Method of Moments Estimation in Linear Regression with Errors in both Variables

by

J.W. Gillard and T.C. Iles

Cardiff University School of Mathematics Technical Paper

October 2005

Cardiff University School of Mathematics, Senghennydd Road, Cardiff, CF24 4AG

Contents

1. Introduction

The problem of fitting a straight line to bivariate (x, y) data where the data are scattered about the line is a fundamental one in statistics. Methods of fitting a line are described in many statistics text books, for example Draper and Smith (1998) and Kleinbaum et al (1997). The usual way of fitting a line is to use the principle of least squares, finding the line that has the minimum sum of the squares of distances of the points to the line in the vertical y direction. This line is called the regression line of y on x. In the justification of the choice of this line it is assumed that deviations of the observations from the line are caused by unexplained random variation that is associated with the variable y. Implicitly it is assumed that the variable x is measured without error or other variation. Clearly, if it is felt that deviations from the line are due to variation in x alone the appropriate method would be to use the regression line of x on y, minimising the sum of squares in the horizontal direction.

The random deviations of the observations from the supposed underlying linear relationship are usually called the errors. Although the word error is a very common term it is an unfortunate choice of word; the variation may incorporate not just measurement error but any other sources of unexplained variation that results in scatter from the line. Some authors have suggested that other terms might be used, disturbance, departure, perturbation, noise and random component being amongst the suggestions. In this report, however, because of the wide use of the word, the variation from the line will be described as error.

In many investigations the scatter of the observations arises because of error in both measurements. This problem is known by many names, the commonest being errors in variables regression and measurement error models. The former name is used throughout this report. Casella and Berger (1990) wrote of this problem, '(it) is so different from simple linear regression ... that it is best thought of as a completely different topic'. There is a very extensive literature on the subject, but published work is mainly in the form of articles in the technical journals, most of which deal with a particular aspect of the problem. Relatively few standard text books on regression theory contain comprehensive descriptions of solutions to the problem. A brief literature survey is given in the next section.

We believe that the errors in variables regression problem is potentially of wide practical application in the analysis of experimental data. One of the aims of this report therefore is to give some guidance for practitioners in deciding how an errors in variables straight line should be fitted. We give simple formulas that a practitioner can use to estimate the slope and intercept of an optimum line together with variance terms that are also included in the model. Very few previous authors have given formulas for the standard errors of these estimators, and we offer some advice regarding these. Indeed, a detailed exposition on the variance covariance matrices for most of the estimators in this report is included in Gillard and Iles (2006).

In our approach we make as few assumptions as are necessary to obtain estimators that are reliable. We have found that straightforward estimators of the parameters and their asymptotic variances can be found using the method of moments principle. This approach has the advantage of being simple to follow for readers who are not principally interested in the methodology itself.

The method of moments technique is described in many books of mathematical statistics, for example Casella and Berger (1990), 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 (1988) 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'. Method of moments estimators can be criticised because they are not uniquely defined, so that if the method is used it is necessary to choose amongst possible estimators to find ones that best suit the data being analysed. This proves to be the case when the method is used 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 distributions describing the observations exist. We also assume here that these distributions are mutually uncorrelated. It is relatively easy to work out the theoretical asymptotic variances and covariances of the estimators using the delta method outlined by Cramer (1946). The information in this report will enable a practitioner to fit the line and calculate approximate confidence intervals for the associated parameters. Significance tests can also be done. A limitation of the formulas is that they are asymptotic results, so they should only be used for moderate or large data sets.

2. Literature Survey

As mentioned above, the errors in variables regression problem is rarely included in statistical texts. There are two texts devoted entirely to the errors in variables regression problem, Fuller (1987) and Cheng and van Ness (1999). Casella and Berger (1990) has an informative section on the topic, Sprent (1969) contains chapters on the problem, as do Kendall and Stuart (1979) and Dunn (2004). Draper and Smith (1998) 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 (1980). In these texts the method of instrumental variables is often given prominence. Instrumental variables are uncorrelated with the error distributions, but are highly correlated with the predictor variable. The extra information that these variables contain enables a method of estimating the parameters of the line to be obtained. Carroll et al (1995) described errors in variables models for non-linear regression, and Seber and Wild (1989) included a chapter on this topic.

