ESTIMATING MEASUREMENT ERROR IN BLOOD PRESSURE, USING STRUCTURAL EQUATIONS MODELLING

by

Lulama Patrick Kepe

Thesis presented in partial fulfillment of the requirements for the degree of Masters in Science at the

Department of Statistics and Actuarial Science

Faculty of Science

Stellenbosch University

Supervisor

Dr Carl Lombard
Medical Research Council

Co-supervisor

Mr. Chris Muller
Department of Statistics and Actuarial Science
Stellenbosch University

2004

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously submitted it in part or in its entirety at any university for a degree.

Signature Date

Summary

Any branch in science experiences measurement error to some extent. This maybe due to conditions under which measurements are taken, which may include the subject, the observer, the measurement instrument, and data collection method. The inexactness (error) can be reduced to some extent through the study design, but at some level further reduction becomes difficult or impractical. It then becomes important to determine or evaluate the magnitude of measurement error and perhaps evaluate its effect on the investigated relationships. All this is particularly true for blood pressure measurement.

The gold standard for measuring blood pressure (BP) is a 24-hour ambulatory measurement. However, this technology is not available in Primary Care Clinics in South Africa and a set of three mercury-based BP measurements is the norm for a clinic visit. The quality of the standard combination of the repeated measurements can be improved by modelling the measurement error of each of the diastolic and systolic measurements and determining optimal weights for the combination of measurements, which will give a better estimate of the patient's true BP. The optimal weights can be determined through the method of structural equations modelling (SEM) which allows a richer model than the standard repeated measures ANOVA. They are less restrictive and give more detail than the traditional approaches.

Structural equations modelling which is a special case of covariance structure modelling has proven to be useful in social sciences over the years. Their appeal stem from the fact that they includes multiple regression and factor analysis as special cases. Multi-type multi-time (MTMT) models are a specific type of structural equations models that suit the modelling of BP measurements. These designs (MTMT models) constitute a variant of repeated measurement designs and are based on Campbell and Fiske's (1959) suggestion that the quality of methods (time in our case) can be determined by comparing them with other methods in order to reveal both the systematic and random errors.

MTMT models also showed superiority over other data analysis methods because of their accommodation of the theory of BP. In particular they proved to be a strong alternative to be considered for the analysis of BP measurement whenever repeated measures are available even when such measures do not constitute equivalent replicates. This thesis focuses on SEM and its application to BP studies conducted in a community survey of Mamre and the Mitchells Plain hypertensive clinic population.

Opsomming

Elke vertakking van die wetenskap is tot 'n minder of meerdere mate onderhewig aan metingsfout. Dit is die gevolg van die omstandighede waaronder metings gemaak word soos die eenheid wat gemeet word, die waarnemer, die meetinstrument en die data versamelingsmetode. Die metingsfout kan verminder word deur die studie ontwerp maar op 'n sekere punt is verdere verbetering in presisie moeilik en onprakties. Dit is dan belangrik om die omvang ven die metingsfout te bepaal en om die effek hiervan op verwantskappe te ondersoek. Hierdie aspekte is veral waar vir die meting van bloeddruk by die mens.

Die goue standaard vir die meet van bloeddruk is 'n 24-uur deurlopence meting. Hierdie tegnologie is egter nie in primêre gesondheidsklinieke in Suid-Afrika beskikbaar nie en 'n stel van drie kwik gebasseerde bloedrukmetings is die norm by 'n kliniek besoek. Die kwaliteit van die standard kombinasie van die herhaalde metings kan verbeter word deur die modellering van die metingsfout van diastoliese en sistoliese bloeddruk metings. Die bepaling van optimale gewigte vir die lineêre kombinasie van die metings lei tot 'n beter skatting van die pasiënt se ware bloedruk. Die gewigte kan berekening word met die metode van strukturele vergelykings modellering (SVM) wat 'n ryker klas van modelle bied as die standaard herhaalde metings analise van variansie modelle. Dié model het minder beperkings en gee dus meer informasie as die tradisionele benaderings.

Strukurele vergelykings modellering wat 'n spesial geval van kovariansie strukturele modellering is, is oor die jare nuttig aangewend in die sosiale wetenskap. Die aanhang is die gevolg van die feit dat meervoudige lineêre regressie en faktor analise ook spesiale gevalle van die metode is. Meervoudige-tipe meervoudige-tyd (MTMT) modelle is 'n spesifieke strukturele vergelykings model wat die modellering van bloedruk pas. Hierdie tipe model is 'n variant van die herhaalde metings ontwerp en is gebaseer op Campbell en Fiske (1959) se voorstel dat die kwaliteit van verskillende metodes bepaal kan word deur dit met ander metodes te vergelyk om sodoende sistematiese en stogastiese foute te onderskei. Die MTMT model pas ook goed in by die onderliggende fisiologies aspekte

van bloedruk en die meting daarvan. Dit is dus 'n goeie alternatief vir studies waar die herhaalde metings nie ekwivalente replikate is nie.

Hierdie tesis fokus op die strukturele vergelykings model en die toepassing daarvan in hipertensie studies uitgevoer in die Mamre gemeenskap en 'n hipertensie kliniek populasie in Mitchells Plain.

Parts of the research described in this thesis were presented at the following conference:

The 50th Annual Conference of the South African Statistical Association, which was held at the Caesar's Gauteng Convention resort from the 3-7 November 2003.

DEDICATION

This work is dedicated to both my mother and grand mother who sacrificed a lot to give me opportunities they never had.

ACKNOWLEDGMENTS

I am forever indebted to...

- My supervisor Dr Carl Lombard for his patience, support and guidance;
- Dr Lize van der Merwe for her guidance and contribution throughout these studies;
- Mr Chris Muller for his guidance and suggestions;
- Prof Stephan Maritz for his suggestions;
- Colleagues for their input and interest;
- My wife, Nomakwezi, for her patience, encouragement and prayers;
- The Stellenbosch University's Department of Statistics and Actuarial Science for financial support;
- The Medical Research Council of South Africa for the study time, computer facilities, and financial support;
- The National Research Foundation (NRF) for financial support towards this research. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to the NRF.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

BLOOD PRESSURE MEASUREMENT

1.1 What is blood pressure?

Arterial blood pressure (BP) is the lateral pressure, or force exerted by the blood on a unit area of the blood vessel wall. This arterial blood pressure constantly changes during the course of the cardiac cycle. The highest pressure in this cycle is the systolic blood pressure (SBP); the lowest is the diastolic blood pressure (DBP). The numerical difference between the two is the pulse pressure. In the *Systeme International de'Unites* the basic unit for pressure is newton per square meter, known as the pascal (Pa). Blood pressure is traditionally measured in millimeters of mercury (mmHg), where 1mmHg=7.5kPa [1].

1.2 Why measure blood pressure?

There is evidence that links high BP to medical conditions, such as stroke, angina, heart attack and renal failure [1][2][3]. In elderly or diseased persons it is common for arteries to lose their elasticity, resulting in the elevation of these persons' BP. People who are overweight, eat salty or fatty foods or those who are physically inactive develop high BP. Other factors that contribute towards high

BP are peripheral vascular resistance, viscosity of the blood, cardiac output and the volume of blood in the artery. Peripheral resistance and cardiac output seem to have the greatest influence on blood pressure [11]. Early detection of high BP is important, because this allows the patient to take preventive measures that include lifestyle changes, such as regular exercising.

1.3 Conditions of measuring blood pressure

Blood pressure is a physiological variable and is affected by various factors, such as anxiety, distention of the urinary bladder, extreme changes in temperature, exertion, pain and even recent smoking or food intake. The ideal BP measurements should thus be taken in a quiet environment and comfortable temperature, after the patient has been seated or lying down for at least five minutes. A single BP reading provides a predictive value of a patient's BP, but this needs to be confirmed by two or more readings taken at one examination. The level of the average of these readings will determine whether or not the patient should be rechecked. For example, a patient is classified as hypertensive according to the American Heart Association (AHA) [1] guidelines for high BP, if the DBP is greater than 100 mmHg for at least three measurements taken on different days.

Potential errors in diagnosing high BP are:

1) false positive diagnostic error

This error occurs when a person is labelled hypertensive, and is treated, when

in fact his or her underlying mean (true) BP is below a chosen cut point

2) false negative diagonistic error

This results in declaring persons normotensive (normal BP) when in fact the underlying true BP is above the high BP level. A false negative error is more serious, as it can lead to fatalities.

1.4 Measuring blood pressure

Blood pressure can either be measured directly or indirectly. Direct measurement of blood pressure is done by inserting a needle or catheter into the arterial tree and connecting it to a calibrated transducer. This provides a continuous measurement of the beat-by-beat arterial pressure. Indirect measurement is done with the use of manometers (*viz.*, sphygmomanometers, aneroid manometers, electronic manometers *etc.*). Indirect methods of measuring BP are based on the occluding-cuff auscultatory technique based on the work of Riva-Rocci (1896), Hill and Barnard (1897) and Korotkoff (1905). The following are techniques for indirect measurement of BP:

1.4.1 Palpatory technique

The palpatory technique is used to estimate, and not measure BP. It involves palpating the radial pulse and rhythm to estimate SBP. Diastolic blood pressure cannot be easily estimated by this technique. Only manual manometers can be

used for this technique.

1.4.2 Auscultatory technique

For this procedure a stethoscope is used. A cuff with a bladder is placed over the brachial artery around the upper arm. The bladder is inflated until its pressure exceeds that of the arterial pressure. This blocks the blood flow and stops the radial pulse from palpitating. When the pressure is released the blood in the artery starts flowing intermittently, producing sharp rythmic and knocking sounds (Korotkoff) with each cardiac beat. As the pressure continues to drop these sounds change in their quality and intensity and gradually fade. They finally disappear when the pressure in the bladder is smaller than that of the artery.

There are five Korotkoff phases:

Phase I: Sharp and clear sounds that are proportional to the intensity of the cuff deflation.

Phase II: Blowing or swishing sounds.

Phase III: Softer thud than *Phase I.*

Phase IV: Softer blowing sounds that disappear

Phase V: Silence

Phase I determines the SBP, whereas *Phase IV* is the DBP for children and *Phase V* the DBP for adults. In patients with hypertension brachial artery sounds

may disappear when the pressure in the cuff is very high and reappear later as the pressure is reduced. This is the so-called *auscultatory gap* and occurs in the latter part of *Phase I* and *Phase II*. This may lead to serious underestimation of the SBP or overestimation of the DBP.

1.5 Measuring devices

Blood pressure measuring devices are divided into two broad categories: manual manometers and automated manometers (comprising of devices for clinical use in hospitals, for self-use and for ambulatory BP). The former type is of simple design with standard components, thus it is assumed that it will be similar in accuracy. The latter type is used for measuring BP in community settings. It is almost impossible to assess the accuracy of these automated manometers.

1.5.1 Manual devices

These include mercury and aneroid manometers. For accurate results, they require the user to be trained and cautious. Techniques for using these are outlined in Section 1.4.2.

1.5.2 Automated devices

In recent years, automated devices are used more frequently. They are battery operated and display the readings digitally. They are either semi-automatic

(inflated manually) or fully automatic (inflated by pressing a button on the machine). They are useful for ambulatory blood pressure monitoring (ABPM). A patient wears the small ABPM device with a belt or strap over the shoulder. This device is usually set to measure BP at pre-set times, thus making it easy to investigate BP over longer periods of time.

1.6 Sources of error in BP measurement

Inaccurate measurements of BP arise either from observer error, defective equipment or failure to standardize the measurement technique.

1.6.1 The observer

An observer who is unable to clearly see the calibration marks on the manometer, or to clearly hear the Korotkoff sounds, or who is unskilled in interpreting the readings, will cause an error to BP measurements. Unskillful application of BP measurement techniques can result in the following errors:

1. Incorrect positioning of the extremity (limb). *Care should be taken that the position of the limb that is used for the BP measurement is at the level of the heart.*

2. Failure to place the mercury column vertically. *The mercury column must be placed at the level of the heart in a vertical position.*

3. Incorrect deflation of the cuff. *Pressure should be lowered at 2 mmHg per*

heartbeat.

4. Recording of the first BP. *Spasm of the artery upon initial compression, anxiety, or apprehension of the subject can cause abnormal readings, because of this the first reading should not be recorded.*

5. Auscultatory gap. *Very low systolic readings can be avoided by recording the BP by the palpatory method.*

6. Incorrect cuff positioning. *Both bulging and ballooning result in very high readings.*

1.6.2 Defective equipment

Faulty equipment includes a number of things, among these are:

1. Faulty air-release valve or tubing connections.

2. Leaking mercury reservoir, resulting in insufficient mercury.

3. Dirty column or bubbles in the column, preventing the flow of mercury.

4. Variations of microphone sensitivity in identifying Korotkoff sounds can cause errors in BP measurement.

1.6.3 Failure to standardize the measurement technique

Periodical review of recent methods of BP measurements are essential for people taking BP measurements. Standardized training for BP reading should in-

clude the most recent information on detection and referral. For example, filmed BP measurements may be used to standardize techniques, test accuracy and reproducibility, and to identify and correct the observer error.

1.7 Discussion

Each individual has a 'true' underlying BP measure around which the observed BP measurements vary by week, day or minute. The true BP estimated at a particular time will depend on the situation surrounding the patient at that time. For example, the true BP under a stressful work environment will differ from that under a relaxed environment.

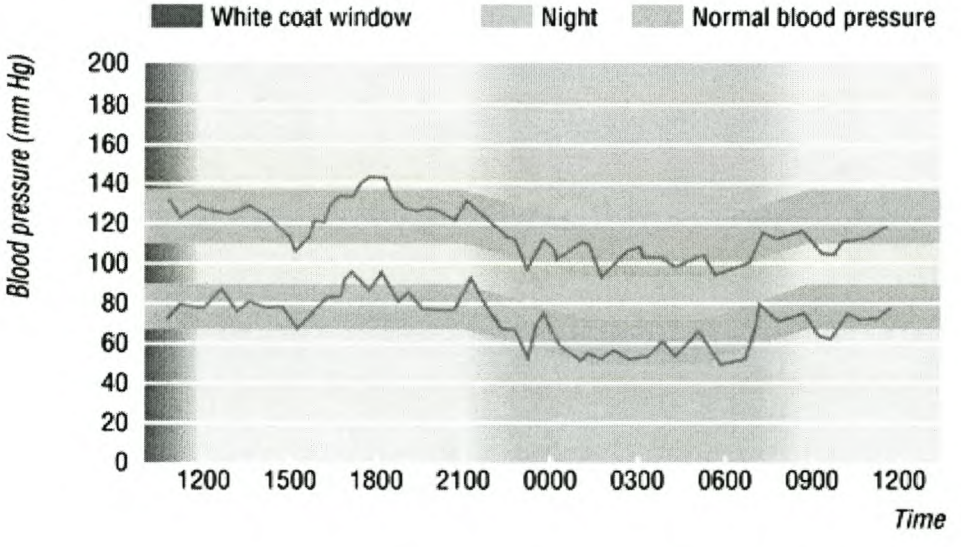

Figure 1.1: Systolic and Diastolic BP monitored over 24 hours

Figure 1.1 shows a typical true BP at different times of the day. Observed

BP measurements, on the other hand will vary according to the measurement technique, variables that are associated with the test subjects and the measurers, as well as the testing apparatus and the environment. The problem of white coat BP measurements (this is a condition in which a subject's BP rises during BP measurement and then falls to normal levels outside the medical environment) can also produce misleading readings. A simple description of the variability in BP measurements is insufficient, and a proper analytical model is needed to quantify the different sources of random variation. One clinical way of looking at the variation of the observed BP measurement from the 'true' one is to divide it into within-visit (between minutes apart) and between-visit (between days or weeks apart) components. The between-visit variability is a larger component of the BP variability and affects the accuracy of BP assessment.

It is known that both random and systematic measurement errors (ME) occur in the measurement of BP [4]. For example, a bladder of inappropriate size for a subject's limb circumference will cause a *systematic error* in the BP measurement. If the bladder is too broad the pressure will be underestimated and if it is too narrow an overestimation of the pressure occurs. On the other hand, BP cuffs from different manufacturers can cause *random errors* if used on the same subject. Recent medical methods of dealing with ME make use of random error variances. A random error variance is usually estimated by computing the correlation between several measurements on the same subject or by estimating the

within-subject variance across several measurements by means of some form of analysis of variances (ANOVA) model.

Random error causes differences even between repeated measurements taken under exactly the same conditions. Systematic error on the other hand becomes apparent only if some conditions are changed across measurements. Such systematic errors can be decomposed into bias (constant across subjects) and variance (variant across subjects) components [4]. Bias here is defined as the extent to which the expected value of an estimator differs from the population parameter. Unlike the random error, bias cannot be reduced or eliminated by increasing the sample size.

The accuracy of BP measurements has always been improved by averaging several measurements. In this thesis other recent methods in medical science that are used for improving the quality of measurements, particularly BP measurements will be investigated. This is especially important in public health research since an error of only a few millimeters of mercury is enough to misclassify a subject as either normotensive or hypertensive.

The research by Batista et al. [4] on structural equations modelling inspired the work done in this thesis. Structural equations models show superiority over other data analysis methods because of their accommodation of the theory of BP. In particular they prove to be a strong alternative to be considered for the analysis of BP measurements whenever repeated measures are available even when such

measures do not constitute equivalent replicates. In this work the focus will be on structural equations models and their application to BP studies conducted in a community survey of Mamre and the Mitchells Plain hypertensive clinic population.

CHAPTER 2

BLOOD PRESSURE STUDIES

2.1 Introduction

Blood pressure measurements are made on a physiological process that is not homogeneous or in a steady state over time. The instruments that are used to measure the systolic and the diastolic BP at a specific moment, produce a unique measurement that cannot necessarily be repeated exactly at the same level. Such repeated measurements are often analyzed as if they were true replicates, which in essence they are not. In this chapter data from two differently designed studies are introduced with the intention of exploring their nature. Through exploratory data analysis it is expected that the structural equation model most suitable for the specific data will be found. The two studies concerned are the Mamre study and the Mitchells Plain study.

2.2 Exploratory data analysis

For each of the studies, Tables containing summary statistics will be presented. The univariate statistics used in the analysis are the mean and the standard deviation (SD) of each of the BP measures in the data set. The mean gives

the average BP readings and the SD indicates how much the BP readings vary around the average. A Table of bivariate statistics, for meaningful pairs of BP readings, is also presented. For each pair, the mean, the difference (DIFF) and the measurement error (ME) (calculated as the square root of the mean square error) is given.

Side-by-side boxplots are drawn for SBP and DBP readings. These plots provide information about the location and the variability of the data and any extreme data values. The minimum and the maximum values are shown and the median is the horizontal line inside the box. The box represents the middle 50% of the data. A rough impression is given of the symmetry of the distribution by examining the symmetry of the boxplot. The whiskers extend from the box as far as the smallest or largest data value or at the most a distance of 1.5 interquartile units. Points further away from the box are outliers and are shown individually.

Scatterplots are drawn to represent the relationship between different BP readings on the same subject at different times. These scatterplots are drawn for all pairs of BP readings with the corresponding mean-difference plots next to them. The scatterplots always have a chronological reading i ($i = 1, 2$) on the y-axis and the reading j ($j > i = 2, 3$) on the x-axis. The difference is always calculated as: reading i-reading j. If the pairs of BP readings lie close to the 45 degree line through the origin, it means that there are small differences between the readings and hence a good reliability. This is desirable since the more reliable the readings

are, the less the measurement error will be.

In the scatterplots of the difference against the mean small homogeneous differences (close to zero) over the mean range for a reliable measure is expected. When the mean difference is above or below the zero line on the mean-difference plot, it indicates some form of systematic bias. The shapes emanating from the mean-difference plots for the data are also interpreted.

- If a "uniform band" like shape is observed, it suggests a consistent or homogeneous measurement error throughout the BP readings.

- If a "funnel" like shape is observed, it suggests a non-consistent or heterogeneous measurement error.

A lowess smooth curve is fitted to the differences in the mean-difference plot. This is a non-parametric regression function calculating the mean expected value using the local data. The lowess curve provides an estimate of the extent and consistency of the differences.

2.3 Mamre study

The first data set, the Mamre study [57], is from a health survey that was carried out in the community of Mamre in 1989. The aim of the Mamre study was to determine the prevalence of hypertension and other risk factors for cardiovascular diseases. The random sample consisted of 975 community members ranging from

Notation		Meaning
SIST	=	systolic BP
DIAS	=	diastolic BP
MSIST	=	mean of SBPs
DSIST	=	difference between SBPs
MDIAS	=	mean of DBPs
DDIAS	=	difference between DBPs

Table 2.1: Notation for BP measures in the Mamre data

14 to 86 years old. Blood pressure readings were taken in a seated position at three time points, five minutes apart in the same visit, using a mercury manometer. Both the SBP and DBP were recorded at each time point. The notation given in Table 2.1 was used for naming the BP measures and transformations in Table 2.3, boxplots [figure 2.1] and scatterplots [figures 2.2 and 2.3].

Univariate descriptive statistics of the Mamre data are given in Table 2.2. There was a small decrease in the mean SBP and DBP over time. The SD indicates that systolic readings varied more than diastolic readings, and that there was a decrease in variation for both the SBP and DBP from the first to the third reading.

Table 2.3 shows pairwise descriptive statistics for the Mamre data. The smallest difference between the SBP readings is between the second and the third systolic readings. There is almost no difference between the second and the third DBP readings. The largest difference (approximately 3.5 mmHg) is between the

	Systolic		Diastolic	
	Mean	SD	Mean	SD
Repeat				
1	133.44	25.65	78.85	13.96
2	131.14	25.24	77.57	13.95
3	129.98	24.87	77.06	13.93

Table 2.2: Summary statistics of the Mamre data

first and the third SBP readings.

The boxplots (figure 2.1) reveal that there is skewness in the systolic BP measures, and less so in the DBP readings.

Figure 2.2 shows the comparisons between the SBP readings. The majority of the differences between the SBP readings fall between -20 and 20 mmHg. There is a slight indication of non-zero differences among the systolic readings according to figure 2.2. The mean-difference plot between the first and the second SBP readings shows a uniform band. The comparison of the third SBP with either the first or the second SBP readings produced a funnel shape. Thus, it appears that involvement of the third systolic reading produces non-consistent measurement error.

Figure 2.3 shows the scatterplots of the diastolic readings against each other with corresponding mean-difference plots. The majority of the differences between the DBP readings fall between -10 and 15 mmHg. The mean-difference plots

Pair		MEAN	DIFF	M.E
SIST1	SIST2	132.29	2.3	25.45
SIST2	SIST3	130.56	1.16	25.06
SIST1	SIST3	131.71	3.46	25.26
DIAS1	DIAS2	78.21	1.28	13.96
DIAS2	DIAS3	77.32	0.51	13.94
DIAS1	DIAS3	77.96	1.79	13.95

Table 2.3: Statistics for pairs of BP readings in the Mamre study

indicate the presence of very slightly non-zero differences among the DBP readings. The comparisons of the DBP readings mostly produced a uniform band. There is more variation between the first and the third readings, than there is between consecutive pairs. There is an indication of a homogeneous measurement error among the DBP readings.

