ERRORS IN VARIABLES REGRESSION: WHAT IS THE APPROPRIATE MODEL?

Thesis submitted to
CARDIFF UNIVERSITY
for the degree of
DOCTOR OF PHILOSOPHY
by

JONATHAN WILLIAM GILLARD

November 2007,
School of Mathematics,
Cardiff University

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in any candidature for any degree.

Signed J. W. Gillard

Date 10/03/08

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references and a bibliography is supplied.

Signed J. W. Gillard

Date 10/03/08

Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and inter-library loan, and for the title and summary to be made available to outside organisations.

Signed J. W. Gillard

Date 10/03/08

Acknowledgements

Without the help and support of a number of people, this thesis would never have existed. I would like to thank my parents for their unconditional support throughout my many years of education, Becky for putting up with me, and all my friends for offering a much needed distraction at the relevant times!.

Special thanks are also due to my supervisor, Mr T. C. Iles for his outstanding supervision over the past few years, and for being an endless source of ideas and suggestions. My thanks are also extended to Dr A. B. J. Nix for the many helpful conversations and thoughts.

Finally, my thanks go to the EPSRC for provision of funding for this research.

Jonathan Gillard

November 2007.

Summary

The fitting of a straight line to bivariate data (x,y) is a common procedure. Standard linear regression theory deals with the situation when there is only error in one variable, either x, or y. A procedure known as y on x regression fits a line where the error is assumed to be associated with the y variable, alternatively, x on y regression fits a line when the error is associated with the x variable. The model to describe the scenario when there are errors in both variables is known as an errors in variables model.

Errors in variables modelling is fundamentally different from standard regression techniques. The problems of model fitting and parameter estimation of a straight line errors in variables model cannot be solved by generalising a simple linear regression model.

Briefly, this thesis provides a unified framework to the fitting of a straight line errors in variables model using the method of moments. Estimators of the line using a higher moments approach have been detailed, and asymptotic variance covariance matrices of a plethora of slope estimators are provided. Simulations demonstrate that these variance covariance matrices are accurate for even small data sets. The topic of prediction is considered, with an estimator for the latent variable presented, as well as advice on the mean value of y given x via both à parametric and non-parametric approach. The problem of residuals in an errors in variables model is described, and some quick solutions given. Some examples are presented towards the end of this thesis to demonstrate how the ideas provided may be applied to real-life data sets, as well as some areas which may demand further research.

List Of Notation

As there is a lot of notation in this thesis, only the most commonly used notations are given here. Some symbols are specific to this thesis, whilst some are used more generally in statistics. All notations are carefully explained at the appropriate place in the text.

ξ_i	An unobserved latent measurement.
η_i	An unobserved latent measurement such that $\eta_i = \alpha + \beta\xi_i$
α	The intercept of a straight line.
β	The slope of a straight line.
δ_i	A random error component with zero mean and variance σ_δ^2
ε_i	A random error component with zero mean and variance σ_ε^2
x_i	An observed measurement on the latent variable ξ_i, $x_i = \xi_i + \delta_i$.
y_i	An observed measurement on the latent variable η_i, $y_i = \eta_i + \varepsilon_i$.
ω_i	Equation error with zero mean and variance σ_ω^2 added to y_i.
μ	Generic symbol for a mean value. Exact definition depends on context, usually $\mu = E[\xi]$.
σ^2	Generic symbol for a variance. Usually $\sigma^2 = Var[\xi]$.
λ	Ratio of error variances $\frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$.
κ	Reliability ratio $\frac{\sigma^2}{\sigma^2+\sigma_\delta^2}$.
$\tilde{e}$	The method of moments estimator for the parameter e.
$\widehat{e}$	The maximum likelihood estimator for the parameter e.
$\bar{x}$	The sample mean of x measurements. Defined similarly for other variables.
s_{xy}	Statistic defined as $\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})(y-\bar{y})$. Statistics such as s_{xx} defined similarly.
n	Sample size.
i	As a subscript, relates to an individual data point, $i = 1, \ldots, n$.
μ_{ri}	The i-th cental moment of the variable r.
$\phi(\cdot)$	Probability density function of the standard Normal distribution.
$\Phi(\cdot)$	Cumulative density function of the standard Normal distribution.
$f_{a,b,c}(a,b,c)$	The joint probability density function of variables a, b and c.

Contents

Chapter 1

Introduction

1.1 Introductory Remarks

The fitting of a straight line to bivariate data (x,y) is a common procedure. Standard linear regression theory deals with the situation when there is only error in one variable, either x, or y. A procedure known as y on x regression fits a line where the error is assumed to be associated with the y variable, alternatively, x on y regression fits a line when the error is associated with the x variable. Both of these regression techniques will be briefly outlined here.

Using y on x regression If the error is associated with the y variable, a suitable linear model could be

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

where $(x_1, y_1), \ldots, (x_n, y_n)$ are our observations, and $\varepsilon_1, \ldots, \varepsilon_n$ are considered to be random error components, each with zero mean and non-zero variance. The parameters α and β may be estimated by minimising some function of these random error components.

Least squares theory as advocated by Carl Freidrich Gauss (1777-1855) and Adrien

Marie Legendre (1752-1833) suggests minimising the sum of the squared error components. In other words, the y on x regression line is obtained by minimising the sum of squares of the vertical discrepancies from the data to the regression line. Mathematically speaking, this involves minimising the quantity

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \tag{1.1}$$

By differentiating (1.1) with respect to each of the parameters, and solving the equations which arise by setting the derivatives to zero we obtain the least squares estimators of the parameters α and β as

$$\begin{aligned} \tilde{\beta}_0 &= \frac{s_{xy}}{s_{xx}} \\ \tilde{\alpha}_0 &= \bar{y} - \widehat{\beta}\bar{x} \end{aligned}$$

where

$$\begin{aligned} s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ s_{xx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ s_{yy} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \end{aligned} \tag{1.2}$$

and $\bar{x}$ and $\bar{y}$ are the usual sample means. It can easily be shown that the sum of squares in (1.1) is minimised with $\alpha = \tilde{\alpha}_0$ and $\beta = \tilde{\beta}_0$ (see for example Draper and Smith [37]). These sample quantities are fundamental to the topic of regression and will appear throughout this thesis.

Using x on y regression Now assume that the error is associated with the x variable. Least squares estimation can still be used, but the sum of the squared horizontal discrepancies from the data to the regression line is minimised instead. The methodology of finding this minimum value is identical to that of y on x regression, and thus details are omitted. Since the assumed relationship takes the form $x = \phi + \theta y$ the slope estimator for this model is $\tilde{\theta} = \frac{s_{xy}}{s_{yy}}$, but comparison with $\tilde{\beta}$ is made by taking the reciprocal so the comparable estimator is

$$\tilde{\beta}_0^* = \frac{s_{yy}}{s_{xy}}$$

The estimators quoted above for both y on x and x on y regression are the best linear unbiased estimators for the given model. This follows from the Gauss-Markov theorem (see Draper and Smith [37]). However, the models that have been considered so far assume that there is a homoscedastic error structure. Models with heteroscedastic error also may occur, and a modified least squares procedure, known as weighted least squares can be used to obtain estimates of the parameters. Such models are not discussed in this thesis, and the reader is again referred to Draper and Smith [37] for more details.

Once the regression line has been obtained, a Student's t test, or an analysis of variance procedure can be used to test the significance of the regression. Confidence intervals can be constructed for the slope and the intercept, and a lack of fit test performed for the chosen model. Further details on this methodology, and of fitting standard regression lines may be found in Draper and Smith [37].

An assumption made in both y on x regression and x on y regression is that error is only present in one variable. In some situations however, it may be possible that there are errors in both variables. This is commonly known as the errors in variables or measurement error model. Casella and Berger in [16] comment that the errors in variables model

> "is so fundamentally different from the simple linear regression ... that it is probably best thought of as a different topic."

This type of model usually occurs when both the x-variable and y-variable are experimentally measured.

Authors such as Kendall and Stuart [67] have shown that the least squares estimate for the slope in y on x regression is biased if applied to an errors in variables model. This emphasises the importance of using modelling appropriately and carefully. Finding the balance between functionality and simplicity is crucial. Krzanowski [69] comments that a model may be constructed to appear mathematically elegant, but unless a user can fully understand and operate the model it is worthless.

All parametric models are developed from making particular assumptions. It is important that these assumptions correspond to the data. Ideally, the data would be allowed to speak for themselves, rather than having a model aggressively forced upon them. Indeed, natural data will never follow a model. Box [11] wrote

> "Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad".

He also stated

> "in nature there was never a normal distribution, there was never a straight line, yet with normal and linear assumptions, known to be false, he can

often derive results which match to a useful approximation those found in the real world".

Tsay [105] took the extreme view that

"Since all statistical models are wrong, the maximum likelihood principle does not apply."

These ideas have recently been addressed by Longford [73]. Committing to a model, and putting

"all our inferential eggs in one unevenly woven basket"

may ignore a disastrous error. This is particulary the case when certain modelling techniques rely on making heavy assumptions - assumptions which may not properly reflect the data in question. For example, James [59], mentions that a normal distribution is commonly assumed for an error term, even though a negative measurement cannot be observed. He quotes blood pressure as such an example of a variable that cannot take negative values.

Although simple linear regression models that are associated with the problem of fitting straight lines to scattered data are inevitably wrong, their widespread use indicates that they are not always, in Box's sense, importantly wrong. However, this thesis describes circumstances where simple linear regression models are importantly wrong; where there are measurement errors in both the x and y variables. In these circumstances a completely different type of model is called for.

1.2 The Linear Errors in Variables Model

Suppose that there are n individuals in a sample with true values (ξ_i, η_i) and observed values (x_i, y_i). It is assumed that there is an underlying linear relationship between ξ_i and η_i

$$\eta_i = \alpha + \beta\xi_i$$

However there is variation in both variables that result in a deviation of the observations from the true values, resulting in a scatter about the underlying straight line. This scatter is represented by the addition of a random error component to the true values. The observations x_i and y_i can be written

$$\begin{aligned} x_i &= \xi_i + \delta_i \\ y_i &= \eta_i + \varepsilon_i = \alpha + \beta\xi_i + \varepsilon_i \end{aligned}$$

These errors, δ and ε are assumed to be independent of ξ. To use the terminology of Carroll et al.[14], ξ_i are latent variables, whilst x_i are surrogate variables.

Errors in variables modelling can be split into two general classifications defined by Kendall [65], [66], as the functional and structural models. The fundamental difference between both models regards the treatment of the $\xi_i's$

The functional model This assumes the $\xi_i's$ to be unknown, but fixed constants μ_i. If

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}\right],$$

this will be referred to as the Normal functional model.

The structural model This model assumes the $\xi_i's$ to be a random sample from a distribution with mean μ and variance σ^2. If

$$\begin{pmatrix} \xi_i \\ \delta_i \\ \varepsilon_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}\right],$$

then this type of structural model will be referred to as the Normal structural model.

An extension of the structural model is the ultrastructural model. The ultrastructural model extends the structural model to a series of subpopulations through which the relationship of the centroids is linear.

The higher central moments of ξ are also needed for work in this thesis, and so the notation is introduced now

$$\begin{aligned} \mu_{\xi 3} &= E[(\xi - \mu)^3] \\ \mu_{\xi 4} &= E[(\xi - \mu)^4]. \end{aligned}$$

The random error components, or errors for short are assumed to have zero means and variances that are independent of the suffix i

$$\begin{aligned} E[\delta_i] &= E[\varepsilon_i] = 0 \\ Var[\delta_i] &= \sigma_\delta^2 \\ Var[\varepsilon_i] &= \sigma_\varepsilon^2. \end{aligned}$$

The higher central moments of the errors are also assumed to exist

$$\begin{aligned} E[\delta_i^3] &= \mu_{\delta 3}, \ E[\delta_i^4] = \mu_{\delta 4} \\ E[\varepsilon_i^3] &= \mu_{\varepsilon 3}, \ E[\varepsilon_i^4] = \mu_{\varepsilon 4}. \end{aligned}$$

The errors are assumed to be mutually uncorrelated such that

$$E[\delta_i\delta_j] = 0, \quad E[\varepsilon_i\varepsilon_j] = 0, \text{ for all i} \neq j$$

$$E[\delta_i\varepsilon_j] = 0 \text{ for all i and j.}$$

It is possible to rewrite the model outlined above as

$$y_i = \alpha + \beta x_i + (\varepsilon_i - \beta\delta_i), \quad i = 1, \ldots, n$$

This highlights the difference between this problem and the standard regression model since the term $\varepsilon - \beta\delta$ is correlated with x. Indeed,

$$Cov[x, \varepsilon - \beta\delta] = E[x(\varepsilon - \beta\delta)] = E[(\xi + \delta)(\varepsilon - \beta\delta)] = -\beta\sigma_\delta^2$$

and is only zero if $\beta = 0$ or $\sigma_\delta^2 = 0$. If $\sigma_\delta^2 = 0$, the model is equivalent to standard y on x regression, and the usual results outlined earlier apply. In addition to this, the error term is clearly dependent on β.

There have been several reviews of errors in variables methods, notably Casella and Berger [16], Cheng and Van Ness [20], Fuller [41], Kendall and Stuart [67] and Sprent [97]. Unfortunately the notation has not been standardised. This thesis closely follows the notation set out by Cheng and Van Ness [20] but for convenience, it has been necessary to modify parts of their notation. All notation will be carefully introduced at the appropriate time. A list of all notation used is available towards the beginning of the thesis.

1.3 Outline of the Thesis and Comments

As stated by Casella and Berger [16], errors in variables modelling is fundamentally different from standard regression techniques. The model fitting and parameter estimation of an errors in variables model is notably different to fitting a simple linear regression model. The array of different methods that have been used to tackle the problem of errors in both variables are described in Chapter 2. In simple regression, the method of least squares and maximum likelihood are closely linked and furnish a unified structure to estimation. In the errors in variables situation it turns out that the method of maximum likelihood is only satisfactory when all random variables in the model ξ, δ and ε are Normally distributed (i.e. the Normal structural model). Then the method of maximum likelihood exactly coincides with the method of moments. Even then some additional information about the parameters, for example knowledge of the ratio of the error variances $\frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ (called λ) is needed. The likelihood method, and the reasons for its unsuitability is described in Chapter 5. There are comparisons with the method of least squares estimation, notably orthogonal regression and the λ known case. However least squares does not provide a framework that covers all possibilities. Fortunately the method of moments is a flexible alternative method of estimation that gives a range of possible estimating equations, each one suited to the exact circumstances in which a model is fitted. It is possible to extend the usual range of method of moment estimating equations by appealing to higher order moments. This is described fully in Chapter 3.

Asymptotic results about the variance covariance structure are as easily obtained as they are in maximum likelihood estimation, these results are given in full in Chapter 3. The delta method (see for example Cramer [28]) allows insights into the exact structure of the variance covariance matrices that the inversion of an information matrix needed in the maximum likelihood approach does not provide. Since the results given are asymptotic ones, it is necessary to establish some guidance on topics such as the minimum sample sizes that are needed for reliance to be placed on the estimators. This is done in Chapter 4 using simulations.

The key application of regression models is often prediction, not just the identification of the model. Here too there are profound differences between errors in variables models and simple linear regression. The distinction is that the variable ξ, is not measured directly, but instead is a latent variable. The measurement of ξ, x, differs from the latent value by an unknown error δ. As a consequence there are several predictions that might be of interest, such as the recovery of the latent data set $\{(\xi_i, \eta_i, i = 1, \ldots, n)\}$ or the average value of y given an x, $E[y|x]$. These turn out to be different, as described in Chapter 6, and the appropriate predictor to use in practise will depend on the circumstances of the investigation.

Just as prediction differs from the case of simple regression so does the notion of residuals. There are several possible definitions of a residual, but here the diagnostic checking of the model is complicated by a phenomenon that has only been briefly been described previously, which was called migration by Nix (pers. comm.), who seems to have been the first to identify the phenomenon. The effect of measurement error in the x measurement is to distort the scatter of data from the true line, but also to make the average value of y at any particular x follow a curve. The average value of y given x follows a straight line only for the Normal structural model. This phenomenon somewhat complicates the usual plot of deviations from the fitted model against x. The interpretation of residuals in the context of errors in variables models is discussed in Chapter 7.

So as to illustrate the practical application of the theory developed in this thesis

Chapter 8 contains a number of case studies based on data from a wide range of disciplines. These applications help in clarifying the way in which errors in variables models can be applied in a practical way, and illustrate that an approach that ignores the measurement errors in x is often fundamentally flawed. Throughout the investigation it has become clear that it is important to consider not just one standard model, as in the case in simple regression, but instead to consider carefully the specific application so as to settle on the correct model for the specific investigation.

Chapter 9 summarises the contents of this thesis, as well as describing some potential further work as a result of the investigations undertaken in this thesis.

Chapter 2

An Overview of Errors in Variables Modelling

2.1 Introductory Remarks

The literature on errors in variables modelling is scattered and wide ranging. It is the aim of this Chapter to bring together some of the main concepts developed to aid with errors in variables modelling, and highlight some similarities between the methods. It is impossible to discuss the entire wealth of literature on errors in variables modelling, and strict attention has been placed on a few key ideas and methods.

The discussion here begins with the historical development of linear errors in variables modelling, and progresses to discuss how some of the available computer packages with statistical capabilities, such as SAS, may be used to aid with fitting an errors in variables model.

2.2 Origins and Beginnings

The author first associated with the errors in variables problem was Adcock [1], [2]. In the late 1800's he considered how to make the sum of the squares of the errors at right angles to the line as small as possible. This enabled him to find what he felt to be the most probable position of the line. Using ideas from basic geometry, he showed that the errors in variables line must pass through the centroid of the data. However, Adcock's results were somewhat restrictive in that he only considered what is commonly referred to as orthogonal regression. Orthogonal regression minimises the orthogonal distances (as opposed to vertical or horizontal distances in standard linear regression) from the data points to the regression line. As will be shown in Chapter 3 this assumes that the error variances σ_δ^2 and σ_ε^2 are equal. Use of the orthogonal regression line has been questioned by some authors, notably Bland [9], on the grounds that if the scale of measurement of the line is changed, then a different line would be fitted. However this is only true if λ is not modified along with the scale of measurement. If λ is modified along with the scale of measurement, the same line is fitted.

Adcock's work was extended a year later by Kummel [70]. Instead of taking equal error variances, he assumed that the ratio $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ was known instead. This method of identifying a line has proved popular and will be mentioned in detail many times in this thesis. Kummel derived an estimate of the line which clearly showed the relation between his and Adcock's work. Kummel argued that his assumption of knowing λ was not unreasonable. He suggested that most experienced practitioners have sufficient knowledge of the error structure to agree a value for this ratio.

The idea of orthogonal regression was included in a book by Deming [33]. He noted that just as the orthogonal projections from the data to the regression line may be taken, so can any other projection. This would then take account of unequal error variances.

Figure 2.1 illustrates how this may be done. A least squares method can then be used to minimise the sum of squares of these oblique distances. Lindley [72] found that adding a weighting factor when minimising the sum of squares of the orthogonal projections, allowed one to minimise projections other than orthogonal. It should be pointed out that all these authors implicitly assumed that the error structure is homoscedastic, otherwise additional weighting factors to allow for the heteroscedasticity would have to be used. For example, a recent paper by Cheng and Riu [18] illustrates how some of the ideas presented in this literature survey may be applied to a model with heteroscedastic errors. They talked about the concept of equation error (discussed later), correlations in the error structure and heteroscedasticity.

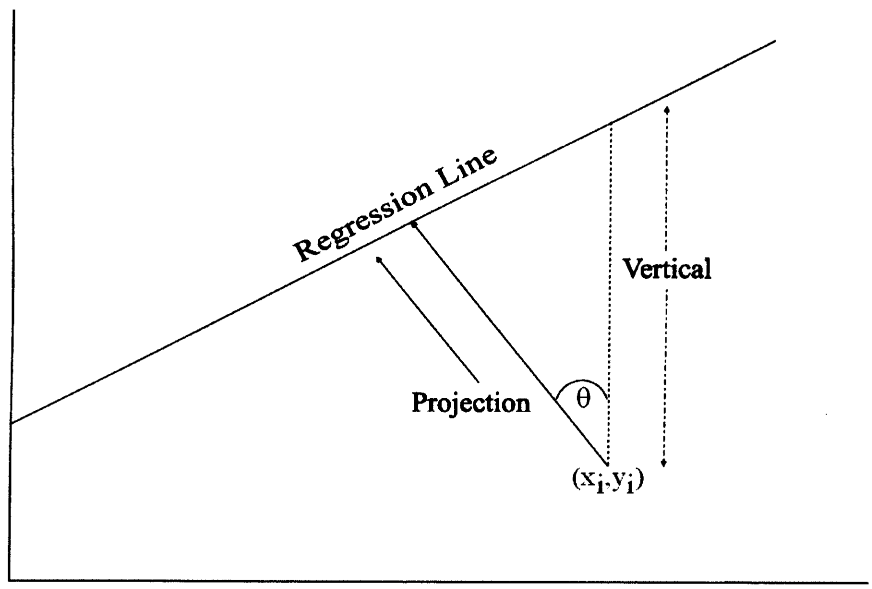


Figure 2.1: Deming's Regression

An early paper on modelling with errors in both variables was by Pearson [86]. He extended the ideas of previous authors to allow the fitting of lines and hyperplanes (when there is more than one predictor) of best fit. Pearson was able to show that the orthogonal regression line lies between the y on x, and x on y regression lines.

2.3 Grouping Methods

A different approach was suggested by Wald [110]. He described a method that did not make any parametric assumptions regarding the error structure. He stressed that there was no justification in making assumptions such as $\lambda = 1$, and that the regression line would not be invariant under transformations of the coordinate system (this criticism has been dealt with in the previous section). Wald suggested splitting the observations into two groups, G_1 and G_2, where G_1 contains the first half of the ordered observations $(x_{(1)}, y_{(1)}), \ldots, (x_{(m)}, y_{(m)})$ and G_2 contains the second half $(x_{(m+1)}, y_{(m+1)}), \ldots, (x_{(n)}, y_{(n)})$, the two halves being determined by the ordered x_i's . An estimator of the slope is then

$$\tilde{\beta}_W = \frac{(y_{(1)} + \ldots + y_{(m)}) - (y_{(m+1)} + \ldots + y_{(n)})}{(x_{(1)} + \ldots + x_{(m)}) - (x_{(m+1)} + \ldots + x_{(n)})}$$

A problem here is that for the estimator to be consistent the grouping should be based on the order of the true values, otherwise, in general, the groups are not independent of the error terms $\delta_1, \ldots, \delta_n$. Wald countered this by proving that, at least approximately, grouping with respect to the observed values is the same as grouping with respect to the true values. Properties of this estimator for finite samples, as well as approximations of the first four moments can be found in Gupta and Amanullah [54].

The idea of grouping the observations was further developed by Bartlett [6]. Instead of separating the ordered observed values into two groups, he suggested that greater efficiency would be obtained by separating the ordered observations into three groups, G_1, G_2 and G_3. G_1 and G_3 are the outer groups, and G_2 is the middle group. Nair and Banerjee [78] show that for a functional model, Bartlett's grouping method provided them with a more efficient estimator of the slope than Wald's method. In Bartlett's method the slope is found from a line through the points $(\bar{x_{G_1}}, \bar{y_{G_1}})$ and $(\bar{x_{G_3}}, \bar{y_{G_3}})$, where $(\bar{x_{G_1}}, \bar{y_{G_1}})$ and $(\bar{x_{G_3}}, \bar{y_{G_3}})$ are the mean points of the observations in G_1 and G_3 respectively. In effect, the observations in G_2 are not used after the data are grouped. Gibson and Jowett [47] offered advice on how to place the data into these three groups to obtain the most efficient estimator of the slope. How the data should be grouped depended on the distribution of ξ. A table summarising their results for a variety of distributions of ξ can be found in the review paper by Madansky [74].

Neyman and Scott [80] suggested another grouping method. The methodology they used is as follows. They suggested fixing two numbers, a and b such that $a \leqslant b$. The numbers a and b must be selected so $P[x \leqslant a] > 0$ and $P[x > b] > 0$. The observations x_i are then divided into three groups, G_1, G_2 and G_3. If $x_i \leqslant a$ those observations are put into G_1, if $a < x_i \leqslant b$ those observations are put into G_2, and if $x_i > b$ those observations are put into G_3. A further two numbers $-c$ and d are then found such that $P[-c \leqslant \delta \leqslant d] = 1$. An estimator of the slope is then given by

$$\tilde{\beta}_{NS} = \frac{\bar{y_{G_3}} - \bar{y_{G_1}}}{\bar{x_{G_3}} - \bar{x_{G_1}}}$$

and is a consistent estimator of β if

$$P[a - c < \xi \leqslant a + d] = P[b - c < \xi \leqslant b + d] = 0.$$

However, whether this condition is one that is obtainable in practice is open to debate.

Grouping methods, in particular Wald's method, have been critised by Pakes [82]. He claimed that the work of Gupta and Amanullah [54] is unnecessary as Wald's estimator is, strictly speaking, inconsistent. Letting $\tilde{\beta_W}$ denote Wald's estimator for the slope, Pakes showed

$$|p\lim \tilde{\beta}_W| = |\beta| \left| \frac{(\bar{x_{G_2}} - \bar{x_{G_1}})}{(\bar{x_{G_2}} - \bar{x_{G_1}}) + E[\delta|\, x \in G_2] - E[\delta|\, x \in G_1]} \right| < |\beta|,$$

which shows that, in general, Wald's estimator will underestimate the value of the true slope.

However, this expression derived by Pakes offers a similar conclusion to that of Neyman and Scott [79]. As long as the horizontal error δ is bounded (δ_i small in relation to the spacing $x_{(i+1)} - x_{(i)}$ for all i) so that the ranks of ξ are at least approximately equal to the ranks of x, then grouping methods should provide a respectable estimator for the slope as the expression $E[\delta|\, x \in G_2] - E[\delta|\, x \in G_1]$ should be negligible.

2.4 Instrumental Variables

Extensive consideration of this method has appeared in the econometrics literature. Essentially, the instrumental variables procedure involves finding a variable w that is correlated with x, but is uncorrelated with the random error component, δ. The estimator for the slope is then

$$\tilde{\beta}_{IV} = \frac{s_{yw}}{s_{xw}},$$

where, s_{yw} and s_{xw} are defined analogously to (1.2). In practice however, it is difficult to obtain a good instrumental variable which meets the aforementioned criteria.

The method of grouping can be put into the context of instrumental variables. Maddala [75] showed that Wald's grouping method is equivalent to using the instrumental variable

$$w_i = \begin{cases} 1 & \text{if } x_i > \text{median}(x_1, \dots, x_n) \\ -1 & \text{if } x_i < \text{median}(x_1, \dots, x_n) \end{cases}$$

and similarly Bartlett's grouping method is equivalent to using

$$w_i = \begin{cases} 1 & \text{for the largest } \frac{n}{3} \text{ observations} \\ -1 & \text{for the smallest } \frac{n}{3} \text{ observations} \\ 0 & \text{otherwise.} \end{cases}$$

2.5 Geometric Mean

Other than grouping the data, or looking for an instrumental variable, another approach is to simply take the geometric mean of the y on x regression line, and the reciprocal of the x on y regression line. This leads to the estimator

$$\tilde{\beta}_{GM} = sgn(s_{xy})\sqrt{\frac{s_{yy}}{s_{xx}}}.$$

There is a geometric interpretation of the line having this slope - it is the line giving the minimum sum of products of the horizonal and vertical distances of the observations from the line (Tessier [103]). However, for the estimate to be unbiased (see Jolicoeur [61] for example), one must assume that

$$\lambda = \beta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}. \tag{2.1}$$

This is due to

$$\tilde{\beta}_{GM} \longrightarrow \sqrt{\frac{\beta^2\sigma^2 + \sigma_\varepsilon^2}{\sigma^2 + \sigma_\delta^2}}.$$

This limit is equal to β if and only if $\lambda = \beta^2$.

It is also worth noting that with $\rho = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$

$$\frac{s_{xy}}{\rho s_{xx}} = \frac{s_{xy}}{s_{xx}}\frac{\sqrt{s_{xx}s_{yy}}}{s_{xy}} = \sqrt{\frac{s_{yy}}{s_{xx}}} = \tilde{\beta}_{GM}.$$

$\tilde{\beta}_{GM}$ is therefore the ordinary y on x slope estimator $\frac{s_{xy}}{s_{xx}}$, scaled by the correlation coefficient between x and y.

A technical criticism of the use of this estimator is that it may have infinite variance (Creasy [29]). This happens when the scatter of the observations is so great that it is difficult to determine if one line or another perpendicular to it should be used to represent the data. As a result, it may be difficult to construct confidence intervals of a respectable finite width. Geometric mean regression has received much attention, primarily in the fisheries literature. Ricker [87] examined a variety of regression methods applied to fish biology, and promoted the use of geometric mean regression. He claimed that in most situations it is superior to grouping methods, and the geometric mean regression line is certainly one of the easiest to fit. In addition, Ricker also warned that regression theory based on assuming that the data are from a Normal distribution may not apply to non-Normally distributed data. Great care must be taken by the statistician to ensure the proper conclusions are obtained from the data.

Jolicoeur [61], again in the fisheries literature, discussed the paper by Ricker. He stated that as geometric mean regression is equivalent to the assumption in equation (2.1) it is difficult to interpret the meaning of the slope, as the error variances σ_δ^2 and σ_ε^2 only contaminate and cannot explain the underlying relationship between ξ and η. Ricker replied to the paper by Jolicoeur in a letter, and claimed that the ratio (2.1) may not be linked to the presence or the strength of the underlying relationship, but the correlation coefficient will always give an idea as to the strength. Ricker reiterated that geometric mean regression is an intuitive approach, and as long as the assumption (2.1) holds, is a perfectly valid regression tool.

Further discussion on this estimator was initiated by Sprent and Dolby [99]. They discouraged the use of geometric mean regression, due to the unrealistic assumption of (2.1). They both however sympathised with practitioners, especially those in fish biology, who do not have any knowledge regarding λ and therefore would be unable to use the methods described in Section 2.2. In addition, they commented that the correlation coefficient might be misleading in an errors in variables model, due to each of the observations containing error. They did however suggest that a correlation coefficient may be useful in determining if a transformation to linearity has been successful.

2.6 Cumulants

Another method of estimation that has been used in errors in variables modelling is the method of moments. This will be described in the following section. A closely related approach to this is using cumulants, which were proposed by Geary [43], [44], [45], [46]. Cumulants can be defined as follows. Assume that X and Y are jointly distributed random variables. Then, provided the expansions are valid in the given domain, the natural logarithm of the joint characteristic function can be written as

$$\psi(t_1, t_2) = \ln[\phi(t_1, t_2)] = \ln[E(e^{it_1X + it_2Y})] = \sum_{r,s=0}^{\infty} \kappa(r, s) \frac{(it_1)^r}{r!} \frac{(it_2)^s}{s!} \qquad (2.2)$$

Here, ψ is the so-called joint cumulant generating function, and, if $r \neq 0$ and $s \neq 0$ then $\kappa(r, s)$ is called the r,s product cumulant of X and Y. The slope can be estimated via the method of cumulants as follows.

Assume that a structural errors in variables model has been selected. Then

$$\begin{aligned} x_i &= \xi_i + \delta_i \\ y_i &= \eta_i + \varepsilon_i \\ \eta_i &= \alpha + \beta\xi_i \end{aligned}$$

where the error laws quoted earlier in this thesis apply. If the true values ξ and η are centered with respect to their true mean, then the intercept vanishes, and we can write the structural relationship in the form

$$\beta\xi - \eta = 0 \tag{2.3}$$

Letting $\kappa_{(x,y)}$ denote the cumulants of (x, y), and $\kappa_{(\xi,\eta)}$ denote the cumulants of (ξ, η) we have

$$\kappa_{(x,y)}(r, s) = \kappa_{(\xi,\eta)}(r, s)$$

This follows from the following important properties of bivariate cumulants (see, for example Cheng and Van Ness [20], Pal [83])

- The cumulant of a sum of independent random variables is the sum of the cumulants.
- The bivariate cumulant of independent random variables is zero.

The joint characteristic function of (ξ, η) is

$$\phi(t_1, t_2) = E[e^{it_1\xi + it_2\eta}] \tag{2.4}$$

It follows from (2.3) and (2.4) that

$$\beta\frac{\partial\phi}{\partial it_1} - \frac{\partial\phi}{\partial it_2} = E[(\beta\xi - \eta)e^{it_1\xi + it_2\eta}] = 0$$

and if we replace the joint characteristic function ϕ by the cumulant generating function ψ we obtain

$$\beta\frac{\partial\psi}{\partial it_1} - \frac{\partial\psi}{\partial it_2} = \frac{1}{\phi}\left(\beta\frac{\partial\phi}{\partial it_1} - \frac{\partial\phi}{\partial it_2}\right) = 0 \tag{2.5}$$

and it follows from (2.2) and (2.5), for all $r, s > 0$

$$\beta\kappa(r+1, s) - \kappa(r, s+1) = 0$$

If $\kappa(r+1, s) \neq 0$ an estimator for the slope is then

$$\tilde{\beta}_C = \frac{\kappa(r, s+1)}{\kappa(r+1, s)}$$

In reality, the cumulants $\kappa(r, s)$ will have to be replaced by their sample equivalents $K(r, s)$. Details of how these sample cumulants may be computed as functions of sample moments are included in Geary [43].

2.7 Method of Moments

Instead of tackling the problem via cumulants, the method of moments can be used. Briefly, this is where a set of estimating equations are derived by equating population moments with their sample equivalents. The method of moments approach shall be considered in Chapter 3 of this thesis, and so only a brief survey of the existing literature is given here. Kendall and Stuart [67] derived the five first and second order moment equations for the structural errors in variables model. However, there are six parameters, $\mu, \alpha, \beta, \sigma_x^2, \sigma_\delta^2$ and σ_ε^2 for the structural model. So in order to proceed with the method of moments, some information regarding a parameter must be assumed known, or more estimating equations must be derived by going to the higher moments. Details on the various assumptions that can be made are included in Cheng and Van Ness [20], Dunn [39], and Kendall and Stuart [67], as well as others. Dunn [39] gave formulas for many of the estimators of the slope that are included in the next Chapter. However, he did not give any information regarding estimators based on higher moments. Neither did he give information about the variances of these estimates. A recent paper by Davidov [31] considered the Normal structural model and commented that the method of moment estimators are equal to the maximum likelihood estimators. Some large sample properties were also offered. Work on the higher order moment estimating equations has been done by Drion [38], and more recently by Pal [83], Van Montfort et al [108], Van Montfort [107] and Cragg [27].

Drion [38], in a paper that is infrequently cited, looked at an estimator that could be derived through the third order non central moment equations for a functional model. Drion computed the variances of all the sample moments that he used, and showed that his estimator of the slope is consistent. Prior to this work, Scott [90] considered the structural model, and also found an estimator based on the third moments. Scott was able to show that if the third central moment of ξ exists, and is non-zero, then the equation

$$F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \bar{y} - b(x_i - \bar{x})]^3 = 0$$

has a root $\hat{b}$ which is a consistent estimate of β. This is because the stochastic limit of $F_{n,1}(b)$ is $(\beta - b)^3\mu_{\xi 3}$, where $\mu_{\xi 3}$ denotes the third central moment of ξ. The estimate of the slope is then a function of the third order sample moments. Scott was able to generalise this result. If the random variable ξ has central moments up to and including order $2m+1$ and if at least one of the first m odd central moments $\mu_{\xi,2k+1}$

$(k = 1, 2, \ldots, m)$, differs from zero, then the equation

$$F_{n,m}(b) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \bar{y} - b(x_i - \bar{x})]^{2m+1} = 0$$

has a root $\hat{b}$ which is a consistent estimate of β. Scott did warn however, that estimators based on the lower order moments are likely to be more precise than those based on higher order moments. Unfortunately, Scott did not provide a method of extracting the root which would provide the consistent estimator.

More recently, Pal [83] further examined the possibilities of the moment equations in a structural model. He stated that in economics, the errors in variables situation cannot be ignored, and as a result, least squares estimation is the wrong way to proceed. Pal derived six possible estimators of the slope, but showed that three of these are functions of the other slope estimators, and concluded that there must be infinitely many consistent estimators which can be obtained by taking different functions of the slope estimators he derived. For each of the six estimators, Pal found their asymptotic variances when the error terms were assumed to follow a Normal distribution. He then went on to consider a variety of regression scenarios, such as $\frac{\sigma_\xi^2}{\sigma^2} \longrightarrow 0$, to offer advice as to which estimator has the smallest variance. The asymptotic efficiency of a particular estimator with respect to the least squares estimator was also provided, for different distributions of ξ. A brief review on the method of cumulants, and how errors in variables modelling might be extended to a multiple linear regression model was included towards the end of the paper.

Van Montfort et al [108] gave a detailed survey on estimators based on third order moments. They provided an optimal estimator of the slope which is a function of three slope estimators. In order to obtain this optimal estimator, the variance covariance matrix of the third order moments if not known, has to be estimated. By replacing the variance covariance matrix with its estimate, the optimal estimator is no longer a function of moments up to order three since moments of order higher than three appear in the estimation of the variance covariance matrix. Van Montfort et al, through a simulation study, demonstrated that the optimal estimator behaves well for a sample size of 50, and is superior to any other third moment estimator. The same study was replicated for a sample size of 25. For this sample size, they stated that the third moment estimators performed badly. A standard assumption is to assume that the errors δ and ε are independent. Van Montfort et al showed that even if δ and ε are linearly related, then their optimal estimator of the slope is still optimal for all consistent estimators of β which are functions of the first, second and third order moments. In addition, the asymptotic properties of the slope estimator are not altered.

A detailed account of alternative approaches to errors in variables modelling was written by Van Montfort [107]. This text included estimation based on third order moments, extensions to polynomial regressions, using characteristic functions and links to the factor analysis model. More details on the asymptotic variances and covariances of the third order moment slope estimators were provided. This text is an extension of the details included in the paper by Van Montfort et al [108].

The most recent account of the use of using higher moments was that by Cragg [27]. He extended the work on the moment equations to include those of the fourth order. A problem with moment based estimators however, is stability. It is well known that as the order of the moment increases they become progressively more difficult to estimate and larger sample sizes are needed to obtain a reliable estimate. Indeed, a paper by

Kagan and Nagaev [64] showed under general conditions that the order of a population moment that can be estimated by the corresponding sample moment in a sample is roughly $\frac{\ln(n)}{2\ln[\ln(n)]}$ and this order is extremely sharp. They offer the warning

> "one should be very careful in using too many sample moments even when the sample size is rather large"

Cragg applied a minimum χ^2 approach to the second, third and fourth moments in order to obtain an efficient general moment estimator. This approach again involves finding an estimated variance covariance matrix of the moments. As Cragg noted, this may be difficult as it will involve the eighth order moments. He suggested avoiding this problem by replacing the variance covariance matrix with some weighting matrix. This will result in less asymptotic efficiency however. In his simulations Cragg used a diagonal weighting matrix with elements $\frac{1}{2}, \frac{1}{15}$ and $\frac{1}{96}$ depending whether the moment equations are based on the second, third or fourth moments respectively. This may be deemed inappropriate as these values correspond to the theoretical variances of the second, third and fourth powers of a Normally distributed variable with zero mean and unit variance, even though a Normal distribution will not be applicable for every structural model.

A somewhat different use of the method of moments was suggested by Dagenais and Dagenais [30]. They proposed a consistent instrumental variable estimator for the errors in variables model based on higher moments. In addition, they showed how a regression model may be tested to detect the presence of errors in both variables. Dagenais and Dagenais illustrated their ideas through a number of numerical simulations and showed that their estimator is superior to the ordinary least squares estimator.

An alternative way of using the method of moments method was presented in a recent paper by Woodhouse [112]. By standardising the data, he presented chart solutions to assist users to readily find estimators of the slope β and gave advice on how the slope may be converted to the original unstandardised data. To assist in the explanation of this method, he also provided a detailed illustration. Woodhouse also commented on the wide ranging applications of errors in variables modelling, from laboratory use, method comparison studies and to make estimates of the constants associated with scientific laws.

2.8 Equation Error

Some authors have stressed the importance of a concept known as equation error. Further details are given by Fuller [41] and Carroll and Ruppert [13]. Equation error introduces an extra term ω_i to each y_i

$$y_i = \eta_i + \omega_i + \varepsilon_i = \alpha + \beta\xi_i + \omega_i + \varepsilon_i$$

Dunn [39] described the additional error term ω_i as

> "(a) new random component (that) is not necessarily a measurement error but is part of y that is not related to the construct or characteristic being measured."

Despite its name, equation error is not intended to model a mistake in the choice of equation used in describing the underlying relationship between ξ and η. Assuming that the equation error terms have a variance σ^2_ω that does not change with the suffix i, and that they are uncorrelated with the other random variables in the model, the practical effect of the inclusion of the extra term is to increase the apparent variance of y by the addition of σ_ω^2. The impact of equation error upon the estimation and fitting of an errors in variables model is discussed in the next Chapter.

2.9 Maximum Likelihood

The vast majority of the papers available on errors in variables modelling have adopted a maximum likelihood approach to estimate the parameters. Only a selection of the large number of papers shall be mentioned here. These papers assumed that either the Normal functional or Normal structural model applied. Lindley [72] was one of the first authors to use maximum likelihood estimation for the errors in variables model. Lindley commented that the likelihood equations are not consistent, unless there is some prior information available on the parameters. He suggested that the most convenient assumption to make is to assume that the ratio λ is known. Estimates of all the relevant parameters are then derived and discussed.

Kendall and Stuart [67] reviewed the topic of estimation in an errors in variables model, but concentrated their efforts on the maximum likelihood principle. They commented that the sample means, variances and covariances form sufficient statistics for a bivariate Normal distribution. As a result, the solutions of the method of moment estimating equations for the unknown parameters $\mu, \alpha, \beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2 are also maximum likelihood solutions, provided that these solutions give admissible estimates (namely, positive estimators for the variances in the model). The conditions to obtain admissible estimates are then outlined. Further details on these conditions, and estimation using the method of moment estimating equations is included in Chapter 3. The essential difficulty is that of having five moment estimating equations, and six parameters to estimate. Kendall and Stuart suggested various 'cases', each which consist of a different assumption regarding a subset of the parameters. Estimators for the parameters are derived for each of these 'cases', and advice is given on how to construct confidence intervals. A brief survey on cumulants, instrumental variables and grouping methods was also included in their work.

A disadvantage of the likelihood method in the errors in variables problem is that it is only tractable if all the distributions describing variation in the data are assumed to be Normal. In this case a unique solution is only possible if additional assumptions are made concerning the parameters of the model, usually assumptions about the error variances. The likelihood approach where the distribution assumed for ξ is different from Normal is touched upon in Chapter 5, where the difficulties of the approach are outlined. Nevertheless, maximum likelihood estimators have certain optimal properties and it is possible to work out the asymptotic variance covariance matrix of the estimators. These were given for a range of assumptions about the error structure but for the case when ξ is Normally distributed by Hood et al [57]. In addition, Hood et al conducted a simulation study in order to determine a threshold sample size to successfully estimate their variance covariance matrix. They concluded that this threshold was approximately 50.

Other papers on the likelihood approach have tended to focus on a particular aspect of the problem. For example, Wong [111] considered the likelihood equations when the error variances were assumed to be known, and equal. This case has attracted much attention, as if both error variances are known, the problem is overidentified - there are four parameters to be estimated from five estimating equations (be it likelihood equations, or moment equations). To simplify the procedure, Wong used an orthogonal parameterisation in which the slope parameter is orthogonal to the remaining parameters. Approximate confidence intervals for the parameters, information on testing hypotheses regarding the slope, and the density function for the slope are also included. Prior to this, Barnett [5] also commented on the inherent difficulties in using the maximum likelihood technique.

Again for the structural model Birch [7] showed that the maximum likelihood estimator for the slope is the same when both error variances are known, and when the ratio of the error variances λ is known. He also commented that the maximum likelihood estimators provided by Madansky [74] are inconsistent, and as a result need to be modified. Some discussion on the admissability conditions was also included.

A key author in this area was Barnett [5]. His paper on the fitting of a functional model with replications commented on the importance of errors in variables modelling in the medical and biological areas. The paper adopted the maximum likelihood technique for estimating the parameters, but no closed form solution could be found. He mentioned that the maximum likelihood method tends to run into computational problems due to the awkward nature of the likelihood equations. Barnett also considered alternative error structures which might be applicable to biological and medical areas.

Most papers concern themselves with homoscedastic errors. Chan and Mak [17] looked at heteroscedastic errors in a linear functional relationship. To find the estimators for the parameters in the model they employed a numerical method to solve a set of non-linear equations iteratively. The asymptotic behaviour of the estimators were considered and an approximate asymptotic variance covariance matrix was found. A procedure for consistently estimating this variance covariance matrix was outlined.

Solari [95] found that the maximum likelihood solution for the linear functional model discussed by many authors was actually a saddle point, and not a maximum. She said that although the point was purely academic, it was still one worth making. A detailed analysis of the form of the likelihood surface was given, and and she concluded that a maximum likelihood solution for the linear functional model does not exist, unless one has some prior distribution to place on a parameter. Solari commented that this problem might appear in other estimation problems. Detailed consideration must be given to see if the maximum likelihood solution is indeed a maximum. Sprent [98] considered Solari's work and further noted the practical implications of her findings.

Copas [23] extended the work of Solari [95]. He showed that when errors made when rounding the observations are considered, then the likelihood surface becomes bounded. This allows for a different consideration of the likelihood surface. An estimate for the model can be found, which is approximately maximum likelihood. In other words, a point close to the global supremum was used instead. Copas' solution for the slope is equivalent to using either the x on y estimate or the y on x estimate. The y on x regression estimate is used if the line corresponding to the geometric mean estimate lies within 45^o of the x-axis. The x on y estimate is used if the geometric mean estimate lies within 45^o of the y-axis. A numerical example was provided to illustrate his suggested methodology, and the likelihood surface for this example was drawn.

Essentially, Copas introduced a modified likelihood function

$$L = \prod_i P_i(x_i)Q_i(y_i) \tag{2.6}$$

where $P_i(x) = P\left(x - \frac{h}{2} \leq \xi_i < x + \frac{h}{2}\right)$ and $Q_i(x) = P\left(y - \frac{h}{2} \leq \beta\xi_i < y + \frac{h}{2}\right)$ (note that Copas' model did not include an intercept). The value h was introduced to allow a discrepancy when $(\xi_i, \beta\xi_i)$ were recorded or measured. The saddle point noted by Solari; according to Copas, is a direct consequence of the likelihood function having singularities at all points within the sets

$$A = \{\beta, \sigma_\delta, \sigma_\varepsilon, \underline{\xi} : \sum(x_i - \xi_i)^2 = 0, \sigma_\delta = 0\}$$

and

$$B = \{\beta, \sigma_\delta, \sigma_\varepsilon, \underline{\xi} : \sum(y_i - \beta\xi_i)^2 = 0, \sigma_\varepsilon = 0\}$$

Copas showed that within these sets A and B his modified likelihood function reduces to the likelihood function for y on x regression and x on y regression respectively. This however is to be expected as set A essentially assumes that there is no horizontal error (δ) present and set B essentially assumes that there is no vertical error (ε) present. In addition, Copas' analyses assume that h is small, which will also imply that the simple linear regression techniques outlined at the front of this thesis are appropriate.

In summary Copas' method is equivalent to using y on x regression if it appears that ξ_i is close to x_i, and x on y regression if $\beta\xi_i$ is close to y_i. The choice of which regression to use depends on the location of the geometric mean regression line. Copas admitted that the y on x and x on y regression estimators do not maximise his likelihood function L. So, as it is well known that y on x and x on y regression are biased, and can only offer a crude approximation to the true line, the method proposed by Copas must be questioned.

An interesting modification of the structural model is the ultrastructural model. Cheng and Van Ness [19] considered this model with no replication. They showed that if one of the error variances are known, the maximum likelihood estimators are not consistent, whilst the method of moments estimators are. Much work on this model was carried out by Dolby [36]. He wrote on the linear functional and structural models, constructing a model which he called a synthesis of the functional and structural relations. Dolby [35] also discussed the linear structural model, giving an alternative derivation of Birch's [7] maximum likelihood solution. Yet another paper which adopts a maximum likelihood approach was that by Cox [26]. He wrote about the linear structural model for several groups of data, in other words, the ultrastructural model. He also provided a method to test various hypotheses regarding the model, and offered an example using head length and breadth measurements.

2.10 Confidence Intervals

Confidence intervals are beyond the scope of this thesis, and only a brief description is given here. Creasy [29] constructed confidence intervals for Lindley's [72] estimate of the slope. Patefield [84] extended her work and showed that her results can be applied to other errors in variables models. On the other hand, Gleser and Hwang [51] claimed that for the majority of linear errors in variables models it is impossible to obtain confidence intervals of finite width for certain parameters. Gleser has been active in writing about errors in variables models. With a number of coauthors, he has written on various aspects of the model. These include the unreplicated ultrastructural model

[49], the limiting distribution of least squares estimates [50], and estimating models with an unknown variance covariance matrix [52].

2.11 SIMEX

SIMEX is the method of Simulation-Extrapolation developed by Cook and Stefanski [101], and makes use of the fact that the standard y on x slope estimator is a biased estimator of the slope of an errors in variables model. For the straight line model, as will be restated in the next Chapter,

$$E\left[\frac{s_{xy}}{s_{xx}}\right] = \left(\frac{\sigma^2}{\sigma^2+\sigma_\delta^2}\right)\beta.$$

This result was derived by Fuller [41]. The SIMEX method works by computing the standard y on x estimator for a number of simulated data sets. In order to use SIMEX, σ_δ^2 must be known. Note that SIMEX is not constrained to straight line models, it may also be applied to multivariate errors in variables models, as well as nonlinear errors in variables models (see for example James [59]). The steps behind this method are illustrated here. The pseudo-code needed to implement this procedure was provided by James [59].

The SIMEX method involves adding increasing amounts of error to the x observations. A standard y on x fit is made to each data set, paying attention to the resulting change in bias. So for some chosen value of $\tau \geq 0$ a new set of observations is calculated as

$$x_{\tau_i} = x_i + \sigma_\delta\sqrt{\tau}Z_i \tag{2.7}$$

where each Z_i is an independently and identically distributed standard Normal random variable.

Carroll et al. [15] recommend that $0 \leq \tau \leq 2$ and about five values in this range are chosen for analysis. Equation (2.7) implies that

$$Var[x_{\tau_i}] = \sigma^2 + (1+\tau)\sigma_\delta^2.$$

By using standard regression methods, a relationship between τ and the average values of β corresponding to each τ may be found. Back-extrapolating to the case $\tau = -1$ then yields the SIMEX slope estimator.

For our straight line model, Carroll et al. [15] show that the regression function to back-extrapolate is of the form

$$\beta = c_1 + \frac{c_2}{c_3 + \tau}$$

where c_1, c_2 and c_3 are constants to be estimated. Carroll et al. show that for the straight line model SIMEX produces the same estimates as the method of moments. The method of moments is discussed in detail in Chapter 3.

Gleser [52] also described a method of estimating the unknown slope by shrinking the observed x_i towards the mean to adjust, on average, for measurement error. Then ordinary least squares regression can be used to obtain an estimator for β that he showed is consistent. This method is discussed later in this thesis.

2.12 Total Least Squares

Total least squares is a method of estimating the parameters of a general linear errors in variables model and was introduced by Golub and Van Loan [53], which is frequently cited in the computational mathematics and engineering literature. Broadly speaking, total least squares may be viewed as an optimisation problem with an appropriate cost function. The standard formulation of the total least squares problem is as follows. Consider a linear measurement error model

$$\mathbf{AX} \approx \mathbf{B}$$

where $\mathbf{A} = \mathbf{A}_0 + \tilde{\mathbf{A}}$ and $\mathbf{B} = \mathbf{B}_0 + \tilde{\mathbf{B}}$. It is assumed that the underlying physical relationship $\mathbf{A}_0\mathbf{X}_0 = \mathbf{B}_0$ exists.

In total least squares estimation, a matrix $\mathbf{D} = [\mathbf{AB}]$ is constructed which contains the measured data, and the parameter matrix $\mathbf{X}$ is to be estimated. There is an assumption that there exists a true unknown value of the data $\mathbf{D}_0 = [\mathbf{A}_0\mathbf{B}_0]$ and a true value of the parameters $\mathbf{X}_0$ such that $\mathbf{A}_0\mathbf{X}_0 = \mathbf{B}_0$. However, the measured data $\mathbf{D}$ depends on some additive error $\tilde{\mathbf{D}} = [\tilde{\mathbf{A}}\tilde{\mathbf{B}}]$ so that $\mathbf{D} = \mathbf{D}_0 + \tilde{\mathbf{D}}$.

The ordinary least squares method gives a solution $\mathbf{X}$ such that the Euclidean norm $||\mathbf{AX} - \mathbf{B}||$ is minimised. The total least squares technique applies a small correction (measured by the Euclidean norm) $\Delta\mathbf{D} = [\Delta\mathbf{A}\Delta\mathbf{B}]$ to the matrix $\mathbf{D}$ such that the equations $(\mathbf{A} + \Delta\mathbf{A})\mathbf{X} = \mathbf{B} + \Delta\mathbf{B}$ are readily solved. Solutions for this system of equations are obtained by computing its singular value decomposition, and this is the precise topic of the paper by Golub and Van Loan [53] mentioned earlier.

The total least squares methodology has been extended to generalised total least squares (where the errors are allowed to be correlated), and more recently element-wise total least squares (which deals with non-identically distributed errors). For a brief review of total least squares and its related methods, see for example Markovsky and Van Huffel [76]. A complete monograph on the topic has been written by Van Huffel and Vandewalle [106]. Cheng and Van Ness [20] noted that total least squares is in its most simple version, orthogonal regression. Hence, this methodology may not be appropriate when there is some different information available on a parameter.

2.13 Structural Equation Modelling

Structural equation modelling (sometimes referred to as covariance structure analysis) is the broad name given to the modelling of a structure specified by a system of equations. These equations specify phenomena in terms of cause and effect variables, and in their most general form can deal with unobservable, latent variables. Johnson and Wichern [60] comment that structural equation models have been successfully applied in the behavioural and social sciences in modelling such latent variables as social status and discrimination in employment. The most common parameterisation for a structural equation model has become to be known as LISREL (Linear Structural Relationships) (see for example Skrondal and Rabe-Hesketh [94]). A computer package has been developed to fit such models, and thus further details are placed in the next section.

2.14 Computer Aided Methods

There are presently a number of computer packages which aid with errors in variables modelling. This section will describe two of these available packages, namely LISREL and SAS.

2.14.1 LISREL

LISREL is an example of a structural equation model, and computer software to implement such a model was created by Joreskog and Sorbom (see for example [62]). To use their notation, the LISREL model is formulated as follows:

$$\underline{\eta} = \mathbf{B}\underline{\eta} + \Gamma\underline{\xi} + \underline{\zeta} \quad (2.8)$$

$$\underline{Y} = \Lambda_y\underline{\eta} + \underline{\varepsilon} \quad (2.9)$$

$$\underline{X} = \Lambda_x\underline{\xi} + \underline{\delta} \quad (2.10)$$

where $\underline{\eta}$ is a $(m \times 1)$ vector, $\mathbf{B}$ is a square $(m \times m)$ matrix, Γ is a $(m \times n)$ matrix, $\underline{\xi}$ is a $(n \times 1)$ vector, $\underline{\zeta}$ is a $(m \times 1)$ vector, $\underline{Y}$ is a $(p \times 1)$ vector, Λ_y is a $(p \times m)$ matrix, $\underline{\varepsilon}$ is a $(p \times 1)$ vector, $\underline{X}$ is a $(q \times 1)$ vector, Λ_x is a $(q \times n)$ matrix, and $\underline{\delta}$ is a $(q \times 1)$ vector. At a first glance, the LISREL model resembles a combination of two factor analysis models, (2.9) and (2.10) into the structural setting of equation (2.8).

The matrix $\mathbf{B}$ is introduced to allow inter-relations between the latent variables of the model to be formed. Similarly, Γ, Λ_x and Λ_y are matrices which contain loadings for the relevant latent variables in the model. $\underline{\delta}$ and $\underline{\varepsilon}$ are the measurement errors according to $\underline{X}$ and $\underline{Y}$ respectively, with $\underline{\zeta}$ representing equation error as discussed earlier.

Our errors in variables model outlined in Section 1.2 may be fitted into a LISREL format as follows. Take $m = n = p = q = 1$, $\mathbf{B} = 0$, $\underline{\zeta} = 0$, $\Gamma = \beta$ and $\Lambda_x = \Lambda_y = 1$. The standard assumption of the LISREL model is to take $E[\underline{\xi}] = E[\underline{\eta}] = 0$. This constrains us to take $\mu = \alpha = 0$ for our model in Section 1.2. The remaining parameters to be estimated are $\beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2.

A LISREL model usually cannot be solved explicitly, and in this scenario an iterative procedure to estimate the parameters is adopted. Essentially, this involves constructing a set of estimating equations for the parameters. The usual methodology is to set the sample variance covariance matrix equal to the theoretical variance covariance matrix. The elements of the theoretical variance covariance matrix are nonlinear functions of the model parameters Λ_x, Λ_y, Γ and the variance covariance matrices of $\underline{\xi}$, $\underline{\zeta}$, $\underline{\delta}$ and $\underline{\varepsilon}$.

The LISREL model, (as in factor analysis), implies a particular structure for the theoretical variance covariance matrix. Johnson and Wichern [60] gave details of the structure, and stated the following identities (they took $\mathbf{B} = 0$ to simplify proceedings)

$$\begin{aligned} E[\underline{Y}\underline{Y}^T] &= \Lambda_y(\Gamma\Phi\Gamma^T + \psi)\Lambda_y^T + \Theta_\varepsilon \\ E[\underline{X}\underline{X}^T] &= \Lambda_x\Phi\Lambda_x^T + \Theta_\delta \\ E[\underline{X}\underline{Y}^T] &= \Lambda_y\Gamma\Phi\Lambda_x^T \end{aligned}$$

where $E[\underline{\xi}\underline{\xi}^T] = \Phi, E[\underline{\delta}\underline{\delta}^T] = \Theta_\delta, E[\underline{\varepsilon}\underline{\varepsilon}^T] = \Theta_\varepsilon$ and $E[\underline{\zeta}\underline{\zeta}^T] = \psi$. It is assumed that the variables ζ, δ and ε are mutually uncorrelated. Also ζ is uncorrelated with ξ, ε is uncorrelated with η and δ is uncorrelated with ξ.

The iteration procedure mentioned above begins with some initial parameter estimates, to produce the theoretical variance covariance matrix which approximates the sample theoretical variance covariance matrix. However, for this estimation procedure to occur, there must be at least as many estimating equations as parameters. Indeed, Johnson and Wichern [60] state that if t is the number of unknown parameters then the condition

$$t \leq \frac{1}{2}(p+q)(p+q+1)$$

must apply to allow estimation of the parameters. For our model of Section 1.2, $t = 4$ (β, σ^2, σ_δ^2 and σ_ε^2) and $\frac{1}{2}(p+q)(p+q+1) = 3$ and so we cannot use the LISREL approach to estimate our parameters unless we assume something further known. This ties in with the thoughts of Madansky [74] who stated that

> "To use standard statistical techniques of estimation to estimate β, one needs additional information about the variance of the estimators."

Also, comparisons may be drawn between LISREL, the method of moments and maximum likelihood, as both of the latter methods also assume that there is some parameter known to allow identifiability of the model.

Applying the LISREL methodology to our model of Section 1.2, we get

$$\begin{aligned} E[\underline{Y}\underline{Y}^T] &= \beta^2\sigma^2 + \sigma_\varepsilon^2 \\ E[\underline{X}\underline{X}^T] &= \sigma^2 + \sigma_\delta^2 \\ E[\underline{X}\underline{Y}^T] &= \beta\sigma^2 \end{aligned}$$

since for our model $\Phi = \sigma^2$, $\psi = 0$, $\Theta_\delta = \sigma_\delta^2$ and $\Theta_\varepsilon = \sigma_\varepsilon^2$. We can now equate the theoretical variance covariance matrix to the sample variance covariance matrix to construct the following three equations

$$\sigma^2 + \sigma_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s_{xx} \tag{2.11}$$

$$\beta^2\sigma^2 + \sigma_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s_{yy} \tag{2.12}$$

$$\beta\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = s_{xy} \tag{2.13}$$

which are identical to the method of moment estimating equations (and subsequently the maximum likelihood estimating equations) (3.3), (3.4) and (3.5) outlined in

Chapter 3.

The first order moment equations $\mu = \bar{x}$ and $\alpha + \beta\mu = \bar{y}$ are missing as the LISREL model assumes the data are centered, so μ and α are taken as known in the assumption $E[\underline{\xi}] = E[\underline{\eta}] = 0$. There are three equations (2.11), (2.12), (2.13) and four parameters to be estimated. Hence, in order to solve these equations explicitly we need to restrict the parameter space by assuming something known (e.g. assume σ_δ^2 known). So LISREL for our model is identical to the method of moments, and thus maximum likelihood. As stated earlier, the method of moments is discussed in Chapter 3.

2.14.2 SAS

Details of how to use SAS procedure NLMIXED to fit linear and nonlinear structural errors in variables models were provided by Patefield [85]. His paper described the methodology behind the procedure, as well as examples as to its implementation. The theory behind the procedure NLMIXED is based upon that of fitting the general linear latent variable model by maximum likelihood.

In a general nonlinear structural model a number of response variables $x_1, \ldots, x_k$ are defined by a smaller number of hidden, latent variables $\xi_1, \ldots, \xi_r$. A sample of n observations is taken, with $\underline{x_i}$ being the vector of observations on $x_1, \ldots, x_k$ and $\underline{\xi_i}$ is the corresponding vector of unobserved latent variables.

Letting $\chi = (\phi, \theta)$ be a vector of parameters and $i = 1, \ldots, n$, Patefield states that the key components to this model are:

1. The conditional distribution of $\underline{y_i}$ given $\underline{\xi_i}$, with probability density function $p_i(\underline{y_i}|\phi, \underline{\xi_i})$.

2. The distribution of the latent variables $\underline{\xi_i}$ with probability density function $q_i(\underline{\xi_i}|\theta)$.

Then the marginal likelihood is the joint distribution of the data taken as a function of χ given by

$$l(\chi) = \prod_{i=1}^{n} \int p_i(\underline{y_i}|\phi, \underline{\xi_i}) q_i(\underline{\xi_i}|\theta)\, d\underline{\xi_i}.$$

The procedure NLMIXED numerically maximises this marginal likelihood using quadrature and an iterative numerical method to give maximum likelihood estimators of χ. However, NLMIXED assumes that the latent variables follow a Normal distribution. The iterative procedure requires starting values for the parameters to be estimated. Patefield recommends that good starting values are found to save on computation time, and to help avoid the problem of $l(\chi)$ having multiple local maxima.

To demonstrate how a linear model may be fitted in SAS, Patefield considered a bivariate data set taken from Fuller [41] of the average number of hen pheasants sighted in August and Spring in Iowa from 1962 to 1976. Fuller decided to model this using the Normal structural model, taking $\lambda = \frac{1}{6}$. As the latent variables are assumed to be Normally distributed then the SAS procedure NLMIXED may be used.

Before the NLMIXED procedure is applied to the data, it has to be manipulated so it takes the structure of a mixed model. All data has to be concatenated into a single response vector of length $2n$. Then for each element of this single response vector two things must be specified:

- The subject classification variable. For the example given in this section, year is suitable.
- A value of indicator variables d_1 and d_2 where $d_1 = 1$ and $d_2 = 0$ for the x values and $d_1 = 0$ and $d_2 = 1$ for the y values.

The procedure NLMIXED may then be implemented. The code Patefield used to implement such a model is included here:

```
proc NLMIXED gconv=1e-9 cov;
parms alpha=0 beta=1 meanxi=0 varxi=1 vare=1;
bounds varxi,vare>=0;
muy=d1*(alpha+beta*xi)+d2*xi;
model y~normal(muy,(d1/6+d2)*vare);
random xi~normal(meanxi,varxi) subject=year;
run;
```

The `gconv` option controls convergence based on the gradients of the log-likelihood. Its default value is 10^{-8}. The `cov` option gives the variance covariance matrix of the estimated parameters as part of its output. The **bounds** statement allows restrictions on the parameters to be set. The obvious constraints here are for the variances to be non-negative.

Patefield commented that the maximum likelihood estimates produced by SAS are the same as those produced by the formulae of Hood et al. [57]. As commented earlier, the maximum likelihood estimators are identical to the method of moments estimators. So for this example SAS produces the same answers as the method of moment estimating equations. The variance covariance matrix outputted by SAS also agrees with the variance covariance matrices derived by Hood et al.

2.15 Review Papers and Monographs

Over the years several authors have written review articles on errors in variables regression. These include Kendall [65], [66], Durbin [40], Madansky [74], Moran [77] and Anderson [3]. Riggs et al [88] performed simulation exercises comparing some of the slope estimators that have been described in the literature. There are two texts devoted entirely to the errors in variables regression problem, Fuller [41] and Cheng and Van Ness [20]. Casella and Berger [16] has an informative section on the topic, Sprent [97] contains Chapters on the problem, as do Kendall and Stuart [67] and Dunn [39]. Draper and Smith [37] on the other hand, in their book on regression analysis, devoted only 7 out of a total of almost 700 pages to errors in variables regression. The problem is more frequently described in econometrics texts, for example Judge et al [63]. In these texts the method of instrumental variables is often given prominence. Carroll et al [14] described errors in variables models for non linear regression, and Seber and Wild [92] included a Chapter on this topic.

Chapter 3

The Method of Moments and the Linear Structural Model

3.1 Introductory Remarks

The method of moments technique is described in many books of mathematical statistics, for example Casella and Berger [16] and DeGroot [32], although here, as elsewhere the treatment is brief. In common with many other mathematical statistical texts, they gave greater attention to the method of maximum likelihood. Bowman and Shenton [10] wrote that

> "the method of moments has a long history, involves an enormous literature, has been through periods of severe turmoil associated with its sampling properties compared to other estimation procedures, yet survives as an effective tool, easily implemented and of wide generality".

In the method of moments estimating equations are derived by equating sample moments to their population equivalents. The population moments are functions of the parameters of the model, so the estimating equations are solved to give estimators of the unknown parameters of the model yielding the so called method of moments estimators. Use of this method of moments has been criticised because method of moments estimators are not uniquely defined. The population moments are all functions of the unknown parameters and, as long as the moments exist, any moment could be used to derive an estimating equation. Thus if this method is used it may be necessary to choose amongst possible estimators to find ones that best suit the data being analysed. This proves to be the case in errors in variables regression theory. Nevertheless the method of moments has the advantage of simplicity, and also that the only assumptions that have to be made are that low order moments of the random variable used as a model for the population exist.

It is relatively easy to work out the theoretical asymptotic variances and covariances of the estimators by a method outlined by Cramer [28]. Cramer's methodology shall be outlined in more detail later. Indeed, after making particular distributional assumptions, the method of moments enables a practitioner to fit the line and calculate approximate confidence intervals for the associated parameters. Approximate significance tests can also be done. A limitation of the formulae is that they are asymptotic results, so they should only be used for moderate or large data sets.

3.2 Restricting the Parameter Space

Consider the structural model outlined in Chapter 1. The method of moments estimating equations follow from equating population moments to their sample equivalents. By using the properties of ξ, δ and ε detailed in Chapter 1, the population moments can be written in terms of parameters of the model. This was also done by Kendall and Stuart [67], and have been repeated by Cheng and Van Ness [20], [39] amongst others.

$$\begin{aligned} E[x] &= E[\xi] = \mu \\ E[y] &= E[\eta] = \alpha + \beta\mu \\ Var[x] &= Var[\xi] + Var[\delta] = \sigma^2 + \sigma_\delta^2 \\ Var[y] &= Var[\alpha + \beta\xi] + Var[\varepsilon] = \beta^2\sigma^2 + \sigma_\varepsilon^2 \\ Cov[x, y] &= Cov[\xi, \alpha + \beta\xi] = \beta\sigma^2 \end{aligned}$$

The method of moments estimating equations are now found by equating the population moments to their sample equivalents

$$\bar{x} = \tilde{\mu} \tag{3.1}$$

$$\bar{y} = \tilde{\alpha} + \tilde{\beta}\tilde{\mu} \tag{3.2}$$

$$s_{xx} = \tilde{\sigma}^2 + \tilde{\sigma}_\delta^2 \tag{3.3}$$

$$s_{yy} = \tilde{\beta}^2\tilde{\sigma}^2 + \tilde{\sigma}_\varepsilon^2 \tag{3.4}$$

$$s_{xy} = \tilde{\beta}\tilde{\sigma}^2 \tag{3.5}$$

Here a tilde is placed over the symbol for a parameter to denote a method of moments estimator. From equations (3.3), (3.4) and (3.5) it can be seen that there is a hyperbolic relationship between the method of moments estimators for σ_δ^2 and σ_ε^2. This was called the Frisch hyperbola by van Montfort [107].

$$(s_{xx} - \tilde{\sigma}_\delta^2)(s_{yy} - \tilde{\sigma}_\varepsilon^2) = (s_{xy})^2 \tag{3.6}$$

This is a useful equation as it relates pairs of estimates $(\tilde{\sigma}_\delta^2, \tilde{\sigma}_\varepsilon^2)$ to the data in question. In point of fact equations for any pair of parameters can be derived, such as $s_{yy} = \tilde{\beta}s_{xy} + \tilde{\sigma}_\varepsilon^2$.

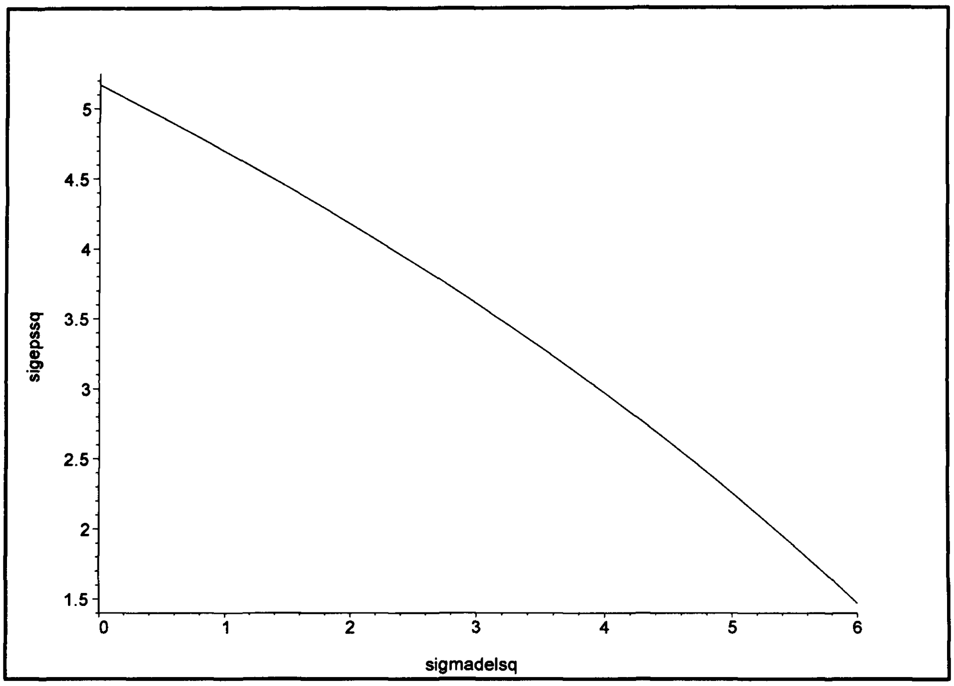


Figure 3.1: An example of a Frisch hyperbola

One of the main problems in fitting an errors in variables model using the method of moments is that of identifiability. It can be seen from equations (3.1), (3.2), (3.3), (3.4) and (3.5) that a unique solution cannot be found for the parameters since there are five equations, but six unknown parameters. One way to proceed with this method is to assume that there is some prior knowledge of the parameters that enables a restriction to be imposed. The method of moments equations under this restriction can then be readily solved.

Another possibility is to derive additional estimating equations based on the higher moments. This is the subject of Section 3.3.

There is a comparison between this identifiability problem in the method of moments and the maximum likelihood approach. The only tractable assumption to obtain a maximum likelihood solution is to assume that the distributions of ξ, δ and ε are all Normal (Normal structural model). Otherwise the algebraic manipulation required becomes an enormous task. This is discussed in subsequent Chapters. If all the distributions are assumed Normal, this leads to the random variable $(x, y)^T$ having a bivariate Normal distribution. This distribution has five parameters, and the maximum likelihood estimators for these parameters are identical to the method of moments estimators based on the moment equations (3.1), (3.2), (3.3), (3.4), and (3.5) above. In this case therefore it is not possible to find unique solutions to the likelihood equations without making an additional assumption, effectively restricting the parameter space. The maximum likelihood approach is discussed in detail in Chapter 5.

3.2.1 Estimators Based on the First and Second Moments

Equation (3.1) immediately yields the intuitive estimator for μ

$$\tilde{\mu} = \bar{x} \tag{3.7}$$

The estimators for the remaining parameters can be expressed as functions of the slope estimator, $\tilde{\beta}$, and other sample moments. An estimator for the intercept may be found by substituting (3.1) into (3.2) and rearranging to give

$$\tilde{\alpha} = \bar{y} - \tilde{\beta}\bar{x} \tag{3.8}$$

This shows just as in simple linear regression, that this solution for the errors in variables regression line passes through the centroid $(\bar{x}, \bar{y})$ of the data.

Equation (3.5) gives

$$\tilde{\sigma}^2 = \frac{s_{xy}}{\tilde{\beta}} \tag{3.9}$$

with $\tilde{\beta}$ and s_{xy} sharing the same sign so that the variance estimate is non-negative. This is a fundamental assumption, referred to frequently in the following presentation.

If the error variance σ_δ^2 is unknown, it may be estimated using (3.3)

$$\tilde{\sigma}_\delta^2 = s_{xx} - \tilde{\sigma}^2 \tag{3.10}$$

Finally, if the error variance σ_ε^2 is unknown, it may be estimated using (3.4)

$$\tilde{\sigma}_\varepsilon^2 = s_{yy} - \tilde{\beta}^2\tilde{\sigma}^2 \tag{3.11}$$

In order to ensure that the estimators for the variances are non negative, admissibility conditions must be placed on the equations. The straightforward conditions are included below

$$\begin{aligned} s_{xx} &> \sigma_\delta^2 \\ s_{yy} &> \sigma_\varepsilon^2 \end{aligned}$$

Other admissibility conditions specific to special cases are described later in this Chapter. Admissibility conditions are discussed in detail by Kendall and Stuart [67], Hood [56], Hood et al [57] and Dunn [39]. Practically speaking, if these admissibility conditions are broken, the choice of a linear structural model must be questioned. More precisely the estimate of the slope must lie between the slopes of the regression lines of y on x and x on y for variance estimates using equations (3.3), (3.4) and (3.5) to be non-negative. This point is demonstrated mathematically here.

$\tilde{\beta}$ and s_{xy} should have the same sign and variances are non-negative. We first deal with the case where $s_{xy} > 0$, hence $\tilde{\beta} > 0$. From equation (3.3) the condition $\tilde{\sigma}_\delta^2 \geq 0 \Rightarrow s_{xx} \geq \tilde{\sigma}^2$. From equation (3.5) this gives $\tilde{\beta}s_{xx} \geq \tilde{\beta}\tilde{\sigma}^2 = s_{xy}$ and so $\tilde{\beta} \geq \frac{s_{xy}}{s_{xx}}$. The right hand side is the slope of the simple linear regression of y on x. From equation (3.4) the condition $\tilde{\sigma_\varepsilon}^2 \geq 0 \Rightarrow s_{yy} \geq \tilde{\beta}^2\tilde{\sigma}^2 = \tilde{\beta}s_{xy}$ from equation (3.5). Thus $\tilde{\beta} \leq \frac{s_{yy}}{s_{xy}}$. The simple linear regression of x on y gives an estimator for the slope of the equation to predict x with y as $\frac{s_{xy}}{s_{yy}}$. However the slope is usually taken to calculate y with x and comparison should be made with the reciprocal of this estimator which is $\frac{s_{yy}}{s_{xy}}$. Hence the result that the errors in variables slope estimator is between the slopes of y on x and x on y regression is shown. If s_{xy} is negative, all inequalities are reversed. In conclusion for negative s_{xy},

$$\frac{s_{yy}}{s_{xy}} \leq \tilde{\beta} \leq \frac{s_{xy}}{s_{xx}},$$

and for positive s_{xy},

$$\frac{s_{xy}}{s_{xx}} \leq \tilde{\beta} \leq \frac{s_{yy}}{s_{xy}}.$$

All of the above estimating equations can be written in terms of sample moments and the slope. Unfortunately there is no single errors in variables slope estimator that can be used in all situations. In order to use the first and second moment estimating equations alone, and to avoid the identifiability problem, the practitioner must decide which restriction of the parameter space is likely to suit the purpose best. Various restrictions and their corresponding slope estimates are discussed below. With one exception, these estimators have been described previously; most were given by Kendall and Stuart [67], Hood et al [57] and, in a method of moments context by Dunn [39].

Intercept α known With this restriction, an estimator for the slope β can be derived using equations (3.1) and (3.2) alone. Substituting (3.1) into (3.2) and rearranging yields

$$\tilde{\beta}_1 = \frac{\bar{y} - \alpha}{\bar{x}}$$

This estimator can be seen to be similar to the ratio of means of grouped data advocated by Wald [110] and Bartlett [6], except that an adjustment is made in the numerator for the intercept. Dunn [39] considered this restriction when $\alpha = 0$. He wrote that this assumption is extremely unsafe as a particular characteristic of the very line that is used as a model is assumed.

Obvious problems occur with this estimator when $\bar{x} \approx 0$. Specific admissibility conditions are

$$\begin{aligned} s_{xx} &> \sigma^2 \\ s_{yy} &> \frac{\bar{y} - \alpha}{\bar{x}} s_{xy} \end{aligned}$$

Error variance σ_δ^2 known Equations (3.3) and (3.5) are used to obtain an estimator for β. Since σ_δ^2 is known, (3.3) can be written in terms of σ^2. It remains to substitute this into (3.5), and rearrange to obtain

$$\tilde{\beta}_2 = \frac{s_{xy}}{s_{xx} - \sigma_\delta^2}$$

This estimate is a modification of the standard y on x regression slope estimator. The modification is to subtract the known error variance σ_δ^2 from s_{xx} in the denominator of the expression. The effects of equation error (outlined earlier in this thesis) have led some authors, notably Dunn [39], to recommend that an estimator be chosen that relies only on information about σ_δ^2. The difficulty of using prior information of error variability in the y variable to estimate the variance σ_ε^2 is that such information may underestimate the variance terms as the contribution made by the equation error term may be overlooked. Dunn's conclusion is that estimators that assume prior knowledge of the error variance σ_δ^2 associated with the measurement of x, are more likely to be reliable in practical applications than those that assume prior knowledge of σ_ε^2.

The admissibility conditions for this estimator are

$$\begin{aligned} s_{xx} &> \sigma_\delta^2 \\ s_{yy} &> \frac{(s_{xy})^2}{s_{xx} - \sigma_\delta^2} \end{aligned}$$

Error variance σ_ε^2 known Writing equation (3.4) in terms of σ^2 and then substituting into equation (3.5) gives an estimate of β as

$$\tilde{\beta}_3 = \frac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}}$$

This estimator is a modification of the reciprocal of the slope of an x on y regression. The modification here is to subtract the known error variance σ_ε^2 from s_{yy} in the numerator of the expression.

The admissibility conditions for this estimator are

$$\begin{aligned} s_{yy} &> \sigma_\varepsilon^2 \\ s_{xx} &> \frac{(s_{xy})^2}{s_{yy} - \sigma_\varepsilon^2} \end{aligned}$$

Reliability ratio $\kappa = \frac{\sigma^2}{\sigma^2+\sigma_\delta^2}$ known The y on x regression estimator for the slope is biased when applied to an errors in variables model. Indeed,

$$E\left[\frac{s_{xy}}{s_{xx}}\right] \approx \frac{\sigma^2}{\sigma^2 + \sigma_\delta^2}\beta$$

as shown by Fuller [41]. The ratio

$$\kappa = \frac{\sigma^2}{\sigma^2 + \sigma_\delta^2}$$

is known as the reliability ratio, and if known, the bias of the slope estimator for y on x regression may be corrected. This suggests an estimator of the form

$$\tilde{\beta}_4 = \frac{s_{xy}}{\kappa s_{xx}}$$

This estimator may be derived from the first and second moment equations by dividing equation (3.5) by equation (3.3). Substituting in the known value for the reliability ratio and rearranging then gives the above estimator. There are no admissibility conditions associated with this estimator.

Ratio $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ known Putting $\sigma_\varepsilon^2 = \lambda\sigma_\delta^2$ and manipulating equations (3.3), (3.4) and (3.5) gives the following quadratic in $\tilde{\beta}$

$$\tilde{\beta}^2 s_{xy} + \tilde{\beta}(\lambda s_{xx} - s_{yy}) - \lambda s_{xy} = 0$$

To ensure that admissible estimates are obtained, the positive root must be taken, and so the slope estimator in this scenario is

$$\tilde{\beta}_5 = \frac{(s_{yy} - \lambda s_{xx}) + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda(s_{xy})^2}}{2s_{xy}}$$

and there are no admissibility conditions. The positive sign is taken for the square root term to ensure that $\tilde{\beta}_5$ and s_{xy} have the same sign.

If λ is taken to be 1, this estimator is the same as that in orthogonal regression outlined towards the beginning of this thesis. If $\lambda \neq 1$ then a different projection from the data point onto the regression line is minimised. In particular, with $\lambda = 1$ and $s_{yy} = s_{xx}$, $\tilde{\beta}_5 = 1$.

Riggs et al. [88], based on their simulation studies, recommended the use of $\tilde{\beta}_5$ but emphasised the importance of having a reliable prior knowledge of the ratio λ. The effects of using an incorrect λ have been discussed by Lakshminarayanan and Gunst [71].

As a point of note, when $\lambda = 1$ the discriminant becomes $(s_{yy} - s_{xx})^2 + 4s_{xy}^2$. By rotating the axes to a new co-ordinate system (u, v) through the transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

it is straightforward to show that

$$(s_{yy} - s_{xx})^2 + 4s_{xy}^2 = (s_{vv} - s_{uu})^2 + 4s_{uv}^2$$

This is an example of a rotationally invariant moment. This was derived by Hu [58]. If the rotation is through an angle whose tangent is equal to the slope of y on x regression line, then $s_{uv} = 0$ and the discriminant reduces to $(s_{vv} - s_{uu})^2$. Thus the discriminant is the square of the difference in variation along the y on x line and that orthogonal to it. In addition, the discriminant measures whether the dispersion of points about the centroid is isotropic or directional. Rotationally invariant moments are commonly used in some aspects of signal and image processing where an object must be analysed independently of its angular orientation. Indeed, Hu stated that the term $\sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}$ may be interpreted as the "slenderness" of the data.

Ratio $\nu = \frac{\lambda}{\beta^2}$ known This dimensionless ratio is used as a restriction to illustrate a link between y on x regression, geometric mean regression and x on y regression. In practise it seems unlikely that this ratio would be known a priori. However, for completeness, details are included here.

Using equations (3.3) and (3.4), then the ratio λ can be written as $\frac{s_{yy}-\beta^2\sigma^2}{s_{xx}-\sigma^2}$. Combining this with equations (3.3), (3.4) and (3.5) yields the following quadratic in $\tilde{\beta}$

$$\tilde{\beta}^2 \nu s_{xx} + \tilde{\beta} s_{xy}(1-\nu) - s_{yy} = 0$$

An estimator for β under the assumption that $\nu = \frac{\lambda}{\beta^2}$ is known is then

$$\tilde{\beta}_6 = \frac{-(1-\nu)s_{xy} + \sqrt{s_{xy}^2(1-\nu)^2 + 4\nu s_{xx}s_{yy}}}{2\nu s_{xx}}$$

Indeed,

$$\tilde{\beta}_6 = \frac{s_{xy}}{s_{xx}} \cdot \left[\frac{(\nu-1) + \sqrt{(\nu-1)^2 + \frac{4\nu}{r^2}}}{2\nu}\right] \tag{3.12}$$

where

$$r^2 = \frac{s_{xy}^2}{s_{xx}s_{yy}}$$

is the Pearson product-moment correlation coefficient between x and y.

Now if $\nu = 1$, this is equivalent to assuming $\lambda = \beta^2$. This is the exact assumption made when using geometric mean regression (as outlined in Chapter 2). Substituting $\nu = 1$ into (3.12) yields

$$\tilde{\beta}_{GM} = sgn(s_{xy})\sqrt{\frac{s_{yy}}{s_{xx}}}$$

ensuring that s_{xy} and $\tilde{\beta}$ have the same sign. Similarly substituting $\nu = \infty$ and $\nu = 0$ into (3.12) yields the slope estimator of y on x and x on y regression respectively, after some algebraic manipulation.

In a method comparisons context, Dunn [39] defined two methods of measurement as equivalent if $\nu = 1$. Indeed, if $\nu = 1$, then this intrinsically implies that the geometric mean line is the line of best fit. A number of different types of equivalences were introduced by Tan and Iglewicz [102], again reported by Dunn [39]

1. Individual equivalence, when $\lambda = \beta = 1$ and $\alpha = 0$

2. Average equivalence, when $\beta = 1$ and $\alpha = 0$

3. Sensitivity equivalence, when $\nu = 1$ and $\alpha = 0$

Both variances σ_δ^2 and σ_ε^2 known For this case, there are four parameters and five moment equations. Therefore this model is underparameterised and any four of the moment equations (3.1) to (3.5) can be used to derive unique estimators. Some possible solutions of the method of moment estimating equations (3.1) to (3.5) are outlined here.

