Nonparametric and Semiparametric Linear Mixed Models

Megan J. Waterman

Department of Defense, U.S. Government, USA.

Jeffrey B. Birch

Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0439, USA.

Abdel-Salam G. Abdel-Salam

Faculty of Economics and Political Science, Cairo University, Egypt.

Abstract

Mixed models are powerful tools for the analysis of clustered data and many extensions of the classical linear mixed model with normally distributed response have been established. As with all parametric models, correctness of the assumed model is critical for the validity of the ensuing inference. An incorrectly specified parametric means model may be improved by using a local, or nonparametric, model. Two local models are proposed by a pointwise weighting of the marginal and conditional variance-covariance matrices. However, nonparametric models tend to fit to irregularities in the data and may provide fits with high variance. Model robust regression techniques estimate mean response as a convex combination of a parametric and a nonparametric model fit to the data. It is a semiparametric method by which incomplete or incorrectly specified parametric models can be improved by adding an appropriate amount of the nonparametric fit. We compare the approximate integrated mean square error of the parametric, nonparametric, and mixed model robust methods via a simulation study and apply these methods to two real data sets: the monthly wind speed data from counties in Ireland and the engine speed data.

KEY WORDS: Semiparametric, Nonparametric, Mixed Effects, Robust.

1 Introduction

Linear mixed (LM) 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 Pierce [2002], West et al. [2006].

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

The general linear mixed regression model (Laird and Ware, 1982) is commonly expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon, \tag{1}$$

where $\mathbf{Y}$ is an (n x 1) vector of responses, $\mathbf{X}$ and $\mathbf{Z}$ are (n x p) and (n x q) fixed model matrices, β is a (p x 1) vector of fixed effects, $\mathbf{b}$ is a (q x 1) vector of random effects, and ϵ is an (n x 1) vector of random disturbances. Additionally, p, the number of parameters estimated, is equal to k+1, where k is the number of regressor variables in the LM model in the presence of the intercept. We assume that the random effects and errors are normal variates with zero expectation and $Var(\mathbf{b}) = \mathbf{B}$, $Var(\epsilon) = \mathbf{R}$, and $Cov(\mathbf{b}, \epsilon) = \mathbf{0}$. Consequently, the conditional variance of $\mathbf{Y}|\mathbf{b}$ is $Var(\mathbf{Y}|\mathbf{b}) = R$ and the marginal variance of $\mathbf{Y}$ is $Var(\mathbf{Y}) = \mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{B}\mathbf{Z}'$.

The estimates of the fixed effects and the prediction of the random effects

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{2}$$

$$\widehat{\mathbf{b}} = \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\beta}) \tag{3}$$

are the solutions to the mixed model estimating equations. The variance-covariance matrix $\mathbf{V}$ is typically unknown and parametrized as $\mathbf{V}(\theta)$. After estimating θ by (restricted) maximum likelihood, ANOVA, or some other method, the estimated variance-covariance matrix $\widehat{\mathbf{V}} = \mathbf{V}(\widehat{\theta})$ is substituted in (2), (3), and in other expressions that depend on θ. We suppress this dependency for brevity.

We are concerned with linear mixed modeling for clustered data. Let i=1,.., s index clusters and write (1) as

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i. \tag{4}$$

Relating equation (4) to (1), $\mathbf{X}$ results from stacking the s matrices $\mathbf{X}_i$. In addition, $\mathbf{Z}_i$ is a submatrix of $\mathbf{X}_i$ of dimension $(n_i \times p_1)$, where $p_1 \leq p$. Then $\mathbf{Z} = diag(\mathbf{Z}_i)$. Each $\mathbf{b}_i$ is $(p_1 \times 1)$ and $\mathbf{b}$ results by stacking the s vectors $\mathbf{b}_i$. Likewise, ϵ_i is $(n_i \times 1)$ and ϵ results by stacking the s vectors ϵ_i. Quantities arising from the marginal distribution of $\mathbf{Y}_i$ will be referred to as population average (PA) quantities; those arising from the distribution of $\mathbf{Y}_i|\mathbf{b}_i$ are referred to as conditional or cluster-specific quantities.

The nonparametric (NP) regression approach to fitting cluster data is more flexible than a purely parametric (P) regression approach. In modeling new data, one often has very little information regarding the appropriate form for the model. While a number of heuristic tools using diagnostic plots are available to help search for this form, it would be useful to let the modeling approach complement this search. One disadvantage of the P approach is that one can easily choose the wrong form for the means 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], 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].

Gy and Zy [2009] developed a robust estimation method for estimation of both the mean and variance components in the generalized partial linear mixed model (GPLMM). The GPLMM can be viewed as a combination of a generalized mixed model and a fully NP model. They introduced a robustified likelihood for the GPLMM and achieve the robust estimation of both the mean and the variance components for the case of outliers existing in the response. Silva and Opsomer [2009] introduced the NP propensity weighting procedure to adjust for unit nonresponse in surveys. Several NP techniques for mixed models have been developed by others including Buskirka and Lohr [2005], Francisco-Fernandez and Opsomer [2005], Lee and Durbán [2009], Morris and Carroll [2006], Opsomer and Francisco-Fernàndez [2010], QingGuo and LongSheng [2009], Wu and Zhang [2002]. Opsomer and Francisco-Fernàndez [2010] utilized the NP approach in finding local departures from a P model.

In this paper the effects of model misspecification on the linear mixed model are analyzed. Two local methods each using a kernel-based weighting scheme, developed to alleviate the bias problem of misspecified P models, will be presented as NP alternatives. Although the local methods result in less bias of fits, there is a tendency for overfitting and, consequently, increased variance of fits. The model robust mixed model, a hybrid combination of the P and local mixed models, is shown to minimize the integrated mean square error of fits when compared to separate P and local methods, while retaining important features of the data.

This paper is organized as follows. Sections 2 and 3 introduce the nonparametric and semiparametric (model robust) mixed models. Section 4 offers a simulation study to compare approximate integrated mean square errors of the P, NP, and model robust methods; these methods are applied in Section 5 to data resulting from monthly wind speed readings taken from twelve locations in Ireland and to the engine speed data. Section 6 contains a discussion of our results.

2 The local mixed models

Let Y_{ij} denote the j^{th} observation from the i^{th} cluster for j=1,..,n_i, i=1,..,s, and $\sum_{i=1}^{s} n_i = n$. The vectors $\mathbf{x}'_{ij}$ and $\mathbf{z}'_{ij}$ are vectors from row j of $\mathbf{X}_i$ and $\mathbf{Z}_i$, the model matrices associated with cluster i for the fixed and random effects, respectively. In this paper, we consider polynomials locally in a single regressor Fan and Gijbels [1996]. Extensions of our method to multiple regressors are straightforward. Our local models are found by pointwise fitting a weighted version of the Laird-Ware model. The weights depend on the point of estimation, a bandwidth, and a kernel function and are constructed with respect to the conditional or marginal distribution of $\mathbf{Y}_i$.

In some applications of the LM model emphasis is placed on estimating and comparing the cluster-specific curves. For example, in the area of quality control, specifically in profile monitoring, the estimated cluster-specific curves (referred to as 'profiles') are compared to determine if any are the result of a process that is 'out-of-control'. On the other hand, in some applications, emphasis is placed on the estimated PA curve and not the estimated cluster-specific curves. For example, two treatments can be compared by comparing their estimated PA curves. As will be demonstrated in the sections to follow, one of our local mixed models favors estimation of cluster-specific curves while the other method favors estimation of the PA curve.

It should be pointed out that in the context considered here other NP regression methods may also be used such as one of the various versions of spline regression Ruppert et al. [2003]. For example, Abdel-Salam et al. [2012] use p-splines to model the cluster-specific curves in a profile monitoring application. We chose to use kernel-based methods in this paper due to the flexibility they provide in emphasizing either the PA curve or the cluster-specific curves.

2.1 The coditional local mixed model

Consider the localized d^{th} order polynomial mixed model for estimation at an arbitrary $\tilde{x}_0$ for the i^{th} cluster

$$Y_{i0} = \tilde{\mathbf{x}}'_{i0}\beta_0 + \tilde{\mathbf{z}}'_{i0}\mathbf{b}_0 + \epsilon_{i0} \tag{5}$$

where Y_{i0} is the response at $\tilde{x}_0$ for the i^{th} cluster, $\tilde{\mathbf{x}}'_{i0} = 1\tilde{x}_{i0}...\ \tilde{x}^d_{i0}$, $\tilde{\mathbf{z}}'_{i0}$ contains the regressors in $\tilde{x}_0$ corresponding to the random effects and ϵ_{i0} represents the "localized" error at $\tilde{x}^{'}_{0i}$. The vectors β_0 and $\mathbf{b}_0$ are the fixed parameter and random effects vectors at $\tilde{x}_0$, respectively. Typically, the order of the polynomial, d, is either 1, 2 or 3. In many applications, including the examples presented in Section 4, d=1 provides a very satisfactory fit. The $(n \times 1)$ vector of localized random errors, ϵ_0, is assumed to be from a multivariate Gaussian distribution with zero mean and variance-covariance matrix $\mathbf{K}_0^{-\frac{1}{2}}\mathbf{R}\mathbf{K}_0^{-\frac{1}{2}}$, where $\mathbf{K}_0^{-\frac{1}{2}}$, the localizing matrix, is an $(n \times n)$ diagonal weight matrix containing the inverse square root of the Nadaraya [1964], Watson [1964] weights at $\tilde{x}_0$. This weighting scheme is motivated by local polynomial regression, a popular NP technique, used for the fixed effect case. We label this approach the conditional local mixed model (CLMM) because the weighting is applied to the variance of $\mathbf{Y}|\mathbf{b}$. We will assume that the distances used in the weights have been standardized appropriately so that the weights will sum to one across a data set, and that the weights have been assigned without regard to cluster.

The estimator $\widehat{\beta}_0^C$ and the predictor $\widehat{\mathbf{b}}_0$ at the point $\tilde{x}_0$ can be found by incorporating the weight matrix in Henderson's joint likelihood expression Henderson [1950] and solving the mixed model equations for estimation at $\tilde{x}_0$. We let $\widetilde{\mathbf{X}}_i$ denote the model matrix appropriate for an $(n_i \times d+1)$ d^{th} order polynomial in one regressor for the i^{th} cluster. Then $\widetilde{\mathbf{X}}$ results from stacking the s matrices $\widetilde{\mathbf{X}}_i$. With $\widetilde{\mathbf{X}}$ of full rank, solutions to the equations yield the estimator

$$\widehat{\beta}_0^C = (\widetilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\mathbf{Y} \tag{6}$$

and the predictor

$$\widehat{\mathbf{b}}_0 = \mathbf{B}\widetilde{\mathbf{Z}}'\mathbf{V}_\mathbf{0}^{*-\mathbf{1}}(\mathbf{Y} - \widetilde{\mathbf{X}}\widehat{\beta}_\mathbf{0}^\mathbf{C}), \tag{7}$$

where $\mathbf{V}_0^* = \mathbf{K}_0^{-\frac{1}{2}}\mathbf{R}\mathbf{K}_0^{-\frac{1}{2}} + \widetilde{\mathbf{Z}}\mathbf{B}\widetilde{\mathbf{Z}}'$. The expressions given above are of similar form to those given by the parametric mixed model except that different estimates and predictions of the parameter vectors are realized at each arbitrary $\widetilde{x}_0$.

