the multi-response segment.

Generally in the likelihood function, the noise variances (typically ${\sigma_\varepsilon}^2$ or $\boldsymbol{\Sigma}$) are considered as nuisance parameters that need to be estimated. One estimation approach is the concentrated likelihood method which starts by estimating these nuisance parameters

based on the parameter values ($\hat{\sigma}_\varepsilon{}^2(\boldsymbol{\theta})$), and then concentrating the likelihood function by substituting these noise estimates into it. As a result the likelihood function becomes mainly a function of $\boldsymbol{\theta}$ and the parameters can be estimated by maximizing this function (Seber and Wild, 1989). For single-response models with *IID* noise, $\boldsymbol{\theta}$ can be estimated independently of $\sigma_\varepsilon{}^2$, however this is not the case if the noise is not *IID*.

Bayesian estimation is also another widely used technique for parameter estimation in which estimates for parameter values are obtained from the knowledge about their prior distribution. The posterior density function $f(\boldsymbol{\theta}, \sigma_\varepsilon{}^2|\boldsymbol{y})$ for parameters is related to their prior density function $f(\boldsymbol{\theta}, \sigma_\varepsilon{}^2)$ based on the following relationship (Seber and Wild, 1989):

$$f(\boldsymbol{\theta}, \sigma_\varepsilon{}^2|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_\varepsilon{}^2) f(\boldsymbol{\theta}, \sigma_\varepsilon{}^2) \quad (2\text{-}11)$$

where $f(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_\varepsilon{}^2)$ is the conditional probability density function of the observations (which is the same as the likelihood function). One of the most common Bayesian methods is the Maximum a Posteriori (MAP) approach in which the parameter estimates are the values of parameters that maximize the posterior likelihood (also referred to as "the mode" of the function). This can be achieved by several means such as numerical methods or through differentiating the posterior density function with respect to the parameters and solving the conditions for a stationary point to come up with the parameter estimates. Since this method uses the likelihood function for parameter estimation, under the assumption of $\varepsilon \sim IID\ N(0, \sigma_\varepsilon{}^2)$, and with a uniform prior distribution for the parameters, the maximum posterior density estimate produces the same results as MLE (Seber and Wild, 1989).

Some of the major considerations in parameter estimation revolve around nonlinearity of the model parameterization. A considerable amount of work can be found in the literature that addresses this issue (e.g., (Beale, 1960); (Bates and Watts, 1988); (Watts, 1994); (Asuero and Bueno, 2011)); yet there is a continuing need for improvements in this field. Nonlinearity of a model arises from the nonlinear dependency of the model function and predictions on the respective parameters. A part of that is due to the way the parameters are expressed in the model (parameter-effects nonlinearity) and might be removed to some extent by re-parameterization or parameter transformations. Another part is due to the nature of the model function and cannot be completely removed (intrinsic nonlinearity). In order to distinguish these two sources of nonlinearity, Bates and Watts (1988) suggested a useful approach. They begin by developing measures for the curvature of the solution locus from second derivatives of the model function with respect to the parameters. Then the curvature is decomposed into a piece that lies in the tangent plane described by the sensitivity matrix (tangential curvature) and a remaining piece that is orthogonal to the tangent plane (normal curvature). Tangential curvature reflects the parameter-effects nonlinearity, whereas normal curvature constitutes the intrinsic nonlinearity (Bates and Watts, 1988). Intrinsic nonlinearity is the more problematic component since it cannot be reduced by re-parameterization. Thus, intrinsic nonlinearity in particular can impact parameter inferences. Curvature measures have been used to represent the extent of nonlinearity of both single- and multi-response systems (Guay, 1995).

When estimating parameters or functions of them, it is important to determine inference regions and provide a measure of uncertainty for the estimated values. Inference

approaches are drawn from estimation methods since the estimation techniques dictates the manner in which the parameter estimates depend on the noise, which is the source of random variation. For example joint and marginal confidence intervals are based on NLS estimations, likelihood regions are derived from MLE, and Highest Posterior Density (HPD) regions are drawn from Bayesian analysis (Bates and Watts, 1988).

Considering the single-response model of equation 2-1, inference can be carried out using the Taylor series linearization approximation of the expectation function (Bates and Watts, 1988):

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\eta}\left(\widehat{\boldsymbol{\theta}}\right) + \boldsymbol{V}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \tag{2-12}$$

where $\boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ is the sensitivity matrix evaluated at the least squares parameter estimates obtained from:

$$\boldsymbol{V}(\widehat{\boldsymbol{\theta}}) = \left.\frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{2-13}$$

Consequently, the approximate $100(1-\alpha)$ confidence interval for each parameter θ_k can be formed based on the least squares method (Bates and Watts, 1988); (Seber and Wild, 1989); (Watts, 1994):

$$\left[\hat{\theta}_k - t_{n-p,\alpha/2} se(\hat{\theta}_k), \hat{\theta}_k + t_{n-p,\alpha/2} se(\hat{\theta}_k)\right] \tag{2-14}$$

where $\hat{\theta}_k$ is the least squares estimate for θ_k, $t_{n-p,\alpha/2}$ represents the upper $\alpha/2$ quantile of the Student's *t* distribution with $n-p$ degrees of freedom, and $se\left(\hat{\theta}_k\right)$ is the estimated standard error of $\hat{\theta}_k$ and can be obtained from the sensitivity matrix $\boldsymbol{V}(\boldsymbol{\theta})$ and the average noise variance estimate (s^2) by means of the following equation:

$$se\left(\hat{\theta}_k\right) = \sqrt{\{(\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{V}(\boldsymbol{\theta}))^{-1}\}_{kk} s^2} \tag{2-15}$$

In this equation the subscript kk refers to the element on the k^{th} row and k^{th} column of the matrix, i.e., the k^{th} diagonal element, and s^2 is the estimated average noise variance that can be obtained from the Mean Squared Errors (MSE) (Bates and Watts, 1988):

$$s^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n-p} \tag{2-16}$$

Based on MLE, parameter likelihood regions can be viewed as parameter values for which the following equation holds:

$$L(\boldsymbol{\theta}, {\sigma_\varepsilon}^2 | \boldsymbol{y}) \geq cL\left(\hat{\boldsymbol{\theta}}, {\sigma_\varepsilon}^2 | \boldsymbol{y}\right) \tag{2-17}$$

where the elements of $\hat{\boldsymbol{\theta}}$ are the MLE parameter estimates. In the case of *IID Normal* noise, this equation becomes:

$$S(\boldsymbol{\theta}) \leq c'\, S(\hat{\boldsymbol{\theta}}) \tag{2-18}$$

The main problem is finding an appropriate value for c' that provides the correct coverage probability (i.e., the probability that the region contains the true parameter values). According to the linearization approximation:

$$S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) \approx \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)' \boldsymbol{V}(\hat{\boldsymbol{\theta}})' \boldsymbol{V}(\hat{\boldsymbol{\theta}}) \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) \tag{2-19}$$

Consequently, it can be shown that,

$$c' = 1 + \frac{p}{n-p} F_{p,n-p} \tag{2-20}$$

Thus, the $100(1-\alpha)$ joint confidence region for the parameters can be built from:

$$\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)' \boldsymbol{V}(\hat{\boldsymbol{\theta}})' \boldsymbol{V}(\hat{\boldsymbol{\theta}}) \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) \leq p s^2 F_{p,n-p,\alpha} \tag{2-21}$$

where $F_{p,n-p,\alpha}$ is the upper α quantile of the *F* distribution with p and $n-p$ degrees of freedom. The inequality of equation 2-21 results in elliptical joint confidence regions (Watts, 1994).

It is a common practice to estimate parameters based on linear approximations to the model. However, it should be noted that for nonlinear models, the Sum of Squared Errors (SSE) surface is not quadratic with respect to the parameters. Therefore, the results of a linearization approach which assumes this surface to be quadratic could vary from the actual case. If the derivatives of the expectation function with respect to parameters at the linearization point are relatively insensitive to changes in the parameter values (i.e., the first derivatives show only small changes and the second derivatives are very small), linearization may provide a good approximation to the nonlinear models. Thus, the accuracy of linearization-based estimations depends on the curvature of the expectation surface, and uniformity of parameter steps (Bates and Watts, 1991); (Watts, 1994).

The joint confidence regions for parameter estimates of a nonlinear model deviate from the usual elliptical shapes and as a result, the linearized approximate confidence intervals will not always produce satisfactory results. Moreover, symmetric confidence intervals (provided by linearization) may not accurately reflect the nature of the uncertainty in the parameter estimates, which might be more accurately reflected by asymmetric likelihood intervals. Due to the potential inaccuracy of linearized least squares confidence regions, likelihood regions have been used as an alternative for nonlinear inferences, but they still include some degree of approximation around the choice of significance levels (see for example, (Fraser, 1993)). The Bayesian approach has also been applied to deal with nonlinear inference regions, primarily through the concept of HPD regions (Bates and Watts, 1988); (Seber and Wild, 1989). An HPD region reflects a section R of the parameter space with $1-\alpha$ probability of the parameters belonging to that section ($pr(\theta \in R) = 1-\alpha$) and if $\theta_1 \in R$ and $\theta_2 \notin R$, then $p(\theta_1|\boldsymbol{y}) \geq p(\theta_2|\boldsymbol{y})$. The objective

of the inference problem is to make the region to be as small as possible. HPD regions are often generated using empirical means, such as Monte Carlo simulation and Markov Chain Monte Carlo (MCMC) methods (e.g., (Jitjareonchai et al., 2006)). They can be a good alternative if all the numerical steps involved can be carried out correctly and reliably. Likelihood parameter intervals will be further discussed later on in the context of profiling.

2.3 Parameter Estimation and Inference Methods for Multi-Response Models

For a multi-response system with *p* parameters, *m* responses and *n* experimental runs, the matrix representation of the model is (Bates and Watts, 1988):

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{2-22}$$

with $\boldsymbol{Y}$ being the response matrix:

$$\boldsymbol{Y} = [y_{ij}]\ ;\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{2-23}$$

and $\boldsymbol{\eta}(\boldsymbol{\theta})$ being the matrix of expectation functions evaluated at the run conditions, defined as:

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = [\eta_{ij}(\boldsymbol{\theta})] \tag{2-24}$$

$$\eta_{ij}(\boldsymbol{\theta}) = f_j(\boldsymbol{x}_i, \boldsymbol{\theta});\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{2-25}$$

For parameter estimation purposes, it is sometimes convenient to use the vector representation of the model by arranging the observations and the corresponding model functions and noise terms in single vectors as opposed to matrices. This can be done by introducing $\boldsymbol{y}_{vec}$, $\boldsymbol{\eta}_{\boldsymbol{vec}}(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_{\boldsymbol{vec}}$ as:

$$\boldsymbol{y}_{vec} = vec(\boldsymbol{Y}) \tag{2-26}$$

$$\boldsymbol{\eta}_{\boldsymbol{vec}}(\boldsymbol{\theta}) = vec(\boldsymbol{\eta}(\boldsymbol{\theta})) \tag{2-27}$$

$$\boldsymbol{\varepsilon_{vec}} = vec(\boldsymbol{\varepsilon}) \tag{2-28}$$

where the "*vec*" operator stacks the elements of a matrix into a vector. The following formulation has been chosen for the problem in this thesis which groups the observations by run (e.g., (McLean and McAuley, 2012)):

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nm} \end{bmatrix} = \begin{bmatrix} f_1(\boldsymbol{x}_1, \boldsymbol{\theta}) \\ \vdots \\ f_m(\boldsymbol{x}_1, \boldsymbol{\theta}) \\ \vdots \\ f_1(\boldsymbol{x}_n, \boldsymbol{\theta}) \\ \vdots \\ f_m(\boldsymbol{x}_n, \boldsymbol{\theta}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1m} \\ \vdots \\ \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nm} \end{bmatrix} \tag{2-29}$$

Thus the vector representation of the model becomes:

$$\boldsymbol{y}_{vec} = \boldsymbol{\eta_{vec}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon_{vec}} \tag{2-30}$$

For these systems the normality assumption takes the form $\boldsymbol{\varepsilon}^{(i)} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, indicating that the noise associated with responses at run i $\left(\boldsymbol{\varepsilon}^{(i)}\right)$ has a *Multivariate Normal* distribution within each run. The within-run noise covariance matrix for run i $(\boldsymbol{\Sigma_i})$ is defined as:

$$\boldsymbol{\Sigma_i} = \mathrm{E}\left\{\boldsymbol{\varepsilon}^{(i)\prime}\boldsymbol{\varepsilon}^{(i)}\right\} \tag{2-31}$$

where E{} is the expected value of the function within the brackets. If the noise for each run is considered to be independent from the noise in other runs, for the vector representation of the model (equation 2-30) the within-run noise covariance matrix can be estimated by:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{e}^{(i)\prime}\boldsymbol{e}^{(i)}\right) \tag{2-32}$$

where $\boldsymbol{e}^{(i)}$ are the residuals for run i. Subsequently, the estimated overall noise covariance matrix $\widehat{\boldsymbol{\Omega}}$ consists of a block diagonal matrix of $\widehat{\boldsymbol{\Sigma}}$ from all the runs.

Multi-response parameter estimation typically requires iterative methods which are nonlinear optimization problems even if the expectation function is linear with respect to the parameters. This inherent nonlinearity arises from the estimation of noise covariance matrix and accounting for the co-dependencies between different types of responses (Bates and Watts, 1987); (Seber and Wild, 1989); (Guay, 1995). As a result, the estimation and inference methods based on linear approximation might lead to inaccurate outcomes. There are two common methods for parameter estimation in multi-response models, namely the Generalized Least Squares (GLS) and the Determinant Criterion (DC) (Seber and Wild, 1989); (Box and Draper, 1965).

The GLS estimation method was first proposed by Zellner (1962) for linear models and later extended to nonlinear models by Gallant (1975). This estimation method arises naturally from a conditional likelihood formulation under certain assumptions and can be considered as an extension of the Weighted Least Squares (WLS) method to multi-response models (Gallant, 1975). In this method parameter estimates are the values that minimize the SSE function $\boldsymbol{S}(\boldsymbol{\theta})$ defined as (Seber and Wild, 1989):

$$S(\boldsymbol{\theta}) = \left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})) \quad (2\text{-}33)$$

In order to get reliable parameter estimates from this method, there should be a sufficient number of runs. This means that the number of runs should be larger than the number of parameters and greater than/or equal to the number of responses (i.e., $n > p$ and $n \geq m$) because otherwise the estimation problem will run into singularity issues (Seber and Wild, 1989); (Kang and Bates, 1990).

The likelihood function for the vector representation of the model (equation 2-30) under the assumption of *MVN* distribution for noise with zero mean and the overall noise covariance matrix $\boldsymbol{\Omega}$ can be written as (Seber and Wild, 1989):

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right)'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right) \quad (2\text{-}34)$$

Comparing equations 2-33 and 2-34 reveals that if the noise terms have a $MVN(\mathbf{0}, \boldsymbol{\Omega})$ distribution with known $\boldsymbol{\Omega}$, the estimations from both GLS and MLE methods lead to the same results.

Very often in chemical or biological models, $\boldsymbol{\Omega}$ is not known and its estimation process can become a challenge in using the GLS method. Co-dependencies of the noise terms might exist between the experimental runs (causing *between-run* covariance), as well as among multiple responses within each run (leading to *within-run* covariance). Usually between-run covariances are assumed to be negligible, but it is wise to keep the within-run covariance terms in the model in order to take advantage of the information provided by the noise covariance matrix (Kang and Bates, 1990); (Soo and Bates, 1996).

Chapter 3 of this thesis is devoted to the GLS-based profiling method where estimation of $\boldsymbol{\Omega}$ and the assumptions involved are fully described and a method for this estimation process is proposed. At this point, no particular assumptions are made about the within-run or between-run covariance (i.e., about whether the noise terms are co-dependent either across or between runs). The form of the GLS objective function can be derived using a likelihood approach in which the noise is assumed to follow an *MVN* distribution. Under this assumption, the least squares function with a quadratic form can be used as the objective function.

Another commonly used approach for parameter estimation in multi-response models is the Determinant Criterion (DC) which was first introduced by Box and Draper (1965). Considering the matrix representation of the multi-response problem with p parameters, m responses and n experimental runs $\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ (equation 2-22), the residual matrix $\boldsymbol{Z}(\boldsymbol{\theta})$ can be obtained from:

$$\boldsymbol{Z}(\boldsymbol{\theta}) = \boldsymbol{Y} - \boldsymbol{\eta}(\boldsymbol{\theta}) \tag{2-35}$$

In the DC approach, the parameter estimates are the values of parameters that minimize the Determinant function $D(\boldsymbol{\theta})$ defined as (Box and Draper, 1965):

$$D(\boldsymbol{\theta}) = |\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})| \tag{2-36}$$

In order to obtain proper results from this approach, the number of experiments should be greater than or equal to the number of responses $(n \geq m)$ since otherwise the residual matrix will be rank deficient. The number of runs should also be larger than the number of parameters $(n > p)$ to avoid zero values for determinant by fitting one response perfectly regardless of the remaining responses (Bates and Watts, 1985). Using this approach, the DC-based average noise variance estimate (s'^2) can be obtained from (Soo and Bates, 1996):

$$s'^2 = \frac{D(\widehat{\boldsymbol{\theta}}_{DC})}{(n-p)} \tag{2-37}$$

where $\widehat{\boldsymbol{\theta}}_{DC}$ is the vector of DC-optimal parameter estimates. The overall noise covariance matrix $\boldsymbol{\Omega}$ is implicitly estimated in the DC approach by:

$$\widehat{\boldsymbol{\Omega}}_{\boldsymbol{DC}} = \frac{1}{n}\boldsymbol{Z}\left(\widehat{\boldsymbol{\theta}}_{DC}\right)'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}_{DC}) \tag{2-38}$$

Although the DC method seems to be more straightforward than GLS for parameter inference due to implicit calculation of the noise covariance matrix, the Determinant

function will be a higher-order polynomial in the nonlinearly parameterized model, leading to computations that can be very demanding, with a higher potential of convergence to local minima in DC than GLS (Bates and Watts, 1987); (Kang and Bates, 1990). These problems become more pronounced with small data sets and in most cases as the number of experimental runs increases, more reliable DC inference results are obtained (Routray and Deo, 2005). Therefore, in spite of a relatively complicated procedure for estimating the noise covariance matrix in GLS, this method is widely used for parameter inference.

The DC approach can be derived using a likelihood approach, under the assumption that the noise is *MVN* identically distributed between the runs. Within a given run, the noise need not be independent and the DC approach will account for covariance between responses. The DC approach can be derived by concentrating the likelihood, i.e., by estimating the within-run covariance as a function of the parameters, substituting this into the likelihood function, and then obtaining the expression for maximizing the likelihood function. The DC approach is discussed in detail in Chapter 4.

Another class of multi-response models encountered frequently in Econometrics are the Seemingly Unrelated Regression (SUR) models (also referred to as 'classical multivariate linear regression models') (Zellner, 1962); (Beauchamp and Cornell, 1966). In an SUR model, the dependence of each response on the regressor is parameterized independently of the other responses, i.e., no response shares parameters in common with the other responses. SUR models are a multivariate generalization of the univariate linear regression and most of the statistical inference methods for such models are direct extensions of those for the univariate linear models and provide exact results (Kang and

Bates, 1990). The coupling between responses is introduced by covariance between the noise that appears in each response. SUR models are always nonlinear even if their parameterization is linear due to covariance nuisance parameters. The GLS method was initially used for parameter estimation and inference in SUR models by Zellner (1962). A linear SUR multi-response model can be defined as (Zellner, 1962):

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{e} \tag{2-39}$$

with **B** being a $p \times m$ parameter matrix:

$$\boldsymbol{B} = \left[b_{kj}\right]; (k = 1,2, \dots, p,\ j = 1,2, \dots, m) \tag{2-40}$$

Several researchers have confirmed that for linear SUR multi-response models with known noise covariance matrix, even though the maximum likelihood function contains the noise covariance matrix, the corresponding maximum likelihood parameter estimates are identical to the least squares estimates from the separate maximum likelihood method. In other words the following objective function:

$$LL(\boldsymbol{B}) = -\frac{1}{2} tr\{(\boldsymbol{Y} - \boldsymbol{XB})\boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{XB})'\} \tag{2-41}$$

is incapable of taking advantage of the information provided by the noise covariance matrix (Lee and Liu, 2012); (Rothman et al., 2010); (Breiman and Friedman, 1997). Therefore, in order to improve parameter estimation for these types of models they have suggested using some form of penalized log-likelihood functions instead. For example, Lee and Liu (2012) have proposed a Plug-in Joint Weighted LASSO (PWL) (where LASSO stands for Least Absolute Shrinkage and Selection Operator) and Rothman et al. (2010) have suggested another method known as Multivariate Regression with Covariance Estimation (MRCE) for such cases. These techniques are advantageous for SUR models and are discussed more in Chapter 3. However, this type of model where

each expectation function is individually parameterized is not very common in chemical and biological processes. For the general multi-response model considered throughout this research, the SSE function is sensitive to $\boldsymbol{\Omega}$ and there is no need to add penalty terms to the likelihood function to include the impact of the noise covariance terms.

Following the estimation process, parameter inference for multi-response models can be achieved by means of the GLS or DC methods. Using the GLS approach, the approximate $100(1-\alpha)$ standard marginal inference interval for parameter θ_k is given by (Seber and Wild, 1989):

$$\left[\hat{\theta}_k - t_{n-p;\alpha/2}\sqrt{\left\{\left(\boldsymbol{V}(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V}(\hat{\boldsymbol{\theta}})\right)^{-1}\right\}_{kk}}, \hat{\theta}_k + t_{n-p;\alpha/2}\sqrt{\left\{\left(\boldsymbol{V}(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V}(\hat{\boldsymbol{\theta}})\right)^{-1}\right\}_{kk}}\right] \quad (2\text{-}42)$$

where $\hat{\theta}_k$ is the least squares parameter estimate for θ_k and $\boldsymbol{V}(\hat{\boldsymbol{\theta}})$ is the sensitivity matrix evaluated at the least squares parameter estimates $\left(\boldsymbol{V}(\hat{\boldsymbol{\theta}}) = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)$. The degrees of freedom associated with the *t* distribution in this equation and the *F* distribution that follows (equation 2-43) are conventionally considered to be $n-p$. The subscript kk refers to the element on the k^{th} row and k^{th} column of the matrix, i.e., the k^{th} diagonal element.

Subsequently, the approximate $100(1-\alpha)$ standard joint inference region for parameter estimates can be constructed from (Seber and Wild, 1989):

$$(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\boldsymbol{V}(\hat{\boldsymbol{\theta}})'\boldsymbol{V}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \leq ps^2F_{p,n-p;\alpha} \quad (2\text{-}43)$$

In this equation s^2 is the estimated average noise variance which can be obtained from $s^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n-p}$ (equation 2-16).

Using the Determinant approach, an approximate $100(1-\alpha)$ standard marginal inference interval for parameter estimate $\hat{\theta}_k$ can be obtained from (Bates and Watts, 1988):

$$\hat{\theta}_k - t_{n-p;\alpha/2}s'\sqrt{2\{\mathbf{H}^{-1}\}_{kk}}, \hat{\theta}_k + t_{n-p;\alpha/2}s'\sqrt{2\{\mathbf{H}^{-1}\}_{kk}} \quad (2\text{-}44)$$

where $\mathbf{H}$ is the Hessian of $|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|$ evaluated at $\widehat{\boldsymbol{\theta}}$. The approximate $100(1-\alpha)$ joint inference region would be (Bates and Watts, 1988):

$$\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)'\frac{\mathbf{H}}{2}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right) \leq ps'^2F_{p,n-p;\alpha} \quad (2\text{-}45)$$

The Determinant function can also be directly used to get more accurate joint inference intervals (Beale, 1960); (Polic et al., 2004):

$$|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})| - \left|\boldsymbol{Z}\left(\widehat{\boldsymbol{\theta}}\right)'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})\right| \leq ps'^2F_{p,n-p;\alpha} \quad (2\text{-}46)$$

An extensive literature review on the parameter inference in multi-response systems is provided in section 2.7.

As mentioned before, nonlinearity is inherent in multi-response estimation and it affects not only parameter estimation but also parameter inference. As a result, the standard inference methods mentioned above are sometimes unreliable and inefficient. These issues have led to the application of profiling for generating more reliable parameter inference intervals. This method is introduced in the next section and its application for inference is discussed in the following sections.

2.4 Statistical Profiling

Due to the inadequacies of linearization-based approximation methods, Bates and Watts (1988) introduced statistical profiling as an alternative method for parameter inference.

This method consists of a set of graphical techniques to generate inference results and likelihood intervals for parameters and functions of them by studying the likelihood function. Over the years this method has been pursued by a number of researchers and now it is by no means restricted to the ideas initially proposed. Further work on the subject has led to generalized profiling methods which allow application of profiling to functions of the parameters (e.g., (Chen and Jennrich, 1996); (Quinn et al., 2000); (Qi and Zhao, 2010); (Gerhard, 2010)).

Profiling was initially defined by Bates and Watts (1988) for single-response models in terms of the sum of squares function. Other researchers, notably Chen and Jennrich (1996) defined a more general form of profiling in terms of likelihood functions. The presentation that follows uses the approach described by Bates and Watts (1988). Additional comments about the likelihood treatment are provided later in the section. One of the major purposes of this type of profiling is to obtain inference results from a meaningful description of the sum of squares surface. Profiling is performed about the location of the least squares parameter estimates ($\widehat{\boldsymbol{\theta}}$) and examines the value of the corresponding SSE function and the behavior of this function near the parameter estimates $\widehat{\boldsymbol{\theta}}$. In order to achieve this, each parameter is fixed at certain values around its least squares estimate and all the other parameters are estimated from the minimization of the SSE function conditional on the fixed value for the profiling parameter. The progression of the deviation of the corresponding SSE values from $S(\widehat{\boldsymbol{\theta}})$ provides a useful tool for developing likelihood intervals (Watts, 1994).

The graphical outcomes of this method, such as profile *t* plots and profile traces, contain useful information on the nonlinearity of the model with respect to the parameters, the

behavior of the profiling function in the neighborhood of the least squares parameter estimates, and the co-dependencies of parameter estimates corresponding to the form of the expectation function, relevant experimental design and the dataset. This information can be applied to developing models and inference intervals (Watts, 1994).

Unlike the standard Wald inference intervals that are constructed around the least squares parameter estimates (by adding and subtracting a specific constant times the approximate standard error of the parameter estimates), the profile-based intervals can be asymmetrical which can offer more representative results for nonlinearly parameterized models (Gerhard, 2010).

For a typical single-response model with *p* parameters and *n* experimental runs described by $\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ (equation 2-1), profiling can be done for each parameter (θ_k) by defining the profile *t* function $\tau(\theta_k)$ in terms of the Sum of Squared Errors (SSE) function as follows (Bates and Watts, 1988):

$$\tau(\theta_k) = sign\left(\theta_k - \hat{\theta}_k\right)\sqrt{\frac{\tilde{S}(\theta_k) - S(\hat{\boldsymbol{\theta}})}{s^2}} \qquad (2\text{-}47)$$

where θ_k is the profiling parameter and $\hat{\theta}_k$ is the least squares estimate of it. $S(\hat{\boldsymbol{\theta}})$ represents the value of the SSE function at the least squares parameter values and $\tilde{S}(\theta_k)$ is the value of the SSE function at the least squares estimates of the parameters conditional on a fixed value of θ_k. $\tilde{S}(\theta_k)$ can also be written as $S(\theta_k, \tilde{\boldsymbol{\theta}}_{-k}(\theta_k))$ where $\boldsymbol{\theta}_{-k}$ represents the vector containing the model parameters except for the profiling parameter θ_k. As mentioned before, the denominator s^2 is an estimate of the noise variance which can be obtained from $s^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n-p}$ (equation 2-16). An algorithm for generating the profile *t* function and the graphical profiling outputs is summarized in

Table 2-1. Please note that in this thesis the ˆ symbol denotes the least squares estimates, while the ˜ symbol denotes a conditional estimate based on a specific value of the profiling parameter.

Table 2-1: An algorithm for obtaining profile *t* plots and profile traces

1	Start with initial guesses for parameters and generate the least squares parameter estimates ($\widehat{\boldsymbol{\theta}}$).
2	Calculate the value of the SSE function at the least squares parameter estimates ($S(\widehat{\boldsymbol{\theta}})$).
3	Choose one of the model parameters (θ_k) as the profiling parameter and fix it at specific values (preferably around its least squares estimate to study that neighborhood).
4	Determine the least squares estimates for all the other parameters ($\widetilde{\boldsymbol{\theta}}_{-k}$) conditional on the fixed values of the profiling parameter θ_k. This can be done by means of an iterative method such as Gauss-Newton using the parameter estimates from the previous iteration each time until a convergence criterion is met.
5	Once the conditional parameter estimates are obtained from step 4, determine the corresponding value of the SSE function ($\tilde{S}(\theta_k)$).
6	Calculate $\tau(\theta_k)$ from $\tau(\theta_k) = sign\left(\theta_k - \hat{\theta}_k\right)\sqrt{\frac{(\tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}))}{s^2}}$ (equation 2-47).
7	Plot $\tau(\theta_k)$ versus θ_k to obtain profile *t* plots to investigate the behavior of profile *t* function in the neighborhood of the least squares parameter estimates.
8	Plot elements of $\widetilde{\boldsymbol{\theta}}_{-k}$ versus θ_k (or the studentized equivalents of them) to obtain profile traces for exploring the co-dependencies among the parameter estimates.

A plot of $\tau(\theta_k)$ versus θ_k is a profile *t* plot which provides useful information on the nonlinearity of the estimation problem as well as parameter inference. For a linear model, the SSE function is a quadratic function of the parameters and therefore the resulting profile *t* plot is a straight line. Otherwise, the plot is curved and the amount of curvature provides a measure of nonlinearity of the model. Since one-to-one parameter transformations do not impact the SSE function, the parameters can be conveniently transformed to the Studentized Parameter $\delta(\theta_k)$ by (Bates and Watts, 1988):

$$\delta(\theta_k) = (\theta_k - \hat{\theta}_k)/se(\hat{\theta}_k) \quad (2\text{-}48)$$

where $se(\hat{\theta}_k)$ is the estimated standard error of the parameter estimate $\hat{\theta}_k$ and can be obtained from the sensitivity matrix $\boldsymbol{V}(\boldsymbol{\theta})$ and the average noise variance estimate (s^2) by means of $se(\hat{\theta}_k) = \sqrt{\{(\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{V}(\boldsymbol{\theta}))^{-1}\}_{kk}s^2}$ (equation 2-15).

For a linear model the plot of $\tau(\theta_k)$ versus $\delta(\theta_k)$ is a straight line with unit slope that passes through the origin, whereas for a nonlinear model this plot is curved due to the non-quadratic SSE function. Departure of the $\tau(\theta_k)$ curve from the forty five degree reference line (with unit slope) reveals nonlinearity in the parameters as well.

Pairwise plots of any of the remaining parameters (elements of $\widetilde{\boldsymbol{\theta}}_{-k}$) versus θ_k and vice versa are called profile traces, and provide information on co-dependencies between the estimates of two parameters caused by model structure and experimental design. If the pairwise curves for two parameters versus each other are perpendicular at their intersection, the two parameter estimates are not locally co-dependent. Otherwise, the sharpness of the angle at the intersection provides a measure of the co-dependency between the two parameter estimates (Bates and Watts, 1988); (Bates and Watts, 1991); (Routray and Deo, 2005). For a more thorough analysis of the parameter co-

dependencies, the profile traces can be plotted on the same graph of likelihood contours for two parameters to show how the co-dependencies between parameters translates into profile traces. If the parameters are independent, contours are wide and profile traces are perpendicular. If there are substantial co-dependencies, the contours are long and thin and the profile traces fall very close to each other. Elliptical likelihood contours indicate a linear model and thus the corresponding profile traces should be straight lines (Bates and Watts, 1988). It should be noted that the joint likelihood inference regions can also be developed for pairs of parameters using profile traces (see for example, (Bates and Watts, 1988)).

The applicability of profile traces in practice depends on the complexity of the model and the purpose of profiling. If the model has fewer parameters or the modeler is interested in the co-dependencies among a particular subset of parameters, profile traces could provide a good insight. For complex models with a large number of parameters generating and summarizing all the profile trace plots might not be very efficient given that for a model with p parameters there would be $p(p-1)$ profile trace curves. However, it should be noted that the information required to generate the profile traces is a by-product of the profile *t* calculations and readily available. No additional calculations are required for profile traces. Overall, profile *t* plots are very useful and are the primary focus of this research.

In the linear case with *IID Normal* noise, $\tau(\theta_k)$ has a Student's *t* distribution with $n-p$ degrees of freedom (t_{n-p}). An approximate $100(1-\alpha)$ likelihood interval for θ_k consists of parameter values corresponding to the following:

$$-t_{n-p;\alpha/2} \leq \tau(\theta_k) \leq t_{n-p;\alpha/2} \qquad (2\text{-}49)$$

These profile-based parameter inference intervals are exact for linear systems but approximate for nonlinear systems due to the different distribution of the profiling objective function resulting from the nonlinearity of the parameterization. Other elements of approximation for inference regions are associated with the significance level (α), and also arise from the estimated variance of the parameter estimates (see for example, (Fraser, 1993); (Reid and Fraser, 2010)). Unlike the standard Wald inference intervals, the profile-based intervals can be asymmetrical which might provide a more representative summary reflecting the inference results for some nonlinear models (Gerhard, 2010).

2.5 Applications of Profiling

Profile plots and traces are valuable graphical tools for revealing the extent of nonlinearity of the model and the co-dependencies among the parameters. If the plots show negligible nonlinearity, linear approximations can be used knowing that the results will be fairly reliable. On the other hand, if the profile plots indicate higher degrees of nonlinearity and co-dependency between parameters, it might be wise to use parameter transformations before applying linearization-based inference as long as the new parameters behave more linearly, or consider nonlinear methods (Watts, 1994). Thus, profiling can assist effective re-parameterizations and help visualize the effect of certain modifications to the model (e.g., taking more variables into account) on parameter co-dependencies or the sum of squares surface (Brooks et al., 1994). In brief, profiling enables the researcher to avoid re-parameterization and work with the original model parameterization which is usually more meaningful, but if one chooses to use model

transformations they can still benefit from the fact that profile-based inference regions remain invariant through model transformations.

Furthermore, profile traces can be used to reveal the co-dependencies between the parameter estimates. Profile traces that fall very close to each other or even overlap, suggest a strong co-dependency between the parameter estimates and thus it is better to work with joint inference regions in such cases (Bates and Watts, 1988).

Another area where profiling has been successfully applied is sensitivity analysis and experimental design. Profile plots are used in experimental design by revealing which parameter estimates lack proper data and how changes in the number of experimental runs or observation and the respective degrees of freedom could lead to more appropriate confidence intervals (Watts, 1994); (Raue et al., 2011).

Sulieman et al. (2001) investigated these applications by studying profile plots to draw conclusions on the nonlinearity of the system and the amount of co-dependency between the parameter estimates. They introduced the Profile-based Sensitivity Coefficient (PSC) to find the intervals where each parameter has the most impact on the model, taking into account co-dependencies between the impacts of different parameters on the responses. For a single-response model using the least squares method, PSC is defined as (Sulieman et al., 2001):

$$PCS_k(x_0) = \frac{\partial f(x_0,\boldsymbol{\theta})}{\partial \theta_k} - \frac{\partial f(x_0,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-k}} \left(\frac{\partial^2 S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-k} \partial \boldsymbol{\theta}_{-k}'} \right)^{-1} \frac{\partial^2 S(\boldsymbol{\theta})}{\partial \theta_k \partial \boldsymbol{\theta}_{-k}} \Big|_{\widetilde{\theta}_{-k}} \tag{2-50}$$

Note that θ_k is the profiling parameter and $\widetilde{\boldsymbol{\theta}}_{-k}$ contains all the other parameters estimated conditionally based on θ_k. The first term in this equation is the conventional marginal sensitivity coefficient, while the second term represents a correction that accounts for the co-dependencies. Later on they extended their method for multi-response

systems using the Determinant Criterion, and established the following definition for PCS (Sulieman et al., 2004):

$$PCS_k(x_0) = \frac{\partial \boldsymbol{G}_0'}{\partial \theta_k} + \frac{\partial \boldsymbol{G}_0'}{\partial \boldsymbol{\theta}_{-k}}\Big|_{\tilde{\theta}_{-k}} \frac{\partial \tilde{\boldsymbol{\theta}}_{-k}}{\partial \theta_k} \tag{2-51}$$

where $\boldsymbol{G}_0$ is the vector of response variables evaluated at x_0. Plots of the absolute value of $PCS_k(x_0)$ versus θ_k have maximum values where the models are most sensitive to θ_k. The results contribute to experimental design by suggesting more measurements where required.

As mentioned before, the standard confidence intervals provided by linearization methods (Wald intervals) may not be very accurate for nonlinear models. Also the linearization-based, ellipsoidal uncertainty regions are inadequate in many nonlinear cases. This becomes a greater issue as models deviate more from linearity or the parameter values become closer to stability/convertibility boundaries. An example of the latter case is when the numerator and/or the denominator of the time series transfer function have roots close to unit circle. In such cases profiling methods have been used to provide more reliable likelihood regions (Quinn et al., 2005). A profile likelihood contains valuable information around the estimated values for parameters, so even though construction of likelihood intervals by profiling might require more computations than the linearization approach, it is still beneficial (van Ewijk and Hoekstra, 1994); (Gerhard, 2010). Overall, I believe that applying profiling for parameter inference in nonlinear multi-response models is a very advantageous method with reliable outcomes. Thus, one the most important objectives of this research is to investigate profile plots and their distributions in order to extend this particular application of profiling.

2.6 Profiling for Multi-Response Models

As mentioned previously, the idea of profiling was proposed by Bates and Watts (1988) for single-response models and continued by several researchers. However, most of these studies targeted single-response models (e.g., (Chen and Jennrich, 1996)). Parameter likelihood intervals for multi-response systems have been explored by a number of researchers but not in a profiling context (e.g., (Zellner, 1962); (Bard, 1974); (Bates and Watts, 1985); (Kang and Bates, 1990); (Oxby et al., 2003)). Therefore, the main objective of this research is to find appropriate ways to extend the application of profiling for parameter inference to multi-response models. This can be done based on the common estimation approaches for multi-response models namely, the GLS and DC methods that were explained in section 2.3.

Soo and Bates (1996) built on the profiling concept of Bates and Watts by means of the DC approach for parameter estimation. They substituted the SSE function in the single-response profile *t* function with the Determinant function and came up with the following DC-based profile *t* function for multi-response models (Soo and Bates, 1996):

$$\tau_{DC}(\theta_k) = sign\left(\theta_k - \hat{\theta}_k\right)\frac{\sqrt{\frac{\tilde{D}(\theta_k) - D(\hat{\boldsymbol{\theta}})}{}}}{\sqrt{\frac{D(\hat{\boldsymbol{\theta}})}{\nu}}} \qquad (2\text{-}52)$$

In this equation ν is equal to $n - p$ for the multi-response models, or is calculated from the spline method and the number of knots that is used for approximating the model functions for nonparametric regression.

Raue et al. (2009) investigated profiling in multi-response systems by means of the GLS method, focusing on the cases with known diagonal noise covariance matrix. Following the previous works on parameter likelihood intervals (such as (Venzon and Moolgavkar,

1988); (Murphy and van der Vaart, 2000)), they used the following likelihood function $L(\boldsymbol{\theta})$ to estimate the parameters for a system with p parameters, m responses and n experimental runs (Raue et al., 2009):

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{{\sigma_j}^2} (y_j(x_i, \boldsymbol{\theta}) - f_j(x_i, \boldsymbol{\theta}))^2 \quad (2\text{-}53)$$

$$\widehat{\boldsymbol{\theta}} = \min_{\theta}[L(\boldsymbol{\theta})] \quad (2\text{-}54)$$

In equation 2-53, ${\sigma_j}^2$ are the measurement error variances for response j which are assumed to be known. Thus, noise terms are distributed as $\varepsilon_j \sim N(0, {\sigma_j}^2)$ with the assumption of *IID* between-run noise. The profile likelihood function ($L_{PL}(\theta_k)$) is constructed by minimizing the following objective function at fixed values of θ_k:

$$L_{PL}(\theta_k) = \min_{\theta}[L(\boldsymbol{\theta})] \quad (2\text{-}55)$$

A more detailed review of these references and other works regarding parameter inference in multi-response models is provided in the next section.

In this research the concept of GLS-based profiling has been revisited for a more general and realistic case which involves an unknown noise covariance matrix. In this thesis profile likelihood plots are used rather than the traditional profile *t* plots. Profile likelihood plots are sections of the likelihood function (without the square root that is used in profile *t* function), and summarize the change in the likelihood (sum of squared residuals) as the profiling parameter moves away from its least squares estimate. Profile likelihood plots for linear models have a quadratic form which might be more familiar to researchers in this field and make it easier to identify nonlinear models based on the deviation of the profile likelihood curve from the symmetric parabola. Furthermore, since profile likelihood functions do not contain the square root of the SSE function, they are potentially more robust and could avoid some of the numerical issues in the calculation of

the profiling function especially in instances where there are local minima and the likelihood function is not monotonic.

Looking back at the definition of profile *t* function for single-response models (equation 2-47), one way to extend this concept to multi-response models is to insert the values of the SSE function and the estimated average noise variance for the multi-response model in this equation. These values can be determined from the GLS estimation approach and substituted into equation 2-47. Note that as previously mentioned, the SSE function for a multi-response model is obtained from $S(\boldsymbol{\theta}) = \left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y_{vec}} - \boldsymbol{\eta}_{vec}(\boldsymbol{\theta}))$ (equation 2-33) where $\boldsymbol{\Omega}$ is the overall noise covariance matrix (Seber and Wild, 1989). Since $\boldsymbol{\Omega}$ is already incorporated into the SSE function, further division by the estimated average noise variance (s^2) will not be necessary.

Definition: The "**GLS-based profile likelihood function**" $PL_S(\theta_k)$ takes the following form:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \qquad (2\text{-}56)$$

Also based on the DC-based profile *t* function defined by Soo and Bates (1996) (equation 2-52), the following definition is developed.

Definition: The "**DC-based profile likelihood function**" $PL_D(\theta_k)$ is defined as:

$$PL_D(\theta_k) = \frac{\tilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}})}{(s')^2} \qquad (2\text{-}57)$$

As before, the Determinant function $D(\boldsymbol{\theta})$ can be obtained from $D(\boldsymbol{\theta}) = |\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|$ (equation 2-36) and the estimated average noise variance (s'^2) can be obtained from $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}})}{(n-p)}$ (equation 2-37).

Plots of $PL_S(\theta_k)$ or $PL_D(\theta_k)$ versus the profiling parameter θ_k are called the profile likelihood plots and similar to the profile *t* plots, they contain a great amount of information about the model that can be used in several applications such as parameter inference and parameter estimability analysis.

2.7 Literature Review on Parameter Inference in Multi-Response Models

For parameter inference purposes in multi-response systems, it is important to determine the distribution of the profile likelihood function and the associated degrees of freedom. An extensive amount of work can be found in the literature that addresses the issue of generating parameter likelihood intervals. However, most of these studies target single-response models (e.g., (Chen and Jennrich, 1996) and (Beale, 1960)). The modest number of research papers that focus on parameter likelihood intervals in multi-response models mainly vary based on the choice of estimation method (GLS or DC), and whether the noise variance terms are known or not (e.g., (Box and Draper, 1965); (Bates and Watts, 1987); (Kang and Bates, 1990)). Estimation of the noise covariance matrix is particularly important in multi-response models because the distribution of the estimated variance/covariance terms and the assumptions involved impact the distribution of the likelihood and SSE functions. Different approaches for dealing with the covariance matrix in GLS inference are reviewed and discussed in the following paragraphs. The literature review indicates that there is a continuing need for improvements in this field.

In the discussion that follows, likelihood inference and profiling approaches for single-response nonlinear models are first reviewed. Inference techniques for linear and nonlinear multi-response estimation problems are then discussed, summarizing the work

of Draper and Smith, and Bates and Watts, in particular. The discussion then returns to how single-response profiling methods have been extended for the DC case to provide a profiling approach for nonlinear multi response models.

Beale (1960) investigated the likelihood regions for nonlinear single-response cases with p parameters and concluded that the $100(1-\alpha)$ likelihood region for parameters could be provided by:

$$S(\theta) - S(\hat{\theta}) \leq ps^2 F_{p,\nu;\alpha} \tag{2-58}$$

with $F_{p,\nu;\alpha}$ being the upper α quantile of the F distribution with p and ν degrees of freedom. ν is associated with the noise variance estimate (s^2). Therefore, if the noise variance is unknown and estimated by $s^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n-p}$ (equation 2-16), then the degrees of freedom is equal to $\nu = n - p$. Otherwise, if the noise variance σ^2 is known, the likelihood region can be obtained from:

$$S(\theta) - S(\hat{\theta}) \leq \sigma^2 \chi^2_{p;\alpha} \tag{2-59}$$

in which $\chi^2_{p;\alpha}$ represents the upper α quantile of the χ^2 distribution with p degrees of freedom.

The profiling method of Bates and Watts for parameters (explained in section 2.4) was further developed by Chen and Jennrich (1996) as the "generalized profiling" method. Generalized profiling is an extension to profiling that sets a function of one or more parameters at a specific value ($g(\boldsymbol{\theta}) = c$) rather than fixing just one parameter value. Based on this generalization, profiling for each parameter θ_k would be a special case of the generalized profiling in which $g(\boldsymbol{\theta}) = \theta_k$ (Chen and Jennrich, 1996). The concept of using profiling for building likelihood intervals for single-response models was first

introduced by Bates and Watts (1988) and developed further by Chen and Jennrich (1996).

The likelihood ratio (LR) with a *Standard Normal* distribution had been previously used for inference but not in a profiling context (Barndorff-Nielsen, 1986). If $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{c}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ conditional on $g(\boldsymbol{\theta}) = c$, the LR is defined as:

$$\mathrm{LR}(c) = \frac{L(\widehat{\boldsymbol{\theta}}_{\boldsymbol{c}})}{L(\widehat{\boldsymbol{\theta}})} \tag{2-60}$$

The likelihood interval for each parameter (θ_k) of a single-response model based on the likelihood ratio can be obtained from the following equation with $g(\boldsymbol{\theta}) = \theta_k$ (Chen and Jennrich, 1996):

$$I = \{c: -2log\mathrm{LR}(c) \leq \chi^2_{1;\alpha}\} \tag{2-61}$$

Chen and Jennrich (1996) defined the Signed Root Deviance (SRD) function as:

$$z^*(c) = sign\big(c - g(\widehat{\boldsymbol{\theta}})\big)\sqrt{-2ln\left(\frac{L(\boldsymbol{\theta})}{L(\widehat{\boldsymbol{\theta}})}\right)} \tag{2-62}$$

which is a generalization of the profile *t* function of Bates and Watts (1988). They stated that this function has a *Standard Normal* distribution when the noise is *IID Normal*, leading to the following likelihood intervals for parameters:

$$I = \{c: -z_{\alpha/2} \leq z^*(c) \leq z_{\alpha/2}\} \tag{2-63}$$

with $z_{\alpha/2}$ being the upper $\alpha/2$ quantile of the *Standard Normal* distribution. Like Bates and Watts, they estimated noise variances by using MSE from the nonlinear regression fit. When $g(\boldsymbol{\theta}) = \theta_k$, the results of equations 2-62 and 2-63 correspond to the profile *t* function by Bates and Watts (1988). A plot of $z^*(c)$ versus c is called the Signed Root Deviance Profiling (SRDP) plot. SRDP is useful in constructing LR intervals for $g(\boldsymbol{\theta})$

and providing the standard linear approximation intervals (or Wald intervals). The standard and LR intervals can be compared by means of SRDP plots in order to find appropriate transformations of $g(\boldsymbol{\theta})$ that bring these two values closer to each other. This process can be viewed as a series of optimizations for constructing the best profile (Chen and Jennrich, 1996). As discussed in detail by Quinn et al. (2000) this method has some restrictions regarding the types of functions it can handle and might fail to produce accurate results.

