Variation Transmission in Multi-Stage Industrial Processes

by

Rekha Agrawal

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 1997

0-612-21327-7

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

The subject of variation analysis is of interest in manufacturing processes where items are being produced in large quantity and pass through many operations or stages before they are completed. After the final operation, they must meet certain specifications. The issue is to discover how variation in the product characteristics at the final stage of the process can be reduced. With that goal in mind, it is useful to understand how the variation is conveyed through the process.

Multivariate normality is assumed as the underlying model for the measured product. Methods are given for analysing variance transmission under this model, both when a general multivariate normal holds, and in a more restricted case, when a first order autoregressive structure is appropriate.

Inevitably, there will be measurement error in the data collected on the process. It is shown that this measurement error can severely hinder attempts to characterize the process, and should be incorporated explicitly in an analysis. A naive estimation method is introduced and shown to work well.

It may be less expensive, in some instances, to collect large amounts of sample data after each stage, and then track only a few items through the process. Methods are given of incorporating cross-sectional data into the analysis. Also discussed is how to do this when the problem is compounded by measurement error.

Finally, some consideration is given to the issue of multivariate data.

Acknowledgements

I would first like to thank my supervisors, Jock MacKay and Jerry Lawless. I have not only learned a great deal about research from them, they also gave me an enormous amount of support. Their patience was boundless.

Thanks to Greg Bennett, Stefan Steiner and Jim Whitney for their input. There are many people in the statistics department and on the fifth floor who made my days much more cheery. Thanks to them also.

I would like to thank Patrick Maidorn, for his constant support. He put up with me even in my most "stressed out" times. Matt Schonlau deserves a big thanks for "sticking with me" throughout my entire time here at Waterloo. I would also like to acknowledge Andreas Sashegyi for his remarkable ability to find something positive in any situation.

Finally, I can't express how grateful I am to my mom, dad and brother for their belief in me, and for their unwavering love and support. They are my pillars of strength.

Dedicated to the memory of

Chelsea Anne Pichach

With you a part of me hath passed away;
For in the peopled forest of my mind
A tree made leafless by this wintry wind
Shall never don again its green array.
Chapel and fireside, country road and bay,
Have something of their friendliness resigned;
Another, if I would, I could not find,
And I am grown much older in a day.
But yet I treasure in my memory
Your gift of charity, and young heart's ease,
And the dear honor of your amity;
For these once mine, my life is rich with these.
And I scare know which part may greater be -
What I keep of you, or you rob from me.

- George Santayana

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis deals with the subject of analysing variation in an industrial process. This subject is of interest in many industrial processes in which items are being produced in large quantity. These items generally pass through many operations before they are completed. After the final operation, they must meet certain specifications. The issue of interest is to discover how variation in the product at the final stage of the process can be reduced. A process that has little variation in its final product is a cost efficient one, since few parts will be scrapped due to failure to meet specifications. Also, to produce high quality products it is important to minimize variation in key quality characteristics. For instance, suppose a consumer buys a new car, and discovers that although the vehicle has just been purchased, it is noisy to drive in because the doors of the car do not close tightly. Thus, wind can be heard traveling around the car, and the consumer finds the drive unpleasant. This particular problem would be eliminated if the manufacturer consistently made vehicles with doors that close tightly in their frames. Variation analysis is useful in determining how to do this.

The key to reducing variation in the final product is to have an understanding of how much of that variation arises at each stage of a process. If some data can be tracked through the course of the process, then statistical methods can be used to determine those stages that are the largest contributors to the variation. Statistical insight into this problem helps to focus engineering efforts. The aim is a less variable product.

1.1 Description of Problem

We consider the problem of an industrial process producing items that should conform to certain target values. These targets may relate to the dimensions of the product, or they may relate to other characteristics such as, for example, roundness, flatness or smoothness. These characteristics are referred to as quality characteristics, because they are the measure of performance of the product. For more on quality characteristics, see Montgomery (1985), Moen et al. (1991), Nair (1992) or Roy (1990). The products of the process will naturally vary about the given values, and this variation may be costly to the manufacturer if it results in customer dissatisfaction (Provost, 1990). It is therefore desirable to minimize the variation of the process. For some characteristics, variation only needs to be reduced to the point where the product will meet specification. For other characteristics, any reduction in variation is desirable, even after the product conforms to specification. These types of characteristics are called key quality characteristics (KQC's).

Often, the industrial process in question consists of many serial or parallel stages

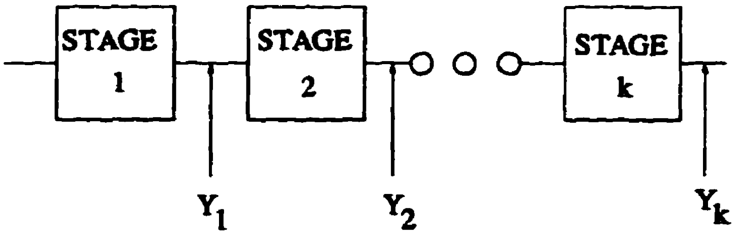

Figure 1.1: Stages of the process

that the items pass through before they are completed, as illustrated in Figure 1.1. This complicates the issue of minimizing variation in the quality characteristics, because it is no longer clear where the variation in the product at the final stage is coming from. Consider a two stage process, for instance, where a measurement in the same characteristic is taken before an operation and then again after it (the final stage). Consider a given amount of variation in the process before the operation. Then, any combination of three things can happen. The operation might simply transmit the variation, in which case the variation of the product at the final stage is determined by the variation in the product at the first stage. It is also possible that the operation adds to the variation. The variation present at the first stage will be of little importance if the variation added at the operation is large comparatively. Another possibility is that the operation may in fact "remove" the variation that was present in the process at the first stage. In that case, the variation at the first stage of the process is not relevant to the variation at the final stage. Clearly, the scenarios given above can be generalized to any number of stages.

We can illustrate these situations graphically through the use of scatter plots and sequence plots. If we track items through the two stage process and plot their measurements, we get a scatter plot of the data. If this scatter plot appears as

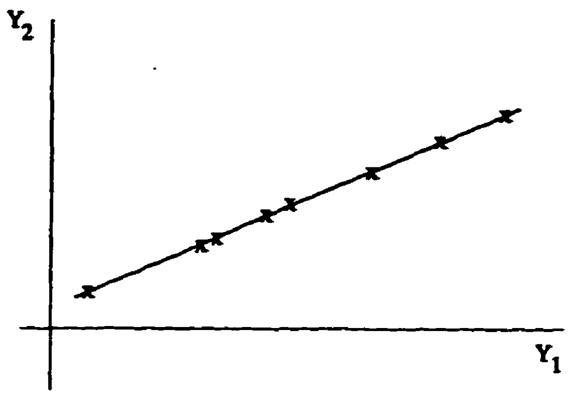

Figure 1.2: Perfect transmission

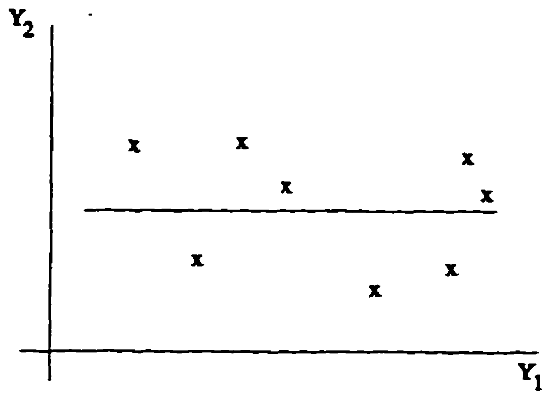

Figure 1.3: Total added variation

shown in Figure 1.2, then we have perfect transmission of variance, since all the variation in Y_2 is due to variation in Y_1. Clearly, the slope of the regression is relevant here - if we could "flatten" the line so that it is horizontal, there would be no variation in Y_2. This corresponds to removing variation from the first stage.

If, on the other hand, a scatter plot of the data revealed the figure shown in 1.3, we would have total added variation in Y_2. This is because none of the variation in Y_2 is due to variation in Y_1.

The more likely situation is that some combination of the above two situations occurs, as illustrated in Figure 1.4.

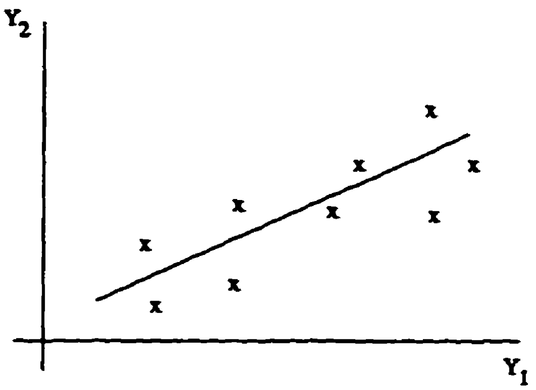

Figure 1.4: Both added and transmitted variation

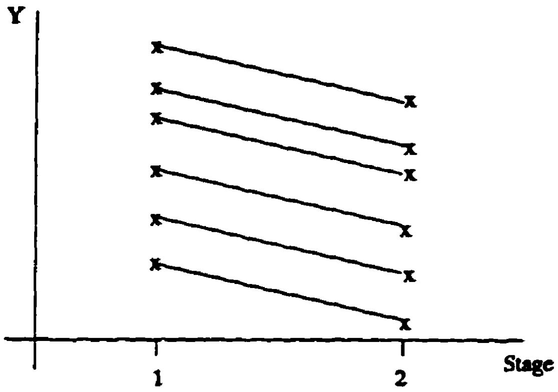

Figure 1.5: Perfect transmission of variance - sequence plot

Sequence plots can also be used to illustrate these situations. These are plots of the time sequence for each item tracked through the process. In the case of perfect transmission, this plot would appear as illustrated in Figure 1.5, where all the lines are parallel. When there is total added variation, this plot would appear as in Figure 1.6, where the lines on the plot all cross.

Assuming it is possible to track at least some items through the process and make measurements after each stage, those stages that are contributing the most to the variation in the final product can be identified. This contributes to a deeper understanding of the variation in the process and how it affects variation in the

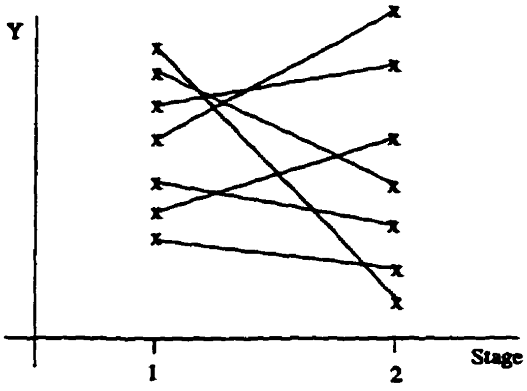

Figure 1.6: Total added variation - sequence plot

quality characteristic. Potentially, an intervention could be made that reduces the variation contributed at the key sources, and ultimately results in a more cost effective process and lower variation in the quality characteristic.

1.2 Examples

1.2.1 Piston Example

This example will be used throughout this thesis to provide a numerical illustration of the concepts described.

A piston is a part in an automobile located in the engine cylinder, the basic framework of the engine. The piston is essentially a cylinder closed at the top and open at the bottom, where it is connected to a rod. The piston moves in a vertical motion in the engine cylinder, pushing out exhaust on the upstroke, and intaking fuel on the downstroke (Crouse, 1970).

A study was done on 96 pistons as they were passing through a production line. Each of the 96 pistons studied had 53 observations recorded on it. The process is illustrated schematically in Figure 1.7. The quality characteristics of the piston were four diameters, located at a height of 4 mm, 10 mm, 36.7 mm and 58.7 mm. These diameters were measured after each operation in the process, denoted in Figure 1.7 by Y1 - Y7. It should be noted that all diameters were measured in millimetres, to a precision of 0.001 millimetres, or 1 micron.

The following is a breakdown of the measurements on a piston.

(1) Piston number.

(2) Die number - A piston is produced from one of six possible dies. Each die produced an equal number of pistons (Z1).

(3) Week number - This study was done over a two week period; 48 pistons were produced in each week (Z2).

(4) Path 270 machine number - At operation 270, there were two different machines that the piston could have come through. An equal number of pistons went through each machine (Z3).

(5) Path 290 machine number - The same situation occurred at operation 290 (Z4).

(6-20) Covariates - 15 attributes were measured on the pistons before production (X1-X15).

(21,22) Op 210 - After operation 210, the diameters of the pistons were measured at 4 mm. and 58.7 mm. At this particular operation, no measurements were made at 10 mm and 36.7 mm (Y1).

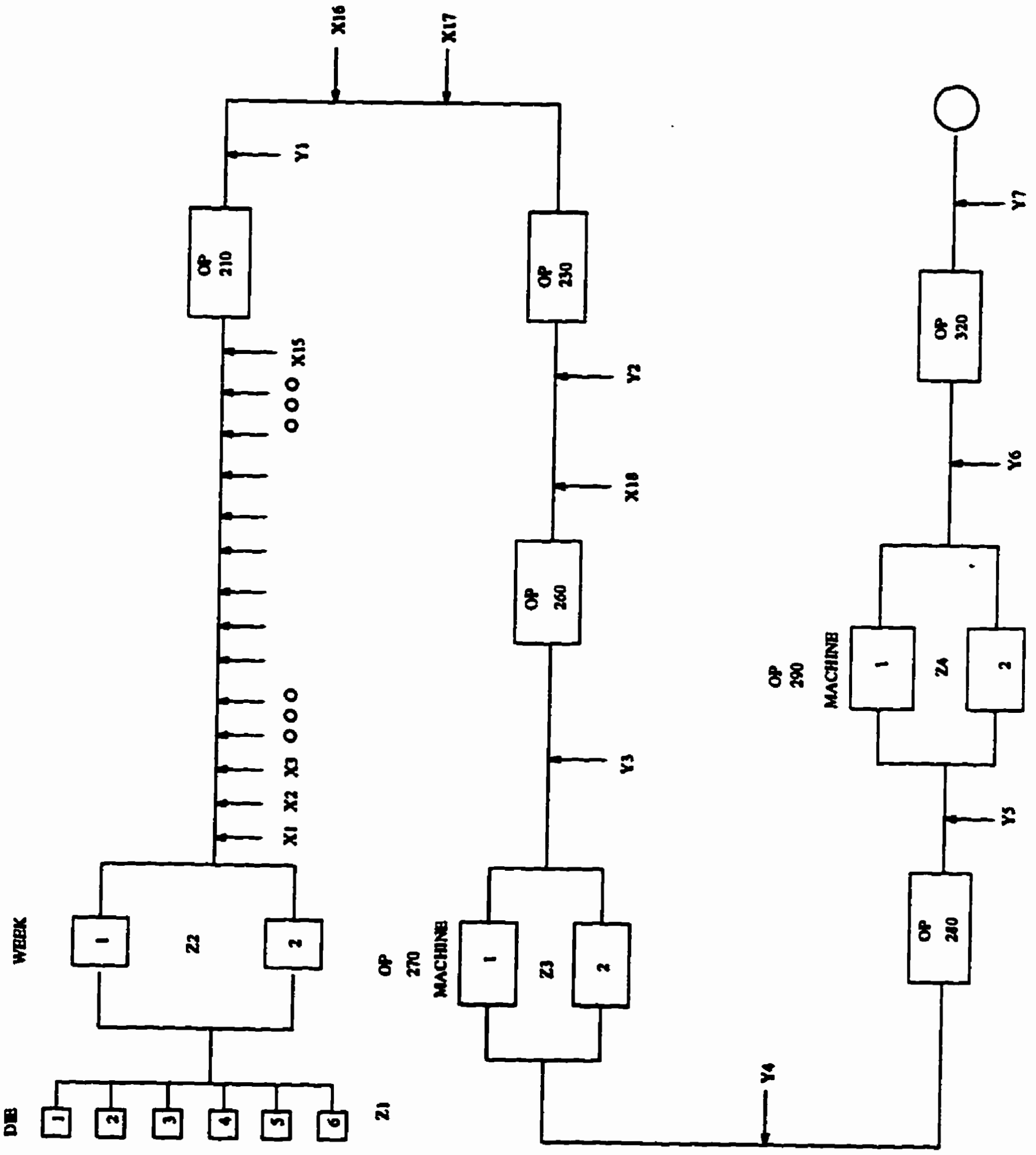

Figure 1.7: Schematic diagram of process. Note that Yi denotes the ith set of four diameter measurements.

(23,24) Covariates - Two attributes were measured after op 210 (X16,X17).

(25-28) Op 230 - Four diameters were measured after op 230: 4 mm, 10 mm, 36.7 mm, and 58.7 mm (Y2).

(29) Covariate - One attribute was measured after op 230 (X18).

(30-33) Op 260 - Four diameters were measured after op 260 (Y3).

(34-37) Op 270 - Four diameters were measured after op 270 (Y4).

(38-41) Op 280 - Four diameters were measured after op 280 (Y5).

(42-45) Op 290 - Four diameters were measured after op 290 (Y6).

(46-49) Op 320 - Four diameters were measured after op 320 (Y7).

(50-53) Op 320 - Four diameters were measured after op 320 using a different gauge (Y7F).

In any subsequent analysis, when a measurement was required for the final diameter of the piston, the first set of measurements (46-49) was used instead of the second set (50-53), because the former was deemed to be more reliable. The second set of measurements was taken from a different measurement machine than the others.

This is an example of the type of multi-stage industrial process described above. It is of interest to identify the stages of the process that are major contributors of variation in the final diameters.

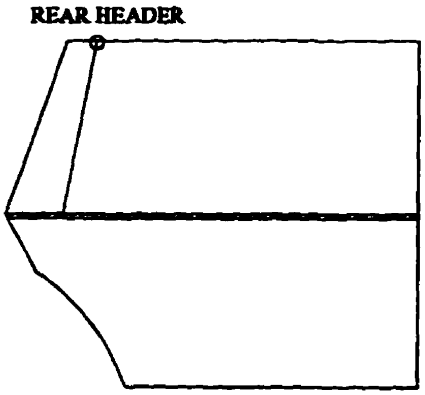

Figure 1.8: Location of rear header on door

1.2.2 Door Hanging Example

Another example that will be used occasionally is some car assembly door hanging data. Here, thirteen cars were tracked through a seven operation process, and the flushness of the rear header was measured on the rear door. This is an in-out measurement which can either be above or below the target value. To see where this location is on the door, see Figure 1.8.

The seven operations that the cars went through were the following:

1. Door hanging

2. Paint

3. Door hardware installation

4. Striker installation

5. Striker fit

6. Seals and chassis

7. Final fit

For each car, a measurement was taken on the rear header after each of the above operations.

Again, this is an example of the type of multi-stage industrial process we are interested in. In fact, several flushness measurements were taken on each car. Note that the geometry of the car door might lead us to consider several quality characteristics. The flushness measurements themselves are clearly quality characteristics. The difference between measurements on the top of the car door and on the bottom will indicate how the door is tilted in that plane, and hence might also be a quality characteristic of interest. Similarly for the difference between measurements made on the left of the door and on the right.

1.3 Statistical Issues and Problems

The subject of reducing variation is discussed throughout the quality literature. See, for example, Joiner and Gaudard (1990), Pyzdek (1990), and Nolan and Provost (1990). References on this issue that more closely resemble our approach, however, are Lawless, MacKay and Robinson (1996), Hamada and Lawless, Wu et al. (1994), Xie et al. (1994) and Knof and Farrow (1996).

To illustrate the types of issues we address, consider a simplified situation in which there is a single operation. A measurement X is made before the operation and a measurement Y is made afterwards, where X and Y are not necessarily measurements of the same thing. This can be thought of as illustrated in Figure

1.1, where there are two stages. It will then be true that

$$Var(Y) = Var_X(E(Y|X)) + E_X(Var(Y|X))$$

The first term of the above equation can be interpreted as the variation in Y explained by X. The second term can be interpreted as the unexplained variation in Y. If we assume that the measurement of X carries all relevant information about the variation of Y at that stage, then we can also interpret this equation as the following: the first term is the variation transmitted to Y from the first stage and the second term is the variation added to Y after the first stage. Clearly, to compute the relevant expectations and variances, models are needed for $f(X)$ and $f(Y|X)$, where f denotes a probability function in the discrete case or a probability distribution function in the continuous case. At the very least, the first two moments of these functions will be needed. With these models in hand, the variation in Y can be broken down as desired.

In the general problem, we consider a k-stage process, with upstream measurements $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_{k-1}$ and the final quality characteristic measurements $\mathbf{Y}$. All of the upstream and quality characteristic measurements may be vectors, and need not be measurements of the same characteristic at each stage. It is possible, for example, that some of the $\mathbf{X}_i$ are measurements of the quality characteristic at an earlier stage of the process, while others may be measurements of completely different attributes of the process. Ideally, we would use all of this information to understand how variation in $\mathbf{Y}$ is propagated, and how it might be reduced.

This thesis will focus on the problem when the same quality characteristics are measured at each stage. Thus, the measurements are $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$. There are many statistical issues associated with this problem. Finding appropriate models to describe the data is the first issue. Associated with this are issues of model fitting and assessment.

The data are often observed with some measurement error. This is another issue of importance, because often the error involved can be substantial and ignoring it can seriously mislead the investigator. Methodology needs to be developed to explicitly handle this error.

Another issue of interest is related to data collection. It frequently occurs that while tracking items through a process is expensive, measuring large numbers of items after each stage is considerably less expensive. Methodology that uses this type of "cross-sectional" data in the analysis would be useful.

Missing data is another relevant statistical issue. Frequently, not all data can be taken on all the items after each stage. This is especially true when the data are collected using automatic methods, such as coordinate measurement machines. Methods are needed that use the data that are available as efficiently as possible.

Another thing that occurs often in these situations is that the data that are collected are correlated cross-sectionally, and so multivariate methods are needed in the analysis. Analysing the quality characteristics one variable at a time is not sufficient. The diameters measured on the pistons are an example of such data. Although models can be developed to take account of correlated data, the difficulty occurs when trying to relate analysis done with these models back to the original process.

This thesis is outlined as follows: the remainder of this chapter introduces the univariate AR(1) model and some methods of analysis; Chapter two discusses methods of analysis of variance transmission in the presence of measurement error; Chapter three discusses how to handle the data when the longitudinal data are supplemented with cross-sectional data; Chapter four discusses non AR(1) normal models; Chapter five provides some discussion on multivariate data and Chapter six presents conclusions and ideas for future research.

Most of this thesis will focus on univariate measurements. This is applicable methodology when there is only one quality characteristic of interest, or when there are more, but they are uncorrelated.

1.4 The AR(1) Model

1.4.1 The Model

The use of the AR(1) model was proposed by Lawless, MacKay and Robinson (1996), following work by Robinson.

As a first step in addressing the identification of key sources of variation in an industrial process, we consider a two stage process in which there is a unique dimension of interest. We will assume Y_1 and Y_2 to be random variables from a bivariate normal distribution. In that case, we can represent them as follows:

$$\begin{aligned} Y_1 &\sim N(\mu_1, \sigma_1^2) \\ Y_2|Y_1 &\sim N(\alpha_2 + \beta_2 Y_1, \sigma_{2A}^2). \end{aligned}$$

Then, by the conditional variance formula,

$$\begin{aligned} Var(Y_2) &= E(Var(Y_2|Y_1)) + Var(E(Y_2|Y_1)) \\ &= \sigma_{2A}^2 + \beta_2^2 Var(Y_1). \qquad (1.1) \end{aligned}$$

The interpretation of equation (1.1) is that the first term is the amount of variation added due to the operation, whereas the second term is the amount of variation present in Y_1 that is transmitted through to Y_2. If β_2 is close to one, almost all of the variation present in Y_1 will be transmitted to Y_2. Conversely, if β_2 is close to zero, then the variation in Y_2 is due almost entirely to the variation added at the operation.

We now expand the process to three stages. The AR(1) model assumption specifies that the conditional distribution of a particular variable, Y_i, given all the earlier ones, $Y_1, \ldots, Y_{i-1}$, is normal, with a mean which is a function only of the previous variable, Y_{i-1}, and a constant variance. We will subsequently refer to this model as the first order autoregressive model (AR(1)), due to its similarity to the time series model of the same name. In this case,

$$\begin{aligned} Y_1 &\sim N(\mu_1, \sigma_1^2) \\ Y_2|Y_1 &\sim N(\alpha_2 + \beta_2 Y_1, \sigma_{2A}^2) \\ Y_3|Y_2, Y_1 &\sim N(\alpha_3 + \beta_3 Y_2, \sigma_{3A}^2) \end{aligned}$$

Again using the conditional variance formula, we get that

$$\begin{aligned} Var(Y_3) &= \sigma_{3A}^2 + \beta_3^2 Var(Y_2) \\ &= \sigma_{3A}^2 + \beta_3^2\sigma_{2A}^2 + \beta_3^2\beta_2^2 Var(Y_1) \qquad (1.2) \end{aligned}$$

The first term in equation (1.2) is the variance added due to operation two. The second term is the variation added at operation one and transmitted through to Y_3, and the final term is the variation transmitted from Y_1.

Notice that if we are interested in collapsing both of the above operations into one single operation, we could consider the effect of that combined operation. It can be shown that

$$\begin{aligned} E(Y_3|Y_1) &= E_{Y_2|Y_1}(E(Y_3|Y_2, Y_1)) \\ &= \alpha_3 + \beta_3(\alpha_2 + \beta_2 Y_1). \end{aligned}$$

Also,

$$\begin{aligned} Var(Y_3|Y_1) &= E_{Y_2|Y_1}(Var(Y_3|Y_2, Y_1)) + Var_{Y_2|Y_1}(E(Y_3|Y_2, Y_1)) \\ &= \sigma_{3A}^2 + \beta_3^2\sigma_{2A}^2 \end{aligned}$$

Obviously, the above expressions give the same results for the unconditional variance of Y_3 as was found in (1.2).

This type of calculation can similarly be carried out on any number of serial operations. In the general case, there are k stages and under the AR(1) model, it is assumed that

$$\begin{aligned} Y_1 &\sim N(\mu_1, \sigma_1^2) \\ Y_i | Y_1, \ldots, Y_{i-1} &\sim N(\alpha_i + \beta_i Y_{i-1}, \sigma_{iA}^2) \qquad i = 2, \ldots, k. \end{aligned} \tag{1.3}$$

Equivalently, we could write

$$\begin{aligned} Y_1 &\sim N(\mu_1, \sigma_1^2) \\ Y_i | Y_1, \ldots, Y_{i-1} &\sim N(\alpha_i^* + \beta_i (Y_{i-1} - \mu_{i-1}), \sigma_{iA}^2) \qquad i = 2, \ldots, k. \end{aligned}$$

This form is more convenient when discussing targeting, since $\mathrm{E}(Y_i) = \alpha_i^* = \mu_i$. The variance formulae are not affected by this alternate form.

The AR(1) model states that the current measurement is a function of the previous measurement only. This is often a reasonable situation from a physical point of view. Some reasons for which this might not hold, however, are that the multivariate normal model and the linear form of $E(Y_i|Y_{i-1})$ may not be valid. Further, if there are several correlated variables and key ones are not observed, then the observed measurements may not conform to an AR(1) model.

A useful "marginal" re-parameterization of the AR(1) model is the following:

$$\begin{aligned} Y_i &\sim N(\mu_i, \sigma_i^2) \\ \rho_{i-2,i} &= \rho_{i-2,i-1} * \rho_{i-1,i}. \end{aligned}$$

Here, ρ_{ij} represents the correlation between measurements at stage i and j. In this parameterization, the variance partition of a k-stage process is expressed as

$$\begin{aligned} Var(Y_k) &= \sigma_k^2(1-\rho_{k-1,k}^2)+\sigma_k^2\rho_{k-1,k}^2(1-\rho_{k-2,k-1}^2)+\cdots \\ &+ \sigma_k^2\rho_{k-1,k}^2\cdots\rho_{2,3}^2(1-\rho_{1,2}^2)+\sigma_k^2\rho_{k-1,k}^2\cdots\rho_{2,3}^2\rho_{1,2}^2. \end{aligned} \tag{1.4}$$

Dividing by the total variance σ_k^2 gives

$$1=(1-\rho_{k-1,k}^2)+\rho_{k-1,k}^2(1-\rho_{k-2,k-1}^2)+\cdots\rho_{k-1,k}^2\cdots\rho_{2,3}^2\rho_{1,2}^2. \tag{1.5}$$

This form indicates the proportion of the variance of the final product that is contributed at each stage. The proportions of variance are generally of more interest than the components themselves.

It should be noted here that if the AR(1) model is appropriate, collapsing operations one and two into a combined operation will result in a variance partition equivalent to that given by equation (1.2). That is, the first two terms of equation (1.2) will sum to give the added variation of the combined operation.

Since, in the case of the AR(1) model considered here, all partitions of variance of the final response attribute consistent contributions of variation to the previous stages, evaluating the effect of an intervention in the process is relatively straightforward. This assumes that the AR(1) structure is not affected by the intervention. For instance, if in the case of three stages, the variation added at operation two (σ_{3A}^2) is reduced by one half, then this reduces the first term in equation (1.2) by one half. Also, if some intervention could be made that changes the slope of Y_3 on Y_2 (β_3) to one half its value, then both the second and third terms in equation

(1.2) reduce by one quarter. This approach corresponds to "removing" variation at operation two. These scenarios suggest different ways of reducing the variation in Y_3.

For references on this type of model used in longitudinal data analysis, see, for example, Diggle et al. (1994).

1.4.2 Maximum Likelihood Estimation

Since the data in this situation are n items that are tracked through a k-stage process and measured after each stage, we will write y_{ab} to denote the bth item's measurement after the ath stage. In that case, the maximum likelihood estimates for the AR(1) model parameters are (Lawless, MacKay and Robinson, 1996):

$$\begin{aligned}
\hat{\mu}_1 = \bar{y}_1 \qquad \hat{\sigma}_1^2 &= \frac{1}{n}\sum_{k=1}^{n}(y_{1k}-\bar{y}_1)^2 \\
\hat{\beta}_i = \frac{S_{y_{i-1}y_i}}{S_{y_{i-1}y_{i-1}}} & \qquad i = 2,\ldots,k \\
\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_i\bar{y}_{i-1} & \qquad i = 2,\ldots,k \\
\hat{\sigma}_{iA}^2 = \frac{S_{y_iy_i}}{n} - \hat{\beta}_i\frac{S_{y_{i-1}y_i}}{n} & \qquad i = 2,\ldots,k
\end{aligned} \tag{1.6}$$

where

$$\begin{aligned}
\bar{y}_i &= \frac{1}{n}\sum_{k=1}^{n} y_{ik} \\
S_{y_iy_i} &= \sum_{k=1}^{n}(y_{ik}-\bar{y}_i)^2 \\
\text{and} \qquad S_{y_{i-1}y_i} &= \sum_{k=1}^{n}(y_{i-1,k}-\bar{y}_{i-1})(y_{i,k}-\bar{y}_i).
\end{aligned}$$

In the marginal parameterization, they are

$$\begin{array}{lll} \hat{\mu}_i = \bar{y}_i & \hat{\sigma}_i^2 = \frac{1}{n} S_{y_i y_i} & i = 1, \ldots, k \\ \hat{\rho}_{i-1,i} = \frac{S_{y_{i-1} y_i}}{\sqrt{S_{y_{i-1} y_{i-1}} S_{y_i y_i}}} & & i = 2, \ldots, k \end{array}$$

Using these estimates, the total estimated variance at any stage can be exactly partitioned into its components.

It is not true, however, that the estimated components of two or more operations will sum to give the component of a combined operation. For example, the estimated added variation in the process between Y_1 and Y_3 would be

$$\hat{\sigma}_{3A}^2 + \hat{\beta}_3^2 \hat{\sigma}_{2A}^2 = \frac{S_{y_3 y_3}}{n} - \frac{S_{y_2 y_1}^2}{n S_{y_1 y_1}} \left(\frac{S_{y_2 y_3}^2}{S_{y_2 y_2}^2}\right).$$

If operations one and two are grouped together, however, and Y_2 is not observed, then we would estimate the added variation between Y_1 and Y_3 to be

$$\frac{S_{y_3 y_3}}{n} - \frac{S_{y_1 y_3}^2}{n S_{y_1 y_1}}.$$

This indicates that while the true components of variance added at operations one and two should sum to that component added by the super operation with an AR(1) model, the estimates of these components do not. If the AR(1) model is correct and the sample size is moderately large, though, these estimates should roughly add.

1.4.3 Diagnostics

Various graphical and formal methods can be used to determine the adequacy of the AR(1) model.

Since the AR(1) model implies that the marginal distributions of each of the stages must be normal, the observed values from each stage can be plotted using a QQ plot (see Johnson et al., 1988, p. 146). If any of these plots reveal substantial departures from normality, the AR(1) model should be rejected.

If not, however, then the assumption of linearity of the conditional means should be verified. This would require plotting all combinations of the stages pairwise to see if the linear assumption is reasonable.

Plots of the residuals should also be made to see if the first order autoregressive relationship holds. Hence, the residuals of the Y_i vs Y_{i-1} regression should be plotted against all previous stages, $Y_{i-2}, \ldots, Y_2, Y_1$. If these plots indicate any relationship between the residuals and the variables $Y_{i-2}, \ldots, Y_2, Y_1$, then the AR(1) model is not applicable, since Y_i would then be a function of something other than just Y_{i-1}. Different methodology will be required in this case.

The assumption of constant variance can be verified by plotting the residuals against their predicated values. For example, outward-opening funnel shapes on these plots indicate that the variance is changing with the mean. Details are given in Montgomery and Peck (1992), p. 74 or Draper and Smith (1981), p. 147.

To formally test univariate normality, a number of tests have been developed. Popular test are the Shapiro-Wilks statistic, and tests of skewness and kurtosis. See, for example, Madansky (1988).

The bivariate normality of consecutive stages can be tested by generating elliptical contours of the bivariate density with the estimated parameters, and comparing the proportion of sample observations lying inside these contours to a theoretical values. See Jobson (1991), p. 115.

The first order autoregressive nature of the data can be tested using the extra sums of squares principle. Using this method, the model

$$Y_i = \alpha + \beta Y_{i-1} + \epsilon$$

can be tested against the model

$$Y_i = \alpha + \beta_1 Y_1 + \beta_2 Y_2 + \ldots + \beta_{i-1} Y_{i-1} + \epsilon$$

If the smaller model is not adequate, then the AR(1) assumption is not valid. See, for example, Montgomery and Peck (1992), p. 139, or Draper and Smith (1981), p. 97.

A likelihood ratio test could also be done to test the AR(1) model against a more general multivariate normal model. Details on how to do this for a larger class of models are given in the next chapter.

1.4.4 Missing Data

It sometimes happens in industrial processes that all the desired measurements on a part are not taken at all the stages. When this occurs, methods of estimating distribution parameters are needed that make use of all the available data. If some data are missing in a serial process, but the process is thought to adhere to an AR(1) model, sets of data from adjacent stages can be considered pairwise, and the bivariate normal distribution parameters estimated.

The EM algorithm (Little and Rubin, 1987) can be used to derive estimates of the parameters, assuming the data are missing at random. For a bivariate normal distribution when both variables may contain missing values, this calculation involves dividing the data into three groups: (1) units in which the first variable is observed but the second is not, (2) units in which both variables are observed and (3) units in which the first variable is missing but the second is observed. For further details, see Little and Rubin (1987), page 132. Fong and Lawless (1996) give a general solution to this problem.

Chapter 2

The AR(1) Model with Measurement Error

In industrial processes, the measurement system involved in determining the quantities of interest is an important issue. With the technological developments of recent years, machines are being used that are capable of repeating measurements to a remarkable precision. This is not the only factor, however, that is relevant when considering the error involved in determining the true dimension. Usually, experiments have been done to determine the precision of the measurement system, taking into account factors such as different operators and positioning inside a measurement machine, as well as the machine itself. The term measurement error refers to the error that occurs as a result of all of these different sources of variability. It is an issue of concern when dealing with data of all sorts, and has been addressed by authors such as Fuller (1987), Seber (1977) and Johnson (1972).

To add measurement error to the AR(1) model introduced in the previous chapter, suppose the measured values of the characteristic of interest are $X_1, X_2, \ldots, X_k$ where

$$X_i = Y_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2_{\epsilon_i}) \tag{2.1}$$

We will mostly assume that the variances $\sigma^2_{\epsilon_i}$ are known. When we do discuss estimating $\sigma^2_{\epsilon_i}$, however, we will assume that the data used to do this are independent of the process data.

The process $(X_1, X_2, \ldots, X_k)$ is no longer AR(1) if $k > 2$. In fact, the conditional distribution of $X_i|X_1, \ldots, X_{i-1}$ depends on all of $X_1, \ldots, X_{i-1}$.

Given observations $(X_1, X_2, \ldots, X_k)$ on n items, the goal is to estimate the proportions of variance (1.5). The fact that we no longer observe the Y_i due to the presence of measurement error substantially complicates this problem.

This chapter is outlined as follows: the first section discusses the effects of measurement error if ignored, the second section elaborates on the estimation problem and introduces an alternative method to maximum likelihood, the third section provides some model checking techniques, the fourth section discusses approaches to use when the measurement error is estimated instead of known exactly, the fifth section describes solutions to the missing data problem, and the last section describes using these techniques for the piston example.

2.1 Effects of Measurement Error if Ignored

We review the effects of ignoring measurement error (Lawless et al., 1996), since this will motivate what follows. To demonstrate the effect of ignoring measurement error in the identification of the variance proportions (1.5), consider first a two stage process in which

$$X_1 \sim N(\mu_1, \sigma_1^2 + \sigma_{e_1}^2) \qquad X_2 \sim N(\alpha_2 + \beta_2\mu_1, \sigma_{2A}^2 + \beta_2^2\sigma_1^2 + \sigma_{e_2}^2) \tag{2.2}$$

with $Cov(X_1, X_2) = \beta_2\sigma_1^2$. This comes from (1.3) and (2.1).

Suppose n items are tracked through the process so that we have data (x_{1j}, x_{2j}; j=1, $\ldots, n$) and we estimate the variance components assuming that the AR(1) model is appropriate, that is, assuming $\sigma_{e_1} = \sigma_{e_2} = 0$. Then the maximum likelihood estimates given earlier are

$$\hat{\sigma}_1^2 = \frac{S_{x_1x_1}}{n}, \qquad \hat{\beta}_2 = \frac{S_{x_1x_2}}{S_{x_1x_1}}, \qquad \hat{\sigma}_2^2 = \frac{S_{x_2x_2}}{n},$$

$$\text{where} \qquad S_{x_ix_j} = \sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

Note that as $n \to \infty$, $\frac{S_{x_ix_j}}{n} \to Cov(X_i, X_j)$, where "$\to$" denotes convergence in probability, so that

$$\hat{\sigma}_2^2 \to \sigma_2^2 + \sigma_{e_2}^2 \qquad \hat{\beta}_2 \to \beta_2\frac{\sigma_1^2}{\sigma_1^2 + \sigma_{e_1}^2} \qquad \hat{\sigma}_1^2 \to \sigma_1^2 + \sigma_{e_1}^2.$$

In the partition (1.5),

$$1 = \frac{\sigma_{2A}^2}{\sigma_2^2} + \frac{\beta_2^2\sigma_1^2}{\sigma_2^2},$$

the estimates are such that as $n \to \infty$

$$\frac{\hat\beta_2^2\hat\sigma_1^2}{\hat\sigma_2^2} \to \frac{\beta_2^2\sigma_1^2}{\sigma_2^2}(\frac{\sigma_1^2}{\sigma_1^2+\sigma_{\epsilon_1}^2})(\frac{\sigma_2^2}{\sigma_2^2+\sigma_{\epsilon_2}^2}).$$

Hence, the variation transmitted from stage one is underestimated. Since the estimates of the proportions must also sum to one, this implies that the variation added at stage two is overestimated. If the measurement system contributes 20% of the variation in X_1 and X_2, then the asymptotic bias is substantial.

