Profile Monitoring with Fixed and Random Effects using Nonparametric and Semiparametric Methods

by

Abdel-Salam G. Abdel-Salam

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Dr. Jeffrey B. Birch, Chairman
Dr. Willis A. Jensen
Dr. Eric P. Smith
Dr. William H. Woodall

October 14, 2009
Blacksburg, Virginia

Keywords: Model Robust Profile Monitoring, Parametric, Nonparametric, P-splines, Semiparametric, Model Robust Regression, Model Misspecification, T^2 Control Chart.

Profile Monitoring with Fixed and Random Effects using Nonparametric and Semiparametric Methods

Abdel-Salam G. Abdel-Salam
Virginia Polytechnic Institute and State University, 2009
Advisor: Jeffrey B. Birch, Ph.D.

Abstract

Profile monitoring is a relatively new approach in quality control best used where the process data follow a profile (or curve) at each time period. The essential idea for profile monitoring is to model the profile via some parametric, nonparametric, and semiparametric methods and then monitor the fitted profiles or the estimated random effects over time to determine if there have been changes in the profiles. The majority of previous studies in profile monitoring focused on the parametric modeling of either linear or nonlinear profiles, with both fixed and random effects, under the assumption of correct model specification.

Our work considers those cases where the parametric model for the family of profiles is unknown or at least uncertain. Consequently, we consider monitoring profiles via two techniques, a nonparametric technique and a semiparametric procedure that combines both parametric and nonparametric profile fits, a procedure we refer to as model robust profile monitoring ($MRPM$). Also, we incorporate a mixed model approach to both the parametric and nonparametric model fits. For the mixed effects models, the $MMRPM$ method is an extension of the $MRPM$ method which incorporates a mixed model approach to both parametric and nonparametric model fits to account for the correlation within profiles and to deal with the collection of profiles as a random sample from a common population.

For each case, we formulated two Hotelling's T^2 statistics, one based on the estimated random effects and one based on the fitted values, and obtained the corresponding control limits. In addition, we used two different formulas for the estimated variance-covariance matrix: one based on the pooled sample variance-covariance matrix estimator and a second one based on the estimated variance-covariance matrix based on successive differences.

A Monte Carlo study was performed to compare the integrated mean square errors ($IMSE$) and the probability of signal of the parametric, nonparametric, and semiparametric approaches. Both correlated and uncorrelated errors structure scenarios were evaluated for varying amounts of model misspecification, number of profiles, number of observations per profile, shift location, and in- and out-of-control situations. The semiparametric ($MMRPM$) method for uncorrelated and correlated scenarios was competitive and, often, clearly superior with the parametric and nonparametric over all levels of misspecification. For a correctly specified model, the $IMSE$ and the simulated probability of signal for the parametric and the $MMRPM$ methods were identical (or nearly so). For the severe model misspecification case, the nonparametric and $MMRPM$ methods were identical (or nearly so). For the mild model misspecification case, the $MMRPM$ method was superior to the parametric and nonparametric methods. Therefore, this simulation supports the claim that the $MMRPM$ method is robust to model misspecification.

In addition, the $MMRPM$ method performed better for data sets with correlated error structure. Also, the performances of the nonparametric and $MMRPM$ methods improved as the number of observations per profile increases since more observations over the same range of X generally enables more knots to be used by the penalized spline method, resulting in greater flexibility and improved fits in the nonparametric curves and consequently, the semiparametric curves.

The parametric, nonparametric and semiparametric approaches were utilized for fitting the relationship between torque produced by an engine and engine speed in the automotive industry. Then, we used a Hotelling's T^2 statistic based on the estimated random effects to conduct Phase I studies to determine the outlying profiles. The para-

metric, nonparametric and seminonparametric methods showed that the process was stable. Despite the fact that all three methods reach the same conclusion regarding the "in-control" status of each profile, the nonparametric and $MMRPM$ results provide a better description of the actual behavior of each profile. Thus, the nonparametric and $MMRPM$ methods give the user greater ability to properly interpret the true relationship between engine speed and torque for this type of engine and an increased likelihood of detecting unusual engines in future production. Finally, we conclude that the nonparametric and semiparametric approaches performed better than the parametric approach when the user's model is misspecified. The case study demonstrates that, the proposed nonparametric and semiparametric methods are shown to be more efficient, flexible and robust to model misspecification for Phase I profile monitoring in a practical application.

Thus, our methods are robust to the common problem of model misspecification. We also found that both the nonparametric and the semiparametric methods result in charts with good abilities to detect changes in Phase I data, and in charts with easily calculated control limits. The proposed methods provide greater flexibility and efficiency than current parametric methods used in profile monitoring for Phase I that rely on correct model specification, an unrealistic situation in many practical problems in industrial applications.

بسم الله الرحمن الرحيم

In the Name of Allah, the Most Gracious, the Ever Merciful

I would like to dedicate this dissertation to my parents, my wife, my sweet daughter Salma, my lovely sons (Mohamed and Ali), and to all my family.

ABDEL-SALAM GOMAA ABDEL-SALAM

Fall 2009

Acknowledgments

Praise is due to Allah, the beneficent, the merciful. First and foremost, I would like to thank God for his blessings and help that He has bestowed on me through out my journey in life. Prophet Mohammad (Peace and Blessing be upon Him) said: *"He who does not thank the people, does not thank Allah the Almighty."* In this spirit, I will thank as many people as my memory enables me, and for those I mistakenly forget, please forgive my shortcoming.

I cannot emphasize how much I was lucky to work with Dr. Jeffrey B. Birch. Dr. Birch is an accomplished and extraordinary academician. He is a leader in his field. But that is not who Jeffrey B. Birch is; I think this is only a small part of who this individual is. Dr. Birch is the *advisor*. Yes, I think Dr. Birch is the *advisor*; this is how you define an advisor. He does not only monitor the academic and research progress of his students. Rather, he tries to make sure they succeed in their careers, and he actually helps to the best of his ability and beyond. He honestly points to his students their strengths and things they need to work on; he never mentions weaknesses! He advises on the writing style, problem approach, research methodology, career search, job interviews, all the way to how to carry yourself and answer questions in the professional world. He is an amazing person all around. Dr. Jeffrey B. Birch is a demonstration of how you can be a great academic with an extremely busy schedule, while remaining a successful father and husband, and an amazing boss who can touch the lives of all his subordinates. Truly, he is a one of a kind family man, with unsurpassed social and personal skills. Being around him is so much fun, and watching him carry himself in any situation is just mesmerizing. Dr. Birch is student oriented, often chatting with us, and asking about our families.

I am deeply affected by Dr. Birch, and I really wish I could become an advisor like him one day.

I would like to thank my committee members: Dr. Willis A. Jensen, Dr. Eric P. Smith and Dr. William H. Woodall. They were very helpful and always provided positive input to further improve this work. I thank Dr. Woodall for the opportunity to work with him on an other research topic. It was his initial presentation early in my Ph.D. program that lead me down this path of research which has been so very enjoyable for me. I also thank the other members of my committee for their attention for detail and numerous marking of my manuscripts have hopefully helped me to be a better writer. I thank Dr. Oliver Schabenberger for bringing my attention to the penalized spline regression as a nonparametric technique. Also, I would like to extend my deepest appreciation to Dr. Jensen and his co-authors (Amirhossein Amiri and Reza Kazemzadeh) for allowing me to apply our proposed methods on their Automotive Industry data set.

I have had the pleasure of knowing many people in Virginia Tech, starting with the Department of Statistics, Department of Civil and Environmental Engineering and the Virginia Tech Transportation Institute (VTTI), where the environment was so friendly and everyone was willing to help with everything. It has been a great pleasure working with the faculty, staff and students in the Department of Statistics at Virginia Tech during my PhD. program of study. I gratefully acknowledge all of them for their great support and encouragement. I also appreciate the friendship of all the graduate students at Virginia Tech. I also would like to extend my deepest appreciation to my professors, staff and colleagues at Faculty of Economics and Political Science, Cairo University for their great helping for me and my family.

Furthermore, I like to thank the Department of Statistics, Virginia Tech for supporting me financially through GTAs. My gratitude is due to the Faculty of Economics and Political Science (Cairo University) and to Dr. Suaad Al Sabah for supporting me financially throughout my stay in the USA. Without this generous support this work would not have been done. My family and I are deeply grateful for all the Muslim community in Blacksburg.

I thank my parents, parents-in-law, brothers, sisters and all my family. They have all been very encouraging and have done their best to support us in this endeavor from very long distance. To my mother, father, grandfather and grandmother, I apologize for there is no way to say thank you. However, I can promise, and so far I have kept my promises, that I will absolutely do everything possible to raise my kids with the same ethics, standards and dedication your raised me with.

And, the sweet comes at the end. I can not thank enough my wonderful wife (*Sherin*), and my beautiful three children: *Salma, Mohamed* and *Ali*, for their love and support. My wife gave up a more comfortable, secure situation so that I could pursue this degree. She has put up with me many hours and nights of my studying, even when she was not feeling well, allowing us to complete our goal together. I would also like to dedicate this dissertation to my greatest gifts, my wife and my kids.

Abdel-Salam G. Abdel-Salam

Virginia Polytechnic Institute and State University

October 2009

Contents

List of Tables

List of Figures

Acronyms

blup	Best Linear Unbiased Predictor
CLM	Conditional Local Mixed
CS	Cluster Specific
CV	Cross Validation
df	Degrees of Freedom
eblup	Estimated Best Linear Unbiased Prediction
GCV	Generalized Cross Validation
GLS	Generalized Least Squares
HDS	Historical Data Set
Ker	Kernel Regression
LLR	Local Linear Regression
LM	Linear Mixed
LPR	Local Polynomial Regression
MC	Monte Carlo
ML	Maximum Likelihood
MMRPM	Mixed Model Robust Profile Monitoring
MRR	Model Robust Regression
MRPM	Model Robust Profile Monitoring
NLM	Nonlinear Mixed
NP	Nonparametric
OLS	Ordinary Least Squares
Par	Parametric

PA	Population Average
PS	Penalized Spline
REML	Restricted Maximum Likelihood
SAS	Statistical Analysis Software
SLP	Simple Linear Profile
SPC	Statistical Process Control
SSE	Sum of Squared Errors
UCL	Upper Control Limit

Nomenclature

$\mathbf{b}$	A $(q \times 1)$ vector of unknown random effects parameters
$\mathbf{b}_i$	A vector of random effects that represents the $eblups$ for the i^{th} profile from the PA
B	Variance-covariance matrix for $\mathbf{b}$, $B = diag(D)$
β	A $(p \times 1)$ vector of unknown fixed parameters for PA
β_i	A $(p \times 1)$ vector of unknown fixed parameters for the i^{th} profile
D	Variance-covariance matrix for $\mathbf{b}_i$, $D = \sigma_b^2 I$
d	Degree for the polynomial regression
df_1	The df for χ^2 distribution, $df_1 = tr(H_{CS,i}^{MRPM})$
df_2	The df for χ^2 distribution, $df_2 = tr(H_{CS,i}^{MMRPM})$
ϵ_{ij}	The error value for j^{th} observation from the i^{th} profile (group or cluster)
ϵ_i	An $(n_i \times 1)$ vector of errors
ϵ	An $(n \times 1)$ stacked vector of errors
$f(.)$	True PA profile function
$f_i(.)$	True i^{th} profile function
$g_i(.)$	The i^{th} profile smooth function
γ	The misspecification parameter
h	Bandwidth
H^{Ker}	Smoother matrix for the kernel regression
H^P	Smoother (hat) matrix for the parametric fit
H^{LPR}	Smoother matrix for the local polynomial regression
H^{LLR}	Smoother matrix for the local linear regression
H^C	Smoother matrix for the CLM model

H^{MRPM}	Smoother matrix for the $MRPM$ model
H^{MMRPM}	Smoother matrix for the $MMRPM$ model
K	Number of knots in p-spline
K_1	Number of knots in p-spline for the PA profile
K_2	Number of knots in p-spline for the CS profiles
L_1	The df for χ^2 distribution, $L_1 = tr(H^{LLR}_{CS,i})$
L_2	The df for χ^2 distribution, $L_2 = tr(H^{C}_{CS,i})$
L_{LR}	The likelihood or restricted likelihood value
λ	Mixing parameter
$\hat{\lambda}_{CS}$	Estimated mixing parameter for the CS fits
$\hat{\lambda}_{PA}$	Estimated mixing parameter for the PA fits
m	Number of profiles in HDS
m^*	Number of remaining profiles in HDS after removing the outliers profiles
n_i	Sample size for the i^{th} profile
n	Total sample size for the HDS
n'	Sample size for comparing m profiles, when n_i is the same for all profiles
p	Number of fixed effects
q	Number of random effects
R	Variance-covariance matrix for ϵ
R_i	Variance-covariance matrix for ϵ_i
$\hat{S}_P$	Moment estimator of V
$\hat{S}_D$	Estimated variance-covariance matrix based on the successive difference
$\mathbf{t}_i$	A vector of random effects represent the coefficients for spline component in the i^{th} profile
$T^2_{P,i}$	Hotelling T^2 statistics for the i^{th} parametric fitted profile
$T^2_{Par1,i}$	Hotelling T^2 statistics for the i^{th} profile based on the fitted values parametrically
$T^2_{Par2,i}$	Hotelling T^2 statistics for the i^{th} profile based on the estimated random effects parametrically
$T^2_{NP,i}$	Hotelling T^2 statistics for the i^{th} NP fitted profile
$T^2_{NP1,i}$	Hotelling T^2 statistics for the i^{th} profile based on the fitted values

	nonparametrically
$T^2_{NP2,i}$	Hotelling T^2 statistics for the i^{th} profile based on the estimated random effects nonparametrically
$T^2_{LLR,i}$	Hotelling T^2 statistics for the i^{th} LLR fitted profile
$T^2_{C,i}$	Hotelling T^2 statistics for the i^{th} CLM fitted profile
$T^2_{MRPM,i}$	Hotelling T^2 statistics for the i^{th} $MRPM$ fitted profile
$T^2_{MMRPM,i}$	Hotelling T^2 statistics for the i^{th} $MMRPM$ fitted profile
$T^2_{MMRPM1,i}$	Hotelling T^2 statistics for the i^{th} profile based on the fitted values semiparametrically
$T^2_{MMRPM2,i}$	Hotelling T^2 statistics for the i^{th} profile based on the estimated random effects semiparametrically
$T^2_{PS1,i}$	Hotelling T^2 statistics based on p-spline fits for the i^{th} mixed effects profile and $\hat{S}_P$
$T^2_{PS2,i}$	Hotelling T^2 statistics based on p-spline fits for the i^{th} mixed effects profile and $\hat{S}_D$
$T^2_{PSF1,i}$	Hotelling T^2 statistics based on p-spline fits for the i^{th} fixed effects profile and $\hat{S}_P$
$T^2_{PSF2,i}$	Hotelling T^2 statistics based on p-spline fits for the i^{th} fixed effects profile and $\hat{S}_D$
$\mathbf{u}$	A($K_1 \times 1$) vector of random effects in p-spline for the PA profile
V	Variance-covariance matrix of the response, $\mathbf{y}$, $Var(\mathbf{y}) = V = ZBZ^T + R$
$\hat{\mathbf{v}}_i$	A vector of successive differences between $\mathbf{a}_{i+1}$ and $\mathbf{a}_i$
W_0	A diagonal matrix for the kernel weights
x_{ij}	The fixed effects value for j^{th} observation in the i^{th} explanatory
x^*_l	A specific point for x, $l = 1,2,\ldots,n'$
x_0	An arbitrary point for x
$\mathbf{x}_i$	A vector of n_i stacked values of a single regressor
$\tilde{\mathbf{x}}_l$	The l^{th} row vector of the model matrix for a d^{th} order polynomial
X_i	An ($n_i \times p$) matrix of regressor variables for the fixed effects
X	An ($n \times p$) stacked matrix of m X_i
$\tilde{X}$	The model matrix for a d^{th} order polynomial
y_{ij}	The response value for j^{th} observation from the i^{th} profile (group or cluster)
$\mathbf{y}_i$	An ($n_i \times 1$) vector of response for the i^{th} profile
$\mathbf{y}$	An ($n \times 1$) stacked vector of responses

$\hat{\mathbf{y}}^{P}_{i,-i}$	A vector of parametric fitted values for the i^{th} profile with the i^{th} profile deleted
$\hat{\mathbf{y}}^{NP}_{i,-i}$	A vector of NP fitted values for the i^{th} profile with the i^{th} profile deleted
$\hat{\mathbf{y}}^{P}_{PA}$	A vector of parametric fitted values for the PA profile
$\hat{\mathbf{y}}^{NP}_{PA}$	A vector of NP fitted values for the PA profile
$\hat{\mathbf{y}}^{LLR}_{PA}$	A vector of LLR fitted values for the PA profile
$\hat{\mathbf{y}}^{C}_{PA}$	A vector of CLM fitted values for the PA profile
$\hat{\mathbf{y}}^{PS}_{PA}$	A p-spline fitted value for the PA profile
$\hat{\mathbf{y}}^{LLR}_{PA,l}$	A LLR fitted value for the PA profile at x^*_l
$\hat{\mathbf{y}}^{C}_{PA,l}$	A CLM fitted value for the PA profile at x^*_l
$\hat{\mathbf{y}}^{MRPM}_{PA,l}$	A $MRPM$ fitted value for the PA profile at x^*_l
$\hat{\mathbf{y}}^{MMRPM}_{PA,l}$	A $MMRPM$ fitted value for the PA profile at x^*_l
$\hat{\mathbf{y}}^{P}_{CS,i}$	A vector of parametric fitted values for the i^{th} profile
$\hat{\mathbf{y}}^{NP}_{CS,i}$	A vector of NP fitted values for the i^{th} profile
$\hat{\mathbf{y}}^{LLR}_{CS,i}$	A vector of LLR fitted values for the i^{th} profile
$\hat{\mathbf{y}}^{C}_{CS,i}$	A vector of CLM fitted values for the i^{th} profile
$\hat{\mathbf{y}}^{LLR}_{CS,i,l}$	A LLR fitted value for the i^{th} profile at x^*_l
$\hat{\mathbf{y}}^{C}_{CS,i,l}$	A CLM fitted value for the i^{th} profile at x^*_l
$\hat{\mathbf{y}}^{PSF}_{CS,i}$	A fixed effects p-spline fitted value for the i^{th} profile
$\hat{\mathbf{y}}^{PS}_{CS,i}$	A mixed p-spline fitted value for the i^{th} profile
$\hat{\mathbf{y}}^{MRPM}_{CS,i,l}$	A $MRPM$ fitted value for the i^{th} profile at x^*_l
$\hat{\mathbf{y}}^{MMRPM}_{CS,i,l}$	A $MMRPM$ fitted value for the i^{th} profile at x^*_l
Z_i	An $(n_i \times q)$ matrix of regressor variables for the random effects
$\mathbf{z}_i$	A vector of n_i stacked values of a single random effects
Z	An $(n \times q)$ stacked matrix of m Z_i

‘

Chapter 1

Introduction and Motivation

Profile monitoring is relatively a new approach in statistic process control (SPC) that is important when the product or process quality is best represented by a profile at each time period. Profile monitoring combines the concept of fitting profiles using regression techniques and the notion of separating special cause variability from common cause variability in quality control (Mahmoud and Woodall, 2004).

1.1 Phase I and Phase II

The analysis of profile monitoring can be performed into two phases, Phase I and Phase II. The aim of the analysis in Phase I (retrospective analysis) is to obtain the estimated in-control-limits by analyzing a historical data set (HDS) to gain understanding of the process variation, to determine the process stability, and to remove the outlying samples when setting control charts for Phase II analysis. In the Phase I HDS, trends, step change, outliers and any other types of instability may have adverse effect on the resulting Phase II control limits. One important activity in Phase I is to separate in-control and out-of-control data within historical observations. The performance of Phase I control chart methods are often measure in terms of the probability of signal, which is the probability of getting at least one charted statistic outside the control limits.

Phase II (future monitoring) consists of monitoring future observations utilizing control limits calculated from Phase I to determine if the process continues to be stable. The performance of the control chart in Phase II is often measured by the average run length, which is the number of samples taken until the first out-of-control signal. For

more details regarding the differences between the analysis for Phase I and Phase II in SPC see Sullivan (2002), Mahmoud and Woodall (2004), and Montgomery (2005).

In Phase I analysis, profile monitoring can be accomplished mainly using two approaches: the first is the parametric approach where the focus is on monitoring the estimated parameters of each profile. The second is the nonparametric approach where the focus is on monitoring the estimated profiles instead of monitoring estimated parameters.

1.2 Profile Monitoring Literature

In most of SPC applications, the quality of a process or product is characterized by univariate or multivariate quality characteristics. However, in some applications the quality of a process or product is characterized by a relationship between a response variable and one or more explanatory variables. Kang and Albin (2000) refer to this relationship as *a profile*.

In general, the act of using various techniques to statistically monitor the process or product profiles is known as *profile monitoring* (Woodall et al., 2004; Woodall, 2007). Profile monitoring follows the basic approach of functional data analysis in that the collection of observed data for all the process variables are treated as a single profile, rather than as merely a sequence of individual observations (Ramsay and Silverman, 2005).

Woodall et al. (2004) and Woodall (2007) presented a literature review on this subject and introduced a general framework for process monitoring using profile data. They introduced a general strategy for monitoring more complicated parametric models than the simple linear regression (SLR) model, including, for example, applications to nonlinear models, and they discuss several NP methods, such as wavelets and splines, to fit each profile. Practical applications of profile monitoring have been reported by many researchers including Stover and Brill (1998), Kang and Albin (2000), Kim et al. (2003), Mahmoud and Woodall (2004), Woodall et al. (2004), Wang and Tsung (2005), Zhou et al. (2006), Jeong et al. (2006), Chang and Gan (2006), Wu (2007), Zou et al. (2007), Staudhammer et al. (2007), Shao and Wu (2007), Bersimis et al. (2007), Hawkins and

Maboudou-Tchao (2007), Mahmoud (2008), Kazemzadeh et al. (2008), Kusiak et al. (2009), Chicken et al. (2009), Shiau et al. (2009), Zhang et al. (2009), Saghaei et al. (2009), and Wei et al. (2010).

Several control chart techniques have been developed by researchers for monitoring parametric simple linear profiles (SLP) in both Phase I and Phase II applications, $e.g.$, Noorossana et al. (2008), Kazemzadeh et al. (2009a), and Soleimani et al. (2009). Kang and Albin (2000) proposed two methods for monitoring SLP. The first method uses a bivariate T^2 control chart to monitor parameters of the parametric regression line representing the SLP. The second method uses exponentially weighted moving average ($EWMA$) and R control charts to monitor the mean and the variance of residuals, respectively. Kim et al. (2003) coded the x-values to average zero to make the parameter estimates uncorrelated and then applied three separate $EWMA$ control charts. Application of three separate $EWMA$ control charts not only improves the average run length (ARL) performance but also helps interpretation of out-of-control signals.

Mahmoud and Woodall (2004) recommended the use of a univariate control chart to monitor error standard deviation in conjunction with a global F-test to monitor the regression coefficients in Phase I. Zou et al. (2006) and Mahmoud et al. (2007) proposed methods based on likelihood ratio statistics to detect changes in the parameters of SLP in Phase I and Phase II respectively.

Jensen et al. (2008) proposed use of linear mixed (LM) models to account for autocorrelation within a linear profile using a parametric approach. Noorossana et al. (2008) investigated the effect of autocorrelation between linear profiles on the performance of T^2 control chart proposed by Kang and Albin (2000). In addition, they proposed three methods based on a time series approach to eliminate the autocorrelation effect.

Nonlinear profile applications were studied by Jin and Shi (1999), Lada et al. (2002), Walker and Wright (2002), Ding and Zhou (2006), Gupta et al. (2006), Williams et al. (2007a), Williams et al. (2007b), and Jensen and Birch (2009). Lada et al. (2002) and Ding and Zhou (2006) investigated a general class of nonlinear profiles using techniques such as dimension reduction, wavelet transformations and independent analysis. Williams et al. (2007b) developed the multivariate T^2 control chart to monitor profiles which can

be represented by a parametric nonlinear regression model. They applied three approaches to estimate the variance-covariance matrix: the sample variance-covariance matrix, the successive difference estimator, and the minimum volume ellipsoid estimator. For each approach, the control limits were obtained based on the beta distribution, the chi-square distribution and by simulation, respectively.

Williams et al. (2007a) used the nonlinear regression approach of Williams et al. (2007b) to monitor dose-response profiles used in high-throughput screening. They focused on monitoring the parameters of the nonlinear regression model. A 4-parameter logistic regression model was used to represent the profiles. Jensen and Birch (2009) introduced nonlinear mixed (NLM) models for the correlated and uncorrelated data. They showed that NLM models could have significant advantages over nonlinear regression models when the data are correlated or even uncorrelated within profiles.

Not all of the previous work in profile monitoring has been completed using parametric models. Several researchers have relied on NP regression or data-driven methods, such as wavelet thresholding, spline regression and local polynomial regression for monitoring profiles Kazemzadeh et al. (2008), Zou et al. (2009b), Zou et al. (2009a), Qiu and Zou (2009) and Kazemzadeh et al. (2009b). Reis and Saraiva (2006), Jeong et al. (2006), Zou et al. (2007), and Chicken et al. (2009) explored the nonparametric wavelet models and constructed the control charts based on a subset of wavelet coefficients. Winistorfer et al. (1996) utilized the spline smoothing regression to fit the vertical density profiles (VDP) of pressed wood panels and tested for significance differences between each pair of these profiles. Zou et al. (2008) presented an alternative nonparametric approach for profile monitoring but they assumed the measures within each profile are independent, which might restricted their applications.

Walker and Wright (2002) used an additive model to assess the sources of variation active on the VDP data. Their model contained a B-spline to smooth the profile data and a parametric portion to incorporate other sources of variation. Williams et al. (2007b) replicated the spline fits to each VDP and proposed several detection criterion based on the deviation of each fitted spline from the average spline using dissimilarity metrics suggested by researchers at Boeing Commercial Airplane Group (1998). Colosimo and

Pacella (2007) introduced methods for monitoring roundness profiles of some manufactured items.

Zou et al. (2009b) proposed a technique for Phase II analysis by integrating the multivariate exponentially weighted moving average procedure with the generalized likelihood ratio test based on local linear regression. This technique can be used for monitoring on-line changes in both the regression relationship and the variation of the profile.

Recently, Wei et al. (2010) proposed a nonparametric $L-1$ regression location-scale model to screen the shape of the profiles in Phase II analysis. Their method is robust against heavy-tailed distributions that do not have finite second moments. Qiu et al. (2010) introduced a nonparametric procedure by incorporating local linear kernel smoothing in the exponentially weighted moving average ($EWMA$) control scheme to perform Phase II profile monitoring.

1.3 Motivation

Much of the previous work on profile monitoring has been based on the assumption that the parametric models for profiles are correctly specified. This is often an unrealistic assumption in practice for many types of applications. For example, in plotting the profile data, the researcher may see features in the scatter plot such as peaks, dips or local wiggles that are not captured by a parametric profile of any type, linear or nonlinear. While these features may be unique to a single or a few profiles and thus could be considered outliers, they can often be features that consistently appear in all the profiles and thus should be captured in the model.

The parametric fixed or mixed effects model may be misspecified in different ways, including an incorrect covariance structure, wrong distributional assumptions, wrong effect classification (either fixed or random), and wrong model matrices.

In these situations, the researcher may still want a "function" to describe the profile, although it may not be a parametric one. NP models are ones where the curve used to describe the profile can not be expressed as a parametric model. An incorrectly specified parametric profile model may be improved by using a NP profile model. Such a NP

model may be used in estimated profiles with greatly reduced bias in estimating the true profile than achieved by using an incorrectly specified parametric model.

The virtues of NP regression models have been discussed extensively in the statistics literature. Competing approaches to NP modeling include, but are not limited, smoothing splines, kernel methods Fan and Gijbels (1995), regression splines and penalized splines Wand (2003).

The NP approach to fitting profile data is more flexible than a purely parametric approach. In modeling new data, one often has very little idea of an appropriate form for the model. We do have a number of heuristic tools using diagnostic plots to help search for this form, but it would be useful to let the modeling approach complement this search. One more disadvantage of the parametric approach is that one can easily choose the wrong form for the model leading to biased estimators, the direct result of model misspecification. The NP approach requires fewer assumptions about the model's form and consequently this approach can be less likely to make serious mistakes in estimation of mean response. The NP approach is particularly useful when little past experience is available. For more details see Faraway (2006), and Hastie et al. (2009).

Even when a specific functional form appears reasonable, the NP model provides a more robust model alternative that can be useful in the process of model checking and validation. In the NP framework the shape of the functional relationship between covariates and the dependent variables is determined by the data, whereas in the fixed or mixed effects parametric framework the shape is determined by the model DeBoor (2001).

Mixed effects models include at least one fixed effect and at least one random effect in addition to the error term. When mixed models arise in practice, the data are often grouped together by a common characteristic. These groups are known as clusters or profiles. Clustered data includes situations such as repeated measures on subjects as well as split-plot experiments where the whole-plot is the cluster. For more details on cluster data see Schabenberger and Gotway (2005), and West et al. (2006).

An example of clustered data is the wind speed data set from Haslett and Raftery

(1989). The data set represents twelve meteorological stations in Ireland. Twelve stations were selected and the average wind speed in knots was measured daily during the years 1961 to 1978. Researchers are interested in the average weekly wind speed over the eighteen years. Each station was randomly chosen from a population of stations so that station is a random effect. Measurements were taken at the same fifty three time points for each station.

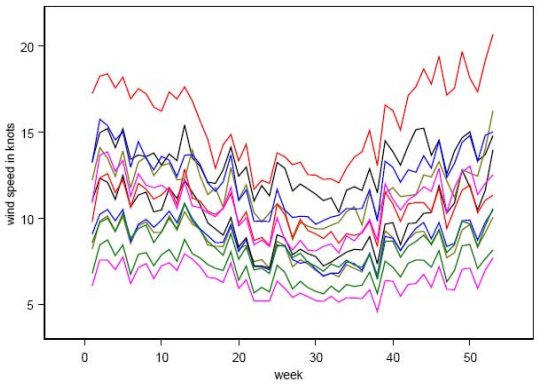

Figure 1.1: Plot of Wind Speed versus Week by Station

Figure (1.1) represents a plot of the wind speed data, where a line connects each wind speed value by week for each station. In this example the cluster is the station. Hence there are twelve clusters with each containing fifty three observations. This is just an example for a longitudinal data set. It is not a appropriate for profile monitoring since there is no time ordering.

One of our motivations for this work is the vertical density profile (VDP) data set from Walker and Wright (2002). The density is measured by a profilometer which uses a laser device to take a series of measurements across the thickness of the board. A profilometer takes multiple measurements on a sample usually a (2×2) inch piece. A baseline sample of twenty-four particleboards, each one consists of 314 measurements, taken every 0.002 inches, are illustrated in Figure (1.2).

From Figure (1.2), one can see that the data set contains complicated profiles, not easily fitted by a parametric approach. Several authors, for example Walker and Wright (2002), and Wei et al. (2010) utilized the nonparametric approach to fit these profiles.

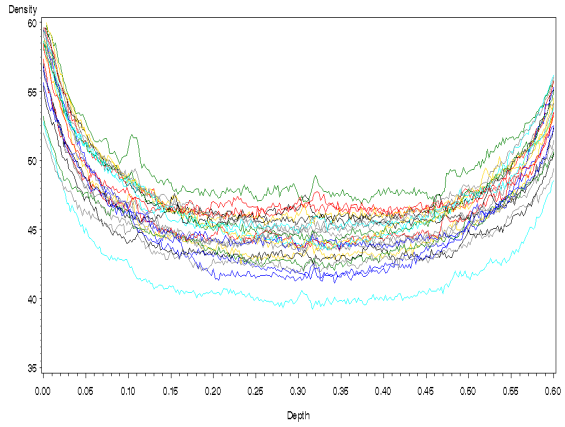

Figure 1.2: Plot of Vertical Density Profile of 24 Particleboards

Another motivation comes from the automotive industry, where one of the most important quality characteristic of the automobile engine is the relationship between the torque produced by an engine and the engine speed in revolutions per minute (RPM). The application is for the engine type $TU3$ which are assembled for a French automobile, the Peugeot. In the study, the engine is run at different RPM values and the corresponding torque values obtained. This means that the torque produced by the engine is considered as response variable and the correspondent speed RPM is considered as the explanatory variables. The profiles that describe the relationship between torque and RPM should be similar, when the manufacturing process is in-control. An engine with mechanical defects or any other issues will result in an outlying profile. Because there are multiple RPM values obtained for each engine it is natural to try to apply a multivariate quality control procedure to discover engines which would not be acceptable. This is a new data set from Amirhossein et al. (2009), and it is reasonable to apply our proposed methods. Therefore, we are going to apply our new methods for profile monitoring on this dataset in Chapter 6. The data set appears in Appendix G.

In this research, we propose two new methods for profile monitoring: a nonparametric method and a semiparametric procedure that combines both parametric and NP profile fits, a procedure we refer to as model robust profile monitoring ($MRPM$). Our methods can be used to monitor a broad category of profiles, either linear or nonlinear, by using the historical profiles to estimate the average (baseline) profile and cluster spe-

cific (CS) profiles both nonparametrically and semiparametrically. We consider two NP approaches, one based on penalized spline regression and one based on local polynomial regression. In our application, we will focus on using penalized spline regression because of some computational aspects.

Moreover, we incorporate a mixed model approach to both the parametric and NP model fits in order to account for the correlation structure within profiles. We name this method as "mixed model robust profile monitoring ($MMRPM$)". As a consequence, we speculate our new methods, both the NP method and the semiparametric method, will result in charts with good abilities to detect changes in Phase I data and have simple to calculate control limits.

Profile monitoring analysis using our methods should provide greater flexibility and efficiency over current parametric methods that rely on correct model specification, an unrealistic situation in many practical problems in industrial applications. In other words, both new methods, for the fixed and random effects models, should be robust to the common problem of model misspecification. We will focus here on Phase I control chart applications.

1.4 Dissertation Layout

The remainder of this dissertation is organized as follows. In Chapter 2, we review briefly the parametric approach for the fixed effects models. Also, we introduce penalized spline regression and local polynomial regression as new nonparametric methods to estimate fixed effects profiles. In addition, we present diagnostic tools to determine outlying profile(s). In Chapter 3 we introduce a model robust profile monitoring ($MRPM$) method for fixed effects profiles. In addition, we present a selection criteria for bandwidth and mixing parameter with its asymptotic properties. Chapter 4 presents the penalized spline and the conditional local mixed model as nonparametric methods for profile monitoring in mixed models. In addition, we extend the ($MRPM$) method to create an $MMRPM$ method for mixed models. Chapter 5 contains the simulation studies done to compare the parametric, nonparametric, and mixed model robust profile mon-

itoring methods. In Chapter 6, we also apply the proposed methods and the parametric quadratic mixed model to the automotive industry data set, and Chapter 7 summarizes the conclusions and lists some future considerations.

Chapter **2**

Parametric and Nonparametric Profile Monitoring for Fixed Effects Models

A fixed effects model contains only constants in its systematic part and one random variable which is the error term (Schabenberger and Pierce, 2002). Suppose we have m profiles from a HDS and that each profile contains n_i observations, where i indicates the i^{th} profile, $i = 1,2,\ldots,m$. Ignoring the fact that each profile is sampled at random from the population of profiles, nature's model can be expressed as

$$y_{ij} = f(x_{ij}) + \epsilon_{ij} \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots n_i. \tag{2.1}$$

where y_{ij} represents the j^{th} observation from the i^{th} profile. $f(x_{ij})$ is any arbitrary mean function representing nature's model as a function in one regressor, x_{ij}, and ϵ_{ij} represents the random error term for the j^{th} observation in the i^{th} profile, where it is assumed $\epsilon_{ij} \sim N(0,\sigma^2_\epsilon)$. Of course, (2.1) can easily be extended to deal with more than one regressor but we will consider only one in this research. For multiple linear regression profile monitoring using a parametric approach see Mahmoud (2008).

Our goals for this chapter are to estimate the fixed effects profiles by suitable parametric and NP techniques. We extend penalized spline regression and local polynomial regression to the area of profile monitoring. These NP methods use only the data themselves to provide these estimates. Furthermore, the T^2 statistic will be obtained and utilized to determine outlying profile(s). The novelty of our methods is in monitoring estimated fixed effects profiles. This allows us to monitor small, but significant departures from the basic shape of the profiles.

This chapter is organized as follows. In the next section, we provide a brief review of the parametric approach for estimating the fixed effects profiles. We review briefly Hotelling's T^2 statistic for monitoring a multivariate process with two different formulas for estimating the variance-covariance matrix in Section 2.2. In Section 2.3 we introduce our NP approach for fixed effects profile monitoring using penalized spline regression and introduce T^2 statistics based on the fitted profiles. In Section 2.4, we present local polynomial regression as an alternative NP method for fixed effects profile monitoring. Moreover, we give an algorithm for determining outlying profiles using T^2 statistic based on local linear regression (LLR) fits. The chapter is summarized in Section 2.6.

2.1 Parametric Estimation for Fixed Effects Profiles

One commonly used form of a parametric model is the linear model of conditional means that can be expressed in matrix form, and is given for the i^{th} profile by

$$\mathbf{y}_i = X_i\beta_i + \epsilon_i \qquad\qquad i = 1,2,\ldots,m. \tag{2.2}$$

where $\mathbf{y}_i$ is a $(n_i \times 1)$ vector of responses for the i^{th} profile, X_i is a $(n_i \times p)$ matrix of the regressor variables associated with the fixed effects and β_i is a $(p \times 1)$ vector of parameters for fixed effects in the i^{th} profile. ϵ_i is a $(n_i \times 1)$ vector of errors assumed to follow a multivariate normal distribution with zero mean vector and R_i variance-covariance matrix, $\epsilon_i \sim MN(0, R_i)$.

It is convenient to stack the responses and the model matrices for the individual profiles. Let $\mathbf{y}$ represent the stacking of $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$, resulting in the $(n \times 1)$ vector $\mathbf{y}$, where $n = \sum_{i=1}^{m} n_i$ is the total sample size. Similarly, let X represent the stacking of the m X_i matrices. Model (2.2) can be written as

$$\mathbf{y} = X\beta + \epsilon. \tag{2.3}$$

where $\mathbf{y}$ is an $(n \times 1)$ stacked vector of responses, X is a $(n \times p)$ stacked model matrix, β is a $(p \times 1)$ vector of unknown fixed parameters, and ϵ is an $(n \times 1)$ vector of random errors. Under the above assumptions for the error term, one interpretation of the fact that

$E(\epsilon) = 0$ is the implication that model (2.2), the so-called "user's model", is the correct model and, therefore, equal to (2.1).

To estimate the parametric fixed effects model, the ordinary least squares (OLS) method or generalized least squares (GLS) method can be used. See Appendix (A.1) for more details. The parametric fit for the estimated population average (PA) profile is

$$\hat{\mathbf{y}}_{PA}^{P} = X\hat{\beta}. \tag{2.4}$$

where $\hat{\beta} = (X^T R^{-1} X)^{-1} X^T R^{-1} \mathbf{y}$ is the fixed effect parameter estimators representing the PA parameters across all profiles. Here, $R = diag(R_i)$ is the $(n \times n)$ variance-covariance matrix for the stacked response observations. If the errors are assumed to be independent and homogeneous, then $R = \sigma^2 I$, where I is the identity matrix and the estimated parameter vector can be easy obtained using the OLS method where $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$.

The parametric fits for each cluster specific (CS) curve or profile are

$$\hat{\mathbf{y}}_{CS,i}^{P} = X_i \hat{\beta}_i \qquad\qquad i = 1, 2, \ldots, m. \tag{2.5}$$

where $\hat{\mathbf{y}}_{CS,i}^{P}$ represents the parametric CS fit for the i^{th} profile, and $\hat{\beta}_i$ indicate the fixed effect parameter estimator for the i^{th} profile, $\hat{\beta}_i = (X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} \mathbf{y}_i$.

Notice that, the PA and CS fits rely on knowledge of the variance-covariance matrix. In practice, R is seldom known. Suppose that $var(\mathbf{y}) = R = \sigma^2 \tilde{R}$ is known up to a scalar σ^2. This constant can be estimated (Schabenberger and Pierce, 2002) as

$$\hat{\sigma}^2 = \frac{1}{n - rank(X)} (\mathbf{y} - X\hat{\beta})^T \tilde{R}^{-1} (\mathbf{y} - X\hat{\beta}). \tag{2.6}$$

Another possibility is that $var(\mathbf{y}) = R$ is completely unknown. An unstructured variance-covariance matrix could be assumed with the estimates obtained by the method of moments, or R may be assumed to be a structured form, such as the $AR(1)$ or the Toeplitz forms, with estimates of unknown parameters obtained in standard ways (Schabenberger and Pierce, 2002). In each case, the estimated variance-covariance matrix ($\hat{R}$) or $\hat{R}_i$ is substituted into the above formulas for $\hat{\beta}$ or $\hat{\beta}_i$.

