REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 57–80

AN OVERVIEW OF LINEAR STRUCTURAL MODELS IN ERRORS IN VARIABLES REGRESSION

Author: Jonathan Gillard
– Cardiff School of Mathematics, Cardiff University,
Senghennydd Road, Cardiff, UK
GillardJW@Cardiff.ac.uk

Received: May 2009 Revised: May 2010 Accepted: May 2010

Abstract:

- This paper aims to overview the numerous approaches that have been developed to estimate the parameters of the linear structural model. The linear structural model is an example of an errors in variables model, or measurement error model that has wide practical use. This paper brings together key concepts from a scattered literature to give an accessible account of existing work on this particular errors in variables model.

Key-Words:

- *errors in variables; regression; measurement error; linear structural model.*

AMS Subject Classification:

- 62-02, 62J05.

1. INTRODUCTION

1.1. Aim of errors in variables modelling

Given a set of variables, a common statistical procedure is to try and find relationships between them. A technique that may aid with this is regression, which can provide an estimate of the formulaic relationship between these variables. The relationships between these variables, if they exist, may be linear or non-linear. Commonly the variables are split into dependent and independent variables, and regression analyses are concerned with writing the dependent variables in terms of some function of the independent variables. Standard regression procedures assume that the independent variables are measured without error, and that the error inherent in the model is associated with the dependent variables only. The theory of fitting such models is plentiful, and is well documented in the literature. An obvious extension to this model is to assume that there is error also present in the independent variables. This has become known as the errors in variables situation. There are errors in the measurement of both the independent and dependent variables, and so usually a different tack is called for. Indeed, in [9] Casella and Berger wrote 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."

1.2. Common applications

Errors in variables models have been successfully applied to a number of different subject areas over the years. Indeed, different ways of solving the problem of having errors in variables have become associated with different subject areas. For example, the method of using instrumental variables had its origins in the economic literature, but this technique is not restricted to economic applications. Use of errors in variables methodology has proved fruitful in areas as diverse as astrostatistics, as a method to cope with astronomical measurement error, in fisheries statistics, as a way of looking at fish stocks, in medical statistics, commonly in method comparison studies when a number of methods of measurement are compared and much more. The errors in variables problem is also one that is inherently interesting theoretically, and a number of theoretical developments in their own right have been made. In particular, numerical analysts have been interested in the development and study of particular types of numerical algorithms to solve the errors in variables problem.

Errors in variables models tend to be appropriate when all variables are experimentally observed. Each variable is then subject to its own inherent measurement error. Despite their common application, errors in variables methodology is still quite neglected in practice. This could be for a number of reasons. Firstly, the literature on errors in variables topics is widely scattered, appearing in a range of journals, in a number of different contexts. Secondly, the notation used for errors in variables models varies tremendously. Thus it is sometimes difficult to read papers from different sources. Finally, there are a number of different approaches to fit an errors in variables model. Some of these will be described in this paper. The aim of this paper is to bring ideas from this widely scattered literature together, and to explain the development of key methodologies and links between them.

For brevity, this paper will focus on the linear structural model which is a commonly fitted errors in variables type model. Section 2 describes the linear structural model. Section 3 outlines the main approaches that have been adopted to estimate the parameters of the linear structural model.

2. AN INTRODUCTION TO THE LINEAR STRUCTURAL MODEL

Consider two variables, ξ and η which are linearly related in the form

$$\eta_i = \alpha + \beta\,\xi_i\ , \qquad i = 1, ..., n\ .$$

However, instead of observing ξ_i and η_i, we observe

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

where δ_i and ε_i are considered to be random error components, or noise.