Probably the earliest work describing a method that is appropriate for the errors in variables problem was published by Adcock (1878). He suggested that a line be fitted by minimising the sum of squares of distances between the points and the line in a direction perpendicular to the line, the method that has come to be known as orthogonal regression. Kummel (1879) took the idea further, generalising to a line that has minimum sum of squares of distances of the observations from the line in a direction other than perpendicular. Pearson (1901) generalised the errors in variables model to that of multiple regression, where there are two or more different x variables. He also pointed out that the slope of the orthogonal regression line is between those of the regression line of y on x and that of x on y. The idea of orthogonal regression was included in Deming's book (1943), and orthogonal regression is sometimes referred to as Deming regression.

Another method of estimation that has been used in errors in variables regression is the method of moments. Geary (1942, 1943, 1948 and 1949) wrote a series of papers on the method, but using cumulants rather than moments in the later papers. Drion (1951), in a paper that is infrequently cited, used the method of moments, and gave some results concerning the variances of the sample moments used in the estimators that he suggested. More recent work using the moments approach has been written by Pal (1980), van Montfort et al (1987), van Montfort (1989) and Cragg (1997). Much of this work centres on a search for optimal estimators using estimators based on higher moments. Dunn (2004) gave formulas for many of the estimators of the slope that we describe later in this report using a method of moments approach. However, he did not give information about estimators based on higher moments and it turns out that these are the only moment based estimators that can be used unless there is some information about the relationship additional to the (x, y) observations. Neither did he give information about the variances of the estimators.

Another idea, first described by Wald (1940) and taken further by Bartlett (1949), is to group the data, ordered by the true value of the predictor variable, and use the means of the groups to obtain estimators of the slope. The intercept is then estimated by choosing the line that passes through the centroid ($\bar{x}, \bar{y}$) of the complete data set. A difficulty of the method, noted by Wald himself, is that the grouping of the data cannot, as may at first be thought, be based on the observed values without making further assumptions. In order to preserve the properties of the random variables underlying the method it is necessary that the grouping be based on some external knowledge of the ordering of the data. In depending on this extra information, Wald's grouping method is a special case of an instrumental variables method, the instrumental variable in this case being the ordering of the true values. Gupta and Amanullah (1970) gave the first four moments of the Wald estimator and Gibson and Jowett (1957) investigated optimum ways of grouping the observations. Madansky (1959) reviewed some aspects of grouping methods.

Lindley (1947) and many subsequent authors approached the problem of errors in variables regression from a likelihood perspective. Kendall and Stuart (1979), Chapter 29, reviewed the literature and outlined the likelihood approach. A disadvantage of the likelihood method in the errors in variables problem is that it is only tractable if all of 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 (1999). The likelihood approach was also used by Dolby and Lipton (1972), Dolby (1976) and Cox (1976) to investigate the errors in variables regression problem where there are replicate measured values at the same true value of the predictor variable.

Lindley and el Sayyad (1968) described a Bayesian approach to the errors in variables regression problem and concluded that in some respects the likelihood approach may be misleading. A description of a Bayesian approach to the problem, with a critical comparison with the likelihood method, is given by Zellner (1980).

Golub and van Loan (1980), van Huffel and Vanderwalle (1991) and van Huffel and Lemmerling (2002) have developed a theory that they have called total least squares. This method allows the fitting of linear models where there are errors in the predictor variables as well as the dependent variable. These models include the linear regression one. The idea is linked with that of adjusted least squares, that has been developed by Kukush et al (2003) and Markovsky et al (2002, 2003).

Errors in variables regression has some similarities with factor analysis, a method in multivariate analysis described by Lawley and Maxwell (1971) and Johnson and Wichern (1992) and elsewhere. Factor analysis is one of a family, called latent variables methods (Skrondal and Rabe-Hesketh, 2004), that include the errors in variables regression problem. Dunn and Roberts (1999) used a latent variables approach in an errors in variables regression setting, and more recently extensions combining latent variables and generalised linear models methods have been devised (Rabe-Hesketh et al, 2000, 2001).