Several authors, *e.g.*, Kang and Albin (2000), Kim et al. (2003), and Mahmoud and Woodall (2004), have utilized the T^2 statistic to determine outlying profiles based on the

estimated parameters and assuming the parametric model was correctly specified. The following Section 2.2 gives a general frame work for the Hotelling's T^2 statistic.

2.2 Hotelling's T^2 Statistic

The T^2 control chart was proposed by Hotelling in 1941. Fuchs and Kenett (1998) and Mason and Young (2002) gave extensive reviews of the use of the T^2 chart. Multivariate control charts are very important to monitor the curve or surface that indicates the quality of a certain product. This curve or surface is called a "*profile*" (Woodall et al., 2004).

To monitor a multivariate process in a way that takes into account the correlations among the variates we may use a Hotelling's T^2 control chart (Tracy et al., 1992). The main aim of multivariate control charts is to determine the existence of the special causes of variation. Specifically, for Phase I analysis the objectives are

1. To determine shifts in the mean function from the estimation of the in-control mean function "baseline".

2. To identify and remove the multivariate outliers.

In order to develop the methodology to monitor parametric, NP, and semiparametric profiles, we first consider the general framework of the multivariate T^2 statistic or Hotelling T^2.

Given a sample of m independent observation vectors to be monitored, $\mathbf{a}_i$ $(i = 1,2,\ldots,m)$, each of dimension k, the general form of the T^2 statistic in Phase I for observation i is

$$T_i^2 = (\mathbf{a}_i - \bar{\mathbf{a}})^T \hat{S}^{-1} (\mathbf{a}_i - \bar{\mathbf{a}}) \qquad i = 1,2,\ldots,m. \tag{2.7}$$

where $\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{ik})^T$ are the m $k-$ variate observations with the sample mean $\bar{\mathbf{a}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{a}_i$, and where $\hat{S}$ is an estimator for the common variance-covariance matrix, S.

There are several alternatives estimators for the common variance-covariance matrix (S). Here we discus two different alternatives. The first choice is to consider $\hat{S}$ as the

pooled sample variance-covariance matrix estimator ($\hat{S}_p$) expressed by

$$\hat{S}_p = \frac{1}{m-1}(\mathbf{a}_i - \bar{\mathbf{a}})^T(\mathbf{a}_i - \bar{\mathbf{a}}). \tag{2.8}$$

A T^2 statistic based on the sample mean vector and sample variance-covariance matrix is widely used but is not very effective in detecting anything more than a single moderately size outlying profile (Vargas, 2003). Its distribution is proportional to the beta distribution for small sample size (Mason and Young, 2002).

Another alternative choice of $\hat{S}$ is one based on successive differences ($\hat{S}_D$), introduced by Hawkins and Merriam (1974) and used by Holmes and Mergen (1993). More recently, Jensen and Birch (2009), and Jensen et al. (2008) used this estimator for Phase I profile monitoring in parametric linear and nonlinear mixed models.

To compute the estimator $\hat{S}_D$ we let

$$\hat{\mathbf{v}}_i = \mathbf{a}_{i+1} - \mathbf{a}_i \qquad\qquad i = 1, 2, \ldots, m-1.$$

we then stack the transpose of these $(m-1)$ successive difference vectors into the $(m-1)\times n'$ matrix $\hat{V}$ as

$$\hat{V} = \begin{bmatrix} \hat{\mathbf{v}}_1^T \\ \hat{\mathbf{v}}_2^T \\ \vdots \\ \hat{\mathbf{v}}_{m-1}^T \end{bmatrix}.$$

The estimator of the variance-covariance matrix is

$$\hat{S}_D = \frac{\hat{V}^T\hat{V}}{2(m-1)}. \tag{2.9}$$

Sullivan and Woodall (1996) showed that using successive differences is effective in detecting sustained step changes in the production process that occur in Phase I data. While the distribution of the T^2 statistic based on successive difference does not have a known form, its asymptotic distribution is $\chi^2_{(df)}$, which is a Chi-square distribution with df is an appropriate degrees of freedom. A T^2-chart can be obtained by plotting the T_i^2 statistics, $i = 1, 2, \ldots, m$, versus i and out-of-control signals will be given for any T_i^2 value exceeding an upper control limit (UCL).

To determine the statistical properties of the T^2-chart it is often assumed that each of the $\mathbf{a}_i$ vectors follows a multivariate normal distribution, $\mathbf{a}_i \sim MN(\mu, \sum)$, with mean μ and variance-covariance matrix $\sum$. This assumption is critical for finding the marginal distribution of T_i^2. Under this assumption, the UCL for Phase I data is based on a beta distribution for small sample sizes, while, for large sample sizes it the follows a χ^2 distribution asymptotically. See Tracy et al. (1992) and Williams et al. (2006) for more details.

In NP literature, Diggle and Verbyla (1998) focused on the estimation of the variance-covariance matrix for longitudinal data using kernel weighted local linear regression based on the variogram and squared residuals. Ruppert et al. (1997) introduced a new way to estimate the variance-function using local polynomial regression in the presence of heteroscedasticity. The present study is the first to consider estimating the variance-covariance matrix using the successive difference approach for the NP profile monitoring.

Standard fixed effect parametric methods are straightforward and easily implemented for profile monitoring, but their use can be problematic. If the parametric profile model is misspecified, the estimates and predictions (and hence the fits) may be biased. One way to reduce this bias is to estimate the profile model nonparametrically. Our NP methods are penalized spline (p-spline) regression and local polynomial regression. For the latter, we especially focus only on local linear regression (LLR).

2.3 P-spline Estimation for Fixed Effects Profiles

In recent years, penalized spline regression (often referred to as p-spline regression) has received renewed attention as a powerful alternative smoothing method. Originally, p-spline regression was suggested by O'Sullivan (1986). The flexibility of p-spline regression and utility of p-spline when applied to a large range of modeling contexts has been demonstrated by Eilers and Marx (1996) and more recently through the book by Ruppert et al. (2003). We refer to Ruppert et al. (2003) for an overview of applications of p-spline to different settings and for discussion of the close connections between p-spline regression with linear mixed (LM) models, discussed further in Wand (2003).

The ability to combine NP regression and mixed model regression with p-spline has recently been used in other contexts. Parise et al. (2001), Coull et al. (2001a), and Coull et al. (2001b) all provide examples of using p-splines in the construction of mixed effect regression models for the analysis of data containing random effects.

The main idea of p-spline regression is to fit the function $f(x_{ij})$ parametrically with a sufficiently flexible spline basis. Instead of simple parametric estimation, however, a penalty is imposed on the spline coefficients to achieve a smooth fit. One technical benefit of this approach is that it reveals a link to LM models, which can be very useful in different applications, see Wand (2003).

The resulting affinity to LM models is advantageous and can be exploited in various ways. In particular, the smoothing or penalty parameter required by p-spline regression can be expressed as the ratio of variances in the mixed model which suggests the application of maximum likelihood (ML) theory for estimation.

P-spline regression is used to estimate an unknown smooth function. The smooth function is represented in a high dimensional spline basis with spline coefficients estimated in a penalized form. P-spline coefficients can be viewed as the best linear unbiased predictions ($blups$) in a mixed model framework, which allows the use of mixed model software for smoothing.

In the following section, we provide a review of p-spline smoothing for NP regression. In addition, we present a framework for LM models and the relationship between p-spline and LM models. We introduce p-spline regression for estimating PA and CS fixed effects profiles in (2.3.5) and (2.3.6), respectively. In (2.3.7), we introduce T^2 statistics for p-spline fits.

2.3.1 Penalized Spline

Splines are generally defined as curves which consist of individual segments which are joined smoothly. The segments are given by polynomials and the locations at which they join are referred to as *knots*. The curve takes the shape which minimizes the energy required for bending, thus giving the smoothest shape possible. Similarly, bivariate

splines are commonly called "thin plate" splines, referring to a thin, bendable sheet of metal used to approximate the surface. There are a variety of different spline functions, and their use is discussed in a large body of literature, see Ruppert et al. (2003) and Karin (2005) for instance.

Consider the model given in (2.1) where the function $f(x_{ij})$ is assumed to be smooth but otherwise unspecified, and needs to be estimated from the dataset. The smooth function could be modeled by a variety of methods including natural cubic splines, B-splines, truncated polynomials, or radial splines.

The main idea of p-spline smoothing is that, $f(x_{ij})$ can be approximated by $f(x_{ij}) \approx C\theta + \delta$, where C is a high dimensional basis chosen in advance. In this form, δ denotes the approximation bias of the spline basis in C. If C is chosen as a sufficiently flexible basis, δ will not contain relevant information and may be dropped from the approximation. That is, we assume the function $f(.)$ to be represented by a high dimensional linear parametric form $C\theta$.

It is convenient to decompose C into a low dimensional component X and a high dimensional component Z (Ruppert, 2002). For instance, $X = (1, x, \ldots, x^p)$ can contain a low dimensional polynomial form while Z is a truncated polynomial basis $Z = [(x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p]$, where $(x)_+^p = x^p$ for $x > 0$ and zero otherwise, and $\kappa_1 < \ldots < \kappa_K$ is a set of K fixed knots. Following Ruppert (2002), K can be chosen to be large but less than the sample size n or $(n - p - 1)$. As a practical choice, Claeskens et al. (2009) suggested choosing K according to $K = min(n/4, 40)$. Alternatively, one may use one of several knot selection routines suggested in Ruppert (2002). In our presentation, to keep the approach simple, K might be fixed using the above rule of thumb for the PA and CS profiles. Once K is chosen, the knots κ_k, $k = 1, 2, \ldots, K$, should be selected to cover the range of $x-$values using quantiles.

In the simplest case, a spline function consists of linear segments. This is often described as *a broken-stick* curve, and is a straight forward extension of the parametric linear regression. Given that K knots, $\kappa_1, \kappa_2, \ldots, \kappa_K$ have been selected, we construct a

linear spline by assuming a parametric regression model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+ + \epsilon_{ij} \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i. \quad (2.10)$$

where β_0 and $(\beta_1, u_1, \ldots, u_k)$ represent the intercept and linear regression coefficients, respectively. This results in a slope of β_1 for the first segment, *i.e.* $\forall\ x_{ij} \leq \kappa_1$, a slope of $\beta_1 + u_1$ for the second segment with $\kappa_1 \leq x_{ij} \leq \kappa_2$, and a slope of $\beta_1 + \sum_{k=1}^{r} u_k$ for the segment between bordered by κ_r and κ_{r+1} and so on.

For splines containing polynomial segments of degree p as an extension for model (2.10), a spline with truncated p power basis is

$$y_{ij} = \beta_0 + \sum_{l=1}^{p} \beta_l x_{ij}^l + \sum_{k=1}^{K} u_{pk} (x_{ij} - \kappa_k)_+^p + \epsilon_{ij} \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i. \quad (2.11)$$

where $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients for the p^{th} order polynomial and $u_{p1}, u_{p2}, \ldots, u_{pK}$ are the coefficients for the truncated polynomial bases, the spline portion of the model. $(x)_+^p$ are known as truncated power functions. Obviously, model (2.10) is just model (2.11) with $p = 1$.

The model (2.11) can yield "very wiggly" estimates of curves, especially if p is low and there are many knots. In addition, the choice of knots can have a substantial effect on estimates. These problems can be alleviated by imposing a penalty to reduce the influence of the spline regression coefficients. This typically yields a smoother curve and the penalty is thus referred to as "roughness" penalty (Ruppert et al., 2003).

2.3.2 P-spline as Penalized Least Squares

The model given by equation (2.10) can be written in a matrix form as

$$\mathbf{y} = X\beta + Z\mathbf{u} + \epsilon. \quad (2.12)$$

where X is a $(n \times 2)$ matrix of model matrices stacked by profile, and Z is a $(n \times K)$ truncated polynomial basis, are written as follow:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{mn_m} \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}, \quad X_i = \begin{bmatrix} 1\ x_1 \\ 1\ x_2 \\ \vdots\ \vdots \\ 1\ x_{n_i} \end{bmatrix}, \quad Z = \begin{bmatrix} (x_1 - \kappa_1)_+ \ldots\ldots (x_1 - \kappa_K)_+ \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ (x_n - \kappa_1)_+ \ldots\ldots (x_n - \kappa_K)_+ \end{bmatrix},$$

and

$$\beta = [\beta_0,\ \beta_1]^T, \quad \mathbf{u} = [u_1, u_2, \ldots, u_K]^T.$$

Let us define $C = [X\ Z]$, $\theta = (\beta^T, \mathbf{u}^T)$ and assume that $\epsilon \sim MN(0, \sigma_\epsilon^2 I)$. The OLS fit can be written as follows:

$$\hat{\mathbf{y}} = C\hat{\theta}. \tag{2.13}$$

where $\hat{\theta}$ minimizes $(\mathbf{y} - C\theta)^T(\mathbf{y} - C\theta)$. By considering a penalty on the coefficients of the truncated polynomial bases, $\sum_{k=1}^{K} u_k^2 < \rho$, where ρ is a constant, then the minimization problem becomes

$$\begin{aligned} &Minimize \quad (\mathbf{y} - C\theta)^T(\mathbf{y} - C\theta) \\ &Subject\ to \quad \mathbf{u}^T \tilde{D} \mathbf{u} \leq \rho, \end{aligned} \tag{2.14}$$

where $\tilde{D}$ is some appropriate $(K \times K)$ matrix. It can be shown using a Lagrange multiplier argument that equation (2.14) is equivalent to choosing θ to minimize

$$SSE(\theta) = (\mathbf{y} - C\theta)^T(\mathbf{y} - C\theta) + \lambda^2(\theta^T D\theta - \rho). \tag{2.15}$$

where $D = diag(0_{(p+1\times 1)}, \tilde{D}_{K\times 1})$ is a block diagonal matrix with diagonals of 0 indicating that the coefficient vector β is non penalized and $\tilde{D}$ indicating the penalty on the coefficients in $\mathbf{u}$. Note that, $\theta^T D\theta = \mathbf{u}^T \tilde{D} \mathbf{u} = \sum_{k=1}^{K} u_k^2$, for a special case when $\tilde{D}_{K\times 1} = I_{K\times 1}$. In addition, the second term in the right hand side in (2.15) is called a roughness penalty because it penalized fits that are too rough, thus yielding a smoother result for some number $\lambda \geq 0$. If λ is larger than it should be, the resulting estimated curve tends to be smoother than it should be, smoothing away interesting features in the data. On the other hand, if λ is smaller than it should be, the resulting estimated curve tends to contain more details than truly appearing in the true mean function. Therefore, the objective in selecting the proper value of λ is to select a value that allows the fitted curve to contain interesting and important details present in the true mean function and to avoid fitting details in the data that are due solely to random error.

Differentiating (2.15) with respect to θ gives

$$\frac{d(SSE(\theta))}{d\theta^T} = C^T(\mathbf{y} - C\theta) + \lambda^2 D\theta, \tag{2.16}$$

which by setting (2.16) equals zero will result in

$$\hat{\theta} = (C^T C + \lambda^2 D)^{-1} C^T \mathbf{y}. \tag{2.17}$$

Hence, the fitted values for the p-spline regression are given as

$$\hat{\mathbf{y}} = \hat{f}(.) = C(C^T C + \lambda^2 D)^{-1} C^T \mathbf{y}. \tag{2.18}$$

2.3.3 Framework for Mixed Models

A useful form of mixed models can be expressed as

$$E(\mathbf{y}|\mathbf{u}) = f(X\beta + Z\mathbf{u}). \tag{2.19}$$

where $\mathbf{y}$ is a response vector, X and Z represent model matrices, β is a fixed effects vector and $\mathbf{u}$, $\mathbf{u} \sim (0, G)$, is a random effects vector. Here f is a '$link$' function and is evaluated element-wise for vector arguments.

A special case is the LM model where f is the identity function

$$\mathbf{y} = X\beta + Z\mathbf{u} + \epsilon, \tag{2.20}$$

where

$$\begin{bmatrix} \mathbf{u} \\ \epsilon \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right).$$

Laird and Ware (1982) provided an excellent overview of the LM models. The Gaussian mixed model is a special case for LM models, where

$$\begin{bmatrix} \mathbf{u} \\ \epsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right). \tag{2.21}$$

Model (2.20) can be written as

$$\mathbf{y} = C\theta + \epsilon \tag{2.22}$$

where $C = [X\, Z]$ and $\theta = (\beta^T, \mathbf{u}^T)$. One way to derive an estimate of β and the prediction of $\mathbf{u}$ is by maximizing the likelihood of the $(\epsilon, \mathbf{u})$ over the unknowns β and $\mathbf{u}$. This leads to the criterion

$$(\mathbf{y} - C\theta)^T R^{-1} (\mathbf{y} - C\theta) + \mathbf{u}^T G^{-1} \mathbf{u} \tag{2.23}$$

Equation (2.23) shows that estimation of β and prediction of $\mathbf{u}$ involves GLS with a penalty term $\mathbf{u}^T G^{-1}\mathbf{u}$. Differentiating (2.23) with respect to θ and setting this derivative equals to zero, will result in

$$\hat{\theta} = (C^T R^{-1} C + M)^{-1} C^T R^{-1} \mathbf{y} \tag{2.24}$$

where $M = \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} \end{bmatrix}$. By doing some algebra on (2.24), the estimator of β is

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y} \tag{2.25}$$

and the $blups$ are

$$\hat{\mathbf{u}} = G Z^T V^{-1} (\mathbf{y} - X\hat{\beta}) \tag{2.26}$$

where $V \equiv Cov(\mathbf{y}) = ZGZ^T + R$. If other regressor variables are available that need to be included in the model as parametric terms, they can by added into X fixed effects matrix, and, if appropriate, their associated random effects included by adding the variables to the Z matrix.

2.3.4 Relationship Between P-spline and LM Models

Consider the model implied by (2.20), with the Gaussian distributional assumption governing both the error vector ϵ and the random effects vector $\mathbf{u}$. If we assumed that $R = \sigma_\epsilon^2 I$ and $G = \sigma_u^2 I$ then the solution for $\hat{\theta}$ as in (2.24), for the LM model, can be written as

$$\begin{aligned} \hat{\theta} = & \quad (\tfrac{C^T C}{\sigma_\epsilon^2} + \tfrac{1}{\sigma_u^2} D)^{-1} \tfrac{C^T}{\sigma_\epsilon^2} \mathbf{y} \\ = & \quad (C^T C + \tfrac{\sigma_\epsilon^2}{\sigma_u^2} D)^{-1} C^T \mathbf{y} \\ = & \quad (C^T C + \lambda^2 D)^{-1} C^T \mathbf{y} \end{aligned} \tag{2.27}$$

where $\lambda^2 = \frac{\sigma_\epsilon^2}{\sigma_u^2}$. One can see that $\hat{\theta}$ in (2.24) or in (2.27) for LM models has the same form as for p-spline in (2.17) with $\lambda^2 = \frac{\sigma_\epsilon^2}{\sigma_u^2}$. Thus, the LM model is a convenient way to think about p-spline regression. This formulation allows for great flexibility in the mixed model fit since both random and fixed effects can be easily added. In general, the mixed model approach can be utilized to obtain a NP fit using p-spline regression for any form of mixed model including the linear, nonlinear and generalized linear mixed models.

2.3.5 P-spline Estimation for The Population Average Profile

The population average (PA) profile is the overall trend in the population. A parametric representation for the PA profile can be considered as given in (2.1), where $f(x_{ij})$ can be a linear or nonlinear parametric function. This function can be approximated using a linear model with one of several forms of basis functions such as the polynomial basis. The truncated polynomial basis can be used here since their simple mathematical form often provides an excellent approximation to complicated models. In the following, we present p-spline regression as a new technique to estimate curves in the area of profile monitoring and then by the using the estimated parameters and predicted random effects ($eblups$) one can be determined the outlying profiles.

Model (2.1) can be approximated using p-spline regression with the polynomial basis of order p as follows

$$y_{ij} = \sum_{l=0}^{p} \beta_l x_{ij}^l + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^p + \epsilon_{ij} \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i \tag{2.28}$$

where $f(x_{ij})$ is approximated by $\beta_0 + \sum_{l=1}^{p} \beta_l x_{ij}^l + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^p$.

Using the matrix notation, model (2.28) can be written as given in (2.12) as

$$\mathbf{y} = X\beta + Z\mathbf{u} + \epsilon \tag{2.29}$$

where $\mathbf{y}$ represents the $(n \times 1)$ response vector using for estimating the PA profile where $K = K_1$ is the assumed number of knots for PA profile, and $\mathbf{u} \sim MN(0, G)$, where $Cov(\mathbf{u}) \equiv G = \sigma_u^2 I$. The ratio $\sigma_u^2/\sigma_\epsilon^2$ controls the amount of smoothing in estimating $f(x_{ij})$ where σ_ϵ^2 measures the within profile variation. By using the ML or $REML$ to estimate these variance components, one can obtained the estimated parameters and the predictions as given in (2.25) and (2.26), respectively.

Then the estimated PA profile using p-spline regression is

$$\hat{\mathbf{y}}_{PA}^{PS} = X\hat{\beta} + Z\hat{\mathbf{u}} \tag{2.30}$$

where $\hat{\mathbf{y}}_{PA}^{PS}$ represents the p-spline estimation for the PA profile.

2.3.6 P-spline Estimation for Cluster Specific Profiles

Assuming that nature's model for the i^{th} profile is given by

$$\mathbf{y}_{cs,i} = f_i(x_{ij}) + \epsilon_i \qquad i = 1,2,\ldots,m \tag{2.31}$$

where $\mathbf{y}_{cs,i}$ is a $(n_i \times 1)$ vector of response for the i^{th} profile, with $\epsilon_i \sim MN(0, \sigma^2_\epsilon I)$, and $f_i(x_{ij})$ is unknown function. The $f_i(x_{ij})$ can be approximated using p-spline regression with a polynomial basis of order p as follows

$$f_i(x_{ij}) \approx \sum_{l=0}^{p} \beta_{il} x_{ij}^l + \sum_{k=1}^{K_2} u_{ipk}(x_{ij} - \kappa_k)_+^p \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i, \tag{2.32}$$

where $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iK_2})^T$ is a random effects vector with $\mathbf{u}_i \sim MN(0,\ \sigma^2_{u_i} I)$, and K_2 is the number of knots for the CS curves. In model (2.32), each CS curve has two parts: a parametric component, a p^{th} order polynomial component in our example, $\beta_{i0} + \sum_{l=1}^{p} \beta_{il} x_{ij}^l$, and a spline component, $\sum_{k=1}^{K_2} u_{ipk}(x_{ij} - \kappa_k)_+^p$. Model (2.32) can be described in the mixed model framework for the i^{th} fixed effects profile as

$$\mathbf{y}_{cs,i} = X_i \beta_i + Z_i \mathbf{u}_i + \epsilon_i \qquad i = 1,2,\ldots,m \tag{2.33}$$

where $\beta_i = [\beta_{i0}, \beta_{i1}, \ldots, \beta_{ip}]^T$, $\mathbf{u}_i = [u_{i1}, u_{i2}, \ldots, u_{iK_2}]^T$. In addition

$$X_i = \begin{bmatrix} 1\ x_{i1}\ \ldots\ldots x_{i1}^p \\ 1\ x_{i2}\ \ldots\ldots x_{i2}^p \\ \vdots\ldots\ldots\vdots \\ 1\ x_{in_i}\ \ldots\ x_{in_i}^p \end{bmatrix}, \quad Z_i = \begin{bmatrix} (x_{i1} - \kappa_1)_+^p\ \ldots\ (x_{i1} - \kappa_{K_2})_+^p \\ \vdots\ldots\ldots\ldots\ldots\vdots \\ (x_{in_i} - \kappa_1)_+^p\ \ldots\ (x_{in_i} - \kappa_{K_2})_+^p \end{bmatrix} \tag{2.34}$$

The estimators and the predictors can be obtained as follows

$$\hat{\beta}_i = (X_i^T V_i^{-1} X_i)^{-1} X_i^T V_i^{-1} \mathbf{y}_i \tag{2.35}$$

and the $blups$ are given by

$$\hat{\mathbf{u}}_i = G_i Z_i^T V_i^{-1} (\mathbf{y}_i - X_i \hat{\beta}_i) \tag{2.36}$$

where $V_i \equiv Cov(\mathbf{y}_i) = Z_i G_i Z_i^T + R_i$, with $G_i = \sigma^2_{u_i} I$ and $R_i = \sigma^2_\epsilon I$. Hence, the estimated i^{th} fixed effects profile is given by

$$\hat{\mathbf{y}}^{PSF}_{CS,i} = X_i \hat{\beta}_i + Z_i \hat{\mathbf{u}}_i \qquad i = 1,2,\ldots,m \tag{2.37}$$

where $\hat{\mathbf{y}}^{PSF}_{CS,i}$ is the p-spline regression estimator for the i^{th} fixed effects profile.
To simplify, the above model can be written as

$$\hat{\mathbf{y}}^{PSF}_{CS,i} = C_i \hat{\gamma}_i \qquad i = 1,2,\ldots,m \tag{2.38}$$

with $C_i = [X_i \; Z_i]$ and $\hat{\gamma}_i = [\hat{\beta}_i \; \hat{\mathbf{u}}_i]^T$.

2.3.7 Determine Outlying Profile(s) Based on P-spline Fits

For the fixed effects profile monitoring using p-spline regression we fit a separate p-spline model to each profile to obtain individual profile parameter estimates. In addition, we fit a p-spline regression for the *PA* profile using the entire data set as given in (2.30). We utilize the $\hat{\gamma}_i$ vectors to calculate the T^2 statistics by assuming that the locations of the regressor values and the number of observations at each location is the same across all m profiles (a common occurrence for profile data obtained in industrial settings) and consequently, $K_1 = K_2$.

Once we have obtained the estimates for (2.30) and (2.38), we compute the T^2 statistic to determine if outlying profiles are present. In other words, we employ the multivariate T^2 statistic to evaluate the hypothesis that $f_i(x_{ij})$ is equal to $f(x_{ij})$, $i = 1,2,\ldots,m$, where $f_i(x_{ij})$ represents true CS curve for the i^{th} profile, and $f(x_{ij})$ represents true PA profile. We simply replace $\mathbf{a}_i$ with $\hat{\gamma}_i$ and $\bar{\mathbf{a}}$ with $\hat{\gamma}$ in equation (2.7) to obtain the T^2 statistics for the p-spline fixed effects fits as

$$T^2_{PSF1,i} = (\hat{\gamma}_i - \hat{\gamma})^T [\frac{\sum_{i=1}^{m} (\hat{\gamma}_i - \hat{\gamma})(\hat{\gamma}_i - \hat{\gamma})^T}{m-1}](\hat{\gamma}_i - \hat{\gamma}) \qquad i = 1,2,\ldots,m \tag{2.39}$$

and

$$T^2_{PSF2,i} = (\hat{\gamma}_i - \hat{\gamma})^T [\frac{\sum_{i=1}^{m-1} (\hat{\gamma}_{i+1} - \hat{\gamma}_i)(\hat{\gamma}_{i+1} - \hat{\gamma}_i)^T}{2(m-1)}](\hat{\gamma}_i - \hat{\gamma}) \qquad i = 1,2,\ldots,m \tag{2.40}$$

where $\hat{\gamma} = (\hat{\beta}, \hat{\mathbf{u}})^T$ with $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ for linear p-spline regression case, and $\hat{\mathbf{u}} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_{K_1})^T$ for the *PA* profile as given in (2.30).

Unusual profile(s) can be determined by comparing $T^2_{PSFr,i}$, $r = 1,2$, with a value from χ^2-distribution where the profile will be declared as outlying if $T^2_{PSFr,i} \geq \chi^2_{df,\alpha}$ for $i = 1,2,\ldots,m$, where α represents the significance level, df represents the degrees of freedom which is equal to the $tr(H^{PSF}_{CS,i})$, $H^{PSF}_{CS,i}$ is the smoother matrix for the fixed effects p-spline regression, $H^{PSF}_{CS,i} = C_i(C_i^T C_i + \lambda^2 D)^{-1} C_i^T$.

2.4 *LLR* Model Estimation for Fixed Effects Profiles

Our aim is to present a simple procedure that maintains robustness to a misspecified parametric model while adequately fitting the profiles. In this section, we introduce *LLR* as an alternative *NP* method for profile monitoring.

As in section (2.1), let y_{ij} represent the response variable for the j^{th} observation in the i^{th} profile. Nature's mean function $f_i(x_{ij})$ in (2.31) may be any arbitrary function, perhaps one with no known parametric form for the i^{th} profile.

Using the notation of (2.1), we have the following alternative form

$$\mathbf{y}_i = f_i(\mathbf{x}_i) + \epsilon_i \qquad i = 1,2,\ldots,m \tag{2.41}$$

where $\mathbf{x}_i$ represents a vector of n_i stacked values of the single regressor x.

An estimate for the mean response $f_i(.)$ at a specific point x_0, where x_0 is a value of the regressor, for the i^{th} profile using certain NP methods, such as kernel regression(Ker) or local polynomial regression (LPR), can be given by a weighted sum of responses as

$$\hat{f}_i(x_0) = \sum_{j=1}^{n_i} w_{0j} y_{ij} \qquad i = 1,2,\ldots,m \tag{2.42}$$

where $0 \leq w_{0j} \leq 1$ and w_{0j} indicates the weight assigned to the j^{th} observation in the i^{th} profile for the estimation of $f_i(.)$ at x_0. The weight may be calculated using the Nadaraya-Watson (1964) kernel weight (Waterman et al., 2007), where

$$w_{0j} = \frac{K\left(\frac{(x_0 - x_j)}{h}\right)}{\sum_j^{n_i} K\left(\frac{(x_0 - x_j)}{h}\right)} \tag{2.43}$$

where $K(u)$ is the kernel function, $u = \frac{(x_0 - x_j)}{h}$, and h is the bandwidth (smoothing parameter), $h > 0$, often an unknown parameter.

A kernel function is a decreasing function of $|u|$. Consequently, observations at locations close to x_0 receive larger weight than observations far from x_0. Kernel functions are often based on symmetric density functions. The most common kernels include the normal, uniform, and Epanechnikov. For more kernel functions see Hardle (1990).

The kernel regression estimate at the regressor locations $x_1, x_2, \ldots, x_n$ can be expressed as

$$\hat{f} = \hat{\mathbf{y}} = H^{\text{ker}}\mathbf{y} \tag{2.44}$$

where $H^{\text{ker}} = (\mathbf{h}_1^{\text{ker}^T}\ \mathbf{h}_2^{\text{ker}^T} \ldots \mathbf{h}_n^{\text{ker}^T})^T$ is the kernel "hat" matrix or smoother matrix. The rows of the smoother matrix are $\mathbf{h}_l^{\text{ker}^T} = (w_{l1}\ w_{l2} \ldots w_{ln})$, where w_{lj}, computed from (2.43) with "0" replaced by the "l", is the kernel weight given to the j^{th} response in the prediction at x_l. One disadvantage of kernel regression is that the estimate of the profile at the boundaries is often biased. LPR improves upon the bias shortcoming of the kernel regression (Fan and Gijbels, 1995).

LPR estimates the mean function at x_0, for any arbitrary value of x by fitting locally a d^{th} order polynomial at x_0 as

$$f(x_0) = \beta_{00} + \beta_{10}x_0 + \beta_{20}x_0^2 + \ldots + \beta_d x_0^d \tag{2.45}$$

where $\beta_{00}, \beta_{10}, \beta_{20}, \ldots, \beta_{d0}$ are fixed unknown parameters. The polynomial is commonly of order $d = 1\ or\ 3$ for local linear regression or for local cubic regression, respectively. The polynomial of order $d = 0$ is kernel regression. However, local linear regression or local cubic regressions are often preferred over kernel regression or LPR of even order as LPR of odd order for mean function estimation have less bias than those of even order (Fan and Gijbels, 1995).

The LPR fit at x_0 can be expressed in matrix form as

$$\hat{f}(x_0) = \tilde{\mathbf{x}}_0^T(\tilde{X}^T W_0 \tilde{X})^{-1}\tilde{X}^T W_0 \mathbf{y} = \mathbf{h}_0^{LPR^T}\mathbf{y} \tag{2.46}$$

where $\tilde{X}$ is a model matrix for a d^{th} order polynomial

$$\begin{bmatrix} 1\ x_1\ x_1^2\ \ldots x_1^d \\ 1\ x_2\ x_2^2\ \ldots x_2^d \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ 1\ x_n\ x_n^2\ \ldots x_n^d \end{bmatrix}$$

and $\tilde{\mathbf{x}}_0^T = [1\ x_0\ x_0^2\ \ldots x_0^d]$. Note that $\tilde{X}$ is used here to distinguish this model matrix from X, the model matrix in the parametric model (2.3). The matrices $\tilde{X}$ and X may contain different regressors and hence may not be identical. The matrix W_0 is a diagonal matrix containing the kernel weights associated with x_0

$$W_0 = diag(w_{01}, w_{02}, \ldots, w_{0n}) = < w_{0j} > \qquad j = 1, 2, \ldots, n$$

In LPR, the estimate of the mean response at x_0 is a weighted sum of the response as in the kernel regression. The difference between kernel and *LPR* regression can be seen in their respective weights. The weights at x_0 for the kernel regression are
$\mathbf{h}_0^{ker^T} = (w_{01}\ w_{02}\ \ldots\ w_{0n})$, whereas the weight vector associated with x_0 for LPR is
$\mathbf{h}_0^{LPR^T} = \tilde{\mathbf{x}}_0^T(\tilde{X}^T W_0 \tilde{X})^{-1}\tilde{X}^T W_0$. In matrix form, the LPR estimate at the regressor values $x_1, \ldots, x_n$ can be given as

$$\hat{f}(.) = H^{LPR}\mathbf{y} \tag{2.47}$$

where $H^{LPR} = (\mathbf{h}_1^{LPR^T}\ \mathbf{h}_2^{LPR^T}\ \ldots\ \mathbf{h}_n^{LPR^T})^T$ is the LPR smoother matrix with
$\mathbf{h}_l^{LPR^T} = \tilde{\mathbf{x}}_l^T(\tilde{X}^T W_l \tilde{X})^{-1}\tilde{X}^T W_l$. For more details see Fan and Gijbels (1995).

A widely used and thoroughly investigated NP regression technique is that of local linear regression (*LLR*) estimation due to its straightforward extension to the multivariate case. The NP fits for the *PA* and *CS* curves using the *LLR* estimation at $x_1^*, x_2^*, \ldots, x_{n'}^*$, where $x_l^*,\ l = 1, 2, \ldots, n'$, are an arbitrary set values of x, typically uniformly spaced over the entire range of x, and n' is a number of arbitrary observations for comparing m profiles, are obtained as follows.
The *LLR PA* curve fit is

$$\hat{\mathbf{y}}_{PA,l}^{LLR} = \tilde{\mathbf{x}}_l^T(\tilde{X}^T R_l^{-1}\tilde{X})^{-1}\tilde{X}R_l^{-1}\mathbf{y} = \mathbf{h}_{PA,l}^{LLR^T}\mathbf{y} \tag{2.48}$$

where $\tilde{\mathbf{x}}_l^T = (1\ x_l^*)$, $\mathbf{h}_{PA,l}^{LLR^T} = \tilde{\mathbf{x}}_l^T (\tilde{X}^T R_l^{-1} \tilde{X})^{-1} \tilde{X} R_l^{-1}$ with $R_l = W_l^{-1/2} R W_l^{-1/2}$.

The *LLR CS* fits for the i^{th} profile at x_l^* are given by

$$\hat{\mathbf{y}}_{CS,i,l}^{LLR} = \tilde{\mathbf{x}}_l^T (\tilde{X}_i^T R_{il}^{-1} \tilde{X}_i)^{-1} \tilde{X}_i R_{il}^{-1} \mathbf{y}_i = \mathbf{h}_{CS,i,l}^{LLR^T} \mathbf{y}_i \tag{2.49}$$

where $\mathbf{h}_{CS,i,l}^{LLR^T} = \tilde{\mathbf{x}}_l^T (\tilde{X}_i^T R_{il}^{-1} \tilde{X}_i)^{-1} \tilde{X}_i R_{il}^{-1}$ with $R_{il} = W_l^{-1/2} R_i W_l^{-1/2}$.

The PA and CS profiles may be graphically represented by calculating the fits for many values of x_l^* covering the range of x and then connecting the fits.

2.5 T^2 Statistic for the LLR Fits

For the *LLR* profile situation given in (2.48) and (2.49), we have two scenarios. The first one is to consider all values x_l^* $l = 1,2,\ldots,n'$, where n' represents the number of $x-$values used for comparison of the m profiles. If all n_i are equal and the same $x-$values are used for all m profiles, then $n' = n_i$ for all $i = 1,2,\ldots,m$, and $x_l^* = x_l$, for $l = 1,2,\ldots,n'$, where x_l represents an observed value of the regressors, x. The second scenario is for the case where $n_i's$ are not equal and there are different observed values for x, then we will select the common values for all profiles and add some arbitrary points to obtain n' points. Consequently, for either case, we obtain $\hat{\mathbf{y}}_{PA}^{LLR}$ a $(n' \times 1)$ vector of fits for the *PA* profile $(\hat{f}(x_i))$, common to all profiles, and obtain $\hat{\mathbf{y}}_{CS,i}^{LLR}$ as the $(n' \times 1)$ vector of fits for the i^{th} profile $(\hat{f}_i(x_i))$, $i = 1,2,\ldots,m$.

We employ the multivariate T^2 statistic to assess stability of the n' observation simultaneously, *i.e.* to evaluate the hypothesis that $f_i(x_{ij})$ is equal to $f(x_{ij})$, $i = 1,2,\ldots,m$.

We simply replace $\mathbf{a}_i$ with $\hat{\mathbf{y}}_{CS,i}^{LLR}$ and $\bar{\mathbf{a}}$ with $\hat{\mathbf{y}}_{PA}^{LLR}$ in equation (2.7) to obtain the T^2 statistics for the *LLR* fits as

$$T_{LLR,i}^2 = (\hat{\mathbf{y}}_{CS,i}^{LLR} - \hat{\mathbf{y}}_{PA}^{LLR})^T \hat{V}^{-1} (\hat{\mathbf{y}}_{CS,i}^{LLR} - \hat{\mathbf{y}}_{PA}^{LLR}) \qquad i = 1,2,\ldots,m \tag{2.50}$$

where $\hat{\mathbf{y}}_{CS,i}^{LLR}$ is the *CS* curve fit for the i^{th} profile using *LLR* model. The $\hat{\mathbf{y}}_{PA}^{LLR}$ represents the *PA* curve fit for the *HDS*, $\hat{V}$ is an $n' \times n'$ appropriate estimated variance-covariance matrix for the $\hat{\mathbf{y}}_{CS,i}^{LLR}$ where $\hat{V}$ can be estimated by either $\hat{S}_P$ or $\hat{S}_D$ as given in (2.8) and (2.9), respectively, and replacing $\mathbf{a}_i$ with $\hat{\mathbf{y}}_{CS,i}^{LLR}$ and $\bar{\mathbf{a}}$ with $\hat{\mathbf{y}}_{PA}^{LLR}$.

A large value of $T^2_{LLR,i}$ indicates an unusual $\hat{\mathbf{y}}^{LLR}_{CS,i}$, which implies that the i^{th} profile is out-of-control. In contrast to the parametric profile monitoring case where the estimated model coefficients are used as $\mathbf{a}_i$ as given in (2.7), we utilize the T^2 statistic to monitor the estimate of the mean response at $x^*_1, x^*_2, \ldots, x^*_{n'}$ for the i^{th} profile.

To define large values of $T^2_{LLR,i}$, we need the distribution of $T^2_{LLR,i}$ which depends partially on the distribution of $\hat{\mathbf{y}}^{LLR}_{CS,i,l}$ at x^*_l for every $l = 1,2,\ldots,n'$. We show that this distribution asymptotically the χ^2-distribution (the proof is given in Appendix (*B*)).

The i^{th} profile is unusual if $T^2_{LLR,i} \geq \chi^2_{(\alpha,L_1)}$ for a significance level α. Hastie and Tibshirani (1990) suggested approximating L_1 as $tr(H^{LLR}_{CS,i})$ where $H^{LLR}_{CS,i}$ is the smoother matrix for the i^{th} profile as defined previously.

2.5.1 Algorithm for Profile Monitoring Based on LLR Fits

Assuming a $\hat{S}$ a $(n' \times n')$ is the common variance-covariance matrix for all profiles, then, we propose the following algorithm for NP profile monitoring for fixed-effects models in Phase I analysis.

1. Select a HDS of m profiles, use the LLR model for each profile at x^*_l with $l = 1,2,\ldots,n'$, to obtain $\hat{\mathbf{y}}^{LLR}_{PA}$ and $\hat{\mathbf{y}}^{LLR}_{CS,i}$, and estimate the associated variance-covariance matrix $\hat{S}$ by using the $\hat{S}_p$ or $\hat{S}_D$ estimator.

2. Calculate the T^2 statistic for each profile as given in (2.50), $T^2_{LLR,i}$ $(i = 1,2,\ldots,m)$. Use $Q-Q$ plots and/or a Kolmogorov-Smirnov test to check that $T^2_{LLR,i} \sim \chi^2$-distribution, approximately.