It is assumed that $E[\delta_i] = E[\varepsilon_i] = 0$ and that $\mathrm{Var}[\delta_i] = \sigma_\delta^2$, $\mathrm{Var}[\varepsilon_i] = \sigma_\varepsilon^2$ for all i. Also the errors δ_i and ε_i are mutually uncorrelated. Thus

$$\begin{aligned} \mathrm{Cov}[\delta_i, \delta_j] \ &= \ \mathrm{Cov}[\varepsilon_i, \varepsilon_j] \ = \ 0\ , \qquad i \neq j\ , \\ \mathrm{Cov}[\delta_i, \varepsilon_j] \ &= \ 0\ , \qquad \forall\, i, j\ . \end{aligned}$$

It is possible to rewrite the above model as

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

This highlights the difference between this problem and the standard regression model. The error term is clearly dependent on β. In addition to this term $(\varepsilon - \beta\delta)$ is correlated with x. Indeed,

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

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 apply. See [16] for details on standard regression models.

There have been several reviews of errors in variables methods, notably [9], [10], [20], [37] and [55]. Unfortunately the notation has not been standardised. This paper closely follows the notation set out by Cheng and Van Ness in [10] but for convenience, it has been necessary to modify parts of their notation. All notation will be carefully introduced at the appropriate time.

Errors in variables modelling can be split into two general classifications defined in [35], and [36], as the functional and structural models. The fundamental difference between these models lies in the treatment of the ξ_i's:

The functional model – This assumes the ξ_i's to be unknown, but fixed constants μ_i.

The structural model – This model assumes the ξ_i's to be a random sample from a random variable with mean μ and variance σ^2. The linear structural model is thus the linear model described above, with the ξ_i's taken in a structural setting.

Due to the wealth of literature available this paper will focus mainly on the linear structural model. It will however prove prudent at times to mention the linear functional model at certain places in the text. Further information on the linear functional model is provided in [26].

3. AN OVERVIEW OF ERRORS IN VARIABLES MODELLING

3.1. Origins and beginnings

The author first associated with the errors in variables problem was Adcock ([1], [2]). In the late 1800s 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 equal error variances ($\sigma_\delta^2 = \sigma_\varepsilon^2$). These ideas are linked to 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 onto the regression line.

Adcock's work was extended a year later by Kummel in [38]. Instead of assuming equal error variances, he assumed that the ratio $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ was known instead. 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. Use of the orthogonal regression line has been questioned by some authors 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 going to be 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.

The idea of orthogonal regression was included in a book by Deming in [15], and so orthogonal regression is sometimes referred to as Deming regression, particularly in the medical literature. 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. The least squares method can then be used to minimise this residual error. This assumes that the error structure is homoscedastic, otherwise this method cannot be used. Lindley in [39] found that adding a weighting factor when minimising the sum of squares of the orthogonal projections, allowed one to minimise projections other than orthogonal.

Another early paper on this subject was by Pearson ([49]). He extended the ideas of previous authors to allow the fitting of lines and hyperplanes 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.

3.2. Grouping methods

A different approach was suggested by Wald in [62]. Wald described a method that did not make an assumption 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 $\big(x_{(1)}, y_{(1)}\big), ..., \big(x_{(m)}, y_{(m)}\big)$ and G_2 contains the second half $\big(x_{(m+1)}, y_{(m+1)}\big), ..., \big(x_{(n)}, y_{(n)}\big)$. An estimate of the slope is then

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

A problem here is that the grouping must be based on the order of the true values, otherwise, in general, the groups are not independent of the error terms $\delta_1, ..., \delta_n$. 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 [29].

The idea of grouping the observations was further developed by Bartlett in [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 [44]) 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 by drawing 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. In [25] advice on how to place the data into these three groups to obtain the most efficient estimate of the slope is given. 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 [40].

Neyman and Scott in [46] 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

$$\tilde{\beta}_{\mathrm{NS}} = \frac{\bar{y}_{G_3} - \bar{y}_{G_1}}{\bar{x}_{G_3} - \bar{x}_{G_1}}$$

and is a consistent estimator of β if

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

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

Grouping methods, in particular Wald's method, have been criticised by Pakes in [47]. He claimed that the work of in [29] is unnecessary as Wald's estimate is, strictly speaking, inconsistent. Letting $\tilde{\beta}_{\mathrm{W}}$ denote Wald's estimate for the slope, Pakes showed

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

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

However, this expression derived by Pakes offers a similar conclusion to that of Neyman and Scott ([45]). As long as the error δ is bounded (or not too significant) 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 $E\big[\delta|x \in G_2\big] - E\big[\delta|x \in G_1\big]$ should be negligible.