The population average (PA) fit at $\widetilde{x}_0$ is simply

$$\widehat{Y}_{PA,0} = \widetilde{\mathbf{x}}_0' \widehat{\beta}_0^C = \sum_k h_{PA,0,k}^C y_k, \tag{8}$$

where y_k is the k^{th} element of $\mathbf{Y}$ and where $h_{PA,0,k}^C$ is the kth element of $\widetilde{\mathbf{x}}_0'(\widetilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\mathbf{V}_0^{*-1}$. The cluster specific (CS) fits at $\widetilde{x}_0$ for the i^{th} cluster (i=1,...,s) are

$$\widehat{Y}_{CS,i,0} = \widetilde{\mathbf{x}}_{i0}' \widehat{\beta}_0^C + \widetilde{\mathbf{x}}_{i0}' \widehat{\mathbf{b}}_{i0}. \tag{9}$$

where $\widehat{\mathbf{b}}_{i0}$ is the subvector of $\widehat{\mathbf{b}}_0$.

2.2 The marginal local mixed model

Localization through weighting can also be accomplished by targeting the marginal variance-covariance matrix. Consider the following model for estimation at $\widetilde{x}_0$ for the i^{th} cluster

$$Y_{i0} = \widetilde{\mathbf{x}}_{i0}' \beta_0 + k_{i0}^{-\frac{1}{2}} \widetilde{\mathbf{x}}_{i0}' \mathbf{b}_0 + \epsilon_{i0} \tag{10}$$

where Y_{i0}, $\widetilde{\mathbf{x}}_{i0}'$, $\widetilde{\mathbf{z}}_{i0}'$, β_0, $\mathbf{b}_0$, and ϵ_{i0} are as defined for CLMM and $k_{i0}^{-\frac{1}{2}}$ is the i^{th} element of $\mathbf{K}_0^{-\frac{1}{2}}$. The vector of localized random errors, ϵ_0, is again assumed to follow a multivariate Gaussian distribution with zero mean and variance-covariance matrix $\mathbf{K}_0^{-\frac{1}{2}}\mathbf{R}\mathbf{K}_0^{-\frac{1}{2}}$. We label this model the marginal local mixed model (MLMM). As in the conditional model, the local influence of an observation is directed by its variance. Observations that contribute more (that is, to have larger weight) to the prediction at $\widetilde{x}_0$ are considered to have smaller variance. The variance, which is transformed to represent the relative weights of observations, is Var[$\mathbf{Y}_i|\mathbf{b}_i$] in the model of Section 2.1 and Var[$\mathbf{Y}_i$] here. The multiplicative involvement of the Nadaraya-Watson weights in (11) accomplish that as

$$\begin{aligned} Var(\mathbf{Y}) &= \mathbf{K}_0^{-\frac{1}{2}}(\mathbf{R} + \widetilde{\mathbf{Z}}\mathbf{B}\widetilde{\mathbf{Z}}')\mathbf{K}_0^{-\frac{1}{2}} \\ &= \mathbf{K}_0^{-\frac{1}{2}}\mathbf{V}\mathbf{K}_0^{-\frac{1}{2}} = \mathbf{V}_0^{**}. \end{aligned}$$

The estimator $\widehat{\beta}_0^M$ at the point $\widetilde{x}_0$ for the marginal local mixed model is

$$\widehat{\beta}_0^M = (\widetilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\mathbf{Y}, \tag{11}$$

and the population average fit at $\widetilde{x}_0$ is

$$\widehat{Y}_{PA,0} = \widetilde{\mathbf{x}}_0' \widehat{\beta}_0^M = \sum_k h_{PA,0,k}^M y_k, \tag{12}$$

with $h_{PA,0,k}^M$ the kth element of $\widetilde{\mathbf{x}}_0'(\widetilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\mathbf{V}_0^{**-1}$. Cluster specific prediction in the marginal model is not appropriate as model (11) is the result of localizing the Var[$\mathbf{Y}_i$] at the cost of incorrectly representing E[$\mathbf{Y}_i|\mathbf{b}_i$]. The random effects regressors in the marginal local mixed model are transformed to correctly weight the marginal variance of the response.

2.3 Bandwidth selection

The kernel weights used in the conditional and marginal local mixed models depend upon a bandwidth parameter (h). A natural criterion for selection is to choose h in order to minimize a function of the squared error of estimation of mean response to account for bias and variance.

Härdle and Marron [1985] and Härdle [1990] provide a rule for bandwidth selection that chooses the asymptotically optimal bandwidth with respect to a number of criteria, including the average squared error, integrated squared error, and the conditional mean integrated squared error. A bandwidth is chosen by minimizing an estimate of some appropriate criterion as the true mean function used in the criteria is unknown.

Plug-in methods, where unknown quantities in the squared error function are replaced with estimates, are very popular. Rule of thumb selectors offer a simple estimate of the bandwidth that is easy to calculate Fan and Gijbels [1995], Härdle and Marron [1995].

We prefer bandwidth estimators based on cross-validation Craven and Wahba [1979], in particular penalized, 'leave-one-out' statistics.

2.3.1 PRESS

The prediction error sum of squares, or PRESS statistic Allen [1974], is defined in the usual regression setting as

$$PRESS = \sum_{i=1}^{n} (Y_i - \widehat{Y}_{i,-i})^2, \tag{13}$$

where Y_i is the i^{th} observation and $\widehat{Y}_{i,-i}$ is the estimate of the regression function at $\tilde{x}_i$ with the i^{th} data point removed. The bandwidth selected is the value h that minimizes the PRESS statistic.

In the mixed effects model with clustered data, the notion of 'leave-one-out' extends to removal of entire clusters, as clusters represent uncorrelated units. Cluster deletion formulas for the parametric mixed model are given by Henderson [1950], Hurtado-Rodriguez [1993]. The estimators for CLMM and MLMM at $\tilde{x}_0$ with the i^{th} cluster deleted, denoted by $\widehat{\beta}^C_{0,-i}$ and $\widehat{\beta}^M_{0,-i}$, respectively, are

$$\widehat{\beta}^C_{0,-i} = \widehat{\beta}^C_0 - (\tilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\mathbf{U}\widehat{\phi}^C_{0,-i} \tag{14}$$

$$\widehat{\beta}^M_{0,-i} = \widehat{\beta}^M_0 - (\tilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\mathbf{U}\widehat{\phi}^M_{0,-i}, \tag{15}$$

where U is the (n x n_i) matrix that contains the identity matrix for the i^{th} cluster and zeros elsewhere, and

$$\widehat{\phi}^C_{0,-i} = (\mathbf{U}'\mathbf{P}^*\mathbf{U})^{-1}\mathbf{U}'\mathbf{P}^*\mathbf{Y},$$

$$\widehat{\phi}^M_{0,-i} = (\mathbf{U}'\mathbf{P}^{**}\mathbf{U})^{-1}\mathbf{U}'\mathbf{P}^{**}\mathbf{Y},$$

$$\mathbf{P}^* = \mathbf{V}_0^{*-1} - \mathbf{V}_0^{*-1}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\mathbf{V}_0^{*-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_0^{*-1},$$

$$\mathbf{P}^{**} = \mathbf{V}_0^{**-1} - \mathbf{V}_0^{**-1}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\mathbf{V}_0^{**-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_0^{**-1}.$$

Then the population average fit at $\tilde{x}_0$ with the i^{th} cluster deleted is

$$\widehat{Y}_{PA,0,-i} = \tilde{\mathbf{x}}_0' \widehat{\beta}_{0,-i}, \tag{16}$$

where $\widehat{\beta}_{0,-i}$ equals $\widehat{\beta}^C_{0,-i}$ and $\widehat{\beta}^M_{0,-i}$ for CLMM and MLMM, respectively. The BLUP $\widehat{\mathbf{b}}_{0,-i}$ of the i^{th} cluster at $\tilde{\mathbf{x}}_0$ in the CLMM is equal to zero[1], so that cluster specific predictions reduce to population average estimation for cluster deletion.

2.3.2 PRESS**

We adopt here to the mixed model scenario a penalized version of PRESS, proposed in the context of fixed effects models for uncorrelated data by Mays et al. [2001]. For the linear mixed model, PRESS** is defined as

$$PRESS^{**} = \frac{PRESS}{n - trace(\mathbf{H}) + (n-d')(\frac{SSE_{max}-SSE_h}{SSE_{max}-SSE_{\bar{y}}})}, \tag{17}$$

where d' is the number of fixed effects parameters in the local mixed model and

$$\mathbf{H} = \mathbf{H}^{CLMM} = \begin{matrix} \tilde{\mathbf{x}}_1'(\tilde{\mathbf{X}}'\mathbf{V}_1^{*-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_1^{*-1} \\ \tilde{\mathbf{x}}_2'(\tilde{\mathbf{X}}'\mathbf{V}_2^{*-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_2^{*-1} \\ \vdots \\ \tilde{\mathbf{x}}_n'(\tilde{\mathbf{X}}'\mathbf{V}_n^{*-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_n^{*-1} \end{matrix}$$

for the conditional local mixed model and

$$\mathbf{H} = \mathbf{H}^{MLMM} = \begin{matrix} \tilde{\mathbf{x}}_1'(\tilde{\mathbf{X}}'\mathbf{V}_1^{**-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_1^{**-1} \\ \tilde{\mathbf{x}}_2'(\tilde{\mathbf{X}}'\mathbf{V}_2^{**-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_2^{**-1} \\ \vdots \\ \tilde{\mathbf{x}}_n'(\tilde{\mathbf{X}}'\mathbf{V}_n^{**-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{V}_n^{**-1} \end{matrix}$$

for the marginal local mixed model. These matrices are the local smoother matrices for the population average. The term SSE_{max} is the sum of squared deviations of the response and the local fit that assigns a constant weight to each response. This would result as $h \to \infty$ and represents the worst possible fit using the nonparametric method. The sum of squares $SSE_{\bar{y}}$ is the accumulation across the regressor locations of the sum of the squared deviations

[1] Removal of some, but not all, of the observations in a cluster results in a BLUP not equal to zero for that cluster. See Hurtado-Rodriguez [1993] for formulas.

of the responses around the mean response and represents the sum of squares error as $h \rightarrow 0$ for the population average model. The expression SSE_h is the sum of squared deviations of the response and the local fit using a specific value of the bandwidth h. As $h \rightarrow 0$, $SSE_h \rightarrow 0$ for CLMM and $SSE_h \rightarrow SSE_{\bar{y}}$ for MLMM. Cluster specific and population average fits are used in the sum of squared calculations for PRESS** for CLMM and MLMM, respectively.