3. Confirm that, there are no unusual profiles in the data set by examining a control chart of the test statistic by comparing $T^2_{LLR,i}$ with the χ^2-value. The profile(s) should be removed if and only if $T^2_{LLR,i} \geq \chi^2_{(\alpha,L_1)}$ for $i = 1,2,\ldots,m$. $\chi^2_{(\alpha,L_1)}$ is called the UCL because it provides an upper bound for plausible values of $T^2_{LLR,i}$ at the α level of significance.

2.6 Chapter Summary

Parametric and NP methods for the fixed effects profile were the topic of this chapter, in which we bridged the gap between p-spline regression and profile monitoring. We briefly presented how p-spline regression can be performed in terms of the mixed models context to take advantage of the existence of software for the mixed model. We introduced p-spline regression and LLR as new techniques to estimate fixed effects profiles. For a misspecified parametric profile model, p-spline regression and LLR offer the user methods to obtain profile fits with less bias than the misspecified parametric profile fits. In addition, we gave diagnostic tools to determine the presence of outlying profile(s) using T^2 statistics.

Although the NP methods may result in less bias for mean response, there is a tendency for over fitting. Model robust regression (MRR), a hybrid combination of the parametric and NP methods, has been shown to minimize the integrated mean square error when compared to the parametric and NP methods, while retaining important features of the data. For these reasons we extend the MRR procedure to the area of profile monitoring in Phase I analysis as described in the following Chapter 3.

Chapter 3

Semiparametric Profile Monitoring for Fixed Effects Models

Model robust profile monitoring ($MRPM$) can be considered as an extension of model robust regression (MRR) to the quality control area. In this chapter, we propose to extend MRR to profile monitoring in Phase I analysis. In addition, we introduce a new method for profile monitoring using the basic idea of MRR for estimating the regression models as a convex combination of the parametric and NP fits via mixing parameter (λ).

The development of $MRPM$ is motivated by the need to improve upon the shortcomings of the parametric method that may result in high bias due to model misspecification. In addition the NP method may introduce high variance of fit possibly resulting from an estimated curve that fits the data too closely.

In $MRPM$, two separate fits are combined to get the final fits for the PA and the m profiles. In $MRPM$, we assume that the user has some information about the underlying profiles from which data have been generated and that a parametric model can be formed that provides a reasonable fit to certain portions of the data but fails to adequately fit the data in other parts. That is, the parametric model has been misspecified. Relying on an NP profile entirely results in loss of information about the profile and possibly subjects the results to highly variable fits.

3.1 Model Robust Regression for Fixed Effects Profiles

Several authors have discussed the semiparametric approach, an approach where parametric and NP information are combined in some way. See (Olkin and Spiegelman, 1987), and (Ke and Wang, 2001) for more details. Olkin and Spiegelman (1987) introduced a semiparametric approach to estimate density functions. Olkin and Spiegelman (1987) combined a parametric and NP estimate by a mixing proportion determined by means of a maximum likelihood approach. Ke and Wang (2001) proposed a class of semiparametric nonlinear mixed effects to combine the advantages of parametric and NP approaches. By using a parametric technique, the parameters give efficient and interpretable data summaries. In an NP approach, the functional form is flexible and is decided by data. Ke and Wang (2001) applied the Laplace method to approximate the log-likelihood function.

Einsporn (1987), Einsporn and Birch (1993), Starnes (1999), Mays et al. (2000), Mays et al. (2001), Waterman (2002), and Waterman et al. (2007) used the basic idea of the semiparametric approach and applied it to linear and LM regression models. Einsporn (1987) and Einsporn and Birch (1993) introduced a semiparametric method for modeling the mean response for the independent, constant error variance case. Their method, model robust regression 1 ($MRR1$), combines a parametric fit to the n observations ($\hat{Y}^P$, a $n \times 1$ vector) and an NP fit ($\hat{Y}^{NP}$) in a convex combination via a mixing parameter $\lambda \in [0,1]$.

The $MRR1$ fit is given as follows

$$\hat{\mathbf{y}}^{MRR1} = (1-\lambda)\hat{\mathbf{y}}^P + \lambda\hat{\mathbf{y}}^{NP} \tag{3.1}$$

From the above model, one can observe that λ should be zero for the correctly specified model where $MRR1$ will reduce to the parametric model. The value of λ should be small if the parametric model is not badly misspecified and be large for a more misspecified model. λ should be close to one for a badly misspecified model. In this case, $MRR1$ will reduce to the NP model. For most practical situations, λ is an unknown and must be chosen as a function of the data.

Mays et al. (2001) showed through a series of Monte-Carlo (MC) simulations that

when using a proper mixing parameter selector, a model specified correctly results in an $MRR1$ estimate equal to or nearly equal to the parametric fit. If the model is badly misspecified, the $MRR1$ estimate is the same or nearly equal to the NP fit. The simulation results also showed that $MRR1$ is superior to a separate parametric model fit or a separate NP model fit under moderate model misspecification in terms of the minimum mean square error of fit criteria.

Previous studies have shown that MRR can often provide a superior fit to either the parametric or the NP fit by using the latter to correct for the inadequacy of a misspecified parametric model fit while retaining the data information contained in the parametric model itself. MRR can be successfully extended to the profile monitoring in Phase I analysis. In general, our $MRPM$ fit for the i^{th} profile is given by

$$\hat{\mathbf{y}}_i^{MRPM} = (1-\lambda)\hat{\mathbf{y}}_i^{P} + \lambda\hat{\mathbf{y}}_i^{NP} \qquad i = 1,2,\ldots,m \tag{3.2}$$

where $\hat{\mathbf{y}}_i^{P}$ represents the parametric fit and $\hat{\mathbf{y}}_i^{NP}$ is the NP fit for the i^{th} profile.

In the following sections, we introduce a new method for Phase I model robust profile monitoring analysis. We name the method "$MRPM$" where both the parametric fits and the NP model fits are obtained only for the fixed effects model.

3.2 MRPM Method for Fixed Effects Models

MRPM is a convex combination of the parametric and NP fits for models containing only fixed effects and no random effects. $MRPM$ can be considered as an extension of $MRR1$ to the area of profile monitoring. The fixed effects model is commonly used in SPC to describe the production process. The fixed effects model for the i^{th} profile as given in (2.41) can be estimated using the MRR technique.

The $MRPM$ fit for the PA profile can be obtained by combining the parametric and NP fits for the PA profile as follows

$$\hat{\mathbf{y}}_{PA}^{MRPM} = (1-\lambda)\hat{\mathbf{y}}_{PA}^{P} + \lambda\hat{\mathbf{y}}_{PA}^{NP} = [(1-\lambda)H_{PA}^{P} + \lambda H_{PA}^{NP}]\mathbf{y} \tag{3.3}$$

where $\hat{\mathbf{y}}_{PA}^{P}$ is the PA fit from the parametric linear fixed model using an appropriate method, such as the OLS method or GLS method, and $\hat{\mathbf{y}}_{PA}^{NP}$ is the p-spline fit or the LLR

fit for the PA curve with $H_{PA}^{P} = X(X^T R^{-1} X)^{-1} X^T R^{-1}$ and H_{PA}^{NP} are the hat matrices for the parametric fit and the smoother matrix for the NP method, respectively. The mixing parameter (λ) is an element between zero and one. A value of $\lambda = 1$ produces an $MRPM$ fit equal to the NP fit; $\lambda = 0$ results in an $MRPM$ fit equal to the parametric fit, while values of $\lambda \in (0,1)$ give $MRPM$ fits that are convex combinations of the other two fits.

The $MRPM$ fits for the i^{th} profile is

$$\hat{\mathbf{y}}_{CS,i}^{MRPM} = (1-\lambda)\hat{\mathbf{y}}_{CS,i}^{P} + \lambda \hat{\mathbf{y}}_{CS,i}^{NP} \qquad i = 1,2,\ldots,m \tag{3.4}$$

where $\hat{\mathbf{y}}_{CS,i}^{MRPM}$ represents the $MRPM$ fit for the i^{th} profile with $\hat{\mathbf{y}}_{CS,i}^{P}$ and $\hat{\mathbf{y}}_{CS,i}^{NP}$ represent the parametric and NP fits for the i^{th} profile, respectively. The $MRPM$ fits should fall between the NP model and the parametric fits since it is a convex combination of them for $\lambda \in (0,1)$.

The $MRPM$ T^2 statistics are given by

$$T_{MRPM,i}^2 = (\hat{\mathbf{y}}_{CS,i}^{MRPM} - \hat{\mathbf{y}}_{PA}^{MRPM})^T \hat{S}^{-1} (\hat{\mathbf{y}}_{CS,i}^{MRPM} - \hat{\mathbf{y}}_{PA}^{MRPM}) \qquad i = 1,2,\ldots,m \tag{3.5}$$

where $\hat{S}$ can be estimated by either the pooled sample variance-covariance matrix ($\hat{S}_p$) or the successive difference estimator ($\hat{S}_D$) for the CS fits, as given in (2.8) and in (2.9). The unusual profile(s) can be determined by comparing $T_{MRPM,i}^2$ to the χ^2-distribution where the profile will be marked as outlying if $T_{MRPM,i}^2 \geq \chi_{(df_1,\alpha)}^2$ for $i = 1,2,\ldots,m$, where α represents the significance level and df_1 represents the degrees of freedom and is equal to $df_1 = tr(H_{CS,i}^{MRPM}) = (1-\lambda)\, tr(H_{CS,i}^{P}) + \lambda\, tr(H_{CS,i}^{NP})$, where $\hat{\mathbf{y}}_{CS,i}^{MRPM}$ can be rewritten as $\hat{\mathbf{y}}_{CS,i}^{MRPM} = \left[(1-\lambda)H_{CS,i}^{P} + \lambda\, H_{CS,i}^{NP}\right] \mathbf{y}_i$.

Our proposed method is expected to give estimates with smaller mean square error of fits especially for misspecified models by using an optimal bandwidth (h) for NP fits and an appropriate mixing parameter value (λ).

3.2.1 Bandwidth Selection

Generally, bandwidth (h), required by the LLR method, is found by minimizing some reasonable optimality criterion as a function of h such as mean squared error, cross-validation (CV), or generalized cross-validation (GCV) as in Hardle and Marron (1985), Loader (1999), and Zhang and Chen (2007).

Mays et al. (2001) introduced the penalized CV technique by incorporating a penalty function into the CV expression. One penalty function, $PRESS^*$, is computed as

$$PRESS^* = \frac{PRESS}{n - tr(H^{NP})}$$

where the $PRESS$ statistic, the CV statistic, is an estimate of the average squared prediction error given by

$$PRESS = \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_{i,-i})^2$$

where $\hat{\mathbf{y}}_{i,-i}$ is the estimate of the model function at x_i with the i^{th} observation removed. The penalty, $n - tr(H^{NP})$, is an adjustment, preventing $PRESS^*$ from being minimized for h that is "too small".

In the cluster correlated data case, it is recommended to remove an entire profile rather than a single observation, since it is the profile that is of interest. H^{NP} represents the smoother matrix for the NP. The elements along the diagonal of H^{NP} are large for small h. This increases $PRESS^*$, which in turn limits the choosing of the bandwidth away from very small values.

Mays et al. (2001) showed that bandwidths chosen by $PRESS^*$ are often too large for the uncorrelated fixed effect model. They introduced $PRESS^{**}$ to overcome the problem of large bandwidths. The $PRESS^{**}$ is given by

$$PRESS^{**} = \frac{PRESS}{n - tr(H^{NP}) + (n-d)[\frac{SSE_{max} - SSE_h}{SSE_{max}}]} \tag{3.6}$$

where SSE_{max} is the largest sum square of errors (SSE) over all possible values of h. SSE_h is the SSE associated with a certain value of bandwidth, and d is the number of the estimated parameters at each x_0 by the NP. The value of $(n-d)[\frac{(SSE_{max} - SSE_h)}{SSE_{max}}]$ approaches zero for large bandwidths. $PRESS^{**}$ guards against overly small and large bandwidths. Therefore, we use $PRESS^{**}$ in our methods to get the optimal bandwidth.

3.2.2 Mixing Parameter Selection

The mixing parameter (λ) is an unknown quantity and should be estimated from the data. The value of λ can be obtained from equation (3.2), ignoring the subscript i, for a

fixed value of h as follows:

$$(\hat{\mathbf{y}}^{MRPM} - \hat{\mathbf{y}}^{P}) = \lambda(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P}).$$

The above expression is similar to a no-intercept simple regression model with a regression coefficient λ. The least square estimator ($\hat{\lambda}$) for λ, can be expressed as

$$\hat{\lambda} = \frac{(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P})^T(\hat{\mathbf{y}}^{MRPM} - \hat{\mathbf{y}}^{P})}{(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P})^T(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P})}.$$

This formula is an implicit expression for $\hat{\lambda}$ and $\hat{\mathbf{y}}^{MRPM}$ and since $\hat{\lambda}$ is needed to compute $\hat{\mathbf{y}}^{MRPM}$, Mays et al. (2001) suggested an explicit formula for $\hat{\lambda}$ as

$$\hat{\lambda} = \frac{(\hat{\mathbf{y}}^{NP}_{i,-i} - \hat{\mathbf{y}}^{P}_{i,-i})^T(\mathbf{y} - \hat{\mathbf{y}}^{P})}{(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P})^T(\hat{\mathbf{y}}^{NP} - \hat{\mathbf{y}}^{P})}, \tag{3.7}$$

where $\hat{\mathbf{y}}^{NP}_{i,-i}$ and $\hat{\mathbf{y}}^{P}_{i,-i}$, respectively, represent the NP and parametric fits of the mean response at x_i without the i^{th} observation.

Waterman et al. (2007) modified the formula (3.7) of $\hat{\lambda}$ for a cluster correlated data. They used $\hat{\mathbf{y}}^{NP}_{i,-i}$ and $\hat{\mathbf{y}}^{P}_{i,-i}$ to indicate NP and parametric fits , respectively, for the i^{th} cluster without the i^{th} cluster.

We formulate the mixing parameter estimation ($\hat{\lambda}$) in the same way as in (Waterman et al., 2007). The $\hat{\lambda}$ formula for the CS fits is expressed as

$$\hat{\lambda}_{CS} = \frac{(\hat{\mathbf{y}}^{NP}_{i,-i} - \hat{\mathbf{y}}^{P}_{i,-i})^T(\mathbf{y} - \hat{\mathbf{y}}^{P}_{CS})}{(\hat{\mathbf{y}}^{NP}_{CS} - \hat{\mathbf{y}}^{P}_{CS})^T(\hat{\mathbf{y}}^{NP}_{CS} - \hat{\mathbf{y}}^{P}_{CS})}, \tag{3.8}$$

and

$$\hat{\lambda}_{PA} = \frac{(\hat{\mathbf{y}}^{NP}_{i,-i} - \hat{\mathbf{y}}^{P}_{i,-i})^T(\mathbf{y} - \hat{\mathbf{y}}^{P}_{PA})}{(\hat{\mathbf{y}}^{NP}_{PA} - \hat{\mathbf{y}}^{P}_{PA})^T(\hat{\mathbf{y}}^{NP}_{PA} - \hat{\mathbf{y}}^{P}_{PA})} \tag{3.9}$$

where $\hat{\lambda}_{CS}$ is used for estimating CS curves and $\hat{\lambda}_{PA}$ is utilized for fitting the PA profile using the $MRPM$ method. We showed the asymptotic properties for the mixing parameter for the PA profile. In addition we proved that, the convergence rate for the entire proposed robust estimate model will converge as quickly as the slowest of the two component estimates. For more details see Appendix (C.1).

3.3 Chapter Summary

In this chapter, we presented a new method for fixed effects profile monitoring using a semiparametric approach by extending MRR to the area of profile monitoring. In addition, bandwidth and mixing parameter selection criterion were presented. Moreover, we presented a formula for a T^2 statistic to detect outlying profile(s). In Chapter 4, we extend this work for mixed models profile monitoring.

Chapter 4

Profile Monitoring for Mixed Models

A mixed model is defined as a model that contains at least one fixed effect and at least two random effects, including the error term. Mixed models are widely used in many disciplines. Henderson (1950), for example, one of the first authors on mixed models, applied them to estimation problems in the animal sciences. Laird and Ware (1982) generalized Henderson's work, and their formulation is commonly known as the Laird-Ware model (Laird and Ware, 1982).

This chapter is presented as follows. In the next section, we briefly review the parametric approach for LM model profile monitoring. In Section 4.2, we extend p-spline regression to mixed model profile monitoring. Section 4.3 gives the conditional local mixed model as an alternative nonparametric technique for fitting mixed profiles and utilizing these fits to determine outlying profiles. Moreover, we extend our $MRPM$ method to mixed models with its corresponding T^2 statistic in Section 4.4. This chapter is summarized in Section 4.5.

4.1 Parametric Mixed Models

Mixed models have received a great deal of attention in the statistical literature in the last few decades. The flexibility, interpretability, and parsimony of linear models enables them to handle a broad variety of problems in such areas as population pharmacokinetics, bioassay, agricultural growth processes, longitudinal count data or repeated binary outcomes, and continuous or discrete repeated measurements data (balanced or unbalanced) (Vonesh et al., 1996; Pinheiro and Bates, 2000).

Mixed models permit two types of correlation for the measurement within a cluster (profile). The first correlation within the cluster comes from the random effects, while the second comes from the within-cluster variance-covariance matrix of errors (R). Mixed models can be either nonlinear or linear mixed models depending on the relationship between the response variable with the fixed and random effects.

4.1.1 Parametric Nonlinear Mixed Models

The nonlinear mixed (NLM) model can be expressed as

$$\mathbf{y}_i = f_i(x_{ij};\, \beta, \mathbf{b}_i) + \epsilon_i \qquad\qquad i = 1, 2, \ldots, m \tag{4.1}$$

where $\mathbf{y}_i$ represents a NLM response vector for the i^{th} profile, β represents a vector of fixed effects common for all profiles with a known X, the design matrix, $\mathbf{b}_i$ is an $(m \times 1)$ vector of random effects, and the function $f_i(.)$ is a known vector-valued (possibly) nonlinear function.

In the following section, assume that nature's model is a LM, as presented in (2.20), for simplification purposes only. The LM model contains at least one fixed effect and two random effects and it is linear in these effects. The LM models can be estimated using the parametric, nonparametric (NP) and semiparametric approaches.

4.1.2 Parametric Estimation for Linear Mixed Models

For the fixed effects models of Chapter 2, the vector $\mathbf{y}$ was defined as the sum of the linear predictor and error, where the linear predictor was comprised of fixed effects. The Laird-Ware model includes an additional linear combination of the random coefficients. The literature on LM models often refers to the collection of data that forms a profile as a "cluster" or "subject" depending on the particular application.

In this research, we use the term profile throughout but note that applications of the methods and analysis presented herein are also applicable if the data are presented by clusters or subjects. The LM model allows us to account for the correlation within profiles and to deal with the collection of profiles as a random sample from a common population distribution of profiles.

Nature's LM model for the i^{th} profile can be given by

$$f_i(x_{ij}) = f(x_{ij}) + \xi_i(x_{ij}) + \epsilon_i \qquad i = 1, 2, \ldots, m \tag{4.2}$$

where $f(x_{ij})$ represent the mean response function for all profiles, the population average (PA), $\xi_i(x_{ij})$ represents the random effects for the i^{th} profile, and $\xi_i(x_{ij}) \sim N(0, \sigma^2_\xi)$.

The LM model approach has several advantages over the linear fixed effects model as pointed out by Verbeke and Molenberghs (2000), Vonesh and Carter (1992), Verbeke and Lesaffre (1996), Pinheiro and Bates (2000), and Demidenko (2004). The LM model can be easily fit for balanced and unbalanced data and often provides a better fit than the linear model approach when the number of observations per profile is small. The LM model approach combines information from the profiles to achieve the model fit with fewer parameters than fitting separate regression function for each profile. Additionally, the LM model approach is capable of handling profiles with missing data.

The user's model can be the Laird-Ware model for the i^{th} profile so that

$$\mathbf{y}_i = X_i\beta + Z_i\mathbf{b}_i + \epsilon_i \qquad i = 1, 2, \ldots, m \tag{4.3}$$

where Z_i is a $(n_i \times q)$ matrix of the predictor variables with random effects, $\mathbf{b}_i \sim MN(0, D)$ is a $(q \times 1)$ vector of random effects for the i^{th} profile, and D is a $(q \times q)$ variance-covariance matrix.

The LM model is flexible in that it allows the errors to be independent or correlated. If correlated, R_i is often assumed to be a simple form such as the autoregressive (AR) form or the compound symmetry form where such forms aim to reduce the number of unknown covariance parameters that require estimation. For more details see Schabenberger and Pierce (2002), Seber and Wild (2003), and Demidenko (2004). Similar structure may be assumed for the D matrix, but typically D is restricted to a diagonal matrix and hence the random effects are assumed to be independent. Also, it is assumed that $Cov(\epsilon_i, \mathbf{b}_i) = O$, where O is a $n \times q$ matrix of zeros.

The conditional CS model, based on a fixed value of $\mathbf{b}_i$, can be given by

$$\mathbf{y}_i | \mathbf{b}_i \sim MN(X_i\beta + Z_i\mathbf{b}_i, R_i) \qquad i = 1, 2, \ldots, m \tag{4.4}$$

Also, the marginal expected value of $\mathbf{y}_i$ can be given as

$$E(\mathbf{y}_i) = E[E(\mathbf{y}_i|\mathbf{b}_i)] = X_i\beta \qquad i = 1,2,\ldots,m \tag{4.5}$$

where $E(\mathbf{y}_i)$ is referred to as the "marginal" mean. The marginal variance of $\mathbf{y}_i$ can be obtained by using the conditional expectation and the conditional variance as follows:

$$Var(\mathbf{y}_i) = R_i + Z_i D Z_i^T = V_i \qquad i = 1,2,\ldots,m \tag{4.6}$$

From the above assumptions, the marginal distribution of $\mathbf{y}_i$ is normal with mean $X_i\beta$ and variance-covariance matrix V_i.

It is convenient to stack the responses and the model matrices for the individual profile. Let $\mathbf{y}$ and X be as given in (2.1). Furthermore, let Z represent a block diagonal matrix of dimension $(n \times mq)$ with Z_i along each diagonal block. The estimator of the fixed effects parameters representing the *PA* coefficients for all the profiles is given by $\hat{\beta}_{LM}$, and prediction of the random deviations from the *PA* is given by $\hat{b}_{LM}$. If $V = diag(V_i)$ is assumed known then it can be shown that

$$\hat{\beta}_{LM} = (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{y} \tag{4.7}$$

and the best linear unbiased predictors ($blups$), in stacked-form, are

$$\hat{\mathbf{b}}_{LM} = BZ^T V^{-1}(\mathbf{y} - X\hat{\beta}) \tag{4.8}$$

where $V = var(\mathbf{y}) = ZBZ^T + R$, with $R = diag(R_i)$ and $B = diag(D)$. It can be shown that $\hat{\beta} \sim MN(\beta, (X^T V^{-1} X)^{-1})$ under the assumption of multivariate normality (Schabenberger and Pierce, 2002) and (Demidenko, 2004).

Usually, V is unknown and therefore must be estimated to compute $\hat{\beta}$ and $\hat{b}$. The matrix V can be estimated by one of several methods, including maximum likelihood (ML) or restricted maximum likelihood ($REML$). Schabenberger and Pierce (2002) prefer $REML$ to estimate V because it produces estimators with less bias than estimators obtained using ML. By substituting the estimated values of V and B, $\hat{V}$ and $\hat{B}$, into (4.7) and (4.8) the parameter estimates and estimated best linear unbiased predictors ($eblups$) are obtained.

The estimated parameter vector for the i^{th} profile ($\hat{\beta}_i^P$) is

$$\hat{\beta}_i^P = \hat{\beta}_{LM} + \hat{\mathbf{b}}_i^* \qquad i = 1,2,\ldots,m \tag{4.9}$$

where $\hat{\beta}_{LM}$ represents the PA coefficients for all the profiles. Equation (4.7) can be rewritten as $\hat{\beta}_{LM} = (\sum_{i=1}^m X_i^T \hat{V}_i^{-1} X_i)^{-1}(\sum_{i=1}^m X_i^T \hat{V}_i^{-1}\mathbf{y}_i)$, and $\hat{\mathbf{b}}_i^*$ is a $(p \times 1)$ vector containing the elements of $\hat{\mathbf{b}}_i = \hat{D}Z_i^T \hat{V}_i^{-1}(\mathbf{y}_i - X_i\hat{\beta}_{LM})$ for the columns of Z_i that are equal to those of X_i and zeros otherwise. Consequently $\hat{\mathbf{b}}_i^* = \hat{\mathbf{b}}_i$ when $Z_i = X_i$.

Jensen et al. (2008) used the parametric approach to determine outlying profiles based on the distance of the estimated parameter vector from the center of the group of estimated parameter vectors. They introduced a formula for the T^2 statistic based on the sample mean vector and the estimated variance-covariance matrix $\hat{V}$ for the estimated parameter vector for the i^{th} profile ($\hat{\beta}_i^P$) based on the successive difference estimator of V. The T^2 statistics are

$$T_{P,i}^2 = (\hat{\beta}_i^P - \bar{\hat{\beta}}^P)^T \hat{V}^{-1}(\hat{\beta}_i^P - \bar{\hat{\beta}}^P) \qquad i = 1,2,\ldots,m \tag{4.10}$$

where $\bar{\hat{\beta}}_i^P = \frac{1}{m}\sum_{i=1}^m \hat{\beta}_i^P$, and $\hat{V} = \frac{1}{2(m-1)}\sum_{i=1}^{m-1}(\hat{\beta}_{i+1}^P - \hat{\beta}_i^P)(\hat{\beta}_{i+1}^P - \hat{\beta}_i^P)^T$.

Sullivan and Woodall (1996) showed that using the successive difference estimator ($\hat{V}$) was effective in detecting sustained step changes in the process that occur in Phase I data. Also, they showed that the distribution of $T_{P,i}^2$ follows asymptotically a χ_p^2 distribution for large samples, where "p" the degrees of freedom for the χ^2-distribution, is the number of estimated parameters.

Since the $\sum_{i=1}^m \hat{\mathbf{b}}_i = 0$, Jensen et al. (2008) modified the above formula (4.10) for the LM model as a function of the $eblups$ ($\hat{b}_i$) as follows:

$$T_{Par2,i}^2 = (\hat{\mathbf{b}}_i)^T \hat{V}^{-1}(\hat{\mathbf{b}}_i) \qquad i = 1,2,\ldots,m \tag{4.11}$$

where $\hat{V} = \frac{1}{2(m-1)}\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1} - \hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1} - \hat{\mathbf{b}}_i)^T$.

In addition, the Hotelling's T^2 statistics can be performed based on the fitted values for the cluster specific ($\mathbf{y}_{CS,i}^P$) and for the population average ($\mathbf{y}_{PA}^P$) as follow

$$T_{Par1,i}^2 = (\hat{\mathbf{y}}_{CS,i}^P - \hat{\mathbf{y}}_{PA}^P)^T \hat{V}^{-1}(\hat{\mathbf{y}}_{CS,i}^P - \hat{\mathbf{y}}_{PA}^P) \qquad i = 1,2,\ldots,m \tag{4.12}$$

where $\hat{\mathbf{y}}^{P}_{CS,i}$ is the CS curve fit for the i^{th} profile at $l = 1,2,\ldots,n'$ using the parametric LM model. The PA curve fit is $\hat{\mathbf{y}}^{P}_{PA}$ for the HDS parametrically and $\hat{V}$ is an $n' \times n'$ appropriate estimated variance-covariance matrix for the $\hat{\mathbf{y}}^{P}_{CS,i}$, such as either $\hat{S}_P$ or $\hat{S}_D$. This research showed that the $T^2_{Par,i}$ statistics based on the estimated random effects (as in Equation (4.11)) and based on the fitted values (as in Equation (4.12)) give the same results and they are equivalent. The proofs are given in Appendices E and F using pooled variance-covariance and successive difference, respectively.

Kang and Albin (2000) introduced the parametric T^2 statistic for simple linear profile monitoring. Jensen et al. (2008) introduced the T^2 statistic for the LM model case. Also, their approach is based on the assumption that the fitted model adequately describes the profile data, which is an unrealistic situation in many practical problems in real life applications.

4.2 Profile Monitoring for Mixed Models using P-spline Regression

In this section, we extend the p-spline regression method to estimate mixed models in the profile monitoring context. Also, we introduce diagnostic tools to determine the presence of outlying profile(s).

4.2.1 P-spline Estimation for Mixed Effects Profiles

The most applicable and flexible models are those that allow for the CS curve differences to be NP functions. See Ruppert et al. (2003), and Wegener and Kauermann (2008) for more details. Recall nature's LM model CS expressed by (4.2)

$$y_{ij} = f(x_{ij}) + \xi_i(x_{ij}) + \epsilon_{ij} \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i \tag{4.13}$$

where y_{ij} is a response variable for the j^{th} observation on the i^{th} profile, with $\epsilon_{ij} \sim N(0,\sigma^2_\epsilon)$, $f(x_{ij})$ represents the overall PA profile and $\xi_i(x_{ij})$ is the i^{th} profile smoother function, representing the random difference between the i^{th} CS curve and the PA curve. Both $f(x_{ij})$ and $\xi_i(x_{ij})$ can be approximated by a p-spline regression.

For example, the polynomial basis of order p can be used to approximate $\xi_i(x_{ij})$, other basis can be utilized as well, as

$$\xi_i(x_{ij}) \approx b_{i0} + \sum_{l=1}^{p} b_{il} x_{ij}^l + \sum_{k=1}^{K_2} t_{ipk}(x_{ij} - \kappa_k)_+^p \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n_i \tag{4.14}$$

where p is the order of the polynomial basis with $(b_{i0}, b_{i1}, \ldots, b_{ip})^T \sim MN(0, \sum_b)$, and $t_{ipk_2} \sim N(0,\ \sigma_t^2)$, for i^{th} profile, $k = 1,2,\ldots,K_2$, and K_2 is the number of knots for the CS curves. In model (4.13), each CS profile has four parts: a PA component ($f(x_{ij})$) and the "difference between the i^{th} CS curve and the PA curve" component. And, each of these components can be approximated by p-spline regression curves consisting of a parametric component and a spline component. For example, the PA curve can be approximated by (2.28) or, in matrix form by (2.29). The random "difference" component can be approximated by (4.14), where in model (4.14), the i^{th} profile has two parts: the random parametric component is $b_{i0} + \sum_{l=1}^{p} b_{il} x_{ij}^l$, and the random spline component is $\sum_{k=1}^{K_2} t_{ipk}(x_{ij} - \kappa_k)_+^p$. The coefficient in both components are random.

Utilizing the relationship between mixed models and p-spline regression, the approximation to model (4.13) can be described succinctly in the mixed model framework for the i^{th} profile as

$$\mathbf{y}_{CS,i} = X_i\beta + Z_i\mathbf{u} + X_i\mathbf{b}_i + E_i\mathbf{t}_i + \epsilon_i \qquad i = 1,2,\ldots,m \tag{4.15}$$

where $\beta = [\beta_0, \beta_1, \ldots, \beta_p]^T$, $\mathbf{u} = [u_1, u_2, \ldots, u_{K_1}]^T$, $\mathbf{b}_i = [b_{i0}, b_{i1}, \ldots, b_{ip}]^T$, and $\mathbf{t}_{ip} = [t_{ip1}, t_{ip2}, \ldots, t_{ipK_2}]^T$. It is not necessary for K_1, the number of knots for the PA profile, to be equal to K_2. X_i and Z_i matrices are the same as given in (2.34). In addition we have

$$E_i = \begin{bmatrix} (x_{i1} - \kappa_1^i)_+^p \ \ldots \ (x_{i1} - \kappa_{K_2}^i)_+^p \\ \vdots \qquad\qquad\qquad \vdots \\ (x_{in_i} - \kappa_1^i)_+^p \ \ldots \ (x_{in_i} - \kappa_{K_2}^i)_+^p \end{bmatrix}.$$

The above model (4.15) also can be written in a stacked matrix notation as follows

$$\mathbf{y} = X\beta + Z\mathbf{B} + \epsilon \tag{4.16}$$

where

$$Z = \begin{bmatrix} Z_1 & X_1 & 0 & \dots & 0 & E_1 & \dots & \dots & 0 \\ Z_2 & 0 & X_2 & \dots & 0 & 0 & E_2 & \dots & 0 \\ \vdots & \vdots & \dots\dots & & \vdots & \vdots & \dots\dots & & \vdots \\ Z_m & 0 & 0 & \dots & X_m & \dots\dots & \dots\dots & & E_m \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{u} \\ \mathbf{b} \\ \mathbf{t} \end{bmatrix},$$

with

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{K_1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} \mathbf{t}_{1p} \\ \mathbf{t}_{2p} \\ \vdots \\ \mathbf{t}_{mp} \end{bmatrix},$$

and

$$Cov(\mathbf{B}) \equiv G = \begin{bmatrix} \sigma_u^2 I & 0 & 0 \\ 0 & blockdiag_{1\le i\le m}\Sigma_b & 0 \\ 0 & 0 & \sigma_t^2 I \end{bmatrix} \tag{4.17}$$

where σ_u^2 controls the amount of smoothing to estimate $f(x_{ij})$, σ_b^2 measures the between profiles variation, σ_ϵ^2 measures the within profile variation, and σ_t^2 controls the amount of smoothing done to estimate $\xi_i(x_{ij})$. Now, $\hat{\beta}$ and $\hat{B}$ can be obtained by using the formulas in (2.25) and $\hat{B}$ as in (4.18), respectively. For example,

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{b}} \\ \hat{\mathbf{t}} \end{bmatrix} = GZ^T V^{-1}(\mathbf{y} - X\hat{\beta}) \tag{4.18}$$

with $V = ZGZ^T + \sigma_\epsilon^2 I$ with G as in (4.17).

Then the estimated PA curve using p-spline regression is given by

$$\hat{\mathbf{y}}_{PA}^{PS} = X_i\hat{\beta} + Z_i\hat{\mathbf{u}} \tag{4.19}$$

and the estimated CS curve for the i^{th} profile is

$$\hat{\mathbf{y}}_{CS,i}^{PS} = X_i\hat{\beta} + Z_i\hat{\mathbf{u}} + X_i\hat{\mathbf{b}}_i + E_i\hat{\mathbf{t}}_i \qquad i = 1,2,\dots,m \tag{4.20}$$

where $\hat{\mathbf{y}}_{CS,i}^{PS}$ is the p-spline regression estimator for the i^{th} profile response.

4.2.2 Diagnostic for Outlying Profile(s) in Mixed Effects Models

In the model (4.15), there are three main sources of variability (not counting the error term):

- The variability between profiles, represented in (4.20) by σ_b^2 and σ_t^2.
- The deviation from the linear polynomial model for the *CS* curves, as accounted for the spline functions, represented in (4.20) by σ_u^2 and σ_t^2.
- The deviation from the linear polynomial model for the *PA* profile, represented in (4.20) by σ_u^2.

In the following section, we present two steps to determine outlying profile(s) as follows

1. To determine if there are significance differences between the *CS* profiles and the *PA* profile, the *CS* profile effects.
2. If there are significant differences between the *CS* profiles and the *PA* profile, the next step is to determine the presence of outlying profile(s).

Testing for *CS* Profile Effects

Since both of *PA* and *CS* profiles are modeled via random effects in a *LM* model, the absence of one of the effects is characterized by zero value of the corresponding variance component.

A likelihood ratio test (or restricted likelihood ratio test) to test the presence of the *CS* profile effect (the existence of least one profile different from the *PA* profile) is readily constructed. That is, the test can be written as

$$H_{0,b\&t} : \sigma_b^2 = 0 \;\&\; \sigma_t^2 = 0 \quad VS \quad H_{1,b\&t} : \sigma_b^2 > 0 \; or \; \sigma_t^2 > 0$$

We fit the model two times, once with these random effects included, resulting in the likelihood or restricted likelihood value L_{H_1}, and once without the *CS* profile random effects, giving L_{H_0}, the test statistic is

$$L_{LR} = 2(L_{H_1} - L_{H_0}) \tag{4.21}$$

The likelihood or restricted likelihood ratio statistic (L_{LR}) has an asymptotic distribution which is a $\frac{1}{2}(\chi^2_q + \chi^2_{q+1})$, where q is the number of fixed effects under the null hypothesis (Ruppert, 2002; Durbàn et al., 2005; Wegener and Kauermann, 2008). There are two possible actions that result from this test. First, if we reject the null hypothesis then one should determine if the outlying profile(s) are present using T^2 statistic. Secondly, if we do not reject the null hypothesis then we should use our HDS to estimate the control limits for Phase II analysis.

Determine Outlying Profile(s)

For the p-spline regression approach, in the mixed effects models, there are two methods for determining abnormal profiles using T^2 statistic. The first method, the T^2 statistics based on the fitted profile values as

$$T^2_{NP1,i} = (\hat{\mathbf{y}}^{PS}_{CS,i} - \hat{\mathbf{y}}^{PS}_{PA})^T \hat{V}^{-1} (\hat{\mathbf{y}}^{PS}_{CS,i} - \hat{\mathbf{y}}^{PS}_{PA}) \qquad i = 1,2,\ldots,m \tag{4.22}$$

moreover $\hat{\mathbf{y}}^{PS}_{CS,i}$ is the CS curve fit for the i^{th} profile at $l = 1,2,\ldots,n'$ using the mixed p-spline regression model as given in Equation (4.20). The PA curve fit is $\hat{\mathbf{y}}^{PS}_{PA}$ for the HDS nonparametrically as given in Equation (4.19) and $\hat{V}$ is an $n' \times n'$ appropriate estimated variance-covariance matrix for the $\hat{\mathbf{y}}^{PS}_{CS,i}$, such as either $\hat{S}_P$ or $\hat{S}_D$.

The second method, we use the $\hat{\phi}_i$ vectors to calculate the T^2 statistics, where

$$\hat{\phi}_i = \begin{bmatrix} \hat{\mathbf{b}}_i \\ \hat{\mathbf{t}}_i \end{bmatrix} \qquad i = 1,2,\ldots,m \tag{4.23}$$

with $\hat{\mathbf{b}}_i = [\hat{b}_{i0}, \hat{b}_{i1}]^T$ assuming $p = 1$ and $\hat{\mathbf{t}}_i = [\hat{t}_{i1}, \hat{t}_{i2}, \ldots, \hat{t}_{iK_2}]^T$. If the locations of the regressor values and the number of observations at each location is the same across all m profiles (a common occurrence for profile data obtained from designed experiments), then the T^2 statistic can be based on the estimated predicted random effects, as these represent the possible differences in the m profiles. The T^2 statistics will be denoted by $T^2_{NP,i}$ and $T^2_{NP2,i}$ and are respectively,

$$T^2_{NP,i} = (\hat{\phi}_i - \bar{\phi})^T [\frac{\sum_{i=1}^{m}(\hat{\phi}_i - \bar{\phi})(\hat{\phi}_i - \bar{\phi})^T}{m-1}]^{-1}(\hat{\phi}_i - \bar{\phi}) \qquad i = 1,2,\ldots,m \tag{4.24}$$

and

$$T^2_{NP2,i} = (\hat{\phi}_i - \bar{\phi})^T [\frac{\sum_{i=1}^{m-1}(\hat{\phi}_{i+1} - \hat{\phi}_i)(\hat{\phi}_{i+1} - \hat{\phi}_i)^T}{2(m-1)}]^{-1}(\hat{\phi}_i - \bar{\phi}) \qquad i = 1,2,\ldots,m \tag{4.25}$$

where $\bar{\phi} = \frac{\sum_{i=1}^{m}\hat{\phi}_i}{m}$.

We speculate that unusual profile(s) can be determined by comparing $T^2_{NP,i}$, $T^2_{NP1,i}$ and $T^2_{NP2,i}$ with a value from χ^2-distribution where the profile will be marked as outlying if $T^2_{NP,i} \geq \chi^2_{(df,\alpha)}$ for $i = 1,2,\ldots,m$, where α represents the significance level, df represents the degrees of freedom which is equal to the $tr(H^{PS}_{CS,i})$, and $H^{PS}_{CS,i}$ is the smoother matrix for p-spline regression.

4.3 Profile Monitoring for Mixed Models using Conditional Local Mixed Models

In this section, we consider a second NP method, one based on the concept of local weighting applied to the LM models. There are two approaches for localizing. The first approach is the marginal local mixed (MLM) model. The MLM model contains kernel weights in the variance-covariance matrix of ϵ_0 and as a multiplier of the random effects. The second is the conditional local mixed (CLM) model. The CLM assigns kernel weights to the variance-covariance matrix of the observation vector conditional on the random effects. Unlike the parametric approach and CLM models, the MLM model approach provides only an estimate of the PA curve. For more details see Waterman et al. (2007). In this work, we are concerned with comparing the CS curves using the NP mixed model and consequently, NP estimation of the CS curves will be accomplished exclusively by CLM model estimation. The CLM model will be discussed in the next section.