1. From (3.3), $\sigma^2 = s_{xx} - \sigma_\delta^2$. Substituting into equation (3.5) yields the same estimator as when σ_δ^2 is solely known.

$$\tilde{\beta}_2 = \frac{s_{xy}}{s_{xx} - \sigma_\delta^2}$$

2. From (3.4), $\beta^2\sigma^2 = s_{yy} - \sigma_\varepsilon^2$. Substituting into equation (3.5) yields the same estimator as when σ_ε^2 is solely known.

$$\tilde{\beta}_3 = \frac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}}$$

3. Since both error variances are known, the ratio λ is also known. This yields

$$\tilde{\beta}_5 = \frac{(s_{yy} - \lambda s_{xx}) + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda(s_{xy})^2}}{2s_{xy}}$$

4. Rearranging equation (3.4) in terms of $\beta^2\sigma^2$ and dividing by equation (3.3) gives $\beta^2 = \frac{s_{yy}-\sigma_\varepsilon^2}{s_{xx}-\sigma_\delta^2}$. Upon taking the square root, another estimator for β is

$$\tilde{\beta}_7 = sgn(s_{xy})\sqrt{\frac{s_{yy} - \sigma_\varepsilon^2}{s_{xx} - \sigma_\delta^2}}$$

The estimator $\tilde{\beta}_7$ is a modification of the geometric mean estimator described earlier in this thesis, in that the numerator and denominator are modified by the subtraction of both error variances. Since the assumption is always made that the sign of s_{xy} is the same as the slope β, the $sgn(s_{xy})$ component of $\tilde{\beta}_7$ is included. To this extent (3.5) is used in deriving this estimator. There are clear admissibility conditions for this estimator.

$$\begin{aligned} s_{xx} &> \sigma_\delta^2 \\ s_{yy} &> \sigma_\varepsilon^2 \end{aligned}$$

Once a slope estimator has been obtained, its value may be substituted into equations (3.7) to (3.11) in order to estimate the remaining parameters. All the estimators outlined above are found by restricting the parameter space. If a restriction is not made, then the method of moment equations are inconsistent. This is primarily due to the elementary problem of having six unknown parameters, yet only five moment estimating equations. The admissibility conditions essentially suggest that the errors in variables regression line lies between the y on x and x on y regression lines (see proof earlier). If this is not the case, then negative estimates for some or all of the variances in the model (namely $\sigma^2, \sigma_\delta^2$ and σ_ε^2) may be obtained. This is also applicable to the estimators making use of higher moments described next.

The moment equations (3.1) to (3.5) only use the first and second order central moments. It is possible to extend this set of equations to consider third order moments, and even fourth order moments. This may provide an alternative way of using the method of moments instead of restricting the parameter space. Estimators making use of higher moments are now discussed.

3.3 Estimators Making Use of Higher Moments

3.3.1 Estimators Making Use of the Third Moments

The third order moments are written as follows

$$\begin{aligned}
s_{xxx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3 \\
s_{xxy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y}) \\
s_{xyy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})^2 \\
s_{yyy} &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^3
\end{aligned}$$

In this section, the notation

$$\begin{aligned}
\xi_i^* &= \xi_i-\bar{\xi} \\
\delta_i^* &= \delta_i-\bar{\delta} \\
\varepsilon_i^* &= \varepsilon_i-\bar{\varepsilon}
\end{aligned}$$

is introduced for brevity.

The moment equations based on the third moments are slightly more difficult to derive than the first and second order moment equations. An example is provided below to outline the general approach.

Derivation of Moment Equation for s_{xxy}

$$\begin{aligned}
E[ns_{xxy}] &= E\Big[\sum(x_i-\bar{x})^2(y_i-\bar{y})\Big] \\
&= E\Big[\sum\{(\xi_i^*)+(\delta_i^*)\}^2\{\beta(\xi_i^*)+(\varepsilon_i^*)\}\Big]
\end{aligned}$$

Terms of order n^{-1} are neglected, so the expectations of all the cross products are zero. Moreover because of the assumptions that ξ, δ and ε are mutually uncorrelated, to order n^{-1} terms such as $E[(\xi_i - \bar{\xi})]$ are also zero. Hence $E[ns_{xxy}] = n\beta\mu_{\xi 3}$, where $\mu_{\xi 3} = E[(\xi - \mu)^3]$. This procedure can be replicated for each of the third order central moments yielding the third order moment estimating equations below.

$$s_{xxx} = \tilde{\mu_{\xi 3}} + \tilde{\mu_{\delta 3}} \tag{3.13}$$

$$s_{xxy} = \tilde{\beta}\tilde{\mu_{\xi 3}} \tag{3.14}$$

$$s_{xyy} = \tilde{\beta}^2\tilde{\mu_{\xi 3}} \tag{3.15}$$

$$s_{yyy} = \tilde{\beta}^3\tilde{\mu_{\xi 3}} + \tilde{\mu_{\varepsilon 3}} \tag{3.16}$$

where $\mu_{\delta 3} = E[\delta^3]$, and $\mu_{\varepsilon 3} = E[\varepsilon^3]$ as defined earlier.

Combining the first and second moment equations (3.1) to (3.5), with the third moment equations (3.13) to (3.16) gives nine equations in nine unknown parameters. Hence there exist unique estimators for the unknown parameters. The additional parameters that have been gained are the third moments $\mu_{\xi 3}, \mu_{\delta 3}$, and $\mu_{\varepsilon 3}$. However, it is unlikely in practise that these third moments of the error terms are of as much interest as parameters such as the slope and the intercept of the regression line.

These equations must be treated with care. It is necessary to assume that $\mu_{\xi 3} \neq 0$, and the third sample moments should be significantly different from zero. In other words, in order for estimators based on these equations to be reliable it is necessary that the observed distribution of both x and y are sufficiently skewed. Moreover, the sample sizes needed to accurately compute third order moments will inevitably be larger than those for first and second order moments. It is this requirement that has probably led to the use of third moment estimators receiving relatively little attention in the literature. Papers by authors who have used this approach have been discussed in the literature survey in Chapter 2.

Nevertheless, assuming that one has a sufficiently large sample size and both x and y are skewed, a straightforward slope estimator may be found without assuming anything known a priori about the values taken by any of the parameters. This estimator is obtained by dividing equation (3.15) by equation (3.14)

$$\tilde{\beta}_8 = \frac{s_{xyy}}{s_{xxy}}$$

The estimator of β may be substituted into equations (3.7) to (3.11) to obtain estimators for $\mu, \alpha, \sigma^2, \sigma_\delta^2$ and σ_ε^2. The third moment $\mu_{\xi 3}$ may be estimated from equation (3.14)

$$\tilde{\mu_{\xi 3}} = \frac{s_{xxy}}{\tilde{\beta}_8} = \frac{s_{xxy}^2}{s_{xyy}}$$

Other simple ways of estimating the slope are available if the additional assumptions $\mu_{\delta 3} = \mu_{\varepsilon 3} = 0$ hold. The assumptions hold if the error terms δ and ε are from a symmetric distribution. It still remains the case however that ξ has to be sufficiently skewed to allow the third order sample moments of x and y to be sufficiently different from zero. With these additional assumptions, two further slope estimators may be found. Dividing equation (3.14) by (3.13) yields

$$\tilde{\beta} = \frac{s_{xxy}}{s_{xxx}} \tag{3.17}$$

and dividing equation (3.16) by equation (3.15) gives

$$\tilde{\beta} = \frac{s_{yyy}}{s_{xyy}} \tag{3.18}$$

Estimators (3.17) and (3.18) will receive little attention in this thesis, as estimators that make the least number of assumptions are likely to be of the most practical value. Note that if the estimator $\tilde{\beta}_8$ is to be consistent with equations (3.3), (3.4) and (3.5), and that variance estimates are to be non-negative, it is necessary that $\tilde{\beta}_8$ should lie between the slopes of x on y and y on x regression.

3.3.2 Estimators Making Use of the Fourth Moments

A way of avoiding having to assume that the observations are sufficiently skewed is by using the fourth order moment estimating equations. However, in order to ensure a stable estimate the sample size needed will be larger even than that for estimators using the third order moment equations. Moreover the distributions of x and y need to be sufficiently kurtotic for the fourth moments to be significantly different from zero.

The fourth order central moments are written as follows

$$\begin{aligned}
s_{xxxx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 \\
s_{xxxy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3(y_i - \bar{y}) \\
s_{xxyy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2 \\
s_{xyyy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^3 \\
s_{yyyy} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^4
\end{aligned}$$

The fourth order moment equations can then be derived in a similar manner to the third order moment equations described in the previous section

$$s_{xxxx} = \tilde{\mu_{\xi4}} + 6\tilde{\sigma}^2\tilde{\sigma_\delta}^2 + \tilde{\mu_{\delta4}} \quad (3.19)$$

$$s_{xxxy} = \tilde{\beta}\tilde{\mu_{\xi4}} + 3\tilde{\beta}\tilde{\sigma}^2\tilde{\sigma_\delta}^2 \quad (3.20)$$

$$s_{xxyy} = \tilde{\beta}^2\tilde{\mu_{\xi4}} + \tilde{\beta}^2\tilde{\sigma}^2\tilde{\sigma_\delta}^2 + \tilde{\sigma}^2\tilde{\sigma_\varepsilon}^2 + \tilde{\sigma_\delta}^2\tilde{\sigma_\varepsilon}^2 \quad (3.21)$$

$$s_{xyyy} = \tilde{\beta}^3\tilde{\mu_{\xi4}} + 3\tilde{\beta}\tilde{\sigma}^2\tilde{\sigma_\varepsilon}^2 \quad (3.22)$$

$$s_{yyyy} = \tilde{\beta}^4\tilde{\mu_{\xi4}} + 6\tilde{\beta}^2\tilde{\sigma}^2\tilde{\sigma_\varepsilon}^2 + \tilde{\mu_{\varepsilon4}} \quad (3.23)$$

where $\mu_{\xi4} = E[(\xi - \mu)^4], \mu_{\delta4} = E[\delta^4]$, and $\mu_{\varepsilon4} = E[\varepsilon^4]$ as defined earlier.

Combining these fourth moment equations with the first and second order moment equations results in a set of ten equations, in nine unknowns. The new parameters introduced here are $\mu_{\xi4}, \mu_{\delta4}$, and $\mu_{\varepsilon4}$. Some of these equations are therefore not needed As a result there does not exist a unique estimator for the slope.

In this situation, it makes sense to use the equations which avoid the higher moments of the error terms. This leaves (3.20), (3.21) and (3.22). These equations can be combined with those based on first and second moments to obtain three different slope estimators. All of these are derived here.

Using (3.20) **and** (3.22) Multiply (3.20) by β^2 and subtract (3.22) to give

$$\tilde{\beta}^2 s_{xxxy} - s_{xyyy} = 3\tilde{\beta}\tilde{\sigma}^2(\tilde{\beta}^2\tilde{\sigma}_\delta^2 - \tilde{\sigma}_\varepsilon^2)$$

Also, multiply (3.3) by β^2 and subtract from (3.4) to obtain

$$\beta^2 s_{xx} - s_{yy} = \tilde{\beta}^2\tilde{\sigma}_\delta^2 - \tilde{\sigma}_\varepsilon^2$$

Equation (3.5) is now used, and after some algebraic simplification, the slope estimator $\tilde{\beta}_9$ is derived as

$$\tilde{\beta}_9 = \sqrt{\frac{s_{xyyy} - 3s_{xy}s_{yy}}{s_{xxxy} - 3s_{xx}s_{xy}}} \tag{3.24}$$

Using (3.20) **and** (3.21) Multiply (3.20) by β and subtract (3.21) to obtain

$$\beta s_{xxxy} - s_{xxyy} = 3\beta^2\sigma^2\sigma_\delta^2 - \beta^2\sigma^2\sigma_\delta^2 - \sigma^2\sigma_\varepsilon^2 - \sigma_\delta^2\sigma_\varepsilon^2$$

As stated previously, the distribution of the bivariate random variable $(x, y)^T$ has a mean vector that is equal to $(\mu, \alpha + \beta\mu)^T$ and variance covariance matrix given by the following expression

$$\Sigma = \begin{pmatrix} \sigma^2 + \sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

This variance covariance matrix is estimated by the matrix S.

$$S = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{pmatrix}$$

The determinant of the matrix Σ is $|\Sigma| = \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2$ (which appears in the expression for $\beta s_{xxxy} - s_{xxyy}$) and is therefore estimated by the determinant of S.

$$|\tilde{\Sigma}| = |S| = s_{xx}s_{yy} - (s_{xy})^2$$

Hence we can write

$$\beta s_{xxxy} - s_{xxyy} = 3\beta^2\sigma^2\sigma_\delta^2 + (s_{xy})^2 - s_{xx}s_{yy}$$

Using equations (3.3) and (3.5) and rearranging yields the following estimator for the slope

$$\tilde{\beta*} = \frac{s_{xxyy} - 2(s_{xy})^2 - s_{xx}s_{yy}}{s_{xxxy} - 3s_{xx}s_{xy}}$$

Using (3.21) **and** (3.22) Multiply (3.21) by β and subtract (3.22) to get

$$\beta s_{xxyy} - s_{xyyy} = \beta(\beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2) - 3\beta\sigma^2\sigma_\varepsilon^2$$

Using the estimator for $|\Sigma|$ and equations (3.4) and (3.5) we obtain the following slope estimator

$$\tilde{\beta}^\bullet = \frac{s_{xyyy} - 3s_{xy}s_{yy}}{s_{xxyy} - 2(s_{xy})^2 - s_{xx}s_{yy}}$$

The slope estimators derived above are functionally related, in that

$$\tilde{\beta}_9 = \sqrt{\tilde{\beta}^*\tilde{\beta}^\bullet}$$

It was functional relations of this sort that was covered in the paper by Cragg [27] mentioned in Chapter 2.

There may be a practical difficulty associated with the use of (3.24) if the random variable ξ is Normally distributed. In this case the fourth moment is equal to three times the square of the variance. A random variable for which this property does not hold is said to be kurtotic. A scale invariant measure of the excess of kurtosis is given by the following expression

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

If the distribution of ξ has zero excess of kurtosis the average values of the five sample moments used in equation (3.24) are as follows

$$\begin{aligned}
E[s_{xyyy}] &= 3\beta^3\sigma^4 + 3\beta\sigma^2\sigma_\varepsilon^2 \\
E[s_{xxxy}] &= 3\beta\sigma^4 + 3\beta\sigma^2\sigma_\delta^2 \\
E[s_{xx}] &= \sigma^2 + \sigma_\delta^2 \\
E[s_{yy}] &= \beta^2\sigma^2 + \sigma_\varepsilon^2 \\
E[s_{xy}] &= \beta\sigma^2
\end{aligned}$$

Then it can be seen that the average value of the numerator of equation (3.24) is equal to zero (to order n^{-1}), as is the average value of the denominator. Thus there is an additional assumption that has to be made for this equation to be a reliable estimator, and that is that $\mu_{\xi 4}$ must be different from $3\sigma^4$. In practical terms, both the numerator and the denominator of the right hand side of equation (3.24) must be significantly different from zero.

If a reliable estimate of the slope β can be obtained from (3.24), equations (3.1) to (3.5) enable the intercept α and the variances $\sigma^2, \sigma_\delta^2$ and σ_ε^2 to be estimated. As is the case for $\tilde{\beta}_8$, the slope $\tilde{\beta}_9$ must lie between the slopes of y on x and x on y regression respectively so the variance estimators are non-negative. The fourth moment $\mu_{\xi 4}$ of ξ can then be estimated from (3.20), and the fourth moments $\mu_{\delta 4}$ and $\mu_{\varepsilon 4}$ of the error terms δ and ε can be estimated from equations (3.19) and (3.23) respectively, though estimates of these higher moments of the error terms are less likely to be of practical value.

In this thesis, only $\tilde{\beta}_9$ will be considered due to the length of time taken to construct the variance covariance matrix when the slope estimator involves fourth moments. Similar analysis (that will be shown in subsequent sections of this thesis) can be applied to the other fourth moment estimators outlined above.

3.4 Equation Error

Thus far, much consideration has been given to the estimation of the linear structural model. This section will look at the impact of equation error upon the estimation of the model. Similarly, the application of the estimation procedures to a functional model is looked at in Chapter 5 via a maximum likelihood approach.

Equation error A term for equation error is represented by the addition of a new random component that is associated with the measurement y, thus $y = \alpha + \beta\xi + \omega + \varepsilon$. This has the effect of changing moment equation (3.4) since $Var[y] = Var[\alpha + \beta\xi + \omega + \varepsilon] = \beta^2\sigma^2 + \sigma_\omega^2 + \sigma_\varepsilon^2$; wherein it is assumed that the equation error terms have a homoscedastic variance σ_ω^2 and that they are uncorrelated with the other random variables in the model.

The inclusion of the equation error term then yields the following first and second order moment equations:

$$\begin{aligned} \bar{x} &= \tilde{\mu} \\ \bar{y} &= \tilde{\alpha} + \tilde{\beta}\tilde{\mu} \\ s_{xx} &= \tilde{\sigma}^2 + \tilde{\sigma}_\delta^2 \\ s_{yy} &= \tilde{\beta}^2\tilde{\sigma}^2 + \tilde{\sigma}_\omega^2 + \tilde{\sigma}_\varepsilon^2 \\ s_{xy} &= \tilde{\beta}\tilde{\sigma}^2. \end{aligned}$$

It can be seen that the only equation that is changed is (3.4). The effect of the introduction of equation error in the model is an extra term σ_ω^2 on the right hand side of this equation. In practise, given a data set it is difficult to partition equation error and measurement error. It is presumably for this reason that Dunn [39] makes the recommendation that estimators solely based on the assumption that σ_δ^2 is known are likely to be safer. The estimators of the slope that are directly affected by the presence of equation error are $\tilde{\beta}_3$, $\tilde{\beta}_5$, $\tilde{\beta}_6$ and $\tilde{\beta}_7$.

For σ_ε^2 known, the method of moments slope estimator when equation error is present is

$$\tilde{\beta}_{3e} = \frac{s_{yy} - \sigma_\varepsilon^2 - \sigma_\omega^2}{s_{xy}} = \tilde{\beta}_3 - \frac{\sigma_\omega^2}{s_{xy}}.$$

If equation error is ignored, the slope will be either over, or underestimated by $\frac{\sigma_\omega^2}{s_{xy}}$ depending on the sign of s_{xy}. The magnitude of s_{xy} will also affect the degree to which the slope is over or under estimated. The effect is to move the slope even further from the x on y regression line than is the case if equation error is not present.

Let $u = \frac{s_{yy} - \lambda s_{xx}}{2s_{xy}}$. For $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ known, the method of moments slope estimator when equation error is present is

$$u - \frac{\sigma_\omega^2}{2s_{xy}} + \sqrt{\left(u - \frac{\sigma_\omega^2}{2s_{xy}}\right)^2 + \lambda}$$

where

$$\tilde{\beta}_5 = u + \sqrt{(u^2 + \lambda)}.$$

So again the term $\frac{\sigma_\omega^2}{s_{xy}}$ distinguishes $\tilde{\beta}_{5e}$ from $\tilde{\beta}_5$, and the magnitude of s_{xy} affects the degree to which the slope is over or under estimated.

The assumption $\nu = \frac{\lambda}{\beta^2}$ known was included for completeness, and to extend the concept of geometric mean regression in a previous section. In a similar manner, the method of moments slope estimator when equation error is present is

$$\tilde{\beta}_{6e} = \frac{s_{xy}}{s_{xx}} \cdot \left[\frac{(\nu - 1) + \sqrt{(\nu - 1)^2 + \frac{4\nu}{r^2} - \frac{4\nu s_{xx}\sigma_\omega^2}{s_{xy}^2}}}{2\nu}\right].$$

The difference between $\tilde{\beta}_6$ and $\tilde{\beta}_{6e}$ being the additional term $-4\frac{\nu s_{xx}\sigma_\omega^2}{s_{xy}^2}$ in the square root (r is the Pearson product-moment correlation introduced earlier).

When $\nu = 1$,

$$\tilde{\beta}_{6e} = \frac{s_{xy}}{s_{xx}} \cdot \left[\sqrt{\frac{s_{xx}s_{yy}}{s_{xy}^2} - \frac{\sigma_\omega^2 s_{xx}}{s_{xy}^2}} \right] < \tilde{\beta}_6$$

and so the assumption $\lambda = \beta^2$ does not reduce the above estimator to that of geometric mean regression. However, as $\nu \longrightarrow \infty$,

$$\tilde{\beta}_{6e} \longrightarrow \frac{s_{xy}}{s_{xx}}$$

as does $\tilde{\beta}_6$.

In the case where both σ_δ^2 and σ_ε^2 are known there are four estimators for the slope. One solution is the same as σ_δ^2 known, and is robust to equation error. A second solution is $\tilde{\beta}_{3e}$ as above, and upon taking the ratio of these error variances, a third solution is $\tilde{\beta}_{5e}$ Finally, another method of moments slope estimator when equation error is present is

$$\tilde{\beta}_{7e} = \sqrt{\frac{s_{yy} - \sigma_\varepsilon^2}{s_{xx} - \sigma_\delta^2} - \frac{\sigma_\omega^2}{s_{xx} - \sigma_\delta^2}}.$$

It is necessary to assume that $s_{xx} - \sigma_\delta^2 > 0$ for variance estimates to be non-negative, and so $\tilde{\beta}_{7e} < \tilde{\beta}_7$. Thus if equation error is ignored, the slope will be underestimated if $\tilde{\beta}_7$ is used.

3.5 Variances and Covariances of the Estimators

A common misunderstanding regarding the method of moments is that there is a lack of asymptotic theory associated with the method. This however is not true. Cramer [28] and subsequently other authors such as Bowman and Shenton [10] detailed an approximate method commonly known as the delta method (or the method of statistical differentials) to obtain expressions for variances and covariances of functions of sample moments. The method is sometimes described in statistics texts, for example DeGroot [32], and is often used in linear models to derive a variance stabilisation transformation (see Draper and Smith [37]). The delta method is used to approximate the expectations, and hence also the variances and covariances of functions of random variables by making use of a Taylor series expansion about the expected values. The derivation of the delta method is included below.

Consider a first order Taylor expansion of a function of a sample moment x, $f(x)$ where $E[x] = \mu$,

$$f(x) \approx f(\mu) + (x - \mu)f'(\mu). \tag{3.25}$$

Upon taking the expectation of both sides of (3.25) the first order approximation

$$E[f(x)] \approx f(\mu)$$

is found. Additionally,

$$\begin{aligned} Var\left[f(x)\right] = E\left[\{f(x) - E[f(x)]\}^2\right] &\approx \{f'(\mu)\}^2 E[(x-\mu)^2] \\ &= \{f'(\mu)\}^2\, Var[x] \\ &= \left\{\overline{\frac{\partial f}{\partial x}}\right\}^2 Var[x] \end{aligned}$$

The notation

$$\left\{\overline{\frac{\partial f}{\partial x}}\right\} = \left.\frac{\partial f}{\partial x}\right|_{x=E[x]}$$

was introduced by Cramer to denote a partial derivative evaluated at the expected values of the sample moments.

This can be naturally extended to functions of more than one sample moment. For a function $f(x,y)$

$$Var[f(x,y)] \approx \left\{\overline{\frac{\partial f}{\partial x}}\right\}^2 Var[x] + \left\{\overline{\frac{\partial f}{\partial y}}\right\}^2 Var[y] + 2\left\{\overline{\frac{\partial f}{\partial x}}\right\}\left\{\overline{\frac{\partial f}{\partial y}}\right\} Cov[x,y]$$

and for a function of p sample moments, $x_1, \ldots, x_p$,

$$Var[f(x_1, \ldots, x_p)] \approx \nabla^T V \nabla$$

where

$$\nabla^T = \left(\overline{\frac{\partial f}{\partial x_1}}, \ldots \overline{\frac{\partial f}{\partial x_p}}\right)$$

is the vector of derivatives with each sample moment substituted for its expected value, and

$$V = \begin{pmatrix} Var[x_1] & Cov[x_1,x_2] & \ldots & Cov[x_1,x_p] \\ \vdots & & \ddots & \vdots \\ Cov[x_1,x_p] & Cov[x_2,x_p] & \ldots & Var[x_p] \end{pmatrix}$$

is the $p \times p$ matrix containing the variances of and covariances between sample moments. Covariances between functions of sample moments can be derived in a similar manner.

Indeed for two functions of two sample moments x and y

$$\begin{aligned} Cov[f(x,y), g(x,y)] \approx\ & \left\{\overline{\frac{\partial f}{\partial x}}\right\}\left\{\overline{\frac{\partial g}{\partial x}}\right\} Var[x] + \left\{\overline{\frac{\partial f}{\partial y}}\right\}\left\{\overline{\frac{\partial g}{\partial y}}\right\} Var[y] \\ & + \left[\left\{\overline{\frac{\partial f}{\partial x}}\right\}\left\{\overline{\frac{\partial g}{\partial y}}\right\} + \left\{\overline{\frac{\partial f}{\partial y}}\right\}\left\{\overline{\frac{\partial g}{\partial x}}\right\}\right] Cov[x,y] \end{aligned}$$

Essentially, use of this method requires the prior computation of the variance of each relevant sample moment, and the covariances between each sample moment. For each of the restricted cases (in Section (3.2.1)), the following variances and covariances are used. The variances and covariances needed to compute the asymptotics for the higher moment based estimators will be stated later on in this Chapter.

$$\begin{aligned}
Var[\bar{x}] &\approx \frac{\sigma^2 + \sigma_\delta^2}{n} && (3.26)\\
Var[\bar{y}] &\approx \frac{\beta^2\sigma^2 + \sigma_\varepsilon^2}{n} && (3.27)\\
Cov[\bar{x}, \bar{y}] &\approx \frac{\beta\sigma^2}{n} && (3.28)\\
Var[s_{xx}] &\approx \frac{(\mu_{\xi 4} - \sigma^4) + (\mu_{\delta 4} - \sigma_\delta^4) + 4\sigma^2\sigma_\delta^2}{n} && (3.29)\\
Var[s_{xy}] &\approx \frac{\beta^2(\mu_{\xi 4} - \sigma^4) + \sigma^2\sigma_\varepsilon^2 + \beta^2\sigma^2\sigma_\delta^2 + \sigma_\delta^2\sigma_\varepsilon^2}{n} &&\\
Var[s_{yy}] &\approx \frac{\beta^4(\mu_{\xi 4} - \sigma^4) + (\mu_{\varepsilon 4} - \sigma_\varepsilon^4) + 4\beta^2\sigma^2\sigma_\varepsilon^2}{n} &&\\
Cov[\bar{x}, s_{xx}] &\approx \frac{\mu_{\xi 3} + \mu_{\delta 3}}{n} &&\\
Cov[\bar{y}, s_{xx}] &\approx \frac{\beta\mu_{\xi 3}}{n} &&\\
Cov[\bar{x}, s_{xy}] &\approx \frac{\beta\mu_{\xi 3}}{n} &&\\
Cov[\bar{y}, s_{xy}] &\approx \frac{\beta^2\mu_{\xi 3}}{n} &&\\
Cov[\bar{x}, s_{yy}] &\approx \frac{\beta^2\mu_{\xi 3}}{n} &&\\
Cov[\bar{y}, s_{yy}] &\approx \frac{\beta^3\mu_{\xi 3} + \mu_{\varepsilon 3}}{n} &&\\
Cov[s_{xx}, s_{xy}] &\approx \frac{\beta(\mu_{\xi 4} - \sigma^4) + 2\beta\sigma^2\sigma_\delta^2}{n} && (3.30)\\
Cov[s_{xx}, s_{yy}] &\approx \frac{\beta^2(\mu_{\xi 4} - \sigma^4)}{n} &&\\
Cov[s_{xy}, s_{yy}] &\approx \frac{\beta^3(\mu_{\xi 4} - \sigma^4) + 2\beta\sigma^2\sigma_\varepsilon^2}{n} &&
\end{aligned}$$

Expressions (3.26), (3.27) and (3.28) follow from the definition of the linear structural model. To show how these may be derived, the algebra behind expressions (3.29) and (3.30) shall be outlined.

Derivation of $Var[s_{xx}]$ Since ξ_i and δ_i are uncorrelated we can write

$$\begin{aligned} E[s_{xx}] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\{(\xi_i^*)+(\delta_i^*)\}^2\right] \\ &= \frac{1}{n}E\left[\sum_{i=1}^{n}(\xi_i^*)^2+2\sum_{i=1}^{n}(\xi_i^*)(\delta_i^*)+\sum_{i=1}^{n}(\delta_i^*)^2\right] \\ &\approx \sigma^2+\sigma_\delta^2. \end{aligned}$$

The above result also follows from the method of moment estimating equation stated earlier, $s_{xx} = \sigma^2 + \sigma_\delta^2$.

$$\begin{aligned} E[(s_{xx})^2] &= \frac{1}{n^2}E\left[\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2\right] \\ &= \frac{1}{n^2}E\left[\left\{\sum_{i=1}^{n}(\xi_i^*+\delta_i^*)^2\right\}^2\right] \\ &\approx \frac{1}{n^2}\Big(n(\mu_{\xi 4}+6\sigma^2\sigma_\delta^2+\mu_{\delta 4})+n(n-1)(\sigma^4+2\sigma^2\sigma_\delta^2+\sigma_\delta^4)\Big) \end{aligned}$$

Hence it follows that

$$\begin{aligned} Var[s_{xx}] &= E[(s_{xx})^2]-E^2[s_{xx}] \\ &\approx \frac{(\mu_{\xi 4}-\sigma^4)+(\mu_{\delta 4}-\sigma_\delta^4)+4\sigma^2\sigma_\delta^2}{n} \end{aligned}$$

Derivation of $Cov[s_{xx}, s_{xy}]$

$$E[s_{xx}s_{xy}] = \frac{1}{n^2}E\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\times\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]$$

Now, $(x_i - \bar{x}) = (\xi_i^*) + (\delta_i^*)$ and $(y_i - \bar{y}) = \beta(\xi_i^*) + (\varepsilon_i^*)$. Substituting these into the above summation, and multiplying out leads to

$$E[s_{xx}s_{xy}] \approx \frac{1}{n^2}\Big(n(\beta\mu_{\xi 4}+\beta\sigma^2\sigma_\delta^2+2\beta\sigma^2\sigma_\delta^2)+n(n-1)(\beta\sigma^4+\beta\sigma^2\sigma_\delta^2)\Big)$$

Hence,

$$\begin{aligned} Cov[s_{xx}, s_{xy}] &= E[s_{xx}s_{xy}]-E[s_{xx}]E[s_{xy}] \\ &\approx \frac{\beta(\mu_{\xi 4}-\sigma^4)+2\beta\sigma^2\sigma_\delta^2}{n}. \end{aligned}$$

3.5.1 Constructing the Variance Covariance Matrices

For each restricted case, and for the estimators of the slope based on the higher moments a variance covariance matrix can be constructed. As there are six parameters in the linear structural model $\mu, \alpha, \beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2 the maximum size of the variance covariance matrix is 6×6. If the parameter space is restricted, then the size of the variance covariance matrix will decrease in accordance with the number of assumed parameters.

It is possible to use the delta method in order to construct 'shortcut' formulae or approximations to enable quicker calculation of each element of the variance covariance matrix. These shortcut formulae depend on the variance of the slope estimator and the covariance of the slope estimator with a first or second order sample moment. In some cases the variances and covariances do not depend on the slope estimator used, and as a result are robust to the choice of this estimator. These shortcut formulae are stated below, and repeating the style of the previous section, an example of how one is derived will be given. For brevity, the notation $|\Sigma| = \sigma_\delta^2\sigma_\varepsilon^2 + \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2$ is introduced. This is the determinant of the variance covariance matrix of the bivariate distribution of x and y that was introduced earlier in this Chapter.

Firstly, the shortcut formulae for the variances will be considered. $Var[\tilde{\alpha}]$ will be the example derivation provided.

$$
\begin{aligned}
Var[\tilde{\mu}] &\approx \frac{\sigma^2+\sigma_\delta^2}{n}\\
Var[\tilde{\alpha}] &\approx \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2+\sigma_\varepsilon^2}{n} + 2\mu(\beta Cov[\bar{x},\tilde{\beta}] - Cov[\bar{y},\tilde{\beta}])\\
Var[\tilde{\sigma}^2] &\approx \frac{\sigma^4}{\beta^2}Var[\tilde{\beta}] + \frac{|\Sigma|+\beta^2(\mu_{\xi 4}-\sigma^4)}{\beta^2 n} - \frac{2\sigma^2}{\beta^2}Cov[s_{xy},\tilde{\beta}]\\
Var[\tilde{\sigma}_\delta^2] &\approx \frac{\sigma^4}{\beta^2}Var[\tilde{\beta}] + \frac{|\Sigma|+\beta^2(\mu_{\delta 4}-\sigma_\delta^4)}{\beta^2 n} + \frac{2\sigma^2}{\beta}\left(Cov[s_{xx},\tilde{\beta}] - \frac{Cov[s_{xy},\tilde{\beta}]}{\beta}\right)\\
Var[\tilde{\sigma}_\varepsilon^2] &\approx \beta^2\sigma^4 Var[\tilde{\beta}] + 2\beta\sigma^2(\beta Cov[s_{xy},\tilde{\beta}] - Cov[s_{yy},\tilde{\beta}]) + \frac{\beta^2|\Sigma|+(\mu_{\varepsilon 4}-\sigma_\varepsilon^4)}{n}
\end{aligned}
$$

Derivation of $Var[\tilde{\alpha}]$

$$
\begin{aligned}
Var[\tilde{\alpha}] = Var[\bar{y}-\tilde{\beta}\bar{x}] &\approx \left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{y}}}\right\}^2 Var[\bar{y}] + \left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\tilde{\beta}}}\right\}^2 Var[\tilde{\beta}] + \left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{x}}}\right\}^2 Var[\bar{x}]\\
&+ 2\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{y}}}\right\}\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\tilde{\beta}}}\right\}Cov[\bar{y},\tilde{\beta}] + 2\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{x}}}\right\}\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\tilde{\beta}}}\right\}Cov[\bar{x},\tilde{\beta}]\\
&+ 2\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{x}}}\right\}\left\{\overline{\frac{\partial\tilde{\alpha}}{\partial\bar{y}}}\right\}Cov[\bar{x},\bar{y}]\\
&\approx \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2+\sigma_\varepsilon^2}{n} + 2\mu(\beta Cov[\bar{x},\tilde{\beta}] - Cov[\bar{y},\tilde{\beta}])
\end{aligned}
$$

A similar shortcut formula was provided in the paper by Hood et al [57]. As outlined in the literature survey towards the beginning of this thesis, they investigated the Normal structural model. They then used the theory of maximum likelihood to obtain the information matrices required for the asymptotic variance covariance matrices for the parameters of the model. Applying various algebraic manipulations to $Var[\tilde{\alpha}]$ they showed that

$$Var[\tilde{\alpha}] \approx \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2+\sigma_\varepsilon^2}{n}$$

The shortcut formula derived above is a generalisation of that derived by Hood et al. to cope with non Normal ξ. Indeed, if (ξ,δ,ε) do follow a trivariate normal distribution, then as $\tilde{\beta}$ is a function only of second order moments (or higher), $\tilde{\beta}$ is statistically independent of the first order sample moments. As a result $Cov[\bar{x},\tilde{\beta}] = Cov[\bar{y},\tilde{\beta}] = 0$ and the shortcut formula derived above collapses to that suggested by Hood et al.

Now, the shortcut formulae for the covariances of $\tilde{\mu}$ with the remaining parameters will be provided.

$$\begin{aligned}
Cov[\tilde{\mu},\tilde{\alpha}] &\approx -\frac{\beta\sigma_\delta^2}{n} - \mu Cov[\bar{x},\tilde{\beta}] \\
Cov[\tilde{\mu},\tilde{\beta}] &\approx Cov[\bar{x},\tilde{\beta}] \\
Cov[\tilde{\mu},\tilde{\sigma}^2] &\approx \frac{\mu_{\xi 3}}{n} - \frac{\sigma^2}{\beta}Cov[\bar{x},\tilde{\beta}] \\
Cov[\tilde{\mu},\tilde{\sigma}_\delta^2] &\approx \frac{\mu_{\delta 3}}{n} + \frac{\sigma^2}{\beta}Cov[\bar{x},\tilde{\beta}] \\
Cov[\tilde{\mu},\tilde{\sigma}_\varepsilon^2] &\approx -\beta\sigma^2 Cov[\bar{x},\tilde{\beta}]
\end{aligned}$$

The shortcut formulae for the covariances of $\tilde{\alpha}$ with all other parameters are listed here.

$$\begin{aligned}
Cov[\tilde{\alpha},\tilde{\beta}] &\approx Cov[\bar{y},\tilde{\beta}] - \beta Cov[\bar{x},\tilde{\beta}] - \mu Var[\tilde{\beta}] \\
Cov[\tilde{\alpha},\tilde{\sigma}^2] &\approx \frac{\mu\sigma^2}{\beta}Var[\tilde{\beta}] + \sigma^2\left(Cov[\bar{x},\tilde{\beta}] - \frac{Cov[\bar{y},\tilde{\beta}]}{\beta}\right) - \frac{\mu}{\beta}Cov[s_{xy},\tilde{\beta}] \\
Cov[\tilde{\alpha},\tilde{\sigma}_\delta^2] &\approx \frac{\mu}{\beta}Cov[s_{xy},\tilde{\beta}] - \frac{\beta\mu_{\delta 3}}{n} - \mu Cov[s_{xx},\tilde{\beta}] - \frac{\mu\sigma^2}{\beta}Var[\tilde{\beta}] \\
&\quad -\sigma^2\left(Cov[\bar{x},\tilde{\beta}] - \frac{Cov[\bar{y},\tilde{\beta}]}{\beta}\right) \\
Cov[\tilde{\alpha},\tilde{\sigma}_\varepsilon^2] &\approx \frac{\mu_{\varepsilon 3}}{n} + \beta\mu\sigma^2 Var[\tilde{\beta}] + \beta\sigma^2(\beta Cov[\bar{x},\tilde{\beta}] - Cov[\bar{y},\tilde{\beta}]) \\
&\quad +\mu(\beta Cov[s_{xy},\tilde{\beta}] - Cov[s_{yy},\tilde{\beta}])
\end{aligned}$$

The shortcut formulae for the covariances of $\tilde{\beta}$ with the remaining parameters are listed here.

$$\begin{aligned}
Cov[\tilde{\beta},\tilde{\sigma}^2] &\approx \frac{1}{\beta}Cov[s_{xy},\tilde{\beta}] - \frac{\sigma^2}{\beta}Var[\tilde{\beta}] \\
Cov[\tilde{\beta},\tilde{\sigma}_\delta^2] &\approx Cov[s_{xx},\tilde{\beta}] - \frac{Cov[s_{xy},\tilde{\beta}]}{\beta} + \frac{\sigma^2}{\beta}Var[\tilde{\beta}] \\
Cov[\tilde{\beta},\tilde{\sigma}_\varepsilon^2] &\approx Cov[s_{yy},\tilde{\beta}] - \beta Cov[s_{xy},\tilde{\beta}] - \beta\sigma^2 Var[\tilde{\beta}]
\end{aligned}$$

The shortcut formulae for the covariances of $\tilde{\sigma}^2$ with the remaining parameters are listed here.

$$\begin{aligned}
Cov[\tilde{\sigma}^2,\tilde{\sigma}_\delta^2] &\approx -\frac{\sigma^4}{\beta^2}Var[\tilde{\beta}] + \frac{\sigma^2}{\beta}\left(\frac{2}{\beta}Cov[s_{xy},\tilde{\beta}] - Cov[s_{xx},\tilde{\beta}]\right) + \frac{|\Sigma| - 2\sigma^2\sigma_\varepsilon^2 - 2\sigma_\delta^2\sigma_\varepsilon^2}{\beta^2 n} \\
Cov[\tilde{\sigma}^2,\tilde{\sigma}_\varepsilon^2] &\approx \sigma^4 Var[\tilde{\beta}] - \frac{\sigma^2}{\beta}Cov[s_{yy},\tilde{\beta}] + \frac{|\Sigma| - 2\beta^2\sigma^2\sigma_\delta^2 - 2\sigma_\delta^2\sigma_\varepsilon^2}{n}
\end{aligned}$$

Finally, the covariance between the error variance estimates is

$$Cov[\tilde{\sigma}_\delta^2,\tilde{\sigma}_\varepsilon^2] \approx -\sigma^4 Var[\tilde{\beta}] + \frac{\sigma^2}{\beta}Cov[s_{yy},\tilde{\beta}] - \beta\sigma^2 Cov[s_{xx},\tilde{\beta}] + \frac{|\Sigma| - 2\beta^2\sigma^2\sigma_\delta^2 - 2\sigma^2\sigma_\varepsilon^2}{n}$$

Again, an example derivation is provided.

Derivation of $Cov[\tilde{\beta},\tilde{\sigma}^2]$ We have that

$$\tilde{\sigma}^2 = \frac{s_{xy}}{\tilde{\beta}}.$$

A first order Taylor expansion of $\tilde{\sigma}^2$ around the expected values of s_{xy} and $\tilde{\beta}$ is

$$\tilde{\sigma}^2 = \sigma^2 + (s_{xy} - \beta\sigma^2)\frac{1}{\beta} - (\tilde{\beta} - \beta)\frac{\sigma^2}{\beta}.$$

Hence,

$$Cov[\tilde{\beta},\tilde{\sigma}^2] = E[(\tilde{\beta} - \beta)(\tilde{\sigma}^2 - \sigma^2)] \approx \frac{1}{\beta}Cov[s_{xy},\tilde{\beta}] - \frac{\sigma^2}{\beta}Var[\tilde{\beta}].$$

The complete asymptotic variance covariance matrices for the different slope estimators under varying assumptions are included in the following pages. For ease of presentation, the matrices are expressed as the sum of three components, A, B and C. This presentation has the advantage of making the matrices simpler for a practitioner to use.

The matrix A alone is needed if the assumptions are made that ξ, δ and ε all have zero third moments and zero measure of excess of kurtosis. These assumptions would be valid if all three of these variables are Normally distributed as in the Normal structural model.

The matrix B gives the additional terms that are necessary if ξ has non zero third moment and a non zero measure of kurtosis. It can be seen that in most cases the B matrices are sparse, needing only adjustment for the terms for $Var[\tilde{\sigma}^2]$ and $Cov[\tilde{\mu}, \sigma^2]$. The exceptions are the cases where the reliability ratio is assumed known ($\tilde{\beta}_4$), and slope estimators involving the higher moments.

The C matrix contains additional terms that are needed if the third moments and measures of excess of kurtosis are non zero for the error terms δ and ε. It is likely that these C matrices will prove of less value to practitioners than the A and B matrices. It is quite possible that a practitioner would not wish to assume that the distribution of the variable ξ is normal, or even that its third and fourth moments behave like those of a normal distribution. Indeed, the necessity for this assumption to be made in the likelihood approach may well have been one of the obstacles against a more widespread use of errors in variables methodology. The assumption of normal like distributions for the error terms, however, is more likely to be acceptable. Thus in many applications, the C matrix may be ignored.

As a check on the method employed the A matrices were checked with those given by Hood [56] and Hood et al. [57], where a different likelihood approach was used in deriving the asymptotic variance covariance matrices. In all cases exact agreement with the A matrices was found, although much simplification of the algebra has been found to be possible. As discussed in Chapter 5, the limitation of the likelihood approach is that it is limited to the case where all random variables are assumed to be Normally distributed. The moments approach described in this Chapter does not have this limitation.

3.5.2 The Variance Covariance Matrices

This section contains the variance covariance matrices for each of the slope estimators outlined earlier. The results are stated first, followed by a brief discussion. For brevity, the notation $U = \sigma^2 + \sigma_\delta^2$, $V = \beta^2\sigma_\delta^2 + \sigma_\varepsilon^2$, $e_1 = \mu_{\delta 4} - 3\sigma_\delta^4$, $e_2 = \mu_{\varepsilon 4} - 3\sigma_\varepsilon^4$ and $e_3 = \beta\lambda\mu_{\delta 3} + \mu_{\varepsilon 3}$ shall be used. This notation shall also be carried into the next section. U and V are the variances of x and y respectively. e_1 and e_2 are the excesses of kurtosis for δ and ε respectively.

Intercept α known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_1 = \frac{\bar{y} - \alpha}{\bar{x}}.$$

Since α is assumed to be known, the variance covariance matrix for $\tilde{\mu}, \tilde{\beta}, \tilde{\sigma}^2, \tilde{\sigma}_\delta^2$ and $\tilde{\sigma}_\varepsilon^2$ is required.

$$A_1 = \frac{1}{n}\begin{pmatrix} U & -\frac{\beta\sigma_\delta^2}{\mu} & \frac{\sigma^2\sigma_\delta^2}{\mu} & -\frac{\sigma^2\sigma_\delta^2}{\mu} & \frac{\beta^2\sigma^2\sigma_\delta^2}{\mu} \\ & \frac{V}{\mu^2} & -\frac{\sigma^2}{\beta\mu^2}V & \frac{\sigma^2}{\beta\mu^2}V & -\frac{\beta\sigma^2}{\mu^2}V \\ & & \frac{|\Sigma|}{\beta^2} + \frac{\sigma^4}{\beta^2\mu^2}V + 2\sigma^4 & -\frac{|\Sigma|}{\beta^2} - \frac{\sigma^4}{\beta^2\mu^2}V - 2\sigma^2\sigma_\delta^2 & -|\Sigma| + \frac{\sigma^4}{\mu^2}V + 2\sigma^2\sigma_\varepsilon^2 \\ & & & \frac{|\Sigma|}{\beta^2} + \frac{\sigma^4}{\beta^2\mu^2}V + 2\sigma_\delta^4 & |\Sigma| - \frac{\sigma^4}{\mu^2}V - 2\sigma^2\sigma_\varepsilon^2 \\ & & & & \beta^2|\Sigma| + \frac{\beta^2\sigma^4}{\mu^2}V + 2\sigma_\varepsilon^4 \end{pmatrix}$$

$$B_1 = \frac{1}{n}\begin{pmatrix} 0 & 0 & \mu_{\xi 3} & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & & \mu_{\xi 4} - 3\sigma^4 & 0 & 0 \\ & & & 0 & 0 \\ & & & & 0 \end{pmatrix}$$

$$C_1 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & & 0 & \frac{\sigma^2}{\mu}\mu_{\delta 3} & -\frac{\sigma^2}{\beta\mu}\mu_{\varepsilon 3} \\ & & & -2\frac{\sigma^2}{\mu}\mu_{\delta 3} & \frac{\sigma^2}{\beta\mu}(\mu_{\varepsilon 3} + \beta^3\mu_{\delta 3}) \\ & & & & -2\frac{\beta\sigma^2}{\mu}\mu_{\varepsilon 3} \end{pmatrix}$$

Error variance σ_δ^2 known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_2 = \frac{s_{xy}}{s_{xx} - \sigma_\delta^2}.$$

Since σ_δ^2 is assumed known, the variance covariance matrix for $\tilde{\mu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2$ and $\tilde{\sigma}_\varepsilon^2$ is required.

$$A_2 = \frac{1}{n}\begin{pmatrix} U & -\beta\sigma_\delta^2 & 0 & 0 & 0 \\ & \frac{\mu^2}{\sigma^4}(|\Sigma| + 2\beta^2\sigma_\delta^4) + V & -\frac{\mu}{\sigma^4}(|\Sigma| + 2\beta^2\sigma_\delta^4) & \frac{2\mu\beta\sigma_\delta^2}{\sigma^2}U & \frac{2\mu\beta\sigma_\delta^2}{\sigma^2}V \\ & & \frac{1}{\sigma^4}(|\Sigma| + 2\beta^2\sigma_\delta^4) & -\frac{2\beta\sigma_\delta^2}{\sigma^2}U & -\frac{2\beta\sigma_\delta^2}{\sigma^2}V \\ & & & 2U^2 & 2\beta^2\sigma_\delta^4 \\ & & & & 2V^2 \end{pmatrix}$$

$$B_2 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & \mu_{\xi 3} & 0 \\ & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ & & & \mu_{\xi 4} - 3\sigma^4 & 0 \\ & & & & 0 \end{pmatrix}$$

$$C_2 = \frac{1}{n}\begin{pmatrix} 0 & \frac{\beta\mu\mu_{\delta 3}}{\sigma^2} & -\frac{\beta\mu_{\delta 3}}{\sigma^2} & \mu_{\delta 3} & \beta^2\mu_{\delta 3} \\ & -\frac{2\beta^2\mu\mu_{\delta 3}}{\sigma^2} & \frac{\beta^2\mu_{\delta 3}}{\sigma^2} & -\beta\mu_{\delta 3} & \mu_{\varepsilon 3} - \beta^3\mu_{\delta 3} \\ & & \frac{\beta^2}{\sigma^4}e_1 & -\frac{\beta}{\sigma^2}e_1 & -\frac{\beta^3}{\sigma^2}e_1 \\ & & & e_1 & \beta^2 e_1 \\ & & & & e_2 + \beta^4 e_1 \end{pmatrix}$$

Error variance σ_ε^2 known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_3 = \frac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}}.$$

Since σ_ε^2 is assumed known, the variance covariance matrix for $\tilde{\mu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2$ and $\tilde{\sigma}_\delta^2$ is required. For brevity, the notation $W = (\beta^2|\Sigma| + 2\sigma_\varepsilon^4)$ is introduced.

$$A_3 = \frac{1}{n}\begin{pmatrix} U & -\beta\sigma_\delta^2 & 0 & 0 & 0 \\ & \frac{\mu^2}{\beta^2\sigma^4}W + V & -\frac{\mu}{\beta^2\sigma^4}W & \frac{2\mu}{\beta^3\sigma^2}(\sigma_\varepsilon^2 V + \beta^4\sigma^2\sigma_\delta^2) & -\frac{2\mu\sigma_\varepsilon^2 V}{\beta^3\sigma^2} \\ & & \frac{1}{\beta^2\sigma^4}W & -\frac{2}{\beta^3\sigma^2}(\sigma_\varepsilon^2 V + \beta^4\sigma^2\sigma_\delta^2) & \frac{2\sigma_\varepsilon^2 V}{\beta^3\sigma^2} \\ & & & \frac{2}{\beta^4}(\beta^4 U^2 + V^2 - 2\beta^4\sigma_\delta^4) & -\frac{2}{\beta^4}(\sigma_\varepsilon^2 V + 2\beta^2\sigma_\delta^2\sigma_\varepsilon^2) \\ & & & & \frac{2V^2}{\beta^4} \end{pmatrix}$$

$$B_3 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & \mu_{\xi 3} & 0 \\ & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ & & & \mu_{\xi 4} - 3\sigma^4 & 0 \\ & & & & 0 \end{pmatrix}$$

$$C_3 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & 0 & \mu_{\delta 3} \\ & -\frac{2\mu\mu_{\varepsilon 3}}{\beta\sigma^2} & \frac{\mu_{\varepsilon 3}}{\beta\sigma^2} & -\frac{\mu_{\varepsilon 3}}{\beta^2} & \frac{\mu_{\varepsilon 3}}{\beta^2} - \beta\mu_{\delta 3} \\ & & \frac{1}{\beta^2\sigma^4}e_2 & -\frac{1}{\beta^3\sigma^2}e_2 & \frac{1}{\beta^3\sigma^2}e_2 \\ & & & \frac{1}{\beta^4}e_2 & -\frac{1}{\beta^4}e_2 \\ & & & & \frac{1}{\beta^4}(\beta^4 e_1 + e_2) \end{pmatrix}$$

Reliability ratio $\kappa = \frac{\sigma^2}{\sigma^2+\sigma_\delta^2}$ known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_4 = \frac{s_{xy}}{\kappa s_{xx}}.$$

The variance covariance matrix for $\tilde{\mu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2$ and $\tilde{\sigma}_\delta^2$ is required. For brevity, the notation $\varpi = 1 - \kappa$ is introduced.

$$A_4 = \frac{1}{n}\begin{pmatrix} U & -\beta\sigma_\delta^2 & 0 & 0 & 0 \\ & \mu^2\frac{|\Sigma|}{\sigma^4} + V & -\mu\frac{|\Sigma|}{\sigma^4} & 0 & 2\mu\beta\frac{\varpi}{\sigma^2}|\Sigma| \\ & & \frac{|\Sigma|}{\sigma^4} & 0 & -2\beta\frac{\varpi}{\sigma^2}|\Sigma| \\ & & & 2\sigma^4 & -2\beta^2\kappa\sigma^2\sigma_\delta^2 \\ & & & & 4\beta^2\varpi|\Sigma| + 2\sigma_\varepsilon^4 \end{pmatrix}$$

$$B_4 = \frac{1}{n}\begin{pmatrix} 0 & -\mu\frac{\beta\varpi}{\sigma^2}\mu_{\xi 3} & -\frac{\beta\varpi}{\sigma^2}\mu_{\xi 3} & \mu_{\xi 3} & -\beta^2\varpi\mu_{\xi 3} \\ & 0 & 0 & 0 & 0 \\ & & \frac{\beta^2\varpi^2}{\sigma^4}(\mu_{\xi 4} - 3\sigma^4) & \frac{\beta\kappa\varpi}{\sigma^2}(\mu_{\xi 4} - 3\sigma^4) & -\frac{\beta^3\varpi^2}{\sigma^2}(\mu_{\xi 4} - 3\sigma^4) \\ & & & \kappa^2(\mu_{\xi 4} - 3\sigma^4) & -\beta^2\kappa\varpi(\mu_{\xi 4} - 3\sigma^4) \\ & & & & \beta^4\varpi^2(\mu_{\xi 4} - 3\sigma^4) \end{pmatrix}$$

$$C_4 = \frac{1}{n}\begin{pmatrix} 0 & \mu\frac{\beta\kappa}{\sigma^2}\mu_{\delta 3} & -\frac{\beta\kappa}{\sigma^2}\mu_{\delta 3} & \kappa\mu_{\delta 3} & \beta^2\kappa\mu_{\delta 3} \\ & -2\mu\frac{\beta^2\kappa}{\sigma^2}\mu_{\delta 3} & \frac{\beta^2\kappa}{\sigma^2}\mu_{\delta 3} & -\beta\kappa\mu_{\delta 3} & -\beta^3\kappa\mu_{\delta 3} + \mu_{\varepsilon 3} \\ & & \frac{\beta^2\kappa^2}{\sigma^4}e_1 & -\frac{\beta\kappa^2}{\sigma^2}e_1 & -\frac{\beta^3\kappa^2}{\sigma^2}e_1 \\ & & & \kappa^2 e_1 & \beta^2\kappa^2 e_1 \\ & & & & \beta^4\kappa^2 e_1 + e_2 \end{pmatrix}$$

Ratio of the error variances $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\delta^2}$ known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_5 = \frac{(s_{yy} - \lambda s_{xx}) + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda(s_{xy})^2}}{2s_{xy}}.$$

The variance covariance matrix for $\tilde{\mu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2$ and $\tilde{\sigma}_\delta^2$ is required.

$$A_5 = \frac{1}{n}\begin{pmatrix} U & -\beta\sigma_\delta^2 & 0 & 0 & 0 \\ & \mu^2\frac{|\Sigma|}{\sigma^4} + V & -\mu\frac{|\Sigma|}{\sigma^4} & \frac{2\mu\beta}{(\beta^2+\lambda)\sigma^2}|\Sigma| & 0 \\ & & \frac{|\Sigma|}{\sigma^4} & -\frac{2\beta}{(\beta^2+\lambda)\sigma^2}|\Sigma| & 0 \\ & & & 2\sigma^4 + \frac{4|\Sigma|}{(\beta^2+\lambda)} & -\frac{2\sigma_\delta^2\sigma_\epsilon^2}{(\beta^2+\lambda)} \\ & & & & 2\sigma_\delta^4 \end{pmatrix}$$

$$B_5 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & \mu_{\xi 3} & 0 \\ & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ & & & \mu_{\xi 4} - 3\sigma^4 & 0 \\ & & & & 0 \end{pmatrix}$$

$$C_5 = \frac{1}{n}\begin{pmatrix} 0 & \frac{\mu\lambda\beta}{(\beta^2+\lambda)\sigma^2}\mu_{\delta 3} & -\frac{\lambda\beta}{(\beta^2+\lambda)\sigma^2}\mu_{\delta 3} & \frac{\lambda}{(\beta^2+\lambda)}\mu_{\delta 3} & \frac{\beta^2}{(\beta^2+\lambda)}\mu_{\delta 3} \\ & -2\frac{\mu\beta}{(\beta^2+\lambda)\sigma^2}e_3 & \frac{\beta}{(\beta^2+\lambda)\sigma^2}e_3 & -\frac{e_3}{(\beta^2+\lambda)} & \frac{\beta}{(\beta^2+\lambda)}e_3 - \beta\mu_{\delta 3} \\ & & \frac{\beta^2 e_2+\lambda^2\beta^2 e_1}{(\beta^2+\lambda)^2\sigma^4} & -\frac{(\beta e_2+\lambda^2\beta e_1)}{(\beta^2+\lambda)^2\sigma^2} & \frac{\beta e_2-\lambda\beta^3 e_1}{(\beta^2+\lambda)^2\sigma^2} \\ & & & \frac{e_2+\lambda^2 e_1}{(\beta^2+\lambda)^2} & -\frac{(e_2+\lambda\beta^2 e_1)}{(\beta^2+\lambda)^2} \\ & & & & \frac{e_2+\beta^4 e_1}{(\beta^2+\lambda)^2} \end{pmatrix}$$

Both variances σ_δ^2 and σ_ε^2 known The method of moments estimator for the slope based on this assumption is

$$\tilde{\beta}_7 = sgn(s_{xy})\sqrt{\frac{s_{yy} - \sigma_\varepsilon^2}{s_{xx} - \sigma_\delta^2}}.$$

The variance covariance matrix for $\tilde{\mu}, \tilde{\alpha}, \tilde{\beta}$ and $\tilde{\sigma}^2$ is required. For brevity, the notation

$$T = \frac{|\Sigma|}{\sigma^4} + \frac{(\beta^2\sigma_\delta^2 - \sigma_\varepsilon^2)}{2\beta^2\sigma^4}$$

is introduced.

$$A_7 = \frac{1}{n}\begin{pmatrix} U & -\beta\sigma_\delta^2 & 0 & 0 \\ & \mu^2 T + V & -\mu T & \frac{\beta\mu\sigma_\delta^2}{\sigma^2}(U+\sigma^2) \\ & & T & -\frac{\beta\sigma_\delta^2}{\sigma^2}(U+\sigma^2) \\ & & & \mu_{\xi 4} - 3\sigma^4 \end{pmatrix}$$

$$B_7 = \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & \mu_{\xi 3} \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 2V^2 \end{pmatrix}$$

$$C_7 = \frac{1}{n}\begin{pmatrix} 0 & \frac{\mu\beta\mu_{\delta 3}}{2\sigma^2} & -\frac{\beta\mu_{\delta 3}}{2\sigma^2} & \mu_{\delta 3} \\ & -\frac{\beta^2\mu}{\sigma^2}\mu_{\delta 3} - \frac{\mu}{\beta\sigma^2}\mu_{\varepsilon 3} & \frac{\beta^2}{2\sigma^2}\mu_{\delta 3} + \frac{\mu\varepsilon 3}{2\beta\sigma^2} & -\beta\mu_{\delta 3} \\ & & \frac{\beta^2}{4\sigma^4}e_1 + \frac{e_2}{\beta^2\sigma^4} & -\frac{\beta}{2\sigma^2}e_1 \\ & & & e_1 \end{pmatrix}$$

3.5.3 Description of Matrices

There are some common themes and patterns which run through the variance covariance matrices. For each of the A matrices for example, $Var[\tilde{\mu}] = \frac{U}{n}$, $Var[\tilde{\alpha}] = \mu^2 Var[\tilde{\beta}] + \frac{V}{n}$, and $Cov[\tilde{\mu}, \tilde{\alpha}] = -\beta\sigma_\delta^2$. $\tilde{\mu}$ is also uncorrelated with $\tilde{\beta}$, and the variance estimators. $Var[\tilde{\beta}]$ and $Cov[\tilde{\alpha}, \tilde{\beta}]$ are different in each case. Patterns between rows and columns of the A matrices were reported by Hood et al.[56].

As can be seen the matrix B, reflecting skewness and kurtosis in the distribution of ξ, is generally sparse, although the B_4 matrix is more complicated. For the cases other than the reliability ratio κ known there are only corrections for $Var[\tilde{\sigma}^2]$ and $Cov[\tilde{\mu}, \tilde{\sigma}^2]$.

The C_1 matrix is more sparse than any other C matrix. $Var[\tilde{\sigma}_\delta^2]$ and $Var[\tilde{\sigma_\varepsilon}^2]$ depend on the skewness and kurtosis of δ and ε. $Cov[\tilde{\sigma_\delta}^2, \tilde{\sigma_\varepsilon}^2]$ depends on the skewness of δ and ε. The remaining covariances involving $\tilde{\sigma_\delta}^2$ depend solely on the skewness of δ, whilst the remaining covariances involving $\tilde{\sigma_\varepsilon}^2$ depend solely on the skewness of ε.

For the C_2 matrix, each of the variances and covariances involving $\tilde{\mu}$ and $\tilde{\alpha}$ are affected by skewness in δ and ε but not kurtosis. The variances and covariances involving $\tilde{\beta}$, $\tilde{\sigma}^2$ and $\tilde{\sigma_\varepsilon}^2$ are affected by kurtosis in δ and ε but not skewness.

We have an identical pattern for the matrices C_3, C_4 and C_5, except variances and covariances of $\tilde{\sigma_\delta}^2$ replace variances and covariances of $\tilde{\sigma_\varepsilon}^2$. Each of the variances and covariances involving $\tilde{\mu}$ and $\tilde{\alpha}$ are affected by skewness in δ and ε but not kurtosis. The variances and covariances involving $\tilde{\beta}$, $\tilde{\sigma}^2$ and $\tilde{\sigma_\delta}^2$ are affected by kurtosis in δ and ε but not skewness. C_5 has a much more complicated structure than the other C

matrices.

3.5.4 Variances and Covariances for Higher Moment Estimators

The methodology underlying the derivation of the asymptotic variances and covariances for estimators based on higher moments is identical to that outlined previously. However, the algebraic expressions for the variances and covariances of higher moment based estimators are longer and more cumbersome than those for the restricted parameter space. As a result, the full variance covariance matrices for higher moment estimators will not be reported here. However, the expressions needed to work out the full variance covariance matrices for the slope estimator based on third moments will be provided. These expressions can then be substituted into the shortcut formulae to derive the full variance covariance matrices.

Estimator based on Third Moments The estimator for the slope β based on the third order moments derived earlier is

$$\tilde{\beta}_8 = \frac{s_{xyy}}{s_{xxy}}.$$

In order to use the shortcut equations outlined in Section 3.5.1, the quantities $Cov[\bar{x}, \tilde{\beta}_8]$, $Cov[\bar{y}, \tilde{\beta}_8]$, $Cov[s_{xx}, \tilde{\beta}_8]$, $Cov[s_{xy}, \tilde{\beta}_8]$ and $Cov[s_{yy}, \tilde{\beta}_8]$ are needed. Further, to obtain these quantities, the covariances between each of the first and second order moments ($\bar{x}$, $\bar{y}$, s_{xx}, s_{xy}, s_{yy}) and the third order moments that occur in $\tilde{\beta}_8$ (s_{xxy}, s_{xyy}) must be obtained. Also, the variances of these third order moments must be obtained, as well as the covariance between them.

Using the method illustrated in deriving $Var[s_{xx}]$ and $Cov[s_{xx}, s_{xy}]$, the required covariances between the first, second order and third order moments are:

$$
\begin{aligned}
Var[s_{xxy}] &= \frac{\beta^2(\mu_{\xi 6} - \mu_{\xi 3}{}^2) + 6\beta^2\mu_{\xi 4}\sigma_\delta{}^2 + \mu_{\xi 4}\sigma_\varepsilon{}^2 + 4\beta^2\mu_{\xi 3}\mu_{\delta 3}}{n} \\
&+ \frac{\beta^2\sigma^2\mu_{\delta 4} + \mu_{\delta 4}\sigma_\varepsilon{}^2 + 6\sigma^2\sigma_\delta{}^2\sigma_\varepsilon{}^2}{n} \\
Var[s_{xyy}] &= \frac{\beta^4(\mu_{\xi 6} - \mu_{\xi 3}{}^2) + 6\beta^2\mu_{\xi 4}\sigma_\varepsilon{}^2 + \beta^4\mu_{\xi 4}\sigma_\delta{}^2 + 4\beta\mu_{\xi 3}\mu_{\varepsilon 3}}{n} \\
&+ \frac{\sigma^2\mu_{\varepsilon 4} + \sigma_\delta{}^2\mu_{\varepsilon 4} + 6\beta^2\sigma^2\sigma_\delta{}^2\sigma_\varepsilon{}^2}{n} \\
Cov[\bar{x}, s_{xxy}] &= \frac{\beta\left(\mu_{\xi 4} + 3\sigma^2\sigma_\delta{}^2\right)}{n} \\
Cov[\bar{x}, s_{xyy}] &= \frac{\beta^2\mu_{\xi 4} + \sigma_\delta{}^2\sigma_\varepsilon{}^2 + \beta^2\sigma^2\sigma_\delta{}^2 + \sigma^2\sigma_\varepsilon{}^2}{n} \\
Cov[\bar{y}, s_{xxy}] &= \frac{\beta^2\mu_{\xi 4} + \sigma_\delta{}^2\sigma_\varepsilon{}^2 + \beta^2\sigma^2\sigma_\delta{}^2 + \sigma^2\sigma_\varepsilon{}^2}{n} \\
Cov[\bar{y}, s_{xyy}] &= \frac{\beta\left(\beta^2\mu_{\xi 4} + 3\sigma^2\sigma_\varepsilon{}^2\right)}{n} \\
Cov[s_{xx}, s_{xxy}] &= \frac{\beta\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + 5\beta\mu_{\xi 3}\sigma_\delta{}^2 + 4\beta\sigma^2\mu_{\delta 3}}{n} \\
Cov[s_{xy}, s_{xxy}] &= \frac{\beta^2\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + 3\beta^2\mu_{\xi 3}\sigma_\delta{}^2 + \sigma_\varepsilon{}^2\left(\mu_{\xi 3} + \mu_{\delta 3}\right) + \beta^2\sigma^2\mu_{\delta 3}}{n} \\
Cov[s_{yy}, s_{xxy}] &= \frac{\beta^3\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + \sigma^2\mu_{\varepsilon 3} + \sigma_\delta{}^2\mu_{\varepsilon 3} + \beta^3\mu_{\xi 3}\sigma_\delta{}^2 + 2\beta\mu_{\xi 3}\sigma_\varepsilon{}^2}{n} \\
Cov[s_{xx}, s_{xyy}] &= \frac{\beta^2\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + 2\beta^2\mu_{\xi 3}\sigma_\delta{}^2 + \beta^2\sigma^2\mu_{\delta 3} + \mu_{\xi 3}\sigma_\varepsilon{}^2 + \mu_{\delta 3}\sigma_\varepsilon{}^2}{n} \\
Cov[s_{xy}, s_{xyy}] &= \frac{\beta^3\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + 3\beta\mu_{\xi 3}\sigma_\varepsilon{}^2 + \sigma^2\mu_{\varepsilon 3} + \sigma_\delta{}^2\mu_{\varepsilon 3} + \beta^3\mu_{\xi 3}\sigma_\delta{}^2}{n} \\
Cov[s_{yy}, s_{xyy}] &= \frac{\beta^4\left(\mu_{\xi 5} - \sigma^2\mu_{\xi 3}\right) + 5\beta^2\mu_{\xi 3}\sigma_\varepsilon{}^2 + 4\beta\sigma^2\mu_{\varepsilon 3}}{n} \\
Cov[s_{xxy}, s_{xyy}] &= \frac{\beta^3\mu_{\xi 6} - \beta^3\mu_{\xi 3}{}^2 + 3\beta\mu_{\xi 4}\sigma_\varepsilon{}^2 + 3\beta^3\mu_{\xi 4}\sigma_\delta{}^2}{n} \\
&+ \frac{\mu_{\xi 3}\mu_{\varepsilon 3} + \beta^3\mu_{\xi 3}\mu_{\delta 3} + 9\beta\sigma^2\sigma_\delta{}^2\sigma_\varepsilon{}^2 + \mu_{\delta 3}\mu_{\varepsilon 3}}{n}
\end{aligned}
$$

By using the methodology outlined at the beginning of Section 3.5, we can now obtain the variance of our slope estimator $\tilde{\beta}_8$, and the covariances of our slope estimator with the first and second order moments.

$$
\begin{aligned}
Var[\tilde{\beta}_8] &= \frac{\beta^2\mu_{\xi4}{\sigma_\varepsilon}^2 + \beta^4\mu_{\xi4}{\sigma_\delta}^2 + 2\beta\mu_{\xi3}\mu_{\varepsilon3} + \sigma^2\mu_{\varepsilon4} + {\sigma_\delta}^2\mu_{\varepsilon4} - 6\beta^2\sigma^2{\sigma_\delta}^2{\sigma_\varepsilon}^2}{\beta^2\mu_{\xi3}^2 n} \\
&\quad + \frac{2\beta^4\mu_{\xi3}\mu_{\delta3} + \beta^4\sigma^2\mu_{\delta4} + \beta^2\mu_{\delta4}{\sigma_\varepsilon}^2 - 2\beta\mu_{\delta3}\mu_{\varepsilon3}}{\beta^2\mu_{\xi3}^2 n} \\
Cov[\bar{x}, \tilde{\beta}_8] &= \frac{|\Sigma| - 3\beta^2\sigma^2\sigma_\delta^2}{\beta\mu_{\xi3}n} \\
Cov[\bar{y}, \tilde{\beta}_8] &= \frac{3\sigma^2\sigma_\varepsilon^2 - |\Sigma|}{\mu_{\xi3}n} \\
Cov[s_{xx}, \tilde{\beta}_8] &= \frac{-3\beta^2\mu_{\xi3}{\sigma_\delta}^2 - 3\beta^2\sigma^2\mu_{\delta3} + \mu_{\xi3}{\sigma_\varepsilon}^2 + \mu_{\delta3}{\sigma_\varepsilon}^2}{\beta\mu_{\xi3}n} \\
Cov[s_{xy}, \tilde{\beta}_8] &= \frac{2\beta\mu_{\xi3}{\sigma_\varepsilon}^2 + \sigma^2\mu_{\varepsilon3} + {\sigma_\delta}^2\mu_{\varepsilon3} - 2\beta^3\mu_{\xi3}{\sigma_\delta}^2 - \beta\mu_{\delta3}{\sigma_\varepsilon}^2 - \beta^3\sigma^2\mu_{\delta3}}{\beta\mu_{\xi3}n} \\
Cov[s_{yy}, \tilde{\beta}_8] &= \frac{3\beta\mu_{\xi3}{\sigma_\varepsilon}^2 + 3\sigma^2\mu_{\varepsilon3} - {\sigma_\delta}^2\mu_{\varepsilon3} - \beta^3\mu_{\xi3}{\sigma_\delta}^2}{\mu_{\xi3}n}
\end{aligned}
$$

If Normal errors δ and ε are assumed, then the variance of $\tilde{\beta}_8$ may be simplified to

$$Var[\tilde{\beta}_8] = \frac{1}{\beta^2\mu_{\xi3}^2 n}\left[\beta^2\mu_{\xi4}(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2) + 3\sigma_\varepsilon^4(\sigma^2 + \sigma_\delta^2) + 3\beta^2\sigma_\delta^4(\sigma_\varepsilon^2 + \beta^2\sigma^2) - 6\beta^2\sigma_\delta^2\sigma_\varepsilon^2\right].$$

The $\mu_{\xi3}$ in the denominator emphasises the importance of having skewed ξ.

For any distribution of ξ, and for any assumed distribution concerning the error terms δ and ε, $Var[\tilde{\alpha}]$ is relatively straightforward in terms of $Var[\tilde{\beta}_8]$:

$$Var[\tilde{\alpha}] = \mu^2 Var[\tilde{\beta}_8] + \frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{n} + \frac{2\mu}{\mu_{\xi3}n}\left[2|\Sigma| - 3\sigma^2(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)\right],$$

and similarly

$$Cov[\tilde{\alpha}, \tilde{\beta}_8] = -\mu Var[\tilde{\beta}_8] + \frac{1}{\mu_{\xi3}n}\left[|\Sigma| - 3\sigma_\delta^2\sigma_\varepsilon^2\right].$$

The formulas for $Var[\tilde{\alpha}]$ and $Cov[\tilde{\alpha}, \tilde{\beta}_8]$ can be seen to be fairly neat, and are easily estimated by using the method of moment estimating equations (3.1) to (3.5).

Formulas for the variances and covariances of the variance estimators are not as straightforward. These, however, are less likely to be of interest than the ones given above. Nevertheless, we now have each of the components needed to use the short-cut formulae to obtain the following variance covariance matrix for the parameters $\mu, \alpha, \beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2 when the estimator $\tilde{\beta}_8$ is used. The complete variance covariance matrix will not be reported here, but a practitioner now has the tool to compute it if needed. In addition, the Maple programme described towards the end of this Chapter does have the capability to create the entire variance covariance matrix for $\mu, \alpha, \beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2 when the estimator $\tilde{\beta}_8$ is used, if it is needed.

Estimator based on Fourth Moments The estimator for the slope β based on fourth order moments derived earlier is

$$\tilde{\beta}_9 = \sqrt{\frac{s_{xyyy} - 3s_{xy}s_{yy}}{s_{xxxy} - 3s_{xx}s_{xy}}}$$

In order to use the shortcut equations outlined in Section 3.5.1, the quantities $Cov[\bar{x}, \tilde{\beta}_9]$, $Cov[\bar{y}, \tilde{\beta}_9]$, $Cov[s_{xx}, \tilde{\beta}_9]$, $Cov[s_{xy}, \tilde{\beta}_9]$ and $Cov[s_{yy}, \tilde{\beta}_9]$ are needed. Further, to obtain these quantities, the covariances between each of the first and second order moments ($\bar{x}$, $\bar{y}$, s_{xx}, s_{xy}, s_{yy}) and the fourth order moments that occur in $\tilde{\beta}_9$ (s_{xyyy}, s_{xxxy}) must be obtained. Also, the variances of these fourth order moments must be obtained, as well as the covariance between them. For brevity of algebra, variances and the covariances of the sample moments are only presented here.

Using the method illustrated in deriving $Var[s_{xx}]$ and $Cov[s_{xx}, s_{xy}]$, the required covariances between the first order and fourth order moments are:

$$
\begin{aligned}
Cov[\bar{x}, s_{xyyy}] &= \frac{\beta^3\mu_{\xi 5} + 3\beta\mu_{\xi 3}\sigma_\varepsilon^2 + \beta^3\mu_{\xi 3}\sigma_\delta^2 + \mu_{\varepsilon 3}(\sigma^2 + \sigma_\delta^2)}{n} \\
Cov[\bar{y}, s_{xyyy}] &= \frac{\beta^4\mu_{\xi 5} + 6\beta^2\mu_{\xi 3}\sigma_\varepsilon^2 + 4\beta\sigma^2\mu_{\varepsilon 3}}{n} \\
Cov[\bar{x}, s_{xxxy}] &= \frac{\beta\mu_{\xi 5} + 6\beta\mu_{\xi 3}\sigma_\delta^2 + 4\beta\sigma^2\mu_{\delta 3}}{n} \\
Cov[\bar{y}, s_{xxxy}] &= \frac{\beta^2\mu_{\xi 5} + 3\beta^2\mu_{\xi 3}\sigma_\delta^2 + \beta^2\sigma^2\mu_{\delta 3} + \sigma_\varepsilon^2(\mu_{\xi 3} + \mu_{\delta 3})}{n}
\end{aligned}
$$

The required covariances between the second order and fourth order moments are:

$$
\begin{aligned}
Cov[s_{xx}, s_{xyyy}] &= \frac{\beta^3\mu_{\xi 6} + 2\,\beta^3\mu_{\xi 4}{\sigma_\delta}^2 + 3\,\beta\,\mu_{\xi 4}{\sigma_\varepsilon}^2 + \mu_{\xi 3}\mu_{\varepsilon 3} + \beta^3\mu_{\xi 3}\mu_{\delta 3} + \mu_{\delta 3}\mu_{\varepsilon 3}}{n} \\
&\quad + \frac{6\,\beta\,\sigma^2{\sigma_\delta}^2{\sigma_\varepsilon}^2 - \sigma^2\beta^3\mu_{\xi 4} - 3\,\sigma^4\beta\,{\sigma_\varepsilon}^2}{n} \\
Cov[s_{xy}, s_{xyyy}] &= \frac{\beta^4\mu_{\xi 6} + 6\,\beta^2\mu_{\xi 4}{\sigma_\varepsilon}^2 + \beta^4\mu_{\xi 4}{\sigma_\delta}^2 + 4\,\beta\,\mu_{\xi 3}\mu_{\varepsilon 3} + 3\,\beta^2\sigma^2{\sigma_\delta}^2{\sigma_\varepsilon}^2}{n} \\
&\quad + \frac{\sigma^2\mu_{\varepsilon 4} + {\sigma_\delta}^2\mu_{\varepsilon 4} - \beta^4\sigma^2\mu_{\xi 4} - 3\,\beta^2\sigma^4{\sigma_\varepsilon}^2}{n} \\
Cov[s_{yy}, s_{xyyy}] &= \frac{\beta\,(\beta^4\mu_{\xi 6} + 9\,\beta^2\mu_{\xi 4}{\sigma_\varepsilon}^2 + 10\beta\,\mu_{\xi 3}\mu_{\varepsilon 3} + 5\,\sigma^2\mu_{\varepsilon 4})}{n} \\
&\quad - \frac{\beta(\beta^4\sigma^2\mu_{\xi,4} + 3\,\beta^2\sigma^4{\sigma_\varepsilon}^2 + 3\,{\sigma_\varepsilon}^4\sigma^2)}{n} \\
Cov[s_{xx}, s_{xxxy}] &= \frac{\beta\,(\mu_{\xi 6} + 9\,\mu_{\xi 4}{\sigma_\delta}^2 + 10\,\mu_{\xi 3}\mu_{\delta 3} + 5\,\sigma^2\mu_{\delta 4})}{n} \\
&\quad - \frac{\beta(\sigma^2\mu_{\xi 4} + 3\,\sigma^4{\sigma_\delta}^2 + 3\,{\sigma_\delta}^4\sigma^2)}{n} \\
Cov[s_{xy}, s_{xxxy}] &= \frac{\beta^2\mu_{\xi 6} + 6\,\beta^2\mu_{\xi 4}{\sigma_\delta}^2 + \mu_{\xi 4}{\sigma_\varepsilon}^2 + 4\,\beta^2\mu_{\xi 3}\mu_{\delta 3} + \mu_{\delta 4}{\sigma_\varepsilon}^2 + \beta^2\sigma^2\mu_{\delta 4}}{n} \\
&\quad + \frac{6\,\sigma^2{\sigma_\delta}^2{\sigma_\varepsilon}^2 - \beta^2\sigma^2\mu_{\xi 4} - 3\,\beta^2\sigma^4{\sigma_\delta}^2}{n} \\
Cov[s_{yy}, s_{xxxy}] &= \frac{\beta^3\mu_{\xi 6} + 3\,\beta^3\mu_{\xi 4}{\sigma_\delta}^2 + 2\,\beta\,\mu_{\xi 4}{\sigma_\varepsilon}^2 + \beta^3\mu_{\xi 3}\mu_{\delta 3} + \mu_{\xi 3}\mu_{\varepsilon 3} + \mu_{\delta 3}\mu_{\varepsilon 3}}{n} \\
&\quad + \frac{6\,\beta\,\sigma^2{\sigma_\delta}^2{\sigma_\varepsilon}^2 - \sigma^2\beta^3\mu_{\xi 4} - 3\,\beta^3\sigma^4{\sigma_\delta}^2}{n}
\end{aligned}
$$

Let

$$\begin{aligned}
u_1 &= 15\beta^4\sigma_\varepsilon^2 + \beta^6\sigma_\delta^2 \\
u_2 &= 15\beta^4\sigma_\delta^2\sigma_\varepsilon^2 + 15\beta^2\mu_{\varepsilon 4} - \beta^6\mu_{\xi 4} - 6\beta^4\sigma^2\sigma_\varepsilon^2 \\
u_3 &= 20\beta^3\sigma_\delta^2\mu_{\varepsilon 3} + 6\beta\mu_{\varepsilon 5} \\
u_4 &= 15\beta^2\sigma_\delta^2\mu_{\varepsilon 4} + \mu_{\varepsilon 6} - 9\beta^2\sigma^2\sigma_\varepsilon^4 \\
v_1 &= 15\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2 \\
v_2 &= 15\sigma_\delta^2\sigma_\varepsilon^2 + 15\beta^2\mu_{\delta 4} - \beta^2\mu_{\xi 4} - 6\beta^2\sigma^2\sigma_\delta^2 \\
v_3 &= 20\mu_{\delta 3}\sigma_\varepsilon^2 + 6\beta^2\mu_{\delta 5} \\
v_4 &= 15\mu_{\delta 4}\sigma_\varepsilon^2 + \beta^2\mu_{\delta 6} - 9\beta^2\sigma^2\sigma_\delta^4.
\end{aligned}$$

Then, the variances of the fourth order moments are:

$$\begin{aligned}
Var[s_{xyyy}] &= \frac{\beta^6\mu_{\xi 8} + \mu_{\xi 6}u_1 + \mu_{\xi 4}u_2 + 20\beta^3\mu_{\xi 5}\mu_{\varepsilon 3} + \mu_{\xi 3}u_3 + \sigma^2 u_4 + \sigma_\delta^2\mu_{\varepsilon 6}}{n} \\
Var[s_{xxxy}] &= \frac{\beta^2\mu_{\xi 8} + \mu_{\xi 6}v_1 + \mu_{\xi 4}v_2 + 20\beta^2\mu_{\xi 5}\mu_{\delta 3} + \mu_{\xi 3}v_3 + \sigma^2 v_4 + \mu_{\delta 6}\sigma_\varepsilon^2}{n}
\end{aligned}$$

Making the substitutions

$$\begin{aligned}
c_1 &= 6\beta^4\sigma_\delta^2 + 6\beta^2\sigma_\varepsilon^2 \\
c_2 &= 4\beta\mu_{\varepsilon 3} + 4\beta^4\mu_{\delta 3} \\
c_3 &= \beta^4\mu_{\delta 4} + \mu_{\varepsilon 4} - 3\beta^2\sigma^2\sigma_\varepsilon^2 - 3\beta^4\sigma^2\sigma_\delta^2 + 30\beta^2\sigma_\delta^2\sigma_\varepsilon^2 \\
c_4 &= 15\beta^2\mu_{\delta 3}\sigma_\varepsilon^2 + 24\beta\sigma_\delta^2\mu_{\varepsilon 3} \\
c_5 &= 10\beta\sigma^2\mu_{\delta 3}\mu_{\varepsilon 3} + 6\sigma^2\sigma_\delta^2\mu_{\varepsilon 4} + 6\beta^2\sigma^2\mu_{\delta 4}\sigma_\varepsilon^2 + \mu_{\delta 4}\mu_{\varepsilon 4} - 9\beta^2\sigma^4\sigma_\delta^2\sigma_\varepsilon^2
\end{aligned}$$

then the covariance between the two fourth moments in question is

$$Cov[s_{xyyy}, s_{xxxy}] = \frac{\beta^4\mu_{\xi 8} + \mu_{\xi 6}c_1 + \mu_{\xi 5}c_2 - \beta^4\mu_{\xi 4}^2 + \mu_{\xi 4}c_3 + \mu_{\xi 3}c_4 + c_5}{n}$$

For brevity of presentation, it is at this point the algebra for the variances and covariances for the fourth moment estimator is left. The expressions for the fourth moment estimator are particularly cumbersome. It is worth noting however, that $Var[\tilde{\beta}_9]$ depends on the sixth moment of ξ, $\mu_{\xi 6}$. Obtaining reliable estimators of this high order moment may be difficult. The variance of $Var[\tilde{\beta}_8]$ depends on the fourth moment of ξ, which may be estimated using the moment equations (3.19) to (3.23). However, a Maple 11 program has been created that allows the algebraic manipulation of the variance covariance matrices for all the estimators discussed in this Chapter, including the estimator of the slope based on fourth moments, $\tilde{\beta}_9$.

Further details and examples of how to use this Maple 11 program are included in Appendix A. This program enables a user to both manipulate the expressions of the variance covariance matrices, and to substitute numerical values into the variance covariance matrices without unnecessary effort. It is hoped that this program provides help for anyone wishing to theoretically analyse the variance covariance matrices and for practical use.

Chapter 4

Simulations

4.1 Introductory Remarks

The previous Chapter has introduced the method of moments as a method of estimating the parameters of the errors in variables model. This Chapter will use simulation to gain a deeper understanding of the estimators introduced earlier.

The typical questions regarding sample size and asymptotic results are considered in this Chapter, as well as the effect of having a small sample. A feature with errors in variables modelling not present with simple linear regression is that there are strict admissibility conditions. This is a key concept, as if the admissibility conditions are broken, then negative variance estimates may be obtained. As demonstrated in the previous Chapter, a simple form of the admissibility conditions is to ensure that the errors in variables slope estimator lies in between the slope estimators of x on y and y on x regression respectively.

4.2 Simulation to Assess Bias

From a maximum likelihood perspective, Hood [56] performed large scale simulations to ensure that her parameter estimates were asymptotically unbiased as predicted by standard asymptotic likelihood theory. As stated in the previous Chapter, the method of maximum likelihood provides identical estimators to the method of moments if the Normal structural model is assumed, so Hood's work provides guidance on the behaviour of the parameter estimators derived in Chapter 3 where the distribution of ξ, δ and ε are taken to be Normal as in the Normal structural model.

This section will investigate the effect of manipulating the distribution of ξ away from Normal. In particular both the use of the uniform and chi distribution (two degrees of freedom) for ξ will be investigated, as well as the Normal functional model, and comparisons drawn with the results of Hood for the Normal structural model. For completeness of presentation, some of the simulations ran by Hood to assess bias in the Normal structural model will be replicated and discussed, although this thesis details some different slope estimators from those described by Hood.

For brevity, the parameter β will be chosen for analysis. All other parameter estimators (apart from μ) may be written as functions of sample moments and β, thus the bias in β largely determines the bias in the other estimators. For example, $\tilde{\alpha} = \bar{y} - \tilde{\beta}\bar{x}$ and so an underestimated slope results in an overestimated intercept, and vice versa.

Hood chose the parameter settings $\mu = 1$, $\sigma = 2$, $\sigma_\delta = \sigma_\varepsilon = 1$, $\alpha = 0$ and $\beta = 1$ for her simulations. Her motivation for the choice of these settings for the true line was based on the premise that most method comparison studies expect some sort of identity between methods. It is these parameter settings that shall be used in the majority of the simulations that follow.