Suppose we expand this to a process with three stages. If we ignore the measurement error, then we would use the estimates

$$\hat\sigma_1^2 = \frac{S_{x_1x_1}}{n}, \qquad \hat\sigma_{2A}^2 = \frac{S_{x_2x_2}}{n} - \hat\beta_2\frac{S_{x_1x_2}}{n}, \qquad \hat\sigma_{3A}^2 = \frac{S_{x_3x_3}}{n} - \hat\beta_3\frac{S_{x_2x_3}}{n},$$

$$\hat\beta_2 = \frac{S_{x_1x_2}}{S_{x_1x_1}}, \qquad \hat\beta_3 = \frac{S_{x_2x_3}}{S_{x_2x_2}}.$$

Then, the proportions of variance contributed according to (1.5) are

$$1 = \frac{\hat\sigma_{3A}^2}{\hat\sigma_3^2} + \frac{\hat\beta_3^2\hat\sigma_{2A}^2}{\hat\sigma_3^2} + \frac{\hat\beta_3^2\hat\beta_2^2\hat\sigma_1^2}{\hat\sigma_3^2}.$$

Using the above estimates,

$$\begin{aligned}
\frac{\hat\sigma_{3A}^2}{\hat\sigma_3^2} &\to \frac{\sigma_{3A}^2}{\sigma_3^2}(\frac{\sigma_3^2}{\sigma_3^2+\sigma_{\epsilon_3}^2}) + \frac{\sigma_{\epsilon_3}^2}{\sigma_3^2+\sigma_{\epsilon_3}^2} + \frac{\beta_3^2\sigma_2^2\sigma_{\epsilon_2}^2}{(\sigma_2^2+\sigma_{\epsilon_2}^2)(\sigma_3^2+\sigma_{\epsilon_3}^2)}, \\
\frac{\hat\beta_3^2\hat\sigma_{2A}^2}{\hat\sigma_3^2} &\to \frac{\beta_3^2\sigma_{2A}^2}{\sigma_3^2}(\frac{\sigma_3^2}{\sigma_3^2+\sigma_{\epsilon_3}^2})(\frac{\sigma_2^2}{\sigma_2^2+\sigma_{\epsilon_2}^2})^2(\frac{\sigma_2^2+\sigma_{\epsilon_2}^2-\beta_2^2\sigma_1^2(\frac{\sigma_1^2}{\sigma_1^2+\sigma_{\epsilon_1}^2})}{\sigma_2^2-\beta_2^2\sigma_1^2}), \\
\frac{\hat\beta_3^2\hat\beta_2^2\hat\sigma_1^2}{\hat\sigma_3^2} &\to \frac{\beta_3^2\beta_2^2\sigma_1^2}{\sigma_3^2}(\frac{\sigma_3^2}{\sigma_3^2+\sigma_{\epsilon_3}^2})(\frac{\sigma_2^2}{\sigma_2^2+\sigma_{\epsilon_2}^2})^2(\frac{\sigma_1^2}{\sigma_1^2+\sigma_{\epsilon_1}^2}).
\end{aligned} \tag{2.3}$$

While it is clear that the proportion of variance transmitted from the first stage is underestimated, the direction of bias for the other two proportions is not obvious. In fact, the bias of the variance added at the third stage is always positive, which can be seen by writing it in the marginal parameterization. In this form,

$$\frac{\hat{\sigma}_{3A}^2}{\hat{\sigma}_3^2} \to \frac{\sigma_2^2\sigma_3^2(1-\rho_{23}^2)+\sigma_2^2\sigma_{\epsilon_3}^2+\sigma_3^2\sigma_{\epsilon_2}^2+\sigma_{\epsilon_2}^2\sigma_{\epsilon_3}^2}{(\sigma_2^2+\sigma_{\epsilon_2}^2)(\sigma_3^2+\sigma_{\epsilon_3}^2)}.$$

2.2 Estimation

2.2.1 Two Stages

In the situation described above, it is possible to develop maximum likelihood estimates to take account of the measurement error. Recall that the distribution of (X_1, X_2) was given in equation (2.2). X_1 and X_2 have a bivariate normal distribution, and there are five functionally independent unknown parameters, μ_1, σ_1, α_2, β_2, σ_{2A} in the model. Equivalently, we may take the parameters to be $\mathrm{E}(X_1)$, $\mathrm{Var}(X_1)$, $\mathrm{E}(X_2)$, $\mathrm{Var}(X_2)$, and $\mathrm{cov}(X_1, X_2)$. The maximum likelihood estimates of these parameters are (Larsen et al., 1986)

$$\hat{E}(X_1) = \bar{x}_1, \qquad \hat{E}(X_2) = \bar{x}_2, \qquad \hat{Var}(X_1) = \frac{S_{x_1x_1}}{n}, \qquad \hat{Var}(X_2) = \frac{S_{x_2x_2}}{n}$$

$$\hat{Cov}(X_1X_2) = \frac{S_{x_1x_2}}{n}.$$

We then get the following maximum likelihood estimates for the original parameters by the invariance property, assuming that $\sigma^2_{e_i}$ are known:

$$\hat{\mu}_1 = \bar{x}_1 \qquad \hat{\sigma}_1^2 = \frac{S_{x_1x_1}}{n} - \sigma^2_{e_1} \qquad \hat{\beta}_2 = \frac{S_{x_1x_2}}{S_{x_1x_1} - n\sigma^2_{e_1}}$$
$$\hat{\alpha}_2 = \bar{x}_2 - \hat{\beta}_2\bar{x}_1 \qquad \hat{\sigma}^2_{2A} = \frac{S_{x_2x_2}}{n} - \frac{nS^2_{x_1x_2}}{(S_{x_1x_1} - n\sigma^2_{e_1})} - \sigma^2_{e_2}$$

assuming these estimates are greater than or equal to zero. If they are not, then some investigation should be made into whether the data are representative of the process, and also whether the measurement error variance is appropriate. A simplistic solution is to set the variance estimates to zero.

In the marginal parameterization for this model (μ_1, σ_1, μ_2, σ_2, ρ_{12}),

$$\hat{\mu}_1 = \bar{x}_1 \qquad \hat{\mu}_2 = \bar{x}_2 \qquad \hat{\sigma}_1^2 = \frac{S_{x_1x_1}}{n} - \sigma^2_{e_1} \qquad \hat{\sigma}_2^2 = \frac{S_{x_2x_2}}{n} - \sigma^2_{e_2}$$
$$\hat{\rho}_{12} = \frac{S_{x_1x_2}}{\sqrt{(S_{x_1x_1} - n\sigma^2_{e_1})(S_{x_2x_2} - n\sigma^2_{e_2})}}$$

These estimates are intuitive; we estimate the variance of Y_i, for example, by estimating the observed variance and subtracting the measurement error variance.

Exact distributional properties can be determined for $\hat{\mu}_1$ and $\hat{\sigma}_1^2$ above, since

$$\begin{aligned} X_1 &\sim N(\mu_1, \sigma_1^2 + \sigma^2_{e_1}), \\ \bar{x}_1 &\sim N(\mu_1, \frac{\sigma_1^2 + \sigma^2_{e_1}}{n}) \end{aligned}$$

and

$$\frac{\sqrt{n}(\bar{x}_1 - \mu_1)}{\{\frac{S_{X_1X_1}}{n-1}\}^{\frac{1}{2}}} \sim t_{n-1}.$$

Inferences can be made accordingly. Also

$$\frac{S_{x_1x_1}}{\sigma_1^2 + \sigma_{\epsilon_1}^2} \sim \chi_{n-1}^2$$

and since $\sigma_{\epsilon_1}^2$ is presumed known, inferences can be made about σ_1^2. For instance, a $100(1-\alpha)\%$ confidence interval for σ_1^2 is

$$[\frac{S_{x_1x_1}}{\chi_{1-\alpha/2,n-1}^2} - \sigma_{\epsilon_1}^2, \frac{S_{x_1x_1}}{\chi_{\alpha/2,n-1}^2} - \sigma_{\epsilon_1}^2]$$

assuming the left hand side is greater than zero. If not, it can be replaced by zero. Finding exact distributions for the remaining parameters proves to be more difficult. If we condition on the x_{1i}, we find that

$$E(\hat{\beta}_2 | x_{11}, x_{12}, \ldots, x_{1n}) = \frac{\beta_2 \sigma_1^2 S_{x_1x_1}}{(\sigma_1^2 + \sigma_{\epsilon_1}^2)(S_{x_1x_1} - n\sigma_{\epsilon_1}^2)}$$

which shows that the estimator is biased. Further,

$$Var(\hat{\beta}_2 | x_{11}, x_{12}, \ldots, x_{1n}) \approx \frac{S_{x_1x_1}(\sigma_{2a}^2 + \sigma_{\epsilon_2}^2 + \beta_2^2\sigma_1^2(\frac{\sigma_{\epsilon_1}^2}{\sigma_1^2+\sigma_{\epsilon_1}^2}))}{(S_{x_1x_1} - n\sigma_{\epsilon_1}^2)^2}$$

Since the exact distribution is difficult to specify, simulations were done on two variables,

$$Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{Var(\hat{\beta}_2)}}$$

Here, the denominator is simply the variance of $\hat{\beta}_2$ with the maximum likelihood estimates substituted for the real values. Further

$$Z_2 = \frac{\hat{\beta}_2 - c\beta_2}{\sqrt{\hat{Var}(\beta_2)}}$$

where c is a correction for the bias, i.e.

$$c = \frac{\hat{\sigma}_1^2 S_{x_1x_1}}{(\hat{\sigma}_1^2 + \sigma_{e_1}^2)(S_{x_1x_1} - n\sigma_{e_1}^2)}$$

These simulations showed that the interval [-1.96, 1.96] had a coverage frequency fairly close to 95%, which shows that a normal approximation may be useful. There was no discernable difference between the coverage frequencies of Z_1 and Z_2.

For the asymptotic properties of $\hat{\alpha}_2$ and $\hat{\beta}_2$, see Fuller (1987), p. 15.

Recall that we are interested in the estimates of the proportions of the variance of Y_2, which in terms of μ_1, μ_2, σ_1, σ_2 and ρ_{12} is

$$1 = (1 - \hat{\rho}_{12}^2) + \hat{\rho}_{12}^2.$$

It is possible to get approximate variance estimates for these proportions, by observing that the cross product matrix has a Wishart distribution (Mardia et al., 1979)

$$S = \begin{bmatrix} S_{x_1x_1} & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2x_2} \end{bmatrix} \sim W_2(\Sigma, n-1)$$

where

$$\Sigma = \begin{bmatrix} \sigma_1^2 + \sigma_{\epsilon_1}^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 + \sigma_{\epsilon_2}^2 \end{bmatrix}$$

This gives us that (Magnus and Neudecker, 1979):

$$E(S) = (n-1)\Sigma$$

and that

$$Var \begin{bmatrix} S_{x_1x_1} \\ S_{x_1x_2} \\ S_{x_2x_2} \end{bmatrix} = \left[\begin{array}{c|c|c} 2(n-1)(\sigma_1^2 + \sigma_{\epsilon_1}^2)^2 & 2(n-1)\rho_{12}\sigma_1\sigma_2 * & 2(n-1) \\ & (\sigma_1^2 + \sigma_{\epsilon_1}^2) & \rho_{12}^2\sigma_1^2\sigma_2^2 \\ \hline 2(n-1)\rho_{12}\sigma_1\sigma_2 * & (n-1)\{(\sigma_1^2 + \sigma_{\epsilon_1}^2) * & 2(n-1)\rho_{12}\sigma_1 * \\ (\sigma_1^2 + \sigma_{\epsilon_1}^2) & (\sigma_2^2 + \sigma_{\epsilon_2}^2) + \rho_{12}^2\sigma_1^2\sigma_2^2\} & \sigma_2 * (\sigma_2^2 + \sigma_{\epsilon_2}^2) \\ \hline 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2 & 2(n-1)\rho_{12}\sigma_1\sigma_2 * & 2(n-1) * \\ & (\sigma_2^2 + \sigma_{\epsilon_2}^2) & (\sigma_2^2 + \sigma_{\epsilon_2}^2)^2 \end{array}\right] \tag{2.4}$$

Hence we can conclude that

$$Var((1 - \hat{\rho}_{12}^2)) \approx F * V * F^T$$

where

$$F = [\frac{(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2}, \frac{-2(n-1)\rho_{12}\sigma_1\sigma_2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}}, \frac{(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}^2\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}}]$$

and V is the variance-covariance matrix given in (2.4). Approximate variances for the components of variance can be found analogously, and are given in Appendix A.

2.2.2 Three or More Stages

Maximum likelihood estimation

Maximum likelihood estimates do not have closed form expressions for models with more than two stages. The number of functionally independent parameters in an AR(1) k-stage process observed with measurement error is 3k-1 (two parameters for the initial stage and three more for every additional stage). The number of independent parameters in a general multivariate normal, however, is $k + \frac{k(k+1)}{2}$ (k parameters for the mean, and $\frac{k(k+1)}{2}$ variance-covariance parameters). In the case when k=2, these values are the same and the parameterization (μ_1, σ_1, α_1, β_1, σ_{A1}) is equivalent to (E(X_1), Var(X_1), E(X_2), Var(X_2),Cov(X_1, X_2)). For $k > 2$ the general multivariate normal has more parameters, and a one to one mapping between the two sets of parameters does not exist.

If, in the three stage case, we presume the existence of an underlying AR(1) process (1.3) for Y_1, Y_2, Y_3, but that what we observe is X_1, X_2, X_3, given by (2.1), we can parameterize the joint distributions of these variables as follows:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim MVN \left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \Sigma_y = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\rho_{23}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{12}\rho_{23}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix} \right)$$

and

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim MVN\left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \Sigma_{\mathbf{x}} = \begin{bmatrix} \sigma_1^2 + \sigma_{e_1}^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\rho_{23}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 + \sigma_{e_2}^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{12}\rho_{23}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 + \sigma_{e_3}^2 \end{bmatrix}\right) \tag{2.5}$$

In this case, the proportions of the variance of Y_3 can be expressed as

$$1 = (1 - \rho_{23}^2) + \rho_{23}^2(1 - \rho_{12}^2) + \rho_{23}^2\rho_{12}^2 \tag{2.6}$$

where the first term is the proportion of variation added at the third stage, the second term is the proportion of variation added at the second stage and transmitted to the third stage, and the third term is the proportion of variation transmitted from the first stage of the process.

The goal here is to estimate these three proportions based on independent observations (x_{1j}, x_{2j}, x_{3j}), $j = 1, \ldots, n$. This involves estimating the eight unknown parameters in the distribution (2.5).

The multivariate normal likelihood of (X_1, X_2, X_3) can be expressed as (Johnson et al., 1988)

$$\begin{aligned} L(\mu, \Sigma_x) &= \frac{1}{(2\pi)^{3n/2}|\Sigma_x|^{n/2}} \exp\{-\text{trace}[\Sigma_x^{-1}(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T)]/2 \\ &\quad -n/2(\bar{\mathbf{x}} - \mu)^T\Sigma_x^{-1}(\bar{\mathbf{x}} - \mu)\} \end{aligned}$$

where $\mathbf{x_j} = (x_{ij}, x_{2j}, x_{3j})^T$. If we write

$$\sum_{j=1}^{n}(\mathbf{x_j} - \bar{\mathbf{x}})(\mathbf{x_j} - \bar{\mathbf{x}})^T = \begin{bmatrix} S_{x_1x_1} & S_{x_1x_2} & S_{x_1x_3} \\ S_{x_1x_2} & S_{x_2x_2} & S_{x_2x_3} \\ S_{x_1x_3} & S_{x_2x_3} & S_{x_3x_3} \end{bmatrix} = S_{xx}$$

then the log-likelihood can be written as

$$\begin{aligned} l(\mu, \Sigma_x) &= -(3n/2)\log(2\pi) - (n/2)\log|\Sigma_x| \\ &-(1/2)trace\{\Sigma_x^{-1}S_{xx}\} - (n/2)(\bar{x} - \mu)^T\Sigma_x^{-1}(\bar{x} - \mu) \qquad (2.7) \end{aligned}$$

It is known (Johnson et al., 1988) that for any Σ, this likelihood is maximized with respect to μ by $\hat{\mu}_i = \bar{x}_i$, (i = 1,2,3). It remains, then, to determine the values of the three variance parameters and the two correlation parameters that will maximize the likelihood. There is not a closed form algebraic expression for any of these estimates, and they must be determined numerically. This is computer intensive and time consuming. If confidence intervals for variance components are also desired, additional computation will be needed. In the next section, we present a simpler method that performs very well.

Naive estimates

Simple estimates for a k-stage process can be obtained by using the two stage maximum likelihood estimates obtained earlier for each pair of consecutive stages.

This leads to the following estimates for the k-stage case:

$$\tilde{\mu}_i = \bar{x}_i, \qquad \tilde{\sigma}_i^2 = \frac{S_{x_i x_i}}{n} - \sigma_{e_i}^2 \qquad\qquad i = 1, 2, \ldots, k$$
$$\tilde{\rho}_{i,i+1} = \frac{S_{x_i x_{i+1}}}{\sqrt{(S_{x_i x_i} - n\sigma_{e_i}^2)(S_{x_{i+1} x_{i+1}} - n\sigma_{e_{i+1}}^2)}} \qquad\qquad i = 1, 2, \ldots, k-1. \quad (2.8)$$

Proving consistency of these estimators is straightforward. That the $\tilde{\mu}_i$ converge in probability to μ_i is an application of the weak law of large numbers. Similarly, for

$$\tilde{\sigma}_i^2 = \frac{S_{x_i x_i}}{n} - \sigma_{e_i}^2,$$

it is true that

$$\begin{aligned} \frac{S_{x_i x_i}}{n} &\rightarrow Var(X_i) \\ &= \sigma_i^2 + \sigma_{e_i}^2 \end{aligned}$$

Hence, $\tilde{\sigma}_i^2 \rightarrow \sigma_i^2$. Finally, since it is true that

$$\begin{aligned} \frac{S_{x_i x_{i+1}}}{n} &\rightarrow Cov(X_i, X_{i+1}) \\ &= \rho_{i,i+1}\sigma_i\sigma_{i+1} \end{aligned}$$

and that

$$\sqrt{(\frac{S_{x_i x_i}}{n} - \sigma_{e_i}^2)(\frac{S_{x_{i+1} x_{i+1}}}{n} - \sigma_{e_{i+1}}^2)} \rightarrow \sigma_i\sigma_{i+1}$$

we get

$$\begin{aligned}\tilde{\rho}_{i,i+1} &= \frac{\frac{S_{x_i x_{i+1}}}{n}}{\sqrt{(\frac{S_{x_i x_i}}{n} - \sigma^2_{\epsilon_i})(\frac{S_{x_{i+1} x_{i+1}}}{n} - \sigma^2_{\epsilon_{i+1}})}} \\ &\rightarrow \rho_{i,i+1}\end{aligned}$$

Note that the above calculations are general, and hold for any number of stages.

In simulations, it was found that the distributions of the estimated square roots of the individual variance proportions could be well approximated by normal distributions. This is also true for the estimates of the square roots of the variance components. Hence, it is useful to find confidence intervals in this metric. Calculating the asymptotic variance of these quantities can be done analogously to the method for the results shown in the previous section. See Appendix A for the approximate variances of the square root of the proportions and the components of variance at each of the three stages. An approximate 98% confidence interval can be computed using the formula

$$\text{estimated proportion} \pm 2.33\sqrt{\hat{Var}(\text{estimated proportion})} \qquad (2.9)$$

where $\hat{Var}$(estimated proportion) is found using the approximate formula and replacing the true values of the parameters by their estimates.

Parametric bootstrap calculations can also be used to get approximate confidence intervals. Once estimates for the parameters of the model have been found, these values can be used as the "true" values in generating N "bootstrap" samples of size n, the original sample size. Estimates of the variance components can be computed from each of the N samples, and confidence intervals calculated from them. For example, to get a 90% confidence interval, we could take N=99, and select the 5th and 95th values of the ordered estimates as the lower and upper limit for each variance component. For more details on parametric bootstrapping to compute confidence intervals, see Efron and Tibshirani (1986).

2.2.3 Simulation Results

We would like to know how the naive estimators compare to the maximum likelihood estimators. In addition, we want to know how well confidence intervals for variance components perform in terms of giving close to the stated coverage. These questions were addressed in a simulation study in which a three stage process was considered.

Since the values of the variances at the three stages do not change the properties of the estimators, they were set to always be one. For the same reason, the means at all three stages were set to zero. The variables that were manipulated were ρ_{12}, ρ_{23} and σ_{ϵ_i}. In this simulation, the measurement error was set to be the same at all stages, since this often occurs when the same characteristic is measured at each stage of the process. Three levels for each of ρ_{12} and ρ_{23} were used, $\sqrt{0.2}$, $\sqrt{0.5}$ and $\sqrt{0.8}$. These values were chosen because they provide a wide range of different values (see (1.4)) being added and transmitted through the process. Hence, values of the first variance proportion in (2.6) range from 0.2 to 0.8, while values of the second and third variance proportions range from 0.04 to 0.64. In this case, the true values of the proportions are the same as the components. Please see Table 2.1 for the exact quantities.

ρ_{12}	Component	$\rho_{23} = \sqrt{0.2}$	$\rho_{23} = \sqrt{0.5}$	$\rho_{23} = \sqrt{0.8}$
$\sqrt{0.2}$	First	0.80	0.50	0.20
	Second	0.16	0.40	0.64
	Third	0.04	0.10	0.16
$\sqrt{0.5}$	First	0.80	0.50	0.20
	Second	0.10	0.25	0.40
	Third	0.10	0.25	0.40
$\sqrt{0.8}$	First	0.80	0.50	0.20
	Second	0.04	0.10	0.16
	Third	0.16	0.40	0.64

Table 2.1: Actual values of the three variance components in (1.4) in the simulation runs for a three stage process.

Two levels of σ_{ϵ_i} were chosen, 0.1 and 0.3, for i = 1,2,3. At the high level of measurement error, the ratio $\sigma_{\epsilon_i}/\sigma_i$ is 30%. This level of measurement error would be unacceptable in some applications in industry; anything higher would call for a different measurement system. Note that even at the low measurement error level, and in the case of three stages, the bias in estimation resulting from ignoring the measurement error can be substantial. Bias here refers to the difference between the mean of a variance proportion estimate in large samples, as given in (2.3), and the true value of the variance proportion. Table 2.2 reproduces the variance proportions of Table 2.1 for each scenario and shows the bias that results if measurement error is ignored.

These combinations of three levels for ρ_{12} and ρ_{23} and two levels for σ_{ϵ_i} were used for an 18 run simulation. At each run, 99 samples (X_1, X_2, X_3) of 99 units were created using the given set of values of ρ_{12}, ρ_{23} and σ_{ϵ_i} as true parameters. For each sample, both the maximum likelihood estimates and the naive estimates

ρ_{23}	ρ_{12}	Comp.	Actual	$\sigma_\epsilon = 0.1$	$\sigma_\epsilon = 0.3$
				Bias	Bias
$\sqrt{0.2}$	$\sqrt{0.2}$	First	0.8	0.00394	0.0317
		Second	0.16	-0.00238	-0.0200
		Third	0.04	-0.00156	-0.0117
	$\sqrt{0.5}$	First	0.8	0.00394	0.0317
		Second	0.1	-3.88e-05	-0.00251
		Third	0.1	-0.00390	-0.0292
	$\sqrt{0.8}$	First	0.8	0.00394	0.0317
		Second	0.04	0.00230	0.0150
		Third	0.16	-0.00624	-0.0467
$\sqrt{0.5}$	$\sqrt{0.2}$	First	0.5	0.00985	0.0792
		Second	0.4	-0.00595	-0.0500
		Third	0.1	-0.00390	-0.0292
	$\sqrt{0.5}$	First	0.5	0.00985	0.0792
		Second	0.25	-9.71e-05	-0.00627
		Third	0.25	-0.00975	-0.0729
	$\sqrt{0.8}$	First	0.5	0.00985	0.0792
		Second	0.1	0.00576	0.0375
		Third	0.4	-0.0156	-0.117
$\sqrt{0.8}$	$\sqrt{0.2}$	First	0.2	0.0158	0.127
		Second	0.64	-0.00952	-0.0800
		Third	0.16	-0.00624	-0.0467
	$\sqrt{0.5}$	First	0.2	0.0158	0.127
		Second	0.4	-0.000155	-0.0100
		Third	0.4	-0.0156	-0.117
	$\sqrt{0.8}$	First	0.2	0.0158	0.127
		Second	0.16	0.00921	0.0600
		Third	0.64	-0.0250	-0.187

Table 2.2: Bias of simulation proportions when measurement error is ignored.

were found, and the three variance components were calculated. Then, 99 bootstrap samples were created using each set of estimates as the real parameters. The lowest and the highest values of the estimated variance components from these bootstrap samples were used to specify 98% confidence intervals for the components for each sample. Only 99 bootstrap samples were done here to keep the time limitations of the simulation feasible. In an industrial setting, computing more bootstrap samples, for example 1000, are recommended.

The results of the simulation are given in Tables 2.3 - 2.8. Table 2.3 shows the average value of the maximum likelihood estimates and the naive estimates for each run, for the first variance proportion. Also included are the standard deviation estimates of the run. Tables 2.4 and 2.5 show the same for the second and third variance proportions, respectively. Tables 2.6 - 2.8 gives the coverage frequencies of the bootstrap-based confidence intervals for both the maximum likelihood estimates and the naive estimates for all three variance components ("Raw") and the proportions ("Proportion"). Recall that this theoretical coverage frequency is 98%. No major discrepancies in coverage frequency are seen.

These results indicate that the performances of the naive estimates and the maximum likelihood estimates are virtually indistinguishable. In fact, the estimates are very close to each other in most cases. This can be seen in Figures 2.1 - 2.9, which show the naive estimates plotted against the maximum likelihood estimates for each of the variance components and for all runs. The top row of plots on these graphs is the run at the low measurement error level, and the bottom row of plots is the run at the high measurement error level. The Y=X line has been

ρ_{23}	ρ_{12}	Estimate	σ_ϵ =L	σ_ϵ = H	Real Value
L	L	Mle	0.792 (0.072)	0.790 (0.092)	0.8
		Naive	0.792 (0.072)	0.790 (0.092)	
	M	Mle	0.790 (0.079)	0.789 (0.079)	0.8
		Naive	0.790 (0.079)	0.790 (0.080)	
	H	Mle	0.783 (0.076)	0.792 (0.078)	0.8
		Naive	0.783 (0.076)	0.793 (0.080)	
M	L	Mle	0.503 (0.072)	0.491 (0.087)	0.5
		Naive	0.503 (0.072)	0.490 (0.087)	
	M	Mle	0.499 (0.074)	0.487 (0.079)	0.5
		Naive	0.499 (0.074)	0.486 (0.080)	
	H	Mle	0.499 (0.075)	0.500 (0.094)	0.5
		Naive	0.499 (0.075)	0.501 (0.096)	
H	L	Mle	0.204 (0.039)	0.196 (0.057)	0.2
		Naive	0.204 (0.039)	0.196 (0.056)	
	M	Mle	0.205 (0.042)	0.190 (0.058)	0.2
		Naive	0.205 (0.042)	0.189 (0.058)	
	H	Mle	0.205 (0.043)	0.197 (0.054)	0.2
		Naive	0.205 (0.044)	0.195 (0.054)	

Table 2.3: Average of 99 values of first component *of proportion estimates* for each run. The figures in brackets represent the estimated standard deviation for these values. Sample size is 99.

ρ_{23}	ρ_{12}	Estimate	σ_ϵ =L	σ_ϵ = H	Real Value
L	L	Mle	0.163 (0.054)	0.167 (0.074)	0.16
		Naive	0.163 (0.054)	0.167 (0.074)	
	M	Mle	0.104 (0.040)	0.101 (0.045)	0.10
		Naive	0.104 (0.040)	0.101 (0.045)	
	H	Mle	0.044 (0.017)	0.041 (0.021)	0.04
		Naive	0.044 (0.017)	0.041 (0.021)	
M	L	Mle	0.396 (0.063)	0.402 (0.074)	0.4
		Naive	0.396 (0.063)	0.402 (0.074)	
	M	Mle	0.253 (0.049)	0.256 (0.052)	0.25
		Naive	0.253 (0.049)	0.256 (0.052)	
	H	Mle	0.100 (0.022)	0.094 (0.027)	0.10
		Naive	0.100 (0.022)	0.094 (0.027)	
H	L	Mle	0.634 (0.064)	0.645 (0.076)	0.64
		Naive	0.634 (0.064)	0.646 (0.076)	
	M	Mle	0.401 (0.052)	0.393 (0.073)	0.40
		Naive	0.401 (0.052)	0.391 (0.073)	
	H	Mle	0.163 (0.034)	0.158 (0.049)	0.16
		Naive	0.164 (0.034)	0.157 (0.050)	

Table 2.4: Average of 99 values of second component *of proportion estimates* for each run. The figures in brackets represent the estimated standard deviation for these values. Sample size is 99.

ρ_{23}	ρ_{12}	Estimate	σ_ϵ =L	σ_ϵ = H	Real Value
L	L	Mle	0.045 (0.027)	0.043 (0.025)	0.04
		Naive	0.045 (0.027)	0.043 (0.026)	
	M	Mle	0.107 (0.043)	0.110 (0.044)	0.10
		Naive	0.107 (0.043)	0.109 (0.044)	
	H	Mle	0.173 (0.062)	0.167 (0.064)	0.16
		Naive	0.173 (0.062)	0.167 (0.066)	
M	L	Mle	0.100 (0.040)	0.107 (0.045)	0.10
		Naive	0.100 (0.040)	0.108 (0.045)	
	M	Mle	0.247 (0.062)	0.257 (0.066)	0.25
		Naive	0.247 (0.062)	0.258 (0.068)	
	H	Mle	0.401 (0.068)	0.406 (0.084)	0.40
		Naive	0.401 (0.068)	0.405 (0.086)	
H	L	Mle	0.162 (0.060)	0.159 (0.067)	0.16
		Naive	0.162 (0.060)	0.159 (0.066)	
	M	Mle	0.395 (0.067)	0.417 (0.080)	0.40
		Naive	0.395 (0.067)	0.419 (0.080)	
	H	Mle	0.632 (0.054)	0.645 (0.071)	0.64
		Naive	0.631 (0.054)	0.648 (0.074)	

Table 2.5: Average of 99 values of third component *of proportion estimates* for each run. The figures in brackets represent the estimated standard deviation for these values. Sample size is 99.

ρ_{23}	ρ_{12}	Estimate	Raw	Frequency	Proportional	Frequency
			σ_ϵ =L	σ_ϵ = H	σ_ϵ =L	σ_ϵ = H
L	L	Mle	98	91	96	95
		Naive	97	91	98	91
	M	Mle	97	98	95	98
		Naive	98	97	97	98
	H	Mle	91	97	95	98
		Naive	92	94	97	97
M	L	Mle	96	96	96	95
		Naive	94	94	99	98
	M	Mle	97	95	99	96
		Naive	97	95	99	97
	H	Mle	97	91	98	95
		Naive	97	91	95	97
H	L	Mle	96	95	97	95
		Naive	98	96	98	96
	M	Mle	95	93	95	96
		Naive	97	94	97	99
	H	Mle	97	96	95	97
		Naive	98	97	98	97

Table 2.6: Coverage frequency for first component for each run. Note that these figures are not given in percentages - they are the actual number of intervals that cover the real value out of 99 trials. (Coverage interval should be 98%). Sample size is 99.

ρ_{23}	ρ_{12}	Estimate	Raw	Frequency	Proportional	Frequency
			σ_ϵ =L	σ_ϵ = H	σ_ϵ =L	σ_ϵ = H
L	L	Mle	97	94	96	95
		Naive	98	95	99	95
	M	Mle	97	94	95	97
		Naive	98	96	97	97
	H	Mle	96	96	96	98
		Naive	98	96	97	98
M	L	Mle	96	97	95	97
		Naive	95	95	95	97
	M	Mle	97	95	95	98
		Naive	99	97	97	98
	H	Mle	97	96	97	96
		Naive	95	96	97	98
H	L	Mle	97	93	95	97
		Naive	97	92	96	97
	M	Mle	94	96	96	97
		Naive	96	93	97	99
	H	Mle	93	92	97	95
		Naive	95	94	97	94

Table 2.7: Coverage frequency for second component for each run. Note that these figures are not given in percentages - they are the actual number of intervals that cover the real value out of 99 trials. (Coverage interval should be 98%.) Sample size is 99.

ρ_{23}	ρ_{12}	Estimate	Raw	Frequency	Proportional	Frequency
			σ_ϵ =L	σ_ϵ = H	σ_ϵ =L	σ_ϵ = H
L	L	Mle	92	93	94	95
		Naive	96	95	98	96
	M	Mle	96	98	98	99
		Naive	99	98	99	98
	H	Mle	97	97	96	97
		Naive	99	97	96	98
M	L	Mle	97	97	98	97
		Naive	96	98	96	96
	M	Mle	94	96	94	96
		Naive	94	99	94	98
	H	Mle	97	98	98	93
		Naive	96	96	94	98
H	L	Mle	96	95	94	98
		Naive	98	95	97	95
	M	Mle	97	96	94	96
		Naive	96	97	98	95
	H	Mle	95	95	97	94
		Naive	93	97	97	96

Table 2.8: Coverage frequency for third component for each run. Note that these figures are not given in percentages - they are the actual number of intervals that cover the real value out of 99 trials. (Coverage interval should be 98%.) Sample size is 99.

added for reference. The naive estimates appear to be closest to the maximum likelihood estimates when the largest amount of variation is added at the end of the process. An interesting feature that can be seen is that regardless of the ρ_{12} or ρ_{23} values, the naive estimates are closer to the maximum likelihood estimates when the measurement error is low, as compared to when it's high. This is expected, since we know that the estimators are the same when there is no measurement error.

Overall, the data suggest that in the three stage case, the naive estimates can be substituted for the maximum likelihood estimates in many situations likely to be encountered in practice. There is little justification for spending time computing the maximum likelihood estimates, when the naive estimates can be found faster and without the use of optimization methods.

Other simulations were done to check the coverage frequencies of the confidence intervals given in equation (2.9) for various values of the true parameters. For a given set of true parameters, 1000 data sets of sample size 99 were generated. For each data set, the naive estimates of the square root of the variance components and proportions were found. Their approximate variances were calculated using these estimates, and a 98% confidence interval was computed using equation (2.9). Then, the coverage frequency for that set of real parameters was found by counting how many of the 1000 intervals actually contained the true parameters. See Tables 2.9 - 2.11 for these values. Overall, the coverage frequencies achieved were very close to 98%. This suggests that the approximate variance formulas given in the appendix are useful in finding confidence intervals, which further strengthens the argument for using the naive estimates.

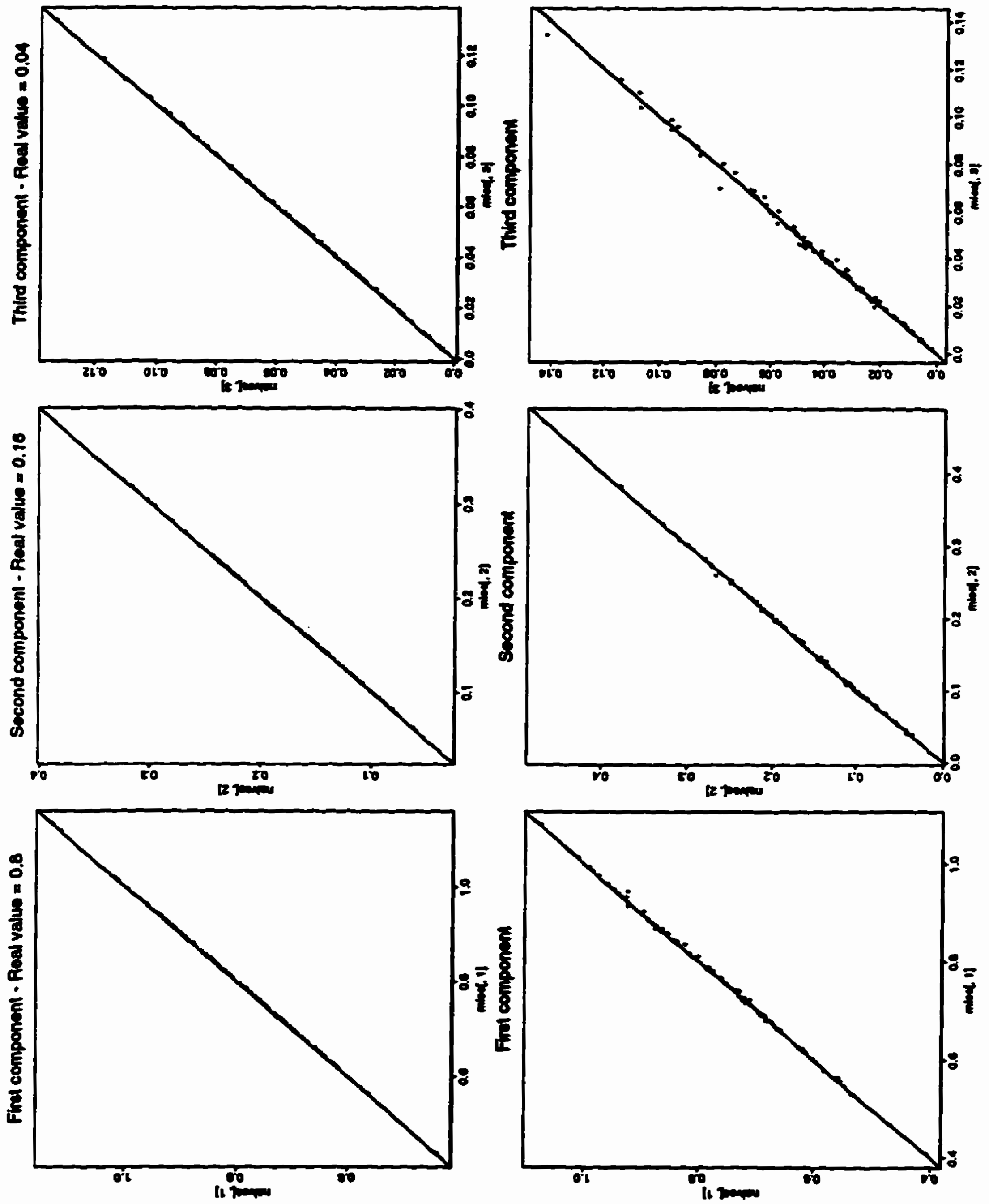

Figure 2.1: Figures for $\rho_{12} = \sqrt{0.2}$ and $\rho_{23} = \sqrt{0.2}$.

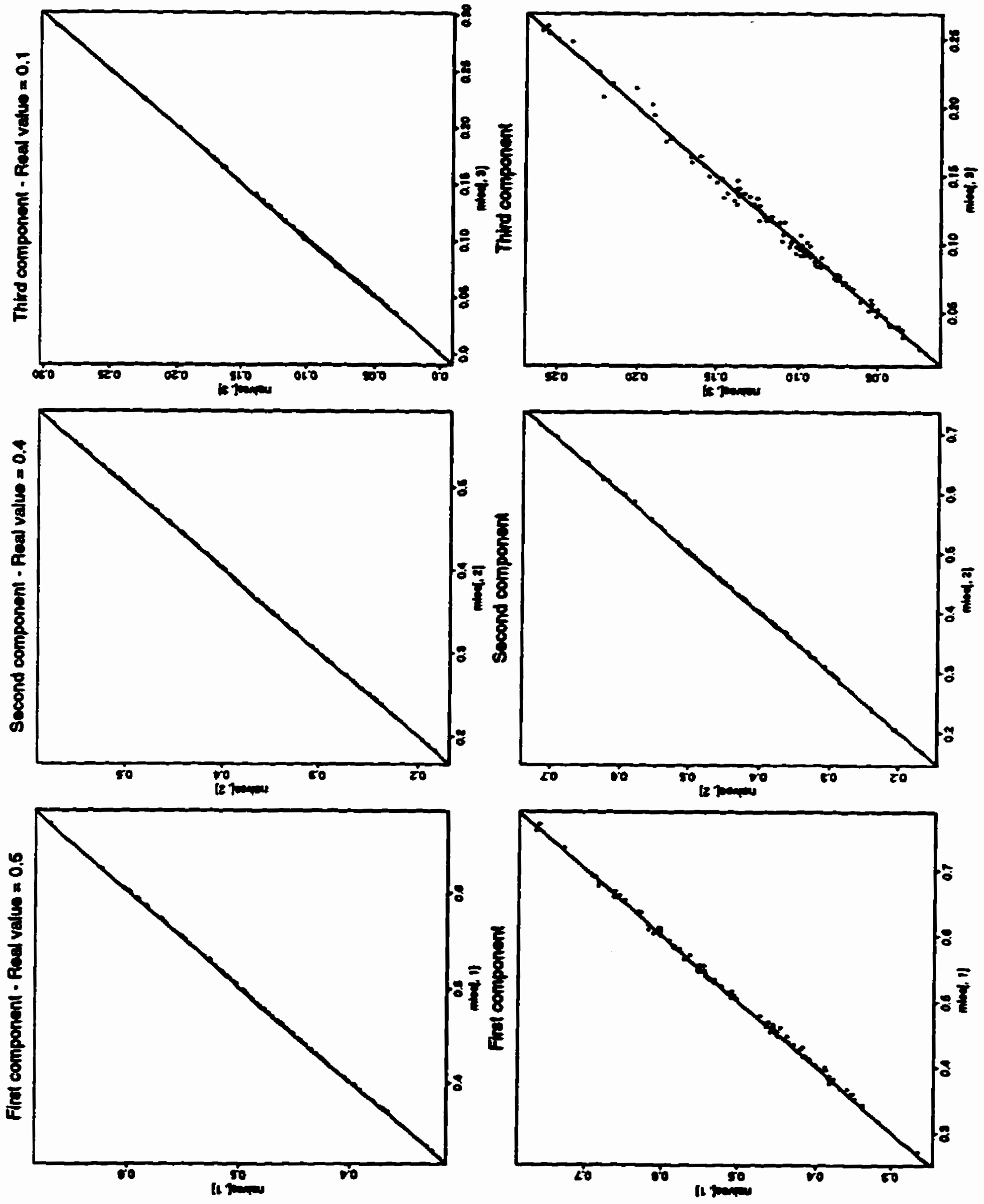

Figure 2.2: Figures for $\rho_{12} = \sqrt{0.2}$ and $\rho_{23} = \sqrt{0.5}$.