4.3.1 *CLM* Model Estimation for Mixed Effects Profiles

Consider the localized *LM* model at x_0 for the entire HDS

$$\mathbf{y} = \tilde{X}\beta_0^C + \tilde{Z}\mathbf{b}_0^C + \tilde{\epsilon}_0 \qquad i = 1,2,\ldots,m \tag{4.26}$$

where $\tilde{X}$ is a $(n \times p)$ matrix of model matrices stacked by profile, and $\tilde{Z}$ is an $(n \times mq)$ block diagonal matrix with random effects model matrices for each profile on the diagonals. Notice that the regressors in the parametric mixed model may differ from the regressors in the NP mixed model, and so a ($\tilde{}$) represents an NP counterpart. The β_0^C represents a vector of unknown fixed parameters called the "localized fixed effects", and $\mathbf{b}_0$ represents a vector of random effects or the "localized random effects", $\mathbf{b}_0^C \sim MN(0, \tilde{B})$. The $\tilde{\epsilon}_0$ is a $(n \times 1)$ vector of error vectors stacked by profile. Also $\tilde{\epsilon}_0 = K_0^{-1/2}\epsilon_0$, and $\tilde{\epsilon}_0 \sim MN(0, K_0^{-1/2}\tilde{R}K_0^{-1/2})$ where $\tilde{R}$ is a $(n \times n)$ block diagonal matrix, and K_0 is an $(n \times n)$ block diagonal weight matrix , and the j^{th} element of the diagonal block K_0 can be calculated as in (2.43). In addition, the random effects $\mathbf{b}_0$ are assumed to be normally distributed with a mean zero vector and a variance-covariance matrix $\tilde{B}$, which is a block diagonal matrix, $\tilde{B} = diag(D)$, and $Cov(\tilde{\epsilon}_0, \mathbf{b}_0^C) = 0$. The subscript "0" indicates that *LM* model has been localized for obtaining the fit at x_0, and the "C" indicates that the *CLM* model is used. This model (4.26) can be written for the i^{th} profile by including a subscript i on $Y, \tilde{X}, \tilde{Z}, \mathbf{b}_0^C$ and $\tilde{\epsilon}_0$.

In general, the local mixed model is a pointwise fit, obtained for each point x_0 in the range of x. Thus, there exists $\beta_0^C, \mathbf{b}_0^C$ and $\tilde{\epsilon}_0$ for each value of x_0. Furthermore, each profile curve can then be graphically represented by computing the pointwise fit at many different values with $x_0 = x_l$, $l = 1,2,\ldots,n'$, over the range of the data and connecting these many fits with straight line segments. This *NP* approach differs from the parametric linear model fit as given by Ding and Zhou (2006), Jensen and Birch (2009) and Jensen et al. (2008), where the profiles were obtained from a single β and $\mathbf{b}$ for all profiles contained in the *HDS*.

The previous distribution assumptions at x_0 can be expressed by

$$\begin{bmatrix} \tilde{\epsilon}_0 \\ \mathbf{b}_0 \end{bmatrix} \sim MN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_0^{-1/2}\tilde{R}K_0^{-1/2} & 0 \\ 0 & \tilde{B} \end{bmatrix} \right)$$

The estimator $\hat{\beta}_0^C$ and the estimated predictor $\hat{\mathbf{b}}_0^C$ at the point x_0, can be found using Henderson's joint likelihood expression (see Appendix (D.1) for details). The results are

$$\hat{\beta}_0^C = (\tilde{X}^T V_0^{-1} \tilde{X})^{-1} \tilde{X}^T V_0^{-1} \mathbf{y} \tag{4.27}$$

and

$$\hat{\mathbf{b}}_0^C = \tilde{B}\tilde{Z}^T V_0^{-1} (\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \tag{4.28}$$

where $V_0 = K_0^{-1/2} \tilde{R} K_0^{-1/2} + \tilde{Z}\tilde{B}\tilde{Z}^T$. Of course, in practice V_0 must be estimated by $\hat{V}_0$ and used in computing (4.27) and (4.28). Note that, the $'C'$ in (4.27) and (4.28) indicates that β and $\mathbf{b}$ are obtained using the CLM model. In addition, $\hat{\beta}_0^C$ is an unbiased estimator for β_0^C if the local model is correct. That is, $f_i(x_i) = X_i \beta_j^C$. (See Appendix (D.2) for details.)

Therefore, the local polynomial regression (LPR) fit of order d of the PA curve fit at an arbitrary value of the regressor, x_0, can be expressed as

$$\hat{\mathbf{y}}_{PA,0}^C = \mathbf{x}_0^T \hat{\beta}_0^C = \mathbf{x}_0^T (\tilde{X}^T \hat{V}_0^{-1} \tilde{X})^{-1} \tilde{X}^T \hat{V}_0^{-1} \mathbf{y} = \mathbf{h}_{PA,0}^{C^T} \mathbf{y} \tag{4.29}$$

where $\mathbf{x}_0^T = (1\ x_0 \ldots x_0^d)$, and $\mathbf{h}_{PA,0}^{C^T} = \mathbf{x}_0^T (\tilde{X}^T \hat{V}_0^{-1} \tilde{X})^{-1} \tilde{X}^T \hat{V}_0^{-1}$. If x_0 represents a data point, say the j^{th} data point, x_j, then $\mathbf{h}_{PA,0}^C$ becomes $\mathbf{h}_{PA,j}^C$ where j replaces 0 in (4.29), and represents the j^{th} row of the smoother matrix H_{PA}^C, $j = 1, 2, ..., n'$. Consequently, the $n' \times 1$ vector of the PA fits at the n' data points may be expressed as $\hat{\mathbf{y}}_{PA}^C = H_{PA}^C \mathbf{y}$, where $H_{PA}^{C^T} = [\mathbf{h}_{PA,1}^{C^T} \mathbf{h}_{PA,2}^{C^T} .... \mathbf{h}_{PA,n'}^{C^T}]^T$.

The PA profile can be graphically displayed by calculating the PA fits for several values ($l = 1, 2, \ldots, n'$) covering the range of x. Then the fits can be connected to obtain the plot for the PA profile.

Now, the LPR fit of order d of the i^{th} profile at x_l^*, $l = 1, 2, ..., n'$, is given by

$$\hat{\mathbf{y}}_{CS,i,l}^C = \mathbf{x}_l^T \hat{\beta}_l^C + \mathbf{z}_l^T \hat{\mathbf{b}}_{i,l}^C = \hat{\mathbf{y}}_{PA,l} + \mathbf{z}_{i,l}^T D \tilde{Z}_i^T \hat{V}_l^{-1} (\mathbf{y}_i - \tilde{X}_i \hat{\beta}_l^C) = \mathbf{h}_{CS,i,l}^{C^T} \mathbf{y} \tag{4.30}$$

where $\mathbf{b}_{i,l}^C \sim MN(0, V_l^*)$ Appendix (D.2.1) gives more details, and for mean and variance for $\hat{\mathbf{y}}_{CS,i,l}^C$ see Appendix (D.2.2). The $\mathbf{h}_{CS,i,l}^{C^T}$ represents the l^{th} row vector of the local

smoother matrix ($H^C_{CS,i}$), $l = 1,2,...,n'$, which is

$$\begin{bmatrix} \mathbf{x}_1^T(\tilde{X}_i^T\hat{V}_1^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T V_1^{-1} - \mathbf{z}_1^T D\tilde{Z}_i^T\hat{V}_1^{-1}\tilde{X}_i(\tilde{X}_i^T\hat{V}_1^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T V_1^{-1} + \mathbf{z}_1^T D\tilde{Z}_i^T\hat{V}_1^{-1} \\ \mathbf{x}_2^T(\tilde{X}_i^T\hat{V}_1^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T V_1^{-1} - \mathbf{z}_2^T D\tilde{Z}_i^T\hat{V}_2^{-1}\tilde{X}_i(\tilde{X}_i^T\hat{V}_2^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T V_2^{-1} + \mathbf{z}_2^T D\tilde{Z}_i^T\hat{V}_2^{-1} \\ \vdots \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad \vdots \\ \mathbf{x}_{n'}^T(\tilde{X}_i^T\hat{V}_1^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T V_1^{-1} - \mathbf{z}_{n'}^T D\tilde{Z}_i^T\hat{V}_{n'}^{-1}\tilde{X}_i(\tilde{X}_i^T\hat{V}_{n'}^{-1}\tilde{X}_i)^{-1}\tilde{X}_i^T\hat{V}_{n'}^{-1} + \mathbf{z}_{n'}^T D\tilde{Z}_i^T\hat{V}_{n'}^{-1} \end{bmatrix} \tag{4.31}$$

where the row vectors $\mathbf{x}_l^T$, and $\mathbf{z}_l^T$ are the l^{th} rows of the $\tilde{X}$ and $\tilde{Z}$ matrices, and $\hat{V}_l^{-1}$ is the CLM model variance-covariance matrix for estimation at $\mathbf{x}_l$.

The *CS* fits for the i^{th} profile can be graphically displayed by calculating the *CS* fits for several values, $l = 1,2,...,n'$, covering the range of x. Then the fits can be connected to display the *CS* profile, $i = 1,2,\ldots,m$. As in the parametric mixed models, the *CS* fits differ from every profile unless the predictor $\hat{\mathbf{b}}_0^C$ is equal to the zero vector for all $\mathbf{x}_0$.

4.3.2 T^2 Statistic Based on CLM Model Fits

We employ the multivariate T^2 statistic to assess stability of the n' observation simultaneously, *i.e.* to evaluate the hypothesis that $f_i(x_i) = f(x_i)$, $i = 1,2,\ldots,m$.

We simply replace $\mathbf{a}_i$ with $\hat{\mathbf{y}}^C_{CS,i}$ and $\bar{\mathbf{a}}$ with $\hat{\mathbf{y}}^C_{PA}$ in Equation (2.7) to obtain the T^2 statistic for the CLM profile fits as follows:

$$T^2_{C,i} = (\hat{\mathbf{y}}^C_{CS,i} - \hat{\mathbf{y}}^C_{PA})^T\hat{V}^{-1}(\hat{\mathbf{y}}^C_{CS,i} - \hat{\mathbf{y}}^C_{PA}) \qquad i = 1,2,\ldots,m \tag{4.32}$$

where $\hat{\mathbf{y}}^C_{CS,i}$ is the *CS* curve fit for the i^{th} profile at $l = 1,2,\ldots,n'$ using CLM model. $\hat{\mathbf{y}}^C_{PA}$ represents the *PA* curve fit for the *HDS* and $\hat{V}$ is an $n' \times n'$ appropriate estimated variance-covariance matrix for the $\hat{\mathbf{y}}^C_{CS,i}$, such as either $\hat{S}_P$ or $\hat{S}_D$. As stated previously, a large value of $T^2_{C,i}$ indicates an unusual $\hat{\mathbf{y}}^C_{CS,i}$, suggesting that the i^{th} profile is out-of-control.

4.3.3 Algorithm for LM Models Profile Monitoring using CLM Model Fits

We assume that $\hat{V}$ is a $(n' \times n')$ common variance-covariance matrix for all profiles. Then, we propose the following algorithm for NP profile monitoring for LM models in

Phase I analysis.

1. Select a HDS of m profiles, and fit the CLM model curves for each profile at x_l^*, $l = 1,2,\ldots,n'$. Use these curves to obtain $\hat{\mathbf{y}}_{PA}^{C}$ and $\hat{\mathbf{y}}_{CS,i}^{C}$, and estimate the associated variance-covariance matrix $\hat{V}$ by using $\hat{S}_p$ or $\hat{S}_D$ estimator.

2. Calculate the T^2 statistic for each profile as in (4.32), $T_{C,i}^2$ $(i = 1,2,\ldots,m)$. Use $Q-Q$ plots and/or a Kolmogorov-Smirnov test to check that $T_{C,i}^2 \sim \chi^2$-distribution, approximately.

3. Confirm that, there are no unusual profiles in the data set by examining a control chart of the test statistic by comparing $T_{C,i}^2$ to an appropriately chosen χ^2-value. We remove those profiles where $T_{C,i}^2 \geq \chi_{(\alpha,L_2)}^2$, where $\chi_{(\alpha,L_2)}^2$ is the UCL at the α level of significance and $L_2 = tr(H_{CS,i}^{C})$.

Despite the possibility of less bias for the mean response resulting from using the NP methods for mixed models, there is a tendency for over fitting as in the fixed effects model. Mixed model robust regression ($MMRR$), a hybrid combination of the parametric LM and NP mixed models, has been shown to minimize the integrated mean square error when compared to the parametric LM and NP mixed method, while retaining important features of the data (Waterman et al., 2007). For these reasons, we extended the $MMRR$ procedure to the area of profile monitoring in Phase I analysis as discussed in the following section. This second approach is named "mixed model robust profile monitoring ($MMRPM$)" and is obtained by incorporating a mixed model approach to both the parametric and the NP model fits.

4.4 Mixed Model Robust Profile Monitoring Method

The proposed $MMRPM$ method is an adaptation of the $MRPM$ method used with the fixed effects models. Also, $MMRPM$ can be considered an extension for $MMRR$ to the area of profile monitoring. Specifically, the $MMRPM$ fit for the PA profile is

$$\hat{\mathbf{y}}_{PA}^{MMRPM} = (1-\hat{\lambda})\hat{\mathbf{y}}_{PA}^{P} + \hat{\lambda}\hat{\mathbf{y}}_{PA}^{NP} \tag{4.33}$$

where $\hat{\mathbf{y}}_{PA}^{P}$ is the PA fit from the parametric LM model using an appropriate method, and $\hat{\mathbf{y}}_{PA}^{NP}$ is an NP mixed model fit for the PA profile using either p-spline regression or CLM Model.

The $MMRPM$ fit for the i^{th} CS profile is

$$\hat{\mathbf{y}}_{CS,i}^{MMRPM} = (1-\hat{\lambda})\hat{\mathbf{y}}_{CS,i}^{P} + \hat{\lambda}\hat{\mathbf{y}}_{CS,i}^{NP} \qquad i = 1,2,\ldots,m \tag{4.34}$$

where $\hat{\mathbf{y}}_{CS,i}^{P}$ is the CS fit from the parametric LM model using an appropriate method for the i^{th} profile, and $\hat{\mathbf{y}}_{CS,i}^{NP}$ is an NP mixed model fit for the i^{th} profile.

The $MMRPM$ T^2 statistic for the separate semiparametric LM models, using the sample mean ($\hat{\mathbf{y}}_{PA}^{MMRPM}$), the CS profiles ($\hat{\mathbf{y}}_{CS,i}^{MMRPM}$), and the estimated variance-covariance matrix ($\hat{V}$) for the CS profiles, is denoted by $T_{MMRPM,i}^2$. The $MMRPM$ T^2 statistics are given by

$$T_{MMRPM1,i}^2 = (\hat{\mathbf{y}}_{CS,i}^{MMRPM} - \hat{\mathbf{y}}_{PA}^{MMRPM})^T \hat{V}^{-1} (\hat{\mathbf{y}}_{CS,i}^{MMRPM} - \hat{\mathbf{y}}_{PA}^{MMRPM}) \qquad i = 1,2,\ldots,m \tag{4.35}$$

where ($\hat{V}$) can be replaced by ($\hat{S}_p$) or ($\hat{S}_D$) as given in (2.8) and in (2.9), respectively, for the CS fits. The unusual profile(s) can be determined by comparing $T_{MMRPM,i}^2$ with the appropriate value from the χ^2-distribution. A profile will be marked as outlying if $T_{MMRPM,i}^2 \geq \chi_{(df_2,\alpha)}^2$ for $i = 1,2,\ldots,m$, where df_2 represents the degrees of freedom which is equal to the $tr(H_{CS,i}^{MMRPM})$, $df_2 = tr(H_{CS,i}^{MMRPM}) = (1-\lambda)\ tr(H_{CS,i}^{P}) + \lambda\ tr(H_{CS,i}^{NP})$, where $H_{CS}^{P} = (I - ZBZ^T\hat{V}^{-1})X(X^T\hat{V}^{-1}X)^{-1}X^T\hat{V}^{-1} + ZBZ^T\hat{V}^{-1}$, and $H_{CS,i}^{NP}$ is the smoother matrix for the NP model, with $H_{CS,i}^{CLM}$ as given in (4.31) for the CLM model.

Another version for the $MMRPM$ T^2 statistics based on a convex combination of the estimated random effects from the parametric and nonparametric approaches. Let $\hat{\psi}_i$ refers to the convex combination of the estimated random effects where

$$\hat{\psi}_i = \begin{bmatrix} (1-\hat{\lambda})\hat{\mathbf{b}}_i \\ \hat{\lambda}\hat{\phi}_i \end{bmatrix} \qquad i = 1,2,\ldots,m \tag{4.36}$$

with $\hat{\mathbf{b}}_i$ is the estimated random effects parametrically as in Equation (4.8) and $\hat{\phi}_i$ represents the estimated random effects via mixed p-spline as given in Equation (4.23). The mixing parameter (λ) can be estimated as in Equation (3.9). Then the $MMRPM$ T^2

statistics can be obtained based on $\hat{\psi}_i$ as these give the possible differences in the m profiles. We refer to as $T^2_{MMRPM2,i}$ which is given as

$$T^2_{MMRPM2,i} = (\hat{\psi}_i - \bar{\psi})^T [\frac{\sum_{i=1}^{m-1}(\hat{\psi}_{i+1} - \hat{\psi}_i)(\hat{\psi}_{i+1} - \hat{\psi}_i)^T}{2(m-1)}]^{-1}(\hat{\psi}_i - \bar{\psi}) \qquad i = 1,2,\ldots,m \quad (4.37)$$

where $\bar{\psi} = \frac{\sum_{i=1}^{m} \hat{\psi}_i}{m}$.

The unusual profile(s) can be determined by comparing $T^2_{MMRPM2,i}$ with a value from χ^2-distribution where the profile will be marked as outlying if $T^2_{MMRPM2,i} \geq ((1-\hat{\lambda}) * UCL_{Par} + \hat{\lambda} * UCL_{NP})$ for $i = 1,2,\ldots,m$, where UCL_{Par} represents the upper control limit for the parametric method, and UCL_{NP} represents the upper control limit for the nonparametric method. The estimated mixed parameter is $\hat{\lambda}$ as in Equation (3.8).

4.5 Chapter Summary

In this chapter, we developed two new methods for mixed effects profile monitoring using the NP and the semiparametric approaches. For the NP approach we introduced the p-spline regression and the CLM model to estimate mixed effects profile fits, and then by utilizing T^2 statistics we can determined the outlying profile(s). Our semiparametric method is the $MMRPM$ as an extension of $MRPM$ to mixed models. In addition, we presented its corresponding T^2 statistic. Chapter 5 gives the results of a Monte-Carlo simulation study to compare the proposed methods to the parametric approach.

Chapter 5

A Monte-Carlo Study

5.1 Introduction

To better understand, compare, and observe the practical performance of the methods developed in this work, they will be evaluated via simulated results from a Monte Carlo study. This is done by generation of data from an assumed model and application of the estimation methods to the generated correlated and uncorrelated data sets.

Statistical properties such as simulated integrated mean square error ($SIMSE$) and simulated probability of signal for the in-control and an out-of-control situations will be obtained. The simulation study will also help identify situations in which the parametric, nonparametric (NP), and mixed model robust profile monitoring ($MMRPM$) methods perform well and when they do not. As mentioned in Chapter 4, mixed effects models treat the collection of profiles as a random sample from a population of profiles and thus accounts for the first stage of the sampling process. The second stage, that of obtaining data within each profile, is represented by the random error terms. Also, Jensen et al. (2008) showed the superiority of linear mixed (LM) models to account for autocorrelation within a linear profile using a parametric approach over fitting separate linear models. Moreover, Jensen and Birch (2009) showed that NLM models could have significant advantages over nonlinear regression models when the data are correlated or even uncorrelated within profiles.

5.2 Simulation Models

Monte Carlo simulation methods are employed to generate a specific number of correlated and uncorrelated data sets and calculates the desired criteria. In the examples studied here, the model is similar to that used in Mays et al. (2001) and Waterman et al. (2007). The data generated from the cluster specific (*CS*) model is

$$y_{ij} = (5+b_{i1})x_{ij} + (2+b_{i2})(x_{ij}-5.5)^2 + \gamma[10\sin(\frac{\pi(x_{ij}-1)}{2.25}) + b_{i3}] + \epsilon_{ij}$$
$$i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n \tag{5.1}$$

where y_{ij} is the simulated response for the j^{th} observation from the i^{th} profile at x_{ij}. The single regressor X takes on integer values from one to ten, inclusive. The random effects are b_{i1}, b_{i2} and b_{i3}, which are generated independently from the normal distribution with mean zero and variance 0.5. The random errors ϵ_{ij} are assumed to follow a normal distribution with mean zero and variance-covariance matrix (R).The variance-covariance matrix R will take on several forms. For the uncorrelated data case, R will equal $\sigma^2 I$, where σ^2 error variance and I is the identity matrix. For the correlated data case, R will take the form of the first-order autocorrelation matrix, $AR(1)$, with $corr(\epsilon_{ij}, \epsilon_{ij+1}) = \rho$.

The population average (*PA*) model can be expressed as

$$y_{ij} = 5x_{ij} + 2(x_{ij}-5.5)^2 + \gamma[10\sin(\frac{\pi(x_{ij}-1)}{2.25})] + \epsilon_{ij} \tag{5.2}$$

The parameter γ is the misspecification parameter. That is, the user will assume that the data are generated from the quadratic model

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{ij} + (\beta_2 + b_{i2})x_{ij}^2 + \epsilon_{ij} \tag{5.3}$$

The random effects are b_{i0}, b_{i1} and b_{i2}, which are generated independently from the normal distribution with mean zero and variance 0.5. The random effect b_{i0} was always estimated as zero from our Monte Carlo simulation. But the true model is the one given in Equation (5.1), where the trigonometric component times γ, the misspecification parameter, serves as the amount of model misspecification. The user's model equals the

true model at γ equals zero, the case where there is no model misspecification. Otherwise, the model has been misspecified, and the degree of misspecification increases with γ.

We consider various amounts of misspecification ranging from γ equal to 0 (no misspecification), 0.25, 0.5, 0.75 and 1. A plot of the *PA* models using different values for γ is given in Figure 5.1. The smooth black line ($\gamma = 0$) occurs when the user's model equals

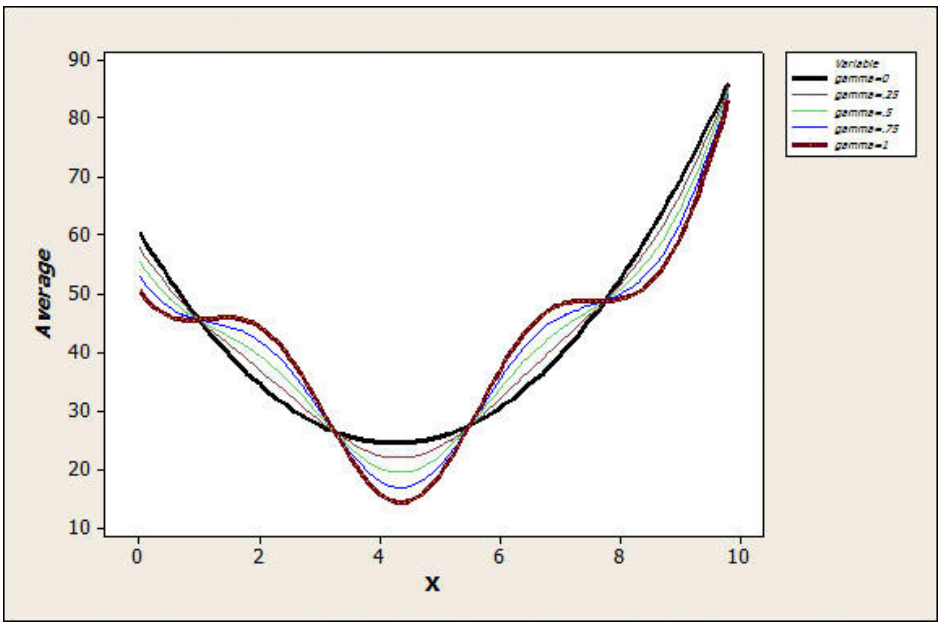

Figure 5.1: Plot of PA underlying models (γ is the misspecification parameter)

the true model. The sold maroon curve represents the most misspecification in the *PA*, when $\gamma = 1$. The larger disparity between the $\gamma = 0$ and $\gamma = 1$ models should be reflected in the simulated integrated mean square error ($SIMSE$) and simulated probability of signal from the simulation study.

Three different variance-covariance structure for the random errors (ϵ_{ij}) were considered. The first structure considered was independence, with the variance of the errors equal to 16 ($i.e.\ \sigma^2_\epsilon = 16$). Therefore, $R = \sigma^2_\epsilon I = 16 * I$. The second covariance structure was a first order auto-regressive ($AR(1)$) model where ρ represents the amount of autocorrelation between error terms for successive equally spaced observations. Here, ρ equals 0.2 as a small autocorrelation. The last covariance structure considered was the

$AR(1)$ model with $\rho = 0.8$ and $\sigma_\epsilon^2 = 16$. Thus, the three variance-covariance structures cover three correlation ranges: uncorrelated (independent), low correlation ($AR(1)$ with $\rho = 0.2$), and the high correlation case ($AR(1)$ with $\rho = 0.8$).

We assumed that there is no misspecification in the variance-covariance structure, which means that, if the random errors are generated from an $AR(1)$ variance-covariance structure, then the parametric model is the quadratic model with an $AR(1)$ structure for R.

In this research, uncorrelated and correlated data sets are generated from different scenarios. The parameter estimation techniques required by the nonparametric (NP) and semiparametric ($MMRPM$) methods are iterative processes. Thus, the fitting of a number of models and computing the probability of signals for each method requires substantial computing resources. To determine the minimum number of required simulation runs that provides sufficient precision of Monte Carlo averages, we examined the standard errors of the simulated mean square errors for the case where the errors are independent and the parametric model correctly specified with $m = 30$ and $n = 10$. In Figure 5.2, each line represents the Monte Carlo standard error of the average $SIMSE$ for the three different methods plotted against the number of Monte Carlo simulation runs of 100, 250, 1,000 and 10,000. The green, the black, and the red lines represent the Monte Carlo standard error of the average $SIMSE$ for the parametric, the NP, and the $MMRPM$ methods, respectively. As expected, the black and the green lines are coincident since the $SIMSE$ from the $MMRPM$ is nearly the same for the parametric method for the correctly specified model since the average of $\hat{\lambda}$, the average of the estimated mixing parameter, is nearly equal to zero, see Table 5.1.

From Figure 5.2, as the number of runs increases, the standard error decreases and levels off. Clearly, more than 1,000 runs could be advantageous, but computing time becomes excessive at this point, as 10,000 runs requires fifty times more computing time as 1,000 runs. Hence, it is determined that 1,000 runs is satisfactory, acknowledging that this will retain Monte Carlo variability closer to the desired level. Also, the absolute relative efficiency based on the Monte Carlo MSE was approximately within 0.2% for 1,000 compared to the 10,000 runs, which still allows a valid comparison of the methods.

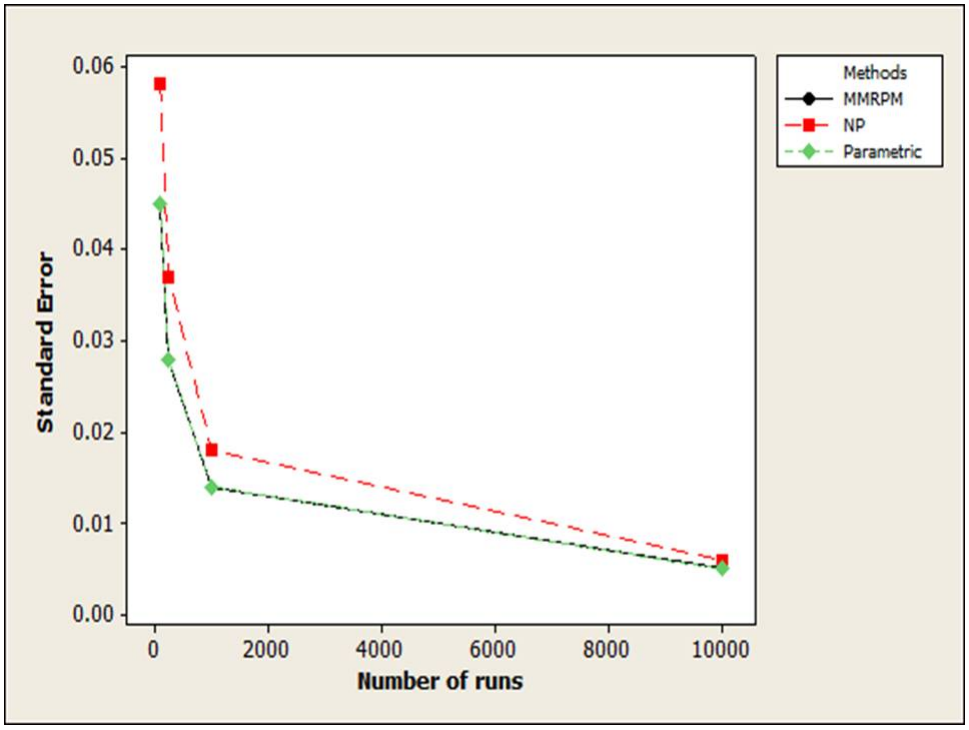

Figure 5.2: Plot of standard error of the average $SIMSE$ versus the number of Monte Carlo simulation runs for the $m = 30$, $n = 10$ case

The parametric, nonparametric, and semiparametric approaches were utilized to fit the cluster specific profiles. The parametric approach used the quadratic model with three random effects. The nonparametric approach used the mixed penalized splines regression model as described in Section 4.2. The semiparametric approach utilized our $MMRPM$ method, combining the parametric and nonparametric models via the mixing parameter λ, as described in Section 4.4.

The general procedure for the simulation study used to compare the parametric, nonparametric, and semiparemtric approaches for profile monitoring is as follows.

Multivariate data that follow a mixed model structure were generated where the random errors and the random coefficients follow some specific structure. This is done by generating univariate normal data. For the correlated data case, the Cholesky decomposition was used to transform the generated univariate data to multivariate data. The data are fit with a second order polynomial mixed model (parametric approach) using proc $GLIMMIX$ from SAS®. The mixed penalized splines regression (nonparametric

approach) was fit using proc $GLIMMIX$ from $SAS^{\textcircled{R}}$. We include both correlated and uncorrelated errors in our comparisons. The $MMRPM$ method fits calculating via a $SAS^{\textcircled{R}}$ macro, written by the author and available from the author upon request.

Six T_i^2 statistics, two each for the parametric, nonparametric and semiparametric methods based on the fitted values and based on the estimated random effects, are computed for each Monte Carlo replication. Once the distribution of each T_i^2 statistic is obtained, an upper control limit (UCL) corresponding to an overall probability of false alarm (α) may be calculated. Hence, we need the joint distribution of the T_i^2 statistics. However, the m T_i^2 values within each method are correlated, since each T_i^2 statistic, ($i = 1, 2, ..., m$) is based on the same estimated mean and variance-covariance matrix, thus making the joint distribution of the T_i^2 value difficult to obtain.

As an alternative, Mahmoud and Woodall (2004), Williams et al. (2006), and Jensen et al. (2008) suggested using an approximated joint distribution assuming the T_i^2 statistic values are independent.

We follow their suggestion here. Let α be the probability of a false alarm for any individual T_i^2 statistic, then the approximate overall probability of a false alarm for a sample of m independent statistics is

$$\alpha_{overall} = 1 - (1 - \alpha)^m$$

Thus, for a given overall probability of a false alarm, we use

$$\alpha = 1 - (1 - \alpha_{overall})^{\frac{1}{m}} \tag{5.4}$$

in calculation of $UCLs$. Mahmoud and Woodall (2004), Williams et al. (2006), and Jensen et al. (2008) found that $UCLs$ based on this approach performed well.

When using the asymptotic chi-squared distribution of each T_i^2, Opsomer et al. (2008) and Jensen et al. (2008), the UCL is given by

$$UCL_{\chi^2} = \chi^2_{(1-\alpha,df)} \tag{5.5}$$

where $\chi^2_{(1-\alpha,df)}$ is the $(1-\alpha)$ quantile of a $\chi^2_{(df)}$ distribution, and df represents the degrees of freedom. The df for the parametric approach is given as the number of random

effects in the estimated parametric model. For the nonparametric approach, the df is the number of random effects plus the number of knots. The simulation results show that using the chi-square distribution as an approximation to the actual distribution of each T_i^2 along with above mention values for the df work very well for the parametric and nonparametric approaches. However, for the $MMRPM$ method, the simulation results show that the chi-square approximation works well when the value of df is adjusted to depend on the degree of correlation existing among the error terms. That is, for independent and low correlated errors ($\rho = 0$ or 0.2), df is calculated as the degrees of freedom for the parametric approach plus the degrees of freedom for the nonparametric approach. For high correlated errors, df is computed as a convex combination of the parametric and nonparametric degrees of freedom, where $\hat{\lambda}$ is the coefficient for the nonparametric degrees of freedom. In this study, the probability of signal for the in-control dataset is 0.05, the nominal value.

The actual probability of signal is calculated as the proportion of simulated datasets where a signal occurred. That is, a signal is given when at least one of the mT^2 statistics exceeds the control limit. For the parametric approach, we consider the quadratic polynomial regression model with two random effects, one for linear and one for the quadratic term. Notice that, the true model, as in Equation (5.1), contains three random effects, but when the user's model includes all three random effects, the third random effect can not be estimated for the data situations considered in our simulation. For the nonparametric approach, the mixed linear penalized splines regression model with one random effect on the slope is considered.

To keep the number of scenarios reasonably manageable, the study considers different values for the number of profiles ($m = 30, 60$), profile size ($n = 10, 20$), $\rho = (0, 0.2, 0.8)$, and varying degrees of misspecification ($\gamma = 0, 0.25, 0.5, 0.75, 1$). For the $n = 10$ and $n = 20$ cases, the design points are selected as equally space along the interval from 1 to 10.

The simulated integrated mean square error ($SIMSE$) was calculated as

$$SIMSE = \frac{1}{m}(\hat{\mathbf{Y}}_{CS,i} - \mu_{CS,i})^T(\hat{\mathbf{Y}}_{CS,i} - \mu_{CS,i}) \tag{5.6}$$

at 46 equally spaced design points (values 1 to ten by 0.2), as an arbitrary number of points picked large enough to detect the differences in curves that may occurred between the minimum and the maximum values for X. The simulated integrated mean square error ($SIMSE$) quantifies the amount by which an estimator ($\hat{\mathbf{Y}}_{CS,i}$), for the i^{th} profile, differs from the true mean response value of the quantity being estimated ($\mu_{CS,i}$). The $SIMSE$ values were calculated for the CS using the parametric, nonparametric, and $MMRPM$ approaches.

In the results that follow, the $SIMSE$ and the probability of signal for the in-control and the out-of-control situations for the uncorrelated and correlated scenarios with different combinations of m, n, and different levels of misspecification were calculated.

5.3 Uncorrelated Data

Our initial study is for an uncorrelated error situation, where $\rho = 0$. We show here the $SIMSE$ and the probability of signal for in-control and out-of-control scenarios with various degrees of misspecified parametric models.

Table 5.1 contains the estimated average mixing parameter ($\bar{\hat{\lambda}}$) and the $SIMSE$ values corresponding to a given degree of misspecification for the independence case. The first and second columns give the number of profiles (m) and profile size (n), respectively. The third column contains the degree of misspecification (γ). The average mixing parameter ($\bar{\lambda}$) is given in the fourth column. Columns five through seven contain the $SIMSE$ from the parametric (Par) using the mixed polynomial regression model, nonparametric (NP) via mixed linear penalized splines regression and semiparametric ($MMRPM$) methods, respectively.

The values in bold represent the smallest $SIMSE$ values. The values in brackets represent the Monte Carlo standard errors using 1,000 replications. All of the results in Table 5.1 were as expected. As the number of observations per profile (n) increases, the $SIMSE$ for each method decreases, as more observations will result in estimates that will be more precise. There is a smaller decrease for the $SIMSE$ from the parametric CS fits especially for the misspecified model. Of most interest is the behavior of $SIMSE$

Table 5.1: $SIMSE$ and average $\hat{\lambda}$ across m, n, and γ. Monte Carlo standard errors in parenthesis. Best values in bold.

m	n	γ	$\bar{\hat{\lambda}}$	Par	NP	MMRPM
30	10	0.00	0.03	2.51 (0.45)	4.04 (0.57)	**2.50** (0.45)
		0.25	0.10	5.45 (0.46)	7.29 (0.65)	**5.40** (0.45)
		0.50	0.57	14.42 (0.53)	12.21 (0.81)	**11.86** (0.61)
		0.75	0.92	29.72 (0.78)	9.58 (0.94)	**9.55** (0.92)
		1.00	1.00	51.71 (1.28)	**11.15** (1.02)	**11.15** (1.02)
30	20	0.00	0.02	**1.25** (0.22)	2.07 (0.32)	**1.25** (0.22)
		0.25	0.20	4.24 (0.25)	5.04 (0.35)	**4.15** (0.25)
		0.50	0.96	13.26 (0.28)	**8.71** (0.46)	**8.71** (0.46)
		0.75	0.98	28.38 (0.37)	**5.44** (0.47)	**5.44** (0.47)
		1.00	1.00	49.73 (0.53)	**6.65** (0.50)	**6.65** (0.50)
60	10	0.00	0.02	**2.50** (0.32)	4.04 (0.42)	**2.50** (0.32)
		0.25	0.10	5.47 (0.35)	7.44 (0.45)	**5.43** (0.33)
		0.50	0.56	14.39 (0.36)	12.20 (0.56)	**11.80** (0.42)
		0.75	0.92	29.65 (0.52)	9.58 (0.87)	**9.57** (0.66)
		1.00	1.00	51.57 (0.85)	**11.16** (0.73)	**11.16** (0.73)
60	20	0.00	0.01	**1.26** (0.17)	2.07 (0.22)	**1.26** (0.17)
		0.25	0.20	4.26 (0.17)	5.03 (0.25)	**4.17** (0.17)
		0.50	0.97	13.27 (0.20)	**8.70** (0.33)	**8.70** (0.33)
		0.75	0.98	28.38 (0.27)	**5.44** (0.35)	**5.44** (0.35)
		1.00	1.00	49.69 (0.39)	**6.61** (0.36)	**6.61** (0.36)

as a function of γ. For $\gamma = 0$, the no model misspecification case, the $SIMSE$ for the parametric and the $MMRPM$ methods are identical (or nearly so).

We see that the average of $\hat{\lambda}$ is near zero and thus the $MMRPM$ method results are identical to the parametric method results (or nearly so). For large values of γ, the most severe model misspecification case, the nonparametric and $MMRPM$ methods are identical (or nearly so). As indicated by the average $\hat{\lambda}$, the $MMRPM$ method is mostly or all composed of the nonparametric method. For intermediate values of γ, the mild model misspecification case, we see that the $MMRPM$ method performs better than either the parametric or the nonparametric method. Thus, the claim that the $MMRPM$ method is robust to model misspecification is supported by this simulation. Additionally, we see that the main advantage of the $MMRPM$ over either the parametric or the nonparametric methods occurs when the user's model is partially correct and

provides a reasonable but not wholly satisfactory fit to the data.

Next, consider the simulated probability of signal for the in-control scenario. The $T^2_{Par1,i}$, $T^2_{NP1,i}$, and $T^2_{MMRPM1,i}$ as in Equations (4.12), (4.23) and (4.35), respectively, are calculated from the parametric polynomial mixed model, the mixed linear penalized splines, and the mixed model robust profile monitoring methods, respectively, using the fitted values for the cluster specific profile at n' observations for each profile. Also, $T^2_{Par2,i}$, $T^2_{NP2,i}$, and $T^2_{MMRPM2,i}$ as in Equations (4.11), (4.25) and (4.37), respectively, are calculated based on the estimated random effects from all three approaches; parametric, nonparametric, and semiparametric, respectively.

Furthermore, we included in this study two versions for estimating the variance-covariance matrix; one using the pooled variance-covariance matrix (V_P) as given in Equation (2.8), and the other based on successive difference variance-covariance matrix (V_D), as given in Equation (2.9). A simulation study, not presented here, showed that the T^2 statistics using V_P are not efficient in detecting the step shift, regardless of the estimating method, whether parametric, or nonparametric, or semiparametric. This result is consistent with those from Sullivan and Woodall (1996), and Vargas (2003). Thus, this simulation study focuses only on the T^2 statistics based on V_D.

Table 5.2 shows the proportion of the 1,000 datasets that had a signal on the control charts for various T^2 statistics for the in-control scenario and correct model specification ($\gamma = 0$). The UCL were calculated based on the approximate chi-squared distributions with the df chosen as described before. The first and second columns give the number of profiles (m) and the number of observations per profile (n), respectively. The third column presents the true autoregressive coefficient values (ρ). Columns four, five and six contain the simulated probability of signal from the T^2 statistics using the fitted values from the parametric ($T^2_{Par1,i}$), the nonparametric ($T^2_{NP1,i}$), and the semiparametric ($T^2_{MMRPM1,i}$) approaches, respectively. The simulated probability of signals from the T^2 statistics based on the estimated parameter values utilizing the parametric ($T^2_{Par2,i}$), the nonparametric ($T^2_{NP2,i}$), and the semiparametric ($T^2_{MMRPM2,i}$) approaches are given in the seventh, eighth, and ninth columns, respectively.