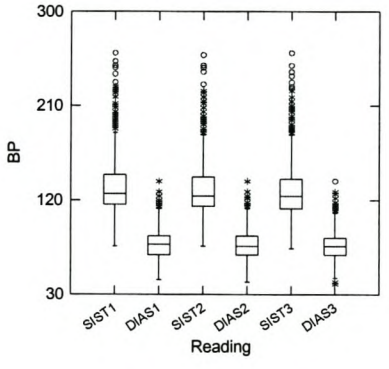

Figure 2.1: Boxplots of the six consecutive BP readings in the Mamre study

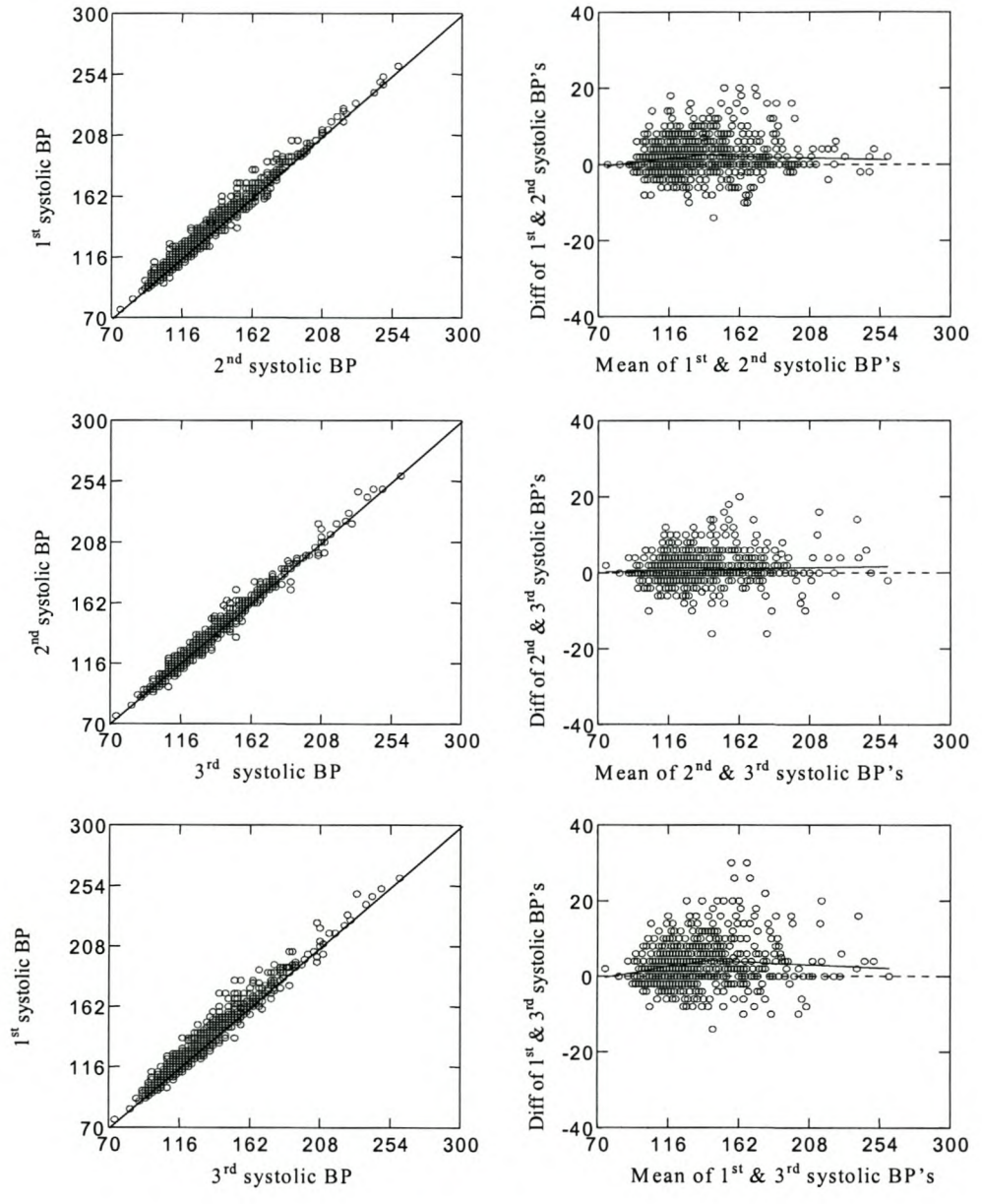

Figure 2.2: Scatterplots and mean-difference plots for comparison of systolic BP readings in the Mamre study

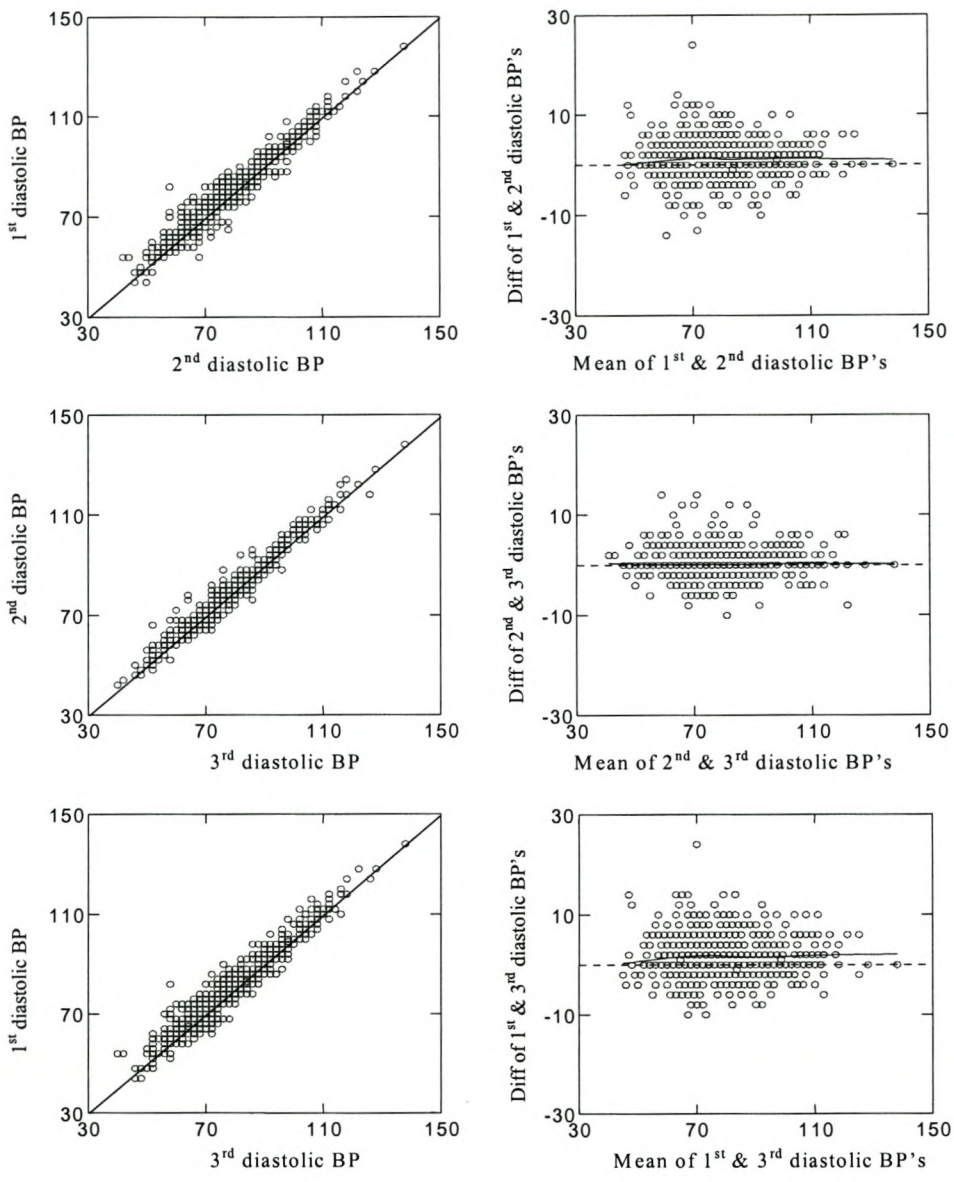

Figure 2.3: Scatterplots and mean-difference plots for comparison of diastolic BP readings in the Mamre study

2.4 Mitchells Plain study

The second data set is from a community health centre in the Cape Peninsula. The aim of this Mitchells Plain study [56] was to assess the treatment status, knowledge and experience of hypertensive patients. Two hundred and two hypertensive patients were first interviewed, and their BP was then measured electronically with a digital machine (Dinamap Vital Signs monitor) (D) and then manually with a mercury manometer (HG). The Dinamap Vital Signs monitor is a fully automatic monitor and was set to give readings at two minute intervals. After three sets of readings were recorded, a mercury manometer was used to take three more sets of BP readings at two minute intervals.

The descriptive statistics for the Mitchells Plain study are displayed in Table 2.4. The statistics indicate that there is a decrease in the mean value of the SBP over all six readings and a decrease in the DBP readings within each measuring method. A jump occurs in the DBP readings with the switch from the digital to the mercury machine. Whereas there is no clear pattern of the variance of the BP readings within a machine, the digital machine gave slightly higher variances than the mercury machine. It can also be seen that the digital machine gave, on average, higher readings for the SBP than the mercury manometer.

In figure 2.4 boxplots of the BP readings from different machines are displayed. There seems to be many high outliers and no low ones, indicating positive skewness in the data. This is understandable, given that these data are taken from

		Systolic		Diastolic	
Method	Repeat	Mean	SD	Mean	SD
Digital	1	169.00	29.99	90.77	14.80
	2	165.00	29.50	89.19	14.32
	3	162.30	30.01	88.46	14.99
Mercury	1	158.36	28.46	91.11	14.04
	2	156.30	28.69	90.28	14.16
	3	154.61	28.49	89.42	14.44

Table 2.4: Summary of descriptive statistics for the Mitchells Plain data

people with high BP. It can also be seen that the mercury manometer gave more outliers than the digital machine for the SBP reading. There were no clear differences between readings within these machines when measuring DBP. Variations among the systolic readings are almost double those of the diastolic BP readings, confirming what was found with the summary statistics of Table 2.4. There is a slight decrease of the median through the digital and mercury systolic readings. There is no evidence of such a decrease among the diastolic readings.

Figure 2.5 contains scatterplots and mean-difference plots showing the relationship between all pairs of digital systolic BP measurements. The mean-difference plots show that the differences lie between -30 to 30 mmHg. The largest difference between the SBP readings, is between the first and the third, which is approximately 7 mmHg, and the average difference between the SBP

Pair		MEAN	DIFF	M.E
Dig. SBP1	Dig. SBP2	167.00	4.00	29.75
Dig. SBP2	Dig. SBP3	163.65	2.7	29.80
Dig. SBP1	Dig. SBP3	165.65	6.7	30.00
Dig. DBP1	Dig. DBP2	89.98	1.58	14.56
Dig. DBP2	Dig. DBP3	88.83	0.73	14.66
Dig. DBP1	Dig. DBP3	89.62	2.31	16.90

Table 2.5: Summary statistics for pairs of digital BP measures in the Mitchells Plain study

readings is just above 4 mmHg. There is a slight increase in the differences between data points from adjacent pairs to those that are further apart (first and the third measures) and an indication of the non-zero differences (the first measurement is often larger than the second, as indicated by the majority of the points being above the zero line and the fitted lowess curve). All three mean-difference plots in figure 2.5 show a uniform band, suggesting the presence of a consistent measurement error.

Figure 2.6 shows the relationship between pairs of mercury SBP readings. The mean-difference plots show that the differences lie between -30 to 30 mmHg. In fact, Table 2.6 shows that the largest difference is between the first and the third readings, approximately 3.8 mmHg. The average difference for these mercury systolic readings is approximately 2.5 mmHg. There is also a slight increase in

Pair		MEAN	DIFF	M.E
Mer. SBP1	Mer. SBP2	157.33	2.06	28.58
Mer. SBP2	Mer. SBP3	155.46	1.69	28.59
Mer. SBP1	Mer. SBP3	156.49	3.75	28.48
Mer. DBP1	Mer. DBP2	90.70	0.83	14.10
Mer. DBP2	Mer. DBP3	89.85	0.86	14.30
Mer. DBP1	Mer. DBP3	90.27	1.69	14.24

Table 2.6: Summary statistics for pairs of mercury BP in the Mitchells Plain study

the spread of the data points from adjacent pairs to those that are further apart (first and the third measures). The mean-difference plots for all these comparisons has almost a uniform band shape, indicating a homogeneous measurement error.

Figure 2.7 shows pairs of mercury DBP readings against each other. The mean-difference plots show that there are very small differences between the readings. The lowess curve also confirms this in that it lies very close to the zero line. The majority of the differences fall between -15 to 15 mmHg. Table 2.6 shows that the average difference is approximately 1 mmHg. The largest difference which is between the first and the third readings is almost 1.7 mmHg. Both pairwise comparisons between the second reading with either the first or the third readings produced a difference of less than 1 mmHg. There is consistency in the reading of the DBP using the mercury manometer.

Relationships between the digital DBP measures are shown in figure 2.8. The

mean-difference plots show that the differences lie between -20 to 20 mmHg. Table 2.5 reveals that there is almost no difference between the second and the third diastolic readings. The smallest difference is between the second and the third readings, approximately 0.7 mmHg. The largest difference (which is approximately 2.3 mmHg) is between the first and the third readings. The average of the differences between the digital diastolic readings is almost 1.5 mmHg. From the plots in figure 2.8 there is an indication of small differences between readings. The first diastolic reading is clearly higher than the rest (differences fall mostly above the zero line). The mean-difference plots reveal a uniform band, thus there is a consistent measurement error among digital diastolic readings.

Figure 2.9 shows the comparison of the SBP readings of the different machines. One can see from the mean-difference plots that not only is there an indication of the presence of large differences between the readings, but also that these differences appear to be non-constant (see the fitted curve). The mean-difference plots show that the differences lie between -20 and 30 mmHg. These mean-difference plots reveal a funnel shape form. Table 2.7 shows that the smallest difference is between the third systolic readings, which is approximately 7.7 mmHg, and the largest difference is between the first readings and is approximately 10.6 mmHg. The average of the differences is approximately 9 mmHg. Hence, there may exist a heterogeneous measurement error between these pairs.

Figure 2.10 reveals how the DBP measurements from the different machines

Pair		MEAN	DIFF	M.E
Dig. SBP1	Mer. SBP1	163.68	10.64	29.23
Dig. SBP2	Mer. SBP2	160.65	8.70	29.10
Dig. SBP3	Mer. SBP3	158.46	7.69	29.25
Dig. DBP1	Mer. DBP1	90.94	-0.34	14.42
Dig. DBP2	Mer. DBP2	89.74	-1.09	14.24
Dig. DBP3	Mer. DBP3	88.94	-0.96	14.72

Table 2.7: Summary statistics for pairs of BP measures from different machines in the Mitchells Plain study

compare. The mean-difference plots show that the differences lie between -20 and 20 mmHg. It appears that there is no obvious difference between the two machines when measuring the DBP. This is supported by the lowess smooth curves since they are very close to zero. In Table 2.7 the means of the digital diastolic readings are shown to be smaller than those of the mercury diastolic readings. The average of the differences between the diastolic readings from the different machines is less than 1 mmHg, thus confirming that there were very small differences due to the machine when measuring the diastolic BP. There is an indication of homogeneous measurement error.

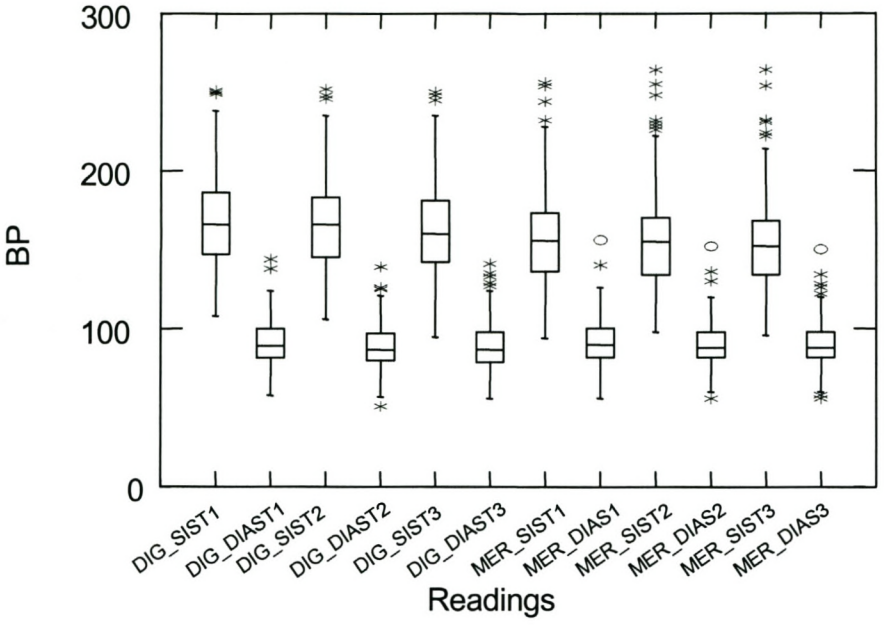

Figure 2.4: Boxplots representing all twelve consecutive BP readings in the Mitchells Plain study

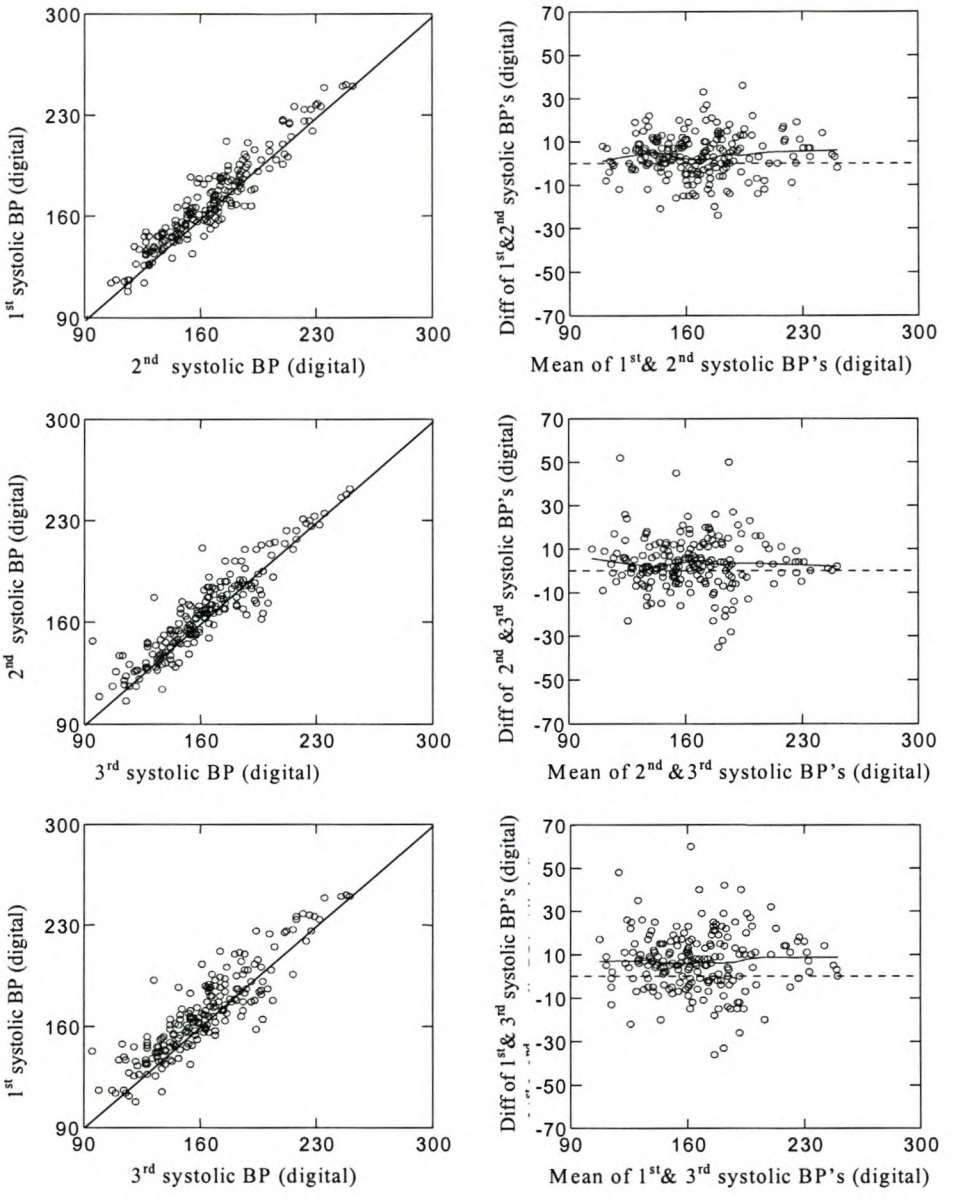

Figure 2.5: Scatterplots and mean-difference plots for comparison of digital systolic BP readings in the Mitchells Plain study

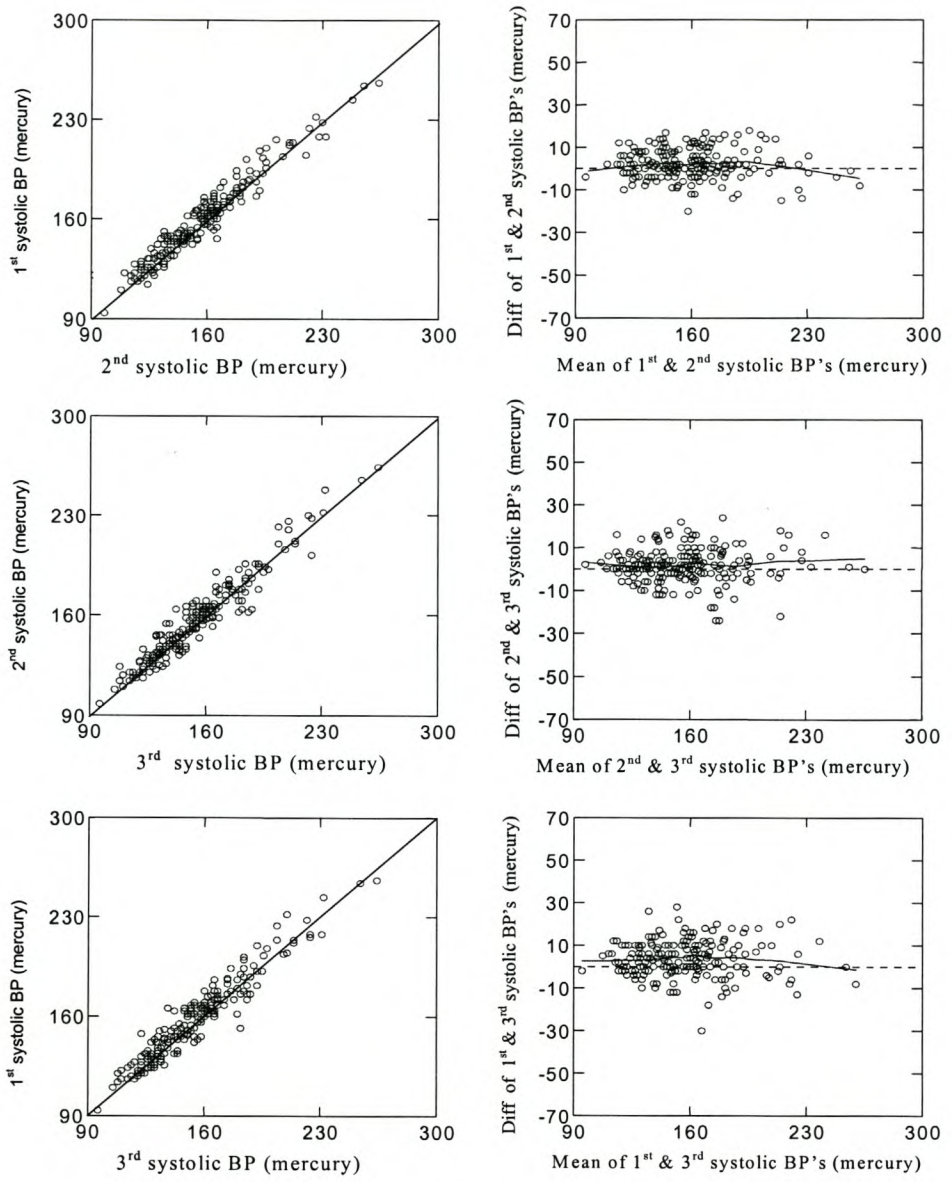

Figure 2.6: Scatterplots and mean-difference plots for comparison of mercury systolic BP readings in the Mitchells Plain study

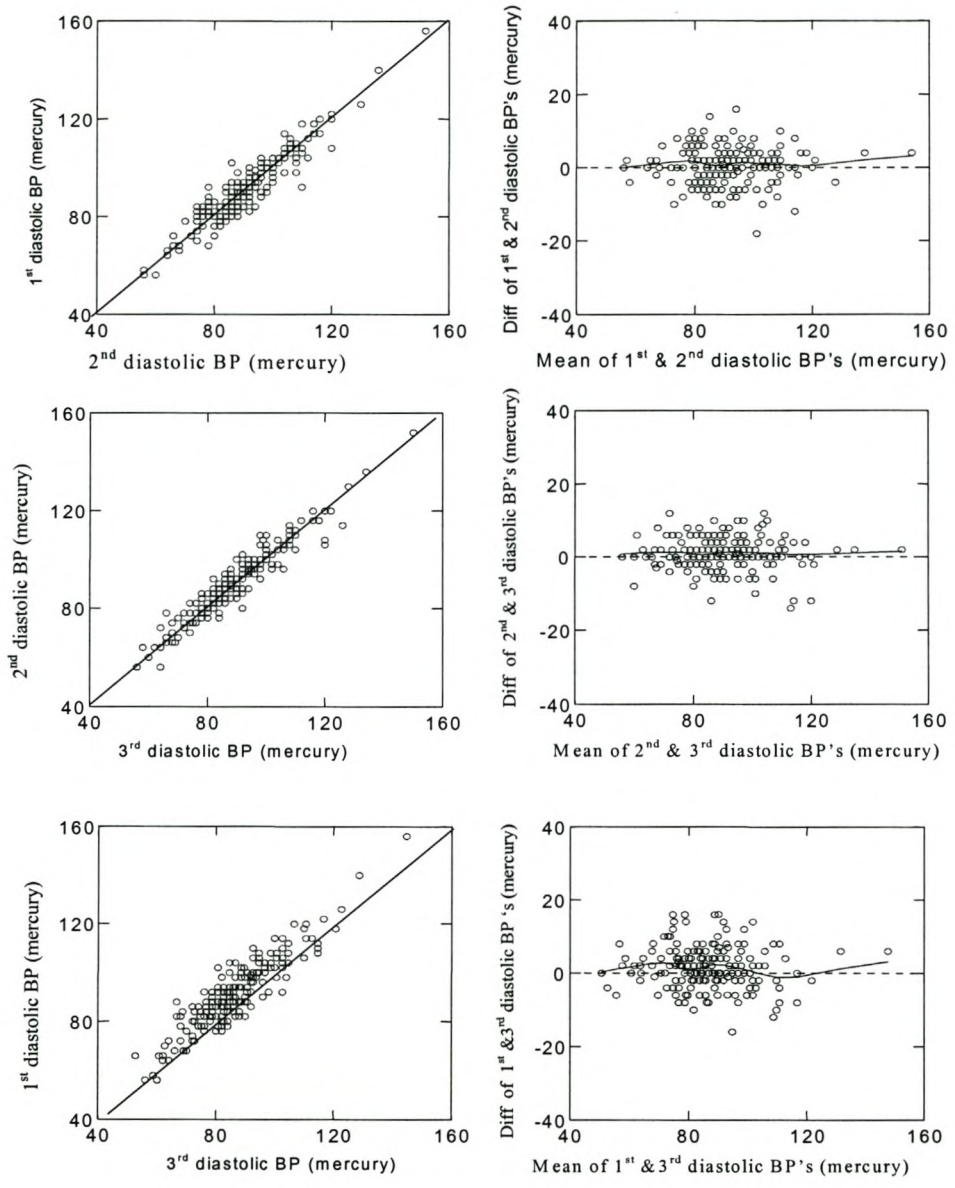

Figure 2.7: Scatterplots and mean-difference plots for comparison of mercury diastolic BP readings in the Mitchells Plain study