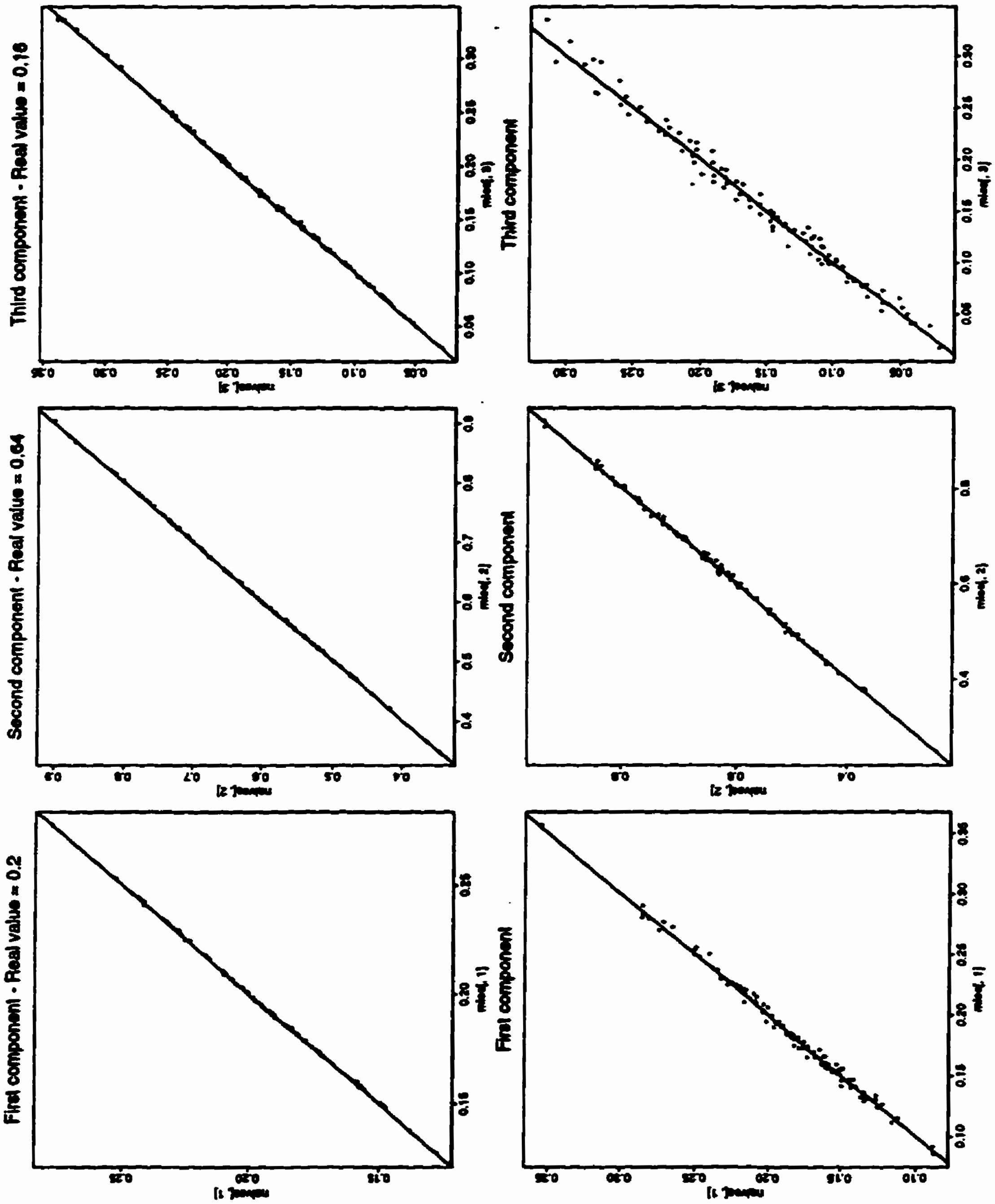

Figure 2.3: Figures for $\rho_{12} = \sqrt{0.2}$ and $\rho_{23} = \sqrt{0.8}$.

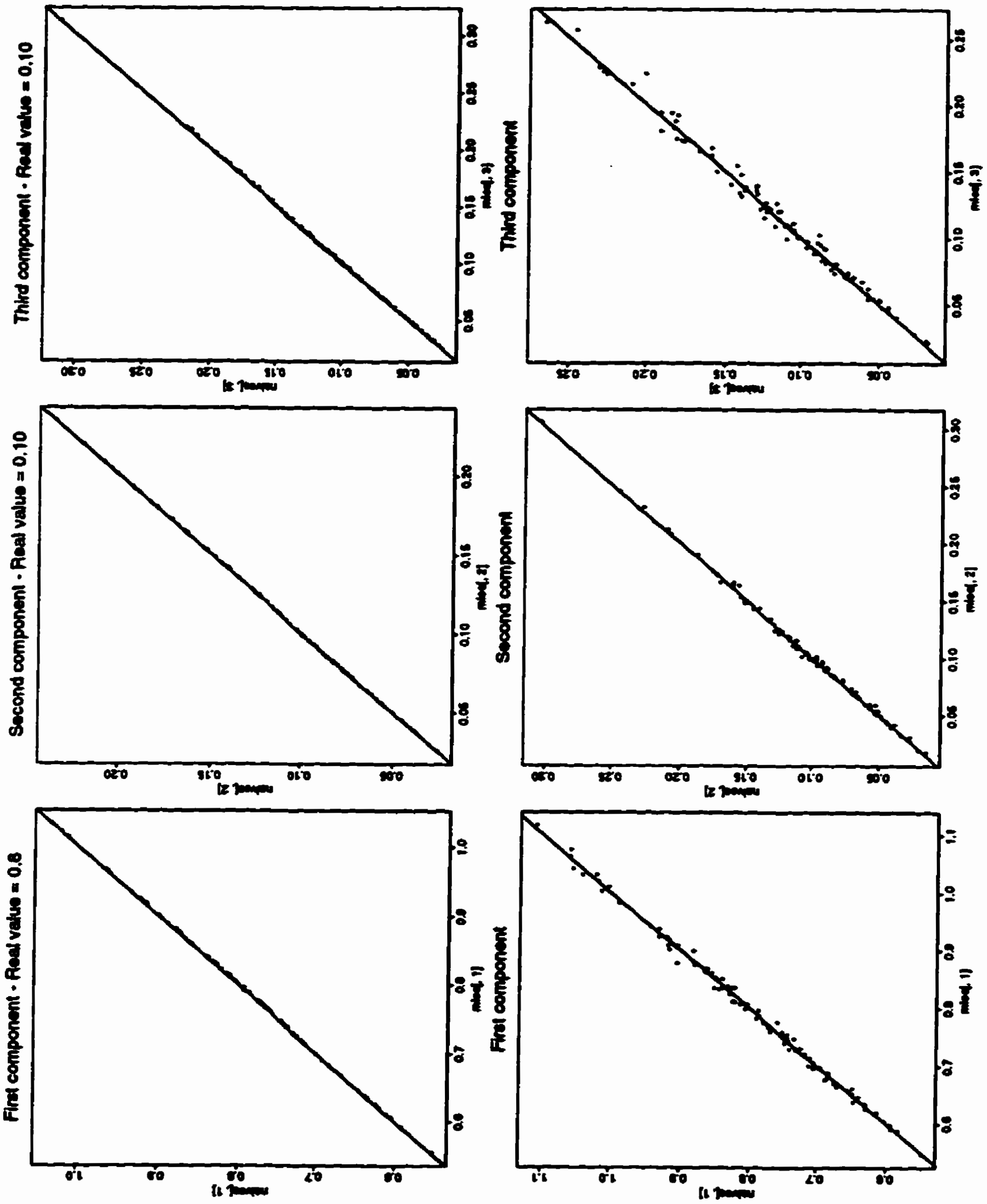

Figure 2.4: Figures for $\rho_{12} = \sqrt{0.5}$ and $\rho_{23} = \sqrt{0.2}$.

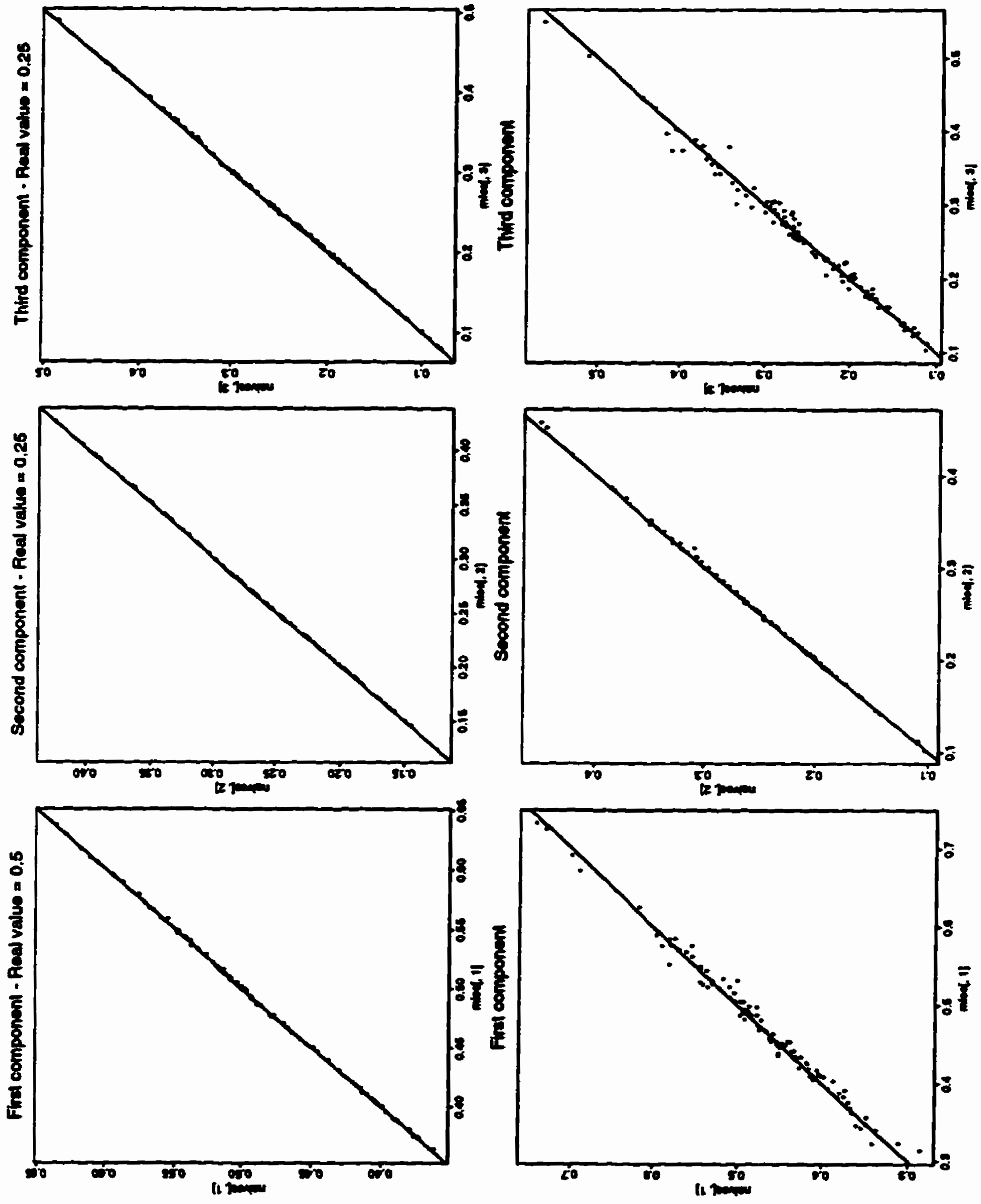

Figure 2.5: Figures for $\rho_{12} = \sqrt{0.5}$ and $\rho_{23} = \sqrt{0.5}$.

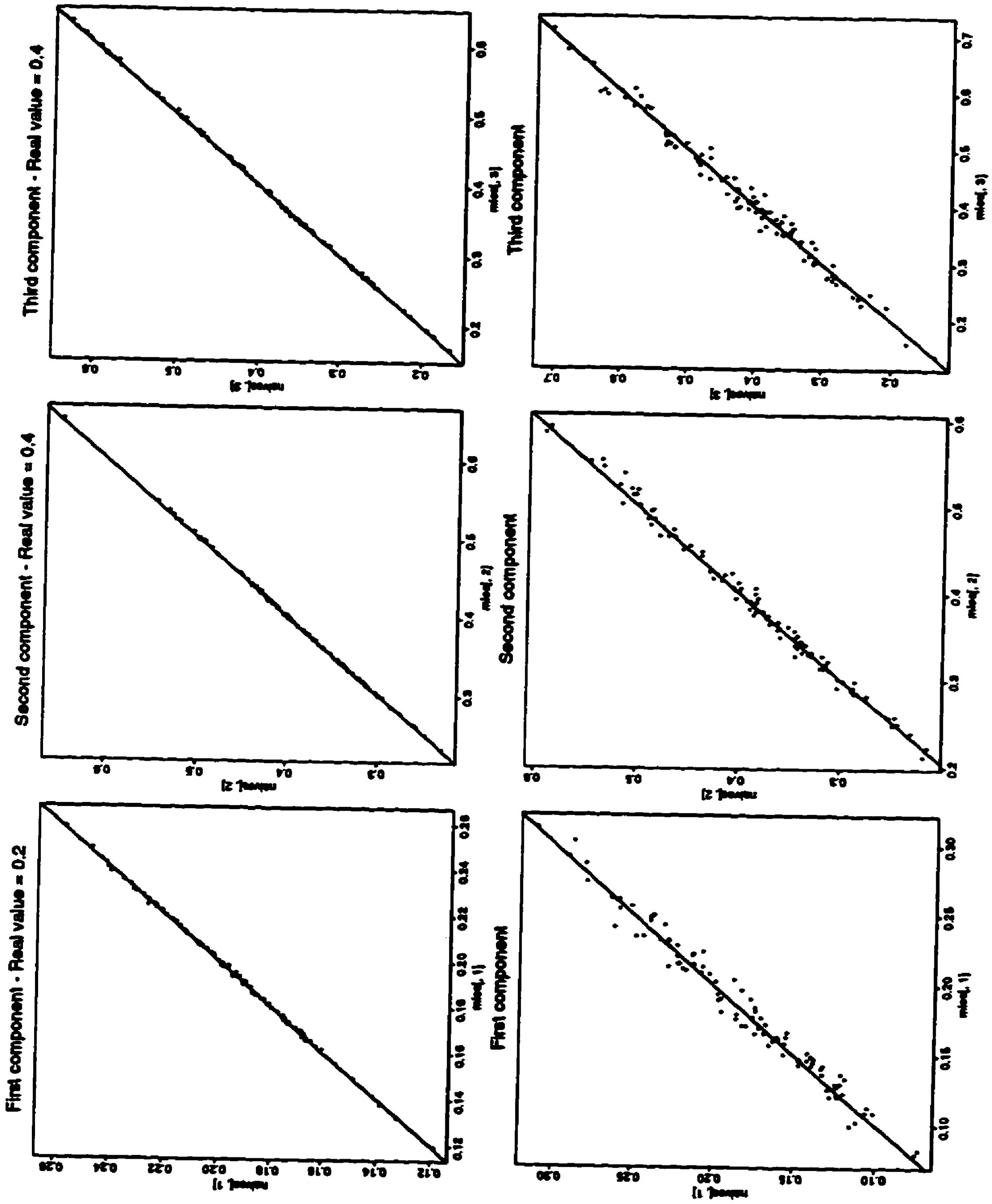

Figure 2.6: Figures for $\rho_{12} = \sqrt{0.5}$ and $\rho_{23} = \sqrt{0.8}$.

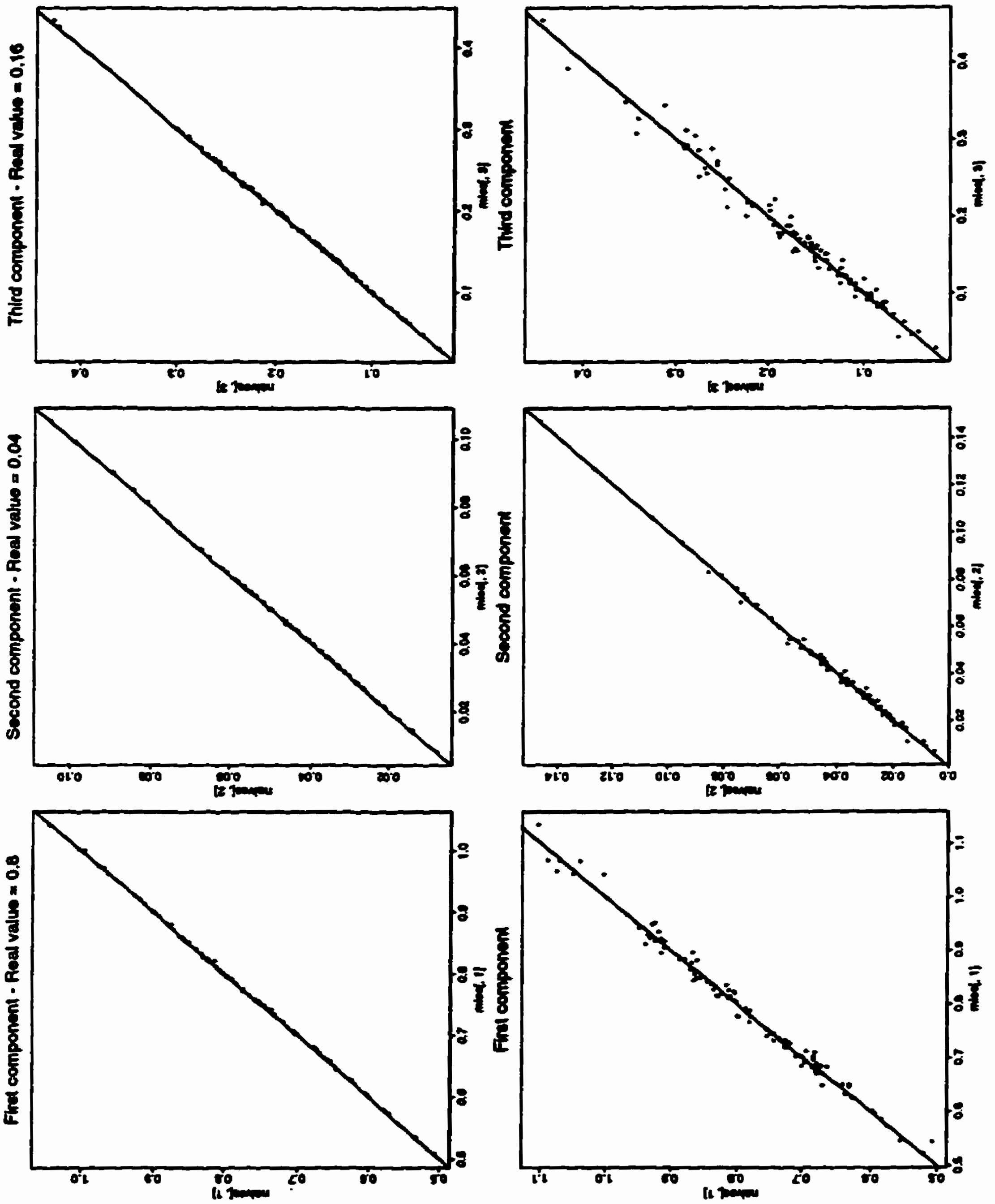

Figure 2.7: Figures for $\rho_{12} = \sqrt{0.8}$ and $\rho_{23} = \sqrt{0.2}$.

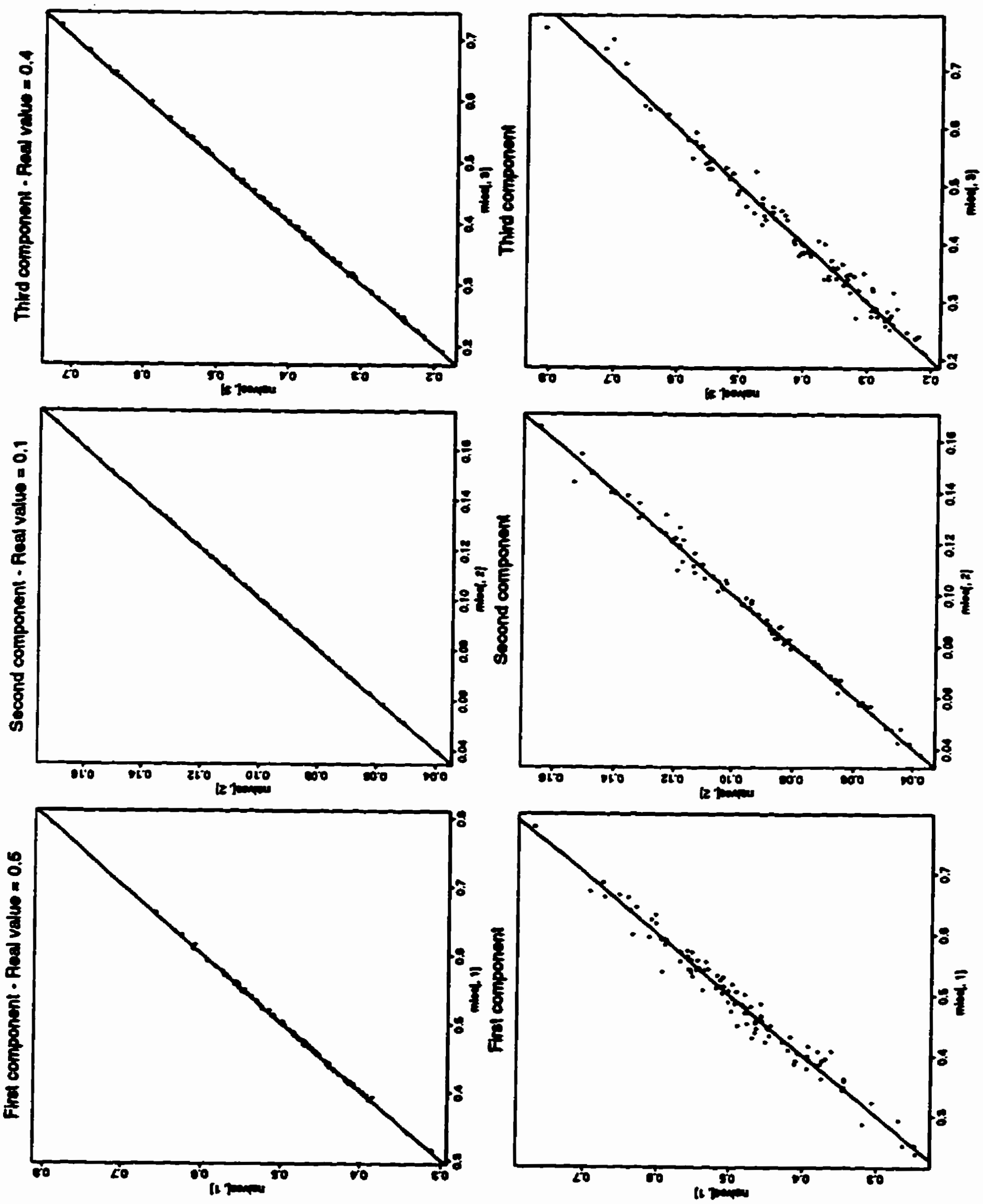

Figure 2.8: Figures for $\rho_{12} = \sqrt{0.8}$ and $\rho_{23} = \sqrt{0.5}$.

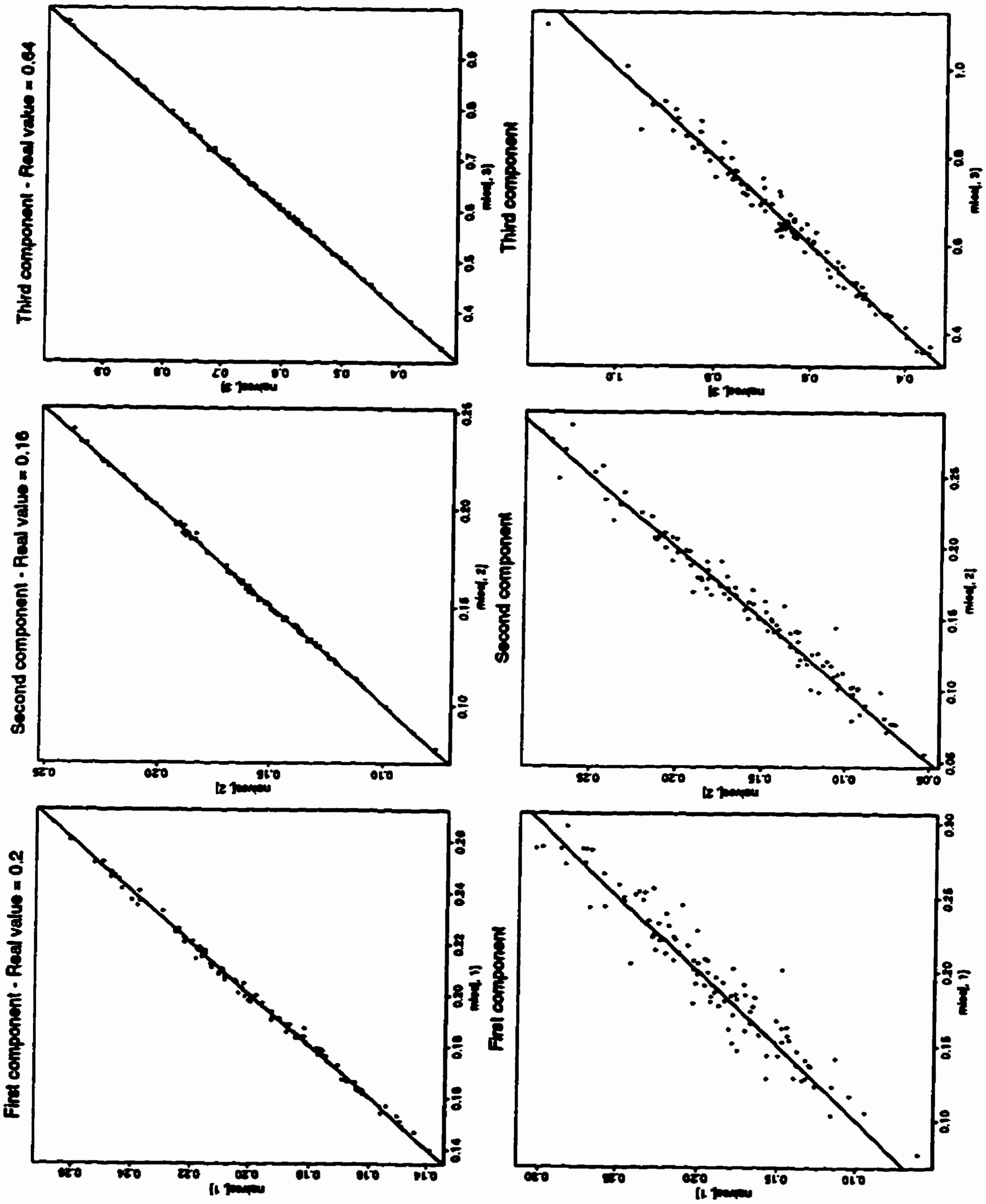

Figure 2.9: Figures for $\rho_{12} = \sqrt{0.8}$ and $\rho_{23} = \sqrt{0.8}$.

ρ_{23}	ρ_{12}	Component		Proportion	
		σ_ϵ =L	$\sigma_\epsilon =$ H	σ_ϵ =L	$\sigma_\epsilon =$ H
L	L	96.9	96.8	95.8	96.9
L	M	97.0	97.2	97.2	96.5
L	H	96.9	97.2	97.1	97.4
M	L	96.5	97.7	98.1	97.3
M	M	96.9	97.3	98.1	96.7
M	H	96.8	97.1	98.3	96.9
H	L	96.2	97.3	97.2	97.8
H	M	96.6	97.9	96.6	97.8
H	H	97.2	97.7	98.1	98.4

Table 2.9: Coverage frequencies of confidence intervals using approximate variance formulas for the square root of the first variance component. Numbers are percentages of 1000 simulations. Theoretical coverage frequency is 98%. Sample size is 99.

ρ_{23}	ρ_{12}	Component		Proportion	
		σ_ϵ =L	$\sigma_\epsilon =$ H	σ_ϵ =L	$\sigma_\epsilon =$ H
L	L	97.3	98.1	97.3	98.0
L	M	97.6	97.6	98.0	98.1
L	H	97.6	97.8	98.2	98.1
M	L	97.4	97.7	97.4	97.1
M	M	97.8	97.2	97.7	96.8
M	H	96.8	97.0	96.5	98.1
H	L	96.6	97.0	97.1	97.7
H	M	96.6	96.7	96.7	97.0
H	H	96.8	97.7	97.5	98.3

Table 2.10: Coverage frequencies of confidence intervals using approximate variance formulas for the square root of the second variance component. Numbers are percentages of 1000 simulations. Theoretical coverage frequency is 98%. Sample size is 99.

ρ_{23}	ρ_{12}	Component		Proportion	
		σ_ϵ =L	σ_ϵ = H	σ_ϵ =L	σ_ϵ = H
L	L	96.3	96.9	96.7	96.9
L	M	97.6	98.1	97.6	97.9
L	H	98.0	98.4	97.5	98.3
M	L	96.8	97.3	97.6	97.9
M	M	96.3	97.0	97.7	96.7
M	H	98.4	97.1	98.3	97.4
H	L	96.8	98.3	96.6	97.6
H	M	97.4	96.7	96.9	97.3
H	H	98.0	97.1	97.5	97.6

Table 2.11: Coverage frequencies of confidence intervals using approximate variance formulas for the square root of the third variance component. Numbers are percentages of 1000 simulations. Theoretical coverage frequency is 98%. Sample size is 99.

It seems that both the bootstrapping and the approximate variance formulas are satisfactory methods of finding confidence intervals for the sample size considered here (n=99). For small sample sizes, however, one might expect the bootstrap method to be more accurate.

2.3 Model Checking

It is important to check whether observed data are consistent with an AR(1) process with known measurement error. As indicated in (2.5), this model implies that the observed measurements follow a multivariate normal distribution.

As a first step in evaluating the multivariate normal assumption, the normality of the univariate marginal distributions should be checked, as for the AR(1) model. If the marginal distributions do not seem normal, then the multivariate normal assumption can be rejected. If they do seem normal, however, the assumption of the linearity of the conditional means should be verified. That is, plots should be made of all Y_i vs Y_j for $i > j$. Again, if this assumption is contradicted, the multivariate normal assumption should be rejected.

Residual plots can also be done for the AR(1) model with measurement error. To see how, write

$$\begin{aligned} Y_i &= \alpha_i + \beta_i Y_{i-1} + \mu_i \\ X_i &= Y_i + \epsilon_i \end{aligned}$$

Since we are assuming independence of μ_i and ϵ_i, of all μ_i, μ_j when $i \neq j$, and similarly for ϵ_i, we get that

$$X_i = \alpha_i + \beta_i X_{i-1} - \beta_i \epsilon_{i-1} + \epsilon_i + \mu_i$$

Let

$$R_i = -\beta_i \epsilon_{i-1} + \epsilon_i + \mu_i$$

Using the calculated estimates for α_i and β_i, we can estimate R_i by

$$\hat{R}_i = X_i - \hat{\alpha}_i - \hat{\beta}_i X_{i-1}$$

These estimated residuals should be independent of all previous values, i.e. X_1, X_2, ..., X_{i-2}. Hence, plots of these residuals against these stages should reveal no discernible trends.

Other, more formal tests can be applied to test for multivariate normality (Looney, 1995).

We can test the adequacy of the AR(1) model or the AR(1) with measurement error model within a normal model via a likelihood ratio test, as follows. Under a general multivariate normal structure, the maximum likelihood estimates are (Johnson, 1988)

$$\hat{\mu} = \bar{X} \qquad \hat{\Sigma} = \frac{S_{xx}}{n} \tag{2.10}$$

and so the maximized log-likelihood takes the form

$$\begin{aligned} l(\hat{\Omega}) &= -\frac{n}{2} log|\frac{S_{xx}}{n}| - \frac{np}{2} log(2\pi) - \frac{1}{2}\text{trace}\{(\frac{S_{xx}}{n})^{-1} S_{xx}\} \\ &= -\frac{n}{2} log|\frac{S_{xx}}{n}| - \frac{np}{2} - \frac{np}{2} log(2\pi) \end{aligned}$$

Under the constraint of being an AR(1) process with measurement error, the maximized log-likelihood takes the form

$$l(\hat{\omega}) = -\frac{n}{2} log|\hat{\Sigma}_x| - \frac{1}{2}\text{trace}\{\hat{\Sigma}_x^{-1} S_{xx}\} - \frac{np}{2} log(2\pi) \tag{2.11}$$

where Σ_x is of the form given in (2.5), and an estimate of it has been found by optimizing (2.7). From the theory of the likelihood ratio test,

$$-2(l(\hat{\omega}) - l(\hat{\Omega})) \approx \chi^2_{\frac{k(k-3)}{2}+1}$$

In simulations for the case k=3, it was found that the distribution of the statistic given above could not be distinguished from χ^2_1, for sample sizes as small as 30.

This was true even when $l(\hat{\omega})$ was approximated by evaluating (2.11) using naive estimates. This means that a simple approximate test can be carried out for the AR(1) model with measurement error without needing to compute the maximum likelihood estimates for the model.

Using the above likelihood expression, the deviance residuals can be examined to see if any observations are particularly influential. See, for example, Williams (1987).

2.4 Uncertainty in the Measurement Error Variance

At this point, we will discuss how the results given above can be modified to take into account uncertainty in the measurement error variance.

We will assume that the data taken to estimate the measurement error are independent of the process data. Further, we will assume that out of the experiment, we have an estimate of σ_ϵ^2, $\hat{\sigma}_\epsilon^2$, such that

$$\begin{aligned} E(\hat{\sigma}_\epsilon^2) &= \sigma_\epsilon^2 \\ \text{and} \qquad Var(\hat{\sigma}_\epsilon^2) &= V_\epsilon \end{aligned}$$

If the estimate is not unbiased, minor adjustments can be made to the following procedures.

The naive estimates described earlier can then be modified by replacing the known measurement error with the estimate given above. The maximum likelihood estimates can be computed also by replacing the estimate above with the known measurement error in the likelihood (2.7).

The delta method can be used here to get approximate variance formulas for the proportions and components. For example, in a two stage process where the measurement error is the same at both stages, the approximate variance formula can be computed by noting that

$$1-\tilde{\rho}_{12}^2 = 1 - \frac{S_{x_1x_2}^2}{(S_{x_1x_1} - n\hat{\sigma}_\epsilon^2)(S_{x_2x_2} - n\hat{\sigma}_\epsilon^2)}$$

is a function of four random variables, $S_{x_1x_1}$, $S_{x_1x_2}$, $S_{x_2x_2}$ and $\hat{\sigma}_\epsilon^2$, and that the last is independent of the first three. When the gradient is taken with respect to each of these variables, and the expected values of $S_{x_1x_1}$, $S_{x_1x_2}$, $S_{x_2x_2}$ and $\hat{\sigma}_\epsilon^2$ are substituted into these expressions, we get that the resulting vector is

$$\begin{aligned}
F \;=\; & [\frac{(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{\{(n-1)\sigma_1^2-\sigma_\epsilon^2\}^2\{(n-1)\sigma_2^2-\sigma_\epsilon^2\}}, \\
& \frac{-2(n-1)\rho_{12}\sigma_1\sigma_2}{\{(n-1)\sigma_1^2-\sigma_\epsilon^2\}\{(n-1)\sigma_2^2-\sigma_\epsilon^2\}}, \\
& \frac{(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{\{(n-1)\sigma_1^2-\sigma_\epsilon^2\}\{(n-1)\sigma_2^2-\sigma_\epsilon^2\}^2}, \\
& \frac{-n(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{\{(n-1)\sigma_1^2-\sigma_\epsilon^2\}\{(n-1)\sigma_2^2-\sigma_\epsilon^2\}} \\
& *\{\frac{1}{\{(n-1)\sigma_1^2-\sigma_\epsilon^2\}}+\frac{1}{\{(n-1)\sigma_2^2-\sigma_\epsilon^2\}}\}]
\end{aligned}$$

.

Further, the variance-covariance matrix of these variables is

$$\Sigma = \left[\begin{array}{c|c|c|c} 2(n-1)(\sigma_1^2+\sigma_\epsilon^2)^2 & \begin{array}{c} 2(n-1)\rho_{12}\sigma_1\sigma_2* \\ (\sigma_1^2+\sigma_\epsilon^2) \end{array} & \begin{array}{c} 2(n-1) \\ \rho_{12}^2\sigma_1^2\sigma_2^2 \end{array} & 0 \\ \hline \begin{array}{c} 2(n-1)\rho_{12}\sigma_1\sigma_2* \\ (\sigma_1^2+\sigma_\epsilon^2) \end{array} & \begin{array}{c} (n-1)\{(\sigma_1^2+\sigma_\epsilon^2)* \\ (\sigma_2^2+\sigma_\epsilon^2)+\rho_{12}^2\sigma_1^2\sigma_2^2\} \end{array} & \begin{array}{c} 2(n-1)\rho_{12}\sigma_1* \\ \sigma_2(\sigma_2^2+\sigma_\epsilon^2) \end{array} & 0 \\ \hline 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2 & \begin{array}{c} 2(n-1)\rho_{12}\sigma_1\sigma_2* \\ (\sigma_2^2+\sigma_\epsilon^2) \end{array} & \begin{array}{c} 2(n-1) \\ (\sigma_2^2+\sigma_\epsilon^2)^2 \end{array} & 0 \\ \hline 0 & 0 & 0 & V_\epsilon \end{array}\right]$$

Hence, the approximate variance of this variance proportion can be calculated as

$$Var(1-\tilde{\rho}_{23}^2) \approx F * \Sigma * F^T$$

Similar calculations can be done for the variance components, and in the case of more stages.

Note that the method described here is not the only method of collecting data on measurement error. For instance, such data can be collected while gathering the process data, simply by measuring each part twice. Methods of analysis in this case have yet to be investigated.

2.5 Missing Data

Missing data can be handled in the situation when the data adhere to an AR(1) model with measurement error. The procedure used to do this is a generalized version of that used for the simple AR(1) model.

In this case, the observed X's are treated as a general multivariate normal, as given in the three stage case by (2.5). In the EM algorithm, all the data are used to estimate the expectations of the E step. A numerical optimization is then required for the M step. Details of this calculation are given in Little and Rubin (1987), p. 142, and have been used in Hamada and Lawless.

The problem with the procedure described above is that it is very computationally intensive. Fong and Lawless (1996) use a Kalman filtering approach to facilitate the use of the EM algorithm, and find this approach to be more efficient.

The naive estimates can also be used to estimate the parameters in the case where some data are missing. The method of doing this would be simply be to estimate the parameters over the data that are available. For example, in the case of estimating a variance parameter, we would use all the data that are available for that stage and estimate the variance as the sum of squares of that data divided by the amount of data and subtract the measurement error variance. Correlation parameters between adjacent stages could similarly be calculated over all the data in which both stages were observed. This approach is much simpler than implementing the EM algorithm described above.

2.6 Piston Example

For the piston example described in the previous chapter, the variance of the final stage will be partitioned using an AR(1) model with measurement error. To simplify the process, it will be reduced to three stages, namely Y_4, Y_5, and Y_7. It was believed that these stages were not changing the diameter of interest at all. Further, only one of the diameters will be considered here, the diameter at a height of 4 mm. The known measurement error standard deviation is approximately $5*10^{-4}$ mm, or 0.5 microns, at each stage. This gives an estimated ratio of $\frac{\sigma_\epsilon}{\sigma_3} = 22\%$.

Engineering knowledge of the process indicated that the normal AR(1) model with measurement error should adequately describe it. Various model checks were used to determine the adequacy of this model. The data are essentially discrete over the range in which they were measured, which affects the normality assumption. Still, the QQ plots at each stage did not reveal any significant departures. The deviance residuals of three pistons proved to be particularly influential. Scatter plots of pairs of measurements also showed these three points as outliers, and the sequence plot revealed that this might be because their second measurements were faulty (Y_5). Hence, these outliers were removed from the subsequent analysis, although some investigation should be done to seek causes for why these particular pistons may have differed from the rest. See figures 2.10 - 2.12 for plots of the data. Note that Figure 2.12 is not a good example of a sequence plot, since the individual items are difficult to trace due to the discrete nature of the data. Still, it is apparent from this plot why the three pistons specified are outliers.

The goal of this study is to determine how the variation at the final stage of the process can be attributed to variation transmitted from upstream. When the measurement error is ignored, the proportions of variance contributed according to the AR(1) model are 0.256 at the third stage, 0.244 at the second stage, and 0.500 at the first stage. Using the naive estimates introduced in section three and the

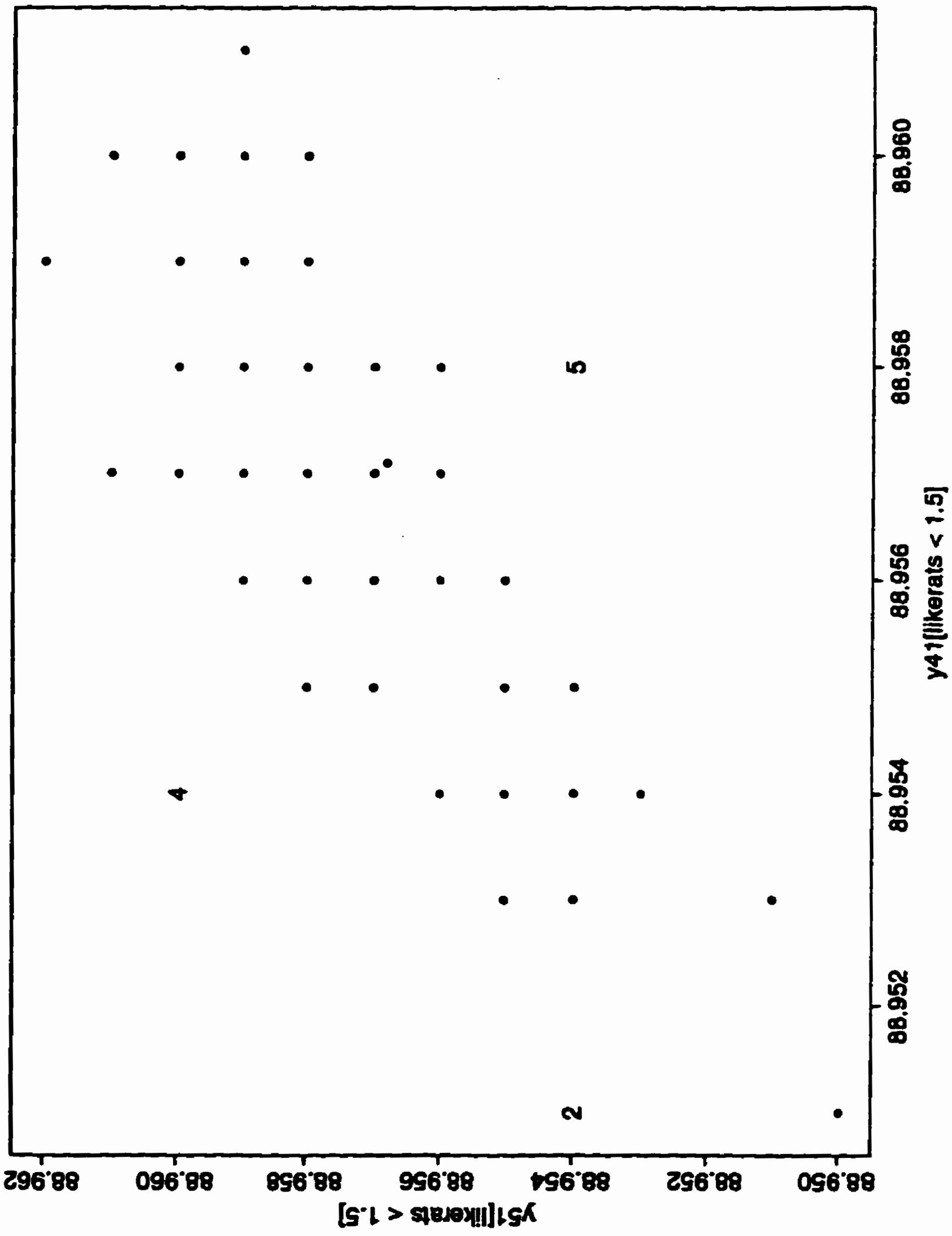

Figure 2.10: Second stage of piston data plotted against the first stage. The numeric values indicate the three outliers.