Over the years several authors have written review articles on errors in variables regression. These include Kendall (1951), Durbin (1954), Madansky (1959), Moran (1971) and Anderson (1984). Riggs et al (1978) performed simulation exercises comparing some of the slope estimators that have been described in the literature.

3. Statistical Assumptions

The notation in the literature for the errors in variables regression problem differs from author to author. In this report we use a notation that is similar to that used by Cheng and van Ness (1999), and that appears to be finding favour with other modern authors. It is, unfortunately, different from that used by Kendall and Stuart (1979), and subsequently adopted by Hood (1998) and Hood et al (1999).

We suppose that there are n individuals in the sample with true values (ξ_i, η_i) and observed values (x_i, y_i). It is believed that there is a linear relationship between the two variables ξ and η.

$$\eta_i = \alpha + \beta\xi_i \tag{1}$$

However, there is variation in both variables that results in a deviation of the observations (x_i, y_i) from the true values (ξ_i, η_i) resulting in a scatter about the straight line. This scatter is represented by the addition of random errors representing the variation of the observed from the true values.

$$x_i = \xi_i + \delta_i \tag{2}$$

$$y_i = \eta_i + \varepsilon_i = \alpha + \beta\xi_i + \varepsilon_i \tag{3}$$

The errors δ_i and ε_i are assumed to have zero means and variances that do not change with the suffix i.

$$E[\delta_i] = 0, \ \mathrm{Var}[\delta_i] = \sigma_\delta^2$$

$$E[\varepsilon_i] = 0, \ \mathrm{Var}[\varepsilon_i] = \sigma_\varepsilon^2.$$

We assume that higher moments also exist.

$$E[\delta_i^3] = \mu_{\delta 3}, \ E[\delta_i^4] = \mu_{\delta 4}$$

$$E[\varepsilon_i^3] = \mu_{\varepsilon 3}, \ E[\varepsilon_i^4] = \mu_{\varepsilon 4}.$$

We also assume that the errors are mutually uncorrelated and that the errors δ_i are uncorrelated with ε_i.

$$E[\delta_i\delta_j] = 0, \ E[\varepsilon_i\varepsilon_j] = 0 \ \ (i \neq j)$$

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

Some authors have stressed the importance of a concept known as equation error. Further details are given by Fuller (1987) and Carroll and Ruppert (1996). Equation error introduces an extra term on the right hand side of equation (3).

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

Dunn (2004) described the additional equation error term ω_i as '(a) new random component (that) is not necessarily a measurement error but is a part of y that is not related to the construct or characteristic being measured'. It is not intended to model a mistake in the choice of equation to describe the underlying relationship between ξ and η. Assuming that the equation error terms have a variance σ_ω^2 that does not change with 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.

We do not consider in this report methods for use where there may be expected to be serial correlations amongst the observations. Sprent (1969) included a section on this topic and Karni and Weissman (1974) used a method of moments approach, making use of the first differences of the observations, assuming that a non zero autocorrelation is present in the series of observations.

In much of the literature on errors in variables regression a distinction is drawn between the case where the ξ_i are assumed to be fixed, albeit unknown, quantities and the case where ξ_i are assumed to be a random sample from a population. The former is known as the functional and the latter the structural model. Casella and Berger (1990) described the theoretical differences in these two types of model. Using the approach adopted in this report it is not necessary to make the distinction. All that is assumed is that the ξ's are mutually uncorrelated, are uncorrelated with the errors and that the low order moments exist. Neither the problem of estimation of each individual ξ_i in the functional model nor the problem of predicting y is investigated in this report. Whether the ξ's are assumed to be fixed or a random sample we find only estimators for the low order moments.

