Profile Monitoring via Linear Mixed Models

WILLIS A. JENSEN, JEFFREY B. BIRCH, and WILLIAM H. WOODALL

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

Profile monitoring is a relatively new technique in quality control used when the product or process quality is best represented by a profile (or a curve) at each time period. The essential idea is often to model the profile via some parametric method and then monitor the estimated parameters over time to determine if there have been changes in the profiles. Previous modeling methods have not incorporated the correlation structure within the profiles. We propose the use of linear mixed models to monitor the linear profiles in order to account for the correlation structure within a profile. We consider various data scenarios and show using simulation when the linear mixed model approach is preferable to an approach that ignores the correlation structure. Our focus is on Phase I control chart applications.

KEY WORDS: eblups, linear mixed models, multivariate statistical process control, phase I, T^2 statistic.

Introduction

Control charts are known to be effective tools for monitoring the quality of processes and are applied in many industries. Data occur sequentially in time and often data are reduced to a statistic or two which represent the current state of the process. Phase I of the monitoring scheme consists of determining whether or not historical data indicate a stable (or in-control) process. Phase II consists of monitoring future observations using control limits calculated from Phase I data to determine if the process continues to be in-control. A more recent field of research in quality control known as profile monitoring deals with functional data or a curve (the "profile") collected at regular time intervals. Profile monitoring is becoming more prevalent because of the large amounts of data available with today's more sophisticated data collection systems. It combines the idea of fitting models from regression with the idea of separating common cause variability from special cause variability in quality control. For a detailed overview to the concepts, examples of its application, and a review of the literature, see Woodall et al. (2004).

Kang and Albin (2000), Kim, Mahmoud, and Woodall (2003), Mahmoud and Woodall (2004), and Wang and Tsung (2005) all considered monitoring of linear profiles. The idea of these previous methods is to model the linear profiles using some parametric method and then monitor the estimated parameters over time to determine if the profile change. Our approach builds on this basic idea. Because the parameter estimators in our case may be correlated, it is convenient to monitor them using a multivariate control method such as one based on the T^2 statistic.

The previous work on profile monitoring has been based on the assumption that the profiles are independent of each other and that the random errors associated with the measurements within a profile are also independent of each other. This is often an unrealistic

assumption in practice for many types of data. For example, profiles may exhibit spatial correlation if they represent measurements of the physical dimensions of an object. They may exhibit serial correlation if the observations within a profile are collected over time. Therefore, we propose the use of mixed models to monitor the profiles in order to account for the correlation structure within profiles and show using simulation situations when the mixed model approach is preferable. We will focus here on Phase I control chart applications.

We first discuss the linear mixed model formulation and the properties of the resulting estimators. We review briefly the literature that has proposed methods for detecting outlying data in the linear mixed model and present our proposed method. We explain the setup of the simulation studies performed and show our results for several different data scenarios. We conclude with an example of our proposed method applied to linear calibration data.

Model Formulation

The linear mixed model (LMM) is very flexible and capable of fitting a large variety of datasets. It is widely used for repeated measures data or longitudinal studies where data are grouped. The form of the LMM that we use is that of Laird and Ware (1982) which can be considered an extension of the classical linear model. The literature on linear mixed models will often refer to the collection of data that forms a profile as a cluster or subject, depending on the particular application. We use the term profile throughout but note that applications of the methods and analysis presented here apply if the data are represented by clusters or subjects. The LMM allows us to account for the correlation within profiles and to consider the profiles as a random sample from a common population distribution, which may be more realistic.

If we have m profiles of data, each of which has n_i measurements, where i refers to the

i^{th} profile, we can fit a separate linear model to each profile. The model fit in matrix form is given by

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \text{ for } i = 1, 2, \ldots, m \tag{1}$$

where $\mathbf{y}_i$ is a n_i by 1 vector of responses for profile i, $\mathbf{X}_i$ is a n_i by p matrix of the regressor variables associated with the fixed effects, $\boldsymbol{\beta}_i$ is the p by 1 parameter vector of fixed effects for the i^{th} profile, and $\boldsymbol{\epsilon}_i \sim MN(\mathbf{0}, \mathbf{R}_i)$ is the n_i by 1 vector of errors where $\mathbf{R}_i$ is a n_i by n_i positive definite matrix. If the errors are assumed to be independent, then $\mathbf{R}_i = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix and the estimates of the parameters can be easily obtained using least squares (LS) methods. The estimated parameter vector for each profile is given by

$$\widehat{\boldsymbol{\beta}}_{i,LS} = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y_i} \text{ for } i = 1, 2, \ldots, m. \tag{2}$$

In contrast, the LMM has random effects in addition to the fixed effects of the classical linear model and is given by

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \text{ for } i = 1, 2, \ldots, m, \tag{3}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects that is the same for all profiles, $\mathbf{Z}_i$ corresponds to a n_i by q matrix of the predictor variables with random effects, $\mathbf{b}_i \sim MN(\mathbf{0}, \mathbf{D})$ is a q by 1 vector of random effects for the i^{th} cluster where $\mathbf{D}$ is a q by q positive definite matrix. The model is flexible enough to allow the errors to be independent or correlated. If correlated, $\mathbf{R}_i$ is often assumed to be a simple form such as compound symmetry (CS) or autoregressive (AR) in order to reduce the number of covariance parameters that need to be estimated. For more details on the various types of correlated error structures that can be assumed for $\mathbf{R}_i$, see Littell et al. (1996) or Schabenberger and Pierce (2002). Similar structure can be imposed on $\mathbf{D}$, but here we restrict $\mathbf{D}$ to be a diagonal matrix. Thus the random effects are assumed to be uncorrelated with each other.

In addition, we assume that $cov(\boldsymbol{\epsilon}_i, \mathbf{b}_i) = \mathbf{0}$, which means that the random effects and the random errors are uncorrelated, resulting in the conditional model given by

$$\mathbf{y}_i|\mathbf{b}_i \sim MN(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i). \tag{4}$$

Furthermore, we assume that $\mathbf{Z}_i$ is either a subset of or equal to the $\mathbf{X}_i$ matrix, so any columns in the $\mathbf{Z}_i$ matrix are also contained in the $\mathbf{X}_i$ matrix. The case where $\mathbf{Z}_i = \mathbf{X}_i$ is referred to as the random coefficients model (Demidenko, 2004) because all the fixed effects have a corresponding random effect. This restriction of $\mathbf{Z}_i$ being contained within $\mathbf{X}_i$ does not eliminate any of the forms of this model that are in common practice. For examples of cases where this restriction is used see Waternaux, Laird, and Ware (1989), Lesaffre, Asefa, and Verbeke (1999), Longford (2001), or Xu (2003).

The corresponding marginal model is given by

$$\mathbf{y}_i \sim MN(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i) \text{ for } i = 1, 2, \ldots, m, \tag{5}$$

where $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i$ is a n_i by n_i positive definite matrix.

The model allows for two levels of correlation for the measurements within a profile. The first results from the random effects which cause all the measurements within a profile to be correlated to each other. The second results from the within-profile variance-covariance matrix, $\mathbf{R}_i$. Vonesh and Chinchilli (1997, p. 256) noted that in some applications it makes sense to consider both levels and give some references where both levels are used. Chi and Reinsel (1989) also recommended the use of both levels of correlation where needed.

If a particular application has only the first level of correlation, the response is distributed as $\mathbf{y}_i \sim MN(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i)$ where $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}$. If the application only uses the second level of correlation (that is, a fixed effects model where $\mathbf{D}$ is a null matrix and $\mathbf{R}_i$ is non-diagonal) then a time series model can often be fit to account for serial correlation among

the responses. Time series models can be more restrictive because they often require equally spaced data. A LMM that uses neither of the two levels of correlation is simply the classical general linear model in (1) because $\mathbf{Z}_i = \mathbf{0}$ and $\boldsymbol{\epsilon}_i \sim MN(\mathbf{0}, \sigma^2\mathbf{I})$.

On occasion, a time series model will be fit to multiple profiles but such models often require a large number of observations per profile to ensure that the model obtained will be representative of the data. On the other hand, the LMM is usually preferable when there are multiple profiles and there are a smaller number of observations per profile which may/or may not be time ordered. With the LMM one seeks to pool information from multiple profiles in order to improve estimates and subsequent inference while with a time series model one does not usually attempt to pool information.

Data Scenarios

Profile monitoring data can be classified into several different scenarios depending on the number of observations per profile and where those observations are located within the profile. For example, all of the profiles can have measurements at the same location for all profiles. We refer to this scenario as balanced, whether or not the locations are equally spaced from each other. This implies that $n_i = n$ for $i = 1, 2, \ldots, m$, and that the values of the regressors (and consequently, $\mathbf{X}_i$ and $\mathbf{Z}_i$) are the same for all profiles.

Unbalanced data refers to the situation where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are not necessarily the same for all the profiles. They may not even have the same number of rows per profile, which indicates an unequal number of observations per profile. We believe that profile monitoring applications are more likely to have balanced data (equally or unequally spaced) because the emphasis in quality control is consistent data collection to ensure that changes in the responses are a result of changes in the quality characteristics being monitored. Nonetheless,

we will consider both balanced and unbalanced data in our simulation studies. In order to simplify our comparisons, balanced and unbalanced data will both have the same number of observations for all the profiles, that is $n_i = n$ for $i = 1, 2, \ldots, m$.

Estimation in the LMM

Under the distributional assumptions of the marginal model in (5), the fixed-effect parameter estimators representing the population average of all the profiles is given by $\widehat{\boldsymbol{\beta}}_{MIX}$, and the estimates of the random deviations from that population average are given by $\widehat{\mathbf{b}}_i$. If $\mathbf{V}_i$ (and consequently $\mathbf{D}$ and $\mathbf{R_i}$) are assumed known then it can be shown that

$$\widehat{\boldsymbol{\beta}}_{MIX} = \left(\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right), \tag{6}$$

and the best linear unbiased predictors ("blups") as

$$\widehat{\mathbf{b}}_i = \mathbf{D}\mathbf{Z}_i' \mathbf{V}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{MIX} \right). \tag{7}$$

It can be shown that $\widehat{\boldsymbol{\beta}}_{MIX} \sim MN\left[\boldsymbol{\beta}, (\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}\right]$ under the assumption of multivariate normality (Schabenberger and Pierce, 2002).

In practice, $\mathbf{V}$ is not known and therefore must be estimated prior to obtaining $\widehat{\boldsymbol{\beta}}_{MIX}$ and $\widehat{\mathbf{b}}_i$. $\mathbf{V}$ can be estimated via maximum likelihood (ML) or restricted maximum likelihood (REML) and an iterative algorithm. REML is often preferred (Schabenberger and Pierce, 2002, p. 437) because it produces estimators with less bias than estimators obtained using ML. The estimates obtained from ML and REML are often very similar to each other and can sometimes be asymptotically equivalent (Demidenko, 2004, p. 146). Nonetheless, we utilize REML for all of our simulation studies.

Once the solution is obtained, the estimated values of $\widehat{\mathbf{V}}^{-1}$ and $\widehat{\mathbf{D}}$ are then placed in (6) and (7) to obtain the parameter estimates. If a consistent estimate of $\mathbf{V}$ is used, the

distribution of $\widehat{\boldsymbol{\beta}}_{MIX}$ will be asymptotically normal (Demidenko, 2004), that is $\widehat{\boldsymbol{\beta}}_{MIX} \overset{a}{\sim} MN\left[\boldsymbol{\beta}, (\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}\right]$. The blups from (7) are referred to estimated best linear unbiased predictors ("eblups") when an estimated variance-covariance matrix is used. As noted by Verbeke and Lesaffre (1996) and Ritz (2004), the distribution of the eblups does depend on the distribution of both the $\mathbf{b}_i$'s and the $\boldsymbol{\epsilon}_i$'s. In particular, the eblups will be normally distributed as long as the random effects and errors follow a multivariate normal distribution although the distribution of the eblups is not necessarily the same as that of the blups.

The estimated parameter vector for the i^{th} profile is given by

$$\widehat{\boldsymbol{\beta}}_{i,MIX} = \widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i \text{ for } i = 1, 2, \ldots, m, \tag{8}$$

which can be found using (6) and (7). Thus rather than using the actual data in the T^2 statistic we use the estimates of the model coefficients obtained from the data. We are reducing the problem of detecting changes in the data profiles to detecting changes in the parameters that summarize the profiles. This is a more efficient approach because we are monitoring a smaller number of parameters rather than a larger number of data observations. Of course, our approach is based on the assumption that the fitted model adequately describes the profile data.