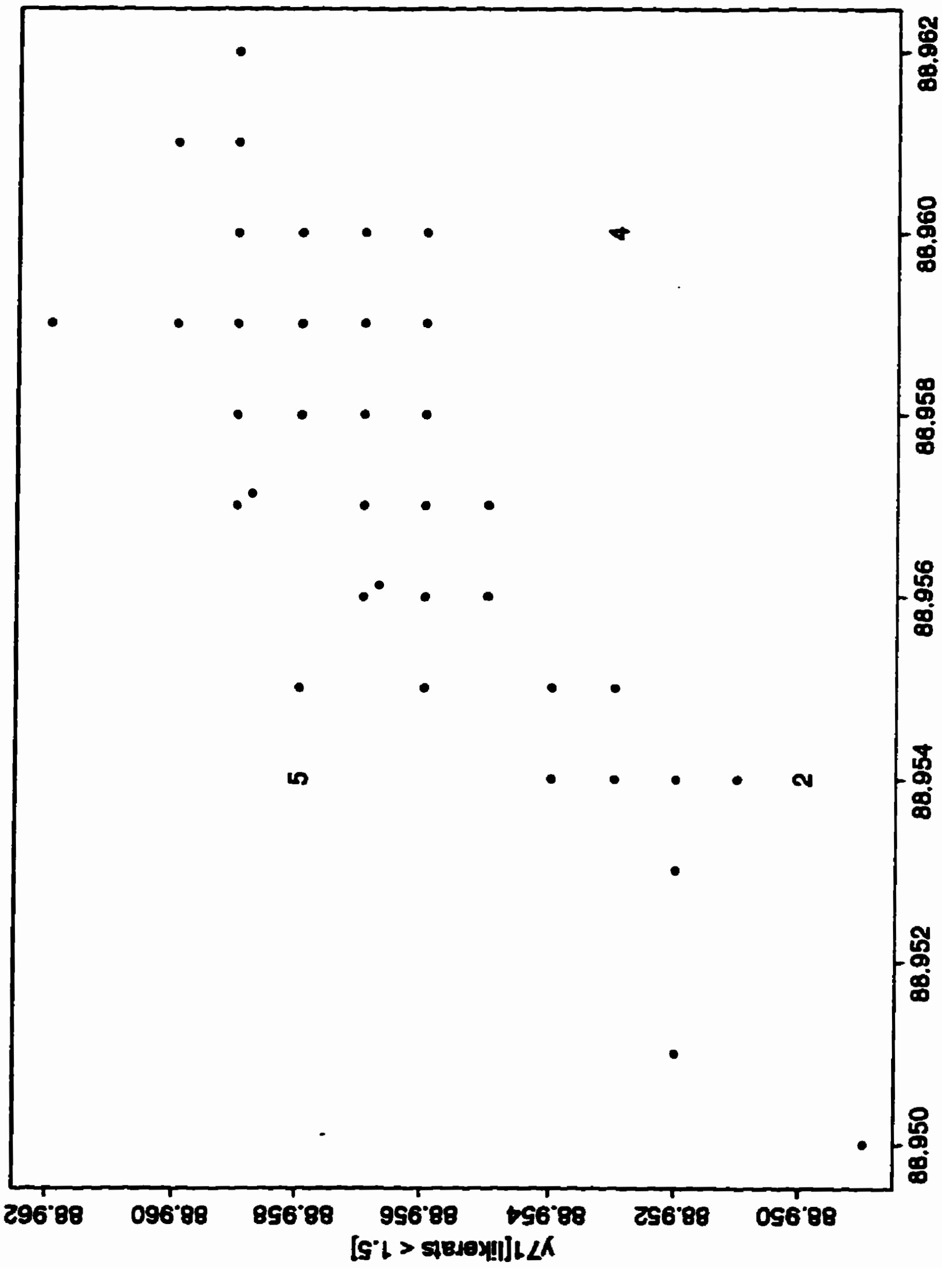

Figure 2.11: Third stage of piston data plotted against the second stage. The numeric values indicate the three outliers.

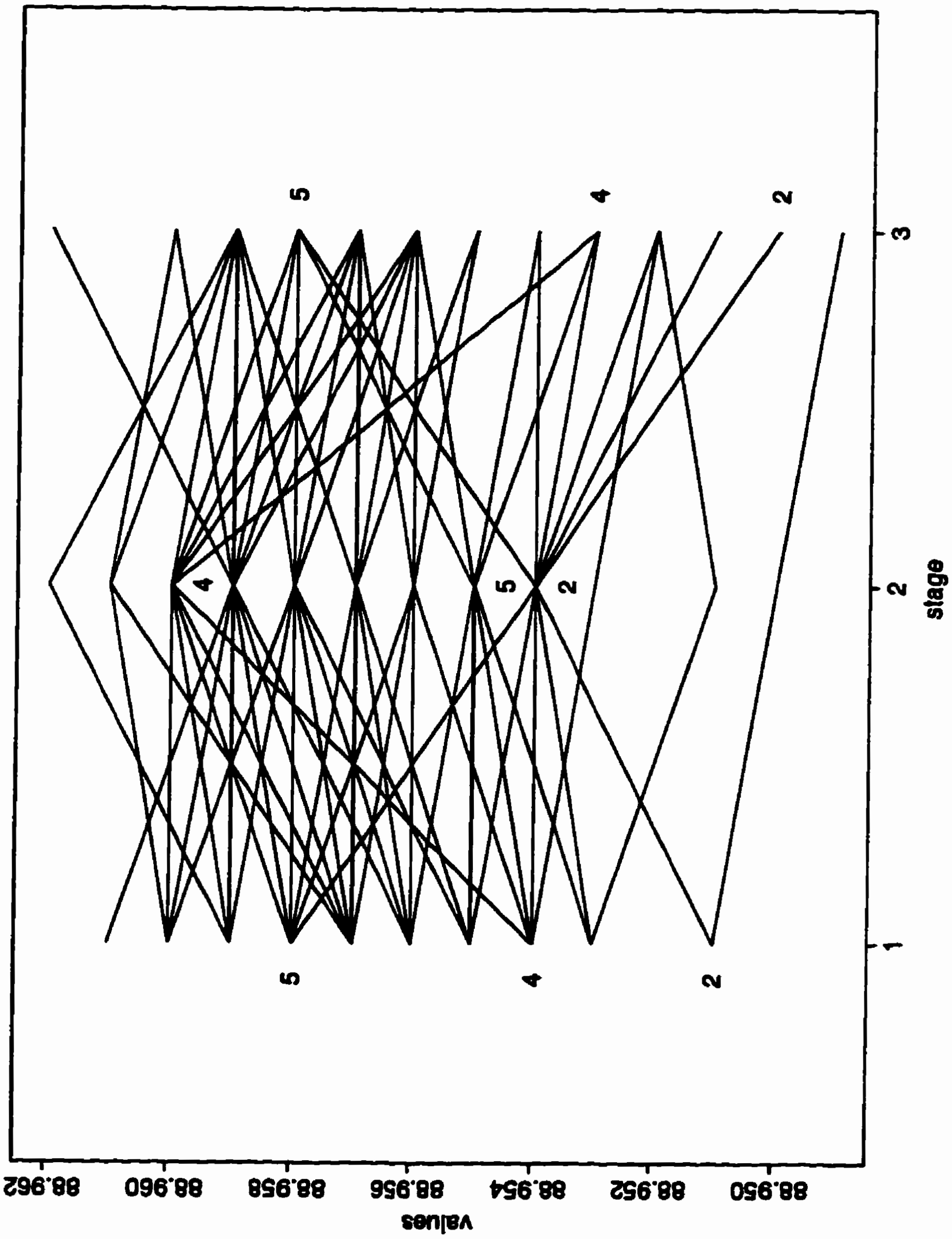

Figure 2.12: Sequence plot of the three stages of the piston data. The numeric values indicate the three outliers.

known measurement error variance, however, we find instead that the proportion of variance contributed is 0.181 at the third stage, 0.206 at the second stage and 0.613 at the first stage. Taking into account the measurement error not only gives a more accurate impression of where the variation is coming from, but also allows a more accurate interpretation of how an intervention in the process will affect the variation at the final stage.

Both the bootstrap technique and the approximate variance method described earlier were used to find 98% confidence intervals for these proportions. In the first case, 1000 bootstrap samples were simulated using the naive estimates as the real values, and new estimates for the proportions were computed. The 10th and the 990th ordered values were then found to give the following confidence intervals

$$\begin{aligned}
\text{Prop. from 3rd stage}: &\quad (0.088, 0.300)\\
\text{Prop. from 2nd stage}: &\quad (0.105, 0.322)\\
\text{Prop. from 1st stage}: &\quad (0.466, 0.750)
\end{aligned}$$

In the case of the approximate variance method, the naive estimates were substituted into the equations in Appendix A and (2.9) to give the confidence intervals

$$\begin{aligned}
\text{Prop. from 3rd stage}: &\quad (0.092, 0.299)\\
\text{Prop. from 2nd stage}: &\quad (0.115, 0.323)\\
\text{Prop. from 1st stage}: &\quad (0.475, 0.769)
\end{aligned}$$

The two sets of confidence intervals agree well. The main conclusion is that the first stage contributes most of the variation.

The analysis done using the maximum likelihood estimates yielded the same conclusions as that done with the naive estimates.

Finally, it should be mentioned that the likelihood ratio test given previously was carried out, and was found to yield

$$l(\hat{\omega}) = 1423.91$$

under the assumption of the AR(1) model with measurement error. Under the full model,

$$l(\hat{\Omega}) = 1424.83$$

Using the approximating chi-square distribution on one degree of freedom, this gives a p-value of 0.173, indicating no reason to reject the measurement error model. The likelihood ratio test was also done for the AR(1) model without measurement error, and was found to give a likelihood of 1421.54, which when compared to the full model gives a p-value of 0.010, suggesting that this model does not describe the data adequately.

We conclude that while more than half of the variation at the final stage is transmitted from the first stage, 40% of it still comes from subsequent stages. This somewhat contradicts previous knowledge of the process, and provides new opportunities for variation reduction.

Chapter 3

Cross-sectional and Longitudinal Data

The methods described thus far to deal with the variation analysis problem have focused solely on one data collection scheme, namely tracking items through the process. As has been mentioned earlier, this type of data collection is expensive. Furthermore, it often happens that large amounts of data are available after each stage of the process, either because measurement systems gather these data in routine monitoring, or simply because they are cheap to collect. Of significant interest, then, is to determine how this type of data can be used in the partitioning of the variance at the last stage of the process.

Statistically, this is a missing data problem, although we prefer to think of the cross-sectional data as supplemental data. Clearly, the variation analysis problem cannot be handled without some longitudinal data, since estimates of the correlation between stages of the process would not be available. We will consider situations, then, in which some longitudinal information is available, but this is augmented with data that have been sampled after each stage.

This chapter is divided into two sections, an analysis section and a design section. In the former, we discuss issues of estimation given data of this sort, first for two stages, and then for three or more stages. An AR(1) model is assumed. We propose two naive sets of estimators, and give their properties. Some simulation results are also given. Also discussed is how these estimators could be modified to include measurement error. In the design section of this chapter, we discuss the issue of how much information is available in the cross-sectional data.

3.1 Analysis

3.1.1 Two Stages

Estimation

Consider a two stage process, in which n observations are made on items tracked through both stages of the process, m items are sampled at the first stage, and l items are sampled at the second stage. The three groups of observations are denoted as S12, S1 and S2, respectively. Since this is a two stage process, it is necessarily AR(1). The goal is to determine the variance components of interest, $\sigma_2^2(1-\rho_{12}^2)$, and $\sigma_2^2\rho_{12}^2$.

Various methods of estimating these components come to mind. The first method is simply to ignore the supplemental data, and to estimate the variance components from only the longitudinal data, as given by the estimates in (1.6).

We will subsequently refer to these estimates as MLES(S12). These estimates are considered to be a baseline against which to measure other estimation techniques.

The second method that could be used is to compute the maximum likelihood estimates from these data. We can write down the likelihood here as

$$\begin{aligned} L(\theta) &= \Pi^{S12} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\{\frac{-1}{2(1-\rho_{12}^2)}[(\frac{Y_{1i}-\mu_1}{\sigma_1})^2 - \\ & 2\rho_{12}(\frac{Y_{1i}-\mu_1}{\sigma_1})(\frac{Y_{2i}-\mu_2}{\sigma_2}) + (\frac{Y_{2i}-\mu_2}{\sigma_2})^2]\} \, \Pi^{S1} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\{\frac{-(Y_{1i}-\mu_1)^2}{2\sigma_1^2}\} \\ & \Pi^{S2} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\{\frac{-(Y_i-\mu_2)^2}{2\sigma_2^2}\} \qquad (3.1) \end{aligned}$$

The maximum likelihood estimates, even in the simple two stage case, must be computed numerically. These estimators will be referred to as MLES(S12-S2).

Fairly obvious naive estimators can be constructed. The first set will make use of all the data available at any stage to estimate the marginal parameters at that stage, and only the longitudinal data to estimate the correlations. Hence, we have

$$\begin{aligned} \tilde{\mu}_1(S12,S1) &= \sum_{S12,S1} \frac{Y_{1i}}{(m+n)} \\ \tilde{\mu}_2(S12,S2) &= \sum_{S12,S2} \frac{Y_{2i}}{(l+n)} \\ \tilde{\sigma}_1(S12,S1)^2 &= \frac{1}{(m+n)} \sum_{S12,S1} (Y_{1i} - \tilde{\mu}_1(S12,S1))^2 \\ \tilde{\sigma}_2(S12,S2)^2 &= \frac{1}{(l+n)} \sum_{S12,S2} (Y_{2i} - \tilde{\mu}_2(S12,S2))^2 \\ \tilde{\rho}_{12} &= \frac{\sum_{S12}(Y_{1i} - \hat{\mu}_1(S12))(Y_{2i} - \hat{\mu}_2(S12))}{\sqrt{\sum_{S12}(Y_{1i} - \hat{\mu}_1(S12))^2 \sum_{S12}(Y_{2i} - \hat{\mu}_2(S12))^2}} \end{aligned}$$

where $\tilde{\rho}_{12}$ is the same as in (1.6). These estimates will be referred to as the naive estimates for the full data (NEFS). Note that it would have been possible to use other forms for the correlation estimate; for example

$$\rho'_{12} = \frac{\frac{1}{n}\sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))(Y_{2i} - \tilde{\mu}_2(S12, S2))}{\tilde{\sigma}_1(S12, S1)\tilde{\sigma}_2(S12, S2)}$$

This form, however, does not guarantee a correlation estimate between -1 and 1, which leads to difficulty in interpretation.

Although we would expect the NEFS to perform better than the MLES(S12), their usefulness appears to be limited by the fact that the correlation estimate is the same correlation estimate used in the MLES(S12). The next set of estimators proposed uses the same estimates as the NEFS for the marginal parameters, μ_i and σ_i, but develops a more intricate method for estimating ρ_{12}. This new estimate of ρ_{12} only uses the longitudinal data, as with the two previous estimators, but in a different form. The form of this estimate comes from writing down the likelihood for the data, as in (3.1), taking the logarithm and solving for ρ_{12}. If we label functions of the data as follows:

$$E = \sum_{S12} Y_{1i}, \quad F = \sum_{S12} Y_{1i}^2, \qquad G = \sum_{S12} Y_{2i}, \quad H = \sum_{S12} Y_{2i}^2$$
$$I = \sum_{S12} Y_{1i}Y_{2i}$$

then the equation to be solved is the one in which the following function is set to zero:

$$\begin{aligned} h(\rho_{12}) &= n\rho_{12}(1-\rho_{12}^2) + 2\rho_{12}(-\frac{F}{2\sigma_1^2} + \frac{\mu_1 E}{\sigma_1^2} - \frac{\mu_1^2 n}{2\sigma_1^2} \\ &- \frac{H}{2\sigma_2^2} + \frac{\mu_2 G}{\sigma_2^2} - \frac{\mu_2^2 n}{2\sigma_2^2}) + (1+\rho_{12}^2)(\frac{I - \mu_2 E - \mu_1 G + n\mu_1\mu_2}{\sigma_1\sigma_2}) \end{aligned} \quad (3.2)$$

This gives a cubic equation in ρ_{12} which can be solved to give three roots. Two of these roots are complex conjugates and the other one is real. The real root has the following form: if we set

$$\alpha = -\frac{F}{2\sigma_1^2} + \frac{\mu_1 E}{\sigma_1^2} - \frac{\mu_1^2 n}{2\sigma_1^2} - \frac{H}{2\sigma_2^2} + \frac{\mu_2 G}{\sigma_2^2} - \frac{\mu_2^2 n}{2\sigma_2^2} \quad (3.3)$$

$$\beta = \frac{I - \mu_2 E - \mu_1 G + n\mu_1\mu_2}{\sigma_1\sigma_2} \quad (3.4)$$

then

$$\rho_{12} = P^{\frac{1}{3}} + \frac{1}{9}Q + \frac{\beta}{3n} \quad (3.5)$$

where

$$\begin{aligned} P &= \frac{1}{27}\frac{\beta(18n^2 + 9n\alpha + \beta^2)}{n^3} \\ &+ \frac{\sqrt{-3n^4 - 18n^3\alpha + 33n^2\beta^2 - 36n^2\alpha^2 + 24n\alpha\beta^2 + 3\beta^4 - 24n\alpha^3 - 3\alpha^2\beta^2}}{9n^2} \end{aligned} \quad (3.6)$$

and

$$Q = \frac{3n^2 + 6n\alpha + \beta^2}{n^2 P^{\frac{1}{3}}} \quad (3.7)$$

The estimated quantities of μ_i and σ_i are substituted into the above expressions to give estimates for α, β, P and Q.

We can show that when the estimates of μ_i and σ_i are substituted in, the equation given by (3.2) must always have a root between -1 and +1, and thus the above correlation estimate must also have this property. To do this, first define

$$\begin{aligned}
\mathbf{U} &= (Y_{11} - \tilde{\mu}_1(S12, S1), \ldots, Y_{1n} - \tilde{\mu}_1(S12, S1)) \\
\mathbf{V} &= (Y_{21} - \tilde{\mu}_2(S12, S2), \ldots, Y_{2n} - \tilde{\mu}_2(S12, S2))
\end{aligned}$$

where $Y_{ij} \in S1$, and define the usual metrics on **U** and **V**. Hence,

$$\begin{aligned}
\mathbf{U} * \mathbf{V} &= \sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))(Y_{2i} - \tilde{\mu}_2(S12, S2)) \\
||U|| &= \sqrt{\sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))^2} \\
||V|| &= \sqrt{\sum_{S12}(Y_{2i} - \tilde{\mu}_2(S12, S2))^2}
\end{aligned}$$

Then, by the Cauchy-Schwartz inequality,

$$|U * V| \leq ||U|| * ||V||$$

Assume now that $|\tilde{\alpha}| < |\tilde{\beta}|$. Then

$$\begin{aligned}
&\frac{\sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))^2}{2\tilde{\sigma}_1(S12, S1)^2} + \frac{\sum_{S12}(Y_{2i} - \tilde{\mu}_2(S12, S2))^2}{2\tilde{\sigma}_2(S12, S2)^2} \\
&\qquad < \frac{|\sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))(Y_{2i} - \tilde{\mu}_2(S12, S2))|}{\tilde{\sigma}_1(S12, S1)\tilde{\sigma}_2(S12, S2)} \\
&\leq \frac{\sqrt{\sum_{S12}(Y_{1i} - \tilde{\mu}_1(S12, S1))^2 \sum_{S12}(Y_{2i} - \tilde{\mu}_2(S12, S2))^2}}{\tilde{\sigma}_1(S12, S1)\tilde{\sigma}_2(S12, S2)}
\end{aligned}$$

This would mean that

$$\frac{\sum_{S12}(Y_{1i}-\tilde{\mu}_1(S12,S1))^2}{2\tilde{\sigma}_1(S12,S1)^2}+\frac{\sum_{S12}(Y_{2i}-\tilde{\mu}_2(S12,S2))^2}{2\tilde{\sigma}_2(S12,S2)^2}$$
$$-\frac{\sqrt{\sum_{S12}(Y_{1i}-\tilde{\mu}_1(S12,S1))^2\sum_{S12}(Y_{2i}-\tilde{\mu}_2(S12,S2))^2}}{\tilde{\sigma}_1(S12,S1)\tilde{\sigma}_2(S12,S2)}<0$$

or

$$(\frac{\sqrt{\sum_{S12}(Y_{1i}-\tilde{\mu}_1(S12,S1))^2}}{\sqrt{2}\tilde{\sigma}_1(S12,S1)}-\frac{\sqrt{\sum_{S12}(Y_{2i}-\tilde{\mu}_2(S12,S2))^2}}{\sqrt{2}\tilde{\sigma}_2(S12,S2)})^2<0$$

which isn't possible. We conclude from this that $|\tilde{\alpha}| \geq |\tilde{\beta}|$.

Coming back to (3.2), we note that

$$\begin{aligned} h(-1) &= -2\hat{\alpha}+2\hat{\beta} \\ h(1) &= 2\hat{\alpha}+2\hat{\beta} \end{aligned}$$

Since $|\tilde{\alpha}| \geq |\tilde{\beta}|$, it is possible to show that for all combinations of $\tilde{\alpha}$ and $\tilde{\beta}$ being positive or negative, we find that h(-1) has a different sign than h(1). We conclude that the function given by (3.2) must have a root between -1 and 1. The estimate of the correlation thus also possesses this feature.

This set of estimators will subsequently be referred to as the semi-naive estimators (SNES). These have the advantage that they are in closed form, despite the fact that they are less intuitive than the NEFS.

The next section describes some properties of the NEFS and the SNES.

Properties of the NEFS and the SNES

Proving consistency of the naive estimators, when the amount of marginal data is a fixed multiple of the longitudinal data, and the amount of longitudinal data increases to infinity, is straightforward. Here, we'll show consistency of the semi-naive estimate of ρ_{12}, given by (3.5). We'll assume that the number of longitudinal observations is n.

Note that we can write the estimate of α in equation (3.3) as

$$\begin{aligned}\dot{\alpha} &= -\frac{\sum_{S12}(Y_{1i}-\hat{\mu}_1(S12))^2}{2\tilde{\sigma}_1(S12,S1)^2} - \frac{n(\hat{\mu}_1(S12)-\tilde{\mu}_1(S12,S1))^2}{2\tilde{\sigma}_1(S12,S1)^2} \\ &- \frac{\sum_{S12}(Y_{2i}-\hat{\mu}_2(S12))^2}{2\tilde{\sigma}_2(S12,S2)^2} - \frac{n(\hat{\mu}_2(S12)-\tilde{\mu}_2(S12,S2))^2}{2\tilde{\sigma}_2(S12,S2)^2}\end{aligned}$$

from which we conclude that

$$\frac{\dot{\alpha}}{n} \xrightarrow{P} -1.$$

Similarly, we write

$$\begin{aligned}\dot{\beta} &= [\sum_{S12}(Y_{1i}-\hat{\mu}_1(S12))(Y_{2i}-\hat{\mu}_2(S12)) + \\ &\quad n(\hat{\mu}_1(S12)-\tilde{\mu}_1(S12,S1))(\hat{\mu}_2(S12)-\tilde{\mu}_2(S12,S2))]/\tilde{\sigma}_1(S12,S1)\tilde{\sigma}_2(S12,S2)\end{aligned}$$

from equation (3.4), which gives that

$$\frac{\dot{\beta}}{n} \xrightarrow{P} \rho_{12}$$

Substituting these into the expressions for $\dot{P}$ and $\dot{Q}$ as given by (3.6) and (3.7), we find that

$$\begin{aligned} \dot{P} &\xrightarrow{P} \frac{1}{3^3}(\rho_{12} + \sqrt{3})^3 \\ \text{and} \qquad \dot{Q} &\xrightarrow{P} \frac{3(-3 + \rho_{12}^2)}{\rho_{12} + \sqrt{3}} \end{aligned}$$

from which we get that $\dot{\rho}_{12} \xrightarrow{P} \rho_{12}$.

The delta method can be used here to get approximate variances for the variance components. For example, in the case where m=l=kn, we find that using the NEFS,

$$\begin{aligned} Var(\bar{\sigma}_2(S12, S2)^2(1 - \bar{\rho_{12}}^2)) \approx [4\{(k+1)n - 1\}^2 \rho_{12}^2 \sigma_2^4 (1 - \rho_{12}^2)^2 \\ -2\{(k+1)n - 1\} \rho_{12}^2 \sigma_2^4 (1 - \rho_{12}^2)^2 (n-1) \\ +2\{(k+1)n - 1\} \sigma_2^4 (1 - \rho_{12}^2)^3 (n-1)]/(k+1)^2 n^2 (n-1) \end{aligned} \tag{3.8}$$

Substituting $k = 0$ into this expression gives

$$\frac{2(n-1)\sigma_2^4(1 - \rho_{12}^2)^2}{n^2} \tag{3.9}$$

which is the approximate variance for the first component when the longitudinal data are ignored, i.e. using the MLES(S12). Similar calculations can be done for the second variance component, and also for the variance components estimated using the SNES. Please see the next section for further details.

3.1.2 Three or More Stages

We consider now a situation in which there are three or more stages in the process. The data will still be considered in groups, so that $S12\ldots k$ contains the longitudinal data and that Sk contains the marginal data on stage k. We will not consider the situation in which there is incomplete longitudinal data. The data are assumed to adhere to an AR(1) model.

Estimation and Properties

The estimates given in equation (1.6) can be used as estimates that do not use the extra cross-sectional data, MLES($S12\ldots k$). This is also true in the case of three or more stages in the process. The maximum likelihood estimates for the full data can be found by optimizing the likelihood, analogous to that shown in the two stage case. For a k-stage process, this likelihood will be the product of a k-variate constrained multivariate normal and k univariate normal parts, to account for the cross-sectional data. Estimates will need to be computed numerically.

The naive estimates can be easily generalized to three or more stages. Again, estimates of the marginal parameters for a stage can be estimated from all the data available at that stage, and correlation estimates for consecutive stages can be estimated from the longitudinal data. Recall that since we are assuming an AR(1) model, the correlation parameters for stages that are not consecutive are just products of the correlations between consecutive stages. The semi-naive estimates can be generalized in the same way to three or more stages.

The naive and semi-naive estimates are clearly consistent for a process with three or more stages, under the same conditions as they were for the two stage process. Furthermore, approximate variance formulas can be found for the variance components. As in the measurement error situation, it was found that the square roots of these components more closely approximated normality than the components themselves. For the naive estimates and in a three stage process, it was found that

$$\begin{aligned}
Var(\sqrt{\tilde{\sigma}_3(S123,S3)^2(1-\tilde{\rho}_{23}^2)}) &\approx \frac{\sigma_3^2(1-\rho_{23}^2)(2\rho_{23}^2nk+n-1)}{2(n-1)(k+1)n} \\
Var(\sqrt{\tilde{\sigma}_3(S123,S3)^2\tilde{\rho}_{23}^2(1-\tilde{\rho}_{12}^2)}) &\approx \sigma_3^2(1-\rho_{12}^2)[2k\rho_{12}^2\rho_{23}^4n+\rho_{23}^2 \\
-4\rho_{23}^2nk+2n-2+2nk+2\rho_{23}^4nk &- \rho_{23}^2n)]/2(n-1)(k+1)n \\
\text{and} \qquad Var(\sqrt{\tilde{\sigma}_3(S123,S3)^2\tilde{\rho}_{23}^2\tilde{\rho}_{12}^2} &\approx \sigma_3^2[2\rho_{23}^2n+2\rho_{12}^2n+2\rho_{12}^2nk+2\rho_{23}^2nk \\
-6\rho_{23}^2\rho_{12}^2nk-2\rho_{23}^2-2\rho_{12}^2 &- 3\rho_{12}^2\rho_{23}^2n+2\rho_{23}^4\rho_{12}^4nk+3\rho_{12}^2\rho_{23}^2] \\
&\quad /2(n-1)(k+1)n \qquad (3.10)
\end{aligned}$$

when n is the size of S123 and the marginal data all have the same size, kn. Similar formulas can be found for the semi-naive estimates, but these are extremely lengthy. Maple programs to produce these formulas for a three stage process are given in Appendix B. Similar calculations can be done for a general k stage process, but might be prohibitive when k is large.

Simulation Results

Some simulation studies were performed to investigate the four estimators described. These studies had two goals. The primary one was to compare the four estimators under a variety of different conditions. The secondary goal was to determine if confidence intervals calculated for the NEFS and the SNES actually gave intervals close to their theoretical coverage frequencies.

The first simulation study modeled a three stage process, and was set up such that the amount of cross-sectional data at all three stages was the same. The experimental design had four factors: the correlation between the first and second stage (ρ_{12}), the correlation between the second and third stage (ρ_{23}), the sample size of the longitudinal data (n) and the multiplicative factor of the marginal data (k). The correlation factors were run at the values $\sqrt{0.2}$, $\sqrt{0.5}$ and $\sqrt{0.8}$, the longitudinal sample size had two values, 20 and 50, and k could take the values 1, 2 or 5. The result was a 54 run simulation.

At each run, 100 samples were generated randomly, and each of the four estimators was used to compute the estimates of the three variance components. Averages and standard deviations for each of the components and each run are given in Tables C.1 - C.6 of Appendix C. An estimate of the mean square error of each estimator was computed, also for each run and each component. These values are plotted in Figures 3.1 - 3.3. The mean square error of each estimator averaged over 27 scenarios for each n and the 100 runs is given in Table 3.1. This was done for each component and for each value of n.

The results of this simulation indicate that irrespective of how much cross-sectional data are available, a sample size of 20 for the longitudinal data gives point estimates that are too imprecise to be of any practical value. This seems to be especially true when the variation is roughly equally divided among all stages;

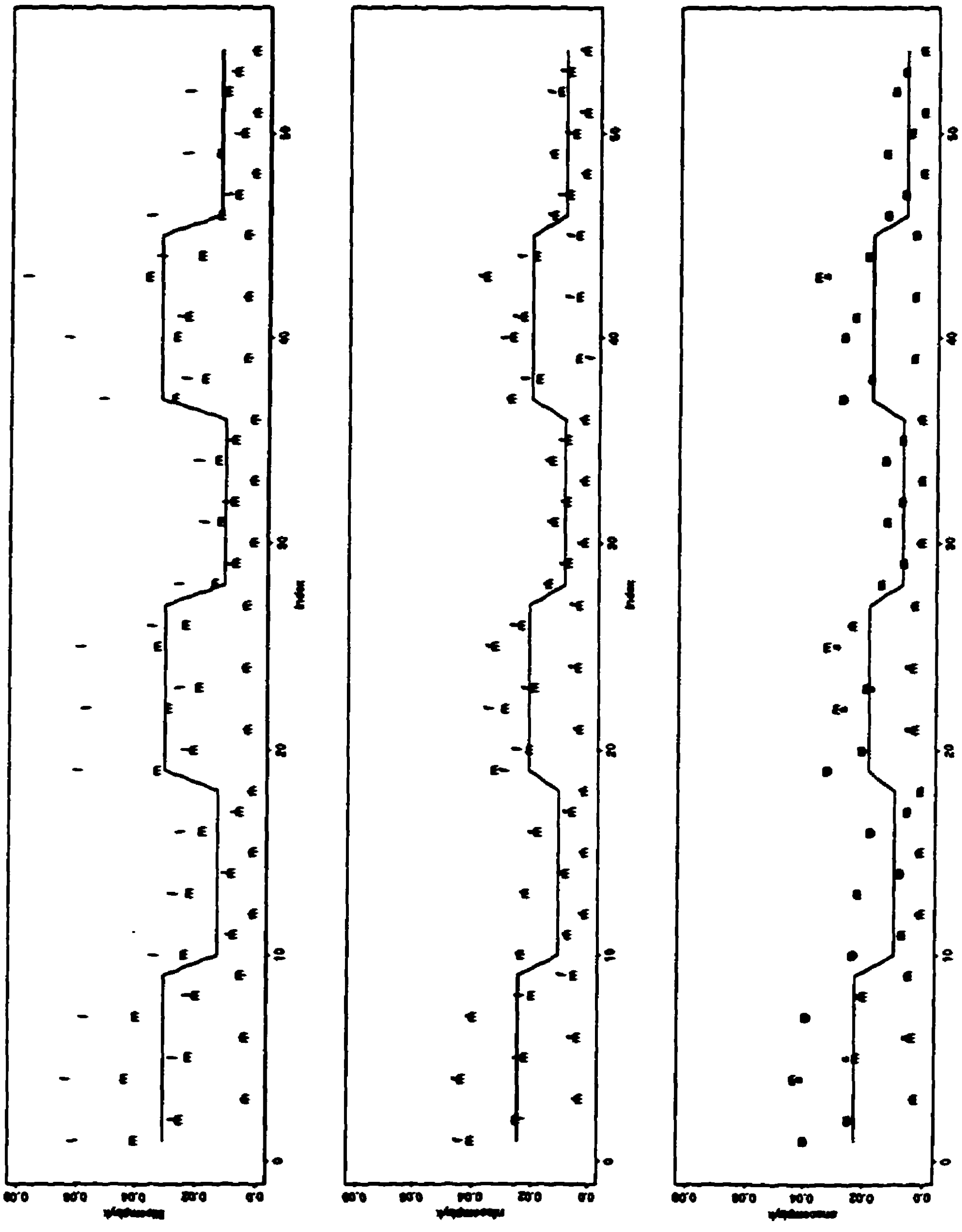

Figure 3.1: MSEs of the first variance component. The "m" on each plot represents the value of the run with estimator MLES(S1-S4). The first plot shows the MLES(S12), the second the NEFS, and the last plot the SNES. The lines on the plot represent averages over k and n. For example, the line from observations 1 to 9 represents the average for k=1 and n=20, while the line from observations 10 to 18 represents the average for k=1 and n=50.

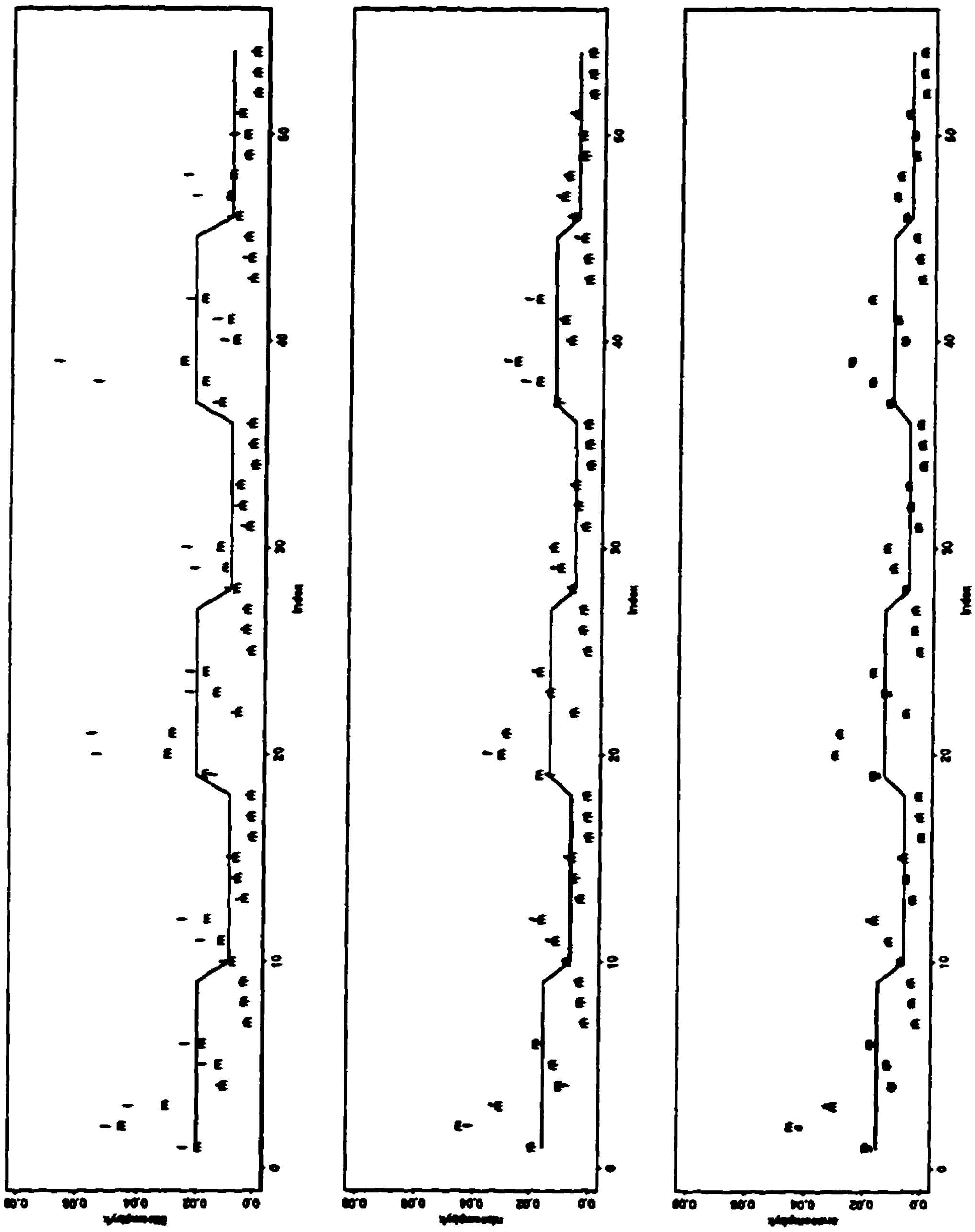

Figure 3.2: MSEs of the second variance component.

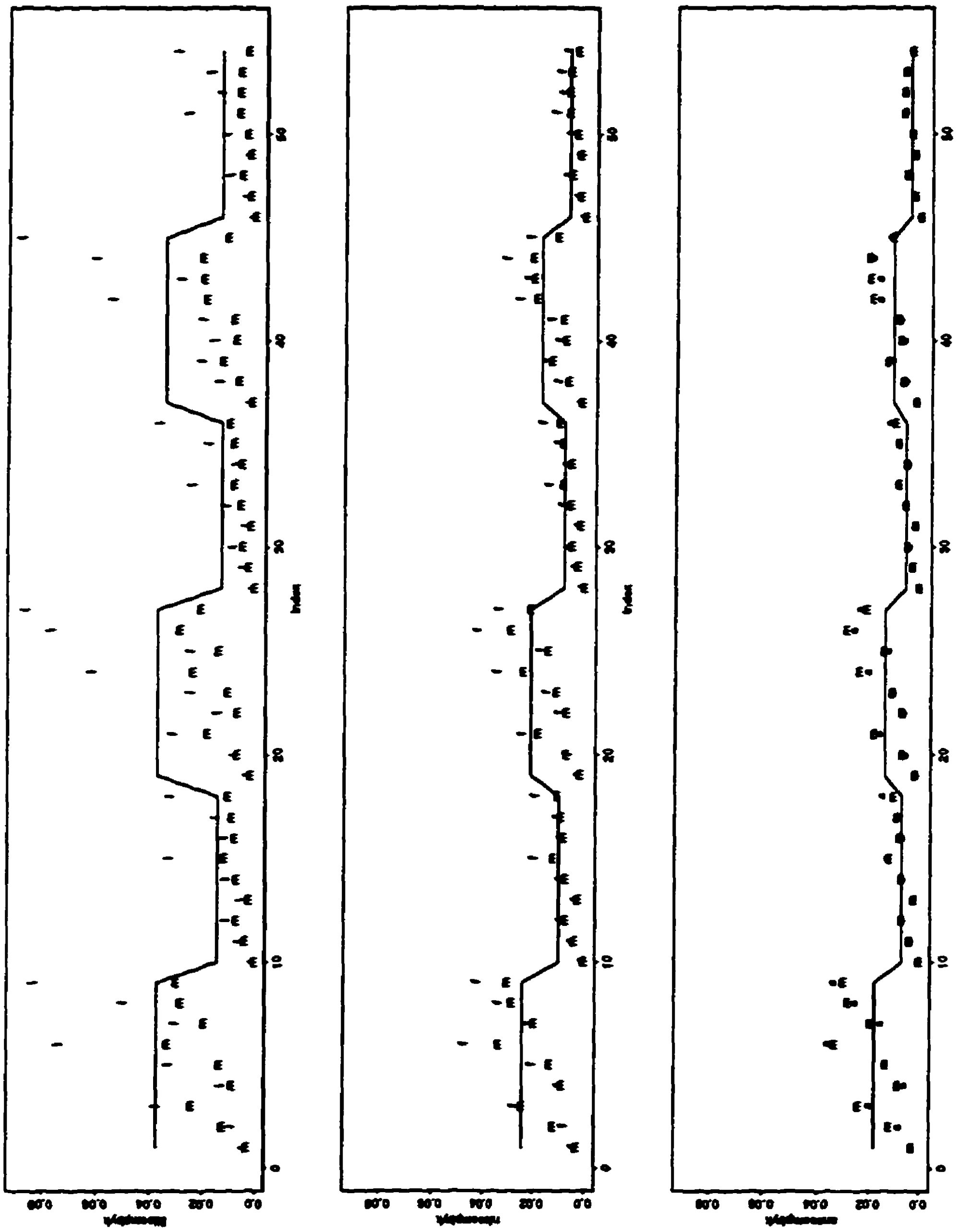

Figure 3.3: MSEs of the third variance component.