The denominator in the PRESS** statistic functions as the penalty term, simultaneously providing protection against choosing h too large or too small. The term $n - trace(\mathbf{H})$, where $\mathbf{H}$ equals either $\mathbf{H}^{CLMM}$ or $\mathbf{H}^{MLMM}$, penalizes small bandwidths as this term approaches 0 as $h \rightarrow 0$ and $n - d'$ as $h \rightarrow \infty$. The second term in the denominator ($(n - d')$ multiplied by the sum of squares ratio) protects against large bandwidths, as the second term approaches $n - d'$ as $h \rightarrow 0$ and 0 as $h \rightarrow \infty$.

3 The semiparametric mixed model

We assume that the user has some knowledge regarding the underlying model from which the data have been generated, but the model fails over a portion of the data; it has been misspecified in functional form. Relying on a nonparametric model entirely results in a loss of information about the model. The nonparametric model also has a tendency to produce highly variable fits. A combination of fits may be advantageous; a nonparametric portion 'corrects' areas of poor parametric estimation while retaining the information about the true model contained in the parametric model.

Burman and Chaudhuri [1992], and Mays et al. [2000] developed a semiparametric method termed Model Robust Regression 1 (MRR1) by Mays et al. [2000]. $MRR1$ has been successfully applied to a broad variety of models and situations including linear normal-theory based regression Mays et al. [2001], logistic regression Nottingham and Birch [2000], simultaneous modeling of the mean and variance functions Pickle et al. [2008], Robinson et al. [2010], generalized estimating equations Clark [2002] and the multi-response optimization problem Wan and Birch [2011].

The MRR1 fit is a convex combination

$$\widehat{\mathbf{Y}}^{MRR1} = (1 - \lambda)\widehat{\mathbf{Y}}^{P} + \lambda\widehat{\mathbf{Y}}^{NP}, \tag{18}$$

where $\widehat{\mathbf{Y}}^{P}$ and $\widehat{\mathbf{Y}}^{NP}$ are the parametric (P) and nonparametric (NP) fits, respectively. The mixing parameter $\lambda \in [0, 1]$ determines the proportion of the NP fit that contributes to the model robust fit. For a correctly specified P model, λ will be zero and the MRR1 model reduces to the P model; for P models that are grossly misspecified, λ will be one or close to one and the MRR1 model is the same or nearly equal to the NP model. Simulation results in Mays et al. [2001] show that MRR1 has smaller average mean square error of fits than separate P and NP fits under low to moderate model misspecification. Under no model misspecification, MRR1 is equivalent, or nearly so, to the P fit. Under a high degree of misspecification MRR1 is equivalent, or nearly so, to the NP fit.

The rationale is that this convex combination should take advantage of the low variance and low bias of the P and local fits, respectively, to decrease mean square error. A simulation study in Section 4 investigates this claim for the mixed model.

Semiparametric modeling can be extended to the mixed model setting. The proposed Mixed Model Robust Regression (MMRR) fit is an adaptation of the MRR1 fit for use with the mixed model. In MMRR, two separate fits are combined to get the final fits for the PA and the s clusters. Specifically, the MMRR fit is of the same form as in (18) where $\hat{\mathbf{Y}}^P$ is the fit from the parametric linear mixed model and $\hat{\mathbf{Y}}^{NP}$ is a local mixed model fit.

As in the P and local models, the MMRR fit can be a population average fit or the cluster specific fits. There are two population average fits for MMRR. One combines the population average fit for the P and the conditional local mixed model; the second combines the fit for the P and the marginal local mixed model. The cluster specific fits for MMRR utilizes the cluster specific P fits in combination with the cluster specific conditional local mixed model fits.

The mixing parameter λ measures the degree of parametric model misspecification and must be estimated from the data. Notice that (18) can be written as

$$(\widehat{\mathbf{Y}}^{MMRR} - \widehat{\mathbf{Y}}^P) = \lambda(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P) \tag{19}$$

yielding the least square estimate

$$\widehat{\lambda} = \frac{(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)'(\widehat{\mathbf{Y}}^{MMRR} - \widehat{\mathbf{Y}}^P)}{(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)'(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)}. \tag{20}$$

Thus, $\widehat{\lambda}$ is an estimate of the slope parameter in a no-intercept model with response $(\widehat{\mathbf{Y}}^{MMRR} - \widehat{\mathbf{Y}}^P)$ and explanatory variable $(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)$. For the uncorrelated, fixed effects model, Burman and Chaudhuri [1992] and Mays et al. [2001] found the optimal data driven estimate of the mixing parameter, an estimate of the value which minimizes the distance between the model robust estimate and the true regression function, as

$$\widehat{\lambda} = \frac{(\widehat{\mathbf{Y}}^{NP}_{i,-i} - \widehat{\mathbf{Y}}^P_{i,-i})'(\mathbf{Y} - \widehat{\mathbf{Y}}^P)}{(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)'(\widehat{\mathbf{Y}}^{NP} - \widehat{\mathbf{Y}}^P)}. \tag{21}$$

$\widehat{\mathbf{Y}}^P_{i,-i}$ and $\widehat{\mathbf{Y}}^{NP}_{i,-i}$ are the P and NP estimates of the mean response at x_i computed without the point (x_i, y_i). In the cluster correlated mixed model, $\widehat{\mathbf{Y}}^P_{i,-i}$ and $\widehat{\mathbf{Y}}^{NP}_{i,-i}$ are replaced with the P and NP fits for the i^{th} cluster with the i^{th} cluster removed. Burman and Chaudhuri [1992] had suggested the substitution of $\widehat{\mathbf{Y}}^P_{i,-i}$ and $\widehat{\mathbf{Y}}^{NP}_{i,-i}$ as a precaution against favoring the NP fit. Notice that $\hat{\mathbf{Y}}^{MMRR}$ is unknown and depends on λ. But $\hat{\mathbf{Y}}^{MMRR}$ approaches $E(\mathbf{Y})$ as the sample size increases, so $\mathbf{Y}$ is used in place of $\hat{\mathbf{Y}}^{MMRR}$ in the estimate.

For the conditional local mixed model, emphasis is placed upon CS estimation. Thus, the fits used in choosing $\hat{\lambda}$ and h for MMRR estimation using the conditional local mixed model will be cluster specific. The conditional local mixed model also yields a PA fit. The $\hat{\lambda}$ used in computing the population average MMRR fit using the CLMM will be the same $\hat{\lambda}$ used in computing the cluster specific MMRR fits using the CLMM. Thus, the mean square error for the population average MMRR estimate will not be optimal when using the CLMM population average fit. However, because we are primarily interested in CS fits when using the CLMM, $\hat{\lambda}$ based on CS fits will be used for all mixed model robust regression estimates that use the conditional local mixed model. PA fits are used in the estimate of λ and h for MMRR using the marginal local mixed model, as the marginal local mixed model is only appropriate for the PA.

4 Simulation study

A Monte Carlo simulation study was conducted with data generated from the CS model

$$Y_{ij} = (2+b_{i1})(X_j-5.5)^2 + (5+b_{i2})X_j + 10\gamma\left[sin\left(\frac{\pi(X_j-1)}{2.25}\right)\right] + \epsilon_{ij}, \quad (22)$$

where Y_{ij} is the simulated response for the i^{th} cluster at X_j. The regressor takes on integer values from one to ten and standardized to be between zero and one, inclusively.[2] The random effects, b_{i1} and b_{i2}, are generated independently from normal distributions with mean zero and variance 0.50.

Three variance-covariance structures for the random errors ϵ_{ij} were considered. The first variance-covariance structure was independence with the variance of the errors equal to 16. First-order autoregressive models (AR(1)) with ρ=0.20 and ρ=0.80 were also used with σ^2=16.

The user's model is

$$Y_{ij} = (2+b_{i1})(X_j-5.5)^2 + (5+b_{i2})X_j + \epsilon_{ij}. \quad (23)$$

The true model is given in (22). The trigonometric component then serves as the misspecification with γ as the misspecification parameter. Values of γ equal to 0 (no misspecification), 0.25, 0.50, 0.75, and 1.0 will be used in the study. A plot of the population average models versus γ is given in Figure 1. The smooth parabola, indicated by the dashed line, occurs with no model misspecification and the solid curve represents the most misspecification at γ=1. The large disparity between the γ=0 and γ=1 models should be reflected in the MSE results from the simulation study. It is assumed that there is no parametric misspecification in the variance-covariance structure. That is, if the random errors are generated from an AR(1) variance-covariance structure, the parametric model is the quadratic model in (23) with an AR(1) structure for **R**.

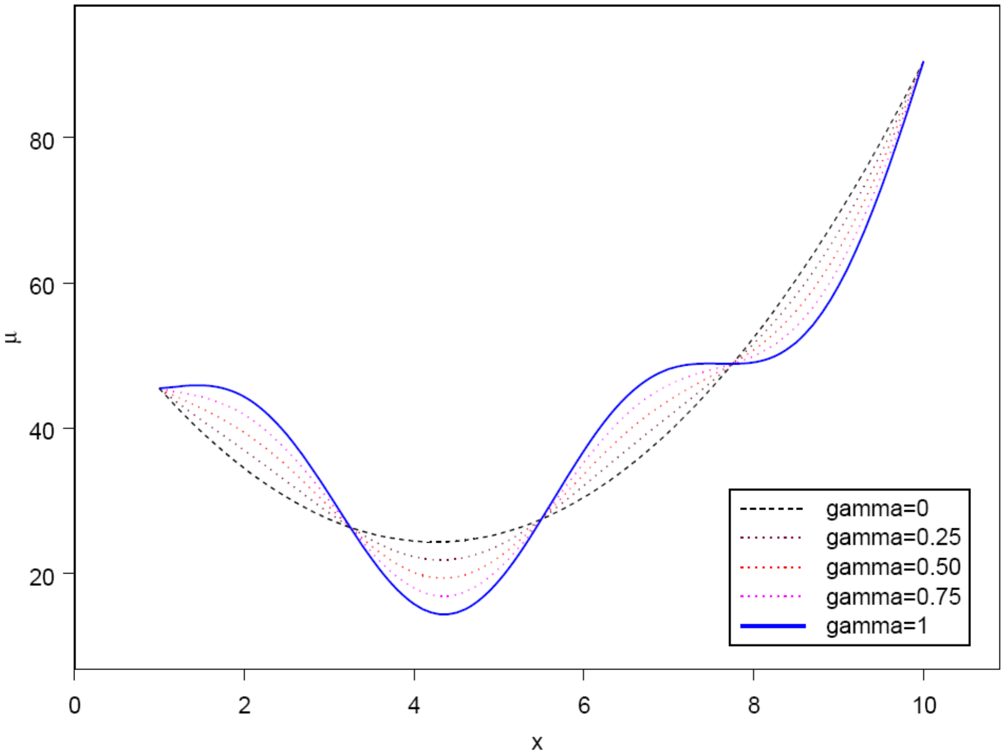
The local model used in the analysis is the local linear mixed model (CLMM or MLMM) with a random intercept:

$$Y_{i0} = \beta_{00} + \tilde{x}_0\beta_{10} + b_{i0} + \epsilon_{i0}. \quad (24)$$