When obtaining the estimates of $\mathbf{V}$ in the LMM, non-convergence of the iterative algorithm can occur. For most simple problems non-convergence is rare but is more common when the data are unbalanced, the variance components in $\mathbf{V}$ are small and/or the model has been misspecified (Verbeke and Molenberghs, 2000, Chapter 5.6). For all our simulation studies we tracked the frequency of non-convergence and found it to be small or non-existent. To reduce the frequency of non-convergence, it is often recommended to use good starting values for the fixed parameters and components of the variance-covariance matrix. These starting values can be obtained via graphical methods (Schabenberger and Pierce, 2002). In

some situations where the non-convergence was more likely to be present, we used in our simulations the known parameter values as starting values of the iterative algorithm as was done by Hartford and Davidian (2000) for nonlinear models. This reduces the frequency of non-convergence just as would occur if a knowledgeable researcher were to spend a sufficient amount of time exploring, cleaning, and appropriately analyzing a single dataset.

Check of Model Assumptions

When a parametric model is fit to profile data, it is important to know if the model fits the data well and if the model assumptions are met. If so, then the parameter estimates obtained will be a good representation of the profile and the estimates can then be used to determine if the Phase I data are in control. Goodness-of-fit techniques and other checks of the model assumptions for the LMM such as those discussed in Verbeke and Mohlenberghs (2000, Ch.4) and Demidenko (2004) can be used. If there is a not a good fit of the LMM to the data, then determining which profiles are outlying is a risky activity and should be used with caution.

Diagnostics in LMM

Diagnostic methods to detect outliers and influential points have been proposed in LMM but they are not well developed. The need for better or more utilized diagnostics for models with random effects and/or correlated errors has been noted by a number of authors, including, Ghosh and Rao (1994), Verbeke and Molenberghs (2000), Tan, Ouwens, and Berger (2001), Houseman, Ryan, and Coull (2004) and Haslett and Dillane (2004). Diagnostic methods are needed to detect outlying profiles as well as outlying observations within profiles. Our focus is determining outlying profiles rather than observations within a profile.

As noted by Langford and Lewis (1998), once the outlying profile is determined, it can be examined for outlying observations.

There are a wide variety of methods for determining outlying profiles in LMM. A case deletion method was proposed by Banerjee and Frees (1997). A local influence approach was proposed by Lesaffre and Verbeke (1998) extended the local influence approach of Cook (1986) to the LMM. A parametric bootstrap approach was proposed by Longford (2001). Tan, Ouwens, and Berger (2001) considered a Cook's distance measure and found that it does not work well in determining the correct outlying profile unless it is modified. Demidenko and Stukel (2005) derived alternative forms of leverage and Cook's distance measures.

Our approach determines outlying profiles based on the distance of the estimated parameter vector from the center of the group of estimated parameter vectors and is most like that of Waternaux, Laird, and Ware (1989), who proposed to detect outlying profiles by using the Mahalanobis distances of the eblups. They proposed to calculate

$$T^2_{varbi,i} = \widehat{\mathbf{b}}_i \, 'Var(\widehat{\mathbf{b}}_i - \mathbf{b}_i) \, \widehat{\mathbf{b}}_i \text{ for } i = 1, 2, \ldots, m, \tag{9}$$

where $Var(\widehat{\mathbf{b}}_i - \mathbf{b}_i)$ is given by (Harville, 1976; and Laird and Ware, 1982)

$$Var(\widehat{\mathbf{b}}_i - \mathbf{b}_i) = \mathbf{D} - \mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D} + \mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D}'. \tag{10}$$

To use the expression in (10) one must replace $\mathbf{D}$ and $\mathbf{V}_i$ by their estimates. Waternaux, Laird, and Ware (1989) proposed to use a Q-Q plot of the values for $T^2_{varbi,i}$ to detect outliers. We calculate the Mahalanobis distance as in (9) with different estimators of the variance-covariance matrix. We will evaluate the method of Waternaux, Laird, and Ware (1989) in our simulation studies to determine its efficacy.

T^2 Statistic for LMM

For most control chart applications, where the profiles occur at regular time periods, the data collection is well controlled as if from a designed experiment. Thus the number of measurements per profile will often be the same and at the same locations along the profile.

Using (8) the two T^2 statistics based on the classical estimators are given by

$$T^2_{1,i,MIX} = (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})'\mathbf{S}^{-1}_{1,MIX}(\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX}) \text{ for } i = 1, 2, \ldots, m, \tag{11}$$

where

$$\mathbf{S}_{1,MIX} = \frac{\sum_{i=1}^{m}(\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})(\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})'}{m-1}, \tag{12}$$

and where

$$\overline{\boldsymbol{\beta}}_{MIX} = \frac{\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}_{i,MIX}}{m}. \tag{13}$$

$T^2_{1,i,MIX}$ will be proportional to a beta distribution if the matrix $\mathbf{V}_i$ is known, and asymptotically proportional to a beta distribution if a consistent estimate of $\mathbf{V}_i$ is obtained. However, as shown by Sullivan and Woodall (1996), $T^2_{1,i,MIX}$ is not effective in detecting sustained step changes in the mean vector, nor is it effective in detecting multiple outliers (Vargas, 2003). While $T^2_{1,i}$ has been shown to be effective in detecting a single moderately-sized multivariate outlier as shown in Figure 1 of Vargas (2003), a single arbitrarily large outlier or step change can render the $T^2_{1,i}$ statistic useless. We concur with the conclusions of Sullivan and Woodall (1996) and Vargas (2003) and do not recommend the use of the $T^2_{1,i}$ statistic for Phase I analysis when outliers or step changes are present.

An alternative is to base the T^2 statistic on the sample mean and the variance-covariance matrix of the successive differences between vectors (Holmes and Mergen, 1993). If $\mathbf{v}_i$ is the vector of the i^{th} successive difference, then an unbiased estimator of the variance-covariance

matrix is

$$\mathbf{S}_2 = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} \mathbf{v}_i \mathbf{v}_i^{'}. \tag{14}$$

Thus, in our application we have

$$T^2_{2,i,MIX} = (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})' \mathbf{S}^{-1}_{2,MIX} (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX}) \text{ for } i = 1, 2, \ldots, m, \tag{15}$$

where

$$\mathbf{S}_{2,MIX} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\widehat{\boldsymbol{\beta}}_{i+1,MIX} - \widehat{\boldsymbol{\beta}}_{i,MIX})(\widehat{\boldsymbol{\beta}}_{i+1,MIX} - \widehat{\boldsymbol{\beta}}_{i,MIX})'. \tag{16}$$

This statistic is analogous to the use of the moving range to construct an univariate Shewhart Individuals chart. Sullivan and Woodall (1996) showed that using successive differences is effective in detecting sustained step changes in the process that occur in Phase I data. While the distribution of $T^2_{2,i,MIX}$ does not have a simple closed form, its asymptotic distribution is χ^2_p. A discussion of the various approximate distributions and the preferred χ^2_p approximation for large samples is given in Williams et al. (2006). However, like $T^2_{1,i,MIX}$, $T^2_{2,i,MIX}$will not be effective in detecting multiple multivariate outliers (Vargas, 2003).

In the Appendix we show that $\sum_{i=1}^{m} \widehat{\mathbf{b}}_i = \mathbf{0}$ as long as the $\mathbf{Z}_i$ matrix is contained within the $\mathbf{X}_i$ matrix. This is true even if the $\mathbf{Z}_i$ and $\mathbf{X}_i$ matrices are different from profile to profile. In addition we show in the Appendix that when the eblups sum to zero, $\overline{\boldsymbol{\beta}}_{MIX} = \widehat{\boldsymbol{\beta}}_{MIX}$.

In addition, we show in the Appendix that the T^2 statistics shown in (11) and (15) can be expressed as a function of the eblups no matter the form of $\mathbf{X}_i$ and $\mathbf{Z}_i$. They are given respectively by

$$T^2_{1,i,MIX} = (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})'(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})}{m-1} \right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{ for } i = 1, 2, \ldots, m, \tag{17}$$

and

$$T^2_{2,i,MIX} = (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m-1} (\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)'(\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)}{2(m-1)} \right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{ for } i = 1, 2, \ldots, m. \tag{18}$$

These expressions simplify the computation of the T^2 statistics when fitting a LMM to the data.

A naive approach to compare m linear profiles is to ignore the correlation structure and the random effects and obtain the model parameters using the LS estimator of the classical linear model (LM) even though the data follow the model (3). We refer to this naive approach as the LS approach to distinguish it from the LMM approach. In the LS approach, the fixed parameters are estimated separately for each profile and are obtained using the expression in (2).

When taking the LS approach, we have

$$T^2_{1,i,LS} = (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})' \mathbf{S}^{-1}_{1,LS} (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS}) \text{ for } i = 1, 2, \ldots, m, \tag{19}$$

where

$$\mathbf{S}_{1,LS} = \frac{\sum_{i=1}^{m} (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})(\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})'}{m-1}, \tag{20}$$

and where

$$\overline{\boldsymbol{\beta}}_{LS} = \frac{\sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_{i,LS}}{m}, \tag{21}$$

and we have

$$T^2_{2,i,LS} = (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})' \mathbf{S}^{-1}_{1,LS} (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS}) \text{ for } i = 1, 2, \ldots, m, \tag{22}$$

where

$$\mathbf{S}_{2,LS} = \frac{1}{2(m-1)} \sum_{i=1}^{m} (\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})(\widehat{\boldsymbol{\beta}}_{i,LS} - \overline{\boldsymbol{\beta}}_{LS})'. \tag{23}$$

The LMM approach has several advantages over the LS approach, some of which were noted by Verbeke and Mohlenberghs (2000). First, the LMM can be easily fit for balanced and unbalanced data and is better than the LS approach when the number of observations per profile is small. The LMM approach combines information from the profiles to achieve

the model fit with fewer parameters than fitting separate regression models for each profile. Second, the LMM approach is capable of handling profiles with missing data, even for situations where the number of observations for a particular profile is less than the number of parameters that would be needed to fit a regression model to that individual profile.

Simulation Study Setup

We now explain the general procedure for the simulation studies used to compare methods. Multivariate data that follow a linear mixed model structure are generated where the random errors follow some specific structure. This is accomplished by generating univariate normal data and using the Cholesky decomposition to transform the generated univariate data to multivariate data. The data are fit with a LMM using proc mixed of $SAS^{®}$ with the correct model specification. We included both correlated and uncorrelated errors in our comparisons.

The control limit is established using the appropriate percentiles of the beta or χ^2 distributions so that the probability of signal for the in-control data is .05, the nominal value. The actual probability of signal is estimated by the proportion of datasets where there was a signal. That is, a signal occurs when at least one of the m T^2 statistics exceeds the control limit.

Here we consider the case of simple linear regression with a random slope and intercept so we have $p = 2$ and $\mathbf{X}_i = \mathbf{Z}_i$. For the studies performed, a total of $10,000$ datasets (Monte Carlo repetitions) were generated for each combination of interest unless otherwise noted. We can assume without loss of generality that $\boldsymbol{\beta} = [0, 1]$. This assumption results from the regression equivariance property of the estimators in both the classical LM and the LMM, as discussed in the Appendix.

Balanced Data

Our initial study is for balanced data situations. Here a commonly used error structure is an AR(1) model where ρ represents the amount of autocorrelation between successive, equally spaced observations. We varied n and ρ but held fixed $m = 30$, $\sigma_0^2 = .1$, $\sigma_1^2 = .1$, and $\sigma^2 = .1$. Thus $\mathbf{D} = diag(\sigma_0^2, \sigma_1^2)$ represents the variability in the slopes and intercepts from profile to profile. The statistics $T_{1,i}^2$ and $T_{2,i}^2$ are calculated from the LMM approach (the "right way" in that both the random effects and the correct correlation structure are accounted for) and the LS approach (the "naive way" where both the random effects and the correlation structure are ignored). Table 1 shows the proportion of the generated datasets that had a signal on the control charts for the various T^2 statistics. We also included in our study the version of the T^2 statistic shown in (9) as proposed by Waternaux, Laird, and Ware (1989) with $Var(\widehat{\mathbf{b}}_i)$ instead of $Var(\widehat{\mathbf{b}}_i - \mathbf{b}_i)$ because the use of $Var(\widehat{\mathbf{b}}_i - \mathbf{b}_i)$ resulted in extremely large probabilities of signal for in-control data. $Var(\widehat{\mathbf{b}}_i)$ is given in Laird and Ware (1982) by

$$Var(\widehat{\mathbf{b}}_i) = \mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D} - \mathbf{D}\mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{X}_i(\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D}', \tag{24}$$

with $\mathbf{D}$ and $\mathbf{V}_i$ replaced with their estimated values.