Component	n	Estimator	$\widehat{MSE}$
1	20	MLES(S123)	0.0311
		NEFS	0.0224
		MLES(S123-S3)	0.0202
		SNES	0.0200
	50	MLES(S123)	0.0125
		NEFS	0.0096
		MLES(S123-S3)	0.0084
		SNES	0.0083
2	20	MLES(S123)	0.0207
		NEFS	0.0139
		MLES(S123-S3)	0.0134
		SNES	0.0131
	50	MLES(S123)	0.0086
		NEFS	0.0056
		MLES(S123-S3)	0.0051
		SNES	0.0053
3	20	MLES(S123)	0.0367
		NEFS	0.0211
		MLES(S123-S3)	0.0160
		SNES	0.0146
	50	MLES(S123)	0.0142
		NEFS	0.0085
		MLES(S123-S3)	0.0061
		SNES	0.0061

Table 3.1: Estimated mean squared errors for the estimators and the components

that is, no stage dominates.

It can be seen that the MLES(S123) do worse then the other estimators uniformly over the scenarios.

Another thing to note is that the MSEs of the NEFS for each run are very close to those of the MLES(S123-S3). The SNES have mean square error values that are almost the same as the full maximum likelihood estimates. This relationship is reflected in the actual estimates themselves. Plots of the estimates show that the NEFS are not as close to the MLES(S123-S3) as the SNES. These latter are extremely close to the MLES(S123-S3) for runs in which most of the variation is being added at the last stage of the process. For runs in which this is not true, the SNES are further away from the MLES(S123-S3).

Another point that can be observed from this simulation is that differences in the mean square error due to an increase in k are less pronounced than differences due to an increase in n.

A further study was carried out to see how closely the confidence intervals found using the approximate variance formulas for the NEFS and the SNES gave their theoretical coverage frequencies. This simulation was also carried out in 27 runs, at the same factor levels as the previous study, except that the only n value used was n = 50. Here, 2500 samples for each run were generated and the NEFS and the SNES computed for each. 98, 95 and 90% confidence intervals were found using the approximate variance formulas discussed earlier. The percentage of confidence intervals that contained the true value was then computed. Tables C.7 - C.12 that give the results of this simulation are also in Appendix C.

For the NEFS, the observed coverage frequencies are close to the theoretical values, although they tend to underestimate them in general. For the SNES, however, the approximate variance formulas are less reliable. The intervals produced by this method are too conservative, and generally give coverage frequencies much higher than their theoretical values.

Further simulation was done to investigate how parametric bootstrapping performed as a method of producing confidence intervals compared to the approximate variance formulas. This simulation was done only at three combinations of the factor levels, because of the amount of computer time required. The values of the runs were $\rho_{12} = \sqrt{0.8}$, $\rho_{23} = \sqrt{0.2}$ and k=1; $\rho_{12} = \sqrt{0.8}$, $\rho_{23} = \sqrt{0.8}$ and k=2; and $\rho_{12} = \sqrt{0.2}$, $\rho_{23} = \sqrt{0.8}$ and k=5. Those three runs were chosen to be such that most of the variation in the process was coming from a single stage. At each run, 1000 samples were generated and confidence intervals were produced using both methods. Tables resulting from this simulation are given in Tables C.13 and C.14 of Appendix C. These tables indicate that for the NEFS, the approximate variance formulas are comparable to the bootstrap method for generating confidence intervals. For the SNES, the bootstrap is a more reliable method, although it can give values far from the theoretical values in some cases.

3.1.3 Adding Measurement Error

As was illustrated in the previous chapter, measurement error is an important issue in industrial processes. One question that comes to mind at this point is "How would measurement error be taken into account when cross-sectional data are present?" The estimates introduced earlier can be extended to explicitly deal with this.

In the case of estimators that only make use of the longitudinal data, the situation is analogous to that described in chapter two. Hence, the naive estimates given in that chapter by equation (2.8) could be used.

If the maximum likelihood estimates are desired in this situation, then the full likelihood can be written out and optimized numerically. For example, in the case of a three stage process, if we denote

$$\Sigma = \begin{bmatrix} \sigma_1^2 + \sigma_{\epsilon_1}^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\rho_{23}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 + \sigma_{\epsilon_2}^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{12}\rho_{23}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 + \sigma_{\epsilon_3}^2 \end{bmatrix}$$

then the likelihood is

$$\frac{1}{(2\pi)^{3n/2+m/2+l/2+r/2}|\Sigma|^{n/2}(\sigma_1^2+\sigma_{\epsilon_1}^2)^{m/2}(\sigma_2^2+\sigma_{\epsilon_2}^2)^{l/2}(\sigma_3^2+\sigma_{\epsilon_3}^2)^{r/2}}$$
$$exp\{-1/2trace[\Sigma^{-1}S_{xx}] - n/2(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu)$$
$$-\frac{\sum_{S1}(X_{1i}-\mu_1)^2}{2(\sigma_1^2+\sigma_{\epsilon_1}^2)} - \frac{\sum_{S2}(X_{2i}-\mu_2)^2}{2(\sigma_2^2+\sigma_{\epsilon_2}^2)} - \frac{\sum_{S3}(X_{3i}-\mu_3)^2}{2(\sigma_3^2+\sigma_{\epsilon_3}^2)}\}$$

where n, m, l and r are the sizes of S123, S1, S2 and S3, respectively, and S_{xx} and $\bar{x}$ are the usual matrix of cross-products and vector of averages. When the measurement error variance is assumed known, then there are eight parameters in this distribution that need to be estimated.

The naive estimates and the semi-naive estimates introduced earlier can be extended in a natural way to take measurement error into account. For the naive estimates, for example, we will continue to estimate the mean for a given stage using all the data, but now the estimate of the variance will be the observed variance over all the data at that stage minus the measurement error variance. The correlation parameters can then be estimated as given by (2.8). The semi-naive estimates can be modified in the same way as the naive estimates have been for the marginal parameters. Then, the correlation parameters will still be estimated using equation (3.5), but now the modified marginal parameters will be substituted into the expression.

A numerical simulation has not yet been done to determine how well these modified estimators perform.

3.2 Design Issues

When considering the issue of cross-sectional data supplementing the longitudinal data, it would be helpful to quantify the relative value of the different data. Such information could be used at the design stage of a study. Clearly such relationships will depend on how expensive it is to collect the data, but given that constraint, optimizing the amount of information that can be gained for a specified cost is desirable. This section addresses that question.

3.2.1 Known Marginal Parameters

We will begin this discussion by investigating a two stage process, and considering what happens when the marginal parameters, μ_1, μ_2, σ_1 and σ_2 are known. This situation is the limit of having large amounts of cross-sectional information.

Assume that we track n items through the process, and that we estimate the correlation parameter between the two stages by

$$\hat{\rho}_{12} = \frac{S_{y_1 y_2}}{\sqrt{S_{y_1 y_1} S_{y_2 y_2}}}$$

Then we are interested in knowing what the gain in precision is if we estimate the variance components of interest by $\sigma_2^2(1 - \hat{\rho}_{12}^2)$ and $\sigma_2^2\hat{\rho}_{12}^2$, that is using the known value of σ_2^2, over estimating the variance at the second stage with the data collected. Note that the estimate of $\hat{\rho}_{12}^2$ is not the maximum likelihood estimate when the marginal information is known.

As mentioned earlier, it can be shown that when σ_2 is estimated from the n observations, the approximate variance of the first component is given by (3.9), and that of the second is

$$\frac{2(n-1)\rho_{12}^2\sigma_2^4(2 - \rho_{12}^2)}{n^2}$$

When the known value of σ_2 is used to estimate these components, they both have the same variance, given by

$$\frac{4\sigma_2^4\rho_{12}^2(1 - \rho_{12}^2)^2}{n-1} \tag{3.11}$$

Figure 3.4 shows a plot of the ratio of the standard deviation of the estimate with the marginal information to the standard deviation of the estimate without the marginal information. The peculiar feature that this plot reveals is that for values of ρ_{12} such that $|\rho_{12}| > \sqrt{0.5}$, the standard deviation of the first component estimate is actually greater when the marginals are known. The explanation appears to be that the extra information is being used inefficiently there. Another interesting feature that this plot reveals is that the ratio of the standard deviations for the first component is highest for those values of ρ_{12} at which the ratio for the second component is lowest.

Similar calculations can be done in this case when the maximum likelihood estimate is used for ρ_{12}. Figure 3.5 shows the plot of the ratio of the standard deviation of the variance components with the known marginal information to that estimating σ_2 with the longitudinal data. In this case, we see that the standard deviation is never greater using the marginal information than without it, which is what we would expect intuitively.

3.2.2 Limited Cross-sectional Data

The situation described above, in which we know the marginal parameters of the distribution exactly, may seldom occur in practice. In what follows, we will assume that the amount of marginal data collected is the same at each stage, and is given by kn.

In this scenario, we can calculate the approximate variances of the NEFS and compare them to the approximate variances of the MLES(S12). When n is large, this should give a good indication of the amount of information to be gained in the cross-sectional data, when using those estimates.

The approximate variances for the components of a two stage process using the MLES(S12) were given in the previous section. The variance of the first component using the NEFS is given by (3.8). If we make the approximation $(k+1)n-1 \approx$

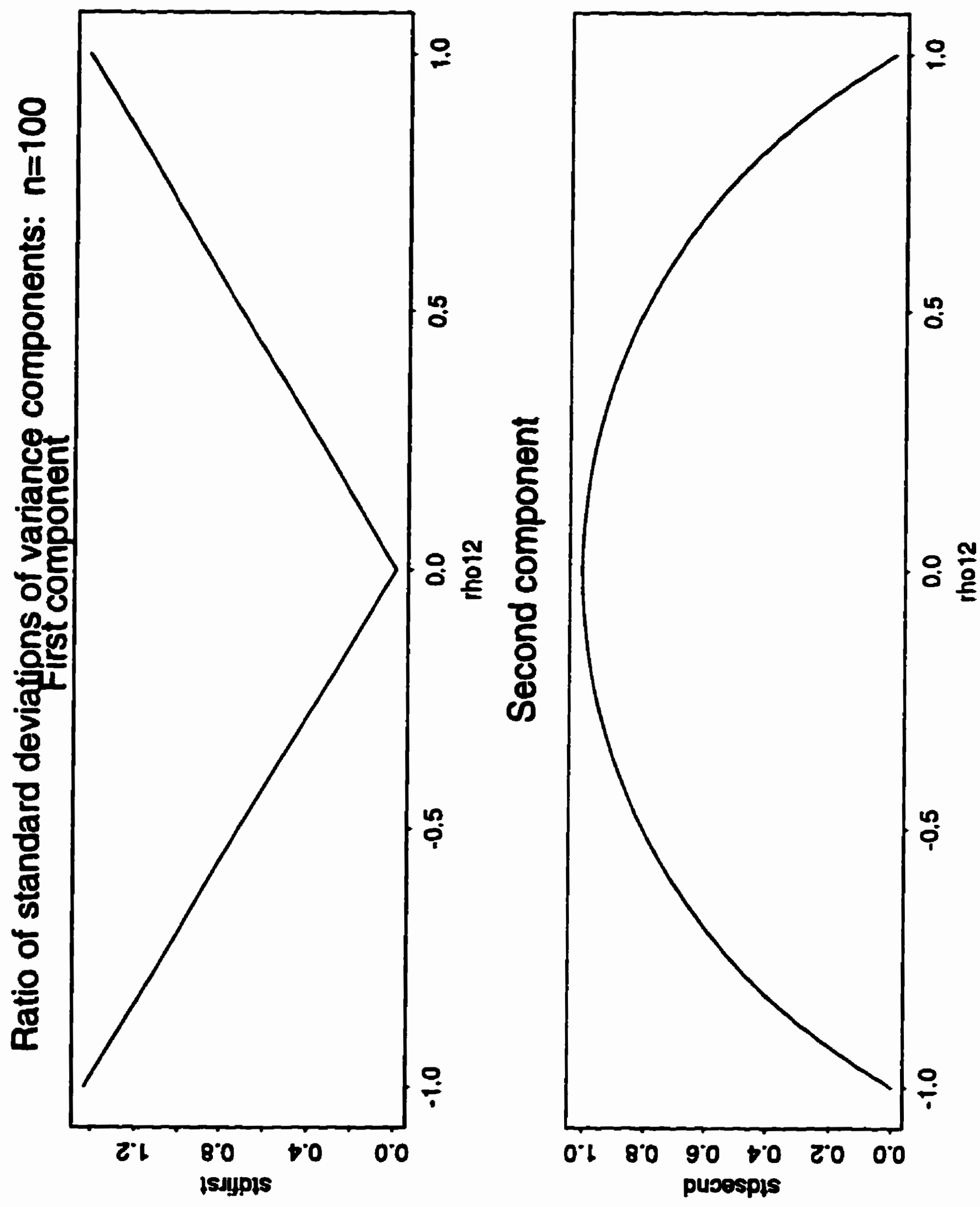

Figure 3.4: Ratio of standard deviations of variance components with known marginal information, using the naive estimate, vs. estimating marginal information

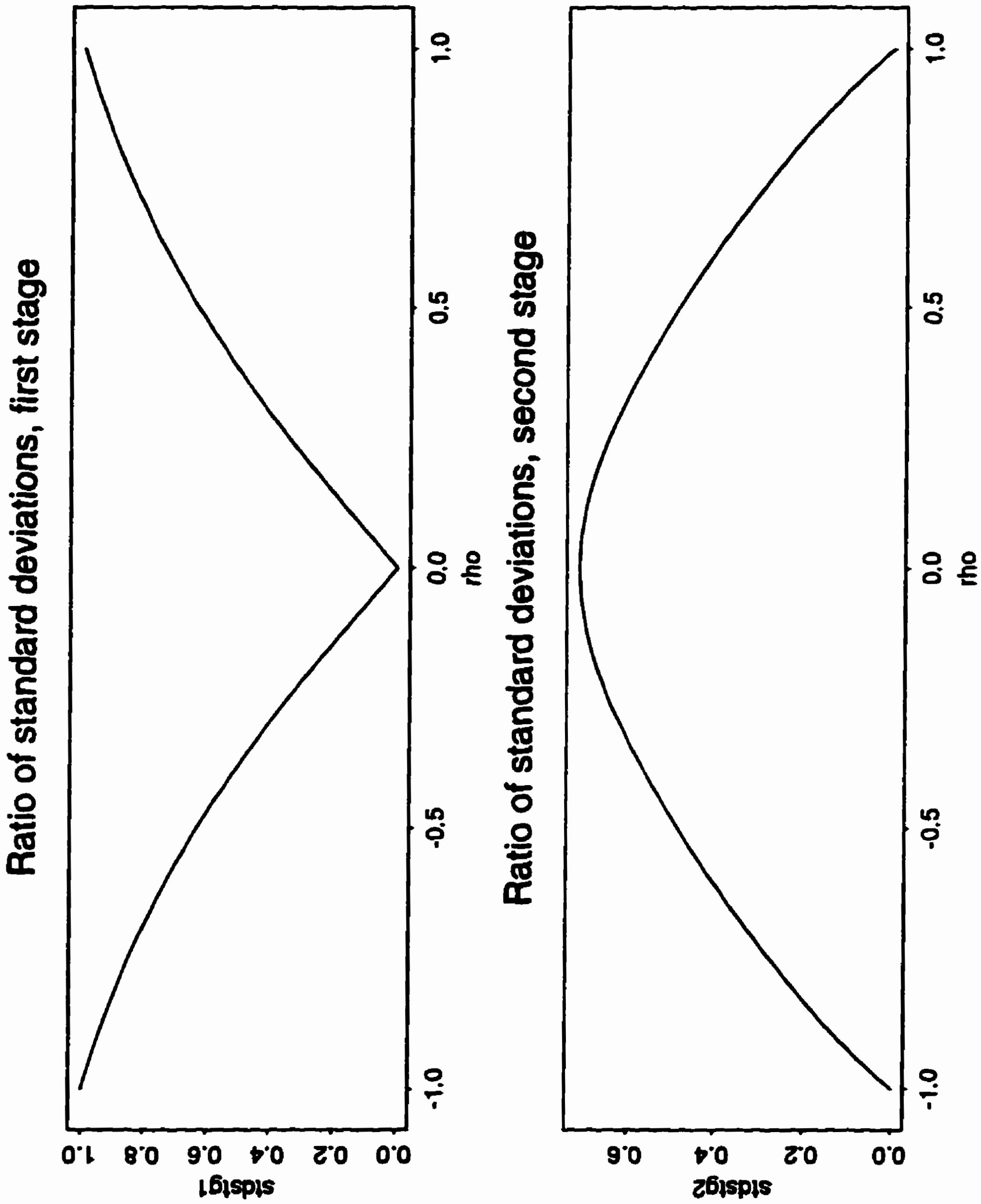

Figure 3.5: Ratio of standard deviations of variance components with known marginal information, using the maximum likelihood estimate, vs. estimating marginal information

$(k+1)n$, then we can write that equation as

$$\frac{4\rho_{12}^2\sigma_2^4(1-\rho_{12}^2)^2}{n-1}+\frac{2\sigma_2^4(1-\rho_{12}^2)^2(1-2\rho_{12}^2)}{(k+1)n}$$

where the first term is the variance calculated when the marginals are known, and the second term is a correction factor. Notice that this second term is negative if $|\rho_{12}| > \sqrt{0.5}$, and so the variance expression increases with increasing k in that region. This behavior can be seen in Figure 3.6, which is a plot of the ratio of the standard deviation of the first component estimated with the NEFS against those estimated with the MLES(S12), for different values of ρ_{12}, and for large n.

Similarly, the variance of the second component using the NEFS can be written as

$$\frac{4\rho_{12}^2\sigma_2^4(1-\rho_{12}^2)^2}{n-1}+\frac{2\sigma_2^4\rho_{12}^4(3-2\rho_{12}^2)}{(k+1)n}$$

In this case, the second term is always positive, and so this variance is always decreasing as a function of k. This can be seen in Figure 3.7.

Some interesting features are revealed in these plots. Note that when ρ_{12} is small, the first component of variance will be larger than the second, and therefore it will be the component of most interest. This is also the situation in which a large gain in precision can be made by using the NEFS over the MLES(S12) to estimate this component. Similarly, when ρ_{12} is large, the second component is the dominant one. Again, this is exactly when the most gain in precision is to be had by using the NEFS to estimate this component. Since the first component in this case will have a small value, a relative loss in precision in estimating it may not be

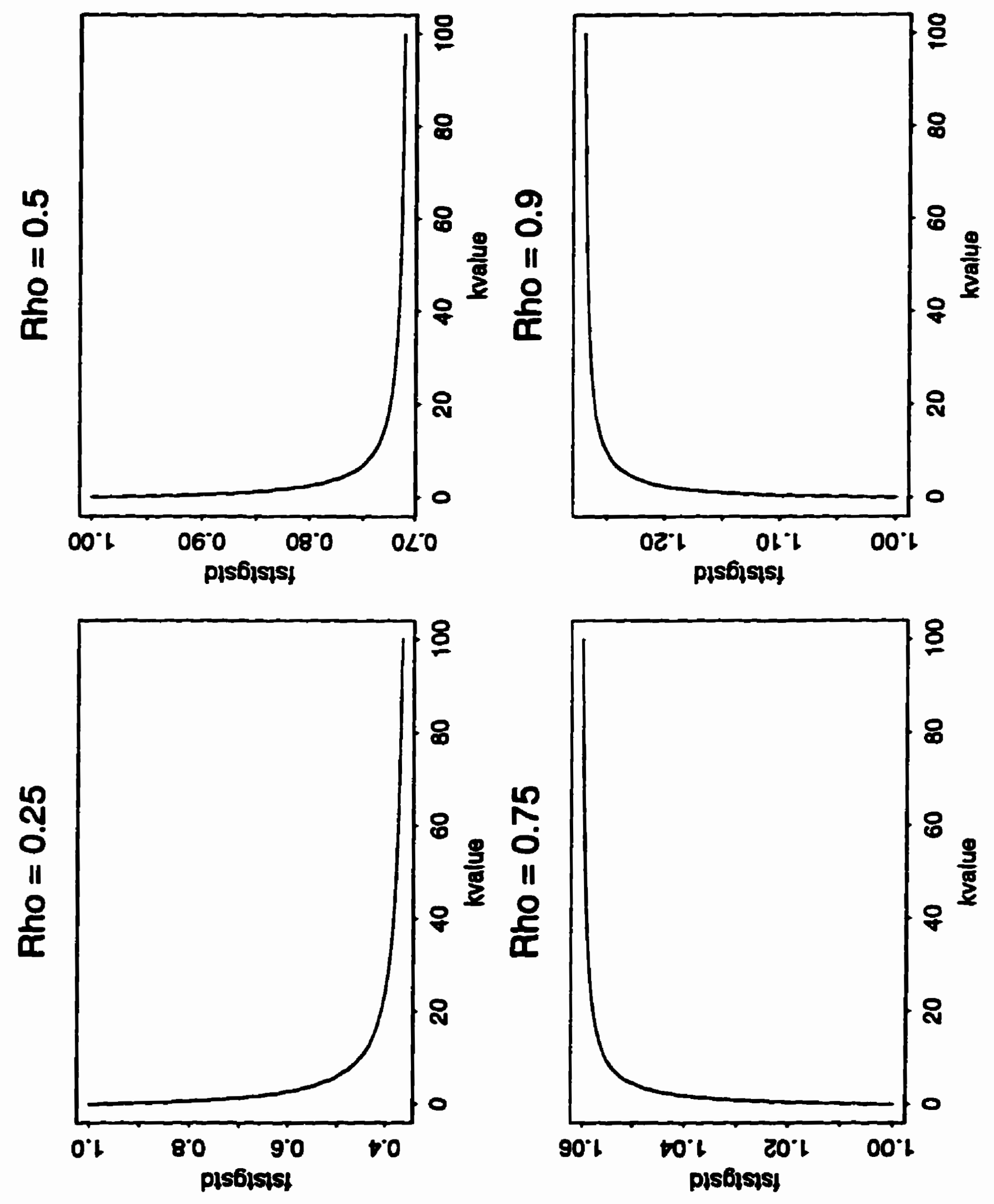

Figure 3.6: Ratio of standard deviations of estimates of first component using cross-sectional information, and with naive estimates

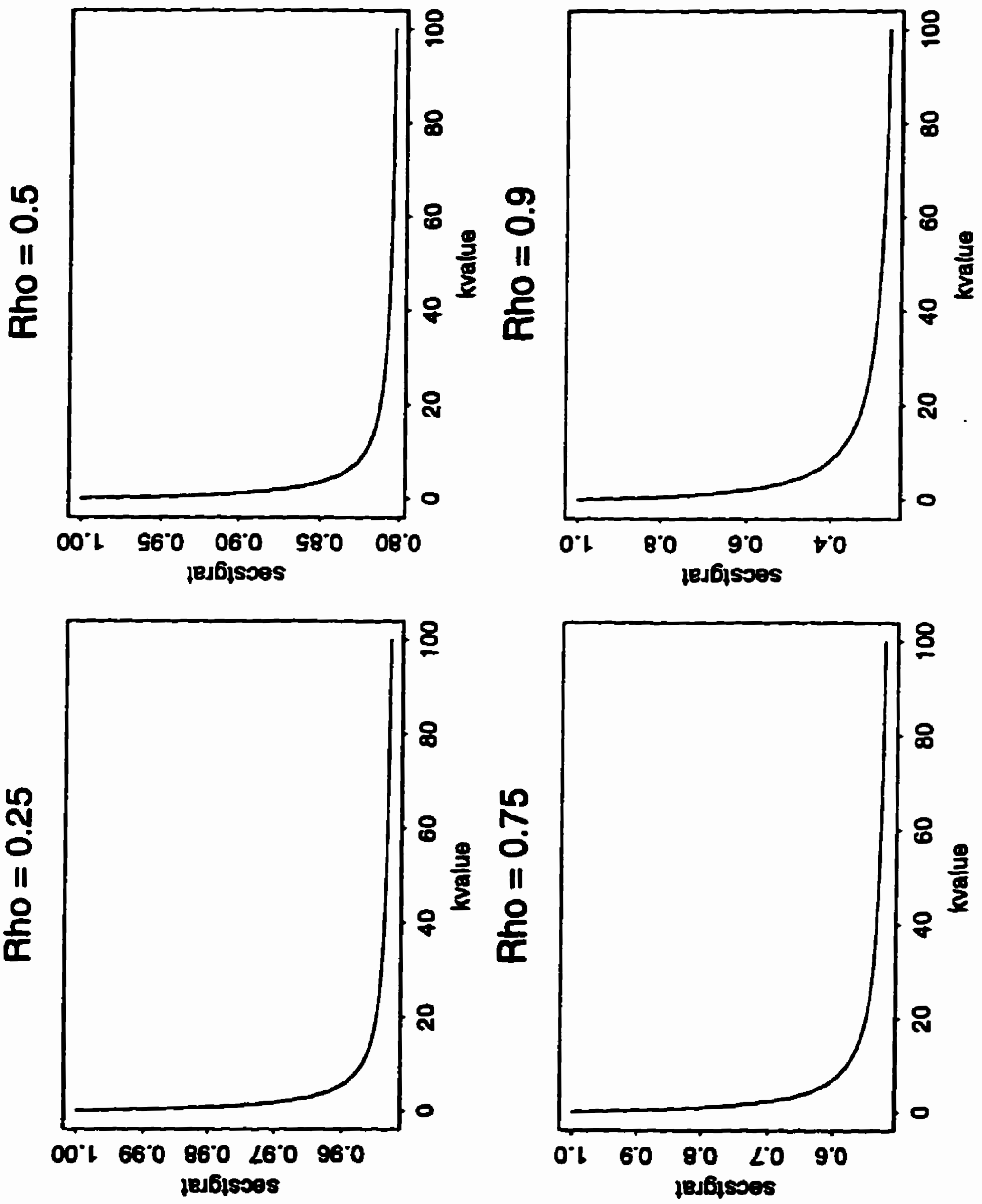

Figure 3.7: Ratio of standard deviations of estimates of second component using cross-sectional information, and with naive estimates

bothersome.

Roughly speaking, it seems that the steepest gain (or loss) in efficiency occurs between k=0 and k=10. By k=20, it appears that most of the gain has been made. At this point, we are approaching the limit in which we know the marginal parameters. These are points to keep in mind when designing this type of experiment.

3.2.3 General Recommendations

It has been shown to be less than straightforward to answer the question "How much information can be gained from the marginal data in this type of variance analysis?" In a practical situation, the answer to that question will always depend on how expensive it is to collect cross-sectional data as compared to longitudinal data. The results given here indicate that if cross-sectional data are about as expensive as longitudinal data, then the latter are more valuable in this analysis. In the more likely case that the cross-sectional data are substantially less expensive, it appears that there is some gain in collecting as much cross-sectional data as longitudinal, or even twice as much. After twice as much cross-sectional data has been collected, the rate of gain seems small. This assumes that there is a reasonable amount of longitudinal data collected. A sample size of 20 for the longitudinal data, for example, is probably too small. The tables given in appendix C can be used as guidelines for these types of decisions. If the investigator is interested in using the SNES, different data collection scenarios could be investigated in a simulation study that uses bootstrapping to give an estimate of the amount of precision that might be expected, for different values of the distributional parameters.

Chapter 4

The General Multivariate Normal Model

The models that have been introduced thus far to deal with the variance transmission problem are the AR(1) model and the AR(1) model with measurement error. Using these models, it is a simple task to assess the effect of an intervention in the process, either by reducing variation added at a certain stage, or by reducing the slope of the regression of a certain stage on the previous stage. In the case of a process that cannot be described by these models, it becomes a difficult task to assess the effect of any given stage on the variation in the final response.

For example, in the case of the piston process introduced in the first chapter, suppose we use an AR(1) model to model the last four stages. We therefore partition the variance of the last stage into four components: transmitted from the first stage (y_4), added at the second stage (y_5), added at the third stage (y_6) and added at the final stage (y_7). We could also combine the last two stages (i.e. pretend that

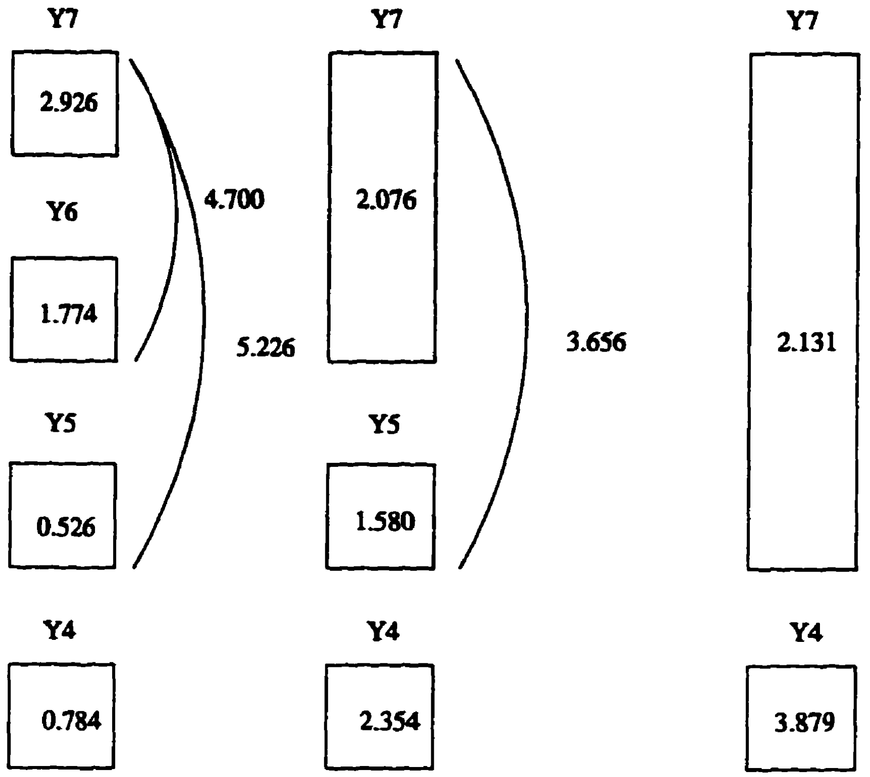

Figure 4.1: Original breakdown of variation by stage for piston data using AR(1) models. Figure 4.1(A) (left) assumes all stages are observed, Figure 4.1(B) (middle) assumes y_6 is not observed and Figure 4.1(C) (right) assumes y_5 and y_6 are not observed.

we don't observe y_6), and partition the variance into three components. Finally, we could combine the last three stages of the process and partition the variance at the final stage into two components: transmitted from the first stage and added between the first and last stage. If we do this we get values as shown in Figure 4.1.

This figure clearly shows a dilemma in taking effective action for variance reduction. When the variance is partitioned into four components, it seems that very little variation is being transmitted from y_4. However, when the variance is decomposed into two components, transmitted from y_4 and added after y_4, it seems that most of the variation is being transmitted from y_4. These present two contradictory messages about how effective reducing the variation in Y_4 would be in reducing the variation at the final stage. The problem is due to the fact that the AR(1) model is not adequate for these data.

This chapter discusses methods of determining the effect of a given stage on the variation in the final product, when the data are assumed to follow a general multivariate normal model. That is, Y_1, ..., Y_k can be modeled as a k-variate multivariate normal. This effect will be analysed by proposing interventions to the process at that stage. In general, there are two types of interventions that can be made at a given stage: the variance at that stage can be reduced or the slope of the regression of that stage on previous stages can be reduced. Methodology will be given to study these interventions in the case of a four stage process, under certain assumptions outside of the model. These methods will be used to analyse data from the piston production process and the car door hanging process.

In what follows, we will often make use of the four variable multivariate normal formulae. If the variables y_1, y_2, y_3 and y_4 are multivariate normal, then their joint probability distribution can be written as

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^2|\Sigma|^{1/2}} exp\{-\frac{1}{2}(\mathbf{Y}-\mu)^T\Sigma^{-1}(\mathbf{Y}-\mu)\} \qquad (4.1)$$

$$\text{where} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix},$$

$$\begin{aligned} \mu &= E(\mathbf{Y}) \\ \text{and} \qquad \Sigma &= E((\mathbf{Y}-\mu)(\mathbf{Y}-\mu)^T). \end{aligned}$$

Recall that in this case, if we partion $\mathbf{Y}$, μ and Σ as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\text{and} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where

$\mathbf{Y}_1$ and μ_1 are q x 1 vectors (q ¡ 4)

$\mathbf{Y}_2$ and μ_2 are (4-q) x 1 vectors

Σ_{11} is a q x q matrix

Σ_{12} is a q x (4-q) matrix

Σ_{21} is a (4-q) x q matrix

and Σ_{22} is a (4-q) x(4-q) matrix,

then $\mathbf{Y}_1$ is a q-variable multivariate normal with mean μ_1 and variance-covariance matrix Σ_{11}. Further, using the above partition of $\mathbf{Y}$, the conditional distribution of $\mathbf{Y}_2$ given $\mathbf{Y}_1 = \mathbf{y}_1$ is

$$N_{4-q}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \tag{4.2}$$

(See Johnson, 1988.)

Since we will be making frequent use of the above formulae, we will adopt the notation that Σ with a subscript of numbers separated by a slash will indicate a conditional variance-covariance matrix. Also, σ with the same type of subscripts will denote an element of this matrix. Hence, for example,

$$\begin{aligned} \Sigma_{234|1} &= Var(\begin{pmatrix} y_2 \\ y_3 \\ y_4 \end{pmatrix}|y_1), \\ \sigma_{23|1} &= Cov(y_2, y_3|y_1), \\ \text{and} \qquad \sigma_{22|1} &= Var(y_2|y_1). \end{aligned}$$

Also, subscripts of Σ that are numbers separated by a comma will indicate those rows and columns of Σ. For instance, $\Sigma_{23,4}$ will indicate a 2x1 vector given by the second and third rows of Σ, and its fourth column. This notation will also apply for μ. Hence, μ_{34} will denote the third and fourth components of the μ vector.

4.1 Intervention at the first stage

4.1.1 Methodology

Consider a four stage process, where the stages produce measurements y_1, y_2, y_3 and y_4, in order. It is assumed that these four variables are multivariate normal. Then, their joint probability distribution can be written as in equation 4.1. We are interested in the conditional distribution of $\mathbf{y}_{234} = (y_2, y_3, y_4)^T$ given y_1, which can be constructed in the manner of equation 4.2. For future reference, the variance-covariance matrix in this conditional distribution will be denoted $\Sigma_{234|1}$.

To determine the effect of an intervention at the first stage in this process, we will change the marginal distribution of y_1, and assume that the above conditional distribution remains the same. This assumption seems reasonable intuitively, but can only be verified experimentally. Thus, suppose that the marginal distribution is modified to be $f(y_1) \sim N(\mu_1, \tau\sigma_1^2)$. Since we are not changing the mean of y_1 nor the conditional distribution of $\mathbf{y}_{234}$ given y_1, the means of $\mathbf{y}_{234}$ will not change either. The changes of interest, then, are the variances of all variables, as well as their correlations. Having modified the distribution of y_1, it would be worthwhile to decompose the variation of the final diameter (y_4) into various stages, to determine if these breakdowns reflect the cause of the reduction in variation.

One way of accomplishing these goals is to reconstruct the multinormal distribution of the four variables, with its new parameters. This can be done using conditional variance formulae. For example, since the conditional distribution of $\mathbf{y}_{234}$ given y_1 is assumed constant, then the conditional distribution of $y_2|y_1$, namely

$$y_2|y_1 \sim N(\mu_2 + \rho_{12}(\sigma_2/\sigma_1)(y_1 - \mu_1), \sigma_2^2(1 - \rho_{12}^2)),$$

is constant as well. Hence, with the new marginal distribution of y_1,

$$\begin{aligned} Var(y_2) &= E(Var(y_2|y_1)) + Var(E(y_2|y_1)) \\ &= E(\sigma_2^2(1 - \rho_{12}^2)) + Var(\mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)) \\ &= \sigma_2^2(1 - \rho_{12}^2) + \rho_{12}^2(\frac{\sigma_2}{\sigma_1})^2\tau\sigma_1^2 \\ &= \sigma_2^2(1 - \rho_{12}^2) + \rho_{12}^2\sigma_2^2\tau \end{aligned}$$

where ρ_{12} is the old correlation between y_1 and y_2 and σ_2^2 is the old variation of y_2. The variances of y_3 and y_4 can be calculated analogously.

All the covariances involved can also be calculated. As an example, consider the following:

$$\begin{aligned}
Cov(y_1, y_2) &= E[Cov(y_1, y_2|y_1)] + Cov[E(y_1|y_1), E(y_2|y_1)] \\
&= 0 + Cov[y_1, \mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)] \\
&= Cov[y_1, \rho_{12}\frac{\sigma_2}{\sigma_1}y_1] \\
&= \rho_{12}\frac{\sigma_2}{\sigma_1}Var(y_1) \\
&= \rho_{12}\sigma_2\sigma_1\tau
\end{aligned}$$

where here σ_1 and σ_2 denote the old variances of y_1 and y_2 respectively, and as before, ρ_{12} is the old correlation between y_1 and y_2. Similar calculations can be done for all other covariance terms involving y_1.

To calculate the covariance of y_2 and y_3, we do the following:

$$Cov(y_2, y_3) = E[Cov(y_2, y_3|y_1)] + Cov[E(y_2|y_1), E(y_3|y_1)]$$

By definition,

$$Cov(y_2, y_3|y_1) = \sigma_{23|1}$$

Hence,

$$Cov(y_2, y_3) = \sigma_{23|1} + Cov[\mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), \mu_3 + \rho_{13}\frac{\sigma_3}{\sigma_1}(y_1 - \mu_1)]$$

$$
\begin{aligned}
&= \sigma_{23|1} + \rho_{12}\rho_{13}\frac{\sigma_2\sigma_3}{\sigma_1^2}Var(y_1) \\
&= \sigma_{23|1} + \rho_{12}\rho_{13}\sigma_2\sigma_3\tau
\end{aligned}
$$

where here again all σ_i represent old standard deviation values and the ρ_{ij} represent old correlations. Analogously,

$$Cov(y_2, y_4) = \sigma_{24|1} + \rho_{12}\rho_{14}\sigma_2\sigma_4\tau$$

and

$$Cov(y_3, y_4) = \sigma_{34|1} + \rho_{13}\rho_{14}\sigma_3\sigma_4\tau.$$

We have therefore found all the parameters in the new multivariate normal distribution of $\mathbf{y}$. Denote the new variance-covariance matrix that has been constructed Σ_{new}. Now we are interested in partitioning the variance of the final response, y_4, into various components. That is, we wish to partition the variance at the final stage into components that can be attributed to upstream stages. As implied earlier, when we have a multivariate normal distribution, any two variables from that distribution are bivariate normal, with a variance-covariance matrix given by the appropriate portions of the multivariate normal variance-covariance matrix. For example,

$$\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} \sim N(\theta, \Gamma)$$

where

$$\theta = (\mu_3, \mu_4)^T$$

$$\Gamma = \Sigma_{new34,34}$$

From this, the conditional distribution of y_4 on y_3 can be determined. In fact, the regression coefficient of y_4 on y_3 will be

$$\Gamma_{1,2}/\Gamma_{1,1}$$

and the conditional variance of y_4 on y_3 will be

$$\begin{array}{lrcl} & \Gamma_{2,2} & * & (1-\rho^2) \\ \text{where} & \rho & = & \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}\Gamma_{22}}} \end{array}$$

All other bivariate conditional distributions can be calculated in the same way. Any partitions of variance can therefore be found.

4.1.2 Examples

Piston Data

The above methodology was tried on the first response of the piston production process. Here, the four variables of interest were denoted y_4, y_5, y_6, and y_7. The equations shown above were used to determine the effect of reducing the variation in y_4 by 50%. All sample values were replaced in the equations for true values. In this way, it was determined that

$$\Sigma = \begin{pmatrix} 4.769 & 3.855 & 3.578 & 4.301 \\ 3.855 & 5.208 & 3.258 & 4.523 \\ 3.578 & 3.258 & 4.797 & 3.846 \\ 4.301 & 4.523 & 3.846 & 6.010 \end{pmatrix}$$

(in microns^2). Hence,

$$\Sigma_{567|4} = \begin{pmatrix} 2.092 & 3.655 & 1.050 \\ 3.655 & 2.112 & 6.193 \\ 1.050 & 6.193 & 2.131 \end{pmatrix}$$

Following the computations as above, and recalling that we are setting $\tau = 0.5$, we compute Σ_{new} to be

$$\begin{pmatrix} 2.385 & 1.928 & 1.789 & 2.151 \\ 1.928 & 3.650 & 1.812 & 2.788 \\ 1.789 & 1.812 & 3.455 & 2.233 \\ 2.151 & 2.788 & 2.233 & 4.070 \end{pmatrix}$$

Thus, to get three different partitions of the variance of y_7, we can compute the parameters for the relevant conditional distributions to be as shown in Table 4.1.