The simplified Gaussian kernel function, $K(u)$, is used throughout our simulations, where, $K(u) = e^{-u^2}$. Our extensive Monte Carlo studies indicate that PRESS provides the best mean square error of fit results when using MLMM, while PRESS** provides best results when using CLMM. Consequently, the bandwidth selectors used in the study were PRESS for MLMM population average estimation (PA MLMM and PA MMRR MLMM) and PRESS** for CLMM population average (PA) and cluster specific (CS) estimation (PA CLMM, PA MMRR CLMM, CS CLMM, and CS MMRR CLMM). A golden section search was performed over the bandwidth range [0.05, 0.30]. These values were selected to minimize the distance covered by the search method as the bandwidths chosen in every scenario fell within these bounds. The estimate of the mixing parameter was found using formula (21). Because no bounds are imposed by this formula, the estimate of the mixing parameter was set to zero or one if the solution to (21) was negative or greater than one, respectively.

[2] The model studied here is similar to the model of Mays et al. [2001], except that in this work the cluster correlated, random coefficient case is considered.

Figure 1: Plot of Population Average Underlying Models (where γ is the misspecification parameter)



Both the bandwidth and the mixing parameter were found by summing over the design points. Using the estimated bandwidth and mixing parameter for a given data set, the integrated mean square error of fit was approximated by calculating the mean square error at 46 points (1 to 10 by 0.20), an arbitrary value selected to adequtely charactrize each CS curve. The mean square errors were calculated for the parametric (population average and cluster specific), CLMM (population average and cluster specific), the MLMM (population average), and the mixed model robust regression models (population average using CLMM, population average using MLMM, and cluster specific using CLMM).

We are interested in the approximate integrated MSE (INTMSE) as a function of cluster size, correlation, and γ. To keep the number of scenarios manageable, s=5 and s=20 clusters per data set are examined for different variance-covariance structures and degrees of misspecification.

Since the data are correlated and parameter estimation is an iterative process, fitting a large number of models requires substantial computing resources. To examine the practically feasible number of needed simulation runs that also provided sufficient precision of Monte-Carlo averages, we examined the standard errors of Monte-Carlo mean square errors. As the number of runs increased, the standard error decreased and leveled off around 250 runs. We decided on 250 simulation runs, attempting to balance between computing time and Monte-Carlo variability. All calculations were perform using *SAS*, with a strong reliance on proc mixed. All programs are available from the authors upon request.

4.1 Varying cluster size

This section presents results for the model in (22) under the assumption of within-cluster independence. It is expected that as the number of clusters increases, the INTMSE for the local population average should decrease since more observations at a given $\tilde{x}_0$ will result in estimates that will be more precise. For small bandwidths, this would mean that non-negligible weight would be given to those observations at $\tilde{x}_0$ only, resulting in a local population average fit that connects the mean response at each value of the regressor. The INTMSE values for the local cluster specific fits should be unaffected by the addition of clusters.

In this work, five clusters will be considered our "small" number of clusters, and twenty clusters our "large" number of clusters. Tables 1 and 2 contain the simulated INTMSE values summed over clusters using an independence within-cluster variance structure over regressor location/cluster size combination. Columns 2 through 6 contain population average (PA) results, while columns 7 through 9 contain cluster specific (CS) results. The bolded value is the minimum population average and cluster specific INTMSE for the given γ value.

Tables 1 and 2 suggest that MMRR-MLMM should be used for population average estimation, while MMRR-CLMM be utilized for cluster specific prediction. These model robust procedures clearly minimize the INTMSE over the entire range of γ. For example, for the PA fit, we note that MMRR-MLMM is very close to the optimal parametric method (when γ=0) in INTMSE for very small values of γ, is very close to MLMM for very large values of γ, and is superior to both for intermediate vales of γ. We note similar results for the CS fit and the MMRR-CLMM technique. MMRR-MLMM and MMRR-CLMM are robust to model misspecification for fitting the PA and CS curves, respectively.

A finer grid between γ=0 and γ=0.30 provides a range of values where parametric, local, and model robust procedures are optimal, information useful to determine the degree of misspecification where model robust procedures would be the most beneficial.

Table 1. Mixed Model Robust Regression using Independence

(10 design points and 5 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**13.83**	16.09	15.76	13.87	14.00	**2.65**	7.54	2.73
0.05	**13.95**	16.10	15.81	13.96	14.08	**2.77**	7.55	2.84
0.10	14.30	16.15	15.82	**14.27**	14.36	**3.12**	7.62	3.16
0.15	14.88	16.19	15.87	**14.74**	14.78	3.71	7.73	**3.66**
0.20	15.70	16.25	15.96	15.31	**15.23**	4.55	7.88	**4.29**
0.25	16.75	16.36	16.03	15.94	**15.60**	5.62	8.06	**4.99**
0.30	18.03	16.45	16.03	16.54	**15.87**	6.93	8.28	**5.73**
0.50	25.51	16.77	**16.19**	18.28	16.33	14.63	9.20	**8.35**
0.75	40.11	17.26	**16.57**	19.35	16.69	29.87	**10.48**	10.67
1.00	60.56	17.69	**17.14**	20.33	17.23	51.35	**11.86**	12.68

Table 2. Mixed Model Robust Regression using Independence

(10 design points and 20 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**3.37**	4.16	4.08	3.38	3.43	**2.49**	9.56	**2.49**
0.05	**3.49**	4.16	4.08	**3.49**	3.56	**2.60**	9.62	2.61
0.10	3.84	4.17	4.09	3.81	**3.79**	2.95	9.79	**2.94**
0.15	4.43	4.17	4.11	4.28	**3.97**	3.54	10.17	**3.44**
0.20	5.24	4.16	4.12	4.84	**4.08**	4.36	10.51	**4.08**
0.25	6.30	4.19	**4.16**	5.42	**4.16**	5.42	10.73	**4.81**
0.30	7.58	4.22	**4.19**	5.96	4.22	6.72	10.80	**5.56**
0.50	15.06	4.48	**4.42**	7.50	4.45	14.35	11.04	**8.40**
0.75	29.67	4.92	**4.84**	8.96	4.93	29.57	12.84	**11.69**
1.00	50.12	5.47	**5.45**	10.52	5.51	51.42	**14.24**	14.64

A cross-over point is defined as the value of the misspecification parameter (γ) at which the minimum INTMSE value switches from parametric to model robust estimation or from model robust to local estimation. For both PA and CS prediction, the minimum INTMSE value changes from the parametric to the model robust method at a γ value between 0.05 and 0.10. The second cross-over point (from model robust to local estimation) occurs much earlier for the population average for both cluster sizes. The second population average cross-over point occurs between γ=0.20 and γ=0.30, whereas the second cluster specific cross-over point occurs for large misspecification–

a γ value between 0.75 and 1.0. In addition, the cross-over occurs earlier for larger cluster sizes, a result consistent with Clark [2002].

4.2 AR(1) Correlation structure

The INTMSE values for the correlated data cases appear in Tables 3-6 for $s = 5$ and $s = 20$. The pattern of the cross-over points in γ across the covariance structures appears to be similar. This pattern suggests that PA mixed model robust regression outperforms MLMM for small amounts of model misspecification, whereas CS mixed model robust regression generally works well for all levels of misspecification, and is outperformed by CLMM only at the extreme cases, as expected.

Table 3. Simulated INTMSE values using AR(1) with ρ=0.20

(10 design points and 5 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**13.77**	16.22	15.71	13.79	13.91	**3.74**	9.04	3.77
0.25	16.73	16.38	15.94	15.99	**15.57**	6.74	9.55	**6.05**
0.50	25.53	16.77	**16.09**	18.69	16.28	16.00	10.65	**9.56**
0.75	40.18	17.18	**16.43**	20.15	16.59	31.81	**11.82**	12.12
1.00	60.99	17.57	**16.99**	21.30	17.14	54.01	**13.05**	14.34

Table 4. Simulated INTMSE values using AR(1) with ρ=0.20

(10 design points and 20 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**3.42**	4.06	3.97	**3.42**	3.46	**3.34**	10.72	**3.34**
0.25	6.34	4.09	**4.06**	5.42	4.08	6.34	11.80	**5.66**
0.50	15.12	4.36	**4.31**	7.66	4.36	15.74	12.20	**9.36**
0.75	29.77	4.77	**4.72**	9.72	4.81	32.31	13.98	**13.04**
1.00	50.28	5.33	**5.31**	11.88	5.37	55.81	**15.03**	16.30

Table 5. Simulated INTMSE values using AR(1) with ρ=0.80

(10 design points and 5 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**14.26**	17.07	16.31	14.44	14.80	**6.48**	14.54	6.60
0.25	17.30	17.21	16.33	16.97	**16.29**	9.41	14.80	**9.14**
0.50	26.24	17.51	**16.41**	19.62	16.83	18.60	15.35	**13.11**
0.75	40.96	17.71	**17.01**	20.44	17.26	34.04	15.90	**15.38**
1.00	61.50	18.06	**17.61**	21.45	17.87	55.93	**16.52**	17.13

Table 6. Simulated INTMSE values using AR(1) with ρ=0.80

(10 design points and 20 clusters. Best values in bold)

γ	PA Parm.	PA CLMM	PA MLMM	PA MMRR CLMM	PA MMRR MLMM	CS Parm.	CS CLMM	CS MMRR CLMM
0.00	**3.68**	4.36	4.25	**3.68**	3.75	**5.78**	14.67	**5.78**
0.25	6.63	4.38	**4.32**	5.62	4.33	8.85	15.04	**8.40**
0.50	15.46	4.64	**4.57**	7.29	4.62	18.13	15.39	**12.28**
0.75	30.13	5.01	**4.99**	8.78	5.03	33.89	16.37	**15.09**
1.00	50.66	5.62	**5.58**	10.72	5.60	56.55	**17.00**	17.56

Notice that the population average MLMM and cluster specific CLMM methods robust models are extremely competitive. For γ=0, the parametric method should have the smallest INTMSE. The model robust procedures obtain INTMSE values very close to the parametric INTMSE values. For γ=1, the local methods should have the smallest INTMSE, and the model robust procedures obtain INTMSE values very close to the local values. As γ increases from zero to one, the INTMSE values for the mixed model robust procedures are either the minimum value or are close in value to the 'winning' INTMSE values.