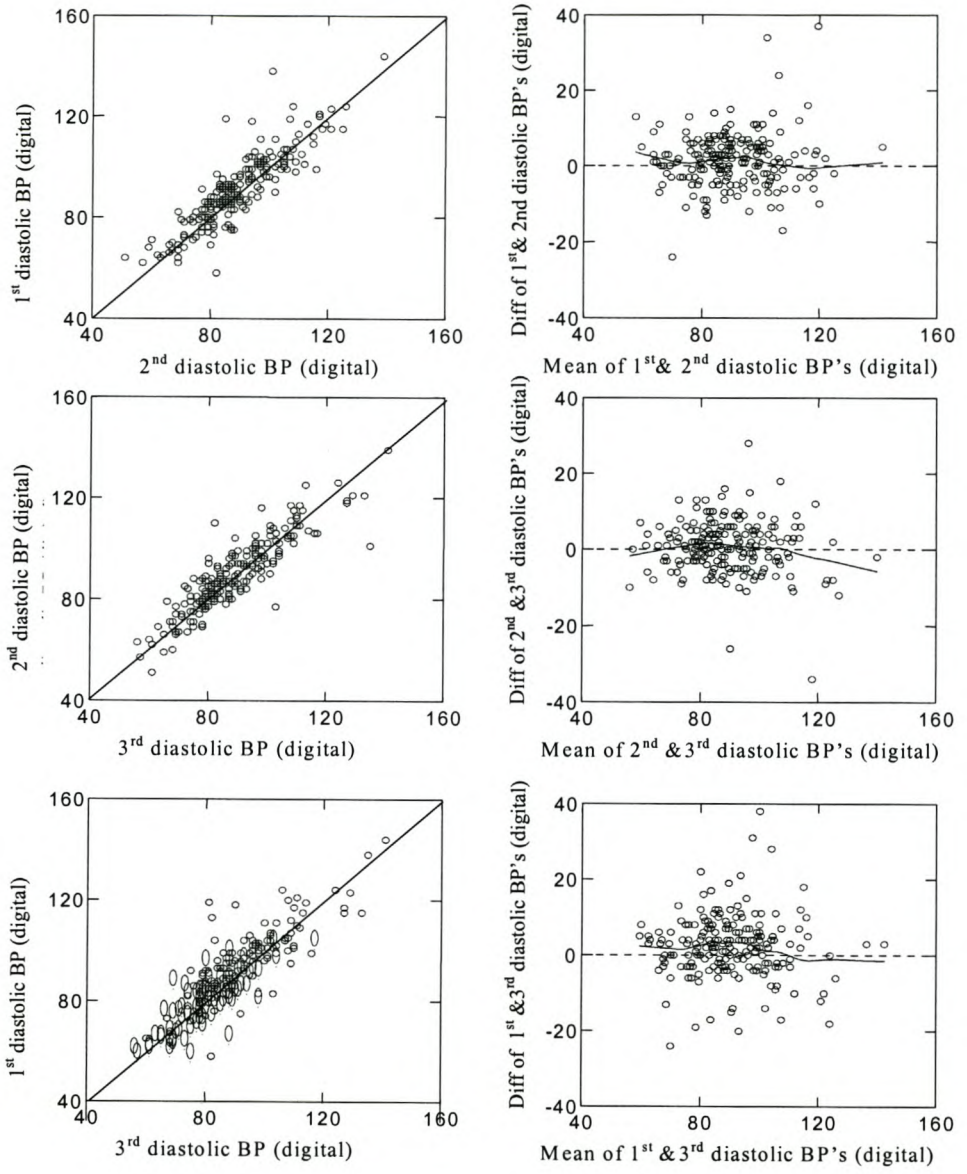

Figure 2.8: Scatterplots and mean-difference plots for comparison of digital diastolic BP readings in the Mitchells Plain study

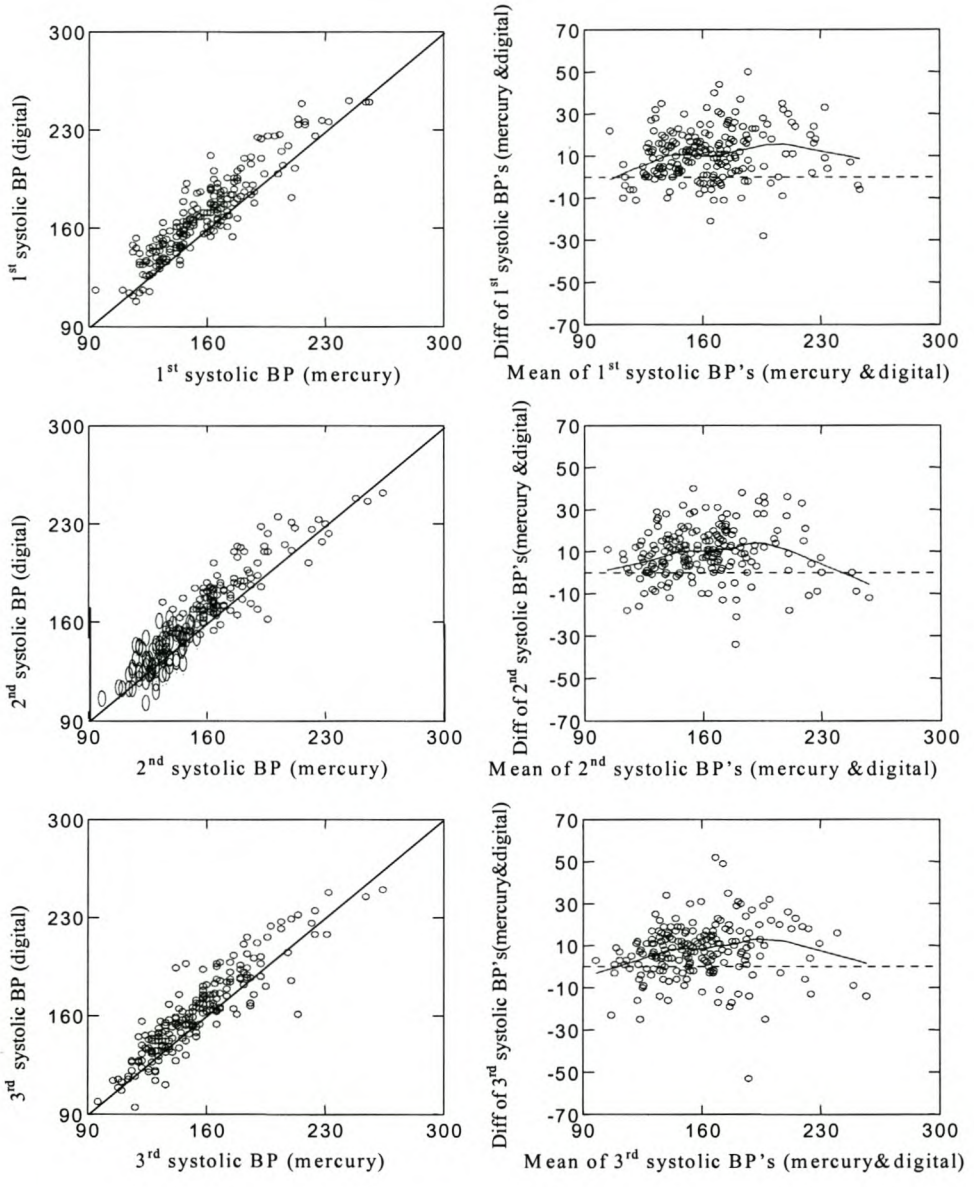

Figure 2.9: Scatterplots and mean-difference plots for comparison of mercury and digital systolic BP readings in the Mitchells Plain study

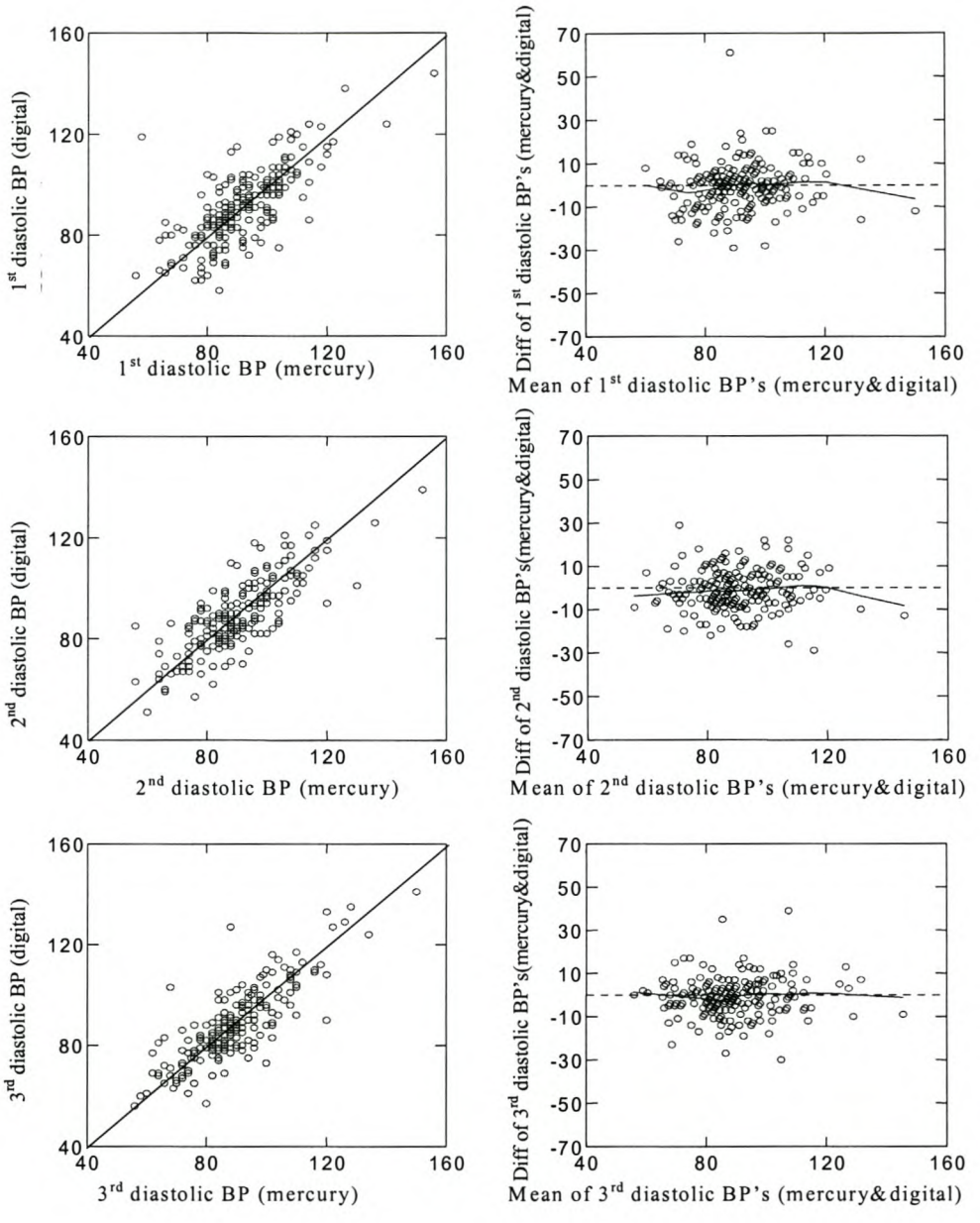

Figure 2.10: Scatterplots and mean-difference plots for comparison of mercury and digital diastolic BP readings in the Mitchells Plain study

2.5 Summary

- For both the studies and all BP readings, there was larger differences between the first and the third BP readings than between the consecutive pairs.

- The Mitchells Plain data showed a slight time-related bias in the SBP readings for both methods.

- Time ordered differences between DBP readings showed no bias for any of the studies. Therefore, DBP readings are closer to being "replicates" than SBP readings in the two studies.

- In the Mitchells Plain data the differences between SBP measurements taken by different methods reveal a mixture of machine and time-related bias (figure 2.9).

- Data will be modelled within a machine in the Mitchells Plain study when fitting the structural equation models in Chapter 6, thus homogeneous measurement error will be assumed. Hence aspects such as the calibration of the machines were not considered and are thus deemed unnecessary in this respect.

- These properties will be incorporated into the model for estimation of BP measurement error in Chapter 5.

CHAPTER 3

MEASUREMENT ERROR

3.1 Introduction

Every measure taken by a clinician or a scientist will have a shadow component, termed, error, which represents the difference between the observed value and the unknown true value. For example an observed BP reading may differ from the true underlying value. Theoretically the true value remains unknown, thus it becomes crucial to estimate errors associated with observed values as a means of gaining confidence in devices and the techniques used. It must be kept in mind that errors can be either random or systematic. Determining the distribution of the random errors is of interest, and can be achieved by taking repeated measurements of the same phenomenon. If repeated measurements are accurate and reliable (standard deviation of the distribution of errors is small), then the measurement error is small, and the observed values are assumed to be close approximations of the true values. Under such conditions one can regard measuring devices and techniques employed as reliable.

3.2 Replicated measurements

Replicated measurements arise when several measurements of the same entity are taken of the same subject (individual) during the same session. Such measurements are viewed as independent and equivalent replicates with the same expected value. Blood pressure measures do not satisfy these conditions even if they are taken at one visit, since the true underlying BP varies with time (see figure 1.1). Thus BP measures cannot be replicates of each other. Furthermore, replicates are interchangeable whereas BP readings cannot be even under the same conditions.

3.3 Repeated measurements

Repeated measurements occur when several measurements of the same entity are taken of the same subject at different times and/or conditions. An example is various BP readings taken at one visit or once at different visits. In the models for this thesis the BP measurements at one visit are regarded as repeated. A reproducibility study with repeated measurements of exposure variables (for example, observed BP measures) can be deployed to estimate random measurement error. On the other hand to evaluate systematic measurement error, a validation study of the device used against a "gold standard" or another device known to be superior to it is required. In BP measurement, for example, a sphygmomanometer's validity can be measured by comparing its readings with intra-arterial pressures taken directly by inserting a catheter or a needle.

3.4 Statistical Model

Suppose BP measurements are taken on more than one patient at one visit. The observed BP measures will be subject to error just as described above. Let X_{ij} be the observed value j of the BP of patient i. The classical unbiased additive model for the observed BP measurements, if replication of measures is assumed is:

$$X_{ij} = \mu_i + \epsilon_{ij} \text{ for } i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m \tag{3.1}$$

where μ_i is the true underlying measurement of the BP of patient i and ϵ_{ij} are assumed to be independent random errors from a distribution with a zero expectation. The ϵ_{ij} are assumed to be independent and are normally distributed with $E(\epsilon_{ij})^2 = \sigma^2$.

3.4.1 Estimation of error for *m=2*

Suppose two BP measurements are made on the same day with the same device, the observed measurements can be modelled as follows :

$$X_{ij} = \mu_i + \epsilon_{ij} \text{ for all } i = 1, 2, 3, \ldots, n;\ j = 1, 2 \tag{3.2}$$

Using the method of moments' estimation, let

$$\widehat{\epsilon}_{ij} = X_{ij} - \widehat{\mu}_i \quad i = 1, 2, 3, \ldots, n;\ j = 1, 2 \tag{3.3}$$

$\sum_{j=1}^{2} \widehat{\epsilon_{ij}} = 0$ then the following equations can solved

$X_{i1} - X_{i2} = \widehat{\epsilon}_{i1} - \widehat{\epsilon}_{i2}$ and $\widehat{\epsilon}_{i1} + \widehat{\epsilon}_{i2} = 0$ simultaneously, and it follows that

$X_{i1} + X_{i2} = 2\widehat{\mu}_i + 0$ which leads to:

$$\widehat{\mu}_i = \frac{\sum_{j=1}^{2} X_{ij}}{2} = \overline{X} \tag{3.4}$$

The main interest here is to estimate the error variance, where $E(\epsilon_{ij} - 0)^2 = \sigma^2$ and its estimator is be given by

$$s_i^2 = \frac{\sum_{j=1}^{2} \widehat{\epsilon}_{ij}^2}{2-1} \quad i = 1, 2, \ldots, n \tag{3.5}$$

3.4.2 Estimation of error for $m > 2$

The above equation is extendable to values of m beyond two, in fact in general it is

$$s_i^2 = \frac{\sum_{j=1}^{m} (X_{ij} - \overline{X}_{i.})^2}{n-1} = \frac{\sum_{j=1}^{m} \widehat{\epsilon_{ij}}^2}{m-1} \quad for\ i = 1, 2 \ldots, n\ and\ j = 1, 2, \ldots, m \tag{3.6}$$

For illustration purposes the following data, which is a sample from the Mamre study described in Chapter 2 is introduced.

IDNO	DIAS1	DIAS2	DIAS3	MEAN	SD	VARIANCE
1	65	78	64	69.00	7.81	61.00
2	86	88	86	86.67	1.15	1.33
3	78	78	78	78.00	0.00	0.00
4	96	96	94	95.33	1.15	1.33
5	84	80	78	80.67	3.06	9.36
6	88	86	86	86.67	1.15	1.33
7	92	88	96	92.00	4.00	16.00
8	72	72	68	70.67	2.31	5.33
9	68	68	68	68.00	0.00	0.00
10	94	92	92	92.67	1.15	1.33

Table 3.1: Repeated DBP measurements from a sample of ten subjects.

For a case with three measurements ($m = 3$).

For subject with IDNO=1 in Table 3.1

$$\widehat{\mu}_1 = \frac{65 + 78 + 64}{3} = 69 \tag{3.7}$$

also

$$s_1^2 = \frac{(-4)^2 + (9)^2 + (-5)^2}{2} = 61 \tag{3.8}$$

An extension of the above s_i^2 is used in the ANOVA of repeated measurements.

Here several measurements of the same quantity are measured, for example, BP measurements taken at the same visit. The standard deviation of repeated measurements on the same subject enables one to measure the size of the measurement error. In such a case it can be assumed that this SD is the same for all subjects (or variances are equal and hence homogeneous). This can be checked by plotting each individual subject's SD against their means and to look for obvious relations. The main exception would be a case where measurement error depends on the size of the measurement, a situation that will not be discussed here. This particular SD of repeated measurements is known as the within-subject standard deviation, that will be denoted by s_w and is given by:

$$s_w = \sqrt{\frac{\sum_{i=1}^{n} s_i^2}{n}} \tag{3.9}$$

where $s_i^2 = \frac{\sum_{j=1}^{m} \widehat{\epsilon_{ij}}^2}{m-1}$ are variances of each of the n subjects.

In an ANOVA setting s_w^2 (within-subject variance) is the so-called residual mean square. Measurement error can be quoted as s_w. This method proves to be very useful since it automatically takes care of the case where subjects have different numbers of observations (when there are missing or extra values). This can be illustrated by an example.

Using the above data from Table 3.1 (n=10) s_w, for $n = 10$, is calculated as

follow:

$$s_w = \sqrt{\frac{\sum_{i=1}^{n} s_i^2}{n}} = \sqrt{\frac{(7.81)^2 + (1.15)^2 + \cdots + (1.15)^2}{10}} = \sqrt{9.699} = 3.11 \qquad (3.10)$$

By making the usual one-way ANOVA assumptions, such as independence of errors, homogeneity of variances, and normality of errors a one way ANOVA can be described as follow. The subjects' mean square represents the between groups sum of squares and the residual means square is the within-group sum of squares.

The ANOVA Table (Table 3.2) decomposes the variance of the data into two components, which are the between-group component and the within-group component. The F ratio for groups, which in this case is 32.5, is a ratio of the between-group to the within-group estimate. Since the p-value which is the upper tail of the F distribution whose numerator degrees of freedom are those associated with between-groups and whose denominator degrees of freedom are those associated with the within-groups is far less than 0.05, there is a statistically significant difference between the means of the various groups at 5% significance level.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Var ratio	p-Value
Subjects (between-grps)	9	2838.987	315.443	32.5	< 0.0001
Residual (within-grps)	20	193.980	9.699		
Total	29	3032.967			

Table 3.2: One way ANOVA table:

3.4.3 The joint model for the systolic and diastolic blood pressures

Assuming that the SBP and DBP taken at one visit can be modelled in the manner described above (unbiased additive error model), let $X_{s_{ij}}$ denote the *jth* observed SBP measure and $X_{d_{ij}}$ the *jth* observed DBP measure for subject i, then the joint model of the pair $(X_{s_{ij}}, X_{d_{ij}})$ will be a bivariate additive error model. Let μ_i be the true underlying BP and let δ_i represent the amplitude difference from the true mean and assume that the $\epsilon's$ are independent of each other. The underlying SBP is modelled as $\mu_i + \delta_i$ and the underlying DBP as $\mu_i - \delta_i$. Then:

$$X_{s_{ij}} = \mu_i + \delta_i + \epsilon_{s_{ij}} \; i = 1, 2, \ldots, n; j = 1, 2, \ldots, m \tag{3.11}$$

and

$$X_{d_{ij}} = \mu_i - \delta_i + \epsilon_{d_{ij}} \; i = 1, 2, \ldots, n; j = 1, 2, \ldots, m \tag{3.12}$$

and subsequently the joint model will be

$$\begin{bmatrix} X_{s_{ij}} \\ X_{d_{ij}} \end{bmatrix} = \begin{bmatrix} \mu_i + \delta_i \\ \mu_i - \delta_i \end{bmatrix} + \begin{bmatrix} \epsilon_{s_{ij}} \\ \epsilon_{d_{ij}} \end{bmatrix} \tag{3.13}$$

Hence

$$E \begin{bmatrix} X_{s_{ij}} \\ X_{d_{ij}} \end{bmatrix} = \begin{bmatrix} \mu_i + \delta_i \\ \mu_i - \delta_i \end{bmatrix} \tag{3.14}$$

and also

$$cov \begin{bmatrix} X_{s_{ij}} \\ X_{d_{ij}} \end{bmatrix} = \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_d \\ \rho\sigma_s\sigma_d & \sigma_d^2 \end{bmatrix} \tag{3.15}$$

where ρ is the correlation between the SBP and DBP and σ_s^2 and σ_d^2 are variances of the respective random errors. Here μ_i and δ_i are constant for each subject and only the random errors $(\epsilon'_{dij}s)$ and $(\epsilon'_{sij}s)$ are varying.