Normal structural model Figure 4.1 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for the Normal structural model. The bias in $\tilde{\beta}_5$ and $\tilde{\beta}_7$ is similar, and for clarity results for $\tilde{\beta}_7$ have been omitted. The parameter settings were identical to those of Hood. The different colours represent different slope estimators, as described by the following table.

Colour	Estimator of Slope
Red	$\tilde{\beta}_1 = \frac{\bar{y} - \alpha}{\bar{x}}$
Blue	$\tilde{\beta}_2 = \frac{s_{xy}}{s_{xx} - \sigma_\delta^2}$
Green	$\tilde{\beta}_3 = \frac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}}$
Black	$\tilde{\beta}_4 = \frac{s_{xy}}{\kappa s_{xx}}$
Brown	$\tilde{\beta}_5 = \frac{(s_{yy} - \lambda s_{xx}) + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda(s_{xy})^2}}{2s_{xy}}$
Pink	$\tilde{\beta}_7 = sgn(s_{xy})\sqrt{\frac{s_{yy} - \sigma_\varepsilon^2}{s_{xx} - \sigma_\delta^2}}$

This representation shall be used throughout this simulation Chapter. It can be seen that the biases present at the sample size $n = 20$ diminish as the sample size increases. Indeed, $\tilde{\beta}_3$ displays only a small underestimation of the true slope, and is thus close to being unbiased even for a relatively small sample of $n = 20$. It can also be seen that $\tilde{\beta}_4$ is virtually unbiased across the whole range of sample sizes. Knowledge of the reliability ratio enables a user to correct for the bias in the simple y on x estimator. This was discussed in the previous Chapter. $\tilde{\beta}_2$ displays a positive bias, but becomes approximately unbiased for larger n. $\tilde{\beta}_1$ seems to behave the most erratically, but starts to settle down for $n > 60$. The scale of Figure 4.1 suggests that all the slope estimators investigated have performed rather well, as a whole. For $n > 40$ the bias in all cases, except that of $\tilde{\beta}_1$, is less than 3%.

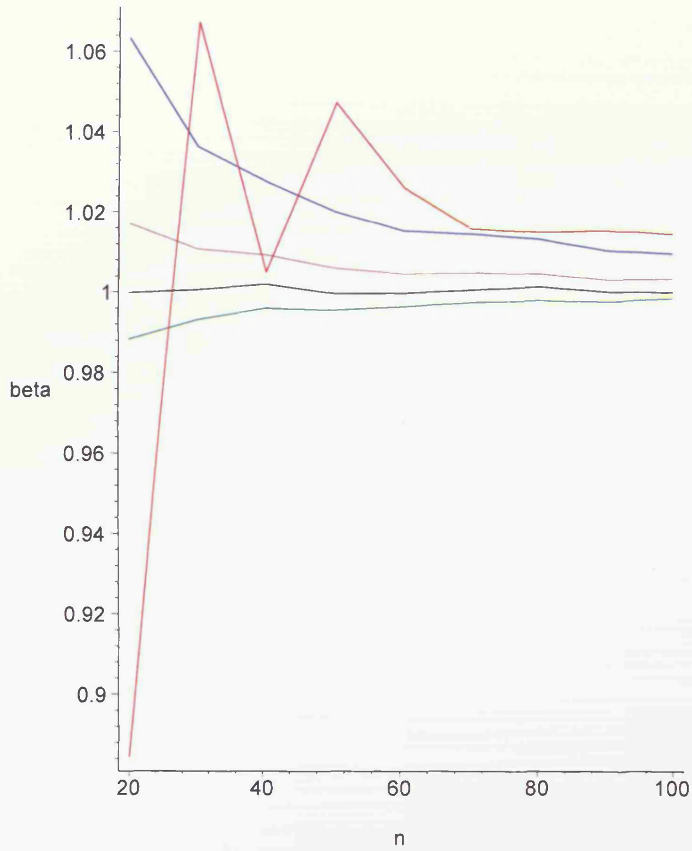


Figure 4.1: Estimate of β against sample size for the Normal structural model.

It appears that $\tilde{\beta}_2$ demonstrates a slight positive bias, and $\tilde{\beta}_3$ demonstrates a slight negative bias. Hood explained that this is likely to be a result of the sample quantities s_{xx}, s_{xy} and s_{yy} not being corrected for bias. Biases may be removed in sample variances and covariances by taking the denominator to be $(n-1)$ as opposed to n.

For example the slope estimator $\tilde{\beta}_2$ may be written

$$\tilde{\beta}_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2}.$$

As $n\sigma_\delta^2$ is being subtracted in the denominator, instead of $(n-1)\sigma_\delta^2$, then $\tilde{\beta}_2$ will have a slight positive bias. Furthermore,

$$\begin{aligned}
\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 - (n-1)\sigma_\delta^2} &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2 + \sigma_\delta^2} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2}\left[\frac{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2}{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2 + \sigma_\delta^2}\right] \\
&= \tilde{\beta}_2\left[\frac{1}{1 + \frac{\sigma_\delta^2}{\sum(x_i - \bar{x})^2 - n\sigma_\delta^2}}\right] \\
&\approx \tilde{\beta}_2\left[\frac{1}{1 + \frac{\sigma_\delta^2}{(n-1)\sigma^2 - \sigma_\delta^2}}\right]
\end{aligned}$$

since $E[\sum(x_i - \bar{x})^2] \approx (n-1)(\sigma^2 + \sigma_\delta^2)$.

To correct for the small sample bias in $\tilde{\beta}_2$, $\tilde{\beta}_2$ can be divided by

$$1 + \frac{\sigma_\delta^2}{(n-1)\sigma^2 - \sigma_\delta^2} = \frac{(n-1)\sigma^2}{(n-1)\sigma^2 - \sigma_\delta^2}.$$

As an example, for $n = 40$, from Figure 4.1 $\tilde{\beta}_2 \approx 1.026$. For the parameters chosen for Figure 4.1 then

$$\frac{(n-1)\sigma^2}{(n-1)\sigma^2 - \sigma_\delta^2} = 1.006.$$

Dividing the original value for $\tilde{\beta}_2$ by 1.006 gives 1.01988, yielding a result closer to the value of the true slope.

Similarly, the slope estimator $\tilde{\beta}_3$ may be written

$$\tilde{\beta}_3 = \frac{\sum(y_i - \bar{y})^2 - n\sigma_\varepsilon^2}{\sum(x_i - \bar{x})(y_i - \bar{y})}$$

and this time, as $n\sigma_\varepsilon^2$ is being subtracted in the numerator, instead of $(n-1)\sigma_\varepsilon^2$, then $\tilde{\beta}_3$ will have a slight negative bias. This positive and negative bias noted in $\tilde{\beta}_2$ and $\tilde{\beta}_3$ respectively will occur regardless of the distribution of ξ since the correction that should be made for the bias in the corrected sums of squares and sums of products is the small sample correction, and the correction is not dependent upon any given distribution.

By considering the small sample correction on $\tilde{\beta}_3$ we have

$$\begin{aligned}
\frac{\sum(y_i-\bar{y})^2-(n-1)\sigma_\varepsilon^2}{\sum(x_i-\bar{x})(y_i-\bar{y})} &= \frac{\sum(y_i-\bar{y})^2-n\sigma_\varepsilon^2+\sigma_\varepsilon^2}{\sum(x_i-\bar{x})(y_i-\bar{y})} \\
&= \frac{\sum(y_i-\bar{y})^2-n\sigma_\varepsilon^2}{\sum(x_i-\bar{x})(y_i-\bar{y})}\left[1+\frac{\sigma_\varepsilon^2}{\sum(y_i-\bar{y})^2-n\sigma_\varepsilon^2}\right] \\
&\approx \tilde{\beta}_3\left[1+\frac{\sigma_\varepsilon^2}{(n-1)\beta^2\sigma^2-\sigma_\varepsilon^2}\right]
\end{aligned}$$

since $E[\sum(y_i-\bar{y})]\approx(n-1)(\beta^2\sigma^2+\sigma_\varepsilon^2)$.

To correct for the small sample bias in $\tilde{\beta}_3$, $\tilde{\beta}_3$ may be multiplied by

$$1+\frac{\sigma_\varepsilon^2}{(n-1)\beta^2\sigma^2-\sigma_\varepsilon^2}=\frac{(n-1)\beta^2\sigma^2}{(n-1)\beta^2\sigma^2-\sigma_\varepsilon^2}. \tag{4.1}$$

As an example, for $n=40$, from Figure 4.1 $\tilde{\beta}_3\approx 0.992$. For the parameters chosen for Figure 4.1 then

$$\frac{(n-1)\beta^2\sigma^2}{(n-1)\beta^2\sigma^2-\sigma_\varepsilon^2}=1.00645.$$

Multiplying together the original value for $\tilde{\beta}_2$ by 1.00645 gives 0.9984, yielding a result noticeably closer to the value of the true slope.

On the other hand, as $\tilde{\beta}_4$ may be written

$$\tilde{\beta}_4=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\kappa\sum(x_i-\bar{x})}$$

then removing biases in the sample quantities s_{xx} and s_{xy} is irrelevant as they would cancel upon taking the ratio. This explains why $\tilde{\beta}_4$ is virtually unbiased across the range of sample sizes.

For the parameters chosen by Hood, $\lambda = 1$ and $\tilde{\beta}_5$ is written

$$\tilde{\beta}_5 = \frac{(s_{yy} - s_{xx}) + \sqrt{(s_{yy} - s_{xx})^2 + 4(s_{xy})^2}}{2s_{xy}}$$

and again removing biases in the sample quantities s_{xx}, s_{xy} and s_{yy} would be irrelevant as they would cancel out.

Uniform ξ, Normal errors In order to make a comparison with the results from simulations of the Normal structural model, the parameters a and b of the support for the distribution of ξ were chosen such that

$$\begin{aligned} E[\xi] &= \frac{a+b}{2} = 1 \\ Var[\xi] &= \frac{(b-a)^2}{12} = 2. \end{aligned}$$

This yields $a = 1 - 2\sqrt{3}$ and $b = 1 + 2\sqrt{3}$. Figure 4.2 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for uniform ξ and Normal errors. The bias in $\tilde{\beta}_5$ and $\tilde{\beta}_7$ is similar, and for clarity results for $\tilde{\beta}_7$ have been omitted. Again the slight positive and negative bias is present in $\tilde{\beta}_2$ and $\tilde{\beta}_3$ respectively, and $\tilde{\beta}_1$ behaves the most erratically. $\tilde{\beta}_4$ is virtually unbiased across the whole range of sample sizes as it is for simulations based on the Normal structural model. Indeed, it appears that Figure 4.2 is very similar to Figure 4.1, the only difference being the performance of $\tilde{\beta}_1$. In Figure 4.1 $\tilde{\beta}_1$ is negatively biased for $n = 20$, and then positively biased for $n > 20$. In Figure 4.1 $\tilde{\beta}_1$ is positively biased across the range of sample sizes.

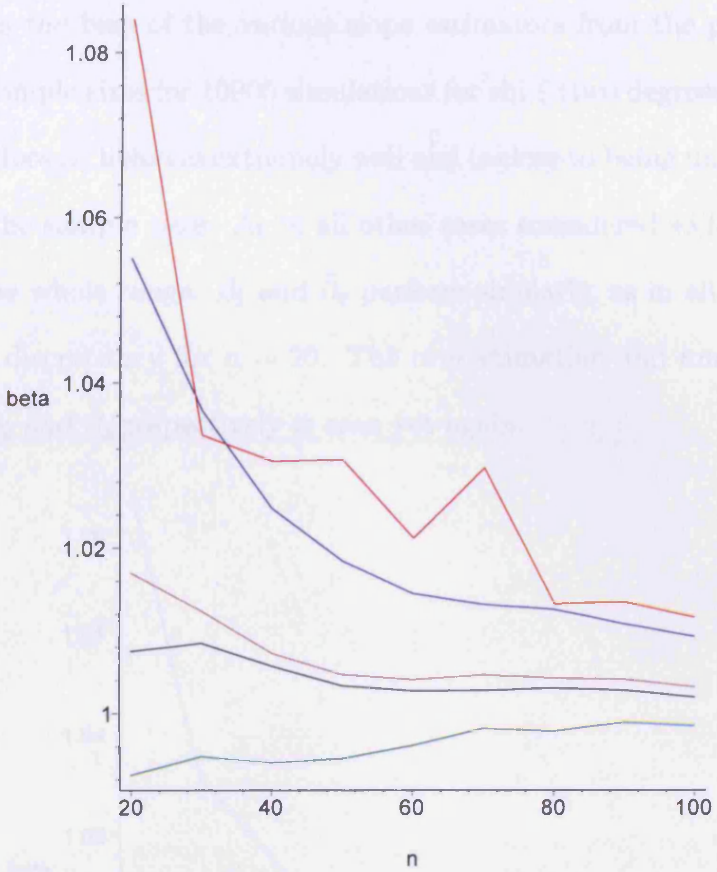


Figure 4.2: Estimate of β against sample size for the structural model with uniform ξ and Normal errors.

Chi ξ (two degrees of freedom), Normal errors The variance of the chi distribution with two degrees of freedom is $2 - \frac{\pi}{2}$. So comparisons can be made with the previous simulations, σ_δ and σ_ε were rescaled so that the same reliability ratio (for both the x and y measurements) as previously is obtained.

Solving the following equations with $\beta = 1$

$$\frac{(2 - \frac{\pi}{2})}{(2 - \frac{\pi}{2}) + \sigma_\delta^2} = 0.8$$
$$\frac{(2 - \frac{\pi}{2})}{(2 - \frac{\pi}{2}) + \sigma_\varepsilon^2} = 0.8$$

gives $\sigma_\delta^2 = \sigma_\varepsilon^2 = 0.1073$.

Figure 4.3 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for chi ξ (two degrees of freedom) and Normal errors. Here $\tilde{\beta}_1$ behaves extremely well and is close to being unbiased across the whole range of the sample sizes. As in all other cases considered so far $\tilde{\beta}_4$ is virtually unbiased over the whole range. $\tilde{\beta}_5$ and $\tilde{\beta}_7$ perform similarly, as in all other cases, but there is a slight discrepancy for $n = 20$. The overestimation and underestimation by the estimators $\tilde{\beta}_2$ and $\tilde{\beta}_3$ respectively is seen yet again.

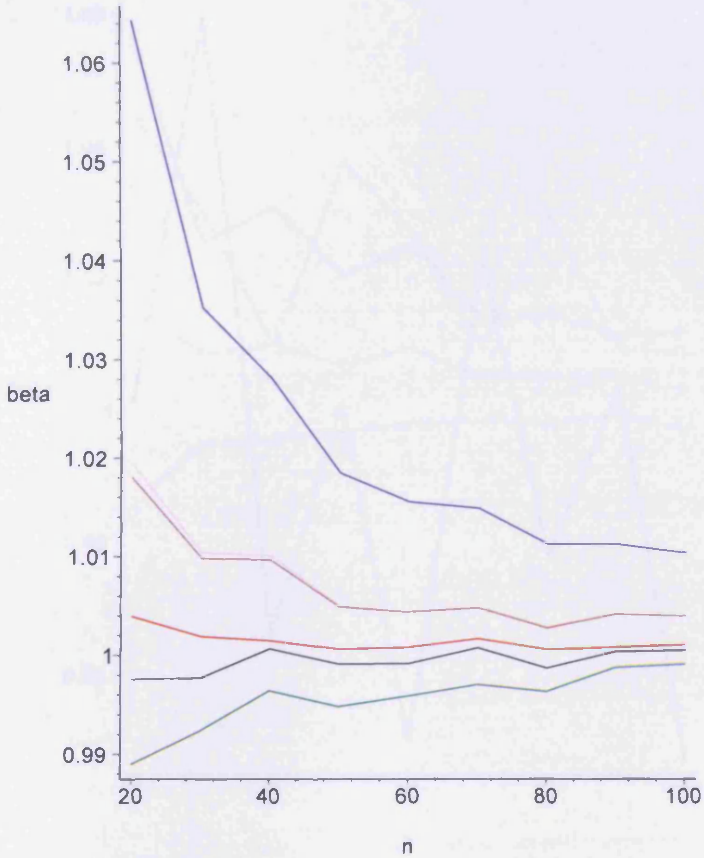


Figure 4.3: Estimate of β against sample size for the structural model with chi ξ (two degrees of freedom) and Normal errors.

Normal functional model The previous Chapter demonstrated that the estimators of the unknown parameters may also be applied to a functional model. For this simulation, a sample of ξ was generated from a Normal distribution. This sample was then fixed, and random Normal errors were added for each simulation. The parameters chosen were again made identical to those of Hood.

Figure 4.4 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for the Normal functional model. Different from previous simulations, $\tilde{\beta}_4$ behaves the most erratically. One reason for

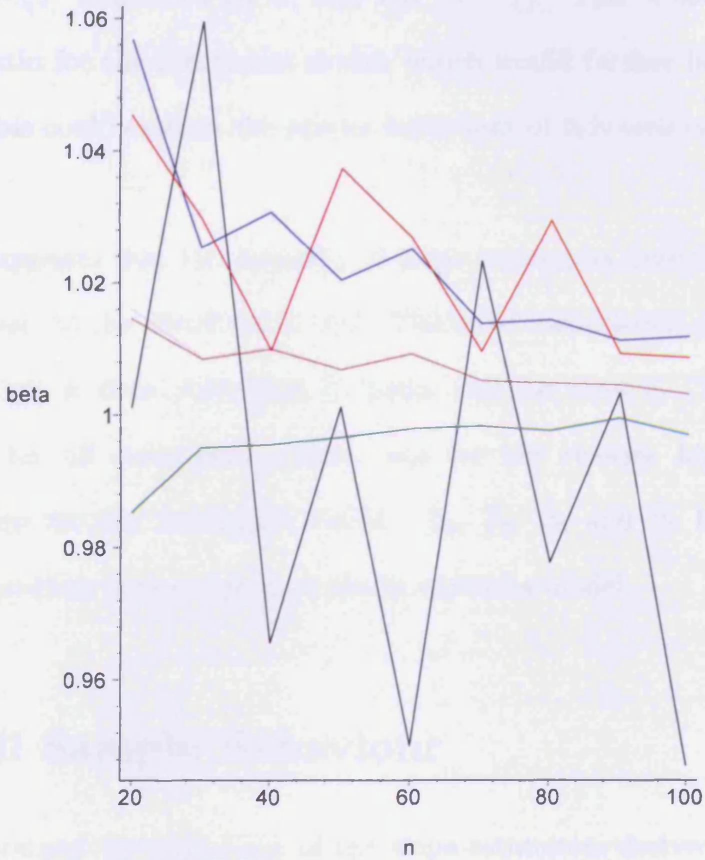


Figure 4.4: Estimate of β against sample size for the Normal functional model.

this could lie in the interpretation of the reliability ratio for a functional model. In a structural model, the reliability ratio is defined to be

$$\kappa = \frac{\sigma^2}{\sigma^2 + \sigma_\delta^2}$$

where each ξ_i is drawn from a random variable with mean μ and variance σ^2. For the functional model however, each ξ is assumed to be a fixed unknown constant. The equivalent reliability ratio for the functional model is therefore

$$\frac{s_{\xi\xi}}{s_{\xi\xi} + \sigma_\delta^2} = \frac{\frac{1}{n}\sum(\xi_i - \bar{\xi})^2}{\frac{1}{n}\sum(\xi_i - \bar{\xi})^2 + \sigma_\delta^2},$$

where $s_{\xi\xi}$ will be formally defined in next previous Chapter. The quantity $s_{\xi\xi}$, which is a measurement of the dispersion of the ξ_i about their mean, is subject to bias as the sum $\sum(\xi_i - \bar{\xi})^2$ is divided by n, and not $(n-1)$. This would yield a bias in the reliability ratio for the functional model, which would further lead to bias in the estimator $\tilde{\beta}_4$. This could explain the erratic behaviour of this estimator.

In summary, it appears that the majority of slope estimators derived in the previous Chapter are robust to the distribution of ξ. The most inconsistent estimator appears to be $\tilde{\beta}_1$, although it does seem that it works well for skew ξ. The estimator $\tilde{\beta}_4$ performed well for all structural models, but for the reasons highlighted earlier, performed weakly for the functional model. $\tilde{\beta}_2$, $\tilde{\beta}_3$, $\tilde{\beta}_5$ and $\tilde{\beta}_7$ have a consistent performance regardless of the type of errors in variables model.

4.3 Small Sample Behaviour

To further understand the behaviour of the slope estimators derived in the previous Chapter, it is interesting to note the bias for small sample sizes.

Normal structural model Figure 4.5 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for the Normal structural model. The parameter settings chosen were the same as for Figure

4.1. $\tilde{\beta}_1$ again performs least favourably, giving an estimate of the slope of more than

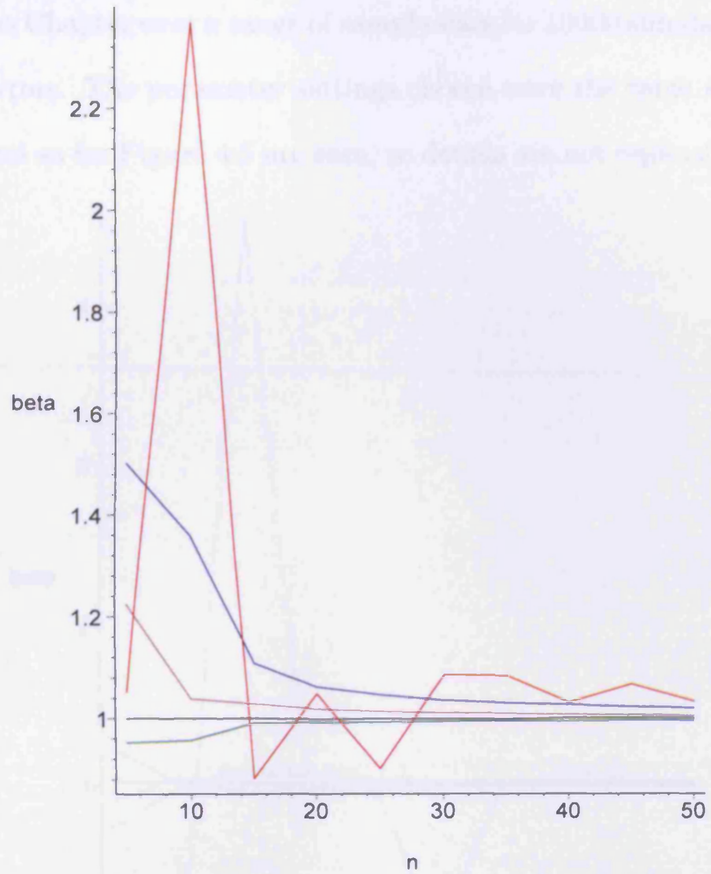


Figure 4.5: Estimate of β against sample size for the Normal structural model.

double its true value for $n = 10$. $\tilde{\beta}_4$ is yet again robust to the sample size, providing an approximately unbiased estimator for $n = 5$. The bias in $\tilde{\beta}_5$ and $\tilde{\beta}_7$ is again indistinguishable. The same positive and negative bias in $\tilde{\beta}_2$ and $\tilde{\beta}_3$ respectively is present.

Uniform ξ, Normal errors Figure 4.6 shows the bias of the various slope estimators from the previous Chapter over a range of sample sizes for 10000 simulations for uniform ξ and Normal errors. The parameter settings chosen were the same as for Figure 4.2. The same features as for Figure 4.5 are seen, so details are not replicated here. A point

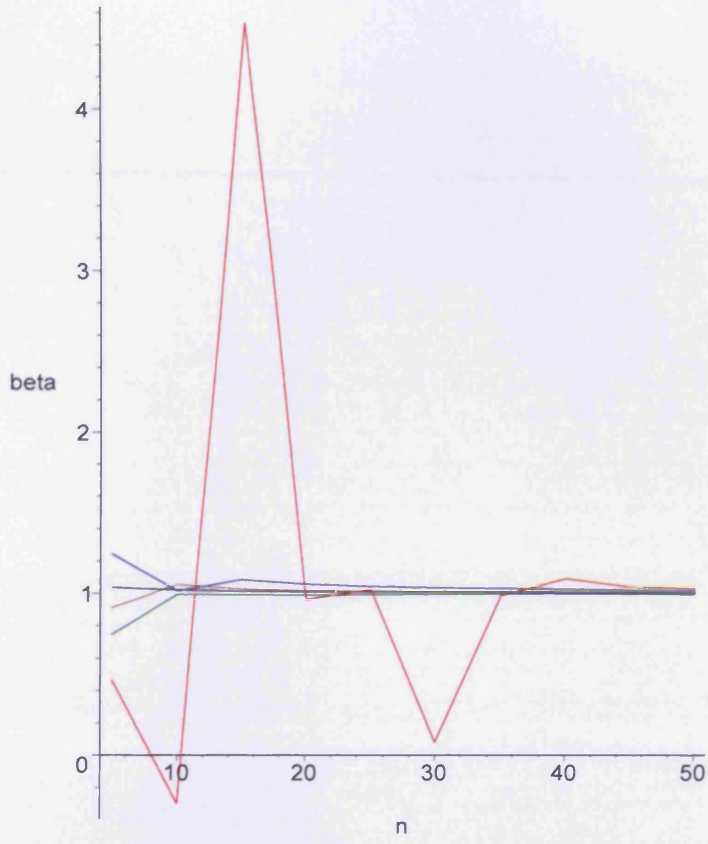


Figure 4.6: Estimate of β against sample size for the structural model with uniform ξ and Normal errors.

worthy of note however, is that for a sample size of $n = 15$ $\tilde{\beta}_1$ is more than four times the true value of the slope. The extreme behaviour of $\tilde{\beta}_1$ for some structural models implies that for some sample sizes it would be difficult to obtain positive variance estimates using this estimator. As derived in the previous Chapter, to ensure positive variance estimators, the estimated β must lie between the slopes of y on x and x and y regression respectively. For small sample sizes it is more likely that the errors in variables estimator of the slope will lie outside of this range. A simulation study looking at the number of times the slope estimators lie outside of this range is included in this Chapter.

Chi ξ (two degrees of freedom), Normal errors Figure 4.7 shows the bias of the various slope estimators from the previous Chapter over a range of small sample sizes for 10000 simulations for chi ξ (two degrees of freedom) and Normal errors. The parameter settings chosen were the same as for Figure 4.3. Some interesting features

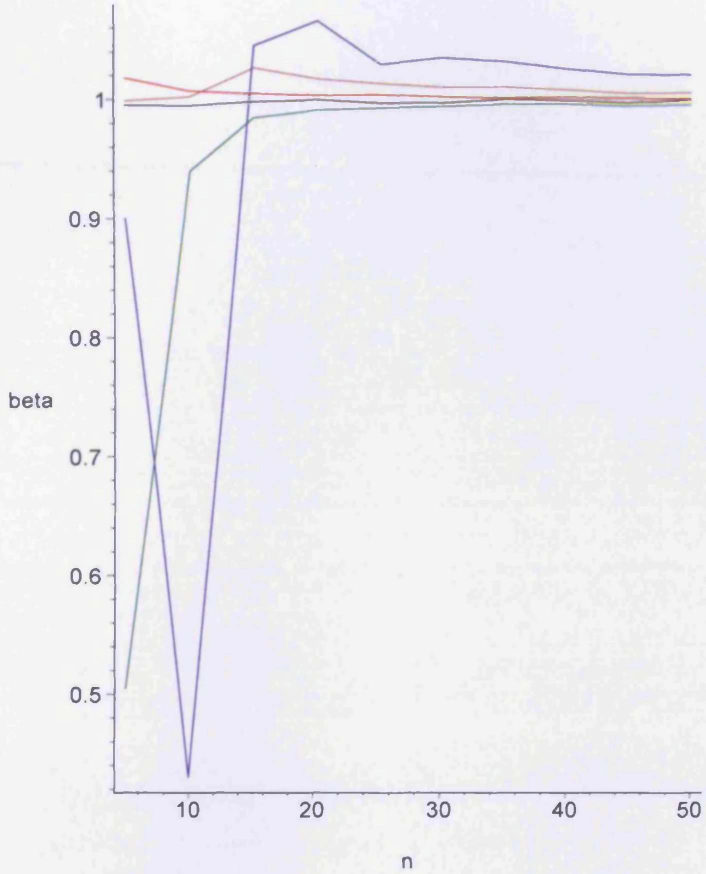


Figure 4.7: Estimate of β against sample size for the structural model with chi ξ (two degrees of freedom) and Normal errors.

are present in this simulation. For the first time in all simulations conducted so far, $\tilde{\beta}_2$ has demonstrated an appreciable negative bias for $n < 15$. $\tilde{\beta}_1$ performs very well and is robust to changes in the sample size. This could be due to the large number of simulations around the origin for this highly skewed distribution of ξ. The migration effect of the measurement error in the x observations (which will be introduced in

Chapter 7) pushes data in the left hand tail further left. For the value of μ chosen in these simulations, the increased volume of data around the origin is likely to make the slope estimator which is a function of the first two sample moments $\bar{x}$ and $\bar{y}$ more reliable.

Normal functional model Figure 4.8 shows the bias of the various slope estimators from the previous Chapter over a range of small sample sizes for 10000 simulations for the Normal functional model. The parameter settings chosen, and the method of simulation were the same as for Figure 4.4. Figure 4.4 and Figure 4.8 are very similar,

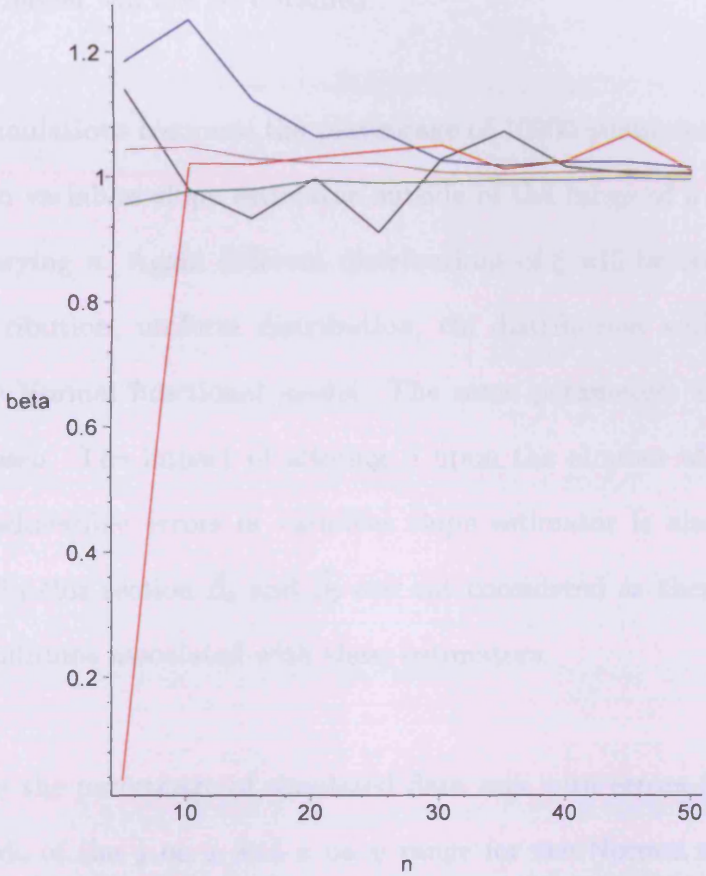


Figure 4.8: Estimate of β against sample size for the Normal functional model.

the only difference being the poor performance of $\tilde{\beta}_1$ for $n = 5$. The erratic behaviour of $\tilde{\beta}_4$ is still present.

4.4 Breaking of Admissibility Conditions

As written in the previous Chapter, if the errors in variables slope estimator does not lie between the slopes of y on x and x on y regression then positive estimators of the variances of the model will not be obtained.

The following simulations compute the percentage of 10000 simulated data sets which have an errors in variables slope estimator outside of the range of y on x and x on y regression for varying n. Again different distributions of ξ will be considered, namely the Normal distribution, uniform distribution, chi distribution with two degrees of freedom and the Normal functional model. The same parameters as in the previous section were chosen. The impact of altering β upon the number of data sets failing to produce an admissible errors in variables slope estimator is also considered. In all the pictures in this section $\tilde{\beta}_4$ and $\tilde{\beta}_5$ are not considered as there are no specific admissibility conditions associated with these estimators.

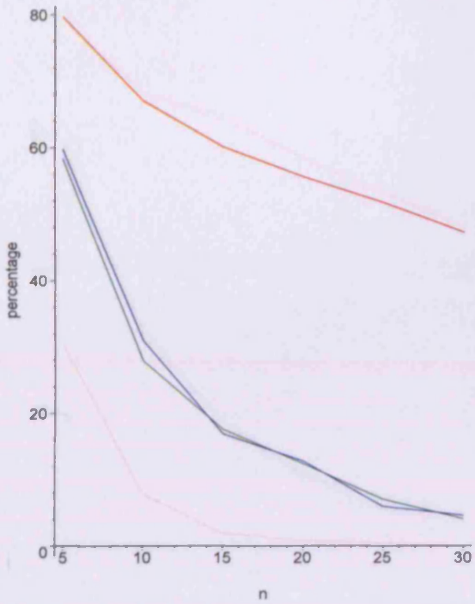
Figure 4.9 shows the percentage of simulated data sets with errors in variables slope estimators outside of the y on x and x on y range for the Normal structural model. For all values of β, $\tilde{\beta}_1$ produces the most inadmissible slope estimators. When $\beta = 1$, at least 70% of the simulated data sets of size 5 produced an inadmissable $\tilde{\beta}_1$. This percentage decreases as the sample size gets larger, but only to about 50% when $n = 30$. As β increases, the number of inadmissable slope estimators produced by $\tilde{\beta}_1$ does increase slightly, but it changes only slowly across the range of β. An estimator which is greatly affected by the value of β is $\tilde{\beta}_2$. When $\beta = 1$, the number of inadmissible estimators produced by $\tilde{\beta}_2$ is indistinguishable from the number produced by $\tilde{\beta}_3$. However, as β increases, the number of inadmissable slope estimators from using $\tilde{\beta}_2$ increases, and for small samples is comparable to the number of inadmissable slope estimators produced using $\tilde{\beta}_1$. On the other hand, as β increases, the number of inadmissable slope estimators from using $\tilde{\beta}_3$ decreases. When $\beta = 4$, less than 10% of the data sets produce inadmissable $\tilde{\beta}_3$ even for a sample size of only 5. $\tilde{\beta}_4$ and $\tilde{\beta}_7$ perform more poorly for large β. When $\beta = 1$, $\tilde{\beta}_7$ starts to produce no inadmissable slope estimators as the sample sizes grows to 30. However when $\beta = 4$, approximately 50% of the samples have an inadmissable slope estimator when $\tilde{\beta}_7$ is used, even for a sample size of 30.

Figure 4.10 shows the percentage of simulated data sets with errors in variables slope estimators outside of the y on x and x on y range for a structural model with uniform ξ and Normal errors. Figure 4.10 can be seen to look similar to 4.9, and so further discussion is not made here.

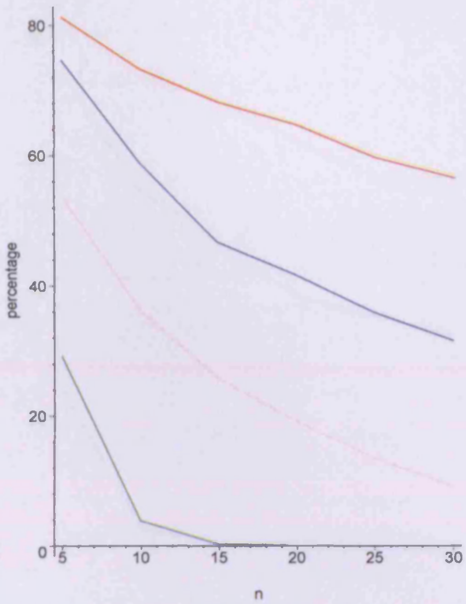
Figure 4.11 does display some features not present in Figures 4.9 or 4.10. Figure 4.11 shows the percentage of simulated data sets with errors in variables slope estimators outside of the y on x and x on y range for a structural model with chi ξ (two degrees of freedom) and Normal errors. $\tilde{\beta}_3$ is the best performing across the range of β. For $\beta = 1$, the number of inadmissible estimators produced by $\tilde{\beta}_2$ and $\tilde{\beta}_3$ is indistinguishable. However, as β grows larger then the number of inadmissible estimators produced by $\tilde{\beta}_2$ and $\tilde{\beta}_1$ is indistinguishable. On the other hand, the number of inadmissible slope estimators produced by $\tilde{\beta}_3$ decreases. $\tilde{\beta}_7$ performs similarly as in Figures 4.9 and 4.10.

Figure 4.12 shows the percentage of simulated data sets with errors in variables slope estimators outside of the y on x and x on y range for a Normal functional model. As in previous simulations it can be seen that in general the output for a Normal functional model is more erratic than for any other errors in variables model. The most erratic estimator is $\tilde{\beta}_1$. For example, when $\beta = 2$, as the sample size increases, the number of inadmissible estimators produced by $\tilde{\beta}_1$ also increases. A common feature as with all other simulations considered here is that for $\beta = 1$, the number of inadmissible estimators produced by $\tilde{\beta}_2$ and $\tilde{\beta}_3$ is again indistinguishable. However as β grows larger, then the number of inadmissible estimators produced by $\tilde{\beta}_3$ decreases.

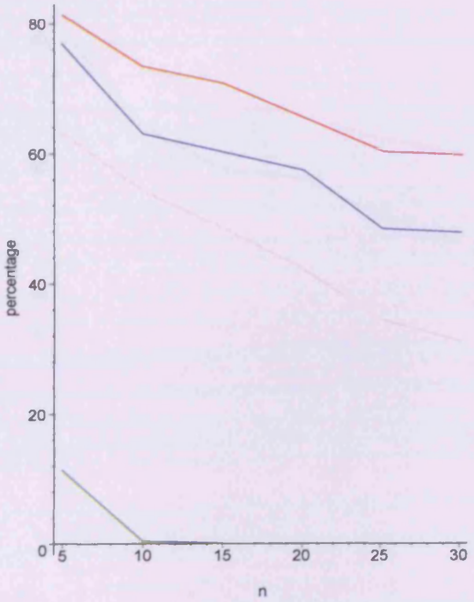
Normal structural model



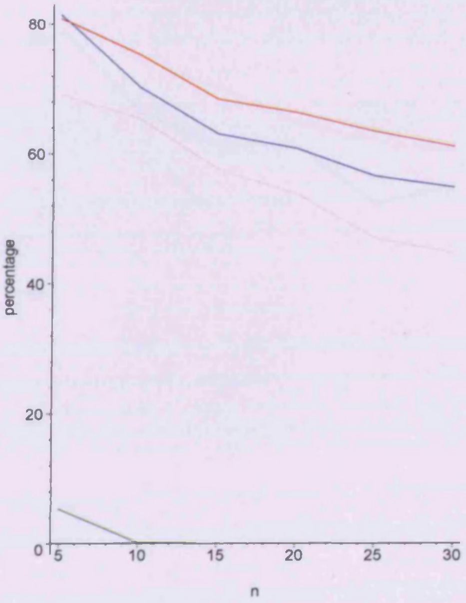
(a) $\beta = 1$



(b) $\beta = 2$



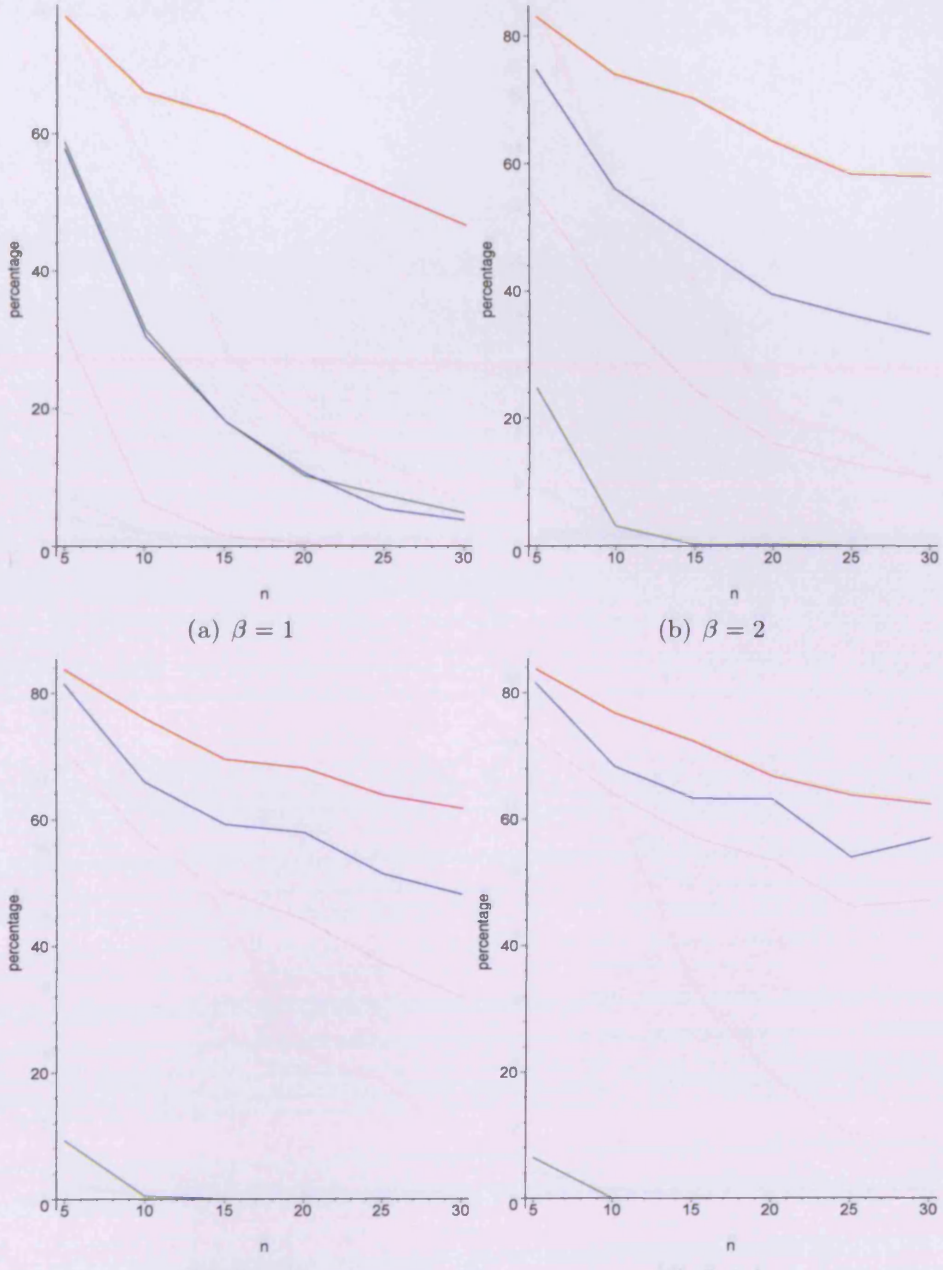
(c) $\beta = 3$



(d) $\beta = 4$

Figure 4.9: Percentage of simulated data sets with errors in variables slope estimators outside of y on x and x on y range for the Normal structural model.

Uniform ξ, Normal errors

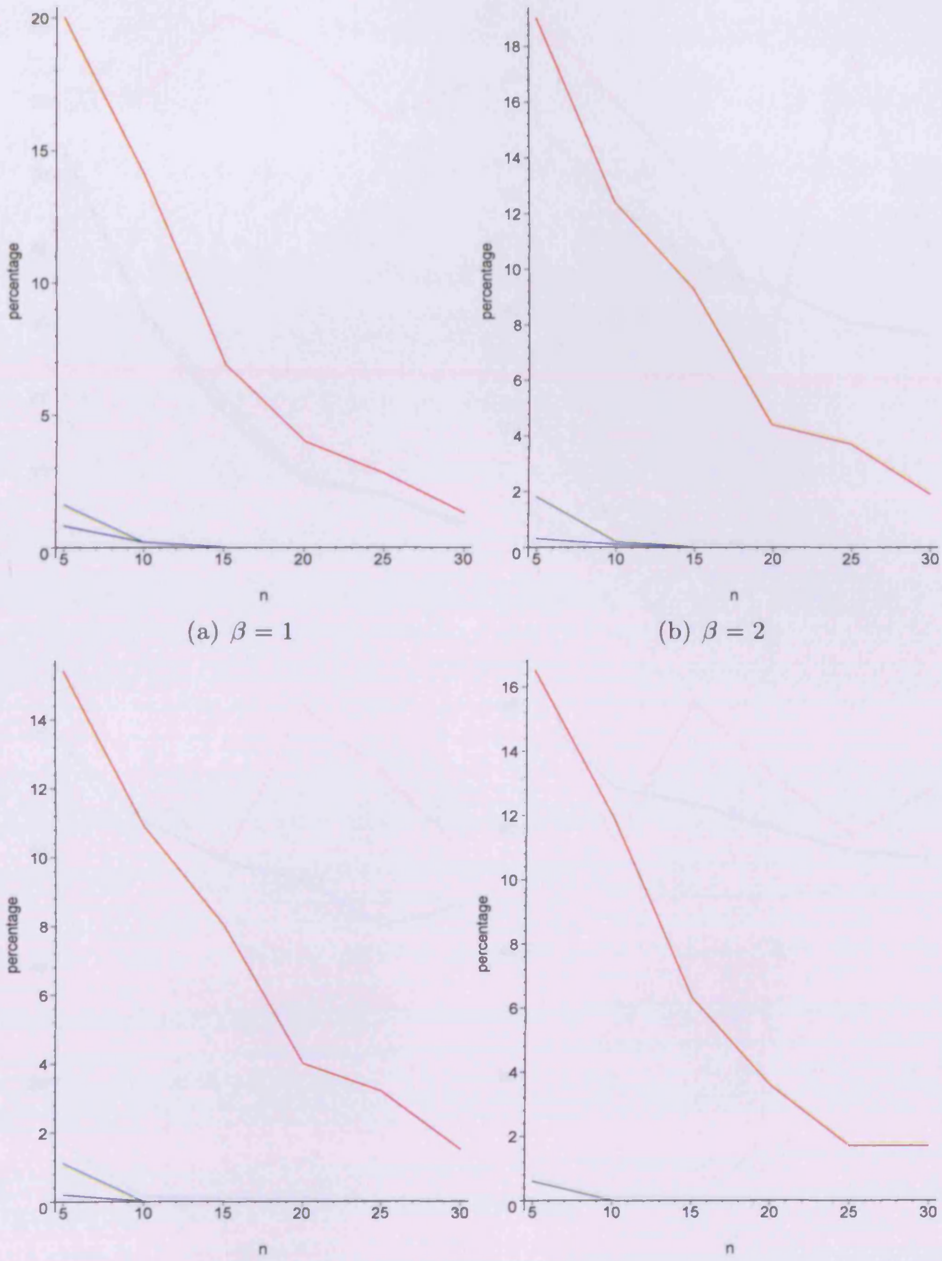


(a) $\beta = 1$ (b) $\beta = 2$

(c) $\beta = 3$ (d) $\beta = 4$

Figure 4.10: Percentage of simulations with errors in variables estimator outside of y on x and x on y range for a structural model with uniform ξ and Normal errors.

Chi ξ, Normal errors

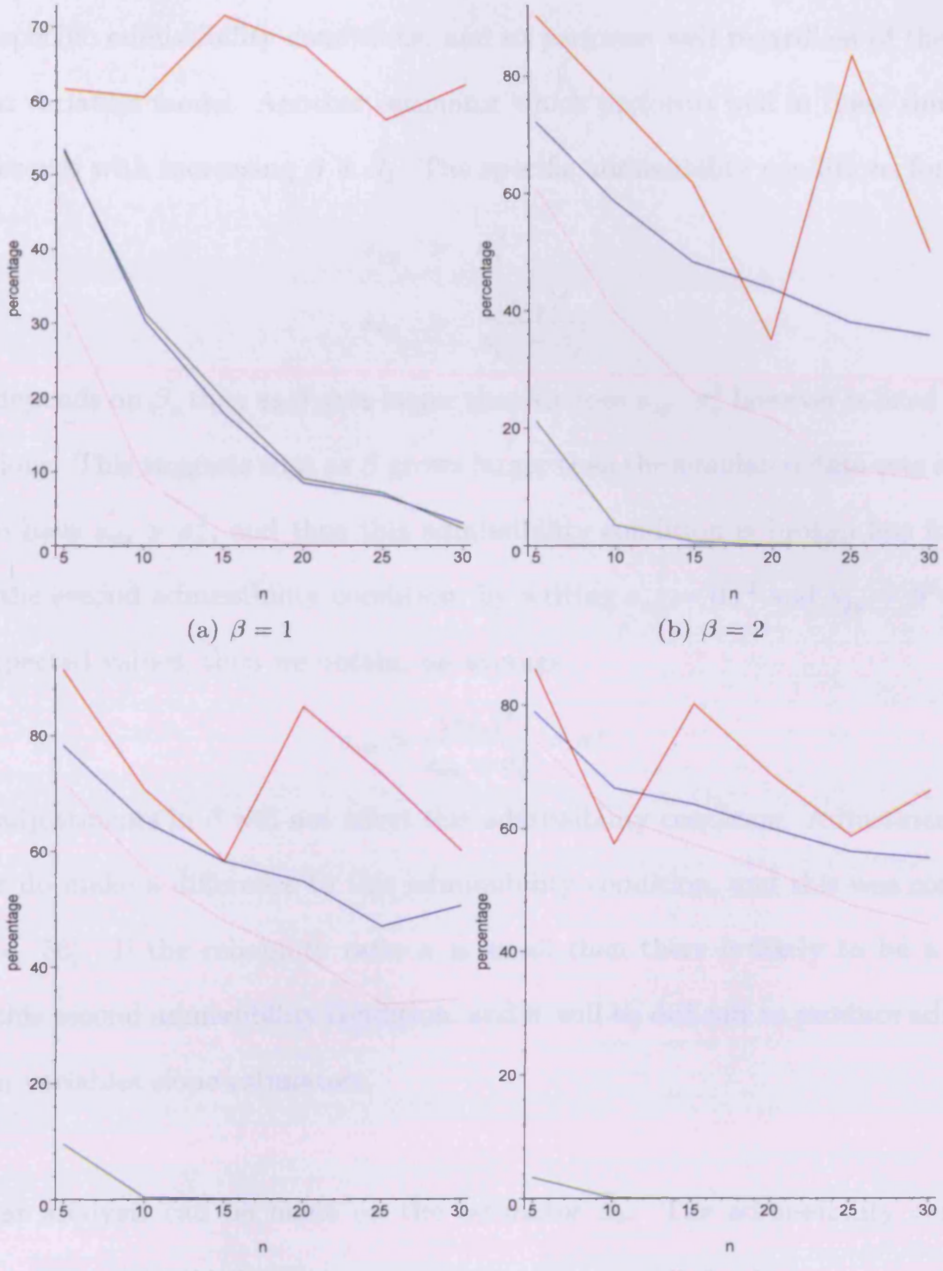


(a) $\beta = 1$ (b) $\beta = 2$

(c) $\beta = 3$ (d) $\beta = 4$

Figure 4.11: Percentage of simulations with errors in variables estimator outside of y on x and x on y range for a structural model with chi ξ (two degrees of freedom) and Normal errors.

Normal functional model



(a) $\beta = 1$ (b) $\beta = 2$

(c) $\beta = 3$ (d) $\beta = 4$

Figure 4.12: Percentage of simulations with errors in variables estimator outside of y on x and x on y range for the Normal functional model.

Admissibility conditions reflect the intuitive requirement that variance estimators must be positive. The behaviour of some of the estimators may be explained by looking deeper into the specific admissibility conditions for each estimator. For example $\tilde{\beta}_5$ has no specific admissibility conditions, and so performs well regardless of the type of errors in variables model. Another estimator which performs well in these simulations and improves with increasing β is $\tilde{\beta}_3$. The specific admissibility conditions for $\tilde{\beta}_3$ are

$$\begin{aligned} s_{yy} &> \sigma_\varepsilon^2 \\ s_{xx} &> \frac{(s_{xy})^2}{s_{yy} - \sigma_\varepsilon^2}. \end{aligned}$$

As s_{yy} depends on β, then as β gets larger then so does s_{yy}. σ_ε^2 however is fixed in these simulations. This suggests that as β grows larger then the simulated data sets are more likely to have $s_{yy} > \sigma_\varepsilon^2$, and thus this admissibility condition is broken less for larger β. For the second admissibility condition, by writing $s_{xy} = \beta\sigma^2$ and $s_{yy} = \beta^2\sigma^2 + \sigma_\varepsilon^2$, their expected values, then we obtain, on average

$$s_{xx} > \frac{(s_{xy})^2}{s_{yy} - \sigma_\varepsilon^2} = \sigma^2.$$

and so adjustments in β will not affect this admissibility condition. Adjustments in σ^2 however do make a difference to this admissibility condition, and this was considered by Hood [56]. If the reliability ratio κ is small then there is likely to be a conflict within this second admissibility condition, and it will be difficult to produce admissible errors in variables slope estimators.

A similar analysis can be made on the estimator $\tilde{\beta}_2$. The admissibility conditions specific to this estimator are

$$\begin{aligned} s_{xx} &> \sigma_\delta^2 \\ s_{yy} &> \frac{(s_{xy})^2}{s_{xx} - \sigma_\delta^2}. \end{aligned}$$

The first admissibility condition does not depend on β as s_{xx} is independent of β. Again if the reliability ratio is small then there is likely to be a conflict within this admissibility condition. For the second admissibility condition, by writing $s_{xx} = \sigma^2 + \sigma_\delta^2$ and $s_{xy} = \beta\sigma^2$ then we obtain, on average

$$s_{yy} > \frac{(s_{xy})^2}{s_{xx} - \sigma_\delta^2} = \frac{\beta^2\sigma^2}{\beta^2\sigma^2 + \sigma_\varepsilon^2}$$

where $\frac{\beta^2\sigma^2}{\beta^2\sigma^2+\sigma_\varepsilon^2}$ is the reliability ratio in the y measurement. In the previous simulations, σ^2 and σ_ε^2 remained fixed, and β was increased. This means that the reliability ratio for the y measurement decreased, causing potential conflict with this second admissibility condition. This could explain the poor performance, in general, of $\tilde{\beta}_2$ as β was taken larger.

4.5 Variance Covariance Matrices

As detailed in Chapter 3, a number of shortcut formulae for each element of the variance covariance matrices of the varying estimators of the slope were derived. The shortcut formulae demonstrated that most elements of the variance covariance matrices are functions of $Var[\tilde{\beta}]$. For example, for a Normal structural model

$$Var[\tilde{\alpha}] = \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{n}.$$

As a result, the following simulation study into the variance covariance matrices of the varying estimators only considers $Var[\tilde{\beta}]$.

The variance covariance matrices are asymptotic results, and should be used for moderate sample sizes. The aim of the simulations here is to provide guidance on the minimum sample size needed to obtain reliable estimators of the variance covariance matrices. As has been the theme with this Chapter thus far, a number of different errors in variables model types will be considered. As the performance of the different slope estimators differs as the distribution of ξ alters, then so do the variance covariance matrices.

In general, as shown in Figures 4.13 to 4.16 the theoretical expressions for the variance of $\tilde{\beta}$ tend to be larger than the sample variances. The scale of the Figures however do suggest that in general there is close agreement between the theoretical variances and sample variances of $\tilde{\beta}$.

Figure 4.13 compares theoretical (using formulae of Chapter 3) and sample variances of different slope estimators for a Normal structural model under varying sample sizes. To compute the sample variances, 100,000 simulations were run. The parameter settings used were $\alpha = 0$, $\beta = 1$, $\mu = 1$, $\sigma = 2$, $\sigma_\delta = 1$ and $\sigma_\varepsilon = 1$. The estimator with the most erratic sample variance for small sample sizes was $\tilde{\beta}_1$. The sample variance for a sample size of 10 was greater than 70, but did settle down to the value of the theoretical variance as soon as the sample size was made larger than approximately 50. The values of the theoretical variances for this simulation study of the Normal structural model were:

$$\begin{aligned}
nVar[\tilde{\beta}_1] &= 2 \\
nVar[\tilde{\beta}_2] &= 0.6875 \\
nVar[\tilde{\beta}_3] &= 0.6875 \\
nVar[\tilde{\beta}_4] &= 0.5625 \\
nVar[\tilde{\beta}_5] &= 0.5625 \\
nVar[\tilde{\beta}_7] &= 0.5625.
\end{aligned}$$

$\tilde{\beta}_1$ has the largest theoretical variance, and this coincides with previous simulations

(Figures 4.1 and 4.5) that demonstrate its erratic behaviour, particularly for small samples. Thus there will not be close agreement of the theoretical and sample variances of $\tilde{\beta}_1$ for small samples. $\tilde{\beta}_4$, $\tilde{\beta}_5$ and $\tilde{\beta}_7$ share the smallest theoretical variance, with $\tilde{\beta}_2$ and $\tilde{\beta}_3$ sharing the same theoretical variance. This again is in agreement with previous simulations. It can be seen that when the sample size is approximately 50, there is little difference between the sample variances and theoretical variances.

Figure 4.14 compares theoretical and sample variances of different slope estimators for a structural model with uniform ξ and Normal errors under varying sample sizes. The parameter settings used were $\alpha = 0$, $\beta = 1$, $a = 1 - 2\sqrt{3}$, $b = 1 + 2\sqrt{3}$, $\sigma_\delta = 1$ and $\sigma_\varepsilon = 1$. Again the estimator with the most erratic sample variance for small sample sizes was $\tilde{\beta}_1$. The sample variance for a sample size of 10 was greater than 500, decreased to approximately 10 for a sample size of 20 but then rose to approximately 150 for a sample size of 30 before settling down to the value of the theoretical variance as soon as the sample size was made larger than approximately 50. The values of the theoretical variances for this particular construction of the structural model were:

$$\begin{aligned} nVar[\tilde{\beta}_1] &= 2 \\ nVar[\tilde{\beta}_2] &= 0.6875 \\ nVar[\tilde{\beta}_3] &= 0.6875 \\ nVar[\tilde{\beta}_4] &= 0.5145 \\ nVar[\tilde{\beta}_5] &= 0.5625 \\ nVar[\tilde{\beta}_7] &= 0.5625. \end{aligned}$$

The theoretical variance of $\tilde{\beta}_4$ is smaller here than for the Normal structural model. In Chapter 3, the theoretical variance covariance matrices were partitioned into the sum of the matrices, the A matrix, the B matrix and the C matrix. The matrix A alone is needed if the assumptions are made that ξ, δ and ε all have zero third moments and zero measure of excess of kurtosis. These assumptions would be valid if all three of these variables are normally distributed as in the Normal structural model. The matrix B gives the additional terms that are necessary if ξ has non zero third moment and a non zero measure of kurtosis. It can be seen that in most cases the B matrices are sparse, needing only adjustment for the terms for $Var[\tilde{\sigma}^2]$ and $Cov[\tilde{\mu}, \sigma^2]$. The exceptions are the cases where the reliability ratio is assumed known ($\tilde{\beta}_4$), and slope estimators involving the higher moments. The C matrix contains additional terms that are needed if the third moments and measures of excess of kurtosis are non zero for the error terms δ and ε and these additional terms are not applicable for the simulations in this Chapter. So as ξ is considered to be a random variable that follows a uniform distribution, then the corrections given by the B matrix must be made. By using the formulae of Chapter 3, the deduction of -0.048 from the previous variance of $\tilde{\beta}_4$ for the Normal structural model must be made. This yields $nVar[\tilde{\beta}_4] = 0.5145$. $\tilde{\beta}_4$ has the smallest theoretical variance, with $\tilde{\beta}_5$ and $\tilde{\beta}_7$ close to this value.

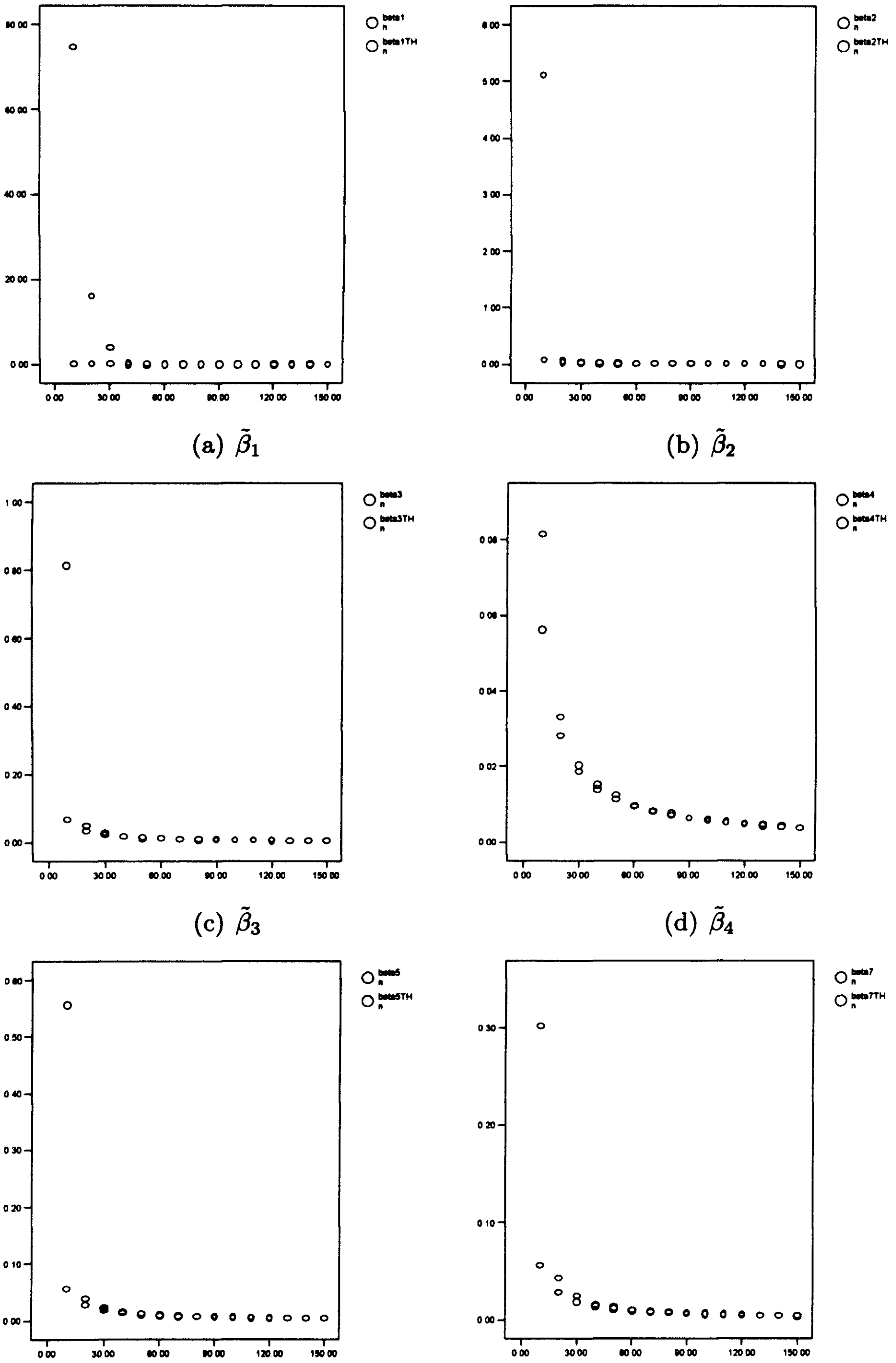
Figure 4.15 compares theoretical and sample variances of different slope estimators for a structural model with chi ξ (two degrees of freedom) and Normal errors under varying sample sizes. The parameter settings used were $\alpha = 0$, $\beta = 1$, $\sigma_\delta = \sigma_\varepsilon = 0.1073$. As two degrees of freedom for the chi distribution were chosen, then $\mu = \sqrt{\frac{\pi}{2}}$ and $\sigma^2 = 2 - \frac{\pi}{2}$. The values of the theoretical variances for this particular construction of the structural model were:

$$
\begin{aligned}
nVar[\tilde{\beta}_1] &= 0.13662 \\
nVar[\tilde{\beta}_2] &= 0.68740 \\
nVar[\tilde{\beta}_3] &= 0.68740 \\
nVar[\tilde{\beta}_4] &= 0.68480 \\
nVar[\tilde{\beta}_5] &= 0.56240 \\
nVar[\tilde{\beta}_7] &= 0.56240.
\end{aligned}
$$

As was demonstrated in Figures 4.3 and 4.7, the best performing estimator was $\tilde{\beta}_1$. This estimator even performed well for small sample sizes. The theoretical variance for this estimator is a lot smaller than for the other estimators of the slope. As can be seen from the scale of the graph, there is close agreement between the sample variances and the theoretical variances of $\tilde{\beta}_1$ across the entire range of sample sizes considered. $\tilde{\beta}_2$ has an exceptionally large variance of 50 for a the smallest sample considered, but does settle down to the value of the theoretical variance as the sample size increases. As in the previous simulation, as ξ is taken to follow a chi distribution with two degrees of freedom, the value $nVar[\tilde{\beta}_4]$ is slightly altered by the correction terms in the B matrix. If the correction terms present in the B matrix are ignored, then the value $nVar[\tilde{\beta}_4] = 0.68740$ is obtained. As can be seen from Figure 4.15 this would distort the close agreement between the sample and theoretical variances observed. In general however, the values for the sample variances and the theoretical variances are virtually indistinguishable across the range of sample sizes and slope estimators.

Figure 4.16 compares theoretical and sample variances of different slope estimators for a Normal functional model with parameter settings $\alpha = 0$, $\beta = 1$, $\sigma_\delta = 1$ and $\sigma_\varepsilon = 1$. For all 100,000 simulations, the same $\xi_i's$ were used. The set of $\xi_i's$ were generated from a Normal distribution with a mean 1 and standard deviation 2. These are the parameter settings that were used for the Normal structural model. The same theoretical variances as for the Normal structural model were used. It can be seen that the sample variances are not as stable as those for the Normal structural model, but the values are roughly similar to those of the Normal structural model. As the sample size gets larger, then the sample variances do tend to the values of the theoretical variances. Again, the estimator with the most erratic sample variances is $\tilde{\beta}_1$. The results for the Normal functional model are very similar to the results for the Normal structural model and thus a detailed analysis is not given here.

Normal structural model



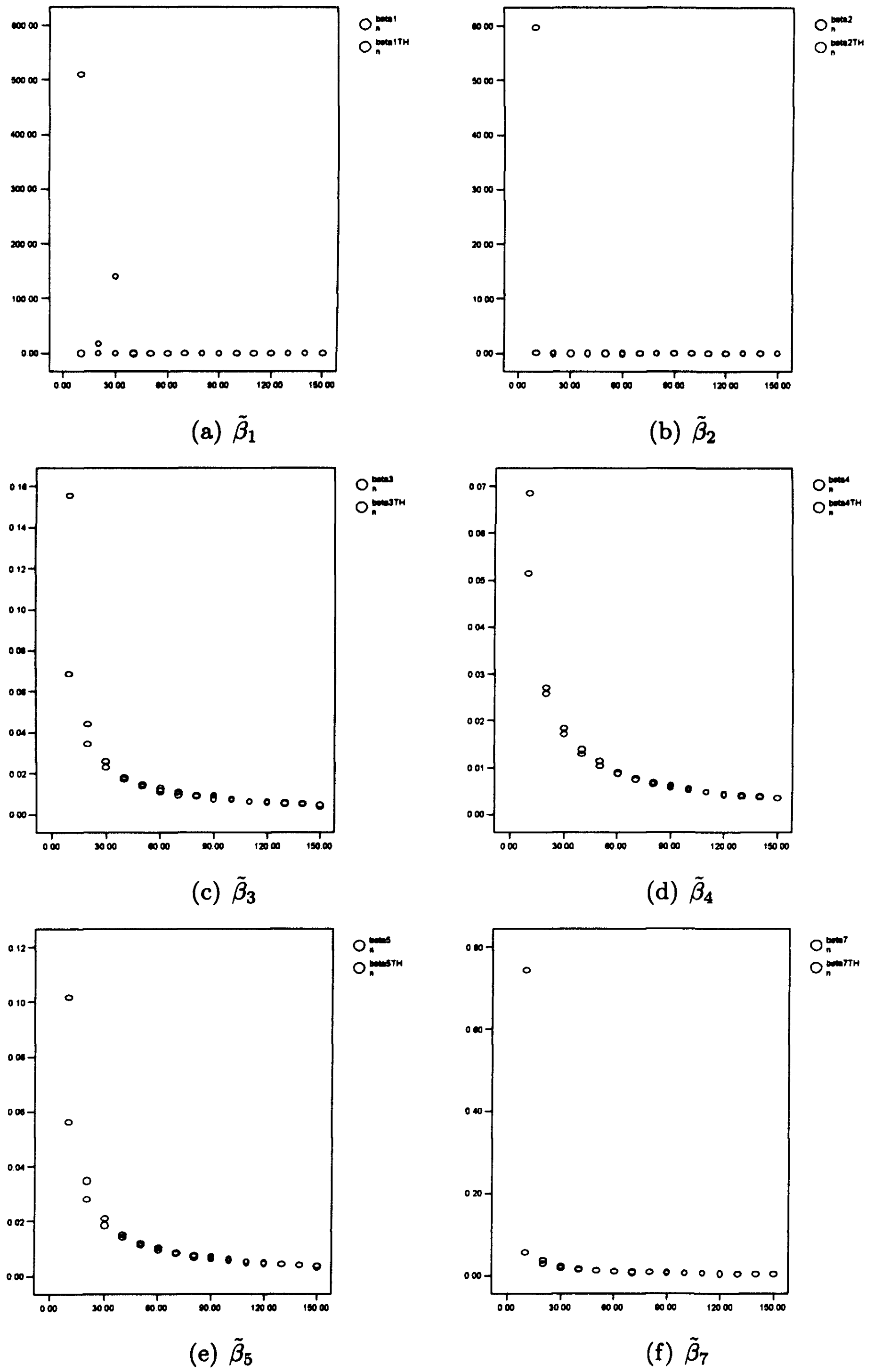
(a) $\tilde{\beta}_1$ (b) $\tilde{\beta}_2$

(c) $\tilde{\beta}_3$ (d) $\tilde{\beta}_4$

(e) $\tilde{\beta}_5$ (f) $\tilde{\beta}_7$

Figure 4.13: Comparing theoretical and sample variances of different slope estimators for a Normal structural model. Sample variances are in blue, theoretical variances are in green.

Uniform ξ, Normal errors



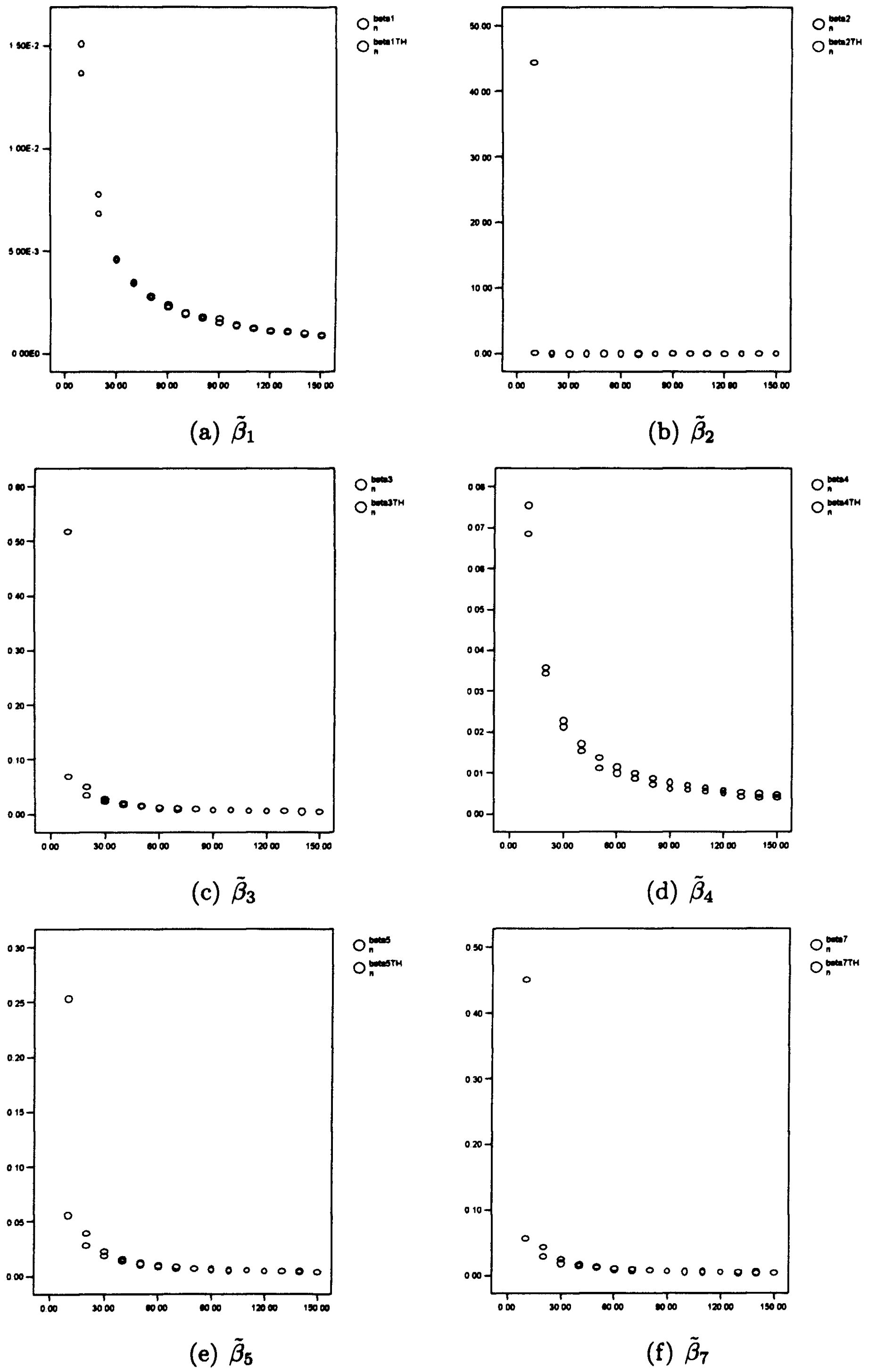
(a) $\tilde{\beta}_1$ (b) $\tilde{\beta}_2$

(c) $\tilde{\beta}_3$ (d) $\tilde{\beta}_4$

(e) $\tilde{\beta}_5$ (f) $\tilde{\beta}_7$

Figure 4.14: Comparing theoretical and sample variances of different slope estimators for a structural model with uniform ξ, and Normal errors. Sample variances are in blue, theoretical variances are in green.

Chi ξ, Normal errors



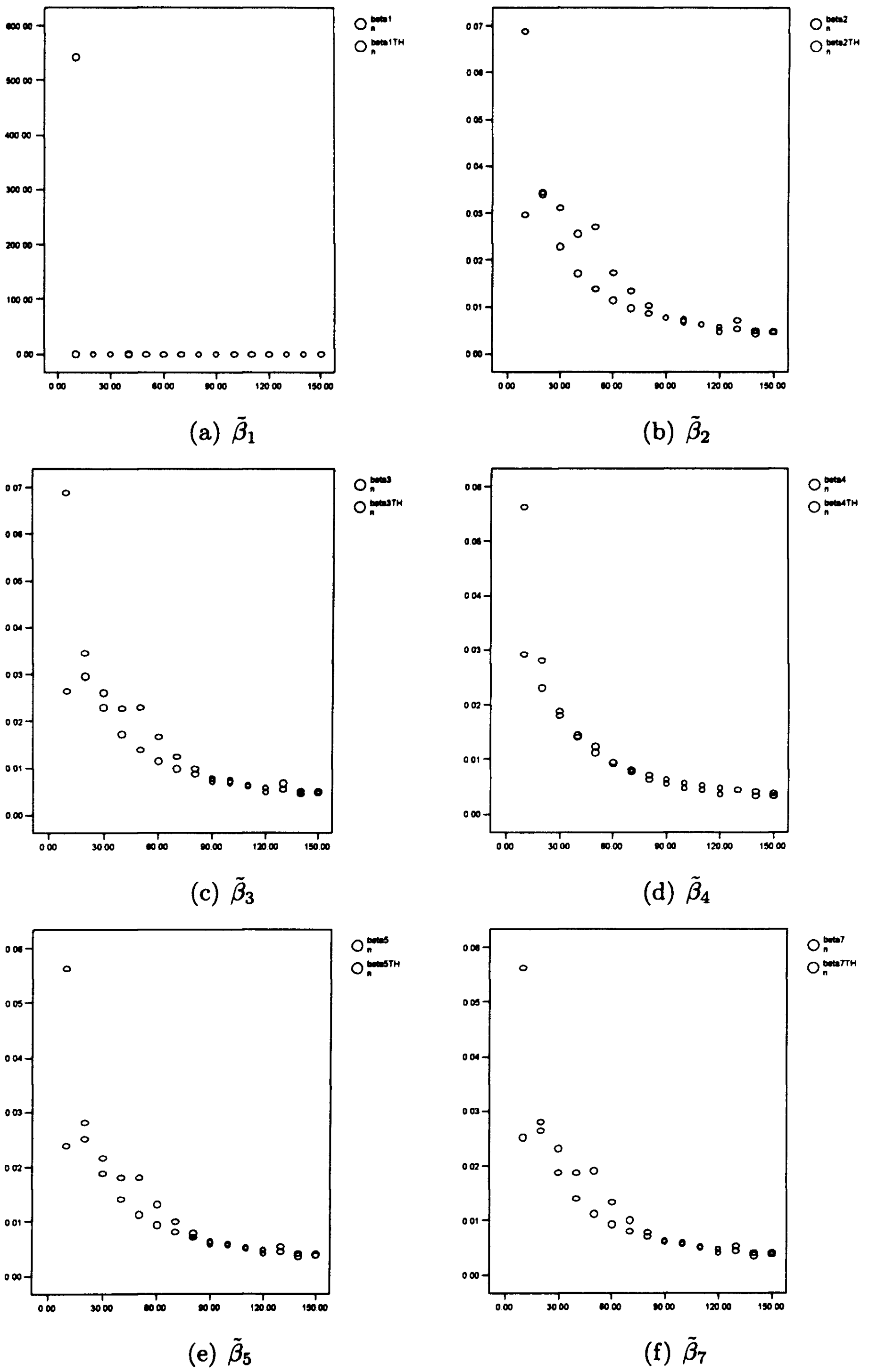
(a) $\tilde{\beta}_1$ (b) $\tilde{\beta}_2$

(c) $\tilde{\beta}_3$ (d) $\tilde{\beta}_4$

(e) $\tilde{\beta}_5$ (f) $\tilde{\beta}_7$

Figure 4.15: Comparing theoretical and sample variances of different slope estimators for a structural model with chi ξ (two degrees of freedom), and Normal errors. Sample variances are in blue, theoretical variances are in green.

Normal functional model



(a) $\tilde{\beta}_1$ (b) $\tilde{\beta}_2$

(c) $\tilde{\beta}_3$ (d) $\tilde{\beta}_4$

(e) $\tilde{\beta}_5$ (f) $\tilde{\beta}_7$

Figure 4.16: Comparing theoretical and sample variances of different slope estimators for a Normal functional model. Sample variances are in blue, theoretical variances are in green.

In the paper by Hood et al. [57], they state that a sample size of 50 is needed for the asymptotic results to be used with reasonable precision. It can be seen from the simulation study in this section, that this is a good general guideline. However, this number may be reduced in some circumstances. For example, when ξ is taken to follow a chi distribution with two degrees of freedom, for sample sizes greater than 30, the results for the sample and theoretical variances are virtually indistinguishable. There is even close agreement for sample sizes smaller than this. This is particularly the case where the estimator $\tilde{\beta}_1$ is used for a structural model with ξ following a chi distribution with two degrees of freedom. So for particular slope estimators, and particular constructs of errors in variables models it may be possible to lower this threshold provided by Hood et al. However as a safe general guideline, a sample size of 50 seems to be a useful threshold.

4.6 Estimator $\tilde{\beta}_8$

As detailed in Chapter 3, in order to use the estimator of the slope based on the third order moments, $\tilde{\beta}_8$ the distribution of both x and y must be sufficiently skewed. Moreover the sample sizes needed to accurately compute third order moments will inevitably be larger than those for first and second order moments. The aim of this section is to provide some advice as to the use of $\tilde{\beta}_8$. To model the skewness of ξ, a chi distribution with k degrees of freedom shall be used. As the number of degrees of freedom increases, then the chi distribution becomes more symmetric. Thus the performance of $\tilde{\beta}_8$ may be monitored as the degrees of freedom increases and the distribution of ξ becomes less skewed.

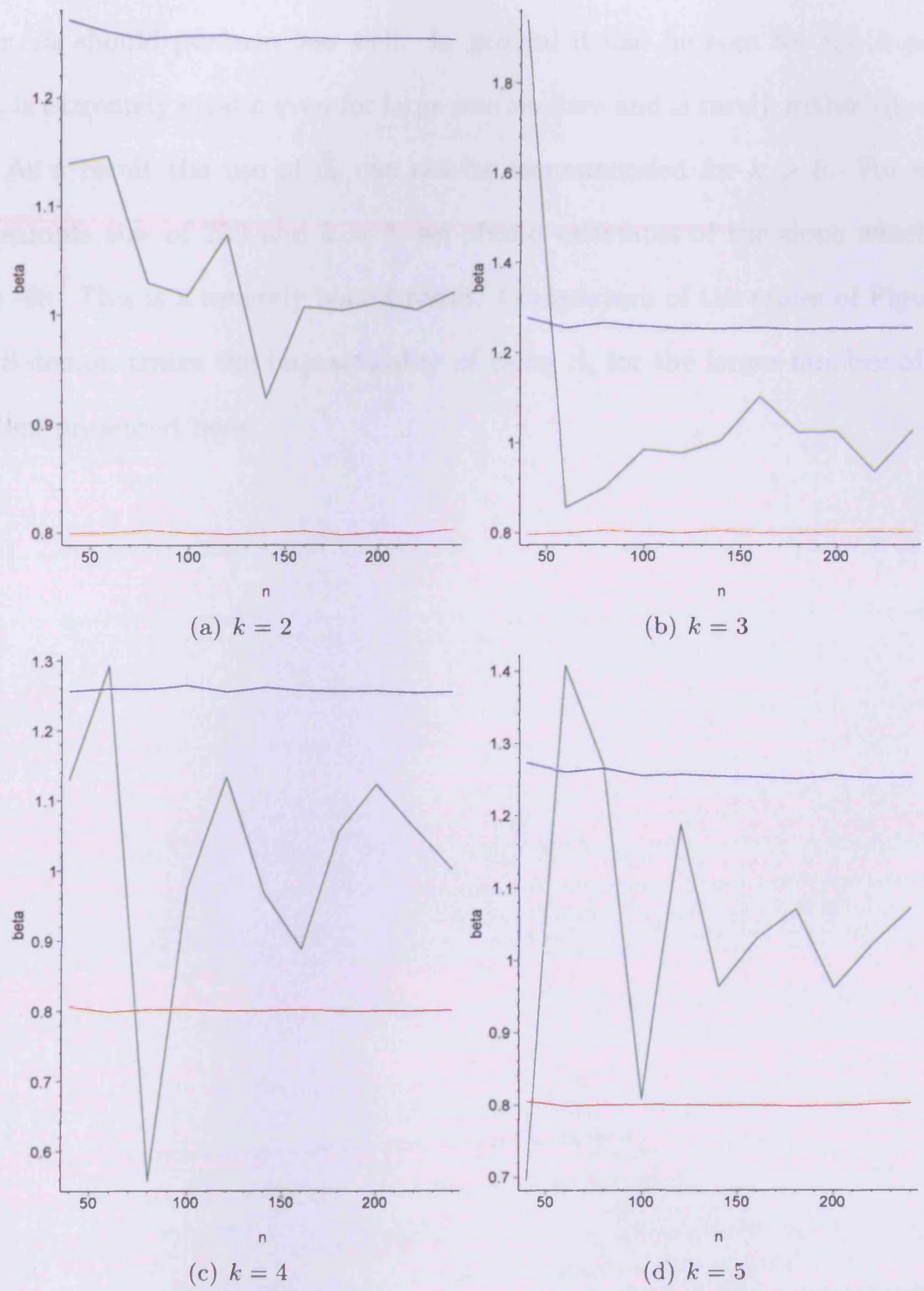
The sample size needed to estimate $\tilde{\beta}_8$ is also considered, as well as if the estimator $\tilde{\beta}_8$ lies outside of the range of slope estimators of y on x and x on y regression respectively. If $\tilde{\beta}_8$ lies outside of this range then admissible estimators of the variances will not be found.

The parameter settings chosen for the following simulations were $\alpha = 0$, $\beta = 1$. As the number of degrees of freedom increases, then so does the variance of the chi distribution. In order to compare results across the range of the degrees of freedom the reliability ratio for both the x and y measurement has been set to 0.8, and σ_δ^2 and σ_ε^2 derived accordingly. For completeness, the variances of the chi distribution with k degrees of freedom and the values of the error variances needed to maintain a reliability ratio in the x and y measurement of 0.8 are included in the following table:

k	Variance of chi distribution with k degrees of freedom	$\sigma_\delta^2 = \sigma_\varepsilon^2 =$
2	0.42920	0.10730
3	0.45352	0.11338
4	0.46571	0.11643
5	0.47293	0.11823
6	0.47767	0.11942
7	0.48101	0.12025
8	0.48349	0.12087
9	0.48541	0.12135
10	0.48692	0.12173

Figure 4.17 contains simulations of $\tilde{\beta}_8$ for a small number of degrees of freedom of the chi distribution. These simulations are therefore for particularly skewed ξ, and one expects $\tilde{\beta}_8$ to perform well. For all the degrees of freedom simulated here, $\tilde{\beta}_8$ performs well. For the most skewed ξ, $\tilde{\beta}_8$ only displays a small amount of bias for small sample sizes, but becomes virtually unbiased at a sample size of around 200. This also appears to be the case for the larger degrees of freedom as displayed in Figure 4.17. As k gets larger, then $\tilde{\beta}_8$ becomes more erratic, and for sample sizes less than 100 tends to stray

outside the range of slopes of y on x and x on y regression. However, at a sample size of again 200, $\tilde{\beta}_8$ settles down.