3.3. 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 estimate for the slope is then

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

where, s_{yw} and s_{xw} are the usual second order sample moments defined as

$$s_{ab} = \frac{1}{n} \sum_{i=1}^{n} (a_i - \bar{a}) \, (b_i - \bar{b}) \, ,$$

and $\bar{a} = n^{-1} \sum_{i=1}^{n} a_i$ is the sample mean. 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. In [41] it was showed that Wald's grouping method is equivalent to using the instrumental variable

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

and similarly Bartlett's grouping method is equivalent to using

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

An idea using the ranks of the x_i was proposed by Durbin in [19]. He suggested an estimator of the form

$$\tilde{\beta}_{\mathrm{D}} = \frac{\sum_{i=1}^{n} i \, y_{(i)}}{\sum_{i=1}^{n} i \, x_{(i)}}$$

where $\big(x_{(1)}, y_{(1)}\big), \big(x_{(2)}, y_{(2)}\big), ..., \big(x_{(n)}, y_{(n)}\big)$ are the ordered observations. However, as with grouping methods, it is unlikely that the ranks of the observed data will match the ranks of the true data. So as in Wald's method this estimate is inconsistent.

3.4. 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 estimate

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

There is a geometric interpretation of the line having this slope — it is the line giving the minimum sum of products of the horizontal and vertical distances of the observations from the line (Teissier [58]). However, for the estimate to be unbiased (see [32] for example), one must assume that

$$\lambda = \beta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2} \, . \tag{3.1}$$

This is due to

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

A technical criticism of the use of this estimator is that it may have infinite variance (Creasy [13]). 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 in [50] 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 in [32], again in the fisheries literature, discussed the paper by Ricker. He stated that as geometric mean regression is equivalent to the assumption in equation (3.1) 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 (3.1) 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 (3.1) holds, is a perfectly valid regression tool.

Further discussion on this estimate was initiated by Sprent and Dolby initially in [57]. They discouraged the use of geometric mean regression, due to the unrealistic assumption of (3.1). They both however sympathised with practitioners, especially those in fish biology, who do not have any knowledge regarding λ. 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.

An alternative way of looking at geometric mean regression was provided by Barker *et al.* in [4]. Instead of looking at it as a geometrical average, it can be derived in its own right by adopting a so-called least triangles approach. This is where the sum of the areas of the right-angled triangles formed from the horizontal discrepancies from the data point to the regression line, the vertical discrepancies from the data point to the regression line, and the regression line itself, are minimised. They also showed a connection between geometric mean regression and the correlation coefficient, thus refuting the claim by Sprent and Dolby made in [57] that the correlation coefficient has little value in errors in variables modelling.

3.5. Cumulants

Another method of estimation that has been used in errors in variables modelling is the method of moments. A closely related approach to this is using cumulants, which were proposed by Geary in the series of papers [21], [22], [24], [23]. 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

$$\psi(t_1,t_2)=\ln\big[\phi(t_1,t_2)\big]=\ln\big[E\big(e^{it_1X+it_2Y}\big)\big]=\sum_{r,s=0}^{\infty}\kappa(r,s)\,\frac{(it_1)^r}{r!}\,\frac{(it_2)^s}{s!}\,. \tag{3.2}$$

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.

If the true values ξ and η are centred with respect to their true mean, then

the intercept vanishes, and we can write the structural relationship in the form

$$\beta\xi - \eta = 0\,. \tag{3.3}$$

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 [10], [48]):

- 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

$$\phi(t_1, t_2) = E\big[e^{it_1\xi + it_2\eta}\big]\,. \tag{3.4}$$

It follows from (3.3) and (3.4) that

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

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

$$\beta\frac{\partial\psi}{\partial it_1} - \frac{\partial\psi}{\partial it_2} = \frac{1}{\phi}\left(\beta\frac{\partial\phi}{\partial it_1} - \frac{\partial\phi}{\partial it_2}\right) = 0 \tag{3.5}$$

and it follows from (3.2) and (3.5), for all $r, s > 0$

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

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

$$\tilde{\beta}_{\mathrm{C}} = \frac{\kappa(r, s+1)}{\kappa(r+1, s)}\,.$$

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 [21].