If two pairs of BP measurements are given, namely $\begin{pmatrix} X_{s_{i1}} \\ X_{d_{i1}} \end{pmatrix}$ and $\begin{pmatrix} X_{s_{i2}} \\ X_{d_{i2}} \end{pmatrix}$, it could be estimated that:

$$\widehat{\mu}_i = \frac{X_{s_{i1}} + X_{d_{i1}} + X_{s_{i2}} + X_{d_{i2}}}{4} = \frac{\overline{X}_{s_i} + \overline{X}_{d_i}}{2} \tag{3.16}$$

and

$$\widehat{\delta}_i = \frac{X_{s_{i1}} - X_{d_{i1}} + X_{s_{i2}} - X_{d_{i2}}}{4} = \frac{\overline{X}_{s_i} - \overline{X}_{d_i}}{2} \tag{3.17}$$

The estimated error for the SBP would be:

$$\widehat{\epsilon}_{s_{ij}} = X_{s_{ij}} - \widehat{\mu} - \widehat{\delta}_i = X_{s_{ij}} - \overline{X}_{s_i} \tag{3.18}$$

and the variance is estimated by

$$\widehat{\sigma}^2_{s_i} = \frac{\sum_{j=1}^{2} \widehat{\epsilon}^2_{s_{ij}}}{2-1} = \frac{\sum_{j=1}^{2} (X_{s_{ij}} - \overline{X}_{s_i})^2}{2-1} \quad i = 1, 2, \ldots, n \tag{3.19}$$

For m pairs there would be

$$\widehat{\mu}_i = \frac{X_{s_{i1}} + X_{d_{i1}} + X_{s_{i2}} + X_{d_{i2}} + \ldots + X_{s_{im}} + X_{d_{im}}}{2m} = \frac{\overline{X}_{s_i} + \overline{X}_{d_i}}{2} \tag{3.20}$$

and

$$\widehat{\delta}_i = \frac{X_{s_{i1}} - X_{d_{i1}} + X_{s_{i2}} - X_{d_{i2}} + \ldots + X_{s_{im}} - X_{d_{im}}}{2m} = \frac{\overline{X}_{s_i} - \overline{X}_{d_i}}{2} \tag{3.21}$$

and hence the variance estimate for SBP measurements would be given by

$$\widehat{\sigma}^2_{s_i} = \frac{\sum_{j=1}^{m} \widehat{\epsilon}^2_{s_{ij}}}{m-1} = \frac{\sum_{j=1}^{m} (X_{s_{ij}} - \overline{X}_{s_i})^2}{m-1} \quad for \quad i = 1, 2, \ldots, n \tag{3.22}$$

A symmetrical argument is used for the estimation of error for DBP measurements:

$$\widehat{\epsilon}_{d_{ij}} = X_{d_{ij}} - \widehat{\mu}_i + \widehat{\delta}_i = X_{d_{ij}} - \overline{X}_{d_i} \tag{3.23}$$

leading to

$$\widehat{\sigma}^2_{d_i} = \frac{\sum_{j=1}^{m} \widehat{\epsilon}^2_{d_{ij}}}{m-1} = \frac{\sum_{j=1}^{m} (X_{d_{ij}} - \overline{X}_{d_i})^2}{m-1} \quad for\ i = 1, 2, \ldots, n \tag{3.24}$$

One pair of observed SBP/DBP will not be enough to estimate the BP error of measurement, because it is not possible to estimate variation with only one observation. Having obtained both $\widehat{\sigma}^2_{s_i}$ and $\widehat{\sigma}^2_{d_i}$ an obvious estimate of ρ from the data of subject i is:

$$\widehat{\rho} = \frac{\sum_{j=1}^{m} (X_{s_{ij}} - \overline{X}_{s_i})(X_{d_{ij}} - \overline{X}_{d_i})/(m-1)}{\widehat{\sigma}_{s_i}\widehat{\sigma}_{d_i}} \quad i = 1, 2, \ldots, n \tag{3.25}$$

Allow the mean BP to be given by

$$Y_{ij} = \frac{X_{s_{ij}} + X_{d_{ij}}}{2} = \frac{\mu_i + \delta_i + \epsilon_{s_{ij}} + \mu_i - \delta_i + \epsilon_{d_{ij}}}{2} = \mu_i + \epsilon^*_{ij}. \tag{3.26}$$

then Y_{ij} will be the mean of the *j-th* BP pair of patient *i*. Now

$E(Y_{ij}) = \mu_i$

and

$$var(Y_{ij}) = var(\mu_i + \epsilon^*_{ij}) = var(\epsilon^*_{ij}) \tag{3.27}$$

where $\epsilon^*_{ij} = \frac{\epsilon_{s_{ij}} + \epsilon_{d_{ij}}}{2}$, is random with a mean of zero and a variance σ^2_ϵ and μ_i is fixed. Giving

$$\sigma_\epsilon^2 = var(\frac{\epsilon_{s_{ij}} + \epsilon_{d_{ij}}}{2}) = \frac{1}{4}\{var(\epsilon_{s_{ij}}) + var(\epsilon_{d_{ij}}) + 2cov(\epsilon_{s_{ij}}, \epsilon_{d_{ij}})\} = \frac{\sigma_s^2}{4} + \frac{\sigma_d^2}{4} + \frac{\rho\sigma_s\sigma_d}{2} \tag{3.28}$$

which can be estimated by

$$\widehat{\sigma}_\epsilon^2 = \frac{\widehat{\sigma}_s^2}{4} + \frac{\widehat{\sigma}_d^2}{4} + \frac{\widehat{\rho}\widehat{\sigma}_s\widehat{\sigma}_d}{2} \tag{3.29}$$

or by

$$\widehat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \overline{Y}_{i.})^2}{n(m-1)} \tag{3.30}$$

It can be seen from equation 3.28 that the greater the covariance between the SBP and DBP is, the greater the measurement error is. For fixed σ_s^2 and σ_d^2 the measurement error would reduce if ρ is negative. All these parameters are estimated by SEM in chapter 5.

3.4.4 Random effects model

Assume the following simple model:

$$X_{ij} = \mu + a_i + \epsilon_{ij} \;\; a_i \sim (0, \sigma^2); \;\; \epsilon_{ij} \sim (0, \sigma_e^2) \tag{3.31}$$

Equivalently

$$X_{ij} = \mu_i + \epsilon_{ij} \text{ for } i = 1, 2 \ldots, n; j = 1, 2, \ldots, m \tag{3.32}$$

Let μ_i be random with a mean μ and a variance σ^2 and let ϵ_{ij} also be random with a mean of zero and a variance of σ_e^2. Assume also that μ_i and ϵ_{ij} are independent, then the above model is a random effects model and the expected values of the observed measurements are given by:

$$E(X_{ij}) = \mu \tag{3.33}$$

and their variances are:

$$Var(X_{ij}) = \sigma^2 + m\sigma_e^2 \tag{3.34}$$

Let the Within-sum of squares be denoted by WSS and the Between-sum of squares by BSS, then according to Searle [52]:

$$SSE = WSS = \sum_{i=1}^{n} \sum_{j=1}^{m} (X_{ij} - \overline{X}_{i.})^2 \tag{3.35}$$

and

$$BSS = m \sum_{i=1}^{n} (\overline{X}_{i.} - \overline{X}_{..})^2 \tag{3.36}$$

and the total sum of squares (SST) is

$$SST = SSE + BSS = \sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - \overline{X}_{..})^2 \tag{3.37}$$

Hence the variance components can be estimated as follow:

$$\widehat{\sigma}^2 = \frac{(MSB - MSE)}{n} \tag{3.38}$$

where

$$MSB = \frac{BSS}{n-1} \tag{3.39}$$

and

$$MSE = \frac{SSE}{n(m-1)} \tag{3.40}$$

and also

$$\widehat{\sigma}_e^2 = MSE \tag{3.41}$$

is an estimate of measurement error. Using the data in Table 3.1 the following sum of squares can be calculated:

For $n = 10$; $m = 3$

let $N = mn = 30$ then $\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}^2 = 204589$ and $\widehat{\mu} = \overline{X}_{..} = 81.967$

hence

$$SSE = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}^2 - m\sum_{i=1}^{n}\overline{X}_{i.}^2 = 193.980 \tag{3.42}$$

Therefore, the measurement error is estimated as:

$$\widehat{\sigma}_e^2 = MSE = \frac{193.98}{20} = 9.699 \tag{3.43}$$

and the total sum of squares as:

$$SST = 3032.967 \tag{3.44}$$

Hence

$$BSS = m\sum_{i=1}^{n} \overline{X}_{i.}^2 - N\overline{X}_{..}^2 = 2838.987 \tag{3.45}$$

leading to

$$MSB = \frac{2838.987}{9} = 315.443 \tag{3.46}$$

The between-subject variance component estimate is:

$$\widehat{\sigma}^2 = \frac{(315.443 - 9.699)}{3} = 101.91 \tag{3.47}$$

3.4.5 Mixed effects model

Assume the following model:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}\ i = 1, \ldots, n; j = 1, \ldots, m; k = 1, \ldots, l \tag{3.48}$$

where

μ is the general mean

α_i is the effect of level i of factor A

β_j is the effect of level j of factor B

γ_{ij} is the interaction effect of level i of factor A and level j of factor B

ϵ_{ijk} is the error term.

Let α_i be a fixed effect, β_j be a random effect with a mean of zero and a variance of σ^2_β, γ_{ij} be a random effect with a mean of zero and a variance of σ^2_γ, and ϵ_{ijk} be a random effect with a mean of zero and a variance of σ^2_e. Assume also that none of the random effects are correlated. Thus, the above model is a two-way mixed effects model with interaction.

The sum of squares are calculated as follow:

the total sum of squares (SST):

$$SST = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l} X^2_{ijk} \tag{3.49}$$

the mean sum of squares:

$$SSM = N\overline{X}^2_{...} \tag{3.50}$$

where $N = nml$

the sum of squares due to factor A (SSA):

$$SSA = ml \sum_{i=1}^{n} (\overline{X}_{i..} - \overline{X}_{...})^2 \tag{3.51}$$

the sum of squares due to factor B (SSB):

$$SSB = nl \sum_{j=1}^{m} (\overline{X}_{.j.} - \overline{X}_{...})^2 \tag{3.52}$$

the sum of squares due to the interaction ($SSAB$)

$$SSAB = l \sum_{i=1}^{n} \sum_{j=1}^{m} (\overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...})^2 \tag{3.53}$$

and the sum of squares due to error (SSE)

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{l} (X_{ijk} - \overline{X}_{ij.})^2 \tag{3.54}$$

The mean squares are then calculated by dividing the sum of squares by the corresponding degrees of freedom, for example:

$$MSB = \frac{SSB}{m-1} \tag{3.55}$$

$$MSAB = \frac{SSAB}{(n-1)(m-1)} \tag{3.56}$$

and

$$MSE = \frac{SSE}{nm(l-1)} \tag{3.57}$$

Hence the variance components are estimated as follow:

$$\widehat{\sigma}_e^2 = MSE \tag{3.58}$$

$$\widehat{\sigma}_\beta^2 = \frac{(MSB - MSAB)}{nl} \tag{3.59}$$

$$\widehat{\sigma}_\gamma^2 = \frac{(MSAB - MSE)}{nl} \tag{3.60}$$

3.4.6 Discussion

There are different ways of estimating these variance components, and maximum likelihood among others is used. As shown above a one-way fixed effects model and a one-way random effects model end up with the same ANOVA Table. Thus, the same estimate of measurement error is achieved. Yet a distinction must be noted in that, in a fixed effects model, factors have fixed levels consisting of a series of identifiable populations (or samples), each with its own mean (μ_i). The interest here lies in estimating the mean of each of those distinct populations (or samples) hence the *best linear unbiased estimator* (BLUE). On the other hand, in a random effects model setting, random factors have levels from a single population that are investigated. Interest here may lie in the variability within the population from which the sample came (variance component), or perhaps in the prediction of the mean of a particular level, hence the *best linear unbiased*

CHAPTER 4

STRUCTURAL EQUATION MODELLING

4.1 Introduction

Epidemiological methods have traditionally been seen as the best to follow when studying public health problems. Most epidemiological research has been focused on establishing the etiology of disease, but recently the view developed that diseases are the result of a complex mix of social, economic, political and environmental factors. In Chapter 1, for example, we mentioned a few factors that affect BP measurement, and those that lead to the condition of being hypertensive. As public health has broadened its focus on medical and behavioural problems to incorporate a more socio-environmental approach, some questions that public-health researchers are asking have become more complex. Social science offers a range of analytical methods that have evolved to deal with these complex questions asked by public-health researchers. One such tool is structural equation modelling (SEM). The general theory of SEM is partially covered in this chapter. It is impossible to cover the theory in detail, since SEM is a broad concept. Relevant topics will be discussed later when investigating measurement error in the taking of BP measurements.

4.2 History and development of SEM

The history of SEM can be traced back to the early work of Spearman (1904) on factor analysis, and Sewall Wright (1918) on path analysis. Goldberger (1972) modified the work of Wright and used his ideas in economics. People like Jöreskog (1973), Keesling (1972), Wiley (1973) and others built up from the existing foundations of path analysis and they developed it to the existing general structural equations. In the 1960s and early 1970s the conceptual synthesis of latent variable and measurement models were developed in sociology. The late 1970s and early 1980s saw a development of estimation procedures, which led to the development of computer software. The work of Jöreskog and Sörbom on LISREL (Linear Structural Relations) programs popularized SEM even more in social sciences. Bentley (1985) and others developed other computer software that performed SEM and were "user-friendly". According to Bollen [9] LISREL still remains the most popular software for peforming SEM.

4.3 What is structural equation modelling?

Structural equation modelling or SEM is a very general statistical modelling technique. Factor analysis, path analysis and multiple regression all represent special cases of SEM. Although SEM is similar to multiple regression and analysis of variance in some way, it differs from them markedly. The similarity lies in the ability to examine relationships, but regression and ANOVA can only examine

one relationship at a time, whereas SEM can examine complex relationships all at once. SEM is a confirmatory rather than an exploratory technique, that is, a researcher is more likely to use it to determine whether a certain model is valid or not, rather than using SEM to "find" a suitable model. The basic idea here is to test whether variables are interrelated through a set of linear relationships by examining their variances and covariances.

In SEM interest usually focuses on concepts or the so-called latent constructs (abstract, hard to measure psychological variables), for example "attitude toward a specific treatment", "socio-economic status of an individual", or "quality of life". Latent constructs vary in their degree of abstractness, the above mentioned examples fall in a category of highly abstract constructs. There are less abstract constructs such as income, age, BP (Bollen, [9]). No distinction will be made between the less and highly abstract latent variables in this thesis as theory reveals that they all can be treated in a similar manner. Latent variables or latent constructs are assumed here to be one and the same thing, hence throughout this thesis the term latent construct will be used. These can either be exogenous (independent) or endogenous (dependent), where the latter are caused or predicted by any other variables in the model. Variables called manifest or indicator variables and sometimes even proxies are associated with latent constructs, such variables are observed and are used to measure these latent constructs. The term indicator will be used throughout this thesis. Theory and practice reveal that there will be

measurement error involved in measuring these latent variables, and SEM takes this into account. Through indicator variables the concept of measurement error is then introduced into the model estimation.

SEM is made up of two types of models, namely structural and measurement models. The structural model relates latent variables only (endogenous constructs to exogenous constructs), whereas the measurement model accounts for the measurement of latent constructs through manifest indicators. A measurement model is a sub-model in SEM that:

1. *specifies the manifest indicators for each latent variable*

2. *assesses the reliability of each latent variable for use in causal relationships, and*

3. *measures the variance extracted by each of the latent constructs.*

Measurement error is not always caused by inaccurate responses, but also occurs when abstract concepts are used. SEM accounts for the measurement error in the measurement model. By explicitly modelling measurement error, SEM users seek to derive unbiased estimates for the relations between latent constructs. The measurement model is very similar to factor analysis and is often referred to as confirmatory factor analysis (CFA) in the literature, (Bollen, [9]). Measurement is recognized as difficult and error-prone.

The primary goal of factor analysis is to explain covariances or correlations

between indicators by means of relatively few constructs. Thus, factor analysis can be classified as a data reduction technique. There are two approaches to factor analysis, namely the exploratory and the confirmatory approach. Both confirmatory factor analysis (CFA) and exploratory factor analysis (EFA) are submodels in SEM where only the measurement model is applicable. Some differences between CFA and EFA are that in EFA:

- a detailed model, relating constructs to indicators, is not specified in advance,
- the number of constructs is not determined before the analysis,
- all constructs influence all indicators,
- measurement errors are assumed to be uncorrelated,
- underidentification is common,

On the other hand, in CFA:

- a model is constructed in advance,
- the number of constructs is set by the analyst,
- the analyst determines whether or not a given construct influences a particular indicator, and
- some parameters may be fixed or be estimated empirically, and measurement errors may be correlated.

In this work EFA will not be demonstrated, but a suitable data set for CFA will be analyzed in Chapter 5. General models for CFA are given and discussed in Section 4.5, and these appear to accommodate the theory of BP (as shown in Chapter 5). Because of the limitations mentioned above EFA reflects an inability to accommodate theoretical knowledge. On the other hand CFA overcomes shortcomings found in EFA, and thus, indicates being a more powerful tool in research.

4.4 Methodology of SEM

Hair *et al.* [27] recommend the following step-by-step approach to ensure that structural models are correctly specified and the results are valid.

These steps are:

1. *Develop a theory-based model, which explains how variables are inter-related.*

2. *Construct a corresponding path diagram of the causal relationships between all the variables.*

3. *Convert the path diagram into a set of structural equations and measurement equations.*

4. *Collect the data for analysis.*

5. *Estimate the proposed model.*

6. *Do SEM statistical model evaluation.*

7. *Modify the model if necessary and if it is theoretically justified return to step6.*

8. *Interpret the results.*

These steps will now be explained using the description given by Hair *et al.* [27].

4.4.1 Development of a theory-based model

Like most multivariate techniques, SEM is based on causal relationships, where a change in one variable results in a change of another. Causal relationships can take many forms and meanings, from the strict causation found in physical processes, such as a chemical reaction to less well defined relationships encountered in behavioural research. Justification of causality between variables does not lie in the analytical method used but in the theory. Thus, it is advised that SEM be employed when theory guides the analysis. A scientist may come up with the theory on which the model is based, but it is up to the analyst to specify the model and then test it with the data at hand. Theory may be based on empirical research from academic sources or be derived from practical experience. The most critical error in the development of a theory-based model is the omission of one or more key predictive variables, this is called specification error. This will cause a bias with respect to the importance of the other variables.

The following example (*The Stability of alienation*) will be used to illustrate

a theory-based model and some other SEM concepts. In the study by Wheaton et al. [63] conducted on 932 people in 1967 and 1971, attitudes such as alienation were studied in relation to background variables, such as education and occupation (Jöreskog,[31]).

The graphical display of the model is given in Figure 4.1. Alienation (the latent construct) is represented at two time points by indicator measurements anomia (lack of social standards) (*anomi67, anomi71*) and powerlessness (*powrl67, powrl71*). The latent constructs of alienation *(alnt67, alnt71)* are linked in the structural part of the model to the exogenous latent construct socio-economic status (*ses*) which is represented by education (*eductn*) and socio-economic index (*sei*). This example will be referred to as simply the "*stability*" when cited in the following sections. This example is used below in the construction of a path diagram and random error terms $(\varepsilon_1, \ldots, \varepsilon_4)$ or (δ_1, δ_2) are included to reveal imperfection of indicator variables in constructing latent variables.

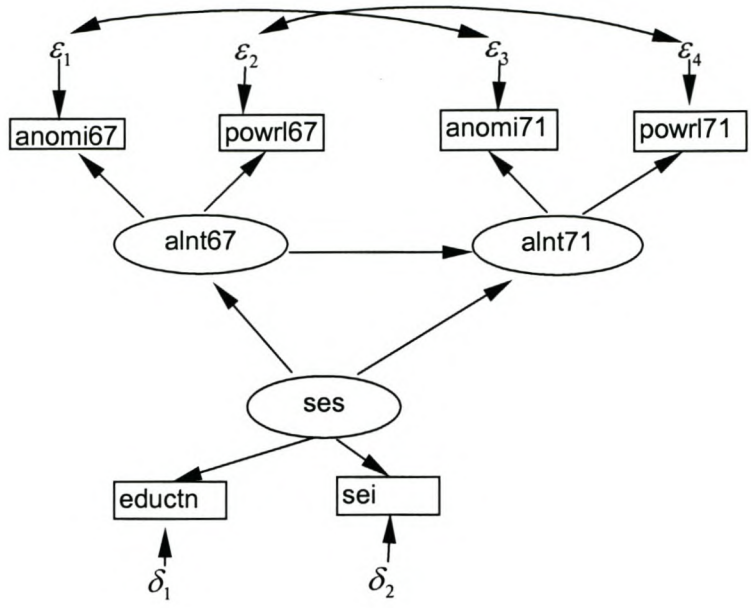

Figure 4.1: Relationships between variables in the Stability of Alienation example

In the following Section the way a path diagram is constructed from a theory based model will be discussed.

4.4.2 Constructing a path diagram

A series of causal relationships can be represented graphically by a method called path diagrams. These path diagrams are similar to flow charts that show variables interconnected with lines that indicate causal flow. When a path diagram is carefully drawn it will not only communicate the basic conceptual ideas about the given model effectively, but it will also represent the exact corresponding algebraic equations of the model and the assumptions regarding errors. The

following general conventions are used in path diagrams:

1. Observed manifest variables or indicators are enclosed in squares or rectangles. Latent variables are enclosed in circles or ellipses, in the *stability* example we have:

2. A one-way straight arrow between two variables indicates a direct influence.

3. A direct influence of one variable on another is indicated by a one-way arrow, so that an absence of an arrow means that there is no assumed direct relationship.

4. A curved two-way arrow indicates a correlation between the two variables.

The above-shown correlation arrow is sometimes referred to as a covariance path. These covariance paths are permitted only between exogenous variables. Covariance paths indicate an association or correlation between the two variables that are pointed. This association may be due to both variables depending on another third variable or the variables may have a causal relationship that still remains unspecified. An absence of a curved two-way arrow between two exogenous constructs means then that there is no assumed correlation between them. In SEM path diagrams are critically important as they provide means of arriving at algebraic equations. It is a common practise for SEM users to write error terms $(\varepsilon_1, \ldots, \varepsilon_4)$ or (δ_1, δ_2) without a circle or ellipse.

4.4.3 Converting a path diagram into a set of structural equations

Once the theoretical model has been developed and path diagrams have been drawn to illustrate the causal relationships, the next move would be to specify a series of equations which define:

- the structural equations linking latent constructs,
- the measurement model defining which manifest variables measure which latent variables,

and also

- a set of matrices indicating all hypothesized variances and covariances of the latent constructs.

This is illustrated in Section 4.4.3.5 by using the **stability** example. To simplify the models, all variables are written as deviations from their means.

4.4.3.1 The structural model

The structural model links latent constructs to one another. Here each endogenous construct (a construct with one or more straight arrow pointing to it) is the dependent variable in one equation. The exogenous constructs (construct with no straight arrow pointing to it) are independent variables. Each equation will contain at least one endogenous variable and one or more exogenous variables with an error term. The fundamental model where only variances (and not means) are estimated is presented below. Mathematically it can be shown as follows:

Suppose that there are m exogenous constructs and n endogenous constructs. The basic **structural** model is given by:

$$\begin{aligned} \eta &= B\eta + \Gamma\xi + \zeta \\ \therefore \quad & (I - B)\eta = \Gamma\xi + \zeta \end{aligned} \tag{4.1}$$

which is a model relating only latent constructs (endogenous to exogenous), where:

$B : n \times n$ is a matrix of coefficients of the endogenous latent variables. The main diagonal of B must be fixed at zero, otherwise η would have a direct effect on itself.

$\eta : n \times 1$ is a vector of endogenous latent variables

$\Gamma : n \times m$ is a matrix of coefficients of the exogenous latent variables

$\xi : m \times 1$ is a vector of exogenous latent variables

$\zeta : n \times 1$ is a vector of residuals

Φ : is the covariance matrix of ξ

Ψ : is the covariance matrix of ζ

The covariance matrix of η is a function of B, Γ, Φ and Ψ and does not have a special symbol. This is shown in Section 4.4.3.3.

4.4.3.2 The measurement model

The measurement model specifies how latent variables or hypothetical constructs are measured in terms of the observed manifest variables or indicators. The procedure is similar to factor analysis, but is much more powerful. Most indicators of constructs contain sizeable amounts of measurement error and the measurement model takes this measurement error into account. Ignoring measurement error leads to inconsistent estimators and inaccurate assessment of the

relationship between the underlying latent constructs. Suppose there are p exogenous construct indicators and q endogenous construct indicators. The basic equations for the measurement model are:

$$\begin{aligned} x &= \Lambda_x \xi + \delta \\ y &= \Lambda_y \eta + \varepsilon \end{aligned} \tag{4.2}$$

$x : p \times 1$ is a vector of indicators of the exogenous construct ξ

$y : q \times 1$ is a vector of indicators of the endogenous construct η

$\delta : p \times 1$ is a vector of measurement errors of x

$\varepsilon : q \times 1$ is a vector of measurement errors of y

$\Lambda_x : p \times m$ is a coefficient matrix relating x to ξ

$\Lambda_y : q \times n$ is a coefficient matrix relating y to η.

Λ_x and Λ_y are known as loading matrices and their elements (say λ_i) are known as the indicator loadings.

4.4.3.3 The hypothesized correlation/covariance matrices

The covariance matrices of errors are $\Theta_\delta = E(\delta\delta')$ and $\Theta_\varepsilon = E(\varepsilon\varepsilon')$. The main diagonals of these matrices are error variances associated with corresponding indicators. For example, Θ_δ is a $p \times p$ matrix with error variances of the x indicators

on the diagonal and the off-diagonal entries are their covariances. Naturally, when indicators are not correlated these covariance matrices become diagonal matrices.

A researcher specifies the correlations or covariances between exogenous latent variables or between endogenous latent constructs. It is common for the exogenous latent constructs to be correlated because of their sharing influence on the endogenous latent constructs. The following assumptions are required for solving the SEM equations in 4.1:

1. $E(\eta) = E(\xi) = E(\zeta) = E(\varepsilon) = E(\delta) = 0$,

2. ζ is uncorrelated with ξ,

3. ε is uncorrelated with η and ξ,

4. ζ, ε and δ are mutually uncorrelated, and

5. *I-B* is non-singular, where I:$n \times n$ is the identity matrix.

These assumptions are used in determining the implied covariance matrices for the exogenous x indicators and endogenous y indicators. From point *one* above and equations 4.2, it follows that the expected value of these indicators is zero. Thus indicators considered here are deviations from their means. Details of the mathematics involved in the derivations of these are provided by Bollen [9],[10] and Hayduk [29]. The $(n+m) \times (n+m)$ covariance matrix Σ of the stacked vector z where $z = \begin{bmatrix} y \\ x \end{bmatrix}$ can be represented as:

$$\widehat{\Sigma} = \begin{bmatrix} C(\Gamma\Phi\Gamma' + \Psi)C' + \Theta_\epsilon & C\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma C' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{bmatrix} \tag{4.3}$$

where $C = \Lambda_y(I - B)^{-1}$.