Figure 4.2 gives the appropriate decomposition of variance. The total variation of y_7 went from 6.010 microns^2 originally to 4.070 microns^2, a reduction of approximately a third. This is the effect that would have been predicted by multiplying "box y_4" in Figure 4.1(C) by 0.5. Doing the same for Figures 4.1(A) and 4.1(B),

Distribution	Slope	Variation Added (microns^2)
$y_7 \mid y_6$	0.646	2.627
$y_6 \mid y_5$	0.496	2.555
$y_5 \mid y_4$	0.808	2.092
$y_7 \mid y_5$	0.764	1.941
$y_7 \mid y_4$	0.902	2.131

Table 4.1: Parameters for the relevant conditional distributions of piston data

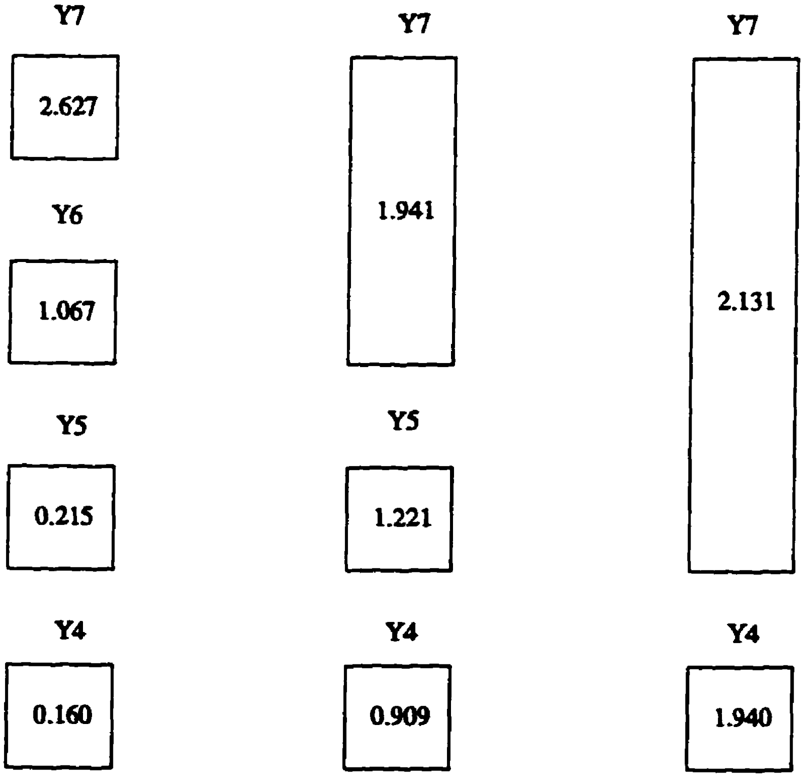

Figure 4.2: Result of reducing the variation of y_4 by 50%

however, would have underestimated the effect of this intervention. As can be seen from Figure 4.2, the last breakdown of variance, Figure 4.2(C), accurately identifies the source of the reduction as the first stage.

Door Hanging Data - AR(1) model

Another data set on which this methodology was used was a car door hanging process. A test of these data reveals that an AR(1) model is adequate. It would be interesting to look at these data in two ways: one in which the AR(1) model is imposed, and the other in which it is not. In the former case, the effect of the intervention on the variance at the last stage can be calculated quickly. Of interest is whether this effect will be the same as estimated in the latter case.

When an AR(1) model is imposed on these data, the following variance matrix is found:

$$\Sigma = \begin{bmatrix} 0.666 & 0.536 & 0.221 & 0.163 \\ 0.536 & 0.870 & 0.360 & 0.265 \\ 0.221 & 0.360 & 0.392 & 0.289 \\ 0.163 & 0.265 & 0.289 & 0.911 \end{bmatrix}$$

This gives the breakdown shown in Figure 4.3. Since an AR(1) model has been assumed, all three partitions of variance are equivalent. This will be true for subsequent analyses as well, and so only the first breakdown will be shown.

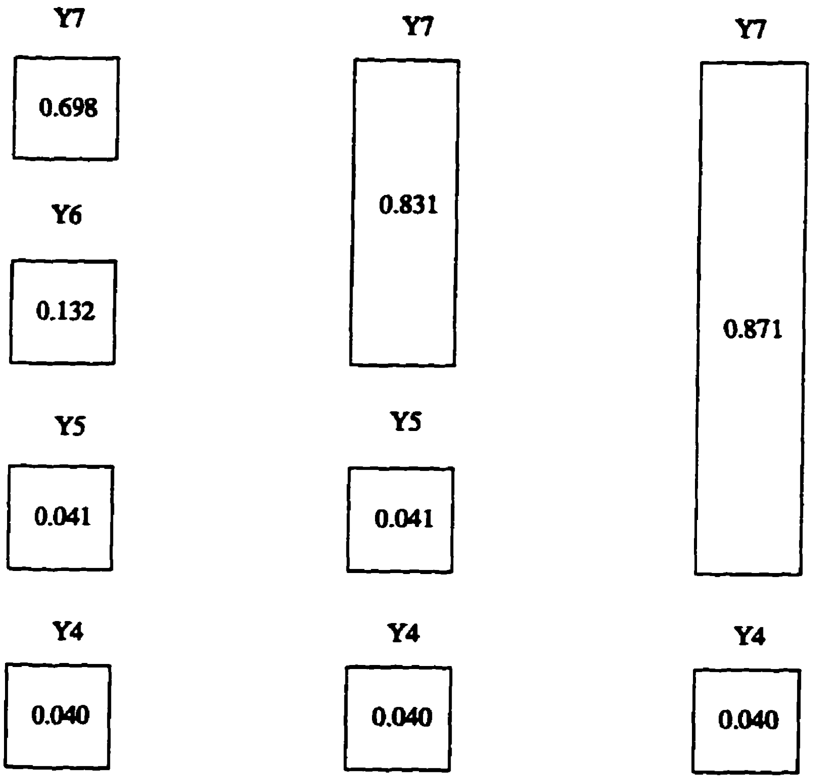

Figure 4.3: Breakdown in door hanging data, with the AR(1) model imposed. Hence, $\rho_{57} = \rho_{56}\rho_{67}$ and $\rho_{47} = \rho_{45}\rho_{56}\rho_{67}$ by construction.

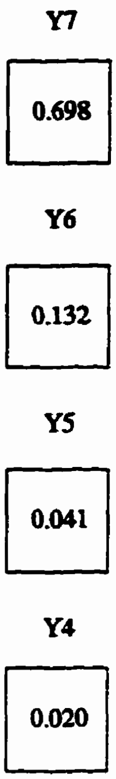

Figure 4.4: Effect of reducing the variation of y_4 by 50%, with the AR(1) model

When the variation in y_4 is reduced by 50%, we get

$$\Sigma_{new} = \begin{bmatrix} 0.333 & 0.268 & 0.111 & 0.082 \\ 0.268 & 0.655 & 0.271 & 0.200 \\ 0.111 & 0.271 & 0.355 & 0.262 \\ 0.082 & 0.200 & 0.262 & 0.891 \end{bmatrix}$$

Hence the final variation has reduced from 0.911 to 0.891 mm^2. The appropriate breakdown of variance is given in Figure 4.4. This is exactly what we would get if we multiply the "y_4 box" in Figure 4.3 by 0.5, as expected.

Door Hanging Data - No model restriction

We can now analyse the same data without imposing the AR(1) model. When this was done, the variance-covariance matrix was found to be

$$\Sigma = \begin{bmatrix} 0.666 & 0.536 & 0.100 & 0.111 \\ 0.536 & 0.870 & 0.360 & 0.128 \\ 0.100 & 0.360 & 0.392 & 0.289 \\ 0.111 & 0.128 & 0.289 & 0.911 \end{bmatrix}$$

Notice that this is very similar to the variance-covariance matrix found previously. The original breakdown of variation here is given in Figure 4.5. Notice that now, the three partitions of variance are not all equal.

When the variation in y_4 was reduced by 50%, it was found that

$$\Sigma_{new} = \begin{bmatrix} 0.333 & 0.268 & 0.050 & 0.056 \\ 0.268 & 0.655 & 0.319 & 0.083 \\ 0.050 & 0.319 & 0.384 & 0.281 \\ 0.056 & 0.083 & 0.281 & 0.902 \end{bmatrix}$$

which gives a decomposition of variance as shown in Figure 4.6. Although the final variation here is very close to that found above, it could only have been predicted from Figure 4.5(C). The effect of this intervention would have been overestimated using Figure 4.5(A) and underestimated using Figure 4.5(B).

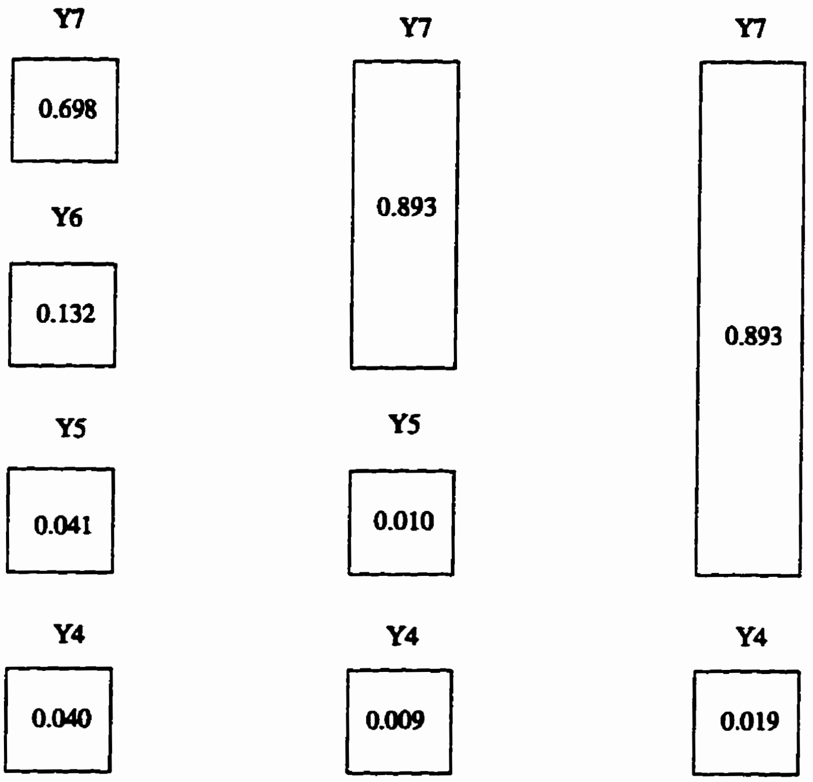

Figure 4.5: Original breakdown of variation in door hanging data, using the AR(1) model: Figure 4.5(A) (left) assumes all stages are observed, Figure 4.5(B) (middle) assumes y_6 is not observed and Figure 4.5(C) (right) assumes y_5 and y_6 are not observed.

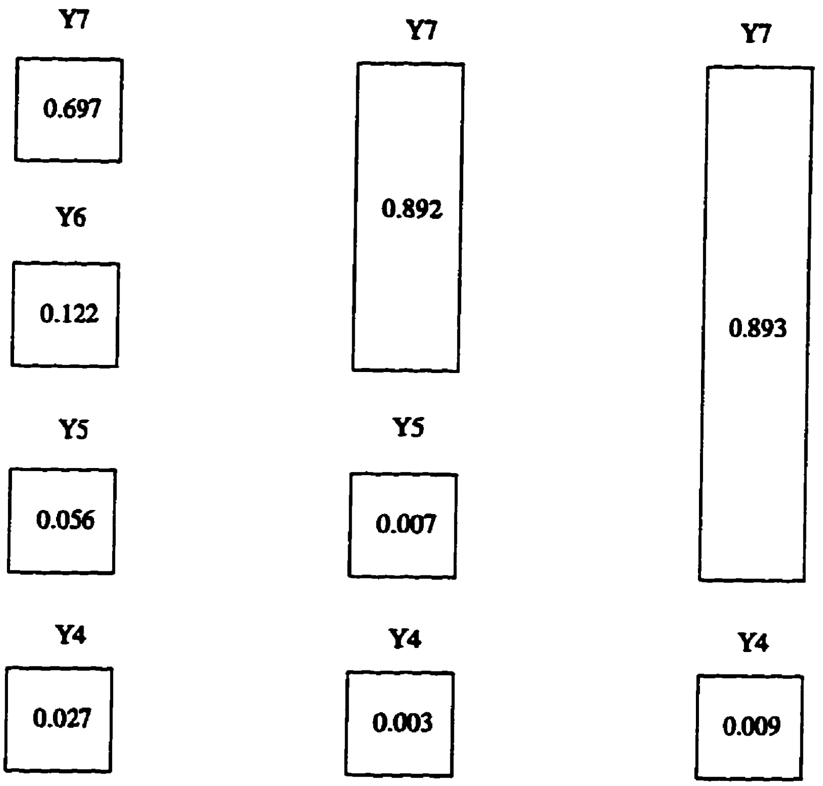

Figure 4.6: Effect of reducing the variation of y_4 by 50%

4.2 Intervention at the second stage

4.2.1 Reducing the added variation

Again, we start with the assumption that the measurements are multivariate normal, with a probability density as given in equation 4.1. Now, let

$$\mathbf{y}_{12} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad \mathbf{y}_{34} = \begin{pmatrix} y_3 \\ y_4 \end{pmatrix}$$

Then, the conditional distribution of $\mathbf{y}_{34}$ given $\mathbf{y}_{12}$ can be found by equation 4.2 and is given by

$$f(\mathbf{y}_{34}|\mathbf{y}_{12}) \sim N(\mu_{34} + \Sigma_{34,12}\Sigma_{12,12}^{-1}(\mathbf{y}_{12} - \mu_{12}), \Sigma_{34,34} - \Sigma_{34,12}\Sigma_{12,12}^{-1}\Sigma_{12,34})$$

The conditional variance-covariance matrix will subsequently be referred to as $\Sigma_{34|12}$. For this analysis, it will be assumed that the above conditional distribution of (y_3, y_4) given (y_1, y_2) does not change when we intervene in the process at the second stage. We will further assume that the marginal distribution of y_1 remains constant, but that the conditional distribution of $y_2|y_1$ changes. Hence, the marginal distribution of y_1 will be

$$y_1 \sim N(\mu_1, \sigma_1^2).$$

Whereas it was true that

$$y_2|y_1 \sim N(\mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), \sigma_2^2(1-\rho_{12}^2)),$$

we will now consider that

$$y_2|y_1 \sim N(\mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), \tau\sigma_2^2(1-\rho_{12}^2)).$$

If τ is less than one, we are considering the situation in which stage two is adding less variation to the process. This may happen if the process at stage two is adjusted based on y_1.

To determine what happens now, note that the variance of y_1 has not changed, but that the variance of y_2 has, according to the following calculations:

$$\begin{aligned} Var(y_2) &= E(Var(y_2|y_1)) + Var(E(y_2|y_1)) \\ &= E(\tau\sigma_2^2(1-\rho_{12}^2)) + Var(\mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)) \\ &= \tau\sigma_2^2(1-\rho_{12}^2) + \rho_{12}^2\frac{\sigma_2^2}{\sigma_1^2}Var(y_1) \\ &= \tau\sigma_2^2(1-\rho_{12}^2) + \rho_{12}^2\sigma_2^2 \\ &= \sigma_2^2(\tau - \tau\rho_{12}^2 + \rho_{12}^2) \end{aligned}$$

Also,

$$\begin{aligned} Cov(y_1, y_2) &= E(Cov(y_1, y_2|y_1)) + Cov(E(y_1|y_1), E(y_2, y_1)) \\ &= Cov(y_1, \mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)) \end{aligned}$$

$$\begin{aligned} &= \rho_{12}\frac{\sigma_2}{\sigma_1}Var(y_1) \\ &= \rho_{12}\sigma_2\sigma_1 \end{aligned}$$

where here σ_2 denotes the old standard deviation of y_2.

We can now calculate the variances of y_3 and y_4. This can be done by applying the conditional variance formula to vectors in the following way:

$$\begin{aligned} Var\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} &= E(Var\left(\begin{array}{c|c} y_3 & y_1 \\ y_4 & y_2 \end{array}\right)) + Var(E\left(\begin{array}{c|c} y_3 & y_1 \\ y_4 & y_2 \end{array}\right)) \\ &= \Sigma_{34|12} + Var(\mu_{34} + \Sigma_{34,12}\Sigma^{-1}_{12,12}\mathbf{y}_{12} - \Sigma_{34,12}\Sigma^{-1}_{12,12}\mu_{12}) \\ &= \Sigma_{34|12} + \Sigma_{34,12}\Sigma^{-1}_{12,12}Var(\mathbf{y}_{12})\Sigma^{-1}_{12,12}\Sigma_{12,34} \qquad (4.3) \end{aligned}$$

Since we have constructed the variance-covariance matrix of $\mathbf{y}_{34}$ above, we get that

$$Var\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} = \Sigma_{34|12} + \Sigma_{34,12}\Sigma^{-1}_{12,12}\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2(\tau - \tau\rho_{12}^2 + \rho_{12}^2) \end{pmatrix}\Sigma^{-1}_{12,12}\Sigma_{12,34}$$

It remains only to find the covariance terms between $\mathbf{y}_{12}$ and $\mathbf{y}_{34}$. This can be done in the manner of the following calculation:

$$\begin{aligned} Cov(\mathbf{y}_{12}, \mathbf{y}_{34}) &= \begin{bmatrix} Cov(y_1, y_3) & Cov(y_1, y_4) \\ Cov(y_2, y_3) & Cov(y_2, y_4) \end{bmatrix} \\ &= E(Cov(\mathbf{y}_{12}, \mathbf{y}_{34}|\mathbf{y}_{12})) + Cov(E(\mathbf{y}_{12}|\mathbf{y}_{12}), E(\mathbf{y}_{34}|\mathbf{y}_{12})) \\ &= Cov(\mathbf{y}_{12}, \mu_{34} + \Sigma_{34,12}\Sigma^{-1}_{12,12}\mathbf{y}_{12} - \Sigma_{34,12}\Sigma^{-1}_{12,12}\mu_{12}) \\ &= Var(\mathbf{y}_{12})\Sigma^{-1}_{12,12}\Sigma_{12,34} \end{aligned}$$

After computing the above quantities, it is possible to reconstruct the variance-covariance of the new multivariate normal distribution of (y_1, y_2, y_3, y_4). Denote this matrix Σ_{new}. As was done in the previous section, we can now construct the bivariate distributions that we may be interested in to partition the variance of y_4 into various components.

4.2.2 Examples

Piston Data

For the piston example, we will use the above methodology to determine the effect of reducing the variation added at the second stage. For these data, it was found that

$$\Sigma_{67|45} = \begin{pmatrix} 2.048 & 4.359 \\ 4.359 & 1.604 \end{pmatrix}$$

(again in microns^2). Letting $\tau = 0.5$, that is having reduced the variation added at this stage by 50%, we found that

$$\Sigma_{new} = \begin{pmatrix} 4.769 & 3.855 & 3.578 & 4.301 \\ 3.855 & 4.162 & 3.075 & 4.002 \\ 3.578 & 3.075 & 4.765 & 3.755 \\ 4.301 & 4.002 & 3.755 & 5.747 \end{pmatrix}$$

This resulting partitions of variance of y_7 are given in Figure 4.7. Hence, the total variation in y_7 has been reduced to 5.747 microns^2, a reduction not nearly as significant as that seen in the previous section.

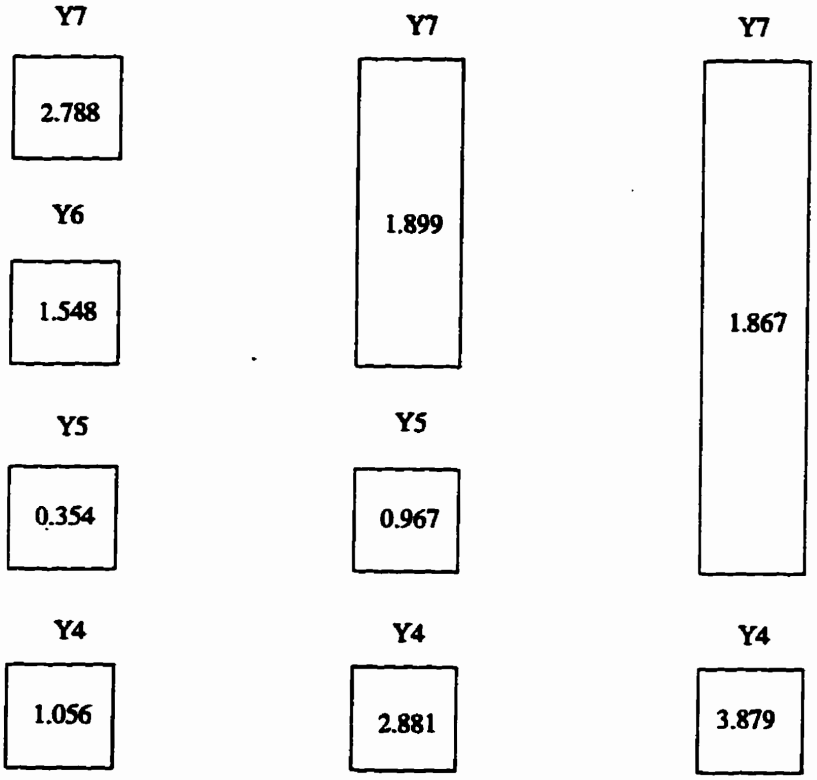

Figure 4.7: Effect of reducing the variation added at y_5 by 50%

If we were to predict the effect of this intervention from Figure 4.1, Figure 4.1(A) would have predicted it accurately, whereas Figure 4.1(B) would have overestimated the reduction in variance. It is unclear whether the former is a lucky coincidence, since the breakdown resulting from multiplying the y_5 box in Figure 4.1(A) is not the same as that shown in Figure 4.7(A). In this figure, none of the three partitions clearly shows the source of the reduction of variation, although the last one does correctly identify it as not having originated from y_4.

Door hanging data - AR(1) model

In the case of the door hanging data with an AR(1) model imposed, the consequence of reducing the added variation at the second stage by 50% is to decrease the final variation from 0.911 mm^2 to 0.891 mm^2. The breakdown of the variation appears in Figure 4.8. This is the breakdown achieved by multiplying the "y_5 box" in Figure 4.3 by 0.5, exactly as expected.

Door hanging data - No model restriction

For the same data with no model restrictions, the effect of reducing the added variation at the second stage by 50% is to give a final variation of 0.910 mm^2. In other words, such an intervention has essentially no effect. This could not have been predicted from any of the breakdowns given in Figure 4.5.

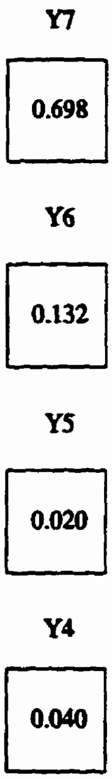

Figure 4.8: Effect of reducing the variation added at y_5 by 50%, AR(1) model

4.2.3 Changing the slope of the second stage on the first stage

We have seen how to analyse the effect of reducing the variation added at the second stage on the variation of the final stage. Another way of intervening at the second stage, however, is to change the slope of the second stage on the first stage. If we reduce this slope, the effect should be to reduce the variation transmitted from the first stage, and hence to reduce the overall variation. To analyse this situation, we will assume the same situation as above, namely that the conditional distribution of (y_3, y_4) on (y_1, y_2) remains constant. We will again assume that the marginal distribution of y_1 hasn't changed, but this time we will assume that the conditional distribution of y_2 on y_1 has changed in the following way:

$$y_2|y_1 \sim N(\alpha(\mu_2 - \rho_{12}\frac{\sigma_2}{\sigma_1}\mu_1) + \tau\rho_{12}\frac{\sigma_2}{\sigma_1}y_1, \sigma_2^2(1-\rho_{12}^2))$$

Note that this change will affect the means of y_2, y_3 and y_4, but we will assume that the process can subsequently be retargeted.

The following can be derived easily

$$\begin{aligned} Var(y_1) &= \sigma_1^2 \\ Var(y_2) &= \sigma_2^2(1-\rho_{12}^2+\tau^2\rho_{12}^2) \\ Cov(y_1, y_2) &= \tau\rho_{12}\sigma_1\sigma_2 \end{aligned}$$

These three equations can be used to construct the variance-covariance matrix of

$\mathbf{y}_{34}$, and the equations above relating to quantities involving y_3 and y_4 can be used here. Hence, for example,

$$Var\begin{pmatrix} y_3 \\ y_4 \end{pmatrix} = \Sigma_{34|12} + \Sigma_{34,12}\Sigma_{12,12}^{-1}\begin{pmatrix} \sigma_1^2 & \tau\rho_{12}\sigma_1\sigma_2 \\ \tau\rho_{12}\sigma_1\sigma_2 & \sigma_2^2(1-\rho_{12}^2+\tau^2\rho_{12}^2) \end{pmatrix}\Sigma_{12,12}^{-1}\Sigma_{12,34}$$

analogously to equation 4.3 and

$$Cov(\mathbf{y}_{12}, \mathbf{y}_{34}) = Var(\mathbf{y}_{12})\Sigma_{12,12}^{-1}\Sigma_{12,34}$$

The construction of Σ_{new} and the subsequent partitioning of the variation of y_4 proceeds as usual.

4.2.4 Examples

Piston Data

The above methodology gives the following for the piston example when $\tau = 0.5$

$$\Sigma_{new} = \begin{pmatrix} 4.769 & 1.928 & 3.241 & 3.334 \\ 1.928 & 2.871 & 1.676 & 2.397 \\ 3.241 & 1.676 & 4.315 & 2.885 \\ 3.334 & 2.397 & 2.885 & 4.461 \end{pmatrix}$$

which results in the partition of variance shown in Figure 4.9. Here the final variance has been reduced to 4.461 microns^2, which is comparable to the change that occurred when the variance of y_4 was reduced by 50%. In Figure 4.9, the last

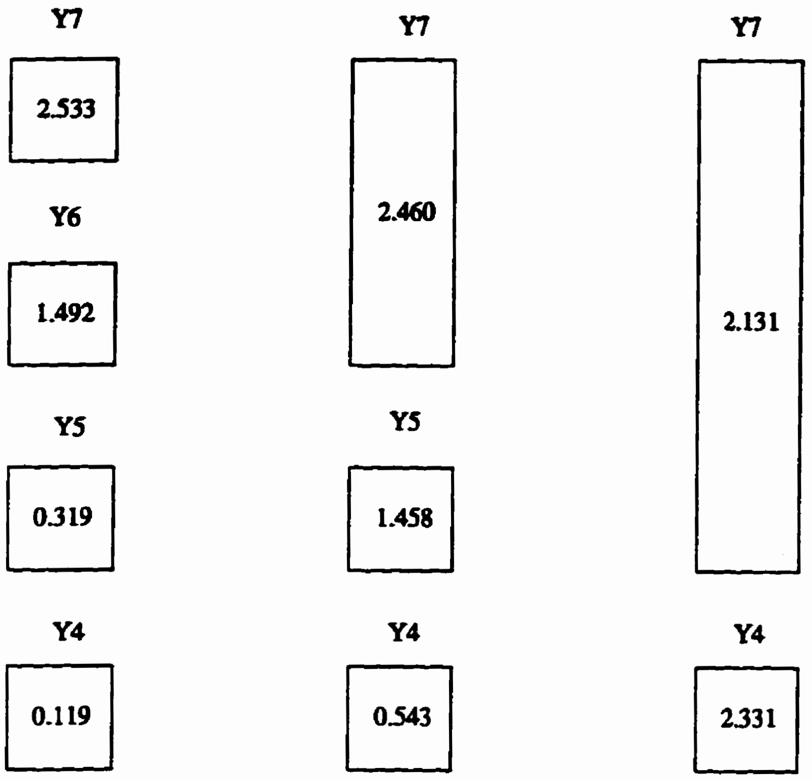

Figure 4.9: Effect of reducing the slope of y_5 on y_4 by 50%

partition of variance accurately reflects the source of the reduction.

To predict the effect of this intervention from Figure 4.1, we would have multiplied the y_4 box in Figures 4.1(A) and 4.1(B) by $\frac{1}{4}$. (It would have been unclear how to predict the effect of changing the slope of y_5 on y_4 from Figure 4.1(C)). Neither of these two values would have produced the 4.461 microns^2 found here. Furthermore, the two values found, 4.245 microns^2 and 5.422 microns^2, are quite different from each other.

Door Hanging Data - AR(1) model

When the slope of the second stage on the first stage is decreased, the effect in this case is to decrease the final variation to 0.881 mm^2. The resulting breakdown of variation is that found by multiplying the y_4 box in Figure 4.3 by $(\frac{1}{2})^2$, or by 0.25.

Door Hanging Data - No model restriction

The same intervention is predicted to have less of an effect in reducing the variation when the AR(1) model is not imposed. Here, the final variation is 0.904 mm^2. Figure 4.5(C) comes close to predicting this value when the y_4 box is multiplied by $\frac{1}{4}$, but the prediction given by Figure 4.5(A) overestimates the amount of reduction occurring.

4.3 Intervention at the third stage

We can consider intervening in the process at the third stage, as we have done for the two previous stages. Again in this case we can investigate two types of interventions: reducing the added variation or reducing the slope of the regression of the third stage on either, or both, of the two previous stages. The calculations required to investigate these types of scenarios are similar to those shown for other stages, and will not be given here in the interest of brevity. Both the piston data and the door hanging data were used to investigate the following scenarios: reducing the variation added at the third stage by one half, reducing the slope of the third stage on the second stage by half while keeping the slope of the third stage on the first stage constant, and reducing the slope of the third stage on the first stage by half while keeping the slope of the third stage on the second stage constant. A summary of these results is given in Table 4.2. One curious result that can be seen in this table is that in some scenarios investigated for the door hanging data, the effect of an intervention is to cause the variance at the final stage to increase. Variation in these estimates has not been discussed, however, and it could be that the increase in variance is not significant.

4.4 Conclusions

When the data of a process can be modeled adequately with an AR(1) model, it is easy to assess the effect of an intervention in the process. When an AR(1) model does not fit the data, it can be seriously misleading to use it to assess how an intervention might effect the variance at the final stage. In this case, an appropriate model might be the full multivariate normal model. Assessing the effect of an intervention is less intuitive than with the more restricted model, but can be done by making some assumptions regarding the conditional distributions of subsequent stages.

For the piston example discussed, the most significant change could be made by either reducing the variation of the first stage, or by reducing the slope of the second stage on the first stage. Other interventions would not be as efficient in reducing the variation of the final response. For the door hanging example, the most significant change could be made by reducing the variation added at the third stage. See Table 4.2 for a summary of these changes. In both cases, these recommendations can now

Intervention	Value	Piston Data	Door Hanging with Model Restriction	Door Hanging No Model Restriction
Total Variance		6.010	0.911	0.911
50% reduction in var'n of 1st stage	Variance Percent Decrease	4.070 32.3	0.891 2.2	0.902 1.0
50% reduction in var'n added at 2nd stage	Variance Percent Decrease	5.747 4.4	0.891 2.2	0.910 0.1
50% reduction in slope of 2nd stage on 1st	Variance Percent Decrease	4.461 25.8	0.881 3.3	0.904 0.8
50% reduction in var'n added at 3rd stage	Variance Percent Decrease	5.934 1.3	0.845 7.2	0.757 16.9
50% reduction in slope of 3rd stage on 2nd	Variance Percent Decrease	5.843 2.8	0.851 6.5	0.946 -3.8
50% reduction in slope of 3rd stage on 1st	Variance Percent Decrease	5.473 8.9	0.911 0	0.995 -9.2

Table 4.2: Summary of effect of interventions at various stages

be passed on to the engineers in charge of the process, in the hope of improving quality.

Chapter 5

Multivariate Data

With the growing complexity of processes seen in industry, and the availability of machines to take many measurements on the process quickly, multivariate data are becoming the norm. Methods are required that can handle correlated data and make use of all its features. The variance transmission problem is to identify those opportunities that have the greatest potential for variation reduction. Since the data are multivariate, variance reduction might be desirable equally at all measurements, or it may be that variance reduction is more valuable at some measurements than at others, or there may be an interaction between various measurements. The priority of variation reduction at different measurements can be quantified by a loss function.

In this chapter, we review three papers that address issues relevant to multivariate data in multi-stage processes, and discuss some of the issues involved in this analysis. We also suggest some other approaches that might be taken, and discuss issues that have yet to be addressed.

We will assume through this discussion that a multivariate AR(1) normal model is appropriate for the data. Hence the model can be written as

$$\begin{aligned} \mathbf{Y}_1 &= \mu_1 + \epsilon_1 \\ \mathbf{Y}_i &= \mathbf{A}_i + \mathbf{B}_i \mathbf{Y}_{i-1} + \epsilon_i \qquad i = 2, \ldots, k \end{aligned} \tag{5.1}$$

where $\mathbf{Y}$ is a vector of m measurements. Also $\mathbf{A}_i$ is a vector and $\mathbf{B}_i$ is a matrix, generalizations of α_i and β_i, respectively. The total variance matrix of $\mathbf{Y}_i$ will be denoted by Σ_i and the added variance at that stage is $\Sigma_{i,A}$, i.e. $\Sigma_{i,A} = \mathrm{Var}(\epsilon_i)$. The variance transmitted from previous stages is given by Σ_i - $\Sigma_{i,A}$. These are now $m * m$ matrices.

5.1 Review

The three papers that will be reviewed in this section are Lawless, MacKay and Robinson (1996), Fong and Lawless (1996) and Xie, Yang and He (1994). Lawless et al. deal with multivariate data by considering each measurement separately, and using the univariate AR(1) model to analyse variance transmission. Fong and Lawless use the generalized AR(1) model given in equation (5.1), and allow for missing data and measurement error. Xie et al. use two approaches in their paper: they first define loss functions that they use at the various stages of the process, and then they consider principle components analysis.

The approaches described above will be demonstrated on some hood fitting data. This is data in which the hoods on 19 cars were measured at four places:

two in the front and two in the rear, one along each side. There was no missing data, and measurement error will be ignored for the subsequent analyses. Each measurement represents a deviation from nominal. There were four stages involved in the installation of the hood: 1) hanging the hood (HANG), 2) painting the hood and the rest of the car (PAINT), 3) installing hardware such as the hood latch (HARD) and 4) adjusting or "finessing" the hood for better fit (FIN). In this case, variation reduction is equally important at all four of the measurements.

Lawless, MacKay and Robinson use a univariate analysis for each of the measurements of interest when dealing with a multivariate data situation. This means considering each measurement independently and modeling it with a univariate AR(1) model. When this is done for the hood data, the results are as given in Table 5.1. The results indicate that for the two front measurements, most of the variation is coming from the finesse stage, while for the two rear measurements, most of the variation is coming from the HANG stage. It should be pointed out, however, that for the left front measurement, the variance at the third stage is roughly twice the variance at the final stage. Hence, while all of the variation present at the final stage is added there, this is an improvement over eliminating the last stage altogether. In all cases, very little variation is contributed by the two intermediate stages of the process.

The strength of this method of analysis is its interpretability. The results given here can be applied directly to the process. The drawback is, of course, that this method does not take into account the correlation between the measurements. Hence, caution needs to be exercised in intervening in the process to effect one

Stage	Left Front	Right Front	Left Rear	Right Rear
FIN	0.556	0.398	0.105	0.041
HARD	0	0.083	0.051	0.014
PAINT	0	0.007	0.023	0.016
HANG	0	0.013	0.725	0.857
TOTAL	0.556	0.500	0.903	0.929

Table 5.1: Variance transmission of hood data using univariate analysis

measurement, since such an intervention may have unforeseen results on other measurements. For example, it is conceivable that in attempting to reduce the variation added at the finesse stage, some adjustment is made that makes the rear measurement values at this stage less dependent on those values at the previous stage. This would result in the fortunate situation in which variation transmitted from previous stages would be reduced, and the variation at the HANG stage need not be adjusted.

Fong and Lawless deal with a multivariate AR(1) model in their analysis, and use a Kalman filtering approach to handle missing data and measurement error. This approach is more efficient in terms of computer time than using, for example, a simplex search algorithm to compute maximum likelihood estimates. Assuming that measurement error is negligible for the hood data, we get the results given in Table 5.2 using the multivariate AR(1) model.

Although this approach makes use of the full structure of the multivariate data, the results are hard to interpret. Note that none of the estimated correlations between measurements is extremely high. They are not, however, negligible and since we are interested in reducing the variance at all the measurements, it is difficult

Stage	$\hat{\Sigma}_i$	$\hat{\Sigma}_{i,A}$
HANG	$\begin{bmatrix} 0.69 & 0.05 & -0.21 & -0.07 \\ 0.05 & 0.38 & 0.25 & -0.03 \\ -0.21 & 0.25 & 1.16 & 0.01 \\ -0.07 & -0.03 & 0.01 & 0.60 \end{bmatrix}$	
PAINT	$\begin{bmatrix} 0.82 & -0.13 & -0.33 & -0.13 \\ -0.13 & 0.22 & 0.38 & -0.11 \\ -0.33 & 0.38 & 1.23 & 0.01 \\ -0.13 & -0.11 & 0.01 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.09 & -0.57 & 0.35 & 0.54 \\ -0.57 & 0.07 & -0.33 & -0.22 \\ 0.35 & -0.33 & 0.03 & 0.25 \\ 0.54 & -0.22 & 0.25 & 0.01 \end{bmatrix}$
HARD	$\begin{bmatrix} 1.12 & -0.28 & -0.02 & 0.19 \\ -0.28 & 0.52 & 0.29 & -0.50 \\ -0.02 & 0.29 & 1.21 & -0.02 \\ 0.19 & -0.50 & -0.02 & 0.79 \end{bmatrix}$	$\begin{bmatrix} 0.25 & -0.12 & 0.32 & -0.22 \\ -0.12 & 0.29 & -0.30 & -0.36 \\ 0.32 & -0.30 & 0.06 & 0.46 \\ -0.22 & -0.36 & 0.46 & 0.01 \end{bmatrix}$
FIN	$\begin{bmatrix} 0.56 & -0.39 & -0.38 & -0.26 \\ -0.39 & 0.50 & 0.38 & -0.22 \\ -0.38 & 0.38 & 0.90 & -0.20 \\ -0.26 & -0.22 & -0.20 & 0.93 \end{bmatrix}$	$\begin{bmatrix} 0.46 & -0.40 & -0.29 & -0.40 \\ -0.40 & 0.34 & 0.03 & 0.38 \\ -0.29 & 0.03 & 0.08 & 0.52 \\ -0.40 & 0.38 & 0.52 & 0.03 \end{bmatrix}$

Table 5.2: Multivariate AR(1) model results. The off diagonals are correlations, while the diagonal elements of the matrices are variances

to tell what type of intervention would be most beneficial. More will be said about this in the next section.

Xie et al. (1994) take a different approach to the analysis of data from a multistage multivariate process. They define two effects in such a process: the certain effect, which results in the same deformation pattern on each item at each stage, and the uncertain effect, which is essentially a random effect on each item. They also reduce the dimension of the data by using the geometry of the product to define sections. The certain effect is quantified by the mean square of the sample mean deviation (MSMD). Let Y_{ijk} denote the deviation from nominal of the i-th point of the k-th item at the j-th stage ($i = 1, \ldots, n$; $j = 1, \ldots, L$; $k = 1, \ldots, m$). Then

$$MSMD = \frac{1}{n_p} \sum_{i \in S_p} (\frac{1}{m} \sum_k Y_{ijk})^2$$

where S_p is the measuring point set of section p; n_p is the number of points in that set. Similarly, the uncertain effect is quantified by the average variance of the deviation (AVD), and is given by

$$\begin{aligned} AVD &= \frac{1}{n_p} \sum_{i \in S_p} (\frac{1}{m}) \sum_k (Y_{ijk} - \bar{Y}_{ij*})^2 \\ \text{where} \qquad \bar{Y}_{ij*} &= \frac{1}{m} \sum_k Y_{ijk} \end{aligned}$$

Both the MSMD and the AVD are computed at each stage of the process. If we define the average loss at a stage and at the p-th section to be

$$\frac{1}{n_p} \sum_{i \in S_p} \frac{1}{m} \sum_k Y_{ijk}^2$$

then the average loss is the sum of the MSMD and the AVD.

Once these values have been calculated, the authors continue by doing a variety of principle components analyses. They first look at the principle components for each section and each stage. They then look at the principle components analysis on $\bar{Y}_{ij*}$ and $(Y_{ijk} - \bar{Y}_{ij*})$, to determine modes of variation in the certain effect and in the uncertain effect. They contrast this to the principle components analysis given by combining all of the data together.

Overall, this approach seems to be ad hoc. It contributes little towards an understanding of the process. For example, the principle components analysis of the certain effect groups the data over the different stages together. Thus, while the first principle component of such an analysis allows us to determine a direction in which a large amount of the variation is occurring, it does not explain where this variation is coming from. The same can be said for the principle component analysis of the uncertain effect. Conversely, while the principle components analysis done at each stage allows determination of the variation modes at each stage, it does not distinguish between certain and uncertain effects. Further, there is no way of determining whether the variation mode at a certain stage is being transmitted through to the final stage. Hence, it is very difficult to relate these results back to the process in a meaningful way.