3.6. 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 is considered in detail in [27], and so only a brief survey of the existing literature is given here. Kendall and Stuart in [37] 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 [10], [18], and [37], as well as others. Dunn in [18] gave formulae for many of the estimators of the slope that are included in [27]. However he did not give any information regarding estimators based on higher moments. Neither did he give information about the variances of these estimates. Work on the higher order moment estimating equations has been done in [17], and more recently in [48], [61], [60] and [12]. Drion, in a paper that is infrequently cited [17], looked at an estimate 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 estimate of the slope is consistent. Prior to this work, Scott in [52] considered the structural model, and also found an estimate 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

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

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 $\mu_{\xi,2k+1}$ $(k = 1, 2, ..., m)$ differs from zero, then the equation

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

has a root $\hat{b}$ which is a consistent estimate of β. Scott did warn however, that estimates 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 will provide the consistent estimate.

More recently, Pal in [48] 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 estimates, and concluded that there must be infinitely many consistent estimates which can be obtained by taking different functions of the slope estimates he derived. For each

of the six estimates, 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 $\frac{\sigma_\delta^2}{\sigma^2} = 0$, to offer advice as to which estimator has the smallest variance. The asymptotic efficiency of a particular estimate with respect to the least squares estimate was also provided, for different distributions of ξ. A brief review of 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.* in [61] gave a detailed survey on estimators based on third order moments. They provided an optimal estimate of the slope which is a function of three slope estimates. In order to obtain this optimal estimate, the variance-covariance matrix 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 lower than three appear in the estimation of the variance-covariance matrix. Van Montfort *et al.*, through a simulation study, demonstrated that the optimal estimate 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 estimates 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 estimate are not altered.

A detailed account of alternative approaches to errors in variables modelling was written by Van Montfort in [60]. 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 estimates were provided. This text is an extension of the details included in [61].

The most recent account on using higher moments was that by Cragg in [12]. 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 will be needed to obtain a reliable estimate. 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. 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 on 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 in [14]. 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 estimate.

3.7. 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 the pairs of observations (x_i, y_i) are jointly normally and identically distributed. Lindley was one of the first authors to use maximum likelihood estimation for the errors in variables model in [39]. 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 again in [37] 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 $\mu, \alpha, \beta, \sigma_x^2, \sigma_\delta^2$ 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 estimating using the method of moment estimating equations is included in [27]. More detail was given on the problem of having five moment estimating equations, and six parameters to estimate. They suggested various 'cases', each of which consist of a different assumption regarding a subset of the parameters. Estimates 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. 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 by Hood *et al.* in [30]. 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 in [63] 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 about regarding the slope, and the density function for the slope are also included. Prior to this, Barnett also commented on the inherent difficulties in using the maximum likelihood technique in [5].

Again for the structural model, Birch in [7] showed that the maximum likelihood estimate 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 estimates provided by Madansky in [40] 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.

Solari in [54] 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 considered Solari's work and further noted the practical implications of her findings in [56].

Copas in [11] extended the work of Solari [54]. He showed that when 'rounding-off' errors for 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° of the x-axis. The x on y estimate is used if the geometric mean estimate lies within 45° 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

$$L = \prod_i P_i(x_i)\, Q_i(y_i)$$

where $P_i(x) = P\big(x - \frac{h}{2} \leq \xi_i < x + \frac{h}{2}\big)$ and $Q_i(x) = P\big(y - \frac{h}{2} \leq \beta\xi_i < y + \frac{h}{2}\big)$ (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

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

and

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

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 such as y on x and x on y regression 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.