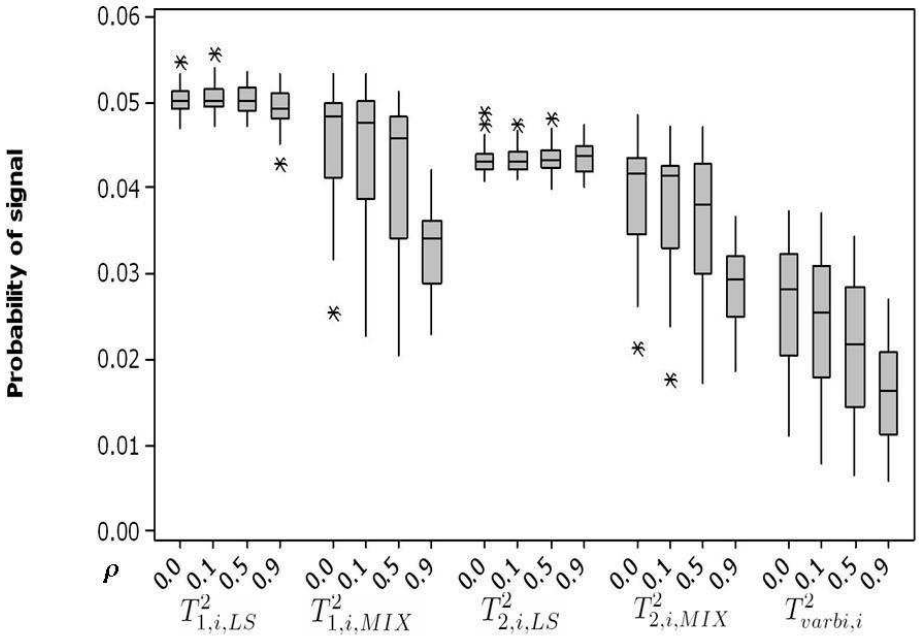
We see here that for the in-control situation, it appears that the more complicated mixed model analysis makes little difference in terms of a probability of a signal when the data are balanced and equally spaced. This is true for other values of m, n, and ρ not shown here. We note that the $T_{2,i,LS}^2$ and $T_{2,i,MIX}^2$ statistics give slightly smaller probability of signal than the nominal .05 level and that the statistics based on the LMM approach have slightly smaller probabilities than those based on the LS approach. The probability of a signal is very low for the $T_{varbi,i}^2$ statistic.

Table 1: Proportion of datasets with a signal for in-control data for the balanced, equally spaced data situation.

m	n	ρ	$T^2_{1,i,LS}$	$T^2_{1,i,MIX}$	$T^2_{2,i,LS}$	$T^2_{2,i,MIX}$	$T^2_{varbi,i}$
30	5	0	0.0507	0.0481	0.0464	0.0436	0.0193
30	5	0.1	0.0515	0.0465	0.0468	0.0427	0.0159
30	5	0.5	0.0506	0.0445	0.0438	0.0377	0.0134
30	5	0.9	0.0515	0.0334	0.0450	0.0274	0.0100
30	10	0	0.0524	0.0514	0.0413	0.0407	0.0215
30	10	0.1	0.0526	0.0506	0.0419	0.0426	0.0193
30	10	0.5	0.0496	0.0464	0.0456	0.0410	0.0159
30	10	0.9	0.0495	0.0355	0.0468	0.0317	0.0103
60	5	0	0.0497	0.0491	0.0441	0.0434	0.0298
60	5	0.1	0.0503	0.0495	0.0432	0.0425	0.0278
60	5	0.5	0.0524	0.0471	0.0456	0.0418	0.0263
60	5	0.9	0.0512	0.0342	0.0460	0.0307	0.0177
60	10	0	0.0497	0.0496	0.0450	0.0449	0.0303
60	10	0.1	0.0502	0.0503	0.0458	0.0456	0.0310
60	10	0.5	0.0504	0.0490	0.0451	0.0433	0.0281
60	10	0.9	0.0494	0.0407	0.0411	0.0331	0.0204
90	5	0	0.0492	0.0490	0.0426	0.0425	0.0349
90	5	0.1	0.0501	0.0499	0.0417	0.0403	0.0318
90	5	0.5	0.0518	0.0479	0.0444	0.0429	0.0299
90	5	0.9	0.0489	0.0340	0.0459	0.0310	0.0210
90	10	0	0.0531	0.0531	0.0445	0.0445	0.0374
90	10	0.1	0.0532	0.0527	0.0445	0.0440	0.0372
90	10	0.5	0.0528	0.0510	0.0445	0.0455	0.0345
90	10	0.9	0.0487	0.0415	0.0439	0.0368	0.0261

We performed similar studies with differences in the σ_0^2, σ_1^2, and σ^2 values used to generate the data in order to investigate their impact. We generated data for all possible combinations of 2 levels for each of $n = (5, 10)$, $\sigma_0^2 = (.1, 1)$, $\sigma_1^2 = (.1, 1)$, and $\sigma^2 = (.1, 1)$, 3 levels for $m = (30, 60, 90)$ and 4 levels for $\rho = (0, .1, .5, .9)$. Thus we have expanded the 24 combinations shown in Table 1 to a total of $4 * 3 * 2 * 2 * 2 * 2 = 192$ combinations. Rather than present the full table of results, we summarize the results by showing in Figure 1 the boxplots of the 5 T^2 statistics. Each boxplot contains all the combinations of m, n, σ_0^2, σ_1^2, and σ^2 and represents 48 combinations of settings, each of which has $10,000$ generated

Figure 1: Boxplots of the probability of signal for various combinations of m, n, σ_0^2, σ_1^2, and σ^2 for the different versions of the T^2 statistic.



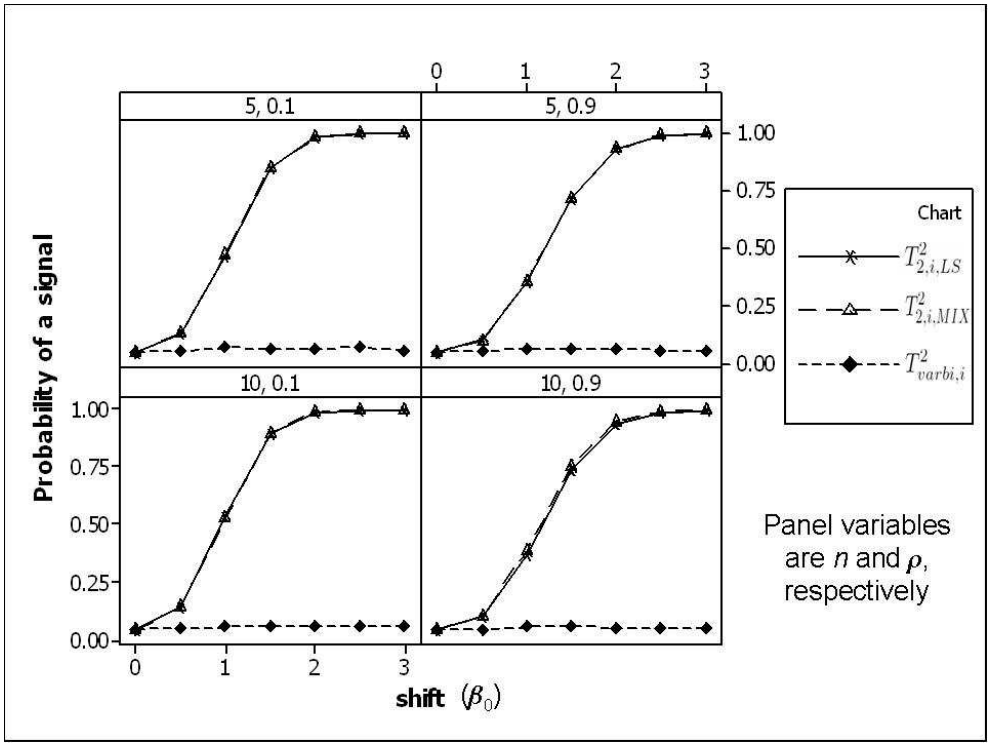
datasets.

We note that for the T^2 statistics based on the LS approach, there is little variability in the probability of signal for different values of m, n, σ_0^2, σ_1^2, and σ^2. There is more variability in the probability of signal for the T^2 statistics based on the LMM approach and for the $T^2_{varbi,i}$ statistic. Thus changing the values of m, n, σ_0^2, σ_1^2, and σ^2 to generate the data will have little impact on the estimated probability of signal for the LMM approach but no impact for the LS approach. Even with this variability, the probability of signal is still low, usually less than .05, but not much less than .05.

We next consider the probability of signal for data that comes from an out-of-control process. These power studies were performed by introducing step changes in the mean

vector, $\boldsymbol{\beta}$. Because of the small difference in the in-control results we expect only a slight improvement (if any) in the ability of detecting changes when using linear mixed models for balanced data. Because the probability of signal is not always .05 for the in-control data, the power studies were based on simulated control limits to ensure that the probability of a signal for in-control data will be the same for all the charts and close to the nominal .05 level.

Figure 2: Probability of signal for different values of n and ρ, for the $T^2_{2,i,LS}$, $T^2_{2,i,MIX}$, and $T^2_{varbi,i}$ charts where $m = 30$, $\sigma^2 = .1$, $\sigma^2_0 = .1$ and $\sigma^2_1 = .1$.



For the m profiles of data, the first l of them are generated from the in-control distribution with $\boldsymbol{\beta} = [0, 1]$ and the last $m-l$ are generated from the same distribution and same settings of the design factors, except that $\boldsymbol{\beta} = [\beta_0, 1]$. Thus we have introduced a step change in the intercept causing the last $m - l$ profiles to be shifted away from the first l profiles.

Figure 2 shows some of the results of the power studies. Here $m = 30$, $\sigma^2 = .1$, $\sigma_0^2 = .1$ and $\sigma_1^2 = .1$ where the step change occurred after the fifth profile. We do not show the results for the charts based on $T^2_{1,i,LS}$ and $T^2_{1,i,MIX}$ because it is known that these statistics will not perform well in detecting step changes (Sullivan and Woodall, 1996; Vargas, 2003).

The curves for $T^2_{2,i,MIX}$ and $T^2_{2,i,LS}$ practically coincide, indicating that the two methods will perform similarly in detecting the step change in the intercept. This is true regardless of the amount of correlation in the errors or the number of observations per profile. On the other hand $T^2_{varbi,i}$ performs poorly, with little ability to signal the shift. This is because the expression for $Var(\widehat{\mathbf{b}}_i)$ in (24) is only correct if all the data are in-control and come from the same distribution. When a step change is present the estimated value of $Var(\widehat{\mathbf{b}}_i)$ is inflated reducing the ability to detect that change.

Similar simulation results, not shown here, were obtained when the step change occurred at some other point, or when the shift occurred in the slope. Thus the conclusions stated here for the out-of-control data will hold for changes in the slope or intercept, no matter the value of l.

Similar results were obtained when the data are unequally spaced and are not presented here. From the simulation study results we conclude that when the data are balanced (equally or unequally spaced) as will often be the case for control charts applications, there appears to be no advantage in modeling correlation and/or including random effects.

Unbalanced Data

Because we found little difference using LS and LMM for balanced data, we considered unbalanced data. This consists of cases where $\mathbf{X}_i$ and $\mathbf{Z}_i$, although equal to each other for the same profile, are different from profile to profile. For simplicity, we kept $n_i = n$ for all

Table 2: Proportion of datasets with a signal for in-control data for unbalanced data situation.

m	n	ρ	$T^2_{1,i,LS}$	$T^2_{1,i,MIX}$	$T^2_{2,i,LS}$	$T^2_{2,i,MIX}$	$T^2_{varbi,i}$
30	5	0	0.3746	0.0907	0.2635	0.06797	0.0199
30	10	0	0.0672	0.0604	0.0517	0.0467	0.0214
30	15	0	0.0543	0.0551	0.0436	0.0426	0.0203
30	5	0.1	0.1452	0.0712	0.0931	0.0563	0.0166
30	10	0.1	0.0608	0.0556	0.0441	0.0414	0.0142
30	15	0.1	0.0525	0.0542	0.0412	0.0393	0.0163
30	5	0.5	0.0895	0.0590	0.0587	0.0492	0.0202
30	10	0.5	0.0535	0.0511	0.0395	0.0412	0.0176
30	15	0.5	0.0508	0.0517	0.0401	0.0405	0.0206
30	5	0.9	0.0559	0.0505	0.0431	0.0426	0.0208
30	10	0.9	0.0509	0.0498	0.0391	0.0392	0.0190
30	15	0.9	0.0507	0.0511	0.0415	0.0412	0.0211
60	5	0	0.3452	0.0880	0.2772	0.0702	0.0288
60	10	0	0.0912	0.0602	0.0739	0.0507	0.0302
60	15	0	0.0653	0.0596	0.0522	0.0456	0.0292
60	5	0.1	0.2541	0.0764	0.1977	0.0585	0.0236
60	10	0.1	0.0852	0.0595	0.0701	0.0471	0.0254
60	15	0.1	0.0591	0.0545	0.0447	0.0433	0.0286
60	5	0.5	0.1282	0.0641	0.0927	0.0462	0.0262
60	10	0.5	0.0576	0.0504	0.0487	0.0437	0.0298
60	15	0.5	0.0549	0.0559	0.0436	0.0439	0.0307
60	5	0.9	0.0551	0.0514	0.0433	0.0399	0.0303
60	10	0.9	0.0485	0.0488	0.0414	0.0398	0.0306
60	15	0.9	0.0507	0.0536	0.0420	0.0450	0.0294
90	5	0	0.3937	0.1070	0.3524	0.0850	0.0337
90	10	0	0.1119	0.0727	0.0963	0.0598	0.0360
90	15	0	0.0631	0.0594	0.0545	0.0532	0.0359
90	5	0.1	0.2796	0.0799	0.2428	0.0665	0.0295
90	10	0.1	0.0943	0.0642	0.0798	0.0544	0.0350
90	15	0.1	0.0562	0.0516	0.0464	0.0431	0.0270
90	5	0.5	0.1282	0.0647	0.1070	0.0536	0.0316
90	10	0.5	0.0687	0.0583	0.0553	0.0495	0.0378
90	15	0.5	0.0503	0.0468	0.0434	0.0396	0.0305
90	5	0.9	0.0568	0.0525	0.0508	0.0483	0.0366
90	10	0.9	0.0545	0.0550	0.0468	0.0464	0.0363
90	15	0.9	0.0471	0.0466	0.0402	0.0389	0.0322

the profiles. Similar to the balanced data study, we considered different combinations of m, n, and ρ where $\sigma^2 = .1$, $\sigma_0^2 = .1$ and $\sigma_1^2 = .1$. The setup is the same as for balanced data but now, the locations along the profile where data are collected were randomly generated within a fixed interval. Once the locations were generated, they were held fixed for each of the Monte Carlo repetitions.