The MSMD and AVD values were calculated for the hood data. Here, the first section was defined to be the front two measurements, and the rear measurements were defined to be the second section. Plots of these values are given in Figures 5.1 and 5.2. These plots indicate that there is large variation in the certain effect at

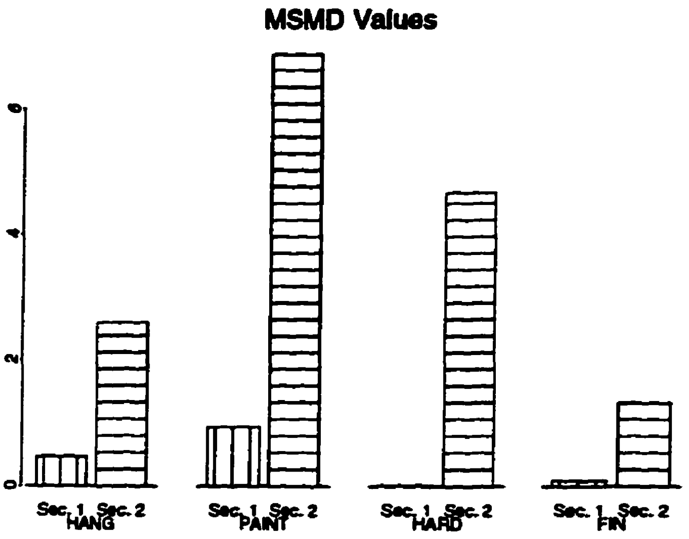

Figure 5.1: Calculated values for the hood data

section two and at the PAINT stage. This implies that there are factors at that stage that are having a large impact on the deformation of the rear of the hood. There is relatively little mean deviation for the first section. This means that there are no large factors that are affecting the process at the front of the hood in a consistent manner. Further, there appears to be about the same amount of variation in the uncertain effect in both sections and at all stages. The implication is that there is something to be gained from focusing on reducing variation at previous stages. Notice, though, that there is no consideration given to how variation at previous stages affects the variation at the last stage. This omission could seriously mislead the investigator. The results from this analysis should be compared to those found using the multivariate AR(1) model.

Generally speaking, there appears to be more work necessary in understanding the multivariate multi-stage problem. The next section proposes other approaches

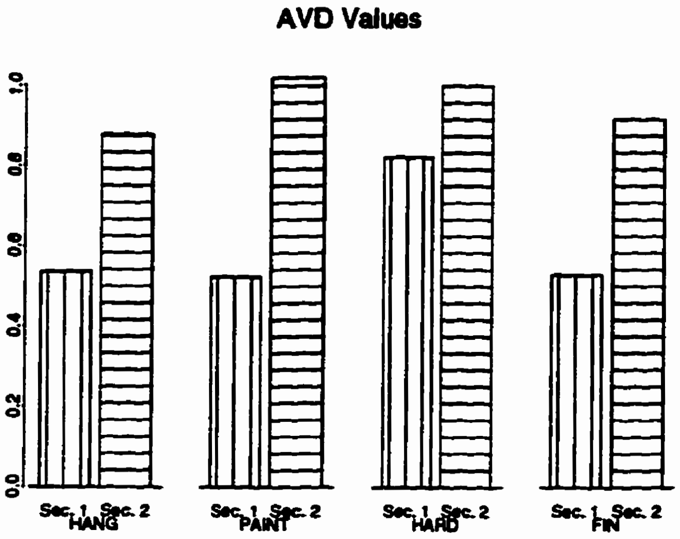

Figure 5.2: Calculated values for the hood data

that might be considered.

5.2 Other Approaches

5.2.1 Modeling the Intervention

In the case of multivariate data that adhere to the generalized AR(1) model, an approach can be taken that models the effect of an intervention and considers this effect with a univariate loss function. For example, consider a two stage process in which bivariate data are observed. Then we can describe this situation as

$$\mathbf{Y}_1 = \begin{bmatrix} Y_{11} \\ Y_{21} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{11} \\ \mu_{21} \end{bmatrix}, \Sigma_1 = \begin{bmatrix} \sigma^2_{1,1} & \sigma^2_{12,1} \\ \sigma^2_{12,1} & \sigma^2_{2,1} \end{bmatrix} \right)$$

$$\mathbf{Y}_2|\mathbf{Y}_1 = \begin{bmatrix} Y_{12} \\ Y_{21} \end{bmatrix} \Big| \begin{bmatrix} Y_{11} \\ Y_{21} \end{bmatrix} \sim N\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \mathbf{Y}_1, \right.$$
$$\left. \Sigma_{2A} = \begin{bmatrix} \sigma^2_{1,2A} & \sigma^2_{12,2A} \\ \sigma^2_{12,2A} & \sigma^2_{2,2A} \end{bmatrix} \right)$$

Here,

$$Var(\mathbf{Y}_2) = BVar(\mathbf{Y}_1)B^T + \Sigma_{2A}$$

We can define in this case a loss function that penalizes the variance of each measurement in $\mathbf{Y}_2$ equally; for example, the average of the variances. This gives us that $L = \frac{1}{2}\text{trace}(Var(\mathbf{Y}_2))$.

We can study various interventions in the process and their effect on the above loss. First, consider reducing the variance at the first stage, which we can model in a general way as changing Σ_1 to

$$\Sigma_1^* = \begin{bmatrix} \tau_1\sigma^2_{1,1} & \tau_{12}\sigma^2_{12,1} \\ \tau_{12}\sigma^2_{12,1} & \tau_2\sigma^2_{2,1} \end{bmatrix}$$

The values of τ_1, τ_{12} and τ_2 will be determined by the way in which the intervention will occur, and the engineering perspective as to what these values should be. We will assume that the conditional distribution of the second stage on the first stage will be unaffected. The variance of $\mathbf{Y}_2$ will be

$$\Sigma_{2new} = B\Sigma_1^*B^T + \Sigma_{2A}$$

The loss is now $L = \frac{1}{2}\text{trace}(\Sigma_{2new})$.

Another intervention in the process could be in the way that $\mathbf{Y}_2$ depends on $\mathbf{Y}_1$. Hence, consider changing B to B^*, where

$$B^* = \begin{bmatrix} \tau_{11}b_{11} & \tau_{12}b_{12} \\ \tau_{21}b_{21} & \tau_{22}b_{22} \end{bmatrix}$$

Again, the values that τ will take should be determined by engineering knowledge. In this case, the loss changes to $L = \frac{1}{2}\text{trace}(B^*\Sigma_1 B^{*T} + \Sigma_{2A})$. Similar calculations can be done if we consider changing Σ_{2A}.

We can apply this methodology to the hood data introduced earlier. Recall that the variance matrices for these data at each stage are given in Table 5.2. If we take the average of the variances at the last stage to be the loss function, then the current loss is 0.722. Consider first changing the variance-covariance matrix added at the last stage, $\Sigma_{FIN,A}$, to

$$\begin{bmatrix} 0.5\sigma_{1,A}^2 & 0.25\sigma_{12,A}^2 & 0.25\sigma_{13,A}^2 & 0.25\sigma_{14,A}^2 \\ 0.25\sigma_{12,A}^2 & 0.5\sigma_{2,A}^2 & 0.25\sigma_{23,A}^2 & 0.25\sigma_{24,A}^2 \\ 0.25\sigma_{13,A}^2 & 0.25\sigma_{23,A}^2 & 0.5\sigma_{3,A}^2 & 0.25\sigma_{34,A}^2 \\ 0.25\sigma_{14,A}^2 & 0.25\sigma_{24,A}^2 & 0.25\sigma_{34,A}^2 & 0.5\sigma_{4,A}^2 \end{bmatrix}$$

This corresponds to reducing the marginal variances of the four measurements by one half, but not changing the correlations between them. If we do this, the loss is reduced to 0.609.

Another intervention to consider is changing $B_{FIN,HARD}$ to

$$\begin{bmatrix} 0.5b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & 0.5b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & 0.5b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & 0.5b_{44} \end{bmatrix} .$$

This intervention has the effect of reducing the loss to 0.392.

A third kind of intervention is changing the variance at the previous stage. We will do this in the same way as we did for the final stage, namely by reducing the variances by one half and the covariances by one quarter. This intervention reduces the loss to 0.469. It should perhaps be pointed out here that if this intervention had proven to be the most effective, the variance transmission methodology could be used to determine the best way of reducing variation at this stage.

Of the interventions considered here, the most effective proved to be by changing the conditional expectation of $\mathbf{Y}_2$ on $\mathbf{Y}_1$. Given the results of the univariate analysis shown earlier, this is somewhat surprising, since the two front measurements had almost all of their variation added at the last stage. Clearly there are other interventions that could have been considered. In a practical situation, the types of interventions investigated should be dictated by engineering knowledge of the process.

5.2.2 Principle Components Analysis

Another approach that might be used to deal with this problem is applying the AR(1) model to principle components of the data. The principle components will be different at each stage. However, if we can isolate a few modes of variation at the final stage, and if these are interpretable, then determining those stages which are contributing to the modes of variation will be useful. Yang (1996) gives an example from the automotive industry of the use of principle components to reduce variation.

A principle components analysis was done for the hood data, and the results are given in Table 5.3. The first two principle components explain 78% of the variation at the final stage. Suppose we use these two components to create two new variables, COMB1 and COMB2, where these are linear combinations of the original variables, given by

$$
\begin{aligned}
COMB1 &= -LF + RF + 2LR - RR \\
\text{and} \qquad COMB2 &= -5LF + RF + 2LR + 8RR
\end{aligned}
$$

where LF, RF, LR and RR are the left front, right front, left rear and right rear measurements, respectively. These two variables have a very small correlation at the final stage.

Now, the AR(1) analysis of variation transmission can be applied to the new variables. Figure 5.3 shows the scatter plots for the first variable over the four stages, and a bar plot of the variation added and transmitted from the different

	PC1	PC2	PC3	PC4
Std. Dev	1.15	1.02	0.66	0.50
Rotation	-0.29	-0.53	0.37	0.71
	0.41	0.09	-0.68	0.59
	0.74	0.22	0.61	0.15
	-0.44	0.81	0.15	0.35

Table 5.3: Principle components analysis of hood data

stages. Figure 5.4 shows the same for the second variable. Both these plots indicate that the first and last stages are the best opportunities for variance reduction. Note that the above analysis is only useful if the engineers on the process can interpret the new variables created, COMB1 and COMB2, in a meaningful way. COMB1, for example, appears to be a measure of the tilt of the hood on the diagonal axis.

Clearly, there is some difficulty in dealing with multi-stage multivariate processes. Issues such as how to use principle component regression in variance transmission analysis have yet to be investigated. In general, there seems to be a trade-off between being able to use all of the available data and simplicity of interpretation.

5.3 Discussion

More work needs to be done in the area of multi-stage multivariate processes. Some graphical methods of portraying data in these cases would be very useful, especially if they could be used as a diagnostic tool for model checking. Also useful would be methods that deal with departures from the AR(1) model, such as the general multivariate normal model. The intervention modeling approach could perhaps

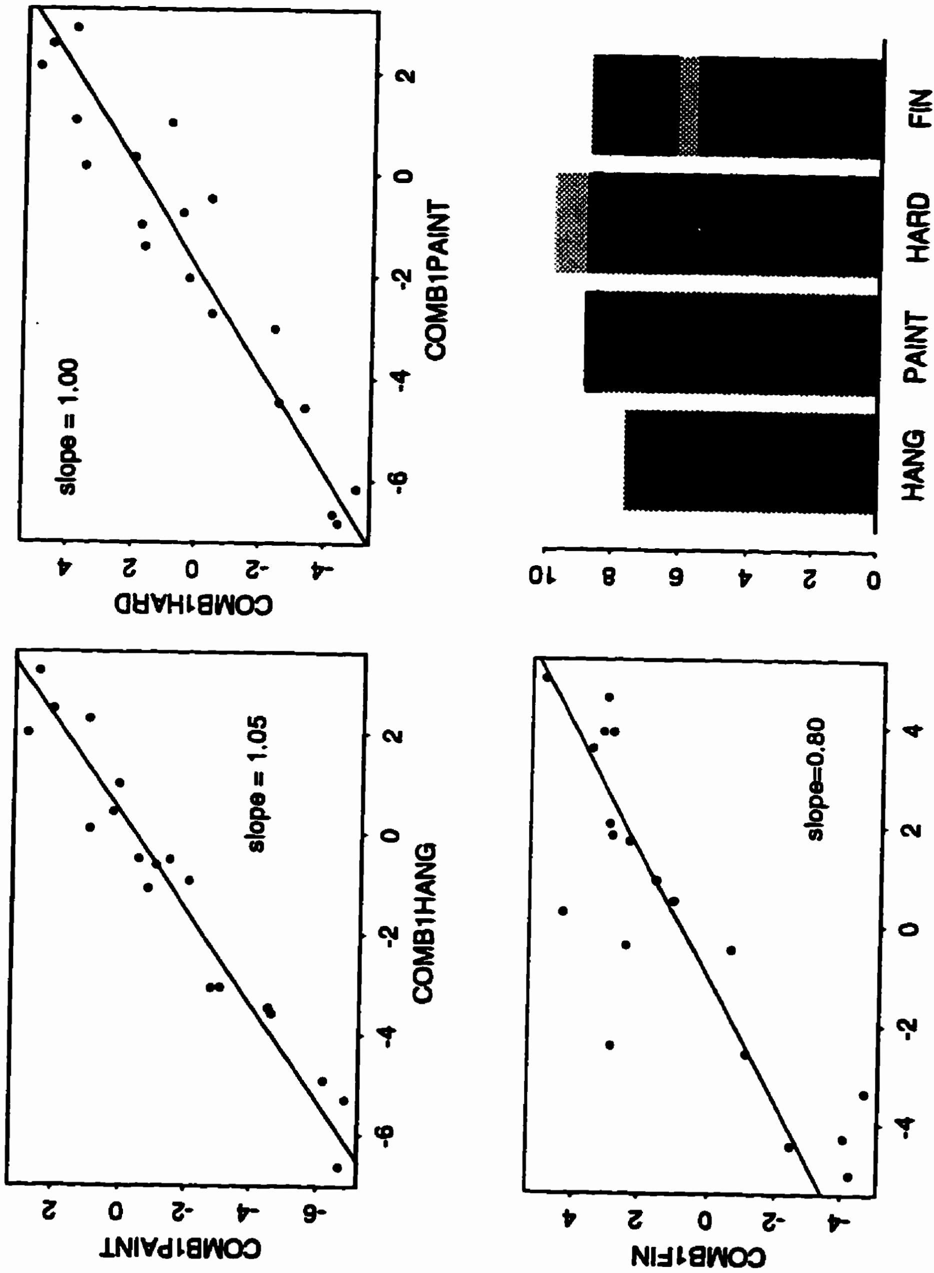

Figure 5.3: Variance transmission for COMB1

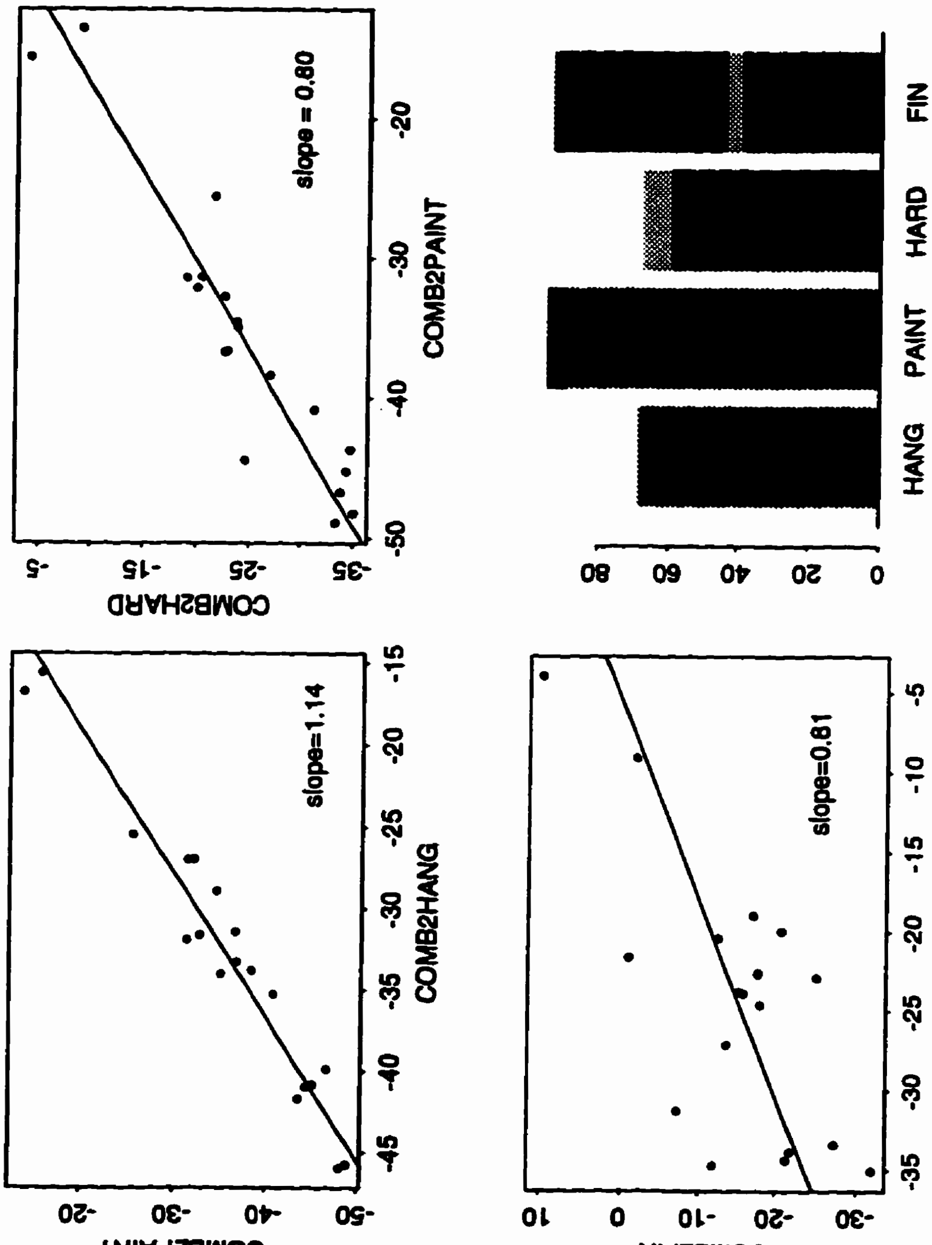

Figure 5.4: Variance transmission for COMB2

be used in these situations. Finally, methods that take cross-sectional data into account are needed. Naive approaches that are relatively simple to understand and calculate, and yet efficient in the statistical sense, would be ideal.

Chapter 6

Discussion

6.1 Conclusions

Variance transmission analysis provides a useful tool for prioritization of variation reduction efforts in multi-stage processes. A first order autoregressive model was introduced by Lawless, MacKay and Robinson (1996), who demonstrated how to partition the variance at the last stage of the process into components attributable to the upstream stages. They discuss the need for data in which items have been tracked through the process, and measurements have been made after each stage.

It was shown that when the data are observed with measurement error, the analysis using the AR(1) model gives biased results. A naive method of estimation that explicitly takes into account the measurement error was introduced. This method was shown to work well when compared to maximum likelihood estimation. Methods of finding confidence intervals for the variance components of interest were also investigated.

Frequently, cross-sectional data are available on the process in addition to the longitudinal data. This type of data is usually less expensive to get than longitudinal data, and may be collected automatically. Methods of estimating the variance components of interest in this situation are investigated. A discussion is given about designing studies when these two modes of data collection are available.

A more general multivariate normal model is also used to model data from multi-stage processes. It is found that in this case, variance transmission analysis is less straightforward then when the more restrictive AR(1) model is imposed. Here, a certain type of intervention in the process is modeled, and the resulting effect on the variance at the last stage is of interest. This method assumes that the certain conditional distributions are unaffected by the intervention.

Finally, a discussion is given about methods of handling multivariate data in multi-stage processes. Some approaches are reviewed and some suggestions are made for other approaches that might be investigated.

6.2 Further Research

Many issues remain to be investigated in this variance transmission problem.

One such issue is the question of non-normal data. It could happen that the data collected from multi-stage processes are binary, categorical, discrete or have a continuous distribution that is not normal. The piston data illustrate a simple example of how this might happen. At the final stage, a measurement could be recorded on the piston that was not the value of the diameters, but rather a measurement of 0 if the piston met specifications or 1 if it did not. In this case, we would have binary measurements at the final stage and continuous measurements upstream. Methods of dealing with such situations need to be investigated. This type of data may also be available in large quantities as cross-sectional data.

Another issue that merits further consideration is loss functions. When a part does not meet specifications, then the way in which it is deviant may be relevant. For example, in the piston process, it might be that if the diameters of interest are too large, then the piston can be reworked, but if they are too small, the piston must be scrapped. Likely the cost of rework will be less than the cost of scrap. This induces a natural loss function on the process and then the issue of interest is not variance transmission, but the way in which upstream measurements affect the expected loss at the final stage.

For example, if Y_k is the product at the final stage, and Y_i is an upstream measurement, then we are interested in minimizing E(L(Y_k)). Note that

$$E(L(Y_k)) = E_{Y_i}[E_{Y_k}(L(Y_k|Y_i))]$$

Suppose that we let the loss function be

$$L(Y_k) = (Y_k - m)^2$$

where m is the target value at the final stage. Then

$$E(L(Y_k)) = E_{Y_i}[Var(Y_k|Y_i) + \{E(Y_k|Y_i) - m\}^2]$$

Clearly, this idea can be extended to include more upstream measurements as well as multivariate measurements.

Loss functions also arise as a method of reducing the dimension of multivariate data. This is discussed by Pignatiello (1993). Methods of handling such situations are required.

Covariates in this type of analysis need to be investigated further. In the piston example, at operations 270 and 290 where there were two machines operating in parallel, the machines become covariates in the process. Differences due to targeting or in the variances at these machines may be affecting the variance at the final stage. Lawless, MacKay and Robinson (1996) discuss covariates briefly, but a more systematic methodology is required.

Appendix A

Approximate Variance Formulae for Naive Estimates with Measurement Error

The purpose of this appendix is to give the approximate variance estimates of various variance components and proportions. These approximate variance estimates are computed by finding the expected values and variances of the random variables of which they are functions. These are then used in a first order Taylor series expansion of the function.

The variance estimate of the first variance component in a two stage process is

$$Var(\tilde{\sigma}_2^2(1 - \tilde{\rho}_{12}^2)) \approx F * \Sigma * F^T$$

where

$$F = [\frac{1}{n}, \frac{-2(n-1)\rho_{12}\sigma_1\sigma_2}{n\{(n-1)\sigma_1^2 - \sigma_{e_1}^2\}}, \frac{(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{n\{(n-1)\sigma_1^2 - \sigma_{e_1}^2\}^2}]$$

and

$$\Sigma = \left[\begin{array}{c|c|c} 2(n-1)(\sigma_2^2+\sigma_{e_2}^2)^2 & 2(n-1)\rho_{12}\sigma_1\sigma_2 (\sigma_2^2+\sigma_{e_2}^2) & 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2 \\ \hline 2(n-1)\rho_{12}\sigma_1\sigma_2 (\sigma_2^2+\sigma_{e_2}^2) & (n-1)\{(\sigma_1^2+\sigma_{e_1}^2)(\sigma_2^2+\sigma_{e_2}^2)+\rho_{12}^2\sigma_1^2\sigma_2^2\} & 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)\rho_{12}\sigma_1\sigma_2 \\ \hline 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2 & 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)\rho_{12}\sigma_1\sigma_2 & 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)^2 \end{array}\right]$$

This estimate for the second component in a two stage process is

$$Var(\tilde{\sigma}_2^2\tilde{\rho}_{12}^2) \approx G * \Gamma * G^T$$

where

$$G = [\frac{2(n-1)\rho_{12}\sigma_1\sigma_2}{n\{(n-1)\sigma_1^2 - \sigma_{e_1}^2\}}, \frac{-(n-1)^2\rho_{12}^2\sigma_1^2\sigma_2^2}{n\{(n-1)\sigma_1^2 - \sigma_{e_1}^2\}^2}]$$

Also,

$$\Gamma = \left[\begin{array}{c|c} (n-1)\{(\sigma_1^2+\sigma_{e_1}^2)(\sigma_2^2+\sigma_{e_2}^2)+\rho_{12}^2\sigma_1^2\sigma_2^2\} & 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)\rho_{12}\sigma_1\sigma_2 \\ \hline 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)\rho_{12}\sigma_1\sigma_2 & 2(n-1)(\sigma_1^2+\sigma_{e_1}^2)^2 \end{array}\right]$$

The variance of the square root of the first proportion is

$$Var(\sqrt{1-\tilde{\rho}^2_{23}}) \approx \frac{1}{4f}\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{12} & v_{22} & v_{23} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

where

$$\begin{aligned}
f &= 1-\frac{(n-1)^2\rho^2_{23}\sigma^2_2\sigma^2_3}{\{(n-1)\sigma^2_2-\sigma^2_{\epsilon_2}\}\{(n-1)\sigma^2_3-\sigma^2_{\epsilon_3}\}} \\
u_1 &= \frac{(n-1)^2\rho^2_{23}\sigma^2_2\sigma^2_3}{\{(n-1)\sigma^2_2-\sigma^2_{\epsilon_2}\}\{(n-1)\sigma^2_3-\sigma^2_{\epsilon_3}\}^2} \\
u_2 &= \frac{-2(n-1)\rho_{23}\sigma_2\sigma_3}{\{(n-1)\sigma^2_2-\sigma^2_{\epsilon_2}\}\{(n-1)\sigma^2_3-\sigma^2_{\epsilon_3}\}} \\
u_3 &= \frac{(n-1)^2\rho^2_{23}\sigma^2_2\sigma^2_3}{\{(n-1)\sigma^2_2-\sigma^2_{\epsilon_2}\}^2\{(n-1)\sigma^2_3-\sigma^2_{\epsilon_3}\}} \\
v_{11} &= 2(n-1)(\sigma^2_3+\sigma^2_{\epsilon_3})^2 \\
v_{12} &= 2(n-1)\rho_{23}\sigma_2\sigma_3(\sigma^2_3+\sigma^2_{\epsilon_3}) \\
v_{13} &= 2(n-1)\rho^2_{23}\sigma^2_2\sigma^2_3 \\
v_{22} &= (n-1)\{(\sigma^2_2+\sigma^2_{\epsilon_2})(\sigma^2_3+\sigma^2_{\epsilon_3})+\rho^2_{23}\sigma^2_2\sigma^2_3\} \\
v_{23} &= 2(n-1)(\sigma^2_2+\sigma^2_{\epsilon_2})\rho_{23}\sigma_2\sigma_3 \\
v_{33} &= 2(n-1)(\sigma^2_2+\sigma^2_{\epsilon_2})^2
\end{aligned}$$

Also,

$$Var(\sqrt{\tilde{\rho}^2_{23}(1-\tilde{\rho}^2_{12}}) \approx \frac{1}{4f}W * X * W^T$$

where f is a scalar, W is a vector of elements w_i and X is a symmetric matrix of elements x_{ij}, and these are given as follows:

$$\begin{aligned}
f &= \frac{(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^3}{\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} - \\
&\quad \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} \\
w_1 &= \frac{-(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}^2} + \\
&\quad \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}^2} \\
w_2 &= \frac{2(n-1)\rho_{23}\sigma_2\sigma_3}{\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} - \\
&\quad \frac{2(n-1)^3\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2^3\sigma_3}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} \\
w_3 &= \frac{-(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} + \\
&\quad \frac{2(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^3\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} \\
w_4 &= \frac{-2(n-1)^3\rho_{12}\rho_{23}^2\sigma_1\sigma_2^3\sigma_3^2}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} \\
w_5 &= \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}^2\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2-\sigma_{\epsilon_3}^2\}} \\
x_{11} &= 2(n-1)(\sigma_3^2+\sigma_{\epsilon_3}^2)^2 \\
x_{12} &= 2(n-1)\rho_{23}\sigma_2\sigma_3(\sigma_3^2+\sigma_{\epsilon_3}^2) \\
x_{13} &= 2(n-1)\rho_{23}^2\sigma_2^2\sigma_3^2 \\
x_{14} &= 2(n-1)\rho_{12}\rho_{23}^2\sigma_1\sigma_2\sigma_3^2 \\
x_{15} &= 2(n-1)\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_3^2 \\
x_{22} &= (n-1)\{(\sigma_2^2+\sigma_{\epsilon_2}^2)(\sigma_3^2+\sigma_{\epsilon_3}^2)+\rho_{23}^2\sigma_2^2\sigma_3^2\} \\
x_{23} &= 2(n-1)\rho_{23}\sigma_2\sigma_3(\sigma_2^2+\sigma_{\epsilon_2}^2)
\end{aligned}$$

$$
\begin{aligned}
x_{24} &= (n-1)\{\rho_{12}\rho_{23}\sigma_1\sigma_2^2\sigma_3 + \rho_{12}\rho_{23}\sigma_1\sigma_3(\sigma_2^2 + \sigma_{\epsilon_2}^2)\} \\
x_{25} &= 2(n-1)\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2\sigma_3 \\
x_{33} &= 2(n-1)(\sigma_2^2 + \sigma_{\epsilon_2}^2)^2 \\
x_{34} &= 2(n-1)\rho_{12}\sigma_1\sigma_2(\sigma_2^2 + \sigma_{\epsilon_2}^2) \\
x_{35} &= 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2 \\
x_{44} &= (n-1)\{(\sigma_1^2 + \sigma_{\epsilon_1}^2)(\sigma_2^2 + \sigma_{\epsilon_2}^2) + \rho_{12}^2\sigma_1^2\sigma_2^2\} \\
x_{45} &= 2(n-1)\rho_{12}\sigma_1\sigma_2(\sigma_1^2 + \sigma_{\epsilon_1}^2) \\
x_{55} &= 2(n-1)(\sigma_1^2 + \sigma_{\epsilon_1}^2)^2
\end{aligned}
$$

Finally,

$$Var(\sqrt{\tilde{\rho}_{23}^2\tilde{\rho}_{12}^2}) = \frac{1}{4f}Y * X * Y^T \tag{A.1}$$

where X is the same matrix that appeared in the previous equation, f is a scalar and Y is a vector of elements as follows:

$$
\begin{aligned}
f &= \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}} \\
y_1 &= \frac{-(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}^2} \\
y_2 &= \frac{2(n-1)^3\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2^3\sigma_3}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}} \\
y_3 &= \frac{-2(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^3\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}} \\
y_4 &= \frac{2(n-1)^3\rho_{12}\rho_{23}^2\sigma_1\sigma_2^3\sigma_3^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}} \\
y_5 &= \frac{-(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{\{(n-1)\sigma_1^2 - \sigma_{\epsilon_1}^2\}^2\{(n-1)\sigma_2^2 - \sigma_{\epsilon_2}^2\}^2\{(n-1)\sigma_3^2 - \sigma_{\epsilon_3}^2\}}
\end{aligned}
$$

The variance formulae for the variance components in a three stage process will now be given. For the first component, this formula is

$$Var(\sqrt{\tilde{\sigma}_3^2(1-\tilde{\rho}_{23}^2)} \approx \frac{1}{4f} H * \Omega * H^T$$

where f is a scalar, H is a vector of elements and Ω is a symmetric matrix as:

$$\begin{aligned}
f &= \frac{(n-1)\sigma_3^2 - \sigma_{e_3}^2}{n} - \frac{(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{n\{(n-1)\sigma_2^2 - \sigma_{e_2}^2\}} \\
h_1 &= \frac{1}{n} \\
h_2 &= \frac{-2(n-1)\rho_{23}\sigma_2\sigma_3}{n\{(n-1)\sigma_2^2 - \sigma_{e_2}^2\}} \\
h_3 &= \frac{(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{n\{(n-1)\sigma_2^2 - \sigma_{e_2}^2\}^2} \\
\Omega_{11} &= 2(n-1)(\sigma_3^2 + \sigma_{e_3}^2)^2 \\
\Omega_{12} &= 2(n-1)\rho_{23}\sigma_2\sigma_3(\sigma_3^2 + \sigma_{e_3}^2) \\
\Omega_{13} &= 2(n-1)\rho_{23}^2\sigma_2^2\sigma_3^2 \\
\Omega_{22} &= (n-1)\{(\sigma_2^2 + \sigma_{e_2}^2)(\sigma_3^2 + \sigma_{e_3}^2) + \rho_{23}^2\sigma_2^2\sigma_3^2\} \\
\Omega_{23} &= 2(n-1)(\sigma_2^2 + \sigma_{e_2}^2)\rho_{23}\sigma_2\sigma_3 \\
\Omega_{33} &= 2(n-1)(\sigma_2^2 + \sigma_{e_2}^2)^2
\end{aligned}$$

The second variance component has an approximate variance given by

$$Var(\sqrt{\tilde{\sigma}_3^2\tilde{\rho}_{23}^2(1-\tilde{\rho}_{12}^2)} \approx \frac{1}{4f} K * \Phi * K^T$$

where

$$
\begin{aligned}
f &= \frac{(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{n\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}} - \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
k_1 &= \frac{2(n-1)\rho_{23}\sigma_2\sigma_3}{n\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}} - \frac{2(n-1)^3\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2^3\sigma_3}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
k_2 &= \frac{-(n-1)^2\rho_{23}^2\sigma_2^2\sigma_3^2}{n\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2} + \frac{2(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^3}\\
k_3 &= \frac{-2(n-1)^3\rho_{12}\rho_{23}^2\sigma_1\sigma_2^3\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
k_4 &= \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}^2\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
\Phi_{11} &= (n-1)\{(\sigma_2^2+\sigma_{\epsilon_2}^2)(\sigma_3^2+\sigma_{\epsilon_3}^2)+\rho_{23}^2\sigma_2^2\sigma_3^2\}\\
\Phi_{12} &= 2(n-1)\rho_{23}\sigma_2\sigma_3(\sigma_2^2+\sigma_{\epsilon_2}^2)\\
\Phi_{13} &= (n-1)\{\rho_{12}\rho_{23}\sigma_1\sigma_2^2\sigma_3+\rho_{12}\rho_{23}\sigma_1\sigma_3(\sigma_2^2+\sigma_{\epsilon_2}^2)\}\\
\Phi_{14} &= 2(n-1)\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2\sigma_3\\
\Phi_{22} &= 2(n-1)(\sigma_2^2+\sigma_{\epsilon_2}^2)^2\\
\Phi_{23} &= 2(n-1)\rho_{12}\sigma_1\sigma_2(\sigma_2^2+\sigma_{\epsilon_2}^2)\\
\Phi_{24} &= 2(n-1)\rho_{12}^2\sigma_1^2\sigma_2^2\\
\Phi_{33} &= (n-1)\{(\sigma_1^2+\sigma_{\epsilon_1}^2)(\sigma_2^2+\sigma_{\epsilon_2}^2)+\rho_{12}^2\sigma_1^2\sigma_2^2\}\\
\Phi_{34} &= 2(n-1)\rho_{12}\sigma_1\sigma_2(\sigma_1^2+\sigma_{\epsilon_1}^2)\\
\Phi_{44} &= 2(n-1)(\sigma_1^2+\sigma_{\epsilon_1}^2)^2
\end{aligned}
$$

Finally, the approximate variance of the third variance component is

$$
Var(\sqrt{\tilde{\sigma}_3^2\tilde{\rho}_{23}^2\tilde{\rho}_{12}^2} \approx \frac{1}{4f}L * \Phi * L^T
$$

where Φ is the matrix given in the previous expression, and f and L are as follows:

$$\begin{aligned}
f &= \frac{(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
l_1 &= \frac{2(n-1)^3\rho_{12}^2\rho_{23}\sigma_1^2\sigma_2^3\sigma_3}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
l_2 &= \frac{-2(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^3}\\
l_3 &= \frac{2(n-1)^3\rho_{12}\rho_{23}^2\sigma_1\sigma_2^3\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}\\
l_4 &= \frac{-(n-1)^4\rho_{12}^2\rho_{23}^2\sigma_1^2\sigma_2^4\sigma_3^2}{n\{(n-1)\sigma_1^2-\sigma_{\epsilon_1}^2\}^2\{(n-1)\sigma_2^2-\sigma_{\epsilon_2}^2\}^2}
\end{aligned}$$

Appendix B

Approximate Variance Formulae for Semi-Naive Estimates: Three Stages

The purpose of this appendix is to give results for the approximate variance formulae for the semi-naive estimates of the cross-sectional data. Because the actual formulae are lengthy, the Maple programs (Char et al, 1985) that were used to calculate them will be given instead. The variance-covariance matrix of the random variables in these expressions were found. The gradients for all the vectors were then calculated at the mean values of these random variables. The appropriate approximate variance formula was then given by the product of the transpose of the gradient, the variance-covariance matrix and the gradient. The formulae given here are for the variance-covariance matrix, as well as for the gradients of the square roots of the three components in a three stage process. They are analogous to the equations given in (3.10) for the naive estimates.