We see from Table 5.2 that for the in-control situation, it appears that the $T^2_{Par1,i}$ and

Table 5.2: Proportion of data sets with a signal for in-control scenario using the chi-squared distribution based on df (degrees of freedom). The nominal value is 0.05.

			T^2 based on the fitted values			T^2 based on the $eblups$		
m	n	ρ	$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
30	10	0.0	0.034	0.037	0.02	0.034	0.037	0.03
		0.2	0.035	0.032	0.01	0.035	0.034	0.02
		0.8	0.034	0.042	0.034	0.034	0.035	0.037
30	20	0.0	0.035	0.04	0.013	0.035	0.049	0.034
		0.2	0.035	0.046	0.01	0.035	0.044	0.024
		0.8	0.036	0.038	0.037	0.036	0.037	0.041
60	10	0.0	0.047	0.054	0.01	0.047	0.053	0.031
		0.2	0.049	0.048	0.003	0.049	0.047	0.019
		0.8	0.047	0.048	0.047	0.047	0.040	0.047
60	20	0.0	0.047	0.044	0.01	0.047	0.043	0.030
		0.2	0.046	0.046	0.004	0.046	0.043	0.010
		0.8	0.039	0.042	0.039	0.039	0.042	0.039

$T^2_{Par2,i}$ statistics for the parametric approach based on the fitted values and the estimated random effects, respectively, give the same probability of signal for the in-control situation. This is true for all values of m, n, and ρ. The results are expected since the two $T^2_{Par,i}$ statistics are identical, as stated in Chapter 4 and proved in Appendix F.

We note that the use of $T^2_{Par1,i}$, $T^2_{Par2,i}$, $T^2_{NP1,i}$ and $T^2_{NP2,i}$ statistics gives probability of signal closer to the nominal α level ($\alpha = 0.05$) and that the statistics based on the semi-parametric approach have smaller probabilities, sometimes much smaller, than those based on the parametric and nonparametric approaches. Furthermore, the $T^2_{MMRPM1,i}$, and $T^2_{MMRPM2,i}$ statistics give probability of signal much smaller than the nominal α level for all values of m. Hence, more work is needed in this area to determine a better approximation for the UCL.

Next we consider the simulated probability of signal for the data that is generated from an out-of-control process for the uncorrelated errors case. The power studies were performed by introducing a step change (δ) in the mean vector, β, with different levels of model misspecification parameter (γ). Since the estimated probability of signal is not always 0.05 for the in-control data, the power studies were based on a simulated control limit to ensure that the probability of signal for in-control data will be the same for all

charts and equal to the nominal 0.05 level. The simulated control limits are given in Table 5.7.

For the generated data for m profiles, the first l of them were generated from the in-control distribution using the model as given in Equation (5.1) with $\beta = (\beta_1, \beta_2)^T = (5,2)^T$, and the last $m-l$ were generated from the same model and same settings of the design factors, except that $\beta = (\beta_1, \beta_2^*)^T = (5, 2+\delta)^T$, where $\delta = 1, 2$, and 4, with $\delta = 1$ referring to a small shift and $\delta = 4$ representing a large shift. Therefore, we have introduced a step change in the mean coefficient vector (for the quadratic term and consequently in the linear term), causing the last $m-l$ profiles to be shifted away from the first l profiles with different sizes for the shift. In addition, several values of l have been tried, and we found that the probability of signal did not depend much on the value of l. Here, we present the results for $l = 20$ when $m = 30$ and $l = 40$ when $m = 60$.

Tables 5.3, 5.4 , 5.5 , and 5.6 give the simulated probability of the out-of-control signal by utilizing our six T^2 statistics for different combinations of m and n. In these tables, the first column gives the degree of misspecification (γ) and the size of the shift is given in the second column. The third through the fifth columns give the simulated probability of signal of the out-of-control situation using the parametric (Par), nonparametric (NP), and semiparametric ($MMRPM$) methods based on the fitted values for the cluster specific profiles, respectively. The simulated probability of signal for the out-of-control scenario obtained via Par, NP, and $MMRPM$ methods based on the estimated random effects ($eblups$) are given in the sixth, seventh, and eighth columns, respectively.

Table 5.3: Simulated probability of signal for out-of-control scenario for independent data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 10$ and $l = 20$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.183**	0.176	0.142	**0.183**	0.117	0.153
	2	**0.569**	0.520	0.429	**0.569**	0.258	0.438
	4	**0.979**	0.961	0.904	**0.979**	0.376	0.837
0.25	1	0.184	0.352	**0.400**	0.184	0.274	**0.454**
	2	0.571	0.715	**0.723**	0.571	0.454	**0.777**
	4	0.979	**0.983**	0.976	**0.979**	0.626	0.971
0.50	1	0.184	**0.610**	0.600	0.184	0.526	**0.697**
	2	0.568	**0.874**	0.861	0.568	0.692	**0.930**
	4	0.983	**0.998**	0.994	0.983	0.921	**0.984**
0.75	1	0.182	**0.951**	0.852	0.182	0.926	**0.950**
	2	0.564	**0.996**	0.968	0.564	0.973	**0.991**
	4	0.978	**1.000**	**1.000**	0.978	0.996	**0.998**
1.00	1	0.185	**0.949**	0.856	0.185	**0.932**	0.834
	2	0.558	**0.996**	0.966	0.558	0.972	**0.992**
	4	0.975	**1.000**	**1.000**	0.975	**0.994**	0.993

Comparison of Table 5.3 to 5.5 and Table 5.4 to 5.6 shows that as the number of profiles (m) increases, the simulated probability of signal of the out-of-control situation increases for all the methods that are presented here for most of the γ and shift values, especially for the moderate and large shifts. The nonparametric method based on the fitted values gives a higher probability of signal than the one based on the estimated random effects, because the number of knots has been selected subjectively, based on the scatter plot for the raw data set.

Table 5.4: Simulated probability of signal for out-of-control scenario for independent data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 20$ and $l = 20$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.183**	0.141	0.132	**0.183**	0.110	0.151
	2	**0.572**	0.474	0.388	**0.572**	0.218	0.413
	4	**0.982**	0.958	0.893	**0.982**	0.323	0.843
0.25	1	0.184	0.310	**0.409**	0.184	0.237	**0.429**
	2	0.570	0.665	**0.751**	0.570	0.393	**0.796**
	4	0.982	0.988	**0.993**	0.982	0.584	**0.996**
0.50	1	0.183	**0.539**	0.401	0.183	**0.441**	0.417
	2	0.566	**0.847**	0.757	0.566	0.634	**0.695**
	4	0.983	**1.000**	**1.000**	0.983	**0.990**	0.986
0.75	1	0.185	**0.919**	0.839	0.185	**0.903**	0.813
	2	0.566	**0.991**	0.977	0.566	**0.958**	0.908
	4	0.982	**1.000**	**1.000**	0.982	**0.991**	0.989
1.00	1	0.188	**0.920**	0.843	0.188	**0.898**	0.876
	2	0.564	**0.993**	0.975	0.564	**0.957**	0.943
	4	0.982	**1.000**	**1.000**	0.982	**0.995**	0.989

Another important observation from the above Tables 5.3, 5.4, 5.5 and 5.6 is that the performances of the $MMRPM$ methods based on either the fitted values or the $eblups$ are much closer to each other and are superior to the parametric and nonparametric methods for the various degrees of model misspecification. In addition, the performance of the nonparametric method gets better as the number of observations per profile increases, since the nonparametric fits become more accurate.

By comparing Tables 5.5 with 5.6, it can be observed that, as n increases, the simulated probability of signal for the out-of-control situation increases over most of the combinations of γ and shift.

Table 5.5: Simulated probability of signal for out-of-control scenario for independent data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 10$ and $l = 40$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.159**	0.130	0.125	**0.159**	0.087	0.134
	2	**0.627**	0.532	0.462	**0.627**	0.225	0.471
	4	**0.999**	0.998	0.982	**0.999**	0.429	0.967
0.25	1	0.159	0.281	**0.363**	0.159	0.206	**0.405**
	2	0.622	0.721	**0.790**	0.622	0.417	**0.849**
	4	**0.999**	0.998	0.998	**0.999**	0.671	**0.999**
0.50	1	0.155	0.492	**0.594**	0.155	0.406	**0.615**
	2	0.620	0.871	**0.912**	0.620	0.640	**0.947**
	4	0.999	0.999	**1.000**	0.999	0.964	**1.000**
0.75	1	0.156	**0.907**	0.812	0.156	0.858	**0.925**
	2	0.621	**0.988**	0.974	0.621	0.948	**0.999**
	4	0.999	**1.000**	**1.000**	0.999	0.999	**1.000**
1.00	1	0.154	**0.900**	0.811	0.154	**0.862**	0.844
	2	0.616	**0.988**	0.983	0.616	**0.955**	0.954
	4	0.999	**1.000**	**1.000**	0.980	**0.999**	0.994

Figures 5.3 and 5.4 show some results of the power studies for the independent dataset using T^2 statistics based on the estimated random effects and the fitted values, respectively. Here $m = 60$ and $n = 10$ where the step change occurred after the $l = 40$ $(\frac{2}{3}m)$ profile.

The figures show that the curves corresponding to the use of the $MMRPM$ method are as good as the parametric approach for the correctly specified model. In other words, the results from using of the semiparametric $(MMRPM)$ and parametric (Par) approaches practically coincide, indicating that the two methods will perform similarly in detecting the step change for the correctly specified parametric model. Furthermore, for the completely misspecified model $(\gamma = g = 1)$, the $MMRPM$ and the nonparametric methods have the same performance. For the small and moderate misspecified models, the semiparametric method is superior to both the parametric and nonparametric methods for any m and n combinations. In addition, the nonparametric method is better than the

Table 5.6: Simulated probability of signal for out-of-control scenario for independent data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 20$ and $l = 40$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.157**	0.135	0.117	**0.157**	0.118	0.147
	2	**0.631**	0.553	0.454	**0.631**	0.268	0.470
	4	**0.999**	0.997	0.993	**0.999**	0.502	0.970
0.25	1	0.156	0.291	**0.403**	0.156	0.264	**0.417**
	2	0.635	0.736	**0.820**	0.635	0.482	**0.877**
	4	0.999	**1.000**	**1.000**	0.999	0.756	**1.000**
0.50	1	0.160	**0.490**	0.407	0.160	0.497	**0.517**
	2	0.630	**0.880**	0.815	0.630	0.711	**0.741**
	4	0.999	**1.000**	**1.000**	**0.999**	0.901	0.996
0.75	1	0.153	**0.904**	0.850	0.153	**0.902**	0.890
	2	0.630	**0.989**	0.982	0.630	**0.979**	0.969
	4	0.999	**1.000**	**1.000**	**0.999**	**0.999**	0.997
1.00	1	0.154	**0.904**	0.850	0.154	**0.904**	0.897
	2	0.628	**0.988**	0.983	0.628	**0.979**	0.971
	4	0.999	**1.000**	**1.000**	**0.999**	**0.999**	0.998

parametric method for the moderate and large misspecification.

Similar simulation results, not shown here, were obtained for different shift sizes and shift locations. We found the conclusions stated here for the out-of-control data will hold, no matter the value of l nor the size for the step change (δ).

Figure 5.5 shows the simulated probability of signal for various combinations of m and n for the T^2 statistics obtained via the parametric method (blue line), the nonparametric method (red line), and the $MMRPM$ method (green line). The horizontal axis is the shift (δ), and the four large panels show results with various combinations of m and n. Within the larger panels are smaller panels which show the different degrees of misspecification. This figure gives the general picture for the power in detecting the step change using our three T^2 statistics based on the fitted values. One can see the same conclusion as mentioned in Figure 5.4.

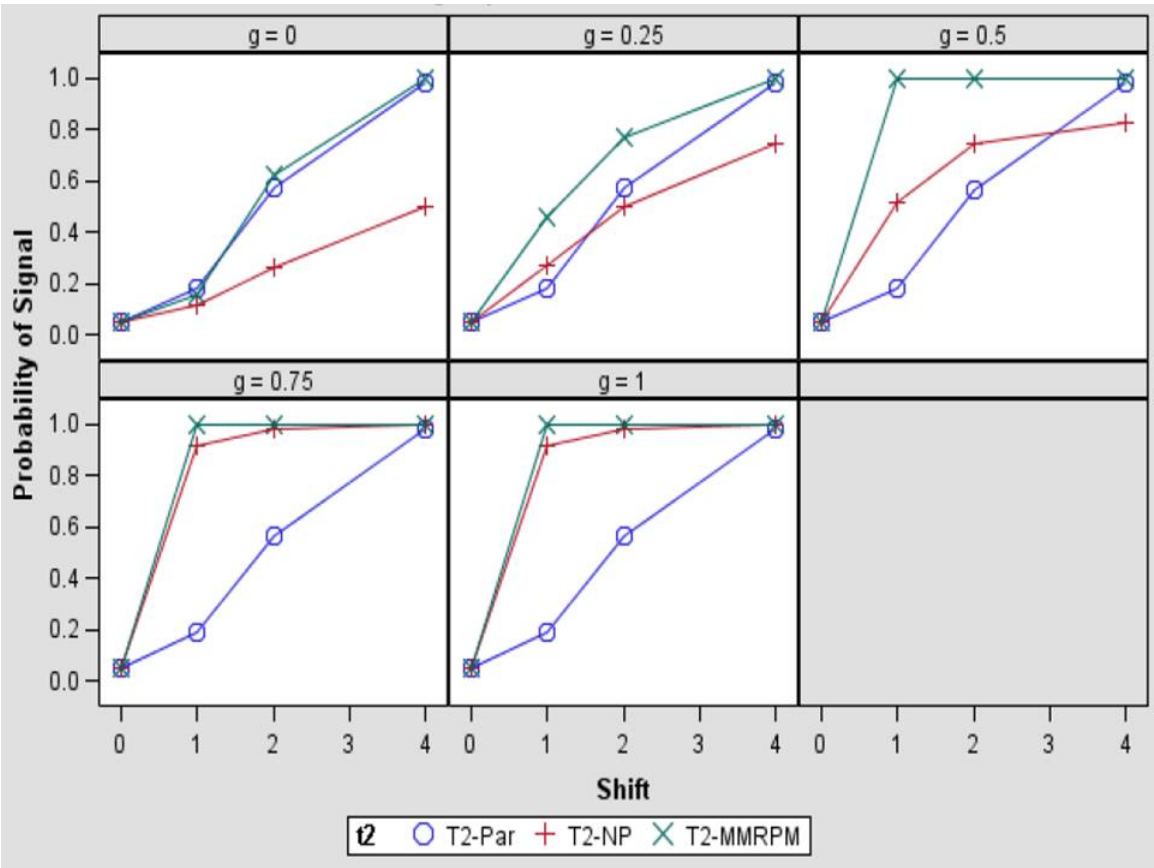

Figure 5.3: Simulated probability of signal for out-of-control scenario for independent data set with $m = 60$ and $n = 10$, for different degrees of misspecification (g), for the T^2- based on $eblups$.

From the simulation study results for the uncorrelated data scenario, we conclude that the method that has the best fits, as measured by $SIMSE$, has the highest probability of signal values. Consequently, the parametric has the highest probability of signal values when γ is zero, regardless of the size of the shift. The $MMRPM$ is competitive with the parametric, but nonparametric is not competitive. When γ is large, $MMRPM$ and nonparametric are very similar in probability of signal values across all values of shift, and values of n, m, and type of T^2. The parametric has very poor probability of signal values, especially for small shifts, being competitive only if the size of the shift is large. When γ is intermediate, $MMRPM$ is clearly superior over nonparametric and parametric, although when the size of the shift is large, all methods give very similar

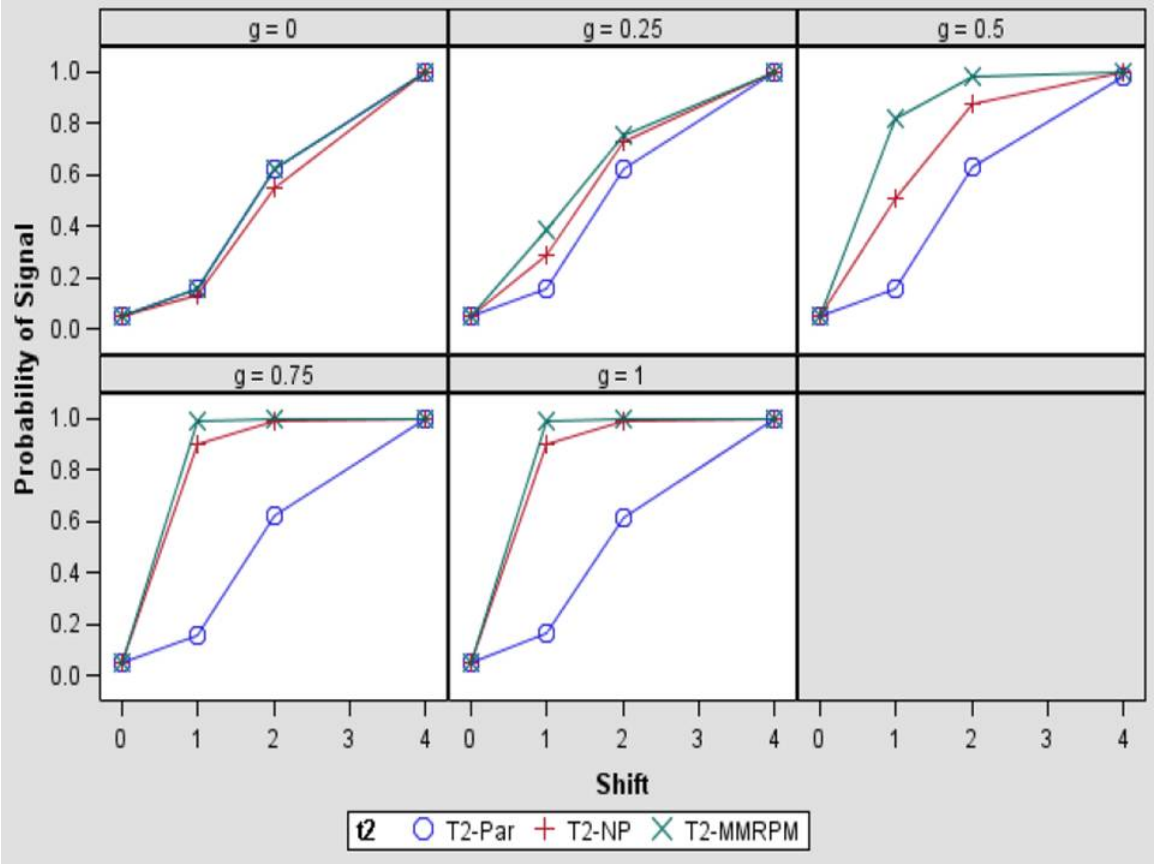

Figure 5.4: Simulated probability of signal for out-of-control scenario for independent data set with $m = 60$ and $n = 10$, for different degrees of misspecification (g), for the T^2- based on the fitted values.

values of probability of signal. Apparently, if the size of the shift is large enough, all these methods work about the same, even if the model fits the data poorly which means that the shift can still be detected. Once again, the main advantage of the $MMRPM$ is its ability to utilize information contained in the parametric model with information picked up by the nonparametric model to work well in terms of fitting the data and well in terms of making decisions about the data regardless of the degree of model misspecification in the parametric model. The impact of changing m or n, over the values we have chosen is minimal compared to the shift effect and the γ effect.

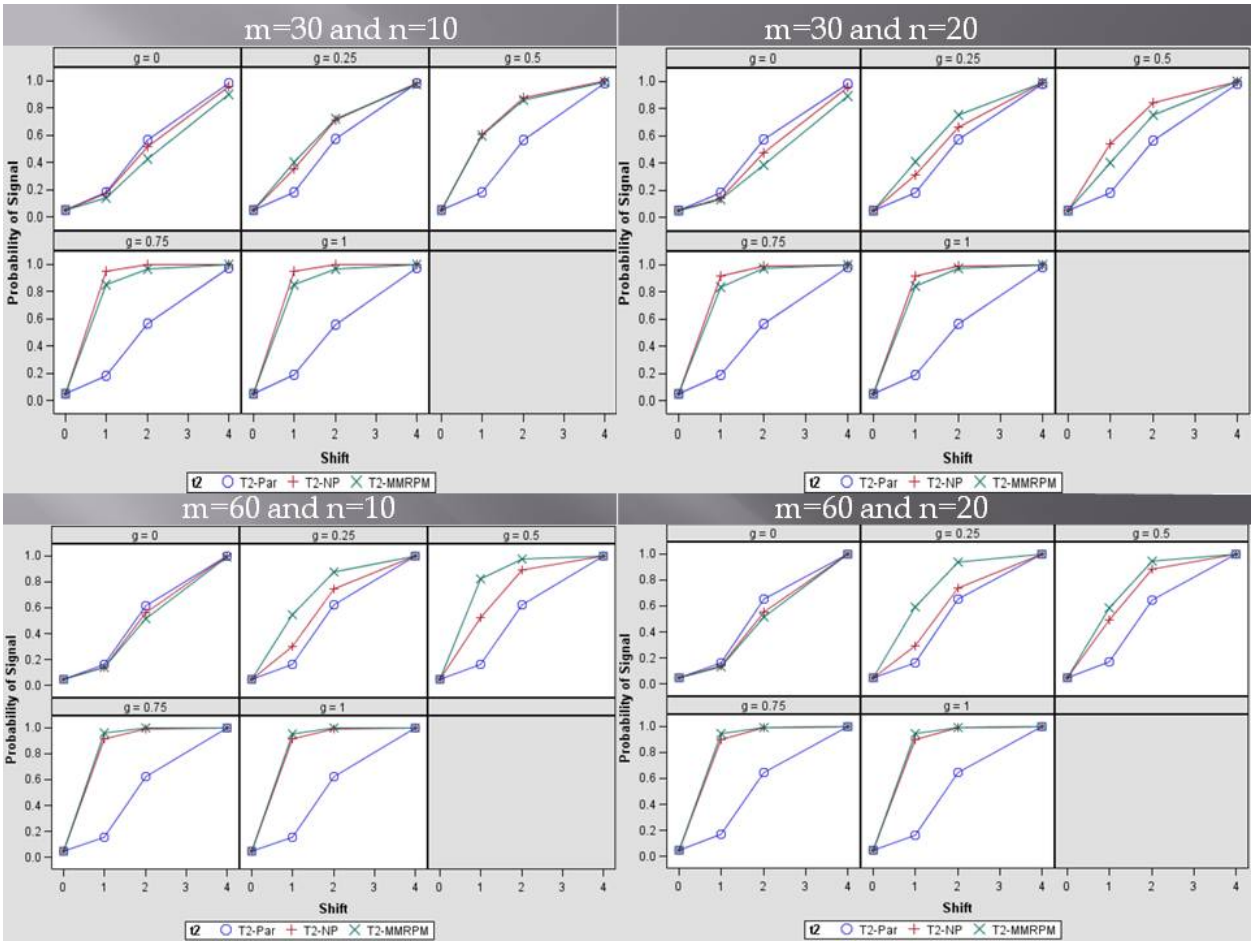

Figure 5.5: Simulated probability of signal for independent data set, for different degrees of misspecification (g) and all m , and n combinations, for the T^2-statistics based on the fitted values.

5.4 Correlated Structure

In this section, the $SIMSE$ and the simulated probability of signal for the out-of-control scenario are estimated for the correlated error structure situation. The $SIMSE$ from estimating cluster specific curves are obtained using the parametric (Par), the nonparametric (NP), and the semiparametric ($MMRPM$) methods. In addition, the simulated probability of signal for the out-of control scenario utilizing both T^2 statistics are calculated. In this study, thirty profiles will be considered our "Small" number of profiles, and sixty profiles our "Large" number of profiles with 10 and 20 observations

Table 5.7: Simulated 95% cutoff values for T^2 statistics using the successive difference variance-covariance matrices.

m	n	ρ	T^2 based on the fitted values			T^2 based on the $eblups$		
			$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
30	10	0.00	11.974207	14.325748	15.779216	11.974207	14.443294	17.939123
		0.20	12.069087	14.420169	15.256804	12.069087	14.516554	16.643221
		0.80	12.091304	14.662319	12.120153	12.091304	14.428173	12.197468
30	20	0.00	12.053110	14.784817	15.874070	12.053110	14.933427	18.282131
		0.20	12.163374	14.977524	14.932630	12.163374	14.780853	16.659744
		0.80	12.145840	14.482509	12.200920	12.145840	14.566426	12.446697
60	10	0.00	14.005429	16.634808	17.612195	14.005429	16.786829	19.842679
		0.20	14.006457	16.380257	15.579720	14.006457	16.513157	17.473977
		0.80	14.064860	16.529242	14.064860	14.064860	16.119014	14.0649
60	20	0.00	14.014220	16.470786	17.142925	14.014220	16.151852	19.701466
		0.20	13.840394	16.414987	15.666122	13.840394	16.242655	17.450160
		0.80	13.707113	16.169692	13.707113	13.707113	16.143457	13.707113

for both cases. Tables 5.8 and 5.9 give the $SIMSE$ and the Monte Carlo standard errors for the correlated errors structure assuming $AR(1)$ with ρ equal 0.2 and 0.8, respectively. The $AR(1)$ with ρ equals 0.2 will be considered our "Low" autocorrelation and ρ equals 0.8 our "High" autocorrelation.

One or two $SIMSE$ values are in bold in each row. These bold values are the minimum cluster specific $SIMSE$ for the given γ value. As expected, the $SIMSE$ decreases for the nonparametric and $MMRPM$ estimations as the number of observations per profile increases. Also, $MMRPM$ estimations have the smallest $SIMSE$. In Tables 5.8 and 5.9, for example, the nonparametric method has the smaller $SIMSE$ values than the parametric method for $\gamma = 0.5$, 0.75 and 1 for different combinations of m and n. The $MMRPM$ method has the smallest $SIMSE$ for all different values of γ. This verifies our observation that the $MMRPM$ method performs at least as well as the parametric and nonparametric methods, especially, for $\gamma = 0$ and 1. These results are very similar to those from the independent data set.

There are some key differences between the uncorrelated and correlated scenarios. On average, the $SIMSE$ values, for correlated error cases, increase as the correlation increases. For example, Tables 5.1, 5.8 and 5.9, in the $m = 30$ with $n = 20$ case, the

Table 5.8: Simulated IMSE and average $\hat{\lambda}$ using correlated data set with $\rho = 0.2$ across m, n, and γ. Monte Carlo standard errors in parenthesis. Best values in bold.

m	n	γ	$\bar{\hat{\lambda}}$	Par	NP	MMRPM
30	10	0.00	0.01	**3.25** (0.45)	5.36 (0.57)	**3.25** (0.45)
		0.25	0.05	6.16 (0.59)	8.68 (0.85)	**6.14** (0.59)
		0.50	0.49	15.00 (0.62)	13.86 (1.03)	**12.96** (0.73)
		0.75	0.89	30.11 (0.81)	11.34 (1.16)	**11.18** (1.10)
		1.00	0.99	51.86 (1.27)	**12.91** (1.23)	12.92 (1.23)
30	20	0.00	0.01	**1.74** (0.33)	2.91 (0.45)	**1.74** (0.33)
		0.25	0.15	4.73 (0.33)	6.00 (0.49)	**4.67** (0.33)
		0.50	0.85	13.72 (0.36)	9.87 (0.62)	**9.85** (0.57)
		0.75	0.95	28.79 (0.45)	6.95 (0.65)	**6.90** (0.63)
		1.00	1.00	50.06 (0.59)	**8.16** (0.67)	**8.16** (0.67)
60	10	0.00	0.00	**3.24** (0.42)	5.35 (0.57)	**3.24** (0.42)
		0.25	0.05	6.14 (0.41)	8.65 (0.61)	**6.13** (0.41)
		0.50	0.48	14.97 (0.41)	13.85 (0.72)	**12.97** (0.51)
		0.75	0.88	30.04 (0.54)	11.34 (0.82)	**11.21** (0.78)
		1.00	1.00	51.72 (0.84)	**12.92** (0.88)	**12.92** (0.88)
60	20	0.00	0.00	**1.76** (0.25)	2.91 (0.32)	**1.76** (0.25)
		0.25	0.15	4.75 (0.25)	6.00 (0.35)	**4.69** (0.25)
		0.50	0.85	13.73 (0.27)	**9.87** (0.44)	**9.87** (0.41)
		0.75	0.95	28.79 (0.32)	6.95 (0.47)	**6.92** (0.47)
		1.00	1.00	50.04 (0.42)	**8.16** (0.50)	**8.16** (0.50)

$SIMSE$ value for the parametric model at $\gamma = 0$ for $\rho = 0$, 0.2 and 0.8 are 1.25, 1.74 and 6.25, respectively. Notice that, the mean square error is the sum of the squared bias plus the variance of the fit when conditioned on the values of the random effects. In the $\gamma = 0$ case, the bias is zero for the parametric and $MMRPM$ methods, hence the mean square error is the variance of the fit. If the sample size n remains fixed, and ρ increases, the effective sample size decreases (as n remains fixed), so the variance must increase as ρ increases.

In virtually all of the $m = 30$ profiles cases, for moderate and large values of γ, the nonparametric and $MMRPM$ have the smaller $SIMSE$. In addition, $MMRPM$ appears to have the smallest $SIMSE$ value on average over different degrees of model misspecification regardless of the sample size.

Also, of interest is the mixing parameter (λ) for different amounts of misspecifica-

Table 5.9: Simulated IMSE and average $\hat{\lambda}$ using correlated data set with $\rho = 0.8$ across m, n, and γ. Monte Carlo standard errors in parenthesis. Best values in bold.

m	n	γ	$\bar{\hat{\lambda}}$	Par	NP	MMRPM
30	10	0.00	0.00	**7.92** (1.75)	12.13 (2.28)	**7.92** (1.75)
		0.25	0.01	**10.68** (1.72)	15.30 (2.36)	**10.68** (1.72)
		0.50	0.39	19.07 (1.64)	20.25 (2.44)	**17.95** (1.82)
		0.75	0.89	33.41 (1.54)	16.41 (2.44)	**16.11** (2.28)
		1.00	0.99	54.10 (1.56)	**17.97** (2.46)	**17.97** (2.44)
30	20	0.00	0.00	**6.25** (1.32)	9.64 (1.65)	**6.25** (1.32)
		0.25	0.05	9.17 (1.31)	12.91 (1.70)	**9.16** (1.31)
		0.50	0.54	17.96 (1.29)	17.35 (1.72)	**15.87** (1.33)
		0.75	0.85	32.66 (1.26)	14.30 (1.84)	**13.74** (1.67)
		1.00	0.95	53.43 (1.25)	15.51 (1.84)	**15.40** (1.79)
60	10	0.00	0.00	**7.91** (1.24)	12.13 (1.66)	**7.91** (1.24)
		0.25	0.01	**10.67** (1.22)	15.29 (1.72)	**10.67** (1.22)
		0.50	0.38	19.04 (1.15)	20.23 (1.78)	**17.96** (1.30)
		0.75	0.88	33.35 (1.07)	16.44 (1.79)	**16.16** (1.66)
		1.00	0.99	53.98 (1.05)	**18.01** (1.80)	**18.01** (1.79)
60	20	0.00	0.00	**6.33** (0.96)	9.69 (1.22)	**6.33** (0.96)
		0.25	0.05	**9.25** (0.95)	12.95 (1.26)	**9.25** (0.95)
		0.50	0.54	18.02 (0.93)	17.39 (1.34)	**15.94** (1.05)
		0.75	0.85	32.72 (0.91)	14.37 (1.37)	**13.82** (1.24)
		1.00	0.95	53.47 (0.91)	15.54 (1.38)	**15.45** (1.33)

tion, profile size, number of profiles and variance-covariance structure. It is of course expected that, as the degree of misspecification increases, the average estimate of λ should increase, as more misspecification indicates a greater need for nonparametric fit. For $\gamma = 0$, the average estimate of λ should be close to zero, as the model has been correctly specified and the $MMRPM$ fit should be equal to the parametric fit.

In all Tables 5.1, 5.8 and 5.9, the average estimate of λ ($\bar{\hat{\lambda}}$) is performing as expected. When γ equals zero (one), $\bar{\hat{\lambda}}$ gets closer to zero (one) for every combination of m, n and ρ. As the number of profiles increases, the $\bar{\hat{\lambda}}$ approaches the desired value more slowly as ρ increases.

Two power studies were performed using $AR(1)$ structure on the random errors with a low correlation ($\rho = 0.2$) and a high correlation ($\rho = 0.8$). As in the independence case, the random effects were assumed to be independent. In addition, the paramet-

ric approach utilizing the second order polynomial regression model with two random effects, the nonparametric approach utilizing the mixed linear penalized splines regression and the semiparametric approach using our $MMRPM$ method are evaluated.

One concern in the correlated data scenario is whether the misspecification term influences the estimated simulated probability of signal for the out-of-control situations to the same degree as for the uncorrelated scenario.

The simulated probability of signal for the out-of-control situation for the correlated data cases appear in Tables 5.10 − 5.17. In all of these tables, the first column gives the misspecification degrees and the second column contains the step shift sizes. The third through the fifth columns presents the simulated probability of signal from the parametric (Par), the nonparametric (NP), and the $MMRPM$ methods based on the fitted values, respectively. Columns six through the last present the probabilities of signal from the three methods, Par, NP, and $MMRPM$ based on the estimated random effects (eblups).

Table 5.10: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 10$, $l = 20$ and $\rho = 0.2$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.172**	0.171	0.106	**0.172**	0.110	0.126
	2	**0.560**	0.516	0.364	**0.560**	0.250	0.348
	4	**0.978**	0.961	0.898	**0.978**	0.379	0.845
0.25	1	0.176	0.328	**0.411**	0.176	0.264	**0.500**
	2	0.559	0.697	**0.700**	0.559	0.451	**0.764**
	4	0.978	**0.981**	0.978	**0.978**	0.612	0.900
0.50	1	0.177	0.592	**0.659**	0.177	0.515	**0.805**
	2	0.556	0.868	**0.898**	0.556	0.693	**0.966**
	4	0.978	0.997	**0.997**	0.978	0.839	**1.000**
0.75	1	0.179	**0.942**	0.896	0.179	0.929	**0.982**
	2	0.577	**0.994**	0.985	0.577	0.970	**0.998**
	4	0.979	**1.000**	**1.000**	0.979	0.995	**0.999**
1.00	1	0.177	**0.940**	0.891	0.177	**0.930**	0.829
	2	0.550	**0.993**	0.985	0.550	**0.973**	0.909
	4	0.876	**1.000**	**1.000**	0.976	**0.995**	0.980

Table 5.11: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 20$, $l = 20$ and $\rho = 0.2$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.182**	0.131	0.115	**0.182**	0.114	0.150
	2	**0.554**	0.465	0.393	**0.554**	0.222	0.412
	4	**0.982**	0.953	0.913	**0.982**	0.342	0.912
0.25	1	0.179	0.275	**0.514**	0.179	0.242	**0.591**
	2	0.553	0.639	**0.815**	0.553	0.405	**0.904**
	4	0.982	0.987	**0.999**	0.982	0.596	**0.998**
0.50	1	0.177	0.506	**0.511**	0.177	0.468	**0.768**
	2	0.553	0.821	**0.826**	0.553	0.645	**0.946**
	4	0.981	0.997	**0.998**	0.981	0.827	**0.997**
0.75	1	0.173	0.912	**0.916**	0.173	0.906	**0.983**
	2	0.552	0.988	**0.989**	0.552	0.969	**0.993**
	4	0.982	**1.000**	**1.000**	0.982	0.992	**0.998**
1.00	1	0.173	0.918	**0.921**	0.173	**0.905**	0.815
	2	0.552	0.989	**0.990**	0.552	**0.966**	0.886
	4	0.982	**1.000**	**1.000**	0.982	**0.993**	0.983

Table 5.12: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 10$, $l = 40$ and $\rho = 0.2$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.159**	0.141	0.136	**0.159**	0.096	0.112
	2	**0.617**	0.562	0.514	**0.617**	0.247	0.408
	4	**0.999**	0.997	0.995	**0.999**	0.460	0.975
0.25	1	0.160	0.303	**0.547**	0.160	0.236	**0.576**
	2	0.620	0.746	**0.873**	0.620	0.447	**0.897**
	4	**0.999**	0.998	**0.999**	**0.999**	0.709	0.996
0.50	1	0.162	0.523	**0.824**	0.162	0.449	**0.864**
	2	0.622	0.889	**0.979**	0.622	0.671	**0.989**
	4	0.999	0.999	**1.000**	0.999	0.999	**1.000**
0.75	1	0.157	0.918	**0.959**	0.157	0.880	**0.988**
	2	0.622	0.991	**0.997**	0.622	0.966	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	0.997	**1.000**
1.00	1	0.157	0.916	**0.956**	0.157	**0.880**	0.835
	2	0.624	0.992	**0.996**	0.624	**0.968**	0.927
	4	0.999	**1.000**	**1.000**	**0.999**	0.997	**0.999**

Table 5.13: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 20$, $l = 40$ and $\rho = 0.2$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.162**	0.137	0.132	**0.162**	0.113	0.138
	2	**0.651**	0.556	0.517	**0.651**	0.257	0.446
	4	**0.999**	0.996	0.996	**0.999**	0.476	0.984
0.25	1	0.165	0.289	**0.589**	0.165	0.260	**0.653**
	2	0.650	0.740	**0.935**	0.650	0.469	**0.966**
	4	0.999	**1.000**	**1.000**	0.999	0.742	**1.000**
0.50	1	0.167	0.493	**0.588**	0.167	0.472	**0.833**
	2	0.649	0.886	**0.942**	0.649	0.694	**0.996**
	4	0.999	**1.000**	**1.000**	0.999	0.901	**1.000**
0.75	1	0.168	0.898	**0.945**	0.168	0.885	**0.992**
	2	0.647	0.991	**0.995**	0.647	0.976	**0.998**
	4	0.999	**1.000**	**1.000**	0.999	0.997	**1.000**
1.00	1	0.164	0.899	**0.943**	0.164	**0.889**	0.804
	2	0.648	0.991	**0.995**	0.648	**0.975**	0.915
	4	0.999	**1.000**	**1.000**	**0.999**	0.996	0.998

Table 5.14: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 10$, $l = 20$ and $\rho = 0.8$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	0.171	0.144	**0.173**	0.171	0.113	**0.174**
	2	0.559	0.492	**0.561**	**0.559**	0.251	0.555
	4	**0.978**	0.953	**0.978**	**0.978**	0.398	0.975
0.25	1	0.172	0.304	**0.438**	0.172	0.273	**0.502**
	2	0.560	0.671	**0.735**	0.560	0.450	**0.748**
	4	0.978	**0.985**	0.984	0.978	0.611	**0.985**
0.50	1	0.173	0.549	**0.863**	0.173	0.516	**0.996**
	2	0.563	0.859	**0.980**	0.563	0.713	**1.000**
	4	0.978	0.996	**1.000**	0.978	0.857	**1.000**
0.75	1	0.174	0.923	**0.998**	0.174	0.932	**0.999**
	2	0.565	0.990	**0.999**	0.565	0.977	**1.000**
	4	0.977	**1.000**	**1.000**	0.977	**1.000**	**1.000**
1.00	1	0.174	0.929	**0.998**	0.174	0.934	**0.993**
	2	0.565	0.989	**1.000**	0.565	0.977	**0.999**
	4	0.976	**1.000**	**1.000**	0.976	0.993	**1.000**

Table 5.15: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 30$, $n = 20$, $l = 20$ and $\rho = 0.8$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.173**	0.146	0.171	**0.173**	0.103	0.161
	2	0.530	0.511	**0.534**	**0.530**	0.236	0.523
	4	**0.982**	0.966	0.981	**0.982**	0.369	0.978
0.25	1	0.169	0.329	**0.758**	0.169	0.250	**0.882**
	2	0.530	0.686	**0.938**	0.530	0.437	**0.963**
	4	0.982	0.991	**0.999**	0.982	0.623	**0.999**
0.50	1	0.170	0.553	**0.824**	0.170	0.481	**0.992**
	2	0.530	0.851	**0.973**	0.530	0.661	**1.000**
	4	0.980	0.997	**0.999**	0.980	0.828	**1.000**
0.75	1	0.169	0.941	**0.997**	0.169	0.920	**1.000**
	2	0.535	0.997	**1.000**	0.535	0.966	**1.000**
	4	0.979	**1.000**	**1.000**	0.979	0.997	**1.000**
1.00	1	0.171	0.941	**0.997**	0.171	0.921	**1.000**
	2	0.536	0.997	**1.000**	0.536	0.967	**1.000**
	4	0.978	**1.000**	**1.000**	0.978	0.996	**1.000**

Table 5.16: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 10$, $l = 40$ and $\rho = 0.8$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.157**	0.131	**0.157**	**0.157**	0.114	**0.157**
	2	**0.623**	0.547	**0.623**	**0.623**	0.264	**0.623**
	4	**0.999**	0.997	**0.999**	**0.999**	0.501	**0.999**
0.25	1	0.160	0.287	**0.385**	0.160	0.271	**0.462**
	2	0.626	0.730	**0.757**	0.626	0.500	**0.771**
	4	**0.999**	**0.999**	**0.999**	**0.999**	0.747	**0.999**
0.50	1	0.160	0.506	**0.822**	0.160	0.513	**0.996**
	2	0.627	0.876	**0.983**	0.627	0.747	**1.000**
	4	0.980	0.997	**0.999**	0.980	0.828	**1.000**
0.75	1	0.160	0.905	**0.993**	0.160	0.917	**1.000**
	2	0.623	0.993	**1.000**	0.623	0.980	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	**1.000**	**1.000**
1.00	1	0.167	0.902	**0.993**	0.167	0.921	**0.996**
	2	0.618	0.993	**1.000**	0.618	0.981	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	**1.000**	**1.000**

Table 5.17: Simulated probability of signal of out-of-control for correlated data set using the six T^2 statistics with different values of misspecification and shifts for $m = 60$, $n = 20$, $l = 40$ and $\rho = 0.8$. Best values in bold.