Because the unbalanced data are unequally spaced, the AR(1) model is no longer reasonable because the correlation would be assumed to be equal between successive observations,

no matter their distance from each other. A more appropriate error structure for unequally spaced data is an exponential model (Schabenberger and Pierce, 2002) that takes into account the distance between measurements and still uses ρ. The exponential model is called the power model in $SAS^{®}$. In the case where the data are equally spaced, the exponential model reduces to the AR(1) model (Schabenberger and Pierce, 2002).

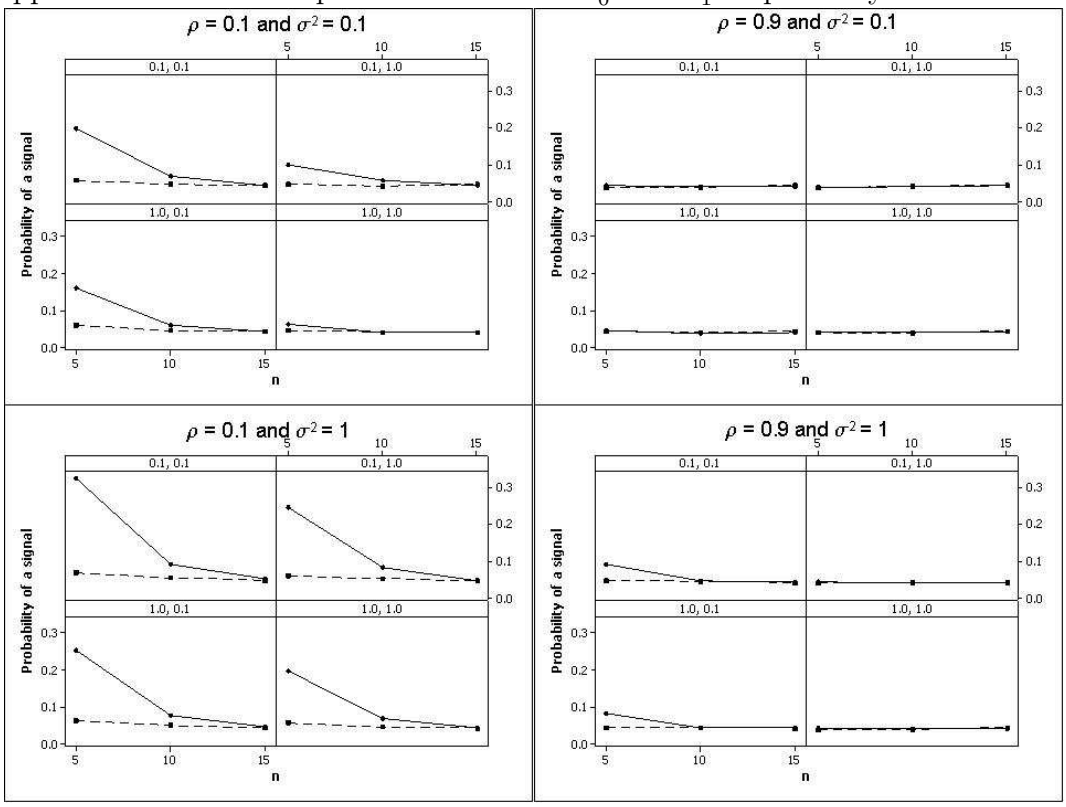
The $T^2_{1,i}$ and $T^2_{2,i}$ values are calculated from LMM and LS approaches and the control limit is obtained using the appropriate percentiles from the beta or chi-square distributions. Table 2 shows the proportion of the $10,000$ datasets that had a signal on the control charts for the various T^2 statistics.

Here we see that in some situations that using the LS approach will result in a much larger probability of signal than the nominal value. This increased probability occurs when there is a smaller number of observations per profile ($n = 5$) and increases as the number of profiles gets larger. It is also higher when the correlation in the errors is smaller. In contrast, using the LMM approach keeps the probability of a signal closer to its nominal .05 level. When the number of observations per profile is larger, there will be little difference between the LS and LMM approaches.

In contrast with the balanced data scenarios of the previous section, the difference in the LS and LMM approaches depends on the values of ρ, σ^2, σ^2_0 and σ^2_1. To see this, consider Figure 3 which shows the probability of signal for various combinations of ρ, σ^2, σ^2_0 and σ^2_1 for the $T^2_{2,i}$ statistic obtained via the LS (solid line) and LMM (the dashed line). Here $m = 60$, the horizontal axis is n, and there are four larger panels that show various combinations of ρ and σ^2. Within the larger panels are smaller panels which show the combinations of σ^2_0 and σ^2_1 respectively.

Here we see that the difference between the LS and LMM approaches is more drastic

Figure 3: Probability of signal for ρ, σ^2, σ_0^2 and σ_1^2 for the $T_{2,i}^2$ statistic where $m = 60$. The solid line represents the probability of signal for the LS approach and the dashed line for the LMM approach. The smaller panel variables are σ_0^2 and σ_1^2 respectively.



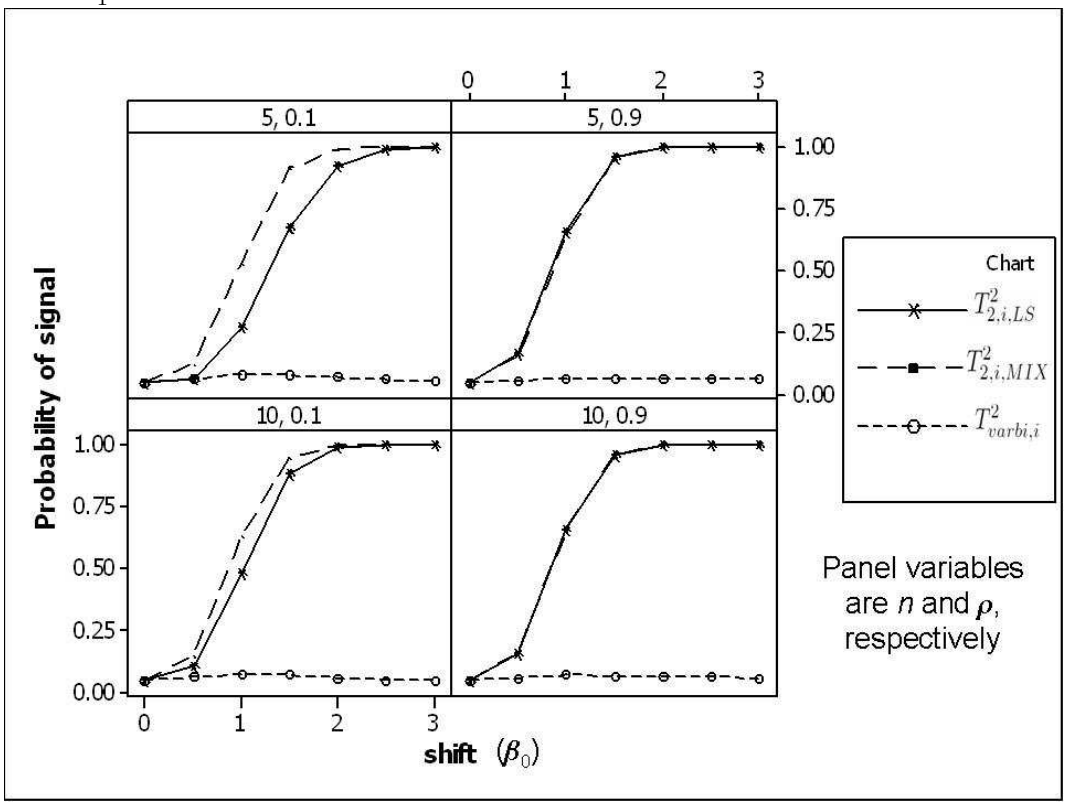
when the correlation is low, the variability in the errors is high, and $n = 5$. When $n = 10$ there will be slight differences in the LS and LMM approaches, at $n = 15$ there will be no difference regardless of the level or correlation or the variability in the random effects and errors. Thus the LMM approach preserves the appropriate Type I error probability of a signal even for an increased amount of error. Similar conclusions hold for other values of m and for $T_{1,i,MIX}^2$ and $T_{2,i,MIX}^2$.

In results not shown here, we have repeated multiple times the simulation study for the unbalanced data scenario with different sets of randomly generated $\mathbf{X}_i$ matrices. We found that the conclusions obtained from Table 2 and Figure 3 still hold with different sets of

randomly generated locations.

In order to do the power studies for this situation, we simulated the control limits to ensure that the charts will have the same probability of signal for in-control data. Figure 4 shows some of the results of the power studies for unbalanced data where $m = 30$, $\sigma^2 = .1$, $\sigma_0^2 = .1$ and $\sigma_1^2 = .1$ and where the step change occurred after the fifth profile. Again we do not show the results for the charts based on $T^2_{1,i,LS}$ and $T^2_{1,i,MIX}$.

Figure 4: Probability of signal for n and ρ, for the $T^2_{2,i,LS}$, $T^2_{2,i,MIX}$, and $T^2_{varbi,i}$ charts where the step change in the intercept, β_0, occurred after the 5^{th} profile and where $m = 30$, $\sigma^2 = .1$, $\sigma_0^2 = .1$ and $\sigma_1^2 = .1$.



We see that just as for in-control data the LMM approach will be superior for smaller amounts of correlation than for a larger level of correlation of the errors. The larger the number of observations per profile, the smaller the difference will be between the LMM

and LS approaches. The $T^2_{varbi,i}$ chart performs poorly just as it did for the balanced data scenario.

At first glance this results may seem counterintuitive. The two approaches perform similarly when the correlation in the errors is higher. Intuition suggests that since the LMM is taking into account the correlation of the errors, it would have higher probabilities of signal for correlated out-of-control data than the LS approach, which is the opposite of what our results show.