There are some key differences between the independence and correlated cases. On average, the INTMSE values increase as the correlation increases. Consider, for example, the case γ=0, where INTMSE reduces to the approximate integrated variance of fit when conditioned on the values of the random effects. If the same size n remains fixed while ρ increases, the effective sample size decreases. Thus, as the correlation increases, the variance of the fits, and hence the mean square error, must increase.

It is clear from the above tables that MMRR-CLMM is superior in CS prediction when measured by INTMSE. For population average estimation, MMRR-MLMM does not achieve the minimum INTMSE value, but its INTMSE value is very close to the minimum value across all values of γ, s, and ρ. Moreover, when the model is correctly specified, MMRR-MLMM always beats MLMM, the preferred PA local method.

4.2.1 Estimation of ρ

One concern in the correlated data case was whether the misspecification term influenced the estimate of ρ. As the estimate of ρ is determined by REML or maximum likelihood (ML), it is very difficult to determine the expected value of the correlation estimate under model misspecification; such estimates, however, can be examined by varying model misspecification in the previous Monte-Carlo study. Five hundred data sets were generated for different values of γ, and the average estimates of ρ from the parametric analysis over the five hundred data sets were calculated. The data were generated from an AR(1) process using $\rho = 0$, $\rho = 0.10$, $\rho = 0.20$, $\rho = 0.33$, $\rho = 0.80$, and $\rho = 0.90$. The estimation of ρ is always made under the assumption that $\gamma = 0$, the misspecified model when $\gamma > 0$. The simulation used 10 design points and 20 clusters using REML estimation (Table 7).

Table 7. Average Estimate of ρ from Parametric Estimation

(10 design points, 20 clusters, and 500 iterations)

γ	$\rho = 0$	$\rho = 0.10$	$\rho = 0.20$	$\rho = 0.33$	$\rho = 0.80$	$\rho = 0.90$
0.00	0.00	0.09	0.19	0.32	0.79	0.89
0.10	0.01	0.10	0.20	0.32	0.76	0.86
0.20	0.05	0.13	0.21	0.32	0.69	0.76
0.25	0.07	0.14	0.22	0.32	0.65	0.71
0.30	0.09	0.16	0.23	0.32	0.61	0.66
0.40	0.13	0.19	0.25	0.32	0.55	0.57
0.50	0.16	0.21	0.26	0.32	0.49	0.50
0.60	0.19	0.23	0.27	0.32	0.45	0.46
0.70	0.21	0.24	0.28	0.32	0.42	0.43
0.75	0.22	0.25	0.28	0.31	0.41	0.41
0.80	0.23	0.25	0.28	0.31	0.40	0.40
0.90	0.24	0.26	0.28	0.31	0.38	0.38
1.00	0.24	0.26	0.28	0.30	0.36	0.36

When the model is correctly specified (γ=0) the estimate of ρ is nearly unbiased. Model misspecification unduly influences the estimate of ρ, as the results indicate that highly (weakly) correlated data appear less (more) correlated for large model misspecification. The value ρ=1/3 provides estimates close to the true value, regardless of γ. This may be an indication that the estimates are converging to a value close to 1/3. We see that model misspecification affects the P model's ability to fit as well as its ability to estimate ρ in an AR(1) model.

5 Application of methods

The P, NP, and model robust methods discussed above can be applied to the wind speed data set from Haslett and Raftery [1989]. The wind speed data is given in the Appendix. Twelve meteorological stations in Ireland were selected and the average wind speed in knots were measured daily during the years 1961 through 1978. This analysis looked at the average weekly wind speeds averaged over the eighteen years. The stations, or clusters, were randomly selected from all such stations in Ireland; consequently, the station is the random effect. Measurements were taken at the same fifty three time points for each station, making a total of 636 observations in the data set. Because of the parabolic trend (the data are shown in Figure 2), a quadratic model was selected as the P model. Two models were considered; a fixed effects model with a quadratic trend in week and a mixed effects model with only a random intercept term. The random intercept term was considered because each cluster had a similar shape. This results in CS fits that are parallel shifts of the PA curve. The within-cluster variation structure for the P model in the wind speed example is assumed to be AR(1) and the between-cluster variation is assumed to be of independent structure. Preliminary analysis showed that the model containing the random intercept is an improvement over the fixed effects model. Thus, our PA curve is estimated as

$$\hat{E}(\mathbf{Y}) = \mathbf{X}\hat{\beta} = \mathbf{X} * (12.6576 \;\; -0.2520 \;\; 0.004549)^T \tag{25}$$

and the estimated variance-covariance matrix is $\hat{\mathbf{R}} + (\hat{\sigma}^2_{b_0})\mathbf{ZZ}' = \hat{\mathbf{R}} + (7.2166)\mathbf{ZZ}'$ where $\hat{\mathbf{R}}_i$ has the estimate of the variance ($\hat{\sigma}^2 = 1.1325$) down the diagonal and the estimated covariances $\hat{c}_{jk} = \hat{\sigma}^2\hat{\rho}^{|j-k|} = (1.1325)(0.5169)^{|j-k|}$ in the (j,k) and (k,j) off-diagonal cells.

The conditional distribution of $\mathbf{Y}|\mathbf{b}$ is assumed to be normal with mean $\mathbf{X}\beta + \mathbf{Zb}$ and variance-covariance matrix $\mathbf{R}$. The cluster specific curves can be estimated as

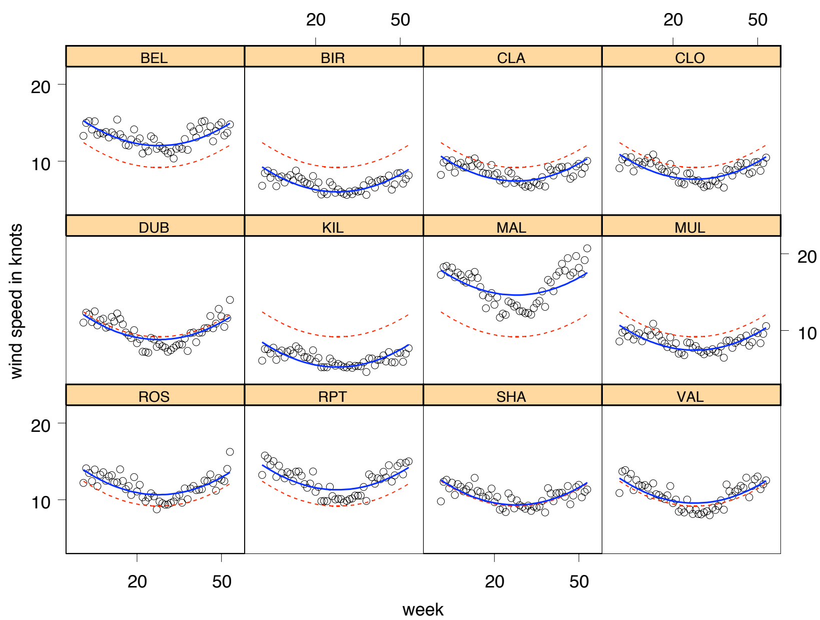
$$\hat{\mathbf{E}}(\mathbf{Y}) = \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{b}} \tag{26}$$

where $\hat{\beta}$ is as above, $\hat{\mathbf{b}} = (2.85 \;\; -3.16 \;\; -1.76 \;\; -1.52 \;\; -0.36 \;\; -3.93 \;\; 5.44 \;\; -1.73 \;\; 1.49 \;\; 2.13 \;\; 0.16 \;\; 0.40)^T$ and the estimated variance-covariance matrix is given by $\hat{\mathbf{R}}$ as above.

A trellis plot of the PA curve and CS curves by cluster (station) appears in Figure 2. The observations in the cluster are represented by the scatterplot. The dotted curve is the PA curve, and the solid curves are the CS curves. The PA curve is the same for every cluster. As shown in the equations and in the plots, the intercepts for the CS curves differ. Thus, the CS fit at each station is a parabola shifted up or down for a particular cluster. Notice that the PA curve fits poorly to some of the clusters, in particular to clusters MAL, KIL, BIR, and MUL. The CS curves are an improvement in fit to each cluster over the PA curve, as to be expected.

At virtually every station, the wind speeds remain relatively constant through January and February, and then diminish during the spring months. This is followed by a drop in wind speed during the middle of the year. This drop remains during the summer months (with a slight increase in wind speed for some stations during July). For some clusters, such as station BIR, the drop in wind speeds during the summer months is minimal. Other clusters,

Figure 2: Parametric Linear Mixed Model (Plot of Population Average and Cluster Specific Curves by Station)



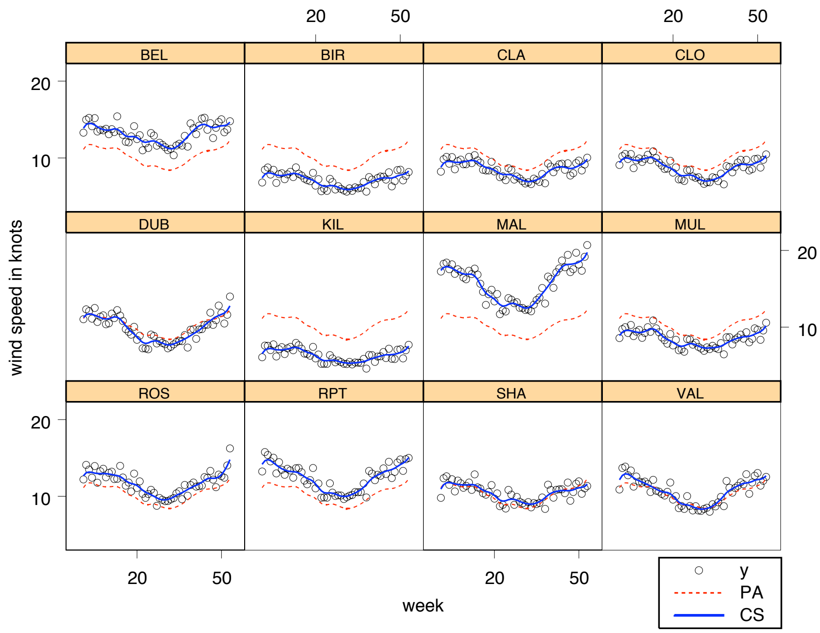
like station MAL, exhibit a steep drop in wind speed. During the fall and winter months, the wind speeds increase and then level off. The proposed P model is unable to model this type of trend, the level speeds in the winter months, combined with the decreased speeds in the summer months. The P model has been misspecified. The bias of the P linear model can be rectified through the use of NP, or local, mixed models.