γ	Shift	T^2 based on the fitted values			T^2 based on the $eblups$		
		$T^2_{Par1,i}$	$T^2_{NP1,i}$	$T^2_{MMRPM1,i}$	$T^2_{Par2,i}$	$T^2_{NP2,i}$	$T^2_{MMRPM2,i}$
0.00	0	0.050	0.050	0.050	0.050	0.050	0.050
	1	**0.182**	0.151	**0.183**	0.182	0.121	**0.183**
	2	0.663	0.585	**0.664**	0.663	0.274	**0.664**
	4	**0.999**	0.997	**0.999**	**0.999**	0.490	**0.999**
0.25	1	0.182	0.300	**0.835**	0.182	0.258	**0.961**
	2	0.660	0.771	**0.981**	0.660	0.487	**0.994**
	4	0.999	**1.000**	**1.000**	0.999	0.739	**1.000**
0.50	1	0.178	0.517	**0.855**	0.178	0.458	**0.997**
	2	0.654	0.899	**0.991**	0.654	0.701	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	0.906	**1.000**
0.75	1	0.178	0.924	**0.996**	0.178	0.890	**1.000**
	2	0.648	0.992	**1.000**	0.648	0.974	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	**1.000**	**1.000**
1.00	1	0.179	0.926	**0.995**	0.179	0.890	**1.000**
	2	0.650	0.993	**1.000**	0.650	0.972	**1.000**
	4	0.999	**1.000**	**1.000**	0.999	**1.000**	**1.000**

Tables 5.10 through 5.13 are for the $AR(1)$ case with $\rho = 0.2$. While Tables 5.14 through 5.17 are for the $AR(1)$ case with $\rho = 0.8$. There are similarities between the uncorrelated case and the correlated errors structure cases. For $\gamma = 0$, the parametric methods should have the highest simulated probability of signal for the out-of-control and the $MMRPM$ procedure obtains the simulated probability of signal values very close to the parametric simulated probability of signal values. For $\gamma = 1$, the nonparametric method should have the highest simulated probability of signal and the $MMRPM$ gives simulated probability of signal very close to the nonparametric values. As γ increases from zero to one, the simulated probability of signal values for the $MMRPM$ method are either the highest values (for low to moderate model misspecification), or are very close in the value to the "best" simulated probability of signal values. Also, it is observed that as the number of profiles (m) increases the simulated probability of signal from the nonparametric and $MMRPM$ methods increase for different shift sizes and various degrees of model misspecification. As the number of observations per each pro-

file (n) increases the power of the $MMRPM$ method to detect the step shift gets higher. In addition, by increasing the degree of autocorrelation between errors the simulated probability of signal from the nonparametric and $MMRPM$ methods increase. While as ρ increases the simulated probability of signal from the parametric method decreases especially for the small and moderate shift sizes. In general, as ρ increases the performance of the nonparametric and $MMRPM$ methods increases for all values of m and n.

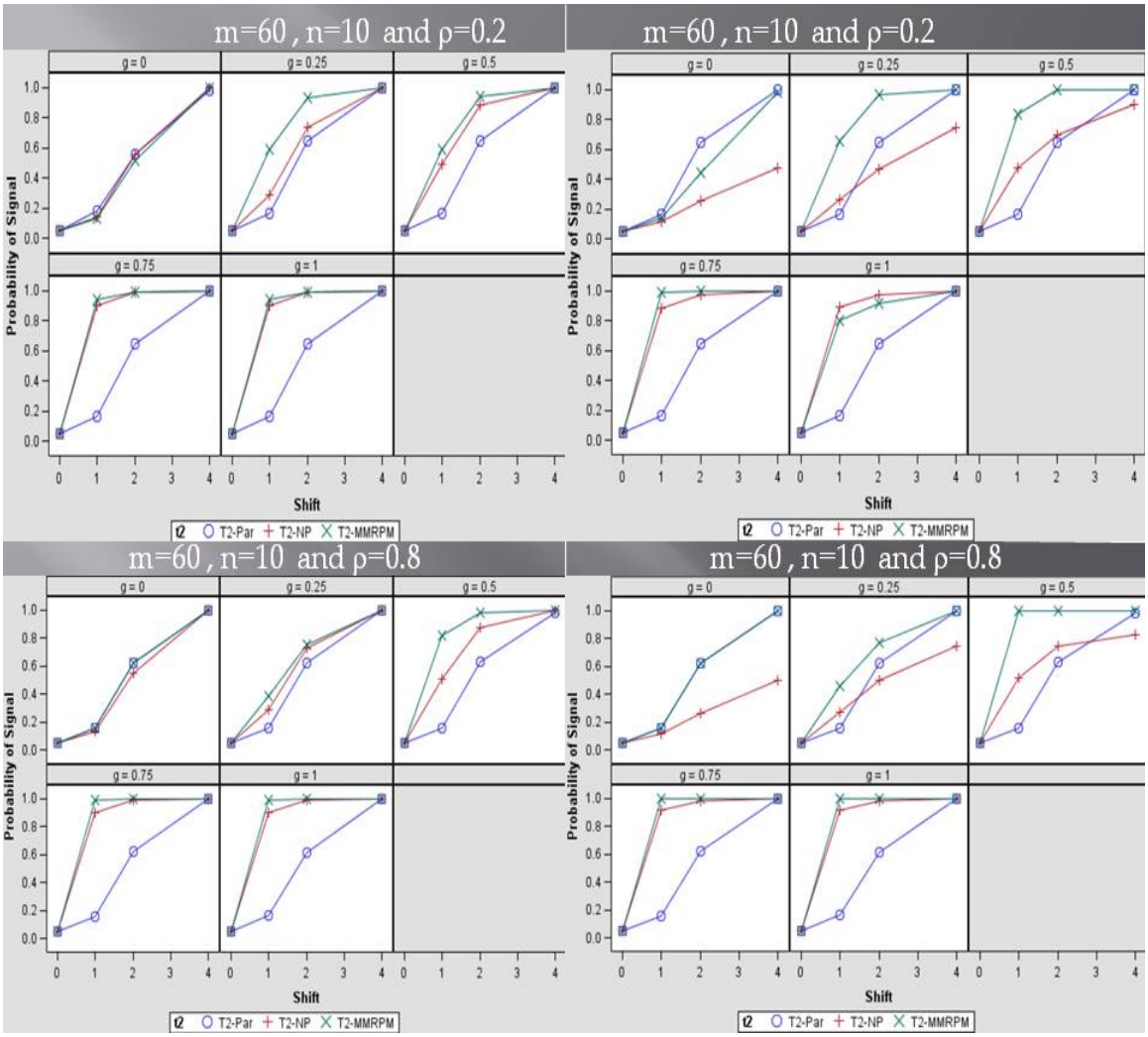

Figure 5.6: Simulated probability of signal for the correlated errors structure dataset, for different degrees of misspecification ($\gamma = g$), for the T^2- based on the fitted values (on the right panel) and the estimated random effects (on the left panel) with $m = 60$, and $n = 10$.

Figure 5.6 shows the simulated probability of signal for the correlated error structure cases with $m = 60$ and $n = 10$ for the T^2 statistics obtained via the parametric method (blue line), the nonparametric method (red line), and the $MMRPM$ method (green line). The horizontal axis is the shift (δ), and four larger panels that show various combinations of ρ and the components that were used to calculate T^2 statistics. On the left side, the power curves from different T^2 statistics based on the fitted values. While on the right hand side gives the T^2 statistics based on the estimated random effects. Within the larger panels are smaller panels which show the different degrees of misspecification and the ρ values. It is clear that, the performance for the $MMRPM$ method is superior to the nonparametric and the parametric methods for the misspecified model. The power curve from $MMRPM$ method is very close to the one from the parametric method for the correctly specified parametric model. For $\gamma = 1$ the power curve from the nonparametric method is very close to the one from the $MMRPM$ method.

In general, when the degrees of misspecification increase the semiparametric and nonparametric approaches show greater superiority over the parametric approach. We see that the semiparametric approach is always at least equivalent to the parametric approach (for the correctly specified model) and often far superior. The $MMRPM$ and the nonparametric approaches are doing a great job for data sets with correlated errors structures, a condition common in longitudinal data sets.

5.5 Chapter Summary

In this Chapter, a Monte Carlo study was performed to compare the simulated integrated mean square errors and the simulated probability of signal of the parametric, nonparametric, and semiparametric approaches. Both correlated and uncorrelated errors structure scenarios were evaluated for varying amounts of model misspecification, number of profiles, number of observations per profile, shift location, and in- and out-of-control situations. The semiparametric ($MMRPM$) method for uncorrelated and correlated scenarios were competitive with the parametric and nonparametric over all levels of misspecification. For a correctly specified model, the $SIMSE$ and the simulated probability of signal for the parametric and the $MMRPM$ methods were identical (or nearly so). Since the average of $\hat{\lambda}$ is zero (or nearly so), the $MMRPM$ method results were the same as the parametric results. For the severe model misspecification case (large values of γ), the nonparametric and $MMRPM$ methods were identical (or nearly so) where the average of $\hat{\lambda}$ was one (or nearly so). For mild model misspecification case, the $MMRPM$ method was superior to the parametric and nonparametric methods. Therefore, this simulation supports the claim that the $MMRPM$ method is robust to model misspecification.

In addition, the $MMRPM$ method performed better for data sets with correlated error structure. Also, the performances of the nonparametric and $MMRPM$ methods improve as the number of observations per profile increases, since more observations over the same range of X provides more information about the behavior of mean response, implying that more knots can be used resulting in greater flexibility in the nonparametric curves and consequently, the semiparametric curves.

Chapter (6) will give another comparative study for the parametric, nonparametric, and semiparametric approaches via a real data set.

Chapter 6

A Case Study

We conclude the presentation of our proposed methods with a "case study" illustrating the use of our procedure applied to a "real world" profile monitoring problem. The simulation studies have shown that our methods hold both theoretical potential and possess good performance characteristics in some restricted data setting. While the features of the simulations appear to be restricted to seemingly simple cases, it is not unusual for the real data sets to satisfy these restrictions. The case study that follows illustrates this point.

In this Chapter, we turn our attention to an application presented in Amirhossein et al. (2009). In the following section, an automobile engine application is presented followed by a Phase I profile monitoring study using the parametric, nonparametric and semiparametric approaches.

6.1 The Automobile Engine Application

This application from Amirhossein et al. (2009) concerns the automotive industry. They want to study the relationship between the torque produced by an engine and the engine speed in revolutions per minute (RPM). The relationship between torque and engine speed can be considered as one of the most important quality characteristic of the automobile engine. The application considered here is for the engine type $TU3$ which are assembled for a French automobile, the Peugeot. In the study, the engine is run at different RPM values and the corresponding torque values obtained. Thus, the torque produced by the engine is considered as response variable and the correspon-

dent speed in RPM is considered as the explanatory variable.

The profiles that describe the relationship between torque and RPM should be similar when the manufacturing process is in-control. An engine with mechanical defects or any other issues hopefully will result in an outlying profile. Because there are multiple RPM values obtained for each engine it is natural to try to apply a multivariate quality control procedure to discover engines which have different profiles from the others.

In this data set, we have a total of twenty-six engines as our historical data set for Phase I analysis. For each engine, 14 engine speed values are set to 1500, 2000, 2500, 2660, 2800, 2940, 3500, 4000, 4500, 5000, 5225, 5500, 5775, and 6000 RPM and the corresponding torque values are reported. Therefore, 14 points are collected for each engine and form the profile of interested. The data set for the full data set (twenty-six profiles) was introduced in Amirhossein et al. (2009) and shown in the Appendix G. The raw data set is shown in Figure 6.1.

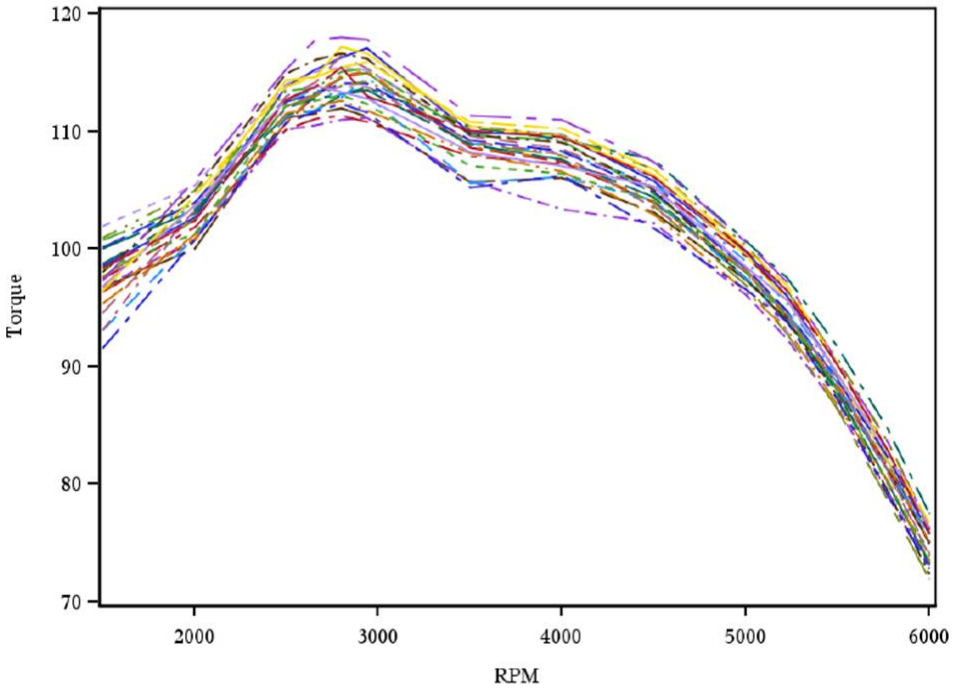

Figure 6.1: The raw data set for 26 automobile engines (Torque vs. RPM)

Figure 6.1 illustrates a typical automobile engine profiles. The data set contains 26

profiles each of which consists of 14 measurements along the profile. There is no replication which means that there is a single measurement at each RPM. The raw data set contains some peaks and dips, not easily captured by the parametric approach. These curves are clearly nonlinear and thus new techniques to monitor these curves are desirable. Therefore, our two new approaches (nonparametric and semiparametric) will be applied to this data set.

In the next section, the relationship between torque and RPM of an engine can be modeled by the parametric, nonparametric and semiparametric ($MMRPM$) methods. In addition, the mean square error (MSE) is calculated for each method.

6.2 Phase I Analysis

In a Phase I analysis, we have m quality profiles in the historical data set (HDS). The main goal of the Phase I analysis is to estimate process parameters based on an in-control process. The proposed approaches that we consider for this data set will account for the autocorrelation within engines. These approaches (parametric, nonparametric and semiparametric) can be considered as methods for data reduction that enable the determination of outlying profiles.

The parametric approach using the mixed second order quadratic model in one regressor is

$$\mathbf{y}_i = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\mathbf{x}_i^* + (\beta_2 + b_{i2})\mathbf{x}_i^{*2} + \epsilon_i \qquad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n \quad (6.1)$$

where $\mathbf{y}_i$ is the vector containing the torque values. The $\mathbf{x}_i^*$ is a vector containing the centered RPM values, $\mathbf{x}_i^* = \mathbf{x}_i - \bar{\mathbf{x}}$. Since the $PRMs$ are fixed for each engine, we denote the RPM, x_{ij} by simply $\mathbf{x}_i$. The ϵ_i is the vector containing the errors for the i^{th} engine. Torque measurements for this data set were taken at $n = 14$ different values for RPM, $\mathbf{x}_i$; $i = 1,2,\ldots,26$. Correspondingly, a series of order pairs (y_{ij}, x_{ij}) ; $j = 1,2,\ldots,14$ results for engine i and forms the automobile engine profile. Because the RPM values are the same for all the profiles we have $\mathbf{x}_i^* = \mathbf{x}^*$ for all values of i. The parameters, β_0, β_1 and β_3 are the fixed effects that are common for all profiles. The b_{i0}, b_{i1} and b_{i3} are the random

effects for the i^{th} profile, and are unique for each profile.

Because the random errors for each profile are likely to be correlated, we must account for their correlation in the analysis. An $AR(1)$ structure for the variance-covariance matrix Σ is utilized to model the errors within a profile, as assumed by Amirhossein et al. (2009) for the linear mixed model (the parametric approach).

The nonparametric approach utilizing the linear mixed penalized spline regression model as described in Section 4.2. In addition, the semiparametric approach via the $MMRPM$ method, as described in Section 4.4, is applied to this data set.

The fitted profiles using the parametric, nonparametric, and $MMRPM$ approaches are illustrated in Figure 6.2 with the raw data as well for the twenty-six engines. Where the sub-figure (a) represents the raw data set with the population average (PA) curve (solid black), the sub-figure (b) gives the parametric fits for the cluster specific (CS) and PA (solid black) profiles, the nonparametric fits for the CS and PA (solid black) profiles are in sub-figure (c), and the $MMRPM$ method CS and PA (solid black) fits are given in sub-figure (d). The MSE values are 4.58, 0.44 and 0.43 for the parametric, nonparametric and $MMRPM$ fits, respectively. Based on the MSE criteria, we can say that the nonparametric and semiparametric approaches are superior to the parametric approach for fitting this data set.

We can see the advantages of nonparametric and semiparametric approaches for fitting this data set more closely by focusing on one profile. Figure 6.3 provides us with the raw data set and the fitted profile using the three approaches for Engine 329. From Figure 6.3, one can see that the parametric approach is not able to capture the main characteristics in the raw data set for this engine, suggesting that the quadratic model is misspecified. On the other hand, the nonparametric and semiparametric approaches capture most of the main features in the data set. Furthermore, the MSE for the parametric, nonparametric and $MMRPM$ fits are 5.02, 0.66 and 0.66, respectively. This result further illustrates the conclusions of the preceding graph.

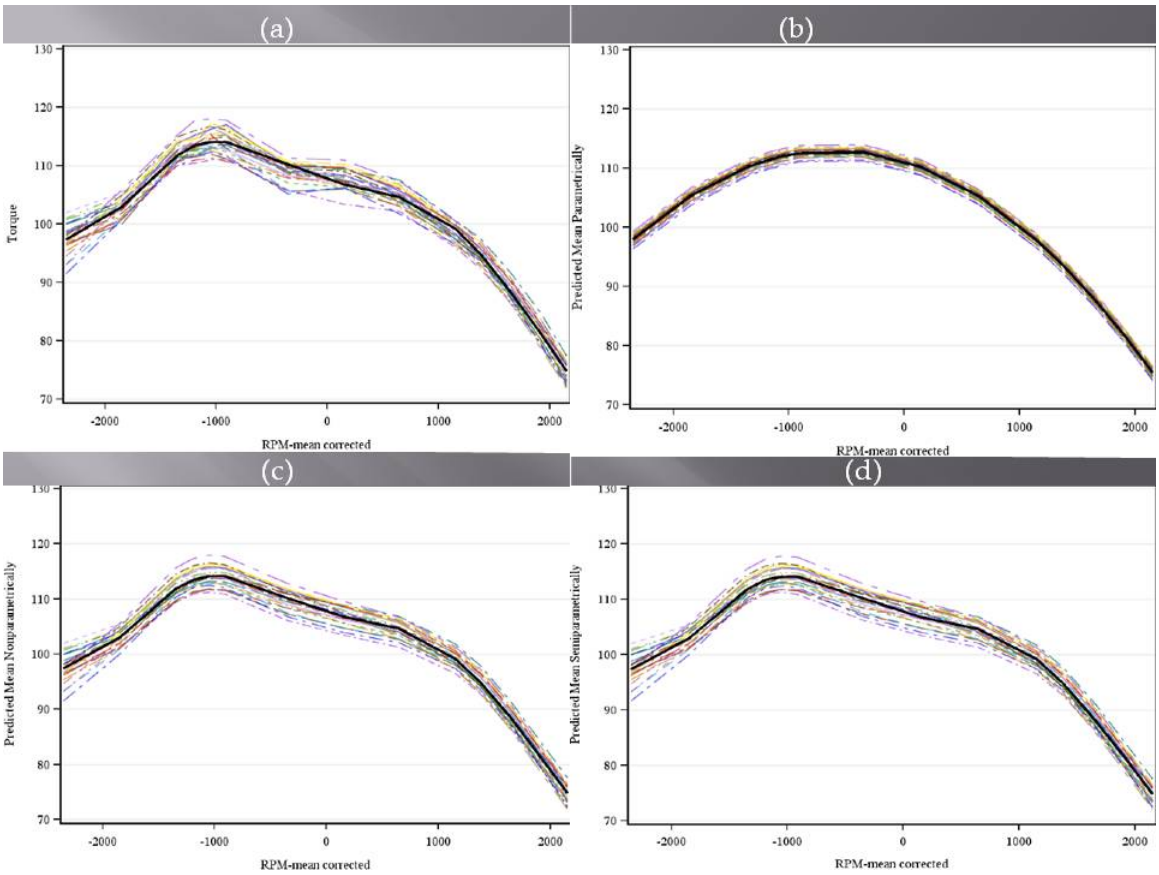

Figure 6.2: (a) The raw data set for 26 automobile engines with the PA curve (solid black). (b) Fitted profiles using a second order polynomial mixed model (Parametric Approach) with the PA curve (solid black) (c) Fitted profiles using linear mixed penalized spline regression (Nonparametric Approach) with the PA curve (solid black) and (d) Fitted profiles using MMRPM method (Semiparametric Approach) with the PA curve (solid black)

In addition, the semiparametric approach relies more on the nonparametric approach than the parametric model indicating that the parametric model is incorrectly specified. The estimated mixing parameter ($\hat{\lambda}$) value is approximately 0.98. Thus, 98% of the semiparametric fit comes from the nonparametric fit while only 2% comes from the parametric fit.

Since all approaches considered in this dissertation assume that the random errors are normally distributed, the first step in the Phase I analysis is to assess the viability of

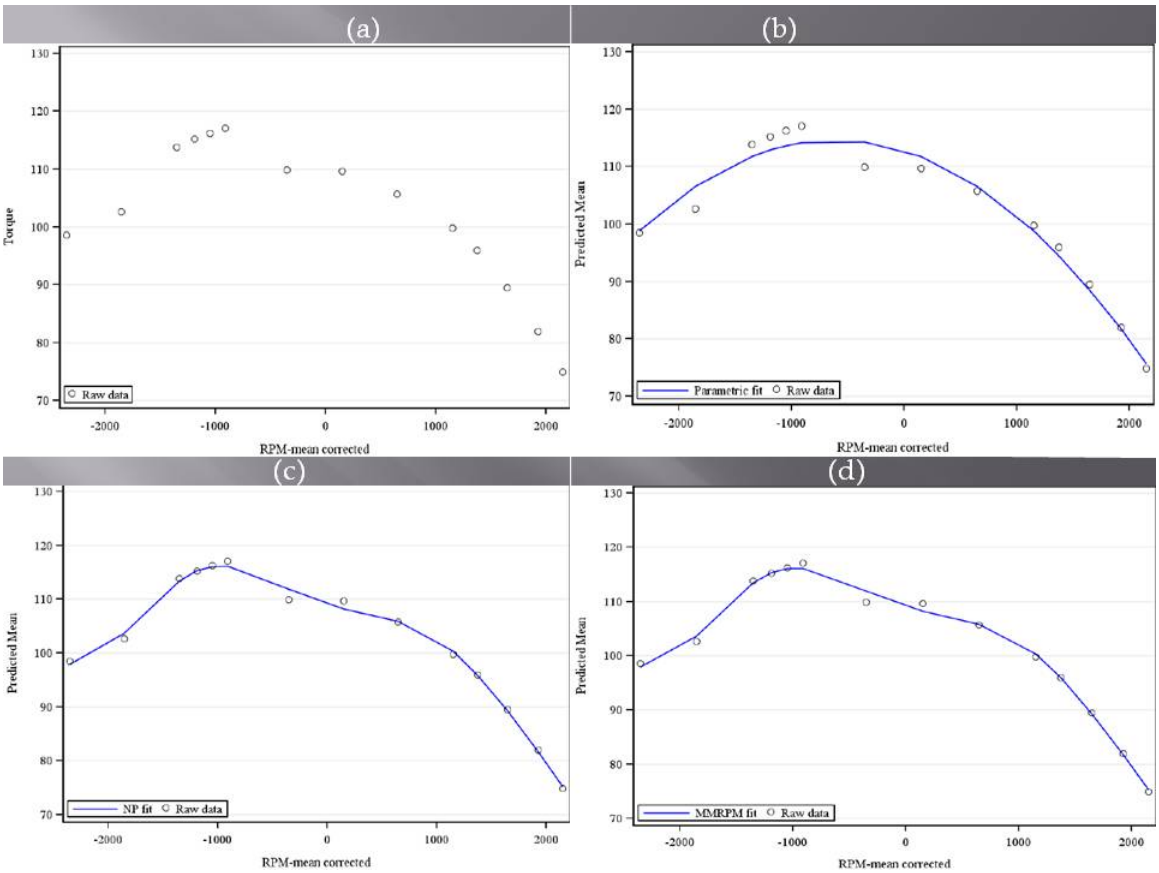

Figure 6.3: (a) The raw automobile engine data set for engine number 329. (b) Fitted profile using a second order polynomial fit for engine number 329 (Parametric Approach) (c) Fitted profile using linear penalized spline regression fit for engine number 329 (Nonparametric Approach) and (d) Fitted profile using MMRPM fit for engine number 329 (Semiparametric Approach)

this assumption. To examine if the normality is a reasonable assumption for the parametric, nonparametric and $MMRPM$ methods, we show in Figure 6.4 a histogram of the residuals from each method as an exploratory measure. The residuals have been computed from each method for the 26 engines using the $AR(1)$ structure for the variance-covariance matrix. Traditional tests of normality, such as the Anderson-Darling test, are not valid here since the residuals are correlated. Therefore, we use these histograms as graphical aids to check the normality assumption.

Figure 6.4, gives the histogram for the residuals from the parametric approach (subfigure (a)). It can be seen that, the residuals are centered around zero but there is some

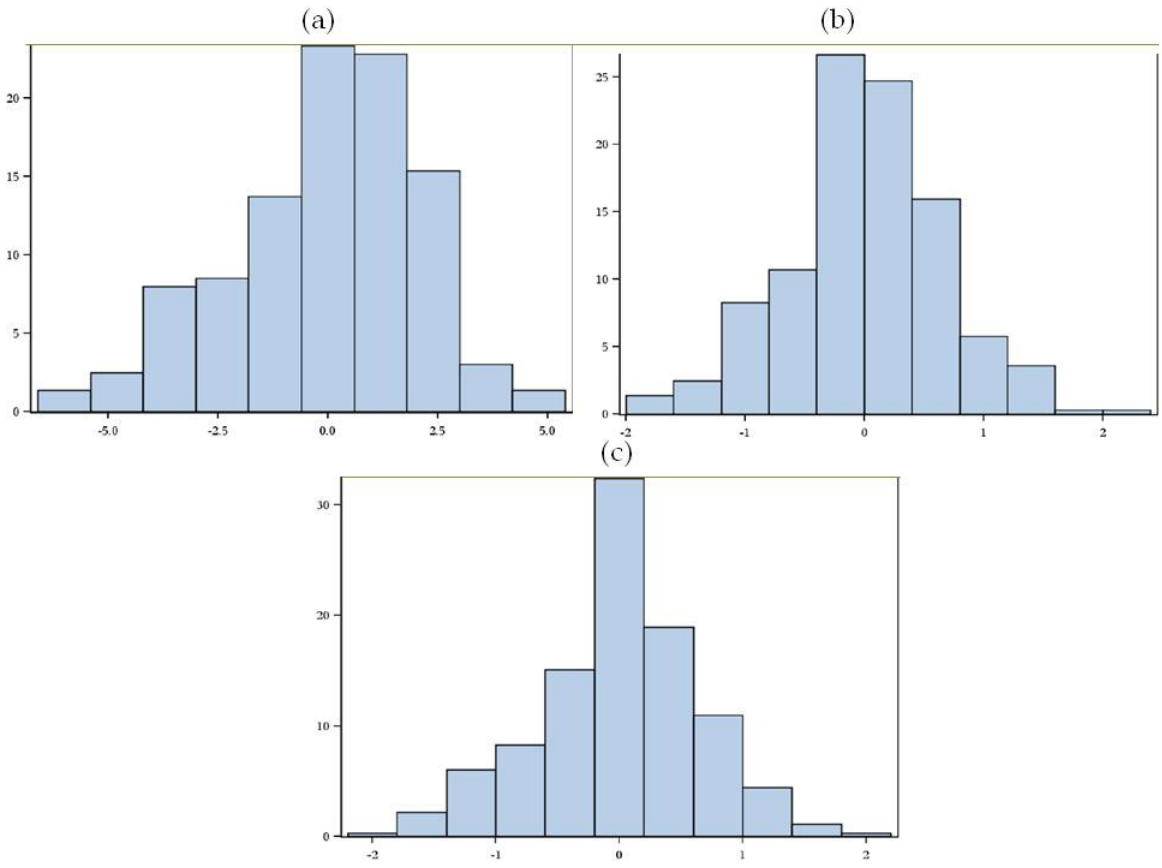

Figure 6.4: (a) Histogram for the residuals from the fitted profiles using the second order polynomial mixed model (Parametric Approach) (b) Histogram for the residuals from the fitted profiles using the linear mixed penalized spline regression (Nonparametric Approach) and (c) Histogram for the residuals from the fitted profiles using the MMRPM method (Semiparametric Approach) with the PA curve (solid black)

amount of left-skewness in the data. This result is in agreement with the one from Amirhossein et al. (2009). The histogram for the residuals from the nonparametric approach (sub-figure (b)) looks very symmetric around zero and the range for the residual values from two to negative two. The residuals from the semiparametric approach (sub-figure (c)) strongly resemble data from a normal distribution with mean zero. Hence, the residuals resulting from the nonparametric and $MMRPM$ methods appear to more strongly support the normality assumption than those from the parametric method.

In the following section, we use a T^2 based control procedure to investigate the stability of the process as well as determine if there are outlying profiles using the parametric, nonparametric and semiparametric approaches.

6.3 The T^2 Control Chart

In order to determine if a profile is an outlier or if the profile has shifted, we propose two methods for each approach: one, by analyzing the vector of the fitted values and the second by utilizing the vector of the estimated random effects from the parametric, nonparametric and $MMRPM$ approaches. In Phase I analysis, we are interested in using the HDS to estimate the mean vector μ_b and variance-covariance matrix Σ_b of the $\hat{\mathbf{b}}_i$ vectors, after we have removed the out-of-control profiles.

One method of identifying out-of-control profile is use of the Hotelling's T^2 statistics, as in Chapter 4. Since the T^2 statistics based on the fitted values and the estimated random effects are identical for the parametric approach (as proved in Appendix F) and nearly equal for the other two approaches, we only focus on the T^2 statistics based on the estimated random effects from each one of our methods as given in Equations (4.11), (4.25), and (4.37), and using the successive differences variance-covariance matrix estimator V_D given in Equation (2.9).

The T^2 control chart consists of a plot of T_i^2 statistics by i for all observations in the HDS. If a T_i^2 statistic exceeds the UCL associated with the chart, then that estimated profile is identified as a possible outlier to be removed from the HDS. In order to calculate the UCL for the T^2 chart, we need the distribution of the T_i^2 statistics. Unfortunately, the exact distribution is unknown. For more details see Jensen et al. (2008), Williams et al. (2006), Durbàn et al. (2005), and Gurrin et al. (2005). However, as shown in Chapter 5, the asymptotic distribution of T_i^2 is a good approximate to the exact distribution whenever the number of remaining samples, m^*, in the HDS is large. For example, m^* should be greater than p^2+3p for the parametric approach. As pointed out in the simulation section, use of the asymptotic distribution for determining the $UCLs$ for the nonparametric and semiparametric is conservative approach, providing $UCLs$ that are

slightly larger than they should be resulting in probability of signals slightly smaller than the nominal value.

For this analysis, the asymptotic distribution of T_i^2 using the parametric approach is χ^2_{df} for all $i = 1,2,\ldots,m$. The UCL is calculated as

$$UCL_{Par} = \chi^2_{(1-\alpha,q)} \tag{6.2}$$

where $\chi^2_{(1-\alpha,q)}$ is the $(1-\alpha)^{th}$ quantile from a χ^2 distribution with q degrees of freedom and q replaced by the number of estimated random effects.

The UCL for the T_i^2 statistics based on the nonparametric estimated random effects is obtained as

$$UCL_{NP} = \chi^2_{(1-\alpha,df_{NP})} \tag{6.3}$$

where $\chi^2_{(1-\alpha,df_{NP})}$ is the $(1-\alpha)^{th}$ quantile from a χ^2 distribution with df_{NP} is replaced by the number of estimated random effects plus the number of knots.

In addition, the estimated first-order autocorrelation, estimated from the $MMRPM$ residuals, is 0.5 approximately. Thus, using our results from the simulation section, the UCL for the T_i^2 statistics based on the $MMRPM$ estimated random effects is obtained as in Equation (6.4), as a convex combination from the parametric UCL and the nonparametric UCL via a mixing parameter $\lambda \in [0,1]$.

$$UCL_{MMRPM} = (1-\hat{\lambda}) * UCL_{Par} + \hat{\lambda} * UCL_{NP} \tag{6.4}$$

where $\hat{\lambda}$ is the estimated mixing parameter. For this analysis, we choose $\alpha_{overall} = 0.05$ which corresponds to

$$\alpha = 1-(1-\alpha_{overall})^{\frac{1}{m}} = 1-(1-0.05)^{\frac{1}{26}} = 0.0019709 \tag{6.5}$$

The upper control limits are calculated as $UCL_{Par} = 12.459$, $UCL_{NP} = 18.942$ and $UCL_{MMRPM} = 18.755$.

We construct the T^2 chart for the $m = 26$ profiles to identify profiles with abnormal estimated random effects values. The T_i^2 values for the parametric, nonparametric and $MMRPM$ as given in Equations (4.11), (4.25), and (4.37) are listed in Table G.

Table 6.1: The engine number and the T^2 statistic values from the parametric, nonparametric and semiparametric approaches.

Engine number	T^2_{Par}	T^2_{NP}	T^2_{MMRPM}
329	0.89252	1.7973	2.3719
449	0.23864	4.6104	4.7524
529	4.11722	7.2894	7.5209
642	2.38670	4.0708	8.2451
724	3.94504	6.8336	7.4838
803	5.40324	6.2442	8.1247
930	0.90897	2.2802	9.8719
1148	0.09433	3.3656	4.6534
1171	0.18307	4.2287	6.0321
1516	6.21119	4.0571	14.0595
1791	8.43671	8.2784	9.3176
2600	0.87051	3.4187	5.0283
3100	0.52440	5.2642	9.4891
3720	1.95266	8.6303	10.4570
4025	1.72572	4.0617	5.5028
4068	4.52936	10.3262	11.4102
4926	3.72849	11.9014	16.5137
5155	2.60457	3.6400	4.9992
6143	7.61174	5.7080	13.2823
6844	2.62916	3.8511	7.0356
7811	1.83053	5.9513	6.5747
8007	0.55750	2.4177	3.6680
8623	1.36062	4.8894	5.2091
9388	5.48545	3.4297	12.9186
9404	6.57477	10.6235	11.6023
10430	0.74884	3.2251	5.1058

The corresponding control charts are given in Figure 6.5.

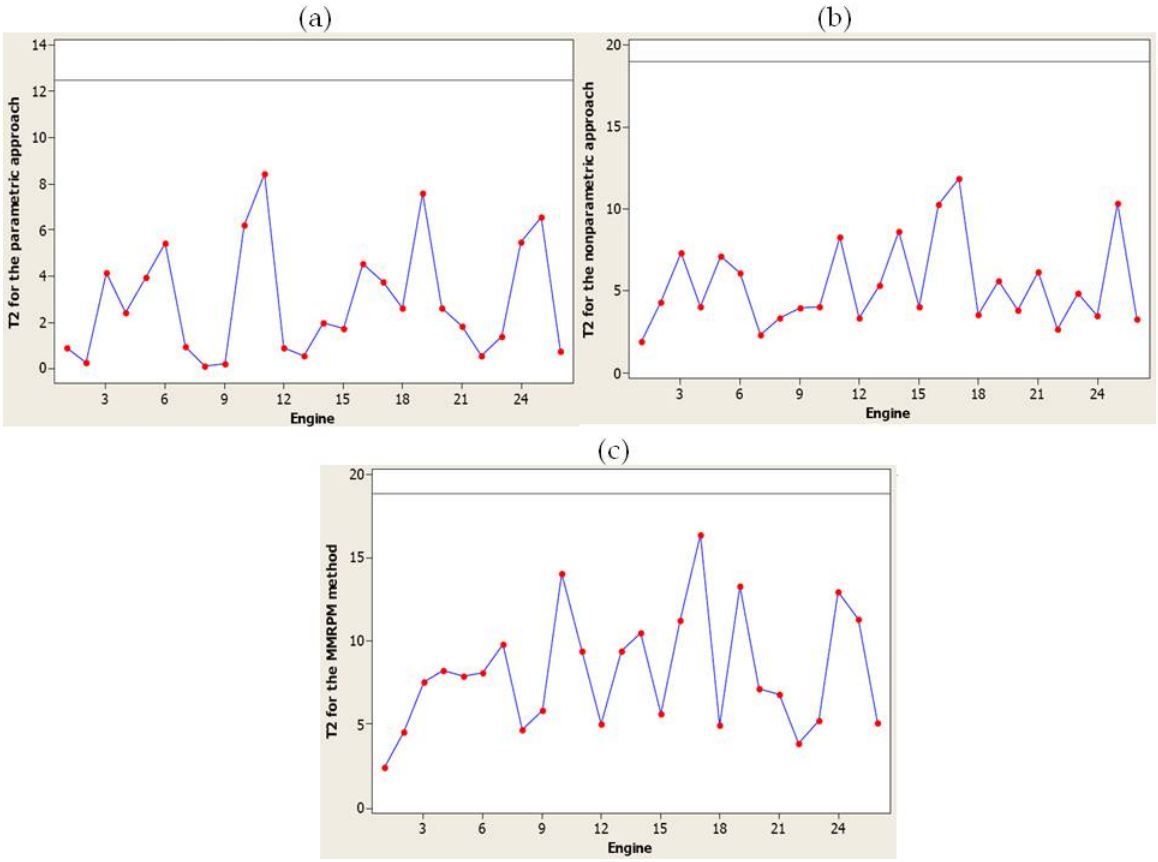

Figure 6.5: (a) T^2 control chart based on the estimated random effects parametrically. (b) T^2 control chart based on the estimated random effects nonparametrically (c) T^2 control chart based on the estimated random effects semiparametrically

From Figure 6.5, we see that the use of the T^2 charts based on the parametric approach (sub-figure (a)) does not produce any signal. This result is in agreement with the one from Amirhossein et al. (2009). In addition, the nonparametric approach (sub-figure (b)) and $MMRPM$ approach (sub-figure (c)) also do not produce a signal. It is interesting to note that the parametric, nonparametric and seminonparametric methods agree on this decision even thought the nonparametric and $MMRPM$ methods give much better fits for the engine profiles in term of the smaller mean square errors.

Our MSE results along with the plots shown in Figure 6.2 indicate that the quadratic parametric model is not able to capture the main features in this data set. Also, the parametric model produce a largest MSE in comparing with the nonparametric and

$MMRPM$ methods. Our simulation results indicate that using the T^2 approach with a misspecified parametric model is not as powerful at detecting step-shifts in profiles away for the normative profiles as the nonparametric or $MMRPM$ method. For these reasons the user should strongly consider using either the nonparametric or $MMRPM$ results for this data. Despite the fact that all three methods reach the same conclusion regarding the "in-control" status of each profile, the nonparametric and $MMRPM$ results provide a better description of the actual behavior of each profile. Thus, the nonparametric and $MMRPM$ methods give the user greater ability to properly interpret the true relationship between engine speed and torque for this type of engine and an increased likelihood of detecting unusual engines in future production.

As a result of this analysis, we conclude that all 26 engines can be used to obtain the parameters estimates on which the Phase II control charts will be based. Our conclusion agrees with that expressed in Amirhossein et al. (2009) where they found no unusual profiles using the quadratic model.