3.8. 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 in [28], 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

$$\mathbf{A}\,\mathbf{X} \simeq \mathbf{B}$$

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 $\mathbf{D} = [\mathbf{A}\,\mathbf{B}]$ 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 $\|\mathbf{A}\,\mathbf{X} - \mathbf{B}\|$ 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 [28] 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 [42]. A complete monograph on the topic has been written and is contained in [59]. Cheng and Van Ness in [10] 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.

3.9. LISREL

As well as total least squares, another method of estimation which had its origins from computational mathematics is LISREL (which stands for Linear Structural Relationships). 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 [33]). To use their notation, the LISREL model is formulated as follows:

$$\underline{\eta} = \mathbf{B}\underline{\eta} + \Gamma\underline{\xi} + \underline{\zeta}\,, \tag{3.6}$$

$$\underline{Y} = \Lambda_y\underline{\eta} + \underline{\varepsilon}\,, \tag{3.7}$$

$$\underline{X} = \Lambda_x\underline{\xi} + \underline{\delta}\,, \tag{3.8}$$

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 an $(m\times 1)$ 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 combines two factor analysis models, (3.7) and (3.8) into the structural setting of equation (3.6).

The linear structural model outlined in Section 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 Chapter 1. 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 Winchern in [31] gave details of the structure, and stated the following identities (they took $\mathbf{B} = 0$ to simplify proceedings)

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

where $E\big[\underline{\xi}\,\underline{\xi}^{\mathrm{T}}\big] = \Phi$, $E\big[\underline{\delta}\,\underline{\delta}^{\mathrm{T}}\big] = \Theta_\delta$, $E\big[\underline{\varepsilon}\,\underline{\varepsilon}^{\mathrm{T}}\big] = \Theta_\varepsilon$ and $E\big[\underline{\zeta}\,\underline{\zeta}^{\mathrm{T}}\big] = \psi$. 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 Winchern 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 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 environment to estimate our parameters unless we assume something further is known. This ties in with the thoughts of Madansky who stated in [40] 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 the linear structural model of Section 2, we get

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

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

$$\sigma^2 + \sigma_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s_{xx}\,, \tag{3.9}$$

$$\beta^2\sigma^2 + \sigma_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s_{yy}\,, \tag{3.10}$$

$$\beta\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = s_{xy}\,, \tag{3.11}$$

which are identical to the method of moment estimating equations (and subsequently the maximum likelihood estimating equations) outlined in [27].

The first order moment equations $\mu = \bar{x}$ and $\alpha + \beta\mu = \bar{y}$ are missing as the LISREL model assumes the data are centred, so μ and α are taken as known in the assumption $E[\underline{\xi}] = E[\underline{\eta}] = 0$. There are three equations (3.9), (3.10), (3.11) 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 is 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 [27].

3.10. Review papers and monographs

Over the years several authors have written review articles on errors in variables regression. These include [35], [36], [19], [40], [43] and [3]. Riggs *et al.* in [51] 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 in [20] and Cheng and Van Ness with [10]. Casella and Berger in their general text [9] has an informative section on the topic, [55] contains chapters on the problem, as do [37] and [18]. Draper and Smith in [16] 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 [34]. In these texts the method of instrumental variables is often given prominence. The text [8] described errors in variables models for non linear regression, and Seber and Wild in [53] included a chapter on this topic.

3.11. Conclusion

The papers described in this presentation are definitive papers that dictated the path of further research in the topic. The sporadic nature of the literature can be seen by looking at the journals from which the papers in this presentation came. Procedures for fitting errors in variables models have been developed in the medical literature, economics literature and statistics literature. There are a plethora of papers available on the linear structural model, and even more on errors in variables in general. The list of references given in this paper are by no means exhaustive, but it is hoped that consolidating some of the key ideas involved in errors in variables modelling into this paper will help stimulate further research into a problem that has existed since the 1800s, and that has interested people in a variety of academic disciplines.

REFERENCES

[1] Adcock, R.J. (1877). Note on the method of least squares, *The Analyst*, **4**(6), 183–184.

[2] Adcock, R.J. (1878). A problem in least squares, *The Analyst*, **5**(2), 53–54.

[3] Anderson, T.W. (1984). Estimating linear statistical relationships, *Ann. Statist.*, **12**, 1–45.