(a) $k = 2$ (b) $k = 3$

(c) $k = 4$ (d) $k = 5$

Figure 4.17: Values of $\tilde{\beta}_8$ for varying degrees of freedom and sample sizes. The y on x slope estimator is in red, and the x on y slope estimator is in blue.

Figure 4.18 contains simulations of $\tilde{\beta}_8$ for a larger number of degrees of freedom of the chi distribution. For the number of degrees of freedom displayed here the distribution of ξ becomes more symmetric, and thus from the theory presented in the previous Chapter, $\tilde{\beta}_8$ should perform less well. In general it can be seen for the k presented here, $\tilde{\beta}_8$ is extremely erratic even for large sample sizes and is rarely within the required range. As a result the use of $\tilde{\beta}_8$ can not be recommended for $k > 5$. For example, with a sample size of 250 and $k = 8$, we obtain estimates of the slope which are on average -40. This is a severely biased result. Comparison of the scales of Figures 4.17 and 4.18 demonstrates the impracticality of using $\tilde{\beta}_8$ for the larger number of degrees of freedom presented here.

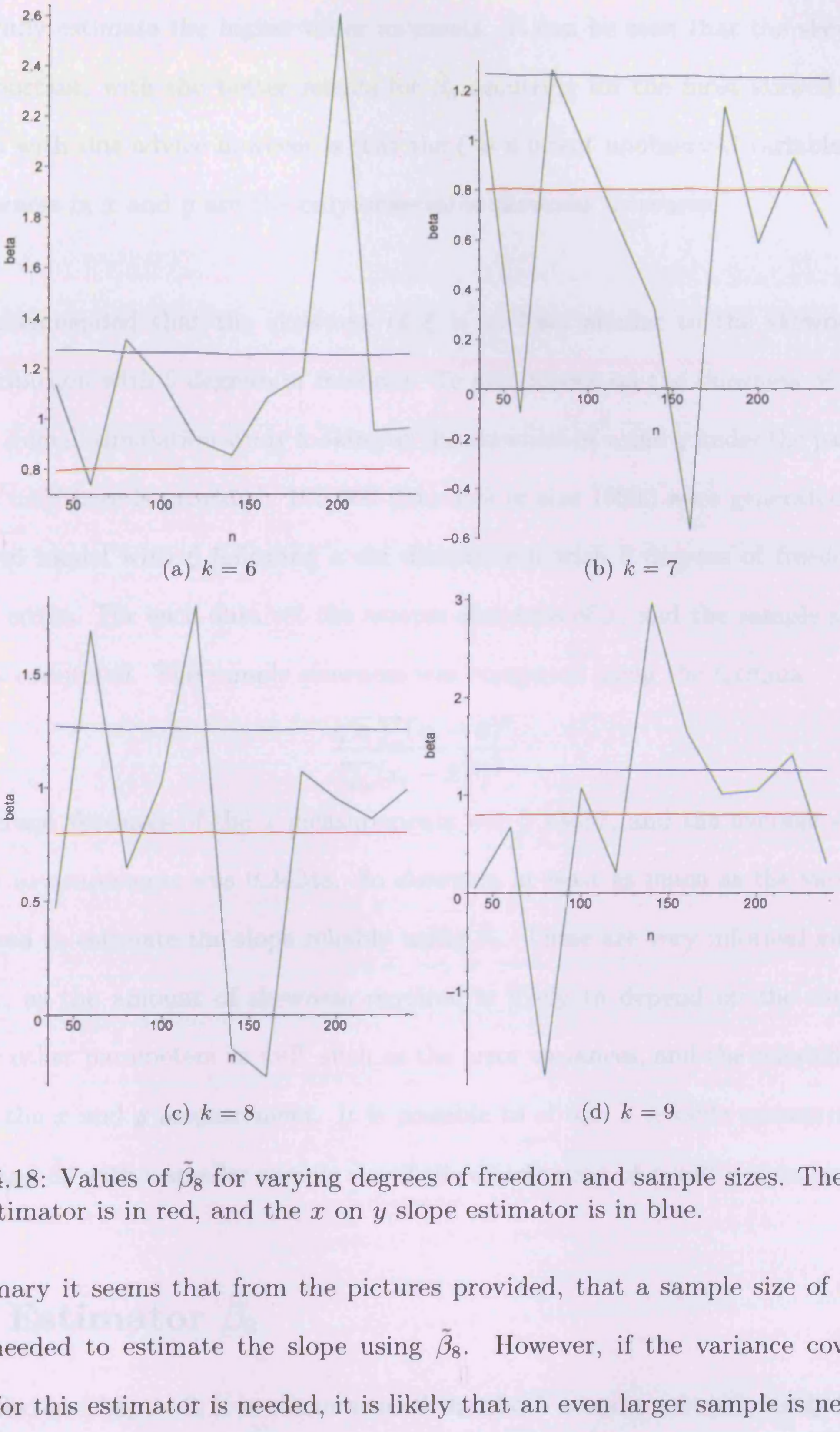


Figure 4.18: Values of $\tilde{\beta}_8$ for varying degrees of freedom and sample sizes. The y on x slope estimator is in red, and the x on y slope estimator is in blue.

In summary it seems that from the pictures provided, that a sample size of at least 150 is needed to estimate the slope using $\tilde{\beta}_8$. However, if the variance covariance matrix for this estimator is needed, it is likely that an even larger sample is needed to successfully estimate the higher order moments. It can be seen that the skewness of ξ is important, with the better results for $\tilde{\beta}_8$ occurring for the most skewed of ξ. A problem with this advice however is that the ξ is a latent unobserved variable, and so the skewness in x and y are the only observable skewness measures.

It is recommended that the skewness of ξ is at least similar to the skewness of a chi distribution with 5 degrees of freedom. To give advice on the skewness of x and y needed, a small simulation study looking at the skewness of x and y under the parameter settings used here is provided. 100,000 data sets of size 10000 were generated from a structural model with ξ following a chi distribution with 5 degrees of freedom, and Normal errors. For each data set the sample skewness of x, and the sample skewness of y was computed. The sample skewness was computed using the formula

$$\frac{\sqrt{n}\sum(x_i-\bar{x})^3}{[\sum(x_i-\bar{x})^2]^{\frac{3}{2}}}$$

The average skewness of the x measurements was 0.23557, and the average skewness of the y measurements was 0.24248. So skewness at least as much as the values here is required to estimate the slope reliably using $\tilde{\beta}_8$. These are very informal guidelines however, as the amount of skewness required is likely to depend on the slope, and possibly other parameters as well, such as the error variances, and the reliability ratio in both the x and y measurement. It is possible to obtain a reliable estimator of the slope using $\tilde{\beta}_8$ with a smaller sample size if the distribution of x and y is more skewed.

4.7 Estimator $\tilde{\beta}_9$

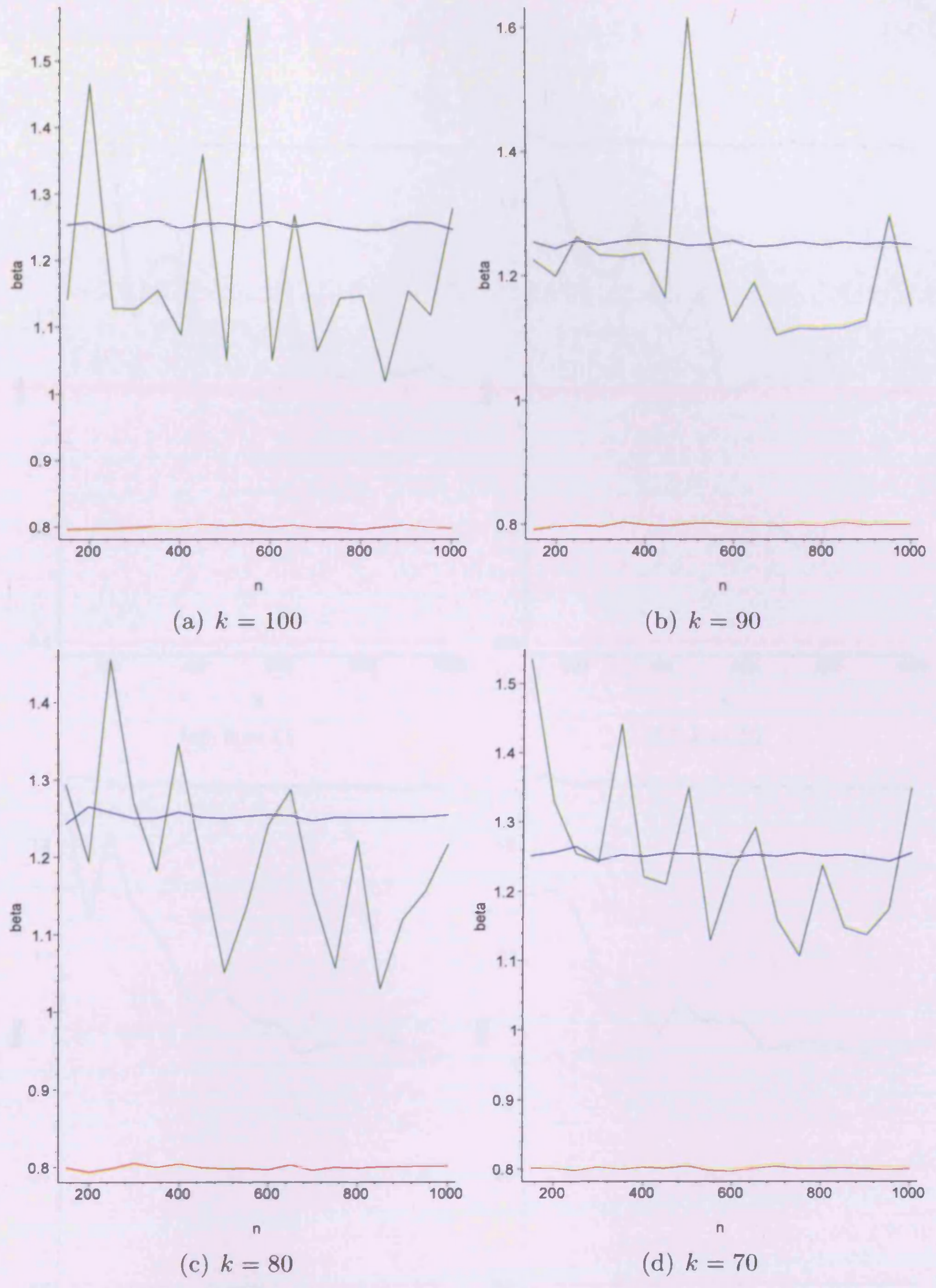
As detailed in Chapter 3, it is recommended that both x and y are sufficiently kurtotic for the estimator $\tilde{\beta}_9$ to be used. Moreover the sample sizes needed to accurately compute fourth order moments will inevitably be larger than those for first and second order moments. The aim of this section is to provide some advice as to the use of $\tilde{\beta}_9$. To model the kurtosis of ξ, Student's t distribution with k degrees of freedom shall be used. As the number of degrees of freedom decreases, then the Student's t distribution becomes more kurtotic. Thus the performance of $\tilde{\beta}_9$ may be monitored as the degrees of freedom decreases and the distribution of ξ becomes more kurtotic. For a large number of degrees of freedom the Student's t distribution is similar to the Normal distribution. The sample size needed to estimate $\tilde{\beta}_9$ is also considered, as well as the number of simulations where the estimator $\tilde{\beta}_9$ lies outside of the range of slope estimators of y on x and x on y regression respectively. If $\tilde{\beta}_9$ lies outside of this range then admissible estimators of the variances will not be found.

The parameter settings chosen for the following simulations were $\alpha = 0$, $\beta = 1$. As the number of degrees of freedom changes, then so does the variance of the Student's t distribution. In order to compare results across the range of the degrees of freedom the reliability ratio for both the x and y measurement has been set to 0.8, and σ_δ^2 and σ_ε^2 derived accordingly. For completeness, the variances of the Student's t distribution with the numbers of degrees of freedom considered in this simulation study, as well as the values of the error variances needed to maintain a reliability ratio in the x and y measurement of 0.8 are included in the following table:

k	Variance of Student's t distribution with k degrees of freedom	$\sigma_\delta^2 = \sigma_\epsilon^2 =$
100	1.02041	0.25510
90	1.02273	0.25568
80	1.02564	0.25641
70	1.02941	0.25735
11	1.22222	0.30556
10	1.25	0.3125
9	1.28571	0.32143
8	1.33333	0.33333
7	1.4	0.35
6	1.5	0.375
5	1.66667	0.41667
4	2	0.5

To demonstrate the importance of having kurtotic data, Figure 4.19 show values of $\tilde{\beta}_9$ assuming that ξ is from a Student's t distribution with a large number of degrees of freedom. The greater the number of degrees of freedom, the less kurtotic the data. Figure 4.19 demonstrates that $\tilde{\beta}_9$ behaves erratically and even for large sample sizes is greater than the x on y slope estimator.

The poor performance even for large sample sizes remains until the number of degrees of freedom is lowered to about $k = 11$. Figure 4.20 contains values of $\tilde{\beta}_9$ for a smaller number of degrees of freedom. As the degrees of freedom are decreased, then the performance of $\tilde{\beta}_9$ improves, although in general there appears to be a slight positive bias. It assumed that this is not due to the absence of small sample corrections in the sample moments as the samples are taken to be rather large. Note that in order to obtain a reliable estimator of β using $\tilde{\beta}_9$ the sample has to be much larger than for any other slope estimator discussed in Chapter 2. The sample size needed does depend however on how kurtotic the data is. For example, Figure 4.21 contains values of $\tilde{\beta}_9$ for an even smaller number of degrees of freedom. When $k = 4$, one may achieve a reliable estimate of β for sample size close to 200, but this would need to increase to a sample size of about 1000 when $k = 10$. Indeed, from inspection, it can be seen



(a) $k = 100$ (b) $k = 90$

(c) $k = 80$ (d) $k = 70$

Figure 4.19: Values of $\tilde{\beta}_9$ for varying degrees of freedom and sample sizes. The y on x slope estimator is in red, and the x on y slope estimator is in blue.

that for every additional degree of freedom added to $k = 4$, the sample size needs to increase further by an additional 100 (approximately).

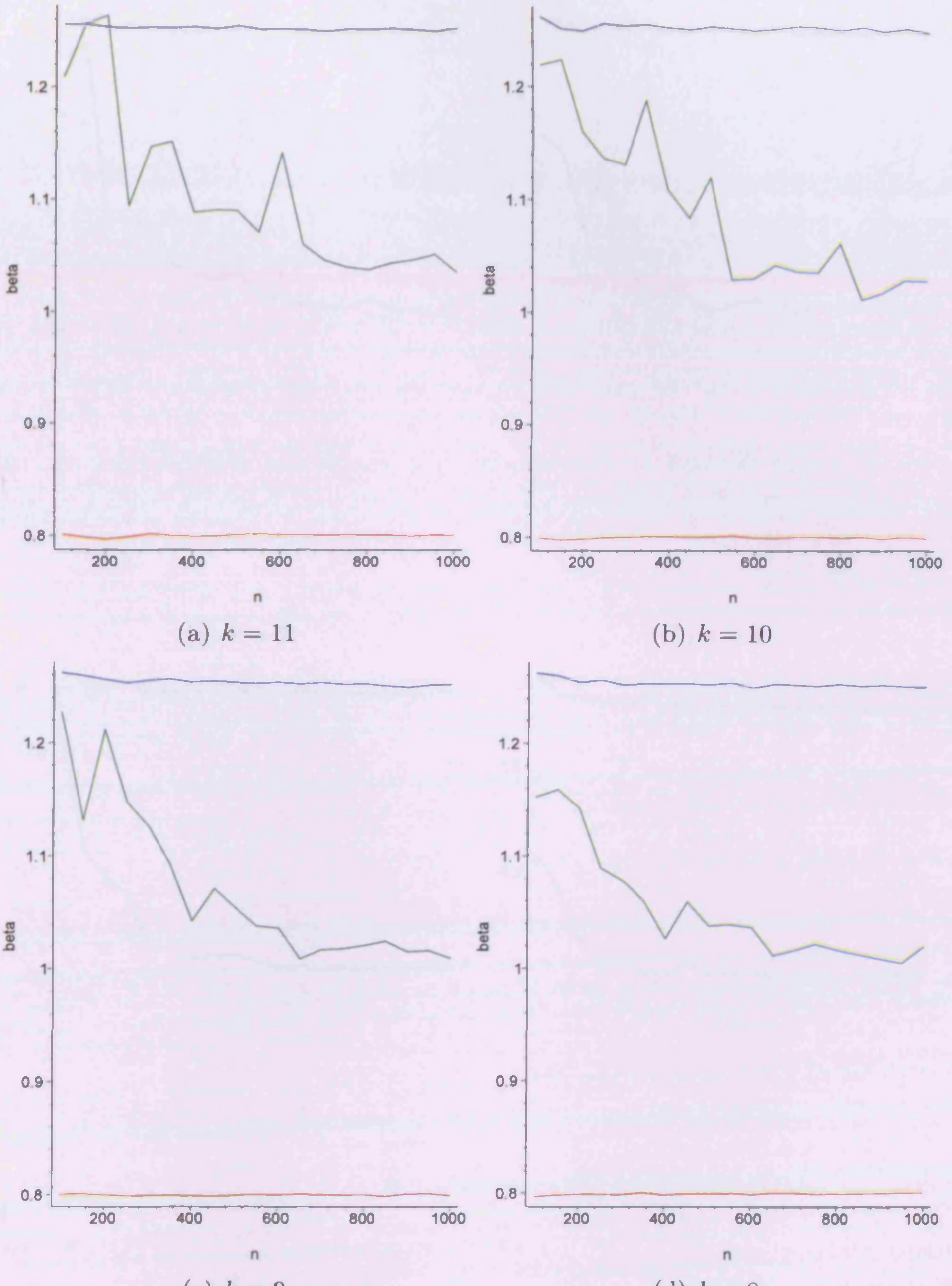
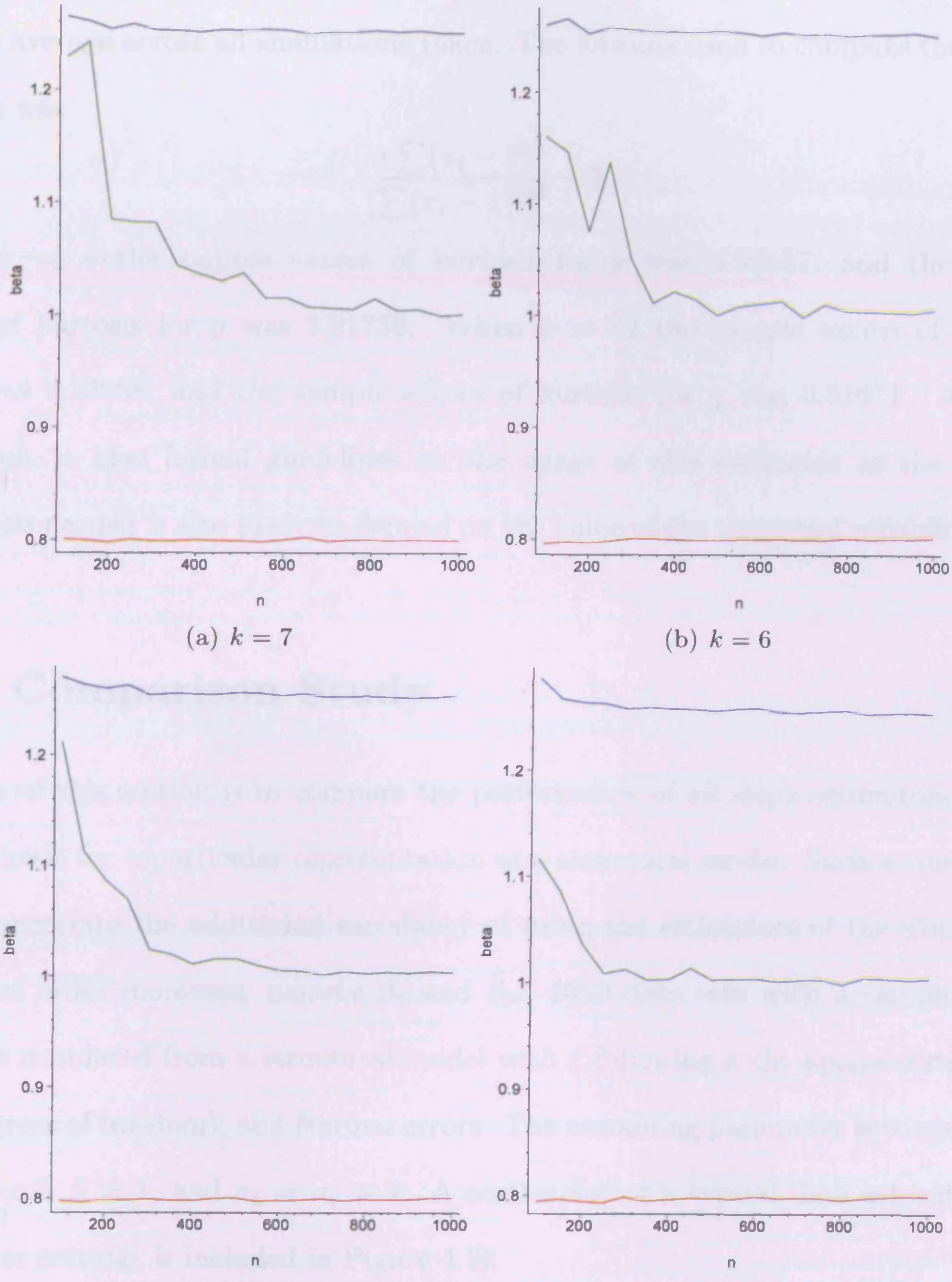


Figure 4.20: Values of $\tilde{\beta}_9$ for varying degrees of freedom and sample sizes. The y on x slope estimator is in red, and the x on y slope estimator is in blue.



(a) $k = 7$ (b) $k = 6$

(c) $k = 5$ (d) $k = 4$

Figure 4.21: Values of $\tilde{\beta}_9$ for varying degrees of freedom and sample sizes. The y on x slope estimator is in red, and the x on y slope estimator is in blue.

For illustrative purposes, the sample excess of kurtosis of x and y for a Student's t distribution with $k = 4$ and $k = 11$ for 100,000 data sets of size 1000 is computed, and the average across all simulations taken. The formula used to compute the sample kurtosis was

$$\frac{n\sum(x_i - \bar{x})^4}{[\sum(x_i - \bar{x})^2]^2} - 3$$

When $k = 4$ the sample excess of kurtosis for x was 3.99657, and the sample excess of kurtosis for y was 3.91756. When $k = 11$ the sample excess of kurtosis for x was 0.33088, and the sample excess of kurtosis for y was 0.51611. Again, it is difficult to give formal guidelines on the usage of this estimator as the amount of kurtosis needed is also likely to depend on the value of the slope and reliability ratio.

4.8 Comparison Study

The aim of this section is to compare the performance of all slope estimators derived in this thesis for a particular representation of a structural model. Such a comparison will demonstrate the additional variability of using the estimators of the slope based on higher order moments, namely $\tilde{\beta}_8$ and $\tilde{\beta}_9$. 1000 data sets with a sample size of 150 were simulated from a structural model with ξ following a chi-square distribution (five degrees of freedom), and Normal errors. The remaining parameter settings chosen were $\alpha = 0$, $\beta = 1$, and $\sigma_\delta = \sigma_\varepsilon = 2$. A scatterplot of a typical data set with these parameter settings is included in Figure 4.22

Figure 4.23 contains histograms of the estimators $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$, $\tilde{\beta}_4$, $\tilde{\beta}_5$, $\tilde{\beta}_7$, $\tilde{\beta}_8$ and $\tilde{\beta}_9$. The scales have deliberately been chosen to be different for each estimator, to demonstrate the differing variation in each estimator. All histograms appear to peak approximately

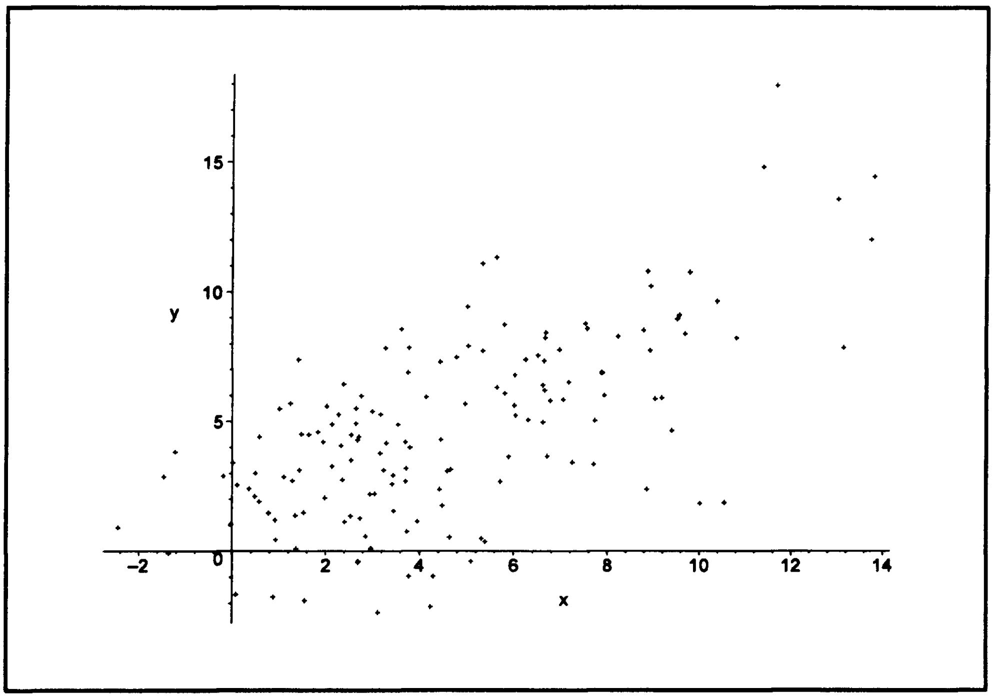
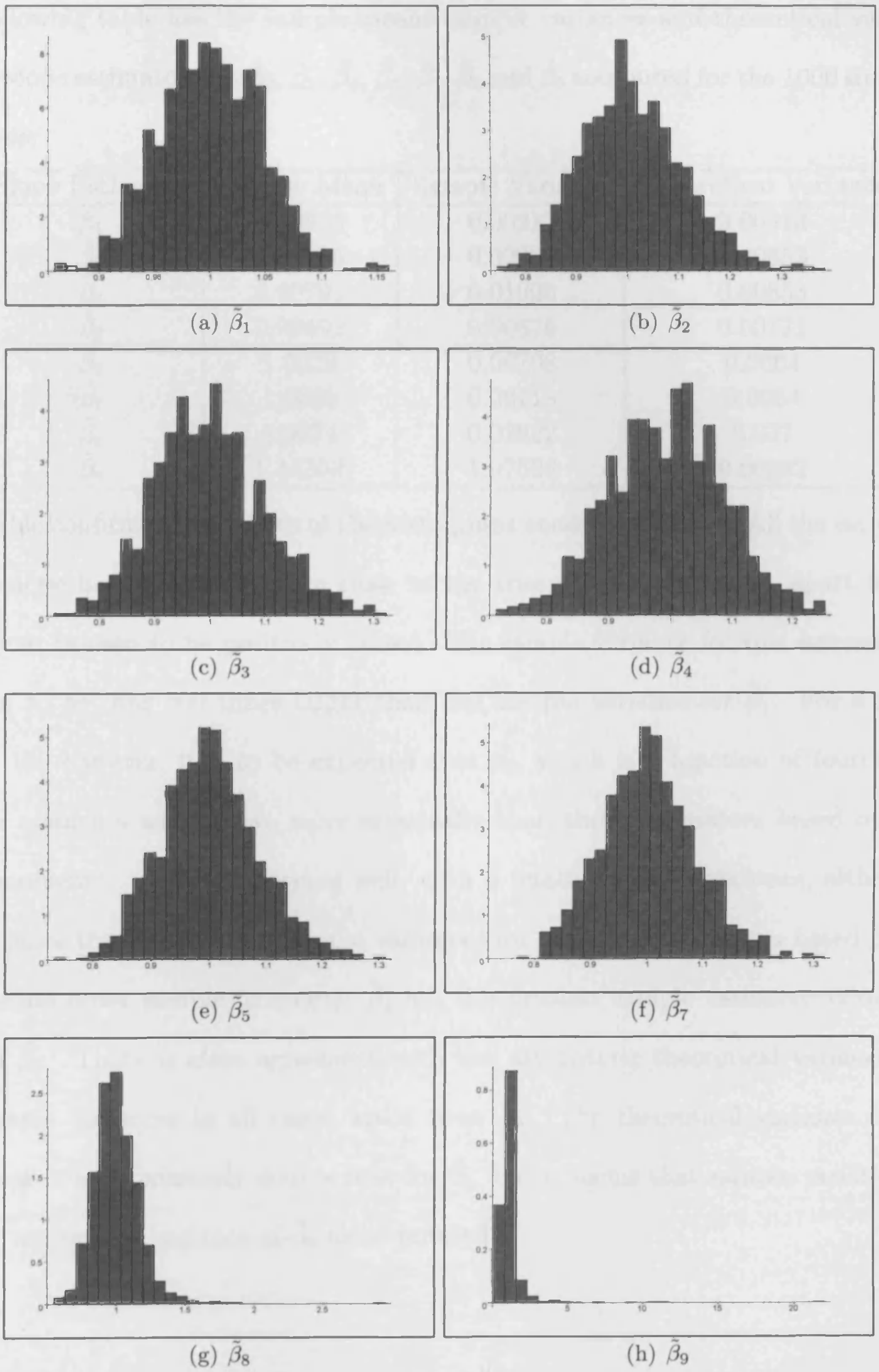


Figure 4.22: A typical scatterplot with ξ following a chi-square distribution (five degrees of freedom), and Normal errors. $\alpha = 0$, $\beta = 1$, and $\sigma_\delta = \sigma_\varepsilon = 2$.

around the true value of the slope $\beta = 1$. The spread of the histogram is minimal for the slope estimator $\tilde{\beta}_1$. As discovered in previous simulations, $\tilde{\beta}_1$ performs surprisingly well when there is skewed ξ. The histograms for $\tilde{\beta}_2$, $\tilde{\beta}_3$, $\tilde{\beta}_5$ and $\tilde{\beta}_7$ are very similar in appearance. For some samples, these slope estimators were estimating the slope as approximately 0.8 and 1.3 respectively. As to be expected, $\tilde{\beta}_8$ and $\tilde{\beta}_9$ perform least favourably. For both of these slope estimators, the peak of the histogram does appear to approximately lie above the true value of the slope $\beta = 1$, but there is much more spread in both of the histograms. Roughly speaking, the histogram for $\tilde{\beta}_8$ demonstrates that for some samples, the estimate of the slope is as extreme as 2.5, whilst the histogram for $\tilde{\beta}_9$ demonstrates that the slope is estimated as 20.



(a) $\tilde{\beta}_1$ (b) $\tilde{\beta}_2$

(c) $\tilde{\beta}_3$ (d) $\tilde{\beta}_4$

(e) $\tilde{\beta}_5$ (f) $\tilde{\beta}_7$

(g) $\tilde{\beta}_8$ (h) $\tilde{\beta}_9$

Figure 4.23: Histograms of different slope estimators for 1000 simulated data sets with a sample size of 150.

The following table has the sample means, sample variances and theoretical variances for the slope estimators $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$, $\tilde{\beta}_4$, $\tilde{\beta}_5$, $\tilde{\beta}_7$, $\tilde{\beta}_8$ and $\tilde{\beta}_9$ computed for the 1000 simulated data sets:

Slope Estimator	Sample Mean	Sample Variance	Theoretical Variance
$\tilde{\beta}_1$	0.99986	0.00212	0.00213
$\tilde{\beta}_2$	1.01056	0.00995	0.00853
$\tilde{\beta}_3$	0.99791	0.01000	0.00853
$\tilde{\beta}_4$	0.99492	0.00876	0.00771
$\tilde{\beta}_5$	1.0029	0.00708	0.0064
$\tilde{\beta}_7$	1.0028	0.00718	0.0064
$\tilde{\beta}_8$	1.0074	0.02922	0.037
$\tilde{\beta}_9$	1.14303	1.07506	0.06982

This table confirms the analysis of the histograms conducted earlier. All the estimators of the slope have a sample mean close to the true value of the slope, apart from $\tilde{\beta}_9$ which can be seen to be positively biased. The sample variance for this estimator can be seen to be over 500 times larger than the sample variance for $\tilde{\beta}_1$. For a sample size of 150 however, it is to be expected that $\tilde{\beta}_9$, which is a function of fourth order sample moments will behave more erratically than those estimators based on lower order moments. $\tilde{\beta}_8$ has performed well, with a relatively small variance, although it is still more than double the sample variances for the slope estimators based on first and second order sample moments. $\tilde{\beta}_1$ has the smallest sample variance, followed by $\tilde{\beta}_5$ and $\tilde{\beta}_7$. There is close agreement with the asymptotic theoretical variances and the sample variances in all cases, apart from $\tilde{\beta}_9$. The theoretical variance for this estimator is approximately double that for $\tilde{\beta}_8$, but it seems that sample variation has caused the sample variance of $\tilde{\beta}_9$ to be inflated.

4.9 Conclusions

All the estimators introduced in the previous Chapter have been investigated using simulations in this Chapter. It appears that the distribution of data is important when considering which estimator to use. For example, from these simulations, it would appear knowing the value of the intercept α would prove more beneficial when estimating the slope when ξ follows a chi distribution with two degrees of freedom, than knowing λ for example. The admissibility conditions are important, but for a sufficiently large sample should not be broken. As Dunn [39] states

> "If the model is correct, however, and the sample size is large enough, then we will get admissible estimates".

$\tilde{\beta}_8$ and $\tilde{\beta}_9$, although initially appealing as they do not require a restriction on the parameter space, do require larger samples than slope estimators based on first and second order moments. The data must also be sufficiently skewed or kurtotic respectively, but as stated, it is difficult to give formal and explicit guidelines as to how much skewness or kurtosis is needed as the situation is likely to depend on a multitude of factors. Nevertheless, it has been demonstrated that given certain conditions are met, $\tilde{\beta}_8$ and $\tilde{\beta}_9$ are perfectly usable.

Chapter 5

Maximum Likelihood

5.1 Introductory Remarks

The application of the maximum likelihood method to an errors in variables model has been briefly discussed in Chapter 2. The literature is silent on the question of maximum likelihood estimators except when $(\xi, \delta, \varepsilon)$ is taken to be trivariate Normal. This Chapter will extend this by providing further algebraic details for particular constructions of the errors in variables model.

This Chapter aims to illustrate the inherent difficulties in the maximum likelihood approach that are naturally avoided by using the method of moments approach advocated in Chapter 3. The complexity of the likelihood function for non-Normal ξ is such that in practice numerical methods would have to be employed to find maximum likelihood estimators, and the thorough investigation of the properties of these estimators would be lengthy and tedious.

In particular it is unlikely that theoretical results concerning the asymptotic variances of the estimators can be derived for anything other than the Normal structural model (this was done by Hood et al.[57]). To form the information matrix it is necessary to find the expectations of the various second derivatives with respect to the unknown parameters. Due to the complex form of the likelihood function, derivation of the second derivatives with respect to the parameters of the model are likely to be intangible. Thus computation of the information matrix, let alone its inverse is likely to be algebraically difficult.

5.2 Normal Structural Model

For the Normal structural model, it is assumed that

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \alpha+\beta\mu \end{pmatrix}, \begin{pmatrix} \sigma^2+\sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2+\sigma_\varepsilon^2 \end{pmatrix}\right]$$

and the log-likelihood of a random sample $\{(x_i, y_i), i = 1, \dots, n\}$ is given by

$$l = -n\ln(2\pi) - \frac{n}{2}\left[\frac{P}{|\Sigma|} + \frac{Q}{|\Sigma|} + \ln(|\Sigma|)\right]$$

(see for example Hood [56]) where

$$\begin{aligned} P &= s_{xx}(\beta^2\sigma^2+\sigma_\varepsilon^2) - 2\beta\sigma^2 s_{xy} + s_{yy}(\sigma^2+\sigma_\delta^2) \\ Q &= (\bar{x}-\mu)^2(\beta^2\sigma^2+\sigma_\varepsilon^2) - 2(\bar{x}-\mu)(\bar{y}-\alpha-\beta\mu)\beta\sigma^2 + (\bar{y}-\alpha-\beta\mu)^2(\sigma^2+\sigma_\delta^2) \end{aligned}$$

and $|\Sigma|$ is the determinant of the variance covariance matrix, a notation introduced in Chapter 3.

The term Q is minimised and equal to 0 when $\mu = \bar{x}$ and $\alpha + \beta\mu = \bar{y}$. These equations are identical to the method of moments estimating equations (3.1) and (3.2). The problem of maximising l is now reduced to minimising

$$l_N = \frac{P}{|\Sigma|} + \ln(|\Sigma|)$$

since α and μ do not appear as parameters in terms other than Q.

Therefore, the parameters remaining to be estimated are β, σ, σ_δ and σ_ε. Taking the partial derivatives of l_N with respect to these parameters and setting them to 0 yields the following set of likelihood equations as derived by Hood et al. [57]:

$$(\beta\sigma^2\sigma_\delta^2 + \beta\sigma^2 s_{xx} - \sigma^2 s_{xy})|\Sigma| - \beta\sigma^2\sigma_\delta^2 P = 0 \quad (5.1)$$

$$(\beta^2\sigma\sigma_\delta^2 + \sigma\sigma_\varepsilon^2 + \beta^2\sigma s_{xx} - 2\beta\sigma s_{xy} + \sigma s_{yy})|\Sigma| - (\beta^2\sigma\sigma_\delta^2 + \sigma\sigma_\varepsilon^2)P = 0 \quad (5.2)$$

$$(\sigma_\delta\sigma_\varepsilon^2 + \beta^2\sigma^2\sigma_\delta + \sigma_\delta s_{yy})|\Sigma| - (\sigma_\delta\sigma_\varepsilon^2 + \beta^2\sigma^2\sigma_\delta)P = 0 \quad (5.3)$$

$$(\sigma_\delta^2\sigma_\varepsilon + \sigma^2\sigma_\varepsilon + \sigma_\varepsilon s_{xx})|\Sigma| - (\sigma_\delta^2\sigma_\varepsilon + \sigma^2\sigma_\varepsilon)P = 0 \quad (5.4)$$

It can immediately be seen that these likelihood equations are not as compact as the method of moments estimating equations (3.1) to (3.5) and thus manipulation of the equations (5.1) to (5.4) is likely to prove more difficult.

Although there are now four likelihood equations in four parameters, these equations are not independent and similar to the method of moments approach of Chapter 3, a restriction on the parameter space must be made in order to make the likelihood equations (5.1) to (5.4) identifiable. This point is illustrated in detail by Hood [56].

As mentioned in Chapter 2, under a restriction the solution of the likelihood equations (5.1) to (5.4) will yield identical solutions to the method of moments estimating equations (3.1) to (3.5). This is because $\bar{x}$, $\bar{y}$, s_{xx}, s_{xy} and s_{yy} form a set of sufficient statistics for the two means, two variances and the covariance of the bivariate Normal distribution (Kendall and Stuart [67]). It is possible however to derive the method of moments estimating equations (3.1) to (3.5) via the method of maximum likelihood.

This has been done by Hood [56] and is illustrated here.

As stated, method of moments estimating equations (3.1) and (3.2) have already been derived from the likelihood function l. It remains to maximise l_N with respect to the terms that form the variance covariance matrix. Keeping the notation

$$P = s_{xx}(\beta^2\sigma^2 + \sigma_\varepsilon^2) - 2\beta\sigma^2 s_{xy} + s_{yy}(\sigma^2 + \sigma_\delta^2)$$

then

$$|\Sigma|\frac{\partial l_N}{\partial(\sigma^2 + \sigma_\delta^2)} = \beta^2\sigma^2 + \sigma_\varepsilon^2 + s_{yy} - \frac{(\beta^2\sigma^2 + \sigma_\varepsilon^2)P}{|\Sigma|} \tag{5.5}$$

$$|\Sigma|\frac{\partial l_N}{\partial(\beta^2\sigma^2 + \sigma_\varepsilon^2)} = \sigma^2 + \sigma_\delta^2 + s_{xx} - \frac{(\sigma^2 + \sigma_\delta^2)P}{|\Sigma|} \tag{5.6}$$

$$-\frac{|\Sigma|}{2}\frac{\partial l_N}{\partial(\beta\sigma^2)} = \beta\sigma^2 + s_{xy} - \frac{\beta\sigma^2 P}{|\Sigma|}. \tag{5.7}$$

Setting these equations to 0 and rearranging we obtain

$$\begin{aligned}\frac{P}{|\Sigma|} &= 1 + \frac{s_{yy}}{\beta^2\sigma^2 + \sigma_\varepsilon^2} \\ &= 1 + \frac{s_{xx}}{\sigma^2 + \sigma_\delta^2} \\ &= 1 + \frac{s_{xy}}{\beta\sigma^2}\end{aligned}$$

and it thus follows that

$$\frac{\beta^2\sigma^2 + \sigma_\varepsilon^2}{s_{yy}} = \frac{\sigma^2 + \sigma_\delta^2}{s_{xx}} = \frac{\beta\sigma^2}{s_{xy}}. \tag{5.8}$$

We may now manipulate these equations to obtain the method of moments estimating equations. Substituting P into (5.6) yields

$$(\sigma^2 + \sigma_\delta^2)|\Sigma| - (\beta\sigma^2)^2 s_{xx} + 2\beta\sigma^2(\sigma^2 + \sigma_\delta^2)s_{xy} - (\sigma^2 + \sigma_\delta^2)^2 s_{yy} = 0$$

and substituting $\beta\sigma^2 = s_{xy}$, and $\beta^2\sigma^2 + \sigma_\varepsilon^2 = s_{yy}$ gives after some manipulation

$$(\sigma^2 + \sigma_\delta^2)^3\frac{s_{yy}}{s_{xx}} - (\sigma^2 + \sigma_\delta^2)^3\frac{s_{xy}^2}{s_{xx}^4} + (\sigma^2 + \sigma_\delta^2)^2\frac{s_{xy}^2}{s_{xx}} - (\sigma^2 + \sigma_\delta^2)^2 s_{yy} = 0$$

which simplifies to $(\sigma^2 + \sigma_\delta^2) = s_{xx}$, the method of moments estimating equation (3.3).

Again substituting back into the identity (5.8) yields the remaining two method of moments estimating equations (3.4) and (3.5); $(\beta^2\sigma^2 + \sigma_\varepsilon^2) = s_{yy}$ and $\beta\sigma^2 = s_{xy}$ respectively.

This implies that the combinations of the parameters $\sigma^2 + \sigma_\delta^2$, $\beta\sigma^2$ and $\beta^2\sigma^2 + \sigma_\varepsilon^2$ may be identified using maximum likelihood, but, unless a restriction is made, the individual parameters β, σ^2, σ_δ^2 and σ_ε^2 may not be identified.

In conclusion then, the method of moments approach of Chapter 2 yields identical estimators to the maximum likelihood approach for the Normal structural model. The variance covariance matrices using the delta method and method of moments also agree with those of Hood et al. [57] who adopted a maximum likelihood approach. The asymptotics of the method of moments estimators derived via the delta method earlier have the advantage of not being solely constrained to the Normal structural model.

5.3 Normal Functional Model

The functional model construct is similar to that of the structural model, but has a crucial difference in the treatment of the latent ξ's. In the functional model, each ξ_i is assumed to be a fixed unknown constant, as opposed to a random variable as in the structural model. A potential problem with this type of model was highlighted by Neyman and Scott [79]. They questioned the use of asymptotic theory for this model since the introduction of each new observation increases the number of unknown parameters to estimate. The dimension of the variance covariance matrix increases as the number of observations increases.

The vast majority of papers on the functional model use a maximum likelihood approach. Some details of this approach are offered here. Assume the Normal functional model applies. For the functional model, there are $(n+4)$ parameters, namely, α, β, σ_δ^2, σ_ε^2 and the n latent ξ_i's. The likelihood function, L (see for example, Hood [56]) may be written as

$$L \propto \sigma_\delta^{-n}\sigma_\varepsilon^{-n} \exp\left[-\frac{1}{2\sigma_\delta^2}\sum_{i=1}^{n}(x_i-\xi_i)^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i-\alpha-\beta\xi_i)^2\right].$$

Differentiating $l = \ln L$ with respect to each of the parameters yields the following $(n+4)$ derivatives, which when equated to 0 give the turning points. In many cases the turning point can be identified as a global maximum, and maximum likelihood estimators are thus obtained. As will shortly be shown, this is not the case here.

$$\frac{\partial l}{\partial \xi_i} = \frac{x_i-\xi_i}{\sigma_\delta^2} + \frac{\beta}{\sigma_\varepsilon^2}(y_i-\alpha-\beta\xi_i) = 0, \qquad i = 1,\ldots,n \tag{5.9}$$

$$\frac{\partial l}{\partial \alpha} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i-\alpha-\beta\xi_i) = 0 \tag{5.10}$$

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\xi_i(y_i-\alpha-\beta\xi_i) = 0 \tag{5.11}$$

$$\frac{\partial l}{\partial \sigma_\delta} = -\frac{n}{\sigma_\delta} + \frac{1}{\sigma_\delta^3}\sum_{i=1}^{n}(x_i-\xi_i)^2 = 0 \tag{5.12}$$

$$\frac{\partial l}{\partial \sigma_\varepsilon} = -\frac{n}{\sigma_\varepsilon} + \frac{1}{\sigma_\varepsilon^3}\sum_{i=1}^{n}(y_i-\alpha-\beta\xi_i)^2 = 0. \tag{5.13}$$

Summing (5.9) over all i using (5.10) we obtain the following relationship at the turning point

$$\sum_{i=1}^{n}(x_i-\xi_i) = 0 \Rightarrow \bar{x} = \bar{\xi}$$

From (5.10) $\alpha = \bar{y} - \beta\bar{\xi}$, therefore from (5.11), we can write an estimator of β as

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})(y_i - \bar{y})}{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}. \tag{5.14}$$

This estimator is of no immediate use since the ξ_i's are unknown. It is, however, interesting to note that if the true ξ values are known exactly the maximum likelihood estimator of the slope is the usual least squares estimator. If each ξ_i is estimated with x_i, then the usual least squares estimate for the slope in a simple linear regression model $\widehat{\beta} = \frac{s_{xy}}{s_{xx}}$ is recovered. However as will be seen in Chapter 6, x_i is not an optimal estimator of ξ_i and so this is ill-advised. In Chapter 6 a number of possible estimators of ξ are considered. Most of these estimators are functions of β and so cannot be used here.

Equations (5.12) and (5.13) can be used to obtain maximum likelihood expressions for the error variances. They are

$$\begin{aligned} \hat{\sigma_\delta^2} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \xi_i)^2 \\ \hat{\sigma_\varepsilon^2} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha - \beta\xi_i)^2. \end{aligned}$$

If equation (5.9) is squared, we have,

$$\frac{(x_i - \xi_i)^2}{\sigma_\delta^4} = \frac{\beta^2}{\sigma_\varepsilon^4}(y_i - \alpha - \beta\xi_i)^2$$

and summing over all i yields $\sigma_\varepsilon^2 = \beta^2\sigma_\delta^2$, or

$$\lambda = \beta^2. \tag{5.15}$$

This is the exact assumption made in the method of moments approach in Chapter 3 when the geometric mean regression slope estimator, $\tilde{\beta}_{GM}$ is used to estimate the slope. Putting $\lambda = \beta^2$ into (5.9) gives us the values of ξ_i at the turning point. These will be discussed in more detail in Chapter 6,

$$\hat{\xi}_i = \frac{1}{2}\left(x_i + \frac{(y_i - \alpha)}{\beta}\right) = \frac{1}{2}\left(x_i + \frac{\sigma_\delta}{\sigma_\varepsilon}(y_i - \alpha)\right).$$

It is issues with estimators derived from these turning points that led authors such as Lindley [72] and Solari [95] to conclude that it is not worth proceeding with the maximum likelihood equation process. Solari for example highlighted that the likelihood function has a saddlepoint. We have no prior knowledge of the parameters β, σ_δ^2 and σ_ε^2, yet (5.15) gives a definite relation between the maximum likelihood estimators which may not be necessarily true in the model specified. Indeed, (5.15) implies that we cannot consistently estimate β, σ_δ^2 and σ_ε^2. This approach is therefore unacceptable for most applications.

Analogously to the structural model, one may not proceed with the functional model unless a further assumption restricting the parameter space is made. An additional problem was highlighted by Lindley [72], in that the maximum likelihood estimators of the error variances are inconsistent. This is highlighted via an example.

Assume that the ratio of the error variances λ is known. This reduces the number of likelihood equations by one, since (5.12) and (5.13) are now replaced with

$$\frac{\partial l}{\partial \sigma_\varepsilon} = -\frac{2n}{\sigma_\varepsilon} + \frac{\lambda}{\sigma_\varepsilon^3}\sum_{i=1}^{n}(x_i - \xi_i)^2 + \frac{1}{\sigma_\varepsilon^3}\sum_{i=1}^{n}(y_i - \alpha - \beta\xi_i)^2 = 0.$$

Assuming that λ is known removes the inconsistency of (5.15). The maximum likelihood estimator of σ_ε^2 is

$$\hat{\sigma_\varepsilon^2} = \frac{1}{2n}\left[\lambda\sum_{i=1}^{n}(x_i - \xi_i)^2 + \sum_{i=1}^{n}(y_i - \alpha - \beta\xi_i)^2\right]. \tag{5.16}$$

For this estimator to be of practical worth, it remains to find estimators of the latent ξ_i, and of the slope β.

From (5.9) we may write

$$(x_i - \xi_i) + \frac{\beta}{\lambda}(y_i - \alpha - \beta\xi_i) = 0$$

and so

$$\hat{\xi}_i = \frac{\lambda}{\lambda + \beta^2}\, x_i + \frac{\beta}{\lambda + \beta^2}\,(y_i - \alpha) = \bar{x} + \frac{1}{\lambda + \beta^2}\,[\lambda(x_i - \bar{x}) + \beta(y_i - \bar{y})]\,.$$

This estimator will be discussed in further detail in Chapter 6. Substituting this estimator into (5.16) gives

$$\begin{aligned}\hat{\sigma_\varepsilon^2} &= \frac{\lambda}{2n(\lambda + \beta^2)} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \\ &= \frac{\lambda}{2n(\lambda + \beta^2)} (s_{yy} - 2\beta s_{xy} + \beta^2 s_{xx}) \qquad (5.17)\end{aligned}$$

since $\alpha = \bar{y} - \beta\bar{\xi} = \bar{y} - \beta\bar{x}$.

Kendall and Stuart [67] is a reference which shows that the sample variances and covariances in the previous expression converge in probability to their expectations. We will now exploit this to show the inconsistency of $\hat{\sigma_\varepsilon^2}$. Letting $s_{\xi\xi}$ denote the variance of the latent ξ_i's $n^{-1}\sum(\xi_i - \bar{\xi})^2$, and $\xrightarrow{p}$ denote convergence in probability, then

$$s_{xx} \xrightarrow{p} s_{\xi\xi} + \sigma_\delta^2 \qquad (5.18)$$

$$s_{yy} \xrightarrow{p} \beta^2 s_{\xi\xi} + \lambda\sigma_\delta^2 \qquad (5.19)$$

$$s_{xy} \xrightarrow{p} \beta s_{\xi\xi}. \qquad (5.20)$$

Substituting these into (5.17) we can show that

$$\hat{\sigma_\varepsilon^2} \xrightarrow{p} \frac{\lambda}{2(\lambda + \beta^2)} \left[\beta^2 s_{\xi\xi} + \lambda\sigma_\delta^2 - 2\beta^2 s_{\xi\xi} + \beta^2(s_{\xi\xi} + \sigma_\delta^2)\right] = \frac{1}{2}\,\sigma_\varepsilon^2,$$

which shows the inconsistency. Kendall and Stuart point out that this is analogous to the correction for degrees of freedom in one-way analysis of variance, and can be rectified by using $2\hat{\sigma_\varepsilon^2}$.

The problems of inconsistency may be avoided by using the method of moments to estimate the parameters of the model. The method of moment estimating equations may be constructed using (5.18), (5.19) and (5.20). The estimating equations are:

$$\bar{x} = \bar{\xi} \tag{5.21}$$

$$\bar{y} = \alpha + \beta\bar{\xi} \tag{5.22}$$

$$s_{xx} = s_{\xi\xi} + \sigma_\delta^2 \tag{5.23}$$

$$s_{yy} = \beta^2 s_{\xi\xi} + \sigma_\varepsilon^2 \tag{5.24}$$

$$s_{xy} = \beta s_{\xi\xi}. \tag{5.25}$$

These equations are similar in appearance to those for the structural model, but have a different interpretation. In equations (5.21) and (5.22), $\bar{\xi}$ is no longer the mean of a random variable, but is the mean of the fixed constants ξ_i. Equations (5.23), (5.24) and (5.25) contain the term $s_{\xi\xi}$ as opposed to σ^2 for the functional model where $s_{\xi\xi} = n^{-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$.

The same problems that arise in the structural model also arise here. In the above construction, there are five equations, but six unknown parameters, μ, $s_{\xi\xi}$, α, β, σ_δ^2 and σ_ε^2. In order to solve these equations, a restriction on the parameter space has to be made. The above estimating equations may be solved once this restriction has been made, and the solutions are identical to those of the structural model. In conclusion the method of moments enables the usual parameters to be estimated without difficulty.

When the distribution of ξ is non-Normal, then the maximum likelihood approach is even more difficult. To illustrate this, a few details on the maximum likelihood approach for a number of examples when ξ follows a distribution other that Normal is included here.

5.4 Uniform ξ, Normal errors

Let ξ be an unobservable latent variable which follows a uniform distribution with finite support $[a, b]$. Then the probability density function of ξ, $f_\xi(\xi)$ can be written

$$f_\xi(\xi) = \frac{1}{(b-a)}$$

with ξ such that $a \le \xi \le b$.

The errors δ and ε are assumed to be mutually uncorrelated, and each follow an independent Normal distribution

$$\begin{aligned} \delta &\sim N(0, \sigma_\delta^2) \\ \varepsilon &\sim N(0, \sigma_\varepsilon^2) \end{aligned}$$

As ξ, δ and ε are mutually uncorrelated, then it follows that the joint p.d.f. of ξ, δ and ε is

$$f_{\xi,\delta,\varepsilon}(\xi, \delta, \varepsilon) = \frac{1}{(b-a)2\pi\sigma_\delta\sigma_\varepsilon} \exp\left(-\frac{\delta^2}{2\sigma_\delta^2} - \frac{\varepsilon^2}{2\sigma_\varepsilon^2} \right)$$

Now, consider the one to one transformation

$$\begin{aligned} x &= \xi + \delta \\ y &= \alpha + \beta\xi + \varepsilon \\ \xi &= \xi \end{aligned}$$

The Jacobian of this transformation is 1, and so the joint p.d.f. of x, y and ξ is

$$f_{x,y,\xi}(x,y,\xi) = \frac{1}{2\pi(b-a)\sigma_\delta\sigma_\varepsilon}\exp\left[-\frac{(\xi-x)^2}{2\sigma_\delta^2}-\frac{(y-\alpha-\beta\xi)^2}{2\sigma_\varepsilon^2}\right]$$

After some simplification and completing the square, we can write the term in the exponential as

$$-\frac{1}{2}\left[A\left(\xi-\frac{B}{A}\right)^2+\left(C-\frac{B^2}{A}\right)\right]$$

where

$$\begin{aligned} A &= \frac{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}{\sigma_\delta^2\sigma_\varepsilon^2} \\ B &= \frac{x}{\sigma_\delta^2}+\frac{\beta(y-\alpha)}{\sigma_\varepsilon^2} \\ C &= \frac{x^2}{\sigma_\delta^2}+\frac{(y-\alpha)^2}{\sigma_\varepsilon^2} \end{aligned}$$

So

$$f_{x,y,\xi}(x,y,\xi) = \frac{1}{\sqrt{2\pi}(b-a)\sigma_\delta\sigma_\varepsilon}\exp\left[-\frac{1}{2}\left(\frac{(y-\alpha-\beta x)^2}{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}\right)\right]\times\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{A}{2}\left(\xi-\frac{B}{A}\right)^2\right]$$

To obtain the joint p.d.f. of x and y, it remains to integrate out the ξ term

$$f_{x,y}(x,y) = \frac{1}{\sqrt{2\pi}(b-a)\sigma_\delta\sigma_\varepsilon}\exp\left[-\frac{1}{2}\left(\frac{(y-\alpha-\beta x)^2}{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}\right)\right]\int_a^b\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{A}{2}\left(\xi-\frac{B}{A}\right)^2\right]d\xi$$

Hence

$$\begin{aligned} f_{x,y}(x,y) &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}}\exp\left[-\frac{1}{2}\left(\frac{(y-\alpha-\beta x)^2}{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}\right)\right] \\ &\times\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(b-\frac{B}{A}\right)\right]-\Phi\left[\sqrt{A}\left(a-\frac{B}{A}\right)\right]\right\} \end{aligned} \qquad (5.26)$$

Here, Φ is the cumulative distribution function of the standard Normal distribution, that is

$$\Phi(u) = \int_{-\infty}^{u}\phi(t)\,dt$$

where

$$\phi(t) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)$$

is the standard Normal probability density function. This notation is now adopted for the rest of the thesis.

The term

$$\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(b-\frac{B}{A}\right)\right]-\Phi\left[\sqrt{A}\left(a-\frac{B}{A}\right)\right]\right\} \tag{5.27}$$

has some important features.

The key feature lies in the consideration of the B/A term. It can be seen that

$$\frac{B}{A}=\frac{\frac{x}{\sigma_\delta^2}+\frac{\beta(y-\alpha)}{\sigma_\epsilon^2}}{\frac{\sigma_\epsilon^2+\beta^2\sigma_\delta^2}{\sigma_\delta^2\sigma_\epsilon^2}}=\frac{\lambda}{\lambda+\beta^2}x+\frac{\beta}{\lambda+\beta^2}(y-\alpha)=\tilde{\xi}.$$

This is the method of moments estimator for the latent, unobserved ξ that has been mentioned earlier, but will be discussed in more detail in the next Chapter and beyond.

This observation allows an analysis of the term (5.27). For ξ close to the center of the distribution, that is $\xi \approx \frac{(a+b)}{2}$, then (5.27) becomes

$$\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(b-\frac{(a+b)}{2}\right)\right]-\Phi\left[\sqrt{A}\left(a-\frac{(a+b)}{2}\right)\right]\right\}$$

which may be written as

$$\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(\frac{(b-a)}{2}\right)\right]+1-\Phi\left[\sqrt{A}\left(\frac{(b-a)}{2}\right)\right]\right\}=\frac{1}{(b-a)}.$$

When ξ is close to a, the left end point of the support, that is $\xi \approx a$, then (5.27) becomes

$$\frac{1}{(b-a)}\left[\int_0^{\sqrt{A}(b-a)}\phi(t)\,dt\right]\approx\frac{1}{2(b-a)},$$

and the identical result holds for ξ close to b. $\sqrt{A}(b-a)$ is likely to be a lot larger than 3 for most applications since $\sqrt{A}$ is of order $1/\sigma_\delta^2$ for moderate β and $(b-a)$, for

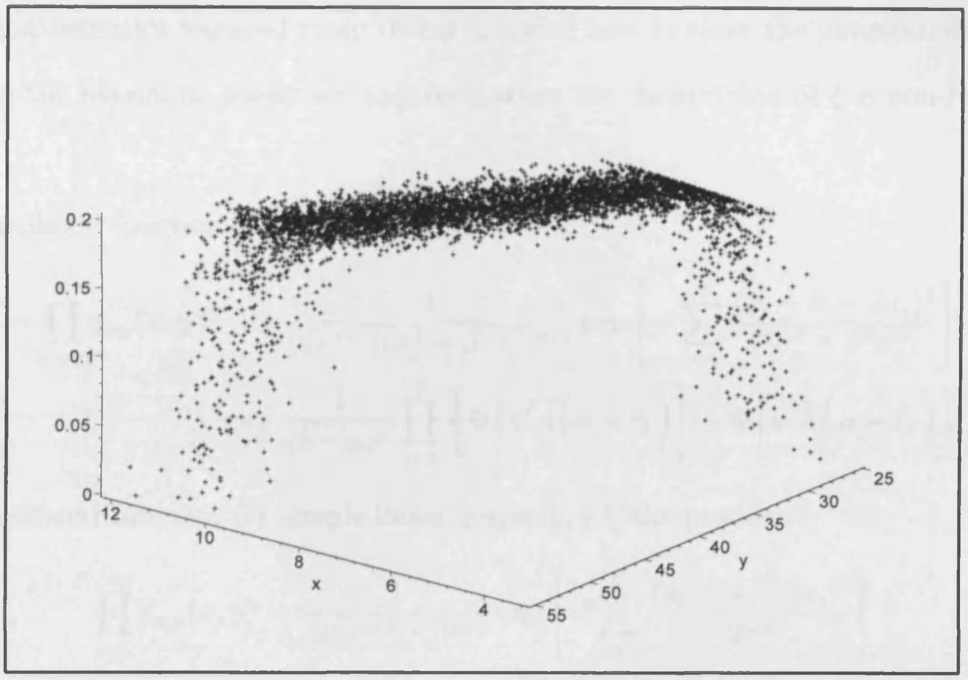


Figure 5.1: Demonstration of behaviour of (5.27) for a simulated data set with uniform ξ

most applications, is large compared to σ_δ^2. If β is large it is immediately seen that $\sqrt{A}(b-a)$ is large. Thus we have the approximation

$$\int_0^{\sqrt{A}(b-a)} \phi(t)\, dt \approx \frac{1}{2}.$$

As an example of the behaviour of (5.27), Figure 5.1 is a plot of (5.27) against x and y for a simulated data set of 5000 points with uniform ξ. The parameters chosen were $a = 5$, $b = 10$, $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 0.7$, $\sigma_\varepsilon = 1$. Note for these parameter settings $\sqrt{A}(b-a) = 132.1429 >> 3$. It can be seen that (5.27) is bounded above by $\frac{1}{10-5} = 0.2$, and that (5.27) gets smaller at the tails of the distribution.

As we have now found the explicit joint probability density function for x and y, we may obtain the likelihood function and attempt to maximise it with respect to the unknown parameters in order to find estimators for these unknown parameters. Some of the mathematics required to do this is included here to show the potential difficulty in using the maximum likelihood approach when the distribution of ξ is non-Normal.

The likelihood function L, is the product

$$\begin{aligned} L = \prod_{i=1}^{n} f_{x,y}(x,y) &= \frac{1}{(2\pi)^{n/2}(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)^{n/2}} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)} \right] \\ &\times \frac{1}{(b-a)^n} \prod_{i=1}^{n} \left\{ \Phi\left[\sqrt{A}\left(b - \tilde{\xi}_i\right)\right] - \Phi\left[\sqrt{A}\left(a - \tilde{\xi}_i\right)\right] \right\} \end{aligned} \tag{5.28}$$

The likelihood function for simple linear regression is the product

$$\prod_{i=1}^{n} f_{x,y}(x,y) = \frac{1}{(2\pi)^{n/2}(\sigma_\varepsilon^2)^{n/2}} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2} \right] \tag{5.29}$$

assuming that the error term ε is Normally distributed.

The likelihood function (5.28) differs from the likelihood function of the simple linear regression model (5.29) by the inclusion of the term after the multiplication sign, and the inflated variation of σ_ε^2 to $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$. It can be seen that the term after the multiplication sign in (5.29) is likely to have an impact on the likelihood function (5.28) due to the investigation of (5.27) earlier but its effect is of the order $1/(b-a)^n$. The term β^2 in the denominator however fundamentally changes the form of the likelihood function (5.28).

As usual, it is more convenient to work with the log likelihood function $l = \ln(L)$.

$$\begin{aligned} l &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2) - \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)} - n\ln(b-a) \\ &+ \sum_{i=1}^{n} \ln\left\{ \Phi\left[\sqrt{A}\left(b - \tilde{\xi}_i\right)\right] - \Phi\left[\sqrt{A}\left(a - \tilde{\xi}_i\right)\right] \right\}. \end{aligned}$$

We can partition this log likelihood into the log likelihood function of the simple linear regression model with inflated variance

$$l_S = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2) - \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)} \tag{5.30}$$

and the additional terms obtained by assuming that ξ follows a uniform distribution

$$l_U = -n\ln(b-a) + \sum_{i=1}^{n} \ln\left\{\Phi\left[\sqrt{A}\left(b - \tilde{\xi}_i\right)\right] - \Phi\left[\sqrt{A}\left(a - \tilde{\xi}_i\right)\right]\right\}.$$

Hence, we may write the log-likelihood as $l = l_S + l_U$. Assuming that the support parameters a and b are known, then the parameters we wish to estimate are $\mu, \alpha, \beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2. It can be seen that maximising the likelihood function will prove difficult, and could only be achieved by numerical search methods. Thus the method of moments must remain the preferred approach. Maximum likelihood would prove more difficult if the parameters a and b were assumed to be unknown.

The term l_U must not be ignored however. As an example, Figure 5.2 contains the log likelihood functions $l = l_S + l_U$, l_S and the log likelihood function of the simple linear regression model with variance σ_ε^2 (equation (5.30) with $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$ replaced by σ_ε^2) where all parameters, except β, are assumed fixed. The parameters used in this simulation are $a = 5$, $b = 10$, $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 1$, and $\sigma_\varepsilon = 1$, with $n = 5000$.

The log likelihood of the simple linear regression model is maximised at a value below the true value of β. The log likelihood functions l and l_S are very flat over the range which includes the true value of β. The lack of a well defined maxima for both of these likelihoods imply that it would be difficult to obtain an accurate slope estimator via maximum likelihood.

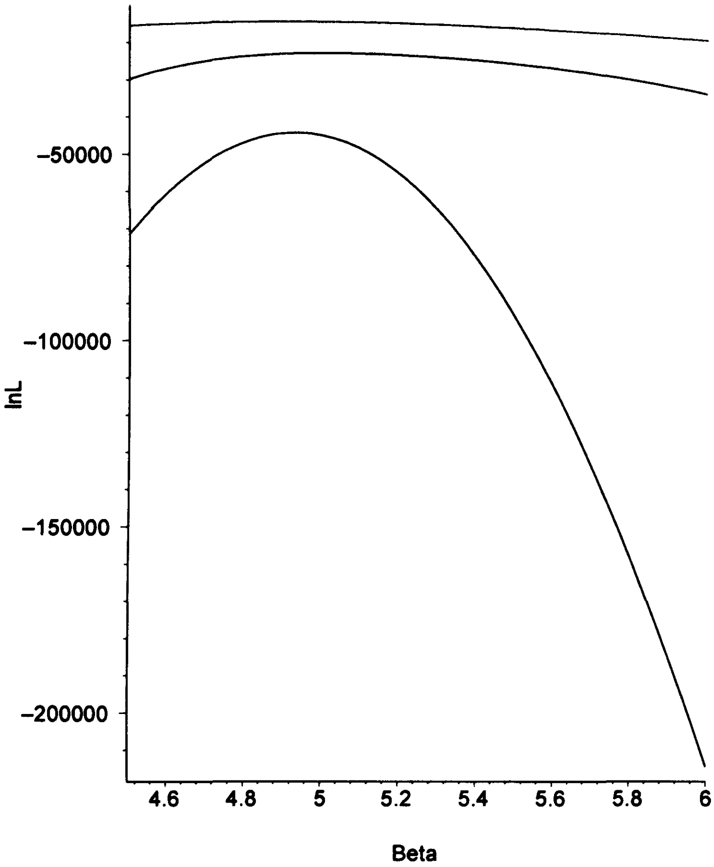


Figure 5.2: Blue curve denotes l, red curve denotes l_S, black curve denotes log likelihood for simple linear regression.

5.5 Chi ξ, Normal errors

As another example of the maximum likelihood approach, let ξ follow a chi distribution with k degrees of freedom. The probability density function of ξ is thus

$$f_\xi(\xi) = \frac{2^{(1-\frac{k}{2})}}{\Gamma(\frac{k}{2})}\xi^{(k-1)} \exp\left(-\frac{\xi^2}{2}\right)$$

with support $0 \leq \xi < \infty$. $\Gamma(t)$ is the standard Gamma function

$$\Gamma(t) = \int_0^\infty x^{(t-1)} \exp(-x)\, dx.$$

The errors δ and ε are assumed to be mutually uncorrelated, and each follow an independent Normal distribution

$$\begin{aligned}\delta &\sim N(0,\sigma_\delta^2)\\ \varepsilon &\sim N(0,\sigma_\varepsilon^2).\end{aligned}$$

As ξ,δ and ε are mutually uncorrelated, then it follows that the joint p.d.f. of ξ,δ and ε is

$$f_{\xi,\delta,\varepsilon}(\xi,\delta,\varepsilon)=\frac{2^{(1-\frac{k}{2})}}{\Gamma(\frac{k}{2})2\pi\sigma_\delta\sigma_\varepsilon}\xi^{(k-1)}\exp\left[-\frac{x^2}{2}-\frac{\delta^2}{2\sigma_\delta^2}-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right].$$

Making the one to one transformation

$$\begin{aligned}x &= \xi+\delta\\ y &= \alpha+\beta\xi+\varepsilon\\ \xi &= \xi\end{aligned}$$

yields

$$f_{x,y,\xi}=\frac{2^{(1-\frac{k}{2})}}{\Gamma(\frac{k}{2})2\pi\sigma_\delta\sigma_\varepsilon}\xi^{(k-1)}\exp\left[-\frac{x^2}{2}-\frac{(x-\xi)^2}{2\sigma_\delta^2}-\frac{(y-\alpha-\beta\xi)^2}{2\sigma_\varepsilon^2}\right]$$

and after completing the square in ξ

$$f_{x,y,\xi}(x,y,\xi)=\frac{2^{(1-\frac{k}{2})}}{\Gamma(\frac{k}{2})2\pi\sigma_\delta\sigma_\varepsilon}\exp\left[-\frac{x^2}{2}\right]\exp\left[-\frac{(y-\alpha-\beta x)^2}{2(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)}\right]\xi^{(k-1)}\exp\left[-\frac{A}{2}\left(\xi-\frac{B}{A}\right)^2\right].$$

Letting $I(k)$ denote the integral

$$I(k)=\int_0^\infty \xi^{(k-1)}\exp\left[-\frac{A}{2}\left(\xi-\frac{B}{A}\right)^2\right]d\xi$$

then the joint probability density function of x and y is given by

$$f_{x,y}(x,y)=\frac{2^{(1-\frac{k}{2})}}{\Gamma(\frac{k}{2})2\pi\sigma_\delta\sigma_\varepsilon}\exp\left[-\frac{x^2}{2}\right]\exp\left[-\frac{(y-\alpha-\beta x)^2}{2(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)}\right]I(k).$$

The integral $I(k)$ may be evaluated for small degrees of freedom. For $k>3$ then the number of terms in the integral grows larger and become difficult to manipulate. The value $k = 1$ is of particular interest as then the chi distribution becomes a half-Normal distribution.

When $k = 1$

$$I(1) = \sqrt{\frac{2\pi}{A}}\Phi\left(\frac{B}{\sqrt{A}}\right).$$

It thus follows that

$$f_{x,y}(x,y) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}} \exp\left[-\frac{(y-\alpha-\beta x)^2}{2(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)}\right] \times \sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right] \Phi\left(\frac{B}{\sqrt{A}}\right).$$

This probability density function differs from that of simple linear regression with inflated variance $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$ by the inclusion of the term

$$\sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right] \Phi\left(\frac{B}{\sqrt{A}}\right). \tag{5.31}$$

As an example of the behaviour of (5.31), Figure 5.3 is a plot of (5.27) against x and y for a simulated data set of 5000 points with chi ξ and $k = 1$. The parameters chosen were $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 0.7$, $\sigma_\varepsilon = 1$.

The log-likelihood function l of a sample $\{(x_i, y_i), i = 1, \ldots, n\}$ may be written $l = l_s + l_c$ where

$$l_c = \frac{n}{2}\ln\left(\frac{2}{\pi}\right) - \sum\frac{x_i^2}{2} + \sum\ln\left[\Phi\left(\frac{B}{\sqrt{A}}\right)\right]$$

It can immediately be seen that the likelihood function may again be partitioned into two components. The first component l_s is that of simple linear regression with inflated variance $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$, and the second term l_c. This second term is different to the term that appears by choosing ξ to be from a uniform distribution as it still depends on x.

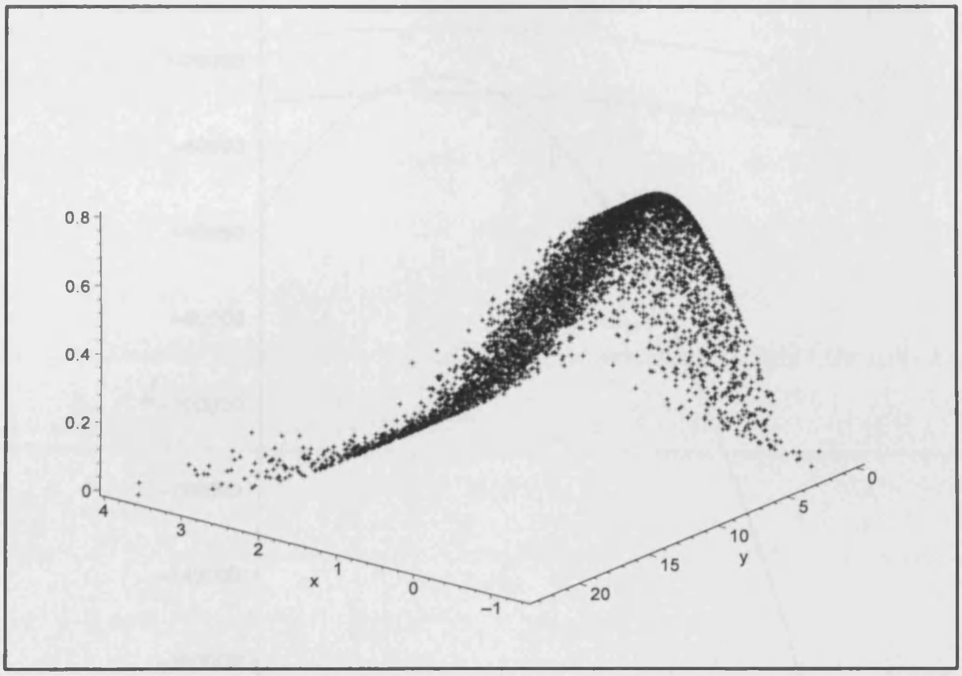


Figure 5.3: Demonstration of behaviour of (5.31) for a simulated data set with chi ξ and $k = 1$

Again the term l_c must not be ignored. As an example, the Figure 5.4 contains the log likelihood functions $l = l_S + l_c$, l_c and the log likelihood function of the simple linear regression model with variance σ_ε^2 (equation (5.30) with $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$ replaced by σ_ε^2) where all parameters, except β, are assumed fixed. The parameters used in this simulation are $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 0.7$, and $\sigma_\varepsilon = 1$, with $n = 5000$.

5.6 Conclusions

It can be seen that the maximum likelihood approach is not as straightforward as the method of moments based approach of Chapter 3. The numerical maximisation of a 6 dimensional likelihood is likely to prove difficult, and as can be seen in the examples investigated here the likelihood surface tends to be flat over a large range of possible β. This implies that constructing a confidence interval for the slope β of respectable

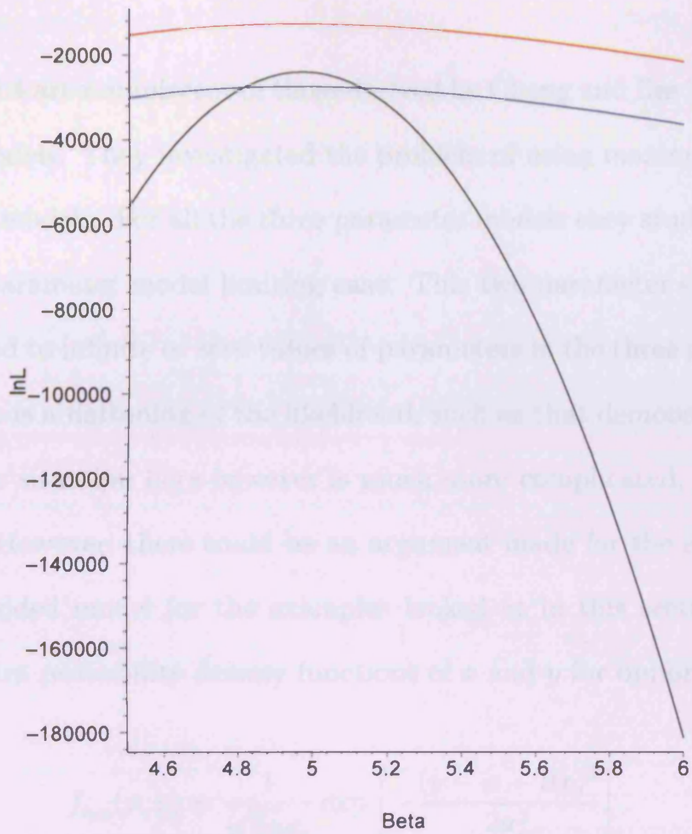


Figure 5.4: Blue curve denotes l, red curve denotes l_S, black curve denotes log likelihood for simple linear regression.

finite width is a difficult task.

For both uniform and chi ξ the inflation of the variance in the denominator of the joint probability functions of x and y has a more profound effect, when compared to the no measurement error case than a simple increase of variation. The presence of β in the term $(\sigma_\epsilon^2 + \beta^2\sigma_\delta^2)^{-\frac{n}{2}}$ has the effect of flattening the likelihood. It can be seen from Figures 5.2 and 5.4 that for the full model the likelihood remains almost constant over a wide range of β. This would make it difficult numerically to estimate β, but more seriously it is highly likely that for any specific data set the value of β identified as the maximum point would lead to estimates of some of the variance terms that are negative.

Figures 5.2 and 5.4 are reminiscent of those derived by Cheng and Iles [21] in their paper on embedded models. They investigated the problem of using maximum likelihood for three parameter models. For all the three parameter models they studied, there was an embedded two parameter model limiting case. This two parameter embedded limited case corresponded to infinite or zero values of parameters in the three parameter model. The effect of this is a flattening of the likelihood, such as that demonstrated in Figures 5.2 and 5.4. The situation here however is much more complicated, with there being six parameters. However, there could be an argument made for the existence of a five parameter embedded model for the examples looked at in this section. As $\sigma_\delta \to 0$, then both the joint probability density functions of x and y for uniform and chi ξ tend to

$$f_{x,y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left[-\frac{(y-\alpha-\beta x)^2}{2\sigma_\varepsilon^2}\right].$$

So it appears that a five parameter model is embedded within the six parameter errors in variables model. This embedded model is not the only one however. For example, taking $\sigma_\varepsilon \to 0$ would lead to the joint probability density function of x and y in accordance with x on y regression. Cheng and Iles demonstrate that when there is an embedded model, the method of maximum likelihood may be unable to identify certain parameters. It could be the case that this phenomenon is present for the errors in variables models discussed in this Chapter, and further explains the flattening of the likelihood functions.

For an errors in variables model, it appears that the combinations of the parameters $(\sigma^2 + \sigma_\delta^2)$, $\beta\sigma^2$ and $(\beta^2\sigma^2 + \sigma_\varepsilon^2)$ are embedded in the model, and they may be estimated from the second order sample moments s_{xx}, s_{xy} and s_{yy} respectively. To estimate the individual parameters σ^2, σ_δ^2 and σ_ε^2 requires additional information in order to recover them from these embedded parameters.

Thus to summarise, the method of maximum likelihood is difficult to apply, is likely to lead to inconsistent estimates of variance parameters, and would not easily lead to expressions for the asymptotic variances. It is for these reasons that for the most part this thesis concentrates on method of moments estimators.

Chapter 6

Prediction

6.1 Introductory Remarks

Cheng and Van Ness [20] commented that

> "Sometimes one constructs a regression model for the purpose of predicting y from x and other times one is more interested in the relationship between y and x"

Chapter 3 of this thesis has already dealt with the latter of their suggestions. Estimators for the linear structural model and corresponding variance covariance matrices have been provided. This allows the practitioner to estimate the relationship between x and y, and, after making some parametric assumptions, to construct approximate confidence intervals and hypothesis tests.

The first purpose of a regression model mentioned by Cheng and Van Ness concerns prediction. In an errors in variables model, there are a number of different prediction based questions that one may ask. Some of these are listed here:

- Given the data set $\{(x_i, y_i),\ i = 1, \ldots, n\}$, how may we estimate the unobserved data set $\{(\xi_i, \eta_i),\ i = 1, \ldots, n\}$?

- Given a ξ, ξ_i say, how does one estimate η_i?

- Given a x, x_i say, how does one estimate y_i?

In addition to the wide variety of prediction questions, there is another difficulty in that the answer for some of these questions depends on whether the model is assumed to be functional or structural. The aim of this Chapter is to clarify the prediction situation and offer some insights into the above questions.

6.2 Estimating y

It is reasonable to assume that a practitioner may wish to estimate a y value, given an x value. There is much confusion in the statistical literature regarding this problem, a point mentioned by Cheng and Van Ness [20]. They write

> "There is an interesting but sometimes misleading statement regarding the prediction of y from x that asserts that the ordinary regression least-squares predictor should be used even when dealing with the ME (measurement error) model"

However, this point is only true under particular circumstances. The main distinction in prediction lies in the inherent differences between the structural and functional model. Both the Normal functional and Normal structural models shall be considered here.

Normal functional model Consider firstly the Normal functional model. Each ξ is considered to be a fixed unknown constant such that $E[x_i] = \xi_i$ and $Var[x_i] = \sigma_\delta^2$.

For a fixed data point, the distribution of the errors $(\delta_i, \varepsilon_i)$ is given by

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}\right]$$

and so

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left[\begin{pmatrix} \xi_i \\ \alpha + \beta\xi_i \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}\right].$$

Note that here, x and y are independent. It thus follows that (using standard properties of multivariate Normal distributions)

$$E[y|x] = \alpha + \beta\xi, \tag{6.1}$$

where ξ is the true latent value that x is used to measure. ξ is an unknown parameter, and each data pair has a different ξ. Since the latent value η measured by y is assumed to be related to ξ by the equation $\eta = \alpha + \beta\xi$, the latent (ξ, η) pairs lie exactly along a straight line.

To estimate (6.1) then unbiased estimators of α and β are needed. To use the simple linear regression estimators here would yield a biased result. Hence, α and β should be estimated by estimators which take into account the errors in both variables. These were discussed in detail in Chapter 3.

Normal structural model The Normal linear structural model has trivariate distribution

$$\begin{pmatrix} \xi \\ \delta \\ \varepsilon \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}\right].$$

Using standard properties of multivariate Normal distributions, the marginal distribution of x and y obtained by transforming to the variables (x, y, ξ) and integrating with respect to ξ is

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \alpha+\beta\mu \end{pmatrix}, \begin{pmatrix} \sigma^2+\sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2+\sigma_\varepsilon^2 \end{pmatrix}\right].$$

It similarly follows from standard results concerning multivariate Normal distributions that

$$E[y|x] = \alpha + \beta\mu + \frac{\beta\sigma^2}{\sigma^2+\sigma_\delta^2}(x-\mu). \tag{6.2}$$

However the expression in (6.2) is often called the regression of y on x and for the Normal structural model is linear in x. This was the point made by Lindley [72]. There is some inconsistency in the literature concerning the use of the word regression. $E[y|x]$ is sometimes called the regression of y on x and is frequently confused with the least squares regression of y on x. A distinction between estimating the parameters of the straight line fit α and β and finding $E[y|x]$ is particularly important for an errors in variables model as they are often two separate constructs.

In practise, one can simply estimate the key components of (6.2) using the method of moment equations stated in Chapter 3. We thus obtain

$$E[y|x] \approx \bar{y} + \frac{s_{xy}}{s_{xx}}(x-\bar{x}).$$

This suggests that the standard least squares estimators are solely needed for prediction in a Normal structural model. However, in the functional model $E[y|x] = E[y] = \alpha+\beta\xi$ and α and β must be estimated via the errors in variables methodology outlined in Chapter 3.

The fact that the least squares estimates appear in $E[y|x]$ for the Normal structural model does not imply that the errors in variables estimates become redundant. The errors in variables estimates are still needed to quantify the relationship between x and y, and also to recover the unobserved data set. These least squares estimates only appear when considering how one may predict y, under a number of conditions, most notably when the distribution of $(\xi, \delta, \varepsilon)$ is assumed to be trivariate Normal. It will be shown later in this Chapter that assuming a non-Normal distribution for ξ the expressions for $E[y|x]$ are more complicated and not as clear cut.

There is then a clear distinction between the Normal functional model and the Normal structural model. For the functional model, if the error laws are considered to be Normal, then the errors in variables fit coincides with the expression for $E[y|x]$. For the structural model, if ξ is taken to be a Normally distributed random variable with Normal error laws, then the least squares line and not the errors in variables fit is used for $E[y|x]$. The algebra combines in such a way for the Normal structural model that the component $\beta\sigma^2(\sigma^2 + \sigma_\delta^2)^{-1}$ appears and this is estimated by the standard least squares regression slope estimator s_{xy}/s_{xx}.

In the following subsections, $E[y|x]$ will be investigated for a variety of distributions of ξ. Lindley [72] gave the explicit conditions for $E[y|x]$ to be linear, but what is the effect upon changing the distribution of ξ to a distribution other than Normal?. This can be investigated parametrically, and nonparametrically.

6.2.1 Parametric Approach

In this subsection, three differing parametric techniques of obtaining $E[y|x]$ shall be discussed. These are:

1. Kendall and Stuart's cumulant approach

2. Manipulating trivariate distributions

3. Cochran's bivariate approach and approximations

Kendall and Stuart's cumulant approach A possible parametric approach was introduced by Kendall and Stuart [67]. Their technique involves constructing expressions for $E[y|x]$ without explicitly determining the joint probability density function $f_{x,y}(x,y)$. Instead they derived a result based on cumulants and derivatives of the probability density function of x, $f_x(x)$. Letting D denote the differential operator $D^j f_x(x) = \frac{d^j f_x(x)}{dx^j}$, and $\kappa_{x,y}(i,j)$ denote the (i,j)-th bivariate cumulant of the joint probability density function of x and y then

$$E[y|x] = \sum_{r=0}^{\infty} \frac{\kappa_{x,y}(r,1)}{r!} \frac{(-D)^r f_x(x)}{f_x(x)}.$$

This approach does assume however that $f_x(x)$ has continuous derivatives at least in the support of x, and that the bivariate cumulants of (x,y) exist.