In a more recent work, Quinn et al. (2005) continued the work of Chen and Jennrich (1996) on generalized profiling, focusing mostly on dynamic models. They applied the following equation to single-response cases to build likelihood intervals for the case of $g(\boldsymbol{\theta}) = c$ (Quinn et al., 2005):

$$-2ln\left(\frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}\right) \leq F_{q,n-p;\alpha} \tag{2-64}$$

For this F distribution, q is determined based on the number of constraints that are imposed on the likelihood function. Thus, when the only constraint is $g(\boldsymbol{\theta}) = c$, the distribution used in equation 2-64 becomes $F_{1,n-p;\alpha}$.

Parameter inference was extended to linear multi-response systems and seemingly unrelated regression (SUR) by Zellner (1962) and further developed by Beauchamp and Cornell (1966). Zellner (1962) applied Aitken's generalized least squares method to SUR, and proposed an $F_{q,N-d;\alpha}$ distribution for building the likelihood intervals, with q being the number of restrictions on the system, N the number of observations (runs), and d the number of independent variables. For the vector representation of the model as indicated in equation 2-30, this terminology would translate into $N = nm$ and $d = mp$. Gallant and Holly (1980) built on the Zellner method and used the GLS approach for

parameter estimation and inference by considering a χ^2 distribution with r degrees of freedom for the likelihood function, with r being the length of the expectation function vector ($r = nm$ for the model formulation of equation 2-30) (Gallant and Holly, 1980). Subsequently Bard (1974) used the GLS approach for general nonlinear multi-response models (not necessarily SUR), and implemented a two-step estimation technique to obtain parameter and noise covariance estimates. The estimation procedure iterates between computing GLS-optimal values of the parameter estimates conditional on a fixed covariance matrix, and then updating the covariance matrix estimate conditional on fixed parameter values. They considered an $F_{p,n-p/m;\alpha}$ distribution for obtaining the likelihood intervals, based on the argument that the model consists of n runs with p/m parameters per equation on average (Bard, 1974). This parameter inference method was also discussed by Seber and Wild (1989), who proposed using the following constrained Log-Likelihood ratio statistic (LR_c) with q constraints and a $\chi^2_{q;\alpha}$ or an F distribution depending on the model (Seber and Wild, 1989):

$$LR_c = 2[L(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Sigma}}) - L(\breve{\boldsymbol{\theta}}_{\boldsymbol{c}}, \widehat{\boldsymbol{\Sigma}})] \tag{2-65}$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are the unconstrained maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ respectively, and $\breve{\boldsymbol{\theta}}_{\boldsymbol{c}}$ represents the restricted estimate of $\boldsymbol{\theta}_{\boldsymbol{c}}$ due to the constraints.

Draper and Smith (1981) discussed an extension of parameter inference methods to nonlinear models based on the DC approach using the following approximate expression for $100(1-\alpha)$ joint confidence region (Draper and Smith, 1981):

$$\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)' \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{-1} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \leq pF_{p,\nu;\alpha} \tag{2-66}$$

with $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ being the parameter covariance matrix. ν is associated with $\widehat{\boldsymbol{\Sigma}}_{\theta}$ which in turn is contingent on the distribution of the estimated overall noise covariance matrix $\widehat{\boldsymbol{\Omega}}$. The

covariance matrix of the parameter estimates is determined by a linearization-based approach, similar to that used in GLS. Therefore, if the noise covariance matrix is estimated based on the DC from $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{DC}} = \frac{1}{n}\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}_{DC})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}_{DC})$ (equation 2-38), then $\nu = n$. The works of Draper and Smith (1981) and Box and Draper (1965) were the basis for many researches that followed regarding parameter inference in multi-response models using the DC approach (Oxby, 1997).

Some of the most substantial work on inference using the DC has been done by Bates et al. They used the DC approach with the average noise variance estimated by $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}}_{DC})}{(n-p)}$ (equation 2-37). Using a quadratic approximation to $\left|\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})\right|$, they came up with the following inequality for estimation of joint likelihood regions (Bates and Watts, 1985):

$$|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})| - \left|\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})\right| \leq ps'^2 F_{p,n-p;\alpha} \qquad (2\text{-}67)$$

where $\boldsymbol{Z}(\boldsymbol{\theta})$ is the residual matrix from matrix predictions using some values of $\boldsymbol{\theta}$, and $\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})$ is the residual matrix evaluated at the DC-optimal estimates. They argued that in the neighborhood of the least squares parameter estimates, the Determinant function is controlled by a single linear combination of the responses. Therefore, in spite of having *nm* pieces of information available (as mentioned by Bard (1974)), the objective function is only sensitive to *n* of them. Equation 2-67 has also been used by Polic et al. (2004) but not in a profiling context, where they applied the DC and assigned values to the noise standard deviation terms based on the existing literature and knowledge of the experiments in their case studies (Polic et al., 2004).

Kang and Bates (1990) continued the work of Bates and Watts, examining both the GLS and the DC methods for parameter estimation and inference. They argued that in theory

DC is a more appropriate method for parameter estimation since it stems from Bayesian methods where the parameter estimates are derived from the posterior density function which accounts for the prior knowledge of the estimates. The GLS approach, on the other hand, estimates parameters based on the conditional likelihood function and estimated noise covariance matrix. However, they acknowledged the potential numerical issues and complexity of the calculations in the DC method and proposed a two-step procedure for improving parameter inference in multi-response models. The first step of their analysis consists of conditioning the generalized sum of squares on $\widehat{\boldsymbol{\Omega}}$ as follows (Kang and Bates, 1990):

$$S(\boldsymbol{\theta}) = tr(\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})) \quad (2\text{-}68)$$

where *tr* stands for the trace of the matrix. The second step involves linear approximation of the model functions $\left(\boldsymbol{f}(\boldsymbol{\theta})\right)$ by:

$$\boldsymbol{f}(\boldsymbol{\theta}) \cong \boldsymbol{f}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{V}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \quad (2\text{-}69)$$

The combination of these steps leads to the following equation for estimation of the parameter estimate covariance matrix:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \boldsymbol{V}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V}(\boldsymbol{\theta}) \quad (2\text{-}70)$$

They suggest doing the parameter inference based on this parameter estimate covariance matrix for example by substituting this matrix into equation 2-66 to obtain the joint confidence regions.

Oxby (1997) compared the DC-based inference methods of Bard (1974) and Bates and Watts (1985) from a theoretical point of view and also applied both methods to a linear multi-response example. Based on the outcomes he concluded that $\nu = n - {}^{p}/_{m}$ is a

better choice since $\nu = n - p$ is too conservative and cannot be applied to all cases (Oxby, 1997).

Extensive work regarding profile likelihoods in systems biology has been carried out in recent years by Raue et al. (e.g., (Raue et al., 2009); (Raue et al., 2010); (Raue et al., 2011). They focus on cases with known diagonal noise covariance matrices consisting of elements of ${\sigma_j}^2$ so that the noise terms are assumed to be distributed as $\varepsilon_{ji} \sim N(0, {\sigma_j}^2)$. Their profiling method was explained in section 2.6. As mentioned in that section, they use the likelihood function $L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{{\sigma_j}^2} (y_j(\boldsymbol{x}_i, \boldsymbol{\theta}) - f_j(\boldsymbol{x}_i, \boldsymbol{\theta}))^2$ (equation 2-53) to estimate the parameters. Under the assumption of normally distributed noise with known diagonal noise covariance matrix (ignoring the co-dependencies between the responses), this objective function becomes similar to a weighted-least-squares case. As a result, they have considered a $\chi^2_{1;\alpha}$ distribution for the profile likelihood plots and have used them for parameter inference and consequently for parameter identification purposes.

Apart from the GLS-based and DC-based methods described in this section, there are also Bayesian methods (such as MCMC), and empirical techniques such as Bootstrap that can be used for generating posterior distributions and inference regions (e.g., (Jitjareonchai et al., 2006)). These techniques generate inference regions and intervals directly through empirical construction of the posterior density function for the parameter estimates, typically requiring many repeated parameter estimations (e.g., typically in the order of many thousands). The primary challenges in using these approaches include proper selection of the "tuning" parameters (e.g., burn-in period, acceptance probability, total number of runs). The total number of runs will depend on how smooth an inference

region is required (see for example, (Jitjareonchai et al., 2006)). The major limitation of these approaches is that they can be computationally intensive, and can be sensitive to convergence of the parameter estimation step. In this regard, they share some of the potential limitations of profiling, although in the case of profiling, fewer runs are likely to be required for generating the profile likelihood information, and the estimation procedure in profiling is started from a neighborhood of the overall maximum likelihood/GLS-optimal estimates. With profiling, the number of required runs depends on the length of the interval that is being studied for each parameter and the choice of the values at which the parameter is fixed at. For instance, for the case studies in this thesis the number of runs were as low as 80 and as high as 400 for generating the profile likelihood plots.

Based on the literature review, many of the current parameter inference techniques for multi-response models run into numerical problems or are challenged by practical considerations such as convergence to local minima. These methods often result in approximate parameter inferences and also yield unsatisfactory results even for cases that seem quite simple with linear model functions such as an example explored by Bates and Watts (1987). The approximation of the inference intervals (including those obtained from profiling methods) also arises from determination of the significance level (α) (see for example (Reid and Fraser, 2010)). Thus, there is ongoing research for reducing approximations and obtaining more reliable results.

This review shows that for single-response models, the *F* distribution is considered for the likelihood function, assuming independent normally distributed noise terms with constant variance. However, in multi-response cases, the responses are co-dependent and

as a result the same distributions will not necessarily lead to appropriate results. A $\chi^2_{1;\alpha}$ distribution has been used for GLS-based profile likelihood function for multi-response models with known noise covariance matrix (Raue et al., 2009) and an *F* distribution is suggested for the DC-based likelihood function (Bates and Watts, 1985). However, the degrees of freedom proposed for this *F* distribution can range from $nm - p$ (e.g., (Bard, 1974)) to $n - p$ (e.g., (Bates and Watts, 1985)). Most recent works (e.g., (Oxby, 1997)) have supported and applied one of these choices. Thus, the distribution of the profile likelihood function using either GLS or DC approaches with unknown noise covariance matrix is still not fully resolved. Finding the appropriate distribution is the key to building profile-based parameter likelihood intervals.

2.8 Summary

In this chapter a background on parameter estimation and inference methods in single- and multi-response models was provided, mentioning some of the shortcomings and inadequacies of the existing methods. The method of statistical profiling was presented as an alternative method for parameter inference in multi-response models. Due to the advantages of profiling, one of the main objectives of this research namely extending this application to multi-response models was introduced.

Consequently, a literature review on parameter inference in multi-response models was provided focusing mainly on GLS-based and DC-based methods. Concluding from this literature review, the distribution of the likelihood function and the corresponding degrees of freedom for multi-response models is still not an entirely resolved issue. Finding the appropriate distribution is the key to building profile-based parameter

likelihood intervals. The distribution of the profile likelihood function is discussed in detail in the next chapter and profile-based parameter inference methods are proposed for multi-response models with known and unknown noise covariance matrices.

Chapter 3

Profiling Based on Generalized Least Squares Method

3.1 Introduction

In Chapter 2, parameter inference in nonlinear multi-response models was identified as one of the main objective of this research, with a particular focus on the method of profiling. As discussed earlier, the Generalized Least Squares (GLS) method is widely used for parameter estimation and inference even though there is a relatively complicated procedure for estimating the noise covariance matrix when it is unknown. However as explained in Chapter 2 section 2.6, most of the published research regarding profiling in multi-response systems has been based on the Determinant Criterion (DC) approach (Soo and Bates, 1996). The few profiling papers that have been based on the GLS method have simplified the problem by assuming that the noise covariance matrix is known and by ignoring the co-dependencies between the noise terms (e.g., (Raue et al., 2009); (Raue et al., 2010)).

In this chapter, I present the idea of profiling based on the GLS method for models with unknown noise covariance matrix, focusing particularly on the estimation of this matrix and its impact on the profiling results. The importance of the noise covariance matrix and the involved assumptions are discussed.

Several approaches for parameter estimation and profiling based on the GLS-method are examined in this chapter. These approaches differ mainly by how the noise covariance is estimated and updated during the profiling procedure and whether or not profiling is done based on a fixed noise covariance matrix. Once these approaches are studied, I propose a GLS-based profiling approach which generates the most reliable results. More

importantly, in order to apply the GLS-based profile likelihood function to parameter inference, the distribution and the associated degrees of freedom for this function are carefully studied and determined.

The proposed GLS-based profiling method for parameter inference is applied to several examples to demonstrate the use and interpretation of these techniques in practice. The results from the case studies validate the suitability of this profiling method.

3.2 Importance of the Noise Covariance Matrix Estimation Process

The vector representation of a multi-response model with *p* parameters, *m* responses and *n* experimental runs is as follows:

$$\boldsymbol{y_{vec}} = \boldsymbol{\eta_{vec}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon_{vec}} \tag{3-1}$$

As explained previously in Chapter 2, for this model parameter estimates can be obtained by means of the GLS estimation method by minimizing the SSE function $S(\boldsymbol{\theta})$ defined as (Seber and Wild, 1989):

$$S(\boldsymbol{\theta}) = \left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})) \tag{3-2}$$

where $\boldsymbol{\Omega}$ is the overall noise covariance matrix. Once the parameters are estimated, it is necessary to determine the uncertainty interval for each parameter estimate. The method of statistical profiling, as explained in Chapter 2 section 2.6, can be very useful in this regard for multi-response models. The profile *t* function for single-response models has been defined as (Bates and Watts, 1988):

$$\tau(\theta_k) = sign\left(\theta_k - \hat{\theta}_k\right)\sqrt{\frac{\tilde{S}(\theta_k) - S(\hat{\boldsymbol{\theta}})}{s^2}} \tag{3-3}$$

where θ_k is the profiling parameter and $\hat{\theta}_k$ is the least squares estimate of it. $S(\widehat{\boldsymbol{\theta}})$ represents the SSE function evaluated at the least squares parameter estimates and $\tilde{S}(\theta_k)$ is the value of the SSE function at the least squares parameter estimates conditional on a fixed value of θ_k. The denominator s^2 is the estimated average noise variance which can be obtained from $s^2 = \frac{S(\widehat{\boldsymbol{\theta}})}{n-p}$ (equation 2-16).

It was explained in Chapter 2 that in order to extend profiling from single-response to multi-response models using the GLS approach, equation 3-3 should be modified to accommodate the SSE function of equation 3-2. However, since $\boldsymbol{\Omega}^{-1}$ is already incorporated into the SSE function, further division of the numerator of equation 3-3 by the estimated average noise variance (s^2) will not be necessary to make the GLS-based profile likelihood function. It was also noted in Chapter 2 that working with the profile likelihood function rather than the traditional profile *t* function can be advantageous due to the familiar quadratic form of the profile likelihood function and simpler computations by avoiding the square root of the SSE function.

In Chapter 2, the "**GLS-based profile likelihood function**" $PL_S(\theta_k)$ was defined as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \qquad (3\text{-}4)$$

In this equation $S(\widehat{\boldsymbol{\theta}})$ represents the SSE function evaluated at the least squares parameter estimates and $\tilde{S}(\theta_k)$ is the value of the SSE function at the least squares estimates of parameters conditional on the fixed value of θ_k.

One of the major considerations in extending the least squares parameter estimation methods from single- to multi-response models is the incorporation of the noise covariance matrix. In single-response estimation, under the *IID Normal* assumption for the noise terms, the least squares estimates of the parameters can be computed

independently of the noise variance by minimizing the sum of squares function $S(\boldsymbol{\theta}) = \big(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\big)'(\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta}))$ (equation 2-3). In contrast, for the general multi-response estimation problem, the parameters and noise variance/covariance estimation problems are linked.

Reflecting on the SSE function for multi-response models (equation 3-2), it can be seen that one of the key features in this equation is the overall noise covariance matrix ($\boldsymbol{\Omega}$) which is not necessarily known. In fact, in most cases this matrix is unknown and its elements impose more parameters on the estimation problem. These elements can be considered as nuisance parameters whose values impact the model parameter estimates. Alternatively, they can be regarded as parameters of interest whose values are required to generate the likelihood function. For a multi-response model with m responses, this means an extra $m(m+1)/2$ elements (for a symmetric within-run noise covariance matrix $\boldsymbol{\Sigma}$) to be estimated, bringing the total number of the unknowns to $p + m(m+1)/2$. The overall noise covariance matrix $\boldsymbol{\Omega}$ can be estimated as a block diagonal matrix of the estimated within-run noise covariance matrix $\widehat{\boldsymbol{\Sigma}}$ from all the runs if the noise is assumed to be *IID Normal* between runs. Either way, one of the main concerns in using the GLS-based profiling is the process of estimating $\boldsymbol{\Omega}$ during parameter estimation and profiling stages.

Sometimes measurement noise variances are considered to be known because a sufficient number of repeated experimental runs are available (Campbell et al., 2012). Also it is possible at times to draw conclusions about the noise terms based on the prior knowledge of the process (Karimi et al., 2012). However, in most cases it is not easy to come up with a reliable noise covariance matrix for the process based on the information from the

dataset and as a result modelers are prompted to make assumptions about the noise variance/covariance values. Since the noise covariance matrix plays a significant role in parameter estimation and inference as well as parameter ranking and sensitivity analyses (for example in building the Fisher Information Matrix) (McLean and McAuley, 2012), any assumptions about $\boldsymbol{\Omega}$ can impact the model development.

Based on the literature, it is common to make assumptions regarding the noise terms to make the estimation practice feasible (Biegler et al., 1986); (Raue et al., 2009); (Karimi et al., 2012). Perhaps one of the most simplifying assumptions is that the noise elements are independent from each other (i.e., there are no co-dependencies among the responses) and the noise variances are known. Thus, a specific variance is assigned to each response variable and stays the same across all runs for that response. The corresponding within-run noise covariance matrix ($\boldsymbol{\Sigma}$) in this case is a diagonal matrix consisting of the noise variances as diagonal entries $\left(\boldsymbol{\Sigma} = diag\left[{\sigma_j}^2\right];\ (j = 1,2,\dots,m)\right)$ (e.g., (Raue et al., 2009)). The overall noise covariance matrix $\boldsymbol{\Omega}$ can be estimated as a block diagonal matrix of n of the within-run noise covariance matrices $\boldsymbol{\Sigma}$. In such cases, if $\boldsymbol{\Omega}$ is specified by its elements as $\boldsymbol{\Omega} = \left[{\omega_j}^2\right]$, the resulting objective function is basically a weighted sum of squares that uses the inverse of these elements as weighting factors:

$$S(\boldsymbol{\theta}) = \sum_{j=1}^{m}\sum_{i=1}^{n}\frac{1}{{\omega_j}^2}\left(y_j(\boldsymbol{x}_i,\boldsymbol{\theta}) - f_j(\boldsymbol{x}_i,\boldsymbol{\theta})\right)^2 \qquad (3\text{-}5)$$

A special case of equation 3-5 would be assuming equal variances for all responses, so that all the weighting factors are equal and the objective function is further simplified to a similar format as that of the single-response systems:

$$S(\boldsymbol{\theta}) = \left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right)'\left(\boldsymbol{y_{vec}} - \boldsymbol{\eta_{vec}}(\boldsymbol{\theta})\right) \qquad (3\text{-}6)$$

Although these assumptions make the estimation problem less complicated, they are not likely to be true in general, and prevent the solution from taking advantage of the information that is available from the dataset (Box and Draper, 1965); (Box et al., 1973). Moreover, it can be seen from equation 3-5 that any assumptions regarding the values of the weighting factors (inverse of the elements of $\mathbf{\Omega}$) can impact the significance of the respective responses relative to each other and lead to vastly diminishing or even eliminating the impact of particular responses on the SSE function and the consequent parameter estimates. For instance under the assumption of a high variance for the first response and a relatively low variance for the second response, the impact of response y_1 on the SSE function will be negligible in comparison to that of response y_2. Thus, response y_1 is effectively left out of the parameter estimations. Also, in chemical processes the model parameters and responses often have physical interpretations and there are restrictions around how they can be co-dependent on each other in reality. Therefore, the modeler should be mindful of the impacts of the assumptions about noise terms on the parameter estimation outcomes.

As mentioned in Chapter 2, the co-dependencies between the responses might exist between the experimental runs (between-run covariance), as well as within each run (within-run covariance). The most common approach for making assumptions about the noise terms is to acknowledge the co-dependencies of different responses in each run (within-run co-dependencies) and consider them to be known, but to assume *IID Normal* noise between the runs (e.g., (Biegler et al., 1986); (Guay and McLean, 1995); (Benabbas et al., 2005)):

$$E\{e_{ij}e_{rs}\} = \begin{Bmatrix} \sigma_{js}, i = r \\ 0\ , i \neq r \end{Bmatrix};\ (i, r = 1,2, \dots, n;\ \ j, s = 1,2, \dots, m) \tag{3-7}$$

where j and s represent the response number. In practice the variance/covariance terms are often unknown and need to be estimated (Box and Draper, 1965). This estimation process can impact the profile likelihood plots and the parameter inferences consequently. Development of the estimated noise covariance matrix and the assumptions involved lead to several approaches for profiling that are discussed in the following section. In the discussion that follows, the within-run noise co-dependencies are taken into account while the between-run co-dependencies are ignored. Note that if there is between-run noise co-dependency, the majority of the expressions can still be used except for the block-diagonal structure for the noise covariance matrix. The more substantial impact of having between-run co-dependency is that it would be very difficult to estimate the overall noise covariance matrix without replicate runs. This is analogous to the situation in single-response estimation when the noise is not homoscedastic.

3.3 Profiling with Unknown Noise Covariance Matrix

In the previous section the importance of the noise covariance matrix in profiling for nonlinear multi-response models was explained. Looking back at the GLS-based profile likelihood function defined in equation 3-4 as $PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}})$, since the SSE function (equation 3-2) includes the overall noise covariance matrix $\boldsymbol{\Omega}$, this matrix can impact the profiling results. One challenge for multi-response model development is that except for the rare cases with sufficient repeated experimental runs and/or reliable prior knowledge about the noise elements, in practice the noise variance and covariance terms are often unknown and need to be estimated (Box and Draper, 1965).

The common assumptions regarding the noise elements and the influence of these assumptions on the profiling outcomes were explained in the previous section. In this section the estimation methodologies for the noise covariance matrix are described. The maximum likelihood estimate for the noise covariance matrix conditional on the parameter estimates is given by (Box and Draper, 1965):

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}) \tag{3-8}$$

with $\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{Y} - \boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})$ being the residual matrix evaluated at the least squares parameter estimates (please refer to equation 2-35). This conditional estimate for the noise covariance matrix has been used to concentrate the likelihood for parameter estimation, leading to the Determinant Criterion (DC) approach that is further discussed in Chapter 4. Alternatively, Beauchamp and Cornell (1966) and Seber and Wild (1989) have suggested estimating the noise covariance matrix by means of a method originally proposed by Zellner (1962) for SUR models. The Zellner method starts by guessing some initial values for parameters, followed by fitting the model for each response *j* individually using the least squares method. Subsequently, the least squares parameter estimates for each response ($\breve{\boldsymbol{\theta}}_j$) are determined and used in the following equation to calculate the residuals $\boldsymbol{e}_j$ for the j[th] response:

$$\boldsymbol{e}_j = \boldsymbol{y}_j - \boldsymbol{\eta}_j(\breve{\boldsymbol{\theta}}_j) \tag{3-9}$$

In the next step, the elements of $\widehat{\boldsymbol{\Sigma}}$ are estimated from these residuals as follows:

$$\hat{\sigma}_{sj} = \frac{1}{n}\boldsymbol{e}_s{}'\boldsymbol{e}_j;\ (s, j = 1,2,\dots,m) \tag{3-10}$$

$$\widehat{\boldsymbol{\Sigma}} = \left[\hat{\sigma}_{sj}\right];\ (s, j = 1,2,\dots,m) \tag{3-11}$$

This $\widehat{\boldsymbol{\Sigma}}$ has been regarded as an appropriate estimate for the noise covariance matrix and used by several researchers (Zellner, 1962); (Beauchamp and Cornell, 1966); (Seber and Wild, 1989).

In this section several approaches for estimating the overall noise covariance matrix are studied and the most reliable one is proposed. These approaches involve calculations based on the Zellner method. In some approaches the matrix obtained from the Zellner method is regarded as an adequate estimate for the with-run noise covariance matrix whereas in others it is just used as an initial estimate and an iterative method is used to arrive at the final $\widehat{\boldsymbol{\Sigma}}$. These approaches also differ based on whether the iteration method is only used during parameter estimation or during both parameter estimation and profiling. This methodology is explained in more detail in the following sections.

3.3.1 Forming the Initial Overall Noise Covariance Matrix

Like most estimation procedures, estimating $\boldsymbol{\Omega}$ usually starts with guessing an initial matrix $\widehat{\boldsymbol{\Omega}}^{(\mathbf{0})}$. If the noise covariance matrix is unknown but the dataset or experimental procedure provides some information about the initial noise variance and covariances, or the literature suggests some approximate values for the within-run or overall noise covariance matrices ($\boldsymbol{\Sigma}$ or $\boldsymbol{\Omega}$) which cannot be entirely trusted, this information can be used to form an initial estimate for $\widehat{\boldsymbol{\Sigma}}^{(0)}$ or $\widehat{\boldsymbol{\Omega}}^{(\mathbf{0})}$ respectively. If an initial *m* by *m* (*m* being the number of responses) within-run noise covariance matrix $\widehat{\boldsymbol{\Sigma}}^{(0)}$ is estimated in this way, then for the vector representation of the model $\boldsymbol{y_{vec}} = \boldsymbol{\eta_{vec}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon_{vec}}$ (equation 3-1) under the assumption of *IID Normal* between-run noise, the initial overall noise covariance matrix $\widehat{\boldsymbol{\Omega}}^{(0)}$ can be formed by the following equation:

$$\widehat{\boldsymbol{\Omega}}^{(0)} = blockdiag_n\, \widehat{\boldsymbol{\Sigma}}^{(0)} \tag{3-12}$$

where $blockdiag_n$ indicates constructing a block diagonal matrix from *n* covariance matrices $\widehat{\boldsymbol{\Sigma}}^{(0)}$.

In cases where reliable conclusions about initial noise variance and covariances cannot be drawn from the dataset and data history, I suggest using the Zellner method to obtain an estimate for the noise covariance matrix and treat it as an initial $\widehat{\boldsymbol{\Sigma}}^{(0)}$ (Zellner, 1962); (Seber and Wild, 1989). Subsequently, the initial overall noise covariance matrix $\widehat{\boldsymbol{\Omega}}^{(0)}$ is estimated as the block diagonal matrix from *n* covariance matrices $\widehat{\boldsymbol{\Sigma}}^{(0)}$ (equation 3-12).

Building $\widehat{\boldsymbol{\Omega}}^{(0)}$ from replicate runs where applicable can improve the results. The difference between the initialization step and the iterative approach discussed later is that responses are fitted individually and variances and covariances are estimated in the initialization step. In the two-step method that follows, the responses are fitted jointly, and the covariance matrix is estimated conditional on the current iterates of the parameter estimates.

3.3.2 Estimating the Overall Noise Covariance Matrix Ω and Performing Profiling

Once the initial $\widehat{\boldsymbol{\Omega}}^{(0)}$ is formed, there are several approaches in the context of the GLS method to estimate $\boldsymbol{\Omega}$ and carry out the profiling process. In this thesis the entire profiling procedure is viewed as a combination of a parameter estimation step followed by a profiling step. Depending on how the initial $\widehat{\boldsymbol{\Omega}}^{(0)}$ is incorporated into these steps, three different scenarios explained in the following lines can be imagined.

GLS Profiling Approach #1 – Fixing the Noise Covariance Matrix at the Initial Values:

In the first approach, the initial noise covariance matrix (obtained from prior knowledge, experimental runs, or methods such as Zellner method) is considered to be an adequate estimate and thus the noise covariance matrix is kept fixed at these initial values during both parameter estimation and profiling stages. Once the parameter estimation is done, profiling is applied to the conditional likelihood function (i.e., the likelihood function conditional on the specified initial values of the noise covariance matrix).

GLS Profiling Approach #2 – Updating the Noise Covariance Matrix during Parameter Estimation, Fixing It during Profiling:

The second approach represents an iterative approach consisting of two conditional estimation loops. This process involves estimating the parameters conditional on current iterates for the noise covariance matrix, followed by a second step in which the noise covariance matrix is re-estimated conditional on the most recent iterates of the parameter estimates. This procedure is alternated until convergence, at which time estimates for both the parameters and the noise covariance matrix are obtained. In other words, in each iteration the parameters are estimated conditional on $\widehat{\boldsymbol{\Sigma}}$ using the GLS parameter estimation method. Then, $\widehat{\boldsymbol{\Sigma}}$ is updated based on the most recent residuals. The final noise covariance matrix obtained from this method is kept fixed during profiling. Thus, the subsequent profiling is conditional on the specified value of the profiling parameter, as well as on the values of the noise covariance matrix which are fixed at the overall GLS-optimal values. The profile traces and plots for this approach should be interpreted accordingly.

GLS Profiling Approach #3 – Updating the Noise Covariance Matrix during Parameter Estimation as well as Profiling:

Since the parameter estimation problem is a joint estimation of parameters and the noise covariance matrix, the full profiling problem could also address the joint estimation of the parameters (less the profiling parameter) and the associated noise covariance matrix. Therefore, in this approach, the two conditional estimation loops of approach #2 are applied to arrive at the overall SSE-optimal estimates for parameters and the noise covariance matrix. For the profiling stage, the $\widehat{\boldsymbol{\Omega}}$ corresponding to the final parameter estimates is used only as an initial estimate and for each set of parameters obtained during the profiling process, $\widehat{\boldsymbol{\Omega}}$ is updated to match the most recent residual values. However this estimation method generally turns out to be an ill-posed and numerically challenging problem.

The main reason for considering the third profiling approach is that when estimating the noise covariance matrix based on the residual values, the entries of this matrix depend on the residual matrix. Therefore, as different sets of parameters are generated during profiling, the residual matrix changes and the new residual values can be used to update the noise covariance matrix. This does not impact the maximum likelihood parameter estimates $\widehat{\boldsymbol{\theta}}$ about which profiling is performed, but can impact the conditional estimates of the other parameters ($\widetilde{\boldsymbol{\theta}}_{-k}$) and the corresponding profile likelihood plots. Thus, in theory iterating on estimates of both parameters and noise covariances should improve the results while providing insight into the extent to which the covariance estimates influence the parameter estimates. However, this approach proved to be impractical for all of the examples that it was applied to, since it led to very poorly conditioned

problems. It appears that since in this method the estimation of $\boldsymbol{\Omega}$ and $\boldsymbol{\theta}$ happen consecutively, the estimation process works in a way that the impact of the changes in the parameter estimates (and hence the residual values) are offset by the estimated values for $\widehat{\boldsymbol{\Omega}}$. As a result, changes in the SSE function are kept at minimum and the value of the profile likelihood function tends towards zero.

Besides the *GLS-based profiling approach #3*, the *DC-based profiling approach* also suffers from the numerical challenges of joint profiling. A good example for illustrating these challenges is Case study 1 presented in Chapter 4. In this case study both *GLS-based profiling approach #2* and *DC-based profiling approach* have been applied to a linearly parameterized multi-response model and dataset generated by (Bates and Watts, 1987). The resulting pairwise GLS-based and DC-based profile traces for $\beta_1$and β_2 for this example are presented. The DC-based profile traces fall very close or almost overlay each other which indicates a higher level of co-dependency between the parameter estimates for the *DC-based profiling* compared to the *GLS-based profiling approach #2*. This is because in the *DC-based profiling approach* both the parameters and the noise covariance matrix estimates are updated implicitly in the iteration loop (as will be explained more thoroughly in Chapter 4). The closeness of the profile plots to each other and the high co-dependencies among the parameter estimates are signs of the ill-conditioning of this estimation problem and the potential numerical issues. Thus,the *GLS-based profiling approach #3* that updates the parameters and noise covariance elements jointly faces the same challenges which make it impractical or less appealing than the other GLS-based approaches.

In this research I opt for *GLS profiling approaches #1 and #2*, which represent profiling conditional on specified values for the noise covariance matrix. In *GLS profiling approach #1*, profiling is conditional on the nominal noise covariance matrix obtained from initial parameter values or prior information about noise co-dependencies. In *GLS profiling approach #2*, profiling is conditional on the covariance matrix estimated at the GLS-optimal estimates of the parameters and noise covariance matrix. A summary of different steps during parameter estimation and profiling along with some possible scenarios that were studied is provided in Figure 3-1.

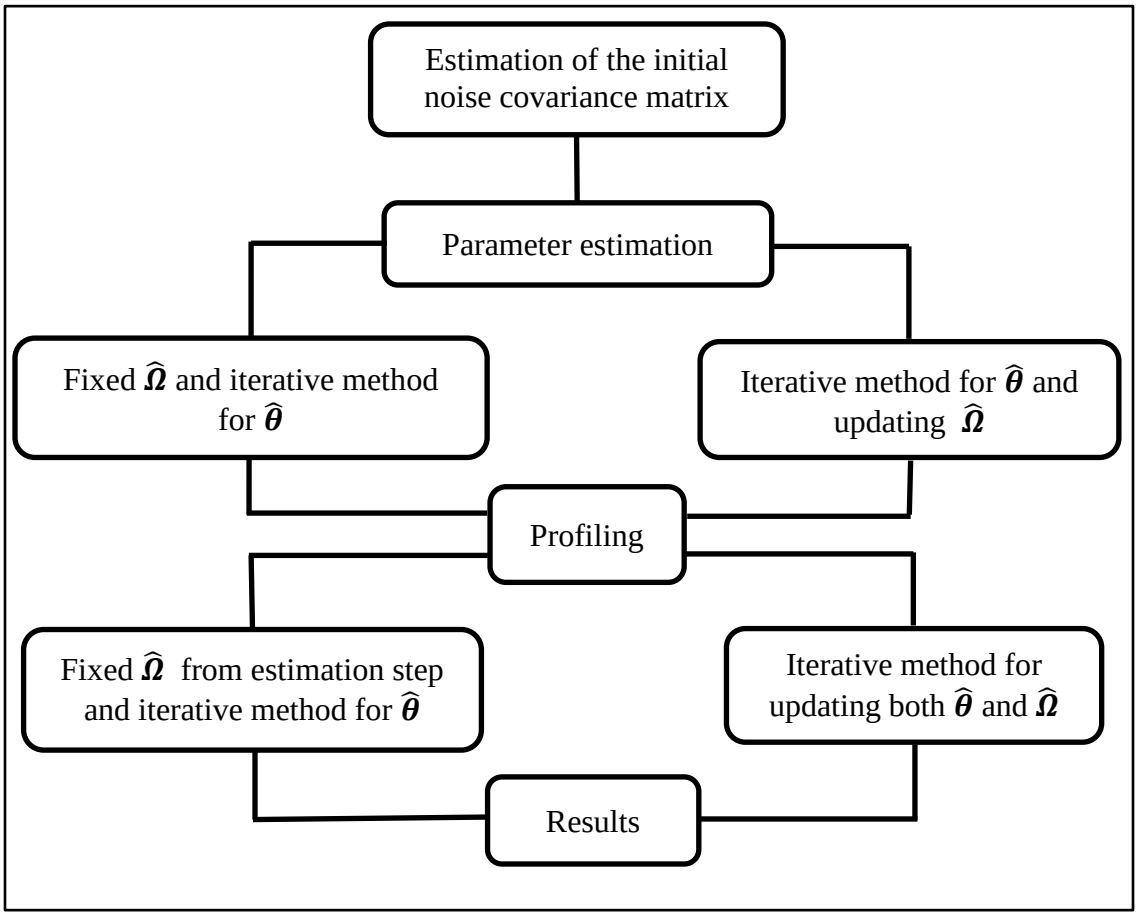

Figure 3-1: Summary of different steps for parameter estimation and profiling approaches with unknown noise covariance matrix based on GLS method

Overall, having studied the three approaches, I propose the *GLS profiling approach #2* as an advantageous profiling method since it accounts for a more reliable $\widehat{\boldsymbol{\Omega}}$ and also avoids ill-conditioned matrices and numerical challenges. This recommended profiling method has been examined on several datasets and the results have confirmed its validity. These case studies are included later in this chapter.

Once the noise covariance matrix is estimated and the profile likelihood function is formed, this function can be applied to parameter inference and estimability analysis. However, for this to be feasible, it is necessary to determine the distribution of the profile likelihood function. This is the objective of the following section.

3.4 Distribution of the GLS-Based Profile Likelihood Function

As mentioned in the previous discussions, the estimation of the noise covariance matrix and its distribution impacts the distribution of the profile likelihood function. Therefore, in order to determine the distribution of the GLS-based profile likelihood function for multi-response models $PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}})$ (equation 3-4), models with known and unknown covariance matrices will be studied individually. The outcomes of this study are required for finding the profile-based parameter intervals since the values of the PL_S function are used in the profile likelihood plots and the distribution of this function determines the reference lines that mark the boundaries of the parameter intervals.

3.4.1 Models with Known Noise Covariance Matrix

If the noise variance and covariance values are known, the within-run noise covariance matrix $\boldsymbol{\Sigma}$ can be formed from these values. Subsequently, under the assumption of *IID*

Normal between-run noise, the overall noise covariance matrix $\mathbf{\Omega}$ can be formed as a block diagonal matrix of $\mathbf{\Sigma}$ over all the runs ($\mathbf{\Omega} = blockdiag_n\, \mathbf{\Sigma}$). If it is further assumed that there are no co-dependencies between the noise terms (neither within-run nor between-run) and the variances are known and are the same for each response across all runs, $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ will be simplified to diagonal matrices ($\mathbf{\Omega} = [{\omega_j}^2]$). The corresponding $S(\boldsymbol{\theta})$ function will be similar to a weighted sum of squares and the likelihood function $L(\boldsymbol{\theta})$ can be expressed as (see for example (Raue et al., 2009)):

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{{\omega_j}^2} (y_j(\boldsymbol{x}_i, \boldsymbol{\theta}) - f_j(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 \quad (3\text{-}13)$$

When the noise covariance matrix embedded in the SSE function is known, the SSE function consists of sums of standard normal distributions (residuals divided by their variance) and from a statistical point of view, this function has a χ^2 distribution (Montgomery et al., 2004). Thus, the least squares SSE function $S(\widehat{\boldsymbol{\theta}})$ for such cases will be distributed as:

$$S(\widehat{\boldsymbol{\theta}}) \sim \chi^2_{nm-p} \quad (3\text{-}14)$$

The right-hand side represents a χ^2 distribution with $nm - p$ degrees of freedom which corresponds to nm pieces of information provided by the responses and p constraints imposed by the parameters. A more comprehensive discussion regarding the degrees of freedom for this distribution is presented in the next section. Likewise the SSE function conditional on a fixed value of the profiling parameter (θ_k) has the following distribution:

$$\tilde{S}(\theta_k) \sim \chi^2_{nm-(p-1)} \quad (3\text{-}15)$$

The degrees of freedom for this distribution reflect the fact that the profiling parameter is fixed and left out of the estimation process, reducing the number of constraints to $p - 1$.

Consequently, the distribution of the GLS-based profile likelihood function can be expressed as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\hat{\boldsymbol{\theta}}) \sim \chi_1^2 \quad (3\text{-}16)$$

This distribution is based on the fact that $\tilde{S}(\theta_k)$ has one more fixed parameter than $S(\hat{\boldsymbol{\theta}})$, resulting in the estimation of one less parameter.

3.4.2 Models with Unknown Noise Covariance Matrix

If the noise covariance matrix is unknown, the Zellner method (described in section 3.3) can be used to determine an estimate for $\hat{\boldsymbol{\Sigma}}$. If $\hat{\boldsymbol{\Sigma}}$ is estimated by the Zellner method (using equations 3-10 and 3-11), its distribution depends on the distribution of the residual matrix. For a multi-response system, each row (i) of the residual matrix represents one run and contains the residuals corresponding to different responses in that run:

$$\boldsymbol{Z}_{(i)}(\boldsymbol{\theta}) = (e_{1(i)}\ e_{2(i)}\ \ldots\ e_{m(i)}) \quad (3\text{-}17)$$

Assuming the model is correct, the noise structure within any given run has an *MVN* distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$:

$$(\boldsymbol{\varepsilon}_{1(i)}\ \boldsymbol{\varepsilon}_{2(i)}\ \ldots\ \boldsymbol{\varepsilon}_{m(i)})' \sim MVN_m(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \quad (3\text{-}18)$$

Since it is assumed that the noise structure between runs is *IID Normal*, the dataset represents a sample of size *n* on an *m-variate Normal* distribution. Based on the definition of the Wishart distribution (Johnson and Wichern, 2007), since elements of $\hat{\boldsymbol{\Sigma}}$ are estimated from the residuals by means of $\hat{\sigma}_{sj} = \frac{1}{n}\boldsymbol{e}_s{}'\boldsymbol{e}_j;\ (s, j = 1,2,\ldots,m)$ (equation 3-10), $\hat{\boldsymbol{\Sigma}}$ asymptotically has a Wishart distribution (regardless of the approach for updating it during parameter estimation and/or the profiling process). The degrees of

freedom for this distribution is equal to that of the residuals which in general is $nm - p$ shown in the discussions that follow:

$$\widehat{\boldsymbol{\Sigma}} \sim W_m(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, nm - p) \tag{3-19}$$

where the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ serves as the scale matrix. The corresponding SSE function contains an estimated noise covariance matrix with a Wishart distribution and therefore its distribution is no longer simply χ^2.

In order to understand the distribution and the associated degrees of freedom for PL_S, it is beneficial to revisit the distribution of the residual sum of squares function and the residuals. The SSE function for such multi-response cases consists of normally distributed residuals and the inverse of the estimated noise covariance matrix, and statistically translates into an *F* distribution. The degrees of freedom of this distribution is the rank of the orthogonal projection operator (Johnson and Wichern, 2007). Using the Gauss-Newton method for parameter estimation of the vector representation of the model (equation 3-1), the increment change in parameter estimates for the i^{th} iteration is given by:

$$\delta\boldsymbol{\theta}^{(i+1)} = \left(\boldsymbol{V}(\boldsymbol{\theta}^{(i)})'\boldsymbol{\Omega}^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})\right)^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{y}_{\boldsymbol{vec}} - \boldsymbol{\eta}_{\boldsymbol{vec}}\left(\boldsymbol{\theta}^{(i)}\right)\right) \tag{3-20}$$

with $\boldsymbol{\theta}^{(i)}$ denoting the parameter estimates for the i^{th} iteration, and $\boldsymbol{V}(\boldsymbol{\theta}^{(i)})$ being the sensitivity matrix defined as:

$$\boldsymbol{V}(\boldsymbol{\theta}^{(i)}) = \left.\frac{\partial\boldsymbol{\eta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}} \tag{3-21}$$

The increment change in the predicted response based on this linearization is:

$$\delta\hat{\boldsymbol{y}}^{(i)} = \boldsymbol{V}(\boldsymbol{\theta}^{(i)})\delta\boldsymbol{\theta}^{(i+1)} = \boldsymbol{V}(\boldsymbol{\theta}^{(i)})\left(\boldsymbol{V}(\boldsymbol{\theta}^{(i)})'\boldsymbol{\Omega}^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})\right)^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})\boldsymbol{\Omega}^{-1}\delta\boldsymbol{y}^{(i)} \tag{3-22}$$

where $\delta \boldsymbol{y}^{(i)} = \boldsymbol{y}_{\boldsymbol{vec}} - \boldsymbol{\eta}_{\boldsymbol{vec}}(\boldsymbol{\theta}^{(i)})$. Therefore, the residual output increment can be written as follows:

$$\delta \boldsymbol{y}^{(i)} - \delta \hat{\boldsymbol{y}}^{(i)} = (\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)})\delta \boldsymbol{y}^{(i)} \tag{3-23}$$

where $\boldsymbol{I}_{nm}$ is the $nm \times nm$ identity matrix and $\boldsymbol{P}^{(i)}$ is the *projection operator* associated with the i^{th} Gauss-Newton iteration:

$$\boldsymbol{P}^{(i)} = \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \left(\boldsymbol{V}\left(\boldsymbol{\theta}^{(i)}\right)' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \right)^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)})' \boldsymbol{\Omega}^{-1} \tag{3-24}$$

Subsequently, the *orthogonal projection operator* can be written as:

$$\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)} = \boldsymbol{I}_{nm} - \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \left(\boldsymbol{V}\left(\boldsymbol{\theta}^{(i)}\right)' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \right)^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)})' \boldsymbol{\Omega}^{-1} \tag{3-25}$$

The objective here is to find the rank of the orthogonal projection matrix. This matrix is idempotent, i.e., (Gentle, 2007):

$$\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right)^{\boldsymbol{2}} = \boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)} \tag{2-26}$$

Using the following matrix identity for idempotent matrices (Gentle, 2007):

$$rank\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right) = tr\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right) \tag{3-27}$$

and using the properties of the trace,

$$tr\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right) = tr(\boldsymbol{I}_{nm}) - tr\left(\boldsymbol{P}^{(i)}\right) \tag{3-28}$$

Since $\boldsymbol{I}_{nm}$ is an $nm \times nm$ identity matrix, $tr(\boldsymbol{I}_{nm}) = nm$, the only unknown here is the trace of $\boldsymbol{P}^{(i)}$ which can be determined by:

$$tr\left(\boldsymbol{P}^{(i)}\right) = tr\left(\boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \left(\boldsymbol{V}\left(\boldsymbol{\theta}^{(i)}\right)' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \right)^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)})' \boldsymbol{\Omega}^{-1} \right) =$$

$$tr(\left(\boldsymbol{V}\left(\boldsymbol{\theta}^{(i)}\right)' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) \right)^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)})' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\boldsymbol{\theta}^{(i)}) = tr\left(\boldsymbol{I}_p\right) = p \tag{3-29}$$

Consequently using the properties of the trace,

$$rank\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right) = nm - p \tag{3-30}$$

As a result, the distribution of the GLS-based profile likelihood function can be expressed as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \sim F_{1,nm-p} \quad (3\text{-}31)$$

The first value for degrees of freedom for this *F* distribution reflects the fact that $\tilde{S}(\theta_k)$ has one more fixed parameter than $S(\widehat{\theta})$. The second degrees of freedom value ($\nu_2 = nm - p$) is based on nm pieces of information and p parameters as concluded from equation 3-30, i.e., the degrees of freedom of the residuals and $\widehat{\boldsymbol{\Sigma}}$. This distribution can be used for marking the reference lines on profile plots to obtain parameter likelihood intervals.