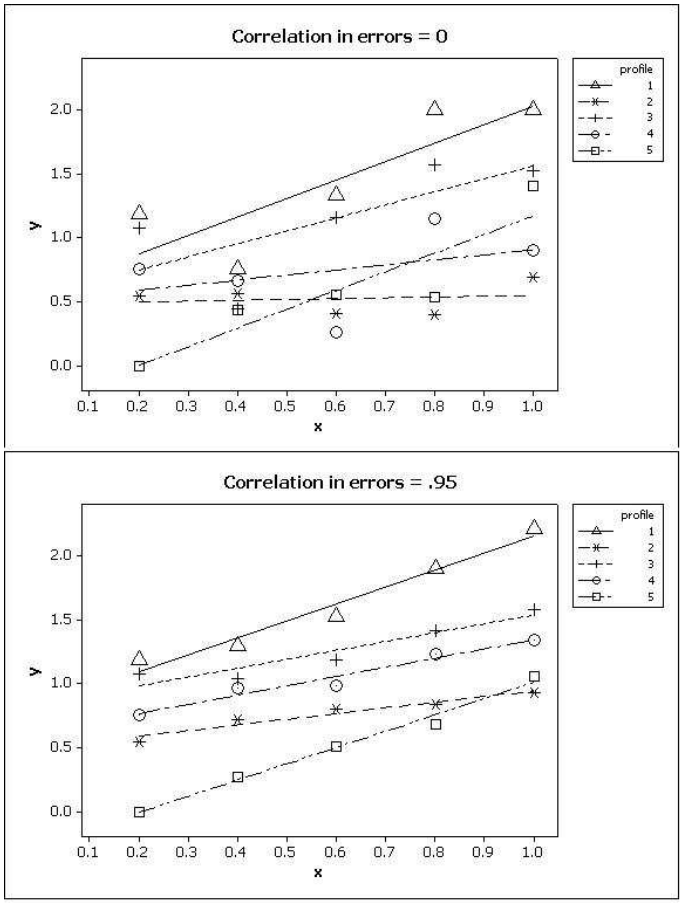
To explain the contradiction, note that an increased amount of correlation in the errors makes the profiles smoother and more similar to each other. To illustrate, consider the top panel of Figure 5 which shows 5 randomly generated profiles from a LMM with no correlation in the errors. The raw data points are shown along with the simple linear regression fits for each profile. The profiles each have 5 measurements with the fixed intercept = 0 and the fixed slope = 1. The random effects of intercept and slope both have variability = .1. Thus we have $\boldsymbol{\beta} = [0, 1]$, $\mathbf{D} = \begin{bmatrix} .1 & 0 \\ 0 & .1 \end{bmatrix}$, $\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & .2 \\ 1 & .4 \\ 1 & .6 \\ 1 & .8 \\ 1 & 1 \end{bmatrix}$, and $\mathbf{R}_i = \sigma^2 \mathbf{I}$ where $\sigma^2 = .1$.

Notice that the data points appear at random on either side of the fitted profile. This is because the independent errors are just as likely to cause a point to be above the line as below the line. Now consider the same previous scenario but now with correlated errors following an AR(1) structure. Thus

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix},$$

where ρ is some number between 0 and 1 and measures the strength of the correlation.

Figure 5: Randomly generated profile data along with simple linear regression fit with no correlation and correlation in the errors.



Contrast with the top panel the bottom panel of Figure 5 which shows the profiles with correlated errors where $\rho = .95$.

Because of the strong correlation, the errors tend to be similar to each other thus dampening the jagged effect of uncorrelated errors. As a result, the fitted profiles tend to appear more similar to each other when there is higher amounts of correlation in the errors. This decreases the frequency of occasions that the profiles are declared different from each other. As a result, both the LS and LMM approaches give a probability of signal close to the nominal value when the correlation is high.

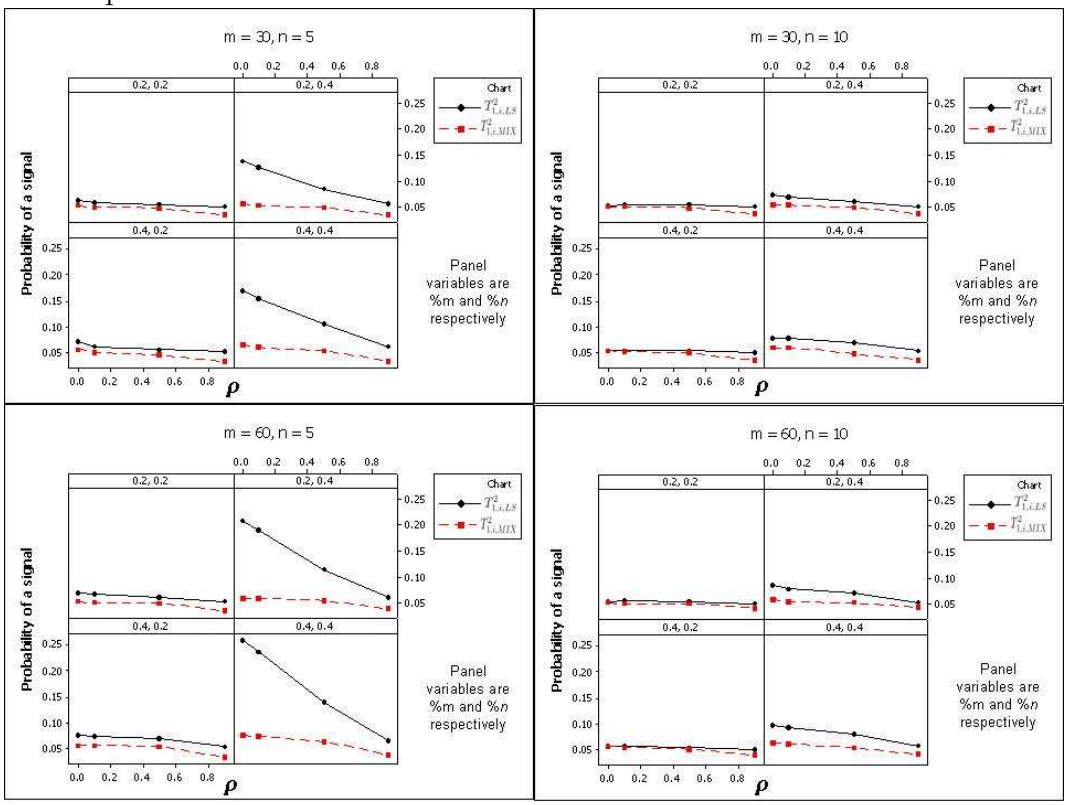
Thus when the data are unbalanced, a mixed model approach will be beneficial, particularly when the number of observations per profile and the amount of correlation are small.

Missing Data

The LMM approach has an advantage over the LS approach when there are missing data. Because the LMM approach pools information together from the profiles, it uses information from the profiles with full data to fit the profiles that have missing data points. The LMM approach can even be used to fit a curve to profiles that would not have enough data points for estimation of its own separate model. For example, profiles with only a single point could not be used when fitting separate simple linear regressions to each profile. But such profiles could be used in the LMM approach. While we do not advocate making a decision about whether or not a profile is outlying based on a single point, we do want to investigate the impact of missing data on the LMM and LS approaches.

We did a simulation study to evaluate the impact of missing data assumed to be missing at random (MAR). While this may be a simplistic assumption, it will serve here to illustrate the differences in the LMM and LS approaches. If the missing data are due to some underlying phenomena, for example, due to dropout in a longitudinal study then the MAR assumption will not be met. More information on the different types of missing data can be found in Verbeke and Molenberghs (2000, Chapters 14-16) and in Little and Rubin (1987). Likelihood based inference is still valid when the data are MAR so no changes are needed in the SAS coding (Vonesh and Chinchilli, 1997, 264). Studying the impact of missing data in profile monitoring requires consideration of both the proportion of profiles that have missing observations, referred to here as %m, and the proportion of observations missing within the profiles, referred to here as %n. For example, if $m = 30$, $n = 5$, $\%m = .2$, and $\%n = .4$,

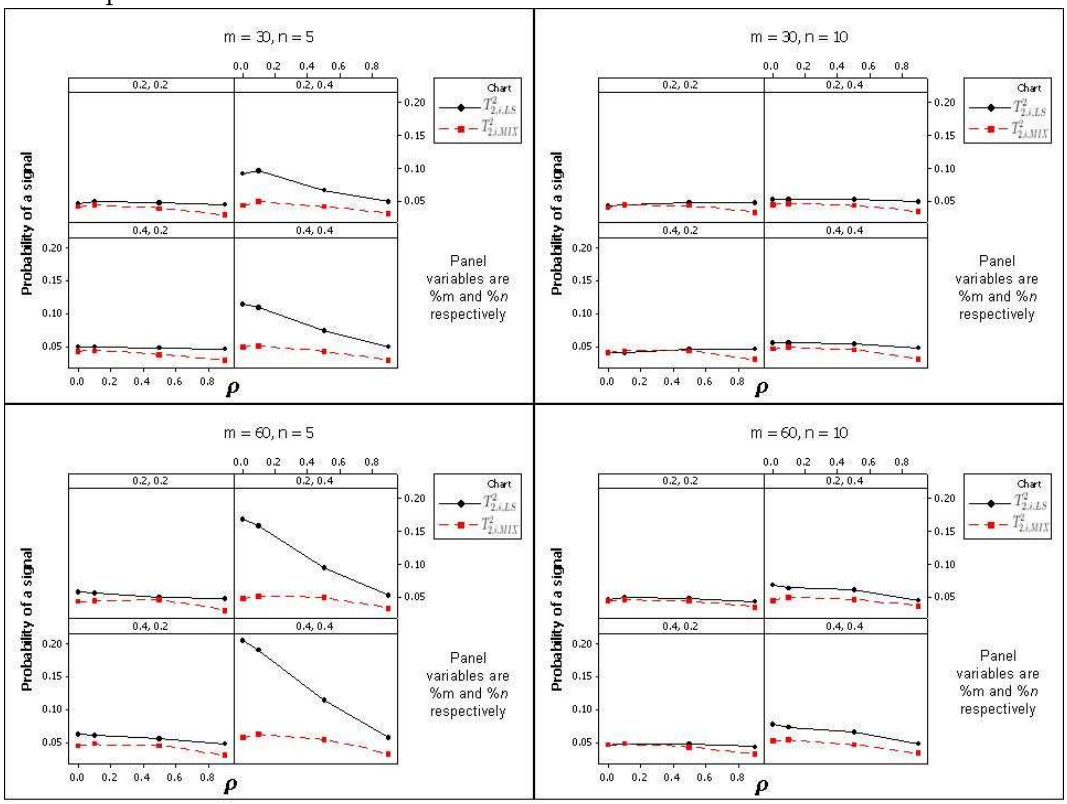
Figure 6: Probability of signal for $T^2_{1,i,LS}$ and $T^2_{1,i,MIX}$ when data are missing at random from the in-control profiles.



there will be 6 profiles that have missing observations and for those 6 profiles, there will be two observations missing for each profile.

We considered the balanced equally spaced data scenario where the data were first generated, then a subset of profiles was selected at random, and then the missing observations were selected at random locations for the subset of profiles. The missing observations occur at different points for the profiles. We show here the results for $\sigma_0^2 = .1$, $\sigma_1^2 = .1$, and $\sigma^2 = .1$ with the control limit obtained from the corresponding beta and chi-square distributions. We considered several values of ρ for the correlation in the errors and also included the case where $\rho = 0$, that is, there is no correlation in the errors. We examined two levels of %m and %n, which were .2 and .4.

Figure 7: Probability of signal for $T^2_{2,i,LS}$ and $T^2_{2,i,MIX}$ when data are missing at random from the in-control profiles.

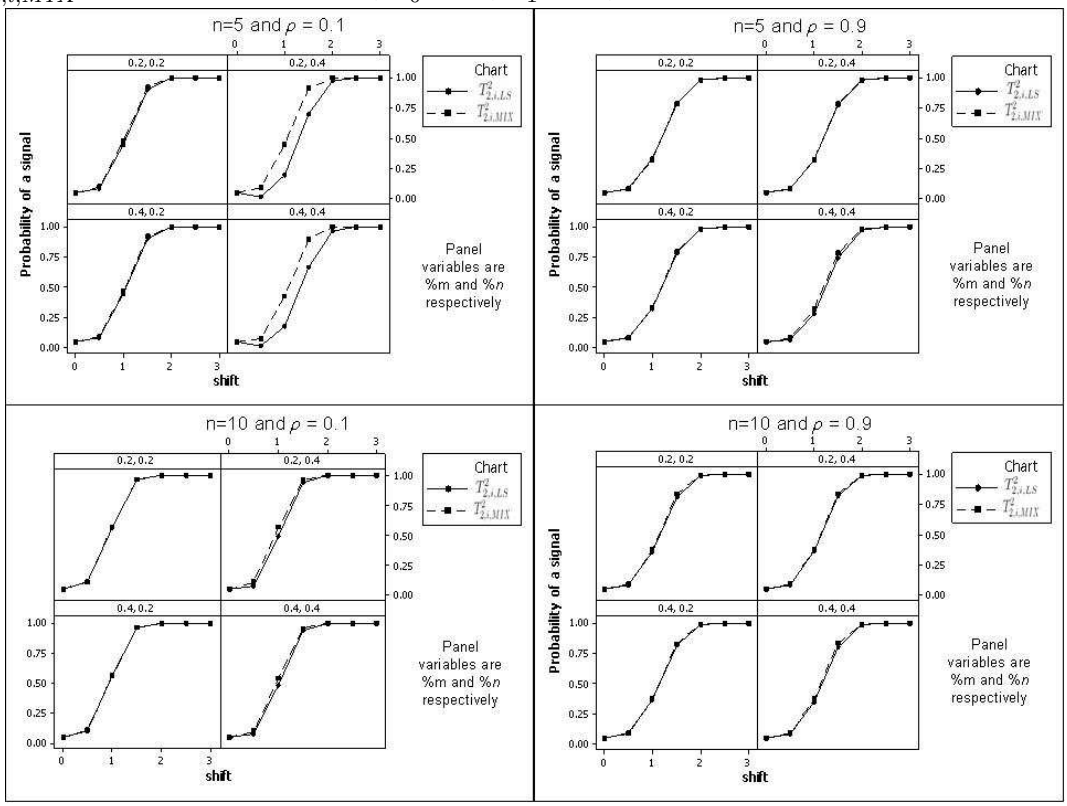


The probability of signal is shown in Figures 6 and 7 for data that comes from an in-control process. We see from the figures that in some cases the T^2 statistics based on the LS approach have a higher probability of signal than the nominal .05 level. The LS approach performs poorly when $n = 5$ and %n is larger. The larger the number of observations per profile then the less drastic will be the impact of missing data. A change in %m only results in small differences between the LMM and LS approaches. As the number of profiles increases the probability of a signal (frequency of a false alarm) increases.