We can apply this methodology to the case when we have a Normal linear structural model as introduced earlier. The joint characteristic function of the standardised variables of $\frac{x-\mu}{\sqrt{\sigma^2+\sigma_\delta^2}}$ and $\frac{y-\alpha-\beta\mu}{\sqrt{\beta^2\sigma^2+\sigma_\epsilon^2}}$ is given by

$$\phi(t_1,t_2) = \exp\left\{-\frac{1}{2}(t_1^2+t_2^2+2\rho t_1 t_2)\right\},$$

where $\rho = \frac{Cov[x,y]}{\sqrt{Var[x]Var[y]}}$.

Due to the definition of bivariate cumulants, we thus have $\kappa_{x,y}(0,1) = 0$, $\kappa_{x,y}(1,1) = \rho$ and $\kappa_{x,y}(r,1) = 0$ for $r > 1$. As the variables have been standardised, the probability density function $f_x(x)$ is the standard Normal probability density function. Hence

$$E[y|x] = \rho x.$$

Returning to the original unstandardised variables gives us the identical result obtained earlier

$$\begin{aligned} E[y|x] &= \alpha + \beta\mu + \rho\frac{\sqrt{\beta^2\sigma^2 + \sigma_\varepsilon^2}}{\sqrt{\sigma^2 + \sigma_\delta^2}}(x - \mu) \\ &\approx \bar{y} + \frac{s_{xy}}{s_{xx}}(x - \bar{x}) \end{aligned}$$

This approach involves computation of the probability density function $f_x(x)$ and the bivariate cumulants of x and y. In the Normal linear structural model, we can use well-known properties of the Normal distribution to readily write down an expression for $E[y|x]$. Upon varying the distribution of ξ, invoking the above theory will not be as straightforward. In particular, finding the bivariate cumulants of x and y under non-Normal ξ is difficult. Moreover, the marginal probability density function, $f_x(x)$ may not in all cases be expressed as a neat closed form expression, and so differentiation may be infeasible.

Manipulating trivariate distributions For distributions of ξ other than Normal, this method is the most algebraically intensive. It involves working with specified distributions of ξ, δ and ε, and transforming these to obtain results regarding x and y.

The main crux of the method is to first obtain the joint probability density function of

ξ, δ and ε, $f_{\xi,\delta,\varepsilon}(\xi,\delta,\varepsilon) = f_\xi(\xi)f_\delta(\delta)f_\varepsilon(\varepsilon)$ and then make the one-to-one transformation

$$\begin{aligned} x &= \xi + \delta \\ y &= \alpha + \beta\xi + \varepsilon \\ \xi &= \xi \end{aligned}$$

which has unit Jacobian.

Integrating out the latent ξ over its support will then yield the marginal joint probability density function of x and y, $f_{x,y}(x,y)$. For some distributions of ξ, it is also possible to compute $f_x(x)$. The expression $E[y|x] = \int y\frac{f_{x,y}(x,y)}{f_x(x)}\,dy$ over the support of y can then be computed. Manipulating trivariate distributions is likely to prove more algebraically intensive than manipulating bivariate distributions. Thus the method that shall be exploited in this thesis is Cochran's bivariate approach, and this is discussed here.

Cochran's bivariate approach and approximations Cochran [22] wrote a paper solely on the problem of constructing $E[y|x]$ for a structural errors in variables model. He reiterates the point made at the beginning of this Chapter, that finding the linear relationship between x and y and predicting y from x are two separate constructs in an errors in variables model and must be considered separately. Indeed he states

> "There are at least two reasons for interest in this regression. The objective may be to obtain a consistent estimate of β for purposes of interpretation or adjustment of covariance. Secondly, the purpose may be to predict y from the fallible x by the regression technique in which the shape of this regression is irrelevant"

Cochran acknowledged the work of Lindley [72] that $E[y|x]$ is linear only for the Normal structural model. However, discussing typical applications of errors in variables model Cochran stated

> "my opinion is that in such applications even the Lindley conditions will not be satisfied, except perhaps by a fluke or as an approximation."

Cochran's tactic to solve the problem of constructing $E[y|x]$ is to reduce the problem from looking at the trivariate distribution of $(\xi, \delta, \varepsilon)$ to look at the bivariate distribution of (ξ, δ). This has two distinct advantages. Firstly, the algebra required in manipulating bivariate distributions is both simpler and neater than for trivariate distributions. Secondly, the results are more general than those using trivariate distributions as no distributional assumptions have to be made regarding the ε error term of the model.

To reduce the problem from trivariate distributions to bivariate distributions, Cochran noted that

$$E[y|x] = E[(\alpha + \beta\xi + \varepsilon)|x] = \alpha + \beta E[\xi|x] = \alpha + \beta R(x) \tag{6.3}$$

where $E[\varepsilon|x] = 0$ and $R(x) = E[\xi|x]$.

For some distributions of ξ and δ, $R(x)$ may be computed directly and substituted into (6.3) to obtain an expression for $E[y|x]$. For other distributions of ξ and δ the algebra needed to obtain $R(x)$ remains difficult and a neat closed form expression for $R(x)$ is not obtainable. In this scenario Cochran offers a method of approximating $R(x)$ by a quadratic or cubic function, depending on the distribution of ξ and δ. The approximation to $E[y|x]$ is obtained by adopting a least squares approach, and is discussed here.

The marginal probability density function $f_x(x)$ is formed by integrating out the latent ξ from the joint probability density function of ξ and δ, $f_{\xi,\delta}(\xi,\delta)$. In symbols

$$f_x(x) = \int f_{\xi,\delta}(\xi,\delta)\, d\xi = \int f_{\xi,\delta}(\xi, x-\xi)\, d\xi$$

and as

$$R(x) = \int \xi \, \frac{f_{\xi,\delta}(\xi, x-\xi)}{f_x(x)}\, d\xi$$

by definition of the conditional expectation, it follows that

$$R(x) f_x(x) = \int \xi f_{\xi,\delta}(\xi,\delta)\, d\xi. \tag{6.4}$$

This identity is the basis for Cochran's approximations to $E[y|x]$.

To estimate the form of $R(x)$, Cochran used the method of least squares. To demonstrate a link with the work of Lindley [72], Cochran initially constrained $R(x)$ to be linear. Then $E[y|x] = \alpha + \beta(c_0 + c_1 x)$ and the result described earlier for the Normal structural model will be recovered. Thus $R(x) \approx c_0 + c_1 x$ with the coefficients c_0 and c_1 determined by the minimisation of an objective function S, weighted by the probability density function $f_x(x)$. This is a continuous form of weighted least squares where each x is weighted by the value of its probability density function $f_x(x)$. So,

$$S = \int \left[R(x) - c_0 - c_1 x\right]^2 f_x(x)\, dx.$$

The minimisation of S may be done using standard calculus:

$$\frac{\partial S}{\partial c_0} = -2 \int \left[R(x) - c_0 - c_1 x\right] f_x(x)\, dx = 0 \tag{6.5}$$

$$\frac{\partial S}{\partial c_1} = -2 \int x \left[R(x) - c_0 - c_1 x\right] f_x(x)\, dx = 0 \tag{6.6}$$

Since using (6.4)

$$\int R(x) f_x(x)\, dx = \int \int \xi f_{\xi,\delta}(\xi,\delta)\, d\xi\, d\delta = \mu$$

then $c_0 = \mu(1 - c_1)$ from (6.5).

Since from (6.4)

$$\int xR(x)f_x(x)\,dx = \int(\xi+\delta)\int \xi f_{\xi,\delta}(\xi,\delta)\,d\xi\,d\delta = \int\int(\xi^2+\xi\delta)f_{\xi,\delta}(\xi,\delta)\,d\xi\,d\delta = \sigma^2+\mu^2,$$

then from (6.6)

$$\sigma^2 + \mu^2 - c_0\mu - c_1(\sigma^2 + \sigma_\delta^2 + \mu^2) = 0$$

giving $c_1 = \frac{\sigma^2}{\sigma^2+\sigma_\delta^2} = \kappa$, where κ is the reliability ratio.

It follows that

$$R(x) \approx \mu\left(1 - \frac{\sigma^2}{\sigma^2 + \sigma_\delta^2}\right) + \left(\frac{\sigma^2}{\sigma^2 + \sigma_\delta^2}\right)x.$$

This is the result Gleser [51] exploited for his method of obtaining an errors in variables fit for when all the random variables of a model are Normally distributed. Gleser's method is discussed in more detail later in this Chapter.

This expression for $R(x)$ may be substituted into (6.3) to obtain the result

$$\begin{aligned} E[y|x] &\approx \alpha + \beta\mu + \frac{\beta\sigma^2}{\sigma^2 + \sigma_\delta^2}(x - \mu) \\ &\approx \bar{y} + \frac{s_{xy}}{s_{xx}}(x - \bar{x}). \end{aligned}$$

This is the same result that has been derived via the other methods described earlier in this Chapter. It agrees with the results that are obtained for the Normal structural model. Cochran however does state that this is only an exact result for the Normal structural model and is a very crude approximation when the distributions of the random variables in the model differ from Normal. As a result, for structural models other than the Normal structural model, Cochran suggests an approximation based on a quadratic or cubic function. In general, it is possible to minimise any objection function of the form

$$S = \int [R(x) - r(x)]^2 f_x(x)\, dx. \tag{6.7}$$

though it is intuitively sensible to keep the estimation of $E[y|x]$ as simple as possible.

Cochran stated that through his own investigations, if ξ or δ follow a skew distribution then $R(x)$ can be approximated well by a quadratic curve. If ξ or δ are symmetrically distributed then he suggested that $R(-x) = -R(x)$ and a cubic curve with zero quadratic term approximation is valid. In order to find these approximations, the method illustrated above is generalised.

To fit a polynomial approximation to $R(x)$ the objective function S changes to

$$S = \int \left[R(x) - \sum_{i=1}^{p} c_i x^i \right]^2 f_x(x)\, dx$$

and there are $(p+1)$ equations that are needed to be solved to find the values of the coefficients $\{c_0, \ldots, c_p\}$. The r-th equation has general form

$$\sum_{i=1}^{p} c_i \mu_{x,i+r} = \int \int \xi(\xi + \delta)^r f_\xi(\xi) f_\delta(\delta)\, d\xi\, d\delta$$

where $\mu_{x,i+r}$ is the $(i+r)$-th central moment of x as defined in Chapter 1. In order to ensure that only the lowest order moments are used, it is essential that p is kept as small as possible. To do this, Cochran gave details only for quadratic and cubic approximations to $R(x)$. In addition, Cochran only provided results for data centered about their mean. The following results have been extended to cope with uncentered data and are thus not in an identical form to the expressions derived by Cochran.

When ξ or δ are skewed then Cochran states $R(x)$ can be approximated by $Q(x)$ where

$$Q(x) = \mu + c_1(x-\mu) + c_2[(x-\mu)^2 - \sigma^2 - \sigma_\delta^2],$$

and defining

$$\triangle = (\mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4})(\sigma^2+\sigma_\delta^2) - (\mu_{\xi 3} + \mu_{\delta 3})^2 - (\sigma^2+\sigma_\delta^2)^3,$$

the coefficients c_1 and c_2 are

$$\begin{aligned} c_1 &= \left[(\mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4})\sigma^2 - (\mu_{\xi 3}+\mu_{\delta 3})\mu_{\xi 3} - (\sigma^2+\sigma_\delta^2)^2\sigma^2\right]\triangle^{-1} \\ c_2 &= \left[(\sigma^2+\sigma_\delta^2)\mu_{\xi 3} - (\mu_{\xi 3}+\mu_{\delta 3})\sigma^2\right]\triangle^{-1}. \end{aligned}$$

When ξ and δ are both symmetrical then Cochran states that $R(x)$ can be approximated by $C(x)$ where

$$C(x) = \mu + c_1(x-\mu) - c_3(x-\mu)^3$$

and defining

$$\triangle = (\mu_{\xi 6} + 15\mu_{\xi 4}\sigma_\delta^2 + 20\mu_{\xi 3}\mu_{\delta 3} + 15\sigma^2\mu_{\delta 4} + \mu_{\delta 6})(\sigma^2+\sigma_\delta^2) - (\mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4})^2,$$

the coefficients c_1 and c_3 are

$$\begin{aligned} c_1 &= \big[(\mu_{\xi 6} + 15\mu_{\xi 4}\sigma_\delta^2 + 20\mu_{\xi 3}\mu_{\delta 3} + 15\sigma^2\mu_{\delta 4} + \mu_{\delta 6})(\sigma^2+\sigma_\delta^2) \\ &\quad -(\mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4})(\mu_{\xi 4} - 3\sigma^2\sigma_\delta^2)\big]\triangle^{-1} \\ c_3 &= \left[(\mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4})\sigma^2 - (\sigma^2+\sigma_\delta^2)\mu_{\xi 4} - 3(\sigma^2+\sigma_\delta^2)\sigma^2\sigma_\delta^2\right]\triangle^{-1}. \end{aligned}$$

Both these approximations $Q(x)$ and $C(x)$ have expected values μ, as

$$E\left[Q(x)\right] = \mu + c_1E[(x-\mu)] - c_2\left[Var[x] - \sigma^2 - \sigma_\delta^2\right] = \mu$$

and

$$E\left[C(x)\right] = \mu + c_1E[(x-\mu)] - c_3E[(x-\mu)^3] = \mu$$

since the use of the approximation $C(x)$ assumes that ξ and δ both follow a symmetrical distribution.

To demonstrate the use of Cochran's method some examples shall be considered here. For some of these examples, it is possible to compute a closed form expression for $R(x)$. Some examples of approximating $R(x)$ shall also be given.

Uniform ξ, Normal δ

Here an exact result can be derived. Let ξ be a random variable from a uniform distribution with support $a \leq \xi \leq b$. The probability density function of ξ on this support is

$$f_\xi(\xi) = \frac{1}{(b-a)}.$$

The error δ is assumed to be Normally distributed with the following probability density function

$$f_\delta(\delta) = \frac{1}{\sqrt{2\pi}\sigma_\delta} \exp\left[-\frac{\delta^2}{2\sigma_\delta^2}\right].$$

As ξ and δ are mutually uncorrelated then

$$f_{\xi,\delta}(\xi,\delta) = \frac{1}{(b-a)\sqrt{2\pi}\sigma_\delta} \exp\left[-\frac{\delta^2}{2\sigma_\delta^2}\right]$$

and it follows that

$$f_{\xi,x}(\xi,x) = \frac{1}{(b-a)\sqrt{2\pi}\sigma_\delta} \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right].$$

The marginal distribution $f_x(x)$ may be computed since

$$f_x(x) = \int_a^b f_{\xi,x}(\xi,x)\,d\xi = \frac{1}{(b-a)\sqrt{2\pi}\sigma_\delta} \int_a^b \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right] d\xi$$

where

$$\int_a^b \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right] d\xi = \sqrt{2\pi}\sigma_\delta \left[\Phi\left(\frac{x-b}{\sigma_\delta}\right) - \Phi\left(\frac{x-a}{\sigma_\delta}\right)\right].$$

Thus

$$f_x(x) = \frac{1}{(b-a)}\left[\Phi\left(\frac{x-b}{\sigma_\delta}\right) - \Phi\left(\frac{x-a}{\sigma_\delta}\right)\right].$$

The conditional probability density function of $\xi|x$ can be written as

$$f_{\xi|x}(\xi|x) = \frac{f_{\xi,x}(\xi,x)}{f_x(x)} = \frac{1}{\sqrt{2\pi}\sigma_\delta}\left[\Phi\left(\frac{x-b}{\sigma_\delta}\right) - \Phi\left(\frac{x-a}{\sigma_\delta}\right)\right]^{-1} \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right]$$

and the conditional expectation is therefore

$$E[\xi|x] = \frac{1}{\sqrt{2\pi}\sigma_\delta}\left[\Phi\left(\frac{x-b}{\sigma_\delta}\right) - \Phi\left(\frac{x-a}{\sigma_\delta}\right)\right]^{-1}\int_a^b \xi \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right] d\xi.$$

The integral $\int_a^b \xi \exp\left[-\frac{(x-\xi)^2}{2\sigma_\delta^2}\right] d\xi$ in the above expression is

$$\sigma_\delta^2\sqrt{2\pi}\left[\phi\left(\frac{x-b}{\sigma_\delta}\right) - \phi\left(\frac{x-a}{\sigma_\delta}\right)\right] + x\sigma_\delta\sqrt{2\pi}\left[\Phi\left(\frac{x-a}{\sigma_\delta}\right) - \Phi\left(\frac{x-b}{\sigma_\delta}\right)\right].$$

Therefore, after some simplification

$$R(x) = x + \sigma_\delta\left[\frac{\phi\left(\frac{x-b}{\sigma_\delta}\right) - \phi\left(\frac{x-a}{\sigma_\delta}\right)}{\Phi\left(\frac{x-b}{\sigma_\delta}\right) - \Phi\left(\frac{x-a}{\sigma_\delta}\right)}\right]$$

and $E[y|x] = \alpha + \beta R(x)$; where α and β are to be estimated by the errors in variables estimators outlined in Chapter 3, and not the standard least squares estimators.

It is important to remember the support of the conditional probability density function $\xi|x$. For the example considered here, the support is given by $a \leq \xi \leq b$. Due to measurement error however, the range of the observed x is usually extended beyond the original support of ξ. So when extrapolation beyond the support of ξ, only those data from the extremities are used. Therefore much care must be made in predicting $E[y|x]$ in the tails of the data, regardless of the distribution of ξ.

Figure 6.1 shows a plot of the least squares line (blue), the errors in variables line (green) and $E[y|x]$ (red) for a simulated data set of 5000 points. The curvature in $E[y|x]$ is readily seen, stressing the point that $E[y|x]$ may not be linear for a structural model other than the Normal structural model. For points close to the mean of the data, it appears that $E[y|x]$ tightly follows the errors in variables line, and then deviates away. For all the examples of deriving $E[y|x]$ in this Chapter the true line is indistinguishable from the errors in variables line and has so been omitted. The value of σ_δ has also been made deliberately large for ease of presentation of the examples.

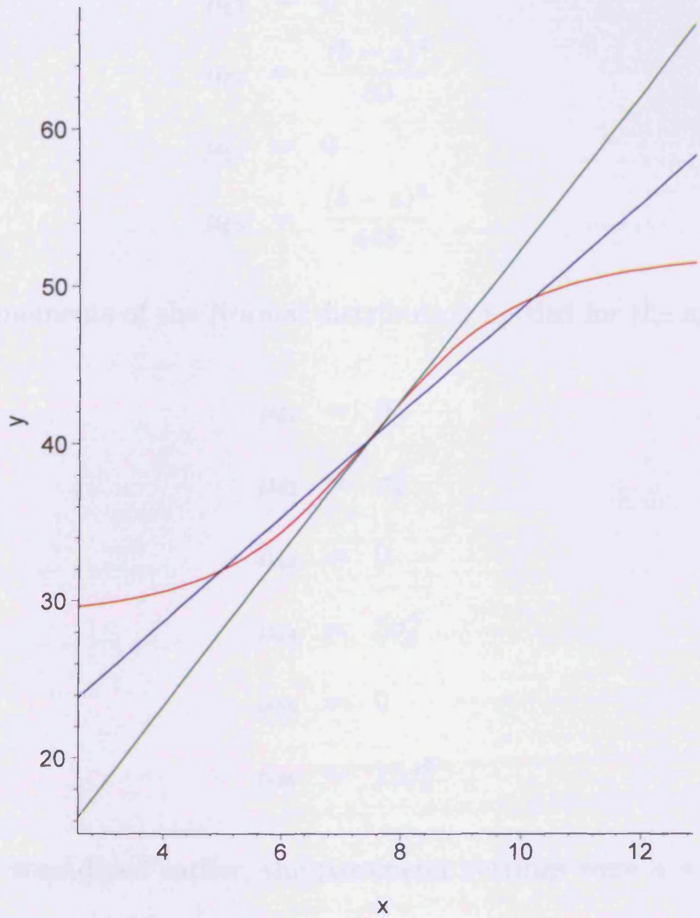


Figure 6.1: Simulated data set with uniform ξ and Normal errors, parameter values are $a = 5$, $b = 10$, $\sigma_\delta = \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$.

Uniform ξ, Normal δ

For comparison purposes Cochran's approximation is given here for this case. Since both ξ and δ have a symmetric distribution, then a cubic approximation to $E[y|x]$ was suggested by Cochran. In order to compute the expression $C(x)$ then the central moments of ξ and δ are needed, up to and including the sixth central moment. The central moments of the uniform distribution needed for the approximation are

$$\begin{aligned}
\mu_{\xi 1} &= \mu = 0 \\
\mu_{\xi 2} &= \sigma^2 = \frac{(b-a)^2}{12} \\
\mu_{\xi 3} &= 0 \\
\mu_{\xi 4} &= \frac{(b-a)^4}{80} \\
\mu_{\xi 5} &= 0 \\
\mu_{\xi 6} &= \frac{(b-a)^6}{448}
\end{aligned}$$

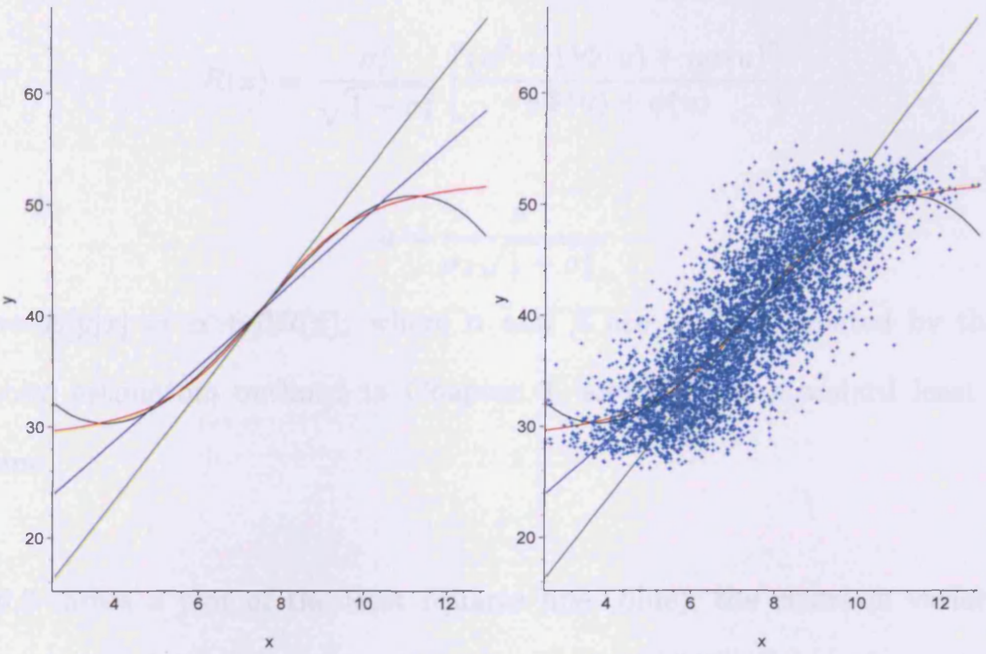
and the central moments of the Normal distribution needed for the approximation are

$$\begin{aligned}
\mu_{\delta 1} &= 0 \\
\mu_{\delta 2} &= \sigma_\delta^2 \\
\mu_{\delta 3} &= 0 \\
\mu_{\delta 4} &= 3\sigma_\delta^4 \\
\mu_{\delta 5} &= 0 \\
\mu_{\delta 6} &= 15\sigma_\delta^6.
\end{aligned}$$

For the example considered earlier, the parameter settings were $a = 5$, $b = 10$, $\sigma_\delta = \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$. For these settings, the approximation $C(x)$ is computed as

$$C(x) = 2.0833 + 0.82674(x - 2.0833) - 0.01998(x - 2.0833)^3$$

Figure 6.2 shows a plot of the least squares line (blue), the errors in variables line (green), the exact expression for $E[y|x]$ (red) and the approximation for $E[y|x] = \alpha + \beta[2.0833 + 0.82674(x - 2.0833) - 0.01998(x - 2.0833)^3]$ (black) for a simulated data set of 5000 points. The approximation is an extremely good fit, and is virtually indistinguishable from the exact result over a large range of the data. In particular, the fit is excellent over the the range $[5, 10]$ which is the original support of the ξ. The approximation however does deviate away from the exact expression for $E[y|x]$ at the tails, beyond the support of ξ.



(a) Without data (b) With data

Figure 6.2: Simulated data set with uniform ξ and Normal errors, parameter values are $a = 5$, $b = 10$, $\sigma_\delta = \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$.

Chi ξ, two degrees of freedom, Normal δ

An exact result may be derived for this case. This was an example also considered by Cochran, and the details are replicated here and discussed. The chi distribution with two degrees of freedom has been chosen as an example of a skew distribution that gives algebraically tractable results. The probability density function of ξ is

$$f_{\xi}(\xi) = \xi \exp\left[-\frac{\xi^2}{2}\right]$$

with support $\xi > 0$. Cochran stated that in this case, the exact result $R(x)$ is simplified to

$$R(x) = \frac{\sigma_\delta^2}{\sqrt{1+\sigma_\delta^2}}\left[\frac{(u^2+1)\Phi(u) + u\phi(u)}{u\Phi(u) + \phi(u)}\right]$$

where

$$u = \frac{x}{\sigma_\delta\sqrt{1+\sigma_\delta^2}}.$$

Therefore $E[y|x] = \alpha + \beta R(x)$; where α and β are to be estimated by the errors in variables estimators outlined in Chapter 3, and not the standard least squares estimators.

Figure 6.3 shows a plot of the least squares line (blue), the errors in variables line (green) and $E[y|x]$ (red) for a simulated data set of 5000 points. The curvature in $E[y|x]$ is again seen, and appears to follow the least squares line deviating most at the left hand tail.

Chi ξ, two degrees of freedom, Normal δ

As an example of the use of Cochran's approximation, details are provided for this case. Since ξ follows a skewed distribution, and δ follows a symmetric distribution then a quadratic approximation to $E[y|x]$ was suggested by Cochran. In order to compute the

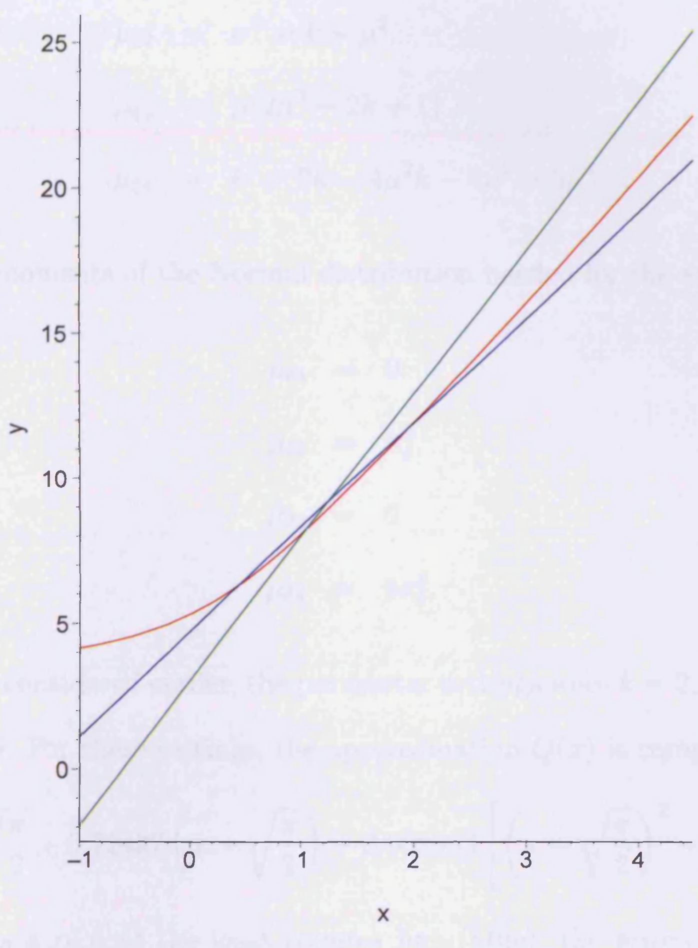


Figure 6.3: Simulated data set with chi ξ and Normal errors, parameter values are $\sigma_\delta = 0.4, \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$.

expression $Q(x)$ then the central moments of ξ and δ are needed, up to and including the fourth central moment. The central moments of the chi distribution with k degrees of freedom needed for the approximation are

$$\begin{aligned}
\mu_{\xi 1} &= \mu = \sqrt{2}\frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \\
\mu_{\xi 2} &= \sigma^2 = k - \mu^2 \\
\mu_{\xi 3} &= \mu(2\mu^2 - 2k + 1) \\
\mu_{\xi 4} &= k^2 + 2k - 4\mu^2 k - 4\mu^2 - 3\mu^4
\end{aligned}$$

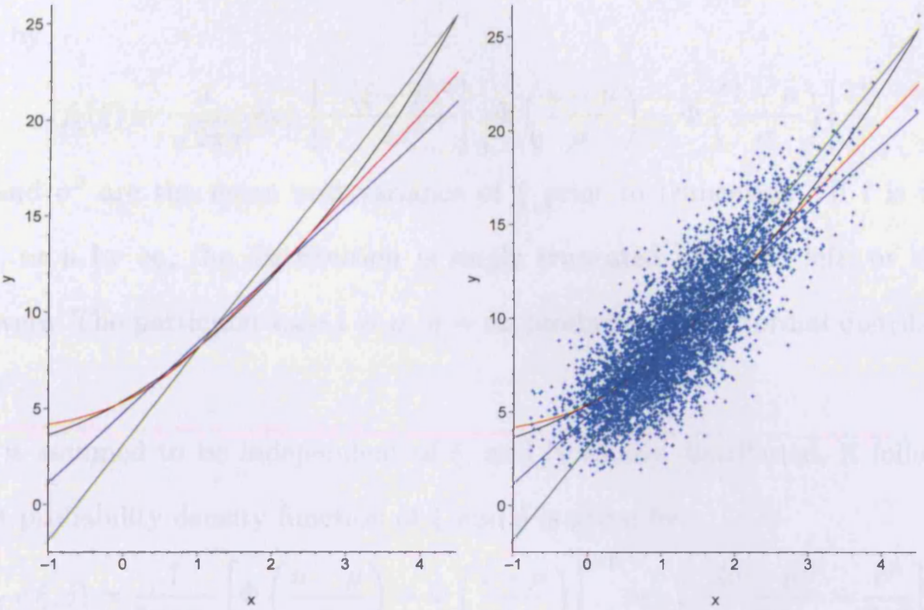
and the central moments of the Normal distribution needed for the approximation are

$$\begin{aligned}
\mu_{\delta 1} &= 0 \\
\mu_{\delta 2} &= \sigma_\delta^2 \\
\mu_{\delta 3} &= 0 \\
\mu_{\delta 4} &= 3\sigma_\delta^4.
\end{aligned}$$

For the example considered earlier, the parameter settings were $k = 2$, $\sigma_\delta = 0.4, \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$. For these settings, the approximation $Q(x)$ is computed as

$$Q(x) = \sqrt{\frac{\pi}{2}} + 0.72427\left(x - \sqrt{\frac{\pi}{2}}\right) + 0.002653\left[\left(x - \sqrt{\frac{\pi}{2}}\right)^2 - 0.58920\right].$$

Figure 6.4 shows a plot of the least squares line (blue), the errors in variables line (green), the exact expression for $E[y|x]$ (red) and the approximation for $E[y|x] = \alpha + \beta\{\sqrt{\frac{\pi}{2}} + 0.698584015(x - \sqrt{\frac{\pi}{2}}) + 0.099151349[(x - \sqrt{\frac{\pi}{2}})^2 - 0.58920]\}$ (black) for a simulated data set of 5000 points. The approximation is indistinguishable from the exact value of $E[y|x]$ at the left part of the data. The approximation deviates from the exact result for $E[y|x]$ most at the right hand tail of the data.



(a) Without data (b) With data

Figure 6.4: Simulated data set with chi ξ and Normal errors, parameter values are $\sigma_\delta = 0.4, \sigma_\varepsilon = 1$, $\alpha = 3$ and $\beta = 5$.

Truncated Normal ξ, Normal δ

An exact result may be derived for $E[y|x]$ with this particular construct of the structural model. In many practical applications, it is unreasonable to assume that the data belong to a distribution with an infinite support. Some data sets might have a natural truncation at one or both ends of the data. For example some experiments may be designed to target a specific range of data and thus there will exist specific cut-off points in the data which may be known in advance. To represent such an application the truncated Normal distribution is chosen to model the distribution of ξ.

Let ξ have a doubly truncated Normal distribution, with a lower truncation point l and upper truncation point u. The probability density function of ξ with support $l \leq \xi \leq u$

is given by

$$f_\xi(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\xi-\mu)^2}{2\sigma^2}\right]\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]^{-1}$$

and μ and σ^2 are the mean and variance of ξ prior to truncation. If l is replaced by $-\infty$, or u by ∞, the distribution is singly truncated from the left, or the right respectively. The particular case $l = \mu$, $u = \infty$ produces a half-Normal distribution.

Since δ is assumed to be independent of ξ, and Normally distributed, it follows that the joint probability density function of ξ and δ is given by

$$f_{\xi,\delta}(\xi,\delta) = \frac{1}{2\pi\sigma\sigma_\delta}\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]^{-1} \exp\left[-\frac{(\xi-\mu)^2}{2\sigma^2} - \frac{\delta^2}{2\sigma_\delta^2}\right].$$

Making the one to one transformation

$$\begin{aligned} x &= \xi + \delta \\ \xi &= \xi \end{aligned}$$

yields the joint probability density function of ξ and x

$$f_{\xi,x}(\xi,x) = \frac{1}{2\pi\sigma\sigma_\delta}\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]^{-1} \exp\left[-\frac{(\xi-\mu)^2}{2\sigma^2} - \frac{(x-\xi)^2}{2\sigma_\delta^2}\right].$$

Letting

$$\begin{aligned} D &= \frac{1}{\sigma^2} + \frac{1}{\sigma_\delta^2} \\ E &= \frac{\mu}{\sigma^2} + \frac{x}{\sigma_\delta^2} \\ F &= \frac{\mu^2}{\sigma^2} + \frac{x^2}{\sigma_\delta^2} \end{aligned}$$

then by completing the square, the marginal probability density function of $f_x(x)$ is given by

$$\begin{aligned} f_x(x) &= \frac{1}{\sqrt{2\pi}\sigma\sigma_\delta}\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]^{-1} \exp\left[-\frac{1}{2}\left(F - \frac{E^2}{D}\right)\right] \\ &\quad \times \int_l^u \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{D}{2}\left(\xi - \frac{E}{D}\right)^2\right] d\xi. \end{aligned}$$

Since

$$\int_l^u \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{D}{2}\left(\xi - \frac{E}{D}\right)^2\right] d\xi = \frac{1}{\sqrt{D}}\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]$$

then the probability density function of x simplifies to

$$\begin{aligned} f_x(x) &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+\sigma_\delta^2}}\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]^{-1} \exp\left[-\frac{(x-\mu)^2}{2(\sigma^2+\sigma_\delta^2)}\right] \\ &\quad \times \left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right] \end{aligned}$$

As the probability density functions $f_{\xi,x}(\xi, x)$ and $f_x(x)$ have been found, the conditional probability function $f_{\xi|x}(\xi|x) = \frac{f_{\xi,x}(\xi,x)}{f_x(x)}$ may be derived.

After some algebraic simplification

$$\begin{aligned} f_{\xi|x}(\xi|x) &= \frac{\sqrt{\sigma+\sigma_\delta^2}}{\sqrt{2\pi}\sigma\sigma_\delta} \exp\left[-\frac{1}{2}\left\{\frac{(\xi-\mu)^2}{\sigma^2} + \frac{(x-\xi)^2}{\sigma_\delta^2} - \frac{(x-\mu)^2}{\sigma^2+\sigma_\delta^2}\right\}\right] \\ &\quad \times \left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1} \end{aligned}$$

and it so follows that

$$E[\xi|x] = \sqrt{\frac{\sigma^2+\sigma_\delta^2}{\sigma^2\sigma_\delta^2}}\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1} \int_l^u \frac{\xi}{\sqrt{2\pi}} \exp\left[-\frac{D}{2}\left(\xi - \frac{E}{D}\right)^2\right] d\xi.$$

This expression may be simplified further since

$$\begin{aligned} \int_l^u \frac{\xi}{\sqrt{2\pi}} \exp\left[-\frac{D}{2}\left(\xi - \frac{E}{D}\right)^2\right] d\xi &= -\frac{1}{D}\left[\phi\left(\frac{Du-E}{\sqrt{D}}\right) - \phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right] \\ &\quad + \frac{E}{D^{(3/2)}}\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]. \end{aligned}$$

and so

$$E[\xi|x] = \frac{\mu\sigma_\delta^2 + x\sigma^2}{\sigma^2+\sigma_\delta^2} - \left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1}\left[\phi\left(\frac{Du-E}{\sqrt{D}}\right) - \phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right].$$

This may be substituted into $E[y|x] = \alpha + \beta E[\xi|x]$ to obtain an expression for $E[y|x]$.

After some simplification then

$$
\begin{aligned}
E[y|x] &= \alpha + \beta\mu + \frac{\beta\sigma^2}{\sigma^2+\sigma_\delta^2}(x-\mu) \\
&\quad -\beta\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1}\left[\phi\left(\frac{Du-E}{\sqrt{D}}\right) - \phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right].
\end{aligned}
$$

This expression is very similar to (6.2), the conditional expectation $E[y|x]$ for the Normal structural model. The difference is the inclusion of the term

$$
-\beta\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1}\left[\phi\left(\frac{Du-E}{\sqrt{D}}\right) - \phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right] \qquad (6.8)
$$

which accounts for the truncation in the data. This term is linked to what is known as the inverse Mills ratio (see for example Tobin [104]). The inverse Mills ratio is the ratio of the probability density function to the cumulative distribution function of the standard Normal distribution.

Additionally,

$$
\frac{Du-E}{\sqrt{D}} = \sqrt{D}\left(u - \frac{E}{D}\right)
$$

where

$$
\frac{E}{D} = \frac{\mu\sigma_\delta^2 + x\sigma^2}{\sigma^2+\sigma_\delta^2} = \mu + \frac{\sigma^2}{\sigma^2+\sigma_\delta^2}(x-\mu).
$$

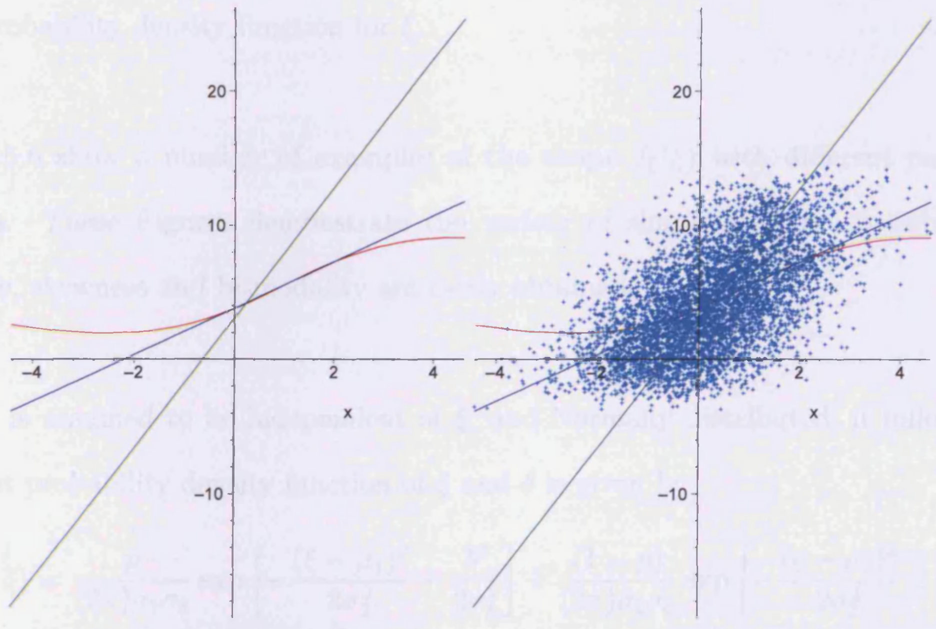
This is the exact result for $E[\xi|x]$ for the Normal structural model that was used by Gleser [51]. The work of Gleser is discussed in more detail later in this Chapter.

By performing a similar analysis that was carried out for the likelihood function of with Normal errors and uniform ξ, the term

$$
-\beta\left[\Phi\left(\frac{Du-E}{\sqrt{D}}\right) - \Phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]^{-1}\left[\phi\left(\frac{Du-E}{\sqrt{D}}\right) - \phi\left(\frac{Dl-E}{\sqrt{D}}\right)\right]
$$

has a bigger impact in the tails of the data. This would make $E[y|x]$ deviate away from the least squares line outside of the support $l \leq \xi \leq u$.

Figure 6.5 shows a plot of the least squares line (blue), the errors in variables line (green) and the exact expression for $E[y|x]$ (red) for a simulated data set of 5000 points. Due to the large σ_δ there is a clear discrepancy between the errors in variables fit and the least squares line. The effect of the measurement error makes the unbiased errors in variables fit seem incorrect for the data shown. The curvature in $E[y|x]$ is demonstrated again, and the extreme tail effects beyond the support of the ξ are readily seen. Between the truncation points $l = -1$ and $u = 2$ the exact result is indistinguishable from the least squares line.



(a) Without data (b) With data

Figure 6.5: $\mu = 0$, $\sigma = 1$, $l = -1$, $u = 2$, $\alpha = 3$, $\beta = 5$ and $\sigma_\delta = \sigma_\varepsilon = 1$

Mixture of two Normals ξ, Normal δ

The mixture of two or more univariate Normal probability density functions results in a probability density function that may be readily manipulated to form a variety of shapes. Since this is an example where the derivation of $E[y|x]$ is tractable, it is included here.

If ξ is assumed to follow a distribution that can be represented by a mixture of two Normal distributions then its probability density function $f_\xi(\xi)$ may be written

$$f_\xi(\xi) = p\phi_1\left(\frac{(\xi - \mu_1)}{\sigma_1}\right) + (1-p)\phi_2\left(\frac{(\xi - \mu_2)}{\sigma_2}\right)$$

where $\phi_i(\xi)$ is the standard Normal probability density function. The parameter p is known as the mixing parameter, and is constrained such that $0 \leq p \leq 1$ to ensure a valid probability density function for ξ.

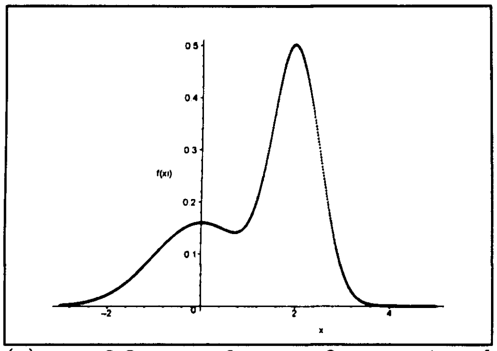
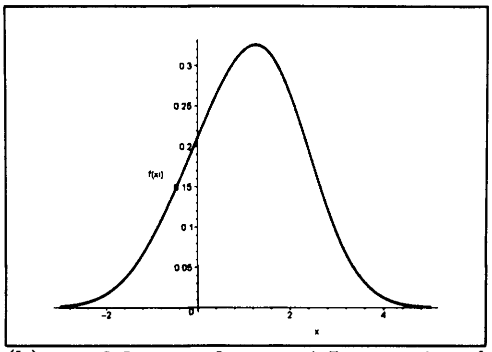
Figure 6.6 show a number of examples of the shape $f_\xi(\xi)$ with different parameter settings. These Figures demonstrate the variety of shapes that are possible. For example, skewness and bi-modality are easily obtained.

Since δ is assumed to be independent of ξ, and Normally distributed, it follows that the joint probability density function of ξ and δ is given by

$$f_{\xi,\delta}(\xi,\delta) = \frac{p}{(2\pi)\sigma_1\sigma_\delta}\exp\left[-\frac{(\xi-\mu_1)^2}{2\sigma_1^2} - \frac{\delta^2}{2\sigma_\delta^2}\right] + \frac{(1-p)}{(2\pi)\sigma_2\sigma_\delta}\exp\left[-\frac{(\xi-\mu_2)^2}{2\sigma_2^2} - \frac{\delta^2}{2\sigma_\delta^2}\right].$$

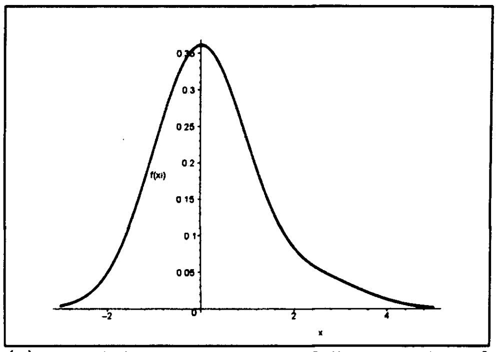
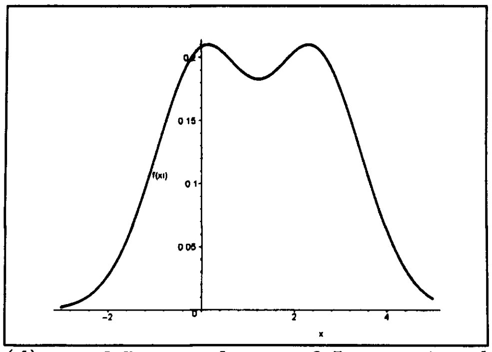
Making the one to one transformation

$$\begin{aligned} x &= \xi + \delta \\ \xi &= \xi \end{aligned}$$

(a) $p = 0.3$, $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1 = 1$ and $\sigma_2 = 0.5$

(b) $p = 0.3$, $\mu_1 = 0$, $\mu_2 = 1.5$, $\sigma_1 = 1$ and $\sigma_2 = 1$

(c) $p = 0.9$, $\mu_1 = 0$, $\mu_2 = 2.5$, $\sigma_1 = 1$ and $\sigma_2 = 1$

(d) $p = 0.5$, $\mu_1 = 0$, $\mu_2 = 2.5$, $\sigma_1 = 1$ and $\sigma_2 = 1$

Figure 6.6: Examples of mixtures of two Normal distributions.

yields the joint probability density function of ξ and x

$$f_{\xi,x}(\xi, x) = \frac{p}{(2\pi)\sigma_1\sigma_\delta} \exp\left[-\frac{(\xi-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\xi)^2}{2\sigma_\delta^2}\right] + \frac{(1-p)}{(2\pi)\sigma_2\sigma_\delta} \exp\left[-\frac{(\xi-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\xi)^2}{2\sigma_\delta^2}\right]$$

Letting

$$\begin{aligned} D_i &= \frac{1}{\sigma_i^2} + \frac{1}{\sigma_\delta^2} \\ E_i &= \frac{\mu_i}{\sigma_i^2} + \frac{x}{\sigma_\delta^2} \\ F_i &= \frac{\mu_i^2}{\sigma_i^2} + \frac{x^2}{\sigma_\delta^2} \end{aligned}$$

for $i = 1, 2$, then the joint probability density function of ξ and x may be written

$$\begin{aligned} f_{\xi,x}(\xi, x) &= \frac{p}{\sqrt{2\pi}\sigma_1\sigma_\delta} \exp\left[-\frac{1}{2}\left(F_1 - \frac{E_1^2}{D_1}\right)\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{D_1}{2}\left(\xi - \frac{E_1}{D_1}\right)^2\right] \\ &\quad + \frac{(1-p)}{\sqrt{2\pi}\sigma_2\sigma_\delta} \exp\left[-\frac{1}{2}\left(F_2 - \frac{E_2^2}{D_2}\right)\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{D_2}{2}\left(\xi - \frac{E_2}{D_2}\right)^2\right]. \end{aligned}$$

Since

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{D_i}{2}\left(\xi - \frac{E_i}{D_i}\right)^2\right] d\xi = \frac{1}{\sqrt{D_i}}$$

then the probability density function of x is

$$f_x(x) = \frac{p}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_1)^2}{2(\sigma_1^2+\sigma_\delta^2)}\right] + \frac{(1-p)}{\sqrt{2\pi}\sqrt{\sigma_2^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_2)^2}{2(\sigma_2^2+\sigma_\delta^2)}\right]$$

The conditional probability density function $f_{\xi|x}(\xi|x)$ is defined as

$$f_{\xi|x}(\xi|x) = \frac{f_{\xi,x}(\xi,x)}{f_x(x)}$$

and since

$$\int_{-\infty}^{\infty} \frac{\xi}{\sqrt{2\pi}} \exp\left[-\frac{D_i}{2}\left(\xi - \frac{E_i}{D_i}\right)^2\right] d\xi = \frac{E_i}{D_i^{(3/2)}}$$

then

$$E[\xi|x] = \frac{\frac{p}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_1)^2}{2(\sigma_1^2+\sigma_\delta^2)}\right] \frac{(x\sigma_1^2+\mu_1\sigma_\delta^2)}{(\sigma_1^2+\sigma_\delta^2)} + \frac{(1-p)}{\sqrt{2\pi}\sqrt{\sigma_2^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_2)^2}{2(\sigma_2^2+\sigma_\delta^2)}\right] \frac{(x\sigma_2^2+\mu_2\sigma_\delta^2)}{(\sigma_2^2+\sigma_\delta^2)}}{\frac{p}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_1)^2}{2(\sigma_1^2+\sigma_\delta^2)}\right] + \frac{(1-p)}{\sqrt{2\pi}\sqrt{\sigma_2^2+\sigma_\delta^2}} \exp\left[-\frac{(x-\mu_2)^2}{2(\sigma_2^2+\sigma_\delta^2)}\right]}.$$

Letting

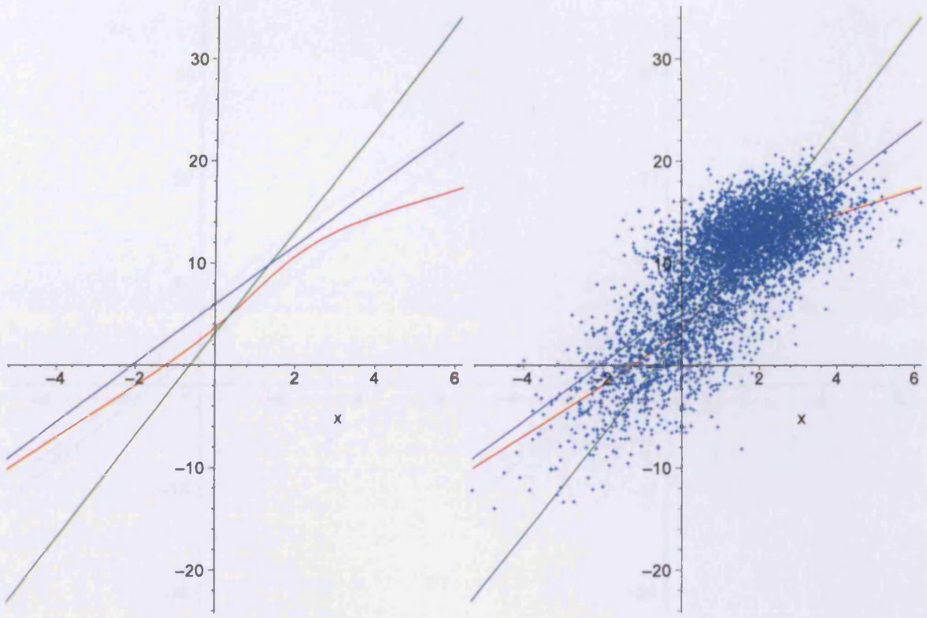
$$\begin{aligned} m_1 &= \frac{p}{(1-p)}\phi\left(\frac{x-\mu_2}{\sqrt{\sigma_2^2+\sigma_\delta^2}}\right)\phi\left(\frac{x-\mu_1}{\sqrt{\sigma_1^2+\sigma_\delta^2}}\right)^{-1} \\ m_2 &= \frac{(1-p)}{p}\phi\left(\frac{x-\mu_1}{\sqrt{\sigma_1^2+\sigma_\delta^2}}\right)\phi\left(\frac{x-\mu_2}{\sqrt{\sigma_2^2+\sigma_\delta^2}}\right)^{-1} \end{aligned}$$

then finally

$$\begin{aligned} E[y|x] &= \alpha + \frac{\beta}{(1+m_1)}\frac{(x\sigma_1^2+\mu_1\sigma_\delta^2)}{(\sigma_1^2+\sigma_\delta^2)} + \frac{\beta}{(1+m_2)}\frac{(x\sigma_2^2+\mu_2\sigma_\delta^2)}{(\sigma_2^2+\sigma_\delta^2)} \\ &= \alpha + \frac{\beta}{(1+m_1)}\mu_1 + \frac{\beta\sigma_1^2}{\sigma_1^2+\sigma_\delta^2}\frac{(x-\mu_1)}{(1+m_1)} + \frac{\beta}{(1+m_2)}\mu_2 + \frac{\beta\sigma_2^2}{\sigma_2^2+\sigma_\delta^2}\frac{(x-\mu_2)}{(1+m_2)} \end{aligned}$$

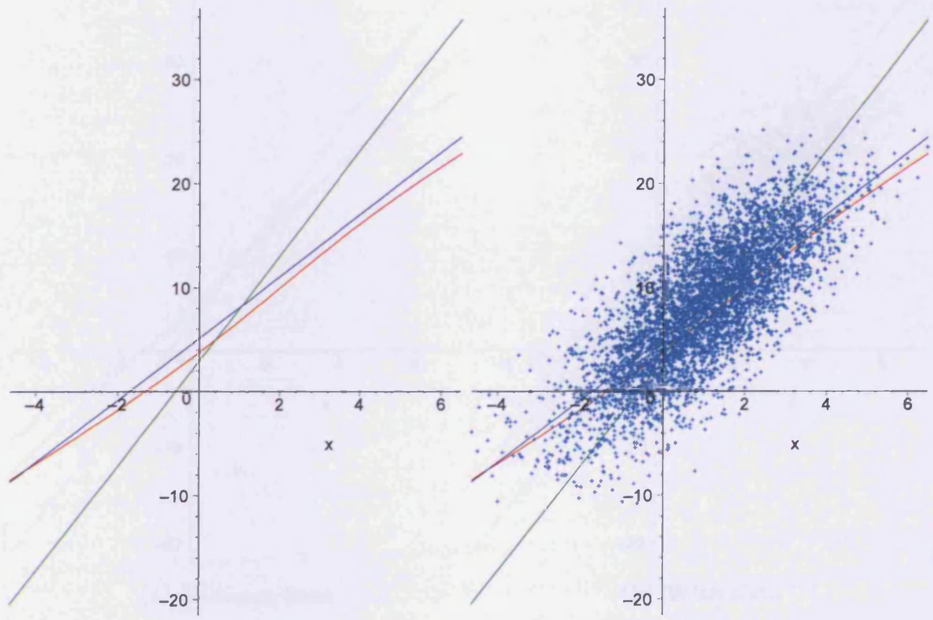
This result can be considered a weighting of the two least squares regressions fitted to each individual mixture distribution. If the mixture distribution is extended to a mixture of M Normal distributions then $E[y|x]$ would become the weighting of the M least squares regressions fitted to each of the M individual mixture distributions.

As an example of this result, the parameter settings used in Figure 6.6 will be used to form example data sets. Figures 6.7 to 6.10 shows the plots of the least squares line (blue), the errors in variables line (green) and $E[y|x]$ (red) for a simulated data set of 5000 points under the parameter settings of Figure 6.6. $E[y|x]$ is obtained by a smoothing of the least squares fit to each mixture component.



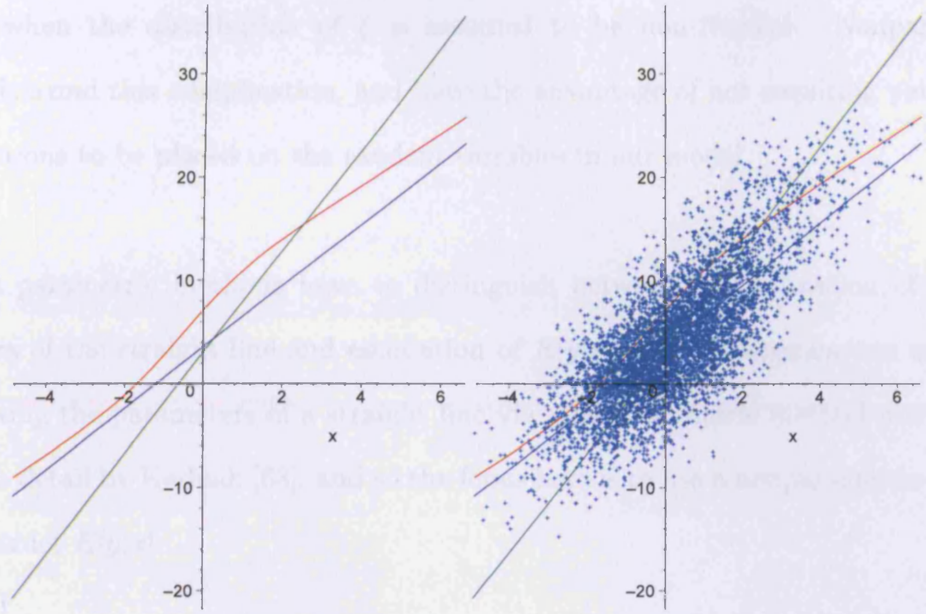
(a) Without data (b) With data

Figure 6.7: $p = 0.3$, $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $\alpha = 3$, $\beta = 5$ and $\sigma_\delta = \sigma_\varepsilon = 1$



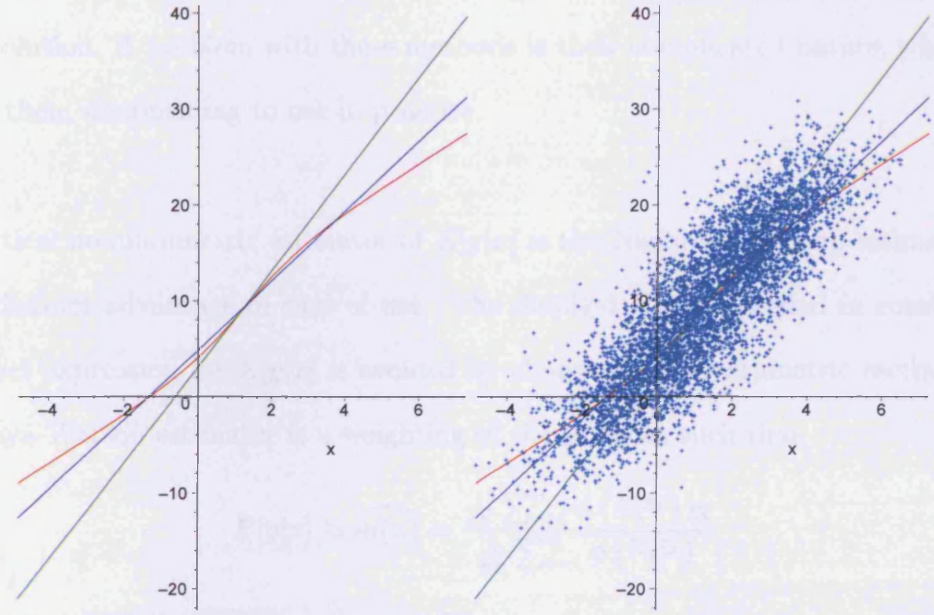
(a) Without data (b) With data

Figure 6.8: $p = 0.3$, $\mu_1 = 0$, $\mu_2 = 1.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\alpha = 3$, $\beta = 5$ and $\sigma_\delta = \sigma_\varepsilon = 1$



(a) Without data (b) With data

Figure 6.9: $p = 0.9$, $\mu_1 = 0$, $\mu_2 = 2.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\alpha = 3$, $\beta = 5$ and $\sigma_\delta = \sigma_\varepsilon = 1$



(a) Without data (b) With data

Figure 6.10: $p = 0.5$, $\mu_1 = 0$, $\mu_2 = 2.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\alpha = 3$, $\beta = 5$ and $\sigma_\delta = \sigma_\varepsilon = 1$

6.2.2 Nonparametric Approach

The previous section has used a parametric approach in deriving explicit and approximate expressions for $E[y|x]$. The algebra behind the derivations is, as seen, sometimes messy when the distribution of ξ is assumed to be non-Normal. Nonparametric methods avoid this complication, and have the advantage of not requiring parametric assumptions to be placed on the random variables in our model.

Just as parametric methods have to distinguish between the estimation of the parameters of the straight line and estimation of $E[y|x]$, so do nonparametric methods. Estimating the parameters of a straight line via a nonparametric method was investigated in detail by Koduah [68], and so the focus here is to use a nonparametric method to construct $E[y|x]$.

Estimating $E[y|x]$ Various algorithms for estimating $E[y|x]$ were discussed by Carroll et al. [15]. These included methods using splines, likelihood methods and deconvolution. A problem with these methods is their complicated nature, which may render them unappealing to use in practise.

A practical nonparametric estimator of $E[y|x]$ is the Nadaraya-Watson estimator and has a distinct advantage of ease of use. The detailed algebra needed in constructing the exact expression for $E[y|x]$ is avoided by choosing a nonparametric method. The Nadaraya-Watson estimator is a weighting of the y values such that

$$E[y|x] \approx \widehat{m(x)} = \frac{\frac{1}{nh}\sum_{i=1}^{n} \phi\left(\frac{x-x_i}{h}\right) y_i}{\frac{1}{nh}\sum_{i=1}^{n} \phi\left(\frac{x-x_i}{h}\right)} \tag{6.9}$$

where h is some suitably chosen bandwidth.

Choice of the bandwidth h, will roughly speaking, alter the apparent smoothness of the fit. Stefanski [100] has shown that for a large number of kernel functions and error distributions, the value of h which minimises mean square error is

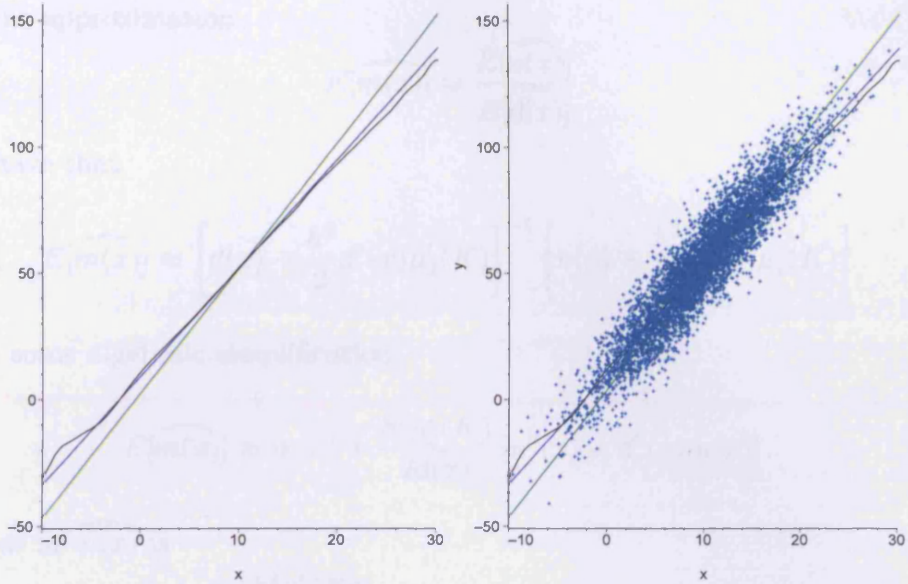
$$h = \sigma_\delta \left[\log(n)\right]^{-\frac{1}{2}}. \tag{6.10}$$

This will be used for some illustrative examples later in this Chapter.

The Nadaraya-Watson estimator (6.9) can be interpreted in a number of ways. Firstly, since the denominator $\frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)$ is the kernel density estimator of the probability density function $f_x(x)$ and the numerator $\frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)y_i$ is the kernel density estimator of $\int y f_{x,y}(x,y)\,dy$ then upon taking the ratio we obtain a kernel density estimator of $E[y|x] = \int y f_{y|x}(y|x)\,dy$. Secondly, and more simply, the Nadaraya-Watson fit may be viewed as a locally weighted average of the observed y values.

As an example, Figure 6.11 shows a simulated data set for the Normal structural model. The parameter settings chosen were $\alpha = 3$, $\beta = 5$, $\mu = 10$, $\sigma = 5$, $\sigma_\delta = 2$, $\sigma_\varepsilon = 1$ and $n = 5000$. Lindley [72] stated that in this case $E[y|x]$ follows the least squares line. Indeed it may be seen that the Nadaraya-Watson estimator closely follows the least squares line over the range of the data. For this example, h was chosen in accordance with equation (6.10).

A problem with the Nadaraya-Watson estimator lies in its limiting properties. As h tends to infinity, then the Nadaraya-Watson estimator tends to $\bar{y}$. This may not be an ideal limiting form. However h should never be chosen so large that this is a concern. In addition, the Nadaraya-Watson estimator tends to behave erratically towards the tails of the data. This is a common trait possessed by most nonparametric methods. In order to produce reliable estimates, nonparametric methods tend to



(a) Without data (b) With data

Figure 6.11: Least squares line (blue), errors in variables line (green) with Nadaraya-Watson estimate (black) for a simulated Normal structural data set.

require a larger sample than their parametric equivalents. Some more properties of the Nadaraya-Watson estimator shall be reported here.

Letting $\widehat{n(x)}$ denote the numerator of (6.9) and $\widehat{d(x)}$ denote the denominator of (6.9) then by expanding in a Taylor series we have (see for example Di Marzio and Taylor [34])

$$E[\widehat{n(x)}] \approx n(x) + \frac{h^2}{2} n''(x)\mu_2'(K)$$

and

$$E[\widehat{d(x)}] \approx d(x) + \frac{h^2}{2} d''(x)\mu_2'(K)$$

where $\mu_2'(K) = \int t^2 \phi(t)\, dt$.

Making the approximation

$$E[\widehat{m(x)}] \approx \frac{E[\widehat{n(x)}]}{E[\widehat{d(x)}]}$$

then we have that

$$E[\widehat{m(x)}] \approx \left[d(x) + \frac{h^2}{2}d''(x)\mu_2'(K)\right]^{-1} \left[n(x) + \frac{h^2}{2}n''(x)\mu_2'(K)\right]$$

and after some algebraic simplification

$$E[\widehat{m(x)}] \approx m(x) + \frac{h^2\mu_2'(K)}{2d(x)}\left[n''(x) - d''(x)m(x)\right].$$

So the bias in $\widehat{m(x)}$ is

$$\frac{h^2\mu_2'(K)}{2d(x)}\left[n''(x) - d''(x)m(x)\right].$$

However, since $n(x) = m(x)d(x)$ then by Leibnitz' rule

$$n''(x) = m''(x)d(x) + 2m'(x)d'(x) + m(x)d''(x)$$

and the bias may be written as

$$\frac{h^2\mu_2'(K)}{2}\left[m''(x) + \frac{2m'(x)d'(x)}{d(x)}\right],$$

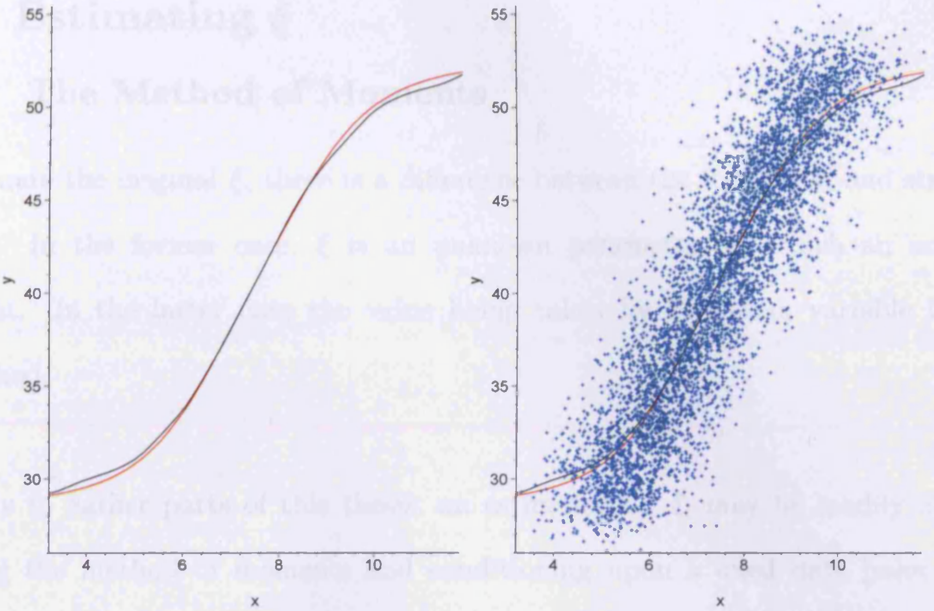
which is of a more interpretable form.

So the bias is dominated by the second derivative $m''(x)$ (close to a turning point) or by the first derivative $m'(x)$ when there are few observations. This point is of particular interest for the Normal structural model. As has been stated throughout this Chapter, Lindley [72] proved that $E[y|x]$ is a straight line if and only if the Normal structural model is assumed. $E[y|x]$ takes a much more complicated form for any other structural model. Hence for the Normal structural model, $m''(x) = 0$, and thus the bias for the Normal structural model is therefore smaller than for any other structural model. This is however the result that $E[y|x]$ follows the standard least squares line, and so nonparametric methods are not needed for the Normal structural model.

As further examples of the Nadaraya-Watson estimator, Figures 6.12 and 6.13 contain the plots of the exact $E[y|x]$ and the estimated $E[y|x]$ by the Nadaraya-Watson method when ξ follows a uniform and chi distribution respectively with Normal errors. When ξ is taken to follow a uniform distribution, the Nadaraya-Watson estimator closely follows the exact result for the main body of the distribution, but deviates perceptibly in the tails. However, the Nadaraya-Watson estimator only deviates greatly from the exact result where data are sparse.

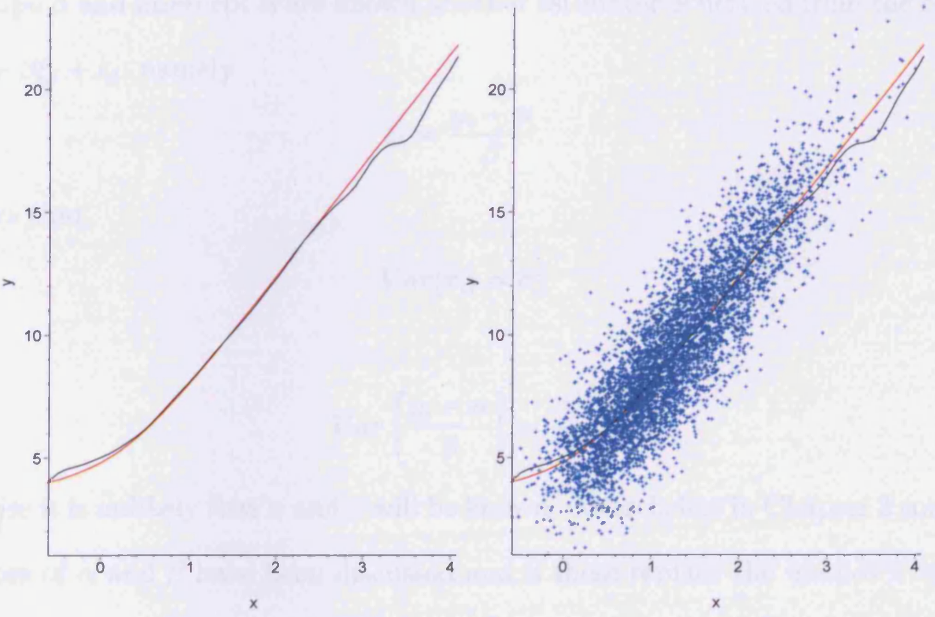
When ξ is taken to be follow a chi distribution with two degrees of freedom there appears to be a closer resemblance between the Nadaraya-Watson estimator and the exact result, but there does remain some deviation in the tails, particularly in the right hand tail where the data are sparse.

These examples illustrate the fact that the Nadaraya-Watson estimator is a serviceable method of approximating to $E[y|x]$ over the range of the support of ξ when data are plentiful. However there are noticeable discrepancies in the tails where data are sparse. In practical applications where the exact distribution of ξ is unlikely to be known, it is clear that the Nadaraya-Watson estimator has much to commend it, but there are difficulties in establishing $E[y|x]$ in the tails.



(a) Without data (b) With data

Figure 6.12: Nadaraya-Watson estimate (black) and exact result (red) for a simulated data set with uniform ξ and Normal errors. Parameter settings are $a = 5$, $b = 10$, $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 0.7$, $\sigma_\varepsilon = 1$ and $n = 5000$.



(a) Without data (b) With data

Figure 6.13: Nadaraya-Watson estimate (black) and exact result (red) for a simulated data set with chi ξ (2 degrees of freedom) and Normal errors. Parameter settings are $\alpha = 3$, $\beta = 5$, $\sigma_\delta = 0.3$, $\sigma_\varepsilon = 1$ and $n = 5000$.

6.3 Estimating ξ

6.3.1 The Method of Moments

To estimate the original ξ, there is a difference between the functional and structural models. In the former case, ξ is an unknown parameter for which an estimator is sought. In the latter case the value being taken by a random variable is being established.

Similarly to earlier parts of this thesis, an estimator for ξ_i may be readily obtained by using the method of moments and conditioning upon a fixed data point. Since $x_i = \xi_i + \delta_i$, a naive method of moment estimators for ξ_i is

$$\tilde{\xi}_i = x_i. \tag{6.11}$$

If the slope β and intercept α are known another estimator is derived from the equation $y_i = \alpha + \beta\xi_i + \varepsilon_i$, namely

$$\tilde{\xi}_i = \frac{y_i - \alpha}{\beta}. \tag{6.12}$$

It follows that

$$Var[x_i] = \sigma_\delta^2$$

and

$$Var\left[\frac{y_i - \alpha}{\beta}\right] = \frac{\sigma_\varepsilon^2}{\beta^2}.$$

In practise it is unlikely that α and β will be known, nevertheless in Chapter 3 consistent estimators of α and β have been discussed and if these replace the unknown values in equation (6.12) a consistent estimator of ξ_i is obtained. Ignoring the variances of these estimators the optimal linear combination (in terms of minimum variance) of equations

(6.11) and (6.12) provides the following approximately unbiased estimator for ξ_i

$$\tilde{\xi}_i = \frac{\lambda}{\lambda + \tilde{\beta}^2} x_i + \frac{\tilde{\beta}}{\lambda + \tilde{\beta}^2}(y_i - \tilde{\alpha}). \tag{6.13}$$

This is the same as the maximum likelihood estimator in the functional model derived by Kendall and Stuart [67] when the ratio of the error variances λ was assumed known. This also was the estimator derived in Chapter 5 via maximum likelihood assuming the Normal functional model.

The approach of using the method of moments can also be used to derive an estimator for η_i. Since $y_i = \eta_i + \varepsilon_i$ and therefore $y_i = \alpha + \beta x_i + (\varepsilon_i - \beta\delta_i)$, two naive method of moment estimators for η_i are

$$\tilde{\eta}_i = y_i \tag{6.14}$$

and, again assuming that α and β are known,

$$\tilde{\eta}_i = \alpha + \beta x_i \tag{6.15}$$

Again, we can find the variances of expressions (6.14) and (6.15)

$$Var[y_i] = \sigma_\varepsilon^2$$

$$Var[\alpha + \beta x_i] = \beta^2\sigma_\delta^2.$$

Thus the optimal linear combination (in terms of minimum variance) of equations (6.14) and (6.15) provides the following approximately unbiased estimator for η_i

$$\tilde{\eta}_i = \frac{\tilde{\beta}^2}{\lambda + \tilde{\beta}^2} y_i + \frac{\lambda}{\lambda + \tilde{\beta}^2}(\tilde{\alpha} + \tilde{\beta} x_i). \tag{6.16}$$

There is some symmetry in the prediction of ξ_i and η_i in that

$$\tilde{\alpha} + \tilde{\beta}\tilde{\xi}_i = \tilde{\alpha} + \frac{\tilde{\beta}\lambda}{\lambda + \tilde{\beta}^2} x_i + \frac{\tilde{\beta}^2}{\lambda + \tilde{\beta}^2} y_i - \frac{\tilde{\beta}^2}{\lambda + \tilde{\beta}^2}\tilde{\alpha} = \frac{\tilde{\beta}^2}{\lambda + \tilde{\beta}^2} y_i + \frac{\lambda}{\lambda + \tilde{\beta}^2}(\tilde{\alpha} + \tilde{\beta} x_i) = \tilde{\eta}_i.$$

This relationship might be expected as the errors in variables model defines $\eta = \alpha + \beta\xi$, however this shows that the naive estimator for η_i (6.16) is not a new estimator different from (6.13), but is merely the point on the true line corresponding to the estimator (6.13) for ξ.

For geometric mean regression discussed in Chapter 2, the assumption made is that $\lambda = \beta^2$. Under this assumption,

$$\begin{aligned} \tilde{\xi}_i &= \frac{1}{2}\left(x_i + \frac{(y_i - \tilde{\alpha})}{\tilde{\beta}}\right) \\ \tilde{\eta}_i &= \frac{1}{2}\left(y_i + \tilde{\alpha} + \tilde{\beta}x_i\right) \end{aligned}$$

so effectively, equal weighting of 1/2 is given to each of the naive moment equations (6.11), (6.12), (6.14) and (6.15). In addition, y on x and x on y simple linear regression are seen as extremes. Consider the factor $\lambda/(\lambda + \beta^2)$. It can be seen that $min\left[\frac{\lambda}{\lambda+\beta^2}\right] = 0$ if and only if $\lambda = 0$ then $\tilde{\eta}_i = y_i$ and $\tilde{\xi}_i = \frac{y_i - \tilde{\alpha}}{\tilde{\beta}}$. Thus if minimum weight is given to x_i this implies that x on y regression is the appropriate tool to use. Also $max\left[\frac{\lambda}{\lambda+\beta^2}\right] = 1$ if and only if $\lambda = \infty$ then $\tilde{\xi}_i = x_i$ and $\tilde{\eta}_i = \tilde{\alpha} + \tilde{\beta}x_i$. This implies that y on x regression is the appropriate tool to use where maximum weight is given to x_i.

In addition, it can be seen from the equation $y_i = \alpha + \beta x_i + (\varepsilon_i - \beta\delta_i)$ that the estimators for ξ_i and η_i given by (6.13) and (6.16) may be written as

$$\begin{aligned} \tilde{\xi}_i &= x_i + \frac{\tilde{\beta}}{\lambda + \tilde{\beta}^2}(\varepsilon_i - \tilde{\beta}\delta_i) = x_i + \frac{\tilde{\beta}}{\lambda + \tilde{\beta}^2}(y_i - \tilde{\alpha} - \tilde{\beta}x_i) \qquad (6.17) \\ \tilde{\eta}_i &= y_i - \frac{\lambda}{\lambda + \tilde{\beta}^2}(\varepsilon_i - \tilde{\beta}\delta_i) = y_i - \frac{\lambda}{\lambda + \tilde{\beta}^2}(y_i - \tilde{\alpha} - \tilde{\beta}x_i). \qquad (6.18) \end{aligned}$$

If the latent data set $\{(\xi_i, \eta_i), i = 1, \ldots, n\}$ has been estimated, it is straightforward to obtain estimates for the values taken by the random variables δ and ε respectively

$$\begin{aligned}\tilde{\delta}_i = x_i - \tilde{\xi}_i &= \frac{-\tilde{\beta}}{\lambda + \tilde{\beta}^2}(y_i - \tilde{\alpha} - \tilde{\beta}x_i) \\ \tilde{\varepsilon}_i = y_i - \tilde{\eta}_i &= \frac{\lambda}{\lambda + \tilde{\beta}^2}(y_i - \tilde{\alpha} - \tilde{\beta}x_i),\end{aligned}$$

and these are the exact terms which appear in the estimating equations (6.17) and (6.18). So it is seen that these method of moments estimators for the latent data set are the observed data set adjusted by the estimated δ and ε. Equations (6.17) and (6.18) can also describe how these estimated values behave depending on where the observed data point lies. Assume $\beta > 0$. If $y_i > \alpha + \beta x_i$ (observed y above true line at observed x) then $\tilde{\xi}_i > x_i$ but $\tilde{\eta}_i < y_i$. If $y_i < \alpha + \beta x_i$ (observed y below true line at observed x) then $\tilde{\xi}_i < x_i$ but $\tilde{\eta}_i > y_i$. If $\beta < 0$, then the obvious alterations are made to the above inequalities. The magnitude of the difference between $\tilde{\xi}_i$ and x_i is greater for those observations most distant from the true line. If $y_i = \alpha + \beta x_i$ then $\tilde{\xi}_i = x_i$. The implications of these statements will be discussed again later in this thesis, particularly in Chapter 7.

The method of moments estimators derived above may also be linked to other slope estimation methods. Consider Figure 6.14. We can find an estimator for the distance $D_i(\lambda)$ by Pythagoras' theorem

$$D_i(\lambda) = \sqrt{\tilde{\delta}_i^{\,2} + \tilde{\varepsilon}_i^{\,2}} = (\sqrt{\lambda^2 + \beta^2})\,\frac{y_i - \alpha - \beta x_i}{\lambda + \beta^2}.$$

The quantity $\frac{D_i(\lambda)}{(\sqrt{\lambda^2+\beta^2})}$ is linked to what has been called orthogonal regression in Chapter 1. For the structural model, the maximum likelihood estimators (and method of moment estimators) when λ is known of α and β are given by

$$(\tilde{\alpha}, \tilde{\beta}) = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{n} \frac{D_i(\lambda)}{(\sqrt{\lambda^2 + \beta^2})}.$$

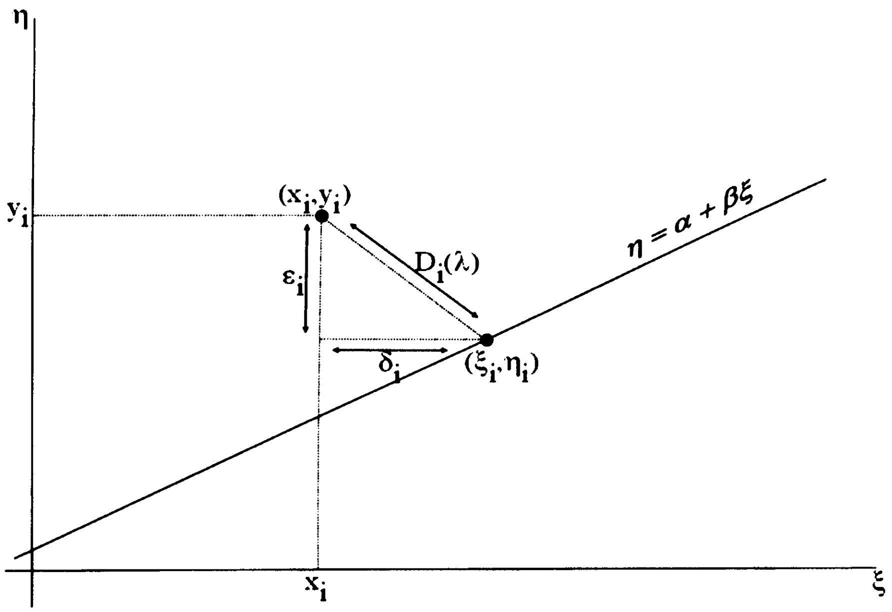


Figure 6.14: Linking to other estimation procedures

The factor $\frac{1}{\sqrt{\lambda+\beta^2}}$ in $D_i(\lambda)$ is a weighting factor which can be varied to give different projections from the data point onto the line. This allows the sum of squares of any projection to be minimised. This is an example of weighted least squares, and has been discussed by many authors, including Lindley [72] and Sprent [96], amongst others. Notice, however, that the weights depend on the slope β, so this is not the form of weighted least squares commonly suggested to allow for heteroscedasticity of the data.

The term $D_i(\lambda)$ is also similar to $\frac{(y_i-\alpha-\beta x_i)}{\sigma_\delta\sqrt{\lambda+\beta^2}}$ which is known as a pivot (see for example Cox [25]). A pivot is defined to be a dimension free function of the data and parameters whose distribution does not depend on any parameters. Pivots are useful in forming hypothesis tests and confidence intervals. Indeed, when σ_δ^2 and σ_ε^2 are known and the data are rescaled so that $\sigma_\delta^2 = \sigma_\varepsilon^2 = 1$, then $\frac{D_i(1)}{\sqrt{1+\beta^2}} \sim \chi_n^2$ if $(\xi, \delta, \varepsilon)$ are considered to be trivariate Normal. This was exploited by Brown [12]. The set

$\left\{(\alpha,\beta)|\frac{D_i(1)}{\sqrt{1+\beta^2}} \le q_{(1-p)}\right\}$ where $P(\chi^2_n \le q_{(1-p)}) = (1-p)$ provides a $(1-p)$ confidence interval for (α,β). Confidence intervals of this form have been discussed further by Okamoto [81] and Cheng and Van Ness [20].

6.3.2 Gleser's Method

The estimation method proposed by Gleser [52] mentioned in Chapter 2 also suggests an estimator for ξ_i. His idea was to first apply a correction to the observed x_i in order to obtain an estimator for ξ_i. Standard regression techniques are then applied to estimate the unknown slope β. The motivation for this is given here. Assuming the Normal structural model,

$$\begin{pmatrix} \xi \\ \delta \\ \varepsilon \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}\right],$$

the joint distribution of (x_i, ξ_i) is

$$\begin{pmatrix} \xi_i \\ x_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_\delta^2 \end{pmatrix}\right].$$

It is seen that the conditional distribution of ξ_i given x_i is

$$\xi_i|x_i \sim N\left(\mu + \frac{\sigma^2}{\sigma^2+\sigma_\delta^2}(x_i-\mu), \frac{\sigma^2\sigma_\delta^2}{(\sigma^2+\sigma_\delta^2)}\right).$$

This suggests another naive estimator for ξ_i given by

$$\tilde{\xi}_i = \tilde{\mu} + \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2+\tilde{\sigma}_\delta^2}(x_i-\tilde{\mu}). \tag{6.19}$$

This is an example of direct shrinkage (see for example Copas [24]). The estimator for ξ_i is taken to be the overall mean μ adjusted for the location of x_i relative to the overall mean, and the multiplicative factor of the reliability ratio. For example,

$$\xi_i - \mu = \frac{\sigma^2}{\sigma^2+\sigma_\delta^2}(x_i-\mu).$$

Thus if $x_i > \mu$ then $(\xi_i - \mu) < (x_i - \mu)$ and if $x_i < \mu$ then $(\xi_i - \mu) > (x_i - \mu)$. Points x_i close to the mean are adjusted to a lesser extent than those further away. The overall effect of Gleser's regression is one of shrinking the data in towards the mean μ.