In summary, the overall noise covariance matrix $\boldsymbol{\Omega}$ is required for GLS-based profiling in nonlinear multi-response models. If this matrix is known, it can be inserted into the equation for the GLS-based profile likelihood function (equation 3-4). Otherwise if the noise covariance matrix is unknown, it is recommended in this thesis to assume that the responses are co-dependent within each run, but *IID Normal* between runs. Under this assumption, the initial within-run noise covariance matrix $\widehat{\boldsymbol{\Sigma}}$ can be estimated based on the Zellner method and the overall noise covariance matrix $\widehat{\boldsymbol{\Omega}}$ can be estimated as a block diagonal matrix of $\widehat{\boldsymbol{\Sigma}}$ for all runs. This estimation process (referred to as *GLS profiling approach #2* and explained in section 3.3) starts with a guessed initial matrix $\widehat{\boldsymbol{\Sigma}}^{(0)}$ which is updated during the parameter estimation based on the latest values for parameter estimates. The corresponding $\widehat{\boldsymbol{\Omega}}$ can be inserted into equation 3-4 to form the GLS-based profile likelihood function $PL_S(\theta_k)$. The PL_S function has a χ_1^2 distribution for cases with known $\boldsymbol{\Omega}$ and an $F_{1,nm-p}$ distribution for the cases with unknown $\boldsymbol{\Omega}$, considering the nm pieces of information available for multi-response systems alongside p constraints

imposed by the parameter estimates. A profile likelihood plot of $PL_S(\theta_k)$ versus θ_k can be used for forming $100(1-\alpha)$ parameter likelihood intervals with reference lines at $\chi^2_{1;\alpha}$ or $F_{1,nm-p;\alpha}$ for cases with known or unknown $\mathbf{\Omega}$ respectively. The points on the plot where the reference line crosses the profile likelihood curve mark the boundaries of the parameter likelihood interval.

3.4.3 Degrees of Freedom Discussion

One of the most important issues when determining the distribution of the profile likelihood function is choosing the correct degrees of freedom. In the argument presented in the previous section, the degrees of freedom associated with the χ^2 distribution of the SSE function $S(\widehat{\boldsymbol{\theta}})$ with known $\mathbf{\Omega}$ and the F distribution of the GLS-based profile likelihood function $PL_S(\theta_k)$ for the case of unknown $\mathbf{\Omega}$ were considered to be $nm-p$ (equations 3-14 and 3-31). This argument accounts for nm pieces of information provided by the responses and p constraints imposed by the parameters for the system. Buzzi Ferraris et al. (1984) have also used a χ^2 distribution with $nm-p$ for $S(\widehat{\boldsymbol{\theta}})$ when $\mathbf{\Omega}$ is known.

Expanding this argument, it should be noted that the $nm-p$ degrees of freedom holds when there are no linear relationships among the responses. In chemical systems it is possible to encounter linear relationships between the expected values of the responses imposed for example by stoichiometry, energy and material balances, or steady-state conditions. For instance, stoichiometry might require that the sum of the expected values of the number of moles of some components in the system must be the same. These linear relationships, if not noticed and accounted for, can cause ill-conditioning in the modelling

process (Box et al., 1973). Also, linear relationships among responses might arise when the observed data are modified by means of normalization or similar adjustment methods to satisfy a stoichiometric or other kind of relationship. Another instance when such dependencies occur is when only some of the responses are actually measured and the rest are calculated from the measured ones by means of auxiliary relationships. These situations should be avoided when possible, but sometimes these linear relationships cannot be avoided due to the analytical procedures or requirements of the equipment. An example of such cases is obtaining chemical composition data from some gas chromatographs where it is only possible to calculate relative percentages. Instances might arise when such details about data collection are not presented to the modeler, compounding the issue of determining and accounting for such linear relationships among the responses.

For parameter estimation and inference purposes, it is important that the linear relations between the responses are determined, and that the response set is reduced to the linearly independent set, and the estimations are carried out only based on the independent responses (Box et al., 1973); (McLean et al., 1979). Methods such as Eigenvalue Analysis or Single Value Decomposition (SVD) can be used to reduce the response set to the largest linearly independent subset. An eigenvalue analysis proposed by Box et al. (1973) for determining and eliminating the linearly dependent responses form the response set has been widely used in this regard. This method enables the researcher to investigate two types of linear dependencies. The first type is the dependency in the data and the second type is the linear dependencies among the expected values of the responses. In this method, the individual average $\bar{y}_j$ for each response across all runs is

calculated and the deviation of each response from its individual average is determined. Subsequently, a deviation matrix (denoted here by $\boldsymbol{C}$) is formed from these deviation values as follows (Box et al., 1973):

$$\boldsymbol{C} = \{c_{ij}\} \tag{3-32}$$

$$c_{ij} = y_{ij} - \bar{y}_j;\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{3-33}$$

Examining the eigenvalues (λ_k) and corresponding eigenvectors (v_k) of the $\boldsymbol{C}'\boldsymbol{C}$ matrix is the key to revealing the potential linear relationships between the responses. The first step is to identify the eigenvalues that are effectively or nearly zero. The eigenvectors corresponding to these eigenvalues depict the linear dependencies among the responses with the elements of these eigenvectors serving as the coefficients of the linear relationship. Thus, if $\lambda_k = 0;\ (k = 1,2,\dots,l_1)$, it can be concluded that there are l_1 linear dependencies among the responses. Uncovering the linear dependencies among the expected values of the responses (second type of dependencies) is more complicated to some extent. If some of the eigenvalues are not near zero but are relatively small in comparison to others, they might be a sign of the second type of dependencies. Box et al. (1973) showed that the eigenvalues that are of the same order of magnitude as $(n-1)v_k'\boldsymbol{\Sigma}v_k$ (with v_k being the eigenvectors corresponding to the mentioned eigenvalues and $\boldsymbol{\Sigma}$ being the within-run noise covariance matrix) indicate linear dependencies among the expected values of the responses. They have mentioned that if $\boldsymbol{\Sigma}$ is not known it needs to be estimated. Thus the success of this analysis depends on having a good estimate for this matrix and this provides another reason for why the estimation of the noise covariance matrix is of great importance throughout the modeling procedure for multi-response systems. The number of such eigenvalues (l_2) indicates the number of the linear

dependencies among the expected values of the responses. Eventually, this analysis reveals a total number of $l = l_1 + l_2$ linear dependencies. Consequently, the response set can be reduced to linearly independent set of m' elements where $m' = m - l$. Note that this is equivalent to performing an unscaled Principal Component Analysis (PCA) on the responses.

From this point on, the estimation problem should be carried out on the reduced response vector. Thus, following the argument in the preceding section for determining the degrees of freedom associated with the GLS-based profile likelihood function, the *orthogonal projection operator* can be written as:

$$\boldsymbol{I}_{nm'} - \boldsymbol{P}^{(i)} = \boldsymbol{I}_{nm'} - \boldsymbol{V}(\boldsymbol{\theta}^{(i)})\left(\boldsymbol{V}(\boldsymbol{\theta}^{(i)})'\boldsymbol{\Omega}^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})\right)^{-1}\boldsymbol{V}(\boldsymbol{\theta}^{(i)})'\boldsymbol{\Omega}^{-1} \tag{3-34}$$

For this idempotent matrix,

$$rank\left(\boldsymbol{I}_{nm'} - \boldsymbol{P}^{(i)}\right) = tr\left(\boldsymbol{I}_{nm'} - \boldsymbol{P}^{(i)}\right) \tag{3-35}$$

Using matrix identities (Gentle, 2007):

$$tr\left(\boldsymbol{I}_{nm'} - \boldsymbol{P}^{(i)}\right) = tr(\boldsymbol{I}_{nm'}) - tr\left(\boldsymbol{P}^{(i)}\right) \tag{3-36}$$

Since $\boldsymbol{I}_{nm'}$ is an $nm' \times nm'$ matrix, $tr(\boldsymbol{I}_{nm'}) = nm'$, therefore:

$$rank\left(\boldsymbol{I}_{nm} - \boldsymbol{P}^{(i)}\right) = nm' - p \tag{3-37}$$

Thus, the GLS-based profile likelihood function is distributed as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \sim F_{1,nm'-p} \tag{3-38}$$

Thus, for such cases there are appropriately $nm' - p$ degrees of freedom for the *F* distribution of $PL_S(\theta_k)$ where m' is the number of independent responses. Consequently, the degrees of freedom argument that was put forward can accommodate situations in which there are dependencies amongst the responses.

In Chapter 2, a class of multi-response models known as Seemingly Unrelated Regression (SUR) models were mentioned in which the dependence of each response on the regressor is parameterized independently of the other responses. It was explained that for parameter estimation in these models to take advantage of the information provided by the noise covariance matrix, the likelihood function needs to include penalty terms. For example, for a linear SUR model with a known noise covariance matrix defined by

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{e} \tag{3-39}$$

with **B** being a $p \times m$ parameter matrix $\left(\boldsymbol{B} = \left[b_{kj}\right]; (k = 1,2, \dots, p,\ j = 1,2, \dots, m)\right)$, the parameters can be estimated by means of a Plug-in Joint Weighted LASSO (PWL) approach in which the parameter estimates are obtained from:

$$\widehat{\boldsymbol{B}} = \text{argmin}_{\boldsymbol{B}}\left[tr\{(\boldsymbol{Y} - \boldsymbol{XB})\boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{XB})'\} + d_1 \sum_{k,j} w_{kj}\left|b_{kj}\right|\right] \tag{3-40}$$

In this equation d_1 is a tuning parameter and w_{ij} are weighting factors for the parameters. If **Ω** is unknown, the estimation becomes a two-step process, starting with estimating the noise covariance matrix for each run $\widehat{\boldsymbol{\Sigma}}$ using methods such as Graphical Least Absolute Shrinkage and Selection Operator (GLASSO) followed by estimating $\widehat{\boldsymbol{B}}$ given $\widehat{\boldsymbol{\Sigma}}$ using equation 3-40. Similar to the estimation approaches suggested earlier in this thesis, this estimation process takes the within-run co-dependencies into account but ignores the between-run co-dependencies. For detailed information about this method please refer to (Lee and Liu, 2012).

Another method for linear SUR models known as Multivariate Regression with Covariance Estimation (MRCE) was proposed by Rothman et al. (2010). In this method **B** and **Ω** are jointly estimated using a penalized normal likelihood function defined as (Rothman et al., 2010):

$$(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Omega}}) = \text{argmin}_{\boldsymbol{B},\boldsymbol{\Omega}}\left[g(\boldsymbol{B}, \boldsymbol{\Omega}) + d_1 \sum_{k' \neq k} |\omega_{k'k}| + d_2 \sum_{k=1}^{p} \sum_{j=1}^{m} \left|b_{kj}\right|\right] \quad (3\text{-}41)$$

in which $g(\boldsymbol{B}, \boldsymbol{\Omega})$ is the standard log-likelihood function, $\omega_{k'k}$ are the off-diagonal elements of the inverse overall noise covariance matrix $\boldsymbol{\Omega}^{-1}$, and d_1 and d_2 are tuning parameters that can be selected by cross-validation methods.

These techniques are advantageous for Linear SUR models. However, since this research does not consider SUR models, for the general model representation of multi-response systems (equation 3-1), the SSE function is sensitive to $\boldsymbol{\Omega}$ and it is not necessary to add penalty terms to the likelihood function to include the impact of the noise covariance terms.

Based on the argument presented in this section, I believe that $nm - p$ represents a valid degrees of freedom for the distributions of $S(\widehat{\boldsymbol{\theta}})$ for multi-response models to allow the objective function to benefit from the information provided by all the responses. However as mentioned in the background chapter (section 2.6), Bates et al. have used $n - p$ degrees of freedom for the arguments regarding the distribution of the DC-based objective functions. For instance, Soo and Bates (1996) have suggested using $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}}_{DC})}{(n-p)}$ for estimating the DC-based average noise variance (s'^2). I consider $nm - p$ to be a more appropriate degrees of freedom for the DC approach as well. Further discussion in this regard is provided in Chapter 4 where the DC approach is explained.

3.5 Case Study Examples

In this section, the application of the proposed profile-based parameter inference method (*GLS-based profiling approach #2*) is demonstrated using several standard multi-

response test examples with noise covariance matrices that are treated as unknown. For these examples, the parameter likelihood intervals are generated from GLS-based profiling and the reference lines for parameter inference are computed under two different degrees of freedom arguments, $nm - p$ and $n - p$. Reference lines corresponding to $F_{1,nm-p;\alpha}$ and $F_{1,n-p;\alpha}$ have been used to construct inference intervals.

Furthermore, in cases where some values for the noise covariance matrix have been reported in the literature, parameter inference is also done by considering a known noise covariance matrix and reference lines based on $\chi^2_{1;\alpha}$ (as per equation 3-16). For these examples, the datasets provided in the literature were originally synthesized by adding noise (with a known covariance matrix) to model predictions based on some known parameter values (referred to as "the *true* parameter values" in the literature). Therefore, the inference results constructed using the known noise covariance matrices are the most representative for the systems. As a result, the inference outcomes of the proposed approach are compared against the results from the *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$.

Lastly, in order to compare the profiling results with an empirical inference method, approximate 95% bootstrap confidence intervals are generated using the "normal approximated interval with bootstrapped bias and standard error method" which is implemented using the *bootci* command in Matlab™ and includes sampling with replacement from the data arguments. A sample of size 1000 is chosen for this purpose, and the GLS approach with updated noise covariance matrix during parameter estimation is applied for obtaining the input parameter estimates. The same case studies will be

studied in Chapter 4 using the Determinant Criterion approach and the results will be compared against the GLS-based results of this section.

For these case studies, some discrepancies and differences between the estimated $\widehat{\boldsymbol{\Sigma}}$ and the known $\boldsymbol{\Sigma}$ reported in the literature are expected based on the following reasons:

- One reason for the potential difference might be that the parameter estimates usually differ from the *true* parameter values that were originally used to generate the dataset because of the realization of the noise present in the simulated data. Thus, $\widehat{\boldsymbol{\Sigma}}$ that is obtained from the residuals will not necessarily be the same as $\boldsymbol{\Sigma}$.
- Another reason for the difference between the known $\boldsymbol{\Sigma}$ and the estimated $\widehat{\boldsymbol{\Sigma}}$ might be that as was noticed with some of the case studies, the known $\boldsymbol{\Sigma}$ reported in the literature was in fact a diagonal matrix consisting of the noise variances. The researchers have assumed variances for the responses based on the available information, and have sometimes taken the conservative approach by assuming large values for the variances. On the other hand, I do not make presumptions about $\widehat{\boldsymbol{\Sigma}}$ and the proposed estimation process allows for estimation of non-diagonal matrices with variance and covariance elements.
- Lastly, the examples studied in this chapter usually consist of relatively small datasets. Therefore, the resulting $\widehat{\boldsymbol{\Sigma}}$ is usually very sensitive to estimation details such as initial parameter values, and the increments used in the iterations.

A more meaningful comparison between the known $\boldsymbol{\Sigma}$ and the estimated $\widehat{\boldsymbol{\Sigma}}$ might only be possible by generating a large dataset by adding noise to model predictions for known parameter values, using a method such as MCMC with a sample size of over 1000. Thus,

for the case studies in this thesis which are not generated in that way, the differences between the two matrices will not come as a surprise.

3.5.1 Case Study 1: Bates and Watts Multi-Response Model and Dataset

For the first case study, a linear and relatively simple model is chosen. However, for this model the estimation problem is still nonlinear due to the noise co-dependencies among multiple responses. This example was introduced by Bates and Watts (1987) and further investigated by Guay (1995) to reveal some of the difficulties that are encountered in multi-response parameter estimation. The system consists of three response variables (y_1, y_2, and y_3), each being linearly related to two parameters (β_1 and β_2) by the following model, with x_{ki} being the k^{th} regressor variable for the i^{th} run (Bates and Watts, 1987); (Guay, 1995):

$$\boldsymbol{\eta}(\boldsymbol{\beta}) = [\beta_1 + \beta_2 x_{1i}, \beta_1 + \beta_2 x_{2i}, \beta_1 + \beta_2 x_{3i}];\ (i = 1,2, \dots, n) \tag{3-42}$$

The within-run noise is assumed to be distributed as $\mathrm{MVN}(0, \boldsymbol{\Sigma})$, and between-run noise is considered to be *IID Normal*. In their work, Bates and Watts (1987) simulated a dataset (presented in Table 3-1) corresponding to this model for the following parameter values (referred to as '*true*' parameter values), adding noise with a known covariance matrix ($\boldsymbol{\Sigma}$):

$$\boldsymbol{\beta}_{true} = [0.5\ 2]' \tag{3-43}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.71 & 0.89 \\ 0.71 & 2 & 0.63 \\ 0.89 & 0.63 & 5 \end{bmatrix} \tag{3-44}$$

Table 3-1: Dataset corresponding to Bates and Watts multi-response example (Bates and Watts, 1987)

x_1	x_2	x_3	y_1	y_2	y_3
0.4076	0.5505	0.2129	-0.2875	0.4633	0.0423
0.8408	0.4085	0.5591	3.5302	3.5414	2.8750
0.5017	0.7217	0.4305	0.5428	1.5101	-1.7595
0.9129	0.4769	0.0228	3.7360	3.4240	2.1991
0.4436	0.6393	0.5761	0.8253	0.2383	1.0350
0.5984	0.9964	0.7149	2.0930	1.0528	2.2779
0.7742	0.1575	0.9322	3.2956	1.5196	4.4792
0.7922	0.5351	0.1233	2.6799	3.3145	-0.7375

They used the DC approach for parameter estimation and reported the following parameter estimates ($\widehat{\boldsymbol{\beta}}$) as well as a local minimum ($\widehat{\boldsymbol{\beta}}_{local}$):

$$\hat{\beta} = [0.41\ 2.55]' \tag{3-45}$$

$$\hat{\beta}_{local} = [1.23 - 0.31]' \tag{3-46}$$

They showed that even for a simple multi-response model such as this, the DC parameter estimates might vary from the true parameter values due to the inherent nonlinearity of the estimation problem, deviation of the determinant contours from the elliptical shape, and possibility of the solution converging to local minima.

For the intended purposes in this thesis, the dataset was treated as a set of "experimental observations" for parameter estimation, ignoring the true parameter values. The noise covariance matrix was considered to be unknown. The parameter estimation and the 95% parameter likelihood results from the proposed approach (*GLS-based profiling approach*

#2), along with *GLS-based profiling approach #1* discussed earlier in this chapter are summarized in Table 3-2 with the corresponding profile plots provided in Figures 3-2 and 3-3. In these figures, the PL_S function has been plotted versus each profiling parameter and reference lines based on $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ are represented by dotted lines and dot-dashed lines respectively.

According to Table 3-2, parameter estimates and inferences from *GLS-based profiling approaches #1 and #2* are fairly close to each other which is to be expected due to the linear structure of the model. Also, Figures 3-2 and 3-3 suggest that the profile plots for the GLS-based approaches are quite symmetric and parabolic, indicating that for this model when the GLS-based profiling is used with a fixed $\boldsymbol{\Omega}$, the estimation problem becomes linear.

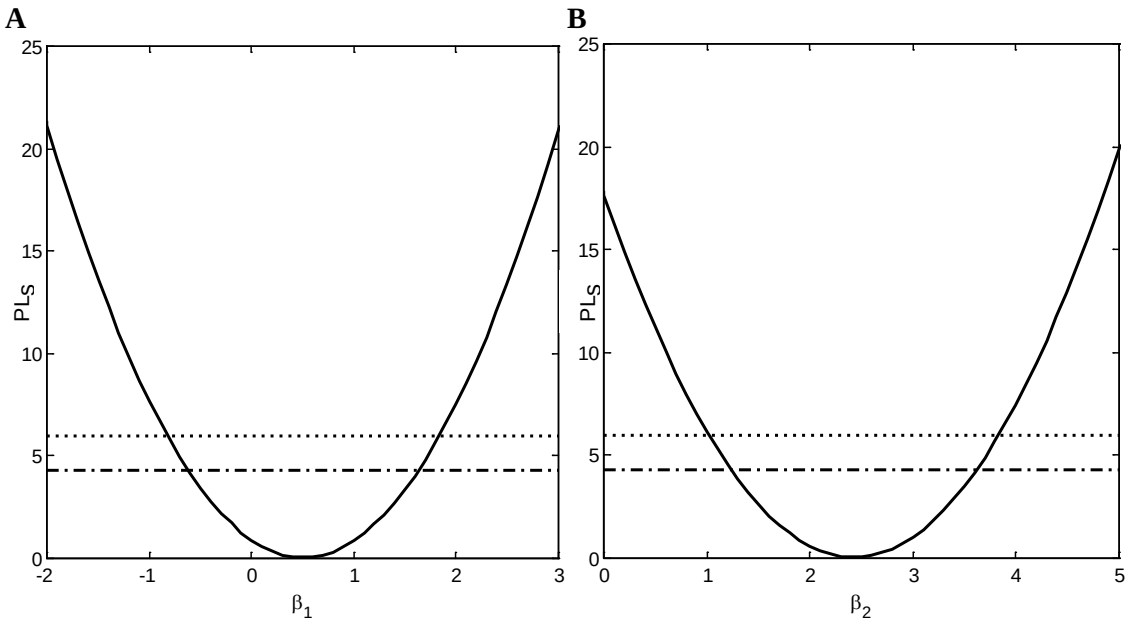

Figure 3-2: Profile likelihood plots for Case study 1 using GLS-based profiling approach #1 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Plot A: PL_S function versus β_1. Plot B: PL_S function versus β_2.

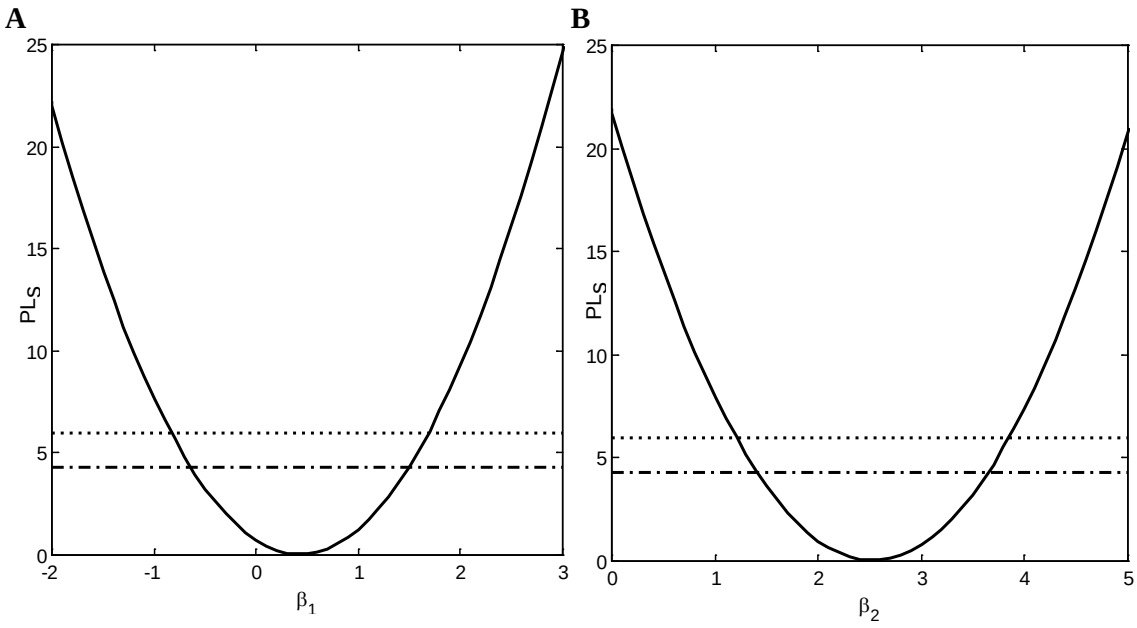

Figure 3-3: Profile likelihood plots for Case study 1 using GLS-based profiling approach #2 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Plot A: PL_S function versus β_1. Plot B: PL_S function versus β_2.

Table 3-2: Parameter estimation and inference results from different approaches for Case study 1 with true parameter values $\boldsymbol{\beta_{true}} = [0.5\ 2]'$

Approach	Parameter estimates	95% Likelihood intervals	
GLS-Based Profiling Approach #1 (with Fixed $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLSapproach1}} = [0.5072\ 2.4249]'$	based on $F_{1,n-p;0.95}$	$-0.84 \leq \beta_{1,GLSapproach1} \leq 1.8$ $1.25 \leq \beta_{2,GLSapproach1} \leq 3.8$
		based on $F_{1,nm-p;0.95}$	$-0.64 \leq \beta_{1,GLSapproach1} \leq 1.6$ $1.2 \leq \beta_{2,GLSapproach1} \leq 3.6$
GLS-Based Profiling Approach #2 (with Iterated $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLSapproach2}} = [0.4280\ 2.5283]'$	based on $F_{1,n-p;0.95}$	$-0.84 \leq \beta_{1,GLSapproach2} \leq 1.7$ $1.2 \leq \beta_{2,GLSapproach2} \leq 3.85$
		based on $F_{1,nm-p;0.95}$	$-0.64 \leq \beta_{1,GLSapproach2} \leq 1.5$ $1.4 \leq \beta_{2,GLSapproach2} \leq 3.6$
GLS-Based Profiling Approach with Known $\boldsymbol{\Sigma}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS,known\ \Sigma}} = [0.3923\ 2.5329]'$	based on $\chi^2_{1;0.95}$	$-0.9 \leq \beta_{1,GLS,known\ \boldsymbol{\Sigma}} \leq 1.75$ $0.8 \leq \beta_{2,GLS,known\ \boldsymbol{\Sigma}} \leq 4.35$

In the next step, parameter inference was carried out one more time using the known covariance matrix of equation 3-44 assumed by Bates and Watts. These results are also listed in Table 3-2 (referred to as *'GLS-based profiling approach with known $\boldsymbol{\Sigma}$'*). The corresponding profile plots of $PL_S(\beta_k)$ versus profiling parameters β_k with known $\boldsymbol{\Sigma}$ are provided in Figure 3-4 with reference lines based on $\chi^2_{1;0.95}$. The profiles of *GLS-based profiling approach with known $\boldsymbol{\Sigma}$* provide a reference standard, and again it can be seen that the profiles are quadratic, indicating that for fixed noise covariance matrix the estimation problem is linear for this example.

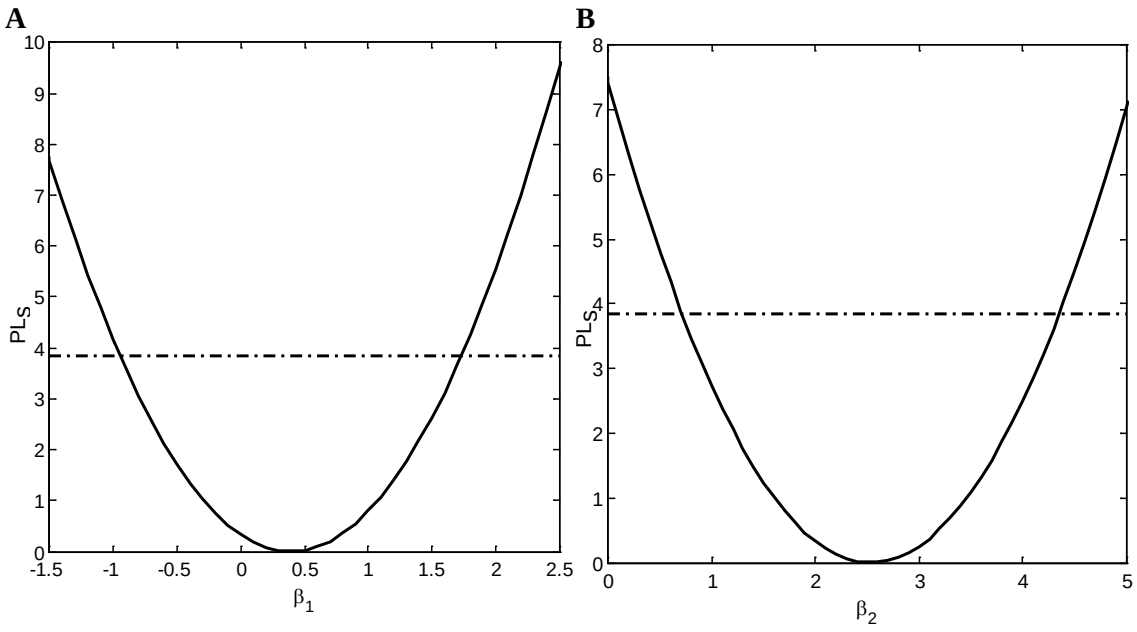

Figure 3-4: Profile likelihood plots for Case study 1 using GLS-based profiling with known **Σ** with reference lines at $\chi^2_{1;0.95}$ (dot-dashed line). Plot A: PL_S function versus β_1. Plot B: PL_S function versus β_2.

It can be seen that likelihood intervals from both *GLS-based profiling approaches #1 and #2* are almost identical and slightly smaller than the reference intervals provided by *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$. The estimated noise covariance matrix using the proposed profiling approach (*GLS-based profiling approach #2*) for this example is:

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.9724 & 1.1355 & 1.2397 \\ 1.1355 & 2.4497 & 1.1575 \\ 1.2397 & 1.1575 & 2.6718 \end{bmatrix} \tag{3-47}$$

This equation indicates that the estimated variances associated with the first and second responses are fairly close to the ones used by Bates and Watts (1987). However, the estimated variance associated with the third response $\left(\hat{\sigma}_3{}^2 = 2.6718\right)$ is noticeably smaller than the value used by Bates and Watts $(\sigma_3{}^2 = 5)$. As a result the profile

likelihood curves grow more rapidly for *GLS-based profiling approach #2* compared to *GLS-based approach with known* $\boldsymbol{\Sigma}$.

In order to compare the profiling results with bootstrap results, the approximate 95% bootstrap confidence intervals were generated from a sample size of 1000 for both parameters using the GLS method with updated noise covariance matrix during parameter estimation:

$$-3.0880 \leq \beta_{1,bootstrap} \leq 1.4298 \quad (3\text{-}48)$$

$$0.1373 \leq \beta_{2,bootstrap} \leq 8.3281 \quad (3\text{-}49)$$

Based on these results, the bootstrap confidence intervals are much wider than the reference intervals provided by *GLS-based approach with known* $\boldsymbol{\Sigma}$ or the intervals obtained from *GLS-based profiling approaches #1 and #2*.

In conclusion, the results of this case study indicate that for such cases, the profiles of the GLS-based approaches are quite similar, even in cases where the covariance matrix is estimated and therefore, there might not be a clear "best" approach for such examples. In fact when $\boldsymbol{\Sigma}$ is fixed for GLS-based profiling, the estimation problem becomes linear. Results from the DC approach will be provided in the next chapter. Since in the DC approach the noise covariance matrix is constantly changing, the estimation problem remains nonlinear.

For this example, the parameter intervals obtained from the proposed profiling approach (*GLS-based profiling approach #2*) are very close to the reference parameter intervals provided by *GLS-based approach with the known* $\boldsymbol{\Sigma}$ from the literature. The results from the proposed profiling approach are also much closer to the reference intervals than the results from the bootstrap confidence intervals. Overall, these results validate the

suitability of the proposed profile-based parameter inference approach for this example. The next case study includes a more nonlinear model with profile plots with pronounced asymmetries.

3.5.2 Case Study 2: Buzzi Ferraris Multi-Response Model and Dataset

This case study demonstrates the application of the proposed GLS-based parameter inference approach to a dataset and multi-response model presented by Buzzi Ferraris et al. (1984). In their work regarding model discrimination, they generated a set of synthetic data for the following kinetic model with two responses (y_1 and y_2), two regressor variables (x_1 and x_2) and four parameters (θ_1, θ_2, θ_3, θ_4) (Buzzi Ferraris et al., 1984):

$$y_{ij} = f_j(\boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_{ij} \tag{3-50}$$

with the model functions defined by:

$$f_1(\boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\theta_1 x_{1i} x_{2i}}{1+\theta_3 x_{1i}+\theta_4 x_{2i}} \tag{3-51}$$

$$f_2(\boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\theta_2 x_{1i} x_{2i}}{1+\theta_3 x_{1i}+\theta_4 x_{2i}} \tag{3-52}$$

The dataset for this example is provided in Table 3-3. In order to create the dataset, they inserted the following parameter values into the two model functions of equations 3-51 and 3-52 and added noise to the responses:

$$\boldsymbol{\theta} = [0.1\ 0.01\ 0.1\ 0.01]' \tag{3-53}$$

Thus, these values are considered to be the *true* parameter values. Ignoring noise co-dependencies, they assumed that the noise was *IID Normal* between runs and that the within-run noise covariance matrix was known, diagonal and equal to:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.35 & 0 \\ 0 & 0.0023 \end{bmatrix} \tag{3-54}$$

Table 3-3: Dataset for Buzzi Ferraris multi-response example (Buzzi Ferraris et al., 1984)

x_1	x_2	y_1	y_2
20	20	3.61	0.43
30	20	5.42	0.44
20	30	5	0.64
30	30	7.5	0.66
25	25	5.73	0.55
25	15	3.80	0.33
25	35	7.30	0.79
15	25	4.90	0.35
35	25	5.90	0.71
55	32.8	9.15	0.93
55	55	13.74	1.34
10.6	55	6	0.7
16	55	8.2	0.84
5	5	0.45	0.04
5	55	3.27	0.44
55	5	1.47	0.125
55	55	14	1.37
10.6	55	5.5	0.55
16	55	8.4	0.7

In this work, the previously discussed approaches (*GLS-based profiling approaches #1 and #2*) were applied to this dataset and model, not taking into account the parameter values and noise covariance matrix used by Buzzi Ferraris et al. (1984). The results for parameter estimation and 95% parameter likelihood intervals from the three approaches

are summarized in Table 3-4 and the respective plots are shown in Figures 3-5 and 3-6. Reference lines based on $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ are represented by dotted lines and dot-dashed lines respectively.

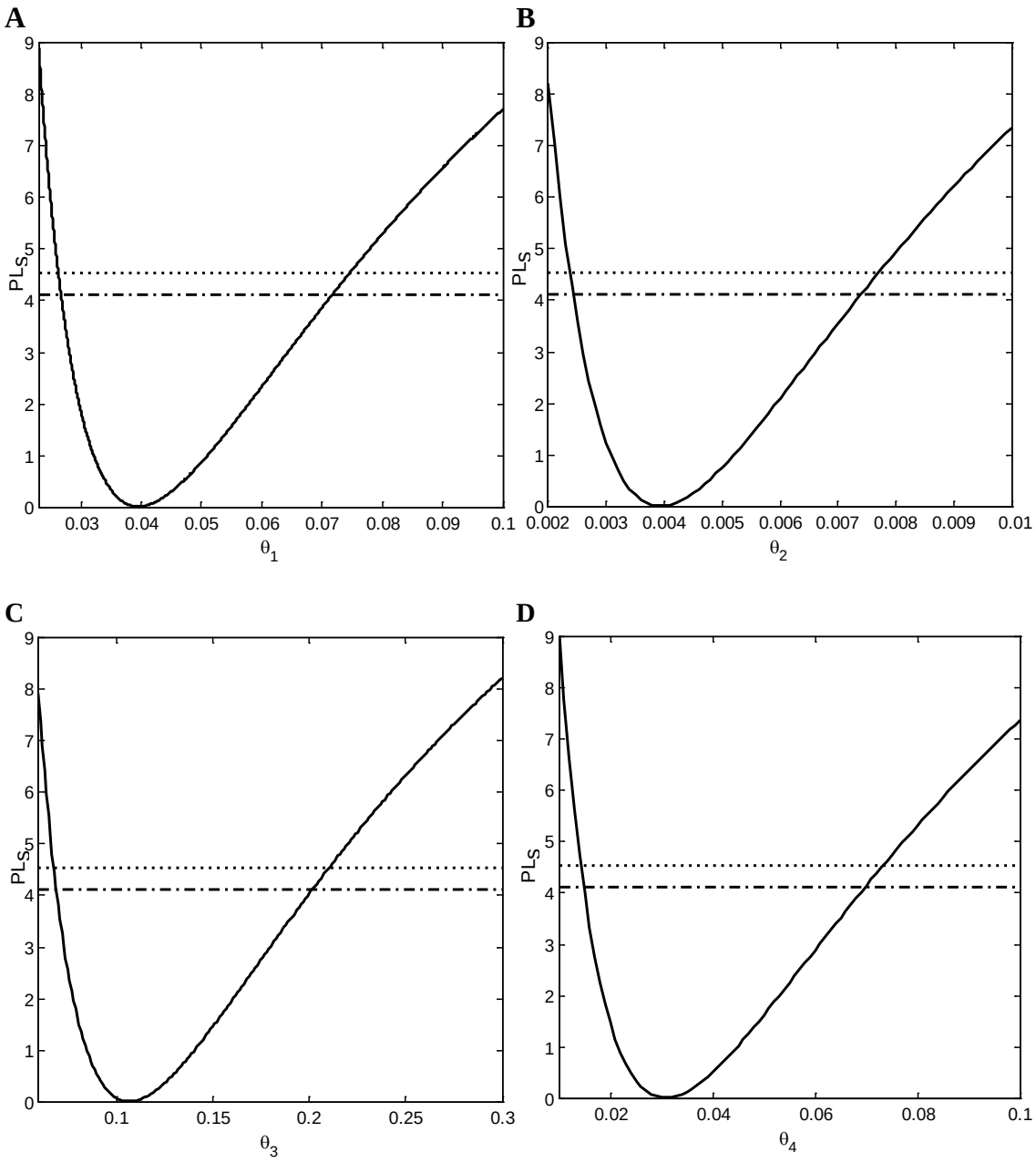

Figure 3-5: Profile likelihood plots for Case study 2 using GLS-based profiling approach #1 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Plot A: PL_S function versus θ_1. Plot B: PL_S function versus θ_2. Plot C: PL_S function versus θ_3. Plot D: PL_S function versus θ_4.

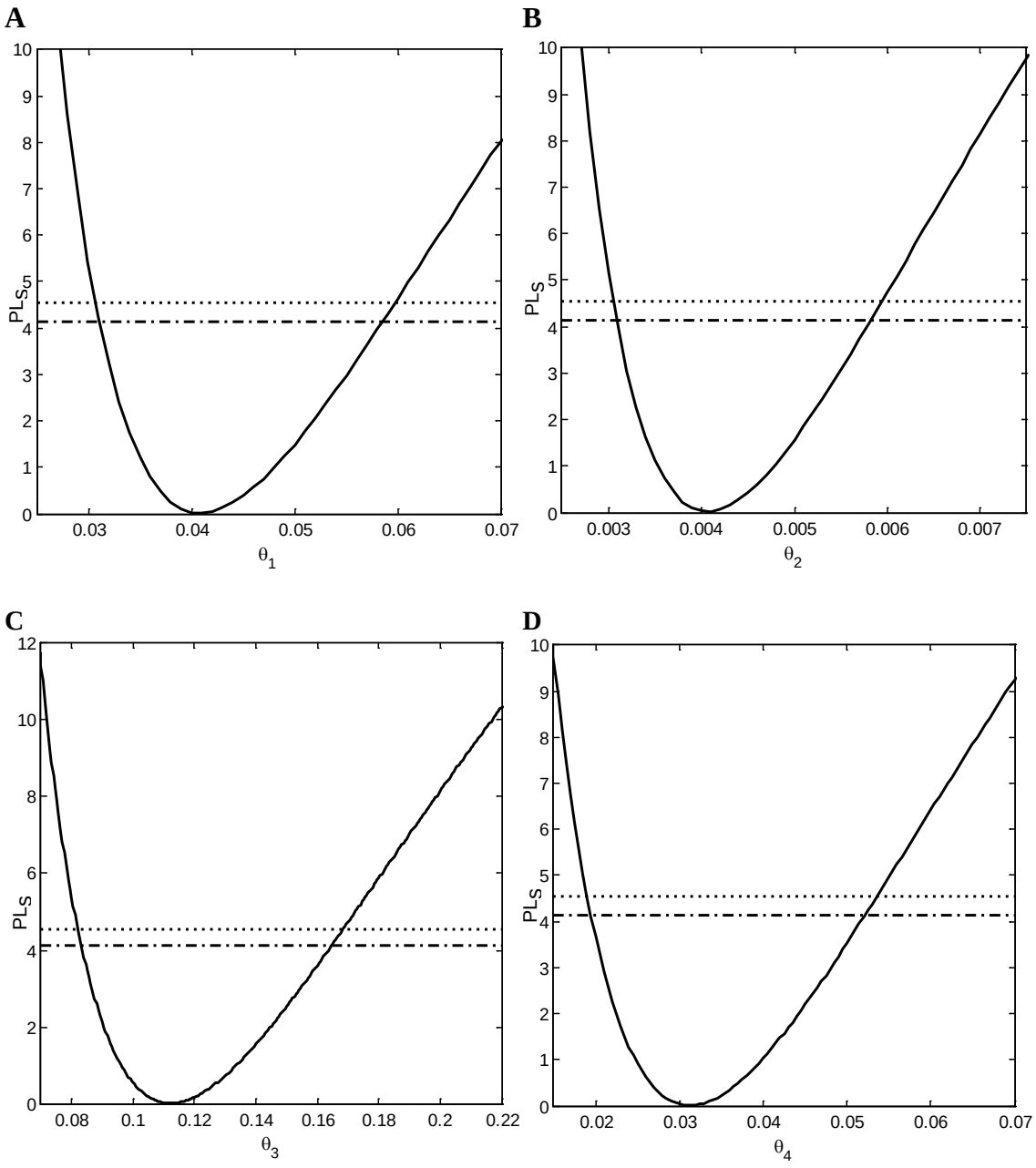

Figure 3-6: Profile likelihood plots for Case study 2 using GLS-based profiling approach #2 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Plot A: PL_S function versus θ_1. Plot B: PL_S function versus θ_2. Plot C: PL_S function versus θ_3. Plot D: PL_S function versus θ_4.

Table 3-4: Parameter estimation and inference results from different approaches for Case study 2 with true parameter values $\boldsymbol{\theta_{true}} = [0.1\ 0.01\ 0.1\ 0.01]'$

Approach	Parameter Estimates	95% Likelihood Intervals	
GLS-Based Profiling Approach #1 (with Fixed $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach1}} =$ $[0.0396\ 0.0039\ 0.1068\ 0.0308]'$	based on $F_{1,n-p;0.95}$	$0.026 \leq \theta_{1,GLSapproach1} \leq 0.075$ $0.0025 \leq \theta_{2,GLSapproach1} \leq 0.0077$ $0.068 \leq \theta_{3,GLSapproach1} \leq 0.22$ $0.014 \leq \theta_{4,GLSapproach1} \leq 0.073$
		based on $F_{1,nm-p;0.95}$	$0.027 \leq \theta_{1,GLSapproach1} \leq 0.072$ $0.0027 \leq \theta_{2,GLSapproach1} \leq 0.0075$ $0.07 \leq \theta_{3,GLSapproach1} \leq 0.2$ $0.015 \leq \theta_{4,GLSapproach1} \leq 0.07$
GLS-Based Profiling Approach #2 (with Iterated $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach2}} =$ $[0.0409\ 0.0041\ 0.1122\ 0.0313]'$	based on $F_{1,n-p;0.95}$	$0.031 \leq \theta_{1,GLSapproach2} \leq 0.059$ $0.0031 \leq \theta_{2,GLSapproach2} \leq 0.0059$ $0.082 \leq \theta_{3,GLSapproach2} \leq 0.17$ $0.018 \leq \theta_{4,GLSapproach2} \leq 0.054$
		based on $F_{1,nm-p;0.95}$	$0.032 \leq \theta_{1,GLSapproach2} \leq 0.058$ $0.0032 \leq \theta_{2,GLSapproach2} \leq 0.0057$ $0.083 \leq \theta_{3,GLSapproach2} \leq 0.17$ $0.019 \leq \theta_{4,GLSapproach2} \leq 0.052$
GLS-Based Profiling Approach with Known $\boldsymbol{\Sigma}$	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLS,known\ \Sigma}} =$ $[0.0419\ 0.0042\ 0.1161\ 0.0316]'$	based on $\chi^2_{1;0.95}$	$0.028 \leq \theta_{1,GLS,known\ \boldsymbol{\Sigma}} \leq 0.074$ $0.0028 \leq \theta_{2,GLS,known\ \boldsymbol{\Sigma}} \leq 0.0074$ $0.075 \leq \theta_{3,GLS,known\ \boldsymbol{\Sigma}} \leq 0.22$ $0.015 \leq \theta_{4,GLS,known\ \boldsymbol{\Sigma}} \leq 0.07$

Subsequently, the model and dataset were studied one more time with the covariance matrix considered to be known (based on equation 3-54) and the GLS approach was

applied for parameter estimation and inference with this matrix. The results are listed in Table 3-4 (referred to as '*GLS-based Profiling with Known* $\boldsymbol{\Sigma}$*'*) and the corresponding plots are provided in Figure 3-7 with reference lines based on $\chi^2_{1;0.95}$. The parameter intervals from *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$ are considered as reference results, and show moderate nonlinearity for this example.

According to Table 3-4, there is just a slight difference between the parameter estimates from *GLS-based profiling approaches #1 and #2*. All approaches tend to underestimate the first and second parameters, and overestimate the fourth parameter. The third parameter is estimated fairly accurately. As expected, the parameter likelihood intervals corresponding to $F_{1,nm-p;\alpha}$ are narrower than those associated with $F_{1,n-p;\alpha}$, because of the additional degrees of freedom.

The profile plots for all approaches are asymmetric, indicating that there is moderate nonlinearity in the parameter estimation problem regardless of the approach that is used. All of the profile plots are less steep on the right hand side, suggesting more noticeable nonlinearity for larger parameter values. This causes a larger difference between the upper likelihood bounds obtained from $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ than the lower bounds. There is a more pronounced difference in the upper inference limits for the *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$ in comparison to *GLS-based profiling approaches #1 and #2*.