[4] Barker, F.; Soh, Y.C. and Evans, R.J. (1988). Properties of the geometric mean functional relationship, *Biometrics*, **44**(1), 279–281.

[5] Barnett, V.D. (1970). Fitting straight lines – The linear functional relationship with replicated observations, *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, **19**, 135–144.

[6] Bartlett, M.S. (1949). Fitting a straight line when both variables are subject to error, *Biometrics*, **5**, 207–212.

[7] Birch, M.W. (1964). A note on the maximum likelihood estimation of a linear structural relationship, *J. Amer. Statist. Assoc.*, **59**, 1175–1178.

[8] Carroll, R.J.; Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

[9] Casella, G. and Berger, R.L. (1990). *Statistical Inference*, Wadsworth & Brooks, Pacific Grove, CA.

[10] Cheng, C.-L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*, Kendall's Library of Statistics, **6**, Arnold, London.

[11] Copas, J.B. (1972). The likelihood surface in the linear functional relationship problem, *J. Roy. Statist. Soc. Ser. B*, **34**, 274–278.

[12] Cragg, J.G. (1997). Using higher moments to estimate the simple errors-in-variables model, *The RAND Journal of Economics*, **28**(0), S71–S91.

[13] Creasy, M.A. (1956). Confidence limits for the gradient in the linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **18**, 65–69.

[14] Dagenais, M.G. and Dagenais, D.L. (1997). Higher moment estimators for linear regression models with errors in the variables, *J. Econometrics*, **76**(1–2), 193–221.

[15] Deming, W.E. (1931). The application of least squares, *Philos. Mag. Ser.*, **7**(11), 146–158.

[16] Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*, Third Ed., Wiley-Interscience, Canada.

[17] Drion, E.F. (1951). 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**, 256–260.

[18] Dunn, G. (2004). *Statistical Evaluation of Measurement Errors*, Second Ed., Arnold, London.

[19] Durbin, J. (1954). Errors in variables, *Rev. Inst. Internat. Statist.*, **22**, 23–32.

[20] Fuller, W.A. (1987). *Measurement error models*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

[21] Geary, R.C. (1942). Inherent relations between random variables, *Proc. R. Irish. Acad. Sect. A*, **47**, 1541–1546.

[22] Geary, R.C. (1943). Relations between statistics: the general and the sampling problem when the samples are large, *Proc. R. Irish. Acad. Sect. A*, **22**, 177–196.

[23] Geary, R.C. (1949). Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown, *Econometrica*, **17**, 30–58.

[24] Geary, R.C. (1949). Sampling aspects of the problem from the error-in-variable approach, *Econometrica*, **17**, 26–28.

[25] Gibson, W.M. and Jowett, G.H. (1957). Three-group regression analysis – Part 1: Simple regression analysis, *Appplied Statistics*, **6**, 114–122.

[26] Gillard, J.W. (2007). *Errors in Variables Regression: What is the appropriate model?* Cardiff University PhD Thesis.

[27] Gillard, J.W. and Iles, T.C. (2005). Method of moments estimation in linear regression with errors in both variables, *Cardiff School of Mathematics Technical Report*.

[28] Golub, G.H. and Van Loan, C.F. (1980). An analysis of the total least squares problem, *SIAM J. Numer. Anal.*, **17**(6), 883–893.

[29] Gupta, Y.P. and Amanullah. (1970). A note on the moments of the Wald's estimator, *Statistica Neerlandica*, **24**, 109–123.

[30] Hood, K.; Nix, A.B.J. and Iles, T.C. (1999). Asymptotic information and variance-covariance matrices for the linear structural model, *The Statistician*, **48**(4), 477–493.

[31] Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc.

[32] Jolicouer, P. (1975). Linear regressions in fishery research: some comments, *J. Fish. Res. Board Can.*, **32**(8), 1491–1494.

[33] Joreskog, K.G. and Sorbom, D. (1984). *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.

[34] Judge, G.G.; Griffiths, W.E.; Carter Hill, R. and Lee, T.-C. (1980). *The Theory and Practise of Econometrics*, Wiley, New York.