The assumptions that we make about the variable ξ are as follows.

$$E[\xi_i] = \mu,\ Var[\xi_i] = \sigma^2 .$$

In some of the work that is described later the existence of higher moments of ξ is also assumed.

$$E[(\xi_i - \mu)^3] = \mu_{\xi 3},\ \ E[(\xi_i - \mu)^4] = \mu_{\xi 4}$$

The variables ξ_i are assumed to be mutually uncorrelated and uncorrelated with the error terms δ and ε.

$$E[(\xi_i - \mu)(\xi_j - \mu)] = 0 \ (i \neq j)$$

$$E[(\xi_i - \mu)\delta_j] = 0 \text{ and } E[(\xi_i - \mu)\varepsilon_j] = 0 \text{ for all i and j.}$$

In order to estimate variances and covariances it is necessary later in this report to assume the existence of moments of ξ of order higher than the fourth. The rth moment is denoted by $\mu_{\xi r} = E[(\xi_i - \mu)^r]$.

4. First and Second Order Moment Equations

The first order sample moments are denoted by $\bar{x} = \frac{\sum x_i}{n}$ and $\bar{y} = \frac{\sum y_i}{n}$.

The second order moments are notated by $s_{xx} = \frac{\sum (x_i - \bar{x})^2}{n}$, $s_{yy} = \frac{\sum (y_i - \bar{y})^2}{n}$ and $s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$.

No small sample correction for bias is made, for example by using (n - 1) as a divisor for the variances rather than n. This is because the results on variances and covariances that we give later on in the report are reliable only for moderately large sample sizes, generally 50 or more, where the adjustment for bias is negligible. Moreover, the algebra needed for the small sample adjustment complicates the formulas somewhat.

The moment equations in the errors in variables setting are given in the equations below. A tilde is placed over the symbol for a parameter to denote the method of moments estimator. We have used this symbol in preference to the circumflex, often used for estimators, to distinguish between method of moments and maximum likelihood estimators.

First order moments:

$$\bar{x} = \tilde{\mu} \tag{4}$$

$$\bar{y} = \tilde{\alpha} + \tilde{\beta}\tilde{\mu} \tag{5}$$

Second order moments:

$$s_{xx} = \tilde{\sigma}^2 + \tilde{\sigma}_\delta^2 \tag{6}$$

$$s_{yy} = \tilde{\beta}^2\tilde{\sigma}^2 + \tilde{\sigma}_\varepsilon^2 \tag{7}$$

$$s_{xy} = \tilde{\beta}\tilde{\sigma}^2 \tag{8}$$

It can readily be seen from equations (6), (7) and (8) that there is a hyperbolic relationship between method of moments estimators $\tilde{\sigma}_\delta^2$ and $\tilde{\sigma}_\varepsilon^2$ of the error variances. This was called the Frisch hyperbola by van Montfort (1989).

$$(s_{xx} - \tilde{\sigma}_\delta^2)(s_{yy} - \tilde{\sigma}_\varepsilon^2) = (s_{xy})^2 \qquad (9)$$

This is a useful equation in that it relates pairs of estimators of σ_δ^2 and σ_ε^2 that satisfy equations (6), (7) and (8). The potential applications of the Frisch hyperbola are dicussed further in Section 9.

One of the difficulties with the errors in variables regression problem is apparent from an examination of equations (4) - (8). There is an identifiability problem if these equations alone are used to find estimators. There are five moment equations of first or second order but there are six unknown parameters. It is therefore not possible to solve the equations to find unique solutions without making additional assumptions. One possibility is to use higher moments, and this is described later in the report. Another possibility is to use additional information in the form of an instrumental variable. A third possibility, and the one that is investigated first, is to assume that there is some prior knowledge of the parameters that enables a restriction to be imposed. This then allows the five equations to be solved.

There is a comparison with this identifiability problem and the maximum likelihood approach. In this approach, the only tractable assumption is that the distributions of δ_i, ε_i and ξ_i, are all Normal. This in turn leads to the bivariate random variable (x, y) 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 five first and second moment equations. In this case therefore it is not possible to find solutions to the likelihood equations without making an additional assumption, resticting the parameter space. The restrictions that we describe in Section 5 below are ones that have been used by previous authors using the likelihood method. The likelihood function for any other distribution than the Normal is complicated and the method is difficult to apply. However the method of moments approach using higher moments and without assuming a restiction in the parameter space, can be used without making the assumption of Normality.