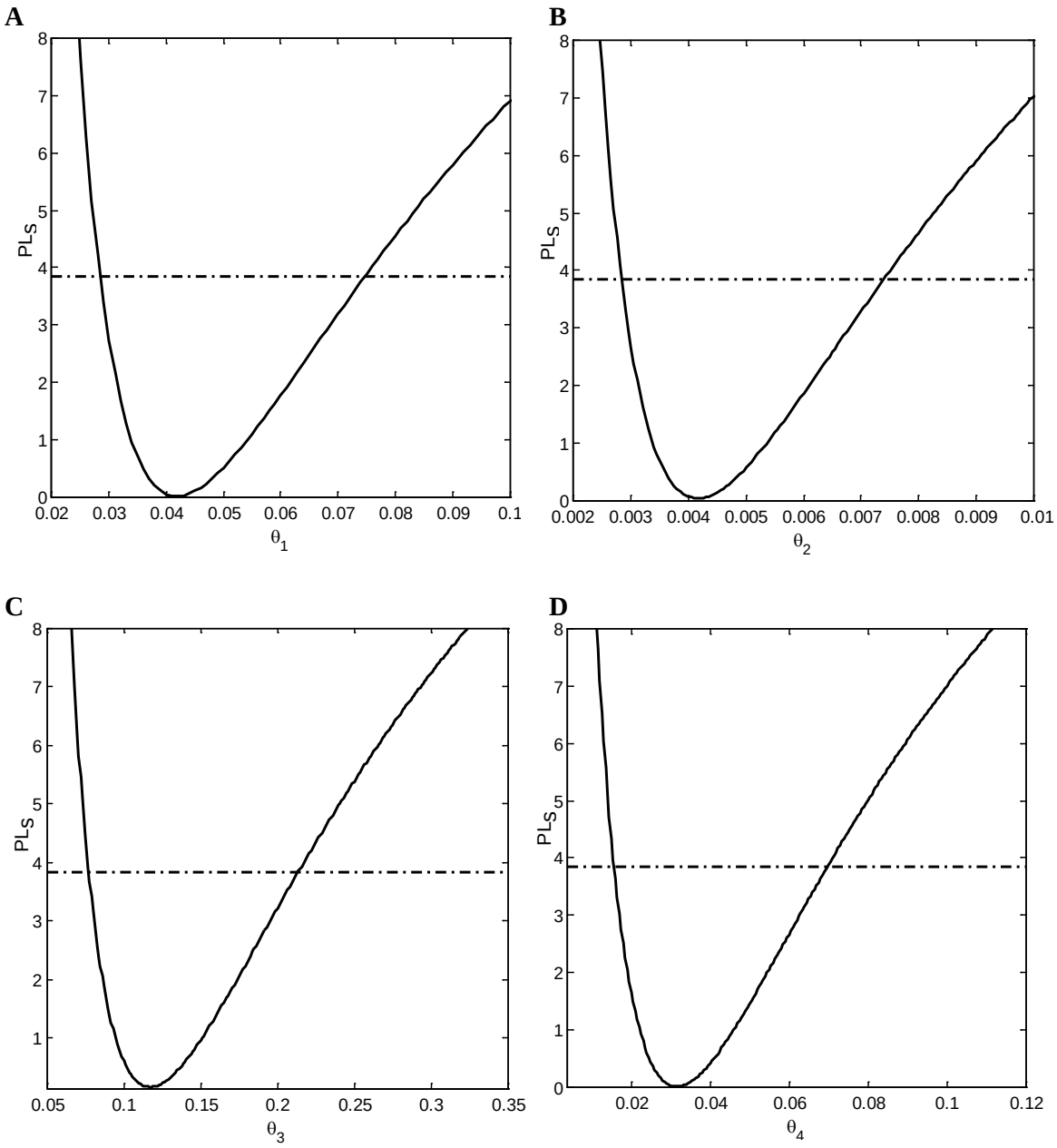

Figure 3-7: Profile likelihood plots for Case study 2 using GLS profiling with known **Σ** with reference lines at $\chi^2_{1;0.95}$ (dot-dashed line). Plot A: PL_S function versus θ_1. Plot B: PL_S function versus θ_2. Plot C: PL_S function versus θ_3. Plot D: PL_S function versus θ_4.

The inference results from the proposed profiling approach (*GLS-based profiling approach #2*) are fairly close to the reference intervals provided by *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$ and thus prove to be reliable for this example.

The estimated noise covariance matrix using the proposed profiling approach (*GLS-based profiling approach #2*) for this example is:

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.1497 & -0.0095 \\ -0.0095 & 0.0034 \end{bmatrix} \tag{3-55}$$

This equation indicates that the estimated variance associated with the second response is relatively close to the value used by Buzzi Ferraris et al. (1984). However, the estimated variance associated with the first response $\left(\hat{\sigma}_1{}^2 = 0.1497\right)$ is smaller than the known value $(\sigma_1{}^2 = 0.35)$. Therefore, the profile likelihood curves grow more rapidly and have greater values for *GLS-based profiling approach #2* compared to *GLS-based approach with known* $\boldsymbol{\Sigma}$. The estimated values for the within-run noise covariances are very small and comply with the results from Buzzi Ferraris et al. (1984) who ignored the within-run co-dependencies.

The comparison between the profiling method and an empirical inference method was done by generating the 95% approximate bootstrap parameter confidence intervals for this example:

$$0.0179 \leq \theta_{1,bootstrap} \leq 0.0421 \tag{3-56}$$

$$0.0019 \leq \theta_{2,bootstrap} \leq 0.0042 \tag{3-57}$$

$$0.0379 \leq \theta_{3,bootstrap} \leq 0.1167 \tag{3-58}$$

$$0.0068 \leq \theta_{4,bootstrap} \leq 0.0337 \tag{3-59}$$

The bootstrap inference intervals are much narrower than the reference intervals provided by *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$. Thus for this example, the proposed approach provides results that are more consistent with the reference results compared to the bootstrapping method.

In conclusion, the results indicate that for cases with asymmetric profile plots such as this example, the approach proposed in this thesis provides parameter estimates and likelihood intervals that are consistent with the reference results from the *GLS-based profiling with known* $\boldsymbol{\Sigma}$. Furthermore, the inference results from the proposed approach match the reference results more closely than the bootstrap method. Results from the DC approach will be provided in the next chapter and the two methods will be compared to each other.

3.5.3 Case Study 3: Beauchamp and Cornell Nonlinear Multi-Response Model and Dataset

This section describes the application of the proposed profile-based parameter inference approach to a multi-response system and dataset consisting of measurements of a radioactive tracer in human body carried out by Galambos and Cornell (1962). The dataset is presented in Table 3-5, with x_i representing measurement times (in hours), y_1 and y_2 representing the proportions of the injected radioactive sulphate tracer (in percentage) in two different compartments of the body.

Beauchamp and Cornell (1966) investigated this dataset and fitted the following model:

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{3-60}$$

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = [f_1(x_i, \boldsymbol{\theta}), f_2(x_i, \boldsymbol{\theta})];\ (i = 1,2, \dots, n) \tag{3-61}$$

with the model functions defined by:

$$f_1(x_i, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 x_i) + (1 - \theta_1) \exp(-\theta_3 x_i) \quad (3\text{-}62)$$

$$f_2(x_i, \boldsymbol{\theta}) = 1 - (\theta_1 + \theta_4) \exp(-\theta_2 x_i) + (\theta_1 + \theta_4 - 1)\exp(-\theta_3 x_i) \quad (3\text{-}63)$$

In these equations x_i are the measurement times and $\boldsymbol{\varepsilon}$ denotes the noise term. The fourth parameter (θ_4) is dependent on the other three by the following equation:

$$\theta_4 = [(\theta_3 - \theta_2)\theta_1(1 - \theta_1)]/[(\theta_3 - \theta_2)\theta_1 + \theta_2] \quad (3\text{-}64)$$

Table 3-5: Dataset for Beauchamp and Cornell's multi-response model (Galambos and Cornell, 1962).

x_i (hours)	y_1	y_2
2	0.84	0.10
3	0.79	0.14
5	0.64	0.21
8	0.55	0.30
12	0.44	0.40
24	0.27	0.54
48	0.12	0.66
72	0.06	0.71

The original dataset included an experiment with a missing response observation that I decided to leave out of the calculations (Galambos and Cornell, 1962). Beauchamp and Cornell (1966) studied the dataset containing missing data points so that the number of runs corresponding to the two responses were not the same and reported the following GLS-based parameter estimates:

$$\widehat{\boldsymbol{\theta}} = [0.06751\ 0.00706\ 0.08393\ 0.39506\ ]' \quad (3\text{-}65)$$

The variance for each response was calculated separately based on the number of data points available for that response leading to the following estimated noise covariance matrix (Beauchamp and Cornell, 1966):

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.00012 & 0.00001 \\ 0.00001 & 0.000032 \end{bmatrix} \tag{3-66}$$

Due to the lack of computer software at the time, Beauchamp and Cornell obtained the initial parameter estimates graphically using semi-logarithmic graph paper. With the improvements in the estimation software, this example seems to be an appropriate example to study in this thesis. In this research, the GLS-based profiling approaches for parameter estimation and inference are applied to this model and dataset ignoring the run with the missing point. The results from *GLS-based profiling approach #1 and #2* are presented in Table 3-6 and the related plots are provided in Figures 3-8 to 3-9.

Table 3-6: Parameter estimation and inference results from different approaches for Case study 3

Approach	Parameter Estimates	95% Likelihood Intervals	
GLS-Based Profiling Approach #1 (with Fixed $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach1}} =$ $[0.5531\ 0.0313\ 0.1713]'$	based on $F_{1,n-p;0.95}$	$0.43 \leq \theta_{1,GLSapproach1} \leq 0.68$ $0.023 \leq \theta_{2,GLSapproach1} \leq 0.039$ $0.13 \leq \theta_{3,GLSapproach1} \leq 0.24$
		based on $F_{1,nm-p;0.95}$	$0.41 \leq \theta_{1,GLSapproach1} \leq 0.66$ $0.025 \leq \theta_{2,GLSapproach1} \leq 0.037$ $0.14 \leq \theta_{3,GLSapproach1} \leq 0.23$
GLS-Based Profiling Approach #2 (with Iterated $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach2}} =$ $[0.1971\ 0.0131\ 0.116]'$	based on $F_{1,n-p;0.95}$	$0.13 \leq \theta_{1,GLSapproach2} \leq 0.26$ $0.01 \leq \theta_{2,GLSapproach2} \leq 0.016$ $0.101 \leq \theta_{3,GLSapproach2} \leq 0.134$
		based on $F_{1,nm-p;0.95}$	$0.142 \leq \theta_{1,GLSapproach2} \leq 0.25$ $0.0106 \leq \theta_{2,GLSapproach2} \leq 0.0155$ $0.103 \leq \theta_{3,GLSapproach2} \leq 0.13$

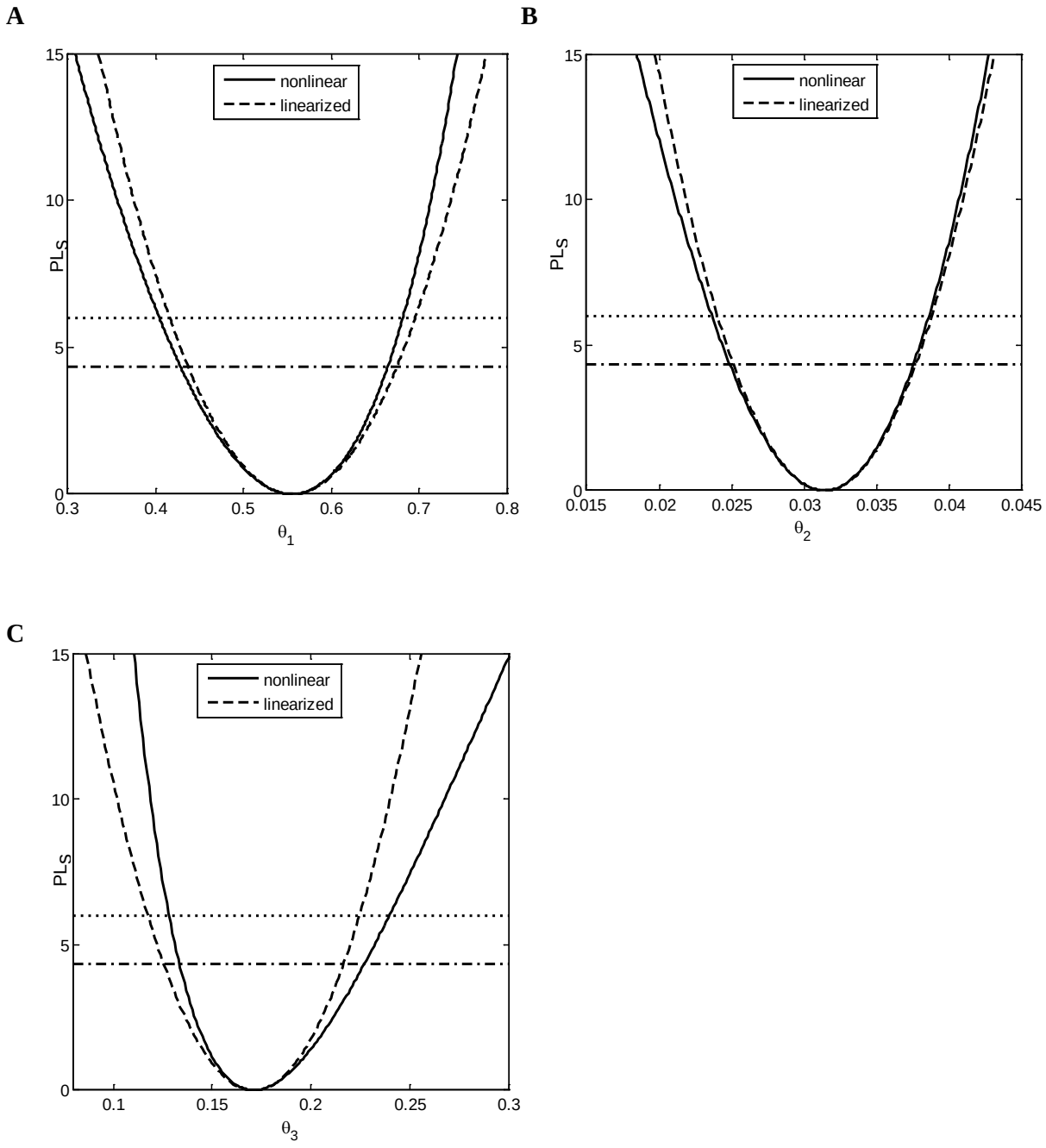

Figure 3-8: Profile likelihood plots for Case study 3 using GLS-based profiling approach #1 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Profile likelihood curve from linear approximation shown by dashed line. Plot A: PL_S function versus θ_1. Plot B: PL_S function versus θ_2. Plot C: PL_S function versus θ_3.

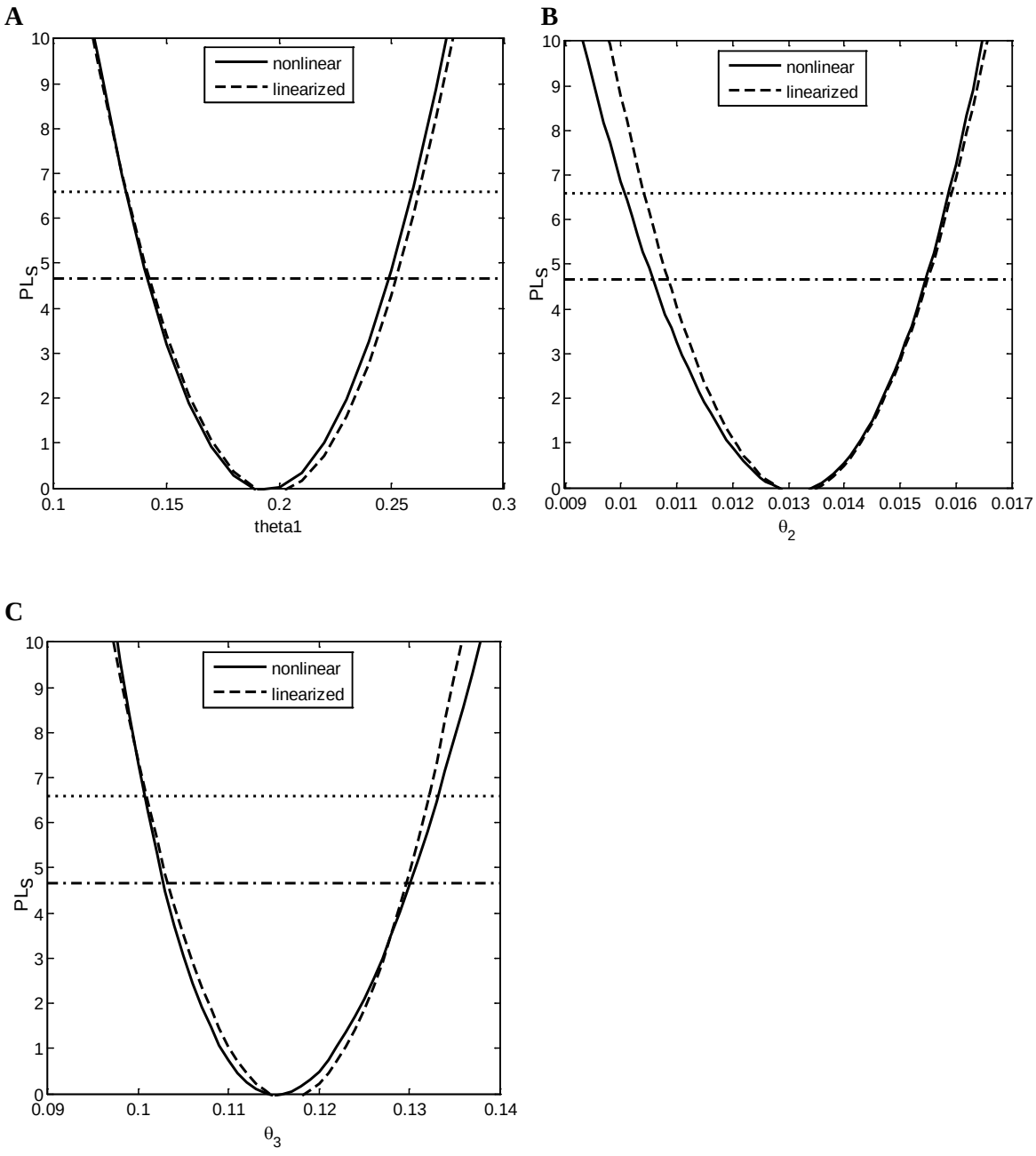

Figure 3-9: Profile likelihood plots for Case study 3 using GLS-based profiling approach #2 with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line). Profile likelihood curve from linear approximation shown by dashed line. Plot A: PL_S function versus θ_1. Plot B: PL_S function versus θ_2. Plot C: PL_S function versus θ_3.

The dashed curves on Figures 3-8 and 3-9 represent the profile likelihood function corresponding to the linear approximation of the model functions by means of a first-order Taylor series:

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\eta}\left(\widehat{\boldsymbol{\theta}}\right) + \boldsymbol{V}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \tag{3-67}$$

where $\boldsymbol{\eta}(\boldsymbol{\theta})$ consists of model functions as follows:

$$\eta_{ij}(\boldsymbol{\theta}) = f_j(x_i, \boldsymbol{\theta});\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{3-68}$$

and the sensitivity matrix $\boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ is evaluated at the least squares parameter estimates. The reason for including the linearization curves in the figures is to provide an example for application of profiling for showing the nonlinearity of the model. The quadratic linearization curves serve as references. The deviation of the nonlinear curves from these ideal quadratic curves shows the extent of nonlinearity of the model with respect to each parameter. This also indicates that the linearization methods and the assumptions involved might not produce sufficiently reliable results for nonlinear systems as seen in the results for *GLS-based profiling approach #1*.

Based on Table 3-6, even though the parameter estimates from the GLS methods differ from those reported by Beauchamp and Cornell (1966), there is a clear improvement in the estimation results (i.e., they fall closer to the parameter estimates provided in the literature) when the noise covariance matrix is updated during the parameter estimation process (*GLS-based profiling approach #2*). Furthermore, the parameter likelihood intervals corresponding to $F_{1,nm-p;\alpha}$ are slightly narrower than those associated with $F_{1,n-p;\alpha}$ due to the additional degrees of freedom.

The estimated within-run covariance matrix using *GLS-based profiling approach #2* for his example is:

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.002 & 0.0016 \\ 0.0016 & 0.0015 \end{bmatrix} \tag{3-69}$$

This matrix indicates relatively small variance and covariance estimates which are larger than the values estimated by Beauchamp and Cornell (1966) (equation 3-66). Since they do not use an iterative method to update the noise covariance matrix, their estimated $\widehat{\boldsymbol{\Sigma}}$ is more consistent with *GLS-based profiling approach #1*. However, since the dataset is fairly small, missing points have a big impact on the estimation outcomes. Thus, it might not be beneficial to compare their estimated $\widehat{\boldsymbol{\Sigma}}$ **(**that is formed from a larger dataset with missing data points) with the $\widehat{\boldsymbol{\Sigma}}$ estimated in this research (that is based on a smaller dataset ignoring the run with the missing data). In other words, since these two matrices are dealing with two different datasets, the differences in the results cannot just be attributed to calculations and estimation techniques.

Figure 3-8 indicates a noticeable difference between the curves obtained directly from profiling and the reference curves obtained from linearization. The difference is particularly pronounced in plot C corresponding to PL_S function with θ_3 as the profiling parameter. Based on Figure 3-9 the differences between the two curves become less evident but still important when using *GLS-based profiling approach #2*. As seen in both figures, the curves obtained from linearization provide more unreliable results as the profiling parameters move away from their least squares estimates.

Overall based on the results, for this example as well the proposed profile-based approach with reference line corresponding to $F_{1,nm-p;0.95}$ provides reliable parameter inference results and updating the noise covariance matrix during profiling leads to more accurate parameter estimation and inference results.

3.6 Discussions and Conclusions

Based on the case studies presented in this chapter, GLS-based profiling appears to be an appropriate method for parameter inference in nonlinear multi-response models. The GLS-based profiling approaches for such models lead to profile likelihood functions that are non-quadratic in parameters. For example, for the Buzzi Ferraris example (Case Study 2), the sum of squares function contains quadratic and cross-product terms of ratios of expressions in the parameters resulting from the model equations.

For all case studies there was a difference between the likelihood intervals obtained from the proposed GLS-based profiling approach using the estimated noise covariance matrix and the results from the case with known noise covariance matrix. This indicates the impact of noise covariance matrix estimation and the assumptions involved on the final inference outcomes, and also confirms the necessity of a good estimation procedure for this matrix. The iterative method for estimating the noise covariance matrix (*GLS-based profiling approach #2*) brought about some improvements in the estimation results which were particularly visible in the Beauchamp and Cornell example (Case Study 3).

In this chapter it was shown that for multi-response models with known noise covariance matrix, the profile-based objective function has a χ_1^2 distribution. For models with unknown noise covariance matrix, an iterative two-step GLS-based profiling approach (referred to as *GLS-based profiling approach #2*) was proposed for estimating the noise covariance matrix and calculating the profile likelihood function. This process involves estimating the parameters conditional on current iterates for the noise covariance matrix, followed by another step in which the noise covariance matrix is re-estimated conditional on the most recent iterates of the parameter estimates. This procedure is alternated until

convergence is reached. The distribution of the corresponding profile likelihood function generated by this method was studied and was shown to follow an $F_{1,nm-p}$ distribution. Determining this distribution is essential for using the profiling method for parameter inference since the reference lines that mark the boundaries of the parameter likelihood interval are formed based on this distribution.

Based on the case studies presented in this chapter, for models with unknown noise covariance matrix, the proposed iterative two-step GLS-based profiling approach with updated noise covariance matrix during parameter estimation with an $F_{1,nm-p}$ distribution for the profile likelihood function generated reliable inference results. The examples confirmed that considering $nm-p$ degrees of freedom as opposed to $n-p$ for the F distribution allowed taking advantage of nm pieces of information provided by the responses and produced likelihood intervals that closely matched the reference intervals.

3.7 Summary

In this section the GLS-based profiling was studied and an approach for obtaining likelihood intervals based on the profile plots was proposed. The importance of the noise covariance matrix and its estimation process and the assumptions involved were discussed. Different distributions for the profile likelihood plots were considered and the most relevant one to nonlinear multi-response models was identified. It was demonstrated that the correct degrees of freedom associated with the distribution of the likelihood function for multi-response models is $nm-p$ as opposed to $n-p$ that has been previously used. This chapter proposed a GLS-based profiling method for nonlinear multi-response models with unknown noise covariance matrix (referred to as *GLS-based*

profiling approach #2) and applied it to parameter inference. It was shown that the profile likelihood function obtained from this method follows an $F_{1,nm-p}$ distribution. It was suggested that the *GLS-based profiling approach #2* with reference lines at $F_{1,nm-p;\alpha}$ generates reliable parameter inference results for such models. The proposed approach was applied to several case studies, all of which confirmed the validity of the method.

The main contributions of Chapter 3 are as follows:

- Proposing a GLS-based profiling method for nonlinear multi-response models with unknown noise covariance matrix (referred to as *GLS profiling approach #2*) and defining the GLS-based profile likelihood function as $PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}})$
- Highlighting the importance of the estimation of noise covariance matrix and the assumptions involved and suggesting a method for estimating and updating the noise covariance matrix during parameter estimation and profiling for the proposed approach
- Identifying the distribution and the corresponding degrees of freedom for the PL_S function, taking into account the estimation process of the noise covariance matrix and suggesting χ_1^2 and $F_{1,nm-p}$ distributions for cases with known or unknown noise covariance matrix respectively
- Exploring the application of the proposed GLS-based profiling method in parameter inference for multi-response models

Chapter 4

Profiling Based on Determinant Criterion Method

4.1 Introduction

In Chapter 2, two common methods for parameter estimation in multi-response models were described, namely the Generalized Least Squares (GLS) and the Determinant Criterion (DC) approaches. The GLS method was studied in detail in the context of profiling in Chapter 3. In this chapter, the DC method is studied more closely, focusing on DC-based profiling method for parameter inference.

It was explained in Chapter 2 section 2.6 that the DC-based profile *t* function for multi-response models has been defined by Soo and Bates (1996) and used for parameter inference. However, there is still some difference of opinion regarding the distribution of the profile likelihood function. In this chapter, the DC-based profile likelihood function is studied and applied to parameter inference by determining the distribution and the corresponding degrees of freedom. The noise characteristics and how the noise covariance matrix is estimated during DC-based profiling for models with unknown $\boldsymbol{\Sigma}$ are also discussed.

The application of the DC-based profiling for parameter inference is demonstrated using several illustrative examples and the outcomes are compared to the reference likelihood intervals from literature where applicable to give an idea of how reliable the DC-based results are. The inference results for these examples are also provided using the proposed GLS-based profiling method (*GLS-based profiling approach #2* explained in Chapter 3).

Lastly, the *DC-based profiling approach* is compared against *GLS-based profiling approach #2* based on the literature and the results obtained in this research, revealing the advantages and disadvantages of each method to assist modelers in choosing the more suitable method based on the type of model, available data and purpose of modeling.

4.2 Formulation of the Determinant Criterion for Parameter Estimation and Inference

As mentioned in Chapter 2, another method used for parameter estimation in multi-response models is the Determinant Criterion (DC) approach. In the development of the DC, estimates for the noise covariance matrix conditional on the parameter estimates are substituted into the likelihood function to form a concentrated likelihood function. Applying the maximum likelihood criterion to this concentrated likelihood function leads to the Determinant Criterion (see for example, (Seber and Wild, 1989)). The matrix representation of a multi-response model with *p* parameters, *m* responses and *n* experimental runs is:

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{4-1}$$

In this equation $\boldsymbol{Y}$ is the $n \times m$ response matrix, and $\boldsymbol{\eta}(\boldsymbol{\theta})$ is the matrix of the expectation functions $\left(\boldsymbol{\eta}(\boldsymbol{\theta}) = [\eta_{ij}(\boldsymbol{\theta})]\right)$, with each element defined by:

$$\eta_{ij}(\boldsymbol{\theta}) = f_j(\boldsymbol{x}_i, \boldsymbol{\theta});\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{4-2}$$

The residual matrix $\boldsymbol{Z}(\boldsymbol{\theta})$ can be obtained from:

$$\boldsymbol{Z}(\boldsymbol{\theta}) = \boldsymbol{Y} - \boldsymbol{\eta}(\boldsymbol{\theta}) \tag{4-3}$$

The parameter estimates for this model consist of values of $\boldsymbol{\theta}$ that optimize some criterion based on the residual matrix $\boldsymbol{Z}(\boldsymbol{\theta})$ in the same way that the least squares parameter estimates minimize the square of the norm of the residual vector $\|\boldsymbol{z}(\boldsymbol{\theta})\|^2$ in single-response models (Bates and Watts, 1988). The criterion depends on the assumptions about the noise covariance terms. For example, if the noise terms are assumed to be *IID Normal* with the same variance σ^2, then the estimation problem becomes very similar to a single-response problem and the parameter estimates can be obtained from minimizing $tr(\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta}))$. However, for the more general case where the disturbance terms are co-dependent and the variances of different responses are not necessarily the same, a noise covariance matrix should be formed and incorporated into the parameter estimation criterion.

The importance of the noise covariance matrix and the common assumptions in its estimation process were discussed in Chapter 3 section 3.2. One of the most appropriate assumptions in this regard is to acknowledge the co-dependency between the noise in different responses in each run (within-run co-dependency), but assuming *IID Normal* noise between the runs (e.g., (Biegler et al., 1986); (Guay and McLean, 1995); (Benabbas et al., 2005)):

$$E\{e_{ij}e_{rs}\} = \begin{Bmatrix} \sigma_{js}, i = r \\ 0\ , i \neq r \end{Bmatrix};\ (i, r = 1,2,\dots,n;\ \ j, s = 1,2,\dots,m) \qquad (4\text{-}4)$$

The within-run noise is assumed to be distributed as $\mathrm{MVN}(0, \boldsymbol{\Sigma})$ and the $m \times m$ within-run noise covariance matrix $\boldsymbol{\Sigma}$ is formed from the σ_{js} elements. Thus, under the assumption of *IID Normal* between-run noise, the log-likelihood function is written as (Bates and Watts, 1988):

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{y}) = -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}tr(\boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}(\boldsymbol{\theta})') \quad (4\text{-}5)$$

Differentiating this expression with respect to elements of the noise covariance matrix (σ_{js}), results in the following:

$$\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\Sigma}|\boldsymbol{y})}{\partial \sigma_{js}} = -\frac{n}{2}\frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \sigma_{js}} - \frac{1}{2}\frac{\partial(\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta}))}{\partial \sigma_{js}} \quad (4\text{-}6)$$

Setting this derivative to zero for optimization purposes leads to the following expression for estimating the within-run noise covariance matrix conditional on the parameter values:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta}) \quad (4\text{-}7)$$

Finally, substituting this estimate into the log-likelihood function expression of equation 4-5, the concentrated log-likelihood function is defined as (Box and Draper, 1965); (Seber and Wild, 1989):

$$L\left(\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\Sigma}}|\boldsymbol{Y}\right)\right) = -\frac{n}{2}\ln|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})| \quad (4\text{-}8)$$

Thus, the DC-optimal parameter estimates are obtained by minimizing the Determinant function $D(\boldsymbol{\theta})$ defined as follows with respect to $\boldsymbol{\theta}$:

$$D(\boldsymbol{\theta}) = |\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})| \quad (4\text{-}9)$$

Similarly, in a Bayesian context the following relationship applies to the marginal posterior density function $p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$:

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\Sigma})p(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \quad (4\text{-}10)$$

where $p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is equivalent to the likelihood function $L(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{Y})$, and $p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is the prior density function of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. In a Bayesian argument by Box and Draper (1965), independent non-informative priors for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are used so that (Box and Draper, 1965); (Seber and Wild, 1989):

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = p(\boldsymbol{\theta})p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(m+1)/2} \quad (4\text{-}11)$$

This equation is combined with the log-likelihood function of equation 4-5 to provide the posterior density function under the assumption that each row of the residual matrix follows an *MVN* distribution. Solving equation 4-11 for the conditional estimate for covariance by differentiating the probability density function with respect to the covariance terms, obtaining conditional estimates for covariance terms, and substituting the estimated covariances into the posterior density function leads to the following expression for the posterior density function (Box and Draper, 1965) :

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto |\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|^{-n/2} \quad (4\text{-}12)$$

Again the parameter estimates are obtained by minimizing $D(\boldsymbol{\theta})$ for the Maximum a Posteriori (MAP) estimates (Bates and Watts, 1988); (Seber and Wild, 1989); (Kang and Bates, 1990). These results have the advantage of being very general in a sense that they do not depend on the linearity of the expectation function, or how the parameters and regressor variables are shared between responses (Box and Tiao, 1973). Phillips (1976) stated that DC is numerically equivalent to an iterated GLS. He showed that by applying an iterative method on the concentrated likelihood function, if $\widehat{\boldsymbol{\Sigma}}$ is updated in each iteration to match the most current residuals, the solution will arrive at the same results as the DC method. Bates and Watts (1984) proposed a generalized Gauss-Newton method for minimizing the Determinant function by deriving the gradient and Hessian of $D(\boldsymbol{\theta})$ explicitly.

For inference, using a quadratic approximation to $D(\boldsymbol{\theta})$, the posterior density function of equation 4-12 is approximated by a multivariate *t* distribution as follows (Bates and Watts, 1984); (Bates and Watts, 1985); (Bates and Watts, 1987):

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \left\{1 + \frac{(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})'\mathbf{H}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})}{2\left|\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})\right|}\right\}^{-n/2} \quad (4\text{-}13)$$

where **H** is the Hessian of $|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|$ evaluated at $\widehat{\boldsymbol{\theta}}$. Thus, as mentioned previously in Chapter 2 section 2.3, using the DC approach, an approximate $100(1-\alpha)$ standard marginal inference interval for parameter θ_k can be obtained from (Bates and Watts, 1988):

$$\hat{\theta}_k - t_{n-p,\alpha/2}s'\sqrt{2\{\mathbf{H}^{-1}\}_{kk}}, \hat{\theta}_k + t_{n-p,\alpha/2}s'\sqrt{2\{\mathbf{H}^{-1}\}_{kk}} \quad (4\text{-}14)$$

where $t_{n-p,\alpha/2}$ represents the upper $\alpha/2$ quantile of the Student's *t* distribution with $n-p$ degrees of freedom and the DC-based estimated average noise standard deviation (s') is the square root of the DC-based average noise variance that is usually obtained from (Bates and Watts, 1988):

$$s'^2 = \frac{D(\widehat{\boldsymbol{\theta}})}{n-p} \quad (4\text{-}15)$$

The approximate $100(1-\alpha)$ joint inference region is given by (Bates and Watts, 1988):

$$\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)'\frac{\mathbf{H}}{2}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \leq ps'^2F_{p,n-p,\alpha} \quad (4\text{-}16)$$

This equation provides inference regions that are elliptical and approximate since they are obtained from linearization. Another way to build the inference regions is to apply the following equation:

$$D(\boldsymbol{\theta}) - D(\widehat{\boldsymbol{\theta}}) \leq ps'^2F_{p,n-p,\alpha} \quad (4\text{-}17)$$

where $F_{p,n-p,\alpha}$ is the upper α quantile of the *F* distribution with p and $n-p$ degrees of freedom. Equations 4-14 and 4-16 can only provide approximate inference results for nonlinear models since they are based on linearization methods and imply symmetric inference intervals or elliptical inference regions which are not necessarily adequate.

Equation 4-17 provides inference regions that are not necessarily elliptical in general. However, the results are still approximate due to the significance level. Overall, these standard inference methods cannot always be trusted to provide adequate results. These issues have prompted the application of profiling for generating more reliable parameter inference intervals (Gerhard, 2010).

In equations 4-14, 4-16 and 4-17 the degrees of freedom associated with the Student's t distribution and F distribution are considered to be $n-p$. This is a result of an argument put forward by Bates et al. suggesting that the estimation problem in multi-response models is only sensitive to n pieces of information from the data in the neighbourhood of the DC-optimal parameter estimates (Bates and Watts, 1985); (Kang and Bates, 1990). However, I believe that $nm-p$ represents a more valid degrees of freedom for the general case that enables the estimation problem to take advantage of nm pieces of information provided by the responses. This was proven in Chapter 3 section 3.4 for the distribution of the GLS-based profile likelihood function. In this chapter, this argument is extended to the *DC-based profiling approach*, and the discussions on the degrees of freedom are presented after the DC-based profiling is described.

4.3 Determinant Criterion in the Context of Profiling

Profiling based on the Determinant Criterion method implicitly accounts for estimation and substitution of the noise covariance matrix since this is embedded in the development of the DC method. Consequently, the DC-based profiling can potentially represent the most accurate picture of the joint estimation of parameters and noise covariances.

However, a number of numerical issues are encountered, hinting at the challenges that also lie in wait for joint profiling on parameters and the covariance matrix.

As mentioned in Chapter 2, Soo and Bates (1996) built on the profiling concept of Bates and Watts by means of the Determinant Criterion (DC) approach for parameter estimation. They substituted the SSE function with the Determinant function and came up with the following DC-based profile *t* function for the multi-response model of equation 4-1 (Soo and Bates, 1996):

$$\tau_{DC}(\theta_k) = sign\left(\theta_k - \hat{\theta}_k\right)\frac{\sqrt{\left(\widetilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}})\right)}}{\sqrt{\frac{D(\widehat{\boldsymbol{\theta}})}{\nu}}} \tag{4-18}$$

where $\widetilde{D}(\theta_k)$ is the Determinant function evaluated at the DC-optimal parameter estimates conditional on the fixed value of θ_k and $D\left(\widehat{\boldsymbol{\theta}}\right)$ is the Determinant function evaluated at the DC-optimal parameter estimates. They considered ν to be equal to $n - p$ for the usual multi-response models. Alternatively, for nonparametric regression which is the focus of their paper, they applied a spline method for approximating the model functions and calculated the degrees of freedom from:

$$\nu = n(m(r + k) + r) \tag{4-19}$$

In this equation n is the number of runs, m denotes the number of responses, and r and k are the number of knots and the order of splines respectively. They have considered a Student's *t* distribution with ν degrees of freedom for this DC-based profile *t* function.

Following the work of Soo and Bates (1996), the "DC-based profile likelihood function" $PL_D(\theta_k)$ is written as:

$$PL_D(\theta_k) = \frac{\widetilde{D}(\theta_k) - D\left(\widehat{\boldsymbol{\theta}}\right)}{(s\prime)^2} \tag{4-20}$$

with s'^2 being the DC-based estimated average noise variance which is often obtained from $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}})}{(n-p)}$ (equation 4-15).

In Chapter 3, three approaches were studied for GLS-based profiling by viewing the profiling procedure as a combination of a parameter estimation step and a profiling step. These approaches differ from each other based on how the noise covariance matrix is estimated and updated during the profiling procedure. Similarly, using the same outlook on the two steps of the profiling procedure, a profiling method can be introduced based on DC which is explained in the following paragraphs.

DC-based profiling approach:

In this approach, the DC-optimal parameter estimates are obtained from the DC estimation method by minimizing the Determinant function $D(\boldsymbol{\theta}) = |\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|$ (equation 4-9). Subsequently, profiling is carried out in the neighborhood of the DC-optimal parameter estimates by means of the DC-based profile likelihood function $PL_D(\theta_k) = \frac{\tilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}})}{(s')^2}$ (equation 4-20). In this approach the noise covariance matrix is estimated implicitly from the residuals $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ (equation 4-7) and as a result the profiling becomes a joint estimation problem on parameters and the noise covariance matrix.

The main objective in this chapter is to investigate the application of the *DC-based profiling approach* in parameter inference for nonlinear multi-response models. In the section that follows, the distribution and the associated degrees of freedom that are currently used for the PL_D function according to the literature are stated, and it is

explained why a distribution with a different degrees of freedom matches this function more closely.

4.4 Distribution of the DC-based Profile Likelihood Function and the Associated Degrees of Freedom

As mentioned in Chapter 2, one of the main applications of profiling is in parameter inference and this technique is particularly beneficial for determining the parameter likelihood intervals in nonlinear models due to its ability to generate asymmetric inference intervals (Bates and Watts, 1988); (Gerhard, 2010). In order to obtain appropriate inference results from this method, it is important to find the correct distribution of the profile likelihood function and the corresponding degrees of freedom. Since the Determinant function is formed from the residuals, the distribution of the DC-based profile likelihood function depends on the distribution of the residuals. Using the matrix representation of the model (equation 4-1), the residual matrix is an $n \times m$ matrix:

$$\boldsymbol{Z}(\boldsymbol{\theta}) = [Z_{ij}(\boldsymbol{\theta})];\ (i = 1,2,\dots,n;\ \ j = 1,2,\dots,m) \tag{4-21}$$

which has an *MVN* distribution with zero mean and within-run noise covariance matrix $\boldsymbol{\Sigma}$ (obtained from equation 4-7). Assuming the model is correct, each row of $\boldsymbol{Z}(\boldsymbol{\theta})$ contains random variables that are jointly distributed as $MVN(\mathbf{0}, \boldsymbol{\Sigma})$, i.e.,

$$\boldsymbol{Z}_i(\boldsymbol{\theta}) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}) \tag{4-22}$$

where i denotes i^{th} the row. Therefore, the distribution of the resulting $D(\boldsymbol{\theta})$ in general should account for this multivariate distribution.

Based on the literature, an F distribution with $n-p$ for the second degrees of freedom is often associated with $D(\widehat{\boldsymbol{\theta}})$ and applied to DC-based estimations (Box and Draper, 1965); (Bates and Watts, 1985). More specifically for profiling, Soo and Bates (1996) have suggested a Student's t distribution with $n-p$ degrees of freedom for their proposed DC-based profile t function (equation 4-18). Thus, the DC-based profile likelihood function (equation 4-20) defined as the square of the profile t function is by analogy distributed as $F_{1,n-p}$. However, based on the earlier discussions, I suggest that this F distribution needs to be modified to accommodate the distribution of the residual matrix for multi-response models (equation 4-22).

Using the $n-p$ degrees of freedom for $D(\widehat{\boldsymbol{\theta}})$ is supported by an argument put forward by Bates and colleagues (Bates and Watts 1985); (Kang and Bates, 1990); (Soo and Bates, 1996). They have argued that in the neighborhood of the DC-optimal parameter estimates, the Determinant function is controlled by a single linear combination of the responses. Therefore, in spite of having nm pieces of information available, this function is only sensitive to n of them and the degrees of freedom is equal to $n-p$. However, this argument might not apply to all multi-response models in general based on the following reasons:

- According to Bates and Watts, the linear combination between the responses arises when the Determinant function is close to zero at its minimum ($D(\boldsymbol{\theta}) \approx 0$), which implies that the residual matrix $\boldsymbol{Z}(\boldsymbol{\theta})$ is rank deficient. With the determinant being close to zero, the determinant behaviour is dominated by the smallest singular value on the argument that it is associated with near-singular

(and zero-determinant) behavior and that it represents linear combination in *n*-dimensional space. However this might not be entirely true for all cases. For parameter estimation the objective is to minimize the Determinant function and this does not necessarily mean that the determinant value corresponding to the parameter estimates is zero. In fact for some of the examples that were studied in this thesis, the value of the Determinant function was very far from zero (e.g., $D(\widehat{\boldsymbol{\theta}})$ =568.5941 for Case study 1). Therefore, the argument that the behavior is dominated by a singular vector associated with an almost zero value (the smallest eigenvalue) does not hold in general.

- Another counter argument to $n-p$ is that zero or close to zero values for the Determinant function imply a rank deficiency in $\boldsymbol{Z}(\boldsymbol{\theta})$ which is not necessarily equal to one. This does not comply with their argument that a "single" linear combination of the responses dominates the Determinant function. In other words, there could be multiple very small (or effectively zero) singular values with associated singular vectors. The null space of $\boldsymbol{Z}(\boldsymbol{\theta})$ could in fact be spanned by several singular vectors, causing a rank deficiency greater than one.
- Furthermore, the assumption of $\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ being singular might cause problems in estimating the noise covariance matrix. If the model is correct and the dataset is large, $\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}}))/n$ is an asymptotically unbiased and consistent estimator for $\boldsymbol{\Sigma}$. For models with m responses and m sources of noise associated with them, $\widehat{\boldsymbol{\Sigma}}$ has to be non-singular and positive definite for the estimation process to be feasible. If $\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ is singular, estimation of $\widehat{\boldsymbol{\Sigma}}$ might lead to a positive semi-definite matrix i.e., $\widehat{\boldsymbol{\Sigma}} \geq 0$. Thus, there is an inherent contradiction in the

assumption of near-singular $\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ and the estimation of the noise covariance matrix.

Based on these reasons, $n-p$ degrees of freedom does not necessary apply to all cases. Considering $nm-p$ degrees of freedom for the distribution allows the model to benefit from nm pieces of information and factors in the p constraints associate with the DC-optimal values for parameter estimates. Thus, I believe that the distribution of the DC-based profile likelihood function is more appropriately represented by $F_{1,nm-p}$.

Finally, it was demonstrated in the previous chapter that the rank of the orthogonal projection operator in GLS for a multi-response model is $nm-p$, reinforcing the conclusion that the correct degrees of freedom is $nm-p$. Since Philips (1976) has shown that the iterated GLS and the DC approaches converge to the same results, one would expect that the degrees of freedom behaviour is the same for the parameter estimation conditional on an estimate of the noise covariance matrix. The difference is that in the case of iterated GLS, the dependence is explicit while in the case of DC, it is implicit.

Expanding the $nm-p$ argument, it should be noted that through physical constraints (for example if responses are mole fractions that sum to unity), the number of the independent responses might be less than *m* (Box et al., 1973); (McLean et al., 1979). Similar to the discussions about the distribution of the GLS-based profile likelihood function, the possibility of having fewer than m independent responses should be considered. The possibility of having linearly dependent responses and the issues associated with it were discussed in Chapter 3 section 3.4.3. The Eigenvalue Analysis by Box et al. (1973) for

determining these linear relationships and reducing the data to a linearly independent set of m' elements was also explained. This method does not guarantee that after the reduction of the linearly dependents responses there won't be any more singularities in the $\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})$ matrix since the singularities depend both on the data and the form of the model. Therefore, the eigenvalue analysis is not sufficient for determining the singularities and the $\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})$ should also be examined for singularities as well (McLean et al., 1979). This can be done by means of an analysis such as Singular Value Decomposition (SVD) in which the residual matrix is decomposed into three matrices as follows (Gentle, 2007):

$$\boldsymbol{Z}(\boldsymbol{\theta}) = \boldsymbol{USV}' \tag{4-23}$$

In this equation $\boldsymbol{U}$ is the output rotation matrix consisting of columns that are orthonormal eigenvectors of $\boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{Z}(\boldsymbol{\theta})'$. $\boldsymbol{V}$ is the input rotation matrix with columns that are orthonormal eigenvectors of $\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ and $\boldsymbol{S}$ is the diagonal matrix of singular values containing the square roots of eigenvalues from $\boldsymbol{U}$ or $\boldsymbol{V}$ in descending order. If $\boldsymbol{Z}(\boldsymbol{\theta})$ is singular, one or more of the eigenvalues of $\boldsymbol{Z}(\boldsymbol{\theta})\boldsymbol{Z}(\boldsymbol{\theta})'$ are equal to zero, causing the $\boldsymbol{S}$ matrix to contain zero values on its diagonal. Based on equation 4-23, these zero elements of $\boldsymbol{S}$ when multiplied by $\boldsymbol{U}$ and $\boldsymbol{V}$, will cancel out the corresponding columns and rows of these matrices. Thus, if the $\boldsymbol{S}$ matrix contains zero or nearly zero elements, the elements of the corresponding eigenvectors can be used as coefficients of the linear relationship that exists among the responses. By eliminating these elements from $\boldsymbol{S}$ and the corresponding columns and rows of $\boldsymbol{U}$ and $\boldsymbol{V}$, the problem will be reduced to a set of linearly independent responses.