The local mixed model used for this data set was the local linear mixed model with a random intercept. PA and CS curves were found for the conditional local mixed model and the PA curve was found for the marginal local mixed model. The between-cluster and within-cluster variation ($\widehat{\mathbf{B}}$ and $\widehat{\mathbf{R}}$) are assumed to be of independent structure. This differs from the within-cluster structure used in the parametric model. Research by Ruppert et al. [2003], Lin and Carroll [2006] and Li and Ruppert [2008] give asymptotic results that suggest the use of the independence structure for local kernel estimation. Both variance-covariance structures were studied for this example and the conclusion was to use independence for the local model due to fewer difficulties with variance component estimation. The independence structure allowed a wider range of bandwidths to be used.

For both the conditional and marginal local mixed models, PRESS and PRESS** both chose a bandwidth of 0.05. It was expected that the bandwidth chosen would be small. The dataset is quite large, so a small bandwidth gives weight to many observations. A small bandwidth is also needed in the conditional local CS analysis to be flexible enough to catch the sudden drop at station MAL.

As in the P linear mixed model, a PA curve and CS curves can be found for the conditional local mixed model. The marginal local mixed model will yield a PA curve. Recall that local linear mixed models were calculated at each value of the regressor $\widetilde{\mathbf{x}}$. Thus, for each $\widetilde{\mathbf{x}}_0$ in the conditional local mixed model, there is an estimate of parameter

Figure 3: Conditional Local Mixed Model with h=0.05 (Plot of Population Average and Cluster Specific Curves by Station)



vector $\hat{\beta}_0$ and a predictor of random effects vector $\hat{\mathbf{b}}_0$. For example, the CLMM population average fit at $\tilde{\mathbf{x}}$=1 using a bandwidth of 0.05 can be expressed, for each of the 12 stations, as

$$\widehat{Y}_{PA,0} = \tilde{\mathbf{x}}_0' \hat{\beta}_0^C = (1 \ \ 1)(10.74 \ \ 0.38)^T \tag{27}$$

and the cluster specific fits at $\tilde{\mathbf{x}}$=1 are

$$\hat{\mathbf{Y}}_{CS,0} = \tilde{\mathbf{x}}_0' \hat{\beta}_0^C + \hat{\mathbf{b}}_0 \tag{28}$$

where $\tilde{\mathbf{x}}_0'$ and $\hat{\beta}_0^C$ are as above, and $\hat{\mathbf{b}}_0 = (2.76 \ \ -3.72 \ \ -2.31 \ \ -1.73 \ \ 0.12 \ \ -4.59 \ \ 6.27 \ \ -2.23 \ \ 1.55 \ \ 3.04 \ \ -0.19 \ \ 1.03)^T$ For each $\tilde{x}_0$ in the marginal local mixed model, there is an estimate of the parameter vector $\hat{\beta}_0$. For a bandwidth of 0.05, the marginal local PA fit at $\tilde{\mathbf{x}}$=1 is

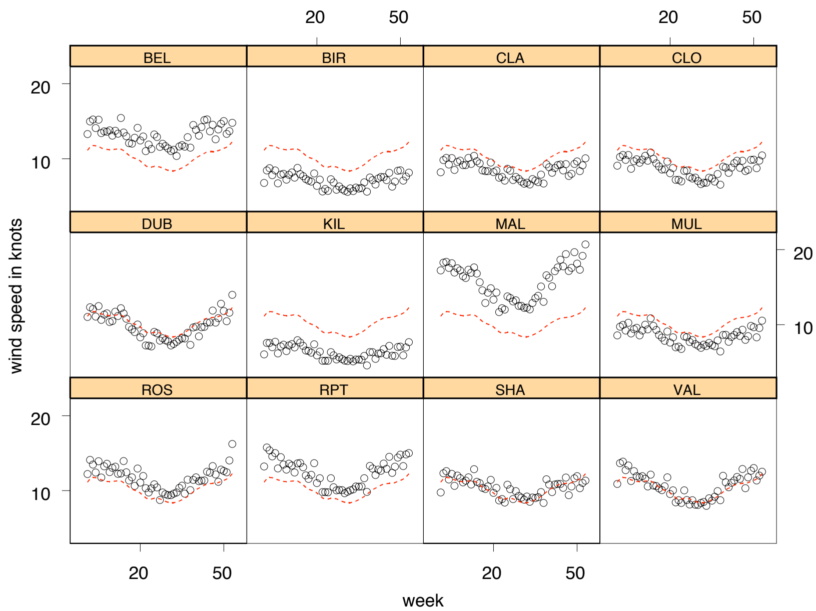
$$\widehat{Y}_{PA,0} = \tilde{\mathbf{x}}_0' \hat{\beta}_0^M = (1 \ \ 1)(10.74 \ \ 0.39)^T \tag{29}$$

Notice that the vector $\hat{\beta}_0$ for the two local models are close, indicating that the PA fits at $\tilde{\mathbf{x}}$=1 for the two local models are almost identical.

Trellis plots by cluster appear in Figures 3 and 4. Figure 3 contains the PA and CS curves for the conditional local mixed model using a bandwidth of 0.05. The PA is the dotted line and the CS fit is the solid line. Figure 4 plots the PA curve by station for the marginal local mixed model.

The PA fits are again the same for every cluster. For some of the clusters, the PA is a poor fit. The CS fits are impressive, however. Figure 5 is a comparison of the CLMM cluster specific fits with a bandwidth of 0.05 and the

Figure 4: Marginal Local Mixed Model with h=0.05 (Plot of Population Average Curve by Station)



parametric CS fits. The local cluster specific fits are tremendously flexible. Because they are fit pointwise, they no longer follow a particular form. In the P model, a random intercept term meant that the CS fits were shifted parabolas; they could never cross. This is not true with the local models. A random intercept term in the local model is also a shift, but it is a shift at a particular point. That shift differs as one moves across the values of the regressor. This allows local fits that potentially could cross.

The local fits are an improvement over the P fit. Notice that the drop in wind speed in midyear is captured in both the PA and CS fits, while capturing the level wind speeds in the winter months. The NP mixed model can capture this trend, whereas the specified P model was unable to model these trends.

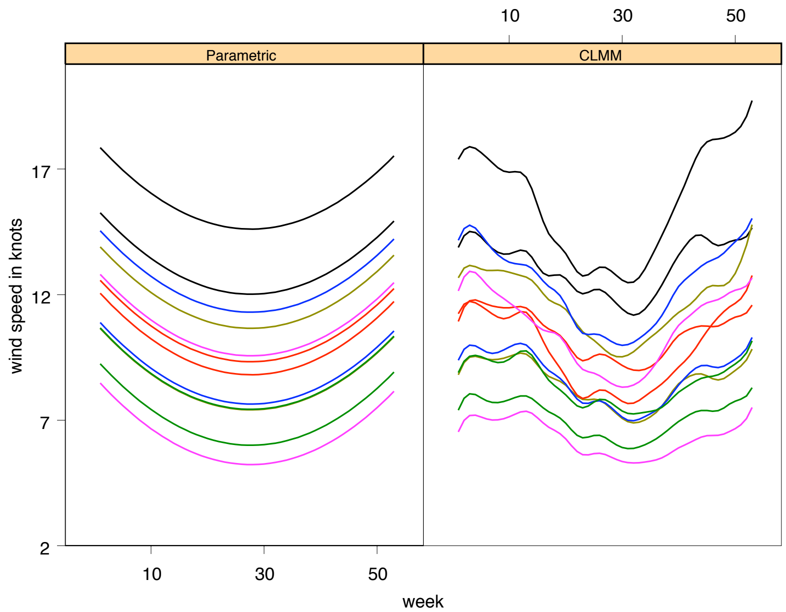
Our mixed model robust regression estimate is

$$\hat{\mathbf{Y}}^{MMRR} = (1-\lambda)\hat{\mathbf{Y}}^{P} + \lambda\hat{\mathbf{Y}}^{NP} \tag{30}$$

For estimating the PA curve, mixed model robust regression based on the conditional local mixed model approach utilizes the P and CLMM population average fits to find the MMRR population average fits. Mixed model robust regression using the marginal local mixed model uses the P and MLMM population average fits. Cluster specific mixed model robust regression uses the P and CLMM cluster specific fits in the calculation of the MMRR cluster specific fits.

For mixed model robust regression using CLMM, the estimate of λ was 0.86, and the estimate for MMRR involving MLMM was 1. A λ of 1 corresponds to a mixed model robust regression fit equal to the local fit, so MMRR using the marginal local mixed model is just the marginal local mixed model fit. The MMRR fit using the conditional local mixed model does not strictly use the conditional local fit, as λ does not equal 1. The estimation of λ

Figure 5: Plot of CLMM and Parametric Cluster Specific Fits(h=0.05)



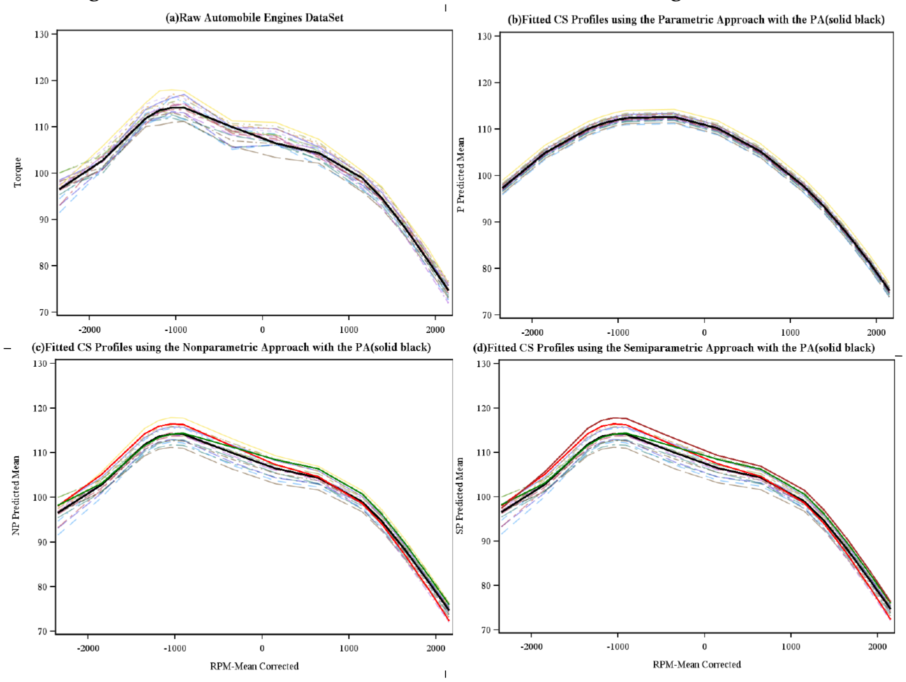
for MMRR using CLMM involves the CS fits, and the estimate of λ less than 1 suggests that the CS fits may benefit from the smoothness of the P regression curve. Because the estimate of λ is so close to one, it would be hard to distinguish between the local and mixed model robust fits in a trellis plot like those given above.