When the model is in terms of deviations from means, SEM consists of a selection of values for the elements of Λ_y, Λ_x,B, Γ, Φ, Ψ, Θ_δ and Θ_ϵ so that $\widehat{\Sigma}$ (sometimes called $\Sigma(\theta)$) matches the covariance matrix of the observed variables. In applications some of these elements are fixed or set equal to assigned values. This is particularly true for elements of Λ_y, Λ_x, B and Γ. This is done to ensure that there are fewer parameters to be estimated than unique moments of the observed indicators. Calculations are illustrated by the example in Section 4.4.3.5 below.

4.4.3.4 Standardized solutions

The Calis procedure of SAS reports residual covariances and standardized residuals which helps to locate the largest residuals and patterns that may appear in the residuals. The standardized residuals (sometimes called "Normalized residuals") are estimates of the number of standard deviations the observed residuals are away from the zero residuals that would be provided by a perfectly fitting model. Each one is calculated by dividing each residual covariance by the square root of its asymptotic variance (Hayduk [29]).

The Calis procedure also reports the standardized loadings. Standardized loadings are defined as the unstandardized loadings multiplied by the ratio of the

standard deviation of the construct (of which the loading is a coefficient) to the standard deviation of its indicator (Bollen [9], p38).

4.4.3.5 Example

In the stability example introduced in Section 4.4.1 from which the path diagram figure 4.1 is drawn, there is one exogenous construct (*ses*), two endogenous constructs (*alnt67* and *alnt71*) and manifest or predictor variables for all these latent variables (both exogenous and endogenous). The structural equations for the endogenous constructs or variables are as follows. From Section 4.4.3.1 and equation 4.1 the *stability* example indicates that

$m = 1$ and $n = 2$

$$\eta = \begin{bmatrix} alnt67 \\ alnt71 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix}, \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \text{ and } \xi = ses$$

thus

$$\begin{bmatrix} alnt67 \\ alnt71 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} alnt67 \\ alnt71 \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} ses + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

and written as equations in Table 4.1.

Equations of the measurement model represent relationships between exogenous indicators (*eductn* and *sei*) and the exogenous (*ses*) construct and are:

Endogenous construct		Exogenous variable	Error
alnt67	=	$\gamma_1 \cdot ses$	$+\zeta_1$
alnt71	=	$\beta_{21} \cdot alnt67 + \gamma_2 . ses$	$+\zeta_2$

Table 4.1: Structural Equations.

Exogenous indicators		Exogenous constructs	Error
eductn	=	$\lambda^x_{11} \cdot ses$	$+ \delta_1$
sei	=	$\lambda^x_{12} \cdot ses$	$+ \delta_2$

Table 4.2: Measurement model for exogenous latent variables

$$x = \begin{bmatrix} eductn \\ sei \end{bmatrix} ; \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \Lambda_x = \begin{bmatrix} \lambda^x_{11} \\ \lambda^x_{12} \end{bmatrix} \text{ and } \xi = ses$$

Substituting these in equations 4.2 leads to the equations in Table 4.2.

The relationship between endogenous indicators (*anomi67, anomi71, powrl67* and *powrl71*) and their latent constructs can be represented as follow:

$$y = \begin{bmatrix} anomi67 \\ powrl67 \\ anomi71 \\ powrl71 \end{bmatrix}, \Lambda_y = \begin{bmatrix} \lambda^y_{11} & 0 \\ \lambda^y_{21} & 0 \\ 0 & \lambda^y_{12} \\ 0 & \lambda^y_{22} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \text{ and } \eta = \begin{bmatrix} alnt67 \\ alnt71 \end{bmatrix}$$

and when substituted into the second sub-equation of equation 4.2 leads to the equations in Table 4.3: Covariance paths between the exogenous constructs ε_1 and ε_3 and between ε_2 and ε_4 are indicated by curved arrows as defined in

Endogenous indicators		Endogenous constructs		Error
anomi67	=	$\lambda^{y}_{11} \cdot alnt67$		$+ \varepsilon_1$
powrl67	=	$\lambda^{y}_{21} \cdot alnt67$		$+ \varepsilon_2$
anomi71	=		$\lambda^{y}_{12} \cdot alnt71$	$+ \varepsilon_3$
powrl71	=		$\lambda^{y}_{22} \cdot alnt71$	$+ \varepsilon_4$

Table 4.3: Measurement model for endogenous latent variables

Section 4.4.2.

Figure 4.2 is a representation of the *stability* example, including loadings and covariance paths.

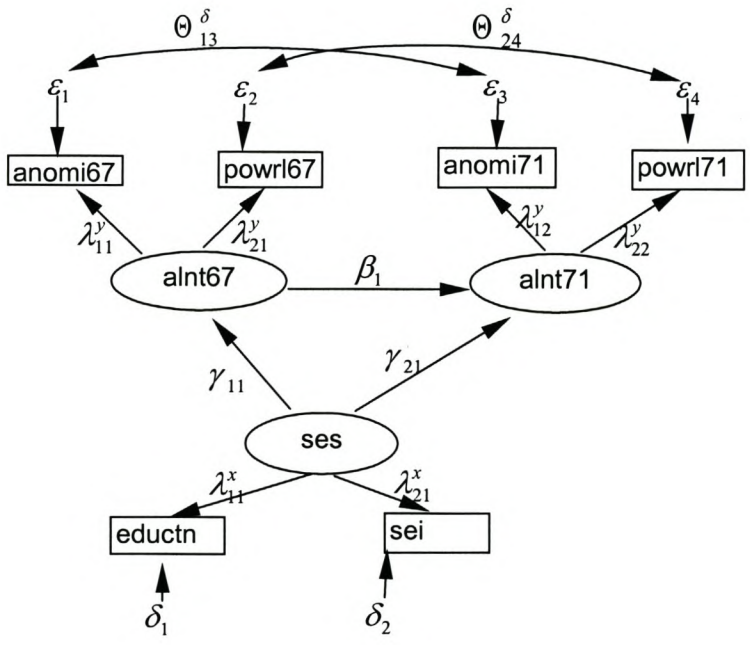

Figure 4.2: Graphical representation of the "stability" example.

4.4.4 Implied moment matrix

Once a model is specified, it implies that the second moments (variances and covariances) of the observed variables are functions of the model parameters. Most structural equation models focus on the implied covariance matrix, $\Sigma(\theta)$ defined in Section 4.4.3.3. For example, for x defined in equation 4.2 the implied covariance matrix is given by

$$\Sigma_x(\theta) = \Lambda_x \Phi \Lambda_x' + \Theta_\delta \tag{4.4}$$

where Λ_x are loading matrices of Section 4.4.3.2, Θ_δ is the diagonal matrix of measurement errors defined in Section 4.4.3.2 and Φ is the matrix of variances and covariances among constructs. $\Sigma_x(\theta)$ is the matrix at the lower right quadrant of $\Sigma(\theta)$. For illustration purpose consider the stability example as formulated in Section 4.4.3.5:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} eductn \\ sei \end{bmatrix}; \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \Lambda_x = \begin{bmatrix} \lambda_{11}^x \\ \lambda_{12}^x \end{bmatrix} \text{ and } \xi = ses$$

then

$$\phi = var(\xi), \Theta_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 \\ 0 & \theta_{\delta_2} \end{bmatrix}$$

From equation 4.4

$$\begin{aligned}\Sigma_x(\theta) &= \begin{bmatrix}\lambda_{11}^x \\ \lambda_{12}^x\end{bmatrix} var(\xi) \begin{bmatrix}\lambda_{11}^x & \lambda_{12}^x\end{bmatrix} + \begin{bmatrix}\theta_{\delta_1} & 0 \\ 0 & \theta_{\delta_2}\end{bmatrix} \\ &= \begin{bmatrix}(\lambda_{11}^x)^2 var(\xi) + \theta_{\delta_1} & \lambda_{12}^x \lambda_{11}^x var(\xi) \\ \lambda_{12}^x \lambda_{11}^x var(\xi) & (\lambda_{12}^x)^2 var(\xi) + \theta_{\delta_2}\end{bmatrix}\end{aligned} \tag{4.5}$$

resulting in the following set of equations:

$$\begin{aligned}var(x_1) &= (\lambda_{11}^x)^2 var(\xi) + \theta_{\delta_1} \\ var(x_2) &= (\lambda_{12}^x)^2 var(\xi) + \theta_{\delta_2} \\ cov(x_1, x_2) &= \lambda_{12}^x \lambda_{11}^x var(\xi)\end{aligned} \tag{4.6}$$

The implied matrix for y is determined as follow:

$$\begin{aligned}\Gamma\Phi\Gamma' + \Psi &= \begin{bmatrix}\gamma_1 \\ \gamma_2\end{bmatrix} \phi \begin{bmatrix}\gamma_1 & \gamma_2\end{bmatrix} + \begin{bmatrix}\psi_{11} & 0 \\ 0 & \psi_{22}\end{bmatrix} \\ &= \begin{bmatrix}\gamma_1^2\phi + \psi_{11} & \gamma_1\gamma_2\phi \\ \gamma_2\gamma_1\phi & \gamma_2^2\phi + \psi_{22}\end{bmatrix}\end{aligned} \tag{4.7}$$

$$C = \Lambda_y(I-B)^{-1} = \begin{bmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ 0 & \lambda_{12}^y \\ 0 & \lambda_{22}^y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta_{21} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ \lambda_{12}^y\beta_{21} & \lambda_{12}^y \\ \lambda_{22}^y\beta_{21} & \lambda_{22}^y \end{bmatrix} \tag{4.8}$$

$$C(\Gamma\Phi\Gamma'+\Psi)C' = \begin{bmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ \lambda_{12}^y\beta_{21} & \lambda_{12}^y \\ \lambda_{22}^y\beta_{21} & \lambda_{22}^y \end{bmatrix} \begin{bmatrix} \gamma_1^2\phi+\psi_{11} & \gamma_1\gamma_2\phi \\ \gamma_2\gamma_1\phi & \gamma_2^2\phi+\psi_{22} \end{bmatrix} \begin{bmatrix} \lambda_{11}^y & \lambda_{21}^y & \lambda_{12}^y\beta_{21} & \lambda_{22}^y\beta_{21} \\ 0 & 0 & \lambda_{12}^y & \lambda_{22}^y \end{bmatrix} \tag{4.9}$$

Hence:

$$\Sigma_y(\theta) = C(\Gamma\Phi\Gamma' + \Psi)C' + \Theta_\varepsilon = [Z]_{ij} \tag{4.10}$$

where $[Z]_{ij}$ is a 4x4 symmetric matrix with the following lower half entries:

$z_{11} = (\lambda_{11}^y)^2 var(\eta_1) + \theta_{\varepsilon_{11}}$

$z_{12} = \lambda_{11}^y\lambda_{21}^y var(\eta_1)$

$z_{22} = (\lambda_{21}^y)^2 var(\eta_1) + \theta_{\varepsilon_{22}}$

$$z_{13} = \lambda_{11}^{y}\lambda_{21}^{y}\{\beta_{21}var(\eta_1) + \gamma_1\gamma_2\phi\} + \theta_{\varepsilon_{13}}$$

$$z_{23} = \lambda_{12}^{y}\lambda_{21}^{y}\{\beta_{21}var(\eta_1) + \gamma_1\gamma_2\phi\}$$

$$z_{33} = \lambda_{12}^{y}\beta_{21}\{\lambda_{12}^{y}\beta_{21}var(\eta_1)+\lambda_{12}^{y}\gamma_1\gamma_2\phi\}+\lambda_{12}^{y}\{\lambda_{12}^{y}\beta_{21}\gamma_1\gamma_2\phi+\lambda_{22}^{y}var(\eta_2)\}+\theta_{\varepsilon_{33}}$$

$$z_{14} = \lambda_{11}^{y}\lambda_{22}^{y}\{\beta_{21}var(\eta_1) + \gamma_1\gamma_2\phi\}$$

$$z_{24} = \lambda_{21}^{y}\lambda_{22}^{y}\{\beta_{21}var(\eta_1) + \gamma_1\gamma_2\phi\} + \theta_{\varepsilon_{24}}$$

$$z_{34} = \lambda_{22}^{y}\beta_{21}\{\lambda_{12}^{y}\beta_{21}var(\eta_1) + \lambda_{12}^{y}\gamma_1\gamma_2\phi\} + \lambda_{22}^{y}\{\lambda_{12}^{y}\beta_{21}\gamma_1\gamma_2\phi + \lambda_{22}^{y}var(\eta_2)\}$$

$$z_{44} = \lambda_{22}^{y}\beta_{21}\{\lambda_{22}^{y}\beta_{21}var(\eta_1)+\lambda_{22}^{y}\gamma_1\gamma_2\phi\}+\lambda_{22}^{y}\{\lambda_{22}^{y}\beta_{21}\gamma_1\gamma_2\phi+\lambda_{22}^{y}var(\eta_2)\}+\theta_{\varepsilon_{44}}$$

where

$$\begin{aligned} var(\eta_1) &= \gamma_1^2\phi + \psi_{11} \\ var(\eta_2) &= \gamma_2^2\phi + \psi_{22} \end{aligned} \tag{4.11}$$

The set of equations for the variances and covariances of the corresponding observed indicators then follow as:

$$
\begin{aligned}
var(y_1) &= (\lambda_{11}^y)^2 var(\eta_1) + \theta_{\varepsilon_{11}} \\
cov(y_1, y_2) &= \lambda_{11}^y \lambda_{21}^y var(\eta_1) \\
var(y_2) &= (\lambda_{21}^y)^2 var(\eta_1) + \theta_{\varepsilon_{22}} \\
cov(y_1, y_3) &= \lambda_{11}^y \lambda_{21}^y \{\beta_{21} var(\eta_1) + \gamma_1\gamma_2\phi\} + \theta_{\varepsilon_{13}} \\
cov(y_2, y_3) &= \lambda_{12}^y \lambda_{21}^y \{\beta_{21} var(\eta_1) + \gamma_1\gamma_2\phi\} \\
var(y_3) &= \lambda_{12}^y \beta_{21} \{\lambda_{12}^y \beta_{21} var(\eta_1) + \lambda_{12}^y \gamma_1\gamma_2\phi\} + \lambda_{12}^y \{\lambda_{12}^y \beta_{21}\gamma_1\gamma_2\phi + \lambda_{22}^y var(\eta_2)\} \\
&\quad + \theta_{\varepsilon_{33}} \\
cov(y_1, y_4) &= \lambda_{11}^y \lambda_{22}^y \{\beta_{21} var(\eta_1) + \gamma_1\gamma_2\phi\} \\
cov(y_2, y_4) &= \lambda_{21}^y \lambda_{22}^y \{\beta_{21} var(\eta_1) + \gamma_1\gamma_2\phi\} + \theta_{\varepsilon_{24}} \\
cov(y_3, y_4) &= \lambda_{22}^y \beta_{21} \{\lambda_{12}^y \beta_{21} var(\eta_1) + \lambda_{12}^y \gamma_1\gamma_2\phi\} + \lambda_{22}^y \{\lambda_{12}^y \beta_{21}\gamma_1\gamma_2\phi + \lambda_{22}^y var(\eta_2)\} \\
var(y_4) &= \lambda_{22}^y \beta_{21} \{\lambda_{22}^y \beta_{21} var(\eta_1) + \lambda_{22}^y \gamma_1\gamma_2\phi\} + \lambda_{22}^y \{\lambda_{22}^y \beta_{21}\gamma_1\gamma_2\phi \\
&\quad + \lambda_{22}^y var(\eta_2)\} + \theta_{\varepsilon_{44}}
\end{aligned} \tag{4.12}
$$

The implied covariance matrix for y and x, namely $\Sigma_{xy}(\theta)$ is determined by

$$
\begin{aligned}
\Sigma_{xy}(\theta) &= C\Gamma\Phi\Lambda_x' \\
&= \begin{bmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ \lambda_{12}^y\beta_{21} & \lambda_{12}^y \\ \lambda_{22}^y\beta_{21} & \lambda_{22}^y \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} var(\xi) \begin{bmatrix} \lambda_{11}^x & \lambda_{12}^x \end{bmatrix} \\
&= \begin{bmatrix} \lambda_{11}^y\lambda_{11}^x\gamma_1\phi & \lambda_{12}^x\lambda_{11}^y\gamma_1\phi \\ \lambda_{11}^x\lambda_{21}^y\gamma_1\phi & \lambda_{12}^x\lambda_{21}^y\gamma_1\phi \\ \lambda_{12}^y\lambda_{11}^x\beta_{21}\gamma_1\phi + \lambda_{12}^y\gamma_2\lambda_{11}^x\phi & \lambda_{12}^y\lambda_{12}^x\beta_{21}\gamma_1\phi + \lambda_{12}^x\lambda_{12}^y\gamma_2\phi \\ \lambda_{22}^y\lambda_{11}^x\beta_{21}\gamma_1\phi + \lambda_{22}^y\lambda_{11}^x\gamma_2\phi & \lambda_{22}^y\lambda_{12}^x\beta_{21}\gamma_1\phi + \lambda_{22}^y\lambda_{12}^x\gamma_2\phi \end{bmatrix} \qquad (4.13)
\end{aligned}
$$

Then the final set of equations follows as:

$$
\begin{aligned}
cov(x_1, y_1) &= \lambda_{11}^y\lambda_{11}^x\gamma_1\phi \\
cov(x_2, y_1) &= \lambda_{12}^x\lambda_{11}^y\gamma_1\phi \\
cov(x_1, y_2) &= \lambda_{11}^x\lambda_{21}^y\gamma_1\phi \\
cov(x_2, y_2) &= \lambda_{12}^x\lambda_{21}^y\gamma_1\phi \\
cov(x_1, y_3) &= \lambda_{12}^y\lambda_{11}^x\beta_{21}\gamma_1\phi + \lambda_{12}^y\gamma_2\lambda_{11}^x\phi \\
cov(x_2, y_3) &= \lambda_{12}^y\lambda_{12}^x\beta_{21}\gamma_1\phi + \lambda_{12}^x\lambda_{12}^y\gamma_2\phi \\
cov(x_1, y_4) &= \lambda_{22}^y\lambda_{11}^x\beta_{21}\gamma_1\phi + \lambda_{22}^y\lambda_{11}^x\gamma_2\phi \\
cov(x_2, y_4) &= \lambda_{22}^y\lambda_{12}^x\beta_{21}\gamma_1\phi + \lambda_{22}^y\lambda_{12}^x\gamma_2\phi \qquad (4.14)
\end{aligned}
$$

These SEM equations (equations 4.6, 4.12 and 4.14) are then solved simultaneously. Furthermore, these relations are critical to the issues of the identification, estimation and fit assessment as they will be discussed below.

4.4.5 Estimation

As shown in the example above, parameter estimation involves solving the equations: $\Sigma = \Sigma(\theta)$ where the matrix on the left is a population matrix and the one on the right is the matrix of free model parameters. In this thesis the CALIS procedure of SAS is used for parameter estimation. There are several estimation methods in SEM and estimation is done under the following assumptions:

1. *independence of observations,*

2. *random sampling of the respondents (indicators),*

3. *linearity of all relationships, and*

4. *multivariate normality.*

The choice of estimation technique is often determined by the distributional properties of the variables being analyzed. The following estimation methods are generally used [9], [27], [28].

4.4.5.1 Maximum Likelihood (ML)

This is the most widely used estimation method in SEM. The fitting function is:

$$F_{ML} = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - (p+q) \tag{4.15}$$

where S is the sample covariance matrix, $(p+q) = tr(I)$ and Σ is the covariance matrix of Section 4.4.3.3.

The covariance matrices used in the ML function (F_{ML}) must be nonsingular to avoid the undefined logarithm of zero. When a model fits perfectly the values of the sample covariance give $F_{ML} = 0$. To verify that $F_{ML} = 0$ when $\widehat{\Sigma} = S$, we substitute $\widehat{\Sigma}$ for $\Sigma(\theta)$ in the above equation.

4.4.5.2 Unweighted least squares (ULS)

The unweighted least squares (F_{ULS}) function minimizes the deviations between the observed elements of the sample covariance matrix and the corresponding elements of the predicted covariance matrix.

4.4.5.3 Generalized least squares (GLS)

Generalized least squares weights observations to correct for unequal variances or nonzero covariances of the disturbances(errors). The function of F_{ULS} is a special case of the generalized least squares' function (F_{GLS}) where all deviations

between the indicators and constructs are given the same weights as if they had the same variance.

4.4.5.4 Two-stage least squares (TSLS),

In the first stage the independent variables are regressed on the instrumental variables. In the second stage the dependent variable is regressed on the predicted values of the independent variables (from the first stage).

4.4.5.5 Generally weighted least squares (WLS),

When non-normality or excessive kurtosis threatens the validity of the ML or GLS significance tests, this method provides an alternative as it accommodates these problems. It makes minimal assumptions about the distribution of the observations.

The last two methods above are popular in econometric procedures. In each case the parameter vector is estimated iteratively by an optimization algorithm.

4.4.6 Sample size and the data format

According to Hair *et al.* [27], sample size is important in the estimation and interpretation of the SEM results:

- A minimum sample size of a hundred is required for the use of ML.
- Very large samples of more than five hundred bring a risk of getting all

goodness-of-fit measures to indicate poor fit due to over-sensitivity in detecting differences in the data.

- The sample size provides a basis for estimating sampling error, which is critical in ML estimation.
- The sample should be large enough to include five observations for each estimated parameter.

4.4.7 Choosing the data for analysis

Input data for the CALIS procedure in SAS can be a covariance matrix, correlation matrix, or the raw data. Inputting a covariance or a correlation matrix has an advantage of using less computer time but will not allow computing of the measures of kurtosis or identification of outliers. When the data has variances of some of the variables much larger than those of others the correlation matrix would be preferred. A correlation matrix is permissible when patterns of relationships among constructs are investigated [27]. Raw data is essential when intercept terms are to be estimated.

4.4.8 Statistical evaluation of a SEM model

A SEM model is specified in accordance to the causality theory which is briefly discussed in Section 4.4.1. Initially hypothesized relationships between the constructs and the indicators are specified. Care must be taken in making sure that

the model is identified (identification is discussed in the next Section). Once the model has been correctly specified, estimation is done by a computer program. The process of estimation is iterative, with the initial part of it concerned solely with working the model fit before the estimation of model parameters and the interpretation of the results. Initial procedures (identification, goodness-of-fit and modification indices) that are of concern in SEM evaluation will be discussed before parameter estimation and the interpretation of the results.

4.4.8.1 Identification

Model identification concerns the question of whether it is possible to determine the parameters of a model from means, variances and covariances of the observed variables uniquely. A model is said to be identified or exactly identified if it has as many linearly independent equations as unknowns. If it has more equations than unknowns it is said to be overidentified. An identified model will produce a unique set of parameter estimates (for a given sample). A model is identified if it is either exactly identified or overidentified. Alternatively an underidentified model has fewer linearly independent equations than unknowns and will not give unique parameter estimates. Clearly, parameter estimates from an underidentified model are useless. A problem of identification in SEM is associated with incorrect modelling of the data.

A necessary but not sufficient condition for the identification of a model is

the assignment of a scale to each construct. One way of doing this is to choose an indicator with a loading equal to one for each construct. The intercept for that same indicator should be set to zero (see Section 4.5). With this scaling, the construct has a metric or scale that is similar to that of the indicator. In the *stability* example of Section 4.4.3.5 for instance, λ_{11}^{y}, λ_{12}^{y} and λ_{11}^{x} could be set equal to one in order to assign scales to η_1, η_2 and ξ respectively.

4.4.8.2 Assessing the identification of the structural model

The output of a computer program can be used to assess whether there is an identification problem. Indications are:

1. *very large standard errors for one or more coefficients,*
2. *the inability of the program to invert the information matrix,*
3. *unreasonable estimates, such as negative error variances,*
4. *high correlations (approx. 0.90) among the estimated coefficients.*

If an identification problem is indicated, then one can check the problem at the following three possible sources:

1. *a large number of estimated coefficients relative to the number of covariances or correlations, indicated by a small number of degrees of freedom (similar to the problem of over-fitting the data in multivariate techniques),*

2. *the use of reciprocal effects (two-way causal arrows between two constructs), and*

3. *failure to fix the scale of a construct.*

A solution is to define more constraints (or fix some parameters) in the model to provide more constraints relative to the number of causal relationships examined. When the model is identifiable then one can examine the fit of the model.

4.4.8.3 Goodness of fit

Goodness of fit is assessed at several levels: first for the overall model and then for the measurement and structural models separately. Once the assumptions (of sections 4.4.3.1, 4.4.3.2, 4.4.3.3) have been checked, the results are then examined for estimated coefficients that exceed acceptable limits, called offending estimates. Common examples of offending estimates are:

1. *negative error variance or non-significant error variances for any construct,*

2. *standardized coefficients exceeding or very close to 1.0, or*

3. *very large standard errors for one or more of the estimates.*

4.4.8.4 Overall model fit

If all assumptions are met and there are no offending estimates one needs to assess one or more of the following goodness of fit measures:

1. *absolute fit,*

2. *incremental fit measure,*

3. *parsimonious fit measures.*

Absolute fit: This determines the degree to which the overall model predicts the observed covariance or correlation matrix. The three absolute measures of fit are:

1. *Chi-squared statistics*

   This is the only goodness of fit measure in SEM with distributional properties. If there is a large chi-square relative to the degrees of freedom then the observed and estimated matrices differ to a large degree. On the other hand low chi-squared values which result in significant levels greater than 0.05 indicate that the difference between the actual and predicted is not significant.