On the other hand, the method of moments estimator behaves differently, as outlined earlier. The estimator for ξ_i is taken to be x_i adjusted for the location of y_i in terms of the true line evaluated at x_i. In other words, the estimator for ξ_i is x_i adjusted for the term $(\varepsilon_i - \beta\delta_i)$. Those points with large residual are pushed further away from the observed x_i, as opposed to those with smaller residuals. Another distinction, is that the method of moments estimator uses both the observed values (x_i, y_i), whilst the Gleser estimator only uses the x_i.

Equation (6.19) only uses the x_i observation to estimate ξ, as opposed to the (x_i, y_i) pair. Gleser's method however, was primarily intended as a tool to estimate the parameters of the model, and not the latent data. By only using the x_i observation, Gleser's method has the advantage of just relying on the reliability ratio being known. In sociology and psychology it is not unusual to have knowledge of the reliability ratio. The measure is also used in genetics where it is called heritability (Hood [56]). To use the method of moments estimator for ξ_i, all the parameters in the model need to be estimated.

6.3.3 Modified Estimator for ξ based on Gleser's Method

As just stated, Gleser's method is primarily a tool to estimate the parameters of the model. It is good in this sense as it only depends on the reliability ratio. In terms of predicting the latent data set however, it only uses the x_i, through the distribution of $\xi_i|x_i$. It is possible to modify Gleser's method to obtain a possibly more reliable estimator for the latent data set. This can be achieved through the consideration of the distribution of $\xi_i|y_i$.

Consider once again the Normal structural model. That is,

$$\begin{pmatrix} \xi \\ \delta \\ \varepsilon \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}\right],$$

and it follows that

$$\begin{pmatrix} \xi \\ y \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \alpha+\beta\mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2+\sigma_\varepsilon^2 \end{pmatrix}\right].$$

Using standard results concerning multivariate Normal distributions we have

$$E[\xi|y] = \mu + \frac{\beta\sigma^2}{\beta^2\sigma^2+\sigma_\varepsilon^2}(y-\alpha-\beta\mu)$$

and

$$Var[\xi|y] = \frac{\sigma^2\sigma_\varepsilon^2}{\beta^2\sigma^2+\sigma_\varepsilon^2}.$$

This suggests a further naive estimator of ξ,

$$\tilde{\xi}_i = \mu + \frac{\tilde{\beta}\tilde{\sigma}^2}{\tilde{\beta}^2\tilde{\sigma}^2+\tilde{\sigma}_\varepsilon^2}(y_i-\tilde{\alpha}-\tilde{\beta}\mu). \tag{6.20}$$

In a similar manner to the method of moments estimator, we may derive an optimal linear combination of the estimators (6.19) and (6.20) of ξ which has minimum variance amongst all other linear combinations. This estimator would then use both the x and y value analogous to the method of moments estimator.

Letting $v_1 = \frac{\sigma^2\sigma_\delta^2}{\sigma^2+\sigma_\delta^2}$ and $v_2 = \frac{\sigma^2\sigma_\varepsilon^2}{\beta^2\sigma^2+\sigma_\varepsilon^2}$, the optimal linear combination of the two estimators given in the previous paragraph is

$$\tilde{\xi} = \frac{v_2}{v_1+v_2}\left[\tilde{\mu} + \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2+\tilde{\sigma}_\delta^2}(x-\tilde{\mu})\right] + \frac{v_1}{v_1+v_2}\left[\tilde{\mu} + \frac{\tilde{\beta}\tilde{\sigma}^2}{\tilde{\beta}^2\tilde{\sigma}^2+\tilde{\sigma}_\varepsilon^2}(y-\tilde{\alpha}-\tilde{\beta}\tilde{\mu})\right]. \tag{6.21}$$

6.3.4 Comparison Study

A comparison study will now be undertaken to compare estimators of ξ in terms of relative errors. For this study a sample of 100 ξ's were generated from a Normal distribution with $\mu = 10$ and $\sigma = 5$. For each of the 5000 simulations, the same ξ's were used to form 5000 different data sets. The parameter settings chosen were $\alpha = 3$, $\beta = 5$, and $\sigma_\varepsilon = 1$. The method of moments estimator (6.13), Gleser estimator (6.19) and modified Gleser estimator (6.21) were used to predict the ξ's for each simulated data set. Relative errors of the form $\frac{|\tilde{\xi}_i - \xi_i|}{|\xi_i|}$ were computed for each data point in each simulation. This allows the computation of the average relative error in estimating each ξ_i for each of the three estimation methods discussed previously. The following table shows the average relative error for each estimation method for some chosen data points throughout the range of the data. Here G denotes the Gleser estimator (6.19), M denotes the method of moments estimator (6.13) and MG denotes the modified Gleser estimator (6.21). Rank denotes the rank of the particular ξ investigated.

Rank	Method	$\sigma_\delta = 0.5$	$\sigma_\delta = 1$	$\sigma_\delta = 2$
1	G	0.02309	0.19368	0.25194
	M	0.01875	0.00383	0.04555
	MG	0.01943	0.00505	0.04558
10	G	0.05864	0.08102	0.1892
	M	0.00752	0.02102	0.03101
	MG	0.00822	0.02051	0.04011
50	G	0.07524	0.14429	0.21111
	M	0.01193	0.07996	0.08991
	MG	0.01545	0.08466	0.09111
90	G	0.04113	0.22237	0.31189
	M	0.01623	0.04360	0.07388
	MG	0.01680	0.05181	0.07995
100	G	0.01777	0.4197	0.5766
	M	0.02959	0.04239	0.06784
	MG	0.03016	0.04649	0.06894

The initial conclusion is that the Gleser estimator appears to be the worst performing in terms of relative error. As σ_δ grows, this estimator becomes more erratic, and tends to provide highly variable estimates for a large number of data points. The method of moments and modified Gleser estimator however, seem to be robust to the change in σ_δ, and both out-perform the Gleser estimator. Indeed the method of moments and modified Gleser estimator perform similarly well in comparison.

The Gleser estimator however, is primarily a tool used to modify the data so that standard linear regression techniques can be applied. It is not surprising it performs badly here as it only uses the x measurement to estimate the ξ. It does however have the advantage that it only requires knowledge of the reliability ratio, κ to be used.

6.4 Conclusions

This Chapter has considered the multifaceted topic of prediction in an errors in variables model. It is essential that the correct prediction question is answered. Does a practitioner wish to find the average y for a given x?, does a practitioner want to uncover the latent data set $\{(\xi_i, \eta_i), i = 1, \ldots, n\}$?. The topic of prediction for an errors in variables model is not as straightforward as that for standard regression models, and presumably this is one reason why the topic is largely neglected in the literature. For example, there are differences in finding $E[y|x]$ depending on whether a functional or structural model is assumed. There are further differences in the structural model depending on the distribution of ξ. It is hoped that this Chapter has clarified the differences in prediction between models, and offered practical advice as to the prediction of ξ, as well as the prediction of y.

Chapter 7

Residuals

7.1 Introductory Remarks

After a regression model has been fitted, various questions are usually asked. Examples of such questions may be:

1. Is the model fitted the correct model?

2. Are there any outliers?

3. Are the distributional assumptions of the model correct?

These questions are typically answered by some sort of residual analysis.

In simple linear regression, much has been documented on residual analysis. Most textbooks on the subject (Draper and Smith [37], and the references therein) contain detailed and informative sections on residual analysis, as well as providing recommendations for the practitioner. The errors in variables situation is not as well documented. In the two main texts on the topic, Cheng and Van Ness [20] and Fuller [41], there is very little information on residuals. This omission is also apparent in the scientific papers and expositions.

A possible reason for this omission is that in the errors in variables setting there is no explicit definition of a residual. The simple linear regression model has a natural concept of a residual which is not as readily obtained when a random error component is included in the x measurement. The aim of this Chapter is to investigate the concept of a residual for our errors in variables model.

7.2 Vertical Residuals

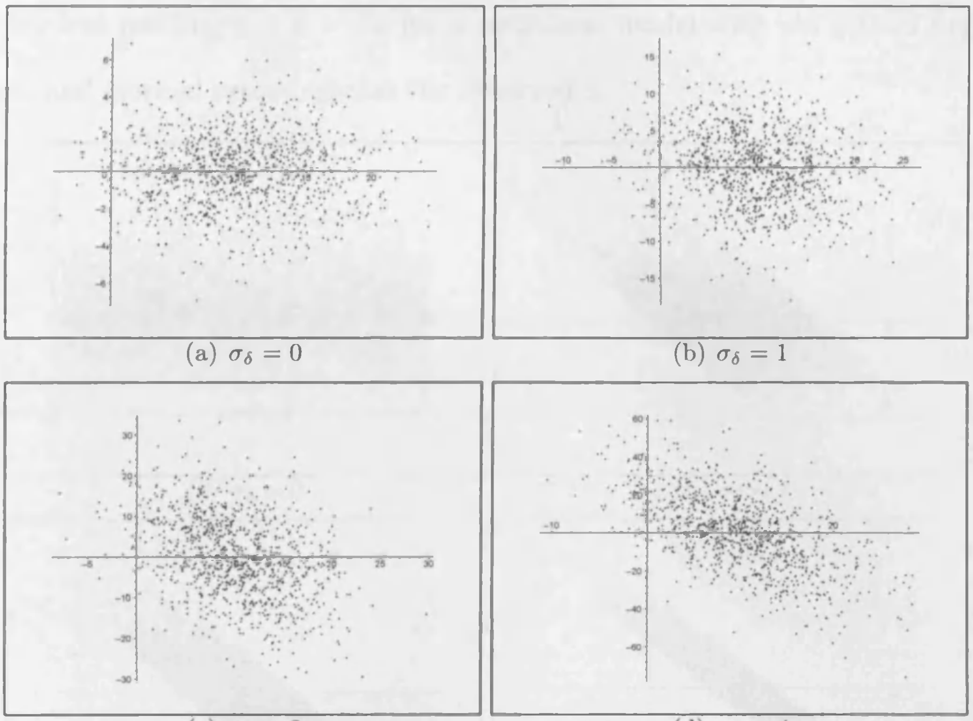
For our model, the problems with performing a standard regression analysis (as outlined in any of the standard textbooks cited earlier) can be seen by attempting to write our errors in variables model in terms of the observed data $\{(x_i, y_i), i = 1, \ldots, n\}$,

$$y = \alpha + \beta x + (\varepsilon - \beta\delta)$$

The problems here, which do not appear in the standard linear regression model are two-fold. Firstly, due to the additional random error component, δ, x is always random. Secondly, the observed x is correlated with the error term $(\varepsilon - \beta\delta)$. Indeed, $Cov[x, \varepsilon - \beta\delta] = -\beta\sigma_\delta^2$.

This latter point poses an immediate problem. A common tool used by practitioners to assess the fit of a regression model is to plot x against the vertical residual $y_i - \tilde{\alpha} - \tilde{\beta}x_i = \varepsilon - \tilde{\beta}\delta$. If there appears to be no trend or pattern in this plot, and the residuals are randomly dispersed around zero, then roughly speaking, this implies that the fitted model has a good fit to the data.

Figure 7.1 has residual plots of the vertical residual $y - \tilde{\alpha} - \tilde{\beta}x$ for a Normal structural model against the observed x. When $\sigma_\delta = 0$, there is no pattern in the residuals, and



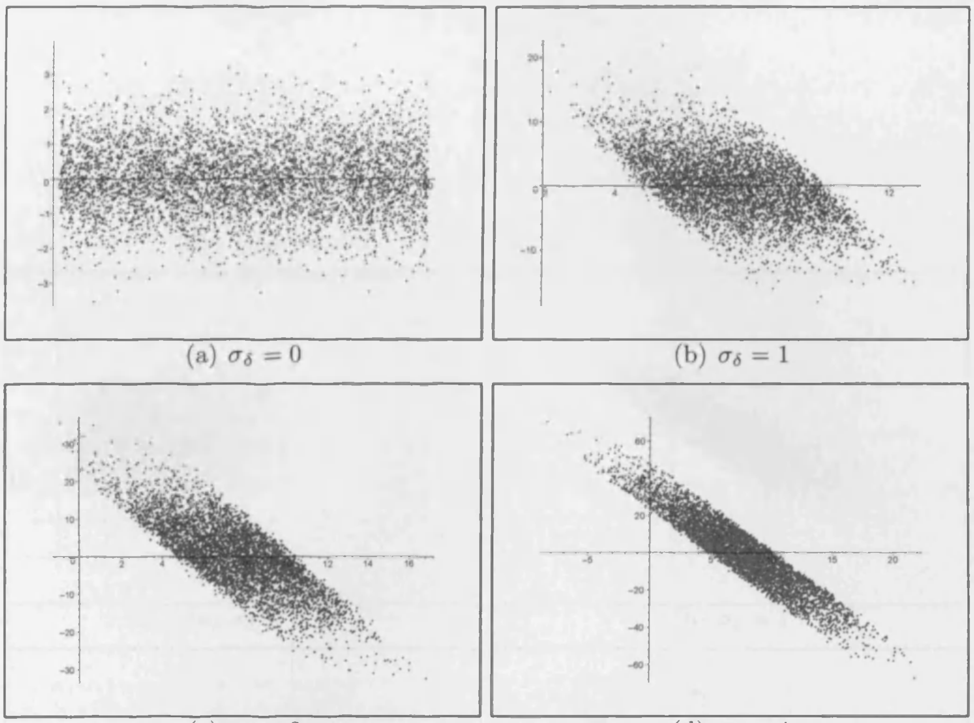
(a) $\sigma_\delta = 0$ (b) $\sigma_\delta = 1$

(c) $\sigma_\delta = 2$ (d) $\sigma_\delta = 4$

Figure 7.1: Vertical residual versus observed x, Normal structural model

they seem to be randomly dispersed about zero. As σ_δ gets larger, the trend between the residuals and the observed x becomes more prominent. This shows that without due care, the residual plots in an errors in variables setting are easily misinterpreted. In other words, in an errors in variables setting, there will always be a trend in the vertical residuals. In the majority of the plots in this Chapter, the scales of each picture are deliberately chosen to be different. This is because the point of note is to investigate the shape and structure of each simulated data set, which would be distorted if all the scales are set to the same structure.

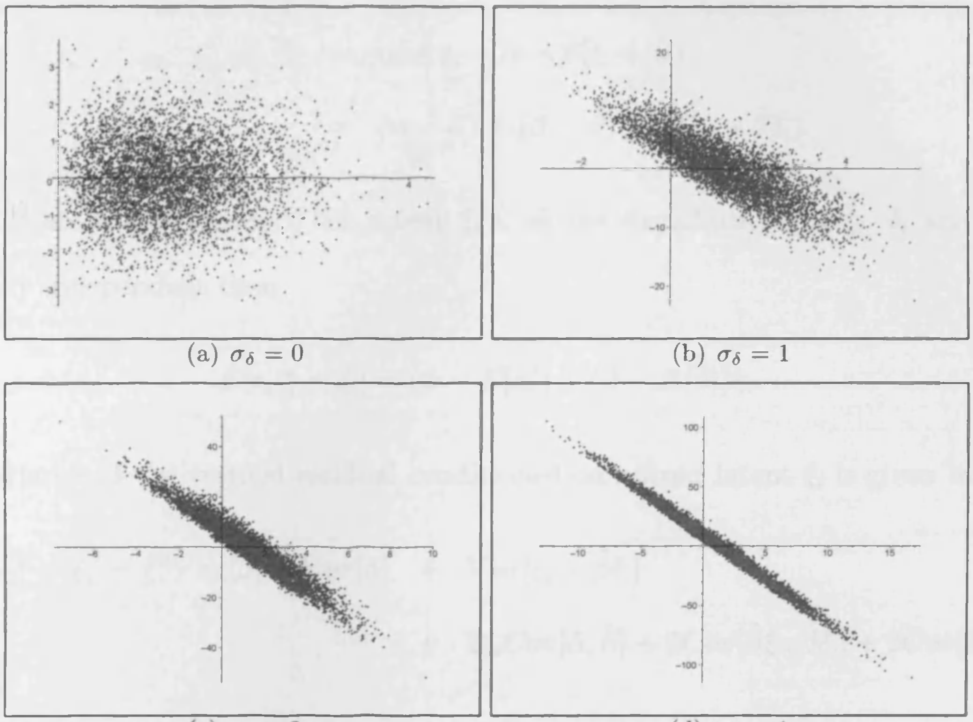
Different trends are observed upon changing the distribution of ξ in a structural setting. Figure 7.2 has residual plots of the vertical residual $y - \tilde{\alpha} - \tilde{\beta}x$ for a structural model with uniform ξ and Normal errors against the observed x. Figure 7.3 has residual plots of the vertical residual $y - \bar{\alpha} - \bar{\beta}x$ for a structural model with chi ξ (two degrees of freedom) and Normal errors against the observed x.



(a) $\sigma_\delta = 0$ (b) $\sigma_\delta = 1$

(c) $\sigma_\delta = 2$ (d) $\sigma_\delta = 4$

Figure 7.2: Vertical residual versus observed x, uniform ξ, Normal errors



(a) $\sigma_\delta = 0$ (b) $\sigma_\delta = 1$

(c) $\sigma_\delta = 2$ (d) $\sigma_\delta = 4$

Figure 7.3: Vertical residual versus observed x, chi ξ with two degrees of freedom, Normal errors

Before proceeding, properties of the vertical residual in an errors in variables setting will be considered. Once the errors in variables model has been fitted, we are provided with the estimated relationship $\tilde{y}(x) = \tilde{\alpha} + \tilde{\beta}x$. So, the vertical residual, r_i say, at the point x_i may intuitively be written as $r_i = y_i - \tilde{y}(x_i)$. If we condition upon a fixed value of ξ, then

$$\begin{aligned}(r_i|\xi = \xi_i) &= \eta_i + \varepsilon_i - \tilde{\alpha} - \tilde{\beta}x_i \\ &= \eta_i + \varepsilon_i - \tilde{\alpha} - \tilde{\beta}(\xi_i + \delta_i) \\ &= (\alpha - \tilde{\alpha}) + (\beta - \tilde{\beta})\xi_i + (\varepsilon_i - \tilde{\beta}\delta_i).\end{aligned}$$

$\tilde{\alpha}$ and $\tilde{\beta}$ are independent of the latent ξ_i's, as the distributions of ξ_i, δ_i and ε_i are mutually independent then

$$E[r_i|\xi = \xi_i] = (\alpha - E[\tilde{\alpha}]) + (\beta - E[\tilde{\beta}])\xi_i.$$

The variance of the vertical residual conditioned on a fixed latent ξ_i is given by

$$\begin{aligned}Var[r_i|\xi = \xi_i] = \xi_i^2 Var[\tilde{\beta}] + Var[\tilde{\alpha}] &+ Var[\varepsilon_i - \tilde{\beta}\delta_i] \\ &+ 2\xi_i Cov[\tilde{\alpha}, \tilde{\beta}] + 2Cov[\tilde{\beta}\xi_i, \tilde{\beta}\delta_i] + 2Cov[\tilde{\alpha}, \tilde{\beta}\delta_i].\end{aligned}$$

It is possible to simplify this expression. Since $E[\delta_i] = 0$ and $\tilde{\beta}$ is assumed to be independent of the error terms δ_i, then

$$Cov[\tilde{\beta}, \tilde{\beta}\delta_i] = E[\tilde{\beta}^2\delta_i] - E[\tilde{\beta}]E[\tilde{\beta}\delta_i] = 0.$$

Similarly,

$$Cov[\tilde{\alpha}, \tilde{\beta}\delta_i] = 0.$$

Finally

$$Var[\varepsilon_i - \tilde{\beta}\delta_i] = \sigma_\varepsilon^2 + Var[\tilde{\beta}]Var[\delta_i] + E^2[\tilde{\beta}]Var[\delta_i] = \sigma_\varepsilon^2 + \sigma_\delta^2 Var[\tilde{\beta}] + \beta^2\sigma_\delta^2.$$

Combining all these results yields the following expression

$$\begin{aligned} Var[r_i|\xi = \xi_i] &= \xi_i^2 Var[\tilde{\beta}] + Var[\tilde{\alpha}] + \sigma_\delta^2 Var[\tilde{\beta}] \\ &\quad +2\xi_i Cov[\tilde{\alpha}, \tilde{\beta}] + \sigma_\varepsilon^2 + \beta^2\sigma_\delta^2 \qquad (7.1) \\ &= Var[\tilde{\beta}]\left\{\xi_i + \frac{Cov[\tilde{\alpha}, \tilde{\beta}]}{Var[\tilde{\beta}]}\right\}^2 \\ &\quad -\frac{Cov^2[\tilde{\alpha}, \tilde{\beta}]}{Var[\tilde{\beta}]} + Var[\tilde{\alpha}] + \sigma_\delta^2 Var[\tilde{\beta}] + \sigma_\varepsilon^2 + \beta^2\sigma_\delta^2. \end{aligned}$$

for the variance of the vertical residual at a fixed ξ.

This expression collapses to that for the simple linear regression model where there is no measurement error in the x observations. Upon taking the standard results $\sigma_\delta^2 = 0$, $Var[\tilde{\beta}] = \frac{\sigma_\varepsilon^2}{s_{\xi\xi}}$, $Var[\tilde{\alpha}] = \sigma_\varepsilon^2 \frac{\sum \xi_i^2}{n s_{\xi\xi}}$ and $Cov[\tilde{\alpha}, \tilde{\beta}] = -\frac{\bar{\xi}\sigma_\varepsilon^2}{s_{\xi\xi}}$ we obtain

$$\begin{aligned} Var[r_i|\xi = \xi_i] &= \frac{\sigma_\varepsilon^2}{s_{\xi\xi}}(\xi_i - \bar{\xi})^2 + \sigma_\varepsilon^2 \frac{\sum \xi_i^2}{n s_{\xi\xi}} - \bar{\xi}^2 \frac{\sigma_\varepsilon^2}{s_{\xi\xi}} + \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 \left\{1 + \frac{1}{n} + \frac{(\xi_i - \bar{\xi})^2}{s_{\xi\xi}}\right\} \qquad (7.2) \end{aligned}$$

which is the corresponding result for simple linear regression.

Equation (7.1) may be further simplified by using the 'shortcut' formulae of Chapter 3,

$$\begin{aligned} Cov[\tilde{\alpha}, \tilde{\beta}] &= Cov[\bar{y}, \tilde{\beta}] - \beta Cov[\bar{x}, \tilde{\beta}] - \mu Var[\tilde{\beta}] \\ Var[\tilde{\alpha}] &= \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{n} + 2\mu(\beta Cov[\bar{x}, \tilde{\beta}] - Cov[\bar{y}, \tilde{\beta}]) \end{aligned}$$

yielding

$$\begin{aligned} Var[r_i|\xi = \xi_i] &= Var[\tilde{\beta}](\xi_i - \mu)^2 + \sigma_\delta^2 Var[\tilde{\beta}] + \left(1 + \frac{1}{n}\right)(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2) \\ &\quad +2(Cov[\bar{y}, \tilde{\beta}] - \beta Cov[\bar{x}, \tilde{\beta}])(\xi_i - \mu) \qquad (7.3) \end{aligned}$$

Expression (7.3) is fundamentally different to that of the simple linear regression model (7.2). For example, in (7.3), we have the larger term $(1+n^{-1})(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)$ instead of $(1+n^{-1})\sigma_\varepsilon^2$ as in (7.2). It is likely that the terms $(\xi_i-\bar{\xi})$ and $(\xi_i-\mu)$ will be similar as μ is an unbiased estimator of $\bar{\xi}$. However, in (7.2) $(\xi_i-\bar{\xi})$ is multiplied by $\frac{\sigma_\varepsilon^2}{s_{\xi\xi}}$ which is again likely to be a lot smaller than $Var[\tilde{\beta}]$ which premultiplies $(\xi_i-\mu)$ in (7.3). For the Normal linear structural model

$$Cov[\bar{y},\tilde{\beta}] = Cov[\bar{x},\tilde{\beta}] = 0.$$

Only upon varying the distribution of ξ from Normal will these covariances be non-zero.

It follows that

$$Var[r_i|\xi=\xi_i] \rightarrow \sigma_\varepsilon^2+\beta^2\sigma_\delta^2$$

as $n\rightarrow\infty$, since as $n\rightarrow\infty$ $Var[\tilde{\beta}]\rightarrow 0$ and $Cov[\bar{y},\tilde{\beta}]-\beta Cov[\bar{x},\tilde{\beta}]\approx 0$. Thus the vertical residuals are more variable towards the tails of the data, but around the mean μ have variance approximately equal to $\sigma_\varepsilon^2+\beta^2\sigma_\delta^2$.

It was results of this form that were exploited by Koduah [68] to obtain a nonparametric errors in variables fit. Koduah derived a local linear nonparametric estimator by taking weighted perpendicular projections from the fitted line to the data. This relates to taking $\lambda=1$ in $\tilde{\beta}_5$ (see Chapter 3).

7.3 Other Residuals

A vertical projection from the regression line on to a data point is not the only projection that can be considered in an errors in variables setting. As mentioned throughout this thesis, for the simple linear regression model a residual is immediately definable. The parameters of this model are then derived by minimising the sum of squares of these residuals. By the well known Gauss-Markov theorem (see Draper and Smith [37]), mimimising the sum of squares provides the best unbiased linear estimators for the parameters of the model.

Even though this methodology cannot be applied to an errors in variables model, some concept of a residual may inherently be found in alternative estimation methods. Some examples are given here.

Geometric mean The concept of the geometric mean slope estimator,

$$\tilde{\beta}_{GM} = sgn(s_{xy})\sqrt{\frac{s_{yy}}{s_{xx}}}$$

has an intuitive interpretation as it is the geometric mean of the slope estimator for y on x and x on y regression respectively.

A different motivation for geometric mean regression was derived by Barker, Soh and Evans [4]. Instead of looking at a geometrical average, they showed that $\tilde{\beta}_{GM}$ may be derived in its own right by adopting a least triangles approach. The least triangles approach aims to minimise the areas of the right-angled triangles formed with the regression line as the hypotenuse, and the vertical and horizonal projections from the data point onto the line as the remaining two sides. Mathematically speaking, this involves finding α and β so that

$$\frac{sgn(\beta)}{2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)\left(x_i - \frac{y_i - \alpha}{\beta}\right)$$

is minimised. The area of these triangles formed from the data point and the regression line may offer a residual which takes into account the extra random error component in the observed x.

Maximum likelihood As given in Chapter 5, the likelihood function of the Normal linear functional model given a sample $\{(x_i, y_i, i = 1, \ldots, n)\}$ may be written:

and this likelihood function is maximised when

$$L \propto \sigma_\delta \sigma_\varepsilon \exp\left[-\frac{1}{2\sigma_\delta^2}\sum_{i=1}^{n}(x_i - \xi_i)^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i - \alpha - \beta\xi_i)^2\right]$$

and this likelihood function is maximised when $\frac{(x_i-\xi_i)^2}{\sigma_\delta^2} + \frac{(y_i-\alpha-\beta\xi_i)^2}{\sigma_\varepsilon^2}$ is minimised for each data point (x_i, y_i).

The term

$$M\left((x,y),(\xi,\alpha+\beta\xi)\right) = \frac{(x-\xi)^2}{\sigma_\delta^2} + \frac{(y-\alpha-\beta\xi)^2}{\sigma_\varepsilon^2}$$

is a metric since $M(a,a) = 0$, $M(a,b) > 0$ if $a \neq b$, $M(a,b) = M(b,a)$ and $M(a,b) \leq M(a,c) + M(b,c)$ for any points a, b and c. Thus $M\left((x,y),(\xi,\alpha+\beta\xi)\right)$ may be considered as a potential residual.

The problem with using M as a residual is the dependence on the latent variable ξ. However, once an initial errors in variables fit has been made, ξ may be estimated as shown in Chapter 6. This residual however, is different from any discussed previously. It is a distance from the observed data point (x_i, y_i) onto the latent data point (ξ_i, η_i).

Leading from this point, the metric M is known as the Mahalanobis distance. The Mahalanobis distance is a useful way of determining similarity of an unknown data set, to a known one. This Mahalanobis distance is regularly used in simple linear regression for outlier detection and leverage analysis. The data point with greatest Mahalanobis distance is known to exert the greatest amount of leverage on the fitted regression line.

Weighted least squares As described in the previous Chapter, weighted least squares is used for an errors in variables model to minimise the sum of squared residuals where a projection other than vertical or horizontal is taken.

Thus the term $D_i(\lambda)$ introduced in Chapter 6 may also be considered as a residual which takes both the errors in the x and y measurement. The expression for $D_i(\lambda)$ is repeated here:

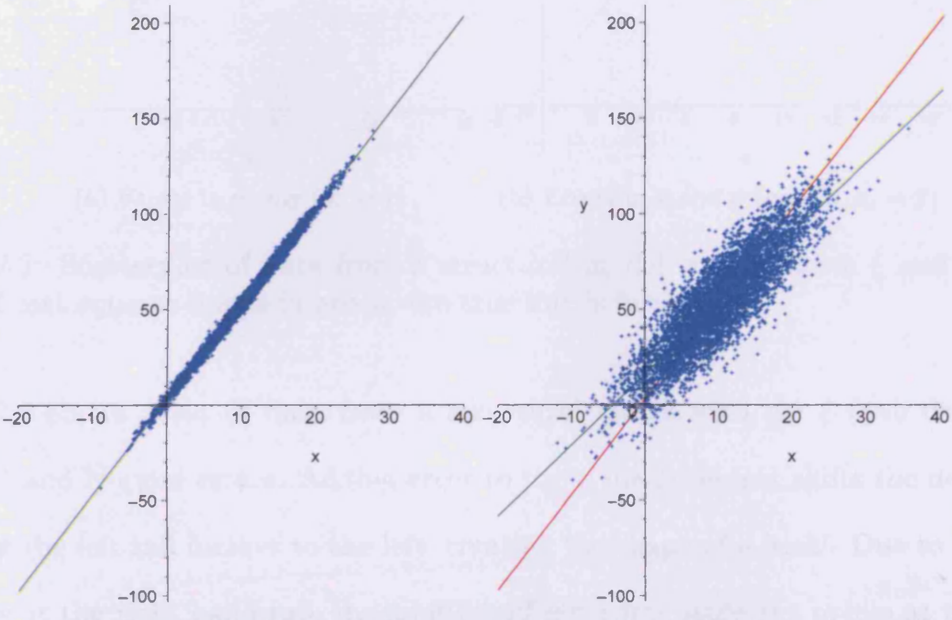
$$D_i(\lambda) = \sqrt{\tilde{\delta}_i^2 + \tilde{\varepsilon}_i^2} = (\sqrt{\lambda^2 + \beta^2})\,\frac{y_i - \alpha - \beta x_i}{\lambda + \beta^2}.$$

This residual is simply a rescaling of the vertical residual, and so there is no benefit in using this residual instead of the vertical residual.

7.4 Migration

It has already been demonstrated that there will be a trend in the vertical residual for an errors in variables model. The introduction of measurement error in the x observation will increase the variability of points about the line. An additional surprising feature of errors in variables modelling is the tendency for data not to be symmetrically distributed around the fitted line, sometimes giving the impression that the fitted line inadequately describes the data in question. This phenomenon has been discovered by Nix (pers.comm.) but investigated by Koduah [68], and describes the movement of the observed data due to the additional error component attributed to the x measurement. For $\beta > 0$ it appears that data are more prevalent above the fitted line at the left hand tail of the data, and are more prevalent below the fitted line at the right hand tail of the data. This not only has important implications for residual analyses, but also for assessing and constructing the reference bands that are commonly used in diagnostic screening (see for example Royston [89]).

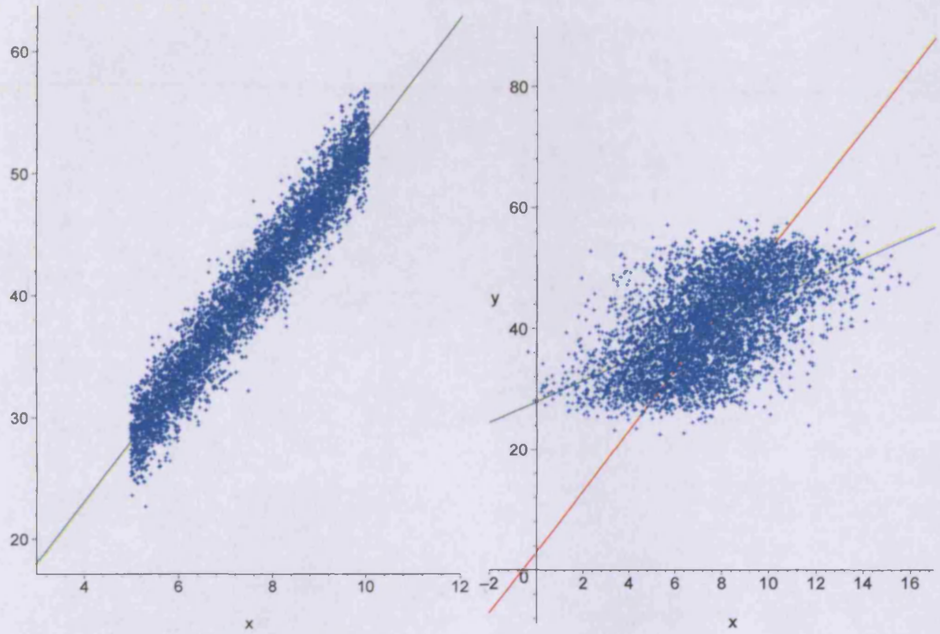
Figure 7.4 shows a set of data generated with Normal ξ. The parameter settings chosen were $\alpha = 3$, $\beta = 5$, $\mu = 10$ and $\sigma = 5$. At any given x, there appears to be a roughly symmetric distribution of y. If measurement error is also added to the x measurement, then this is not the case. The symmetry about the true line of y given x has disappeared, and this is more marked in the tails of the data. The distribution at the left hand tail is asymmetric at any x with larger values of y predominating. At the right hand tail the asymmetry is also present, but is skewed towards lower values of y. The approximate effect of migration is to twist or rotate the data about the true line. This yields the asymmetry of y given x to be more marked at the tails.



(a) Errors in y only ($\sigma_\varepsilon = 2$) (b) Errors in x and y ($\sigma_\delta = 3, \sigma_\varepsilon = 2$)

Figure 7.4: Scatterplot of data from a Normal structural model. Least squares line is in green, the true line is in red.

Figure 7.5 shows a set of data with uniform ξ. Adding error to the x measurement gives a similar migration effect to that of Normal ξ but is not as pronounced. Here the tails appear to further tail off giving a subtle 'z' or saw-tooth structure to the data. The left hand tail has moved horizontally to the left, and the right hand tail has moved horizontally to the right.

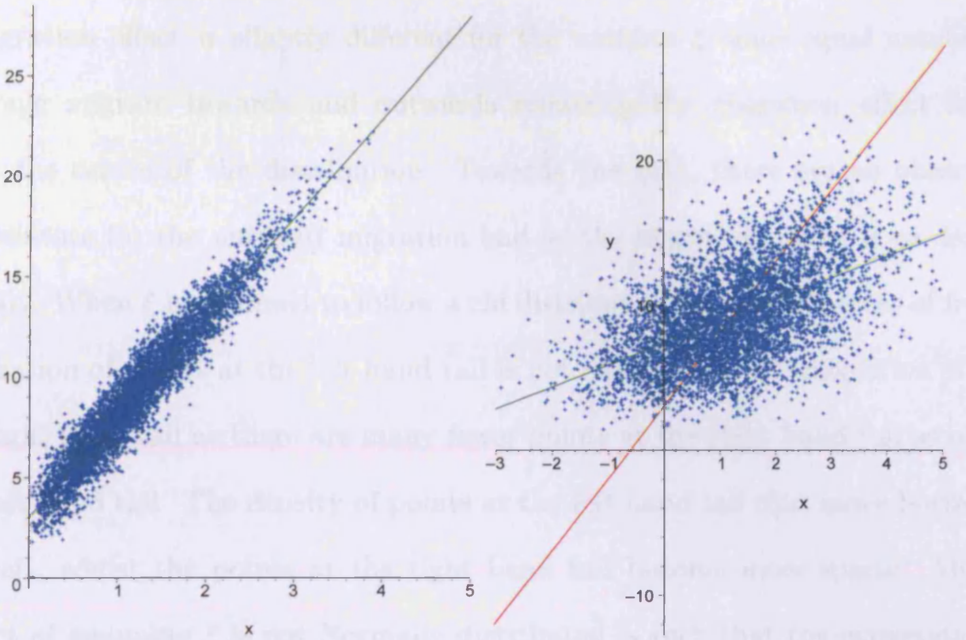


(a) Errors in y only ($\sigma_\varepsilon = 2$) (b) Errors in x and y ($\sigma_\delta = 2, \sigma_\varepsilon = 2$)

Figure 7.5: Scatterplot of data from a structural model with uniform ξ and Normal errors. Least squares line is in green, the true line is in red.

Figure 7.6 shows a set of data from a structural model with chi ξ (two degrees of freedom) and Normal errors. Adding error to the x measurement shifts the density of points at the left tail further to the left, creating the shape of a 'tick'. Due to the lack of points at the right hand tail, the addition of error has made the points at the right hand tail appear more sporadic.

Reasoning for migration can be found by considering the effect of adding measurement



(a) Errors in y only ($\sigma_\varepsilon = 1$) (b) Errors in x and y ($\sigma_\delta = 1, \sigma_\varepsilon = 1$)

Figure 7.6: Scatterplot of data from a structural model with chi ξ (two degrees of freedom) and Normal errors. Least squares line is in green, the true line is in red.

error to an x measurement. As demonstrated in the previous figures, prior to adding measurement error to the x measurement, the distribution of y given any x is at least roughly symmetric about the true line. Once measurement error is added some data have migrated to the left, and some have migrated to the right, approximately half in each direction. For the Normal distribution however there are more observations close to the mean, and fewer in the tails. So more observations migrate outwards than are compensated for by observations moving inwards. At the left hand tail, for observations originally above the true line there are greater numbers migrating left than to the right. Observations below the line migrating outwards in greater numbers than migrating inwards tend to be closer to the true line after adding measurement error to the x. At the right hand tail the opposite effect is seen. This provides the rotating effect that has been demonstrated in Figure 7.4.

The migration effect is slightly different for the uniform ξ since equal numbers will on average migrate inwards and outwards removing the migration effect for data close to the centre of the distribution. Towards the tails, there are no observations to compensate for the outward migration and so the migration effect is as described previously. When ξ is assumed to follow a chi distribution with two degrees of freedom, the migration of points at the left hand tail is not balanced by the migration of points at the right hand tail as there are many fewer points at the right hand tail as opposed to the left hand tail. The density of points at the left hand tail thus move horizontally to the left, whilst the points at the right hand tail become more sparse. Moreover the effect of assuming ξ is not Normally distributed is such that the expectation of y given x is not a straight line, but a curve. This was demonstrated algebraically in the previous Chapter.

A key feature with Figures 7.4, 7.5 and 7.6 is that the true line does not appear to be the best representation of the data. On inspection, it appears that the least squares line provides the best fit, in particular for the Normal structural model. This is the distortion effect of migration, that makes the true line (which is best estimated by the errors in variables line) appear incorrect. This could lead to much confusion, in particular when it comes to residual analysis. A thorough understanding of the effects of migration is essential to fit and check an errors in variables model.

For Normal ξ it is possible to describe this migration effect algebraically. In this section, two different algebraic reasonings for the migration phenomenon with Normal ξ shall be presented. The first relates to the conditional distribution of y given x, and the second relates to the contours of equal probability for a bivariate Normal distribution.

Distribution of $y|x$ For the Normal linear structural model, the conditional distribution of $(y|x = x_0)$ is also Normal (see for example Chapter 6, or DeGroot [32]). Indeed we may write,

$$(y|x = x_0) \sim N\left[\alpha + \beta\mu + \frac{\beta\sigma^2}{\sigma^2 + \sigma_\delta^2}(x_0 - \mu), \left(1 - \frac{\beta^2\sigma^4}{\sigma^2 + \sigma_\delta^2}\right)\sigma_\varepsilon^2\right]$$

and the conditional mean of y given x is (as seen in Chapter 6)

$$E[y|x = x_0] = \alpha + \beta\mu + \frac{\beta\sigma^2}{\sigma^2 + \sigma_\delta^2}(x_0 - \mu).$$

Hence the average of the vertical distance from y to the true line $y_0 = \alpha + \beta x_0$ at $x = x_0$ is

$$-\beta(x_0 - \mu)\frac{\sigma^2}{\sigma^2 + \sigma_\delta^2} = -\beta(x_0 - \mu)\kappa,$$

where κ is the reliability ratio. So when $x_0 < \mu$ the average migration from the true line is positive and when $x_0 > \mu$ the average migration is negative. As mentioned previously, the least amount of migration will occur when $x_0 \approx \mu$. The amount of migration is proportional to β and the reliability ratio κ. In other words the vertical scatter of y values will not be symmetric, especially in the extremes in the range of x. In simple linear regression this migration effect does not occur since the least squares line is an unbiased estimator for the expression for $E[y|x = x_0]$.

This migration effect will clearly have an impact upon the vertical residuals. Again, for the Normal linear structural model we have

$$(y - \alpha - \beta x|x = x_0) \sim N\left[-\frac{\beta\sigma_\delta^2}{\sigma^2 + \sigma_\delta^2}(x_0 - \mu), (1 - \beta^2\sigma_\delta^4)(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^4)\right]$$

Assuming that the slope β is positive, for $x < \mu$ then on average, the vertical residuals will be negative. For $x > \mu$, then on average the vertical residuals will be positive. The effect will be less marked for observations close to the mean, and more marked in the tails. This gives the effect that the observed data has rotated clockwise about the true line. When the slope β is negative, the reverse of the above description applies.

For models other than the Normal structural model, the conditional distribution of $y|x$ can go some way in describing the migration effect. For example, as seen in the previous Chapter when the distribution of ξ is varied away from Normal in a structural model then the expression $E[y|x]$ varies. The figures in Section 6.2 show that the expressions and approximations for $E[y|x]$ follow the migration effect. For the Normal structural model, $E[y|x]$ is the least squares line. The migration effect in the Normal structural model is for the data to rotate about the true line, and then the least squares line seems the best fit of the data. When ξ follows a uniform distribution in the structural model, then the expression and approximation for $E[y|x]$ follow the migration of the data at the tails of the data. This follows for all the examples considered in the previous section.

Equiprobability contours and ellipses For the Normal structural model, the observed data (x, y) create an elliptical shape on the scatterplot of x and y. This was demonstrated in Figure 7.4.

The equation

$$ax^2 - 2bxy + cy^2 = k \tag{7.4}$$

defines an ellipse whose major axis is oblique to the x axis, with centre at $(0, 0)$. The inclination of the ellipse θ, as given in Figure 7.7 is defined by values of a, b and c. In fact

$$\tan(2\theta) = \frac{2b}{c-a} = \frac{2\tan(\theta)}{1-\tan^2(\theta)} \tag{7.5}$$

and thus a quadratic form for $\tan(\theta)$ may be constructed. The roots of the quadratic are the slopes of the major and minor axis of the ellipse.

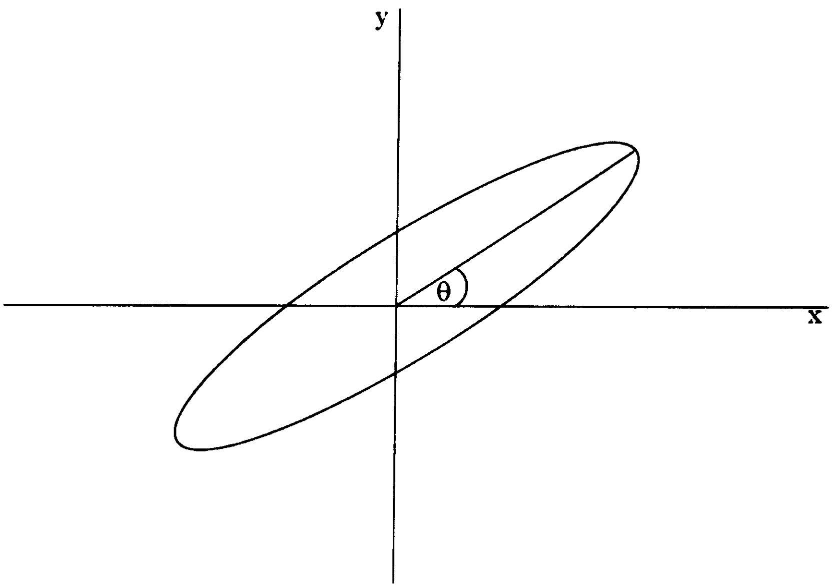


Figure 7.7: An ellipse with inclination θ.

This can be related to the Normal linear structural model by looking at contours of equal probability for a bivariate Normal distribution. For simplicity, we assume the mean to be $(0, 0)$. These contours are elliptical and are described by the form

$$\frac{x^2}{Var[x]} + \frac{y^2}{Var[y]} - \frac{2\rho}{\sqrt{Var[x]}\sqrt{Var[y]}}xy = k$$

for different constants k. The introduction of a random error component in the observed x is to change the variance covariance matrix from $\begin{pmatrix} \sigma^2 & \beta\sigma^2 \\ \beta\sigma^2 & \sigma_\varepsilon^2 + \beta^2\sigma^2 \end{pmatrix}$ to $\begin{pmatrix} \sigma^2 + \sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \sigma_\varepsilon^2 + \beta^2\sigma^2 \end{pmatrix}$. The effect is to rotate the ellipse formed by the contours of equiprobability. Comparing the bivariate Normal distribution with (7.4), $a = \frac{1}{\sigma^2+\sigma_\delta^2}$,

$b = \frac{\beta\sigma^2}{(\sigma^2+\sigma_\delta^2)(\beta^2\sigma^2+\sigma_\varepsilon^2)}$ and $c = \frac{1}{\beta^2\sigma^2+\sigma_\varepsilon^2}$ we may write

$$\tan(2\theta) = \frac{2b}{c-a} = \frac{2\beta\sigma^2}{(\sigma^2+\sigma_\delta^2)-(\beta^2\sigma^2+\sigma_\varepsilon^2)}.$$

As σ_δ^2 increases, $\tan(2\theta)$ decreases and the ellipse rotates, but also expands outwards because of the increase in a. This is exactly the migration effect discussed earlier. In practise seeing an elliptical scatterplot with major axis inclined at an angle θ, it is impossible to totally distinguish how much of this inclination is due to β and how much to σ_δ^2.

The migration effect for the Normal structural model is at least partly corrected for by the method of moments estimator of ξ, $\tilde{\xi}$ described in the previous Chapter. Assume $\beta > 0$. If $y_i > \alpha + \beta x_i$ (observed y above true line at observed x) then $\tilde{\xi}_i > x_i$ but $\tilde{\eta}_i < y_i$. If $y_i < \alpha + \beta x_i$ (observed y below true line at observed x) then $\tilde{\xi}_i < x_i$ but $\tilde{\eta}_i > y_i$. If $\beta < 0$, then the obvious alterations are made to the above inequalities. The magnitude of the difference between $\tilde{\xi}_i$ and x_i is greater for those observations most distant from the true line. In summary, the method of moments estimator of ξ, $\tilde{\xi}$ goes some way to reverse the rotation and distortion away from a straight line effect of migration in order to predict the original, unobserved ξ.

A related issue is how the major axis is related to estimators discussed in Chapter 3. As stated previously, the roots of (7.5) yield the slope of the major and minor axis. The roots of this quadratic are

$$\tan(\theta) = \frac{(a-c) \pm \sqrt{(a-c)^2+4b^2}}{2b}. \qquad (7.6)$$

Let $\tan(\theta_1) = \frac{(a-c)+\sqrt{(a-c)^2+4b^2}}{2b}$ and $\tan(\theta_2) = \frac{(a-c)-\sqrt{(a-c)^2+4b^2}}{2b}$. The equation of the major axis $y = \tan(\theta_i)$ for one of these θ_i, $i = 1, 2$, depending on the inclination of the ellipse. Some algebra shows that $-\frac{1}{\tan(\theta_1)} = \tan(\theta_2)$, thus the two solutions for θ give lines at right angles. In other words, the two lines form the major and minor axis for the ellipse.

We may estimate the slope of the major axis by approximating a, b and c with their moment equivalents, $a = \frac{1}{s_{xx}}$, $b = \frac{s_{xy}}{s_{xx}s_{yy}}$ and $c = \frac{1}{s_{yy}}$. Substituting these into $\tan(\theta_1)$ gives (after some simplification)

$$\tilde{\tan}(\theta_1) = \frac{(s_{yy} - s_{xx}) + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}.$$

This is also the estimator for the slope, $\tilde{\beta}_5$ used when the ratio of the error variances λ is known and equal to 1. The major axis is thus estimated by $y = \tilde{\beta}_5 x$ when $\lambda = 1$. This major axis is the same as the first principal component, whilst the minor axis is the same as the second principal component. This is always true for a bivariate data set (see for example Seal [91].)

Migration for different structural models Some mention has already been made as to the migration effect for models other than the Normal structural model. The expression $E[y|x]$ appears to follow the density of points as they migrate upon the addition of measurement error to the x measurement. The severity of the migration effect depends on the reliability ratio, in that the larger the σ_δ^2 in comparison to σ^2, then the more marked migration effect.

Investigation of the joint probability density function of x and y for different distributions of ξ provide further insights into the migration effect. For example, as seen in Chapter 5 the joint probability density function of x and y for a structural model with uniform ξ and Normal errors is given by

$$\begin{aligned} f_{x,y}(x,y) &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}}\exp\left[-\frac{1}{2}\left(\frac{(y-\alpha-\beta x)^2}{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}\right)\right] \\ &\quad \times\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(b-\tilde{\xi}\right)\right]-\Phi\left[\sqrt{A}\left(a-\tilde{\xi}\right)\right]\right\} \end{aligned}$$

As mentioned in Chapter 5, the term

$$\frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}}\exp\left[-\frac{1}{2}\left(\frac{(y-\alpha-\beta x)^2}{\sigma_\varepsilon^2+\beta^2\sigma_\delta^2}\right)\right]$$

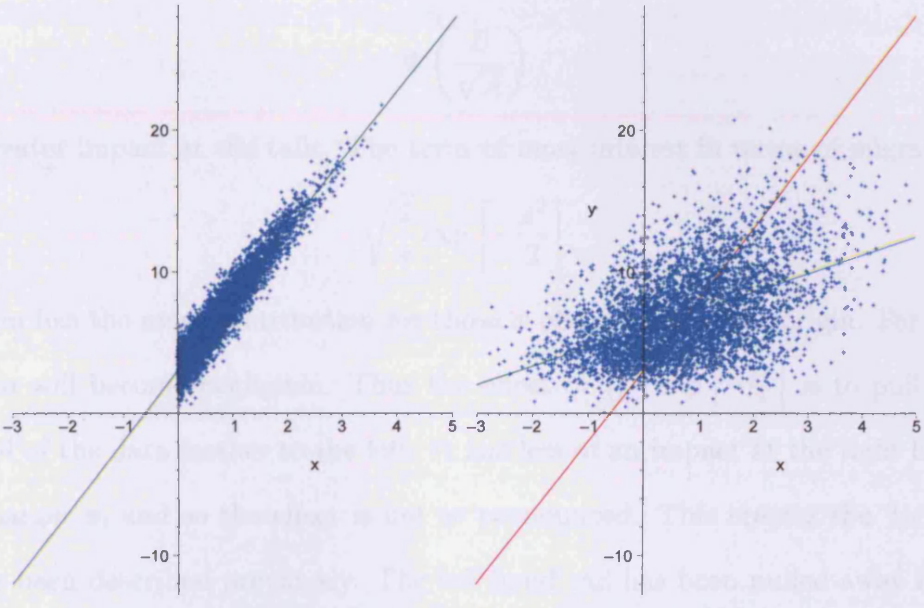
may be viewed as the joint probability density function of x and y of the simple linear regression model, with inflated variance $\sigma_\varepsilon^2+\beta^2\sigma_\delta^2$. As with the Normal structural model, the term in the exponential may be viewed as describing the spread of the observed data $\{(x_i,y_i), i=1,\ldots,n\}$. Instead of an ellipse as in the Normal structural model, if the disturbance term

$$\frac{1}{(b-a)}\left\{\Phi\left[\sqrt{A}\left(b-\tilde{\xi}\right)\right]-\Phi\left[\sqrt{A}\left(a-\tilde{\xi}\right)\right]\right\}$$

is ignored the data for a structural model with uniform ξ and Normal errors are spread around the least squares line, with variance $\beta^2\sigma_\delta^2+\sigma_\varepsilon^2$. The presence of this disturbance term however, has the effect of stretching the tails of the data horizontally away from the mean, as described in Chapter 5. This explains the migration effect seen for the structural model with uniform ξ and Normal errors as explained earlier.

Similar consideration may be given to the example in Chapter 5 of ξ following a chi distribution (one degree of freedom) with Normal errors. As stated earlier, a chi distribution with one degree of freedom is the half Normal distribution. As this distribution is heavily skewed, a migration effect similar as to when ξ follows a chi distribution with two degrees of freedom is observed. The density of points at the left tail moves horizontally to the left, and the points at the right hand tail become more sporadic.

For illustrative purposes, Figure 7.8 is an example of a structural model with chi ξ with Normal errors, with and without measurement error in the x measurement. Again, the 'tick' effect that occurs when ξ follows a chi distribution with one degree of freedom is observed.



(a) Errors in y only ($\sigma_\varepsilon = 1$) (b) Errors in x and y ($\sigma_\delta = 1, \sigma_\varepsilon = 1$)

Figure 7.8: Scatterplot of data from a structural model with chi ξ (one degrees of freedom) and Normal errors. Least squares line is in green, the true line is in red.

As derived in Chapter 5, the joint probability density function of x and y when ξ is assumed to follow a chi distribution with one degree of freedom and the errors are Normally distributed is given by

$$f_{x,y}(x,y) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}} \exp\left[-\frac{(y-\alpha-\beta x)^2}{2(\sigma_\varepsilon^2+\beta^2\sigma_\delta^2)}\right] \times \sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right] \Phi\left(\frac{B}{\sqrt{A}}\right)$$

where

$$A = \frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{\sigma_\delta^2\sigma_\varepsilon^2}$$

$$B = \frac{x}{\sigma_\delta^2} + \frac{\beta(y-\alpha)}{\sigma_\varepsilon^2}.$$

Again ignoring the term

$$\sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right] \Phi\left(\frac{B}{\sqrt{A}}\right)$$

the spread of the points $\{(x_i, y_i), i = 1, \ldots, n\}$ (in the structural model with chi ξ with one degree of freedom and Normal errors) is about the least squares line with inflated variance $\sigma_\varepsilon^2 + \beta\sigma_\delta^2$. As written in Chapter 5, the term

$$\Phi\left(\frac{B}{\sqrt{A}}\right)$$

has a greater impact at the tails. The term of most interest in terms of migration is

$$\sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right]$$

This term has the most contribution for those x observed near the origin. For large x, this term will become negligible. Thus the effect of $\sqrt{\frac{2}{\pi}} \exp\left[-\frac{x^2}{2}\right]$ is to pull the left hand tail of the data further to the left. It has less of an impact at the right hand tail for the larger x, and so the effect is not as pronounced. This creates the 'tick' effect that has been described previously. The left hand tail has been pulled away from the least squares line, whilst the right hand tail still seems to follow the least squares line.

7.5 Vertical Residuals Revisited

The vertical residuals from the errors in variables fit always demonstrate a trend because of the migration phenomenon, which moves the data around the true line. Therefore the vertical residuals from the unbiased estimator of the true line, the errors in variables fit, may be misleading when it comes to detecting outliers, and for model checking. For this reason, the vertical residual from the errors in variables fit may not be suitable for diagnostic analysis of the fitted model.

Based on work from earlier Chapters of this Thesis, there are two ways in which this migration phenomenon might be dealt with. As mentioned previously, the exact expressions for $E[y|x]$ derived in the previous Chapter seem to follow the migration of the data. Thus, the vertical residual from $E[y|x]$ would provide a residual which would not display the trend that occurs when looking at the vertical residual from the errors in variables fit.

Secondly, the migration phenomenon is at least partially corrected for by the method of moments estimator of ξ, $\tilde{\xi}$ (equation (6.13)). Once an errors in variables model has been fitted, then the estimated parameters from the errors in variables fit may be used to estimate the latent ξ values. A simple least squares fit to the data set $\left\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\right\}$ can then be obtained, and vertical residuals from this least squares fit should not display the trend that occurs when looking at the vertical residual from the errors in variables fit.

Some examples of both methods are given here. For the Normal structural model, $E[y|x]$ follows the least squares line. In this scenario then, the vertical residuals from the least squares fit should not be subject to the migration phenomenon. Figure 7.9 for a Normal structural model with parameter settings $\alpha = 3$, $\beta = 5$, $\mu = 10$, $\sigma = 5$, $\sigma_\varepsilon^2 = 1$, $n = 5000$, shows plots of the vertical residual from the errors in variables fit, and from the least squares line ($E[y|x]$) for increasing σ_δ^2. It can be seen that the vertical residuals from the errors in variables line is subject to the migration phenomenon, and the rotation is more marked for large σ_δ. The vertical residual from the least squares line however does not display the migration phenomenon, and is more robust to the increase in σ_δ^2.

As seen in the previous Chapter, the expressions for $E[y|x]$ for structural models other than the Normal structural model are not so simple. Examples of exact expressions for when ξ is assumed to follow a uniform or chi distribution with Normal errors were given and investigated.

Figure 7.10 shows plots of the vertical residual from the errors in variables fit, and from the $E[y|x]$ curve for increasing σ_δ^2, when ξ is assumed to follow a uniform distribution. The parameter settings are $a = 5$, $b = 15$, $\alpha = 3$, $\beta = 5$, $\sigma_\varepsilon = 1$ and $n = 5000$. As σ_δ increases to extreme levels, then a rotated diamond shape becomes more distinct. This is partly due to the migration effect with uniform ξ, the data spreads out at both tails creating an elongated 's' shape. Thus for extreme error, this migration effect is particularly noticed, and so the vertical residual at the tails will be larger than for in the middle of the data.

Figure 7.11 shows plots of the vertical residual from the errors in variables fit, and from the $E[y|x]$ curve for increasing σ_δ^2, when ξ is assumed to follow a chi distribution with two degrees of freedom. The parameter settings are $\alpha = 3$, $\beta = 5$, $\sigma_\varepsilon = 1$ and $n = 5000$. Unlike Figure 7.10 the residual plot from the $E[y|x]$ curve is not as affected by changes in σ_δ.

The second method suggested to correct for migration is the use of the estimator of ξ, $\tilde{\xi}$ (equation (6.13)) that was introduced in the previous Chapter. In summary, the key steps of this method are:

1. Fit the errors in variables line to the data $\{(x_i, y_i), i = 1, \ldots, n\}$, thus obtaining unbiased estimators for α and β.

2. Estimate the latent ξ_i's by using the formula

$$\tilde{\xi}_i = \frac{\lambda}{\lambda + \tilde{\beta}^2} x_i + \frac{\tilde{\beta}}{\lambda + \tilde{\beta}^2} (y_i - \tilde{\alpha}).$$

If λ is unknown, then it may be estimated from the method of moment estimating equations (3.1) to (3.5).

3. Fit the least squares line to the data $\{(\tilde{\xi}_i, y_i), i = 1, \dots, n\}$ thus obtaining a regression equation of the form $y_i = \alpha_0 + \beta_0 \tilde{\xi}_i$

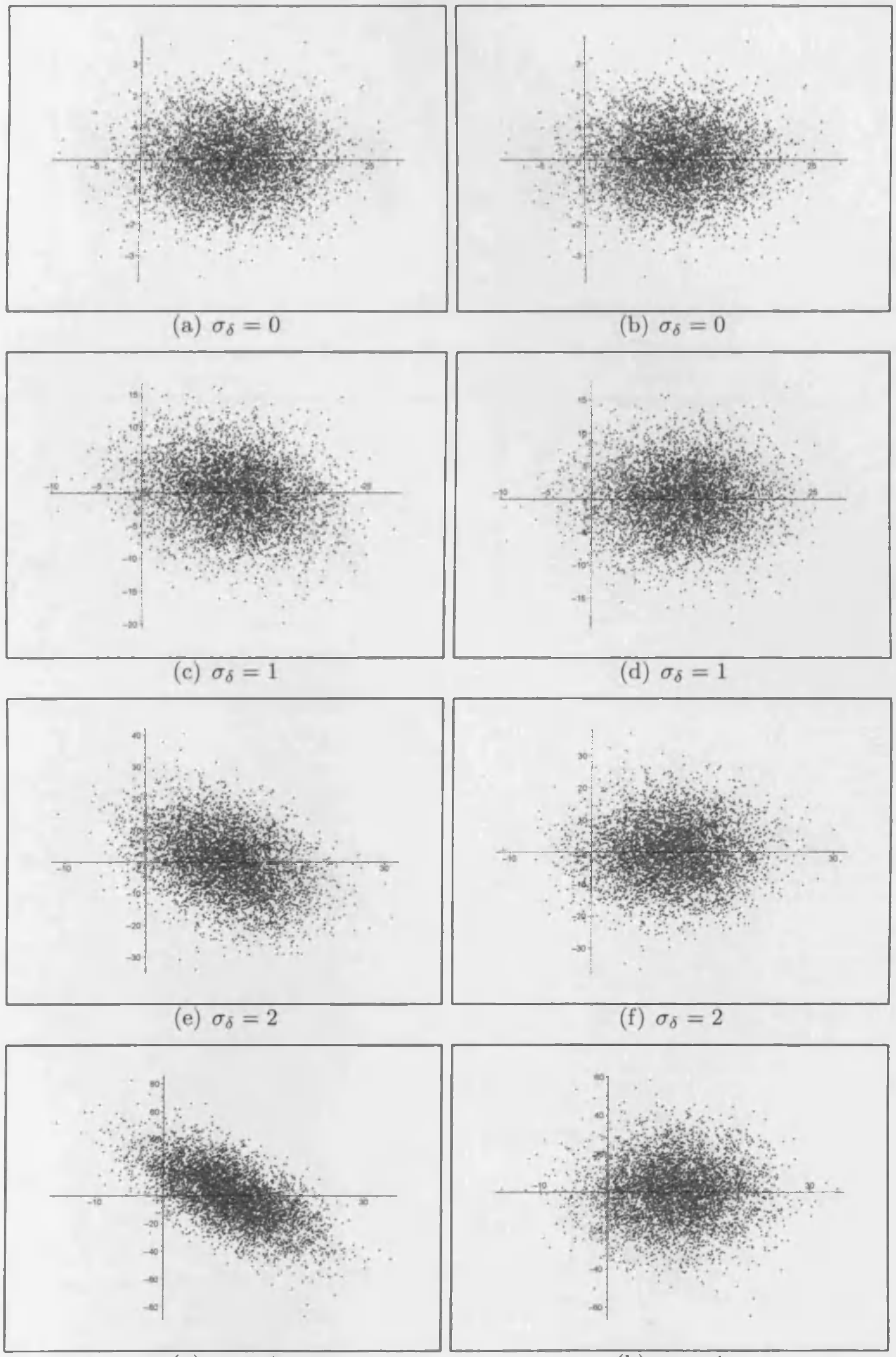
4. Conduct standard residual analysis on the vertical residuals $y_i - \alpha_0 - \beta_0 \tilde{\xi}_i$. Standard residual theory for the simple linear regression model may now be used since $\tilde{\xi}$ is the estimated value of x without measurement error.

For completeness, the vertical residual $y_i - \alpha_0 - \beta_0 \tilde{\xi}_i$ where α_0 and β_0 are the estimators of the intercept and slope of the least squares fit to the data set $\{(\tilde{\xi}_i, y_i), i = 1, \dots, n\}$ is compared to the vertical residual from the errors in variables fit for the same distributions and parameter settings used in Figures 7.9 to 7.11.

Figure 7.12 for a Normal structural model with parameter settings $\alpha = 3$, $\beta = 5$, $\mu = 10$, $\sigma = 5$, $\sigma_\varepsilon^2 = 1$, $n = 5000$, shows plots of the vertical residual from the errors in variables fit, and from the least squares line ($E[y|x]$) for increasing σ_δ^2.

Figure 7.13 shows plots of the vertical residual from the errors in variables fit, and the vertical residual from least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \dots, n\}$ for increasing σ_δ^2, when ξ is assumed to follow a uniform distribution. The parameter settings are $a = 5$, $b = 15$, $\alpha = 3$, $\beta = 5$, $\sigma_\varepsilon = 1$ and $n = 5000$. This plot does not display the same trend of creating diamond like shapes as seen in Figure 7.10, and seems fairly robust to changes in σ_δ.

Figure 7.14 shows plots of the vertical residual from the errors in variables fit, and the vertical residual from least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$ for increasing σ_δ^2, when ξ is assumed to follow a chi distribution with two degrees of freedom. The parameter settings are $\alpha = 3$, $\beta = 5$, $\sigma_\varepsilon = 1$ and $n = 5000$.



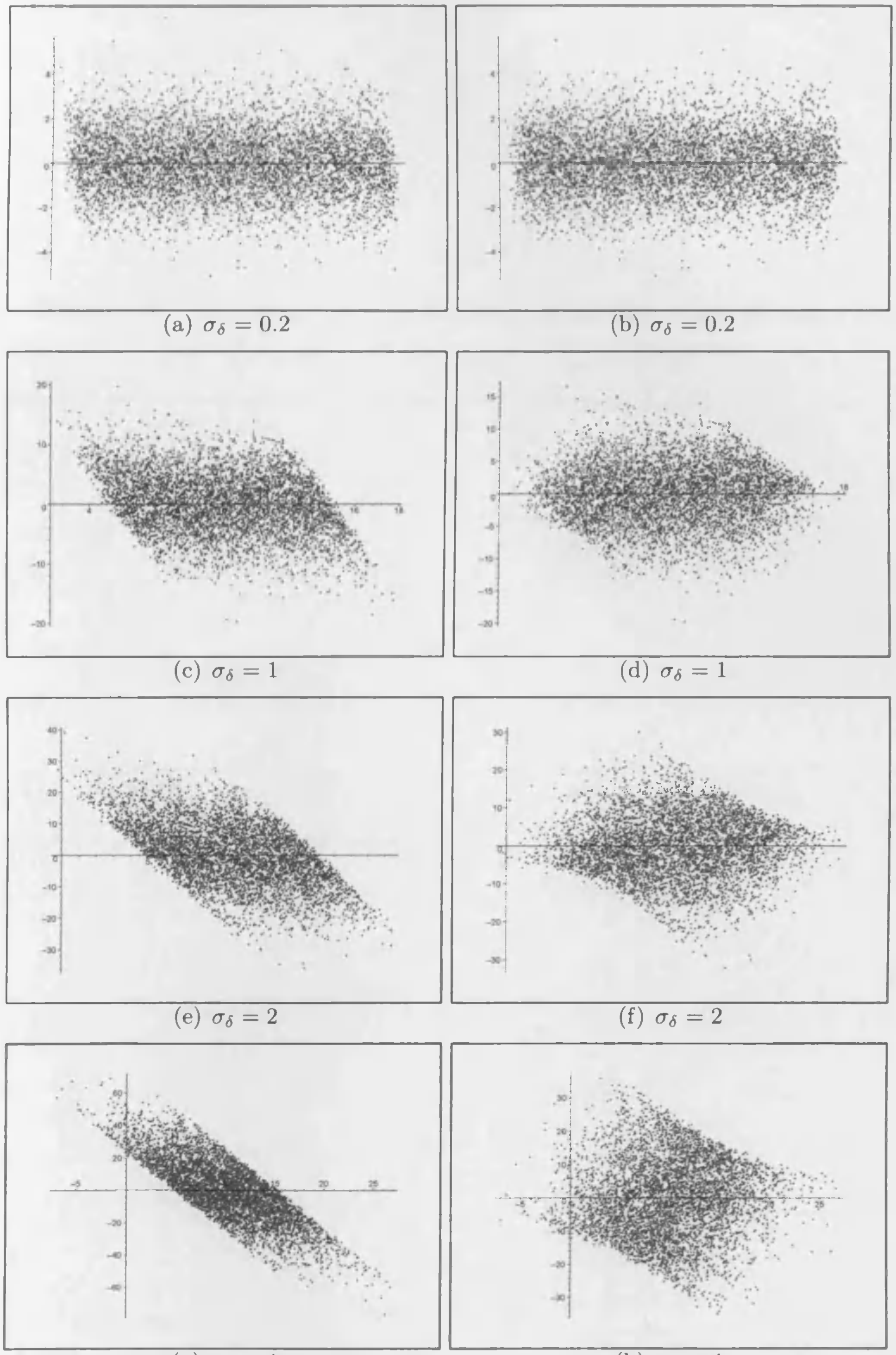
(a) $\sigma_\delta = 0$ (b) $\sigma_\delta = 0$

(c) $\sigma_\delta = 1$ (d) $\sigma_\delta = 1$

(e) $\sigma_\delta = 2$ (f) $\sigma_\delta = 2$

(g) $\sigma_\delta = 4$ (h) $\sigma_\delta = 4$

Figure 7.9: Vertical residual versus observed x, Normal structural model. Vertical residual from errors in variables fit is on the left hand side, vertical residual from least squares fit is on the right hand side.



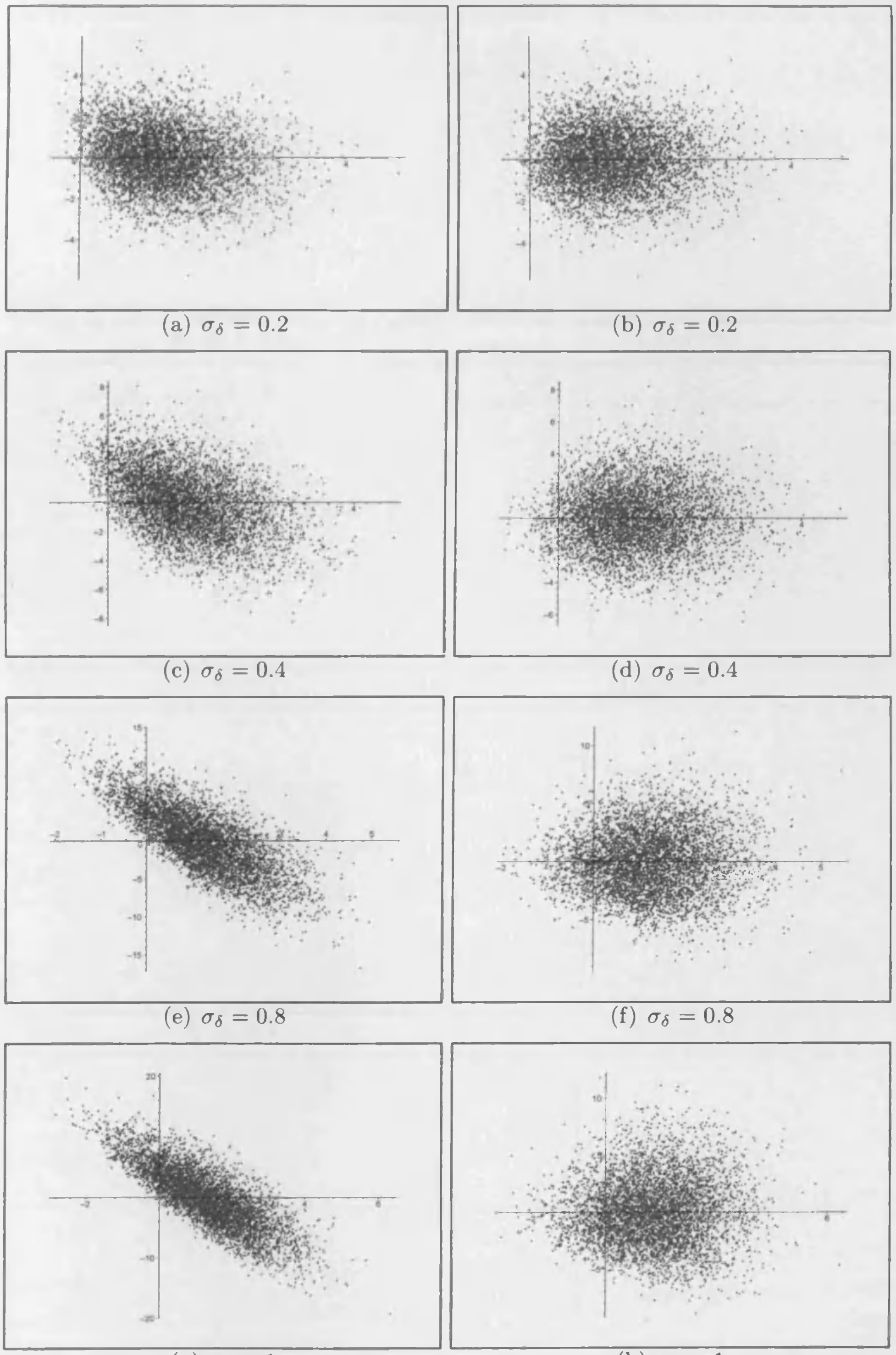
(a) $\sigma_\delta = 0.2$ (b) $\sigma_\delta = 0.2$

(c) $\sigma_\delta = 1$ (d) $\sigma_\delta = 1$

(e) $\sigma_\delta = 2$ (f) $\sigma_\delta = 2$

(g) $\sigma_\delta = 4$ (h) $\sigma_\delta = 4$

Figure 7.10: Vertical residual versus observed x, uniform ξ and Normal errors. Vertical residual from errors in variables fit is on the left hand side, vertical residual from least squares fit is on the right hand side.



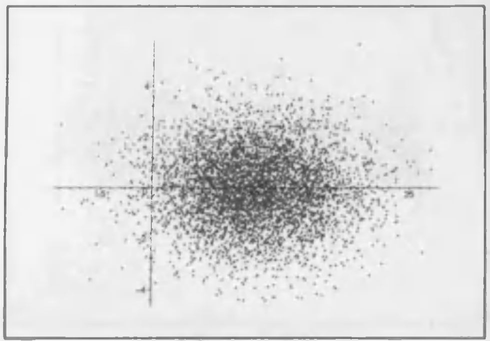
(a) $\sigma_\delta = 0.2$ (b) $\sigma_\delta = 0.2$

(c) $\sigma_\delta = 0.4$ (d) $\sigma_\delta = 0.4$

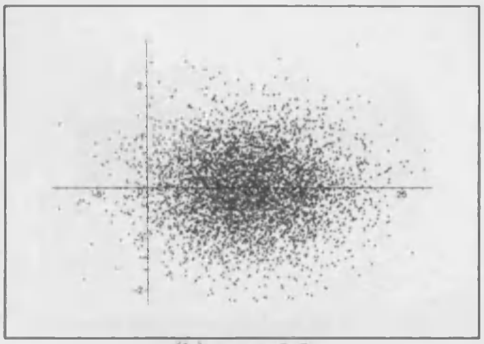
(e) $\sigma_\delta = 0.8$ (f) $\sigma_\delta = 0.8$

(g) $\sigma_\delta = 1$ (h) $\sigma_\delta = 1$

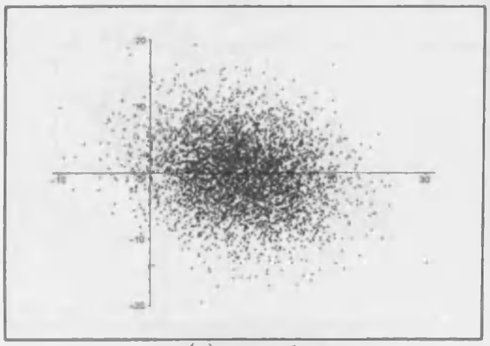
Figure 7.11: Vertical residual versus observed x, chi ξ (two degrees of freedom) and Normal errors. Vertical residual from errors in variables fit is on the left hand side, vertical residual from least squares fit is on the right hand side.



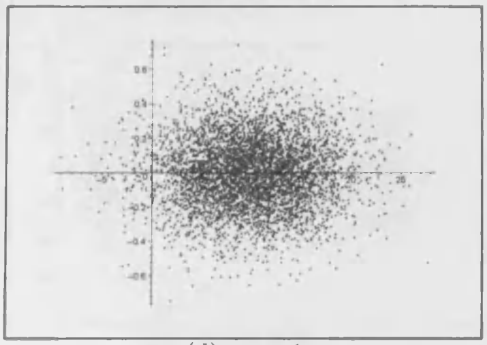
(a) $\sigma_\delta = 0.2$



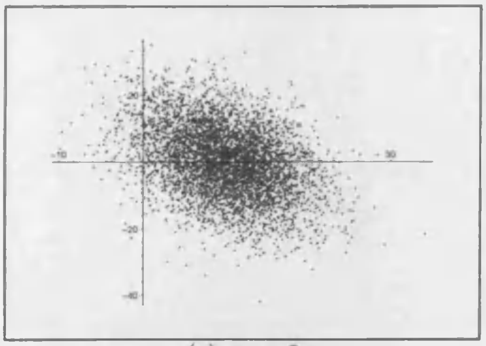
(b) $\sigma_\delta = 0.2$



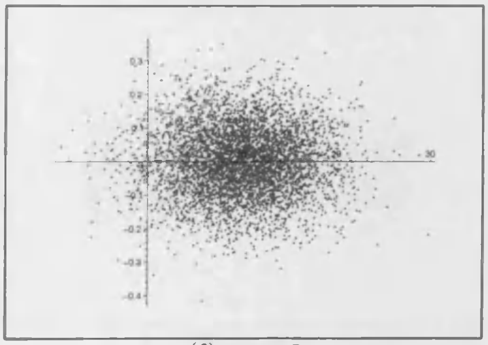
(c) $\sigma_\delta = 1$



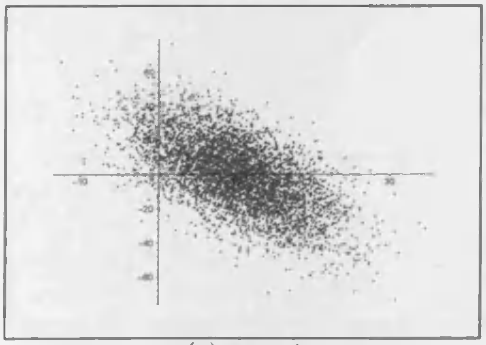
(d) $\sigma_\delta = 1$



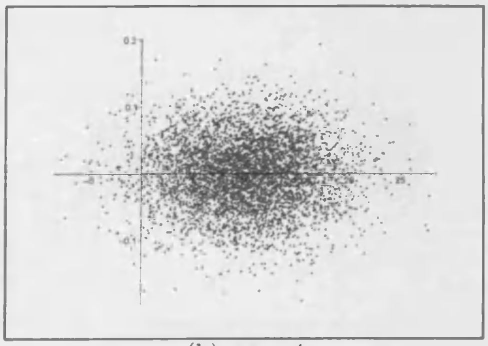
(e) $\sigma_\delta = 2$



(f) $\sigma_\delta = 2$

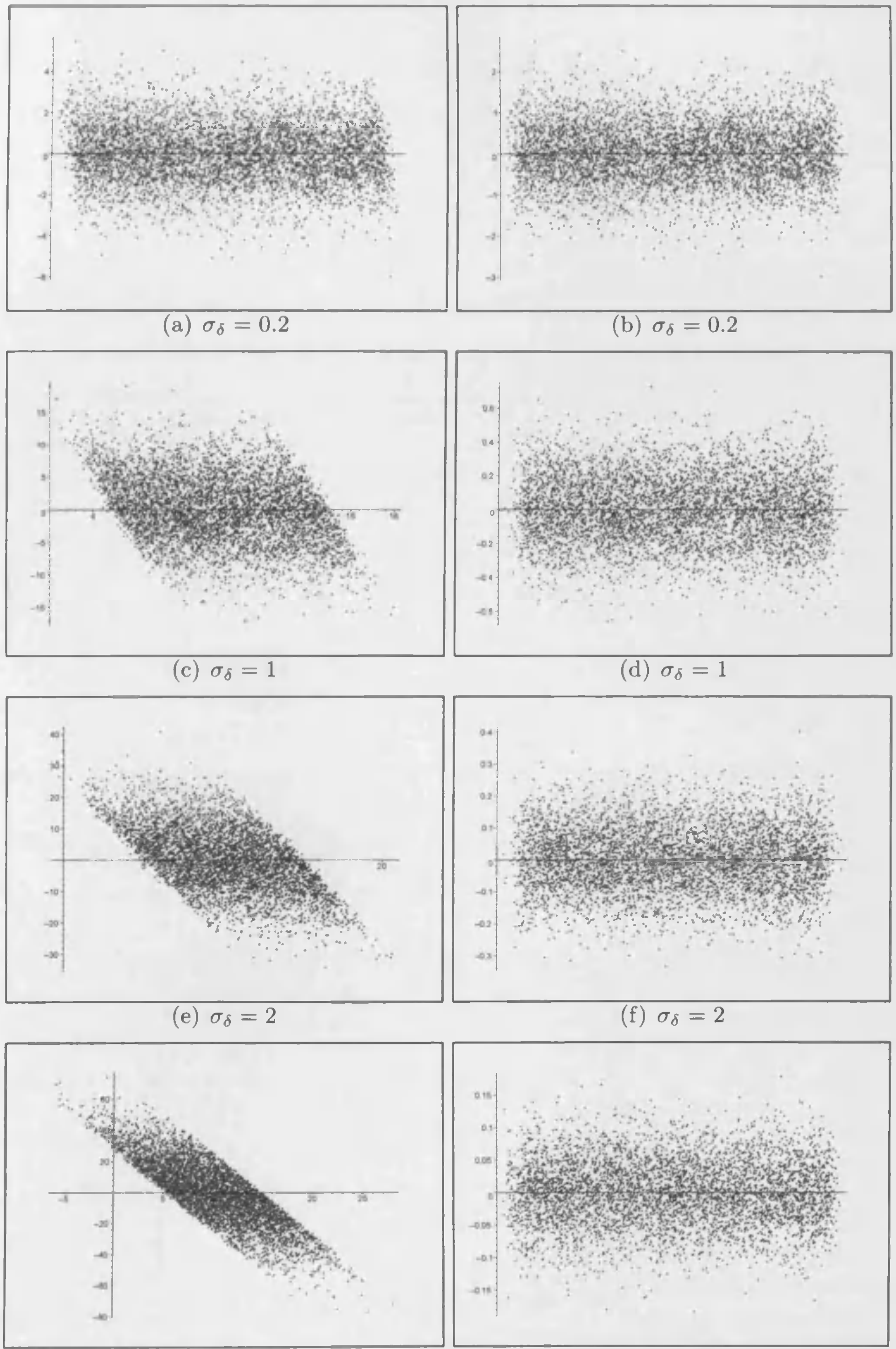


(g) $\sigma_\delta = 4$



(h) $\sigma_\delta = 4$

Figure 7.12: Vertical residual versus observed x, Normal structural model. Vertical residual from errors in variables fit against x is on the left hand side, vertical residual from least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$ against $\tilde{\xi}$ is on the right hand side.



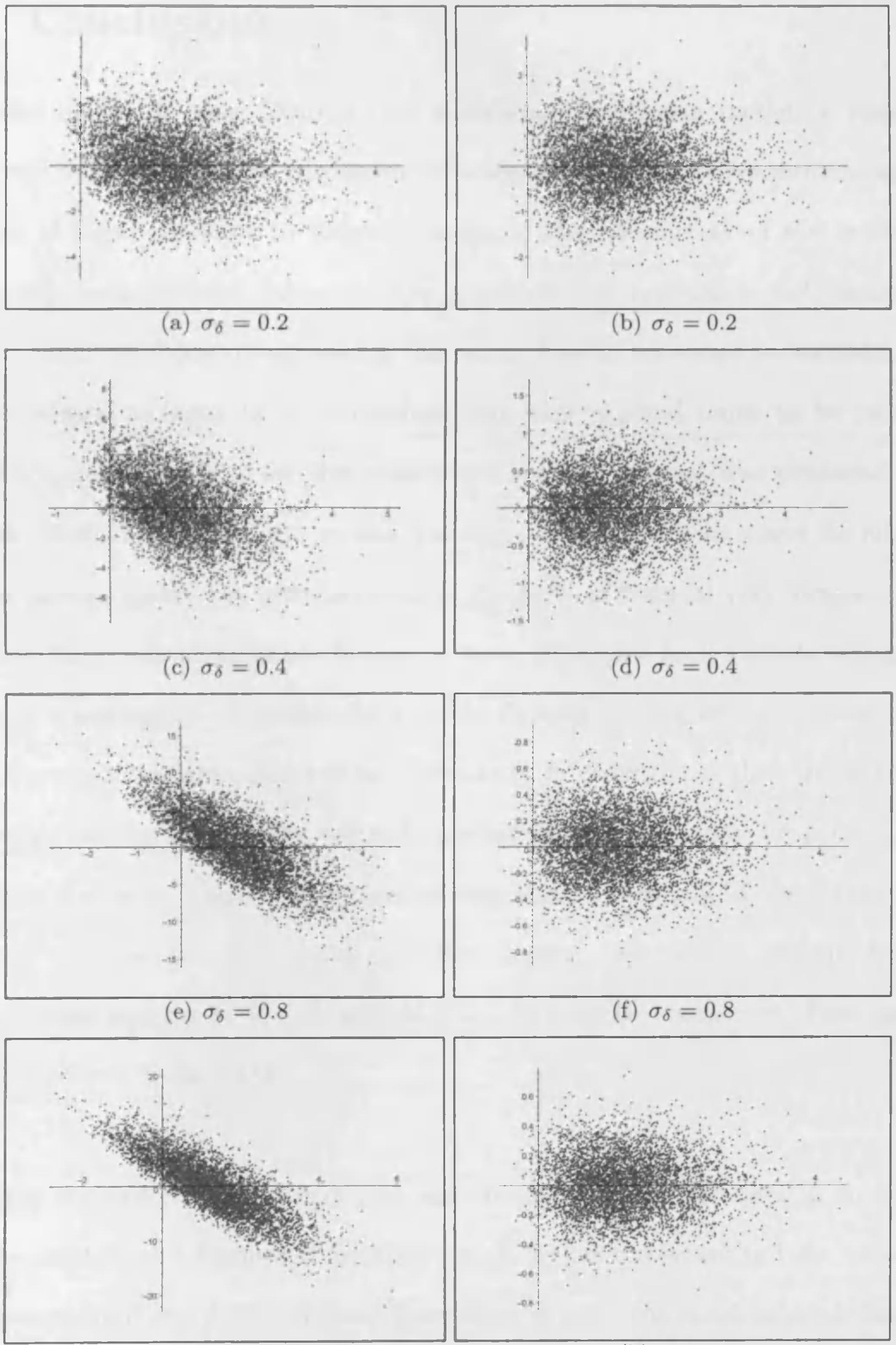
(a) $\sigma_\delta = 0.2$ (b) $\sigma_\delta = 0.2$

(c) $\sigma_\delta = 1$ (d) $\sigma_\delta = 1$

(e) $\sigma_\delta = 2$ (f) $\sigma_\delta = 2$

(g) $\sigma_\delta = 4$ (h) $\sigma_\delta = 4$

Figure 7.13: Vertical residual versus observed x, uniform ξ and Normal errors. Vertical residual from errors in variables fit against x is on the left hand side, vertical residual from least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$ against $\tilde{\xi}$ is on the right hand side.