As we did for other data scenarios, we considered the performance for out-of-control data. As before, we introduce a step change in the intercept and compare the probability of a signal for $T^2_{2,i,LS}$ and $T^2_{2,i,MIX}$ because of their ability to detect step changes. Figure 8 shows the

probability of a signal where $m = 60$, $\sigma_0^2 = .1$, $\sigma_1^2 = .1$, and $\sigma^2 = .1$ for different combinations of ρ, n, $\%m$ and $\%n$.

Figure 8: Probability of signal for n and ρ where data are missing at random, for the $T^2_{2,i,LS}$ and $T^2_{2,i,MIX}$ charts where $m = 60$, $\sigma_0^2 = .1$, $\sigma_1^2 = .1$, and $\sigma^2 = .1$.

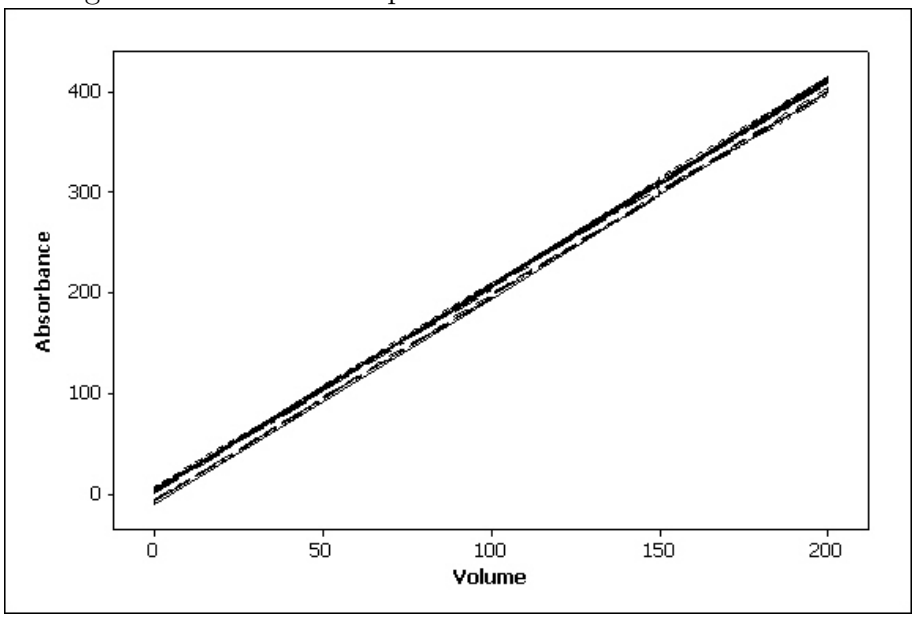


Just as we saw for the in-control data, the biggest difference between the LMM and LS approaches for MAR data occurs when the correlation in the errors is low, the number of observations is small, and when the proportion of missing data is larger. Earlier we noted that for unbalanced data that the difference between the LMM and LS approaches is more drastic when there is an increasing amount of variability in the errors. This is true here. While not shown in Figure 8, when σ^2 increases, the LMM approach will increases its superiority over the LS approach.

Example

To illustrate the control chart procedure discussed here we use the calibration dataset first analyzed in Mestek, Pavlik, and Suchánek (1994) and later analyzed in Mahmoud and Woodall (2004). The data consist of 22 calibration curves each of which relates an absorbance measure of a chemical solution to the volume at which the solution was prepared. The purpose is to determine if the calibration curves are stable over time. There are 5 volumes and 2 replicate measurements for each volume so each profile has a total of 10 measurements. The raw data profiles of the calibration data are shown in Figure 9. We see that the calibration curves are very similar to each other with more variability in the intercepts than in the slopes. These data are balanced, equally spaced, and have no missing observations.

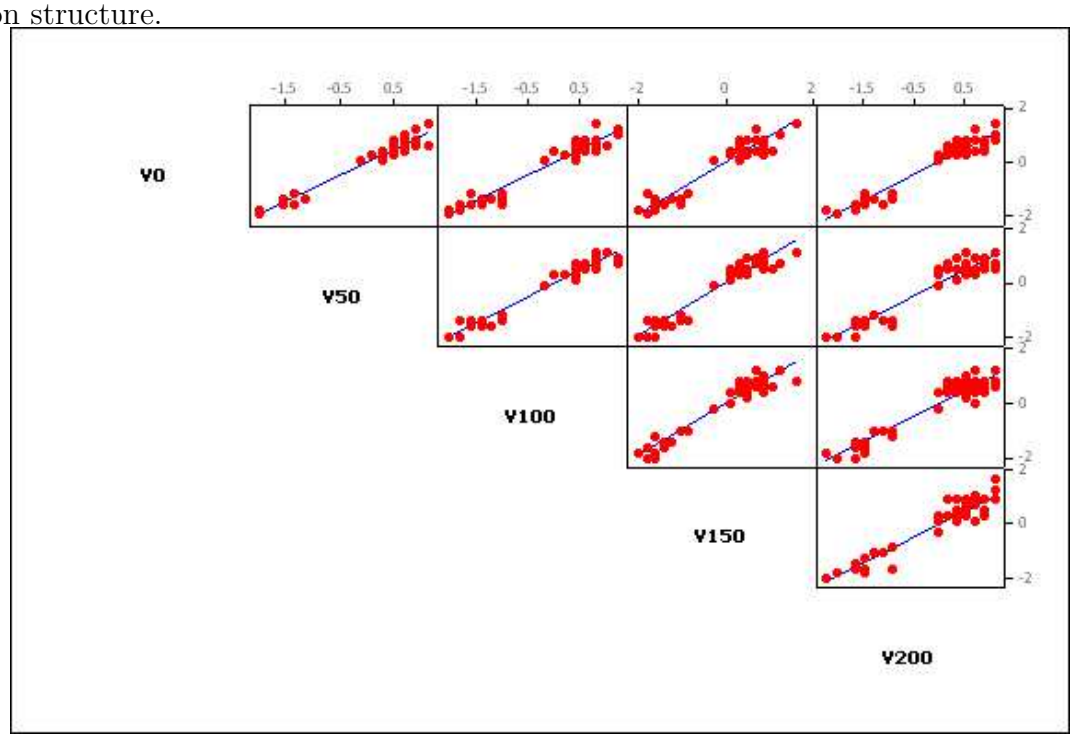
Figure 9: The raw data profiles for the calibration dataset.



We first investigate to see if the measurements within a profile are correlated across the different volumes. To determine the appropriate correlation structure we employed the

graphical methods discussed in Dawson, Gennings, and Carter (1997). After centering and scaling the data by volume we obtained the draftman's display shown in Figure 10.

Figure 10: Draftman's display of the calibration data showing a compound symmetry correlation structure.



Based on the examples shown in Dawson, Gennings, and Carter (1997), we conclude that the calibration dataset has a compound symmetry (CS) structure. This is evident from the positive linear trend in the individual scatterplots on the draftman's display of Figure 10. The strength of the trend is consistent for the plots closest to the diagonal and for the plots in the upper right hand corner. If the strength of the trend were to weaken or decrease for the plots further away from the diagonal, then we would have concluded that the calibration dataset had errors that followed a AR structure.

We then fit the profiles with the LS and LMM approaches. Figure 11 shows the T^2 charts for the LS approach. Figure 12 shows the T^2 charts for the LMM approach. When fitting

Figure 11: T^2 charts for the profiles of the calibration dataset that have been fit by the LS approach.

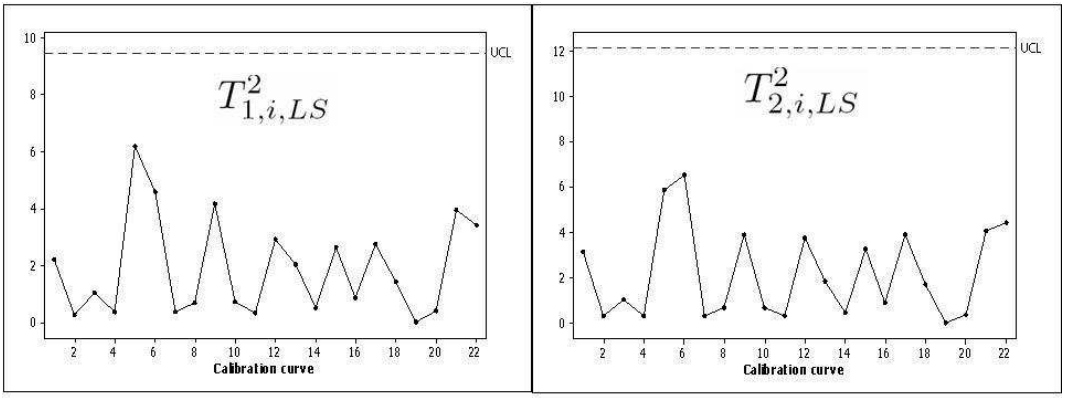
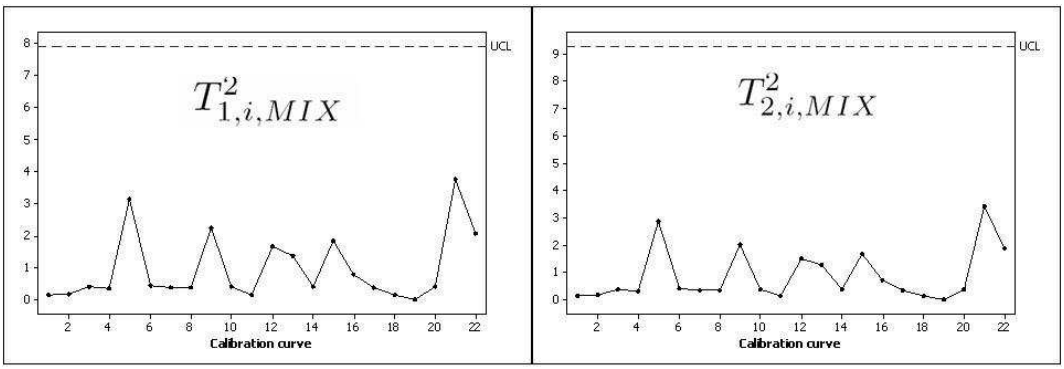


Figure 12: T^2 charts for the profiles of the calibration dataset that have been fit by the LMM approach.



the LMM, we found that the only necessary random effect was that for the intercept. A likelihood ratio test of a random effect on the slope leads us to conclude that a random effect is not needed for the slope because the slopes are so similar to each other. In addition, once the random effect for intercept is included, the estimated errors are no longer correlated with each other, thus we can safely analyze the data with independent errors.

We see that for this dataset, the LS and LMM approaches give the same result. There are no signals on either control charts suggesting that the 22 calibration curves come from an in-control process. All of them can be used to set the control limits for Phase II. We did

not expect to have any drastic differences in the LS and LMM approaches because of the results of our simulation studies showed that there is little difference for balanced, equally spaced data as we have here for the calibration data.

Our conclusions here agree with the results of Staudhammer et al. (2005) who studied the modeling of profiles in lumber manufacturing. They modeled the autocorrelated errors with time series models and concluded that for SPC applications, ignoring the autocorrelation does not make much difference. This is true even though their profiles are rather complicated and there is an obvious autocorrelation in the data. The number of observations that they have per profile is much larger than the sample sizes considered here ($n > 2000$).

Open Questions

In our simulation studies and comparisons we restricted the investigation to certain types of situations and data scenarios. There is a large number of variations that could be considered for the simulations shown here. We discuss briefly some of these variations here.