2. *Goodness of fit index (GFI)*

   This measures ranges in values from 0 (poor fit) to 1 (perfect fit). The data do not fit the model if the GFI is negative or much greater than one.

3. *Root mean square residual (RMSR)*

   This is the root of the mean square residual. If covariances are used it is the average residual covariance, while if a correlation matrix is used, it is then written in terms of an average residual correlation.

Incremental fit measures: These measures compare the proposed model to a comparison model, often referred to as the null model [29]. A null model is the simplest model that can be theoretically justified. The most common example of a null model is a model with a single construct related to all indicators with no measurement error. The following are two of the incremental fit measures:

1. *Normed fit index (NFI)*

   The normed fit index Δ_1, proposed by Bentler and Bonett (1980) is

$$\Delta_1 = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2} = \frac{F_b - F_m}{F_b} \tag{4.16}$$

   where χ_b^2 is the chi-squared estimate for the null model and χ_m^2 being the chi-squared estimate for the proposed model. The best possible fit is when F_m is zero, which results in a Δ_1 of one. The worst fit is when the proposed model is not different to the null model, which results in a Δ_1 of zero.

2. *Non-normed index Delta 2*

   Bollen (1988) proposed a modification of Δ_1 (above) that lessens the dependence of its mean on the sample size (N) and takes into account the degrees of freedom corresponding to the proposed model (df_m).

$$\Delta_2 = \frac{\chi_b^2 - \chi_m^2}{\chi_b^2 - df_m} = \frac{F_b - F_m}{F_b - df_m/(N-1)} \tag{4.17}$$

The recommended value for each of the incremental fit measures is 0.9 or greater.

Parsimonious fit measures: This is used to check if the model fit has been achieved by "overfitting" the data with too many coefficients. The following measures of parsimonious fit are among the used ones:

1. *Adjusted goodness-of-fit Index (AGFI).*

   AGFI is an extension of the GFI defined previously. AGFI is adjusted by the ratio of the degrees of freedom for the proposed model to the degrees of freedom for the null model. The recommended level of acceptance is 0.90 or greater.

2. *Akaike's information criterion (AIC).*

This is a criterion for selecting the best model among a number of candidate models. The final measure of AIC is calculated as follows

$$AIC = \chi^2 - 2df. \tag{4.18}$$

AIC value closest to zero indicate the best fit [9], [51].

4.4.8.5 Measurement Model Fit

Once the overall model fit is evaluated, the measurement of each latent variable can be assessed by:

- examining the indicator loadings (coefficients of constructs which are elements of Λ_x and Λ_y) for statistical significance, and

- assessing the construct's reliability and variance extracted.

The construct's reliability is the measure of internal consistency of the indicators, depicting the degree to which they "indicate" a common construct. It is given by:

$$\text{Construct Reliability}=\frac{\left[\sum(\text{standardized loading } j)\right]^2}{\left[\sum(\text{standardized loading } j)\right]^2+\sum var(\epsilon_j)} \tag{4.19}$$

where the standardized loadings are defined as in Section 4.4.3.4. The sum is over all indicators j loading on the construct and ϵ_j is the measurement error variance of indicator j. A commonly accepted value for reliability is 0.7.

Another measure of reliability is the variance extracted measure. This measure gives the overall amount of variance in the indicators accounted for by a given construct. Higher values of variance extracted occur when the indicators are truly representative of the construct. This measure is given by:

$$\text{Variance extracted} = \frac{\sum(\text{standardized loading } j)^2}{\sum(\text{standardized loading } j)^2+\sum var(\epsilon_j)} \tag{4.20}$$

Guidelines suggest that acceptable variance extracted values should exceed 0.5 (Hair *et al.*, [27]).

4.4.8.6 Structural Model fit

When a structural model is estimated, tests of significance are done on the estimated coefficients. Structural equation modelling provides estimated coefficients with standard errors and t values for each coefficient. Overall coefficient of determination R^2 is also calculated for each endogenous equation. This gives the measure of fit for the entire structural equation, and gives an indication of the amount of variation or correlation of the endogenous variable accounted for by the exogenous variable.

4.4.9 Modification of the model

Modelling data by SEM can never be "exact", but the estimated model will be close to the true model. Hence if the model indicates an adequate fit, one needs to examine possible modifications that will improve both theoretical explanations and the goodness of fit of the model. The necessity for model improvement will be indicated by residuals of the predicted covariance or correlation matrix. Possible modifications to the proposed model will be indicated through the examination of the normalized residuals and the so-called modification indices. Normalized residuals (sometimes known as standardized residuals, of Section 4.4.3.4, [29]) greater than ± 2.0 can be regarded as statistically significant at the 0.05 level. When the normalized residuals are greater than 2 then the proposed model is far off from the predicted one, hence changes must be made in the model. Modification indices

suggest what changes should be.

There are three equivalent modification indices that are usually used, namely *likelihood ratio (LR) test, Lagrangian Multiplier (LM) test and Wald (W) test.* The frequently used ones are the *LM and Wald tests.* The *LM* test estimates the reduction in the model chi-square that would result from freeing a fixed parameter and allowing it to be estimated. However, only exogenous construct are allowed to have an association (correlation), as discussed in section 4.4.2. The *Wald test* on the other hand can be thought of as acting opposite the *LM* test since it identifies paths and covariances that should possibly be deleted from the model. It is possible to have a computer output that does not contain a *Wald test*, and this means that it is probably not possible to drop any of the parameters without significantly hurting the model's fit. If no further modifications are necessary then the estimation of the parameters can be taken as done.

4.5 A model with means and intercepts

The SEM models discussed so far, assume that model indicators and constructs have zero means. The analysis has been restricted to covariance matrices and their deviations from their means. Hence, the use of covariance or correlation matrices for the input data. This assumption restricts the ability of these models in answering certain research questions, for example, means of constructs cannot be estimated. Attempts to model means using SEM, for example, Coles

and Maxwell (1985) and Faulbaum (1987); McArdle and Epstein (1987) have not been widespread despite several discussions of this topic by Jöreskog and Sörbom (1979-1982). Other factors contributing to the lack of use of models with intercepts are the late appearances of the procedures for handling means. Another fundamental reason for the avoidance of means is the fact that it requires changes in the basic model and the input data. In fact, the use of means does not allow the use of covariance or correlation matrices for input data, but the raw data.

In SEM intercepts are required in order to derive the mean (expected values) of dependent variables. A model with means and intercepts is an extension of models represented by equations 4.1 and 4.2. When vectors of intercepts are added to those equations they become:

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta \tag{4.21}$$

and:

$$\begin{aligned} x &= \nu_x + \Lambda_x\xi + \delta \\ y &= \nu_y + \Lambda_y\eta + \varepsilon \end{aligned} \tag{4.22}$$

where α is a $n \times 1$ vector of intercepts for the equation of the endogenous construct, ν_x and ν_y are $p \times 1$ and $q \times 1$ vectors of intercepts for equations of measurement for x and y respectively. Model assumptions are the same as those in Section 4.4.3

except that means of the exogenous construct ξ is taken to be an $n \times 1$ column vector κ, hence, the means of endogenous construct η are given by:

$$\begin{aligned} E(\eta) &= (I - B)^{-1} E(\alpha + \Gamma\xi + \zeta) \\ &= (I - B)^{-1}[\alpha + \Gamma E(\xi)] \\ &= (I - B)^{-1}(\alpha + \Gamma\kappa) \end{aligned} \tag{4.23}$$

since $E(\zeta) = 0$. The expectations of x and y are:

$$\begin{aligned} E(x) &= \nu_x + \Lambda_x \kappa \\ E(y) &= \nu_y + \Lambda_y (I - B)^{-1}(\alpha + \Gamma\kappa) \end{aligned} \tag{4.24}$$

The identification becomes even more complicated because of additional parameters. It will be shown how this type of a model works in the main BP example. Table 4.4, which is borrowed from Bollen [10] shows how common statistical models can be derived from the above model (SEM with means and intercepts). For example, if we assume a scalar, continuous dependent response variable (y), no measurement error in the dependent or explanatory variables, and only dummy explanatory variables, we end up with an ANOVA model. The vectors τ_y and τ_x are vectors of threshold parameters which determine the values taken by y and x respectively.

Statistical Model	ν_y	Λ_y	Θ_ε	ν_x	Λ_x	Θ_δ	τ_y	x	τ_x	α	B	Γ	Ψ	y
ANOVA	0	I	0	0	I	0	-	D	-	S	0	*	S	y^*, scalar
Multiple regression	0	I	0	0	I	0	-	D/C	-	S	0	*	S	y^*, scalar
Probit regression	0	I	0	0	I	0	k-1	D/C	-	S	0	*	S	$1, 2, \cdots, k$
Tobit regression	0	I	0	0	I	0	0	D/C	-	S	0	*	S	$=0$, if $y^* \leq 0$
														$=y^*$ if $y^* > 0$
Classical Econometrics	0	I	0	0	I	0	-	D/C	-	*	*	*	*	y^*
Classical factor analysis	-	-	-	-	*	diag	C	C	-	-	-	-	-	-
(deviation scores)														
CFA	-	-	-	-	*			C	-	-	-	-	-	-

Table 4.4: Common statistical models as special cases of SEM

In Table 4.4 the notation is as follows: I - identity matrix, *diag* - diagonal, y^*-a scalar, D- dummy, C-continuous, S-scalar and *-present.

CHAPTER 5

INVESTIGATING MEASUREMENT ERROR IN SPECIFIC STUDY DESIGNS

5.1 Introduction

The accuracy of BP measurements has always been improved by averaging several measurements [25], but recent medical scientific methods improve such accuracy even more by the introduction of the so-called weighted averages [4]. They also make use of the quality of measurement index, which is the ratio of the variance of the true values (constructs) over the total variance of an indicator. Hence, both the systematic and random errors contribute towards this quality of measurement. The lower their contribution is to the total variance of an indicator, the higher the quality of the measurement is. In this chapter the relationship between the ANOVA and SEM will be investigated. SEM will also be demonstrated as an efficient measurement tool in BP studies. Batista *et al.* [4] modelled BP via the SEM using multi-type multi-trait (MTMT) models, which are a particular case of mean-and-covariance structure confirmatory factor analysis models. These constitute a variant of repeated measurement designs based on Campbell and Fiske's (1959) suggestion that the quality of the measurement instrument

(time, in our BP examples) can be determined by comparing them with other instruments in order to reveal both systematic and random errors. MTMT data allow for the evaluation of the quality of an indicator through the isolation of both the systematic error variance (because of an instrument used) and random error variance from the total variance [4]. The SEM approach outlined in this chapter follows the MTMT structure very closely but with measurement instruments (traits) replaced by observation times, hence the multi-type multi-time (MTMT) models. The BP studies introduced in Chapter 2 will be analyzed by SEM in this chapter. Using these studies specific models for BP measurement which enable estimation of measurement error will be formulated.

5.2 Modelling the data

5.2.1 ANOVA model for the BP measurements

In the *Mamre study* measurements were taken at five-minute intervals during a single visit using a single machine. On the other hand in the *Mitchells Plain study* two machines were used and readings were taken at two-minute intervals. Measurement error will be taken as within-visit variability [4] in both these cases, since measurements are taken at a single visit. The BP measurements will be taken as repeated because of their consistent pattern over time. The underlying SBP will be called systolic blood pressure construct *(SBP)*, the underlying DBP is called diastolic blood pressure construct *(DBP)*, and patients are taken to be

subjects. The BP type is taken as fixed since the entire population of types (DBP and SBP) are included. Time effects are also taken as fixed since they do not constitute a sample of all possible times. Let

α_i be the main effect of type i of BP, taken to be fixed

β_j be the main effect of time j, taken as fixed

γ_k be the main effect of the subject k, taken to be random,

then the following mixed effects model is obtained:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \omega_{ij} + \omega_{ik} + \omega_{jk} + \varepsilon_{ijk} \text{ for } i = 1, 2;\ j = 1, 2, 3 \quad (5.1)$$

where

x_{ijk} is the reading for type i at time j for the subject k,

μ is the overall mean, and is fixed,

ω_{ij} is a term representing the type-time interactions, and is fixed,

γ_k is a random variable to accommodate variation between subjects. This term is independent of ω_{ik}, ω_{jk} and ε_{ijk} with mean of zero and variance $\sigma^2(\gamma_k)$,

ω_{ik} is a random variable to accommodate variation between types within subjects. This term is independent of γ_k, ω_{jk} and ε_{ijk} with a mean of zero and a common variance $\sigma^2(\omega_{ik})$,

ω_{jk} is a random variable to accommodate variation between times within subjects. This term is independent of γ_k, ω_{ik} and ε_{ijk} with a mean of zero and a common variance $\sigma^2(\omega_{jk})$, and

ε_{ijk} is a simple error term with a mean of zero and a common variance $\sigma^2(\varepsilon_{ijk})$.

Then the total pooled variance across subjects and types will be given by:

$$\sigma^2(x_{ijk}) = \pi^2(\alpha_i) + \pi^2(\beta_j) + \sigma^2(\gamma_k) + \pi^2(\omega_{ij}) + \sigma^2(\omega_{ik}) + \sigma^2(\omega_{jk}) + \sigma^2(\varepsilon_{ijk}) \quad (5.2)$$

where π^2 represents the variance of fixed factors, which are not variance components. Consider the six readings (three pairs of DBP and SBP) taken on one subject, if one reading is randomly selected from each subject in the study, there will be a variation according to the type selected which is given by $\pi^2(\alpha_i)$. On the other hand, random effects are represented by σ^2, which are variance components. The variance of the *type* main effects ($\pi^2(\alpha_i)$) has no interpretation with respect to the quality of measurement. The $\pi^2(\beta_j)$ estimates the extent of bias of different time measures to each other, and $\pi^2(\omega_{ij})$ estimates the extent of bias of the different time measures due to *types*. The $\sigma^2(\gamma_k)$ is the contribution of the main effects of subject k to the overall variation, the $\sigma^2(\omega_{ik})$ is the contribution of the type-subject interaction on the variance of the true BP (both systolic and diastolic). The contribution of the main effect of the subject k and the time j is

given by $\sigma^2(\omega_{jk})$, which is mainly the sum of systematic measurement error. The $\sigma^2(\varepsilon_{ijk})$ is a third order interaction (*type* by *time* by *subjects*) variance which is an overall measure of the random error variance for all six of the measurements. All the variance components can be estimated by the ANOVA as was shown in Section 3.4.5.

5.2.2 Random effects model

Make the following substitution in the above mixed effects model: let

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \omega_{ij} \tag{5.3}$$

where μ, α_i, β_j and ω_{ij} are fixed. The μ_{ij} will also be fixed and represents the expectation of measurements of type i at time j. From equation 5.1 follows that a reading for each subject k for type i at time j can be written as

$$x_{ijk} = \mu_{ij} + \gamma_k + \omega_{ik} + \omega_{jk} + \varepsilon_{ijk}. \tag{5.4}$$

The variance decomposition of $x_{ijk} - \mu_{ij}$ conditional on a given type i and time j is

$$\sigma^2(x_{ijk}) = \sigma^2(\gamma_k) + \sigma^2(\omega_{ik}) + \sigma^2(\omega_{jk}) + \sigma^2(\varepsilon_{ijk}). \tag{5.5}$$

The covariance between measurements on the same subject k of the same type i but at different times j and j$'$ is

$$\sigma^2(x_{ijk}, x_{ij'k}) = \sigma^2(\gamma_k) + \sigma^2(\omega_{ik}) \tag{5.6}$$

since the same realization of the same subject random variable γ_k, and the type-within-subject random variable ω_{ik}, is applied to different time realizations j and j′.

The covariance between measurements on the same subject k, at the same time j but different types i and i′ is

$$\sigma^2(x_{ijk}, x_{i'jk}) = \sigma^2(\gamma_k) + \sigma^2(\omega_{jk}) \tag{5.7}$$

since the same realization of the subject random variable γ_k, and the time-within-subject random variable, ω_{jk} is applied to different type realizations (namely, i and i′).

The covariance between measurements on the same subject k, between different times j and j′ and different types i and i′ is

$$\sigma^2(x_{ijk}, x_{i'j'k}) = \sigma^2(\gamma_k) \tag{5.8}$$

since only the same realization of the subject random variable γ_k, is applied to different type (i and i′) and time (j and j′) observations.

The above random effects model can be represented as the simplest SEM model by making the following substitutions

$\xi_{Tik} = (\gamma_k + \omega_{ik})$ represents the true values of DBP and SBP within subjects,

$\xi_{Mjk} = \omega_{jk}$ represents the time-within-subjects effect,

$\delta_{ijk} = \varepsilon_{ijk}$ represents the random error, and

$\tau_{ij} = \mu_{ij}$ be fixed.

Equation 5.4 can then be rewritten as:

$$x_{ijk} = \tau_{ij} + \xi_{Tik} + \xi_{Mjk} + \delta_{ijk} \tag{5.9}$$

where τ_{ij} is an intercept term, ξ_{Tik} is a type construct, ξ_{Mjk} is a time construct and δ_{ijk} is an error term.

Furthermore, let

$$\begin{aligned} var(\xi_{Tik}) &= var(\gamma_k + \omega_{ik}) \\ &= var(\gamma_k) + var(\omega_{ik}) \\ &= \phi_T. \end{aligned} \tag{5.10}$$

$$\begin{aligned} var(\xi_{Mjk}) &= var(\omega_{jk}) \\ &= \phi_M. \end{aligned} \tag{5.11}$$

$$\begin{aligned} var(\delta_{ijk}) &= var(\varepsilon_{ijk}) \\ &= \theta. \end{aligned} \tag{5.12}$$

Then the decomposition of the variance of x_{ijk} is

$$\sigma^2(x_{ijk}) = \phi_T + \phi_M + \theta. \tag{5.13}$$

It has been shown that the simplest SEM model given in equation 5.9 is in fact an ANOVA model. It fails to provide information on errors associated with the individual types or times. Only overall estimates are provided. Thus, such a model is more restrictive than a general MTMT framework as described by Batista *et al.* [4]. The random effects model defined by equation 5.9 is a special case of SEM of equation 4.21, as defined by Bollen [10] with:

- Type constructs with a common variance.
- Time constructs with a common variance.
- Loadings of ξ_{Tik} and ξ_{Mjk} all equal to one.
- There are no correlations between constructs.

5.2.3 Generalization to a SEM model for BP

Equation 5.9 represents the simplest form of the MTMT models [5]. A more general form of that model can be constructed, which will enable the estimation

individual variance components as explained by the type of BP or time or errors by:

- Introducing the type loadings (λ_{Tij}) which will enable estimation of the magnitude of the expected change in the indicator (x_{ijk}) for a unit change in the BP type.
- Set the type variances and covariances free so that they may differ and allow estimation of the variances and covariances.
- Set the time variances free so that they may also differ and thus be estimated.
- Estimate six different error variances individually, one for each time and type combination.

Except for the type covariance, all independence assumptions of model (5.9) are maintained. Incorporating all these aspects in model 5.9 results in the following model

$$x_{ijk} = \tau_{ij} + \lambda_{Tij}\xi_{Tik} + \xi_{Mjk} + \delta_{ijk} \tag{5.14}$$

where

τ_{ij} are intercept terms essential for the estimation of the mean BP,

ξ_{Tik} is a random variable with a mean κ_i, unequal variances $E(\xi_{Tik} - \kappa_i)^2 = \phi_{Tii}$ and covariance $E(\xi_{Tik} - \kappa_i)(\xi_{Ti'k} - \kappa_{i'}) = \phi_{Tii'}$. This variable represents the type construct,

ξ_{Mjk} is a random variable with $E(\xi_{Mjk}) = 0$ and unequal variances ϕ_{Mjj}. This variable represents a time construct, and

δ_{ijk} is random and is associated with error such that, $E(\delta_{ijk}) = 0$ and variances θ_{ij} unequal for the time and type.

Hence, the decomposition of the variance of x_{ijk} is:

$$\begin{aligned} \sigma^2(x_{ijk}) &= \lambda^2_{Tij}\sigma^2(\xi_{Tik}) + \sigma^2(\xi_{Mjk}) + \sigma^2(\delta_{ijk}) \\ &= \lambda^2_{Tij}\phi_{Tii} + \phi_{Mjj} + \theta_{ij}. \end{aligned} \tag{5.15}$$

The generalization from 5.13 to 5.15 shows that the variance of each type and time combination can be individually decomposed. The mean of x_{ijk} is given by:

$$\begin{aligned} E(x_{ijk}) &= \tau_{ij} + \lambda_{Tij}\kappa_i \\ &= \mu^*_{ij}. \end{aligned} \tag{5.16}$$

The covariance between any two measurements of subject k for a given type i, at times j and j$'$ is:

$$\begin{aligned}
\sigma^2(x_{ijk}, x_{ij'k}) &= E(x_{ijk} - \mu^*_{ij})(x_{ij'k} - \mu^*_{ij'}) \\
&= E(\tau_{ij} + \lambda_{Tij}\xi_{Tik} + \xi_{Mjk} + \delta_{ijk} - \tau_{ij} - \lambda_{Tij}\kappa_i) \\
&\quad (\tau_{ij'} + \lambda_{Tij'}\xi_{Tik} + \xi_{Mj'k} + \delta_{ij'k} - \tau_{ij'} - \lambda_{Tij'}\kappa_i) \\
&= \lambda_{Tij}E(\xi_{Tik} - \kappa_i)^2\lambda_{Tij'} \\
&= \lambda_{Tij}\phi_{Tii}\lambda_{Tij'} \qquad (5.17)
\end{aligned}$$

Similarly the covariance between any two measurements of subject k for a given time j, between types i and i′ is calculated as:

$$\begin{aligned}
\sigma^2(x_{ijk}, x_{i'jk}) &= E(x_{ijk} - \mu^*_{ij})(x_{i'jk} - \mu^*_{i'j}) \\
&= \lambda_{Tij}E(\xi_{Tik} - \kappa_i)(\xi_{Ti'k} - \kappa_{i'})\lambda_{Ti'j} + E(\xi^2_{Mjk}) \qquad (5.18) \\
&= \lambda_{Tij}\phi_{Tii'}\lambda_{Ti'j} + \phi_{Mjj}.
\end{aligned}$$

The covariance between any two measurements of subject k, for different types and times is similarly calculated as:

$$\begin{aligned}
\sigma^2(x_{ijk}, x_{i'j'k}) &= E(x_{ijk} - \mu^*_{ij})(x_{i'j'k} - \mu^*_{i'j'}) \\
&= \lambda_{Tij}\phi_{Tii'}\lambda_{Ti'j}. \qquad (5.19)
\end{aligned}$$

Model 5.14 can be written formally as a SEM model of the form of Section 4.5 with the structural model

$$\xi_{Tik} = \kappa_i + \varsigma_{ik} \tag{5.20}$$

where ς_{ik} is a disturbance term, which is random with a mean of zero and a variance ϕ_{Tii} and covariance $\phi_{Tii'}$. The measurement model is:

$$x_{ijk} = \tau_{ij} + \lambda_{Tij}\xi_{Tik} + \xi_{Mjk} + \delta_{ijk}. \tag{5.21}$$

Further generalizations of the model can be done by introducing:

- Time loadings (λ_{Mij}).
- Covariances between time constructs
- Covariance between time and type constructs

None of these extensions were implemented in our SEM model for the estimation and modelling of the measurement error of BP in order to avoid overparameterization. Overparameterization will lead to underidentification of the model. In practise [31] this problem is addressed by constriction or fixation of the parameters. In order to avoid this problem it is common practise to allow a single loading in each group of loadings from a particular construct to be fixed at *one* and also one of the intercepts to be set to *zero* [4],[9]. For example, in our applications with model 5.14 we have set one loading (λ_{Tij}) to *one* for each type in order to identify type variances (ϕ_{Tii}). Similarly, in order to identify the type means

(κ_i) one τ_{ij} must be constrained to zero for each type. Not only will these serve to facilitate the identification of the model and scale the loadings, they will also serve as a testing tool as will be demonstrated in the results below. According to Batista-Foguet *et al* [4] from models of the form of equation 5.14 it is possible to

- Assess the bias as will be discussed below.
- Estimate the quality of the six measurements.

To test biasedness among different time constructs, one is fixed and then compared with rest. This tests if the estimated parameters are systematically different from those fixed in a particular *time* construct. For example one would test if *time two* was biased relative to *time one* in a repeated BP data set by testing;

$$\lambda_{Tij} = 1, \tau_{ij} = 0 \text{ for } i = 1, 2, \ j = 2 \tag{5.22}$$

given that $\lambda_{Tij} = 1$, $\tau_{ij} = 0$ for $i = 1, 2$ and $j = 1$. Tests are usually done by maximum likelihood or any of the other full-information methods that are related to the generalized method of moments family.