(a) $\sigma_\delta = 0.2$ (b) $\sigma_\delta = 0.2$

(c) $\sigma_\delta = 0.4$ (d) $\sigma_\delta = 0.4$

(e) $\sigma_\delta = 0.8$ (f) $\sigma_\delta = 0.8$

(g) $\sigma_\delta = 1$ (h) $\sigma_\delta = 1$

Figure 7.14: Vertical residual versus observed x, chi ξ with two degrees of freedom and Normal errors. Vertical residual from errors in variables fit against x is on the left hand side, vertical residual from least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$ against $\tilde{\xi}$ is on the right hand side.

7.6 Conclusions

As stated earlier in this Chapter, for a standard regression model, a residual is intuitively defined. Modern regression techniques typically involve minimising some function of these residuals in order to estimate the parameters of the model. By introducing measurement error into the x variable, a residual is not immediately defined. Some methods of estimating the parameters of an errors in variables model may be related to some form of residual, and this residual tends to be related to the vertical residual. It is for this reason the vertical residual was exploited in this Chapter. Both the fact that the vertical residual is correlated with x and the migration effect is present makes the interpretation of the vertical residual very dangerous. It is seen that the errors in variables fit can in some situations look visually wrong when drawn on a scatterplot of points. As a result directly looking at the vertical residual from an errors in variables line will be misleading. In order to use the vertical residual, two things can be done. The vertical residual from $E[y|x]$ may be considered, as the $E[y|x]$ seems to follow the pattern of migration. Secondly, if the latent $\xi's$ are estimated from an initial errors in variables fit then the vertical residual from the standard least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$ may be considered. Both methods have been shown to be viable.

The topic of migration is a crucial one and the migration of the data in an errors in variables model has a number of implications. It is thus important to fully understand this phenomena if one is to understand the ethos of errors in variables modelling. It is hoped that the numerous explanations given in this Chapter would enable a practitioner to cope with the migration phenomenon.

Chapter 8

Case Studies and Examples

8.1 Alpha Foeto Protein as a Marker for Down's Syndrome

Down's syndrome is an example of a genetic disorder, which is estimated to have an incidence of 1 per 800 births (see for example Selikowitz [93] and the references therein). The disorder however is not only seen in humans, it has been noted in chimpanzees and mice. Down's syndrome is caused by the presence (either in whole, or in part) of an extra twenty-first chromosome, and is typically associated with both physical and cognitive impairments. Examples of the physical impairments include an almond shape to the eyes, shorter limbs and pure muscle tone, whilst cognitive impairments are mainly associated with mild to moderate learning difficulties. The probability of conceiving a child with Down's syndrome increases with maternal age.

In general, pregnant women may receive a number of prenatal screens. Many of the standard screens can aid with the diagnosis of whether the unborn child is likely to have Down's syndrome. The selection of available screens may be split into examples of invasive and non-invasive screens. Examples of invasive screening include amniocentesis (a small amount of amniotic fluid is taken from the amniotic sac surrounding the fetus, and analysed) and chorionic villus sampling (a sample of placental tissue is obtained, and tested). Both of these procedures however do carry some small risk of disrupting the fetus, thus causing potential complications.

An example of a non-invasive screening method is the measurement of maternal serum alpha foeto protein (AFP) levels. It is known that AFP levels are markers for Down's syndrome, low values generally being associated with the condition. The level of AFP varies with gestational age, and with the health status of the foetus (see for example Koduah [68]).

The motivation for the use of errors in variables methodology for the use of AFP is clear. There is inherent measurement error in the measurement of gestational age and AFP level. Indeed, Selikowitz has stated that one cause of false positives can be incorrect date of pregnancy. Thus the measurement of gestational age is crucial, and a model that can take into account the error inherent in the measurement of gestational age is desirable.

Figure 8.1 contains a typical scatterplot of the natural logarithm of AFP against gestational age in days. This particular data set was analysed in detail by Koduah [68]. The usual screening range for AFP is 15 to 18 weeks, and it is known that the standard deviation for the measurements of gestational age is approximately 2.1 days if measured in days, or is approximately 3.4 days if measured in weeks (see references in Koduah [68]). In the notation of the model used in this thesis then, this suggests that $\sigma_\delta = 2.1$. This information concerning the error variance is enough to compute an errors in variables fit to the scatter of data. The slope estimator $\tilde{\beta}_2$ assumes that the error variance σ_δ^2 is known.

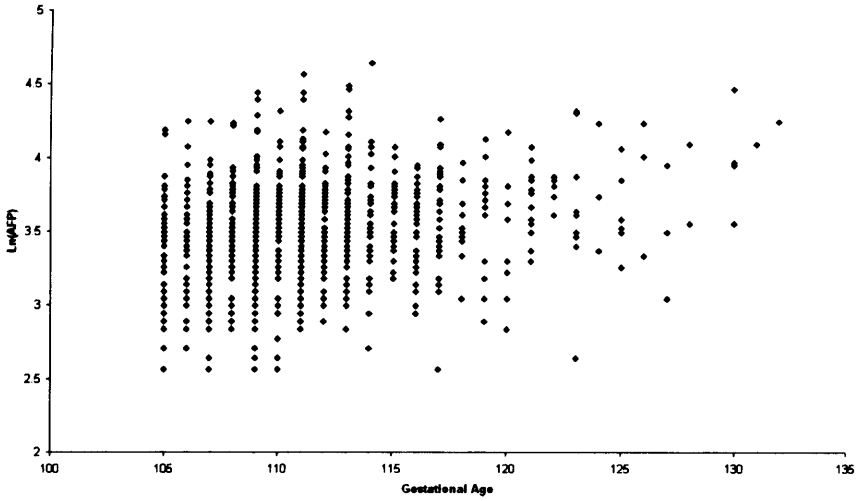


Figure 8.1: Measurement of the natural logarithm of AFP against gestational age (in days).

The following table shows values for the estimated slopes and intercepts via x on y regression, y on x regression, and $\tilde{\beta}_2$. It can be seen that $\tilde{\beta}_2$ does lie in between the values of the slope estimated by x on y and y on x regression, and so no admissibility conditions are broken. This implies that the estimators of the remaining unknown variance parameters σ^2 and σ_ε^2 are positive. It can be seen that $\tilde{\beta}_2$ does align more closely with the estimated slope from y on x regression. Since

$$\tilde{\sigma}^2 = \frac{s_{xy}}{\tilde{\beta}_2} = 18.67873$$

the estimated reliability ratio $\tilde{\kappa}$ for this situation is given by,

$$\tilde{\kappa} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2 + \sigma_\delta^2} = 0.79088.$$

As this is close to 1, then it would suggest that the errors in variables estimator of the slope would align closely with the slope estimated by y on x regression.

Estimator	Estimated Slope	Estimated Intercept
x on y	0.27804	-27.54081
y on x	0.01886	1.38256
$\tilde{\beta}_2$	0.02332	0.88557

Figure 8.2 contains Figure 8.1 with the regression fits described above placed on the scatter of points. The remaining parameters with their estimated values from using

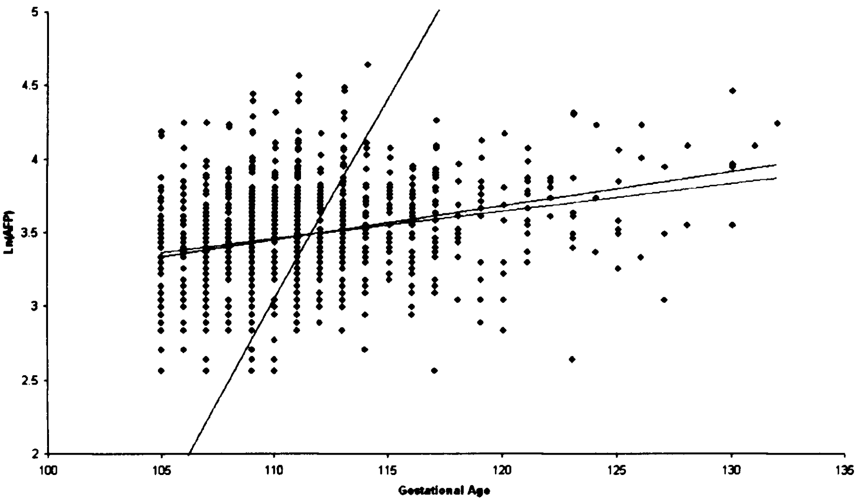


Figure 8.2: Measurement of the natural logarithm of AFP against gestational age (in days), with different regression fits. x on y regression fit is in blue, y on x regression fit is in red, and the errors in variables fit is in green.

the solutions to the equations (3.1) to (3.5) are:

$$\tilde{\mu} = 111.597$$

$$\tilde{\sigma}^2 = 18.67873$$

$$\tilde{\sigma}_\varepsilon^2 = 0.11094$$

The reliability ratio for the natural logarithm of the AFP measurement is estimated as

$$\frac{\tilde{\beta_2}^2 \tilde{\sigma}^2}{\tilde{\beta_2}^2 \tilde{\sigma}^2 + \tilde{\sigma}_\varepsilon^2} = 0.08386$$

and it is thus noted that σ_ε^2 is rather large. The range of $\ln(AFP)$ values at any gestational age is approximately 1.3 but the overall range is only approximately 2.1.

The slope in this example is also very shallow. It is unlikely that measurements of $\ln(AFP)$ will have such a large error variance associated with them, and so presumably there is equation error present, as described in Chapter 3. In other words there must be considerable natural variation in the $\ln(AFP)$ levels of pregnant women. As stated in Chapter 3, the problem of equation error in fitting an errors in variables model is avoided by using an estimator of the slope which does not assume anything concerning the error variance σ_ε^2. As knowledge of the variability in the measurement of gestational age was assumed, the inflated value for σ_ε^2 has no effect upon the estimation of β using $\tilde{\beta}_2$.

Once the parameters of the errors in variables model have been estimated, then it is possible to estimate the true gestational age, and true level of $\ln(AFP)$. As derived in Chapter 6, equation (6.13)

$$\tilde{\xi}_i = \frac{\lambda}{\lambda + \tilde{\beta}^2} x_i + \frac{\tilde{\beta}}{\lambda + \tilde{\beta}^2} (y_i - \tilde{\alpha}),$$

may be used to estimate the true gestational age. The true level of $\ln(AFP)$ may then be estimated using the relation $\tilde{\eta}_i = \tilde{\alpha} + \tilde{\beta}\tilde{\xi}_i$.

Figure 8.3 contains a scatterplot of the estimated true values of gestational age against the observed values of gestational age. The line on the plot is the $y = x$ line. As there is close agreement between the observed gestational age and the estimated true value of gestational age, the points are closely scattered about this line. For this application $\tilde{\xi}_i = 0.97885x_i + 0.90727(y_i - 0.88557)$. Figure 8.3 shows that in general there is close agreement between the observed gestational age and the estimated true gestational age. This is because the reliability ratio of the gestational age is quite large. The observed values of $\ln(AFP)$ are slightly adjusted to estimate the true gestational age, and this adjustment is more marked in the left hand side of the data. Due to the scaling of the data, it is seen that the x_i measurement has more of an influence upon the estimation of the true value of $\ln(AFP)$.

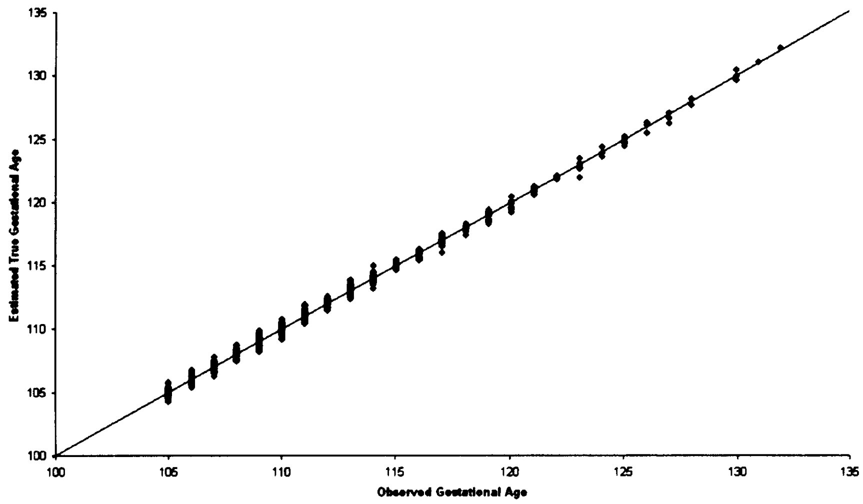


Figure 8.3: Estimated true values of gestational age against observed values of gestational age.

Figure 8.4 contains a scatterplot of the estimated true values of $\ln(AFP)$ against the observed values of $\ln(AFP)$. Again the line on the plot is the $y = x$ line. As can be viewed from the scatter about the line $y = x$, there is not as much of a close agreement between observed and estimated values of $\ln(AFP)$ as was demonstrated in the previous Figure. This is to be expected due to the large variation observed in values of $\ln(AFP)$. From looking at the scatterplot, there is more adjustment of the large observed values of $\ln(AFP)$ than the smaller observed values. In addition, a whole range of different true values of $\ln(AFP)$ is given for the same observed $\ln(AFP)$. For example, for an observed $\ln(AFP)$ level of 2.6068, there are 14 different true values of $\ln(AFP)$ ranging from 3.338 to 3.909. The reason for this is that both the x_i and y_i are used in (6.13). For this particular application, $\tilde{\eta}_i = -0.01873 + 0.02282x_i + 0.021154165y_i$

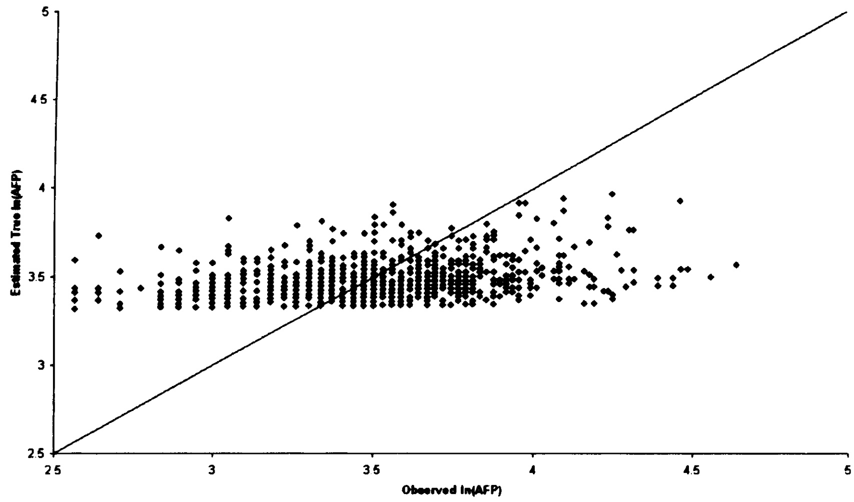


Figure 8.4: Estimated true values of $\ln(AFP)$ against observed values of $\ln(AFP)$.

Gestational age is measured in days however, and so this should be taken into account when estimating the true values of gestational age and $\ln(AFP)$. Figure 8.5 contains a scatterplot of the estimated true values of gestational age rounded to the nearest day, against observed values of gestational age. The line on the scatterplot is the $y = x$ line. It can be seen that in general there is close agreement between the estimated true values of gestational age, and observed gestational age. Many of the estimated true gestational ages match with the observed gestational ages. If they do not match, the estimated true gestational age differs from the observed gestational age by a day. This is closely related to the grouping methods of fitting a straight line discussed in Chapter 2. If the spacings between the x observations are appreciable, then the effects of measurement error in the x observations may be alleviated.

If the estimated true values of gestational age rounded to the nearest day are used to estimate true levels of $\ln(AFP)$, the scatterplot of estimated true levels of $\ln(AFP)$

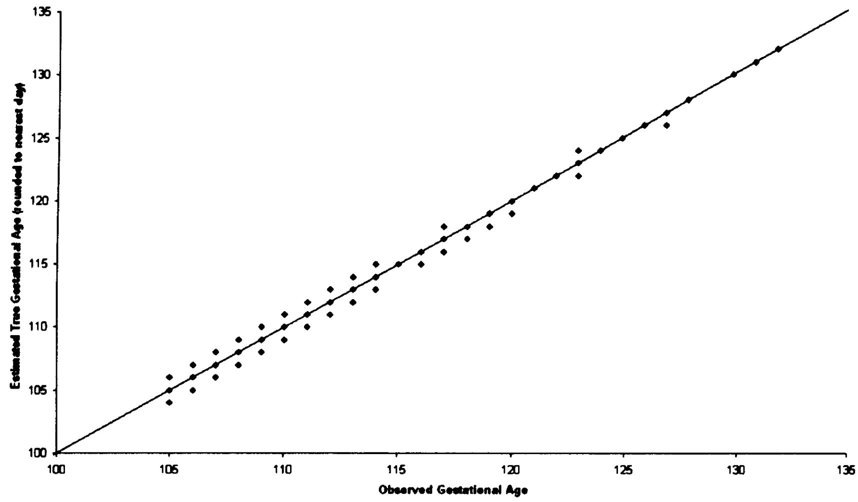


Figure 8.5: Estimated true values of gestational age rounded to the nearest day, against observed values of gestational age.

against observed levels of ln(AFP) becomes that of Figure 8.6. The picture is very similar to Figure 8.4 as there is not too great a difference in the unrounded and rounded gestational age. To demonstrate this Figure 8.7 contains a scatterplot of estimated levels of ln(AFP) when unrounded estimated true values of gestational age (blue diamond) and rounded estimated values of gestational age (red triangle) against observed values of ln(AFP). It can be seen that there is very little difference in the estimated true levels of ln(AFP). There is a small amount of disagreement in the tails of the observed ln(AFP), but close to the $y = x$ line, the observed ln(AFP) and the estimated true levels of ln(AFP) using both rounded and unrounded estimated true gestational age are indistinguishable.

An important topic in terms of screening is residuals. Examination of residuals will allow the identification of pregnant women with noticeably small or large values of ln(AFP). The correct identification of these women is a crucial aim for any screening procedure. As stated in Chapter 7, careful attention should be made when analysing

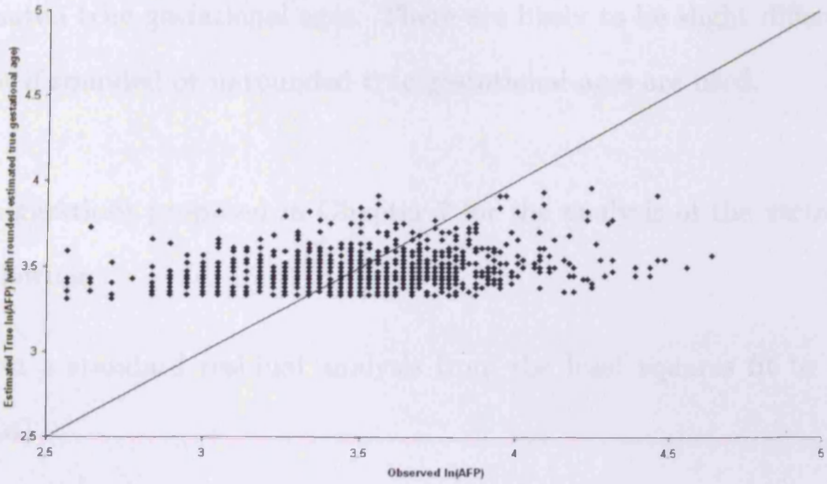


Figure 8.6: Estimated true values of $\ln(AFP)$, against observed values of $\ln(AFP)$, using rounded estimated true gestational age.

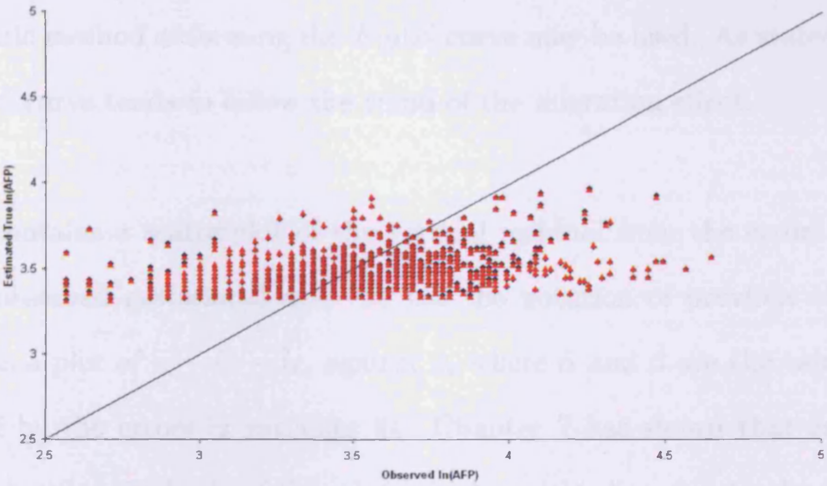


Figure 8.7: Estimated true values of $\ln(AFP)$, against observed values of $\ln(AFP)$, using unrounded and rounded estimated true gestational age.

a vertical residual, as the trend of migration is likely to distort the typical vertical residual scatterplot. For this application, it is difficult to assess migration directly as the distribution of true gestational age is not known. However, from inspection of Figure 8.1 it does seem that the measurement of gestational age is skewed towards the left hand tail. An additional difficulty with this application is in the treatment of the estimated true gestational ages. There are likely to be slight differences in the residual plot if rounded or unrounded true gestational ages are used.

The main suggestions proposed in Chapter 7 for the analysis of the vertical residuals were the following:

1. Perform a standard residual analysis from the least squares fit to $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$.
2. Consider the vertical residual from the $E[y|x]$ curve.

As the distribution of the true gestational age is not known, the Nadaraya-Watson nonparametric method of forming the $E[y|x]$ curve may be used. As stated in Chapter 7, the $E[y|x]$ curve tends to follow the trend of the migration effect.

Figure 8.8 contains a scatterplot of the vertical residual from the errors in variables fit against observed gestational age. To use the notation of previous chapters, the scatterplot is a plot of $y_i - \tilde{\alpha} - \tilde{\beta}x_i$ against x_i where $\tilde{\alpha}$ and $\tilde{\beta}$ are the values of α and β estimated by the errors in variables fit. Chapter 7 has shown that analysing the vertical residual on the basis of this plot may be misleading due to the migration of the observed data from the true values. This scatterplot does highlight some pregnant women with particularly high levels of observed $\ln(AFP)$, namely those who presented themselves for screening at an observed gestational age between 105 days and 115 days. There are some women with noticeably low levels of $\ln(AFP)$, and these women are those that presented for screening after a gestational age of around 120 days.

Figure 8.9 contains scatterplots of the vertical residual from the least squares fit to the

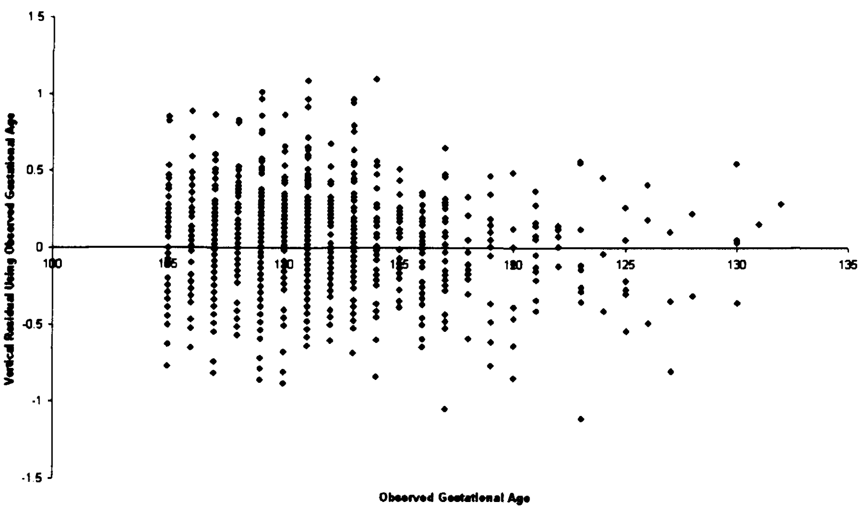
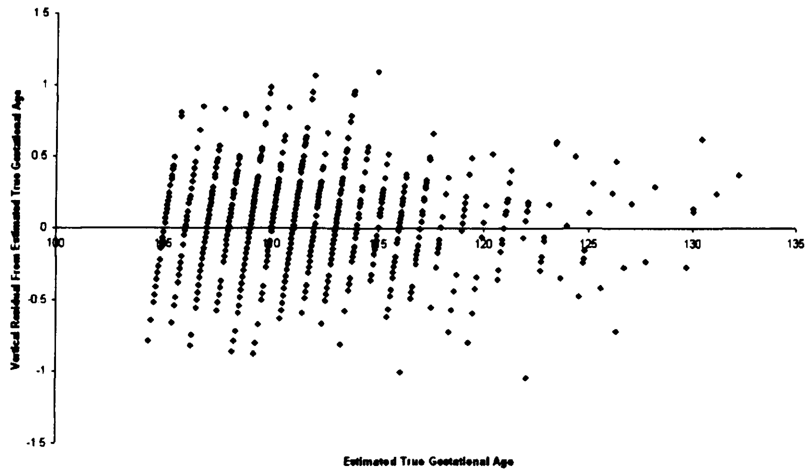
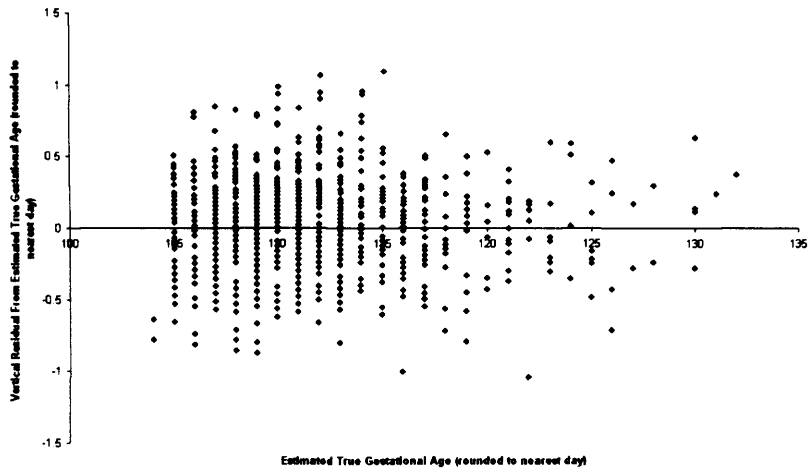


Figure 8.8: Scatterplot of vertical residual from the errors in variables fit against observed gestational age.

estimated true gestational age and observed levels of $\ln(AFP)$, when the estimated true gestational age is both unrounded, and rounded to the nearest day. The general effect of not using the rounded estimated true value of gestational age is to have a twisted residual plot. This is because even though gestational age is measured in days, the estimated true gestational age is on a purely continuous scale. Thus around each day of gestation, there will be some spread of estimated gestational age. This could be misleading, and the twisting effect is removed by rounding to the nearest day. The twisting effect is also added to because of the use of y_i in the estimation of ξ_i, with a greater adjustment for the more extreme y's. The effect is clearly evident in this application due to the multiple values of $\ln(AFP)$ at the same gestational age. This figure demonstrates however, that both residual plots do share the main characteristics and features. Despite this, these residual plots are not recommended for practical use as they have not been corrected for the attenuation of the observed data points that is caused by migration.



(a) Unrounded estimated true gestational age



(b) Rounded estimated true gestational age

Figure 8.9: Scatterplots of the vertical residual from the least squares fit to the estimated true gestational age and observed values of $\ln(AFP)$.

Figure 8.10 contains the scatterplot of the vertical residual from the errors in variables fit against observed gestational age (blue diamond) with a scatterplot of the vertical residual from the least squares fit to the estimated true gestational age and observed levels of $\ln(AFP)$, when the estimated true gestational age is rounded to the nearest day (red square). As can be seen, there are some noticeable discrepancies between the plots. For example, at a gestational age of 105 days, the vertical residuals from the least squares fit are smaller than the vertical residuals from the errors in variables fit. On the other hand, at 116 days, there is an extreme vertical residual from the least squares fit, that is not present with the vertical residual from the errors in variables fit. Therefore there are different conclusions drawn from both residual plots. As stated earlier, the ethos of this application is to highlight women with extreme values of $\ln(AFP)$. The use of the correct residual plot to identify these women is crucial.

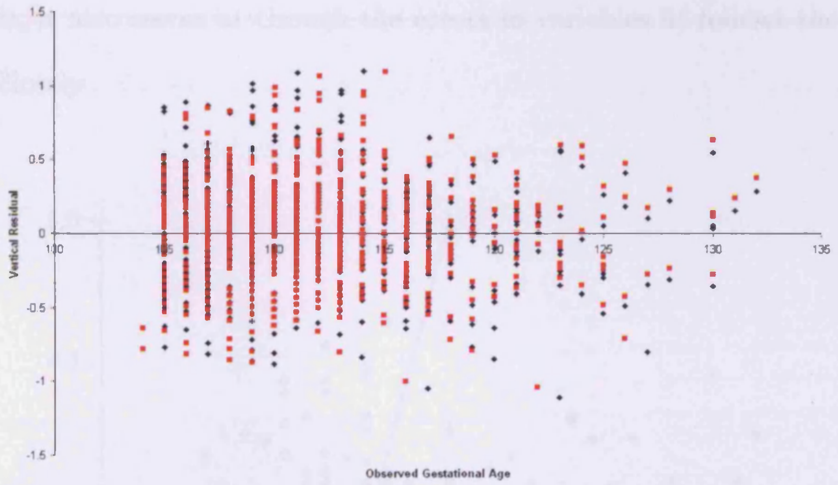


Figure 8.10: Scatterplots of different vertical residuals against observed and estimated gestational age.

As written earlier, an alternative way to look at residuals is to take the vertical residual from the $E[y|x]$ curve. The effect of migration may then be nullified. Figure 8.11 shows the Nadaraya-Watson fit to the data presented in Figure 8.1. The Nadaraya-Watson estimate to $E[y|x]$ is computed as

$$E[y|x] \approx \widehat{m(x)} = \frac{\frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)y_i}{\frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)} \tag{8.1}$$

and h is chosen in accordance with

$$h = \sigma_\delta\left[\log(n)\right]^{-\frac{1}{2}}, \tag{8.2}$$

as discussed in Chapter 6. It is noted that the Nadaraya-Watson fit is not a straight line, but is a curve. As stated by Lindley [72] $E[y|x]$ is only a straight line for the Normal structural model. The Nadaraya-Watson fit is linear over a large range of gestational ages, but does curl up at the right hand tail. This phenomenon is similar to what was experienced in Chapter 6, when the random variable ξ was assumed to follow a skew distribution. It seems that in general, the y on x regression fit follows the Nadaraya-Watson fit closely. But as the errors in variables fit is close to the y on x regression fit, it also seems at though the errors in variables fit follows the Nadaraya-Watson fit closely.

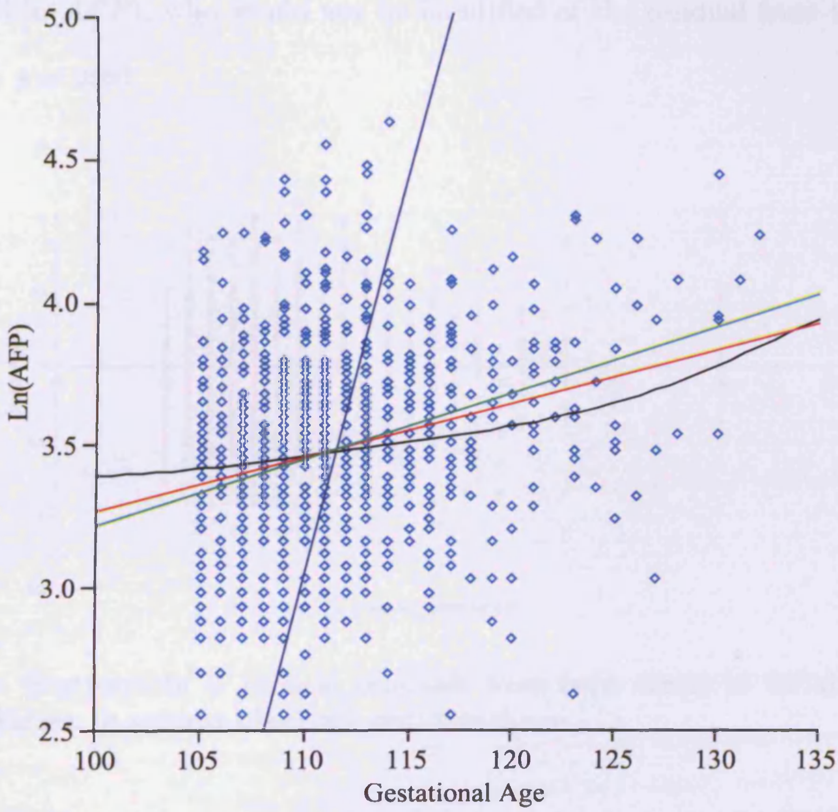


Figure 8.11: Nadaraya-Watson fit to data, with other regression fits. Nadaraya-Watson fit is in black, y on x regression fit is in red, x on y regression fit is in blue and the errors in variables regression fit (using $\tilde{\beta}_2$) is in green.

Figure 8.12 contains the scatterplot of the vertical residual from the errors in variables fit against observed gestational age (blue diamond) with a scatterplot of the vertical residual from the Nadaraya-Watson fit to $E[y|x]$ (purple square). There is close agreement between the two residual plots for small values of gestational age. There is some disagreement however at the right hand tail of the data. For the larger values of observed gestational age there is more of a disagreement as for these values the Nadaraya-Watson and errors in variables fit differ the most. The residuals at the right hand tail from the errors in variables fit are shifted upwards to create the residuals from the Nadaraya-Watson fit. The residuals from the Nadaraya-Watson fit highlight some individuals at the right hand tail of the data who may be considered to have a high level of $\ln(AFP)$, who would not be identified of the residual from the errors in variables fit was used.

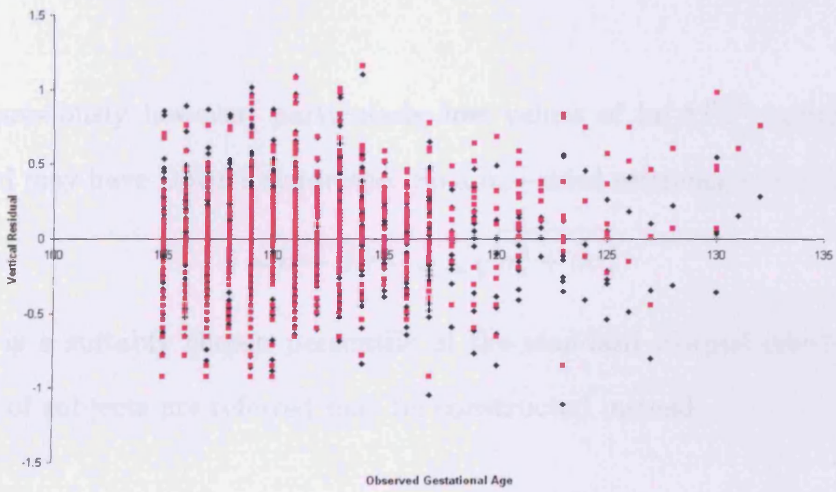


Figure 8.12: Scatterplots of vertical residuals from both errors in variables fit, and Nadaraya-Watson fit against observed gestational age.

An additional aid in the referral of subjects is to construct a reference interval about a chosen baseline. Work on the construction of reference intervals when there are errors in both variables had been conducted by Koduah [68] and thus will not be repeated in detail here. The typical methodology in practise is to use standard regression techniques to estimate the baseline, and then using knowledge of the error variance in the y measurement, draw a reference band consisting of two parallel lines either side of the baseline. These lines are typically constructed to exclude 2.5% in each extreme.

For many applications, as is the case for the $\ln(AFP)$ data, there will be error in both the x and y measurement. Thus standard regression techniques will not give an unbiased estimate of the baseline. Koduah derived the reference interval for an errors in variables model and it may be estimated as

$$\tilde{y} = \tilde{\alpha} + \tilde{\beta} \pm z_{1-\frac{p}{2}}\sqrt{\tilde{\sigma_\varepsilon^2} + \tilde{\beta}\tilde{\sigma_\delta^2}}$$

where $z_{1-\frac{p}{2}}$ is a suitably chosen percentile of the standard Normal distribution such that $100p\%$ of subjects are referred.

As stated previously however, particularly low values of $\ln(AFP)$ suggest that the unborn child may have Down's syndrome. So a one-sided reference interval of the form

$$\tilde{y} = \tilde{\alpha} + \tilde{\beta}x - z_{1-p}\sqrt{\tilde{\sigma_\varepsilon^2} + \tilde{\beta}\tilde{\sigma_\delta^2}}$$

where z_{1-p} is a suitably chosen percentile of the standard Normal distribution such that $100p\%$ of subjects are referred may be constructed instead.

If the usual methodology of fitting a line using standard regression techniques is used, then the corresponding one-sided reference interval is of the form

$$\tilde{y} = \tilde{\alpha_0} + \tilde{\beta}_0 x - z_{1-p}\sqrt{\tilde{\sigma_\varepsilon^2}}$$

where z_{1-p} is a suitably chosen percentile of the standard Normal distribution such that $100p\%$ of subjects are referred, $\tilde{\alpha_0}$ is the least squares estimator of the intercept and $\tilde{\beta}_0$ is the least squares estimator of the slope.

Figure 8.13 contains a scatterplot of the $\ln(AFP)$ data, with y on x and errors in variables regression fits. The one sided reference interval so that 5% of women are referred are also plotted. It can be seen that at the left hand side of the data, different conclusions would be drawn depending on the choice of baseline. The errors in variables line and least squares line are very similar in the main body of the data. The least squares line however lowers coverage in the right hand tail of the data. In terms of the differing results depending on the choice of baseline, Koduah states

> "The observation here reinforces the recommendation made already that the ordinary least squares procedure must not be applied when it is clear that there are measurement errors in the x variable, in this case estimated gestational age."

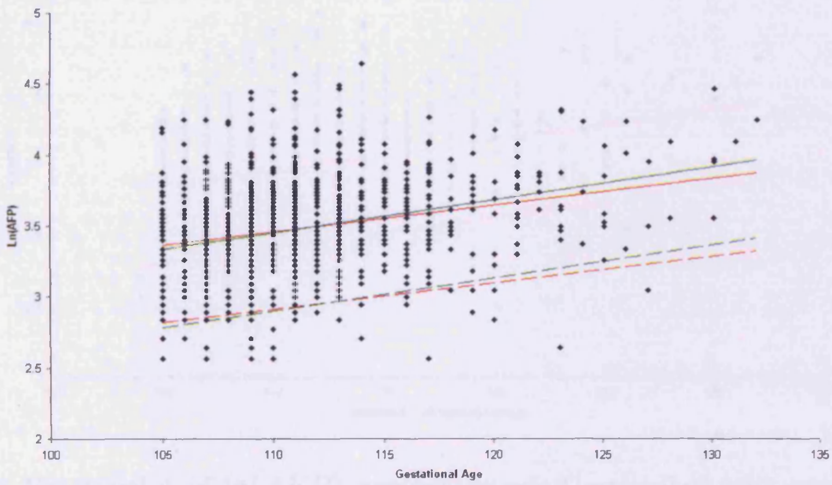


Figure 8.13: Scatterplot of $\ln(AFP)$ data with one-sided reference interval. y on x regression fit is in red, errors in variables fit is in green. Reference interval from y on x line is in red with dashed line. Reference interval from errors in variables fir in in green with dashed line.

One may use standard regression techniques if the true gestational age has been estimated first. As stated earlier, this may be done using the method of moments estimator $\tilde{\xi}$ for the latent ξ. Then the one-sided reference interval may be computed as

$$\tilde{y} = \tilde{\alpha_0} + \tilde{\beta}_0 x - z_{1-p}\sqrt{\tilde{\sigma}_\varepsilon^2}$$

$\tilde{\alpha_0}$ is the least squares estimator of the intercept and $\tilde{\beta}_0$ is the least squares estimator of the slope for the data set $\{(\tilde{\xi}_i, y_i), i = 1, \ldots, n\}$. As an example of this, Figure 8.14 is a scatterplot of $\ln(AFP)$ against rounded estimated true gestational age with the appropriate one-sided reference interval. However, due to the uncertainty in the prediction of the latent true gestational age, it would seem that a reference interval to the data from an errors in variables fit seems the most reliable, and is simpler to implement.

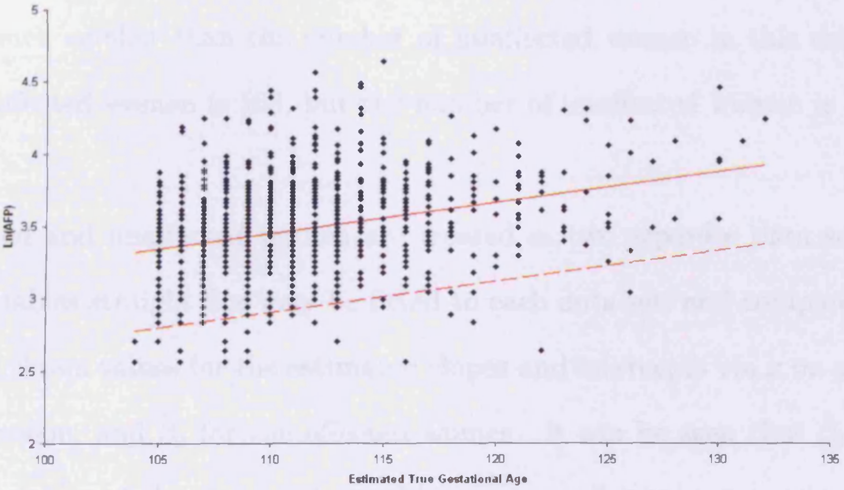


Figure 8.14: Scatterplot of $\ln(AFP)$ against rounded estimated true gestational age with one-sided reference interval. y on x regression fit is in red. Reference interval from y on x line is in red with dashed line.

8.2 Comparison of Affected and Unaffected in Down's Screening

This section is intended to be a small illustration of the comparison between affected (mothers were positively screened as having a Down's syndrome child) and unaffected (mothers were negatively screened as having a Down's syndrome child). Figure 8.15 is a scatterplot of affected women (purple square) gestational age against $\ln(AFP)$ and unaffected women (blue diamond) gestational age against $\ln(AFP)$. The original format of this data was to record gestational age in weeks, and so for consistency of presentation this has been changed to days. The effect of this is to increase the spacings between each stack of observations, at a given gestational age. As stated in the previous section, if gestational age is measured in weeks, then $\sigma_\delta = 3.4$ days. Presumably though, there is a loss of accuracy in measuring gestational age in weeks due to rounding. Due to the prevalence of Down's syndrome, the number of affected women is much smaller than the number of unaffected women in this data set. The number of affected women is 153, but the number of unaffected women is 7468.

If the affected and unaffected women are treated as two separate data sets, then an errors in variables straight line may be fitted to each data set, and compared. The following table shows values for the estimated slopes and intercepts via x on y regression, y on x regression, and $\tilde{\beta}_2$ for the affected women. It can be seen that $\tilde{\beta}_2$ does lie in between the values of the slope estimated by x on y and y on x regression, and so no admissibility conditions are broken. This implies that the estimators of the remaining unknown variance parameters σ^2 and σ_ε^2 are positive. It can be seen that $\tilde{\beta}_2$ does align more closely with the estimated slope from y on x regression. This is because the

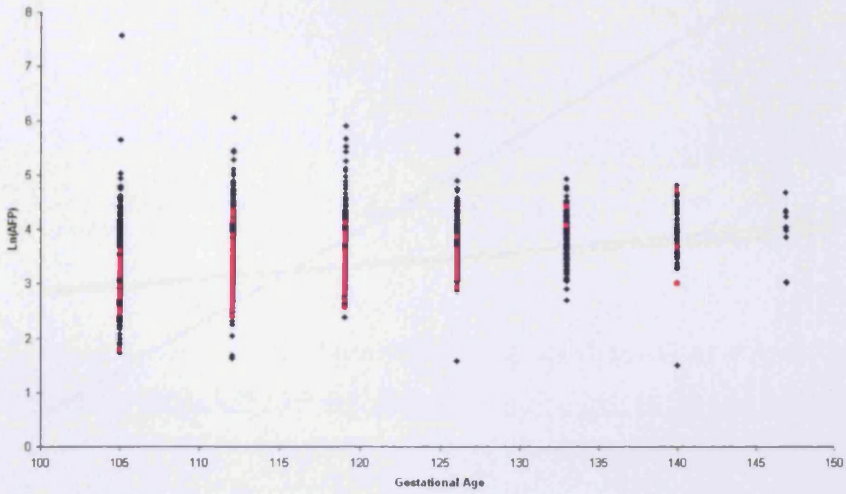


Figure 8.15: Measurement of the natural logarithm of AFP against gestational age (in days) for both affected and unaffected women.

reliability ratio is estimated as

$$\tilde{\kappa} = \frac{\tilde{\sigma^2}}{\tilde{\sigma^2} + \sigma_\delta^2} = 0.77185$$

which is quite large.

Estimator	Estimated Slope	Estimated Intercept
x on y	0.17396	-16.81631
y on x	0.02088	0.84740
$\tilde{\beta}_2$	0.02705	0.13534

The remaining parameters with their estimated values from using the solutions to the equations (3.1) to (3.5) are:

$$\tilde{\mu} = 115.3856$$

$$\tilde{\sigma^2} = 39.10829$$

$$\tilde{\sigma_\varepsilon^2} = 0.15541.$$

Figure 8.16 contains a scatterplot of gestational age against $\ln(AFP)$ for the affected women, with the regression fits as described in the previous table.

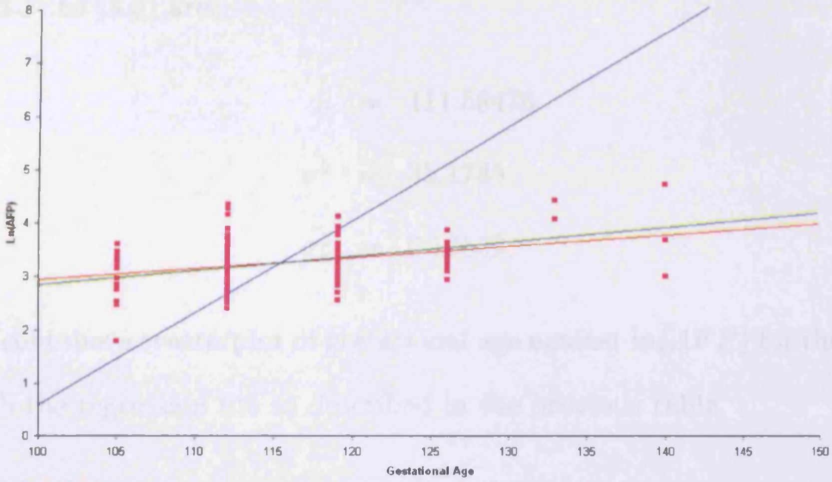


Figure 8.16: Measurement of the natural logarithm of AFP against gestational age (in days), with different regression fits for the affected women. x on y regression fit is in blue, y on x regression fit is in red, and the errors in variables fit is in green.

The following table shows values for the estimated slopes and intercepts via x on y regression, y on x regression, and $\tilde{\beta}_2$ for the unaffected women. It can be seen that $\tilde{\beta}_2$ does lie in between the values of the slope estimated by x on y and y on x regression, and so no admissibility conditions are broken. This implies that the estimators of the remaining unknown variance parameters σ^2 and σ_ε^2 are positive. It can be seen that $\tilde{\beta}_2$ does align more closely with the estimated slope from y on x regression. This is because the reliability ratio is estimated as

$$\tilde{\kappa} = \frac{\tilde{\sigma^2}}{\tilde{\sigma^2} + \sigma_\delta^2} = 0.75265$$

which is quite large.

Estimator	Estimated Slope	Estimated Intercept
x on y	0.18576	-17.27163
y on x	0.01788	1.46107
$\tilde{\beta}_2$	0.02376	0.80530

The remaining parameters with their estimated values from using the solutions to the equations (3.1) to (3.5) are:

$$\tilde{\mu} = 111.58476$$

$$\tilde{\sigma^2} = 35.1745$$

$$\tilde{\sigma^2_\varepsilon} = 0.13539.$$

Figure 8.17 contains a scatterplot of gestational age against $\ln(AFP)$ for the unaffected women, with the regression fits as described in the previous table.

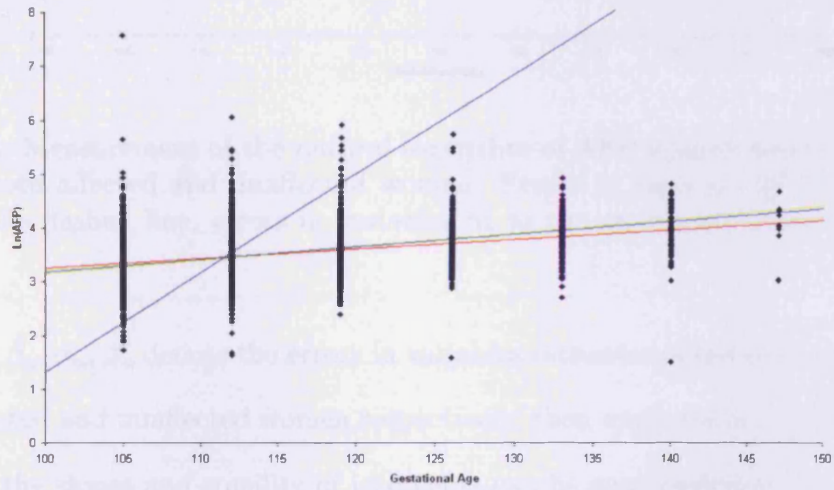


Figure 8.17: Measurement of the natural logarithm of AFP against gestational age (in days), with different regression fits for the unaffected women. x on y regression fit is in blue, y on x regression fit is in red, and the errors in variables fit is in green.

The estimates of the parameters for both the affected and unaffected women are quite similar, and this is to be expected. There is no reason why, for example, the variation of gestational ages observed is different for affected and unaffected women. Figure 8.18 contains the errors in variables fits for both the affected and unaffected women. As stated earlier, lower values of AFP are associated with Down's syndrome, and this is confirmed by the errors in variables fits shown in Figure 8.18. The lines appear to be approximately parallel, as the slope in both fits is approximately the same.

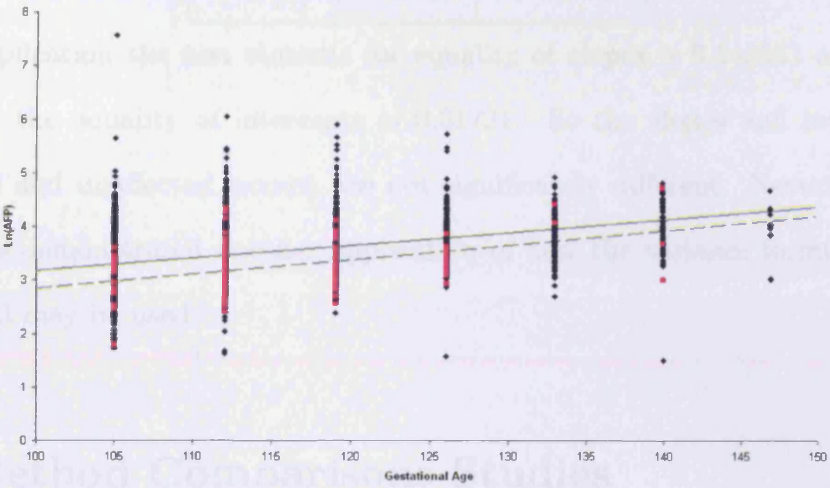


Figure 8.18: Measurement of the natural logarithm of AFP against gestational age (in days) for both affected and unaffected women. Errors in variables fit to the affected women is the dashed line, errors in variables fit to the unaffected women is the bold line.

Letting $\tilde{\alpha_a}$, $\tilde{\beta_a}$, $\tilde{\alpha_u}$, $\tilde{\beta_u}$ denote the errors in variables estimates of the intercept and slope for the affected and unaffected women respectively, then approximate test statistics for equality of the slopes and equality of intercepts can be constructed as

$$\frac{\tilde{\beta_u} - \tilde{\beta_a}}{\sqrt{Var[\tilde{\beta_u}] + Var[\tilde{\beta_a}]}}$$

for the slopes, and

$$\frac{\tilde{\alpha_u} - \tilde{\alpha_a}}{\sqrt{Var[\tilde{\alpha_u}] + Var[\tilde{\alpha_a}]}}$$

for the intercept. These may be compared to the percentage points of the Normal distribution to assess whether the slopes or intercepts are significantly different.

Using the formulae from Chapter 3, the following table contains the approximate variances of the slope and intercept for both the affected and unaffected women.

Variance	Affected	Unaffected
$Var[\tilde{\beta}_2]$	$\frac{0.00549}{153}$	$\frac{0.00542}{7468}$
$Var[\tilde{\alpha}]$	4.45167	0.00904

For this application the test statistic for equality of slopes is 0.543551 and the test statistic for the equality of intercepts is 0.31721. So the slopes and intercepts for the affected and unaffected women are not significantly different. Nevertheless, this example has demonstrated another application of how the variance formulae derived in Chapter 3 may be used.

8.3 Method Comparisons Studies

There are many examples of method comparisons studies in the literature. Essentially, these studies involve the comparison of two measurement techniques. By far the most common method of analysis is to perform y on x regression to quantify the relationship between the two methods, computation of the correlation coefficient, and a specification of the sample size. Method comparison studies are a key example of how the straight line errors in variables methodology developed and discussed in this thesis may be applied.

A number of authors have offered advice as to the statistical approach that should be adopted. Altman and Bland [8] have written a number of papers on method comparison studies. They deemed standard regression techniques inappropriate since errors in both variables attenuate the slope of the line. They propose, as a tool to investigate between method analysis, plotting $(y - x)$ against $(y + x)$. The differences of which may give information about bias and imprecision between methods. The authors do conclude that when the objective is to calibrate one method against the other, regression techniques which take into account errors in both variables may prove valuable.

An alternative statistical procedure was developed by Nix (pers. comm.). Consider the following construction of a method comparisons model. Let an individual be measured with true measurement μ. The two methods of measurement, x and y may be written:

$$\begin{aligned} x &= \mu + b_x(\mu) + \delta \\ y &= \mu + b_y(\mu) + \varepsilon \end{aligned}$$

where $b_x(\mu) = \alpha_x + \beta_x\mu$ and $b_y(\mu) = \alpha_y + \beta_y\mu$ are linear biases in the x and y methods respectively.

What then, is the interpretation of $y = \alpha + \beta x$?. What essentially does this mean in a method comparisons context?.

Since $x = \alpha_x + (\beta_x + 1)\mu + \delta$ and $y = \alpha_y + (\beta_y + 1)\mu + \varepsilon$, eliminating μ yields $(\beta_y + 1)(x - \alpha_x - \delta) = (\beta_x + 1)(y - \alpha_y - \varepsilon)$ and so

$$y = \alpha_y - \frac{(\beta_y + 1)}{(\beta_x + 1)}\alpha_x + \frac{(\beta_y + 1)}{(\beta_x + 1)}x + \left(\varepsilon - \frac{(\beta_y + 1)}{(\beta_x + 1)}\delta\right).$$

The above equation suggests an errors in variables regression form, with $\alpha = \alpha_y - \beta\alpha_x$ and $\beta = \frac{(\beta_y+1)}{(\beta_x+1)}$.

For this method comparisons, errors in variables regression will identify the relationship between the mean levels for each method, but will not identify the bias within each method. However, if one method (say x) is a gold standard, then $\alpha_x = 0$ and $\beta_x = 0$, and one can ultimately identify α_y and β_y. Typically, a new method is compared to a method which is known to be a gold standard for calibration of the new methods.

Altman and Bland would suggest plotting $D = (y - x)$ against $S = (y + x)$ and performing a regression if there appears to be a trend. Using the notation from above, we obtain

$$D = (\alpha_y - \alpha_x) - \left(\frac{(\beta_y - \beta_x)}{(\beta_y + \beta_x + 2)}\right)(\alpha_y + \alpha_x) \;+\; \frac{(\beta_y - \beta_x)}{(\beta_y + \beta_x + 2)}S$$
$$+ \left((\varepsilon - \delta) - \frac{(\beta_y - \beta_x)}{(\beta_y + \beta_x + 2)}(\varepsilon + \delta)\right)$$

Superficially, this is of errors in variables regression form

$$D = \alpha' + \beta' S + (\varepsilon' - \beta'\delta')$$

but with the added difficulty that δ' and ε' are not independent.

From the above relationship, unless both methods have the same relative bias ($\beta_y = \beta_x$), then there will always be a trend between x and y. Since $(y - x)$ against $(y + x)$ gives rise to a more complicated regression model, and seems not to provide any additional benefits over y against x, it would seem that the errors in variables approach is the way to proceed.

As an example of a method comparison study, and some of the questions raised during the statistical analysis of such a study, Figure 8.19 contains a scatterplot of measurements of AFP by an old and a new method (obtained from Nix, again pers. comm.). In other words, a new kit was compared with an old kit. This is the typical context of a method comparison study. The black line on the scatterplot is the $y = x$ line. It can be seen that there is close agreement between the two methods for small values of the old kit, but for larger values it appears that there is some disagreement. In general however, there is a good linear relationship across the range. It could be argued however that there is some heteroscedasticity present in the data. This point shall be dealt with later. As a lot of the data points lie underneath the $y = x$ line, it appears that the new kit gives a smaller measurement of AFP than the old kit. The data is skewed towards the smaller measurements of AFP, and there is one extreme point at the right hand tail of the data, which lies far away from the main body of points.

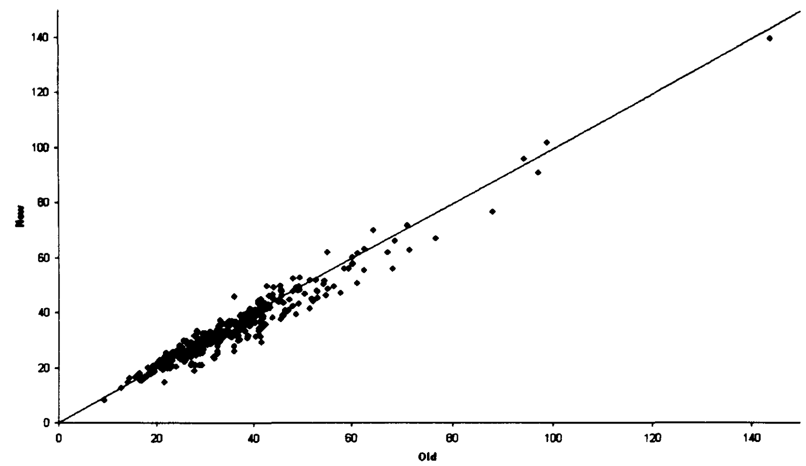


Figure 8.19: A typical method comparison study, a comparison of a new kit with an old kit.

With a method comparison study, there is often good reason to assume that the new method of measurement is likely to be less than or equally variable as the new method of measurement. Thus a good starting point for using errors in variables regression techniques with method comparison studies is to use $\tilde{\beta}_5$ with $\lambda = 1$. A sensitivity plot investigating changes in $\tilde{\beta}_5$ as λ changes can give an idea as to how robust the estimated slope is to modifications in λ.

The following table contains the x on y regression, the y on x regression and the errors in variables regression fit (using $\tilde{\beta}_5$ and $\lambda = 1$) of a straight line to Figure 8.19. All of these fitted lines are different from the $y = x$ line, and so it appears that both methods do not match exactly. $\tilde{\beta}_5$ lies in between the slope as estimated by y on x and x on y regression, and towards the right hand tail of the data, all the straight line regression fits lie below the $y = x$.

Method	Estimated Slope	Standard Error of Slope	Estimated Intercept
y on x regression	0.93360	0.00913	-1.44408
x on y regression	1.06465	0.00971	-1.74311
Errors in variables regression	0.96119	0.01295	0.08371

Figure 8.20 contains a scatterplot of the method comparison data, as well as the straight line fits as described earlier. It can be seen that there is close agreement between all the fits for small values of AFP, but there is an increasingly poor agreement for higher values.

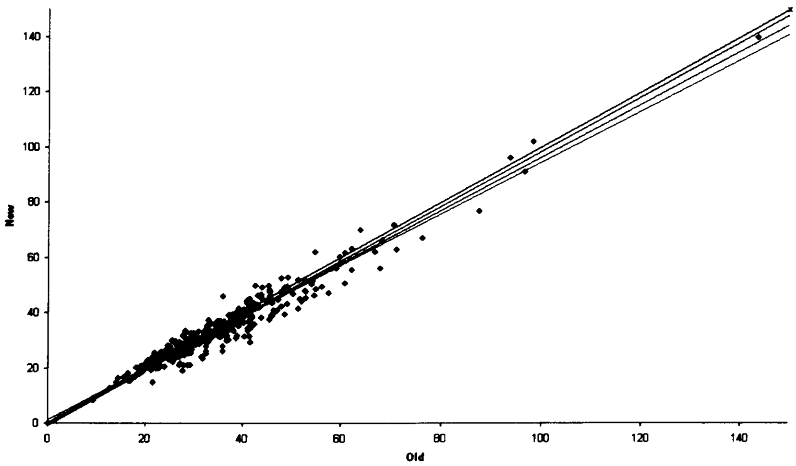


Figure 8.20: Method comparison data with different regression fits.

Data from method comparison studies are often positively skewed, as is the case here. So $\tilde{\beta}_8$ may be of some use as an estimator of the slope for method comparison studies in general. For this application $\tilde{\beta}_8 = 0.98375$, and the corresponding estimate of the intercept is $\alpha = -0.71826$. As the value for $\tilde{\beta}_8$ lies between the values of the slope estimated by y on x and x on y regression, then it is possible, using $\tilde{\beta}_8$ to obtain admissible variance estimates.

As stated earlier, there is usually a firm basis to assume that both methods of measurement have the same variability, or that the new method is less variable than the old method. Thus $\lambda = 1$ seems to be a good starting assumption. The sensitivity of $\tilde{\beta}_5$ to changes in λ may be verified by a sensitivity plot. Figure 8.21 is an example of such a sensitivity plot. This plot shows that the value of $\tilde{\beta}_5$ is robust to small changes in λ. For example, the value of $\tilde{\beta}_5$ recorded to two decimal places remains the same for any λ in the range $0.78 \leq \lambda \leq 1.5$. The value of $\tilde{\beta}_5$ recorded to three decimal places remains the same for any λ in the range $0.98 \leq \lambda \leq 1.04$.

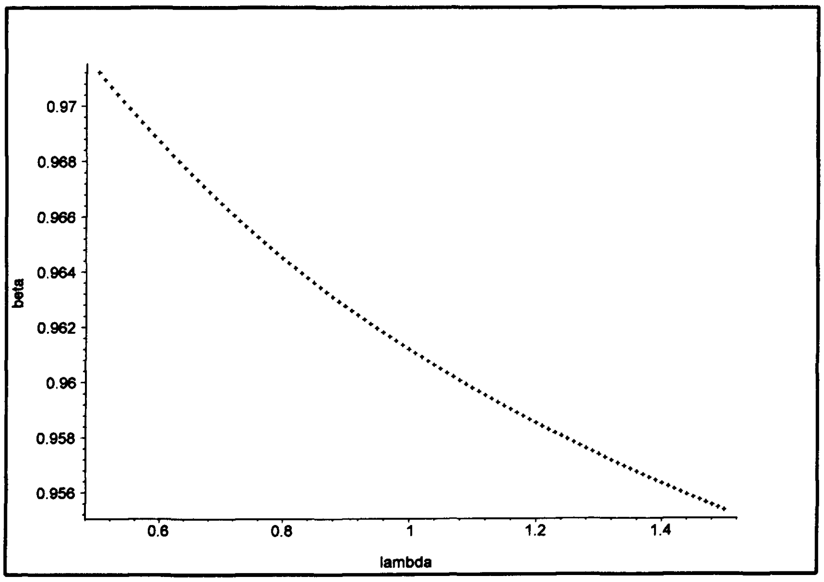


Figure 8.21: Plot of $\tilde{\beta}_5$ against λ.

As again stated earlier, the data plotted in Figure 8.19 seems to demonstrate some heteroscedasticity. In this scenario, Altman and Bland suggest that the logarithm of each variable should be taken. Figure 8.22 contains the scatterplot of the data of Figure 8.19, with the natural logarithm of each variable taken. The bulk of the data lie below the $y = x$ line, suggesting that there is poor agreement of the two methods in general. The log transformation of the data has removed the original skewness of the data, and has appeared to remove at least part of the heteroscedasticity.

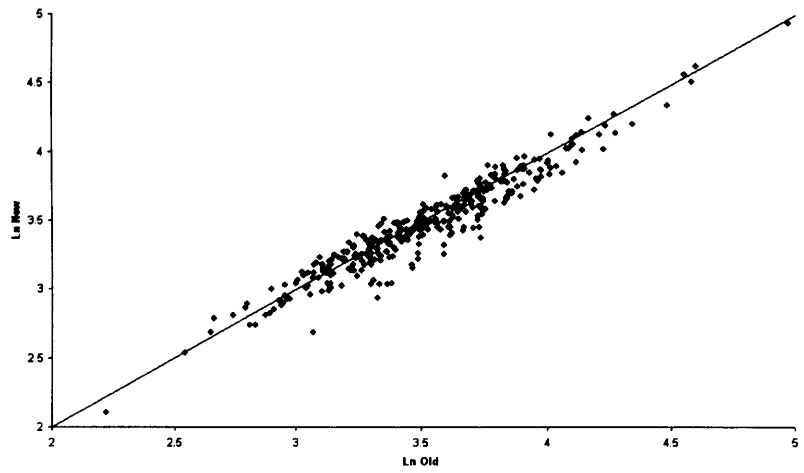


Figure 8.22: A typical method comparison study, a comparison of a new kit with an old kit. The natural logarithm of each variable has been taken.

The following table contains the x on y regression, the y on x regression and the errors in variables regression fit (using $\tilde{\beta}_5$ and $\lambda = 1$) of a straight line to Figure 8.22. All of these fitted lines are different from the $y = x$ line, and so it appears that again the measurement methods do not match exactly. $\tilde{\beta}_5$ lies in between the slope as estimated by y on x and x on y regression, and towards the right hand tail of the data, all the straight line regression fits lie below the $y = x$.

Method	Estimated Slope	Standard Error of Slope	Estimated Intercept
y on x regression	0.94891	0.00482	0.14320
x on y regression	1.02787	0.00515	-0.13352
Errors in variables regression	0.98710	0.00681	0.00936

As mentioned earlier, the transformation to the log domain has removed much of the skewness of the data. The value of $\tilde{\beta}_8$ for the log transformed data is 1.04318, and lies outside of the range of slopes between y on x regression and x on y regression. So $\tilde{\beta}_8$ was a viable estimator for the data in the untransformed domain, but gave an inadmissible estimate in the transformed domain. Figure 8.23 is a sensitivity plot to investigate the robustness of $\tilde{\beta}_5$ to changes in λ. This plot shows that again, the value of $\tilde{\beta}_5$ is robust to small changes in λ. For example, the value of $\tilde{\beta}_5$ recorded to two decimal places remains the same for any λ in the range $0.68 \leq \lambda \leq 1.11$. The value of $\tilde{\beta}_5$ recorded to three decimal places remains the same for any λ in the range $0.99 \leq \lambda \leq 1.03$. The width of these intervals is approximately the same for the untransformed data.

There is a problem with looking at this data in the transformed domain. Assuming the errors in variables fit in the transformed domain we obtain,

$$\ln(New) = 0.00936 + 0.98710 \ln(Old)$$

and converting this back to the untransformed domain we obtain

$$New = 1.0094\, Old^{0.98710}.$$

By transforming the data onto the log domain, the raw data is forced to pass through the origin. This of course, is not ideal as the intercept gives the minimal value of one kit needed for detection in the other kit. This minimal value cannot be safely found in the transformed domain. So while taking logarithms of the data might help

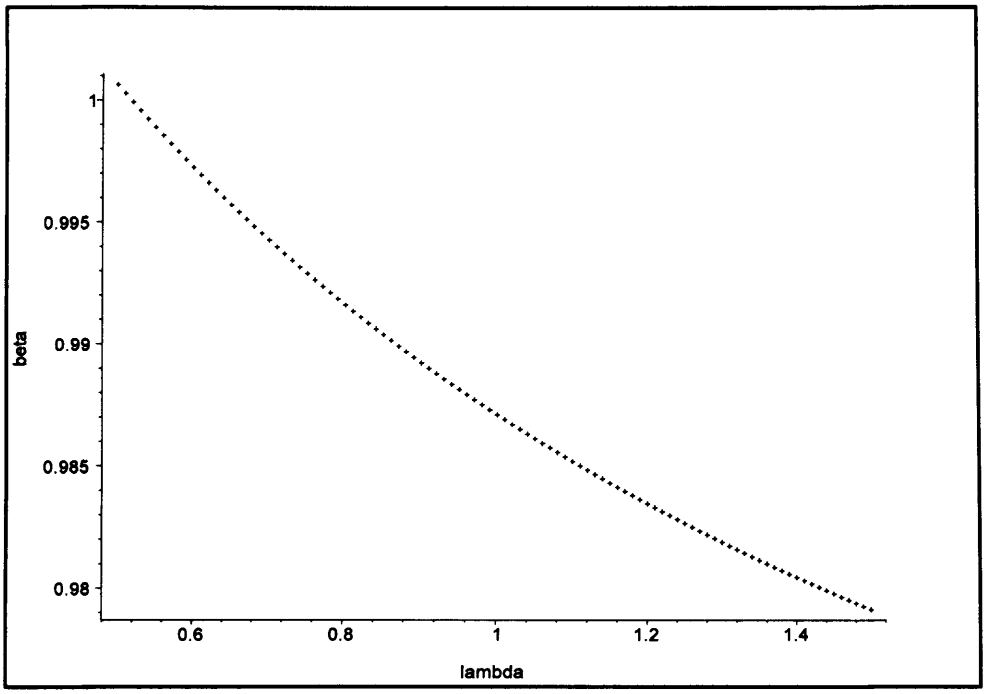
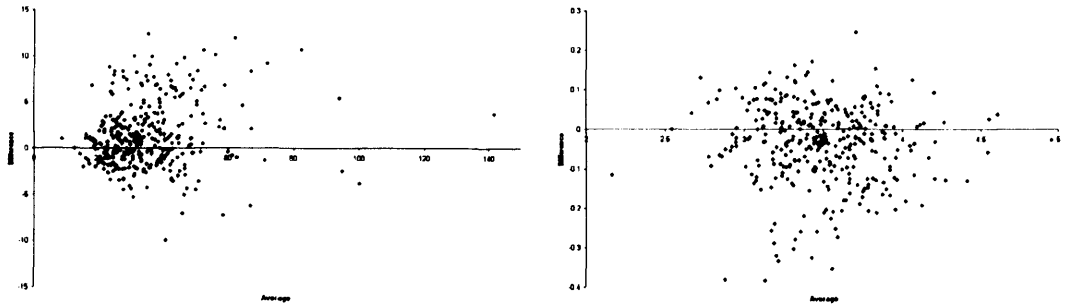


Figure 8.23: Plot of $\tilde{\beta}_5$ against λ for the transformed data.

with heteroscedasticity, for this application, working in the transformed domain has constrained the intercept to be zero, and has removed the possibility of using $\tilde{\beta}_8$.

Altman and Bland [8] advocate the use of the so-called Bland-Altman plot. This is a scatterplot of the difference of the old and new method, plotted against the average value of the old and new method. In other words, for two methods of measurement x and y, the Bland-Altman plot is a scatterplot of $x - y$ against $\frac{x+y}{2}$. The purpose of such a plot for this application is to investigate whether the difference between the old and new method depends of the level of AFP measured. The Bland-Altman plots for both the untransformed and transformed data are shown in Figure 8.24. There does not seem to be much of a pattern in either plot, and so there would appear to be no distinct concentration dependent bias. As mentioned earlier, the untransformed data has shown some heteroscedasticity but the log transform has removed at least part of this.



(a) Untransformed data (b) Transformed data

Figure 8.24: Bland-Altman plots for the AFP method comparison data.

8.4 Functional Regression to Combine Multiple Laser Scans

This section is concerned with the use of a functional regression model to combine multiple laser scans of cDNA microarrays. Microarrays are a powerful and modern method to allow simultaneous analysis of thousands of data points. cDNA microarrays allow the monitoring of expression activities of many genes at the same time. The first stage of the analysis is to estimate the expression levels from the laser scans of the glass slides. A suitable model proposed by Glasbey and Khondoker [48] may be as follows. Let Y_{ij} be the measured response of gene i in scan j. For initial convenience, assume that $Y_{ij} \sim N(\mu_i \beta_j, \sigma_j^2)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Here, μ_i is the expression level for gene i, β_j is the gain setting in laser scan j and σ_j^2 is the variance of the measured response in scan j.

The problem with this model is that of identifiability, and so it is easier to deal with ratios of the form $\frac{Y_{i2}}{Y_{i1}}, \frac{Y_{i3}}{Y_{i1}}$ and $\frac{Y_{i4}}{Y_{i1}}$, where it is assumed that $\beta_1 = 1$. The aim here is to estimate the μ's, β's and σ^2's. A number of differing approaches have been adopted by Glasbey and Khondoker. However it can be seen that the data may be a candidate for a no-intercept slope estimator. Once this slope estimator has been obtained, the other parameters may be estimated. Similarly, due to the large volume of data ($n = 7543$ data points) and its skewness, the third moment slope estimator $\tilde{\beta}_8$ and fourth moment slope estimator $\tilde{\beta}_9$ may be appropriate.

An additional motivation in using ratios of the form $\frac{Y_{ij}}{Y_{i1}}$ is from constructing the moment equations for Glasbey and Khondoker's model

$$Y_{ij} = \mu_i \beta_j \tag{8.3}$$

$$Y_{ij}^2 = \mu_i^2 \beta_j^2 + \sigma_j^2 \tag{8.4}$$

$$Y_{ij} Y_{ik} = \mu_i^2 \beta_j \beta_k. \tag{8.5}$$

Ratios of the form $\frac{Y_{ij}}{Y_{i1}}$ where $\beta_1 = 1$ ultimately allow identification of the other slopes. It is worth noting that different restrictions of the parameter space are possible, such as taking $\beta_2 = 1$ and dealing with ratios of the form $\frac{Y_{ij}}{Y_{i2}}$. As the model is currently formulated, there are $(2m + n)$ parameters and $\frac{mn}{2}(mn + 1)$ second moments. Even $m = n = 2$ gives 10 equations for 6 parameters. This implies that unique estimators are not possible.

Maximum likelihood also fails to give unique answers because of this overparameterisation. For example, one solution of the moment equations (8.3) to (8.5) is $Y_{ij} = \beta_j$, $Y_{ik} = \beta_k$, $\mu_i = 1$ and $\sigma_j^2 = 0$. This solution allows the likelihood function to go to infinity. As there are a multitude of estimators however, there is no need for this solution to be chosen. To obtain unique solutions something about the model must be assumed, and a restriction on the parameter space is the quickest way to do this.

The scatterplot of the data is included in Figure 8.4. The scatterplot of Y_{i2} against Y_{i1} is in blue, the scatterplot of Y_{i3} against Y_{i1} is in purple, and the scatterplot of Y_{i4} against Y_{i1} is in yellow. The skewness of the data in each of the laser scans towards the origin is seen.

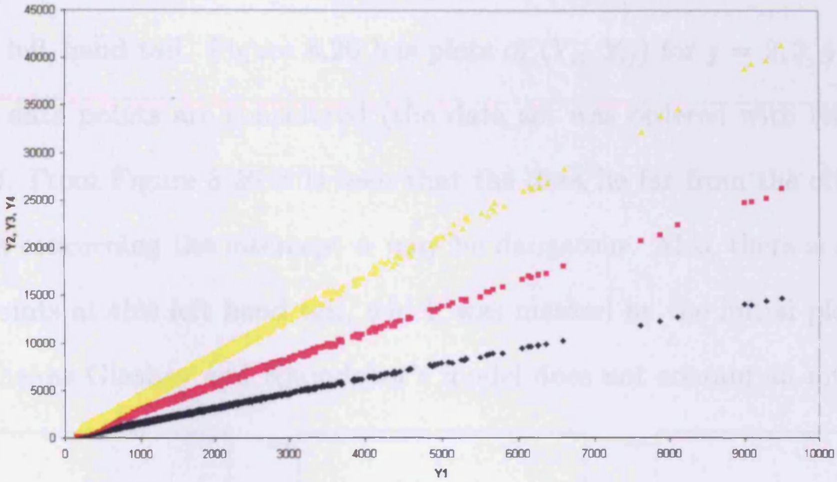


Figure 8.25: Scatterplot of gene expression levels from four laser scans.

The sample skewness and sample excess kurtosis for the measurements on Y_{i1}, Y_{i2}, Y_{i3} and Y_{i4} are given by the following table.

Variable	Skewness	Excess of Kurtosis
Y_{i1}	3.757	24.761
Y_{i2}	3.679	23.784
Y_{i3}	3.668	23.330
Y_{i4}	3.681	23.992

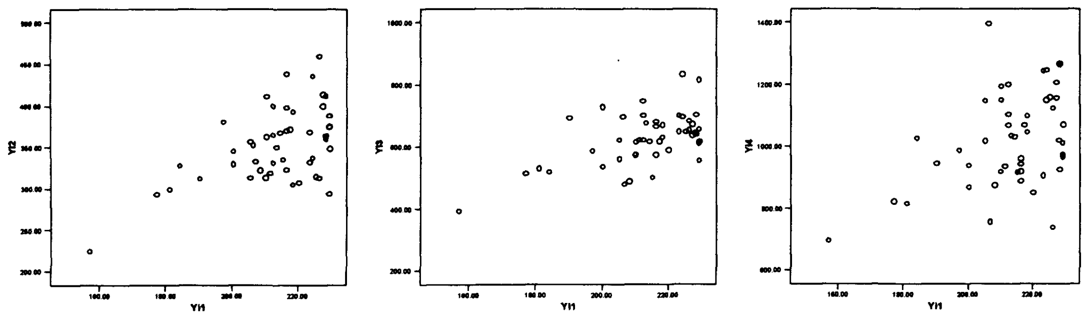
It can be seen that the skewness and kurtosis in each variable is similar and very appreciable.

Y_{i1} is extremely positively correlated with each of the other variables Y_{i2}, Y_{i3} and Y_{i4}. Indeed, the correlation coefficient of Y_{i1} with all the other variables is at least 0.998

and all correlations are significant at the 0.01 level. As the data are so tight, all fitted lines will be very close to each other. This can be seen by just computing the y on x and x on y regression for each pair (Y_{i1}, Y_{ij}), $j = 2, 3, 4$.

Pair	y on x slope	x on y slope
(Y_{i1}, Y_{i2})	1.57077	1.57350
(Y_{i1}, Y_{i3})	2.77004	2.77572
(Y_{i1}, Y_{i4})	4.31012	4.32692

Due to the sheer volume of points at the origin, it is difficult to see the behaviour of the data at the left hand tail. Figure 8.26 has plots of (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$ where only the first 50 data points are considered (the data set was ordered with respect to Y_{i1} beforehand). From Figure 8.26 it is seen that the data lie far from the origin, and so assumptions concerning the intercept α may be dangerous. Also, there is a surprising spread of points at this left hand tail, which was masked by the initial plot in Figure 8.4. Nevertheless Glasbey and Khondoker's model does not contain an intercept, and



(a) (Y_{i1}, Y_{i2}) (b) (Y_{i1}, Y_{i3}) (c) (Y_{i1}, Y_{i4})

Figure 8.26: Scatterplots of the first 50 data points, considering each pair in turn.

so the use of $\tilde{\beta}_1$ is initially appealing. The following table presents the values of $\tilde{\beta}_1$ derived from each pair (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$.

Pair	$\tilde{\beta}_1$
(Y_{i1}, Y_{i2})	1.54815
(Y_{i1}, Y_{i3})	2.72631
(Y_{i1}, Y_{i4})	4.27653

All values of $\tilde{\beta}_1$ lie below the value of the slope estimated by y on x regression. As mentioned in Chapter 3, this is not an appealing feature, as it would lead to negative variance estimates if second order moment equations are used for estimating these variances.

The variances σ_j^2 for $j = 1, 2, 3, 4$ may be estimated from second moment information, that is from using (8.3),

$$\tilde{\sigma_j^2} = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} \mu_i^2\beta_j^2.$$

The obvious admissibility condition here to ensure positive estimators for the variances is that

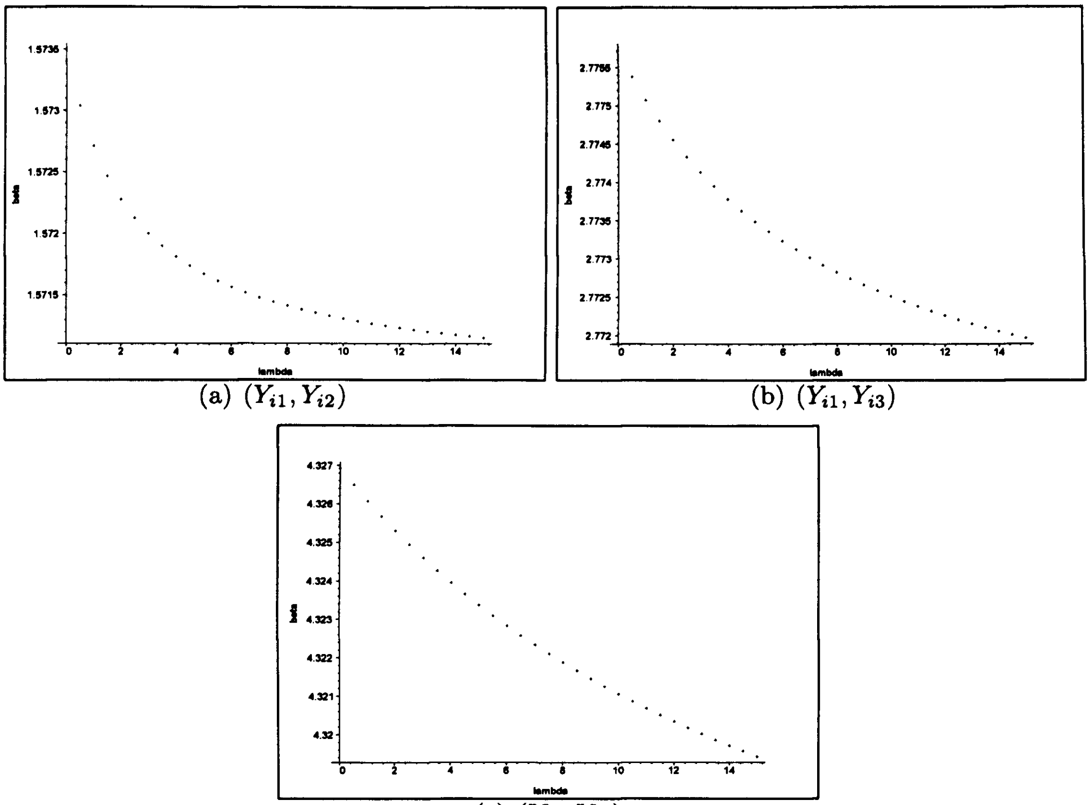
$$\frac{1}{n}\sum_{i=1}^{n} Y_{ij}^2 > \frac{1}{n}\sum_{i=1}^{n} \mu_i^2\beta_j^2.$$

$\tilde{\beta}_5$ is an estimator of the slope which is guaranteed to lie between the slopes of x on y regression and y on x regression. So this estimator may be an improvement on the ones described above. However this estimator assumes that the ratio of error variances, λ is known. This is information that is not explicitly available, nevertheless, some investigation with this slope estimator may shed some light on the fitting of the model. For example, it seems reasonable to assume that for each scan j the error variance is the same. So when pairwise comparisons (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$ are made taking $\lambda = 1$ appears to be a valid assumption. The following table presents the values of $\tilde{\beta}_5$ with $\lambda = 1$ derived from each pair (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$.

Pair	$\tilde{\beta}_1$
(Y_{i1}, Y_{i2})	1.57271
(Y_{i1}, Y_{i3})	2.77507
(Y_{i1}, Y_{i4})	4.32607

It can be seen that the errors in variables estimator of the slope does lie in between the range of slopes given by x on y and y and x regression respectively. The results from using $\tilde{\beta}_5$ seem to be closer to the results from x on y regression than those from y on x regression.