Back to the degrees of freedom argument, once the response set is reduced to the linearly independent set of m' independent responses, the estimation and inference problems can take advantage of nm' pieces of information and as a result the degrees of freedom should be modified to $nm' - p$.

4.5 Illustrative Examples

In this section, the application of DC-based profiling is demonstrated using several standard multi-response examples. For the following examples, parameter inference is done using the *DC-based profiling approach* in which the parameters are estimated based on the DC estimation method and profiling is done by means of the DC-based profile likelihood function defined by equation 4-20. Following the discussion in the previous section, the reference lines for parameter inference are computed from $F_{1,n-p;\alpha}$ as well as $F_{1,nm-p;\alpha}$ to provide a clearer picture of how they compare. The profile plots are used to highlight the strengths and shortcomings of this profiling approach.

For further exploration, since these examples were already studied in Chapter 3 using the GLS method, the estimation results from the proposed iterative GLS-based method (*GLS-based profiling approach #2*) have been transferred here to find out how the two methods compare. For examples in which the dataset in the literature was originally synthesized by adding noise with a specific structure to model predictions based on some known parameter values (referred to as "the *true* parameter values" in the literature), the GLS-based and DC-based parameter estimation methods are compared against each other based on how closely the estimated parameters from each method match the *true* parameter values. For these synthetic examples, it is expected that the inference results

using the known noise covariance which was used for generating the data are the most representative for the system. Hence, the *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$ provides an appropriate reference for comparing the parameter intervals obtained from the DC-based and GLS-based profiling methods. Therefore, these profiling methods are compared based on their agreement with the outcomes of the *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$.

4.5.1 Case Study 1: Bates and Watts Multi-Response Model and Dataset

In this example the system consists of three response variables (y_1, y_2, and y_3), each being linearly related to two parameters (β_1 and β_2) by the following model (Bates and Watts, 1987):

$$\boldsymbol{\eta}(\boldsymbol{\beta}) = [\beta_1 + \beta_2 x_{1i}, \beta_1 + \beta_2 x_{2i}, \beta_1 + \beta_2 x_{3i}];\ (i = 1,2, \dots, n) \tag{4-24}$$

In this equation x_{ki} is the k^{th} regressor variable for the i^{th} run. In their work, Bates and Watts simulated a dataset (presented in Table 3-1) corresponding to this model for the following parameter values (referred to as *true* parameter values), by adding noise with a known within-run covariance matrix ($\boldsymbol{\Sigma}$) and assuming *IID Normal* between-run noise:

$$\boldsymbol{\beta}_{true} = [0.5\ 2]' \tag{4-25}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.71 & 0.89 \\ 0.71 & 2 & 0.63 \\ 0.89 & 0.63 & 5 \end{bmatrix} \tag{4-26}$$

Using this example, they showed some of the difficulties associated with applying the DC estimation to nonlinear multi-response models. Their research revealed that even for a simple multi-response model such as this, the DC-optimal parameter estimates might vary from the true parameter values due to the inherent nonlinearity of the estimation

problem, deviation of the determinant contours from the elliptical shape, and possibility of the solution converging to local minima. For more information on this example and the dataset please refer to Chapter 3, section 3.5.1.

In the present research, the dataset was treated as a set of "experimental observations" for parameter estimation and the *true* parameter values were not used. The parameter estimates and the 95% parameter likelihood results from the *DC-based profiling approach* along with *GLS-based profiling approach #2* (the proposed approach as explained in Chapter 3) are summarized in Table 4-1 and the corresponding DC-based profile plots are provided in Figure 4-1. In this figure, the solid curves are obtained from $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(n-p)}$ (equation 4-15), while the dashed curves are obtained from $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(nm-p)}$. For profile plots of the GLS-based approach please refer to Figure 3-3 of Chapter 3.

While investigating this example in Chapter 3, parameter inference was also carried out using the known noise covariance matrix of equation 4-26 assumed by Bates and Watts (referred to as '*GLS-based profiling approach with known* $\boldsymbol{\Sigma}$'). The results from this approach are also listed in Table 4-1. Since this dataset was originally generated using noise with this known covariance matrix, the results from this approach would provide a reference standard. For the profile likelihood plots corresponding to this approach please refer to Chapter 3, Figure 3-4.

Table 4-1: Parameter estimation and inference results from DC and some GLS profiling approaches for Case Study 1 with true parameter values $\boldsymbol{\beta_{true}} = [0.5\ 2]'$

Approach	Parameter estimates	95% Likelihood intervals	
GLS-Based Profiling Approach #2 (with Iterated $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLSapproach2}} = [0.4280\ 2.5283]'$	based on $F_{1,n-p;0.95}$	$-0.84 \leq \beta_{1,GLSapproach2} \leq 1.7$ $1.2 \leq \beta_{2,GLSapproach2} \leq 3.85$
		based on $F_{1,nm-p;0.95}$	$-0.64 \leq \beta_{1,GLSapproach2} \leq 1.5$ $1.4 \leq \beta_{2,GLSapproach2} \leq 3.6$
DC-Based Profiling Approach	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{DC}} = [0.4086\ 2.5549]'$	based on $F_{1,n-p;0.95}$	$-3 \leq \beta_{1,DC} \leq 2.47$ $-1.45 \leq \beta_{2,DC} \leq 7.8$
		based on $F_{1,nm-p;0.95}$	$-1.1 \leq \beta_{1,DC} \leq 1.5$ $-0.5 \leq \beta_{2,DC} \leq 4.7$
GLS-Based Profiling Approach with Known Σ	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS,known\ \Sigma}} = [0.3923\ 2.5329]'$	based on $\chi^2_{1;\mathbf{0.95}}$	$-0.9 \leq \beta_{1,GLS,known\ \boldsymbol{\Sigma}} \leq 1.75$ $0.8 \leq \beta_{2,GLS,known\ \boldsymbol{\Sigma}} \leq 4.35$

According to Table 4-1, the parameter estimates from the DC approach deviate from the true parameter values even though the example seems relatively simple. This is also the case with the parameter estimates from *GLS-based profiling approach #2*.

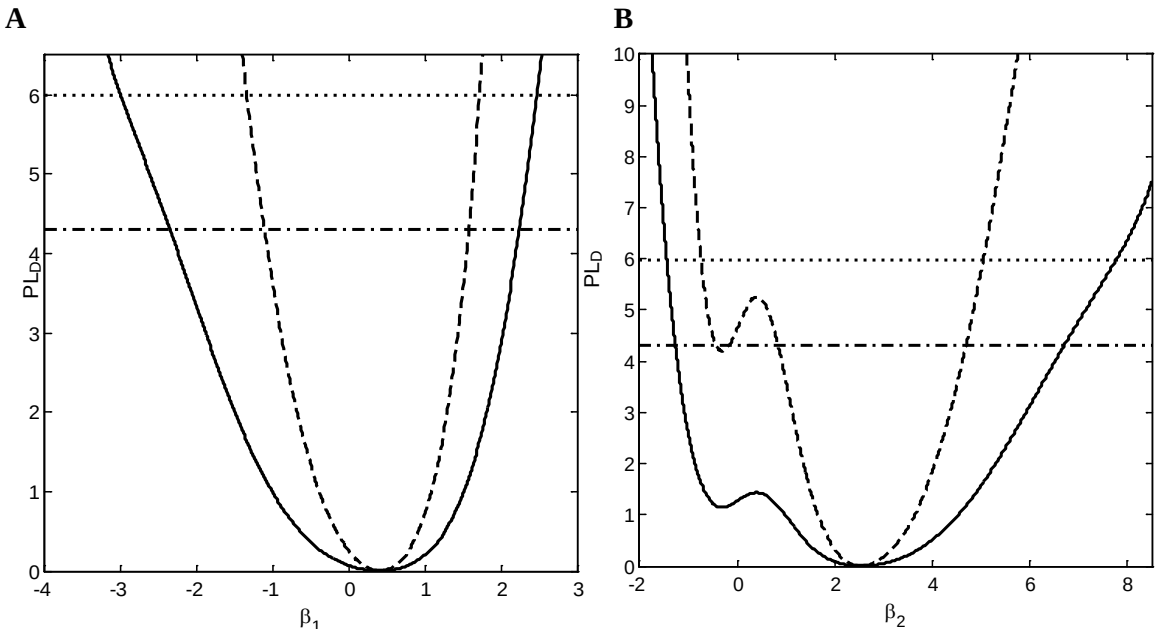

Figure 4-1: Profile likelihood plots for case study 1 using the DC-based profiling approach with n-p degrees of freedom (solid curve and dotted line) and nm-p degrees of freedom (dashed curves and dot-dashed line). Plot A: PL_D function versus β_1. Plot B: PL_D function versus β_2.

Figure 4-1 indicates that the parameter likelihood intervals obtained from the *DC-based profiling approach* with reference lines at $F_{1,nm-p;\alpha}$ are narrower than those associated with $F_{1,n-p;\alpha}$, which is expected given the additional degrees of freedom. Both plots indicate some mild nonlinearity in the model particularly with respect to the second parameter.

Based on the results provided in Table 4-1, the inference intervals from the DC approach are wider than those from the *GLS-based profiling approach #2*. For this synthetic dataset if the results from the *GLS-based approach with known* $\boldsymbol{\Sigma}$ are considered to be the most representative of the system, then the intervals from the *GLS-based profiling approach #2* would appear to be closer to the reference ones and thus more reliable.

As indicated in Chapter 3 Figure 3-4, profile plots of *GLS-based approach with known* $\boldsymbol{\Sigma}$ are exactly quadratic since for this example fixing $\widehat{\boldsymbol{\Sigma}}$ during profiling leads to a linear regression problem. The plots for the DC approach (Figure 4-1) show a mild nonlinearity in the model. However, the asymmetry is not overwhelming, and the lack of monotonicity is a greater concern. Of particular concern is the local minimum in Figure 4-1, plot B, showing the possibility of convergence of the DC method to local minima.

This example particularly provides a clear contrast between the potential numerical challenges of the DC approach relative to those of GLS-based approach. The DC-based profile plots clearly depict the risk of local minima and non-monotonicity. Thus, the proposed profiling approach (*GLS-based profiling approach #2*) appears to be the better choice for this example due to the more monotone plots and also the inference results that are closer to the reference results. Lastly, the DC-based inferences using $nm - p$ degrees of freedom match the reference results more closely than $n - p$, suggesting that $nm - p$ is a more appropriate degrees of freedom for this distribution.

4.5.2 Case Study 2: Buzzi Ferraris Multi-Response Model and Dataset

The second example demonstrates the application of the DC approach to a dataset and multi-response model presented by Buzzi Ferraris et al. (1984). They generated a set of synthetic data for the following kinetic model with two responses (y_1 and y_2), two regressor variables (x_1 and x_2), and four parameters (θ_1, θ_2, θ_3, θ_4):

$$y_{ij} = f_j(\boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_{ij} \tag{4-27}$$

with the model functions defined by:

$$f_1(\boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\theta_1 x_{1i} x_{2i}}{1+\theta_3 x_{1i}+\theta_4 x_{2i}} \quad (4\text{-}28)$$

$$f_2(\boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{\theta_2 x_{1i} x_{2i}}{1+\theta_3 x_{1i}+\theta_4 x_{2i}} \quad (4\text{-}29)$$

The following parameter values (referred to as the *true* parameter values) were inserted into the two model functions of equations 4-28 and 4-29 and noise was added to the model predictions to create the dataset (Buzzi Ferraris et al., 1984):

$$\boldsymbol{\theta}_{true} = [0.1\ 0.01\ 0.1\ 0.01]' \quad (4\text{-}30)$$

Assuming *IID Normal* between-run noise and ignoring the within-run co-dependencies, they used the following diagonal within-run noise covariance matrix (Buzzi Ferraris et al., 1984):

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.35 & 0 \\ 0 & 0.0023 \end{bmatrix} \quad (4\text{-}31)$$

For more information on this example and the dataset please refer to Chapter 3, section 3.5.2 and the dataset provided in Table 3-3.

In this thesis, the *DC-based profiling approach* is applied to this dataset and model, not taking into account the parameter values and noise covariance matrix used by Buzzi Ferraris et al. (1984). The parameter estimates and the 95% parameter likelihood results from this approach along with the results of the proposed *GLS-based profiling approach #2* (explained in Chapter 3) are summarized in Table 4-2. The corresponding DC-based profile plots are provided in Figure 4-2 in which $PL_D(\theta_k)$ is plotted versus each profiling parameter. Reference lines based on $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ are represented by dotted lines and dot-dashed lines respectively. In this figure, the solid curves are obtained from $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(n-p)}$ (equation 4-15), while the dashed curves are obtained from $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(nm-p)}$. For

profile plots of the *GLS-based profiling approach #2* please refer to Figure 3-6 of Chapter 3.

Table 4-2: Parameter estimation and inference results from DC and some GLS profiling approaches for Case Study 2 with true parameter values $\boldsymbol{\theta_{true}} = [0.1\ 0.01\ 0.1\ 0.01]'$

Approach	Parameter Estimates	95% Likelihood Intervals	
GLS-Based Profiling Approach #2 (with Iterated $\widehat{\boldsymbol{\Sigma}}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach2}} =$ $[0.0409\ 0.0041\ 0.1122\ 0.0313]'$	based on $F_{1,n-p;0.95}$	$0.031 \leq \theta_{1,GLSapproach2} \leq 0.059$ $0.0031 \leq \theta_{2,GLSapproach2} \leq 0.0059$ $0.082 \leq \theta_{3,GLSapproach2} \leq 0.17$ $0.018 \leq \theta_{4,GLSapproach2} \leq 0.054$
		based on $F_{1,nm-p;0.95}$	$0.032 \leq \theta_{1,GLSapproach2} \leq 0.058$ $0.0032 \leq \theta_{2,GLSapproach2} \leq 0.0057$ $0.083 \leq \theta_{3,GLSapproach2} \leq 0.177$ $0.019 \leq \theta_{4,GLSapproach2} \leq 0.052$
DC-Based Profiling Approach	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{DC}} =$ $[0.0409\ 0.0041\ 0.1122\ 0.0312]'$	based on $F_{1,n-p;0.95}$	$0.029 \leq \theta_{1,DC} \leq 0.0634$ $0.00295 \leq \theta_{2,DC} \leq 0.00631$ $0.079 \leq \theta_{3,DC} \leq 0.179$ $0.0178 \leq \theta_{4,DC} \leq 0.058$
		based on $F_{1,nm-p;0.95}$	$0.033 \leq \theta_{1,DC} \leq 0.053$ $0.0033 \leq \theta_{2,DC} \leq 0.0053$ $0.09 \leq \theta_{3,DC} \leq 0.148$ $0.022 \leq \theta_{4,DC} \leq 0.046$
GLS-Based Profiling Approach with Known $\boldsymbol{\Sigma}$	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLS,known\ \Sigma}} =$ $[0.0419\ 0.0042\ 0.1161\ 0.0316]'$	based on $\chi^2_{1;\mathbf{0.95}}$	$0.028 \leq \theta_{1,GLS,known\ \Sigma} \leq 0.074$ $0.0028 \leq \theta_{2,GLS,known\ \Sigma} \leq 0.0074$ $0.075 \leq \theta_{3,GLS,known\ \Sigma} \leq 0.22$ $0.015 \leq \theta_{4,GLS,known\ \Sigma} \leq 0.07$

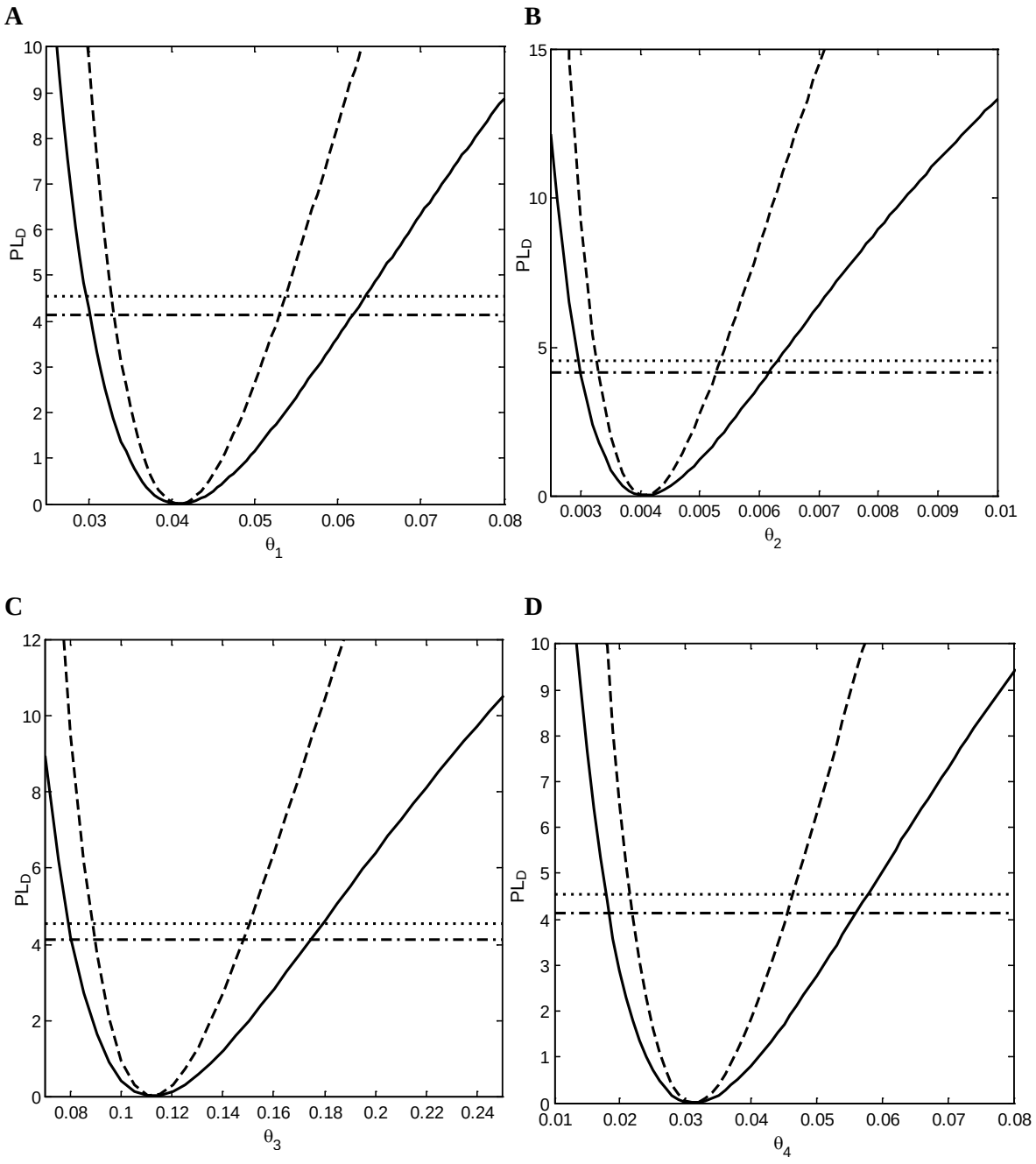

Figure 4-2: Profile likelihood plots for case study 2 using the DC-based profiling approach with n-p degrees of freedom (solid curve and dotted line) and nm-p degrees of freedom (dashed curves and dot-dashed line). Plot A: PL_D function versus θ_1. Plot B: PL_D function versus θ_2. Plot C: PL_D function versus θ_3. Plot D: PL_D function versus θ_4.

According to Table 4-2, the parameter estimates obtained using the DC method are almost identical to those obtained using *GLS-based profiling approach #2*. Both approaches tend to underestimate the first and second parameters, and overestimate the fourth parameter. The third parameter is estimated fairly accurately. Thus, there is no clear preference with the estimation approaches. Similar to previous examples, results from the *GLS-based profiling approach with known* $\boldsymbol{\Sigma}$ (using equation 4-31) are also provided in Table 4-2 to offer a standard reference for inference results.

Based on Figure 4-2, all DC-based profile plots show asymmetry, indicating some moderate nonlinearity in the parameter estimation problem. This was also indicated in the previous chapter by the GLS-based profile likelihood plots (Figures 3-6 and 3-7). As expected, the parameter likelihood intervals corresponding to $F_{1,nm-p;\alpha}$ are narrower than those associated with $F_{1,n-p;\alpha}$, because of the additional degrees of freedom. The profile plots are less steep on the right hand side, indicating more nonlinearity for higher parameter values, and causing a larger difference between the upper likelihood bounds obtained from $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ than the lower bounds for all parameters.

Overall, the results summarized in Table 4-2 suggest that the choice of the degrees of freedom between $nm-p$ or $n-p$ for the distribution of the DC-based profile likelihood function has a noticeable impact on the inference results for this example. The results also indicate that there is no clear advantage to using the *DC-based profiling approach* or the *GLS-based profiling approach #2*. Both of these approaches provide parameter inference intervals that match the reference intervals obtained from the *GLS-based approach with known* $\boldsymbol{\Sigma}$ very closely. One of the main reasons behind the similarity of the results obtained from GLS and DC methods might be the relatively large size of the

dataset for this example. As the size of the dataset increases, the estimations of the parameters and the noise covariance matrix seemingly become less sensitive to the small details of the estimation process and all of the mentioned approaches become relatively successful in estimating parameters and the associated uncertainty intervals. The impact of the size of the dataset on the inference results (not in a profiling context) has also been discussed by Oxby et al. (2003). Based on this case study, the proposed GLS-based profiling approach can provide reliable results which are very close to the reference results and also comparable to the outcomes of the DC-based profiling.

4.5.3 Case Study 3: Beauchamp and Cornell Nonlinear Multi-Response Model and Dataset

In this section a multi-response system and experimental dataset consisting of measurements of a radioactive tracer in human body carried out by Galambos and Cornell (1962) are studied. Beauchamp and Cornell (1966) investigated this dataset and fitted the following model:

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{4-32}$$

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = [f_1(x_i, \boldsymbol{\theta}), f_2(x_i, \boldsymbol{\theta})];\ (i = 1,2, \dots, n) \tag{4-33}$$

with the model functions defined by (Beauchamp and Cornell, 1966):

$$f_1(x_i, \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 x_i) + (1 - \theta_1) \exp(-\theta_3 x_i) \tag{4-34}$$

$$f_2(x_i, \boldsymbol{\theta}) = 1 - (\theta_1 + \theta_4) \exp(-\theta_2 x_i) + (\theta_1 + \theta_4 - 1)\exp(-\theta_3 x_i) \tag{4-35}$$

In these equations x_i are the measurement times, y_1 and y_2 represent the proportions of the injected radioactive sulphate tracer in two different compartments of the body (in percentage), and $\boldsymbol{\varepsilon}$ denotes the noise term. The fourth parameter (θ_4) is dependent on the other three by the following equation:

$$\theta_4 = [(\theta_3 - \theta_2)\theta_1(1 - \theta_1)]/[(\theta_3 - \theta_2)\theta_1 + \theta_2] \quad (4\text{-}36)$$

They reported the following GLS-based parameter estimates (Beauchamp and Cornell, 1966):

$$\widehat{\boldsymbol{\theta}} = [0.06751\ 0.00706\ 0.08393\ 0.39506\ ]' \quad (4\text{-}37)$$

This example was discussed earlier to investigate the application of the GLS method. For more information about this example please see Chapter 3 section 3.5.3 and dataset of Table 3-5. In this section, the *DC-based profiling approach* has been applied to this dataset and model, not taking into account the parameter estimates and the noise covariance matrix used by Beauchamp and Cornell (1966). The parameter estimates and the 95% parameter likelihood results from the *DC-based profiling approach* along with the proposed *GLS-based profiling approach #2* (explained in Chapter 3) are summarized in Table 4-3.

The corresponding DC-based profile plots are provided in Figure 4-3 where $PL_D(\theta_k)$ has been plotted versus each profiling parameter. Reference lines for parameter inference based on $F_{1,n-p;0.95}$ and $F_{1,nm-p;0.95}$ are represented by dotted lines and dot-dashed lines respectively. In this figure, the solid curves are obtained from $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}})}{(n-p)}$ (equation 4-15), while the dashed curves are obtained from $s'^2 = \frac{D(\widehat{\boldsymbol{\theta}})}{(nm-p)}$. For profile plots of the *GLS-based profiling approach #2* please refer to Figure 3-9 of Chapter 3.

Table 4-3: Parameter estimation and inference results from DC-based and GLS-based profiling approach #2 for Case study 3

Approach	**Parameter Estimates**	**95% Likelihood Intervals**	
GLS-Based Profiling Approach #2 (with Iterated $\hat{\Sigma}$)	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{GLSapproach2}} = [0.1971\ 0.0131\ 0.116]'$	based on $F_{1,n-p;0.95}$	$0.13 \leq \theta_{1,GLSapproach2} \leq 0.26$ $0.01 \leq \theta_{2,GLSapproach2} \leq 0.016$ $0.101 \leq \theta_{3,GLSapproach2} \leq 0.134$
		based on $F_{1,nm-p;0.95}$	$0.142 \leq \theta_{1,GLSapproach2} \leq 0.25$ $0.0106 \leq \theta_{2,GLSapproach2} \leq 0.0155$ $0.103 \leq \theta_{3,GLSapproach2} \leq 0.13$
DC-Based Profiling Approach	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{DC}} = [0.0506\ 0.006\ 0.0883]'$	based on $F_{1,n-p;0.95}$	$0.007 \leq \theta_{1,DC} \leq 0.113$ $0.0013 \leq \theta_{2,DC} \leq 0.0098$ $0.065 \leq \theta_{3,DC} \leq 0.106$
		based on $F_{1,nm-p;0.95}$	$0.022 \leq \theta_{1,DC} \leq 0.081$ $0.0032 \leq \theta_{2,DC} \leq 0.0082$ $0.075 \leq \theta_{3,DC} \leq 0.098$

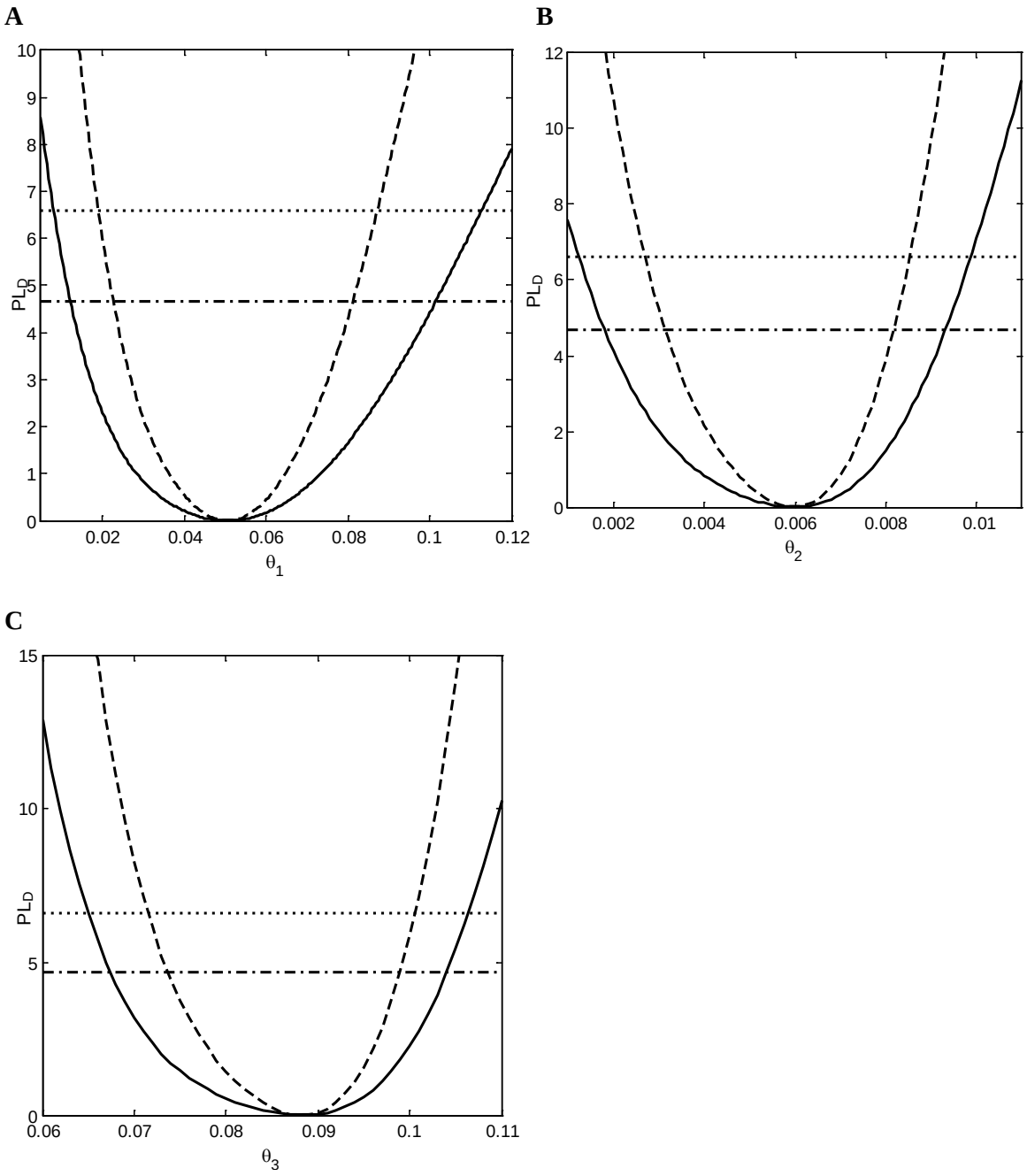

Figure 4-3: Profile likelihood plots for case study 3 using DC approach with n-p degrees of freedom (solid curve and dotted line) and nm-p degrees of freedom (dashed curves and dot-dashed line). Plot A: PL_D versus θ_1. Plot B: PL_D versus θ_2. Plot C: PL_D versus θ_3.

Table 4-3 indicates that the DC approach leads to parameter estimates that are quite similar to those reported by Beauchamp and Cornell (1966). The parameter estimates

from the *GLS-based profiling approach #2* are farther away from the values reported in the literature especially for θ_1 and θ_2 which seems to be due to the numerical conditioning of the problem. As explained in Chapter 3, Beauchamp and Cornell studied the dataset containing missing data points so that the numbers of runs corresponding to the two responses were not the same. All the estimations for parameters and the noise variance/covariance terms were carried out based on the dataset with missing data. As a result, the GLS-based estimations from their method differ from the results obtained in this thesis using *GLS-based profiling approach #2* with the dataset of Table 3-5. Updating the noise covariance matrix during parameter estimation in the GLS-based profiling method is also an important contributor to the difference between the parameter estimates.

The plots presented in Figure 4-3 are reasonably smooth, showing no local minima or numerical issues. The parameter likelihood intervals corresponding to $F_{1,nm-p;\alpha}$ are narrower than those associated with $F_{1,n-p;\alpha}$ due to the additional degrees of freedom. Overall for this small dataset and the corresponding model which consists of exponential terms, there is a noticeable difference between the parameter estimates and inference results from the *DC-based profiling approach* compared to the *GLS-based profiling approach #2*. For this example, the *DC-based profiling approach* seems to provide reasonable parameter inference results and in practice proves to be computationally efficient, i.e., the calculations are faster and more straightforward. Furthermore, the likelihood plots do not show a significant risk of local minima.

4.6 Comparison of GLS and DC Methods for Parameter Estimation and Inference

In the present work, profiling based on GLS and DC methods has been discussed in Chapters 3 and 4 respectively. Using several examples, it is shown that the choice of profiling method impacts the parameter estimation and inference results and that the computational efficiency and the complexity of the estimation process also depend on the chosen profiling method. This raises the question of which method is more appropriate in practice. There is probably not one definitive answer to this question as each method has advantages and disadvantages, and the choice of the profiling method depends on several factors such as the type of model, purpose of modeling, available data and general information about the experiments and process of data collection. In order to help clarify this matter, previous research and point of view of this thesis are mentioned in this section.

According to the examples and discussions earlier in this chapter, the *DC-based profiling approach* is a relatively straightforward process and generates inference outcomes that are comparable to GLS-based results and sometimes even surpass them in matching the reference results (as seen in Case study 3). However, although the DC approach seems less complicated than GLS due to implicit estimation of the noise covariance matrix, there are a few challenges associated with this method that make it less practical in some cases. Calculating the Determinant function for systems with multiple parameters and differential equations is usually computationally demanding and even linearization of the model function does not always improve the complexity of the problem. Furthermore, the DC approach is more prone to falling into local minima and thus generating unsatisfying results (Bates and Watts, 1987); (Kang and Bates, 1990). This was illustrated in Case

study 1 where the profile plots obtained from the *DC-based profiling approach* turned out to be non-monotonic. It should be noted that the GLS method is not free of challenges either since it is very sensitive to the initial parameter values and the choice of iteration method and the associated increments. Local minima are also possible.

In order to obtain proper results from the DC approach, the number of experiments should be greater than or equal to the number of responses. But moreover, if the dataset is relatively small this approach sometimes fails to generate adequate results even when this condition is met. This is due to the fact that for a small dataset, $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\widehat{\boldsymbol{Z}}(\boldsymbol{\theta})'\widehat{\boldsymbol{Z}}(\boldsymbol{\theta})$ (equation 4-7) is a poor estimate for the within-run noise covariance matrix. In most cases, as the number of experimental runs increases, more stable determinants are obtained and it is less likely for the determinant value to converge to a local minimum (Oxby et al., 2003); (Routray and Deo, 2005).

4.6.1 GLS and DC Methods for Parameter Estimation and Inference

Some discussions regarding the comparison of the GLS and DC methods can be found in literature but they are not in a profiling context. For example, Kang and Bates (1990) argue that in theory the DC method gives more precise parameter estimates since in the Bayesian context, which is at the root of this approach, parameter estimates are derived from the marginal posterior density function which accommodates prior knowledge, whereas in the GLS approach the parameter estimates are based on the conditional likelihood. Nonetheless, they acknowledge the complexity of the computations involved in the DC approach which might be an obstacle for widespread use of this method. Thus, the decision about which method to use for parameter estimation depends on the model

and the complexity of the required computations for that model. However with regards to parameter inference, they maintain that the GLS method is generally a more appropriate choice compared to DC. In DC, the parameter covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is approximated by (Kang and Bates, 1990):

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = 2s'^2\{\mathbf{H}^{-1}\}_{kk} \tag{4-38}$$

where s'^2 is the DC-based average noise variance estimate, and $\mathbf{H}$ is the Hessian of $|\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})|$ evaluated at $\widehat{\boldsymbol{\theta}}$. On the other hand in GLS, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is estimated from the noise covariance matrix by means of (Kang and Bates, 1990); (Oxby, 1997):

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \left(\boldsymbol{V}(\widehat{\boldsymbol{\theta}})'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V}(\widehat{\boldsymbol{\theta}})\right)^{-1} \tag{4-39}$$

with the sensitivity matrix $\boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ evaluated at the least squares parameter estimates $\left(\boldsymbol{V}(\widehat{\boldsymbol{\theta}}) = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right)$. Kang and Bates (1990) suggest that equation 4-38 only provides an approximate estimate based on the Hessian matrix; whereas equation 4-39 offers parameter variance/covariance estimates based on the SSE function and accounts for the noise covariance matrix. They also show numerically that the parameter variance/covariance estimates obtained from equation 4-38 are larger than those from equation 4-39. Thus, the GLS method is potentially more appropriate than DC for parameter inference.

In his paper, Phillips (1976) showed that the DC approach generally results in more accurate parameter estimates with the smallest parameter variances. However, Oxby et al. (2003) argued that in practice DC has many shortcomings and is particularly sensitive to the size of the dataset and the perturbations in the data. They proposed a GLS-based approach called the Multivariate Weighted Least Squares (MWLS) which is a two-step

iteration starting with estimating parameters by minimizing the following weighted sum of squares function $\boldsymbol{S}_w(\boldsymbol{\theta})$ (Oxby, 1997); (Oxby et al., 2003):

$$\boldsymbol{S}_w(\boldsymbol{\theta}) = tr(\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{W}\boldsymbol{Z}(\boldsymbol{\theta})) \quad (4\text{-}40)$$

where $\boldsymbol{W}$ is a diagonal weight matrix that can initially be set as the identity matrix. Once the parameter estimates are obtained, the corresponding residuals are used in $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})'\boldsymbol{Z}(\widehat{\boldsymbol{\theta}})$ (equation 4-7) to build the noise covariance matrix. Then the weight matrix is updated as the inverse of the diagonal elements of the noise covariance matrix. These two steps are iterated until convergence.

Oxby et al. (2003) compared their proposed MWLS method with the regular GLS and DC approaches by means of some examples and showed that the size of the dataset has a great impact on the performance of the DC approach and that sometimes for small datasets the DC-based parameter estimates are not as accurate as the GLS-based methods. Using several simulations, they showed that the GLS-based parameter estimate variance decreases linearly as the sample size increases so that the GLS-based parameter estimates become more accurate (with smaller variances) for larger datasets. The same analysis suggested that the parameter estimates from the DC approach were different from the GLS ones but got fairly close to them as the size of the dataset increased. They also explored the distribution of the DC-based parameter estimates. Using a Monte Carlo simulation, one million synthetic datasets were generated from the parameters estimated by MWLS and DC, indicating that the distribution of the DC parameter estimates was farther from normal, and had heavy tails; whereas MWLS parameter estimates were nearly normally distributed. It can be concluded from their results that for small datasets

GLS and MWLS approaches lead to smaller variances for parameter estimates, and are generally more robust and practical for any size of dataset (Oxby et al., 2003).

Based on this short background review, the GLS method seems to be the more appropriate approach for parameter inference and thus it might be best to prioritize GLS-based profiling over DC-based for parameter inference as well.

4.6.2 GLS-Based and DC-Based Profiling Methods

In order to compare the profiling methods based on GLS and DC, the outcomes of applying *GLS-based profiling approach #2* and *DC-based profiling approach* to the previously mentioned case studies are examined side by side. The following figures in this section provide a closer look at the estimation results from DC and GLS approaches for Case studies 1 to 3. The likelihood intervals using *GLS-based profiling approach #2* are depicted by solid lines, while the dashed lines represent the inference intervals from the *DC-based profiling approach*. Based on the discussions regarding the distribution and degrees of freedom of the GLS-based profile likelihood function and DC-based profile likelihood function for multi-response models (Chapter 3 Section 3.4, and Chapter 4 Section 4.4), these likelihood intervals are generated from reference lines based on $F_{1,nm-p;0.95}$. Furthermore, the inference results from *GLS-based approach with known* $\boldsymbol{\Sigma}$ with reference lines at $\chi^2_{1;0.95}$ are also included for reference (shown by dotted lines) where applicable.

4.6.2.1 Case Study 1 (Bates and Watts Example):

Figure 4-4 shows that for the first example, the parameter estimates from all three approaches are very close to each other. The inference results from the *GLS-based profiling approach #2* are almost identical to the reference intervals (provided by *GLS-based approach with known* $\boldsymbol{\Sigma}$). All intervals are almost symmetric and suggest only mild nonlinearity in the model. As mentioned before, for this model the GLS-based profile likelihood plots are symmetric due to the fixed $\widehat{\boldsymbol{\Sigma}}$ during profiling.

For further comparison, the noise covariance matrices obtained from the two methods can also be consulted. The noise covariance matrix obtained from *GLS-based profiling approach #2* is:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{GLS\#2}} = \begin{bmatrix} 0.9724 & 1.1355 & 1.2397 \\ 1.1355 & 2.4497 & 1.1575 \\ 1.2397 & 1.1575 & 2.6718 \end{bmatrix} \tag{4-41}$$

The *DC-based profiling approach* resulted in the following noise covariance matrix:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC}} = \begin{bmatrix} 0.9623 & 1.1305 & 1.2315 \\ 1.1305 & 2.4638 & 1.1662 \\ 1.2315 & 1.1662 & 2.6613 \end{bmatrix} \tag{4-42}$$

Interestingly, both approaches lead to almost similar estimates for the noise covariance matrix.

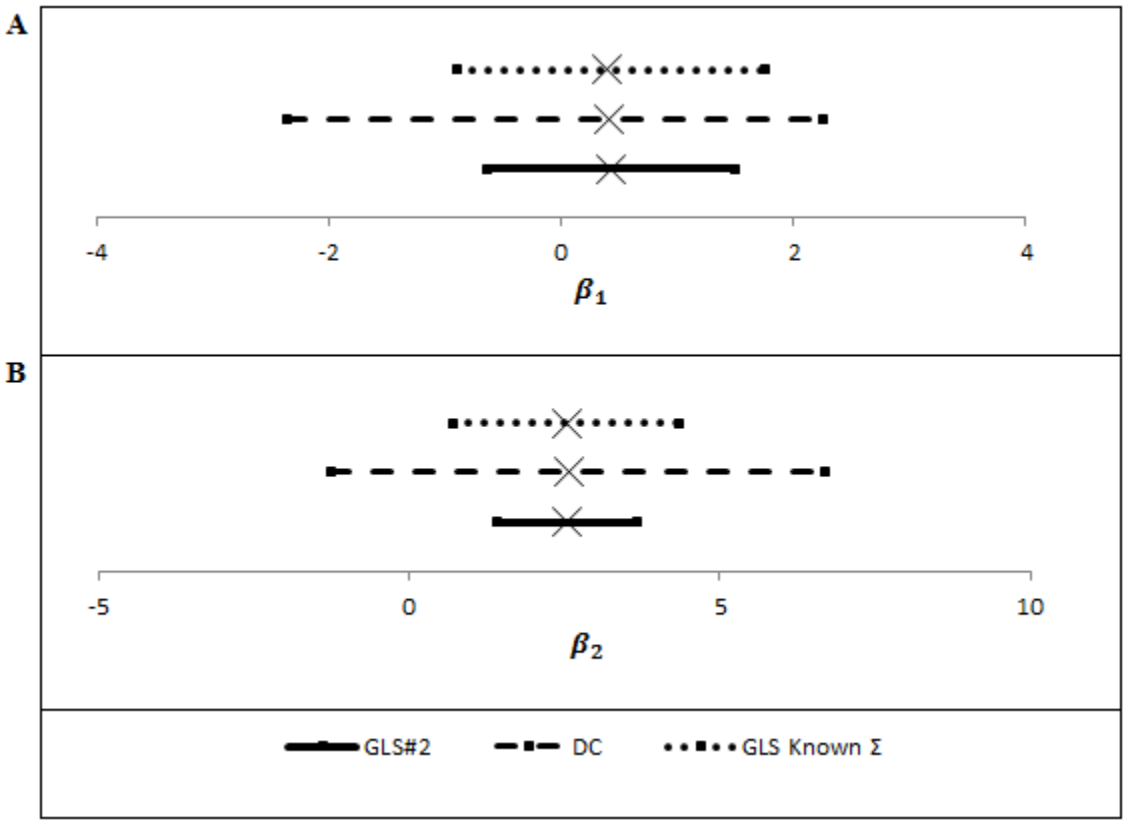

Figure 4-4: Profile-based parameter inference itervals for Case study 1 obtained from *GLS-based profiling approach #2* (solid line), *DC-based profiling* (dashed line) and *GLS-based profiling with known* $\boldsymbol{\Sigma}$ (dotted line) approaches. Plots A and B correspond to β_1 and β_2 respcectively. The corresponding point estimates for parameters from each approach are marked on each line.

Overall based on the comparisons for this example, the GLS-based profiling method proposed in this thesis has a slight advantage over the DC method since it avoids numerical issues and provides adequate parameter estimates and parameter inference intervals that are fairly close to the reference intervals.

4.6.2.2 Case Study 2 (Buzzi Ferraris Example):

For the second example (Figure 4-5), the parameter estimates and inference results from the GLS and DC approaches are very similar to each other. The lines representing the

parameter likelihood intervals from the GLS and DC approaches also fall very close to each other with the DC-based intervals slightly closer to the reference intervals particularly for θ_1 and θ_2. The inference results from both approaches are somewhat narrower than the reference intervals and both approaches yield asymmetric inference intervals which suggest moderate nonlinearity in the model.

The estimated noise covariance matrices from *GLS-based profiling approach #2* and *DC-based profiling approach* are provided in equations 4-43 and 4-44 respectively.

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{GLS\#2}} = \begin{bmatrix} 0.1497 & -0.0095 \\ -0.0095 & 0.0034 \end{bmatrix} \tag{4-43}$$

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC}} = \begin{bmatrix} 0.1499 & -0.0095 \\ -0.0095 & 0.0034 \end{bmatrix} \tag{4-44}$$

Once again, both approaches lead to similar estimates for the noise covariance matrix.

In summary, the results suggest that the DC and GLS approaches are both reliable and fairly comparable for this example and there is no advantage to choosing either one over the other. More importantly, this example emphasizes the significance of obtaining more measurements and shows the impact of larger datasets on improving the estimations.

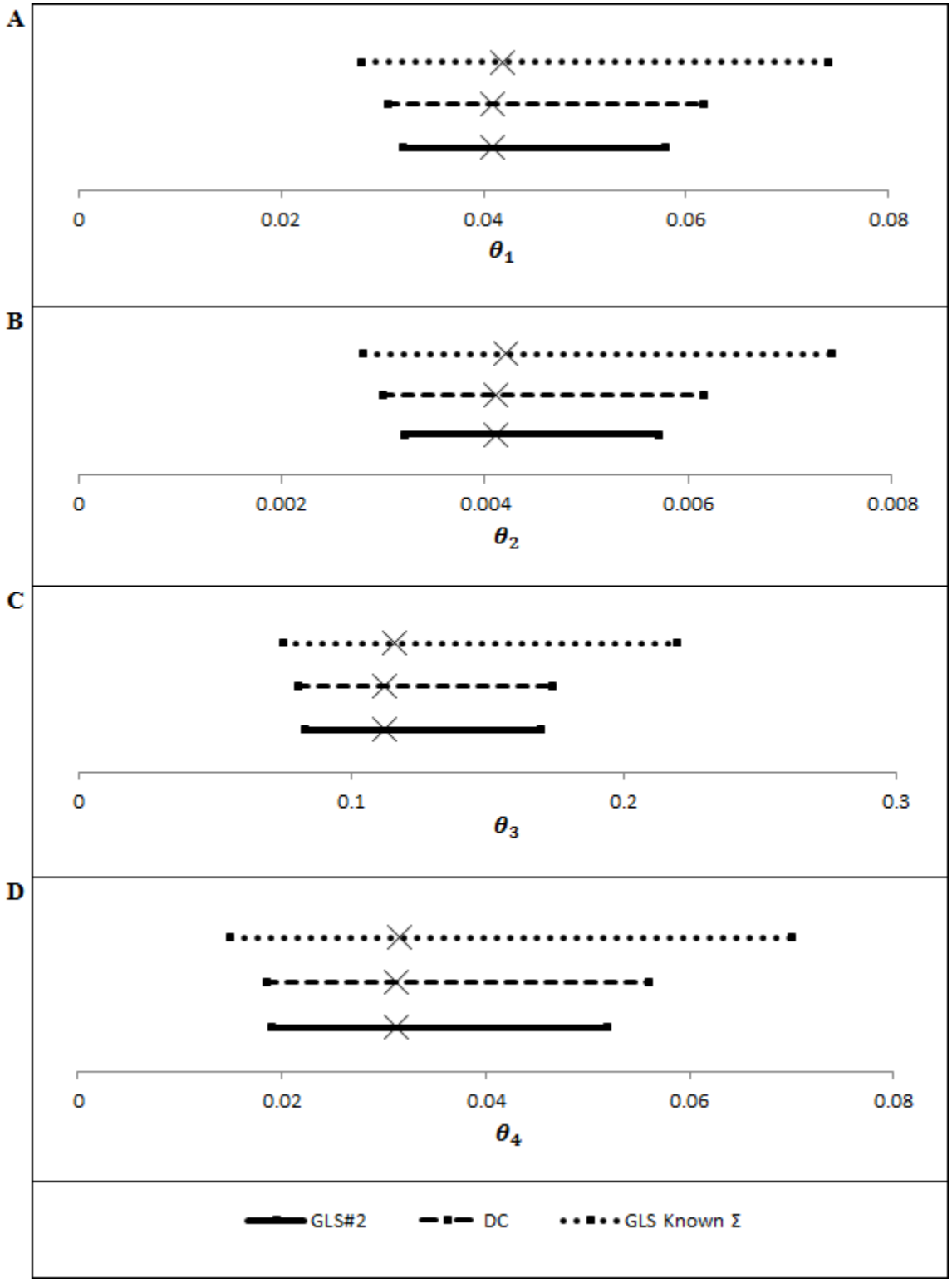

Figure 4-5: Profile-based parameter inference itervals for Case study 2 obtained from GLS approach #2 (solid line), DC (dashed line) and *GLS with known Σ* (dotted line) approaches. Plots A, B, C, and D correspond to θ_1, θ_2, θ_3, and θ_4 respectively. Point estimates for parameters from each approach are shown on the corresponding lines.