The quality of the measurements for these types of models (MTMT models) is given by the so-called standardized loadings λ^s_{Tij} calculated as follows

$$\sqrt{Q_{ij}} = \lambda^s_{Tij} = \sqrt{\frac{\lambda^2_{Tij}\phi^2_{Tii}}{\sigma^2(x_{ijk})}}. \tag{5.23}$$

This measure can be used as a basis for choosing the best quality measurements among the times [4]. Blood pressure theory reveals that measures taken at different times will have different measurement qualities [4].

5.2.4 SEM solution

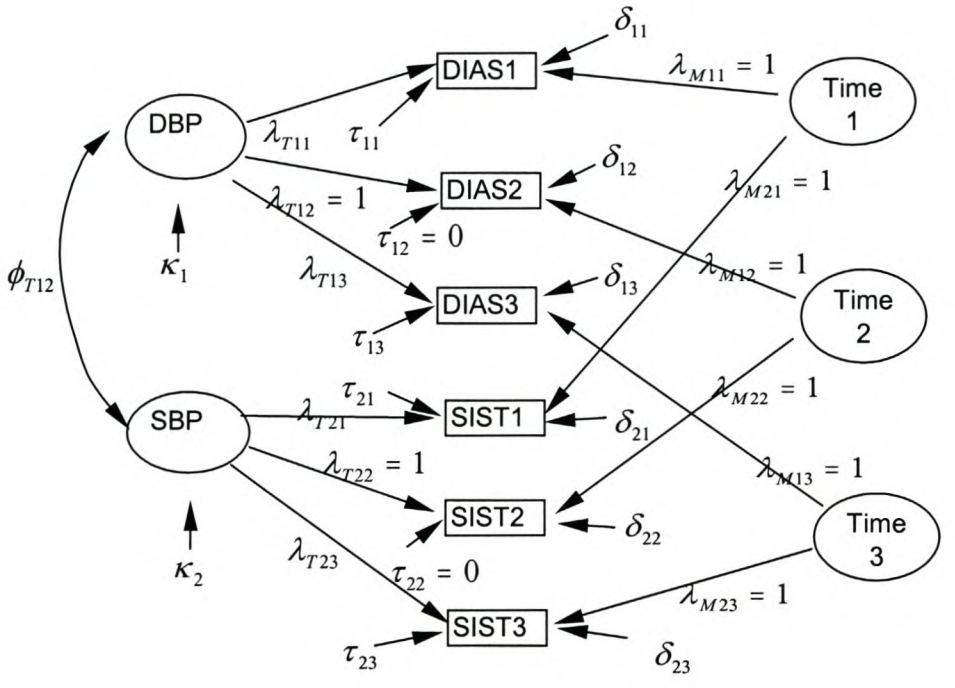

Figure 5.1: Path diagram for the SEM model of equation 5.14

The SEM model given by equation 5.14 is represented pictorially in figure 5.1 according to the conventions of Section 4.4.2. BP readings are represented by rectangles and the BP type and time constructs by ellipses. In figure 5.1, the BP type construct ξ_{Tik} from model 5.14 is represented by DBP and SBP. The

covariance ϕ_{T12} between ξ_{T1k} and ξ_{T2k} is represented by a connecting path. The DBP readings x_{1jk} where $j = 1, 2, 3$ are represented by DIAS1, DIAS2, DIAS3 in the figure 5.1 SIST1, SIST2, SIST3 are the corresponding representations for the SBP readings x_{2jk} for $j = 1, 2, 3$. The time construct ξ_{Mjk} with $j = 1, 2, 3$ has representation by means of TIME1, TIME2, TIME3. All other elements of the model 5.14 are directly indicated with their respective constraints.

The CALIS procedure of SAS [28] was used for the estimation of parameters of the model 5.14. The SAS program for modelling the data is given in the Appendix. The LINEQS model specification was used in the program since it corresponds directly to the path diagram given in figure 5.1 From the path diagram it can also be seen that model parameters were fixed at *time two* ($\lambda_{Ti2} = 1, \tau_{i2} = 0$) since that construct (*time two*) has shown the lowest variances in Chapter 2 (thus indicating it was the most stable). Some variables were given initial values as a means of speeding up the iterative estimation process.

The SEM model 5.14 is applied to each of the following data sets:

- Mamre BP data (Mamre)
- Mitchells Plain digital BP data (MP dig)
- Mitchells Plain mercury BP data (MP mer)

As outlined in Chapter 2 the two studies were done on different populations and the results are presented together in some Tables to facilitate comparisons.

A list of Tables relevant to a specific study and statistical measures are given in Table 5.1. Results will be presented in three main sections, namely:

- Estimated parameters,
- ANOVA versus SEM,
- Goodness of fit.

The specific details of the goodness of fit statistics are given in Section 4.4.8.3. The specific details of the estimated parameters are given in model 5.14 and the standardized loadings are defined in Section 5.2.3, equation 5.23. The residuals discussed in Section 5.2.5.3 are the standardized residuals of Section 4.4.9. The distribution of standardized residuals is illustrated by stem and leaf plots (Tables 5.16, 5.17 and 5.18). In these plots the stem is made of the residual interval (RANGE). The leaf is made up of the number of residuals in that particular interval (FREQ) and the percentage they constitute over the entire population of residuals. These residuals are represented by asterisks (*).

Table 5.7 contains the estimates of variance components as estimated in model 5.4. The random-factor ANOVA was done by PROC MIXED of SAS to obtain the results in Table 5.7, whereas the similar results were obtained by approximating the ANOVA by SEM as described in Section 5.2.2. SAS programs for both the approximation of ANOVA by SEM and ANOVA, using Proc Mixed are in the Appendix. Raw data was used for the estimation according to the discussion of

Model Parameters	Mamre Table	MPdig Table	MP mer Table
λ_{Tij}, τ_{ij}	5.2	5.2	5.2
λ^{s}_{Tij}	5.3	5.3	5.3
θ_{ij}	5.5	5.5	5.5
ϕ_{Mij}	5.4	5.4	5.4
ANOVA versus SEM			
Analysis of variance	5.7	5.7	5.7
Model Fit			
Residual Matrix	5.13	5.14	5.15
Residual Histogram	5.16	5.17	5.18
Goodness of fit statistics	5.8	5.8	5.8
LaGrange multiplier	5.9	5.11	5.12

Table 5.1: Reference table for the Study Results and Statistical Measures

Section 4.5.

5.2.5 Results

5.2.5.1 Estimated parameters

Table 5.2 shows estimates of the unstandardized loadings and intercepts (for all three sets of data) for the free parameters. From this Table we can test for the presence of bias as discussed in Section 5.2.3. We can test the hypotheses : $\tau_{ij} = 0$,

Data	λ_{T11}	λ_{T13}	λ_{T21}	λ_{T23}	τ_{11}	τ_{13}	τ_{21}	τ_{23}
Mamre	0.982	0.993	1.005	0.980	2.673	0.005*	1.669	1.393
MP dig	0.988	1.001	0.989	0.966	2.688*	-0.834*	5.405*	2.375*
MP mer	0.960	0.999	0.986	0.978	4.497	-0.774*	4.266*	1.816*

Table 5.2: Estimates for the unrestricted parameters

$\lambda_{Tij} = 1$. We find that there is a presence of bias between the first and the second diastolic readings in the Mamre data at 5% level of significance ($\tau_{11} = 2.673$ and the t-value= 4.351), whereas there is no indication of bias presence between the second and the third reading ($\tau_{13} = 0.005$ with t-value= 0.012). This confirms what is seen in chapter 2 (see the fitted curve of figure 2.3). Intercept values without the asterix in Table 5.2 are significantly different from *zero* at $\alpha = 0.05$ level of significance. All the loadings are statistically different from *one* at $\alpha = 0.05$ level of significance. For example, $\lambda_{T11} = 0.982$ with a t-value of 126.0 hence we reject the null hypothesis: $\lambda_{Tij} = 1$.

The standardized loadings of Table 5.3 are estimates of the quality of the six measurements. From this Table, *time two* appears to give the best measurement estimates since it has the highest standardized loadings for the SBP readings. For example in the Mamre data the second standardized loading for systolic BP is 1.000 against the first SBP loading (0.988) and the third systolic loading (0.989). On the other hand there is no difference between the second and the third DBP readings. The Mitchells Plain results (both digital and mercury) reveal that the

Data	λ^s_{T11}	λ^s_{T12}	λ^s_{T13}	λ^s_{T21}	λ^s_{T22}	λ^s_{T23}
Mamre	0.966	0.999	0.999	0.988	1.000	0.989
MP dig	0.969	0.999	1.007	0.968	0.999	0.984
MP mer	0.951	0.999	1.008	0.973	0.999	0.988

Table 5.3: Standardized Loadings estimates

Data	ϕ_{M11}	ϕ_{M22}	ϕ_{M33}
Mamre	2.048	0.283	1.436
MP dig	5.543	4.007	14.236
MP mer	5.815	0.555	4.478

Table 5.4: Time variances

second SBP reading is the best whereas among the diastolic BPs the third DBP reading is the best. Thus for all these groups of data the first DBP readings is not as good quality as the second and the third, whereas the second SBP readings have the best quality.

The variance estimates of Table 5.4 indicate that *time two* has lower variances than the other time constructs (0.283 for the Mamre data, 4.007 for MP digital and 0.555 for the MP mercury). This is also confirmed by Table 5.5 since *time two* has the lowest measurement error variances. For example the second diastolic BP measurement error variance in the Mamre data is approximately 1.953 compared to 7.061 and 2.746 for the first and the third measurements respectively. This agrees with the exploratory data analysis results of chapter 2.

Data	θ_{11}	θ_{12}	θ_{13}	θ_{21}	θ_{22}	θ_{23}
Mamre	7.061	1.953	2.746	14.766	2.154	7.678
MP dig	29.302	12.459	16.901	64.293	22.308	98.327
MP mer	10.947	5.101	8.826	22.459	19.126	36.445

Table 5.5: Measurement error variances

Table 5.6 shows construct score regression coefficients for all three data sets. Batista-Foguet *et al.* [4] propose that the factor scores as given in Table 5.6 can be used to calculate a weighted average that gives a better measurement quality for systolic and diastolic BP than an unweighted average. Such a weighted average minimizes the sum of squared measurement errors and thus optimizes measurement quality. From Table 5.6 we see for example that we can estimate the SBP for any subject in the Mamre group as follows:

$$\begin{aligned} systolic &= 0.002 * DIAS1 + 0.013 * DIAS2 - 0.010 * DIAS3 \\ &\quad +0.103 * SIST1 + 0.708 * SIST2 + 0.187 * SIST3. \end{aligned} \quad (5.24)$$

From Table 5.6 there appears to be some negative diastolic score loading when for example, an average systolic BP is estimated by diastolic scores. Besides equation 5.23 measurement quality can also be estimated by an average of the six standardized loadings of Table 5.3 [4]. According to Table 5.6 *time two* has the highest factor scores (except for the mercury diastolic), which also confirms that

Group	Type	DIAS1	DIAS2	DIAS3	SIST1	SIST2	SIST3
Mamre	Systolic	0.002	0.013	-0.010	0.103	0.708	0.187
	Diastolic	0.135	0.553	0.304	-0.003	0.030	-0.023
MP dig	Systolic	0.028	-0.013	0.010	0.227	0.616	0.137
	Diastolic	0.217	0.466	0.255	0.015	0.015	-0.013
MP mer	Systolic	-0.069	0.093	-0.014	0.327	0.453	0.215
	Diastolic	0.185	0.559	0.241	-0.011	0.024	-0.008

Table 5.6: Construct score regression coefficients

time two measures have higher measurement quality than the rest.

5.2.5.2 ANOVA versus SEM

Figure 5.2 is a pictorial representation of the random effects model estimated by an SEM as discussed in Section 5.2.2. A mathematical representation of the model is given by equation 5.9. Table 5.7 shows the SEM approximation of ANOVA, in particular the variance components estimates. The $sub*time$ variance component represents ϕ_M of equation 5.13 and the error variance represents θ. Comparing the ϕ_M of the three data sets in Table 5.7 to the ϕ_{Mjj} of Table 5.4 we observe that the small variances of time two are masked by the ANOVA model. Similarly, this is observed for the comparison of θ versus θ_{ij} of Table 5.5. The common measurement error variance θ masks the smaller measurement error associated with DBP. The θ of 41.208 for the MP digital versus 17.351 for the

MP mercury confirms the larger measurement error associated with the digital instrument found in Table 5.5.

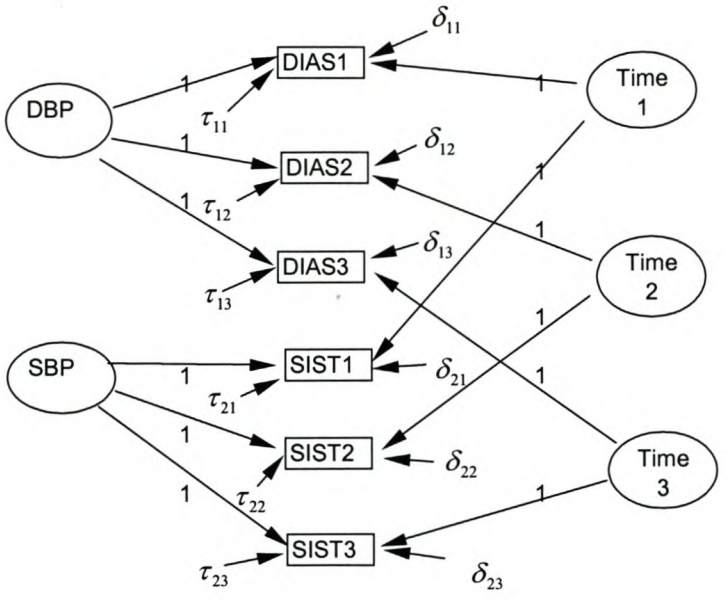

Figure 5.2 Path diagram for the SEM model of equation 5.9

The goodness-of-fit tests show very high chi-squared statistics for all three sets of data indicating that restrictions imposed in order to fit an ANOVA model are not acceptable. Investigating factor scores coefficients for this setting showed that they were 1/3 and 0, thus, unweighted averages. This means that for the estimation of the SBP, a third of each of the observed SBP measures is taken, and none of the observed DBP measurements are used. Similarly for the estimation of the DBP only the observed DBP measures are used. The program for fitting the ANOVA model as an SEM is given in the Appendix.

Data	sub	sub*type	sub*time	Error	$\sqrt{Q_{ijk}}$	χ^2 (17 *d.f.*)
Mamre	252.993	155.481	1.247	6.138	0.982	1198.5
MP dig	261.572	242.929	7.623	41.208	0.912	280.9
MP mer	225.808	261.804	3.561	17.351	0.959	240.5

Table 5.7: ANOVA variance components

5.2.5.3 Goodness of fit

Table 5.8 shows a selection of the goodness-of-fit measures (Section 4.4.8.3) for the three data sets. The GFI measures for the Mamre (0.999), MP digital (0.9874) and MP mercury (0.9966) all agree with residual analyses above that we have good model fit for all of the three data sets. The Chi-square values are fairly low (2.926, 8.9187, 2.4494). One can see from these Chi-squared values that the proposed model did not fit the MP digital data as well as the other data sets did. This was also evident from the distribution of residual plot of Table 5.17. All the p-values for the three data sets suggests that there are nonsignificant differences between the data sets and the proposed model. The Bentley & Bonnet and the Bollen tests were 1.0003 and 1.0001 respectively for the Mamre data, 0.9972 and 0.9993 for the MP digital and 1.0017 and 1.0004 for the MP mercury. These values confirm that all three data sets fitted perfectly since they are far beyond the cut-off point of 0.90.

Table 5.9 shows the ten largest Lagrange Multipliers (LM) for the Mamre data. This Table suggests that a covariance path between DBP construct and

	Mamre	MP dig	MP mer
Goodness of Fit Index (GFI)	0.999	0.9874	0.9966
GFI Adjusted for Degrees of Freedom (AGFI)	0.9952	0.9293	0.9810
Root Mean Square Residual (RMR)	0.711	6.5725	2.1481
Chi-Square	2.926	8.9187	2.4494
Chi-Square DF	5	5	5
Pr > Chi-Square	0.711	0.1124	0.7841
Bentler's Comparative Fit Index	1.000	0.9993	1.0000
Akaike's Information Criterion	-7.074	-1.0813	-7.5506
Bentler & Bonett's (1980) Non-normed Index	1.0003	0.9972	1.0017
Bollen (1988) Non-normed Index Delta2	1.0001	0.9993	1.0004

Table 5.8: Goodness-of-fit Measures for the three data sets

time 3 might improve the model fit more than any other modification (the largest LM is between DBP and TIME3). But since all these modification indices are non-significant at 5% level of significance, these suggestions are ignored. Hence, we assume the model chi-squared cannot be improved for this data set at this level. The *Wald test* (shown in table 5.10) also came out non significant confirming that the proposed model cannot be improved by dropping any of the covariance paths. The right-hand side of this table under the heading "univariate increment" suggests the change in the model that would result from deleting the indicated parameter from the model, when on the other hand the left hand side under "cumulative statistics" estimate how much chi-squared would change if all parameters involved in an identified path were deleted. Tables for the Wald tests for other remaining data sets looked similar and were thus omitted in this report.

In Table 5.11 the ten largest LMs for the MP digital data are displayed. The largest LM here is between DIAS2 and TIME3 with a p-value of 0.0183 (significant). However, this suggestion cannot be accepted since a covariance path is not allowed between a construct and an indicator (see Section 4.4.2). Thus, none of the suggested modifications here either are feasible, thus we assume there can be no improvement of this model. Similarly, the *Wald test* was non-significant.

Similarly, Table 5.12 shows that none of the LMs are significant and furthermore the suggested modifications are not feasible. The largest LM (p-value of 0.2510) is between SIST1 and TIME3, which is non-significant at 5% and not

Pair		Chi-Square	Pr > ChiSq
DBP	TIME3	1.5175	0.2180
SBP	TIME3	1.3147	0.2516
SIST2	TIME3	1.9551	0.3284
DIAS2	TIME1	0.8389	0.3597
DIAS2	TIME3	0.8329	0.3615
DBP	TIME2	0.8093	0.3683
DIAS3	TIME1	0.6004	0.4384
DIAS3	TIME2	0.4916	0.4832
SIST3	TIME2	0.4615	0.4969
SIST2	TIME1	0.3852	0.5348

Table 5.9: 10 Largest Lagrange Multipliers for the Mamre data

	Cumulative Stats			Univariate Increment	
Parameter	Chi-square	DF	Prob	Chi-square	Prob
alf3	0.0001	1	0.9908	0.0001	0.9908
varTime2	1.1665	2	0.5581	1.1664	0.2801

Table 5.10: Wald test for the Mamre data

Pair		Chi-Square	Pr > ChiSq
DIAS2	TIME3	5.5650	0.0183
DIAS2	TIME1	4.7777	0.0288
SIST2	TIME3	4.0921	0.0431
SIST1	TIME2	3.5688	0.0589
SIST3	TIME2	3.0279	0.0818
DBP	TIME3	2.9470	0.0860
SIST1	TIME3	2.2880	0.1304
SIST2	TIME1	1.7200	0.1897
SBP	TIME1	1.4797	0.2238
DBP	TIME1	1.3738	0.2412

Table 5.11: 10 Largest Lagrange Multipliers in the MP digital data

Pair		Chi-Square	Pr > ChiSq
SIST1	TIME3	1.3180	0.2510
DIAS3	TIME2	1.0493	0.3057
SIST2	TIME3	0.9092	0.3403
DIAS3	TIME1	0.9071	0.3409
SIST1	TIME2	0.8064	0.3692
SIST3	TIME2	0.6634	0.4154
DIAS1	TIME2	0.5029	0.4782
DIAS1	TIME3	0.4900	0.4840
DIAS2	TIME3	0.4646	0.4955
DIAS2	TIME1	0.3789	0.5382

Table 5.12: 10 Largest Lagrange Multipliers in the MP mercury data

permissible. The *Wald test* also was not significant.

Table 5.13 shows the residual matrix (Section 4.4.9) for the Mamre data. From this Table the Mamre data seems to have a very good fit since all residuals are far smaller than 2. The average absolute residual (0.0008) and the off-diagonal residual (0.0009) indicate an excellent fit since they are very small. The largest residual is 0.0046, while the smallest is 0.0000.

Table 5.14 shows that the MP digital data also fit the proposed model very well. All residuals are far below the cut-off point of 2. The average residual (0.0026) and the average off-diagonal residual (0.0030) reveal that we also have

	DIAS1	DIAS2	DIAS3	SIST1	SIST2	SIST3	Intercept
DIAS1	0.0006	0.0001	0.0007	0.0046	0.0027	0.0046	0.0000
DIAS2	0.0001	-0.0002	0.0003	0.0014	-0.0006	0.0016	0.0000
DIAS3	0.0007	0.0003	0.0008	0.0002	-0.0022	0.0000	0.0000
SIST1	0.0046	0.0014	0.0002	0.0009	0.0004	0.0001	0.0000
SIST2	0.0027	-0.0006	-0.0022	0.0004	0.0000	-0.0003	0.0000
SIST3	0.0046	0.0016	0.0000	0.0001	-0.0003	-0.0006	0.0000
Intercept	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.13: Standardized Residual Matrix for the Mamre data

	DIAS1	DIAS2	DIAS3	SIST1	SIST2	SIST3	Intercept
DIAS1	-0.0005	-0.0009	0.0016	0.0033	0.0040	0.0153	0.0000
DIAS2	-0.0009	-0.0010	0.0027	-0.0054	-0.0040	0.0107	0.0000
DIAS3	0.0016	0.0027	0.0052	-0.0004	-0.0029	0.0108	0.0000
SIST1	0.0033	-0.0054	-0.0004	0.0009	0.0003	0.0000	0.0000
SIST2	0.0040	-0.0040	-0.0029	0.0003	-0.0006	-0.0005	0.0000
SIST3	0.0153	0.0107	0.0108	0.0000	-0.0005	-0.0002	0.0000
Intercept	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.14: Standardized Residual Matrix for the MP digital data

	DIAS1	DIAS2	DIAS3	SIST1	SIST2	SIST3	Intercept
DIAS1	0.0010	0.0005	0.0009	0.0014	-0.0018	0.0016	0.0000
DIAS2	0.0005	-0.0002	0.0003	0.0003	-0.0031	0.0012	0.0000
DIAS3	0.0009	0.0003	0.0008	0.0045	0.0002	0.0046	0.0000
SIST1	0.0014	0.0003	0.0045	-0.0000	-0.0000	0.0004	0.0000
SIST2	-0.0018	-0.0031	0.0002	-0.0000	-0.0001	0.0004	0.0000
SIST3	0.0016	0.0012	0.0046	0.0004	0.0004	0.0010	0.0000
Intercept	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.15: Standardized Residual Matrix for the MP mercury data

an excellent fit. The largest residual is 0.0040, while the smallest is 0.0000.

Table 5.15 shows that the MP mercury data fitted very well also. The residuals are far less than 2. The average residual (0.0009) and the average off-diagonal residual (0.0010) are also convincingly low. The largest residual is 0.0046, while the smallest is 0.0001.

Table 5.16 shows the distribution of the residuals (for the Mamre data) which is bell-shaped, hence the assumption of normality appears to be supported by this data. Furthermore these residuals are centred around zero (approximately 78.57% were at zero).

The MP digital data also supports the assumption of the normally distributed residuals as can be seen from Table 5.17. Close to 75% of the residuals were at zero. But there appears to be a slight skewness to the right that is revealed by

——Range——		Freq	%	
-0.0023	-0.0011	1	3.6	*
-0.0011	0.0000	5	17.9	*****
0.0000	0.0011	17	60.7	**********
0.0011	0.0023	2	7.1	**
0.0023	0.0034	1	3.6	*
0.0034	0.0046	0	0.0	
0.0046	0.0057	2	7.1	**

Table 5.16: Distribution of Residuals for the Mamre data

this representation.

The MP mercury data is symmetric around zero according to Table 5.18. Approximately 78.58% of the residuals were at zero. Thus this data also appears to support the assumption of normality of the residuals as well.

5.2.5.4 Comparison between the standard method and SEM for the clinical diagnosis of hypertension

In this Section a clinical diagnosis of the hypertension is analyzed. This diagnosis is comprised of the combination of the SBP and DBP. From Section 5.1 we know the best BP measurement estimate is given by the average, which is the mean of the observed BP measurements using the standard method (STD) or can be obtained by the use of weights similarly to those in Table 5.6 if the SEM is

——Range——		Freq	%	
-0.0069	-0.0035	2	7.14	**
-0.0035	0.0000	8	28.57	********
0.0000	0.0035	13	46.43	**********
0.0035	0.0069	2	7.14	**
0.0069	0.0104	0	0.00	
0.0104	0.0138	2	7.14	**
0.0138	0.0173	1	3.57	*

Table 5.17: Distribution of Residuals for the MP digital data

——Range——		Freq	%	
-0.0038	-0.0026	1	3.57	*
-0.0026	-0.0013	1	3.57	*
-0.0013	0.0000	4	14.29	***
0.0000	0.0013	18	64.29	**********
0.0013	0.0026	2	7.14	**
0.0026	0.0038	0	0.00	
0.0038	0.0051	2	7.14	**

Table 5.18: Distribution of Residuals for the MP mercury data

used.