We have not considered the situation where the data are correlated with no random effects where $\mathbf{y}_i \sim MN(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{R}_i)$ and $\mathbf{R}_i$ follows some structure. Nor have we considered models where not all the coefficients are random, where the $\mathbf{Z}_i$ matrix is a subset of the $\mathbf{X}_i$ matrix. We believe that our conclusions obtained here will be similar to those that would hold where not all of the effects are random. We have not considered the many other error structures that could be utilized that depend on a small number of parameters, for example, the CS structure. We believe that regardless of the error structure used, the conclusions obtained here would still hold. The power studies on the out-of-control performance of the T^2 statistic have been performed for step changes in the mean vector where the $T^2_{2,i}$ statistic is superior. We have not considered power studies for outliers (at the profile level and within

the profile level). It seems clear that the high breakdown estimation methods discussed in Vargas (2003) and Jensen, Birch, and Woodall (2006) will be of benefit for this situation. Our studies on missing data assume that data are missing at random. However, for repeated measures data, it is sometimes reasonable to presume that the missing observations depend on a variable that is not observed. For example, if the profiles were to represent human subjects who are measured at repeated time intervals, dropout can occur because subjects are not longer interested in participating. In some cases, the dropout can be the result of an ineffective treatment and the missing observations no longer occurs at random. It is much more difficult to model data that with values that are not missing at random.

Conclusions

To summarize the results, we have found that in all the Phase I scenarios investigated, the LMM approach has either equivalent or superior performance when compared to the LS approach. When the data are balanced, there is little difference between the two approaches but we have found using simulation that the advantage of the LMM over the LS approach is greatest when the data are unbalanced or when there are missing observations. For unbalanced or missing data, the LMM is better for smaller levels of correlation and for a smaller number of observations per profile.

Appendix

Theorem A.1. The sum of the random deviations from the linear mixed model, $\widehat{\mathbf{b}}_i$, is equal to the zero vector when $\mathbf{X}_i = \mathbf{Z}_i$, for $i = 1, 2, \ldots, m$.

Proof. We first set $\mathbf{X}_i = \mathbf{Z}_i$ and noting that we can interchange between the model formulation in terms of the individual profiles and a model formulation using stacked matrices. For example, we have $\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i \mathbf{X}_i = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ and $\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i \mathbf{y}_i = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ where $\mathbf{X}$ is a $(\sum_{i=1}^{m} n_i)$ by p stacked matrix of the $\mathbf{X}_i's$, $\mathbf{V} = \mathbf{Z}\mathbf{B}\mathbf{Z}' + \mathbf{R} = diag(\mathbf{V}_i)$ with $\mathbf{B} = diag(\mathbf{D})$,$\mathbf{R} = diag(\mathbf{R}_i)$, and $\mathbf{Z}$ is a block diagonal matrix containing all the $\mathbf{Z}_i$ matrices. Then using (6) and (7) with the estimated values in place of the known values and doing some manipulation gives

$$\begin{aligned}
\sum_{i=1}^{m} \widehat{\mathbf{b}}_i &= \sum_{i=1}^{m} \left[\mathbf{D}\mathbf{Z}_i' \mathbf{V}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{MIX} \right) \right] \\
&= \sum_{i=1}^{m} \mathbf{D}\mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i - \sum_{i=1}^{m} \mathbf{D}\mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{X}_i \left(\mathbf{X}'\mathbf{V}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= \mathbf{D}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{D}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right) \left(\mathbf{X}'\mathbf{V}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= \mathbf{D}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{D}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= \mathbf{0}. \qquad (25)
\end{aligned}$$

Notice that this proof does not require that $\mathbf{X}_i$ be the same for each profile. Nor does it require that the profiles have the same number of measurements (i.e. n_i does not have to be the same for all the profiles). As a result it is obvious that the average of the random deviations is zero

$$\overline{\mathbf{b}} = \frac{\sum_{i=1}^{m} \widehat{\mathbf{b}}_i}{m} = \mathbf{0}, \qquad (26)$$

and that

$$\overline{\boldsymbol{\beta}}_{MIX} = \frac{\sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_{i,MIX}}{m} = \frac{\sum_{i=1}^{m} (\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i)}{m} = \frac{\sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_{MIX}}{m} = \frac{m\widehat{\boldsymbol{\beta}}_{MIX}}{m} = \widehat{\boldsymbol{\beta}}_{MIX}. \tag{27}$$

In Theorem A.2, we extend the results of Theorem A.1 to the case where the $\mathbf{Z}_i$ matrix is not equivalent to the $\mathbf{X}_i$ matrix but that the columns of $\mathbf{Z}_i$ are contained in $\mathbf{X}_i$, thus $\mathbf{X}_i$ may also have some additional columns not contained in $\mathbf{Z}_i$.

Theorem A.2. The sum of the random deviations from the mixed model, $\widehat{\mathbf{b}}_i$, is equal to the zero vector when the $\mathbf{Z}_i$ matrix is contained within the $\mathbf{X}_i$ matrix.

Proof. We set the partitioned matrix $\mathbf{X}_i = [\mathbf{X}_{1,i}|\mathbf{X}_{2,i}] = [\mathbf{Z}_i|\mathbf{X}_{2,i}]$ where $\mathbf{Z}_i = \mathbf{X}_{1,i}$ and then using (6) and (7) with the estimated values in place of the known values and doing some manipulation gives

$$\begin{aligned}
\sum_{i=1}^{m} \widehat{\mathbf{b}}_i &= \sum_{i=1}^{m} \left[\mathbf{D}\mathbf{Z}_i' \mathbf{V}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{MIX} \right) \right] \\
&= \mathbf{D} \left[\sum_{i=1}^{m} \mathbf{Z}_i' \mathbf{V}_i^{-1} \mathbf{y}_i - \sum_{i=1}^{m} \mathbf{Z}_i' \mathbf{V}_i^{-1} \left[\mathbf{X}_{1,i} | \mathbf{X}_{2,i} \right] \widehat{\boldsymbol{\beta}}_{MIX} \right] \\
&= \mathbf{D} \left[\sum_{i=1}^{m} \mathbf{X}_{1,i}' \mathbf{V}_i^{-1} \mathbf{y}_i - \sum_{i=1}^{m} \mathbf{X}_{1,i}' \mathbf{V}_i^{-1} \left[\mathbf{X}_{1,i} | \mathbf{X}_{2,i} \right] \widehat{\boldsymbol{\beta}}_{MIX} \right] \\
&= \mathbf{D} \left[\sum_{i=1}^{m} \mathbf{X}_{1,i}' \mathbf{V}_i^{-1} \mathbf{y}_i - \sum_{i=1}^{m} \left[\mathbf{X}_{1,i}' \mathbf{V}_i^{-1} \mathbf{X}_{2,i} | \mathbf{X}_{1,i}' \mathbf{V}_i^{-1} \mathbf{X}_{2,i} \right] \widehat{\boldsymbol{\beta}}_{MIX} \right] \\
&= \mathbf{D} \left[\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{y} - \left[\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1 | \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_2 \right] \left(\mathbf{X}' \mathbf{V} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \right],
\end{aligned} \tag{28}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are stacked matrices of $\mathbf{X}_{1,i}$ and $\mathbf{X}_{2,i}$ respectively.

We now turn attention to the partitioned matrix in the right hand side of the expression

and show that

$$\begin{aligned}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1} &= \mathbf{A}^{-1} \\ &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_2 \end{bmatrix}^{-1}. \end{aligned} \tag{29}$$

Then using a common result (2.50 of Rencher, 2000) to take the inverse of the partitioned matrix we have

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1} + (\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1} & -(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1} & \mathbf{B}^{-1} \end{bmatrix},$$

where $\mathbf{B} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = (\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_2) - \mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2$ has an inverse. This inverse of the partitioned matrix also requires the existence of inverses of $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$. These inverses exist because $\mathbf{A}$ has an inverse and by Theorem 2.6F of Rencher (2000) which says the square submatrices of a partitioned matrix have inverses if the whole matrix itself has an inverse. Thus the right hand side of (28) can be written as

$$\left[\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1 | \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\right](\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} =$$

$$\begin{bmatrix} \mathbf{I} + \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1} & -\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1} \\ -(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2)\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1} & (\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2)\mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{X}_2'\mathbf{V}^{-1}\mathbf{y} \end{bmatrix}. \tag{30}$$

Now multiplying this last result of (30) into the earlier expression for the blups in (28) gives

us

$$\begin{aligned}
\sum_{i=1}^{m}\widehat{\mathbf{b}}_i &= \mathbf{D}\left[\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} - \left[\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1|\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\right](\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right] \\
&= \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} - \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} + \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} \\
&\quad -\mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{y} - \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} \\
&\quad +\mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_2\mathbf{B}^{-1}\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{y} \\
&= \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} - \mathbf{D}\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} \\
&= \mathbf{0}.
\end{aligned}$$

Thus the eblups sum to zero for balanced and unbalanced data and/or whether or not $\mathbf{X}_i$ is the same for each profile as long as $\mathbf{Z}_i$ is equal to or contained within $\mathbf{X}_i$.

Theorem A.3. If $\sum_{i=1}^{m}\widehat{\mathbf{b}}_i = \mathbf{0}$, then the $T^2_{1,i,MIX}$ and $T^2_{2,i,MIX}$ statistics in (11) and (15) depend only on the $\widehat{\mathbf{b}}_i$, the random components.

Proof. Starting with (11) and using the results from (21) we have

$$\begin{aligned}
T^2_{1,i,MIX} &= (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})' S^{-1}_{1,MIX}(\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX}) \text{ for } i = 1, 2, \ldots, m \\
&= \left[(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - (\widehat{\boldsymbol{\beta}}_{MIX})\right]' \\
&\quad \left[\frac{\sum_{i=1}^{m}[(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - \widehat{\boldsymbol{\beta}}_{MIX}][(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - \widehat{\boldsymbol{\beta}}_{MIX}]'}{m-1}\right]^{-1} \\
&\quad \left[(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - (\widehat{\boldsymbol{\beta}}_{MIX})\right] \quad (\text{because } \widehat{\boldsymbol{\beta}}_{MIX} = \overline{\boldsymbol{\beta}}_{MIX}) \\
&= \widehat{\mathbf{b}}_i'\left(\frac{\sum_{i=1}^{m}\widehat{\mathbf{b}}_i\widehat{\mathbf{b}}_i'}{m-1}\right)^{-1}\widehat{\mathbf{b}}_i \text{ for } i = 1, 2, \ldots, m. \qquad (31)
\end{aligned}$$

Starting with (15) we have

$$\begin{aligned}
T^2_{2,i,MIX} &= (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX})' S^{-1}_{2,MIX} (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{MIX}) \text{ for } i = 1, 2, \ldots, m \\
&= \left[(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - (\widehat{\boldsymbol{\beta}}_{MIX}) \right]' \\
&\quad \left[\frac{\sum_{i=1}^{m} [(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_{i+1}) - (\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i)][(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_{i+1}) - (\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i)]'}{2(m-1)} \right]^{-1} \\
&\quad \left[(\widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i) - (\widehat{\boldsymbol{\beta}}_{MIX}) \right] \\
&= \widehat{\mathbf{b}}_i' \left[\frac{\sum_{i=1}^{m} (\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)(\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)'}{2(m-1)} \right]^{-1} \widehat{\mathbf{b}}_i \text{ for } i = 1, 2, \ldots, m. \qquad (32)
\end{aligned}$$

Theorem A.4. The $T^2_{1,i,MIX}$ and $T^2_{2,i,MIX}$ statistics in (11) and (15) depend only on the eblups, $\widehat{\mathbf{b}}_i$, and their average, $\overline{\mathbf{b}}$, no matter the value of $\mathbf{Z}_i$ and $\mathbf{X}_i$.

Proof. In the most general situation, we have 3 components that make up the $T^2_{1,i,MIX}$ and $T^2_{2,i,MIX}$ statistics. The first component consisting of both random and fixed effects, comprises the columns of $\mathbf{Z}_i$ and $\mathbf{X}_i$ that are equal to each other, the second component consisting of only fixed effects, comprises the columns in $\mathbf{X}_i$ that are not in $\mathbf{Z}_i$, and the third component consisting of only random effects, comprises the columns in $\mathbf{Z}_i$ that are not in $\mathbf{X}_i$.