4.6.2.3 Case Study 3 (Beauchamp and Cornell Example):

Figure 4-6 contains the comparison of the profiling results for the third example.

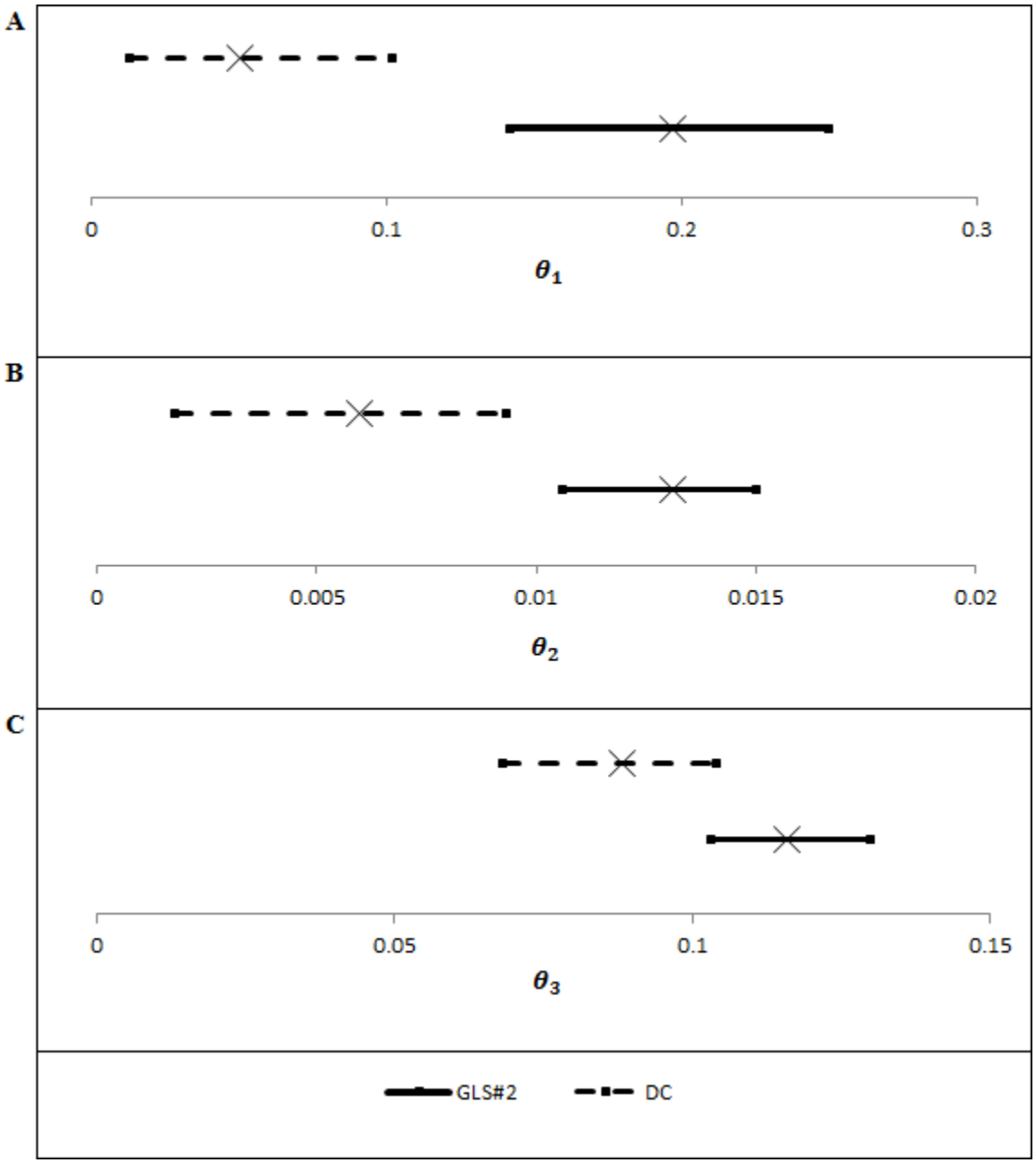

Figure 4-6: Profile-based parameter inference itervals for Case study 3 obtained from GLS-based profiling approach #2 (solid line) and DC (dashed line) approaches. Plots A, B, and C correspond to θ_1, θ_2, and θ_3 respectively. Point estimates for parameters from each approach are shown on the corresponding lines.

For the third example (Figure 4-6) there is a clear difference between parameter estimates and inference intervals from the two methods. The DC approach in this case generates results that match the parameter estimates reported in the literature more closely. Nonetheless, this is not necessarily an advantage because the parameter estimates reported in the literature are obtained based on the GLS method where the noise covariance matrix is estimated from the dataset including an extra run with a missing point. Since the dataset is relatively small, the extra run can have a major impact on the estimations. Furthermore, the parameter estimates reported by Beauchamp and Cornell (1966) were obtained using a fixed estimated noise covariance matrix, whereas the proposed GLS-based profiling method takes advantage of a more accurate $\widehat{\boldsymbol{\Sigma}}$ that is updated during parameter estimation. Thus, some differences between the estimation results were expected.

For comparison purposes, the estimated noise covariance matrices from *GLS-based profiling approach #2* and *DC-based profiling approach* for this example are also provided in equations 4-45 and 4-46 respectively:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{GLS\#2}} = \begin{bmatrix} 0.002 & 0.0016 \\ 0.0016 & 0.0015 \end{bmatrix} \tag{4-45}$$

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC}} = \begin{bmatrix} 0.003 & -0.0001 \\ -0.0001 & 0.00002 \end{bmatrix} \tag{4-46}$$

The main difference between these two estimated matrices is that the DC-based matrix suggests almost no covariance between the noise terms and also estimates a very small variance for the noise associated with the second response. Though they appear small, the GLS-based noise variance and covariance terms are relatively larger than the DC ones for this example.

All in all for this example, it is not easy to choose one method over the other since each method has some advantages and disadvantages and also there is no clear reference to compare the estimation and inference outcomes. However, in practice the DC-based computations for parameter estimation were more straightforward which can be viewed as an advantage.

4.6.3 Profiling about the DC-Optimal Parameter Estimates

To approach the problem from another angle, this thesis focuses on comparing the GLS-based and DC-based profiling methods for parameter inference starting with the same parameter estimates. More specifically, if the estimation problem is viewed as a parameter estimation step followed by a profiling step, in this section the DC approach is used for the parameter estimation step and then applied to both DC-based and GLS-based approaches in the profiling step. The DC parameter estimation method was chosen for the first step provided that it has generated consistent results for the examples and that it has been recommended as the more reliable estimation method in the literature (Kang and Bates, 1990). Thus, profiling is performed around the DC-optimal parameter estimates and the corresponding SSE and Determinant functions are calculated to be used for *GLS-based profiling approach #2* and *DC-based profiling approach* respectively.

In this analysis, both profiling methods start with the same parameter estimates and consequently the same noise covariance matrix that is estimated from the residuals. However, as the profiling method progresses, the estimated noise covariance matrix is kept fixed during the *GLS-based profiling approach #2* while it is automatically updated based on the latest residual values in the *DC-based profiling approach*. Therefore, it is

expected to see some differences between the values of the PL_S and PL_D functions resulting from this approach.

This analysis is applied to Case studies 1 and 2 and the resulting DC-based and GLS-based profile likelihood functions are compared for each example. However, the analysis is not possible for Case study 3 because the $\boldsymbol{\Sigma}$ corresponding to the parameter estimates from the DC approach is not positive definite.

4.6.3.1 Case Study 1 (Bates and Watts Example):

Figure 4-7 shows the profile likelihood plots for Case study 1 (Bates and Watts example) carried out around the DC-optimal parameter estimates $\widehat{\boldsymbol{\beta}}_{\boldsymbol{DC}}$. In this figure the solid lines correspond to DC-based profile likelihood plots and the dashed lines represent the GLS-based ones.

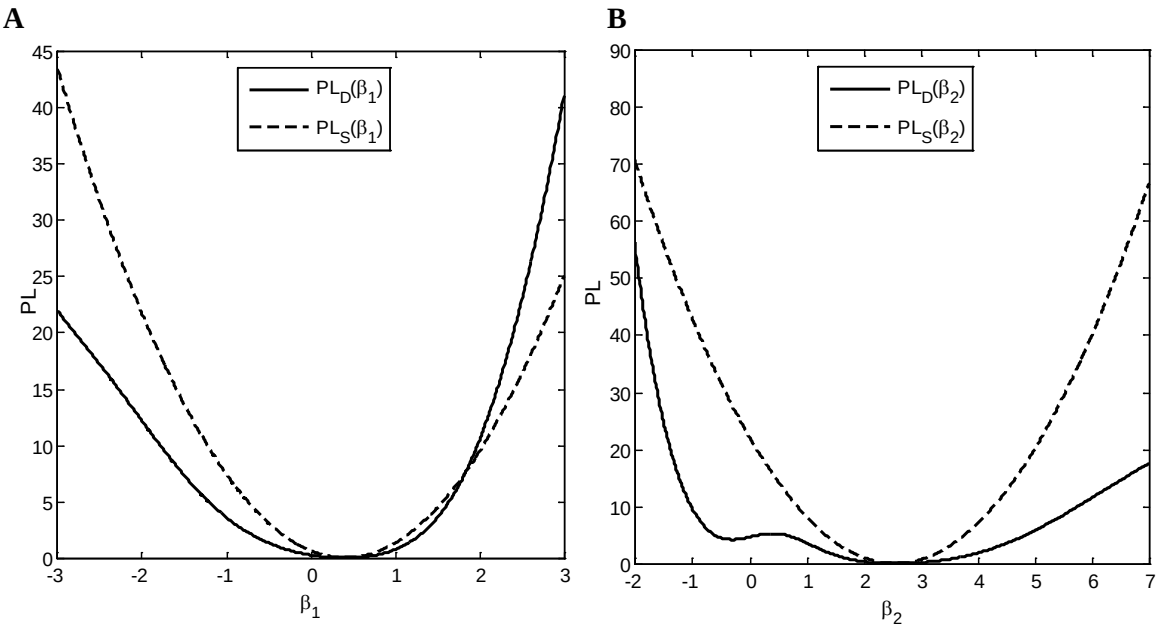

Figure 4-7: Comparison of DC-based and GLS-based profile likelihood plots for case study 1 around the DC-optimal parameter estimates. Plot A: PL_D function versus β_1 (solid curve) and PL_S function versus β_1 (dashed curve). Plot B: PL_D function versus β_2 (solid curve) and PL_S function versus β_2 (dashed curve).

Figure 4-7 indicates a clear difference between the PL_S and PL_D plots for both parameters β_1, β_2. In Figure 4-7 Plot A, $PL_D(\beta_1)$ depicts a mild nonlinearity in the model with respect to β_1 while $PL_S(\beta_1)$ is narrower and quadratic. The difference between the two curves becomes more noticeable for smaller values of β_1 as they move away from the DC-optimal $\hat{\beta}_1$. Figure 4-7 Plot B also depicts a visible difference since $PL_D(\beta_2)$ is much wider than $PL_S(\beta_2)$ and asymmetric, illustrating the nonlinearity of the model with respect to β_2.

An inspection of the noise covariance matrix estimates reveals the main reason for this difference. The initial noise covariance matrix that both profiling procedures use is obtained from the DC-optimal parameter estimates and equals:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC}} = \begin{bmatrix} 0.9623 & 1.1305 & 1.2315 \\ 1.1305 & 2.4638 & 1.1662 \\ 1.2315 & 1.1662 & 2.6613 \end{bmatrix} \quad (4\text{-}47)$$

However, the estimated DC-based noise covariance matrix after profiling with β_1 as the profiling parameter is:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling1}} = \begin{bmatrix} 1.9679 & 1.8234 & 2.3665 \\ 1.8234 & 2.4312 & 1.5652 \\ 2.3665 & 1.5652 & 4.5009 \end{bmatrix} \quad (4\text{-}48)$$

It can be seen that the estimates have particularly changed for $\hat{\sigma}_1{}^2$, $\hat{\sigma}_3{}^2$ and $\hat{\sigma}_{13}$. Similarly, the DC-based profiling with β_2 as the profiling parameter results in the following matrix:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling2}} = \begin{bmatrix} 0.4427 & 0.7482 & 0.7272 \\ 0.7482 & 3.6713 & 1.8723 \\ 0.7272 & 1.8723 & 2.5748 \end{bmatrix} \quad (4\text{-}49)$$

Again there is a visible difference between this matrix and the matrix of equation 4-47 with estimates that are generally smaller except for $\hat{\sigma}_2{}^2$ and $\hat{\sigma}_{23}$.

Based on these results it might appear that the DC-based profiling can successfully update $\widehat{\boldsymbol{\Sigma}}$ and thus reflect the nonlinearity of the system and generate more reliable parameter likelihood intervals. However, it is uncertain whether these estimates provided during profiling for the elements of $\widehat{\boldsymbol{\Sigma}}$ as well as the parameters can be trusted due to the difficulties of this joint estimation. More conclusions in this regard can be made from Figure 4-8 which shows pairwise profile traces for this example. In Figure 4-8 plot A, the solid line represents the DC-based profile trace of $\tilde{\beta}_2$ versus β_1 as the profiling parameter and the dashed line shows the DC-based trace of $\tilde{\beta}_1$ versus β_2. Figure 4-8 plot B contains the GLS-based profile trace of $\tilde{\beta}_2$ versus β_1 as the profiling parameter represented by dot-dashed line and GLS-based profile trace of $\tilde{\beta}_1$ versus β_2 shown by the dotted line. The two plots are overlaid in Figure 4-8 plot C for easier comparison.

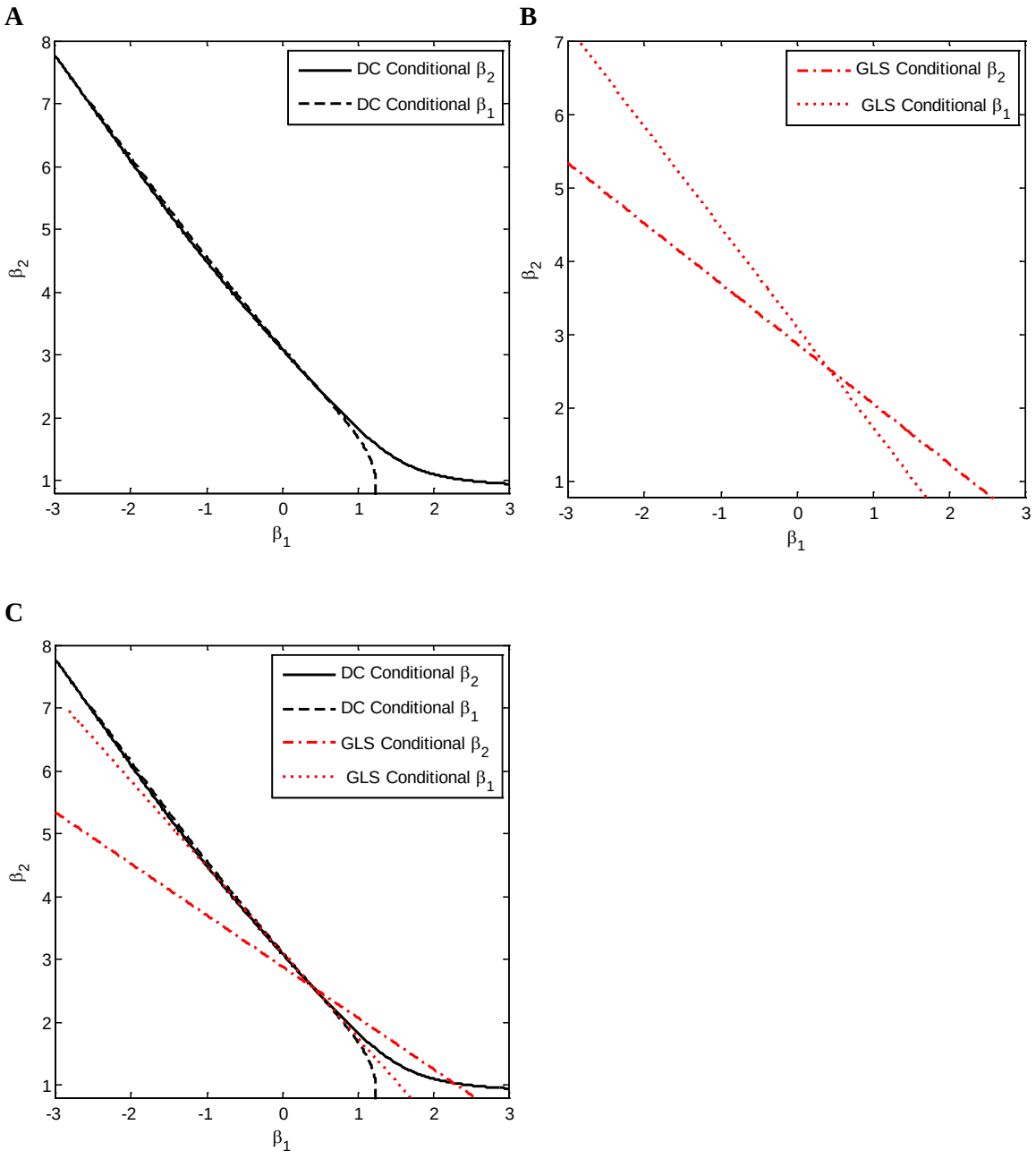

Figure 4-8: Comparison of GLS and DC profile traces for Case study 1 around the DC-optimal parameter estimates. Plot A: $\tilde{\beta}_2$ versus β_1 (solid line) and $\tilde{\beta}_1$ versus β_2 (dashed line) based on DC method. Plot B: $\tilde{\beta}_2$ versus β_1 (dot-dashed line) and $\tilde{\beta}_1$ versus β_2 (dotted line) based on GLS method. Plot C: Overlay of Plots A and B.

Based on Figure 4-8 the DC-based profile traces are curved, implying nonlinearity of the parameter estimation problem. The DC-based profile traces for both parameters fall very

close to each other and even overlap for the most part which is an indication of very strong co-dependencies between the parameter estimates $\hat{\beta}_1$ and $\hat{\beta}_2$. The overlapping of the profile traces in a relatively large interval highlights one of the main challenges of the DC-based profiling which is very poor numerical conditioning of the problem and highly co-dependent parameter estimates. As explained before, this issue is caused by the joint estimation of the parameters and noise variance/covariance terms in DC-based profiling.

The profile traces obtained from the GLS profiling are also not orthogonal, showing co-dependency between the two parameter estimates (Figure 4-8 plot B). However, the traces in this plot only cross each other and don't overlap which is a great advantage for the GLS-based profiling. Overall, this analysis suggests that the GLS-based profiling is a better choice for this example since it avoids potential numerical issues and provides parameter inference intervals that are relatively narrow and reliable.

4.6.4 Case Study 2 (Buzzi Ferraris Example):

The same analysis is done for Case study 2 (Buzzi Ferraris example) where both DC-based and GLS-based profiling are carried out around the DC-optimal parameter estimates. The corresponding profile likelihood plots are shown in Figure 4-9. In this figure, the solid lines correspond to DC-based profile likelihood plots and the dashed lines represent the GLS-based ones.

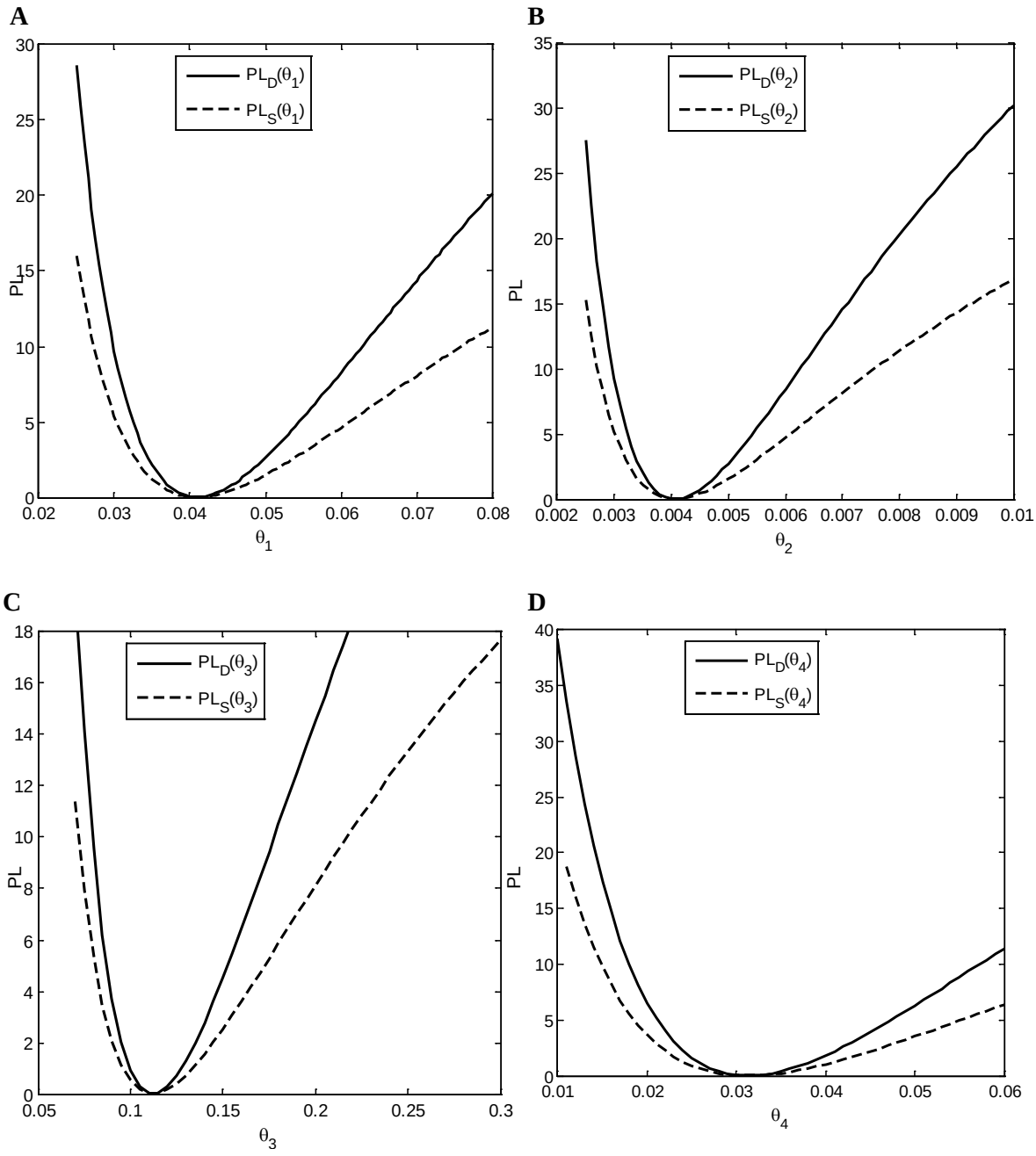

Figure 4-9: Comparison of DC-based and GLS-based profile likelihood plots for case study 2 around the DC-optimal parameter estimates. Plot A: PL_D function versus θ_1 (solid curve) and PL_S function versus θ_1 (dashed curve). Plot B: PL_D function versus θ_2 (solid curve) and PL_S function versus θ_2 (dashed curve). Plot C: PL_D function versus θ_3(solid curve) and PL_S function versus θ_3 (dashed curve). Plot D: PL_D function versus θ_4 (solid curve) and PL_S function versus θ_4 (dashed curve).

For this example, the profile likelihood plots depict moderate nonlinearity of the model with respect to all parameters which is to be expected based on the form of the model functions (equations 4-28 and 4-29). There is a visible difference between the PL_S and PL_D plots for larger values of θ_1 and θ_2 and θ_3 as profiling parameters. For this example, the PL_D plots with $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(nm-p)}$ are narrower than PL_S.

For this example as well, the estimated noise covariance matrices are investigated to reveal the reason for the difference between the values of the PL_S and PL_D functions. The noise covariance matrix obtained from the DC-optimal parameter estimates and used at the start of both profiling procedures is:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC}} = \begin{bmatrix} 0.1499 & -0.0095 \\ -0.0095 & 0.0034 \end{bmatrix} \tag{4-50}$$

However, the final noise covariance matrix after profiling with θ_1 as the profiling parameter is:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling1}} = \begin{bmatrix} 0.1953 & -0.0042 \\ -0.0042 & 0.0040 \end{bmatrix} \tag{4-51}$$

Thus during profiling, the estimated value of the noise covariance $\hat{\sigma}_{12}$ is almost cut in half. Similarly, the DC-based profiling with θ_2 as the profiling parameter results in the following matrix:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling2}} = \begin{bmatrix} 0.1920 & -0.0045 \\ -0.0045 & 0.0040 \end{bmatrix} \tag{4-52}$$

which has a value very close to the previous matrix (equation 4-51). Thus the estimated $\widehat{\boldsymbol{\Sigma}}$ remains relatively unchanged in profiling for θ_2. The DC-based profiling for θ_3 as the profiling parameter generates the following final estimate:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling3}} = \begin{bmatrix} 0.1787 & -0.0056 \\ -0.0056 & 0.0039 \end{bmatrix} \tag{4-53}$$

Again, the change in the estimates from the previous profiling is almost negligible but the elements of this matrix are closer to the initial matrix of equation 4-50. Lastly, the DC-based profiling for θ_4 as the profiling parameter results in:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{DC\ Profiling4}} = \begin{bmatrix} 0.2201 & -0.0025 \\ -0.0025 & 0.0041 \end{bmatrix} \quad (4\text{-}54)$$

Except for ${\sigma_2}^2$, all the other estimates differ significantly from the initial matrix of equation 4-50 which shows the impact of the estimation during this profiling step.

Overall, the estimated $\widehat{\boldsymbol{\Sigma}}$ matrices during the DC-based profiling are different from the initial estimated matrix. However, the form of the model and the relatively small magnitude of the estimated variance/covariance values and their changes during DC-based profiling cause the two profile likelihood curves to fall not too far apart.

For further investigation, the profile traces obtained from both DC and GLS-based profiling methods around the DC-optimal parameter estimates are also consulted. These profile traces are quite similar for the two approaches and show a high level of co-dependency between $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$. There is a milder level of co-dependency between $\hat{\theta}_4$ and any of the other parameter estimates. Due to the large number of profile traces for this example, it was decided in this research to only include the profile traces for the most co-dependent parameter estimates and the least co-dependent parameter estimates in Figures 4-10 and 4-11.

In Figure 4-10 plot A, the solid line represents the DC-based profile trace of $\tilde{\theta}_2$ versus θ_1 as the profiling parameter and the dashed line shows the DC-based trace of $\tilde{\theta}_1$ versus θ_2. Figure 4-10 plot B contains the GLS-based profile trace of $\tilde{\theta}_2$ versus θ_1 and $\tilde{\theta}_1$ versus θ_2 shown by dot-dashed line and dotted line respectively.

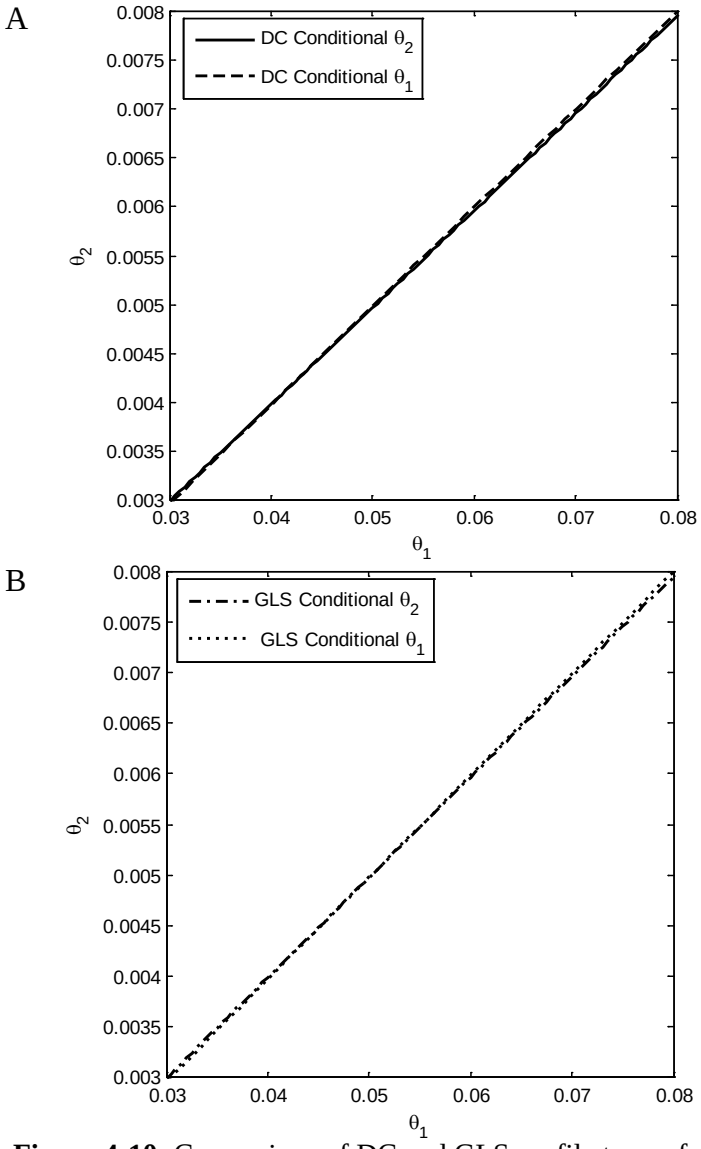

Figure 4-10: Comparison of DC and GLS profile traces for Case study 2 around the DC-optimal parameter estimates. Plot A: $\tilde{\theta}_2$ versus θ_1 (solid line) and $\tilde{\theta}_1$ versus θ_2 (dashed line) based on DC method. Plot B: $\tilde{\theta}_2$ versus θ_1 (dot-dashed line) and $\tilde{\theta}_1$ versus θ_2 (dotted line) based on GLS method.

The profile traces of Figure 4-11 plot A are as follows: the solid line represents the DC-based profile trace of $\tilde{\theta}_4$ versus θ_3 as the profiling parameter, the dashed line shows the DC-based trace of $\tilde{\theta}_3$ versus θ_4 as the profiling parameter. In Figure 4-11 plot B the dot-dashed line shows the GLS-based profile trace of $\tilde{\theta}_4$ versus θ_3 as the profiling parameter,

and the dotted line represents the GLS-based profile trace of $\tilde{\theta}_3$ versus θ_4 as the profiling parameter.

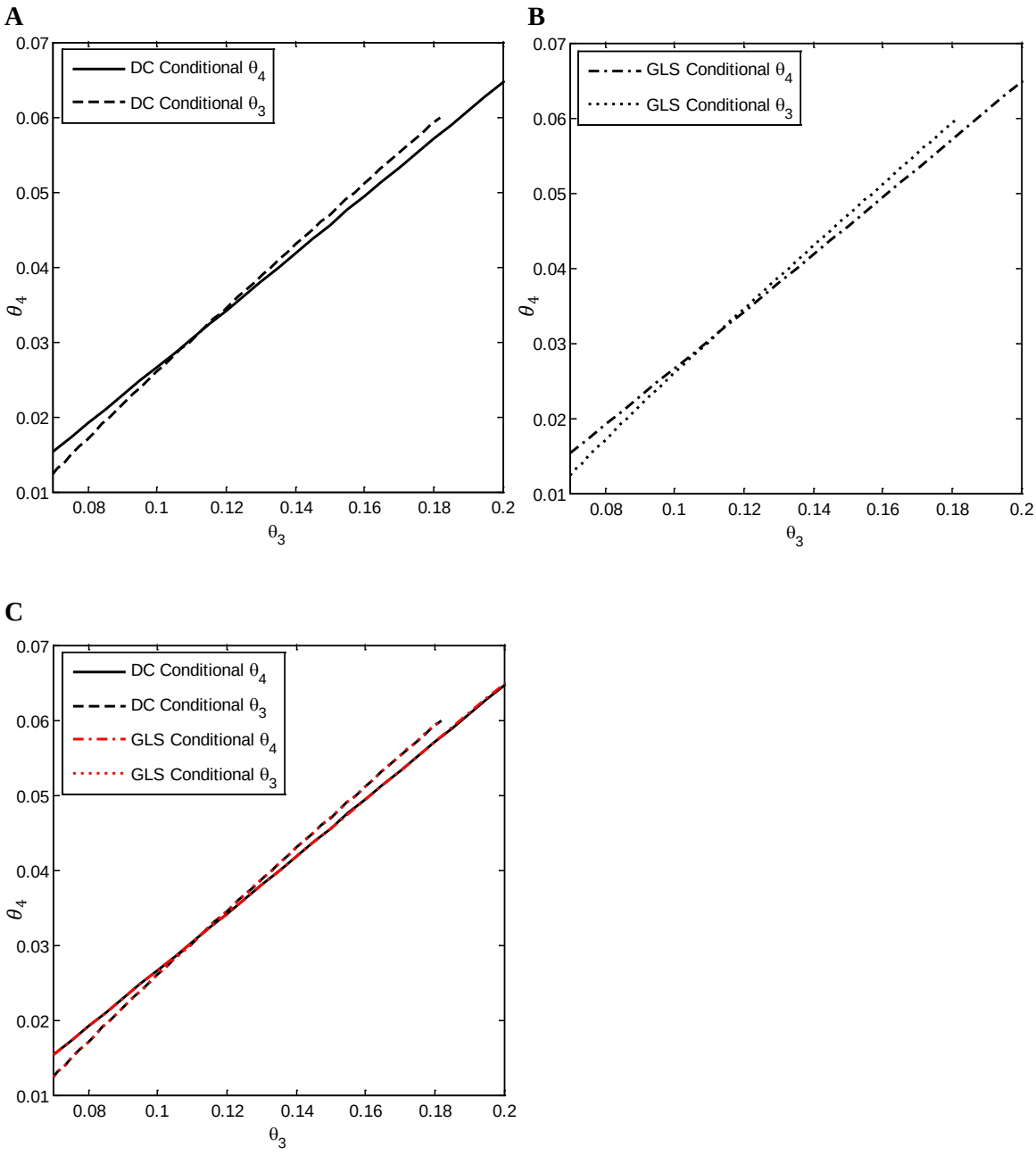

Figure 4-11: Comparison of DC and GLS profile traces for Case study 2 around the DC parameter estimates. Plot A: $\tilde{\theta}_4$ versus θ_3 (solid line) and $\tilde{\theta}_3$ versus θ_4 (dashed line) based on DC method. Plot B: $\tilde{\theta}_4$ versus θ_3 (dot-dashed line) and $\tilde{\theta}_3$ versus θ_4 (dotted line) based on GLS method.

It can be seen in Figures 4-10 that the pairwise profile traces fall fairly close to each other regardless of the profiling method. This is a sign of inevitable high co-dependency between the parameter estimates $\hat{\theta}_1$, $\hat{\theta}_2$. The profile traces of Figures 4-11 are also close, indicating some level of co-dependency between $\hat{\theta}_3$, $\hat{\theta}_4$ which is not as significant as the previous figure.

Overall, the parameter estimates for this example are co-dependent due to the form of the model and/or the method of data collection. Both GLS and DC-based profiling approaches provide reliable results that are not too different from each other in a reasonable neighborhood around the DC-optimal parameter estimates.

4.7 Discussions and Conclusions

In this chapter the Determinant Criterion approach for parameter estimation and profiling was studied and the DC-based profile likelihood function based on the profile *t* function proposed by Soo and Bates (1996) was defined as $PL_D(\theta_k) = \frac{\tilde{D}(\theta_k) - D(\hat{\boldsymbol{\theta}})}{(s')^2}$ (equation 4-20). In this approach the noise covariance matrix is estimated implicitly from the residuals $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Z}(\boldsymbol{\theta})'\boldsymbol{Z}(\boldsymbol{\theta})$ (equation 4-7) and as a result the profiling becomes a joint estimation problem on parameters and the covariance matrix.

The distribution and the corresponding degrees of freedom for this function were discussed and determined in section 4.4. It was shown that considering $nm - p$ degrees of freedom for the F distribution of the PL_D function allows the model to benefit from nm pieces of information and factors in the p constraints associated with the DC-optimal parameter estimates. The application of *DC-based profiling approach* proved to be

successful for parameter inference in nonlinear multi-response systems when an *F* distribution with $nm-p$ degrees of freedom was considered for the DC-based profile likelihood function and the average noise variance was estimated by $s'^2 = \frac{D(\hat{\boldsymbol{\theta}})}{(nm-p)}$.

Once the distribution of the PL_D function was determined, the application of the *DC-based profiling approach* was studied by means of some illustrative examples. For some examples, the dataset was originally synthesized by adding noise with a specific structure to model predictions based on some known parameter values (referred to as "the *true* parameter values" in the literature). This known $\boldsymbol{\Sigma}$ is used here in a GLS-based profiling approach (referred to as '*GLS-based approach with known* $\boldsymbol{\Sigma}$') and the parameter inference results are treated as a reference. The DC-based parameter likelihood intervals were compared against the reference intervals where applicable, and demonstrated relatively good compliance.

Another main focus of this chapter was exploring the differences between the proposed GLS-based profiling approach (*GLS-based profiling approach #2* explained in Chapter 3) and the *DC-based profiling approach* and finding the more appropriate one. The comparison of the GLS and DC estimation methods from a theoretical point of view has been done to some extent in the literature but not in a profiling setting (Phillips, 1976); (Kang and Bates, 1990); (Oxby et al., 2003). Based on the literature review provided in Section 4.6.1, the GLS method seemed to be the more appropriate approach for parameter inference while the DC approach seemed to generate more reliable results for parameter estimation. The theoretical arguments were extended for profiling and also the theoretical expectations were tested by means of some practical examples.

Based on the first case study (Bates and Watts example), *GLS-based profiling approach #2* has an advantage over the DC method since it provides adequate parameter estimates and reliable parameter inference intervals in a sense that they are close to the reference intervals while avoiding numerical issues. This example tangibly depicts a disadvantage of DC-based profiling which is the higher possibility of convergence to local minima. Moreover, when both GLS-based and DC-based profiling are done around the DC-optimal parameter estimates for this example, the DC-based profiling approach leads to poorly conditioned regression and results in highly co-dependent parameter estimates. This is a problem associated with the joint estimation of noise variance/covariance terms and the parameters and can be well depicted by profile traces.

The results of the second case study (Buzzi Ferraris example) suggest that the DC and GLS approaches are both reliable and fairly comparable due to the form of the model as well as the size of the dataset. This example emphasizes the significance of obtaining more measurements and shows the impact of larger datasets on improving the estimation and inference results. Thus, for larger datasets the difference between the two approaches might not be significant in practice. The DC-based approach is less likely to fall into the local minima and the $\widehat{\boldsymbol{\Sigma}}$ estimation in the GLS method is more reliable for bigger datasets.

For the third case study (Beauchamp and Cornell example) *GLS-based profiling approach #2* provides adequate results. However, the *DC-based profiling approach* is more sucessful in providing reasonable parameter inference results while being computationally less costly. Thus, the *DC-based profiling approach* seems like a good choice as long as it avoids the numerical issues.

In summary based on the case studies, the *GLS-based profiling approach #2* is generally a reliable method that provides parameter likelihood intervals that are close to the reference interference results for nonlinear multi-response systems. Therefore, even though the iterated estimation of the noise covariance matrix might appear complicated, the end results are worth the effort. The *DC-based profiling approach* is relatively fast and straightforward in cases where the Determinant function is easy to handle and numerical issues do not distract the solution. For larger datasets there might not be an advantage associated with either of these profiling approaches as the DC-based approach is less likely to fall into local minima and the $\widehat{\boldsymbol{\Sigma}}$ estimation in the GLS method becomes more reliable. Thus, gathering more data where possible is highly recommended.

In conclusion, the choice of the profiling method depends on several factors such as the form of the model, size of the dataset, and purpose of profiling. My overall suggestion is to favor the DC method for parameter estimation if numerical issues can be avoided, and prioritize GLS-based profiling for finding the parameter likelihood intervals. This combination can be successful in generating adequate results with fewer calculations. However, there is a possibility that the noise covariance matrix estimated from the residuals corresponding to DC-optimal parameter estimates is not positive-definite in which case the combination will not yield appropriate results. Therefore, if one choses to use the combined approach, it is important to check that the results are not local, and also test the estimated noise covariance matrix for being positive definite.

For large datasets, both profiling methods have a good chance of providing adequate parameter estimates and inferences. For smaller datasets the results of this research suggest that *GLS-based profiling approach #2* is an appropriate choice.

4.8 Summary

In this chapter the DC method for parameter estimation in nonlinear multi-response models was discussed in detail and DC-based profiling for parameter inference was explored. The distribution of the DC-based profile likelihood function, in particular the corresponding degrees of freedom, was studied concluding that an *F* distribution with $nm - p$ degrees of freedom matches this distribution properly. Therefore, the DC-based profile likelihood plots with reference lines at $F_{1,nm-p;\alpha}$ can be used for parameter inference. This was shown by means of several case studies with nonlinear multi-response models. Eventually the DC approach was compared against *GLS-based profiling approach #2* and based on the current literature and the results obtained in this research, it was suggested that for small datasets the *GLS-based profiling approach #2* is an appropriate choice and it is better to avoid DC for parameter estimation and profiling unless numerical issues can be avoided. However, a combination of DC for parameter estimation and *GLS-based profiling approach #2* for profiling leads to reliable parameter inference results with a relatively straightforward process under certain precautions.

The main contributions of this chapter are as follows:

- Studying the DC-based profiling approach for parameter inference in nonlinear multi-response models and introduction of the DC-based profile likelihood function as $PL_D(\theta_k) = \frac{\tilde{D}(\theta_k) - D(\hat{\boldsymbol{\theta}})}{(s')^2}$ based on the profile *t* function proposed by Soo and Bates (1996).
- Determining the distribution of the DC-based profile likelihood function as $PL_D(\theta_k) \sim F_{1,nm-p}$.

- Comparison of the *DC-based profiling approach* with the proposed *GLS-based profiling approach #2* with the following conclusions:
    - For small datasets the *GLS-based profiling approach #2* is a more appropriate choice and DC should only be used for parameter estimation and profiling if numerical issues can be avoided.
    - Generally a combination of DC for parameter estimation and *GLS profiling approach #2* for profiling leads to reliable parameter inference results with a relatively straightforward process.

Chapter 5

Application of Profiling in Parameter Estimability Analysis

5.1 Introduction

In this chapter, a profile-based exploratory parameter estimability analysis is proposed for nonlinear multi-response models. It is shown that profile plots and profile-based inference intervals can be very useful in determining which parameters of the model are estimable and revealing the potential inestimability issues. Based on the profiling approach offered in the previous two chapters for parameter inference, a profile-based estimability analysis is proposed and its application is demonstrated by means of a case study.

The chapter begins with a review of parameter estimability analysis and its significance in model building. The current methods for this analysis are explained, and some remaining challenges in this analysis are revealed. Then, a four-step profile-based exploratory estimability analysis is proposed which includes studying the profile likelihood plots and profile traces of the model parameters to reveal the co-dependencies among them and make conclusions about their estimability.

Lastly, the proposed method is applied to a case study which consists of a bacterial growth model that is widely used, and a dataset that has been explored previously by researchers in this field due to high potential of inestimability of the system.

5.2 Definition and Importance of Parameter Estimability Analysis

In the previous chapters, several approaches for parameter estimation and inference in nonlinear multi-response systems were discussed. While a number of techniques such as Generalized Least Squares (GLS), Maximum Likelihood Estimation (MLE), and Bayesian methods are used for parameter estimation (e.g., (Bates and Watts, 1988); (Seber and Wild, 1989)), it is important to determine in the first place whether the parameters can be effectively estimated based on the model structure and the available or planned experimental data. The issue of parameter inestimability can be particularly noticeable in biological processes such as pharmacokinetic models, and systems biology models, as well as chemical process models such as polymer reactor models, and combustion models that deal with a large number of parameters and also face some limitations around measurements and data collection (Dochain et al., 1995); (Wu et al., 2008). Due to the lack of proper measurements or suitable experimental conditions, it might not be possible to obtain appropriate estimates for all parameters which potentially reduces the reliability of model predictions (Benyahia et al., 2013).

Identifiability and estimability analyses are common ways to decide whether a unique set of parameters can be estimated for a particular statistical model. *Parameter Identifiability* (also referred to as 'Structural Identifiability') is the ability to uniquely estimate model parameters given a particular model structure. Parameter Identifiability analysis focuses fundamentally on the model structure and studies the possibility of obtaining unique parameter estimates for the model without considering the choice of experimental runs, i.e., the possibility of estimating the parameters given any possible experimental runs. On the other hand, *Parameter Estimability* analysis (also known as 'Practical Identifiability'

or ‘Quantitative Identifiability’) investigates unique values for parameter estimates given the experimental data, and is usually defined in the context of existing or planned experimental runs (Cobelli and DiStefano, 1980). Identifiability and estimability analyses have been recognized and gained importance in a number of fields such as statistics, econometrics, control and systems engineering, and compartmental analysis (Jacquez and Grief, 1985).