6.4 Chapter Summary

In Chapter 6, the relationship between torque produced by an engine and speed in automotive industry was studied. The parametric, nonparametric and $MMRPM$ approaches were utilized for fitting this relationship. The nonparametric and semiparametric approaches were superior to the parametric approach in term of the MSE. We used a Hotelling's T^2 statistic to conduct Phase I studies to determine the outlying profiles based on the estimated random effects. The parametric, nonparametric and semiparametric approaches showed that the process was stable. Finally, we conclude that the nonparametric and semiparametric approaches performed better than the parametric approach when the user's model is misspecified by being more efficient, flexible and robust to model misspecification for Phase I profile monitoring.

In the following Chapter 7, we give the summary and the future works in this area.

Chapter 7

Conclusion, and Outlook on Future Research

7.1 Introduction

This chapter summarizes the important nonparametric and semiparametric results from the previous chapters. Since the work presented here is ongoing research, directions for future research pertaining the nonparametric and semiparametric methods will also be given.

7.2 Conclusion

The field of multivariate SPC is an exciting field with a promising future. As products and process become more competitive and complex it is increasingly more important for the development and evaluation of statistical methodologies to meet those needs.

In this work, we have developed statistical procedures to monitor a product or process whose quality is measured across a continuum where a plot of the resulting response forms a profile. Profile monitoring can be considered a relatively new approach in *SPC* that is important when the product or process quality is best represented by a profile at each time period. Profile monitoring combines the concept of fitting profiles using regression techniques and the notion of separating special cause variability from common cause variability in quality control.

The majority of previous studies in profile monitoring focused on parametric mod-

eling of either linear or nonlinear profiles, with both fixed and random-effects, under the assumption of correct model specification, an unrealistic situation in many practical applications. The specific shape of the profile addressed here is not specified. That is, those profiles that can not be well-modeled by a parametric function.

This research studied the fixed and mixed effects models by introducing two new techniques for profile monitoring in Phase I analysis. The first proposed technique was the nonparametric approach via penalized spline regression to obtain the fitted fixed and mixed effects profiles. In addition, we proposed an alternative nonparametric method based on a localized model to estimate fixed effects profiles, and a localized conditional linear mixed model for fitting the mixed effects profiles.

The nonparametric approach via linear mixed penalized regression and local mixed model methods offer population average and cluster specific profile fits with tremendous flexibility. This flexibility is due in part to the fact that they are fit pointwise and therefore able to model trends that the specified parametric model may be incapable of modeling. The a linear penalized regression model with random effects for the intercept and slope is very effective in fitting many complicated profiles and easy accomplished via many commercial statistics packages. In addition, the local model is typically simple: fitting a local linear or local cubic mixed model with a random intercept at each x_0 value will suffice.

The second proposed technique is a semiparametric procedure in which we combine both parametric and nonparametric profile fits via a mixing parameter to gain advantages from both. For the fixed effects models, the $MRPM$ method is an extension of the MRR (Mays et al., 2001) to the area of profile monitoring. For the mixed effects models, the $MMRPM$ method is an extension of the $MRPM$ method which incorporates a mixed model approach to both parametric and nonparametric model fits to account for the correlation within profiles and to deal with the collection of profiles as a random sample from a common population (Waterman et al., 2007).

For each case, we formulated two Hotelling's T^2 statistics, one based on the estimated random effects and one based on the fitted values, and describe how the corresponding UCL values are obtained. In addition, we used two different formulas for

the estimated variance-covariance matrix: one based on the pooled sample variance-covariance matrix estimator and a second one based on the estimated variance-covariance matrix based on successive differences. Moreover, we presented the bandwidth selection method and the asymptotic properties for the mixing parameter for the fixed effects models based on the localized model.

A Monte Carlo study was performed to compare the simulated integrated mean square errors and the simulated probability of signal of the parametric, nonparametric, and semiparametric approaches. Both correlated and uncorrelated errors structure scenarios were evaluated for varying amounts of model misspecification, number of profiles, number of observations per profile, shift location, and in- and out-of-control situations. The semiparametric ($MMRPM$) method for uncorrelated and correlated scenarios was competitive and, often, clearly superior with the parametric and nonparametric over all levels of misspecification. For a correctly specified model, the $SIMSE$ and the simulated probability of signal for the parametric and the $MMRPM$ methods were identical (or nearly so). Since the average of $\hat{\lambda}$ is zero (or nearly so), the $MMRPM$ method results were the same (or nearly so) as the parametric results. For the severe model misspecification case (large values of γ), the nonparametric and $MMRPM$ methods were identical (or nearly so) where the average of $\hat{\lambda}$ was one (or nearly so). For mild model misspecification case, the $MMRPM$ method was superior to the parametric and nonparametric methods. Therefore, this simulation supports the claim that the $MMRPM$ method is robust to model misspecification.

In addition, the $MMRPM$ method performed better for data sets with correlated error structure. Also, the performance of the nonparametric and $MMRPM$ methods improved as the number of observations per profile increases since more observations over the same range of X generally enables more knots to be used by the penalized spline method, resulting in greater flexibility and improved fits in the nonparametric curves and consequently, the semiparametric curves.

In order to determine if a profile is an outlier or if the profile has shifted, we proposed two methods for each approach: one, by analyzing the vector of the fitted values and the second by utilizing the vector of the estimated random effects from the non-

parametric and semiparametric approaches. In Phase I analysis, we are interested in using the HDS to estimate the mean vector μ_b and variance-covariance matrix Σ_b of the $\hat{\mathbf{b}}_i$ vectors, after we have removed the out-of-control profiles. Furthermore, we included in this study two versions for estimating the variance-covariance matrix; one using the pooled variance-covariance matrix (V_P) as given in Equation (2.8), and the other based on successive difference variance-covariance matrix (V_D), as given in Equation (2.9). A simulation study, not presented here, showed that the T^2 statistics using V_P are not efficient in detecting the step shift, regardless of the estimating method, whether parametric, or nonparametric, or semiparametric. This result is consistent with those from Sullivan and Woodall (1996), and Vargas (2003). Thus, this study focuses only on the T^2 statistics based on V_D.

Real data was used in a case study where the parametric, nonparametric and semiparametric approaches were utilized for fitting the relationship between torque produced by an engine and speed in automotive industry. Then, we used a Hotelling's T^2 statistic to conduct Phase I studies to determine the outlying profiles. Here, we focus our attention to the T^2 statistics based on the estimated random effects as given in Equations (4.11), (4.25), and (4.37), and using the successive differences variance-covariance matrix estimator V_D given in Equation (2.9).

The parametric, nonparametric and seminonparametric methods showed that the process was stable even thought the nonparametric and $MMRPM$ methods give much better fits for the engine profiles in term of the smallest mean square errors. Despite the fact that all three methods reach the same conclusion regarding the "in-control" status of each profile, the nonparametric and $MMRPM$ results provide a better description of the actual behavior of each profile. Thus, the nonparametric and $MMRPM$ methods give the user greater ability to properly interpret the true relationship between engine speed and torque for this type of engine and an increased likelihood of detecting unusual engines in future production. Finally, we conclude that the nonparametric and semiparametric approaches performed better than the parametric approach when the user's model is misspecified. The case study demonstrates that, the proposed nonparametric and semiparametric methods are shown to be more efficient, flexible and robust

to model misspecification for Phase I profile monitoring in a real life application.

In summary, the model robust procedures were either superior to or very close to individual parametric and nonparametric procedures. Thus, our methods are robust to the common problem of model misspecification. We also found that both the nonparametric and the semiparametric methods result in charts with good abilities to detect changes in Phase I data, and in charts with easily calculated though conservative control limits. The proposed methods provide greater flexibility and efficiency than current parametric methods used in profile monitoring that rely on correct model specification.

We have outlined a number of issues that have yet to be addressed for the nonparametric approach for profile monitoring, and consequently for the semiparametric approach in the following Section 7.3.

7.3 Future Research

The work presented here is a start in applying mixed penalized spline regression and mixed model robust ideas to the profile monitoring area. Much additional work needs to be done. The following are suggestions for future work in these areas.

7.3.1 Messy Data

This work considered "nice" datasets, those with no missing observations, no data sparseness, and profiles having the same regressor values (balanced design). For example, most simulated cases assumed that the profile sample size (n) was either 10 or 20, and that the design points are uniformly spaced and the same for all profiles. However, longitudinal studies often involve unequal profile sample sizes, non-uniform spacing of the design points as well as other abnormalities such as outliers, and missing data. Messy datasets, as opposed to nice datasets are often seen in the biological and biomedical sciences. It is expected that our methods can accommodate unequal spacing and different sample sizes as well as other complicating characteristics. It would be interesting to see how our methods perform with such data.

7.3.2 Generalized Mixed Models

Our research assumed that the random effects and random errors are normally distributed. However, normality is a common assumption, though not always appropriate. An important advancement of this work would be the extension of the nonparametric and semiparametric approaches to profile monitoring for those situations where the appropriate distribution is one represented by the generalized mixed model, *i.e.*, situations represented by distributions from the exponential family containing random effects.

7.3.3 Multiple Regression

We have extended profile monitoring to the application of nonparametric methods (p-spline, LLR and CLM) based on one regressor. The nonparametric methods can be extended for the multiple regression case, where $P > 2$. The nonparametric methods can also be extended to the higher order coefficients for a single regressor. This will increase the dimension of the vectors that make up the T^2 statistic. From Chapter 2, p-spline regression can be extended easily to the multiple regressors case by adding more columns to the X and Z matrices. Recall from Chapters 2 and 4, the nonparametric profile fits are obtained by LLR and CLM methods which are weighted by the kernel weights. The kernel weights at a specific point are based on the distance measure $(x_l^* - x_j)$. To extend the nonparametric procedure to multiple regression, it is necessary to base the kernel weights on a multidimensional distance measure. Moreover, our semiparametric methods can be extended to the multiple regressor case as well. Both nonparametric methods considered here can easily be extended to the multiple regressor case. However, computational problems such as convergences issues and existence of $REML$ estimates are more likely to arise when considering the mixed model case for nonparametric profile monitoring with multiple regressors.

7.3.4 Alternative Misspecification

The model misspecification considered in this research is that of the mean model. There are a number of other ways which a model can be incorrect, including the variance-covariance structure or effect classification. That is, what would happen to our methods if the variance-covariance structure specified in either the parametric or nonparametric estimation was incorrect?

7.3.5 The $MMR2$ Estimation for The $MMRPM$ Method

The $MMRPM$ method is based upon the model robust regression-1 ($MMR1$) estimate of Einsporn and Birch (1993). The model robust regression-2 ($MMR2$) estimate of Mays et al. (2001) and Mays and Birch (2002) combines a parametric fit with a fraction of the nonparametric fit of the residuals. An $MMR2$ extension to the mixed model and then to the profile monitoring areas would be of interest. Much of the work for the mixed model robust regression based upon $MMR2$ would simply be a generalization of the work of Mays et al. (2001), along with results given in this work. Mays et al. (2001) found that the $MMR2$ estimate performed slightly better than the $MMR1$ estimate. Although the two methods were competitive, it would be interesting to compare the MSE and POS values of the two mixed model robust profile monitoring methods.

7.3.6 Detecting Outliers

In Phase I analysis, we are concerned with identifying a subset of stable data from the HDS with which to estimate the in-control mean vector and variance-covariance matrix. Two examples of the out-of-control situations that we seek to find are multivariate outliers and sustained shifts in the mean vector. Vargas (2003) studied the power properties of the T^2 statistic based on the minimum volume ellipsoid (MVE) mean and variance-covariance matrix estimators, and found it was very effective in detecting multivariate outliers for Phase I analysis. Although, he found that this statistic was not very efficient in detecting a sustained shift in the mean vector. Jensen et al. (2007), Williams et al. (2006) and Sullivan and Woodall (1996) studied the performance of the T^2 statis-

tic based on the successive difference variance-covariance matrix estimator as given in Equation (2.9) and found that this statistic was very effective in detecting a shift in the mean vector. However, this statistic was not very effective in detecting multivariate outliers. Jensen et al. (2007) and Jensen and Birch (2009) compared the MVE and the minimum covariance determinant (MCD) for detecting multivariate outliers in data. They showed that using MCD is sometimes better that MVE and thus MCD would be used on some profile monitoring cases.

For Phase I analysis, a new chart that combines the T^2 control chart based on our proposed methods and the T^2 chart based on either the MVE or the MCD estimators of Vargas (2003) could be proposed. It is though that this chart combination might outperform any of the existing charts.

7.3.7 Asymptotic Theory

Asymptotic theory for the penalized spline regression method should be considered. Because the penalized spline regression method is new, the nonparametric asymptotic theory has yet to be developed. In chapter 3, convergence rates for the theoretically optimal mixing parameter were given for the local method not for the penalized spline regression method. Also, our data is of course correlated, so future work on the asymptotic theory must take into account the fact that the data are correlated. Asymptotic distribution for the T^2 statistic based on the $MMRPM$ fits should be considered for the future work to have more accurate control limits.

Appendix **A**

Fixed Effects Profiles

A.1 Generalized Least Squares Method

Generalized Least Squares (GLS) utilizes to obtain estimates of the mean response in a fixed effects profile model (2.3). It is based on that the vector of random errors has a zero mean vector and variance-covariance R. Under these assumptions, it is easily shown that

$$E(\mathbf{y}) = X\beta \qquad and \qquad var(\mathbf{y}) = R$$

The least squares estimators found by minimizing the generalized sum of squares

$$(\mathbf{y} - X\dot{\beta})^T R^{-1}(\mathbf{y} - X\dot{\beta}) \tag{A.1}$$

for all $\dot{\beta}$. The derivative of the expression in (A.1) with respect to $\dot{\beta}$ by setting (A.1) equal to zero vector, which gives

$$X^T R^{-1}(\mathbf{y} - X\dot{\beta}) = 0 \tag{A.2}$$

The solution for (A.2) gives

$$\hat{\beta} = (X^T R^{-1} X)^{-1} X^T R^{-1}\mathbf{y} \tag{A.3}$$

which is referred to as the GLS estimator. Notice that, the ordinary least squares (OLS) is a special case of GLS with $R = \sigma^2 I$, where I is the $(n \times n)$ identity matrix.

A.2 Expectation and Variance for Local Linear Regression Estimators

Under the assumption that $\mathbf{y}_i \sim MN(X_i\beta_i^{LLR}, R_i),\ \forall\ i = 1,2,\ldots,m$, it is straightforward to calculate the expected value and the variance of $\hat{\beta}_i^{LLR}$. Since,

$$\hat{\beta}_i^{LLR} = (X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} \mathbf{y}_i \tag{A.4}$$

it follow that,

$$\begin{aligned} E(\hat{\beta}_i^{LLR}) = &\quad E\left\{(X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} \mathbf{y}_i\right\} \\ = &\quad (X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} E(\mathbf{y}_i) \\ = &\quad (X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} X_i \beta_i^{LLR} \\ \Rightarrow \quad E(\hat{\beta}_i^{LLR}) = &\ \beta_i^{LLR}. \end{aligned} \tag{A.5}$$

Applying the variance operation to (A.4), we obtain

$$\begin{aligned} Var(\hat{\beta}_i^{LLR}) = &\quad Var\left\{(X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} \mathbf{y}_i\right\} \\ = &\quad (X_i^T R_i^{-1} X_i)^{-1} \tilde{X}^T X_i^T R_i^{-1} [Var(\mathbf{y}_i)] R_i^{-1} X_i (X_i^T R_i^{-1} X_i)^{-1} \\ = &\quad (X_i^T R_i^{-1} X_i)^{-1} X_i^T R_i^{-1} R_i R_i^{-1} X_i (X_i^T R_i^{-1} X_i)^{-1} \\ \Rightarrow \quad Var(\hat{\beta}_i^{LLR}) = &\ (X_i^T R_i^{-1} X_i)^{-1}. \end{aligned} \tag{A.6}$$

Since $\hat{\beta}_i^{LLR}$ is a linear combination of a multivariate normal distribution variable $\mathbf{y}_i$ at $x_0 = x_j, (j = 1,2,\ldots,n')$, one can conclude that $\hat{\beta}_i^{LLR} \sim MN(\beta_i^{LLR}, (X_i^T R_i^{-1} X_i)^{-1})$.

Appendix **B**

Asymptotic Distribution for T^2 Statistic

To find the distribution of $T^2_{LLR,i}$ which depends partially on the distribution of $\hat{\mathbf{y}}^{LLR}_{CS,i,l}$ at x^*_l for every $l = 1,2,\ldots,n'$. We assume that $\mathbf{y} \sim MN(X\beta^{LLR}_l, R_l)$, where β^{LLR}_l is the true parameter vector at x^*_l, and R_l is the known variance-covariance matrix. The n' fitted values at the observed regressors locations at $\tilde{\mathbf{x}}_l$ are given as $\hat{\mathbf{y}}^{LLR}_{CS,i} = \tilde{X}_i\hat{\beta}^{LLR}_i$. Previously, we showed that $\hat{\mathbf{y}}^{LLR}_{CS,i,l} = \tilde{\mathbf{x}}^T_l\hat{\beta}^{LLR}_{il}$ as in (2.49) where

$$\hat{\beta}^{LLR}_{il} = (\tilde{X}^T_i R^{-1}_{il}\tilde{X}_i)^{-1}\tilde{X}^T_i R^{-1}_{il}\mathbf{y}_i \qquad i = 1,2,\ldots,m \qquad l = 1,2,\ldots,n'$$

Since $\hat{\beta}^{LLR}_{il}$ is a linear combination of a multivariate normal distribution variable $\mathbf{y}$ at x^*_l, it follows that $\hat{\beta}^{LLR}_{il} \sim MN(\beta^{LLR}_{il}, (\tilde{X}^T_i R^{-1}_{il}\tilde{X}_i)^{-1})$. The proof is in Appendix (A.2). Therefore, $\hat{\mathbf{y}}^{LLR}_{CS,i,l}$ follows a normal distribution for the i^{th} profile at x^*_l with mean

$$E(\hat{\mathbf{y}}_{CS,i,l}) = E(\tilde{\mathbf{x}}^T_l\hat{\beta}^{LLR}_{il}) = \mu_{PA,l} \qquad l = 1,2,\ldots,n' \tag{B.1}$$

where $\mu_{PA,l}$ is the PA mean at x^*_l, and the variance for the $\hat{\mathbf{y}}^{LLR}_{CS,i,l}$ can be expressed by

$$Var(\hat{\mathbf{y}}^{LLR}_{CS,i,l}) = \tilde{\mathbf{x}}^T_l(\tilde{X}^T_i R^{-1}_{il}\tilde{X}_i)^{-1}\tilde{\mathbf{x}}_l = s_{ll}$$

where s_{ll} represents the l^{th} diagonal element on the variance-covariance matrix (S).

Since $\hat{\mathbf{y}}^{LLR}_{CS,i,l}$ is a linear combination of variables that follow a normal distribution with mean $\mu_{PA,l}$ and variance s_{ll}, the fitted values for the i^{th} profile at the n' observed regressor locations at $\mathbf{x}_l$, $\hat{\mathbf{y}}^{LLR}_{CS,i}$, which follows a multivariate normal distribution with mean zero vector and variance-covariance matrix S, where $S = diag\{s_{ll}\}$, $l = 1,2,\ldots,n'$.

Hence, $(\hat{\mathbf{y}}^{LLR}_{CS,i} - \mu_{PA})^T \sim MN(0,S)$ exactly. To show this in an easy way, let us consider $S^{1/2}$ as a $n' \times n'$ symmetric standard error matrix of CS fits for the i^{th} profile obtained

using the eigen-decomposition of S. Then, $\omega_i = S^{1/2}(\hat{\mathbf{y}}^{LLR}_{CS,i} - \mu_{PA}) \sim MN(0, I)$. Let

$$T^2_{LLR,i} = \omega_i^T \omega_i \qquad\qquad i = 1, 2, \ldots, m$$

which can be considered as a sum of n' standard normal squared variables. It follows that $T^2_{LLR,i} \sim \chi^2_{(n')}$, where n' is the appropriate degree of freedom, a topic to be discussed this later.

If S is unknown, as is often the case, the exact distribution of $T^2_{LLR,i}$ cannot be determined. The regular solution is to estimate S using a sample variance-covariance matrix $(\hat{S})$ of $\hat{\mathbf{y}}^{LLR}_{CS,i}$. Also, μ_{PA} is estimated with $\hat{\mathbf{y}}_{PA}$ using all the in-control phase I data at x^*_l. The distribution of $T^2_{LLR,i}$ can be approximated by an χ^2-distribution for a large sample size with n' degrees of freedom. See Rencher (2000) and Bowman (2006) for more details.

Suppose that $(\hat{\mathbf{y}}^{LLR}_{CS,i} - \hat{\mathbf{y}}^{LLR}_{PA}) \sim MN(0, \hat{S})$ is an unbiased estimator of the $(\hat{\mathbf{y}}^{LLR}_{CS,i} - \mu_{PA})$ for $i = 1, 2, \ldots, m$ at the n' observed regressors locations, then the T^2 statistics for the LLR fits are given by

$$T^2_{LLR,i} = (\hat{\mathbf{y}}^{LLR}_{CS,i} - \hat{\mathbf{y}}^{LLR}_{PA})^T \hat{S}^{-1} (\hat{\mathbf{y}}^{LLR}_{CS,i} - \hat{\mathbf{y}}^{LLR}_{PA}) \qquad i = 1, 2, \ldots, m \tag{B.2}$$

where $\hat{S}$ is an asymptotic consistent estimator for S as $m \to \infty$.

If the i^{th} profile is in-control then

$$T^2_{LLR,i} \sim \chi^2_{(L_1)} \tag{B.3}$$

where L_1 represents the approximate degrees of freedom.

Appendix **C**

Asymptotic Properties with Proofs for the Mixing Parameter

C.1 Asymptotic Theory for the Mixing Parameter

Let us consider that **y** is a real valued response and **x** is a vector valued predictor. The *PA* profile can be written as

$$y_{ij} = \psi(x_{ij}) + \epsilon_{ij} \qquad i = 1,2,\ldots,m, \qquad j = 1,2,\ldots,n_i \tag{C.1}$$

and the i^{th} profile can be expressed as

$$y_{ij} = \psi_i(x_{ij}) + \epsilon_{ij} \qquad i = 1,2,\ldots,m, \qquad j = 1,2,\ldots,n_i$$

where $\psi(.)$ and $\psi_i(.)$ are the true *PA* and *CS* mean functions, respectively. It is assumed that $\epsilon_i \sim MN(0,R_i)$ is a vector of errors following an independent multivariate normal distribution with mean zero vector and variance-covariance matrix R_i. The values of regressors $\mathbf{x}'s$ are fixed uniformly, taking values in a compact set in $\Re^d$, and $\psi = [\psi(x)_{11},\ldots,\psi(x)_{mn_m}]^T$ is a continuous and differentiable function.

Asymptotic theory is applied for the *PA* model which can be extended to the *CS* models in the same way. "Asymptotic" here means that for fixed values of regressors, the number of clusters tends to infinity $i.e.\ m \to \infty$. In other words, since the cluster is an independent unit, the number of observations increases without limitation through the number of clusters.

Let ζ be the parametric estimate, and η represent the nonparametric estimate. Therefore, our proposed model (C.1) can be written as

$$\hat{\psi} = (1-\lambda)\hat{\zeta} + \lambda\hat{\eta} \tag{C.2}$$

where ζ and η are the population average fits.

For any two functions Q_1 and Q_2, we define the inner product and the norm, as in Burman and Chaudhuri (1992), Starnes (1999), Mays et al. (2001) and Waterman (2002), as follow:

$$< Q_1, Q_2 >= (1/n)\sum_{i=1}^{n} Q_1(.)Q_2(.),$$

where Q_1 and Q_2 are two functions of x_{ij}, and the norm is given by

$$\|Q_1\|^2 =< Q_1, Q_1 >= (1/n)\sum_{i=1}^{n} Q_1(.)Q_1(.),$$

where $\|Q_1\| = (< Q_1, Q_1 >)^{1/2}$.

Let us denote the distance between the unknown regression function and the parametric family of continuous regression models by

$$\gamma_m = inf\left\{\left\|\psi - \zeta(\beta)\right\| \ : \ \beta \in R^d\right\}$$

where γ_m measures the shortest distance between the true model and the parametric fit. If we assume that the infimum is unique, then we can replace β by a unique value β^*

$$\gamma_m = \left\|\psi - \zeta(\beta^*)\right\|,$$

where m represents the number of clusters. Therefore, γ_m depends on the number of clusters, and

$$lim_{m\to\infty}\gamma_m = 0, \tag{C.3}$$

especially when the true model ψ is included in the class of the parametric functions under consideration by the user. As well, we have

$$lim_{m\to\infty}\gamma_m \neq 0. \tag{C.4}$$

is for the misspecified parametric model.

Let δ_m represent the mean square distance between the true model and the NP fit

$$\delta_m = E(\left\|\hat{\eta}(x_{ij}) - \psi\right\|^2)$$

Consider the distance between the proposed model fit and the true model function to be

$$\left\|(1-\lambda)\hat{\zeta} + \lambda\hat{\eta} - \psi\right\|$$

The value λ should be selected to minimize the distance between the unknown regression model ψ and our proposed model fit $\hat{\psi}$. By doing some algebra, based on the definitions of the distance and the norm, the optimal value is

$$\lambda^* = \frac{<\hat{\zeta} - \hat{\eta}, \psi> - <\hat{\zeta} - \hat{\eta}, \hat{\zeta}>}{\left\|\hat{\zeta} - \hat{\eta}\right\|^2}$$

The first term in the numerator is unknown and needs to be estimated. The other terms are known. We utilize the leave-one-out approach to estimate λ^* by using the corresponding data driven optimal mixing parameter ($\hat{\lambda}^*$)

$$\hat{\lambda}^* = \frac{<\hat{\zeta}_{-i} - \hat{\eta}_{-i}, Y - \hat{\zeta}>}{\left\|\hat{\zeta} - \hat{\eta}\right\|^2} \tag{C.5}$$

where $\hat{\zeta}_{-i}$ and $\hat{\zeta}_{-i}$ are the parametric and nonparametric fits, respectively, based on all clusters except the i^{th} cluster.

Results are obtained under the following assumptions:

1. $\left\|\zeta(\hat{\beta},.) - \zeta(\beta^*,.)\right\| = O_p(\pi)$

   This assumption is for the parametric fit behavior. $\zeta(\hat{\beta},.)$ can be linear or nonlinear parametric fitted function. This assumption holds for the linear case. To satisfy the above assumption for the nonlinear parametric function we need the regularity conditions on $\zeta(\beta,.)$. For more details on the regularity conditions see Appendix (C.2.4). Therefore this assumption provides the parametric convergence rate between the optimal parametric fit ($\zeta(\beta^*,.)$) and the user parametric fit ($\zeta(\hat{\beta},.)$).

2. The number of responses plus the number of covariates must not exceed the sample size.

3. $\frac{\left\|\hat{\eta}-\psi\right\|^2-E(\left\|\hat{\eta}-\psi\right\|^2)}{E(\left\|\hat{\eta}-\psi\right\|^2)} \rightarrow zero$ in probability as $m \rightarrow \infty$.

   This assumption guarantees that the distance $\left\|\hat{\eta}-\psi\right\| = O_p(\delta_m)$.

4. $lim_{m\rightarrow\infty}(\pi\delta_m^{-1}) = 0$

   This assumption guarantees that the NP fit $\hat{\eta}$ converges at a rate δ_m slower than the parametric fit $\hat{\zeta}$ convergence rate π.

Fulfilling the above assumptions, we can explain the asymptotic behavior of our proposed model, using the following

Lemma C.1.1 *Assuming that conditions (1)through (4) hold.*
(a) If $lim_{m\rightarrow\infty}(\delta_m)$ *is not equal to zero, then we have* $\left\|\hat{\zeta}-\hat{\eta}\right\| = O_p(1)$.
(b) If δ_m *equal to zero, we have* $\left\|\hat{\zeta}-\hat{\eta}\right\| = O_p(\delta_m)$.

This gives the convergence rate of distance between the parametric and NP estimates, which depends on the status of the user parametric fit.

Lemma C.1.2 *Assuming that conditions (1)through (4) are satisfied.*
(a) If $lim_{m\rightarrow\infty}(\delta_m)$ *is not equal to zero, then we have* $\lambda^* = O_p(\delta_m)$.
(b) If δ_m *is equal to zero, we have* $\lambda^* = O_p(\delta_m^{-1}\pi)$.

This gives the convergence rate of the asymptotically optimal mixing parameter. For a misspecified parametric model this is given in (a) and for a correctly specified parametric model in (b).

Theorem C.1.3 *Assuming that conditions (1)through (4) hold.*
(a) If $lim_{m\rightarrow\infty}(\delta_m)$ *is not equal to zero, then we have* $\left\|(1-\lambda)\hat{\zeta}+\lambda\hat{\eta}-\psi\right\| = O_p(\delta_m)$.
(b) If δ_m *equals zero, we have* $\left\|(1-\lambda)\hat{\zeta}+\lambda\hat{\eta}-\psi\right\| = O_p(\pi)$.

This theorem indicates that the distance between the proposed model estimate using the asymptotically optimal mixing parameter and the true regression model converges

at a faster parametric rate if the user has specified the parametric model correctly; otherwise it will converge using the nonparametric convergence rate. One should notice that, the proposed model still uses λ^* not the estimate $\hat{\lambda}^*$. In fact, using $\hat{\lambda}^*$ will be the same as using λ^* in the asymptotic sense. In other words, the convergence rate for the entire proposed robust estimate model will converge as quickly as the slowest of the two component estimates.

C.2 Proof of Asymptotic Theory for the Mixing Parameter

C.2.1 Proof of Lemma (C.1.1)

In the following, we give the proof for Lemma C.1.1

$$\begin{aligned}
\left\|\hat{\xi}-\hat{\eta}\right\|^2 = & \quad \left\|\hat{\xi}-\zeta-\hat{\eta}+\psi-\psi+\zeta\right\|^2 \\
= & \quad \left\|(\hat{\xi}-\zeta)-(\hat{\eta}-\psi)+(\zeta-\psi)\right\|^2 \\
= & \quad \left\|\hat{\xi}-\zeta\right\|^2+\left\|\hat{\eta}-\psi\right\|^2+\left\|\zeta-\psi\right\|^2-2<\hat{\xi}-\zeta,\hat{\eta}-\psi> \\
& \quad +2<\hat{\xi}-\zeta,\zeta-\psi>-2<\hat{\eta}-\psi,\zeta-\psi> \\
= & \quad C_1+C_2+C_3-2C_4+2C_5-2C_6. \qquad \text{(C.6)}
\end{aligned}$$

where:

$$C_1 = O_p(\pi), \ \ and \ \ C_2 = O_p(\delta_m^2)$$

using assumptions (1) and (3), in addition to Cauchy-Schwartz inequality, we can see that:

$$\begin{aligned}
C_4 &\leq \left\|\hat{\xi}-\zeta\right\|\left\|\hat{\eta}-\psi\right\| = O_p(\delta_m\pi) \\
C_5 &\leq \left\|\hat{\xi}-\zeta\right\|\left\|\zeta-\psi\right\| = O_p(\gamma_m\pi) \\
C_6 &\leq \left\|\hat{\eta}-\psi\right\|\left\|\zeta-\psi\right\| = O_p(\gamma_m\delta_m)
\end{aligned}$$

From the previous, it clear that; C_2 dominates C_1 and C_4, also, C_6 dominates C_5. Hence, ((C.6)) can be written as:

$$\begin{aligned}\left\|\hat{\xi}-\hat{\eta}\right\|^2 = &\quad O_p(\delta_m^2)+\left\|\zeta-\psi\right\|^2+\left\|\zeta-\psi\right\|O_p(\gamma_m)\\ = &\quad O_p(\delta_m^2)+O_p(\gamma_m^2)+O_p(\gamma_m)O_p(\delta_m)\end{aligned}$$

It is very easy now, to see that:

$$\left\|\hat{\xi}-\hat{\eta}\right\|^2=\begin{cases}O_p(1) & if\ \ lim_{m\to\infty}(\delta_m)\neq 0\\ O_p(\gamma_m^2) & if\ \ \delta_m=0\end{cases}$$

Q.E.D.

C.2.2 Proof of Lemma (C.1.2)

To prove this Lemma, let us note that:

$$\lambda^* =< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi > \left\|\hat{\xi}-\hat{\eta}\right\|^{-2}$$

If $lim_{m\to\infty}(\delta_m)\neq 0$. We know that

$$|\lambda^*| = \left|< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi >\right| \left\|\hat{\xi}-\hat{\eta}\right\|^{-2}$$

$$\begin{aligned}Since\quad \left|< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi >\right| \leq &\quad \left|< \hat{\xi}-\zeta,\hat{\eta}-\psi >\right|+\left|< \hat{\eta}-\psi,\psi-\zeta >\right|+\left\|\hat{\eta}-\psi\right\|^2\\ \leq &\quad \left\|\hat{\xi}-\zeta\right\|\left\|\hat{\eta}-\psi\right\|+\left\|\hat{\eta}-\psi\right\|\left\|\psi-\zeta\right\|+\left\|\hat{\eta}-\psi\right\|^2\\ = &\quad O_p(\delta_m)\gamma_m+O_p(\delta_m^2)\end{aligned} \tag{C.7}$$

$$\Rightarrow \left|< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi >\right| = O_p(\delta_m)$$

Then, using Lemma (1), we conclude that

$$|\lambda^*| = \left|< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi >\right| \left\|\hat{\xi}-\hat{\eta}\right\|^{-2} = O_p(\delta_m)$$

So, $\lambda^* = O_p(\delta_m)$.

If δ_m. We observing that

$$|1-\lambda^*| = \left|< \hat{\eta}-\hat{\zeta},\hat{\eta}-\psi >\right| \left\|\hat{\xi}-\hat{\eta}\right\|^{-2}$$

$$
\begin{aligned}
|1-\lambda^*| \leq & \quad \left\|\hat{\eta}-\hat{\zeta}\right\| \left\|\hat{\zeta}-\psi\right\| \left\|\hat{\zeta}-\hat{\eta}\right\|^{-2} \\
\leq & \quad \left[\left\|\hat{\zeta}-\zeta\right\| + 1\left\|\zeta-\psi\right\|\right] \left\|\hat{\zeta}-\hat{\eta}\right\|^{-1} \\
= & \quad \left[O_p(\pi)+\delta_m\right] \left\|\hat{\zeta}-\hat{\eta}\right\|^{-1}
\end{aligned}
$$

using Lemma (1) and $\delta_m = 0$, then $\lambda^* = O_p(\pi\delta_m^{-1})$ Therefore

$$
\lambda^* = \begin{cases} O_p(\delta_m) & if \;\; lim_{m\to\infty}(\delta_m) \neq 0 \\ O_p(\pi\delta_m^{-1}) & if \;\; \delta_m = 0 \end{cases}
$$

Q.E.D.

C.2.3 Proof of the Theorem (C.1.3)

In the following, we give a proof for Theorem C.1.3. For more details see Burman and Chaudhuri (1992)

$$
\begin{aligned}
\left\|(1-\lambda^*)\hat{\zeta}+\lambda^*\hat{\eta}-\psi\right\| \leq & \quad |1-\lambda^*|\left\|\hat{\zeta}-\psi\right\| + |\lambda^*|\left\|\hat{\eta}-\psi\right\| \\
\leq & \quad |1-\lambda^*|\left\{\left\|\hat{\zeta}-\zeta\right\| + \left\|\zeta-\psi\right\|\right\} + |\lambda^*|\left\|\hat{\eta}-\psi\right\| \\
= & \quad |1-\lambda^*|\left\{O_p(\pi)+\gamma_m\right\} + |\lambda^*| O_p(\delta_m)
\end{aligned}
$$

If $lim_{m\to\infty}(\delta_m) \neq 0$. We observing that

$$
\left\|(1-\lambda^*)\hat{\zeta}+\lambda^*\hat{\eta}-\psi\right\| = O_p(\delta_m)\left\{O_p(\pi)+\gamma_m\right\} + O_p(\delta_m)^2 = O_p(\delta_m)
$$

If $\delta_m = 0$. We observing that

$$
\left\|(1-\lambda^*)\hat{\zeta}+\lambda^*\hat{\eta}-\psi\right\| = O_p(\delta_m^{-1}\pi)\left\{O_p(\pi)+0\right\} + O_p(\pi) = O_p(\pi)
$$

Therefore we proved that:

$$
\left\|(1-\lambda^*)\hat{\zeta}+\lambda^*\hat{\eta}-\psi\right\| = \begin{cases} O_p(\delta_m) & if \;\; lim_{m\to\infty}(\delta_m) \neq 0 \\ O_p(\pi) & if \;\; \delta_m = 0 \end{cases}
$$

Q.E.D.

C.2.4 Regularity Conditions

For more details see Gallant (1987); Vonesh and Chinchilli (1997)

1. $f(.)$ is a twice continuous differentiable vector-valued function in $\beta \in \Omega$, which is a compact subspace of $\Re^d$, and the true parameter β^* is an interior point of Ω.

2. Assume that X_i and $\mathbf{b}_i$ originate from compact spaces of Re^t and Re^q respectively, such that their empirical $cdf's \rightarrow cdf's$ as $n \rightarrow \infty$.

3. Let $D_i(.) = [f(.) \;\; F_i(.) \;\; \frac{\partial F_i(.)}{\partial \beta}]$ where $F_i() = \frac{\partial F_i(.)}{\partial \beta}$, it is assumed that the following limits are all converge uniformly in $\beta \in \Omega$.

   - $lim_n \rightarrow \infty [n^{-1} \sum_i F_i(.)^T F_i(.)]$ and $lim_n \rightarrow \infty [n^{-1} \sum_i F_i(.)^T V F_i(.)]$ are converge to positive definite matrices.
   - $lim_n \rightarrow \infty [n^{-1} \sum_i D_i(.)^T D_i(.)]$

4. $lim_n \rightarrow \infty [n^{-1} \sum_i (f_i(.) - f^*(.))^T V^{-1} (f_i(.) - f^*(.)) = G(., V)]$ where $f^*(.)$ represents the function with the true parameters, converges uniformly in $\beta \in \Omega$ with $G(., V) = 0 \Leftrightarrow \beta = \beta^*$, where β^* is true parameter.

5. The ϵ_i are iid with mean 0 and arbitrary positive definite covariance matrix $\sum$.

Appendix **D**

Random Effects Models

D.1 CLM Derivations

For known $\tilde{B}$ and $\tilde{R}$, variance-covariance matrices, the estimator β_0 and the estimated best linear unbiased predictions ($EBLUP_s$) for the random effects (deviations) from the PA at x_0 can be obtained using the ML approach.

The main idea of the ML approach is to maximize the joint localized likelihood function of $\mathbf{b}_0$ and $\tilde{\epsilon}_0$. Where $\mathbf{b}_0$ and $\tilde{\epsilon}_0$ are assumed to be independent and normally distributed. The joint density is

$$f(\mathbf{b}_0,\tilde{\epsilon}_0)=c\times\exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{y}-\tilde{X}\beta_0-\tilde{Z}\mathbf{b}_0\\ \mathbf{b}_0\end{bmatrix}^T\begin{bmatrix}\tilde{R}&0\\0&\tilde{B}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}-\tilde{X}\beta_0-\tilde{Z}\mathbf{b}_0\\ \mathbf{b}_0\end{bmatrix}\right\}$$

where $c=(2\pi)^{-\frac{n+mp_2}{2}}\left|\begin{matrix}\tilde{R}&0\\0&\tilde{B}\end{matrix}\right|^{-1/2}$, $\tilde{R}=diag(\tilde{R}_1,\tilde{R}_2,\ldots,\tilde{R}_n)$ is $(n\times n)$ block diagonal variance-covariance matrix for residuals, and $\tilde{B}=diag(\tilde{D})$ represents a block diagonal variance-covariance matrix for random effects.

By taking the logarithm for the joint density of $\mathbf{b}_0$ and $\tilde{\epsilon}_0$, we get

$$\log f(\mathbf{b}_0,\tilde{\epsilon}_0)=C-(1/2)[(\mathbf{y}-\tilde{X}\beta_0-\tilde{Z}\mathbf{b}_0)^T\tilde{R}^{-1}(\mathbf{y}-\tilde{X}\beta_0-\tilde{Z}\mathbf{b}_0)+\mathbf{b}_0^T\tilde{B}^{-1}\mathbf{b}_0] \tag{D.1}$$

where $C=log(c)$. The first term in the right hand side of (D.1), ignoring C, represents the weighted residuals taking the within-cluster variation into account, while the second term represents the penalty due to random effects $\mathbf{b}_0$ taking the between-clusters variation into account.