Figure 8.27 contains sensitivity plots to see the changes in $\tilde{\beta}_5$ as λ changes for each pair (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$. When $\lambda = 0$ then the value of slope estimated by $\tilde{\beta}_5$ is identical to that estimated by x on y regression, and as $\lambda \to \infty$ then $\tilde{\beta}_5$ tends to the value of the slope estimated by y on x regression. Figure 8.27 shows that when $\lambda = 1$ the estimated slope is very close to that estimated by x on y regression. Indeed, the changes in the slope as λ grows slightly away from 1 are negligible. For example, for the pair (Y_{i1}, Y_{i2}) the estimated slope using $\tilde{\beta}_5$ to two decimal places is 1.57 regardless of the value of λ chosen. As the gain setting of the scan increases ($j = 2, 3, 4$) then a more strict downward linear trend in the estimated value of β as λ increases is observed. Nevertheless, it seems that the value in $\tilde{\beta}_5$ is robust to small changes in λ.



(a) (Y_{i1}, Y_{i2}) (b) (Y_{i1}, Y_{i3})

(c) (Y_{i1}, Y_{i4})

Figure 8.27: Plots of λ against $\tilde{\beta}_5$, considering each pair in turn.

To highlight how robust $\tilde{\beta}_5$ is to small changes in λ, the following table has the value of $\tilde{\beta}_5$ for a particular value of λ for each pair of variables considered. The changes in β only occur in the third decimal place.

Pair	$\lambda = 0$	$\lambda = 1$	$\lambda = 5$
(Y_{i1}, Y_{i2})	1.57350	1.57271	1.57167
(Y_{i1}, Y_{i3})	2.77572	2.77507	2.77348
(Y_{i1}, Y_{i4})	4.27653	4.32607	4.32337

Therefore it seems as though $\tilde{\beta}_5$ with $\lambda = 1$ is a good practical estimator of the slope β for this particular model and data set. As stated earlier, at first sight, it appears that the data presented in Figure 8.4 would be a candidate for estimators of the slope based on third and fourth order moments, namely $\tilde{\beta}_8$ and $\tilde{\beta}_9$. The following table presents the values of $\tilde{\beta}_8$ and $\tilde{\beta}_9$ derived from each pair (Y_{i1}, Y_{ij}) for $j = 2, 3, 4$:

Pair	$\tilde{\beta}_8$	$\tilde{\beta}_9$
(Y_{i1}, Y_{i2})	1.56123	1.55638
(Y_{i1}, Y_{i3})	2.75085	2.74151
(Y_{i1}, Y_{i4})	4.28936	4.28143

Unfortunately, these estimates of the slope do not lie within the range of the slopes estimated by x on y and y on x respectively. Indeed, the values of $\tilde{\beta}_8$ and $\tilde{\beta}_9$ derived for each pair lie below the value of the y on x regression estimator. As the distance between the x on y slope estimator and the y on x estimator is small, then an estimator guaranteed to lie between these values is appealing. Therefore $\tilde{\beta}_5$ seems the obvious choice, and taking $\lambda = 1$ seems to be a valid and correct assumption.

This example has been chosen to illustrate the flexibility of the method of moments approach. An initial analysis of the data indicated that lines through the origin are the best model. However a more thorough exploration of possible estimates has resulted in a model in which somewhat greater confidence can be played.

8.5 Galton and Regression to the Mean

Galton [42] was the first to develop the concept of regression towards mediocrity, or as it is more commonly known at present, regression to the mean. Galton carried out a number of experiments involving both seedlings and humans, and as a result of his investigations introduced the idea of regression to the mean. In terms of his study on seeds from the same species, Galton explained regression to the mean as follows.

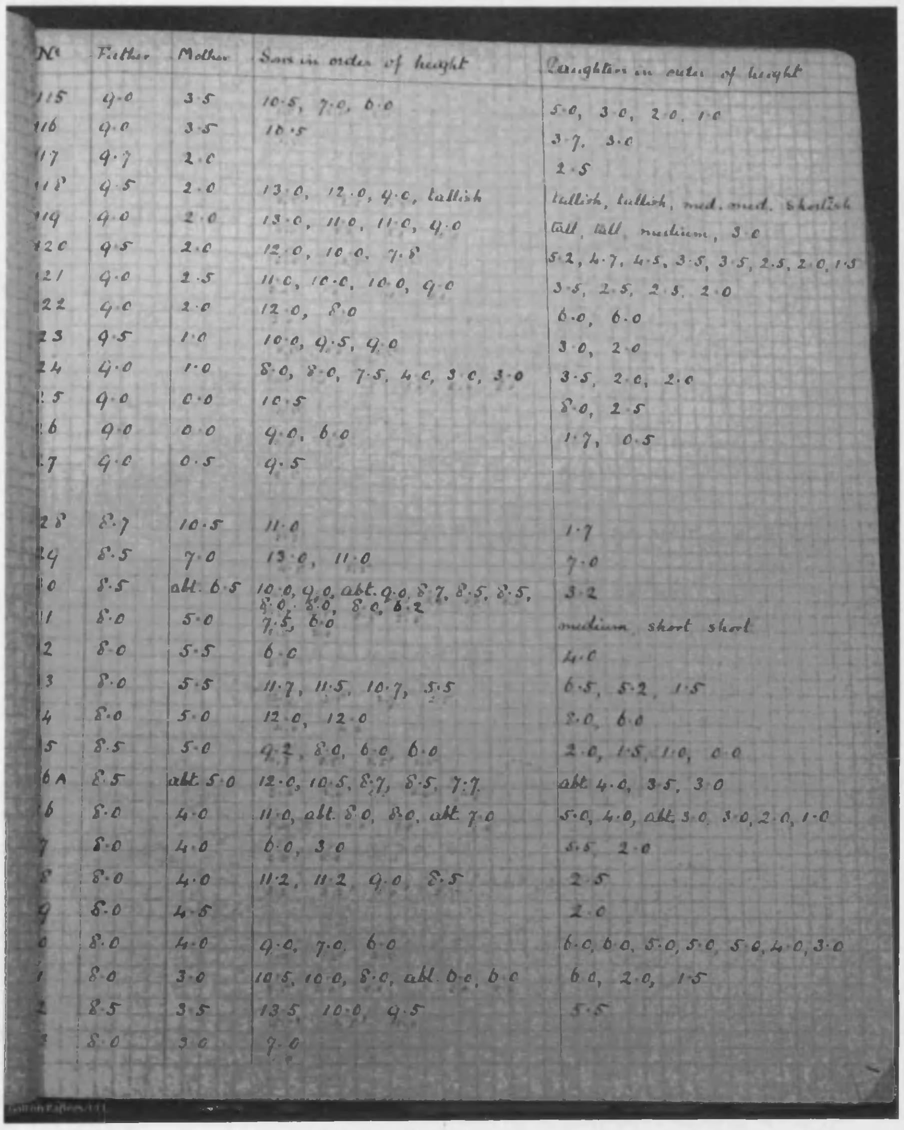
> "It appeared from these experiments that the offspring did not tend to resemble their parents seeds in size, but to always be more mediocre than they- to be smaller than the parents, if the parents were large; to be larger

than the parents, if the parents were very small."

Galton's more famous study involved what he called hereditary stature. He compared the heights of 930 adult children and their parents (205 of them). Even though he had data of the heights of both sons and daughters, his study focused on the heights of the sons. Instead of taking the direct average of both the parents heights, he multiplied each female height by 1.08. Galton offered the following explanation for this.

> "In every case I transmuted the female statures to their corresponding male equivalents and used them in their transmuted form, so that no objection grounded on the sexual difference of stature need to be raised when I speak of averages"

The measurements of the heights of the adult children and parents are likely to be susceptible to measurement error. In addition to this, there is much uncertainty in the data due to the way some measurements have been recorded. Figure 8.28 is an extract from the notebook which contains the data in its originally recorded form. Firstly, it is noted that heights are measured to the nearest inch. Secondly, Galton has used the words 'about', 'medium', 'shortish', 'tall' and 'tallish' to describe the heights of certain individuals. Elsewhere in the data Galton has used the words 'deformed', 'idiotic' and 'middle'. It is thus very difficult to attribute a height to any individual with these entries. Due to the uncertainty in the measurements of all the heights in Galton's data set, a regression technique that can take into account error in both variables is likely to be of some use. Galton's original analysis of the data used standard regression techniques.



Nº	Father	Mother	Sons in order of height	Daughters in order of height
115	9·0	3·5	10·5, 7·0, 6·0	5·0, 3·0, 2·0, 1·0
116	9·0	3·5	10·5	3·7, 3·0
117	9·7	2·0		2·5
118	9·5	2·0	13·0, 12·0, 9·0, tallish	tallish, tallish, med. med. shortish
119	9·0	2·0	13·0, 11·0, 11·0, 9·0	tall, tall, medium, 3·0
120	9·5	2·0	12·0, 10·0, 7·8	5·2, 4·7, 4·5, 3·5, 3·5, 2·5, 2·0, 1·5
121	9·0	2·5	11·0, 10·0, 10·0, 9·0	3·5, 2·5, 2·5, 2·0
122	9·0	2·0	12·0, 8·0	6·0, 6·0
23	9·5	1·0	10·0, 9·5, 9·0	3·0, 2·0
24	9·0	1·0	8·0, 8·0, 7·5, 4·0, 3·0, 3·0	3·5, 2·0, 2·0
5	9·0	0·0	10·5	8·0, 2·5
6	9·0	0·0	9·0, 6·0	1·7, 0·5
7	9·0	0·5	9·5	
28	8·7	10·5	11·0	1·7
29	8·5	7·0	13·0, 11·0	7·0
0	8·5	abt. 6·5	10·0, 9·0, abt. 9·0, 8·7, 8·5, 8·5, 8·0, 8·0, 8·0, 6·2	3·2
1	8·0	5·0	7·5, 6·0	medium, short, short
2	8·0	5·5	6·0	4·0
3	8·0	5·5	11·7, 11·5, 10·7, 5·5	6·5, 5·2, 1·5
4	8·0	5·0	12·0, 12·0	8·0, 6·0
5	8·5	5·0	9·2, 8·0, 6·0, 6·0	2·0, 1·5, 1·0, 0·0
6 A	8·5	abt. 5·0	12·0, 10·5, 8·7, 8·5, 7·7	abt. 4·0, 3·5, 3·0
6	8·0	4·0	11·0, abt. 8·0, 8·0, abt. 7·0	5·0, 4·0, abt. 3·0, 3·0, 2·0, 1·0
7	8·0	4·0	6·0, 3·0	5·5, 2·0
8	8·0	4·0	11·2, 11·2, 9·0, 8·5	2·5
9	8·0	4·5		2·0
0	8·0	4·0	9·0, 7·0, 6·0	6·0, 6·0, 5·0, 5·0, 5·0, 4·0, 3·0
1	8·0	3·0	10·5, 10·0, 8·0, abt. 6·0, 6·0	6·0, 2·0, 1·5
2	8·5	3·5	13·5, 10·0, 9·5	5·5
	8·0	3·0	7·0	

Figure 8.28: An excerpt from Galton's original notebooks containing the data on family heights. 60 inches are to be added to each entry. Taken from http://galton.org/.

A scatterplot of sons heights against midparents heights (in inches) is included in Figure 8.29. A linear trend is apparent in the data, and Hanley [55] argued that nonlinear regressions did not provide significantly better fits than linear ones. This point shall be revisited later.

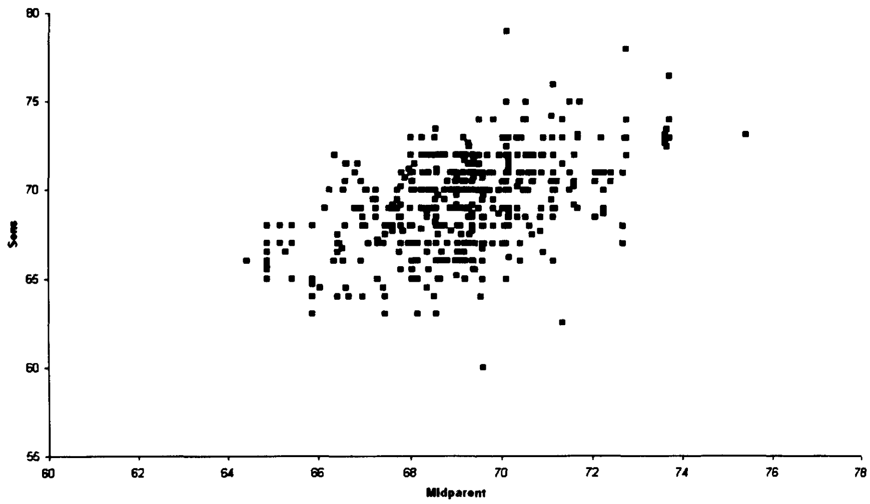


Figure 8.29: Scatterplot of sons heights against midparents heights (in inches).

Letting S denote the height of sons, and M denote the midparent height, y on x regression gives the estimate of the line as

$$S = 19.913 + 0.71328M,$$

and x on y regression gives the estimate of the line as $M = 46.552 + 0.32679S$ and inverting this gives

$$S = -142.45234 + 3.06007M.$$

Figure 8.30 contains the scatterplot of the original data with both the y on x and x on y regression fits. The y on x regression with the slope less than 1 demonstrates the regression to the mean phenomenon that Galton observed.

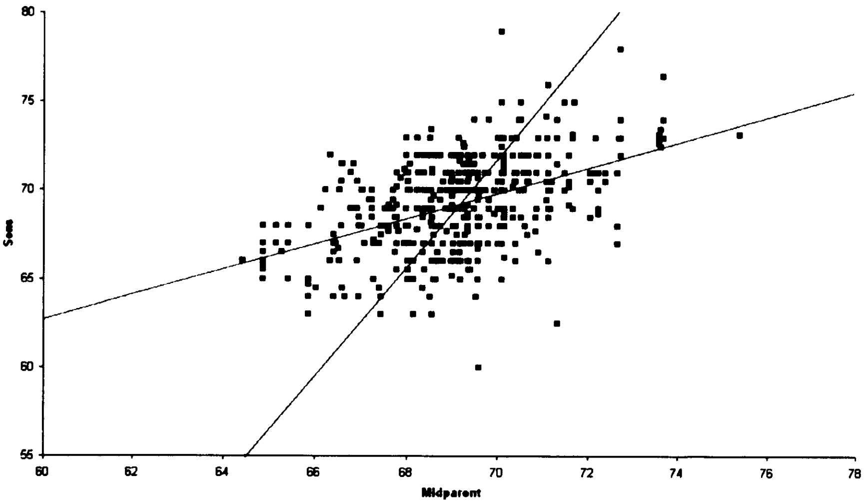


Figure 8.30: Scatterplot of sons heights against midparents heights (in inches) with simple linear regression fits. y on x fit is in red, x on y fit is in blue.

Unless errors in variables estimators of the slope based on higher order moments are used, a restriction on the parameter space has to be made in order to use the standard errors in variables estimators of the slope. For this data set there are a number of options. Due to the construction of the data, the within family error mean square for the sons' heights will give an estimated value for σ^2_ε, and within stature group error mean square for the sons' heights will give another estimated value for σ^2_ε. This gives $\tilde{\sigma}^2_\varepsilon = 4.6$ and $\tilde{\sigma}^2_\varepsilon = 4.242$ respectively. Hanley [55] used the MIXED procedure in SAS as well as WinBUGS and reported the within family error mean square as approximately 4.11, which is slightly lower than the estimate obtained by ANOVA. Hanley however investigated midparent height against son and daughters heights, disregarding sex. Thus a slight discrepancy in the results is to be expected.

Assuming that $\sigma^2_\varepsilon = 4.242$ gives $\tilde{\beta}_3 = 1.17043$, $\tilde{\alpha} = -11.69779$, $\tilde{\sigma}^2 = 1.91796$, and $\tilde{\sigma}^2_\delta = 1.22927$. Assuming that $\sigma^2_\varepsilon = 4.632$ gives $\tilde{\beta}_3 = 1.01095$, $\tilde{\alpha} = -0.67371$, $\tilde{\sigma}^2 = 2.22042$ and $\tilde{\sigma}^2_\delta = 0.92681$. Another method of estimating the slope would be to assume that $\lambda = 2$, as the variable M is the average of two heights. Taking $\lambda = 2$ yields $\tilde{\beta}_5 = 1.41935$ which is larger than the previous errors in variables slopes derived. The remaining parameters have the following estimates; $\tilde{\alpha} = -28.90980$, $\tilde{\sigma^2} = 1.58160$ and $\tilde{\sigma_\delta^2} = 1.56563$.

Figure 8.31 contains the scatterplot of Galton's data, with all regression fits described thus far. All estimated errors in variables regression fits are between the y on x and x on y fits, and so, as seen above all variance estimates are nonnegative.

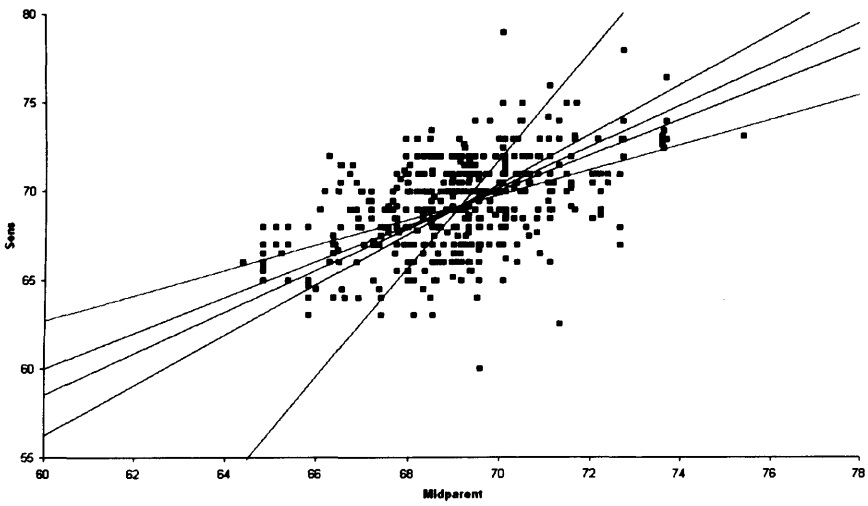
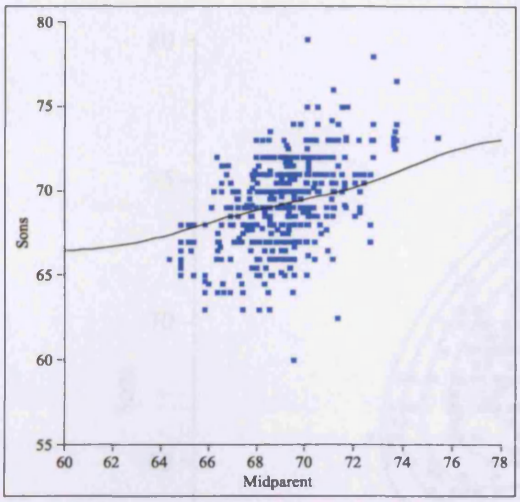


Figure 8.31: Scatterplot of sons heights against midparents heights (in inches) with straight line regression fits. y on x fit is in red, x on y fit is in blue, fit with $\sigma_\varepsilon^2 = 4.242$ is in green, fit with $\sigma_\varepsilon^2 = 4.6$ is in purple and fit with $\lambda = 2$ is in brown.

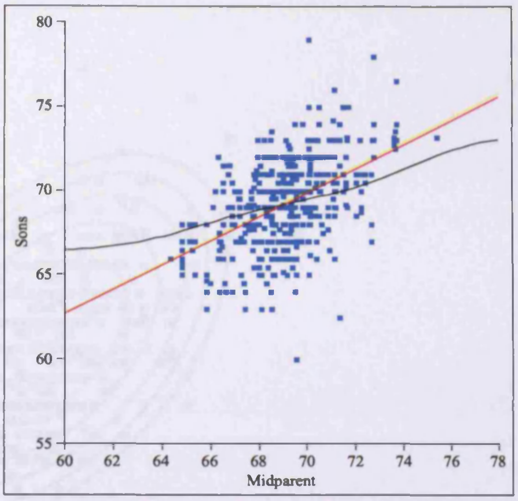
As mentioned earlier, Hanley [55] stated that a nonlinear regression fit to the Galton data is no more beneficial than a linear regression fit. This was in reply to an earlier paper by Wachsmuth et al.[109] who commented on 'Galton's bend' which is what they claimed to be an undiscovered nonlinearity in the data. To demonstrate Galton's bend they fitted a LOESS smoother to the data, and showed that there was a bend in the curve at a midpoint height of approximately 68 inches.

A LOESS smoother fits low degree polynomials to subsets of the data, and each polynomial is fitted by weighted least squares. The degree of the polynomial, and the weights are flexible and may be chosen by the practitioner, but unfortunately, Wachsmuth et al.'s paper does not provide any details concerning the exact details of the LOESS fit. The LOESS fit however is a non-parametric method of establishing $E[y|x]$ (or in this application $E[S|M]$), and as noted in previous Chapters, $E[y|x]$ may not be linear even if the latent data set $\{(\xi_i, \eta_i), i = 1, \ldots, n\}$ do lie exactly on a straight line. $E[y|x]$ is only a straight line for the Normal structural model. Secondly, LOESS is only one such non-parametric method to obtain $E[y|x]$. There are others, most notably the Nadaraya-Watson estimator describes in Chapter 6. The bandwidth for the Nadaraya-Watson estimator may be chosen in accordance with equation (6.10).

Figure 8.32 contains the Nadaraya-Watson fit to $E[y|x]$ with the smallest bandwidth computed from equation (6.10) for the variety of estimated values of σ_δ^2. The bend as described by Wachsmuth at al. at the midpoint height of approximately 68 inches is not observed, and the Nadaraya-Watson estimate of $E[y|x]$ is seen to be approximately linear over the entire range of the data. To reiterate the point made in Chapter 6 however, it is not necessarily the case that $E[y|x]$ follow the y on x line, as demonstrated in this Figure.



(a) Without y on x regression line



(b) With y on x regression line (in red)

Figure 8.32: Nadaraya-Watson fit for the Galton data.

As introduced in Chapter 7, the contours of equal probability are formed by the ellipse defined by the equation

$$\frac{(x-\bar{x})^2}{Var[x]} + \frac{(y-\bar{y})^2}{Var[y]} - \frac{2\rho}{\sqrt{Var[x]}\sqrt{Var[y]}}(x-\bar{x})(y-\bar{y}) = k$$

for different values of k. For the sons and midparent height data this may be written as

$$\frac{(x-69.14678)^2}{3.14723} + \frac{(y-69.23368)^2}{6.86943} - 0.20767(x-69.14678)(y-69.23368) = k. \quad (8.6)$$

As an example of this geometry, Figure 8.33 contains the contours of equal probability for varying values of k.

Using the algebra derived in Chapter 7, the slopes of the major and minor axis are found by solving the quadratic

$$\frac{2\beta}{1-\beta^2} = \frac{0.20767}{0.14557-0.31774}$$

in terms of β, giving the slope of the major axis as 2.12803 and the slope of the minor axis as -0.46992. Figure 8.34 contains a scatterplot of sons heights against midparents

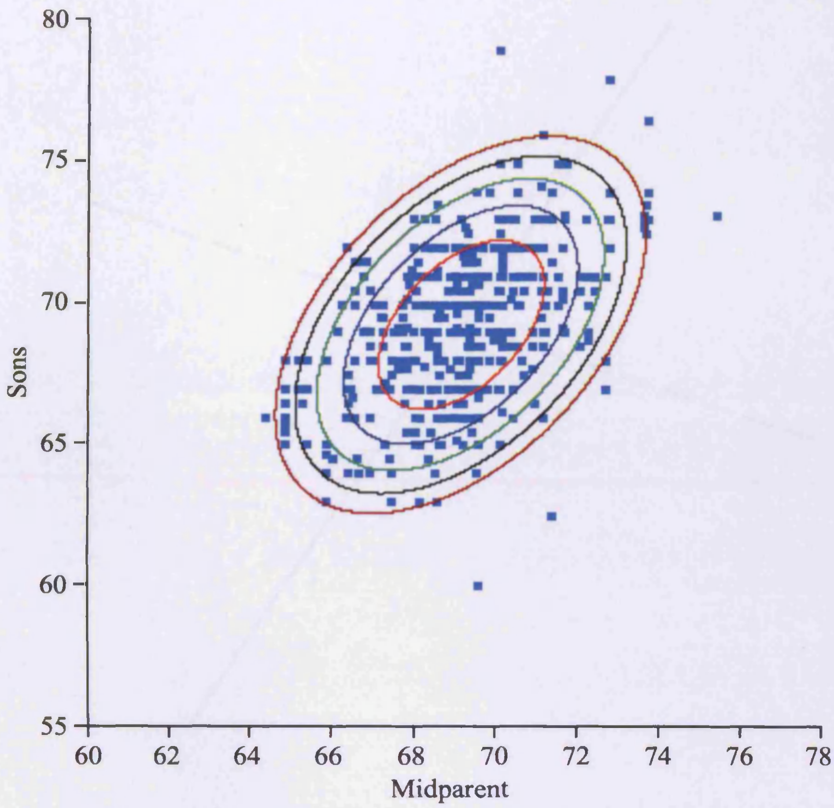


Figure 8.33: Scatterplot of sons' heights against midparents heights (in inches) with ellipses. The inner ellipse is formed with $k = 1$, and the outer ellipse is formed with $k = 5$. Working outwards, the ellipses are formed with $k = 2, 3$ and $k = 4$ respectively.

heights, with ellipse and major and minor axes. Due to the scaling, the major and minor axes do not appear to be at right angles. This is because distances and angles are not preserved under transformations of scale. The major axis is the same line obtained by taking $\lambda = 1$ in $\tilde{\beta}_5$. As stated in Chapter 7, the major and minor axes are identical to the first and second principal component respectively.

As stated earlier in this section, Galton concentrated on the analysis of the sons heights in relation to midparent height. The heights of daughters however were also recorded. A scatterplot of daughters heights against midparents heights (pink diamonds) over-

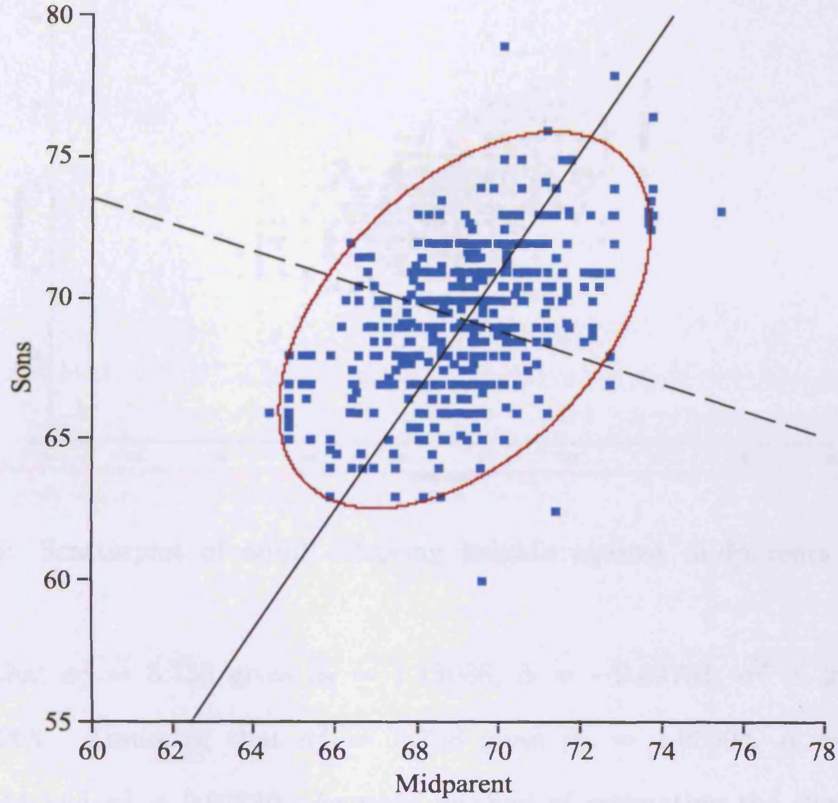


Figure 8.34: Scatterplot of sons heights against midparents heights (in inches) with ellipse ($k = 5$) and major (bold line) and minor axes (dashed line).

layed on the scatterplot of sons heights against midparents heights is included in Figure 8.35. As the mothers heights were scaled by a factor of 1.08, then it is also appropriate to scale the daughters heights by an identical factor. It can be seen that both sets of data are very similar, and share the main features.

Again, due to the construction of the data, the within family error mean square for the daughters' heights will give an estimated value for σ_ε^2, and within stature group error mean square for the daughters' heights will give another estimated value for σ_ε^2. This gives $\tilde{\sigma_\varepsilon^2} = 4.06$ and $\tilde{\sigma_\varepsilon^2} = 3.758$ respectively.

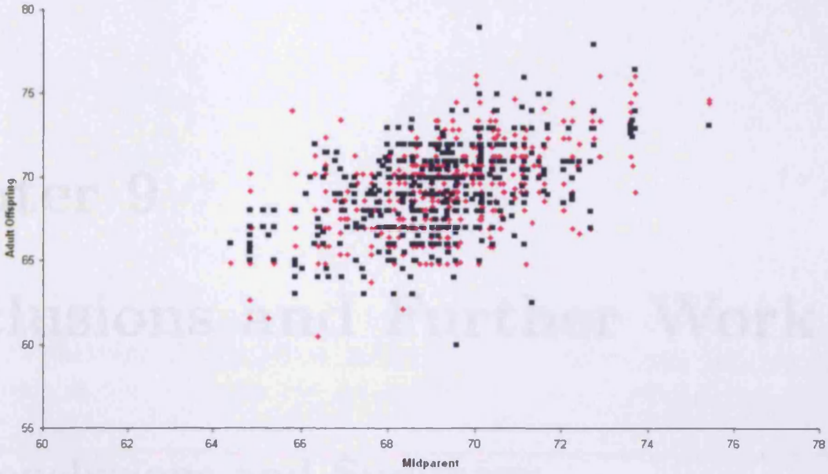


Figure 8.35: Scatterplot of adult offspring heights against midparents heights (in inches).

Assuming that $\sigma_\varepsilon^2 = 3.758$ gives $\tilde{\beta}_3 = 1.13088$, $\tilde{\alpha} = -9.09791$, $\tilde{\sigma^2} = 2.11306$, and $\tilde{\sigma_\delta^2} = 1.23268$. Assuming that $\sigma_\varepsilon^2 = 3.758$ gives $\tilde{\beta}_3 = 1.00894$, $\tilde{\alpha} = -0.65005$, $\tilde{\sigma^2} = 2.36844$ and $\tilde{\sigma_\delta^2} = 0.97730$. Another method of estimating the slope would be to assume that $\lambda = 2$, as the variable M is the average of two heights. Taking $\lambda = 2$ yields $\tilde{\beta}_5 = 1.36668$ which is larger than the previous errors in variables slopes derived. The remaining parameters have the following estimates; $\tilde{\alpha} = -25.43384$, $\tilde{\sigma^2} = 1.74848$ and $\tilde{\sigma_\delta^2} = 1.59726$. All of these estimated values can be seen to be very similar to the corresponding estimates for the data which only considered sons' heights and midparent heights.

Chapter 9

Conclusions and Further Work

9.1 Conclusions and Summary

This section will briefly describe the main results from each of the Chapters in this thesis. Much discussion of the main results in the thesis was included in these Chapters, and so only a brief summary is needed here.

Chapter 1 was an introductory Chapter that set the scene for the remainder of the thesis. As stated, there is a wealth of literature on errors in variables modelling, and to give an idea of some of the approaches adopted by many authors Chapter 2 is a comprehensive literature survey looking at the main approaches. It was demonstrated that many of the methods used to tackle the errors in variables problem are linked, and many were shown to be equivalent to the method of moments. For example, LISREL uses the same method of moments estimating equations derived in Chapter 3, and maximum likelihood for the Normal structural model is equivalent to using the method of moments. Chapter 2 is a useful resource and provides many references for anyone who wishes to investigate some alternative methods for errors in variables modelling discussed in the literature.

The method of moments was used in numerous places throughout this thesis, and was introduced as a method of estimating the parameters of an errors in variables model in Chapter 3. The method of moments approach has a number of distinct advantages. Firstly, the algebra is much simpler for the method of moments than for maximum likelihood. Secondly, the method of moments estimating equations do not depend on the distributions of the random variables in the model. A common misconception regarding the method of moments is that there are no asymptotic results for the estimators. This is not true and, via the delta method, complete variance covariance matrices for each of the slope estimators discussed in Chapter 3 were derived. For the variance covariance matrices not explicitly reported in this thesis, a Maple programme has been created so that they may be computed. The derived variance covariance matrices extend the work of Hood et al. [57] to cope with any distribution of ξ, δ and ε. In addition the derived variance covariance matrices are in a much simpler form than those presented by Hood et al.

Chapter 4 contains many simulations investigating the estimators of the previous Chapter. Particular attention was applied to simulations to assess bias, small sample behaviour, the breaking of admissibility conditions, variance covariance matrices and those estimators of the slope based on higher order moments. The aim of this Chapter was to provide some guidance on the use of the estimators discussed in the previous Chapter, as well as to assess their behaviour.

The key advantage of the method of moments is its simplicity, and in Chapter 5 maximum likelihood estimation is shown to be difficult to initiate algebraically and ineffective as an estimation tool for the examples considered. For these examples, it seems that the only feasible method to maximise the likelihood function would be a numerical one, but the shape of the likelihood function might make it difficult to find a method of solution that converges to the optimum. Some distinctions between the functional and structural models are offered, but again this Chapter was mainly used to demonstrate some of the problems in using maximum likelihood for an errors in variables model.

The topic of prediction was introduced in Chapter 6. In an errors in variables regression model there is a number of different prediction questions. It is important to use the correct technique and to answer the correct question. To predict a y value from a given x, a number of approaches was described in detail. Parametric approaches involved directly computing $E[y|x]$, and an approximation derived by Cochran [22] was given. The non-parametric technique used in this thesis was the Nadaraya-Watson estimator, which was simple to implement. For the small number of examples considered, the Nadaraya-Watson estimator was in close agreement with the exact result. Chapter 6 also detailed some important issues concerning $E[y|x]$. The main issue is that $E[y|x]$ only follows the least squares line for a Normal structural model. For the Normal functional model, $E[y|x]$ follows the errors in variables line. It is also not necessary that $E[y|x]$ is a straight line. For models other than the Normal structural and Normal functional models, $E[y|x]$ will be a curve. To uncover the latent ξ, an optimal linear combination (in terms of minimum variance) was derived from two naive method of moments estimating equations. The estimator was simple in form, and at least partly corrects for the effect of the migration effect that is described in Chapter 7.

Chapter 7 made the point that a residual for an errors in variables model is not explicitly defined as it is in some other estimation methods. In addition, the vertical residual from the errors in variables fit plotted against the observed x will always display a trend, and these residual analyses are made difficult. The concept of migration was introduced, and discussed in detail. The discussion of migration brought together some common themes throughout the thesis, and a number of explanations as to why migration occurs was offered. After the details on migration were given, this enabled a fresh look at residuals to be made, and two approaches to residual analyses were offered. One approach would be to consider the vertical residual from the $E[y|x]$ curve. It was mentioned that the $E[y|x]$ curve seemed to follow the trend of migration, and so the vertical residual from this curve would not be subject to the migration effect. The second method involved estimating the latent $\xi_i's$ once an initial errors in variables fit has been made. Then a residual analysis can performed on the standard least squares fit to $\{(\xi_i, y_i), i = 1, \ldots, n\}$, and this vertical residual would not be subject to the migration effect. The distinction here is whether a curve is sought that smooths the data, in which $E[y|x]$ is appropriate, or whether the true relationship between the variables is sought, in which case an unbiased estimate of the true line is needed.

Finally Chapter 8 offered some applications which would benefit from errors in variables methodology. The examples were chosen to be wide ranging, and the case for using errors in variables methodology in each application was made.

Even though this thesis has considered the topic of straight line fitting, this is still important for a number of reasons. To encourage the proper development of the subject a firm basis for further research must be made. Fitting a straight line is an important topic in its own right, and the development of some of the ideas in this thesis would allow application to more complicated models. There seems to be two approaches in statistics, one is to develop a generic methodology, and then to reduce it to the particular model from hand. Another is to start with a simple model, and then develop the theory on the basis of this model. The latter approach has been adopted in this thesis, and it is felt if this were not the case, then a number of valuable insights would have been lost. The concepts introduced in this thesis are necessary for a full appreciation of errors in variables modelling in general, and the problems mentioned in this thesis of estimation, asymptotics, prediction and residuals would apply to any errors in variables model.

9.2 Further Work and Additional Topics

This section will present some topics that need further investigation as a result of this thesis. For some of these topics, some mathematical details are provided.

Extending the Nadaraya-Watson estimator The Nadaraya-Watson estimator is a non-parametric method of constructing $E[y|x]$. Details on how to implement the estimator were given in Chapter 6. In this thesis, the Nadaraya-Watson estimator has only been applied to linear models, but as it does not depend on the functional form of the data, it may be applied to more complicated non-linear models. An investigation of the performance of the Nadaraya-Watson estimator for non-linear errors in variables models would prove valuable. It is likely that computing exact expressions for $E[y|x]$ for non-linear models will be more difficult than for linear models, and so a reliable and robust non-parametric alternative would be of some use.

An adaptive bandwidth is often used by non-parametric regression methods. For simplicity, an adaptive bandwidth has been omitted from the Nadaraya-Watson estimator. If an adaptive bandwidth was adopted, then it is likely that the properties of the Nadaraya-Watson derived in Chapter 6 would change, but in addition, it would be difficult to explicitly find the changes in these properties with an adaptive bandwidth. As comparisons have shown, the Nadaraya-Watson estimator seems to give a good fit to the exact expression for $E[y|x]$ in the examples considered in Chapter 6. There is sometimes however some slight deviation in the tails, where data are sparse. Adopting an adaptive bandwidth may improve the fit at the tails, improving the general form of the Nadaraya-Watson estimator in this thesis.

The question of how to find the adaptive bandwidth in an errors in variables model has received little attention in the literature. Most non-parametric regression methods that have been applied to an errors in variables model use a fixed bandwidth chosen from some optimality criterion (see for example, Carroll at al. [15]). To improve the fit at the tails, the adaptive bandwidth should depend on the density of points so that data at the tails do not exert as much leverage on the Nadaraya-Watson estimator. Clearly, further investigation in this area is needed.

Quadratic structural regression Some of the theory detailed in this thesis may be applied to more complicated models. For example, as already stated the Nadaraya-Watson estimator of Chapter 6 is not restricted to straight line models. Developing the complexity of the model typically entails introducing more parameters in the model. As an example of extending the model, consider the following quadratic structural regression model.

The observations x_i may again be written as a latent variable adjusted by some random error component so that $x_i = \xi_i + \delta_i$. The fundamental difference between quadratic structural regression and the straight line regression considered in this thesis is in the consideration of the y_i measurements. For quadratic structural regression it is assumed that

$$y_i = \alpha + \beta_1(\xi_i - \mu) + \beta_2(\xi_i - \mu)^2 + \varepsilon_i.$$

These errors, δ and ε are assumed to be independent of ξ, and of each other. As we are assuming the structural model then $E[\xi_i] = \mu$ and $Var[\xi_i] = \sigma^2$.

It was seen in Chapter 3 that the method of moments may be used to estimate the parameters of a straight line errors in variables regression fit. For the quadratic structural regression model there are seven parameters that need to be estimated. They are $\mu, \sigma^2, \alpha, \beta_1, \beta_2, \sigma_\delta^2$ and σ_ε^2. There is an additional parameter not present for an errors in variables straight line fit, and that is β_2 the coefficient of the quadratic term.

The parameter μ may be estimated from the method of moments estimating equation $\bar{x} = \mu$. The sample mean of the y observations gives $\bar{y} = \alpha + \beta_2\sigma^2$. As x_i is defined the same as in Chapter 3 then $s_{xx} = \sigma^2 + \sigma_\delta^2$. The remaining second order moment estimating equations are not as straightforward. For example,

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}\left[(\xi_i - \bar{\xi}) + (\delta_i - \bar{\delta})\right]\left[\beta_1(\xi_i - \bar{\xi}) + \beta_2(\xi_i - \bar{\xi})^2 + (\varepsilon_i - \bar{\varepsilon})\right]$$

and upon taking expectations yields the following method of moments estimating equation $s_{xy} = \beta_1\sigma^2 + \beta_2\mu_{\xi 3}$. If ξ is assumed to follow a symmetric distribution then $\mu_{\xi 3} = 0$, and the method of moments estimating equation (3.5) for the straight line is obtained. Otherwise an additional new parameter $\mu_{\xi 3}$ is introduced into the model.

Using the same method it follows that $s_{yy} = \beta_1^2\sigma^2 + \beta_2^2\mu_{\xi 4} + 2\beta_1\beta_2\mu_{\xi 3} + \sigma_\varepsilon^2$ and another parameter, $\mu_{\xi 4}$ is introduced into the model.

Even if the triple $(\xi, \delta, \varepsilon)$ is assumed to be trivariate Normal as in the Normal structural model, then there are still five equations for seven parameters. In the same manner as Chapter 3, a possible option is to restrict the parameter space. If the triple $(\xi, \delta, \varepsilon)$ is assumed to be trivariate Normal then the method of moments estimating equations become

$$\bar{x} = \mu \tag{9.1}$$

$$\bar{y} = \alpha + \beta_2\sigma^2 \tag{9.2}$$

$$s_{xx} = \sigma^2 + \sigma_\delta^2 \tag{9.3}$$

$$s_{xy} = \beta_1\sigma^2 \tag{9.4}$$

$$s_{yy} = \beta_1^2\sigma^2 + 3\beta_2^2\sigma^4 + \sigma_\varepsilon^2 \tag{9.5}$$

and if σ_δ^2 and σ_ε^2 are assumed known then estimators for the remaining parameters may be found.

For example from (9.3) then $\tilde{\sigma^2} = s_{xx} - \sigma_\delta^2$ and substituting this into (9.4) gives $\tilde{\beta}_1 = \frac{s_{xy}}{s_{xx} - \sigma_\delta^2}$. This is exactly the same as the slope estimator estimated in Chapter 3 for a straight line errors in variables fit when σ_δ^2 was assumed known. Substituting the estimator for σ^2 into (9.5) yields

$$\tilde{\beta}_2 = \sqrt{\frac{\tilde{\beta}_1^2}{3}\left(\frac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}^2} - \frac{\tilde{\beta}_1}{s_{xy}}\right)}.$$

Assuming both σ_δ^2 and σ_ε^2 known may be impractical for many situations, nevertheless these estimators have been included to illustrate that unique solutions may be found from the method of moments estimating equations, for the quadratic structural model.

If the triple $(\xi, \delta, \varepsilon)$ is not assumed to be Normally distributed, and the restriction of having both σ_δ^2 and σ_ε^2 known is not made, then more method of moments estimating equations are needed. These may be obtained by appealing to the higher order moments, in an identical manner to Chapter 3. Indeed in standard quadratic regression s_{xxy} and s_{xxxx} are used. Again, as x_i is defined the same as in Chapter 3 then $s_{xxx} = \mu_{\xi 3} + \mu_{\delta 3}$ which introduces the new parameter $\mu_{\delta 3}$ to the model. The remaining method of moment estimating equations based on third order moments are $s_{xxy} = \beta_1\mu_{\xi 3} + \beta_2\mu_{\xi 4} + \beta_2\sigma^2\sigma_\delta^2$, $s_{xyy} = \beta_1^2\mu_{\xi 3} + \beta_2^2\mu_{\xi 5} + 2\beta_1\beta_2\mu_{\xi 4}$ which introduces the new parameter $\mu_{\xi 5}$ to the model and finally $s_{yyy} = \beta_1^3\mu_{\xi 3} + 3\beta_1^2\beta_2\sigma^4 + 3\beta_2\sigma^2\sigma_\varepsilon^2 + \beta_2^3\mu_{\xi 6} + \mu_{\varepsilon 3}$. This final estimating equation introduces two new parameters to the model, namely $\mu_{\xi 6}$ and μ_{ε_3}. As this equation introduces two new parameters that must be estimated, this equation will be ignored, and moment estimating equations based on fourth order moments derived instead. In addition, it is likely that the sixth central moment of ξ will be difficult to estimate. If this final equation is ignored then there are currently eight method of moments estimating equations and eleven unknown parameters.

By looking to fourth order moments it follows that $s_{xxxx} = \mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + \mu_{\delta 4}$ which introduces the new parameter $\mu_{\delta 4}$ to the model. A fourth order method of moments estimating equation which does not introduce a new parameter to the model is $s_{xxxy} = \beta_1^4\mu_{\xi 4} + \beta_1^3\beta_2\mu_{\xi 5} + 3\beta_1^2\sigma^2\sigma_\delta^2 + 3\beta_1\beta_2\mu_{\xi 3}\sigma_\delta^2 + \beta_2\mu_{\delta 3}\sigma_\varepsilon^2$.

Thus far twelve unknown parameters and ten method of moments estimating equations have been accumulated. One possibility is to assume that the errors δ_i are Normally distributed. Then $\mu_{\delta 3} = 0$ and $\mu_{\delta 4} = 3\sigma_\delta^4$. Under this assumption there are ten unknown parameters and ten method of moments estimating equations. The added bonus of making this assumption is that the form of the estimating equations become simpler. Under this restriction the method of moment estimating equations may be written as follows.

The first order estimating equations are:

$$\begin{aligned} \bar{x} &= \mu \\ \bar{y} &= \alpha + \beta_2\sigma^2 \end{aligned}$$

The second order estimating equations are:

$$\begin{aligned} s_{xx} &= \sigma^2 + \sigma_\delta^2 \\ s_{xy} &= \beta_1\sigma^2 + \beta_2\mu_{\xi 3} \\ s_{yy} &= \beta_1\sigma^2 + \beta_2^2\mu_{\xi 4} + 2\beta_1\beta_2\mu_{\xi 3} + \sigma_\varepsilon^2 \end{aligned}$$

Since $\mu_{\delta 3} = 0$ and $\mu_{\delta 4} = 3\sigma_\delta^4$ then the third and fourth order estimating equations are simplified. The third order estimating equations are:

$$\begin{aligned} s_{xxx} &= \mu_{\xi 3} \\ s_{xxy} &= \beta_1\mu_{\xi 3} + \beta_2\mu_{\xi 4} + \beta_2\sigma^2\sigma_\delta^2 \\ s_{xyy} &= \beta_1^2\mu_{\xi 3} + \beta_2^2\mu_{\xi 5} + 2\beta_1\beta_2\mu_{\xi 4} \end{aligned}$$

and finally the fourth order estimating equations are:

$$\begin{aligned} s_{xxxx} &= \mu_{\xi 4} + 6\sigma^2\sigma_\delta^2 + 3\sigma_\delta^4 \\ s_{xxxy} &= \beta_1^4\mu_{\xi 4} + \beta_1^3\beta_2\mu_{\xi 5} + 3\beta_1^2\sigma^2\sigma_\delta^2 + 3\beta_1^2\sigma^2\sigma_\delta^2 + 3\beta_1\beta_2\mu_{\xi 3}\sigma_\delta^2. \end{aligned}$$

If ξ is assumed to be Normally distributed then $\mu_{\xi 3} = \mu_{\xi 5} = 0$ and these equations would simplify further. There are obviously many questions remaining in fitting a quadratic model. There are different estimating equations that have not been derived here that may be of some use. Also, different restraints on the parameter space may be investigated. Once estimators for a quadratic model have been derived, then asymptotic variance covariance matrices may be constructed using the delta method as illustrated in Chapter 3. The details provided here are merely a starting point for investigating the fitting of quadratic models to data where there are errors in both variables. Nevertheless, it does seem that the method of moments may be able to at least aid with the fitting process, and would be much simpler than maximum likelihood.

Related to the fitting of a quadratic model, there are more obvious extensions of the work presented in this thesis. For example, a more complicated model involving a multiple regression of the form

$$\underline{y} = f(\underline{\xi}) + \underline{\varepsilon}$$

where $\underline{\xi}$, $\underline{y}$ and $\underline{\varepsilon}$ are now vectors containing numerous variables instead of single variables could be investigated for a variety of different functions f. As could be applied to the fitting of a straight line model, further work could be developed by relaxing or introducing various assumptions placed on the model. For example, in this thesis the errors δ and ε were assumed to be independent. Further investigative work could lie in assuming that these errors are no longer independent. Another option with a non-linear model is to transform the data so that a straight line may be fitted. The theory developed in this thesis may be applied. Work on transforming the data to linear form has been conducted by James [59].

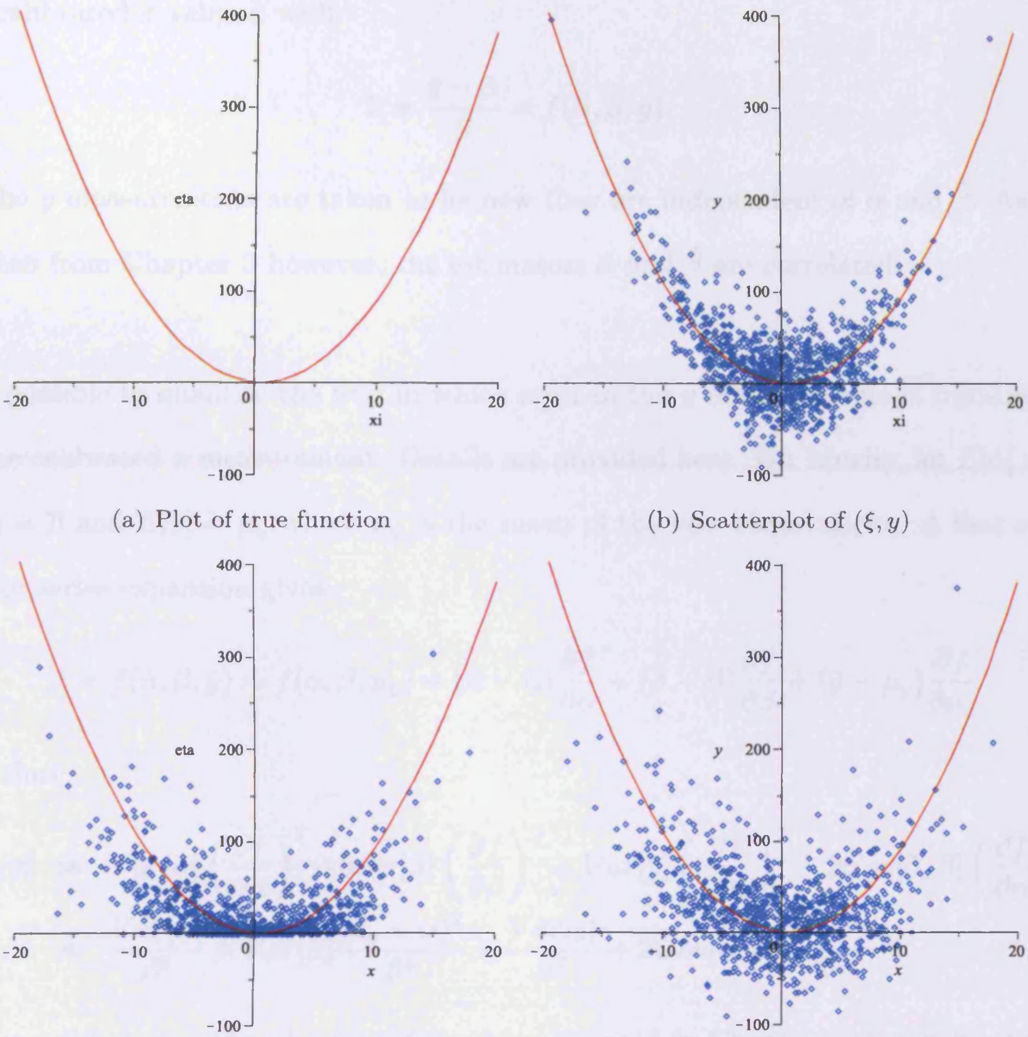
Later in this Chapter, the topic of migration for non-linear models will be discussed. Consider a set of ξ_i's generated from a Normal distribution with mean 0 and variance 25. Let the variable η_i be a quadratic function in ξ_i such that

$$\eta_i = 1 - 5\xi_i + 5\xi_i^2$$

and observations

$$\begin{aligned} x_i &= \xi_i + \delta_i \\ y_i &= \eta_i + \varepsilon_i \end{aligned}$$

are made on the latent variables ξ_i and η_i. Here δ_i and ε_i are generated from independent Normal distributions with mean 0 and variances 9^2 and 30^2 respectively. Figure 9.1 contains plots of the true function, and scatterplots of (ξ, y), (x, η) and (x, y) derived from the above model for 1000 data points. The migration effect here is to increase the scatter of points about the true quadratic, and to fill in the trough of the quadratic. The quadratic structure is still discernable, although the inflection of the quadratic has flattened and dissipated.



(a) Plot of true function (b) Scatterplot of (ξ, y)

(c) Scatterplot of (x, η) (d) Scatterplot of (x, y)

Figure 9.1: Scatterplots of data with true function $\eta_i = 1 - 5\xi_i + 5\xi_i^2$ plotted in red.

Back-calculation The topic of back-calculation and calibration is closely related to prediction. This topic is included here to demonstrate yet another facet of prediction that is largely ignored in the literature. The idea of back-calculation is to find an x (or ξ) given a new y (or even η). An intuitive approach is to use a fitted errors in variables line to obtain the calibrated x measurement. In other words, we interpolate the calibrated x value $\tilde{x}$ with

$$\tilde{x} = \frac{\bar{y} - \tilde{\alpha}}{\tilde{\beta}} = f(\tilde{\alpha}, \tilde{\beta}, \bar{y}).$$

As the y measurements are taken to be new they are independent of $\tilde{\alpha}$ and $\tilde{\beta}$. As can be seen from Chapter 3 however, the estimators $\tilde{\alpha}$ and $\tilde{\beta}$ are correlated.

It is possible to quantify the way in which error in the y measurements is transmitted to the calibrated x measurement. Details are provided here. For brevity, let $E[\tilde{\alpha}] = \alpha$, $E[\tilde{\beta}] = \beta$ and $E[\bar{y}] = \mu_y$ where μ_y is the mean of the new observations. A first order Taylor series expansion gives

$$\tilde{x} = f(\tilde{\alpha}, \tilde{\beta}, \bar{y}) \approx f(\alpha, \beta, \mu_y) + (\tilde{\alpha} - \alpha)\frac{\partial f}{\partial \alpha} + (\tilde{\beta} - \beta)\frac{\partial f}{\partial \beta} + (\bar{y} - \mu_y)\frac{\partial f}{\partial \mu_y}$$

and thus

$$\begin{aligned} Var[\tilde{x}] &\approx Var[\tilde{\alpha}]\left(\frac{\partial f}{\partial \alpha}\right)^2 + Var[\tilde{\beta}]\left(\frac{\partial f}{\partial \beta}\right)^2 + Var[\bar{y}]\left(\frac{\partial f}{\partial \mu_y}\right)^2 + 2Cov[\tilde{\alpha}, \tilde{\beta}]\left(\frac{\partial f}{\partial \alpha}\right)\left(\frac{\partial f}{\partial \beta}\right) \\ &= \frac{Var[\tilde{\alpha}]}{\beta^2} + Var[\tilde{\beta}]\frac{(\mu_y - \alpha)^2}{\beta^4} + \frac{Var[\bar{y}]}{\beta^2} + 2Cov[\tilde{\alpha}, \tilde{\beta}]\frac{(\mu_y - \alpha)}{\beta^3}. \end{aligned}$$

It is possible to use the 'shortcut' formulae presented in Chapter 3 that were used to construct the variance covariance matrices given in that Chapter. The ones of use here are:

$$\begin{aligned} Var[\tilde{\alpha}] &= \mu^2 Var[\tilde{\beta}] + \frac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{n} + 2\mu(\beta Cov[\bar{x}, \tilde{\beta}] - Cov[\bar{y}, \tilde{\beta}]) \\ Cov[\tilde{\alpha}, \tilde{\beta}] &= Cov[\bar{y}, \tilde{\beta}] - \beta Cov[\bar{x}, \tilde{\beta}] - \mu Var[\tilde{\beta}] \end{aligned}$$

If it is assumed that the Normal linear structural model applies, then $Cov[\bar{x}, \tilde{\beta}] = Cov[\bar{y}, \tilde{\beta}] = 0$. Thus,

$$Var[\tilde{x}] = \frac{Var[\tilde{\beta}]}{\beta^2}\left[\mu - \frac{(\mu_y - \alpha)}{\beta}\right]^2 + \frac{Var[\bar{y}]}{\beta^2} + \frac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{\beta^2 n}.$$

Otherwise

$$\begin{aligned} Var[\tilde{x}] &= \frac{Var[\tilde{\beta}]}{\beta^2}\left[\mu - \frac{(\mu_y - \alpha)}{\beta}\right]^2 + \frac{Var[\bar{y}]}{\beta^2} + \frac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{\beta^2 n} \\ &+ \frac{2\mu}{\beta^2}\left(\beta Cov[\bar{x},\tilde{\beta}] - Cov[\bar{y},\tilde{\beta}]\right) + \frac{2(\mu_y - \alpha)}{\beta^3}\left(Cov[\bar{y},\tilde{\beta}] - \beta Cov[\bar{x},\tilde{\beta}]\right). \end{aligned}$$

Here μ_y is the mean of the new y measurements, and would be estimated by the sample mean of the new measurements, and similarly, $Var[y]$ would be estimated by the sample variance of these measurements. The parameters μ, α, β, σ_δ and σ_ε would be estimated by the errors in variables fit.

The above result allows, for a number of replicated y measurements, a range of x to be found that corresponds to these measurements. This result however needs more investigation as to its suitability for an errors in variables model.

Migration in non-linear errors in variables models The migration effect for two linearly related variables with errors in both, is generally to distort the data away from the true line. This gives the impression that visually, the straight line errors in variables fit is not correct. For the Normal structural model for example, the migration effect rotates the data from the true line onto the y on x regression line. As seen in Chapter 7, the migration effect has a number of implications, most notably for residual analyses. Moreover, there is an additional problem in that the migration effect is different for different distributions of ξ (and indeed for different error distributions) and so an appreciation of the migration effect is needed to use errors in variables methodology effectively.

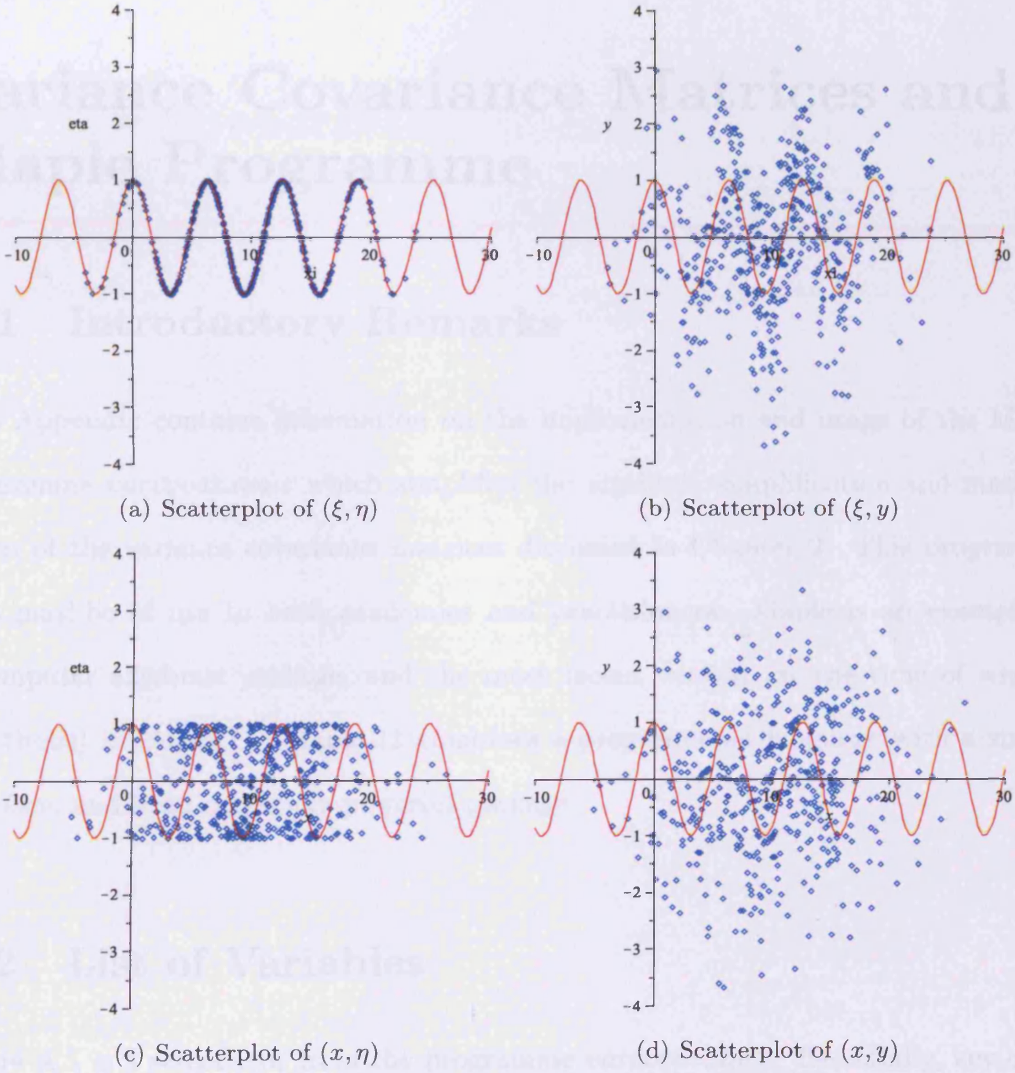
An obvious extension to the work on migration carried out in this thesis is to consider how the migration effect is mitigated in non-linear errors in variables models.

As an example of the migration effect for a particular non-linear model, Figure 9.2 contains scatterplots of (ξ, η), (ξ, y), (x, η) and (x, y) for a structural model with each ξ_i generated from a Normal distribution with $\mu = 10$ and $\sigma^2 = 25$ and

$$\eta_i = \cos(\xi_i).$$

The errors δ and ε were generated from independent Normal distributions with zero mean and variances $\sigma_\delta^2 = 4$ and $\sigma_\varepsilon^2 = 1$. It can be seen that the migration effect here, even for a reliability ratio of 0.8333 is notable. With error only present in y, the cosine wave is still discernable, though the apparent amplitude is considerably greater. The effect of the measurement error in the x is to completely obscure the cosine structure. Then when these two error measurements are combined, the migration effect completely destroys the underlying cosine structure. If faced with such a scatterplot, it is unlikely that a practitioner would believe that a non-linear errors in variables model with $\eta_i = \cos(\xi_i)$ is appropriate. The migration effect has dramatically increased the variation about the true curve, and has distorted both the peaks and the troughs of the cosine wave. The migration effect here is to fill the troughs, dissipate the peaks and to increase the variation of scatter around the cosine wave.

There is clearly much further work needed to fully appreciate the effect of migration. In this thesis, the implications of migration have been discussed in terms of residual analyses. The migration effect however would also distort the initial choice of model to be fitted. To find a general method which corrects for the effects of migration would be difficult as the migration effect depends on a number of factors, such as density of points, and the distribution of the random triple $(\xi, \delta, \varepsilon)$.



(a) Scatterplot of (ξ, η) (b) Scatterplot of (ξ, y)

(c) Scatterplot of (x, η) (d) Scatterplot of (x, y)

Figure 9.2: Scatterplots of data with true function $\eta_i = \cos(\xi_i)$ plotted in red.

Appendix A

Variance Covariance Matrices and Maple Programme

A.1 Introductory Remarks

This Appendix contains information on the implementation and usage of the Maple programme *varcovar.mws* which simplifies the algebraic simplification and manipulation of the variance covariance matrices discussed in Chapter 2. This programme thus may be of use to both academics and practitioners. Maple is an example of a computer algebraic package, and the most recent version (at the time of writing this thesis) is Maple 11. Maple 11 combines a programming language with a simple interface, and is an extremely powerful package.

A.2 List of Variables

Figure A.1 is a screenshot from the programme *varcovar.mws*. Essentially, key components of the variance covariance matrices are stored as short variable names, much similar to the way a calculator can store a number in its internal memory. For example the entire expression for $Var[s_{xx}]$ is stored as $v1$. This is the key advantage of the programme, long cumbersome expressions may be stored as a simple small variable.

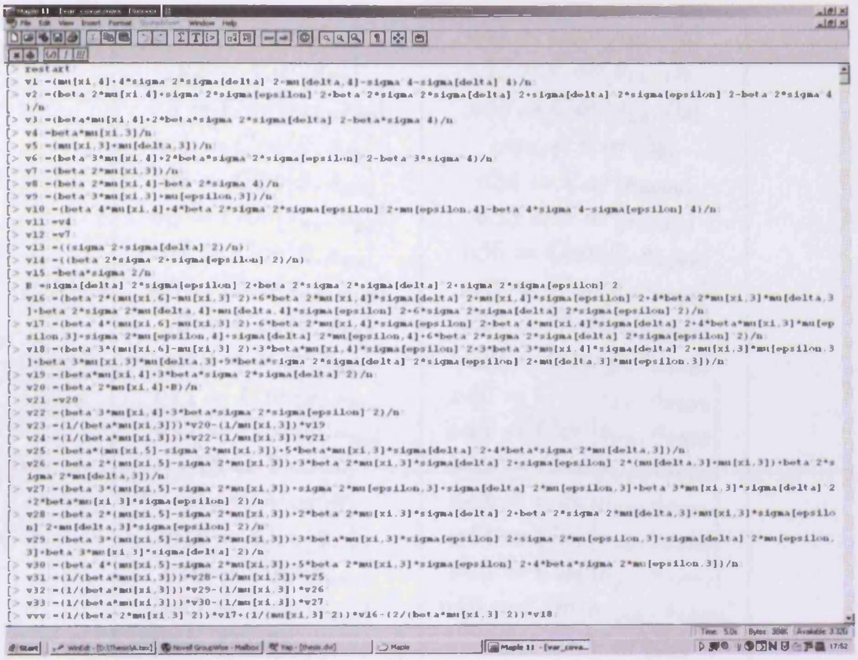


Figure A.1: A screenshot of the programme *varcovar.mws* opened in Maple 11.

Manipulation is therefore less time consuming. This is particularly beneficial for the slope estimators based on the higher order moments. The disadvantage of the programme is that due to the sheer volume of components which make up the variance covariance matrices, there is a large number of variables. The entire list is replicated here.

$v1 = Var[s_{xx}]$	$v31 = Cov[s_{xx}, \tilde{\beta}_8]$
$v2 = Var[s_{xy}]$	$v32 = Cov[s_{xy}, \tilde{\beta}_8]$
$v3 = Cov[s_{xx}, s_{xy}]$	$v33 = Cov[s_{yy}, \tilde{\beta}_8]$
$v4 = Cov[\bar{x}, s_{xy}]$	$vvv = Var[\tilde{\beta}_8]$
$v5 = Cov[\bar{x}, s_{xx}]$	$v34 = Var[s_{xyyy}]$
$v6 = Cov[s_{xy}, s_{yy}]$	$v35 = Var[s_{xxxy}]$
$v7 = Cov[\bar{x}, s_{yy}]$	$v36 = Cov[\bar{x}, s_{xyyy}]$
$v8 = Cov[s_{xx}, s_{yy}]$	$v37 = Cov[\bar{y}, s_{xyyy}]$
$v9 = Cov[\bar{y}, s_{yy}]$	$v38 = Cov[\bar{x}, s_{xxxy}]$
$v10 = Var[s_{yy}]$	$v39 = Cov[\bar{y}, s_{xxxy}]$
$v11 = Cov[\bar{y}, s_{xx}]$	$v40 = Cov[s_{xx}, s_{xyyy}]$
$v12 = Cov[\bar{y}, s_{xy}]$	$v41 = Cov[s_{xy}, s_{xyyy}]$
$v13 = Var[\bar{x}]$	$v42 = Cov[s_{yy}, s_{xyyy}]$
$v14 = Var[\bar{y}]$	$v43 = Cov[s_{xx}, s_{xxxy}]$
$v15 = Cov[\bar{x}, \bar{y}]$	$v44 = Cov[s_{xy}, s_{xxxy}]$
$v16 = Var[s_{xxy}]$	$v45 = Cov[s_{yy}, s_{xxxy}]$
$v17 = Var[s_{xyy}]$	$v46 = Cov[s_{xyyy}, s_{xxxy}]$
$v18 = Cov[s_{xxy}, s_{xyy}]$	$vvvv = Var[\tilde{\beta}_9]$
$v19 = Cov[\bar{x}, s_{xxy}]$	$v47 = Cov[\bar{x}, \tilde{\beta}_9]$
$v20 = Cov[\bar{x}, s_{xyy}]$	$v48 = Cov[\bar{y}, \tilde{\beta}_9]$
$v21 = Cov[\bar{y}, s_{xxy}]$	$v49 = Cov[s_{xx}, \tilde{\beta}_9]$
$v22 = Cov[\bar{y}, s_{xxy}]$	$v50 = Cov[s_{xy}, \tilde{\beta}_9]$
$v23 = Cov[\bar{x}, \tilde{\beta}_8]$	$v51 = Cov[s_{yy}, \tilde{\beta}_9]$
$v24 = Cov[\bar{y}, \tilde{\beta}_8]$	
$v25 = Cov[s_{xx}, s_{xxy}]$	
$v26 = Cov[s_{xy}, s_{xxy}]$	
$v27 = Cov[s_{yy}, s_{xxy}]$	
$v28 = Cov[s_{xx}, s_{xyy}]$	
$v29 = Cov[s_{xy}, s_{xyy}]$	
$v30 = Cov[s_{yy}, s_{xyy}]$	

A.3 Instruction Guide

Prior to performing any algebraic manipulation, the statements storing $v1$, $v2$ etc. need to be activated. To do this, the **ENTER** key must be pressed on each line of the code. This then stores the variables. To display $v1$ for example, we simply type in **v1;**. To perform algebraic manipulation we use conventional mathematics type. For example **(1/beta)*v1;** gives $\frac{Var[s_{xx}]}{\beta}$. If this is combined with the simplify statement, **simplify((1/beta)*v1);**, then Maple will simplify the final answer automatically.

Thus, in combination with the shortcut formulae of Chapter 2, the entire variance covariance matrices may be constructed using this Maple program. To substitute numerical values into an expression, we use the **subs** command. For example, to substitute $\beta = 4$, $\sigma = 2$, and $n = 100$ into $Cov[\bar{x}, \bar{y}]$ we use **subs(beta=4,sigma=2,n=100,v15);**.

The help system in Maple is detailed, and is a much valued resource. It may be accessed by the main toolbar, may also be accessed by the **?** command. For example, typing **?subs;** will open the help page on the **subs** command. For further details on Maple, then the book by Wright [113] is a comprehensive textbook covering many aspects of Maple programming.

Bibliography

[1] ADCOCK, R. J. Note on the method of least squares. *Analyst 4*, 6 (1877), 183–184.

[2] ADCOCK, R. J. A problem in least squares. *Analyst 5*, 2 (1878), 53–54.

[3] ANDERSON, T. W. Estimating linear statistical relationships. *Ann. Statist. 12* (1984), 1–45.

[4] BARKER, F., SOH, Y. C., AND EVANS, R. J. Properties of the geometric mean functional relationship. *Biometrics 44*, 1 (1988), 279–281.

[5] BARNETT, V. D. Fitting straight lines—The linear functional relationship with replicated observations. *J. Roy. Statist. Soc. Ser. C Appl. Statist. 19* (1970), 135–144.

[6] BARTLETT, M. S. Fitting a straight line when both variables are subject to error. *Biometrics 5* (1949), 207–212.

[7] BIRCH, M. W. A note on the maximum likelihood estimation of a linear structural relationship. *J. Amer. Statist. Assoc. 59* (1964), 1175–1178.

[8] BLAND, J. M., AND ALTMAN, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet i* (1986), 307–310.

[9] BLAND, M. *An Introduction to Medical Statistics*, Third ed. Oxford University Press, Oxford, 2000.

[10] BOWMAN, K. O., AND SHENTON, L. R. *Method of Moments.* Encyclopedia of Statistical Sciences (Volume 5). John Wiley & Sons, Inc, Canada, 1985.

[11] BOX, G. E. P. Science and statistics. *J. Amer. Statist. Assoc. 71*, 356 (1976), 791–799.

[12] BROWN, R. L. Bivariate structural relation. *Biometrika 44* (1957), 84–96.

[13] CARROLL, R. J., AND RUPPERT, D. The use and misuse of orthogonal regression in linear errors-in-variables models. *The American Statistician 50*, 1 (1996), 1–6.

[14] Carroll, R. J., Ruppert, D., and Stefanski, L. A. *Measurement Error in Nonlinear Models.* Chapman & Hall, London, 1995.

[15] Carroll, R. J., Ruppert, D., and Stefanski, L. A. *Measurement error in nonlinear models*, vol. 63 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London, 1995.

[16] Casella, G., and Berger, R. L. *Statistical Inference.* Wadsworth & Brooks, Pacific Grove, CA, 1990.

[17] Chan, N. N., and Mak, T. K. Heteroscedastic errors in a linear functional relationship. *Biometrika 71*, 1 (1984), 212–215.

[18] Cheng, C.-L., and Riu, J. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics 48*, 4 (2006), 511–519.

[19] Cheng, C.-L., and Van Ness, J. W. On the unreplicated ultrastructural model. *Biometrika 78*, 2 (1991), 442–445.

[20] Cheng, C.-L., and Van Ness, J. W. *Statistical Regression with Measurement Error.* Kendall's Library of Statistics 6. Arnold, London, 1999.

[21] Cheng, R. C. H., and Iles, T. C. Embedded models in three-parameter distributions and their estimation. *J. Roy. Statist. Soc. Ser. B 52*, 1 (1990), 135–149.

[22] Cochran, W. G. Some effects of errors of measurement on linear regression. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics* (Berkeley, Calif., 1972), Univ. California Press, pp. 527–539.

[23] Copas, J. B. The likelihood surface in the linear functional relationship problem. *J. Roy. Statist. Soc. Ser. B 34* (1972), 274–278.

[24] Copas, J. B. Regression, prediction and shrinkage. *J. Roy. Statist. Soc. Ser. B 45*, 3 (1983), 311–354.

[25] Cox, D. R. *Principles of statistical inference.* Cambridge University Press, Cambridge, 2006.

[26] Cox, N. R. The linear structural relation for several groups of data. *Biometrika 63*, 2 (1976), 231–237.

[27] Cragg, J. G. Using higher moments to estimate the simple errors-in-variables model. *The RAND Journal of Economics 28*, 0 (1997), S71–S91.

[28] Cramer, H. *Mathematical Methods of Statistics.* No. 9 in Princeton Mathematical Series. Princeton University Press, Princeton, 1946.

[29] Creasy, M. A. Confidence limits for the gradient in the linear functional relationship. *J. Roy. Statist. Soc. Ser. B. 18* (1956), 65–69.

[30] Dagenais, M. G., and Dagenais, D. L. Higher moment estimators for linear regression models with errors in the variables. *J. Econometrics 76*, 1-2 (1997), 193–221.

[31] Davidov, O. Estimating the slope in measurement error models - a different perspective. *Statistics and Probability Letters. 71*, 3 (2005), 215–223.

[32] DeGroot, M. H. *Probability and Statistics*, Second ed. Addison-Wesley, 1989.

[33] Deming, W. E. The application of least squares. *Philos. Mag. Ser. 7 11* (1931), 146–158.

[34] Di Marzio, M., and Taylor, C. C. Boosting kernel density estimates: a bias reduction technique? *Biometrika 91*, 1 (2004), 226–233.

[35] Dolby, G. R. A note on the linear structural relation when both residual variances are known. *J. Amer. Statist. Assoc. 71*, 354 (1976), 352–353.

[36] Dolby, G. R. The ultrastructural relation: a synthesis of the functional and structural relations. *Biometrika 63*, 1 (1976), 39–50.

[37] Draper, N. R., and Smith, H. *Applied Regression Analysis*, Third ed. Wiley-Interscience, Canada, 1998.

[38] Drion, E. F. Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. *Indagationes Math. 13* (1951), 256–260.

[39] Dunn, G. *Statistical Evaluation of Measurement Errors*, Second ed. Arnold, London, 2004.

[40] Durbin, J. Errors in variables. *Rev. Inst. Internat. Statist. 22* (1954), 23–32.

[41] Fuller, W. A. *Measurement error models.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1987.

[42] Galton, F. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute 15* (1886), 246–263.

[43] Geary, R. C. Inherent relations between random variables. *Proc. R. Irish. Acad. Sect. A. 47* (1942), 1541–1546.

[44] Geary, R. C. Relations between statistics: the general and the sampling problem when the samples are large. *Proc. R. Irish. Acad. Sect. A. 22* (1943), 177–196.

[45] Geary, R. C. Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica 17* (1949), 30–58.

[46] Geary, R. C. Sampling aspects of the problem from the error-in-variable approach. *Econometrica 17* (1949), 26–28.

[47] Gibson, W. M., and Jowett, G. H. Three-group regression analysis. Part 1: Simple regression analysis. *Appplied Statistics 6* (1957), 114–122.

[48] Glasbey, C. A., and Khondoker, M. R. Efficiency of functional regression estimators for combining multiple laser scans of cDNA microarrays. *To be published* (2007).

[49] Gleser, L. J. A note on G. R. Dolby's unreplicated ultrastructural model. *Biometrika 72*, 1 (1985), 117–124.

[50] Gleser, L. J., Carroll, R. J., and Gallo, P. P. The limiting distribution of least squares in an errors-in-variables regression model. *Ann. Statist. 15*, 1 (1987), 220–233.

[51] Gleser, L. J., and Hwang, J. T. The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Ann. Statist. 15*, 4 (1987), 1351–1362.

[52] Gleser, L. J., and Olkin, I. Estimation for a regression model with an unknown covariance matrix. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics* (Berkeley, Calif., 1972), Univ. California Press, pp. 541–568.

[53] Golub, G. H., and Van Loan, C. F. An analysis of the total least squares problem. *SIAM J. Numer. Anal. 17*, 6 (1980), 883–893.

[54] Gupta, Y. P., and Amanullah. A note on the moments of the Wald's estimator. *Statistica Neerlandica 24* (1970), 109–123.

[55] Hanley, J. A. "Transmuting" women into men: Galton's family data on human stature. *The American Statistician 58*, 3 (2004), 237–243.

[56] Hood, K. *Some Statistical Aspects of Method Comparison Studies.* Ph.d Thesis, Cardiff University, 1998.

[57] Hood, K., Nix, A. B. J., and Iles, T. C. Asymptotic information and variance-covariance matrices for the linear structural model. *The Statistician 48*, 4 (1999), 477–493.

[58] Hu, M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory 8*, 2 (1962), 179–187.

[59] JAMES, G. G. *Non-Linear Errors in Variables Regression Models.* PhD thesis, Cardiff University, 2001.

[60] JOHNSON, R. A., AND WICHERN, D. W. *Applied multivariate statistical analysis.* Prentice-Hall, Inc, 1992.

[61] JOLICOUER, P. Linear regressions in fishery research: some comments. *J. Fish. Res. Board Can. 32*, 8 (1975), 1491–1494.

[62] JORESKOG, K. G., AND SORBOM, D. *LISREL VI Analysis of linear structural relations by maximum likelihood, instrumental variables and least squares methods.* User's Guide, Department of Statistics, University of Uppsala, Uppsala, Sweden, 1984.

[63] JUDGE, G. G., GRIFFITHS, W. E., CARTER HILL, R., AND LEE, T.-C. *The Theory and Practise of Econometrics.* Wiley, New York, 1980.

[64] KAGAN, A., AND NAGAEV, S. How many moments can be estimated from a large sample? *Statistics and Probability Letters. 55*, 1 (2001), 99–105.

[65] KENDALL, M. G. Regression, structure and functional relationship. I. *Biometrika 38* (1951), 11–25.

[66] KENDALL, M. G. Regression, structure and functional relationship. II. *Biometrika 39* (1952), 96–108.

[67] KENDALL, M. G., AND STUART, A. *The Advanced Theory of Statistics Volume Two*, Third ed. Charles Griffin and Co Ltd, London, 1973.

[68] KODUAH, M. *Time-Specific reference intervals when there are errors in both variables.* PhD thesis, Cardiff University, 2004.

[69] KRZANOWSKI, W. J. *An Introduction to Statistical Modelling.* Arnold, London, 1998.

[70] KUMMEL, C. H. Reduction of observed equations which contain more than one observed quantity. *Analyst 6* (1879), 97–105.

[71] LAKSHMINARAYANAN, M. Y., AND GUNST, R. F. Estimation of parameters in linear structural relationships: sensitivity to the choice of the ratio of error variances. *Biometrika 71*, 3 (1984), 569–573.

[72] LINDLEY, D. V. Regression lines and the linear functional relationship. *Suppl. J. Roy. Statist. Soc. 9* (1947), 218–244.

[73] LONGFORD, N. T. Model selection and efficiency - is 'which model...?' the right question? *J. Roy. Statist. Soc. Ser. A. 168*, 3 (2005), 469–472.

[74] MADANSKY, A. The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc. 54* (1959), 173–205.

[75] MADDALA, G. S. *Introduction to Econometrics*, Second ed. Prentice Hall International, Inc, 1988.

[76] MARKOVSKY, I., AND VAN HUFFEL, S. On weighted structured total least squares. In *Large-scale scientific computing*, vol. 3743 of *Lecture Notes in Comput. Sci.* Springer, Berlin, 2006, pp. 695–702.

[77] MORAN, P. A. P. Estimating structural and functional relationships. *Multivariate Analysis 1* (1971), 232–255.

[78] NAIR, K. R., AND BANERJEE, K. S. A note on fitting of straight lines if both variables are subject to error. *Sankhya 6* (1942), 331.

[79] NEYMAN, J., AND SCOTT, E. L. Consistent estimates based on partial consistent observations. *Econometrica 16* (1948), 1–32.

[80] NEYMAN, J., AND SCOTT, E. L. On certain methods of estimating the linear structural relation. *Ann. Math. Statist. 22* (1951), 352–361.

[81] OKAMOTO, M. Asymptotoc theory of Brown–Fereday's method in a linear structural relationship. *J. Jap. Statist. Soc 13* (1983), 53–56.

[82] PAKES, A. On the asymptotic bias of the Wald-type estimators of a straight line when both variables are subject to error. *Int. Econ. Rev. 23* (1982), 491–497.

[83] PAL, M. Consistent moment estimators of regression coefficients in the presence of errors in variables. *J. Econometrics 14* (1980), 349–364.

[84] PATEFIELD, M. Fitting non-linear structural relationships using SAS procedure NLMIXED. *The Statistician 51*, 3 (2002), 355–366.

[85] PATEFIELD, W. M. Confidence intervals for the slope of a linear functional relationship. *Comm. Statist. A—Theory Methods 10*, 17 (1981), 1759–1764.

[86] PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag. 2* (1901), 559–572.

[87] RICKER, W. E. Linear regressions in fishery research. *J. Fish. Res. Board Can. 30* (1973), 409–434.

[88] RIGGS, D. S., GUARNIERI, J. A., AND ADDLEMAN, S. Fitting straight lines when both variables are subject to error. *Life Sciences 22* (1978), 1305–1360.

[89] ROYSTON, P. Constructing time-specific reference ranges. *Statistics in Medicine 10* (1991), 675–690.

[90] SCOTT, E. L. Note on consistent estimates of the linear structural relation between two variables. *Anal. Math. Stat. 21*, 2 (1950), 284–288.

[91] SEAL, H. L. *Multivariate Statistical Analysis for Biologists.* Methuen, London, 1964.

[92] SEBER, G. A. F., AND WILD, C. J. *Nonlinear Regression.* Wiley, New York, 1989.

[93] SELIKOWITZ, M. *Down Syndrome: The Facts*, second ed. Oxford University Press, 1997.

[94] SKRONDAL, A., AND RABE-HESKETH, S. *Generalized latent variable modeling.* Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, FL, 2004. Multilevel, longitudinal, and structural equation models.

[95] SOLARI, M. E. The 'maximum likelihood solution' to the problem of estimating a linear functional relationship. *J. Roy. Statist. Soc. Ser. B 31* (1969), 372–375.

[96] SPRENT, P. A generalized least-squares approach to linear functional relationships. *J. Roy. Statist. Soc. Ser. B 28* (1966), 278–297.

[97] SPRENT, P. *Models in Regression and Related Topics.* Methuen's Statistical Monographs. Matheun & Co Ltd, London, 1969.

[98] SPRENT, P. The saddlepoint of the likelihood surface for a linear functional relationship. *J. Roy. Statist. Soc. Ser. B 32* (1970), 432–434.

[99] SPRENT, P., AND DOLBY, G. R. Query: the geometric mean functional relationship. *Biometrics 36*, 3 (1980), 547–550.

[100] STEFANSKI, L. A. Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters 9*, 3 (1990), 229–235.

[101] STEFANSKI, L. A., AND COOK, J. R. Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association 90*, 432 (1995), 1247–1256.

[102] TAN, C. Y., AND IGLEWICZ, B. Measurement-methods comparisons and linear statistical relationship. *Technometrics 41*, 3 (1999), 192–201.

[103] TEISSIER, G. La relation d'allometrie sa signification statistique et biologique. *Biometrics 4*, 1 (1948), 14–53.

[104] TOBIN, J. Estimation of relationships for limited dependent variables. *Econometrica 26* (1958), 24–36.

[105] TSAY, R. S. Comment on a paper by Chatfield. *J. Bus. Econ. Statist. 11* (1993), 140–142.

[106] VAN HUFFEL, S., AND VANDERWALLE, J. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 1991.

[107] Van Montfort, K. *Estimating in Structural Models with Non-Normal Distributed Varibles: Some Alternative Approaches.* DSWO Press, Leiden, 1989.

[108] Van Montfort, K., Mooijaart, A., and de Leeuw, J. Regression with errors in variables: estimators based on third order moments. *Statist. Neerlandica 41*, 4 (1987), 223–237.

[109] Wachsmuth, A., Wilkinson, L., and Dallal, G. E. Galton's bend: a previously undiscovered nonlinearity in Galton's family stature regression data. *The American Statistician 57*, 3 (2003), 190–192.

[110] Wald, A. The fitting of straight lines if both variables are subject to error. *Ann. Math. Statistics 11* (1940), 285–300.

[111] Wong, M. Y. Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika 76*, 1 (1989), 141–148.

[112] Woodhouse, R. Graphical solutions for structural regression assist errors-in-variables modelling. *Journal of Applied Statistics 33*, 3 (2006), 241–255.

[113] Wright, F. J. *Computing With Maple.* CRC Press, USA, 2002.