Notice that both estimates of lambda are fairly large. This is consistent with our findings given above; the P fit can be poor for some clusters, and there is a considerable difference between the P and NP fits for some clusters. The NP methods were an improvement over the P fits, and the estimates of λ should be close to 1. We also point out computing time for this example required only a few seconds using a moderately equipped PC.

As a second example, we consider the engine data, first presented in Amiri et al. [2010] and later studied by a number of authors, including Abdel-Salam et al. [2012]. The response variable of interest is the torque produced by the engine and it is of interest to model torque as a function of engine speed in revolutions per minute (RPM). In the context of quality control, specifically the subject of profile monitoring, a plot of torque versus RPM for each engine forms a profile. These profiles, from a randomly selected collection of engines, should be similar in appearance if the engines are produced from the same in-control manufacturing process. One issue in the profile monitoring analysis is to determine if any profile in this collection is 'different' from the others, indicating that the 'different' engine was produced by a manufacturing process that is in some sense out-of-control. In the engine data set (which may be found in the Appendix) there are a total of 20 engines. For each engine, the torque values are obtained at 14 different values of RPM: 1500, 2000, 2500, 2660, 2800, 2940, 3500, 4000, 4500, 5000, 5225, 5500, 5775 and 6000 RPM. The upper left plot in Figure 6 displays the 20 engines where straight line segments connect the 14 torque values for each engine. In this analysis, all RPM values are centered by subtracting their mean of 3750

The general approach in profile monitoring is to fit a curve to each profile to obtain the estimated profile curve and then compare these estimated curves by some appropriate control chart technique. Since the profiles are to be represented by the estimated profile curves, it is imperative that these curves provide adequate fits to the each profile.

Figure 6: The Engine Profiles Data Set with the Estimated Curves using the P, NP and SP Techniques



Casual inspection of the profiles for each engine reveals a strong quadratic trend. Thus a reasonable P model is a mixed linear model with torque as the response and the centered RPM and the (centered RPM)^2 as the regressors, along with the intercept. The model includes a normally distributed random coefficient term added to each of the three fixed effect coefficients. Since the 14 torque measurements per engine are taken from the same engine, the random error term (the ϵ term) is assumed to be follow a multivariate normal distribution with a zero mean vector and an AR(1) within engine correlation structure and a zero matrix between engine correlation structure. A plot of the 20 estimated profile curves based on this model is displayed in the upper right plot of Figure 6. It is seen that the quadratic trend is clearly captured by the model but several important characteristics of each engine (for example, the spike in torque at the centered RPM equal to about -1000 and the dip in torque at the centered RPM of about -500) are not captured by the P model. In addition, the individual engines are not as clearly distinguishable as in the profile plot in the upper left.

Since profile monitoring is concerned with comparing estimated profile curves, we use the conditional local mixed model (CLMM) as our NP procedure. Consequently, we fit a first order local mixed model using the centered RPM regressor with random effect coefficients for both the fixed effect intercept and the slope coefficients. Local weighting was provided by the simplified Gaussian kernel function. The results are seen in the lower left plot of Figure 6. Comparing this set of 20 estimated NP profile curves to the profiles in the upper left plot and the estimated P profile curves in the upper right plot reveals that the NP estimates are much closer to matching the individual engine characteristics. Furthermore, the CLMM method gives greater separation between the estimated profile curves than seen in the estimated P profile curves. Our semiparametric (SP) method, MMRR, applied to the engine data results in the estimated profile curves seen in the lower right of Figure 6. This collection of estimated profiles is very similar in appearance to those obtained by the CLMM method. This results it due to the estimated mixing parameter ($\hat{\lambda}$) equal to 0.98, indicating that the MMRR fits are almost entirely composed of the CMMM fit. Indeed, 98% of the MMRR fits are due to the CLMM fits and only 2% to the P fit. Clearly, the P fit is judged as being inadequate for this application.

6 Discussion

The local mixed model methods offer PA and CS 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 P model may be incapable of modeling. The local models are typically simple; fitting a local linear, as done in this paper, or a local cubic mixed model with a random intercept at each $\hat{\mathbf{x}}_0$ value will often suffice.

PRESS should be used as the bandwidth selector for PA estimation. Conversely, PRESS** is the bandwidth selector of choice for CS prediction. The bandwidth selectors are also performing as expected; evidence of this fact is found by comparing the bandwidths selected from PRESS and PRESS** with the optimal bandwidths from the simulation (results not presented here).

The simulation studies indicate that the marginal local mixed model should be used for PA estimation. When using PRESS as the bandwidth selector, the marginal model outperformed the conditional local mixed model in terms of minimizing the INTMSE. In addition, the model robust mixed model using CLMM has larger INTMSE values when estimating the PA curve when compared to the mixed model robust values using MLMM for moderate to large model misspecification. For CS prediction, the conditional local mixed model should be used, as the marginal local mixed model is inappropriate for CS inference.

The mixed model robust methods, using the marginal local mixed model for the PA and the conditional local mixed model for CS estimation, are extremely competitive in terms of minimizing the mean square error. With no misspecification, the parametric model should have the smallest INTMSE; the model robust methods are very close to the P values for the correctly specified model. For low to moderate misspecification (($0 < \gamma < 1$) in the simulation study, for example) the mixed model robust methods often have the smallest mean square error when compared to the P and local methods. When the model is grossly misspecified (for example, when γ=1 in the simulation study), the local methods have the minimum mean square errors, with the mixed model robust mean squares comparable to the local values.

Finally, we can conclude that working with correlated data creates results that may be counterintuitive. Our intuition, often based upon prior work with independent data, was often off the mark due to the lack of consideration of the correlated nature of our data. For example, at first it was counterintutive that the bandwidth in our local models would decrease as the amount of correlation increased. Upon further inspection, we realized that this finding was due to the marginal correlation inherent in the local mixed model. And although we felt that the misspecification term in our simulations would influence the estimate of the correlation, it was unexpected that as γ increased the estimates of ρ in the AR(1) cases either increased or decreased depending upon the magnitude of the correlation; further work indicated that the sinusoidal nature of the misspecification term was the reason for this result.

Appendix : The Wind Speed and Engine Data Sets

I. The Wind Speed Data Set

Week	BEL	BIR	CLA	CLO	DUB	KIL	MAL	MUL	ROS	RPT	SHA	VAL
1	13.29	6.78	8.20	9.07	11.03	6.07	17.24	8.57	12.20	13.22	9.78	10.89
2	14.97	8.45	9.83	10.23	12.32	7.58	18.25	9.75	14.10	15.75	12.34	13.67
3	15.21	8.75	10.13	10.50	12.08	7.56	18.39	9.99	13.48	15.38	12.60	13.85
4	14.10	7.83	9.24	9.86	11.09	7.01	17.56	9.23	12.42	14.55	11.49	12.73
5	15.19	8.44	10.10	10.50	12.49	7.74	18.19	10.27	13.92	15.02	12.23	13.35
6	13.43	6.73	8.53	8.66	10.66	6.19	16.94	8.83	11.75	12.97	10.64	11.27
7	13.67	7.89	9.56	9.66	11.34	7.12	17.52	9.40	13.25	14.45	12.02	12.58
8	13.58	8.00	9.77	9.93	11.51	7.36	17.23	9.61	13.57	13.47	11.69	11.92
9	13.79	7.24	9.17	9.47	10.36	6.49	16.46	8.58	12.48	12.81	11.11	11.78
10	13.08	7.87	9.18	9.88	10.48	7.23	16.23	9.21	13.05	13.59	11.35	11.92
11	13.73	8.18	10.13	10.41	11.74	7.41	17.34	10.00	13.20	13.37	11.79	11.60
12	13.35	7.50	9.31	9.74	11.16	6.97	16.93	9.39	12.24	12.43	10.81	10.56
13	15.42	8.77	10.38	10.85	12.21	7.93	17.63	10.85	12.28	13.65	12.86	12.15
14	13.50	7.94	9.66	9.97	11.50	7.62	16.82	9.84	713.96	13.70	11.18	10.72
15	13.02	7.69	9.38	9.38	10.80	7.19	15.66	9.32	12.45	13.13	11.00	10.62
16	12.09	7.30	8.44	8.98	9.73	6.56	14.58	8.65	11.37	12.05	10.40	10.30
17	12.02	7.089	8.36	8.57	9.38	6.51	12.91	8.27	11.77	11.56	10.21	10.08
18	12.81	6.96	8.35	8.60	9.01	6.27	14.27	7.85	10.64	12.11	10.55	10.59
19	14.12	8.04	9.50	9.63	10.03	7.41	14.86	9.11	12.95	13.68	11.44	11.78
20	12.44	6.40	8.22	8.06	8.33	5.92	13.33	7.61	11.08	11.02	9.69	9.55
21	12.97	7.22	8.43	8.75	8.87	6.42	14.30	8.32	11.94	11.65	10.38	10.03
22	11.00	5.66	7.48	7.20	7.23	5.19	11.67	6.99	10.33	9.81	8.70	8.45
23	11.90	5.98	7.60	7.20	7.22	5.19	12.20	7.09	9.78	9.83	8.88	8.73
24	11.29	5.71	7.09	6.99	7.10	5.20	12.02	6.78	10.30	9.81	8.41	8.35
25	13.24	7.27	8.65	8.37	9.04	6.34	13.80	8.44	10.84	11.67	10.81	10.03

Adapted from Haslett, J. and Raftery, A.E. (1989), "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resources (with Discussion)," Applied Statistics, **38**,1-50.