[35] Kendall, M.G. (1951). Regression, structure and functional relationship – I, *Biometrika*, **38**, 11–25.

[36] Kendall, M.G. (1952). Regression, structure and functional relationship – II, *Biometrika*, **39**, 96–108.

[37] Kendall, M.G. and Stuart, A. (1973). *The Advanced Theory of Statistics*, Volume Two, Third Ed., Charles Griffin and Co. Ltd., London.

[38] Kummel, C.H. (1879). Reduction of observed equations which contain more than one observed quantity, *The Analyst*, **6**, 97–105.

[39] Lindley, D.V. (1947). Regression lines and the linear functional relationship, *Suppl. J. Roy. Statist. Soc.*, **9**, 218–244.

[40] Madansky, A. (1959). The fitting of straight lines when both variables are subject to error, *J. Amer. Statist. Assoc.*, **54**, 173–205.

[41] Maddala, G.S. (1988). *Introduction to Econometrics*, Second Ed., Prentice Hall International, Inc.

[42] Markovsky, I. and Van Huffel, S. (2006). On weighted structured total least squares. In "Large-Scale Scientific Computing", *Lecture Notes in Comput. Sci.*, **3743**, Springer, Berlin, pp. 695–702.

[43] Moran, P.A.P. (1971). Estimating structural and functional relationships, *Multivariate Analysis*, **1**, 232–255.

[44] Nair, K.R. and Banerjee, K.S. (1942). A note on fitting of straight lines if both variables are subject to error, *Sankhya*, **6**, 331.

[45] Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partial consistent observations, *Econometrica*, **16**, 1–32.

[46] Neyman, J. and Scott, E.L. (1951). On certain methods of estimating the linear structural relation, *Ann. Math. Statist.*, **22**, 352–361.

[47] Pakes, A. (1982). On the asymptotic bias of the Wald-type estimators of a straight line when both variables are subject to error, *Int. Econ. Rev.*, **23**, 491–497.

[48] Pal, M. (1980). Consistent moment estimators of regression coefficients in the presence of errors in variables, *J. Econometrics*, **14**, 349–364.

[49] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Philos. Mag.*, **2**, 559–572.

[50] Ricker, W.E. (1973). Linear regressions in fishery research, *J. Fish. Res. Board Can.*, **30**, 409–434.

[51] Riggs, D.S.; Guarnieri, J.A. and Addleman, S. (1978). Fitting straight lines when both variables are subject to error, *Life Sciences*, **22**, 1305–1360.

[52] Scott, E.L. (1950). Note on consistent estimates of the linear structural relation between two variables, *Anal. Math. Stat.*, **21**(2), 284–288.

[53] Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*, Wiley, New York.

[54] Solari, M.E. (1969). The 'maximum likelihood solution' to the problem of estimating a linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **31**, 372–375.

[55] Sprent, P. (1969). *Models in Regression and Related Topics*, Methuen's Statistical Monographs, Matheun & Co Ltd, London.

[56] Sprent, P. (1970). The saddlepoint of the likelihood surface for a linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **32**, 432–434.

[57] Sprent, P. and Dolby, G.R. (1980). Query: the geometric mean functional relationship, *Biometrics*, **36**(3), 547–550.

[58] Teissier, G. (1948). La relation d'allometrie sa signification statistique et biologique, *Biometrics*, **4**(1), 14–53.

[59] Van Huffel, S. and Vanderwalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia.

[60] Van Montfort, K. (1989). *Estimating in Structural Models with Non-Normal Distributed Variables: Some Alternative Approaches*, DSWO Press, Leiden.

[61] Van Montfort, K.; Mooijaart, A. and de Leeuw, J. (1987). Regression with errors in variables: estimators based on third order moments, *Statist. Neerlandica*, **41**(4), 223–237.

[62] Wald, A. (1940). The fitting of straight lines if both variables are subject to error, *Ann. Math. Statistics*, **11**, 285–300.

[63] Wong, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error, *Biometrika*, **76**(1), 141–148.