If a model is identifiable, the model parameters can be potentially uniquely estimated based on the model structure (Holmberg, 1982). However, identifiable parameters might still not be estimable due to insufficient experimental data (McLean and McAuley, 2012). In practice the problem of highly correlated parameter estimates can arise for limited and noisy data. In such circumstances the uniqueness of parameter estimates that was theoretically guaranteed by identifiability might not be valid any longer (Dochain et al., 1995); (Vanrolleghem et al., 1995).

The inestimability problem appears when the predictions are insensitive to some parameter values, or when the effect of one parameter on the model is not independent from the effect of one or more other parameters (McLean and McAuley, 2012). As a result, methods for testing parameter estimability often involve investigating the sensitivity for the model and runs, and exploring the noise covariance matrices to examine the co-dependencies between the noise terms in multi-response models. In practice, inestimability is often due to insufficient experimental data arising from constraints on sample collections, or limited accuracy of the measurements, combined with model structure and parameterization. Parameters of kinetic models in particular are frequently inestimable when a particular set of noisy data does not allow for obtaining

reliable parameter estimates (Vajda et al., 1989). However, this issue is by no means limited to kinetic models.

Probably the best way to resolve parameter inestimability is to collect additional experimental data when possible and/or to change the parameterization of the model (Cobelli and DiStefano, 1980). Alternatively if more experimental runs are not feasible (e.g., due to production constraints), other solutions can be used. The solutions proposed in the literature mostly revolve around simplifying the model by eliminating some terms from the model, or lumping the highly co-dependent parameter estimates together into a single overall parameter, or fixing some parameters at reasonable values and therefore leaving them out of the estimation problem (McLean and McAuley, 2012).

Estimability issues can sometimes be resolved by carrying out more experimental runs in the regions where the data are insufficient. Therefore, in order to avoid estimability issues later on, it is wise to consider estimability analysis during the experimental design stage (Cobelli and DiStefano, 1980); (Raue et al., 2011). There is a strong relationship between parameter estimability and experimental design since the sensitivity matrix and consequently the Fisher Information Matrix are often used in methods for both of these analyses (Yao et al., 2003). The efficiency of different experimental designs for arriving at reliable parameter estimates can be tested based on the determinant of the Fisher Information Matrix which is formed from the sensitivity matrix (Yao et al., 2003). One of the main advantages of estimability analysis is that it can be carried out on planned experimental data. Once the potentially inestimable parameters are determined, the experimental design can be changed to avoid this situation in the actual experimental runs or model re-parameterization or some fixed parameter values can be imposed. The

sensitivity matrix and the Fisher Information Matrix can be modified to accommodate the new suggested experimental conditions. Thus, estimability analysis can provide significant assistance in experimental design.

While identifiability analysis has received a considerable amount of attention (see for example, (Bellman and Astrom, 1970); (Vajda, 1983); (Jacquez and Grief, 1985); (Chappell et al., 1990); (Asprey and Macchietto, 2000); (Ben-Zvi, 2004); (Ben-Zvi et al., 2004)), parameter estimability is still not fully explored. The main focus of this chapter is parameter estimability analysis using profile-based parameter likelihood intervals that were studied in the previous chapters. A review of the current state-of–the-art for parameter estimability analysis is provided in the next section.

5.3 Background on Parameter Estimability Methods

Estimability analysis is a growing field and a number of estimability methods have been proposed (Holmberg, 1982); (Petersen et al., 2001); (Yao et al., 2003); (Rodriguez-Fernandez et al., 2006); (Nikerel et al., 2009). A thorough review of the estimability techniques can be found in the recent paper by McLean and McAuley (2012). Based on their review, current parameter estimability techniques fall into three main categories: methods based on the sensitivity and covariance matrices, methods that require repeated parameter estimation such as Monte Carlo and Bootstrap methods, and finally graphical or visual methods that inspect the sensitivity functions with respect to parameters (McLean and McAuley, 2012).

The sensitivity-based methods usually use the information provided by the approximated Fisher Information Matrix (**F**) defined as follows to draw conclusions about parameter estimability (Petersen et al., 2001):

$$\boldsymbol{F} = \boldsymbol{V}(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\widehat{\boldsymbol{\theta}}) \quad (5\text{-}1)$$

In this equation, $\boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ is the matrix of first-order sensitivities of the model predictions with respect to the parameters evaluated at the least squares parameter estimates, and $\boldsymbol{\Omega}$ is the overall noise covariance matrix. In theory the rank of the Fisher Information matrix indicates the number of estimable parameters and rank deficiency of this matrix can be a sign of inestimability. However, to make a firm conclusion about the parameter estimability in the case of rank deficiency of F, the condition number (ratio of the largest to the smallest singular value) of this matrix should also be consulted. If the condition number is larger than a particular threshold, some parameters are inestimable (Petersen et al., 2001); (Yao et al., 2003).

There are several sensitivity-based methods for assessing the identifiability and estimability of large-scale nonlinear multi-response ODE models including the method proposed by Vajda et al. (1989) who use the eigenvalues of the Fisher Information Matrix (Vajda et al., 1989), and by Yao et al. (2003) who use a sequential orthogonal projection approach for simplifying the model and reducing the number of parameters being estimated (Yao et al., 2003).

A substantial amount of research on identifiability and estimability analyses of multi-response models has been conducted by Vajda et al. (see for example, (Vajda, 1983); (Vajda, 1984); (Vajda et al., 1989); (Chappell et al., 1990)). Their research mostly revolves around nonlinear multi-response kinetic models which are considered to be one

of the major sources of nonlinearity in chemical engineering and frequently happen to have highly co-dependent parameters. In their paper, Vajda et al. (1989) explain that although the SSE function is supposed to take its minimum at the least squares parameter estimates ($\widehat{\boldsymbol{\theta}}$), there might be some other parameter values $\left(\breve{\boldsymbol{\theta}}\right)$ far from the least squares estimates that result in almost the same values for the SEE function so that for small values of ε (Vajda et al., 1989):

$$S\left(\breve{\theta}_k\right) - S(\hat{\theta}_k) \leq \varepsilon \tag{5-2}$$

This region is called the ε-uncertainty region and can be approximated by:

$$\left(\breve{\theta}_k - \hat{\theta}_k\right)' \boldsymbol{V}(\widehat{\boldsymbol{\theta}})' \, \boldsymbol{\Omega}^{-1} \boldsymbol{V}(\widehat{\boldsymbol{\theta}}) \, \left(\breve{\theta}_k - \hat{\theta}_k\right) \leq \varepsilon \tag{5-3}$$

The sensitivity matrix $\boldsymbol{V}(\widehat{\boldsymbol{\theta}})$ in this equation is evaluated at the least squares parameter estimates. If the Fisher Information Matrix (equation 5-1) is nearly singular, it has at least one very small (nearly zero) eigenvalue. They consider an eigenvalue (λ_i) small if $\lambda_i < 100\sigma^2$ where σ^2 is the residual variance. In such instances, the parameter estimates for which the value of $S(\theta)$ is the same are highly co-dependent, the corresponding uncertainty region of equation 5-3 is elongated, and thus those parameters are inestimable. Therefore, Vajda et al. (1989) define and use a modified eigenvalue decomposition or Principal Component Analysis (PCA) on the Fisher Information Matrix to reveal the parameter co-dependencies and unbounded uncertainty regions that may cause inestimability. In order to resolve inestimability, the large eigenvalues and the elements of the corresponding eigenvectors are used to find parameter combinations that can help simplify the model. For example, for two highly co-dependent parameters, they apply logarithmic transformations of parameters which make the predictions sensitive to the ratio of the two parameters as opposed to individual parameters. Based on this

method, the number of large eigenvalues is an indication of the number of estimable parameters in the re-parameterized and simplified model. Overall, this method reveals near singular Fisher Information Matrices and potential parameter co-dependencies which can be used to make decisions about parameter inestimability and offer ways to resolve it (Vajda et al., 1989).

One of the disadvantages of the estimability analysis proposed by Vajda et al. (1989) is the application of parameter transformations. This method produces a list of identifiable transformed parameters rather than the original parameters. Moreover, although this method works well on some examples, Yao et al. (2003) pointed out that it is not very practical for models with large number of parameters since examining all eigenvalues and eigenvectors with large numbers of elements becomes relatively time-consuming and difficult. Alternatively they suggested another method for models with a large number of parameters and responses. Their method is a sequential orthogonal projection approach in which parameters with large sensitivity vectors are identified and set aside. The remaining sensitivity vectors are orthogonalized with respect to these identified vectors. The algorithm proceeds iteratively until a convergence criterion is met as summarized in Table 5-1 (Yao et al., 2003).

Table 5-1: A sensitivity-based methodology for choosing a subset of estimable parameters (Yao et al., 2003)

1	Calculate the magnitude of each column (sum of squares of all elements) of the scaled sensitivity matrix (**V**)
2	Select the parameter corresponding to the largest column as the first estimable parameter
3	Mark the corresponding column as X_L ($L = 1$ for the first iteration)
4	Calculate $\widehat{\boldsymbol{V}}_L = X_L(X_L{}'X_L)^{-1}X_L{}'\boldsymbol{V}$
5	Initiate orthogonalization by calculating the residual sensitivity matrix $\boldsymbol{R}_L = \boldsymbol{V} - \widehat{\boldsymbol{V}}_L$
6	Calculate the magnitude of each column of $\boldsymbol{R}_L$. The parameter corresponding to the largest column is the next estimable parameter
7	Select the column corresponding to the new estimable parameter in **V** and make X_{L+1} from the previous X_L and this new column.
8	Go to the next iteration and repeat steps 4 to 7 until the column with the largest magnitude in the residual sensitivity matrix is smaller than a cut-off value

This orthogonalization approach has the advantage of providing an efficient way to find the original parameters with the most impact on the model predictions. Furthermore, in the case of co-dependent parameters, this method makes it possible to choose the more influential parameters and exclude the others. However, the disadvantages include the potential high sensitivity of the results to initial parameter guesses and the possibility of local results instead of global (McLean and McAuley, 2012); (Benyahia et al., 2013).

Another issue is that this method depends highly on the scaling used for the responses and parameters, which is also the case with the PCA approach of Vajda et al. (1989).

The orthogonalization approach of Yao et al. (2003) has been well received and widely used in different research areas (e.g., (Jayasankar et al., 2009); (Benyahia et al., 2013)). Kou et al. (2005) also used this method to investigate the estimability of the parameters of a model for gas-phase Ethylene/Hexene copolymerization with Metallocene Catalyst with a modification. Furthermore, they suggested that since this estimability analysis depends on the sensitivity matrix which is in turn impacted by the value of the corresponding parameter estimates, the sensitivity matrix and parameter ranking should be updated following each iteration during parameter estimation. In order to overcome the inestimability issues in their case study which contained a large number of parameters, they determined the most estimable subset of parameters and fixed them at their estimated values. Subsequently, the estimation of the less estimable parameters was facilitated further by separating the fairly co-dependent parameters into separate estimation groups. This method led to good results for the estimable parameters and provided reasonable estimates for the less-estimable parameters conditional on the values of the estimable ones (Kou et al., 2005).

One of the main disadvantages of the sensitivity-based methods is that they are based on linearization of the model function which provides results that are not necessarily global. Also, they depend highly on good initial parameter estimates, can be very sensitive to scaling of parameters, require an arbitrary cut-off value for deciding whether a parameter is estimable or not, and overall they rely on reasonably good judgment on the researcher's part (McLean and McAuley, 2012).

Graphical methods have been successfully used to draw conclusions about parameter estimability (e.g., (Holmberg, 1982)). Appropriate plots make it easy to spot regions where model predictions are not impacted by changes in the values of a parameter estimate, and where different parameters have a similar impact on the model predictions. Some of these methods study the trajectories of the sensitivity-based functions with respect to parameters. For example, Holmberg (1982) analyzed plots of the transients of the state variables versus time for different values of parameters, changing one parameter estimate at a time to show how changes in each parameter impacted the shape of the transient plot and consequently how the impact of one parameter could be compensated by another one. They also studied plots of the sensitivity function versus time with respect to each parameter. These sensitivity plots indicate the parameters that have the greatest impact on model predictions at a certain time which is important for parameter ranking. Parameter scaling is very important for these plots to ensure that the small sensitivity values are not a result of the chosen measurement units. Overlaying the corresponding plots for all parameters in one figure makes it possible to observe any trends and co-dependencies that lead to balancing off the effect of one parameter by others. Then, if a set of parameter estimates seems to be problematic based on these plots, they are more closely studied by being plotted against each other to reveal the potential co-dependencies or linear relationships. In order to estimate the parameters, Holmberg (1982) used the least squared errors method with a quadratic objective function, not accounting for the noise covariance matrix with x (concentration of biomass in g/l) and s (concentration of the growth-limiting substrate in g/l) denoting the responses, and t_i (time in hr) representing the regressor variable as follows (Holmberg, 1982):

$$S(t_i) = \sum_{i=1}^{n}\{a_x\big(x_m(t_i) - x(t_i)\big)^2 + a_s\big(s_m(t_i) - s(t_i)\big)^2\} \quad (5\text{-}4)$$

In this equation $x_m(t_i)$ and $s_m(t_i)$ represent the observations for biomass and growth-limiting substrate concentrations (g/l) respectively, $x(t_i)$ and $s(t_i)$ are the model predictions and a_x and a_s as weighting factors for the responses respectively.

Since the measurement noise and the parameter uncertainty play an important role in the estimability analysis, another category of estimability methods relies on repeated parameter estimation by means of techniques such as Monte Carlo, Jackknife or Bootstrap to explore the noise covariance and parameter covariance matrices (e.g., (Efron, 1982); (Nikerel et al., 2009)). In these methods, high level of uncertainty (large variance) in the resulting parameter estimates indicates inestimability. Thus, one issue with such methods is finding an appropriate cut-off value so that if the uncertainty level of the parameter estimates is beyond that threshold, the parameter is considered inestimable. Another concern is the amount of computation that is involved in repeated parameter estimations when applied to relatively complex models (McLean and McAuley, 2012). In their review, McLean and McAuley (2012) have placed profiling in this category since it involves repeated parameter estimations. Yet, profiling has the advantage of providing graphical results that are relatively easy to interpret.

Most of the research in using profiling for estimability analysis has been done by Raue et al. whose research is mostly focused on systems biology (see for example, (Raue et al., 2009); (Raue et al., 2010); (Raue et al., 2011); (Kreutz et al., 2012)). The biological process models that they consider are usually nonlinear, high-dimensional, with insufficient data points which lead to inestimability of parameters. Sets of ordinary differential equations (ODEs) and dynamic models are usually used to describe such

processes (Raue et al., 2010). They address identifiability and estimability analyses by examining profile likelihoods for parameters and they recognize parameter inestimability from likelihood intervals that are infinitely extended with increasing or decreasing parameter values. For parameter estimation in a multi-response model with m output variables and n runs, they use the following objective function (Raue et al., 2009):

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{{\sigma_j}^2} (y_j(\boldsymbol{x}_i, \boldsymbol{\theta}) - f_j(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 \quad (5\text{-}5)$$

In this equation ${\sigma_j}^2$ are the measurement error variances which are assumed to be known. Thus, noise terms are distributed as $\varepsilon_j \sim N(0, {\sigma_j}^2)$ assuming *IID* Normal between-run noise. This leads to a χ^2 distribution for the objective function $L(\boldsymbol{\theta})$. Consequently, profile likelihood function for each parameter is obtained from:

$$L_{PL}(\theta_k) = \min_{\theta}[L(\boldsymbol{\theta})] \quad (5\text{-}6)$$

This equation involves the typical profiling procedure of re-optimizing the objective function $L(\boldsymbol{\theta})$ with respect to all other parameters for fixed values of θ_k and coming up with sets of parameters that minimize the objective function conditional on θ_k. Profile likelihood plots of $L_{PL}(\theta_k)$ versus θ_k are used for estimability analyses. Plots of $L_{PL}(\theta_k)$ versus θ_k that have a minimum but never exceed a certain threshold for increasing or decreasing values of θ_k indicate inestimability. The confidence interval of an inestimable parameter might have an upper or a lower bound and be unlimited on just one side. For building the profile-based parameter confidence intervals for estimability analysis, they use $L_{PL}(\theta_k) - L_{PL}(\widehat{\boldsymbol{\theta}})$ with a threshold of Δ_α which is the $(1-\alpha)$ quantile of a χ_1^2 distribution. The degrees of freedom for this distribution is equal to one based on the one constraint that is caused by the fixed value of the profiling parameter (Raue et al., 2010).

In order to resolve inestimability by planning new experiments, Raue et al. (2010) suggest making new observations where the uncertainties in parameter values have the most impact on the predictions. To find these intervals, the trajectories along the profile likelihood of each parameter should be consulted. Profile traces of several parameters (conditional on the value of the profiling parameter) versus the profiling parameter are also used to indicate the regions of high co-dependencies among parameters and the results are applied for new experimental design (Raue et al., 2010). In order to assess the impact of the new measurement on improving the estimation results, Monte Carlo simulations can be applied (Raue et al., 2009).

Based on this literature review the application of profiling can be extended to parameter estimability analysis which is one of the objectives of this thesis. In the previous chapters a profiling approach was proposed for parameter inference in multi-response systems (*GLS-based profiling approach #2*). The following section will explain how the proposed profiling approach can be applied to parameter estimability. The proposed approach is more general than the previous profiling method by Raue et al. (2009) since it accounts for the noise co-dependencies and accommodates a within-run noise covariance matrix that is not necessarily diagonal. Furthermore, the threshold for determining estimable parameters by means of their likelihood interval is carefully defined based on the distribution of the profile likelihood function, taking the estimated noise covariance matrix into account. Lastly, both GLS-based and DC-based profiling methods are used to compare the results.

5.4 Proposed Profile-Based Exploratory Parameter Estimability Analysis

Previously in this thesis profiling was discussed for the following multi-response model with *p* parameters, *m* responses and *n* experimental runs:

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{5-7}$$

with $\boldsymbol{Y}$ being the $n \times m$ response matrix, $\boldsymbol{\eta}(\boldsymbol{\theta})$ being the matrix of expectation functions and $\boldsymbol{\varepsilon}$ representing noise. The DC-based profiling was carried out on this matrix representation of the model (Chapter 4), whereas for GLS-based profiling (Chapter 3) the responses, model predictions and noises were stacked in vectors leading to the following model representation:

$$\boldsymbol{y_{vec}} = \boldsymbol{\eta_{vec}}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon_{vec}} \tag{5-8}$$

In Chapter 3, the GLS-based profile likelihood function was defined as $PL_S(\theta_k)$ as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \tag{5-9}$$

In this equation $S(\widehat{\boldsymbol{\theta}})$ represents the SSE function evaluated at the least squares parameter estimates and $\tilde{S}(\theta_k)$ is the value of the SSE function at the least squares estimates of parameters conditional on the fixed value of θ_k. Also, the DC-based profile likelihood function $PL_D(\theta_k)$ was described in Chapter 4 as follows:

$$PL_D(\theta_k) = \frac{\widetilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}})}{(s\prime)^2} \tag{5-10}$$

where $\widetilde{D}(\theta_k)$ is the value of the Determinant function at the DC-optimal parameter estimates conditional on the fixed value of θ_k, $D(\widehat{\boldsymbol{\theta}})$ is the Determinant function evaluated at the DC-optimal parameter estimates, and s'^2 is the DC-based estimated average noise variance.

It was shown that using the GLS method for a multi-response model with known noise covariance matrix, a profile likelihood plot of $PL_S(\theta_k)$ versus θ_k with a reference line at

$\chi^2_{1;\alpha}$ generates the $100(1-\alpha)$ likelihood interval for θ_k. For models with unknown noise covariance matrix, an iterative two-step GLS-based profiling approach was proposed in Chapter 3 for estimating the noise covariance matrix and the parameters. For such cases a plot of the GLS-based profile likelihood function $PL_S(\theta_k)$ versus θ_k with a reference line at $F_{1,nm-p;\alpha}$ was used to build the likelihood intervals for θ_k. Similarly based on the discussions of Chapter 4 about the DC method, a profile plot of the DC-based profile likelihood function $PL_D(\theta_k)$ versus θ_k with a reference line at $F_{1,nm-p;\alpha}$ was used for making the DC-based parameter inference. These inference intervals can be used towards parameter estimability since a parameter with an unbounded likelihood interval is not estimable.

In this section, a profile-based exploratory method for parameter estimability analysis is proposed that consists of four steps. The first step consists of a sensitivity-based estimability analysis of the system at nominal parameter values using the orthogonalization approach of Yao et al. (2003). This serves as a screening process which selects a set of potentially estimable parameters. The second step involves forming and studying the profile likelihood plots for all of the parameter selected by step one. The parameters with finite likelihood intervals are likely to be estimable. However, in order to confirm the estimability of the selected parameters, the profile traces of these parameters are studied in the third step to make sure a high level of co-dependency that might lead to inestimability does not exist among them. In the last step, the parameters that are considered to be inestimable are chosen as the profiling parameters and the profile traces of the selected estimable parameters are generated based on fixed values of these profiling parameters to observe any noticeable trends and co-dependencies. Overall, the

first and second steps are the main analysis and the subsequent steps are mostly for testing the results of the main steps and confirming the estimability of the selected parameters. The proposed profile-based exploratory parameter estimability analysis is schematically described in Figure 5-1.

As mentioned before, in profile-based parameter estimability analysis, a flat or nearly flat profile likelihood plot that does not exceed a certain threshold is an indication of inestimability of the profiling parameter (Raue et al., 2009). This concept is illustrated in Figure 5-2 schematically. In this figure, Plot A shows a profile likelihood plot for an estimable parameter where the reference line crosses the profile curve twice, marking the boundaries of the parameter likelihood interval. Figure 5-2 Plot B shows an inestimable parameter for which the profile likelihood plot becomes nearly flat on both sides, resulting in an unbounded likelihood interval for the profiling parameter. Plot C of this figure also depicts inestimability of the parameter by means of a profile likelihood plot that is nearly flat on one side, and does not exceed the threshold marked by the reference line. It should be noted that all plots should have a minimum to ensure that the parameters are identifiable and that the profiling calculations are feasible.

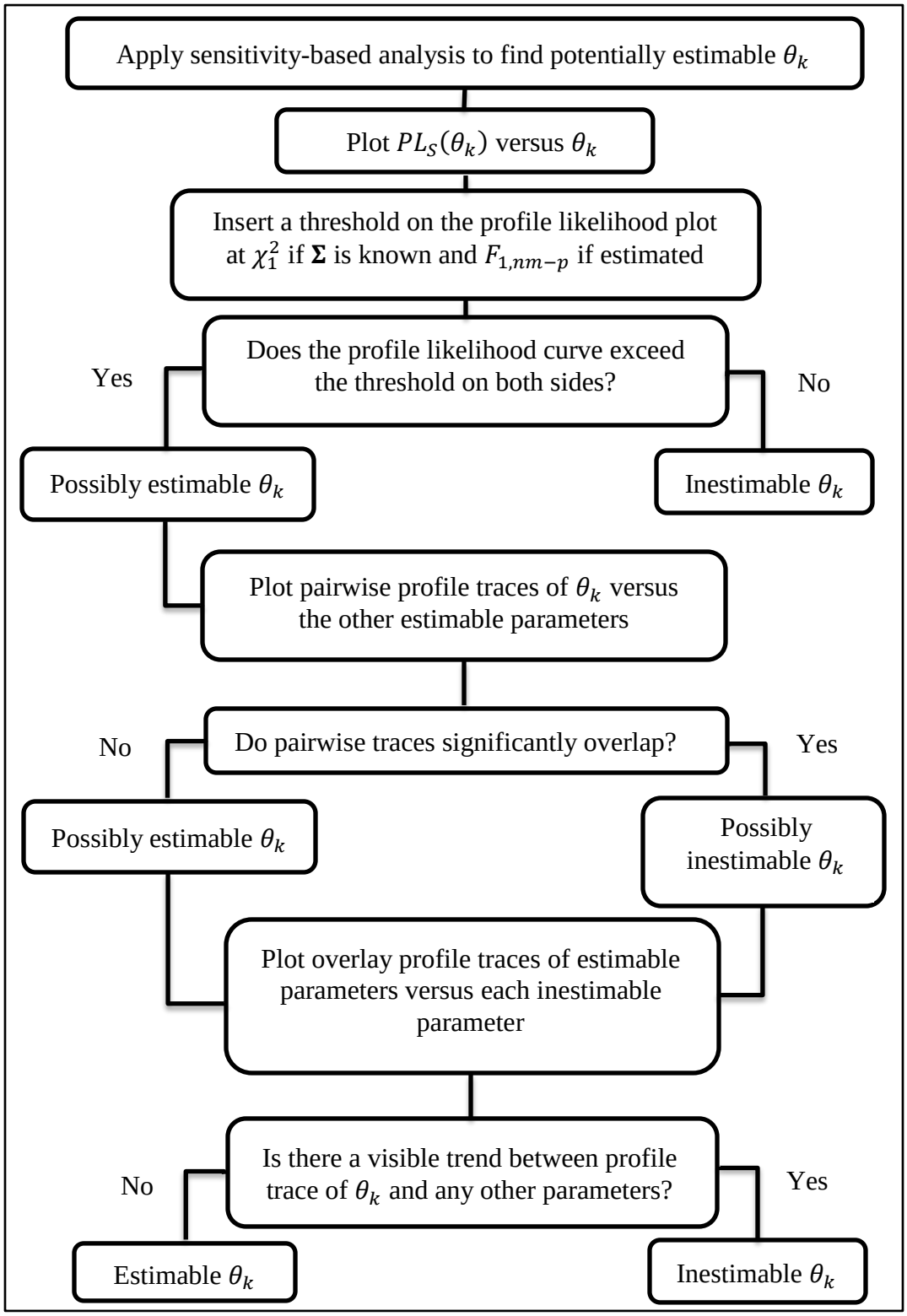

Figure 5-1: Summary of the proposed profile-based estimability analysis

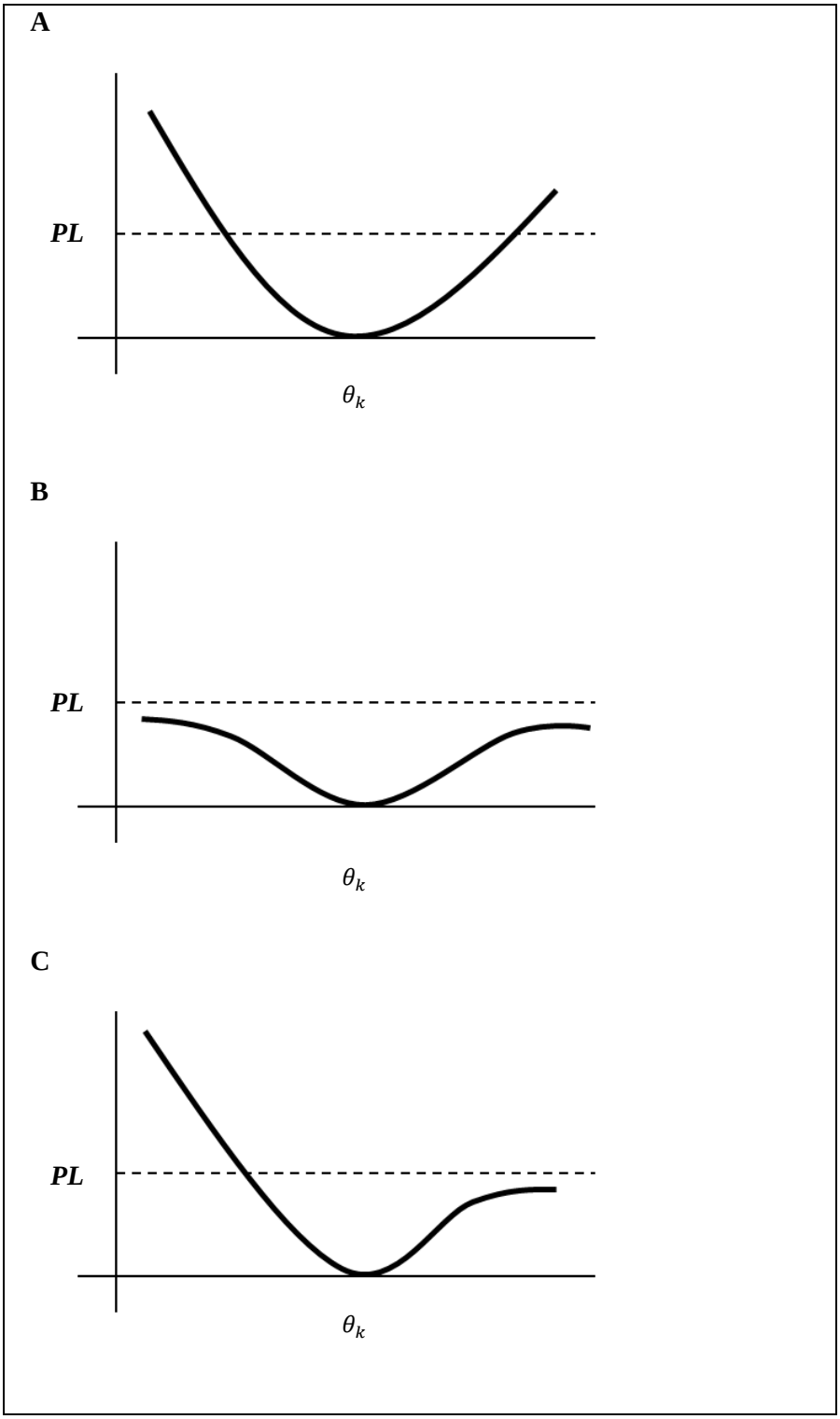

Figure 5-2: Schematic presentation of estimable and inestimable parameters. Plot A: Profile likelihood plot for an estimable parameter with finite likelihood interval on both sides. Plot B: Profile likelihood plot for an inestimable parameter with unbounded likelihood interval on both sides. Plot C: Profile likelihood plot for an inestimable parameter with unbounded likelihood interval on one side (after (Raue et al., 2009)).

The schematic plots of Figure 5-2 can be explained mathematically to clarify this approach. Applying the GLS-based profiling to a nonlinear model with a known or estimated noise covariance matrix, a flat or nearly flat profile likelihood plot of $PL_S(\theta_k)$ versus θ_k can be interpreted as:

$$PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}}) \cong 0 \tag{5-11}$$

Since $S(\widehat{\boldsymbol{\theta}})$ has a specific constant value, for the above equation to be true, the value of the conditional SSE function $\tilde{S}(\theta_k)$ should be equal to the value of $S(\widehat{\boldsymbol{\theta}})$:

$$\tilde{S}(\theta_k) = S\left(\widehat{\boldsymbol{\theta}}\right) = c_1 \tag{5-12}$$

where c_1 is a constant value. Therefore,

$$\tilde{S}(\theta_k) = \left(\boldsymbol{y_{vec}} - \boldsymbol{\eta}(\theta_k)\right)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y_{vec}} - \boldsymbol{\eta}(\theta_k)) = c_1 \tag{5-13}$$

In this equation, $\boldsymbol{y_{vec}}$ is the vector of measured responses and the overall noise covariance matrix $\boldsymbol{\Omega}$ is also fixed at known or estimated values during profiling using *GLS-based profiling approach #2*. Equation 5-13 implies that:

$$\frac{\partial \tilde{S}(\theta_k)}{\partial \theta_k} \approx 0 \tag{5-14}$$

for values of θ_k over a wide interval, i.e., if one wanted to find the optimal values of θ_k, the optimization problem is indeterminate. A range of values of θ_k produce the same SSE which implies inestimability of θ_k. The interesting by-product of this discussion is that profiling implicitly identifies relationships between the parameter estimates that lead to inestimability. These can be seen in the profile traces.

The same argument can be made for the DC-based profiling since a profile plot that becomes flat or nearly flat in some intervals indicates a constant value for the Determinant function. This can be expressed by the following sequence of equations:

$$PL_D(\theta_k) = \frac{\widetilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}})}{(s\prime)^2} \cong 0 \tag{5-15}$$

$$\widetilde{D}(\theta_k) - D(\widehat{\boldsymbol{\theta}}) = 0 \tag{5-16}$$

$$\widetilde{D}(\theta_k) = D(\widehat{\boldsymbol{\theta}}) \tag{5-17}$$

Thus, $\widetilde{D}(\theta_k)$ is constant regardless of the changes in the value of the profiling parameter, indicating the insensitivity of the model predictions to changes in the parameter θ_k.

Based on this discussion, the four-step profile-based parameter estimability analysis unfolds as follows:

In the first step the model is analyzed with respect to all parameters using a sensitivity-based analysis to choose potentially estimable parameters. However, due to the shortcoming of the sensitivity-based analysis these parameters are further studied in the second step where the profile likelihood plots are generated for all of them. The parameters with profile likelihood plots that do not exceed a certain threshold on both sides are considered to be inestimable. I recommend using the same threshold that is used for parameter inference in order to link estimability with parameter inference. Thus, a reference line at $\chi^2_{1;\alpha}$ for GLS-based profile plots with known noise covariance matrix, and at $F_{1,nm-p;\alpha}$ for GLS or DC-based profile plots with unknown noise covariance matrix can serve as a threshold. If the profile likelihood curve goes beyond this reference line on both sides, the profiling parameter is considered to be estimable. Using these thresholds, the likelihood intervals are unbounded on one or both sides for inestimable parameters. For more details on the determination of these reference lines please refer to Chapter 3 section 3.4 and Chapter 4 section 4.4.

The process of these steps leads to the selection of a set of parameters (referred to as the '*selected estimable parameter set*') that are very likely to be estimable. However in order to make well-founded conclusions, I recommend further examination of these parameters in another step. The third step of the proposed approach involves generating the profile traces of the *selected estimable parameter set* to investigate the co-dependencies among them. In this step for any two parameters from the *selected estimable parameter set*, denoted by θ_k and θ_r, pairwise profile traces are formed by plotting $\tilde{\theta}_k$ (the estimated value of θ_k conditional on a fixed value of θ_r) versus θ_r as the profiling parameter and $\tilde{\theta}_r$ versus θ_k. For these parameters the expectation is to obtain profile traces that cross each other but do not overlap in large intervals. Overlapping of the pairwise profile traces is an indication of high level of nonlinear co-dependency between the parameter estimates that might lead to inestimability issues. One of the advantages of including this step is that nonlinear co-dependencies between the parameter estimates are revealed without the need for carrying out any more calculations. All the parameter estimates required for forming the profile traces are already determined during the profiling process of the second step. Moreover, the profile traces are only plotted for the *selected estimable parameter set* as opposed to all of the model parameters which reduces the number of plots that need to be analyzed.

Once the estimable and inestimable parameters are determined, another useful analysis is studying the overlay plot of profile traces of the *selected estimable parameter set* versus fixed values of each of the inestimable parameter. Building and studying these plots is the focus of the fourth step of the proposed approach. If the curves corresponding to two parameters indicate that one parameter behaves in the opposite way as the other one, the

two parameters cannot be estimated separately. For example consider the case where an inestimable parameter denoted by θ_s is chosen as the profiling parameter for this analysis. If the profile trace of $\tilde{\theta}_k$ (the estimated value of θ_k conditional on a fixed value of θ_s) versus θ_s increases in the same way as the profile trace of $\tilde{\theta}_r$ (the estimated value of θ_r conditional on a fixed value of θ_s) versus θ_s decreases, then θ_k has the opposite impact as θ_r on model predictions and thus the two parameters are not estimable. Generally when generating profile traces it is recommended to use the studentized parameters to avoid scaling issues (Bates and Watts, 1988). For this step as well as step three, no extra calculations are required since all the parameter values that are used in for producing the plots are already determined in step two.

If the outcomes of all steps are in favor of the estimability of a parameter, then a firm conclusion can be made about that parameter. Otherwise, there is a possibility of inestimability for that parameter and it is wise to use another method such as a sensitivity-based method to come to a final decision.

As explained, the third step of the analysis mainly focuses on a considerably high degree of co-dependency between the parameter estimates which might cause inestimability issues. Looking back at the discussions in Chapter 4 section 4.6 regarding the comparison of GLS-based and DC-based profiling, the DC-based profiling approach sometimes leads to profile traces that fall very close to each other and even overlap for the most part which is an indication of stronger co-dependencies between the parameter estimates. The GLS-based profiling method is usually less prone to this issue and is more successful at depicting the co-dependencies between the parameter estimates particularly for small datasets. Therefore, I suggest giving priority to GLS-based profiling for the proposed

profiling method unless the DC-based profiling has proven advantageous for the particular example under study.

Regarding the proposed method it should be noted that the first and second steps are the most important and the following steps (examining the profile traces) might or might not be feasible depending on the model and the dataset. The reason is if there are some inestimable parameters in the system, the estimations of the other parameters are also impacted by the inestimable ones and it might be difficult to form meaningful profile traces for the estimable set of parameters since the model predictions are insensitive to the values of the inestimable parameters. In other words, during the estimations for profile traces for the estimable parameters, the value of the inestimable parameters might remain at their initial values, interfering with the conditional estimations of the other parameters. This has been shown in the Case study presented later in this chapter. Thus, the third step of the analysis is advantageous for cases where building the profile traces is feasible. An indication of the feasibility might be to check if all of the parameter values have differed from their initial values at the end of profiling. The main focus in this step should be where the profile traces fall with respect to each other, indicating the co-dependencies between the parameter estimates. This will be discussed in more detail in the Case study provided in section 5.6.

5.5 Advantages and Applications of the Proposed Profile-Based Exploratory Parameter Estimability Analysis

In this section the advantages of using the proposed parameter estimability analysis are described and it is shown how this method overcomes some of the challenges that are

currently encountered in other estimability approaches. Moreover, the application of this analysis for resolving parameter inestimability is explained.

As mentioned in the literature review of section 5.3, one of the challenges in using profiling as well as many other estimability methods is coming up with an appropriate threshold for the objective function to decide between the estimable and inestimable parameters. For instance, in estimability methods that explore the parameter variance/covariance estimates by means of techniques such as Monte Carlo, Jackknife or Bootstrap, the cut-off value pertains to the uncertainty level of the parameter estimates, i.e., if the variance of a parameter estimate is more than the cut-off value, the parameter is deemed inestimable (e.g., (Efron, 1982); (Nikerel et al., 2009)). In profile-based methods the threshold for deciding about estimability is defined for the value of the profile likelihood function and depends on the distribution of this function. I suggest using the same threshold that is used for parameter inference in order to link estimability with parameter inference. By doing so, the calculations for determining the parameter inference intervals can be applied to estimability analysis which significantly reduces the total calculation costs. The proposed estimability method uses a threshold at $\chi^2_{1;\alpha}$ or $F_{1,nm-p;\alpha}$ depending on the distribution of the profile likelihood function. Thus, this method accounts for the noise covariance matrix and its estimation procedure in order to determine the estimability of the model parameters. This is a major advantage of the method proposed in this thesis compared to the existing profile-based method by Raue et al. (2009).

This profile-based estimability analysis is computationally cost-effective since it relates the parameter estimability to parameter inference and by doing so it uses the outcomes of

the parameter inference stage in modeling. This saves a considerable amount of calculation for the estimability phase. Also, all of the steps of the analysis can be done based on the parameter estimates and conditional parameter estimates that are generated in the first and second steps. Therefore, the third and fourth steps do not impose any more calculations and are mostly concerned with analysis of the existing estimations and finding ways to view the results from different angles.

Both GLS-based and DC-based profiling methods can be used for parameter estimability. It is important to adjust the threshold accordingly. As explained before, for the DC-based profiling the threshold is $F_{1,nm-p;\alpha}$ whereas the GLS-based profiling requires a threshold at $\chi^2_{1;\alpha}$ or $F_{1,nm-p;\alpha}$ depending on whether or not the noise covariance matrix is known. However, based on the discussions presented in Chapters 3 and 4, I recommend using the GLS-based profiling especially for small datasets to avoid the potential problems related to the DC method such as local results, highly co-dependent parameter estimates and/or numerical issues.

The proposed profile-based parameter estimability analysis in this thesis is defined in the context of existing datasets, i.e., the response values are required for calculating the profile likelihood function and the corresponding parameter inference intervals. However, if it is preferable to do the estimability analysis before running the actual experiments, this profile-based analysis can be carried out on the planned experimental design. This is accomplished by simulating a dataset from the planned experimental design with nominal parameter estimates with added random noise. The profile-based estimability analysis is feasible based on the synthetic dataset including the generated responses.

One of the main disadvantages of not having actual observations and performing the estimability analysis on the intended experiments is that the analysis is done with nominal parameter values. This might lead to local results or might involve exploring an interval that is far from the least squares parameter estimates. Moreover, the simulation approach involves assumptions about the structure of the added noise which can impact the profiling results significantly as explained in detail in Chapter 3 section 3.2. However, it should be noted that the sensitivity-based estimability methods also suffer from the same disadvantages if they are to be done on planned experiments without actual observations.

One of the main disadvantages of the sensitivity-based methods is that they are based on linearization and thus provide results that are potentially local. While the profile-based analysis studies a wide range of parameter values. Therefore, combining the two methods using the proposed profile-based exploratory parameter estimability analysis enables the researcher to make informed decisions that are more likely to be global. This is one of the most important advantages of this method over the common sensitivity-based methods. The schematic plots of Figure 5-3 can assist in highlighting the difference between these two types of analysis and the issues related to having local results.

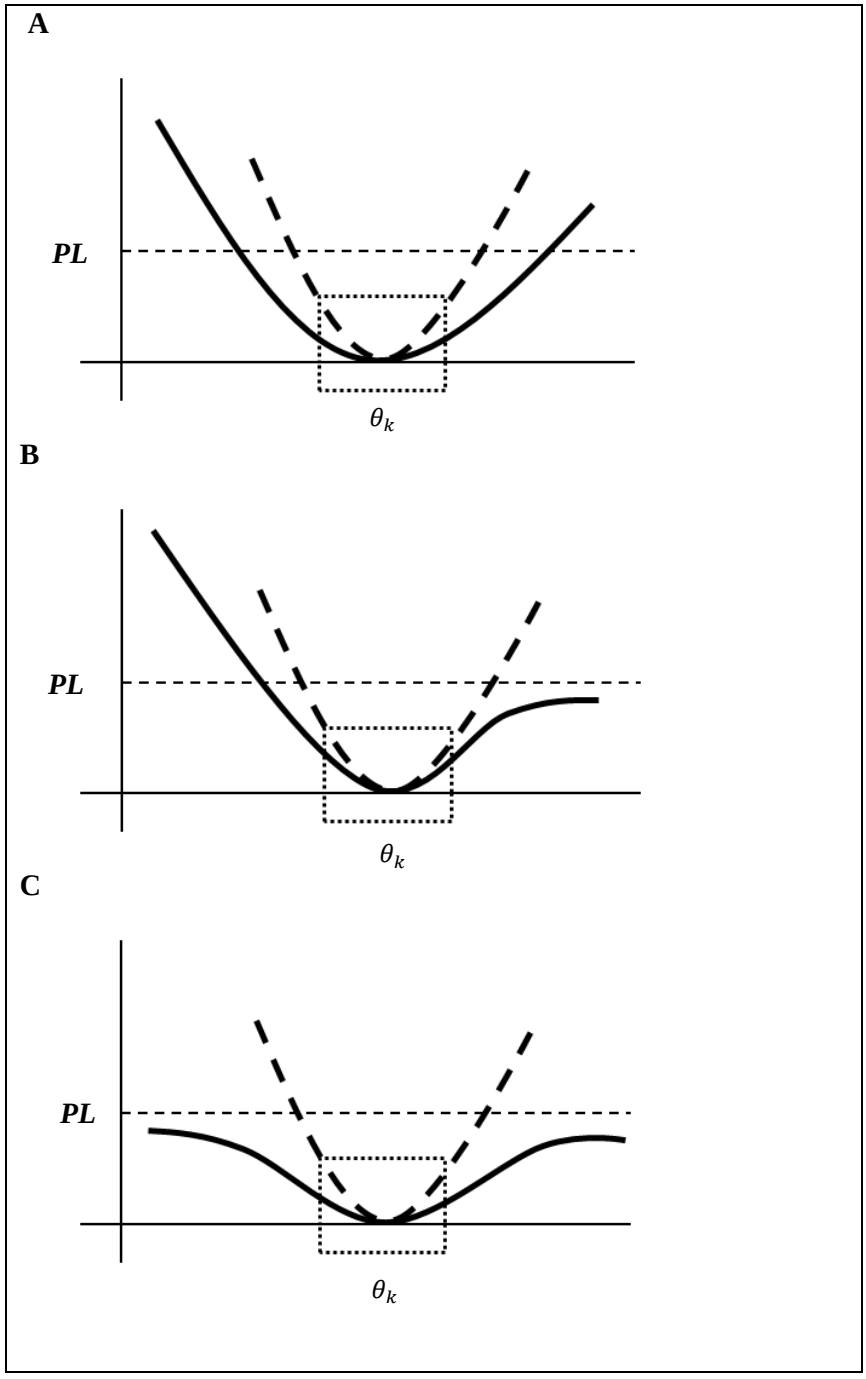

Figure 5-3: Schematic comparison of the profile-based and sensitivity-based estimability analyses. Solid lines: profile likelihood curves, dashed lines: linearization-based curves for sensitivity-based estimability analysis, dotted rectangles: parameter interval for the sensitivity-based analysis. Plot A: Both methods are advantageous. Plot B: The profile-based method produces adequate results while the sensitivity-based method provides local results. Plot C: The sensitivity-based analysis fails to depict the estimability issues.

Each plot in Figure 5-3 contains the profile-based and linearization-based plots for an arbitrary parameter that is being analyzed for estimability. The proposed estimability analysis is performed on the profile-based curves (solid lines), while the common sensitivity-based analysis are carried out on the linearization-based curves (dashed lines). The profile-based method considers the whole range of parameter values shown on the plots, whereas the sensitivity-based method only examines a small range of parameter values around the least squares parameter estimate which is shown by a dashed rectangle.

Figure 5-3 plot A depicts a parameter that is estimable. It can be seen that both methods suggest the estimability of the parameter and are equally successful. Plot B illustrates a parameter that is not estimable as the parameter value becomes large. The profile-based method is able to show the inestimability of the parameter since the profile likelihood plot does not exceed the threshold on the right-hand side. On the other hand, the sensitivity-based method fails to discover the interval where the parameter is inestimable and suggests the local estimability of the parameter. The same issue is presented schematically in Plot C, where the parameter is only estimable in a small neighborhood around its least square estimate. Focusing on the linearization-based plot and limited to a small interval, the sensitivity-based method once again overlooks the possibility of inestimability for this parameter.

Besides determining estimable parameters, the proposed analysis can be beneficial for resolving the inestimability issues as well. As mentioned before, probably the most effective solution for parameter inestimability is to collect additional experimental data when possible (Cobelli and DiStefano, 1980). The profile likelihood plots that are explored in the second step of the analysis show the intervals where the estimations run

into inestimability issues and more experimental runs and/or observations are required. Furthermore, based on the literature review another common way to overcome parameter inestimability issues is fixing some or all of the inestimable parameters and thus leaving them out of the estimations (McLean and McAuley, 2012). This can be achieved in the profile-based analysis by a similar method as the one used in the last step. Each inestimable parameter is chosen as the profiling parameter and fixed at certain values. The profile traces of the remaining parameters versus the fixed parameter can be studied to realize if fixing the parameter has solved the inestimability of the others. More generally, an iterative method can be used whereby one or more inestimable parameters are selected and fixed at certain values and the profile traces of the remaining parameters versus the fixed ones are investigated for signs of inestimability. The iterations continue until the largest set of estimable parameters for the model is determined. Again, all the required parameter values are already calculated in the previous steps and this analysis can be accomplished without imposing extra calculations.