Week	BEL	BIR	CLA	CLO	DUB	KIL	MAL	MUL	ROS	RPT	SHA	VAL
26	12.91	6.85	8.40	8.43	8.85	5.92	13.57	8.35	10.44	10.47	10.21	8.74
27	11.59	5.88	7.17	7.59	7.88	5.40	13.11	7.71	8.76	10.04	8.88	8.16
28	11.99	6.33	7.59	7.38	8.20	5.64	13.25	7.96	9.76	10.11	9.93	8.71
29	11.70	5.96	7.33	7.48	7.94	5.44	12.52	7.43	9.54	10.08	9.28	8.14
30	11.40	5.75	6.97	7.04	7.77	5.21	12.51	7.21	9.38	9.65	9.09	8.09
31	11.03	5.60	6.67	6.62	7.22	5.16	12.25	6.91	9.38	9.87	8.82	8.36
32	11.19	6.04	6.80	6.81	7.45	5.46	12.29	7.34	9.60	10.06	9.22	8.50
33	10.35	5.70	6.58	6.81	7.74	5.12	12.05	7.15	9.77	10.17	8.54	7.98
34	11.62	6.14	7.30	7.46	8.03	5.37	12.91	7.56	10.37	10.57	9.10	8.97
35	11.83	6.03	7.08	7.36	8.19	5.37	13.54	7.27	10.66	10.53	9.00	8.68
36	11.61	6.10	6.88	6.97	8.16	5.32	13.86	7.14	9.61	10.57	9.14	9.20
37	12.87	6.86	7.91	7.96	8.93	5.84	15.10	7.90	11.48	11.65	9.82	9.93
38	11.48	5.61	6.69	6.55	7.33	4.58	13.05	6.44	10.08	9.83	8.35	8.72
39	14.50	7.51	9.27	8.94	9.66	6.37	16.58	8.65	11.45	13.31	11.53	11.99
40	13.87	7.16	8.87	8.83	9.74	6.34	16.26	8.67	11.78	12.96	10.83	11.10
41	13.08	6.58	8.13	8.12	8.46	5.46	15.12	7.73	11.26	12.09	9.81	10.47
42	14.21	7.38	8.76	8.82	9.68	6.16	17.14	8.37	11.32	12.81	10.83	11.07
43	15.15	7.58	9.17	9.41	9.73	6.20	17.64	8.65	11.37	12.67	10.90	11.36
44	15.23	7.56	9.24	9.75	10.28	6.74	18.68	9.02	12.51	13.63	10.90	11.83
45	13.66	7.15	8.44	8.80	10.33	5.98	17.79	8.51	12.40	12.91	10.39	11.57
46	14.54	8.11	9.29	9.74	11.89	7.18	19.40	9.33	13.29	14.50	11.77	12.88
47	12.58	6.30	7.70	8.35	10.30	5.87	17.15	8.00	12.17	12.37	9.40	10.32
48	13.91	6.97	8.00	8.54	10.88	5.84	17.56	8.47	11.19	13.19	10.55	11.78
49	14.68	8.39	9.32	9.86	12.79	7.03	19.69	9.83	12.82	14.41	11.63	12.67
50	15.02	8.47	9.64	9.89	11.82	7.09	18.15	9.67	12.62	14.81	11.93	13.02
51	13.33	7.07	8.35	8.72	10.47	5.92	17.34	8.33	12.44	13.30	10.32	11.38
52	13.70	7.64	9.15	9.76	11.62	6.98	19.16	9.56	14.03	14.82	11.04	12.01
53	14.80	8.16	10.04	10.47	13.97	7.68	20.70	10.55	16.24	15.01	11.34	12.53

II. The Engine Data Set

20 Engine profiles Data Set. Abdel-Salam, Birch and Willis (2012)

RPM (x)	T-E329	T-E449	T-E642	T-E724	T-E803	T-E930	T-E1148	T-E1171	T-E2600	T-E3100
1500	98.53	96.35	96.7	96.75	97.61	100.06	94.55	96.48	96.83	100.07
2000	102.65	100.74	100.05	100.87	102.46	103.6	103.22	100.87	103.78	103.91
2500	113.82	110.67	111.17	110.14	112.18	112.74	112.99	110.81	114.3	112.52
2660	115.26	113.06	111.51	110.48	112.99	113.56	114.18	113.2	114.62	113.25
2800	116.24	114.58	112.01	110.94	114.54	112.85	116.48	114.73	117.19	114.1
2940	117.06	114.98	111.23	111.17	115	114.49	115.33	115.13	116.61	114.1
3500	109.89	108.55	105.64	105.78	108.99	108.95	109.59	108.69	110.43	109.21
4000	109.65	107.41	106.02	103.37	107.95	108.24	108.47	107.55	109.61	108.34
4500	105.72	103.9	103.11	102.23	103.65	105.56	105.27	104.03	106.32	104.87
5000	99.74	97.99	97.4	96.06	96.94	98.92	97.9	98.12	99.44	98.35
5225	95.97	94.27	93.88	92.39	92.78	95.41	94.67	94.39	95.62	94.76
5500	89.47	88.45	88.17	86.54	86.41	89.19	88.23	88.56	89.46	88.93
5775	81.96	81.44	81.18	79.31	78.6	81.85	80.86	81.54	82	82.19
6000	74.9	75	75.03	73.13	71.97	75.09	73.93	75.09	75.83	75.8
RPM (x)	T-E4068	T-E4926	T-E6143	T-E6844	T-E7811	T-E8007	T-E8623	T-E9388	T-E9404	T-E10430
1500	97.98	97.29	93.13	93.11	95.38	98.28	96.79	96.45	91.53	98.37
2000	104.98	105.86	101.02	103.43	101.25	101.29	103.64	104.52	100.72	102.4
2500	114.9	115.25	111.25	112.02	111.53	112.2	112.73	113.78	110.71	112.67
2660	116.06	117.83	111.83	113.2	112.11	112.57	113.92	114.59	111.72	113.76
2800	116.65	117.97	113.27	113.77	112.6	113.06	113.35	115.4	112.29	115.41
2940	116.18	117.77	113.04	113.77	111.76	112.37	112.78	115.86	111.61	113.01
3500	109.65	111.31	105.6	109.15	108.12	107.03	108.2	110.78	105.21	110.08
4000	109.06	110.97	106.15	108.05	106.62	106.37	107.06	110.21	106.22	109.51
4500	105.01	107.37	104.12	103.46	102.92	104.1	105.27	106.75	101.73	106.09
5000	97.43	100.53	97.45	98.26	96.35	98.01	98.47	99.94	96.59	99.84
5225	94.04	97.17	94.68	94.26	93.14	94.21	95.67	96.94	93.78	96.46
5500	87.51	90.47	88.59	89.09	86.75	87.53	89.41	90.24	87.29	90.16
5775	79.36	83.51	81.08	81.06	80.27	80.08	82.57	82.65	78.97	82.74
6000	72.34	76.34	75.77	74.14	73.47	73.9	76.31	76.76	72.8	75.82

References

Abdel-Salam, A.-S. G., Birch, J. B., and Jensen, W. A. (2012). A semiparametric mixed model approach to phase I profile monitoring. *Quality and Reliability Engineering International, DOI: 10.1002/qre.1405*.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16:125–127.

Amiri, A., Jensen, W. A., and Kazemzadeh, R. B. (2010). "A Case Study on Monitoring Polynomial Profiles in the Automotive Industry". *Quality and Reliability Engineering International*, 26-5:509–520.

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.*

Buskirka, T. D. and Lohr, S. L. (2005). Asymptotic properties of kernel density estimation with complex survey data. *Journal of Statistical Planning and Inference*, 128:165–190.

Clark, S. (2002). *Model robust regression based on generalized estimating equations.* PhD thesis, Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, VA.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematics*, 31:377–403.

DeBoor, C. (2001). *A Practical Guide to Splines (Applied Mathematical Sciences).* A Practical Guide to Splines (Applied Mathematical Sciences), Springer,.

Demidenko, E. (2004). *Mixed Models : Theory and Applications.* Mixed Models : Theory and Applications, John Wiley and Sons, New York, NY,.

Fan, J. and Gijbels, I. (1995). Adpative order polynomial fitting: bandwidth robustification and bias reduction. *Journal of Computational and Graphical Statistics*, 4:213–227.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66.* Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66, London: Chapman & Hall/CRC Monographs on Statistics & Applied Probability,.

Faraway, J. J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Extending the Linear Model with *R*: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Chapman & Hall/CRC,.

Francisco-Fernandez, M. and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *The Canadian Journal of Statistics*, 33:279–295.

Gy, Q. and Zy, Z. (2009). Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data. *Biometrics*, 65:52–59.

Härdle, W. (1990). *Applied Nonparametric Regression.* Applied Nonparametric Regression, Cambridge University Press, New York, NY,.

Härdle, W. and Marron, J. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 13:1465–1481.

Härdle, W. and Marron, J. (1995). Fast and simple scatteplot smoothing. *Comp. Statist. Data Anal.*, 20:1–17.

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.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer,.

Henderson, C. (1950). "Estimation of Genetic Parameters". *Annals of Mathematical Statistics*, 21:309–310.

Hurtado-Rodriguez, G. (1993). *Detection of Influential Observations in Linear Mixed Models.* PhD thesis, North Carolina State University, Raleigh, NC.

Lee, D.-J. and Durbán, M. (2009). Smooth-car mixed models for spatial count data. *Computational Statistics & Data Analysis*, 53:2968–2979.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, 95:415–436.

Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems. *J. R. Statist. Soc. B*, 68:69–88.

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.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *J. R. Statist. Soc. B*, 68:179–199.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, 9:141–142.

Nottingham, Q. and Birch, J. (2000). "A Semi-Parametric Approach to Analysing Dose-Response Data". *Statistics in Medicine*, 19:389–404.

Opsomer, J. D. and Francisco-Fernàndez, M. (2010). Finding local departures from a parametric model using nonparametric regression. *Statistical Papers*, 51:69–84.

Pickle, S. M., Robinson, T. J., Birch, J. B., and Anderson-Cook, C. (2008). "A Semi-Parametirc Approach to Robust Parameter Design". *Journal of Statistical Planning and Inference*, 138:114–131.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus*. Mixed Effects Models in S and S-Plus, Springer, New York, Inc.,.

QingGuo, T. and LongSheng, C. (2009). B-spline estimation for varying coefficient regression with spatial data. *Science in China Series A: Mathematics*, 52:2321–2340.

Robinson, T., Birch, J. B., and Starnes, A. (2010). "A Semi-Parametric Approach to Dual Modeling When no Replication Exists". *Journal of Statistical Planning and Inference*, 140:2860–2869.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Semiparametric Regression, Cambridge University Press, Cambridge, NY,.

Schabenberger, O. and Pierce, F. J. (2002). *Contemporary Statistical Models for the Plant and Soil Sciences*. Contemporary Statistical Models for the Plant and Soil Sciences, CRC Press, Boca Raton, Florida,.

Silva, D. N. D. and Opsomer, J. D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Survey Methodology*, 35:165–176.

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.

Wan, W. and Birch, J. B. (2011). "A Semiparametric Technique for the Multi-Response Optimization Problem". *Quality and Reliability Engineering International*, 27:47–59.

Watson, G. (1964). Smooth regression analysis. *Sankhya Series A*, 26:359–372.

West, B., Welch, K. B., and Galecki, A. T. (2006). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Linear Mixed Models: A Practical Guide Using Statistical Software, Chapman & Hall/CRC,.

Wu, H. and Zhang, J.-T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, 97:883–897.