The clinical diagnosis of hypertension is made on the following cutpoints and combinations of diastolic and systolic BP measurements:

- If the *mean* $SBP \leq 140$ mmHg and the *mean* $DBP \leq 90$ mmHg then a subject is normal (NRML)/ has a normal BP.
- If the *mean* $SBP > 160$ mmHg or the *mean* $DBP > 95$ mmHg then a subject is hypertensive (HYPT)/ has hypertension.
- Otherwise the subject is a borderline case (BDRL).

The means of systolic and diastolic BP obtained by SEM and the standard procedure were compared. The following results were obtained:

The Mamre data: The results for the Mamre data in Table 5.19 indicate that there is a slight overestimation of the number of hypertensives by the standard procedure. In fact, there are ten patients that are classified as hypertensive by the standard method, but they are borderline cases according to the SEM. There is only one patient, who is classified as a borderline case by the standard method but is hypertensive according to the SEM.

The Mitchells Plain digital data: The Mitchells Plain digital data displayed in Table 5.20 indicated almost similar results with the two methods. Three patients are classified as hypertensive by the standard method, but they are borderline

		SEM			
		NRML	BDRL	HYPT	TOTAL
STD	NRML	157	1	0	158
	BDRL	8	102	1	111
	HYPT	0	10	697	707
	TOTAL	165	113	698	976

Table 5.19: SEM versus the standard procedure for the Mamre data

		SEM			
		NRML	BDRL	HYPT	TOTAL
STD	NRML	112	4	0	116
	BDRL	4	36	2	42
	HYPT	0	3	41	44
	TOTAL	116	43	43	202

Table 5.20: SEM versus the standard procedure for the MP digital data

cases according to the SEM, on the other hand, classifications are similar for the two methods.

The Mitchells Plain mercury data: The Mitchells Plain mercury data displayed in Table 5.21 show slight overreporting by the standard method. Five of the hypertensives were classified as borderline cases by the SEM, whereas they are hypertensive according to the standard method. Also, eight of the borderline

		SEM			
		NRML	BDRL	HYPT	TOTAL
STD	NRML	91	0	0	91
	BDRL	8	42	0	50
	HYPT	0	5	56	61
	TOTAL	99	47	56	202

Table 5.21: SEM versus the standard procedure for the MP mercury data

cases were classified as normal by the standard method.

5.2.6 Discussion

In order to fit the SEM models, BP types and the different times were modelled as constructs. The BP measures are taken as repeated because of the equal time spacings between them and the measurement error as a within-visit variability. Since the type of BP and measurement times are regarded as constructs the BP measurements followed an MTMT design as described by Campbell and Fiske (1959). The assessment of measurement quality for the six measurements (three DBP and three SBP) showed that the second readings were more stable than the first or the third in all three data sets. This is in line with the theory of BP measurement since it means that a reliable measure is the one that is taken after a patient has been seated a little while. This is what Batista-Foguet *et al.* also observed in their study [4]. Table 5.5 shows for the Mitchells Plain study

that the digital BP readings have larger measurement errors than the mercury BP readings. This was not expected since the digital machine is supposed to be more accurate than the manual one. The community study of Mamre also has a lower measurement error compared to a hypertensive population. This is understandable given that some of the hypertensives are under medication and thus there are bound to be variations in their BP readings.

Groups analyzed had to be homogenous since homogeneity of the population influences measurement quality [4], and hence it affects the measurement error. As a result of this we analyzed the Mitchells Plain data as two groups (separated by the machine). This helped us to avoid analyzing groups with big variations between them (heterogeneity).

SEM models are flexible, in that parameters can be constrained or set free in accordance with the theory of the concept being studied. Tests can be done to see if the data supports such constraints. In our BP data studies, the goodness-of-fit measures informed us that some parameters (loadings and variances) are better off set free than being constrained.

The random-effects ANOVA approach has always been used for the analysis of MTMT data despite its rigid assumptions [4]. This approach proved to be simple and effective for the analyses of these types of data, but lacks the generalization to accommodate the theory of a complex concept like BP. Fitting a random-effects ANOVA to the BP data showed that restrictions implied by this SEM model are

not acceptable. However using a more general model in SEM we could break the overall BP type, time and error variances into individual components.

The comparison of the clinical diagnoses shows that there were no significant differences between the SEM and the standard method. For the mercury-based BP measurements (for both Mamre and MP) the SEM classification is slightly more conservative than the standard procedure.

The measurement error estimate of a true BP as viewed in chapter 3 can now be estimated, by substituting systolic and diastolic measurement error estimates of table 5.5 in equation 3.29. For each of the data the Calis procedure also provides the correlation estimate between SBP and DBP.

CHAPTER 6

CONCLUSION

6.1 Context

Blood pressure studies have always generated a lot of scientific interest over the years. These studies revealed that BP has a highly reproducible profile, with higher values when the patient is awake, mentally and physically active, but much lower values during rest and sleep. In the light of this, ambulatory BP monitoring (ABM) in particular became the subject of main interest (White, 2003 [64]). In fact Clement *et al.* [13] found that 24-hour ambulatory blood pressure predicts the risk of cardiovascular events even after adjustment for classic risk factors including office BP. Thus, ABM can be thought of as a gold standard for the diagnosis of hypertension. However, only first world countries can afford this technology whereas South African clinics and general practitioners still have to use mercury-based system for this purpose in the public-health setting. Hence, the relevance of a good quality BP measurement is still evident.

The study of BP measurement is also very important because high BP is an indicator for other serious illnesses as discussed in Section 1.2. Thus, inaccurate measures of BP will lead to misdiagnoses which can have serious consequences.

Therefore, highly skilled and experienced personnel are needed to monitor BP equipment and as well as taking the measurements. The work presented in the earlier chapters established that there is indeed measurement error present in BP reading. The question that follows is that of the analysis method of measurement that will take into account this measurement error.

It was shown in Section 1.7 that there is more than one type of measurement error present in the BP measurement. These include random error that causes differences between repeated measurements taken under the same measurement conditions, and systematic error which occur if some conditions are changed across the measurements. The systematic error decomposes into variance components (variant across subjects) and bias (constant across subjects) [4]. Recent analysis methods deal with measurement error by;

- Estimating measurement error variances, based on the calculation of correlation between measurements of the same subject or estimation of the within-subject variance in a repeated ANOVA model.

- Calculating composite measures to improve accuracy of BP measurements such as calculating weighted averages that will increase the measurement accuracy.

- Correcting the bias effect of measurement error by using previous information on measurement error variances.

Thus a statistical method that will take into consideration the conditions under which BP is measured will be the most useful in the estimation of measurement error. SEM has been demonstrated as one of such statistical methods.

6.2 Technical discussion

In this work particularly, different times were modelled in a repeated measures scenario as different constructs. This was done since it is well documented in BP measurement theory ([1]-[4]) that time is a significant factor in BP measurement. Although SEM is broad the focus was mainly restricted to the analysis of the longitudinal data. Most applications of SEM use raw data and most analyses of experimental data employ ANOVA or regression techniques. However, both the ANOVA and regression are special cases of structural equation modelling as was shown in Table 4.4 [10]. When applying the SEM technique one is likely to encounter problems such as specification error. This basically means the researcher is using a "wrong model". Hence, caution is needed in using SEM models and making sure that one's model is always based on theory and no important variables are omitted. This is achieved through causation, which is the principle by which cause and effect is established between two variables. This requires that there be sufficient degree of association (correlation) between two variables (constructs), that one construct occurs before the other (i.e., one construct is clearly an outcome of the other), and that there be no other reasonable causes for the

outcome. Although in its strictest terms causation is rarely found, in practise strong theoretical support can make empirical estimation of causation possible as discussed in Section 4.4.1.

Recent work on SEM employs the raw data and expands the ability of SEM to include the estimation of the mean (Bollen, [10]). On the other hand use of covariance or correlation matrices used shorter computer time for the estimation of parameters, whereas the estimation of the mean could not be done [28]. From the results of chapter 5 SEM appeared to be a less restrictive tool for analyzing the data than the traditional ANOVA methods. SEM models give more detailed information about the data, and enable researchers to compute averages of individual measures with optimal measurement quality. Measurement error is allowed to differ and can be quantified for each BP measurement.

Since in SEM each BP measurement is individually evaluated, factor scores, which are estimates of the subjects' true BPs with optimal measurement quality, can allocate the highest weights to the highest quality measurements.

From the results in Section 5.2.5 equal error variances across measurements was not supported by the data, judging by the different measurement quality values for each method (time). This is inline with the fact that BP measurement is affected by a number of factors (for example, time) as discussed in Section 1.3.

6.3 Extension of this work

Batista *et al.* [4] looked at a study design where a single machine is used to measure BP but measurements are taken at different positions or "methods". Although the three measures they obtained for each subject did not constitute replicates due to the change of method, the average of the three measurements was taken as the best estimate of the subject's true BP. Their models are multi-trait multi-method (MTMM) models. In this study focus was at a study design where BP measurements are also taken by a single machine at a single position (namely, sitting) as explained in Section 5.2.6, with equal time intervals between them. These models were named MTMT models.

This work can be extended to study designs where two or more machines are used, which could be done for the MP data. Thus, resulting in the so-called multi-type multi-method multi-time models (MTMMMT). However, caution will be needed for the specification of such a models since there would be many unknown parameters thus, an identification problem might exist.

Factors such as gender, body mass, age are also known to be significant in BP measurement [55] as discussed in Section 1.2, thus the data could have been adjusted for these factors provided the groups formed were large enough. Furthermore groups analyzed might be heterogenous and thus a different methodology to the one used in this thesis might be needed. In SEM correction of bias due to measurement error can also be done on the results of substantive studies by using

external measurement quality estimates.

6.4 Some disadvantages and advantages of SEM

SEM has often been criticized when it comes to its goodness-of-fit measures. Assessing the goodness-of-fit is not straightforward in SEM. To begin with, there is no single statistical test which best describes the "strength" of the model's prediction, instead a number of them are used. Presently there is only one goodness-of-fit measure with known distributional properties, namely the chi-squared statistics. The chi-squared measure is very sensitive to both large and small sample sizes and to departures from multivariate normality of the observed variables. Consequently, any model is likely to be rejected if the sample size is sufficiently large simply because of the differences between the corresponding elements of the observed and predicted covariance matrices.

Other goodness-of-fit measures in Table 5.8 have a recommended threshold of 0.90 and greater. However, it is frequently found that even models with measures of 0.90 can be rejected [9]. Furthermore the theory of SEM reveals that the choice of cutoffs for indices can be influenced by standards set by prior work.

The *RMSEA* measure of Table 5.8 is unfavourable for use when the sample size is small, whereas the comparative fit index (CFI) is preferred. However, CFI is suitable in more exploratory contexts, while RMSEA is more suitable to confirmatory situations. The RMSEA is also good at detecting model misspecifications.

With these in mind, care must then be taken when interpreting goodness-of-fit measures as there are no strict rules to adhere to.

6.5 Final discussion

It has been shown in this and other work (for example, Batista et al. [4]) that the strength of SEM lies in its ability to deal with multiple relationships simultaneously in an easy manner, while still providing statistical efficiency. SEM embodies several multivariate techniques, for example multiple regression, ANOVA and factor analysis (see Table 4.4). It is demonstrated in this thesis that SEM models are a strong alternative to be considered for the analysis of BP measurement whenever repeated measures are available even when such measures do not constitute equivalent replicates. Unlike in social sciences, applications of SEM in BP studies where there are repeated measurements do not often lead to artificial environment with undesirable consequences such as memory effects. Although SEM is a familiar tool in sociology it becomes clear from this work and other recent applications of SEM, like Batista *et al.* [4], that public health can certainly benefit from SEM. In medicine repeated observations constitute natural strategies to reduce measurement error and SEM in such cases offers a number of advantages over other methodologies.

APPENDICES

The SAS program for the SEM as defined by equation 5.14 implemented in Section 5.2.4. The values in brackets are the initial values for the model.

proc calis aug ucov data=mp.sdata pestim maxit200 platcov modification residual stderr;

var dias1 dias2 dias3 sist1 sist2 sist3;

LINEQS

dias1=alf1(-2.7) intercept + Ldias1DBP(1.0) DBP + TIME1 + E1,

dias2= DBP + TIME2 + E3,

dias3=alf3(-2.7) intercept + Ldias3DBP(1.0) DBP + TIME3 + E3,

sist1=alf4(-2.7) intercept + Lsist1SBP(1.0) SBP + TIME1 + E4,

sist2= SBP + TIME2 + E5,

sist3=alf6(-2.7) intercept + Lsist3DBP(1.0) SBP + TIME3 + E6,

DBP=kappa (80.0) intercept + D1,

SBP=kappa (130.0) intercept +D2;

STD

E1-E6=varE1-varE6,

D1-D2=varD1-varD2,

TIME1-TIME3=varTIME1-varTIME3;

COV

D1 D2=CD1D2;

run;

SAS program for SEM estimation of the ANOVA as defined by equation 5.9 implemented in Section 5.2.5.2.

proc calis aug ucov data=mp.sdata pestim maxit200 platcov modification residual stderr;

var dias1 dias2 dias3 sist1 sist2 sist3;

LINEQS

dias1=alf1 intercept + DBP + TIME1 + E1,

dias2= alf2 intercept + DBP + TME2 + E3,

dias3=alf3 intercept + DBP + TIME3 + E3,

sist1=alf4 intercept + SBP + TIME1 + E4,

sist2= alf5 intercept + SBP + TIME2 + E5,

sist3=alf6 intercept + SBP + TIME3 + E6;

STD

E1-E6=varError,

DBP-SBP=varTrait,

```
TIME1-TIME3=varTime;
COV
D1 D2=Ctrait;
run;
```

The SAS program for modelling the variance components in Section 5.2.5.2 using Proc Mixed.

```
proc mixed data=set4 method=REML lognote;
class time type idno;
model pbr=type|time/ DDFM=BW;
random time /subject=idno;
run;
```

BIBLIOGRAPHY

[1] American Heart Association: *Joint recommendations for human blood pressure determination by sphygmomanometers.* New York, AHA. 1988

[2] Armstrong RS. *Nurses' knowledge of error in blood pressure measurement technique.* Int J Nurs Pract. 2002 Jun; **8(3)**:118-126

[3] Arruda-Olson AM, Mahoney DW, Nehra A, Leckel M, Pellikka PA. *Cardiovascular effects of sildenafil during exercise in men with known or probable coronary artery disease: a randomized crossover trial.* JAMA. 2002 Feb 13; **287(6)**:719-25

[4] Batista-Foguet JM, Coenders G, Artés Ferragud M. *Using structural equations models to evaluate the magnitude of measurement error in blood pressure.* Statistics in Medicine, Wiley, New York, 2001; **20**: 2351-2368.

[5] Biemer P, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S: *Measurement errors in surveys.* Wiley: New York,1991

[6] Bland JM, Altman DG. *Education and debate: Measurement Error.* BMJ. 1996; **313**: 744.

[7] Bland JM, Altman DG. *Education and debate: Validating scales and indexes.* BMJ,2002; **324**:606-7

[8] Bock RD, Bargmann RE. *Analysis of covariance structures.* Psychometrica 1966;31:507-534.

[9] Bollen KA. *Structural equations with latent Variables.* New York : Wiley, 1989

[10] Bollen KA. *Structural Equation Models.* Encyclopedia of Biostatistics. pp 4363-4372. P. Armitage and T.Colton (editor-in chief). Sussex, England. John Wiley.1998

[11] Burch GE, DePasquale NP. *Primer of clinical measurement of blood pressure.* New Orleans, LA, 1962

[12] Carroll RJ. Measurement error in Epidemiological Studies. Encyclopedia of Epidemiologic Methods. pp 530-556. P. Armitage and T.Colton (editor-in chief). Sussex, England. John Wiley. 2000.

[13] Clement DL, De Buyzere ML, De Bacquer AD, de Leeuw PW, Duprez DA, Fagard RH, Gheeraert PJ, Missault LH, Braun JJ, Six RO, van der Niepen P, O'Brien E. *Prognostic value of ambulatory blood pressure recordings in patients with treated hypertension.* The New England Journal of Medicine. Vol **348**:2407-2415. June 2003.

[14] Cook NR, Gillman MW, Rosner BA, Taylor JO, Hennekens CH. *Combining annual blood pressure measurements in childhood to improve prediction of young adult blood pressure.* Statistics in Medicine, Wiley, New York, 2000; **19**:2625-2640.

[15] Coon NR. *Estimating predictive values for blood pressure measurements from multivariate regression models with covariates.* Statistics in Medicine, Wiley, New York, 1996; **15**:2013-2038

[16] Cook NR, Rosner BA. *Screening rules for determining blood pressure status in clinical trials: Application to the trials of hypertension prevention.* Am J Epidemiol 1993; **137**:1341-52.

[17] Daniels A, Hoffman M, Lombard CJ, Steyn K, Levitt NS, Katzenellenbogen J. *Blood pressure and social support observations for Mamre, South Africa, during social and political transition.* Journal of Human Hypertension; **13**,689-693 (1999)

[18] Dear KBG, Puterman ML, Dobson A. *Estimating correlations from epidemiological data in the presence of measurement error.* Statistics in Medicine, Wiley, New York,1997; **16**, 2177-2189.

[19] DeShon RP. *Do Structural Equation Model Corrects For Measurement Error?* Research Methods Forum, no **2**,1997.

[20] Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data.* New York: Oxford University Press,1994

[21] Dunn G. *Design and analysis of reliability studies (the statistical evaluation of measurement errors).* Wiley: New York,1989.

[22] Dunn G. *The problem of measurement error in modelling the effect of compliance in a randomized trial.* Statistics in Medicine, Wiley, New York, 1999; **18**:2863-2877.

[23] Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. *Statistical methodology for the concurrent assessment of interrater and intrarater reliability using goniometric measurements as an example.* Phys Ther.1994, **74**: 777-788 ;

[24] Fuller AW, Hidiroglou AM. *Regression estimation after correcting for attenuation.* JASA, 1978; **73**:99-104

[25] Gillman MW, Cook NR, Rosner B, Beckett LA, Evans DA, Keough ME, Taylor JO, Hennekens CH. *Childhood blood pressure tracking correlations corrected for within-person variability.* Statistics in Medicine, Wiley, New York,1992; **11**,1187-1194.

[26] Gleser LJ. *The importance of assessing measurement reliability in multivariate regression.* Journal of American Statistical Association. **87(419)**,1992

[27] Hair JF,Anderson RE, Tatham RL, Black WC. *Multivariate data analysis with readings.* New York: Macmillan Publishing Company, 1992.

[28] Hatcher Larry. *A step-by-step approach to using the SAS system for Factor Analysis and Structural Equation modelling,* Cary, NC: SAS Institute Inc.,1994

[29] Hayduk LA. *Structural Equation modelling with LISREL (essentials and advances).* The John Hopkins University Press: Baltimore and London, 1987

[30] Jaech JL.*Statistical Analysis of Measurement Errors.* Wiley: New York,1985

[31] Jöreskog KG. *Structural Equations Models In The Social Sciences: Specification, Estimation and Testing.* Invited paper for the Symposium on Application of Statistics, Dayton, Ohio, June 14-18,1976.

[32] Küchenhoff H, Carroll RJ. *Segmented regression with errors in predictors: Semi-parametric and parametric methods.* Statistics in Medicine, Wiley, New York,1997; **16**,169-188.

[33] Kupper LL. Encyclopedia of Epidemiologic Methods. pp 530-555. P. Armitage and T.Colton (editor-in chief). Sussex, England. John Wiley. 2000

[34] Lambert PC, Abrams KR, Jones DR, Halligan AWF, Shennan A. *Analysis of ambulatory blood pressure monitor data using a hierarchical model incorporating restricted cubic splines and heterogeneous within-subject variances.* Statistics in medicine, Wiley, New York,2001; **20**:3789-3805.

[35] Littell RC, Pendergast J, Natarajan R. *Modelling covariance structure in the analysis of repeated measures data.* Statistics in Medicine, Wiley, New York,2000; **19**:1793-1819.

[36] Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models,* Cary, NC: SAS Institute Inc., 1996.

[37] Levitt NS, Steyn K, Lambert EV, Reagon G, Lombard CJ, Fourie JM, Rossouw K, Hoffman M. *Modifiable risk factors for type 2 diabetes mellitus in a peri-urban community in South Africa.* Diabetic Medicine, **16**, 946-950 (1999)

[38] Metcalf CA, Hoffman MN, Steyn K, Katzenellenbogen JM, Fourie JM. *Design and baseline characteristics of a hypertension intervention program in a South African village.* Journal of Human Hypertension, **10,**21-26 (1996)

[39] MRC technical report: *Chronic diseases of life style in South Africa, Review of research and identification of essential health research priorities.* 1995

[40] MRC technical report: *Coronary Heart Disease Risk Factor Study in the African population of the Cape Peninsula (BRISK STUDY).* 1991

[41] Muthén BO. Beyond SEM: *General latent variable modelling.* Behaviormetrika, **29(1)**, 81-117, 2002

[42] Nurminen M, Hytönen M, Sala E. *Modelling reproducibility of acoustic rhinometry.* Statistics in Medicine,Wiley, New York, 2000; **19**:1179-1189

[43] O'Brien E, Waeber B, Parati G, Staessen J, Myers MG. *Blood Pressure measuring devices: recommendations of the European Society of Hypertension.* Clinical Review. BMJ Vol **322**, p531-536. march:2001

[44] Oppenheimer L, Kher U. *The impact of measurement error on the comparison of two treatments using a responder analysis.* Statistics in Medicine, Wiley, New York,1999; **18:**2177-2188

[45] Palta M, Lin CY. *Latent variables, Measurement error and methods for an-*

alyzing longitudinal binary and ordinal data. Statistics in Medicine, Wiley, New York, 1999; **18**, 385-396.

[46] Park T, Lee JY. *Covariance models for nested repeated measures data: analysis of ovarian steroid secretion data.* Statistics in Medicine. Wiley, New York, 2002; **21**:143-164.

[47] Pinna DG, Opasich C, Mazza A, Tangenti A, Maestri R, Sanarico M. *Reproducibility of the six-minute test in chronic heart failure patients.* Statistics in Medicine, Wiley, New York,2000; **19**:3087-3094

[48] Ramsey FL, Schafer DW. *The Statistical Sleuth: A course in methods of data analysis.* Duxbury Press, New York, 1996

[49] Rosner B, Polk BF. *Predictive values of routine blood pressure measurements in screening for hypertension.* Am J Epidemiol 1983; **117**:429-42

[50] Rosner B, Polk BF. *The implications of blood pressure variability for clinical screening purposes.* J Chron Dis,1979, vol **32**, pp451-461

[51] SAS Institute Inc., SAS/STAT® *User's guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 3884 pp.

[52] Searle SR. *Linear Models.* Wiley: New York,1971

[53] Singer JD. *Using SAS proc mixed to fit Multilevel Models, Hierarchical Models, and Individual growth Models.* Journal of Ed and behavioral stats,1998; Vol **24**,No 4, pp.323-355

[54] Steiger JH. *Driving fast in reverse- the relationship between software development, theory and education in structural equation modelling.* JASA, March 2001, Vol.**96**. no 453, Review Paper

[55] Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J. *Hypertension in South African adults from the demographic and health Survey,* 1998. Journal of hypertension, 2001. **19**:1717-1725

[56] Steyn K, Levitt N, Fourie J, Rossouw K, Martell R, Stander I. *Treatment status and experiences of hypertension patients at a large health centre in Cape Town.* Ethnicity and disease,Vol **9**, 441-450,1999.

[57] Steyn K, Hoffman M,Levitt NS, Lombard CJ, Fourie JM. *Community-based tobacco control program: The Mamre Study, A demonstration project.* Ethnicity and Disease,2001; **11**:296-302

[58] SYSTAT® 10. Statistics II, SPSS Inc. 2000

[59] Temple NJ, Steyn K, Hoffman M, Levitt NS, Lombard CJ. *The epidemic of Obesity in South Africa: A study in a disadvantaged community.* Ethnicity and Disease, Vol **11**,2001,pp 431-437

[60] Thomas D, Stram D, Dwyer J. *Exposure Measurement Error; Influence on Exposure-Diseases: Relationships and Methods of correction.*Annu Rev Health. 1993. **14**;69-93

[61] Yanez DN, Kronmal AR, Shemanski LR. *The effects of measurement error in response variables and tests of association of explanatory variables in change models.* Statistics in Medicine, Wiley, New York, 1998;**17**:2597-2606.

[62] Wagner JA, Schnoll RA, Gipson MT. *Development of a scale to measure adherence to self monitoring of blood glucose with latent variable measurement.* Diabetes Care, Vol **21**, no7, 1998

[63] Wheaton B, Muthén B, Alwin D, Summers G. *Specification and estimation of panel models incorporating reliability and stability parameters.* In D.R Heise (Ed.), Sociological Methodology. 1977. In press.

[64] White WB. *Ambulatory Blood Pressure Monitoring in Clinical Practice.* The New England Journal of Medicine. Vol **348**:2377-2378. June 2003.

[65] Wu MC, Albert PS, Wu BU. *Adjusting for a drop-out in clinical trials with repeated measures: design and analysis issues.* Statistics in Medicine, Wiley, New York, 2001; **20** :93-108.