The variance-covariance matrix appeared as follows:

```
with(linalg);
bigvar := array(symmetric,1..14,1..14);
bigvar[1,1] := sigma1^2/n;
bigvar[1,2] := rho12*sigma1*sigma2/n;
bigvar[1,3] := rho12*rho23*sigma1*sigma3/n;
bigvar[1,4] := sigma1^2/((k+1)*n);
bigvar[1,5] := rho12*sigma1*sigma2/((k+1)*n);
bigvar[1,6] := rho12*rho23*sigma1*sigma3/((k+1)*n);
bigvar[1,7] := 0;
bigvar[1,8] := 0;
bigvar[1,9] := 0;
bigvar[1,10] := 0;
bigvar[1,11]:= 0;
bigvar[1,12]:= -2*k*mu1*sigma1^2/(k+1);
bigvar[1,13] := -2*k*rho12*sigma1*sigma2*mu2/(k+1);
bigvar[1,14] := -2*k*rho12*rho23*sigma1*sigma3*mu3/(k+1);
bigvar[2,2] := sigma2^2/n;
bigvar[2,3] := rho23*sigma2*sigma3/n;
bigvar[2,4] := rho12*sigma1*sigma2/((k+1)*n);
bigvar[2,5] := sigma2^2/((k+1)*n);
bigvar[2,6] := rho23*sigma2*sigma3/((k+1)*n);
bigvar[2,7] := 0;
bigvar[2,8] := 0;
bigvar[2,9] := 0;
bigvar[2,10]:=0;
bigvar[2,11] := 0;
bigvar[2,12] := -2*k*rho12*sigma1*sigma2*mu1/(k+1);
bigvar[2,13] := -2*k*mu2*sigma2^2/(k+1);
bigvar[2,14] := -2*k*rho23*sigma2*sigma3*mu3/(k+1);
bigvar[3,3] := sigma3^2/n;
bigvar[3,4] := rho12*rho23*sigma1*sigma3/((k+1)*n);
bigvar[3,5] := rho23*sigma2*sigma3/((k+1)*n);
bigvar[3,6] := sigma3^2/((k+1)*n);
bigvar[3,7] := 0;
bigvar[3,8] := 0;
```

```
bigvar[3,9] := 0;
bigvar[3,10] := 0;
bigvar[3,11] := 0;
bigvar[3,12] := -2*k*rho12*rho23*sigma1*sigma3*mu1/(k+1);
bigvar[3,13] := -2*k*rho23*sigma2*sigma3*mu2/(k+1);
bigvar[3,14] := -2*k*sigma3^2*mu3/(k+1);
bigvar[4,4] := sigma1^2/((k+1)*n);
bigvar[4,5] := rho12*sigma1*sigma2/((k+1)^2*n);
bigvar[4,6] := rho12*rho23*sigma1*sigma3/((k+1)^2*n);
bigvar[4,7] := 0;
bigvar[4,8] := 0;
bigvar[4,9] := 0;
bigvar[4,10] := 0;
bigvar[4,11] := 0;
bigvar[4,12] := 0;
bigvar[4,13] := -2*k*rho12*sigma1*sigma2*mu2/(k+1);
bigvar[4,14] := -2*k*rho12*rho23*sigma1*sigma3*mu3/(k+1);
bigvar[5,5] := sigma2^2/((k+1)*n);
bigvar[5,6] := rho23*sigma2*sigma3/((k+1)^2*n);
bigvar[5,7] := 0;
bigvar[5,8] := 0;
bigvar[5,9] := 0;
bigvar[5,10] := 0;
bigvar[5,11] := 0;
bigvar[5,12] := -2*k*rho12*sigma1*sigma2*mu1/(k+1);
bigvar[5,13] := 0;
bigvar[5,14] := -2*k*rho23*sigma2*sigma3*mu3/(k+1);
bigvar[6,6] := sigma3^2/((k+1)*n);
bigvar[6,7] := 0;
bigvar[6,8] := 0;
bigvar[6,9] := 0;
bigvar[6,10] := 0;
bigvar[6,11] := 0;
bigvar[6,12] := -2*k*rho12*rho23*sigma1*sigma3*mu1/(k+1);
bigvar[6,13] := -2*k*rho23*sigma2*sigma3*mu2/(k+1);
bigvar[6,14] := 0;
bigvar[7,7] := 2*(n-1)*sigma1^4;
bigvar[7,8] := 2*(n-1)*rho12*sigma1^3*sigma2;
bigvar[7,9] := 2*(n-1)*rho12^2*sigma1^2*sigma2^2;
```

```
bigvar[7,10] := 2*rho12^2*rho23*sigma1^2*sigma2*sigma3*(n-1);
bigvar[7,11] := 2*rho12^2*rho23^2*sigma1^2*sigma3^2*(n-1);
bigvar[7,12] := 2*(n-1)*sigma1^4;
bigvar[7,13] := 2*(n-1)*rho12^2*sigma1^2*sigma2^2;
bigvar[7,14] := 2*rho12^2*rho23^2*sigma1^2*sigma3^2*(n-1);
bigvar[8,8] := (n-1)*(sigma1^2*sigma2^2 +
rho12^2*sigma1^2*sigma2^2);
bigvar[8,9] := 2*(n-1)*rho12*sigma1*sigma2^3;
bigvar[8,10] := 2*(n-1)*rho12*rho23*sigma1*sigma2^2*sigma3;
bigvar[8,11] := 2*(n-1)*rho12*rho23^2*sigma1*sigma2*sigma3^2;
bigvar[8,12] := 2*(n-1)*rho12*sigma1^3*sigma2;
bigvar[8,13] := 2*(n-1)*rho12*sigma1*sigma2^3;
bigvar[8,14] := 2*rho12*rho23^2*sigma1*sigma2*sigma3^2*(n-1);
bigvar[9,9] := 2*(n-1)*sigma2^4;
bigvar[9,10] := 2*(n-1)*rho23*sigma2^3*sigma3;
bigvar[9,11] := 2*(n-1)*rho23^2*sigma2^2*sigma3^2;
bigvar[9,12] := 2*(n-1)*rho12^2*sigma1^2*sigma2^2;
bigvar[9,13] := 2*(n-1)*sigma2^4;
bigvar[9,14] := 2*rho23^2*sigma2^2*sigma3^2*(n-1);
bigvar[10,10] := (n-1)*(sigma2^2*sigma3^2 +
rho23^2*sigma2^2*sigma3^2);
bigvar[10,11] := 2*(n-1)*rho23*sigma2*sigma3^3;
bigvar[10,12] := 2*(n-1)*rho12^2*rho23*sigma1^2*sigma2*sigma3;
bigvar[10,13] := 2*(n-1)*rho23*sigma2^3*sigma3;
bigvar[10,14] := 2*(n-1)*rho23*sigma2*sigma3^3;
bigvar[11,11] := 2*(n-1)*sigma3^4;
bigvar[11,12] := 2*(n-1)*rho12^2*rho23^2*sigma1^2*sigma3^2;
bigvar[11,13] := 2*(n-1)*rho23^2*sigma2^2*sigma3^2;
bigvar[11,14] := 2*(n-1)*sigma3^4;
bigvar[12,12] := 2*((k+1)*n-1)*sigma1^4;
bigvar[12,13] := 2*rho12^2*sigma1^2*sigma2^2*(n-1) +
4*n^2*k^2*mu1*mu2*rho12*sigma1*sigma2/((k+1)^2*n);
bigvar[12,14] := 2*rho12^2*rho23^2*sigma1^2*sigma3^2*(n-1) +
4*n^2*k^2*mu1*mu3*rho12*rho23*sigma1*sigma3/((k+1)^2*n);
bigvar[13,13] := 2*((k+1)*n-1)*sigma2^4;
bigvar[13,14] := 2*rho23^2*sigma2^2*sigma3^2*(n-1) +
4*n^2*k^2*mu2*mu3*rho23*sigma2*sigma3/((k+1)^2*n);
bigvar[14,14] := 2*((k+1)*n-1)*sigma3^4;
```

The gradient of the square root of the first component was given as follows:

```
gammahat := -(k+1)*n*((sx2x2s1 + n*(u2s1-u2s1s3)^2)/(2*sx2x2s1s3) +
(sx3x3s1 + n*(u3s1-u3s1s4)^2)/(2*sx3x3s1s4));
deltahat := (k+1)^2*n^2*(sx2x3s1 + n*(u2s1-u2s1s3)*(u3s1-u3s1s4))
/(sx2x2s1s3*sx3x3s1s4);
percfour := (3*n^2+6*n*gammahat+deltahat^2)/(n^2*percthree^(1/3));
rho23hat := percthree^(1/3) + (1/9)*percfour + deltahat/(3*n);
firstcomp := sx3x3s1s4/((k+1)*n)*(1-rho23hat^2);
f := sqrt(firstcomp);
gradf := [diff(f,u1s1),diff(f,u2s1),diff(f,u3s1),diff(f,u1s1s2),
diff(f,u2s1s3),diff(f,u3s1s4),diff(f,sx1x1s1),diff(f,sx1x2s1),
diff(f,sx2x2s1),diff(f,sx2x3s1),diff(f,sx3x3s1),diff(f,sx1x1s1s2),
diff(f,sx2x2s1s3),diff(f,sx3x3s1s4)];
u1s1 := mu1;
u2s1 := mu2;
u3s1 := mu3;
u1s1s2 := mu1;
u2s1s3 := mu2;
u3s1s4 := mu3;
sx1x1s1 := (n-1)*sigma1^2;
sx1x2s1 := (n-1)*rho12*sigma1*sigma2;
sx2x2s1 := (n-1)*sigma2^2;
sx2x3s1 := (n-1)*rho23*sigma2*sigma3;
sx3x3s1 := (n-1)*sigma3^2;
sx1x1s1s2 := ((k+1)*n-1)*sigma1^2;
sx2x2s1s3 := ((k+1)*n-1)*sigma2^2;
sx3x3s1s4 := ((k+1)*n-1)*sigma3^2;
evalm(gradf);
```

The gradient of the square root of the second component was given by:

```
alpha := -(k+1)*n*((sx1x1s1 + n*(u1s1-u1s1s2)^2)/(2*sx1x1s1s2)
+ (sx2x2s1 + n*(u2s1-u2s1s3)^2)/(2*sx2x2s1s3));
beta := (k+1)^2*n^2*(sx1x2s1 + n*(u1s1-u1s1s2)*
u2s1-u2s1s3))/(sx1x1s1s2*sx2x2s1s3);
percone := (1/27)*beta*(18*n^2+9*n*alpha+beta^2)/n^3 +
(1/9)*(-3*n^4-18*n^3*alpha+33*n^2*beta^2-36*n^2*alpha^2+
24*n*alpha*beta^2+3*beta^4-24*n*alpha^3-3*alpha^2*beta^2)^(0.5)/n^2;
```

```
perctwo := (3*n^2+6*n*alpha+beta^2)/(n^2*percone^(1/3));
rho12hat := percone^(1/3) + (1/9)*perctwo + beta/(3*n);
gammahat := -(k+1)*n*((sx2x2s1 + n*(u2s1-u2s1s3)^2)/(2*sx2x2s1s3) +
sx3x3s1 + n*(u3s1-u3s1s4)^2)/(2*sx3x3s1s4));
deltahat := (k+1)^2*n^2*(sx2x3s1 + n*(u2s1-u2s1s3)*(u3s1-u3s1s4))
/(sx2x2s1s3*sx3x3s1s4);
percthree := (1/27)*deltahat*(18*n^2+9*n*gammahat +
deltahat^2)/n^3 + (1/9)*(-3*n^4 - 18*n^3*gammahat +
33*n^2*deltahat^2 - 36*n^2*gammahat^2 + 24*n*gammahat*deltahat^2
+ 3*deltahat^4 - 24*n*gammahat^3 - 3*gammahat^2*deltahat^2)
^(0.5)/n^2;
percfour := (3*n^2+6*n*gammahat+deltahat^2)/(n^2*percthree^(1/3));
rho23hat := percthree^(1/3) + (1/9)*percfour + deltahat/(3*n);
seccomp := sx3x3s1s4/((k+1)*n)*rho23hat^2*(1-rho12hat^2);
f := sqrt(seccomp);
gradf := [diff(f,u1s1),diff(f,u2s1),diff(f,u3s1),diff(f,u1s1s2),
diff(f,u2s1s3),diff(f,u3s1s4),diff(f,sx1x1s1),diff(f,sx1x2s1),
diff(f,sx2x2s1),diff(f,sx2x3s1),diff(f,sx3x3s1),diff(f,sx1x1s1s2),
diff(f,sx2x2s1s3),diff(f,sx3x3s1s4)];
u1s1 := mu1;
u2s1 := mu2;
u3s1 := mu3;
u1s1s2 := mu1;
u2s1s3 := mu2;
u3s1s4 := mu3;
sx1x1s1 := (n-1)*sigma1^2;
sx1x2s1 := (n-1)*rho12*sigma1*sigma2;
sx2x2s1 := (n-1)*sigma2^2;
sx2x3s1 := (n-1)*rho23*sigma2*sigma3;
sx3x3s1 := (n-1)*sigma3^2;
sx1x1s1s2 := ((k+1)*n-1)*sigma1^2;
sx2x2s1s3 := ((k+1)*n-1)*sigma2^2;
sx3x3s1s4 := ((k+1)*n-1)*sigma3^2;
evalm(gradf);
```

The gradient of the square root of the third component is given by:

```
alpha := -(k+1)*n*((sx1x1s1 + n*(u1s1-u1s1s2)^2)/(2*sx1x1s1s2) +
(sx2x2s1 + n*(u2s1-u2s1s3)^2)/(2*sx2x2s1s3));
```

```
beta := (k+1)^2*n^2*(sx1x2s1 + n*(u1s1-u1s1s2)*(u2s1-u2s1s3))
/(sx1x1s1s2*sx2x2s1s3);
percone := (1/27)*beta*(18*n^2+9*n*alpha+beta^2)/n^3 +
1/9)*(-3*n^4-18*n^3*alpha+33*n^2*beta^2-36*n^2*alpha^2+
4*n*alpha*beta^2+3*beta^4-24*n*alpha^3-3*alpha^2*beta^2)
^(0.5)/n^2;
perctwo := (3*n^2+6*n*alpha+beta^2)/(n^2*percone^(1/3));
rho12hat := percone^(1/3) + (1/9)*perctwo + beta/(3*n);
gammahat := -(k+1)*n*((sx2x2s1 + n*(u2s1-u2s1s3)^2)/
(2*sx2x2s1s3) + (sx3x3s1 + n*(u3s1-u3s1s4)^2)/(2*sx3x3s1s4));
deltahat := (k+1)^2*n^2*(sx2x3s1 + n*(u2s1-u2s1s3)*
(u3s1-u3s1s4))/(sx2x2s1s3*sx3x3s1s4);
percthree := (1/27)*deltahat*(18*n^2+9*n*gammahat +
deltahat^2)/n^3 + (1/9)*(-3*n^4 - 18*n^3*gammahat +
33*n^2*deltahat^2 - 36*n^2*gammahat^2 + 24*n*gammahat*
deltahat^2 + 3*deltahat^4 - 24*n*gammahat^3 -
3*gammahat^2*deltahat^2)^(0.5)/n^2;
percfour := (3*n^2+6*n*gammahat+deltahat^2)/(n^2*percthree^(1/3));
rho23hat := percthree^(1/3) + (1/9)*percfour + deltahat/(3*n);
thirdcomp := sx3x3s1s4/((k+1)*n)*rho23hat^2*rho12hat^2;
f := sqrt(thirdcomp);
gradf := [diff(f,u1s1),diff(f,u2s1),diff(f,u3s1),diff(f,u1s1s2),
diff(f,u2s1s3),diff(f,u3s1s4),diff(f,sx1x1s1),diff(f,sx1x2s1),
diff(f,sx2x2s1),diff(f,sx2x3s1),diff(f,sx3x3s1),diff(f,sx1x1s1s2),
diff(f,sx2x2s1s3),diff(f,sx3x3s1s4)];
u1s1 := mu1;
u2s1 := mu2;
u3s1 := mu3;
u1s1s2 := mu1;
u2s1s3 := mu2;
u3s1s4 := mu3;
sx1x1s1 := (n-1)*sigma1^2;
sx1x2s1 := (n-1)*rho12*sigma1*sigma2;
sx2x2s1 := (n-1)*sigma2^2;
sx2x3s1 := (n-1)*rho23*sigma2*sigma3;
sx3x3s1 := (n-1)*sigma3^2;
sx1x1s1s2 := ((k+1)*n-1)*sigma1^2;
sx2x2s1s3 := ((k+1)*n-1)*sigma2^2;
sx3x3s1s4 := ((k+1)*n-1)*sigma3^2;
```

```
evalm(gradf);
```

Appendix C

Simulation Results for Cross-sectional and Longitudinal Data

The purpose of this appendix is to give the results of the simulations done in chapter three. Please see that chapter for a complete description of the simulation studies.

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Est.									
L	Real	.80	.50	.20	.80	.50	.20	.80	.50	.20
	MLES	.78	.46	.17	.73	.44	.18	.75	.47	.18
	(S123)	(.25)	(.16)	(.06)	(.24)	(.14)	(.06)	(.22)	(.15)	(.06)
	NEFS	.78	.47	.18	.76	.45	.19	.78	.47	.20
		(.21)	(.15)	(.06)	(.17)	(.15)	(.07)	(.17)	(.15)	(.08)
	SNES	.78	.49	.20	.76	.47	.21	.78	.51	.22
		(.20)	(.15)	(.06)	(.18)	(.14)	(.08)	(.16)	(.14)	(.06)
	MLES	.77	.46	.18	.76	.45	.18	.77	.49	.19
	(S123-S3)	(.20)	(.16)	(.05)	(.18)	(.14)	(.06)	(.17)	(.14)	(.07)
M	Real	.80	.50	.20	.80	.50	.20	.80	.50	.20
	MLES	.71	.45	.17	.70	.45	.18	.75	.49	.17
	(S123)	(.24)	(.16)	(.06)	(.22)	(.15)	(.06)	(.25)	(.16)	(.06)
	NEFS	.72	.49	.19	.75	.46	.21	.78	.52	.21
		(.20)	(.16)	(.08)	(.18)	(.15)	(.08)	(.17)	(.16)	(.088)
	SNES	.73	.51	.21	.76	.48	.23	.79	.54	.21
		(.19)	(.16)	(.08)	(.16)	(.13)	(.07)	(.16)	(.15)	(.07)
	MLES	.72	.48	.18	.75	.46	.19	.78	.52	.18
	(S123-S3)	(.19)	(.15)	(.06)	(.16)	(.13)	(.06)	(.17)	(.15)	(.06)
H	Real	.8	.5	.2	.8	.5	.2	.8	.5	.2
	MLES	.72	.47	.18	.74	.46	.18	.67	.45	.18
	(S123)	(.23)	(.15)	(.07)	(.24)	(.18)	(.06)	(.25)	(.17)	(.06)
	NEFS	.75	.48	.20	.76	.49	.19	.72	.47	.19
		(.20)	(.16)	(.10)	(.18)	(.16)	(.07)	(.18)	(.16)	(.09)
	SNES	.76	.50	.21	.77	.51	.21	.74	.48	.21
		(.19)	(.15)	(.08)	(.17)	(.16)	(.06)	(.17)	(.14)	(.06)
	MLES	.75	.47	.18	.76	.48	.18	.73	.46	.18
	(S123-S3)	(.19)	(.14)	(.07)	(.18)	(.15)	(.06)	(.18)	(.14)	(.06)

Table C.1: Average of 100 values of first component of each run where n = 20. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Estimate									
L	Real	.80	.50	.20	.80	.50	.20	.80	.50	.20
	MLES	.76	.47	.19	.79	.48	.19	.80	.46	.19
	(S123)	(.18)	(.09)	(.04)	(.16)	(.10)	(.04)	(.19)	(.10)	(.04)
	NEFS	.77	.49	.20	.78	.49	.20	.80	.49	.20
		(.15)	(.09)	(.05)	(.12)	(.10)	(.06)	(.12)	(.10)	(.05)
	SNES	.77	.49	.21	.79	.49	.21	.81	.49	.21
		(.15)	(.09)	(.05)	(.12)	(.09)	(.05)	(.11)	(.08)	(.04)
	MLES	.77	.48	.20	.79	.48	.19	.80	.48	.20
	(S123-S3)	(.15)	(.09)	(.04)	(.12)	(.09)	(.04)	(.12)	(.09)	(.04)
M	Real	.80	.50	.20	.80	.50	.20	.80	.50	.20
	MLES	.76	.47	.19	.79	.47	.19	.75	.49	.20
	(S123)	(.16)	(.10)	(.04)	(.14)	(.10)	(.04)	(.15)	(.09)	(.04)
	NEFS	.78	.48	.20	.79	.49	.20	.78	.49	.21
		(.15)	(.10)	(.05)	(.12)	(.10)	(.05)	(.12)	(.09)	(.05)
	SNES	.78	.49	.21	.80	.50	.21	.78	.50	.21
		(.15)	(.09)	(.05)	(.11)	(.09)	(.04)	(.11)	(.08)	(.04)
	MLES	.78	.48	.19	.79	.48	.19	.77	.50	.20
	(S123-S3)	(.15)	(.09)	(.04)	(.11)	(.09)	(.04)	(.11)	(.08)	(.04)
H	Real	.8	.5	.2	.8	.5	.2	.8	.5	.2
	MLES	.76	.47	.19	.77	.48	.19	.76	.47	.19
	(S123)	(.16)	(.08)	(.04)	(.14)	(.09)	(.04)	(.15)	(.09)	(.04)
	NEFS	.77	.49	.19	.80	.49	.20	.78	.49	.20
		(.14)	(.09)	(.05)	(.12)	(.10)	(.05)	(.12)	(.10)	(.05)
	SNES	.78	.49	.20	.80	.50	.21	.78	.49	.20
		(.13)	(.08)	(.04)	(.12)	(.09)	(.05)	(.10)	(.09)	(.04)
	MLES	.77	.48	.19	.80	.49	.20	.78	.49	.19
	(S123-S3)	(.13)	(.08)	(.04)	(.12)	(.09)	(.04)	(.10)	(.09)	(.04)

Table C.2: Average of 100 values of first component of each run where $n = 50$. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Estimate									
L	Real	.16	.40	.64	.16	.40	.64	.16	.40	.64
	MLES	.18	.38	.56	.16	.41	.57	.17	.43	.61
	(S123)	(.16)	(.22)	(.19)	(.12)	(.23)	(.23)	(.12)	(.23)	(.26)
	NEFS	.17	.39	.57	.17	.40	.59	.17	.39	.61
		(.14)	(.20)	(.17)	(.12)	(.19)	(.16)	(.11)	(.15)	(.17)
	SNES	.18	.38	.57	.17	.39	.59	.17	.38	.60
		(.13)	(.20)	(.17)	(.13)	(.18)	(.16)	(.11)	(.13)	(.16)
	MLES	.18	.39	.59	.18	.40	.60	.18	.39	.62
	(S123-S3)	(.14)	(.21)	(.17)	(.13)	(.18)	(.17)	(.11)	(.14)	(.16)
M	Real	.10	.25	.40	.10	.25	.40	.10	.25	.40
	MLES	.11	.22	.37	.09	.25	.37	.11	.23	.33
	(S123)	(.11)	(.13)	(.15)	(.08)	(.15)	(.15)	(.11)	(.12)	(.14)
	NEFS	.11	.23	.38	.10	.24	.40	.11	.23	.38
		(.09)	(.11)	(.13)	(.08)	(.12)	(.14)	(.09)	(.10)	(.15)
	SNES	.11	.23	.38	.10	.24	.40	.11	.24	.39
		(.10)	(.11)	(.12)	(.08)	(.11)	(.13)	(.08)	(.10)	(.14)
	MLES	.11	.23	.39	.10	.24	.40	.11	.24	.39
	(S123-S3)	(.10)	(.11)	(.13)	(.08)	(.12)	(.13)	(.08)	(.10)	(.14)
H	Real	.04	.10	.16	.04	.10	.16	.04	.10	.16
	MLES	.05	.10	.14	.05	.09	.14	.05	.10	.15
	(S123)	(.04)	(.05)	(.05)	(.04)	(.06)	(.05)	(.05)	(.06)	(.06)
	NEFS	.04	.10	.15	.05	.09	.15	.05	.10	.16
		(.04)	(.05)	(.06)	(.04)	(.05)	(.05)	(.04)	(.05)	(.07)
	SNES	.05	.11	.17	.05	.10	.17	.06	.11	.17
		(.05)	(.06)	(.06)	(.04)	(.05)	(.06)	(.04)	(.05)	(.06)
	MLES	.05	.10	.14	.04	.09	.14	.05	.10	.16
	(S123-S3)	(.04)	(.06)	(.06)	(.04)	(.05)	(.05)	(.04)	(.05)	(.06)

Table C.3: Average of 100 values of second component of each run where n = 20. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Estimate									
L	Real	.16	.40	.64	.16	.40	.64	.16	.40	.64
	MLES	.17	.37	.61	.18	.41	.59	.16	.39	.63
	(S123)	(.10)	(.13)	(.16)	(.10)	(.15)	(.15)	(.10)	(.15)	(.16)
	NEFS	.17	.38	.61	.17	.40	.61	.16	.40	.63
		(.09)	(.11)	(.13)	(.08)	(.11)	(.11)	(.09)	(.11)	(.10)
	SNES	.16	.38	.60	.17	.40	.61	.15	.41	.63
		(.08)	(.11)	(.13)	(.08)	(.10)	(.11)	(.08)	(.10)	(.10)
	MLES	.17	.39	.62	.17	.41	.62	.15	.41	.64
	(S123-S3)	(.09)	(.11)	(.13)	(.08)	(.10)	(.11)	(.08)	(.10)	(.09)
M	Real	.10	.25	.40	.10	.25	.40	.10	.25	.40
	MLES	.10	.24	.38	.10	.23	.38	.10	.26	.38
	(S123)	(.07)	(.08)	(.09)	(.06)	(.08)	(.07)	(.06)	(.09)	(.08)
	NEFS	.10	.25	.39	.10	.24	.39	.11	.26	.40
		(.06)	(.07)	(.08)	(.05)	(.07)	(.08)	(.06)	(.07)	(.08)
	SNES	.10	.25	.39	.10	.24	.39	.11	.26	.39
		(.06)	(.07)	(.08)	(.04)	(.07)	(.07)	(.06)	(.06)	(.08)
	MLES	.10	.25	.39	.10	.24	.39	.11	.26	.39
	(S123-S3)	(.06)	(.07)	(.08)	(.05)	(.07)	(.07)	(.06)	(.06)	(.08)
H	Real	.04	.10	.16	.04	.10	.16	.04	.10	.16
	MLES	.05	.10	.15	.04	.10	.15	.04	.10	.15
	(S123)	(.03)	(.03)	(.04)	(.02)	(.04)	(.04)	(.02)	(.03)	(.04)
	NEFS	.05	.10	.15	.04	.10	.16	.04	.10	.16
		(.02)	(.03)	(.04)	(.02)	(.03)	(.04)	(.02)	(.03)	(.03)
	SNES	.05	.11	.16	.04	.11	.16	.04	.10	.16
		(.02)	(.03)	(.04)	(.02)	(.03)	(.04)	(.02)	(.03)	(.03)
	MLES	.05	.10	.15	.04	.10	.16	.04	.10	.16
	(S123-S3)	(.02)	(.03)	(.03)	(.02)	(.03)	(.04)	(.02)	(.03)	(.03)

Table C.4: Average of 100 values of second component of each run where n = 50. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Estimate									
L	Real	.04	.10	.16	.04	.10	.16	.04	.10	.16
	MLES	.05	.12	.18	.05	.11	.19	.05	.14	.17
	(S123)	(.07)	(.10)	(.19)	(.06)	(.10)	(.18)	(.06)	(.12)	(.15)
	NEFS	.05	.11	.18	.05	.11	.19	.05	.13	.17
		(.07)	(.09)	(.17)	(.05)	(.09)	(.16)	(.05)	(.10)	(.13)
	SNES	.05	.11	.16	.05	.10	.18	.05	.11	.17
		(.05)	(.09)	(.14)	(.05)	(.08)	(.13)	(.04)	(.08)	(.11)
	MLES	.05	.13	.18	.05	.11	.20	.05	.12	.18
	(S123-S3)	(.06)	(.11)	(.16)	(.05)	(.09)	(.13)	(.05)	(.09)	(.12)
M	Real	.10	.25	.40	.10	.25	.40	.10	.25	.40
	MLES	.13	.26	.40	.12	.26	.41	.12	.27	.40
	(S123)	(.11)	(.18)	(.27)	(.12)	(.16)	(.25)	(.13)	(.15)	(.24)
	NEFS	.12	.26	.39	.12	.25	.42	.12	.26	.42
		(.10)	(.14)	(.22)	(.10)	(.13)	(.19)	(.10)	(.12)	(.16)
	SNES	.10	.24	.36	.11	.23	.38	.11	.23	.41
		(.08)	(.12)	(.19)	(.08)	(.10)	(.14)	(.08)	(.09)	(.13)
	MLES	.12	.27	.41	.12	.26	.43	.12	.25	.44
	(S123-S3)	(.10)	(.12)	(.18)	(.09)	(.11)	(.16)	(.09)	(.10)	(.14)
H	Real	.16	.40	.64	.16	.40	.64	.16	.40	.64
	MLES	.19	.41	.59	.20	.41	.65	.21	.43	.66
	(S123)	(.18)	(.22)	(.29)	(.16)	(.28)	(.30)	(.17)	(.25)	(.30)
	NEFS	.18	.40	.61	.19	.41	.64	.21	.42	.64
		(.15)	(.18)	(.21)	(.13)	(.21)	(.19)	(.14)	(.18)	(.15)
	SNES	.16	.38	.58	.18	.38	.60	.20	.41	.62
		(.13)	(.16)	(.18)	(.11)	(.16)	(.15)	(.13)	(.14)	(.11)
	MLES	.18	.42	.65	.19	.41	.65	.21	.43	.665
	(S123-S3)	(.14)	(.17)	(.17)	(.12)	(.17)	(.15)	(.14)	(.14)	(.11)

Table C.5: Average of 100 values of third component of each run where n = 20. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	Estimate									
L	Real Value	.04	.10	.16	.04	.10	.16	.04	.10	.16
	MLES	.04	.11	.18	.05	.11	.17	.05	.11	.18
	(S123)	(.03)	(.08)	(.11)	(.04)	(.07)	(.10)	(.04)	(.07)	(.11)
	NEFS	.04	.11	.18	.05	.10	.17	.05	.11	.17
		(.03)	(.07)	(.10)	(.04)	(.07)	(.08)	(.04)	(.06)	(.09)
	SNES	.04	.11	.17	.05	.10	.17	.04	.11	.16
		(.03)	(.06)	(.09)	(.03)	(.06)	(.07)	(.03)	(.05)	(.08)
	MLES	.04	.11	.18	.05	.10	.18	.04	.11	.17
	(S123-S3)	(.03)	(.07)	(.09)	(.03)	(.06)	(.07)	(.03)	(.06)	(.08)
M	Real	.10	.25	.40	.10	.25	.40	.10	.25	.40
	MLES	.11	.26	.42	.11	.25	.40	.10	.27	.40
	(S123)	(.08)	(.11)	(.18)	(.07)	(.11)	(.16)	(.06)	(.11)	(.17)
	NEFS	.11	.26	.42	.11	.25	.40	.10	.26	.40
		(.06)	(.10)	(.14)	(.06)	(.10)	(.12)	(.06)	(.08)	(.11)
	SNES	.10	.26	.41	.10	.24	.39	.10	.25	.40
		(.05)	(.09)	(.12)	(.05)	(.08)	(.10)	(.05)	(.07)	(.08)
	MLES	.11	.27	.43	.11	.25	.41	.11	.26	.41
	(S123-S3)	(.05)	(.09)	(.11)	(.05)	(.08)	(.10)	(.06)	(.07)	(.09)
H	Real	.16	.40	.64	.16	.40	.64	.16	.40	.64
	MLES	.19	.40	.64	.16	.39	.64	.18	.40	.63
	(S123)	(.11)	(.13)	(.18)	(.10)	(.14)	(.19)	(.12)	(.14)	(.18)
	NEFS	.18	.41	.64	.17	.39	.65	.18	.41	.65
		(.10)	(.11)	(.14)	(.09)	(.11)	(.13)	(.10)	(.10)	(.09)
	SNES	.18	.40	.62	.16	.39	.64	.17	.40	.64
		(.08)	(.09)	(.12)	(.07)	(.09)	(.11)	(.08)	(.08)	(.07)
	MLES	.19	.42	.65	.17	.40	.66	.18	.42	.65
	(S123-S3)	(.09)	(.10)	(.11)	(.08)	(.10)	(.10)	(.08)	(.08)	(.06)

Table C.6: Average of 100 values of third component of each run where n = 50. The figures in brackets represent the standard deviation for these values

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	96.84	96.52	96.52	97.36	96.64	97.52	96.80	97.32	96.92
	95	94.16	93.16	92.72	93.84	93.80	94.16	93.68	94.04	93.80
	90	88.92	87.68	88.04	88.48	88.28	88.60	88.68	88.96	88.56
M	98	97.08	96.76	96.32	97.32	97.16	96.64	97.36	96.68	96.60
	95	93.92	93.64	93.04	94.28	93.56	92.56	94.12	93.68	93.04
	90	87.88	88.12	88.20	88.92	88.40	87.28	88.84	87.84	87.84
H	98	97.56	95.64	95.44	98.08	97.12	97.00	97.28	97.12	96.88
	95	94.28	92.80	92.04	94.96	94.00	93.76	94.20	94.16	94.08
	90	89.08	87.08	86.44	89.20	88.76	88.80	89.08	89.16	88.76

Table C.7: Further simulation to check approximate variance formulas. Naive estimates, first component, based on 2500 runs, n=50. Table entries are percent of samples including the true parameter value.

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	97.72	98.00	96.64	97.20	97.88	97.48	97.20	98.04	97.24
	95	94.28	94.80	92.48	93.92	95.24	94.52	93.64	94.52	94.40
	90	89.84	89.00	87.68	88.52	90.32	88.48	88.56	89.72	89.04
M	98	97.92	97.28	96.64	97.60	97.96	96.92	97.72	97.68	96.88
	95	94.48	94.16	93.48	94.08	94.88	93.60	94.16	94.00	93.40
	90	89.16	89.44	88.48	88.68	89.12	88.56	89.32	88.84	87.72
H	98	97.56	96.80	97.08	98.00	97.04	96.52	97.72	97.72	95.96
	95	94.80	93.20	93.64	94.68	93.68	93.16	95.16	95.24	92.36
	90	89.84	88.40	88.48	89.48	88.84	87.76	89.48	90.24	88.08

Table C.8: Further simulation to check approximate variance formulas. Naive estimates, second component, based on 2500 runs, n=50.

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	95.80	97.04	97.32	96.04	97.16	97.64	96.20	97.36	97.00
	95	92.60	93.24	94.40	93.32	94.28	94.28	92.80	94.56	93.44
	90	87.68	87.52	89.72	88.76	89.20	89.52	88.36	89.12	88.80
M	98	97.04	97.32	97.16	97.24	97.32	97.72	97.16	97.24	97.60
	95	94.00	94.24	94.32	93.96	94.16	94.44	94.08	94.08	94.20
	90	88.96	88.64	89.00	88.20	88.56	88.56	89.20	89.60	89.28
H	98	97.76	97.68	97.28	97.68	98.28	97.96	97.88	97.88	97.60
	95	94.68	94.84	94.20	93.76	95.08	94.44	94.64	94.68	94.96
	90	89.72	89.32	89.44	89.04	90.16	89.08	90.08	89.16	90.20

Table C.9: Further simulation to check approximate variance formulas. Naive estimates, third component, based on 2500 runs, n=50.

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	99.68	99.96	100.00	99.48	99.96	100.00	98.60	99.48	100.0
	95	98.84	99.76	100.00	98.24	99.72	100.00	96.44	99.04	99.92
	90	96.12	99.24	100.00	95.32	98.80	100.00	92.08	97.24	99.92
M	98	99.72	100.0	100.00	99.68	99.96	100.00	98.88	99.76	99.96
	95	99.00	99.84	100.00	98.40	99.80	100.00	97.16	99.16	99.88
	90	96.00	99.60	100.00	95.20	98.88	100.00	92.96	96.96	99.72
H	98	99.72	100.0	100.00	99.60	99.88	100.00	98.88	99.92	100.0
	95	98.64	99.80	100.00	98.60	99.72	100.00	96.80	99.16	99.92
	90	96.40	99.24	100.00	95.32	98.76	100.00	92.52	97.72	99.88

Table C.10: Further simulation to check approximate variance formulas. Semi-naive estimates, first component, based on 2500 runs, n = 50

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	96.68	95.72	95.16	95.92	96.84	96.96	96.44	96.72	96.40
	95	92.56	90.44	90.68	92.08	93.12	93.20	93.36	92.72	92.56
	90	87.36	83.76	85.32	86.40	86.36	87.32	88.36	86.68	87.84
M	98	96.76	96.96	96.68	97.00	97.92	97.08	97.40	97.76	97.04
	95	94.04	94.16	93.76	93.28	95.12	93.72	94.20	94.44	92.88
	90	89.16	89.24	89.08	88.84	89.52	89.52	89.08	88.64	88.44
H	98	99.60	99.88	99.76	99.44	99.76	99.88	99.48	99.96	99.92
	95	99.08	99.64	99.48	98.96	99.76	99.72	98.52	99.72	99.52
	90	98.36	99.52	99.28	97.64	99.40	99.36	96.24	98.96	99.04

Table C.11: Further simulation to check approximate variance formulas. Semi-naive estimates, second component, based on 2500 runs, n = 50

k		1			2			5		
ρ_{23}		L	M	H	L	M	H	L	M	H
ρ_{12}	C.I.									
L	98	95.96	96.68	97.36	97.00	97.24	98.24	97.20	97.40	97.40
	95	92.40	93.52	94.36	94.44	94.00	95.72	94.00	94.36	94.20
	90	87.68	87.28	89.64	90.00	89.84	90.52	88.80	88.80	90.12
M	98	97.72	98.80	99.28	97.48	98.96	99.72	97.76	98.28	99.40
	95	94.40	96.32	97.76	94.80	95.96	97.76	94.36	95.36	97.84
	90	89.96	91.44	94.60	90.64	91.28	94.24	90.12	90.96	94.64
H	98	98.60	99.64	99.96	98.80	99.76	100.00	97.44	99.56	100.00
	95	95.96	98.52	99.80	95.48	98.84	99.92	95.12	98.24	99.84
	90	91.56	96.20	99.40	91.04	95.92	99.16	91.20	94.96	99.08

Table C.12: Further simulation to check approximate variance formulas. Semi-naive estimates, third component, based on 2500 runs, n = 50

		Run 1		Run 2		Run 3	
Naives		Coverage	Width	Coverage	Width	Coverage	Width
Approx.							
1st	98	96.9	0.6060	97.1	0.2199	96.5	0.2283
	94	93.4	0.4900	93.2	0.1778	92.4	0.1846
	90	90.0	0.4285	89.7	0.1555	88.4	0.1614
2nd	98	98.0	0.1020	96.7	0.1763	97.2	0.4555
	94	93.1	0.0825	91.4	0.1425	91.9	0.3683
	90	89.3	0.0722	87.4	0.1247	87.6	0.3221
3rd	98	97.3	0.4146	97.5	0.5442	98.1	0.3906
	94	92.6	0.3352	93.5	0.4400	93.6	0.3158
	90	88.3	0.2931	88.7	0.3848	90.5	0.2762
Boot.							
1st	98	96.8	0.6402	97.6	0.2438	97.9	0.2550
	94	91.3	0.4976	93.5	0.1854	93.3	0.1953
	90	87.3	0.4305	88.8	0.1591	88.6	0.1676
2nd	98	98.3	0.1135	97.1	0.1949	96.6	0.4830
	94	94.0	0.0854	92.6	0.1480	92.0	0.3756
	90	89.6	0.0734	88.3	0.1276	87.3	0.3258
3rd	98	97.5	0.4333	97.5	0.5828	99.0	0.3970
	94	93.6	0.3386	92.9	0.4538	95.2	0.3156
	90	88.6	0.2947	88.2	0.3926	90.5	0.2749

Table C.13: Three runs done to compare the approximate variance formulae with the bootstrap, for the naive estimates. Longitudinal sample size = 50.

		Run 1		Run 2		Run 3	
Semi-Naives		Coverage	Width	Coverage	Width	Coverage	Width
Approx.							
1st	98	99.7	0.7307	100.0	0.9799	100.0	0.5983
	94	98.4	0.5908	100.0	0.7922	100.0	0.4837
	90	96.9	0.5167	100.0	0.6929	100.0	0.4231
2nd	98	99.6	0.3034	99.9	0.7381	95.9	0.4328
	94	99.2	0.2453	99.4	0.5967	90.2	0.3499
	90	98.8	0.2145	99.0	0.5219	86.7	0.3060
3rd	98	98.2	0.4098	99.9	0.7654	98.0	0.3470
	94	95.2	0.3313	99.9	0.6188	93.4	0.2805
	90	91.9	0.2898	99.3	0.5412	89.8	0.2453
Boot.							
1st	98	96.4	0.6349	98.4	0.2237	98.2	0.2078
	94	92.4	0.4906	95.9	0.1706	94.0	0.1598
	90	87.5	0.4252	92.1	0.1471	89.6	0.1381
2nd	98	98.4	0.1237	98.2	0.1984	96.3	0.4668
	94	94.3	0.0931	94.1	0.1523	90.6	0.3644
	90	90.3	0.0795	90.8	0.1310	87.0	0.3153
3rd	98	97.2	0.3970	96.2	0.4878	98.5	0.3465
	94	93.1	0.3088	91.6	0.3791	94.0	0.2753
	90	88.0	0.2667	85.3	0.3275	90.5	0.2411

Table C.14: Three runs done to compare the approximate variance formulae with the bootstrap, for the semi-naive estimates. Longitudinal sample size = 50.

Bibliography

[1] Char, Bruce W., Keith O. Geddes, Gaston H. Gonnet, and Stephen M. Watt (1985) *Maple User's Guide. First Leaves: A Tutorial Introduction to Maple and Maple Reference Manual, 4th Edition*, WATCOM Publications Limited, Waterloo, Ontario.

[2] Crouse, William H., (1970) *Automotive Engine Design*. McGraw-Hill Book Company, New York.

[3] Draper, N.R. and H. Smith, (1981) *Applied Regression Analysis*, Wiley and Sons, New York.

[4] Diggle, Peter J., Kung-Yee Liang and Scott L. Zeger (1994) *Analysis of Longitudinal Data*. Oxford University Press.

[5] Efron, B. and R. Tibshirani, (1986) "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy", *Statistical Science*, Vol. 1, No. 1, 54-77.

[6] Fong, Daniel Y.T. and J.F. Lawless, (1996) "The Analysis of Process Variation Transmission with Multivariate Measurements", submitted to *Journal of Quality Technology.*

[7] Fuller, Wayne A., (1987) *Measurement Error Models.* John Wiley and Sons, New York.

[8] Hamada, M.S. and J.F. Lawless, Multivariate Methods for the Assessment of Process Variation Transmission, Prepared for General Motors Research Laboratories.

[9] Jobson, J.D. (1991) *Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design*, Springer-Verlag, New York.

[10] Joiner, Brian L. and Marie A. Gaudard, (1990) "Variation, Management and W. Edwards Deming", *Quality Progress*, December, 29-37.

[11] Johnson, J. (1972) *Econometric Methods*, McGraw-Hill Book Company, New York.

[12] Johnson, Richard A. and Dean W. Wichern (1988) *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey.

[13] Know, Manfred and Malcolm Farrow (1996) "Design of Experiments for Multi-Stage Processes", *Quality and Reliability Engineering International*, Vol 12, 129-132 (1996).

[14] Larsen, Richard J. and Morris L. Marx (1986) *An Introduction to Mathematical Statistics and Its Applications*, Prentice-Hall, New Jersey.

[15] Lawless, J.F., R. J. MacKay and J.A. Robinson (1996), "Analysis of Variation Tranmission in Manufacturing Processes", submitted to *Journal of Quality Technology.*

[16] Little, Roderick J.A. and Donald B. Rubin, (1987) *Statistical Analysis with Missing Data.* Wiley and Sons, New York.

[17] Looney, Stephen W. (1995) "How to Use Tests for Univariate Normality to Assess Multivariate Normality", *The American Statistician*, Vol. 49, No. 1, 64-70.

[18] Mardia, K.V., J.T. Kent and J.M. Bibby, (1979) *Multivariate Analysis.* Academic Press, London.

[19] Magnus, Jan R. and H. Neudecker, (1979) "The Commutation Matrix: Some Properties and Applications", *The Annals of Statistics*, Vol. 7, No. 2, 381-394.

[20] Madansky, Albert, (1988) *Prescriptions for Working Statisticians*, Springer-Verlag, New York.

[21] Moen, Ronald D., Thomas W. Nolan and Lloyd P. Provost (1991) *Improving Quality through Planned Experimentation* McGraw-Hill, Inc.

[22] Montgomery, Douglas C. (1985) *Introduction to Statistical Quality Control* Wiley and Sons, New York.

[23] Montgomery, Douglas C. and Elizabeth A. Peck, (1992) *Introduction to Linear Regression Analysis*, Wiley and Sons.

[24] Nair, Vijayan N. (editor) (1992) "Taguchi's Parameter Design: A Panel Discussion", Technometrics, Vol. 34, No. 2, pp 127-161.

[25] Nolan, Thomas W. and Lloyd P. Provost, (1990) "Understanding Variation", *Quality Progress*, May, 70-78.

[26] Pignatiello, Joseph J. Jr (1993) "Strategies for Robust Multiresponse Quality Engineering" *IIE Transactions* Vol. 25, No. 3, 5-15.

[27] Provost, Lloyd P. and Clifford L. Norman, (1990) "Variation through the Ages", *Quality Progress*, December, 39-44.

[28] Pyzdek, Thomas, (1990) "There's No Such Thing as a Common Cause", *ASQC Quality Congress Transactions* 102-108.

[29] Roy, Ranjit (1990) *A Primer on the Taguchi Method* Van Nostrand Reinhold, New York.

[30] Seber, G.A.F. (1977) *Linear Regression Analysis*, Wiley and Sons, New York.

[31] Williams, D.A. (1987) "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions", *Applied Statistics*, Vol. 36, No. 2, pp. 181-191.

[32] Wu, Shing-Kuo, S. Jack Hu and S.M. Wu, (1994) "A Fault Identification and Classification Scheme for an Automobile Door Assembly Process," *The International Journal of Flexible Manufacturing Systems*, Vol. 6, No. 4, 261-285.

[33] Xie, Weimin, Kai Yang and Yuanzhan He, (1994) *ISSAT conference papers*, "A Multi-stage Multivariate Statistical Approach for the Diagnosis of Sheet Metal Assembly Processes", 102-106.

[34] Yang, Kai (1996), "Improving Automotive Dimensional Quality by Using Principal Component Analysis", *Quality and Reliability Engineering International*, Vol. 12, No. 6, 401-409.