For complex models with a large number of parameters or for cases where the profile plots suggest high co-dependency among the parameters, it is wise to use a sensitivity-based method such as the method suggested by Yao et al. (2003) as a preliminary analysis to narrow down the number of suspected inestimable parameters. If the results of this analysis indicate many inestimable parameters for which the profiling might not be feasible, then the common methods in the literature can be used to reduce the parameters to a smaller set e.g., by fixing some parameters at particular values. The potentially estimable parameters might still only be locally estimable and therefore their estimability should be further explored by the profile-based method to arrive at final conclusions. By

doing so, the analysis moves away from the potentially local results associated with the sensitivity-based method and considers the bigger picture. Furthermore, the results will be more reliable since they have been validated by a combination of methods.

In the following section, application of the proposed estimability approach is demonstrated by means of a case study. The chosen case study consists of a bacterial growth model and dataset which has been explored previously by researchers and proven to contain inestimable parameters.

5.6 Case study 4: Holmberg Example

In this section, in order to demonstrate the application of profiling in parameter estimability analysis a fermentation process and the corresponding Michaelis-Menten equation for describing the bacterial growth behavior is studied. The reason for choosing a fermentation process is that such processes typically result in small datasets with noisy measurements and therefore the corresponding models are prone to having inestimable parameters (Holmberg, 1982). The Michaelis-Menten equation is the most widely used model for describing microbial growth in many kinds of applications such as fermentation processes and is also well-known for having potentially highly correlated parameters in biological applications (Dochain et al., 1995). For bacterial growth this equation is defined by (Monod, 1949):

$$\mu(s) = \frac{\mu_m s}{K_s + s} \tag{5-18}$$

where $\mu(s)$ is the specific growth rate function, μ_m is the maximum specific growth rate (hr^{-1}), s represents the substrate concentration and K_s is the Michaelis-Menten constant (g/l). The properties of the $\mu(s)$ are $\mu(0) = 0$ and also $\mu(s) \to \mu_m$ when $s \to \infty$.

The dataset for fermentation of *Bacillus Thuringiensis* in a batch reactor is extracted from a research work by Holmberg et al. (1980) and presented in Table 5-2. In this table, x and s represent the biomass concentration (g/l) and substrate concentration (g/l) respectively and t represents time (hours). For the intended purposes of this thesis, time was treated as the explanatory variable, while x and s were considered as measured responses. The following model was used in their paper for parameter estimation for this example (Holmberg et al., 1980):

$$\frac{dx}{dt} = \mu(s)x - K_d x \tag{5-19}$$

$$\frac{ds}{dt} = -\frac{1}{Y_{x/s}}\mu(s)x \tag{5-20}$$

with initial conditions being $x_0 = 1$ g/l and $s_0 = 18.75$ g/l. In these equations $Y_{x/s}$ is the yield coefficient and K_d stands for the decay rate coefficient (hr^{-1}). Thus, the parameter vector consists of:

$$\boldsymbol{\theta} = \left[\mu_m \; K_s \; Y_{x/s} \; K_d\right]' \tag{5-21}$$

Parameters μ_m and K_s have biological significance and are often treated as characteristics of the process.

Later on Holmberg (1982) studied the identifiability of this model and generated synthetic data based on the model to investigate parameter estimability using a sensitivity-based graphical analysis. The outcomes showed that for that particular system and dataset, all of the parameters of the batch growth model are identifiable as long as the initial concentrations of biomass (x_0) and the substrate (s_0) are both greater than zero. However, determination of parameters is still difficult due to some estimability issues around μ_m and K_s (Holmberg, 1982).

Table 5-2: Dataset corresponding to Case Study 4, Holmberg example (Holmberg et al., 1980)

t(hours)	x(g/l)	s(g/l)
0	1	18.75
2	1.5625	15.625
4	4.375	14.375
6	6.25	8.75
8	10	3.75
11	13.75	0.5
14	18.75	0.05
17	20.625	0.01

Knowing that identifiability is not an issue for this model, and that there is a significant possibility of having some inestimable parameters in this system, this example was chosen to test the application of the proposed profile-based estimability analysis in this thesis.

Holmberg (1982) showed that the initial concentration of the substrate (s_0) and the value of $\frac{K_s}{s_0}$ are important factors in parameter estimability analysis for this model. Based on their findings and considering equation 5-18, three scenarios can be considered for the estimability analysis (Holmberg, 1982):

Case 1: If $\frac{K_s}{s_0}$ is less than a certain threshold (i.e., $s_0 \gg K_s$), the specific growth rate $\mu(s)$ is dominated by s_0 and remains close to its upper limit $\left(\mu'_m = \frac{\mu_m s_0}{K_s + s_0}\right)$. As a result, $\mu(s)$ is

not heavily impacted by changes in the values of s and the following zero-order model approximation can be used:

$$\mu(s) = \begin{cases} \mu'_m; & if\ s > 0 \\ 0; & if s = 0 \end{cases} \quad (5\text{-}22)$$

In practice the threshold has been considered around 0.1 in their paper.

Case 2: If $\frac{K_S}{s_0}$ is above a certain threshold (i.e., $s_0 < K_s$), the lower and upper limits for $\mu(s)$ will be very close based on equation 5-18 and thus the growth model can be treated as a linear model described by:

$$\mu(s) = \alpha_m s \quad (5\text{-}23)$$

According to their research, $\frac{K_S}{s_0}$ values of at least 4 fall into this category.

Case 3: This case includes all the ratios of $\frac{K_S}{s_0}$ that are in between the previous two cases. In this situation the model is nonlinear throughout the experiment and the corresponding parameters might be inestimable.

They emphasized the influence of sampling frequency and noise on these thresholds and consequently the accuracy of the parameter estimates. According to their conclusions in cases with few and noisy data (which is often the case), these thresholds are very difficult to determine and μ_m and K_s cannot be reliably estimated.

In the present research, the initial parameter guesses are chosen such that they fall into Case 3 category which is supposed to lead to inestimable parameters. However, the following profiling step covers a large range of values for parameters. For parameter estimation, the system of ODEs is solved using a $4^{th}/5^{th}$ order Runge-Kutta approach using a variable step size, as implemented in the standard ode45 library routine in Matlab™. One of the challenges for this example is that the estimation problem is very

sensitive to any small changes in the data points as well as the initial parameter guesses and the Hartley's modified Gauss-Newton increments. This has also been mentioned previously by the other researchers (e.g., (Holmberg, 1982); (Dochain et al., 1995)).

In order to preform estimability analysis, the parameters are estimated using the GLS method and the method explained in Chapter 3 is used to estimate the noise covariance matrix from the residuals. Subsequently, GLS-based profiling is carried out with each parameter as the profiling parameter. The 95% parameter likelihood intervals are built based on the profile likelihood plots where the curve crosses the reference line located at $F_{1,nm-p;0.95}$ (For more details please refer to GLS profiling with unknown noise covariance matrix method as explained in Chapter 3 section 3.4). The results show that all parameters have finite likelihood intervals except for K_s for which the GLS-based profile likelihood plot becomes nearly flat and does not cross the threshold as values of K_s increase. This plot is presented in Figure 5-4. In this plot another reference line at $F_{1,n-p;0.95}$ has been added for the sake of comparison.

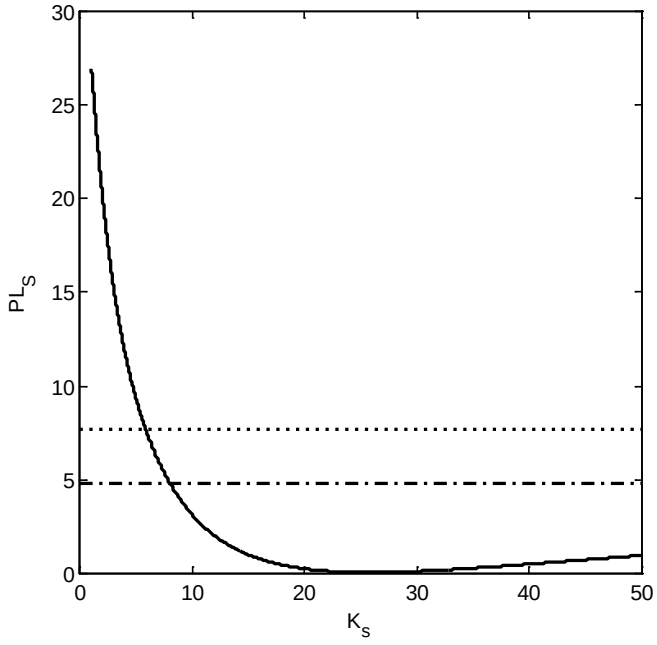

Figure 5-4: Profile likelihood plot of PL_S function versus K_s for case study 4 using the GLS approach with reference lines at $F_{1,n-p;0.95}$ (dotted line) and $F_{1,nm-p;0.95}$ (dot-dashed line).

It can be seen in Figure 5-4 that the right-hand side of the profile likelihood plot becomes relatively flat and does not exceed the threshold. Thus, even though there is a lower limit for the confidence interval for K_s, the unlimited confidence interval on the right-hand side suggests that this parameter is inestimable. Since there is some dispute about the distribution of the profile likelihood plot (as discussed in detail in Chapter 2 section 2.7) a second reference line has been added to the figure at $F_{1,n-p;0.95}$. The estimability results are still valid for this reference line since the profile curve does not cross this line on the right-hand side either.

Profile traces of the potentially estimable parameters are carefully examined to find regions of high co-dependency between the parameters that might lead to inestimability

issues. In these plots, all parameters are studentized based on the following equation to avoid any misjudgments due to scaling issues:

$$\delta(\theta_k) = (\theta_k - \hat{\theta}_k)/se(\hat{\theta}_k) \quad (5\text{-}24)$$

where $se(\hat{\theta}_k)$ is the estimated standard error of the parameter estimate $se(\hat{\theta}_k) = \sqrt{\{(\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{V}(\boldsymbol{\theta}))^{-1}\}_{kk}s^2}$. The results from this step do not suggest any major co-dependencies between the parameter estimates, but do suggest co-dependencies between estimates of μ_m and $Y_{x/s}$. Some of these profile traces are presented in Figure 5-5.

For this example, some of the profile traces of the inestimable parameter K_s are included in Figure 5-5. The main reason is to draw attention to the fact that the profile traces for this parameter barely change (almost horizontal or vertical lines in plots A and C respectively) and show an example of the issue that was discussed at the end of section 5.4. This result indicates that when using profile traces for estimability analysis, it will be important to check whether the estimated parameter values change from initial conditions. In this particular example, setting the value of K_s to the overall SSE-optimal value did not result in a changing profile trace. Consequently, one might in such cases simply set the initial condition at this value for the purposes of computing traces.

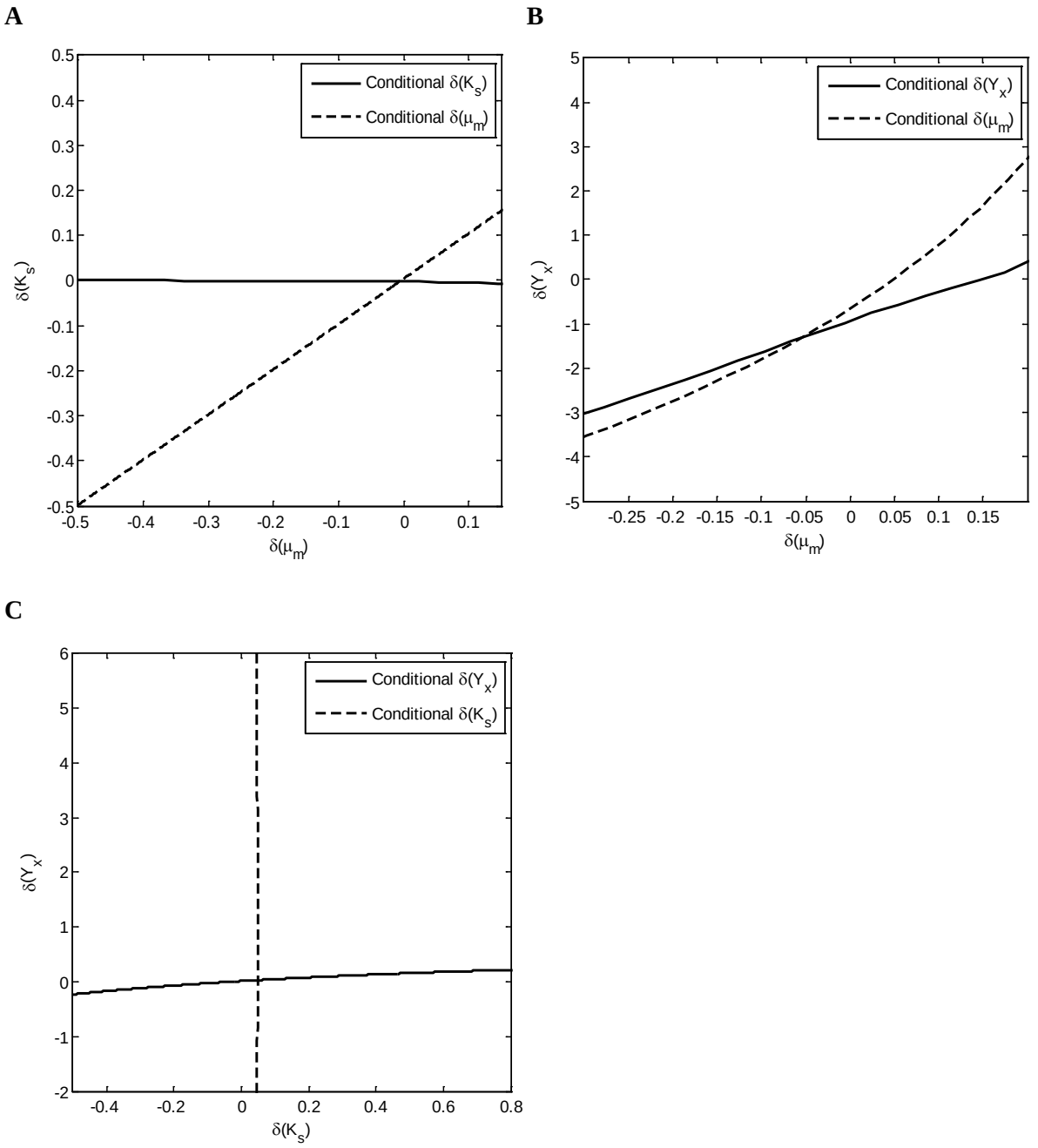

Figure 5-5: Profile traces for case study 4 using the GLS approach. Plot A: pairwise profile traces of $\delta(K_s)$ and $\delta(\mu_m)$ versus each other. Plot B: pairwise profile traces of $\delta(Y_{x/s})$ and $\delta(\mu_m)$ versus each other. Plot C: pirwise profile traces of $\delta(K_s)$ and $\delta(Y_{x/s})$ versus each other.

Since K_s is inestimable, in order to further examine the behavior of parameters in relevance to the inestimable parameter the overlay plot of all parameter estimates conditional on the value of K_s is studied and presented in Figure 5-6. This figure shows no visible trends among the parameter estimates conditional on K_s which supports the estimability of the parameter estimates.

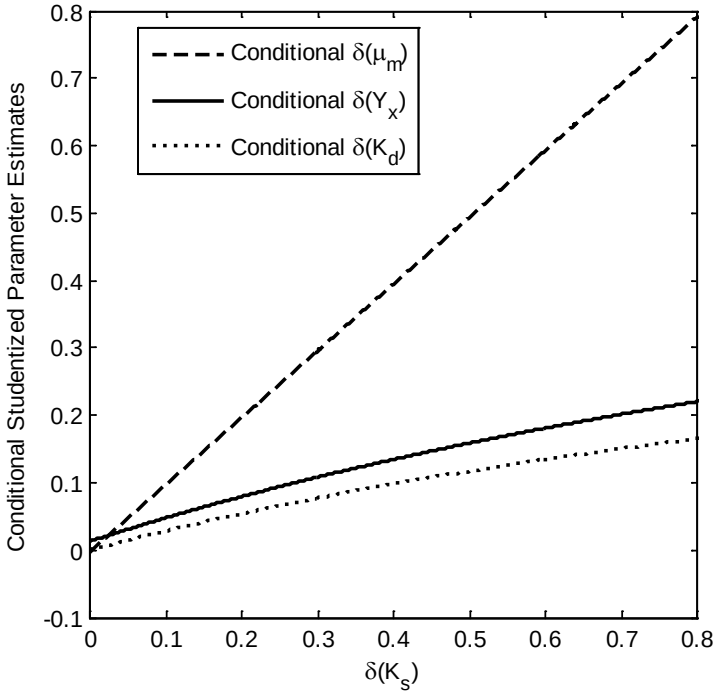

Figure 5-6: Profile traces of all studentized parameters conditional on K_s versus the studentized parameter $\delta(K_s)$. The dashed line corresponds to $\delta(\mu_m)$, solid line represents $\delta(Y_{x/s})$ and dotted line denotes $\delta(K_d)$ respectively.

Based on this profile-based estimability analysis for this example, the parameter K_s can only be estimated from this dataset when the values of K_s are very small compared to the initial substrate concentration ($s_0 \gg K_s$). Otherwise for larger values of K_s the likelihood

interval becomes unlimited on the right-hand side and thus the parameter becomes inestimable. A preliminary suggestion to resolve this issue would be to collect more data at smaller time intervals or start with higher concentrations of substrate to keep the $\frac{K_s}{s_0}$ as small as possible.

5.7 Discussions and Conclusions

In the previous chapters it was shown how profiling can be used to build parameter likelihood intervals. In this chapter the focus is on extending the profiling application to estimability analysis. This idea had been initially presented by Raue et al. (2009) who investigated kinetic biological models using GLS-based profiling with the assumption of known noise variances (Raue et al., 2009).

Based on the literature review that was provided in this chapter, estimability analysis is a necessary step in modelling since insufficient or inaccurate experimental data or highly co-dependent parameter estimates can cause inestimability issues that prohibit adequate estimation of the parameters. In such cases the model predictions are insensitive to some parameter values, or the effect of one parameter on the model is not independent from the effect of one or more other parameters. The background review revealed that a significant amount of work has been done on estimability analysis. The current estimability methods often revolve around investigating the sensitivity and covariance matrices or the Fisher Information Matrix. However, the results of these methods are not always reliable due to issues such as high dependency on the initial guesses for parameters, providing local results, computational costs, and determining the cut-off values to distinguish between estimable and inestimable parameters (McLean and McAuley, 2012).

It was shown in this thesis that profiling could combine the benefits of reliable confidence intervals with easy-to-interpret graphical results for estimability analysis. Unlimited likelihood intervals for parameters that can be built from the profile likelihood plots are a reasonably dependable sign of inestimability of parameters. In the proposed profile-based estimability analysis, the noise covariance matrix can be estimated from the residual matrix and this estimation process is taken into account for coming up with the distribution of the profile likelihood function. This enables the modeler to make reliable parameter likelihood intervals and also use appropriate reference line as the threshold for estimability analysis.

A four-step profile-based exploratory parameter estimability analysis is developed in this research that involves examining the profile likelihood plots and profile traces for the model parameters that are deemed potentially estimable by the sensitivity-based analysis. This method investigates the profile likelihood plots for all parameters and selects a set of parameters that are very likely to be estimable. The subsequent steps (if applicable) further investigate these parameters to confirm the results. The outcomes of all steps provide a firm basis for conclusions about the estimability of the parameters.

Based on this methodology, the GLS-based likelihood plots with a reference line at $\chi^2_{1;\alpha}$ for models with known noise covariance matrix and at $F_{1,nm-p;\alpha}$ for models with unknown noise covariance matrix can be used to determine inestimable parameters. For reasonable accuracy, $\alpha = 0.95$ is an appropriate choice that was used in the case study. The proposed method was successfully applied to a fermentation process and the corresponding Michaelis-Menten equation for describing the bacterial growth behavior. The profile traces, some of which were presented in Figure 5-5, did not show a high level

of co-dependency between parameters. The profile likelihood plots showed inestimability of K_s and the potential need for more experimental runs and/or observations in order to obtain more reliable parameter estimates.

5.8 Summary

In this chapter the parameter estimability analysis and its importance were explained. Current methods for this analysis such as the sensitivity-based methods, visual and graphical techniques and the repeated estimation methods were pointed out. It was shown that profiling can be applied to parameter estimability by studying the likelihood intervals for parameters and drawing conclusions based on the outcomes. A profile-based exploratory parameter estimability analysis method was proposed which consists of examining profile plots and profile traces of the parameters and studying their likelihood intervals. The advantages of using this method such as having a well-defined cut-off value for the profile likelihood function, the capacity to explore a wide range of parameter values and avoiding local results were explained. Furthermore, the application of this profile-based method was showcased by means of an example. The results were in agreement with the estimability results reported by previous researches for this example and confirmed the reliability of the proposed approach in practice.

The main contributions of this chapter are as follows:

- Proposing a four-step profile-based exploratory parameter estimability analysis which involves combining a sensitivity-based estimability method and the profiling method in order to draw reliable conclusions about parameter estimability.

- Defining the thresholds for the value of the profile likelihood function based on the estimation process of the noise covariance matrix to determine the estimability of parameters.
- Linking the estimability analysis to parameter inference and thus reducing the computational costs for the analysis, recommending the application of the GLS-based likelihood plots with a reference line at $\chi^2_{1;\alpha}$ for models with known noise covariance matrix and at $F_{1,nm-p;\alpha}$ for models with unknown noise covariance matrix to determine the estimability of the parameters.
- Applying the profile-based method for resolving inestimability issues by pointing out the regions that require more experimental data, and also by fixing one or more parameters at certain values during profiling and observing the impact of this process on resolving the inestimability of other parameters.
- Improving the computational efficiency of the estimability stage in modeling by using the results from the parameter inference phase as well as using the parameter estimates and conditional parameter estimates generated in the first and second steps of the estimability analysis for the subsequent steps and analysing the results from different angles to extract as much information as possible.

Chapter 6

Contributions and Recommendations

6.1 Contributions

This research explored statistical model-building by means of the profiling method, with the objective of addressing some of the issues regarding inference in nonlinear multi-response models and providing a more comprehensive set of diagnostics for building and assessing such models. I sought ways to extend the application of profiling to nonlinear multi-response models and investigate its application particularly in parameter inference and estimability analysis.

The hope is that this thesis contributes to bridging the gap between statistical methods and engineering techniques by making statistical modeling more accessible and easier to interpret for chemical processes. Overall, the contributions of this thesis can be summarized as follows:

- Proposing a GLS-based profiling method for nonlinear multi-response models with unknown noise covariance matrix (referred to as *GLS-based profiling approach #2*) and defining the GLS-based profile likelihood function as $PL_S(\theta_k) = \tilde{S}(\theta_k) - S(\widehat{\boldsymbol{\theta}})$
- Highlighting the importance of the estimation of noise covariance matrix and the assumptions involved and proposing a method for estimating and updating the noise covariance matrix during parameter estimation and profiling for the proposed approach
- Identifying the distribution and the corresponding degrees of freedom for the PL_S function, taking into account the estimation process of the noise covariance matrix

and suggesting χ_1^2 and $F_{1,nm-p}$ distributions for cases with known or unknown noise covariance matrix respectively

- Exploring and demonstrating the application of the proposed GLS-based profiling method in parameter inference for multi-response models for several standard case studies
- Studying the DC-based profiling approach for parameter inference in nonlinear multi-response models and introducing the DC-based profile likelihood function as $PL_D(\theta_k) = \frac{\tilde{D}(\theta_k) - D(\hat{\boldsymbol{\theta}})}{(s')^2}$ based on the profile *t* function proposed by Soo and Bates (1996).
- Determining the distribution of the DC-based profile likelihood function as $PL_D(\theta_k) \sim F_{1,nm-p}$.
- Comparison of the *DC-based profiling approach* with the proposed *GLS-based profiling approach #2* with the following conclusions:
    - Generally a combination of DC for parameter estimation and *GLS profiling approach #2* for profiling leads to reliable parameter inference results with a relatively straightforward process.
    - For small datasets the *GLS-based profiling approach #2* is a more appropriate choice and DC should only be used for parameter estimation and profiling if numerical issues can be avoided.
- Proposing a four-step profile-based exploratory parameter estimability analysis which involves inspecting profile plots and profile traces of the parameters and their likelihood intervals.

- Defining the thresholds for the value of the profile likelihood function based on the estimation process of the noise covariance matrix to determine the estimability of parameters.
- Linking the estimability analysis to parameter inference and thus reducing the computational costs for the analysis, recommending the application of the GLS-based likelihood plots with a reference line at $\chi^2_{1;\alpha}$ for models with known noise covariance matrix and at $F_{1,nm-p;\alpha}$ for models with unknown noise covariance matrix to determine the estimability of the parameters.
- Applying the profile-based method for resolving inestimability issues by pointing out the regions that require more experimental data, and also by fixing one or more parameters at certain values during profiling and observing the impact of it on resolving the inestimability of other parameters.

6.2 Suggestions for Future Work

Carrying out this research, a number of topics related to profiling were recognized that require further investigation. As mentioned before, statistical profiling has many applications such as parameter inference, estimability analysis, model transformations and re-parameterizations, sensitivity analysis, and experimental design. This research mainly focused on the first two applications (parameter inference and estimability analysis). However, profiling techniques have the potential for the remaining applications in the field of nonlinear multi-response estimation.

For *model transformations and re-parameterizations*, the profile traces provide graphical information about co-dependencies between the parameter estimates over a range of

values. These plots, and the data contained within them, can be used to suggest model re-parameterizations and model transformations. For example, if two profile traces are almost entirely coincident, this indicates the nature of the co-dependency. If the traces are collinear (i.e., linear and coincident), then a linear transformation or re-parameterization might be in order. Alternatively, if the traces are nonlinear and coincident, this suggests that the two estimates are nonlinearly related, and an empirical nonlinear function (e.g., a polynomial or a B-spline) can be used to express one parameter in terms of the other.

Profiling can also be used for *sensitivity analysis*, and indeed Sulieman et al. (2001, 2004) have defined the profile-based sensitivity coefficient (PSC) for both single- and multi-response cases. For the multi-response case, the PSC is defined in terms of the Determinant Criterion, and it would be useful to define the PSC instead using a GLS estimation approach. One difference between a GLS-based PSC and a DC-based PSC will be the treatment of the covariance matrix. The DC-based PSC implicitly takes into account the covariance matrix estimation, as shown in this thesis, whereas in the GLS case, a decision has to be made about whether the covariance matrix is estimated or held fixed. A PSC could also be defined to describe the impact of covariance estimates on predicted response values, which will provide a quantitative basis for understanding and characterizing the extent to which covariance assumptions influence model predictions.

Other researchers (e.g., Raue and colleagues) have suggested and indicated that profiling can be used for *experimental design*. This is also discussed briefly earlier in the thesis. In the work of Raue, experimental design is used to refer to additional responses to be measured, as well as additional experimental runs. However the issue is still not fully resolved for more complex models.

The challenge with profiling is that the profiles describe the likelihood function behavior and the behavior of other parameter estimates as a function of the profiling parameter, and not the run conditions. Incorporating experimental design into profiling could follow a number of different approaches. For example, suppose five additional runs are to be added. These run conditions could be introduced (e.g., as part of an optimization), and the profiles could be generated to see whether the new runs improve the estimability. A "brute force" approach would be introducing runs, and then generating the entire profile analysis again, however it may be possible to examine derivatives of predictions and likelihood functions with respect to run conditions and parameters to see if new measures – not requiring complete computation of the profiling – are possible. An optimization approach for sequential experimental design for a nonlinear multi-response estimation problem was presented previously for predicting molecular-weight distributions in polymer reaction engineering (Thompson et al., 2010). Sensitivity information is used implicitly because the experimental design criteria are generally expressed in terms of the sensitivity matrix at the optimal parameter estimates (e.g., to minimize the volume of the approximate joint confidence region for the parameter estimates). It may be possible to build on profiling, using intermediate information from the existing profiling analysis.

An alternative approach would be to use the "prediction parameter transformation approach" used by Sulieman et al. (2001, 2004), in which the model is re-parameterized so that one or more parameters corresponds to a prediction at a particular run condition.

Other studies have suggested investigating the model trajectories versus time for fixed values of each inestimable parameter to identify the regions where the uncertainty of the parameter has the most impact on the model trajectories and carry out more experiments

in those regions (Raue et al., 2011). Yet this is another area that has the potential to be further explored.

A further area of work is in improving techniques for jointly estimating covariance matrices and parameters, especially in a GLS approach. One of the challenges encountered in this research was the problem of joint profiling over parameter estimates and noise variance/covariance terms which led to very poor numerical conditioning of the problem and highly co-dependent parameter estimates. This impacted the *GLS-based profiling approach #3* as well as the *DC-based profiling approach*. I recommend exploring this estimation further to overcome the challenges of joint profiling over covariances and parameters, and improving the profiling techniques.

This research offered a profile-based method for parameter estimability analysis and developed this application by using combinations of profiling and a sensitivity-based method. There is further work that can be done to expand the role of profiling in estimability analysis. For example, it might be possible to use generalized profiling (as defined by Chen and Jennrich, and extended to multi-response problems) as a way of analyzing estimability problems associated with predictions at particular points or regions. There is significant potential here to develop techniques that can be more targeted to predictions of particular interest.

Finally, it is also recommended to investigate the application of profiling to large-scale problems for parameter inference and estimability analysis and find ways to generate profiles efficiently. For example, spline smoothing could be applied to profiles, assisting in extracting the key features of the profile plots. The ability to apply profiling to large-

scale problems will require efficient computational techniques and automated techniques for interpreting the results.

Bibliography

Asprey, S.P., Macchietto, S., 2000. Statistical Tools for Optimal Dynamic Model Building. Comput. Chem. Eng. 24, 1261–1267.

Asuero, A.G., Bueno, J.M., 2011. Fitting Straight Lines with Replicated Observations by Linear Regression. IV. Transforming Data. Crit. Rev. Anal. Chem. 41, 36–69. doi:10.1080/10408347.2010.523589

Bard, Y., 1974. Nonlinear Parameter estimation. Academic Press.

Barndorff-Nielsen, O.E., 1986. Inference on Full or Partial Parameters Based on the Standardized Signed Log Likelihood Ratio. Biometrika 73, 307. doi:10.2307/2336207

Bates, D.M., Watts, D.G., 1980. Relative Curvature Measures of Nonlinearity. J. R. Stat. Soc. Ser. B 42, 1–25.

Bates, D.M., Watts, D.G., 1984. A Multi-Response Gauss-Newton Algorithm. Commun. Stat. - Simul. Comput. 13, 705–715. doi:10.1080/03610918408812407

Bates, D.M., Watts, D.G., 1985. Multiresponse Estimation With Special Application to to Linear Systems of Differential Equations. Technometrics 27, 329–339.

Bates, D.M., Watts, D.G., 1987. A Generalized Gauss-Newton Procedure for Multi-Response Parameter Estimation. SIAM J. Sci. Stat. Comput. 8, 49–55.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and Its Applications. John Wiley and Sons Inc.

Bates, D.M., Watts, D.G., 1991. Model Building in Chemistry Using Profile t and Trace Plots. Chemom. Intell. Lab. Syst. 10, 107–116. doi:10.1016/0169-7439(91)80039-S

Beale, E.M.L., 1960. Confidence Regions in Non-Linear Estimation. J. R. Stat. Soc. Ser. B 22, 41–88.

Beauchamp, J.J., Cornell, R.G., 1966. Simultaneous Nonlinear Estimation. Technometrics 8, 319–326.

Bellman, R., Astrom, K.J., 1970. On Structural Identifiability. Math. Biosci. 7, 329–339.

Benabbas, L., Asprey, S.P., Macchietto, S., 2005. Curvature-Based Methods for Designing Optimally Informative Experiments in Multiresponse Nonlinear Dynamic Situations. Ind. Eng. Chem. Res. 44, 7120–7131. doi:10.1021/ie040096w

Benyahia, B., Latifi, M.A., Fonteix, C., Pla, F., 2013. Emulsion Copolymerization of Styrene and Butyl Acrylate in the Presence of a Chain Transfer Agent. Part 2: Parameters Estimability and Confidence Regions. Chem. Eng. Sci. 90, 110–118. doi:10.1016/j.ces.2012.12.013

Ben-Zvi, A., 2004. Identifiability of Differential Algebraic Equation Systems. Queen's University.

Ben-Zvi, A., McLellan, P.J., McAuley, K.B., 2004. Identifiability of Linear Time-Invariant Differential-Algebraic Systems. 2. The Differential-Algebraic Approach. Ind. Eng. Chem. Res. 43, 1251–1259. doi:10.1021/ie030534j

Biegler, L.T., Damiano, J.J., Blau, G.E., 1986. Nonlinear Parameter Estimation: A Case Study Comparison. AIChE J. 32, 29–45. doi:10.1002/aic.690320105

Bock, H.G., 1983. Recent Advances in Parameter Identification Techniques for ODE, in: Deuflhard, P., Hairer, E. (Eds.), Numerical Treatment of Inverse Problems in Differential and Integral Equations. Birkhauser, Boston, pp. 95–121.

Box, G.E.P., Draper, N.R., 1965. The Bayesian Estimation of Common Parameters from Several Responses. Biometrika 52, 355–365.

Box, G.E.P., Hunter, W.G., MacGregor, J.F., Erjavec, J., 1973. Some Problems Associated with the Analysis of Multiresponse Data. Technometrics 15, 33–51.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Wiley Clasics Library.

Breiman, L., Friedman, J.H., 1997. Predicting Multivariate Responses in Multiple Linear Regression. J. R. Stat. Soc. Ser. B (Statistical Methodol. 59, 3–54. doi:10.1111/1467-9868.00054

Brooks, I., Watts, D.G., Soneson, K.K., Hensley, P., 1994. Determining Confidence Intervals for Parameters Derived from Analysis of Equilibrium Analytical Ultracentrifugation Data. Methods Enzymol. 240, 459–478.

Buzzi Ferraris, G., Forzatti, P., Emig, G., Hofmann, H., 1984. Sequential Experimental Design for Model Discrimination in the Case of Multiple Responses. Chem. Eng. Sci. 39, 81–85.

Campbell, D.A., 2007. Bayesian Collocation Tempering and Generalized Profiling for Estimation of Parameters from Differential Equation Models. McGill.

Campbell, D.A., Hooker, G., McAuley, K.B., 2012. Parameter Estimation in Differential Equation Models with Constrained States. J. Chemom. 26, 322–332. doi:10.1002/cem.2416

Chappell, M.J., Godfrey, K.R., Vajda, S., 1990. Global Identifiability of the Parameters of Nonlinear Systems with Specified Inputs: A Comparison of Methods. Math. Biosci. 102, 41–73.

Chen, J.S., Jennrich, R.I., 1996. The Signed Root Deviance Profile and Confidence Intervals in Maximum Likelihood Analysis. J. Am. Stat. Assoc. 91, 993. doi:10.2307/2291718

Cobelli, C., DiStefano, J.J., 1980. Parameter and Structural Identifiability Concepts and Ambiguities: A Critical Review and Analysis. Am. J. Physiol. 239, R7–24.

Dochain, D., Vanrolleghem, P.A., van Daele, M., 1995. Structural Identifiability of Biokinetic Models of Activated Sludge Respiration. Water Res. 29, 2571–2578. doi:10.1016/0043-1354(95)00106-U

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. John Wiley and Sons Inc.

Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. The Society for Industrial and Applied Mathematics, Philadelphia.

Fraser, D.A.S., 1993. Directional Tests and Statistical Frames. Stat. Pap. 34, 213–236.

Galambos, J.T., Cornell, R.G., 1962. Mathematical Models for the Study of the Metabolic Pattern of Sulfate. J. Lab. Clin. Med. 60, 53–63.

Gallant, A.R., 1975. Seemingly Unrelated Nonlinear Regressions. J. Econom. 3, 35–50.

Gallant, R., Holly, A., 1980. Statistical Inference in an Implicit , Nonlinear , Simultaneous Equation Mode in the Context of Maximum Likelihood Estimation. Econometrica 48, 697–720.

Gentle, J.E., 2007. Matrix Algebra: Theory, Computations, and Applications in Statistics. Springer, New York.

Gerhard, D., 2010. Simultaneous Small Sample Inference Based on Profile Likelihood. Hannover.

Guay, M., 1995. Curvature Measures for Multiresponse Regression Models. Biometrika 82, 411. doi:10.2307/2337418

Guay, M., McLean, D.D., 1995. Optimization and Sensitivity Analysis for Multiresponse Parameter Estimation in Systems of Ordinary Differential Equations. Comput. Chem. Eng. 19, 1271–1285.

Holmberg, A., 1982. On the Practical Identifiability of Microbial Growth Models Incorporating Michaelis-Menten Type Nonlinearities. Math. Biosci. 62, 23–43.

Holmberg, A., Sievanen, R., Carlberg, G., 1980. Fermentation of Bacillus thuringiensis for exotoxin Production: Process Analysis Study. Biotechnol. Bioeng. 22, 1707–1724.

Jacquez, J.A., Grief, P., 1985. Numerical Parameter Identifiability and Estimability: Integrating Identifiability, Estimability, and Optimal Sampling Design. Math. Biosci. 77, 201–227.

Jayasankar, B.R., Ben-Zvi, A., Huang, B., 2009. Identifiability and estimability study for a dynamic solid oxide fuel cell model. Comput. Chem. Eng. 33, 484–492. doi:10.1016/j.compchemeng.2008.11.005

Jitjareonchai, J.J., Reilly, P.M., Duever, T.A., Chambers, D.B., 2006. Parameter Estimation in the Error-in-Variables Models Using the Gibbs Sampler 84.

Johnson, R.A., Wichern, D.W., 2007. Applied Multivariate Statistical Analysis, 6th ed. Pearson, New York.

Kang, G., Bates, D.M., 1990. Approximate Inferences in Multiresponse Regression Analysis. Biometrika 77, 321. doi:10.2307/2336810

Karimi, H., Schaffer, M.A., McAuley, K.B., 2012. A Kinetic Model for Non-Oxidative Thermal Degradation of Nylon 66. Macromol. React. Eng. 6, 93–109. doi:10.1002/mren.201100051

Kou, B., McAuley, K.B., Hsu, J.C.C., Bacon, D.W., 2005. Mathematical Model and Parameter Estimation for Gas-Phase Ethylene/Hexene Copolymerization With Metallocene Catalyst. Macromol. Mater. Eng. 290, 537–557. doi:10.1002/mame.200400392

Kreutz, C., Raue, A., Timmer, J., 2012. Likelihood Based Observability Analysis and Confidence Intervals for Predictions of Dynamic Models. BMC Syst. Biol. 6.

Lee, W., Liu, Y., 2012. Simultaneous Multiple Response Regression and Inverse Covariance Matrix Estimation via Penalized Gaussian Maximum Likelihood. J. Multivar. Anal. 111, 241–255. doi:10.1016/j.jmva.2012.03.013

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11, 431–441.

McLean, D.D., Pritchard, D.J., Bacon, D.W., Downie, J., 1979. Singularities in Multiresponse Modelling. Technometrics 21, 291–298.

McLean, K.A.P., McAuley, K.B., 2012. Mathematical Modelling of Chemical Processes-Obtaining the Best Model Predictions and Parameter Estimates Using Identifiability and Estimability Procedures. Can. J. Chem. Eng. 90, 351–366. doi:10.1002/cjce.20660

Monod, J., 1949. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 3, 371–394.

Montgomery, D.C., Runger, G.C., Hubele, N.F., 2004. Engineering Statistics, Third. ed. John Wiley and Sons Inc, New York.

Murphy, S.A., van der Vaart, A.W., 2000. On Profile Likelihood. J. Am. Stat. Assoc. 95, 449–465.

Nikerel, I.E., van Winden, W.A., Verheijen, P.J.T., Heijnen, J.J., 2009. Model Reduction and a Priori Kinetic Parameter Identifiability Analysis Using Metabolome Time Series for Metabolic Reaction Networks with Linlog Kinetics. Metab. Eng. 11, 20–30. doi:10.1016/j.ymben.2008.07.004

Oxby, P.W., 1997. Multivariate Weighted Least Squares as a Preferable Alternative to the Deteminant Criterion for Multirespoose Parameter Estimation. Univeristy of Waterloo.

Oxby, P.W., Duever, T.A., Reilly, P.M., 2003. Multivariate Least Squares as an Alternative to the Determinant Criterion for Multiresponse Parameter Estimation, in: Asprey, S.P.,

Macchietto, S. (Eds.), Dynamic Model Development: Methods, Theory and Applications. Elsevier Science Ltd., pp. 63–84.

Pedersen, P. V., 1977. Curve Fitting and Modeling in Pharmacokinetics and Some Practical Experiences with NONLIN and a New Program FUNFIT. J. Pharmacokinet. Biopharm. 5, 513–531.

Petersen, B., Gernaey, K., Vanrolleghem, P.A., 2001. Practical Identifiability of Model Parameters by Combined Respirometric-Titrimetric Measurements. Water Sci. Technol. 43, 347–355.

Phillips, P.C.B., 1976. The Iterated Minimum Distance Estimator and the Quasi-Maximum Likelihood Estimator. Econometrica 44, 449–460.

Polic, A.L., Lona, L.M.F., Duever, T.A., Penlidis, A., 2004. A Protocol for the Estimation of Parameters in Process Models: Case Studies with Polymerization Scenarios. Macromol. Theory Simulations 13, 115–132. doi:10.1002/mats.200300015

Poyton, A.A., Varziri, M.S., McAuley, K.B., McLellan, P.J., Ramsay, J.O., 2006. Parameter Estimation in Continuous-Time Dynamic Models Using Principal Differential Analysis. Comput. Chem. Eng. 30, 698–708. doi:10.1016/j.compchemeng.2005.11.008

Qi, X., Zhao, H., 2010. Asymptotic Efficiency and Finite-Sample Properties of the Generalized Profiling Estimation of Parameters in Ordinary Differential Equations. Ann. Stat. 38, 435–481. doi:10.1214/09-AOS724

Quinn, S.L., Harris, T.J., Bacon, D.W., 2000. Notes on Likelihood Intervals and Profiling. Commun. Stat. - Theory Methods 29, 109–129. doi:10.1080/03610920008832472

Quinn, S.L., Harris, T.J., Bacon, D.W., 2005. Accounting for Uncertainty in Control-Relevant Statistics. J. Process Control 15, 675–690. doi:10.1016/j.jprocont.2005.01.002

Raue, A., Becker, V., Klingmüller, U., Timmer, J., 2010. Identifiability and Observability Analysis for Experimental Design in Nonlinear Dynamical Models. Chaos 20, 0451051–0451058. doi:10.1063/1.3528102

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. Bioinformatics 25, 1923–9. doi:10.1093/bioinformatics/btp358

Raue, A., Kreutz, C., Maiwald, T., Klingmuller, U., Timmer, J., 2011. Addressing Parameter Identifiability by Model-Based Experimentation. IET Syst. Biol. 5, 120–30. doi:10.1049/iet-syb.2010.0061

Reid, N., Fraser, D.A.S., 2010. Mean Loglikelihood and Higher-Order Approximations. Biometrika 97, 159–170. doi:10.1093/biomet/asq001

Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R., 2006. Novel Metaheuristic for Parameter Estimation in Nonlinear Dynamic Biological Systems. BMC Bioinformatics 7, 483. doi:10.1186/1471-2105-7-483

Rothman, A.J., Levina, E., Zhu, J., 2010. Sparse Multivariate Regression With Covariance Estimation. J. Comput. Graph. Stat. 19, 947–962. doi:10.1198/jcgs.2010.09188

Routray, K., Deo, G., 2005. Kinetic Parameter Estimation for a Multiresponse Nonlinear Reaction Model. AIChE J. 51, 1733–1746. doi:10.1002/aic.10446

Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. John Wiley and Sons Inc.

Soo, Y., Bates, D.M., 1996. Multiresponse Spline Regression. Comput. Stat. Data Anal. 22, 619–631.

Sulieman, H., McLellan, P.J., Bacon, D.W., 2001. A Profile-Based Approach to Parametric Sensitivity Analysis of Nonlinear Regression Models. Technometrics 43, 425–433.

Sulieman, H., McLellan, P.J., Bacon, D.W., 2004. A Profile-Based Approach to Parametric Sensitivity in Multiresponse Regression Models. Comput. Stat. Data Anal. 45, 721–740. doi:10.1016/S0167-9473(03)00086-0

Thie, J.A., Smith, G.T., Hubner, K.F., 1997. Linear Least Squares Compartmental-Model-Independent Parameter Identification in PET. IEEE Trans. Med. Imaging 16, 11–6. doi:10.1109/42.552051

Thompson, D.E., McAuley, K.B., McLellan, P.J., 2010. Design of Optimal Sequential Experiments to Improve Model Predictions from a Polyethylene Molecular Weight Distribution Model. Macromol. React. Eng. 4, 73–85. doi:10.1002/mren.200900033

Vajda, S., 1983. Structural Identifiability of Dynamical Systems. Int. J. Syst. Sci. 14, 1229–1247.

Vajda, S., 1984. Analysis of Unique Structural Identifiability via Submodels. Math. Biosci. 71, 125–146.

Vajda, S., Rabitz, H., Walter, E., Lecourtier, Y., 1989. Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic Models. Chem. Eng. Commun. 83, 191–219.

Van Ewijk, P.H., Hoekstra, J.A., 1994. Curvature Measures and Confidence Intervals for the Linear Logistic Model. Appl. Stat. 43, 477–487.

Vanrolleghem, P.A., van Daele, M., Dochain, D., 1995. Practical Identifiability of a Biokinetic Model of Activated Sludge Respiration. Water Res. 29, 2561–2570.

Varziri, M.S., McAuley, K.B., McLellan, P.J., 2008. Approximate Maximum Likelihood Parameter Estimation for Nonlinear Dynamic Models: Application to a Laboratory-Scale Nylon Reactor Model. Ind. Eng. Chem. Res. 47, 7274–7283. doi:10.1021/ie800503v

Venzon, D.J., Moolgavkar, S.H., 1988. A Method for Computing Profile-Likelihood-Based Confidence Intervals. Appl. Stat. 37, 87–94.

Watts, D.G., 1994. Parameter Estimates from Nonlinear Models. Methods Enzymol. 240, 23–36.

Westlake, W.J., 1973. Use of Statistical Methods in Evaluation of In Vivo Performance of Dosage Forms. J. Pharm. Sci. 62, 1579–1589.

Wu, H., Zhu, H., Miao, H., Perelson, A.S., 2008. Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models. Bull. Math. Biol. 70, 785–99. doi:10.1007/s11538-007-9279-9

Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B., Bacon, D.W., 2003. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. Polym. React. Eng. 11, 563–588. doi:10.1081/PRE-120024426

Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. J. Am. Stat. Assoc. 57, 348–368.