The normal equations can be obtain by differentiating (*D*.1) with respect to β_0 and $\mathbf{b}_0$ then equate the results to zero, as follows

$$\begin{aligned}\frac{\partial \log f(.)}{\partial \beta_0^T} = \quad & \tfrac{\partial}{\partial \beta_0^T}[(\mathbf{y}^T\tilde{R}^{-1}\mathbf{y}-\mathbf{y}^T\tilde{R}^{-1}\tilde{X}\beta_0-\mathbf{y}^T\tilde{R}^{-1}\tilde{Z}\mathbf{b}_0-\beta_0^T\tilde{X}\tilde{R}^{-1}\mathbf{y}+\beta_0^T\tilde{X}\tilde{R}^{-1}\tilde{X}\beta_0 \\ & +\beta_0^T\tilde{X}\tilde{R}^{-1}\tilde{Z}\mathbf{b}_0-\mathbf{b}_0^T\tilde{Z}^T\tilde{R}^{-1}\mathbf{y}+\mathbf{b}_0^T\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\beta_0+\mathbf{b}_0^T\tilde{Z}^T\tilde{R}^{-1}\tilde{Z}\mathbf{b}_0+\mathbf{b}_0^T\tilde{B}^{-1}\mathbf{b}_0] \\ \Rightarrow \quad & (-\tfrac{1}{2})[2\tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0-2\tilde{X}^T\tilde{R}^{-1}\mathbf{y}+2\tilde{X}\tilde{R}^{-1}\tilde{Z}\hat{\mathbf{b}}_0]=0 \\ \Rightarrow \quad & \tilde{X}^T\tilde{R}^{-1}\mathbf{y}=\tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0+\tilde{X}^T\tilde{R}^{-1}\tilde{Z}\hat{\mathbf{b}}_0 \qquad \text{(D.2)}\end{aligned}$$

The first partial derivative of $log\, f(.)$ with respect to $\mathbf{b}_0^T$ is

$$\begin{aligned}\frac{\partial \log f(.)}{\partial \mathbf{b}_0^T} &= (\frac{-1}{2})[2\tilde{B}^{-1}\mathbf{b}_0+2\tilde{Z}^T\tilde{R}^{-1}\tilde{Z}\mathbf{b}_0-2\tilde{Z}^T\tilde{R}^{-1}\mathbf{y}+2\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\beta_0] \\ &\Rightarrow \; -[\tilde{B}^{-1}\hat{\mathbf{b}}_0+\tilde{Z}^T\tilde{R}^{-1}\tilde{Z}\hat{\mathbf{b}}_0-\tilde{Z}^T\tilde{R}^{-1}\mathbf{y}+\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0]=0 \\ &\Rightarrow \; \tilde{Z}^T\tilde{R}^{-1}\mathbf{y}=(\tilde{B}^{-1}+\tilde{Z}^T\tilde{R}^{-1}\tilde{Z})\hat{\mathbf{b}}_0+\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0 \qquad \text{(D.3)}\end{aligned}$$

The normal equations (*D*.2) and (*D*.3) can be written in matrix form as

$$\begin{bmatrix} \tilde{X}^T\tilde{R}^{-1}\tilde{X} & \tilde{X}^T\tilde{R}^{-1}\tilde{Z} \\ \tilde{Z}^T\tilde{R}^{-1}\tilde{X} & (\tilde{B}^{-1}+\tilde{Z}^T\tilde{R}^{-1}\tilde{Z}) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\mathbf{b}}_0 \end{bmatrix} = \begin{bmatrix} \tilde{X}^T\tilde{R}^{-1}\mathbf{y} \\ \tilde{Z}^T\tilde{R}^{-1}\mathbf{y} \end{bmatrix}$$

From (*D*.3), $\hat{\mathbf{b}}_0^C$ can be obtained as

$$\hat{\mathbf{b}}_0^C=(\tilde{B}^{-1}+\tilde{Z}^T\tilde{R}^{-1}\tilde{Z})^{-1}\tilde{Z}^T\tilde{R}^{-1}(\mathbf{y}-\tilde{X}\hat{\beta}_0^C)$$

Taking $V_0=\tilde{Z}\tilde{B}\tilde{Z}^T+\tilde{R}$, it can be proved that $G=(\tilde{B}^{-1}+\tilde{Z}^T\tilde{R}^{-1}\tilde{Z})^{-1}=(\tilde{B}-\tilde{B}\tilde{Z}V_0^{-1}\tilde{Z}\tilde{B})$ and then $V_0^{-1}=\tilde{R}^{-1}-\tilde{R}^{-1}\tilde{Z}G\tilde{Z}^T\tilde{R}^{-1}$, see (Harville, 1977), and (Waterman, 2002) for more details. The above formula for $\hat{\mathbf{b}}_0^C$ can be simplified as follows

$$\hat{\mathbf{b}}_0^C=(\tilde{B}-\tilde{B}\tilde{Z}V_0^{-1}\tilde{Z}\tilde{B})\tilde{Z}^T\tilde{R}^{-1}(\mathbf{y}-\tilde{X}\hat{\beta}_0^C)$$

Therefore,

$$\hat{\mathbf{b}}_0^C=\tilde{B}\tilde{Z}^T\tilde{R}^{-1}\mathbf{y}-\tilde{B}\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C-\tilde{B}\tilde{Z}V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T\tilde{R}^{-1}\mathbf{y}+\tilde{B}\tilde{Z}V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C$$

$$\Rightarrow =\tilde{B}\tilde{Z}^T[I-V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T]\tilde{R}^{-1}\mathbf{y}+\tilde{B}\tilde{Z}[V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T-I]\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C$$

$$\begin{aligned}\Rightarrow = \quad & \tilde{B}\tilde{Z}^T[I - V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T - V_0^{-1}\tilde{R} + V_0^{-1}\tilde{R}]\tilde{R}^{-1}\mathbf{y} \\ + \quad & \tilde{B}\tilde{Z}[V_0^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T - I - V_0^{-1}\tilde{R} + V_0^{-1}\tilde{R}]\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C\end{aligned}$$

$$\begin{aligned}\Rightarrow = \quad & \tilde{B}\tilde{Z}^T[I - V_0^{-1}(\tilde{Z}\tilde{B}\tilde{Z}^T + \tilde{R}) + V_0^{-1}\tilde{R}]\tilde{R}^{-1}\mathbf{y} \\ + \quad & \tilde{B}\tilde{Z}[V_0^{-1}(\tilde{Z}\tilde{B}\tilde{Z}^T + \tilde{R}) - I - V_0^{-1}\tilde{R}]\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C\end{aligned}$$

$$\Rightarrow = \tilde{B}\tilde{Z}^T V_0^{-1}\mathbf{y} - \tilde{B}\tilde{Z}V_0^{-1}\tilde{X}\hat{\beta}_0^C$$

$$\hat{\mathbf{b}}_0^C = \tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \tag{D.4}$$

By substituting (*D*.4) into (*D*.2) we get

$$\begin{aligned}\tilde{X}^T\tilde{R}^{-1}\mathbf{y} = \quad & \tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C + \tilde{X}^T\tilde{R}^{-1}\tilde{Z}\hat{\mathbf{b}}_0^C = \tilde{X}^T\tilde{R}^{-1}[\tilde{X}\hat{\beta}_0^C + \tilde{Z}(\tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C))] \\ = \quad & \tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C + \tilde{X}^T\tilde{R}^{-1}\tilde{Z}\tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \\ & -\tilde{X}^T\tilde{R}^{-1}\tilde{R}V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) + \tilde{X}^T\tilde{R}^{-1}\tilde{R}V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \\ = \quad & \tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C - \tilde{X}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) + \tilde{X}^T\tilde{R}^{-1}(\tilde{R} + \tilde{Z}\tilde{B}\tilde{Z}^T)V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \\ = \quad & \tilde{X}^T\tilde{R}^{-1}\tilde{X}\hat{\beta}_0^C - \tilde{X}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) + \tilde{X}^T\tilde{R}^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C) \\ \Rightarrow \quad & \tilde{X}^T V_0^{-1}\tilde{X}\hat{\beta}_0^C = \tilde{X}^T V_0^{-1}\mathbf{y}\end{aligned}$$

If $(\tilde{X}^T V_0^{-1}\tilde{X})$ is nonsingular, then its inverse can exist. Therefore

$$\hat{\beta}_0^C = (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\mathbf{y} \tag{D.5}$$

If the inverse is not exist then $\hat{\beta}_0^C = (\tilde{X}^T V_0^{-1}\tilde{X})^{-}\tilde{X}^T V_0^{-1}\mathbf{y}$.

D.2 Derivations for the Mean and the Variance for CLM Estimators

If we assume that $\mathbf{y} \sim MN(\tilde{X}\beta_0^C, V_0)$.

$$\hat{\beta}_0^C = (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\mathbf{y}.$$

it follows that,

$$\begin{aligned} E(\hat{\beta}_0^C) = &\quad E\left\{(\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\mathbf{y}\right\} \\ = &\quad (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1} E(\mathbf{y}) \\ = &\quad (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\tilde{X}\beta_0^C \\ \Rightarrow &\quad E(\hat{\beta}_0^C) = \beta_0^C. \end{aligned}$$

Applying the variance operator to $\hat{\beta}_0^C$ we obtain

$$\begin{aligned} Var(\hat{\beta}_0^C) = &\quad Var\left\{(\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\mathbf{y}\right\} \\ = &\quad (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}[Var(\mathbf{y})]V_0^{-1}\tilde{X}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} \\ = &\quad (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1} V_0 V_0^{-1}\tilde{X}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} \\ \Rightarrow &\quad Var(\hat{\beta}_0^C) = (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}. \end{aligned}$$

Since $\hat{\beta}_0^C$ is a linear combination of a multivariate normal distribution variable $\mathbf{y}$ at x_l^*, $l = 1,2,\ldots,n'$, it follows that $\hat{\beta}_0^C \sim MN(\beta_0^C, (\tilde{X}^T V_0^{-1}\tilde{X})^{-1})$.

D.2.1 Derivations for the Mean and the Variance for CLM Predictors

We showed that the EBLUP is $\hat{\mathbf{b}}_0^C = \tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C)$. By taking the expectation for $\hat{\mathbf{b}}_0^C$,

$$\begin{aligned} E(\hat{\mathbf{b}}_0^C) = &\quad E\left\{\tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C)\right\} \\ = &\quad \tilde{B}\tilde{Z}^T V_0^{-1}(E(\mathbf{y}) - \tilde{X}E(\hat{\beta}_0^C)) \\ = &\quad \tilde{B}\tilde{Z}^T V_0^{-1}(\tilde{X}\beta_0^C - \tilde{X}\beta_0^C)) \\ \Rightarrow &\quad E(\hat{\mathbf{b}}_0^C) = 0 \end{aligned}$$

where the CLM model is assumed to be correctly specified.

Applying the variance operator to $\hat{\mathbf{b}}_0^C$ we obtain

$$Var(\hat{\mathbf{b}}_0^C) = Var\left\{\tilde{B}\tilde{Z}^T V_0^{-1}(\mathbf{y} - \tilde{X}\hat{\beta}_0^C)\right\}$$

$$\Rightarrow Var(\hat{\mathbf{b}}_0^C) = \quad \tilde{B}\tilde{Z}^T V_0^{-1}[I - (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}]Var(\mathbf{y})[\tilde{B}\tilde{Z}^T V_0^{-1}(I - (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1})]^T$$

$$
\begin{aligned}
&= [\tilde{B}\tilde{Z}^T V_0^{-1} - \tilde{B}\tilde{Z}^T V_0^{-1}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}]\, V_0 [V_0^{-1}\tilde{Z}\tilde{B}^T - V_0^{-1}\tilde{X}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} V_0^{-1}\tilde{Z}\tilde{B}^T]^T \\
&= \tilde{B}\tilde{Z}^T V_0^{-1}\tilde{Z}\tilde{B}^T - \tilde{B}\tilde{Z}^T V_0^{-1}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} V_0^{-1}\tilde{Z}\tilde{B}^T \\
&\quad - \tilde{B}\tilde{Z}^T V_0^{-1}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1}\tilde{Z}\tilde{B}^T + \tilde{B}\tilde{Z}^T V_0^{-1}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} V_0^{-1}\tilde{Z}\tilde{B}^T \\
Var(\hat{\mathbf{b}}_0^C) &= \tilde{B}\tilde{Z}^T V_0^{-1}[I - \tilde{X}(\tilde{X}^T V_0^{-1}\tilde{X})^{-1} V_0^{-1} - (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}\tilde{X}^T V_0^{-1} + (\tilde{X}^T V_0^{-1}\tilde{X})^{-1}]\tilde{Z}\tilde{B}^T = V_0^*.
\end{aligned}
$$

From the above, one can see that $\hat{\mathbf{b}}_0^C \sim MN(0, V_0^*)$.

D.2.2 Derivations for the Mean and the Variance for CLM Estimators/ Predictors for the i^{th} Profile

Since $\hat{\mathbf{y}}_{CS,i,l} = \mathbf{x}_l^T \hat{\beta}_l^C + \mathbf{z}_{i,l}^T \hat{\mathbf{b}}_l^C$ Therefore, $\hat{\mathbf{y}}_{CS,i,l}$ follows a multivariate normal distribution with mean and variance-covariance matrix for the i^{th} profile as the following

$$
\begin{aligned}
E(\hat{\mathbf{y}}_{CS,i,l}) = &\quad E(\mathbf{x}_l^T \hat{\beta}_l^C + \mathbf{z}_{i,l}^T \hat{\mathbf{b}}_l) \\
= &\quad \mathbf{x}_l^T \beta_l^C = \mathbf{y}_{PA,l} = \mu_{PA,l}
\end{aligned}
$$

The variance for the $\hat{\mathbf{y}}_{CS,i,l}$ can be expressed by

$$
Var(\hat{\mathbf{y}}_{CS,i,l}) = Var(\mathbf{x}_l^T \hat{\beta}_l^C + \mathbf{z}_{i,l}^T \hat{\mathbf{b}}_l)
$$

$$
\Rightarrow = Var\left\{[\mathbf{x}_l^T(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T V_l^{-1} + \mathbf{z}_{i,l}^T \tilde{B}\tilde{Z}^T V_l^{-1} - \mathbf{z}_{i,l}^T \tilde{B}\tilde{Z}^T V_l^{-1}\tilde{X}(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T V_l^{-1}]\mathbf{y}\right\}
$$

$$
Var(\hat{\mathbf{y}}_{CS,i,l}) = A V_l A
$$

where $A = [\mathbf{x}_l^T(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T V_l^{-1} + \mathbf{z}_{i,l}^T \tilde{B}\tilde{Z}^T V_l^{-1} - \mathbf{z}_{i,l}^T \tilde{B}\tilde{Z}^T V_l^{-1}\tilde{X}(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T V_l^{-1}]$, and it is easy to show that

$$
Var(\hat{\mathbf{y}}_{CS,i,l}) = \mathbf{x}_l^T(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T + \mathbf{z}_{i,l}^T \tilde{B}\tilde{Z}^T V_l^{-1}[I - \tilde{X}(\tilde{X}^T V_l^{-1}\tilde{X})^{-1}\tilde{X}^T V_l^{-1}]\tilde{Z}\tilde{B}\mathbf{z}_{i,l}^T = s_{ll}
$$

Appendix **E**

The Equality Between the Parametric T^2 Based on Fitted Values and Estimated Parameters Using Moment Variance-Covariance

E.1 The T^2 Statistic Based on the Parametric Fitted Values

From Chapter 4, we know that the T^2 statistics based on the parametric fitted values are given by

$$T^2_{Par1,i} = (\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA})^T \hat{V}^{-}(\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA}) \tag{E.1}$$

where $\hat{V}$ can be estimated using moment estimated variance-covariance $\hat{V}_P$, and $\hat{V}^{-}$ represents the estimated generalized inverse variance-covariance matrix. Let us consider $\hat{\mathbf{Y}}_{PA} = \frac{\sum_{i=1}^{m} \hat{\mathbf{Y}}_{CS,i}}{m}$.

Since

$$\begin{aligned}(\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA}) = & \quad [(X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i) - \textstyle\sum_{i=1}^{m}(X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i)/m] \\ = & \quad X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i - \frac{\sum_{i=1}^{m} X_i\hat{\beta}}{m} - \frac{\sum_{i=1}^{m} Z_i\hat{\mathbf{b}}_i}{m} \\ = & \quad Z_i\hat{\mathbf{b}}_i \end{aligned} \tag{E.2}$$

Notice that, $\frac{\sum_{i=1}^{m} Z_i\hat{\mathbf{b}}_i}{m} = \frac{Z\sum_{i=1}^{m} \hat{\mathbf{b}}_i}{m} = 0$, if and only if all $Z_i = Z$. For example, we assume that all the measurements are taken at the same n locations for each profile which is

common situation in the real industrial applications.

In addition, this means that, $X_i = X$ and $V_i = V$. Also, we have

$$\begin{aligned}\hat{V}_P = & \frac{\sum_{i=1}^{m}(\hat{\mathbf{Y}}_{CS,i}-\hat{\mathbf{Y}}_{PA})(\hat{\mathbf{Y}}_{CS,i}-\hat{\mathbf{Y}}_{PA})^T}{m-1} \\ = & \frac{\sum_{i=1}^{m}(Z_i\hat{\mathbf{b}}_i)(Z_i\hat{\mathbf{b}}_i)^T}{m-1} \\ = & Z(\frac{\sum_{i=1}^{m}\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T}{m-1})Z^T. \end{aligned} \tag{E.3}$$

By substituting from (*E*.2) and (*E*.3) in (*E*.1), then

$$\begin{aligned}T^2_{Par1,i} = & \hat{\mathbf{b}}_i^T Z_i^T [Z(\frac{\sum_{i=1}^{m}\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T}{m-1})Z^T]^- Z_i\hat{\mathbf{b}}_i \\ = & \hat{\mathbf{b}}_i^T Z^T [Z(\frac{\sum_{i=1}^{m}\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T}{m-1})Z^T]^- Z\hat{\mathbf{b}}_i. \end{aligned} \tag{E.4}$$

E.2 The T^2 Statistic Based on the Estimated Random Effects ($EBLUPS$)

From Chapter 4, we have the T^2 statistics based on the estimated random effects values are given by

$$T^2_{Par2,i} = (\hat{\mathbf{b}}_i)^T \hat{V}^-(\hat{\mathbf{b}}_i) \tag{E.5}$$

where $\hat{V}$ can be estimated using moment estimated variance-covariance ($\hat{V}_P$). Such as,

$$\hat{V}_P = \frac{\sum_{i=1}^{m}(\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_i)^T}{m-1} \tag{E.6}$$

By substituting from (*E*.6) into (*E*.5), then

$$T^2_{Par2,i} = \hat{\mathbf{b}}_i^T [\frac{\sum_{i=1}^{m}\hat{\mathbf{b}}_i\hat{\mathbf{b}}_i^T}{m-1}]^-\hat{\mathbf{b}}_i \tag{E.7}$$

By the properties of the generalized inverse, we know that $\hat{\mathbf{b}}_i^T Z^T Z^{T-} = \hat{\mathbf{b}}_i^T$ and $Z^- Z\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_i$. Thus, it follows that (*E*.4) is equal to (*E*.7) and, consequently, $T^2_{Par1,i} \equiv T^2_{Par2,i}$.

Appendix **F**

The Equality Between the Parametric T^2 Based on Fitted Values and Estimated Parameters Using Successive Difference Variance-Covariance

F.1 The T^2 Statistic Based on the Parametric Fitted Values

From Chapter 4, we know that the T^2 statistics based on the parametric fitted values using the successive difference variance covariance matrix are given by

$$T^2_{Par1,i} = (\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA})^T \hat{V}^{-}(\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA}) \tag{F.1}$$

where $\hat{V}$ can be estimated based on successive difference variance-covariance $\hat{V}_D$, and $\hat{V}^{-}$ represents the estimated generalized inverse variance-covariance matrix. Let us consider $\hat{\mathbf{Y}}_{PA} = \frac{\sum_{i=1}^{m} \hat{\mathbf{Y}}_{CS,i}}{m}$.

Since

$$\begin{aligned}(\hat{\mathbf{Y}}_{CS,i} - \hat{\mathbf{Y}}_{PA}) = &\quad [(X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i) - \textstyle\sum_{i=1}^{m}(X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i)/m] \\ = &\quad X_i\hat{\beta} + Z_i\hat{\mathbf{b}}_i - \frac{\sum_{i=1}^{m} X_i\hat{\beta}}{m} - \frac{\sum_{i=1}^{m} Z_i\hat{\mathbf{b}}_i}{m} \\ = &\quad Z_i\hat{\mathbf{b}}_i \end{aligned} \tag{F.2}$$

Notice that, $\frac{\sum_{i=1}^{m} Z_i\hat{\mathbf{b}}_i}{m} = \frac{Z\sum_{i=1}^{m} \hat{\mathbf{b}}_i}{m} = 0$, if and only if all $Z_{i+1} = Z_i = Z$. Here, it is assumed that

all measurements are taken at the same n locations for each profile which is common situation in the real industrial applications. In addition, this means that, $X_{i+1} = X_i = X$ and $V_{i+1} = V_i = V$. Also,

$$\begin{aligned}\hat{V}_D = & \frac{\sum_{i=1}^{m-1}(\hat{\mathbf{Y}}_{CS,i+1}-\hat{\mathbf{Y}}_{CS,i})(\hat{\mathbf{Y}}_{CS,i+1}-\hat{\mathbf{Y}}_{CS,i})^T}{2(m-1)} \\ = & \frac{\sum_{i=1}^{m-1}[(X_{i+1}\hat{\beta}+Z_{i+1}\hat{\mathbf{b}}_{i+1})-(X_i\hat{\beta}+Z_i\hat{\mathbf{b}}_i)][(X_{i+1}\hat{\beta}+Z_{i+1}\hat{\mathbf{b}}_{i+1})-(X_i\hat{\beta}+Z_i\hat{\mathbf{b}}_i)]^T}{2(m-1)} \\ = & Z(\frac{\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)^T}{2(m-1)})Z^T \end{aligned} \tag{F.3}$$

By substituting from (F.2) and (F.3) in (F.1), then

$$\begin{aligned}T^2_{Par1,i} = & \hat{\mathbf{b}}_i^T Z_i^T [Z_i(\frac{\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)^T}{2(m-1)})Z_i^T]^- Z_i\hat{\mathbf{b}}_i \\ = & \hat{\mathbf{b}}_i^T Z^T [Z(\frac{\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)^T}{2(m-1)})Z^T]^- Z\hat{\mathbf{b}}_i \end{aligned} \tag{F.4}$$

F.2 The T^2 Statistic Based on the Estimated Random Effects ($EBLUPS$)

In Chapter 4, we know that the T^2 statistics based on the estimated random effects values using the successive difference variance covariance matrix are given by

$$T^2_{Par2,i} = (\hat{\mathbf{b}}_i)^T \hat{V}^-(\hat{\mathbf{b}}_i) \tag{F.5}$$

where $\hat{V}$ can be estimated using successive difference variance-covariance ($\hat{V}_D$). Such as,

$$\hat{V}_D = \frac{\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)^T}{2(m-1)} \tag{F.6}$$

By substituting from (F.6) into (F.5), then

$$T^2_{Par1,i} = \hat{\mathbf{b}}_i^T [(\frac{\sum_{i=1}^{m-1}(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)(\hat{\mathbf{b}}_{i+1}-\hat{\mathbf{b}}_i)^T}{2(m-1)})]^- \hat{\mathbf{b}}_i \tag{F.7}$$

By the properties of the generalized inverse, we know that $\hat{\mathbf{b}}_i^T Z^T Z^{T-} = \hat{\mathbf{b}}_i^T$ and $Z^- Z\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_i$. Thus, it follows that (F.4) is equal to (F.7) and, consequently, $T^2_{Par1,i} \equiv T^2_{Par2,i}$ using the successive difference variance-covariance matrix.

Appendix **G**

The Automotive Industry Data Set

Table G.1: The Automotive Industry Data 26 Automobile Engine, Torque (T) vs. RPM

RPM (x)	$T-E329$	$T-E449$	$T-E529$	$T-E642$	$T-E724$	$T-E803$	$T-E930$
1500	98.53	96.35	98.77	96.7	96.75	97.61	100.06
2000	102.65	100.74	103.03	100.05	100.87	102.46	103.6
2500	113.82	110.67	111.99	111.17	110.14	112.18	112.74
2660	115.26	113.06	112.78	111.51	110.48	112.99	113.56
2800	116.24	114.58	113.14	112.01	110.94	114.54	112.85
2940	117.06	114.98	113.73	111.23	111.17	115	114.49
3500	109.89	108.55	110.3	105.64	105.78	108.99	108.95
4000	109.65	107.41	109.35	106.02	103.37	107.95	108.24
4500	105.72	103.9	107.61	103.11	102.23	103.65	105.56
5000	99.74	97.99	100.64	97.4	96.06	96.94	98.92
5225	95.97	94.27	97.59	93.88	92.39	92.78	95.41
5500	89.47	88.45	91.68	88.17	86.54	86.41	89.19
5775	81.96	81.44	84.58	81.18	79.31	78.6	81.85
6000	74.9	75	77.48	75.03	73.13	71.97	75.09
RPM (x)	$T-E1148$	$T-E1171$	$T-E1516$	$T-E1791$	$T-E2600$	$T-E3100$	$T-E3720$
1500	98.53	96.35	98.77	96.7	96.75	97.61	100.06
2000	102.65	100.74	103.03	100.05	100.87	102.46	103.6
2500	113.82	110.67	111.99	111.17	110.14	112.18	112.74
2660	115.26	113.06	112.78	111.51	110.48	112.99	113.56
2800	116.24	114.58	113.14	112.01	110.94	114.54	112.85
2940	117.06	114.98	113.73	111.23	111.17	115	114.49
3500	109.89	108.55	110.3	105.64	105.78	108.99	108.95
4000	109.65	107.41	109.35	106.02	103.37	107.95	108.24
4500	105.72	103.9	107.61	103.11	102.23	103.65	105.56
5000	99.74	97.99	100.64	97.4	96.06	96.94	98.92
5225	95.97	94.27	97.59	93.88	92.39	92.78	95.41
5500	89.47	88.45	91.68	88.17	86.54	86.41	89.19
5775	81.96	81.44	84.58	81.18	79.31	78.6	81.85
6000	74.9	75	77.48	75.03	73.13	71.97	75.09

Table G.2: The Automotive Industry Data 26 Automobile Engine, Torque (T) vs. RPM (Cont.)

RPM (x)	$T-E4025$	$T-E4068$	$T-E4926$	$T-E5155$	$T-E6143$	$T-E6844$	$T-E7811$
1500	100	97.98	97.29	100.97	93.13	93.11	95.38
2000	103.27	104.98	105.86	104.96	101.02	103.43	101.25
2500	111.46	114.9	115.25	112.47	111.25	112.02	111.53
2660	111.8	116.06	117.83	113.04	111.83	113.2	112.11
2800	113.02	116.65	117.97	113.91	113.27	113.77	112.6
2940	113.55	116.18	117.77	114.22	113.04	113.77	111.76
3500	108.87	109.65	111.31	110.46	105.6	109.15	108.12
4000	107.6	109.06	110.97	109.66	106.15	108.05	106.62
4500	104.44	105.01	107.37	106.79	104.12	103.46	102.92
5000	97.5	97.43	100.53	100.27	97.45	98.26	96.35
5225	94.04	94.04	97.17	96.48	94.68	94.26	93.14
5500	87.89	87.51	90.47	90.74	88.59	89.09	86.75
5775	80.04	79.36	83.51	83.5	81.08	81.06	80.27
6000	73.18	72.34	76.34	76.38	75.77	74.14	73.47

RPM (x)	$T-E8007$	$T-E8623$	$T-E9388$	$T-E9404$	$T-E10430$
1500	98.28	96.79	96.45	91.53	98.37
2000	101.29	103.64	104.52	100.72	102.4
2500	112.2	112.73	113.78	110.71	112.67
2660	112.57	113.92	114.59	111.72	113.76
2800	113.06	113.35	115.4	112.29	115.41
2940	112.37	112.78	115.86	111.61	113.01
3500	107.03	108.2	110.78	105.21	110.08
4000	106.37	107.06	110.21	106.22	109.51
4500	104.1	105.27	106.75	101.73	106.09
5000	98.01	98.47	99.94	96.59	99.84
5225	94.21	95.67	96.94	93.78	96.46
5500	87.53	89.41	90.24	87.29	90.16
5775	80.08	82.57	82.65	78.97	82.74
6000	73.9	76.31	76.76	72.8	75.82

Bibliography

Amirhossein, A., Jensen, W., and Kazemzadeh, R. (2009). A case study on monitoring polynomial profiles in the automotive industry. *Quality and Reliability Engineering International*, DOI: 10.1002/qre.1071.

Bersimis, S., Psarakis, S., and Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23:517–543.

Boeing Commercial Airplane Group, Materiel Division, P. Q. A. D. (1998). Advanced quality system tools. *AQS D1-9000-1, The Boeing Company: Seattle, WA*.

Bowman, A. W. (2006). Comparing nonparametric surfaces. *Statistical Modelling*, 6:279–299.

Burman, P. and Chaudhuri, P. (1992). A hybrid approach to parametric and nonparametric regression. *Technical Report No. 243, Division of Statistics, University of California-Davis,Davis, CA, USA*.

Chang, T. C. and Gan, F. F. (2006). Monitoring linearity of measurement gauges. *Journal of Statistical Computation and Simulation*, 76:889–911.

Chicken, E., Pignatiello, J. J., and Simpson, J. R. (2009). Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology*, 41:198–215.

Claeskens, G., Kauermann, G., and Opsomer, J. D. (2009). Bootstrapping for penalized spline regression. *Journal of Computational and Graphical Statistics*, 18:126–146.

Colosimo, B. M. and Pacella, M. (2007). On the use of principal component analysis to identify systematic patterns in roundness profiles. *Quality and Reliability Engineering International*, 23:707–725.

Coull, B. A., Ruppert, D., and Wand, M. P. (2001a). Simple incorporation of interactions into additive models. *Biometrics*, 57:539–545.

Coull, B. A., Schwartz, J., and Wand, M. P. (2001b). Respiratory health and air pollution: additive mixed model analyses. *Biostatistics*, 2:337–349.

DeBoor, C. (2001). *A Practical Guide to Splines (Applied Mathematical Sciences)*. Springer.

Demidenko, E. (2004). *Mixed Models : Theory and Applications*. John Wiley and Sons, New York, NY.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, 54:401–415.

Ding, Y., Z. L. and Zhou, S. (2006). Phase *I* analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, 38:199–216.

Durbàn, M., Harezlak, J., Wand, M. P., and Carroll, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, 24:1153–1167.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.

Einsporn, R. L. (1987). *HATLINK: A Link Between Least Squares Regression and Nonparametric Curve Estimation*. PhD thesis, Unpublished doctoral dissertation, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA.

Einsporn, R. L. and Birch, J. B. (1993). Improvement predictions by combining least squares and kernel regression prediction weights. *Technical Report, Department of Statistics, Virginia Polytechnic Institute & State University*, 93-11:21.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:371–394.

Faraway, J. J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC.

Fuchs, C. and Kenett, R. (1998). *Multivariate Quality Control: Theory and Applications*. New York: Marcel Dekker.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: Wiley.

Gupta, S., Montgomery, D. C., and Woodall, W. H. (2006). Performance evaluation of two methods for online monitoring of linear calibration profiles. *International Journal of Production Research*, 44:1927–1942.

Gurrin, L. C., Scurrah, K. J., and Hazelton, M. L. (2005). Tutorial in biostatistics: spline smoothing with linear mixed models. *Statistics in Medicine*, 24:3361–3381.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, New York, NY.

Hardle, W. and Marron, J. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 13:1465–1481.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–338.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing lreland's wind power resource. *Applied Statistics*, 38:1–50.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. New York : Chapman and Hall.

Hawkins, D. M. and Maboudou-Tchao, E. M. (2007). Self-starting multivariate exponentially weighted moving average control charting. *Technometrics*, 49:199–209.

Hawkins, D. M. and Merriam, D. F. (1974). Zonation of multivariate sequences of digitized geologic data. *Mathematical Geology*, 6:263–269.

Henderson, C. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics*, 21:309–310.

Holmes, D. S. and Mergen, A. E. (1993). Improving the performance of the T^2 control chart. *Quality Engineering*, 5:619–625.

Jensen, W. A. and Birch, J. B. (2009). Profile monitoring via nonlinear mixed models. *Journal of Quality Technology*, 41:18–34.

Jensen, W. A., Birch, J. B., and Woodall, W. H. (2007). High breakdown estimation methods for phase *I* multivariate control charts. *Quality and Reliability Engineering International*, 23:615–629.

Jensen, W. A., Birch, J. B., and Woodall, W. H. (2008). Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, 40:167–183.

Jeong, M., Lu, J., and Wang, N. (2006). Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research*, 44:729–744.

Jin, J. and Shi, J. (1999). Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, 41:327–339.

Kang, L. and Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, 32:418–426.

Karin, M. (2005). Random regression analyses using b-splines to model growth of Australian angus cattle. *Genetics selection evolution*, 37:473–500.

Kazemzadeh, R., Noorossana, R., and Amiri, A. (2008). Phase *I* monitoring of polynomial profiles. *Communications in Statistics-Theory and Methods*, 37:1671–1686.

Kazemzadeh, R., Noorossana, R., and Amiri, A. (Scientia Iranica 2009.a). Phase *II* monitoring of autocorrelated polynomial profiles in AR(1) processes. To appear in *the International Journal of Science and Technology*.

Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2009b). Monitoring polynomial profiles in quality control applications. *The International Journal of Advanced Manufacturing Technology*, 42:703–712.

Ke, C. and Wang, Y. (2001). Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association*, 96:1272–1281.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology*, 35:317–328.

Kusiak, A., Zheng, H., and Song, Z. (2009). On-line monitoring of power curves. *Renewable Energy*, 34:1487–1493.

Lada, E. K., Lu, J.-C., and Wilson, J. R. (2002). A wavelet-based procedure for process fault detection. *IEEE Transactions on Semiconductor Manufacturing*, 15:79–90.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.

Loader, C. (1999). *Local Regression and Likelihood*. Springer.

Mahmoud, M. A. (2008). Phase *I* analysis of multiple linear regression. *Communications in Statistics - Simulation and Computation*, 37:2106–2130.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007). A change point method for linear profile data. *Quality and Reliability Engineering International*, 23:247–268.

Mahmoud, M. A. and Woodall, W. H. (2004). Phase *I* analysis of linear profiles with calibration applications. *Technometrics*, 46:380–391.

Mason, R. L. and Young, J. C. (2002). Multivariate statistical process control with industrial applications. *Philadelphia: SIAM*.

Mays, J. E. and Birch, J. B. (2002). Smoothing for small samples with model misspecification: nonparametric and semiparametric concerns. *Journal of Applied Statistics*, 29:1023–1045.

Mays, J. E., Birch, J. B., and Einsporn, R. L. (2000). An overview of model-robust regression. *Journal of Statistical Computation and Simulation*, 66:79–100.

Mays, J. E., Birch, J. B., and Starnes, B. A. (2001). Model robust regression: Combining parametric, nonparametric, and semiparametric methods. *Journal of Nonparametric Statistics*, 13:245 – 277.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control,* 5^{th} *Edition*. John Willy&Sons, Inc.

Noorossana, R., Amiri, A., and Soleimani, P. (2008). On the monitoring of autocorrelated linear profiles. *Communications in Statistics - Theory and Methods*, 37:425–442.

Olkin, I. and Spiegelman, C. H. (1987). A semiparametric approach to density estimation. *Journal of the American Statistical Association*, 82:858–865.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G., and Breidt, F. J. (2008). Nonparametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:265–286.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1:502–518.

Parise, H., Ruppert, D., Ryan, L., and Wand, M. P. (2001). Incorporation of historical controls using semiparametric mixed models. *Applied Statistics*, 50:31–42.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus*. Springer, New York, Inc.

Qiu, P. and Zou, C. (2009). Control chart for monitoring nonparametric profiles with arbitrary design. To appear in *Statistica Sinica*.

Qiu, P., Zou, C., and Wang, Z. (2010). Nonparametric profile monitoring by mixed effects modeling. *Technometrics, (in press)*.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, NY.

Reis, M. and Saraiva, P. (2006). Multiscale statistical process control of paper surface profile. *Quality Technology and Quantitive Management*, 3:263–282.

Rencher, A. C. (2000). *Linear Models in Statistics*. John Wiley and Sons, New York, NY.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational & Graphical Statistics*, 11:735–757.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge, NY.

Ruppert, D., Wand, M. P., Holst, U., and Hossjer, O. (1997). Local polynomial variance-function estimation. *Technometrics*, 39:262–273.

Saghaei, A., Mehrjoo, M., and Amiri, A. (2009). A cusum-based method for monitoring simple linear profiles. *The International Journal of Advanced Manufacturing Technology*, DOI: 10.1007/s00170-009-2063-2.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science.

Schabenberger, O. and Pierce, F. J. (2002). *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton,Florida.

Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression*. Wiley-Interscience, Hoboken,NJ.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35:1773–1801.

Shiau, J.-J. H., Huang, H.-L., Lin, S.-H., and Tsai, M.-Y. (2009). Monitoring nonlinear profiles with random effects by nonparametric regression. *Communications in Statistics - Theory and Methods*, 38:1664–1679.

Soleimani, P., Noorossana, R., and Amiri, A. (2009). Simple linear profiles monitoring in the presence of within profile autocorrelation. *Computers & Industrial Engineering*, 57:1015–1021.

Starnes, B. A. (1999). Asymptotic results for model robust regression. *PhD thesis, Unpublished doctoral dissertation, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA*.

Staudhammer, C., Maness, T. C., and Kozak, R. A. (2007). Profile charts for monitoring lumber manufacturing using laser range sensor data. *Journal of Quality Technology*, 39:224–240.

Stover, F. S. and Brill, R. V. (1998). Statistical quality control applied to ion chromatography calibrations. *Journal of Chromatography A*, 804:37–43.

Sullivan, J. H. (2002). Detection of multiple change points from clustering individual observations. *Journal of Quality Technology*, 34:371–383.

Sullivan, J. H. and Woodall, W. H. (1996). A comparison of multivariate control charts for individual observations. *Journal of Quality Technology*, 28:398–408.

Tracy, N. D., Young, J. C., and Mason, R. L. (1992). Multivariate control chart for individual observations. *Journal of Quality Technology*, 24:88–95.

Vargas, J. A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35:367–376.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91:217–221.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York, NY.

Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, 48:1–17.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, Inc.,New York.

Vonesh, E. F., Chinchilli, V. M., and Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, 52:572–587.

Walker, E. and Wright, S. P. (2002). Comparing curves using additive models. *Journal of Quality Technology*, 34:118–129.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, 18:223–249.

Wang, K. and Tsung, F. (2005). Using profile monitoring techniques for a data-rich environment with huge sample size. *Quality and Reliability Engineering International*, 21:677–688.

Waterman, M. J. (2002). Linear mixed model robust regression. *PhD thesis, Unpublished doctoral dissertation, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA*.

Waterman, M. J., Birch, J. B., and Schabenberger, O. (2007). Linear mixed model robust regression. *Technical Report, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA*, 3:1–39.

Wegener, M. and Kauermann, G. (2008). Examining heterogeneity in implied equity risk premium using penalized splines. *Advances in Statistical Analysis*, 92:35–56.

Wei, Y., Zhao, Z., and Lin, D. K. (2010). A general class of nonparametric L-1 regression with its application to profile control charts. To appear in *Annals of Applied Statistics*.

West, B., Welch, K. B., and Galecki, A. T. (2006). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall/CRC.

Williams, J. D., Birch, J. B., Woodall, W. H., and Ferry, N. M. (2007a). Statistical monitoring of heteroscedastic dose-response profiles from high-throughput screening. *Journal of Agricultural, Biological & Environmental Statistics*, 12:216–235.

Williams, J. D., Woodall, W. H., and Birch, J. B. (2007b). Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International*, 23:925 – 941.

Williams, J. D., Woodall, W. H., Birch, J. B., and Sullivan, J. H. (2006). Distribution of Hotelling's T^2 statistic based on the successive differences covariance matrix estimator. *Journal of Quality Technology*, 38:217–229.

Winistorfer, P. M., Young, T. M., and Walker, E. (1996). Modeling and comparing vertical density profiles. *Wood and Fiber Science*, 28:133–141.

Woodall, W. H. (2007). Current research on profile monitoring. *Producao*, 36:309–320.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, 36:309–320.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability*, 35:2294–2320.

Zhang, J., Li, Z., and Wang, Z. (2009). Control chart based on likelihood ratio for monitoring linear profiles. *Computational Statistics and Data Analysis*, 53:1440–1448.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, 35:1052–1079.

Zhou, S. Y., Sun, B. C., and Shi, J. J. (2006). An SPC monitoring system for cycle-based waveform signals using haar transform. *IEEE Transactions on Automation Science and Engineering*, 3:60–72.

Zou, C., Qiu, P., and Hawkins, D. (2009a). Nonparametric control chart for monitoring profiles using the change point formulation. To appear in *Statistica Sinica*.

Zou, C., Tsung, F., and Wang, Z. (2009b). Monitoring profiles based on nonparametric regression methods. To appear in *Technometrics*.

Zou, C., Wang, Z., and Tsung, F. (2007). A self-starting control chart for linear profiles. *Journal of Quality Technology*, 39:364–375.

Zou, C., Wang, Z., and Tsung, F. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics*, 50:512–526.

Zou, C., Zhang, Y., and Wang, Z. (2006). A control chart based on a change-point model for monitoring linear profiles. *IIE Transactions*, 38:1093–1103.