We note that once $\widehat{\boldsymbol{\beta}}_{MIX}$ and $\widehat{\mathbf{b}}_i$ are obtained in the most general situation that we can partition $\widehat{\boldsymbol{\beta}}_{i,MIX}$ into the 3 components. Thus we have $\widehat{\boldsymbol{\beta}}_{i,MIX} = \widehat{\boldsymbol{\beta}}_{MIX} + \widehat{\mathbf{b}}_i = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{MIX,1} + \widehat{\mathbf{b}}_{i,1} \\ \widehat{\mathbf{b}}_{i,2} \\ \widehat{\boldsymbol{\beta}}_{MIX,2} \end{bmatrix}$ where $\widehat{\boldsymbol{\beta}}_{MIX,1}$ and $\widehat{\mathbf{b}}_{i,1}$ are $ax1$ vectors corresponding to the columns in $\mathbf{X}_i$ and $\mathbf{Z}_i$ that are equivalent, $\widehat{\mathbf{b}}_{i,2}$ is the $cx1$ vector corresponding to the columns in $\mathbf{Z}_i$ that are not in $\mathbf{X}_i$, and $\widehat{\boldsymbol{\beta}}_{MIX,2}$ is a $bx1$ vector corresponding to the columns in $\mathbf{X}_i$ that are not in $\mathbf{Z}_i$. Thus $\mathbf{X}_i$ is a

n_i by $(a+b)$ matrix and $\mathbf{Z}_i$ is a n_i by $(a+c)$ matrix. We also have

$$\begin{aligned}
\overline{\boldsymbol{\beta}}_{i,MIX} &= \frac{\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}_{i,MIX}}{m} \\
&= \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{MIX,1} + \frac{\sum_{i=1}^{m}\widehat{\mathbf{b}}_{i,1}}{m} \\ \frac{\sum_{i=1}^{m}\widehat{\mathbf{b}}_{i,2}}{m} \\ \widehat{\boldsymbol{\beta}}_{MIX,2} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{MIX,1} + \overline{\mathbf{b}}_{i,1} \\ \overline{\mathbf{b}}_{i,2} \\ \widehat{\boldsymbol{\beta}}_{MIX,2} \end{bmatrix} \text{ for } i = 1, 2, \ldots, m, \qquad (33)
\end{aligned}$$

and

$$\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{i,MIX} = \begin{bmatrix} \widehat{\mathbf{b}}_{i,1} + \overline{\mathbf{b}}_{i,1} \\ \widehat{\mathbf{b}}_{i,2} - \overline{\mathbf{b}}_{i,2} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{b}}_i - \overline{\mathbf{b}} \\ 0 \end{bmatrix} \text{ for } i = 1, 2, \ldots, m. \qquad (34)$$

The resulting $T^2_{1,i,MIX}$ statistics is then given by

$$\begin{aligned}
T^2_{1,i,MIX} &= (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{i,MIX})' S^{-1}_{1,MIX} (\widehat{\boldsymbol{\beta}}_{i,MIX} - \overline{\boldsymbol{\beta}}_{i,MIX}) \text{ for } i = 1, 2, \ldots, m \\
&= \begin{bmatrix} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' & 0 \end{bmatrix} \frac{\sum_{i=1}^{m} \begin{bmatrix} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' & 0 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{b}}_i - \overline{\mathbf{b}} \\ 0 \end{bmatrix}}{m-1} \begin{bmatrix} \widehat{\mathbf{b}}_i - \overline{\mathbf{b}} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' & 0 \end{bmatrix} \begin{bmatrix} \frac{\sum_{i=1}^{m}(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})'(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})}{m-1} & 0 \\ 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{b}}_i - \overline{\mathbf{b}} \\ 0 \end{bmatrix}. \qquad (35)
\end{aligned}$$

The previous expression contains a nonsingular matrix because of the column of zeroes. However, because the matrix is block diagonal, it is a simple matter to calculate the generalized inverse by simply taking the inverse of the non-zero portion of the matrix. We can then rewrite the above expression as as

$$\begin{aligned}
T^2_{1,i,MIX} &= \begin{bmatrix} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' & 0 \end{bmatrix} \begin{bmatrix} \left(\frac{\sum_{i=1}^{m}(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})'(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})}{m-1}\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{b}}_i - \overline{\mathbf{b}} \\ 0 \end{bmatrix} \\
&= (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m}(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})'(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})}{m-1}\right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{ for } i = 1, 2, \ldots, m. \qquad (36)
\end{aligned}$$

By similar arguments we can show that

$$\widehat{\boldsymbol{\beta}}_{i+1,MIX} - \widehat{\boldsymbol{\beta}}_{i,MIX} = \begin{bmatrix} \widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i \\ 0 \end{bmatrix}, \tag{37}$$

and

$$T^2_{2,i,MIX} = (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m-1} (\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)'(\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)}{2(m-1)} \right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{for } i = 1, 2, \ldots, m, \tag{38}$$

which is similar to the same result obtained for $T^2_{2,i,MIX}$ in Theorem A.3.

The results of Theorem A.3 and A.4 imply that for a given set of data, we only need to consider the random components when calculating the $T^2_{1,i,MIX}$ and $T^2_{2,i,MIX}$ statistics. This simplifies the calculations needed to determine the properties of multivariate control charts.

Theorem A.5. The estimator of fixed effects and the predictor of the random effects in the LMM are regression equivariant.

Proof. Following the convention of Rousseeuw and Leroy (1987), an estimator, T is regression equivariant, if for the response vector, $\mathbf{y}_i$, the regressors, $\mathbf{x}_i$, and an arbitrary vector, $\mathbf{v}$,

$$T(\mathbf{x}_i; \mathbf{y}_i + \mathbf{x}_i'\mathbf{v}) = T(\mathbf{x}_i; \mathbf{y}_i) + \mathbf{v} \ \forall i. \tag{39}$$

Regression equivariance is important because it allows us to arbitrarily pick parameter values to generate $\mathbf{y}_i$ for simulation studies without loss of generality. For example in a simple linear regression model, the least squares estimators are regression equivariant, which would allow us to pick an arbitrary slope and intercept for simulation studies, and still obtain the same conclusions from the study.

We show here the regression equivariance of the estimator of fixed effects and of the predictor of random effects for the linear mixed model. We use the stacked matrices of Theorem A.1 for the the estimator of the fixed effects and the predictor of random effects.

Suppose the response vector $\mathbf{y}$ was changed by some arbitrary amount that is a function of the regressors, $\mathbf{X}$. The resulting response vector is given by $\widetilde{\mathbf{y}} = \mathbf{y} + \mathbf{X}\mathbf{v}$ where $\mathbf{v}$ is some arbitrary $px1$ vector.

Then the resulting estimator of the fixed effects is given by

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}\mathbf{V}^{-1}(\mathbf{y} + \mathbf{X}\mathbf{v}) \\ &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}\mathbf{V}^{-1}\mathbf{y} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}\mathbf{V}^{-1}\mathbf{X}\mathbf{v} \\ &= \boldsymbol{\beta} + \mathbf{v} \end{aligned} \tag{40}$$

As a result of the definition in (39) the estimator of the fixed effects is regression equivariant.

Now consider the predictor of the random effects. If the response vector is changed by some arbitrary amount so that we have $\widetilde{\mathbf{y}}$ then we can show that the resulting predictor of random effects is given by

$$\begin{aligned} \widetilde{\mathbf{b}} &= \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\widetilde{\mathbf{y}} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \\ &= \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} + \mathbf{X}\mathbf{v} - \mathbf{X}(\boldsymbol{\beta} + \mathbf{v})) \\ &= \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \widehat{\mathbf{b}} \end{aligned} \tag{41}$$

Thus the predicted random effects are unchanged when the response vector has been changed by some arbitrary amount.

REFERENCES

Banerjee, M. and Frees, E.W. (1997). "Influence Diagnostics for Linear Longitudinal Models". *Journal of the American Statistical Association* 92, pp. 999-1005.

Chi, E.M. and Reinsel, G.C. (1989). "Models for Longitudinal Data with Random Effects and AR(1) Errors". *Journal of the American Statistical Association* 84, pp. 452-459.

Cook, R.D. (1986). "Assessment of Local Influence (with discussion)". *Journal of the Royal Statistical Society, Series B* 48, pp. 133-169.

Dawson, K.S., Gennings, C. and Carter, W.H. (1997). "Two Graphical Techniques Useful in Detecting Correlation Structure in Repeated Measures Data". *The American Statistician* 51, pp. 275-283.

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

Demidenko, E. and Stukel, T.A. (2005). "Influence Analysis for Linear Mixed-Effects Models". *Statistics in Medicine* 24, pp. 893-909.

Ghosh, M. and Rao, J.N.K. (1994). "Small Area Estimation: An Appraisal". *Statistical Science* 9, pp. 55-93.

Hartford, A. and Davidian, M. (2000). "Consequences of Misspecifying Assumptions in Nonlinear Mixed Effects Models". *Computational Statistics and Data Analysis* 34, pp. 139-164.

Harville, D.A. (1976). "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects". *The Annals of Statistics* 4, pp. 384-395.

Haslett, J. and Dillane, D. (2004). "Application of 'delete=replace' to Deletion Diagnostics for Variance Component Estimation in the Linear Mixed Model". *Journal of the Royal*

Statistical Society, Series B 66, pp. 131-143.

Holmes, D.S. and Mergen, A.E (1993). "Improving the Performance of the T^2 Control Chart". *Quality Engineering* 5, p. 619-625.

Houseman, E.A., Ryan, L.M., and Coull, B.A. (2004). "Cholesky Residuals for Assessing Normal Errors in a Linear Model with Correlated Outcomes" . *Journal of the American Statistical Association* 99, pp. 383-394.

Jensen, W.A., Birch, J.B., and Woodall, W.H. (2006). "High Breakdown Estimation Methods for Phase I Multivariate Control Charts". Submitted to *Journal of Quality Technology.*

Kang, L., and Albin, S. L. (2000). "On-line Monitoring When the Process Yields a Linear Profile". *Journal of Quality Technology* 32, pp. 418-426.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003). "On the Monitoring of Linear Profiles". *Journal of Quality Technology* 35, pp. 317-328 .

Laird, N.M. and Ware, J.H. (1982). "Random Effects Models for Longitudinal Data". *Biometrics* 38, pp. 963-974.

Langford, I.H. and Lewis, T. (1998). "Outliers in Multilevel Data". *Journal of the Royal Statistical Society, Series A* 161, pp. 121-160.

Lesaffre, E., Asefa, M. and Verbeke, G. (1999). "Assessing the Goodness-of-fit of the Laird and Ware Model - An Example: The Jimma Infant Survival Differential Longitudinal Study". *Statistics in Medicine* 18, pp. 835-854.

Lesaffre, E. and Verbeke, G. (1998). "Local Influence in Linear Mixed Models". *Biometrics* 54, pp. 570-582.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for*

Mixed Models, SAS Institute Inc. Cary, NC.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons, New York, NY.

Longford, N.T. (2001). "Simulation-Based Diagnostics in Random-Coefficient Models". *Journal of the Royal Statistical Society, Series A* 164, pp. 259-273.

Mahmoud, M. A., and Woodall, W. H. (2004). "Phase I Analysis of Linear Profiles with Calibration Applications". *Technometrics* 46, pp. 380-391.

Mestek, O., Pavlik, J. and Suchánek, M. (1994). "Multivariate Control Charts: Control Charts for Calibration Curves" *Fresenius' Journal of Analytical Chemistry* 350, pp. 344-351.

Rencher, A.C. (2000). *Linear Models in Statistics*. John Wiley and Sons, New York, NY.

Ritz, C. (2004). "Goodness-of-fit Tests for Mixed Models". *Scandinavian Journal of Statistics* 31, pp. 443-458.

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

Staudhammer, C.L., Lemay, V.M., Kozak, R.A, and Maness, T.C. (2005). "Mixed-Model Development for Real-Time Statistical Process Control Data in Wood Products Manufacturing". *Forest Biometry, Modelling and Information Sciences* 1, pp. 19-35.

Sullivan, J.H. and Woodall, W.H. (1996). "A Comparison of Multivariate Control Charts for Individual Observations". *Journal of Quality Technology* 28, pp. 398-408.

Tan, F.E.S., Ouwens, J.N. and Berger, M.P.F. (2001). "Detection of Influential Observations in Longitudinal Mixed Effects Regression Models". *The Statistician* 50, pp. 271-284.

Vargas, J.A. (2003). "Robust Estimation in Multivariate Control Charts for Individual Observations". *Journal of Quality Technology* 35, pp. 367-376.

Verbeke, G. and Lessafre, E. (1996). "A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population". *Journal of the American Statistical Association* 91, pp. 217-221,

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York, NY.

Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York, NY.

Wang, K. and Tsung, F. (2005). "Using Profile Monitoring Techniques for a Data-Rich Environment with Huge Sample Sizes". *Quality and Reliability Engineering International* 21, pp. 677-688.

Waternaux, C., Laird, N.M., and Ware, J.H. (1989). "Methods for Analysis of Longitudinal Data: Blood-Lead Concentrations and Cognitive Development". *Journal of the American Statistical Association* 84, pp. 33-41.

Williams, J.D., Woodall, W.H., Birch, J.B., and Sullivan, J.H. (2006). "On the Distribution of T^2 Statistics Based on Successive Differences". to appear in *Journal of Quality Technology*.

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, pp. 309-320.

Xu, R. (2003). "Measuring Explained Variation in Linear Mixed Effects Models". *Statistics in Medicine* 22, pp. 3527-3541.