5. Estimators Based on the First and Second Moments

So that estimating equations stand out from other numbered equations, they are marked by an asterisk. Equation (1) gives the estimator for μ directly

$$\tilde{\mu} = \bar{x} \qquad (10)^*$$

The estimators for all the remaining parameters are easily expressed in terms of the estimator $\tilde{\beta}$ of the slope. Equations (4) and (5) can be used to give an equation for the intercept $\tilde{\alpha}$ in terms of $\tilde{\beta}$.

$$\tilde{\alpha} = \bar{y} - \tilde{\beta}\bar{x} \qquad (11)^*$$

Thus the fitted line in the (x, y) plane passes through the centroid $(\bar{x}, \bar{y})$ of the data, a feature that is shared by the simple linear regression equations.

Equation (8) yields an equation for σ^2, with $\tilde{\beta}$ always having the same sign as s_{xy}.

$$\tilde{\sigma}^2 = \frac{s_{xy}}{\tilde{\beta}} \qquad (12)^*$$

If the error variance σ_δ^2 is unknown, it is estimated from equation (6).

$$\tilde{\sigma}_\delta^2 = s_{xx} - \tilde{\sigma}^2 \qquad (13)^*$$

Finally if $\tilde{\sigma}_\varepsilon^2$ is unknown, it is estimated from equation (7) and the estimator for β.

$$\tilde{\sigma}_\varepsilon^2 = s_{yy} - \tilde{\beta}^2\tilde{\sigma}^2 \qquad (14)^*$$

Since variances are never negative there are restriction on permissible parameter values, depending on the values taken by the sample second moments. These conditions are often called admissibility conditions. The straightforward conditions, enabling non negative variance estimates to be obtained are given below.

$$s_{xx} > \sigma_\delta^2$$

$$s_{yy} > \sigma_\varepsilon^2$$

Alone, these conditions are not sufficient to ensure that the variance estimators are non negative. The errors in variables slope estimator must lie between the y on x and x on y slope estimators $\frac{s_{xy}}{s_{xx}}$ and $\frac{s_{yy}}{s_{xy}}$ respectively.

Other admissibility conditions, relevant in special cases, are given in Table 1. Admissibility conditions are discussed in detail by Kendall and Stuart (1979), Hood (1998), Hood et al (1999) and Dunn (2004).

We now turn to the question of the estimation of the slope. There is no single estimator for the slope that can be used in all cases in errors in variables regression. Each of the restrictions assumed on the parameter space to to get around the identifiability problem discussed above is associated with its own estimator of the slope. In order to use an estimator based on the first and second order moments alone it is necessary for the practitioner to decide on the basis of knowledge of the investigation being undertaken which restriction is likely to suit the purpose best.

Table 1 summarises the simplest estimators of the slope parameter β derived by assuming a restriction on the parameters. With one exception these estimators have been described previously; most were given by Kendall and Stuart (1979), Hood et al (1999) and, in a method of moments context, by Dunn (1989).

Table 1:

Estimators of the slope parameter b based on first and second moments

Restriction	**Estimator**	**Admissibility Conditions**
Intercept α known	$\tilde{\beta}_1 = \dfrac{\bar{y} - \alpha}{\bar{x}}$	$\bar{x} \neq 0$
Variance σ_δ^2 known	$\tilde{\beta}_2 = \dfrac{s_{xy}}{s_{xx} - \sigma_\delta^2}$	$s_{xx} > \sigma_\delta^2$ $s_{yy} > \dfrac{(s_{xy})^2}{s_{xx} - \sigma_\delta^2}$
Variance σ_ε^2 known	$\tilde{\beta}_3 = \dfrac{s_{yy} - \sigma_\varepsilon^2}{s_{xy}}$	$s_{yy} > \sigma_\varepsilon^2$ $s_{xx} > \dfrac{(s_{xy})^2}{s_{yy} - \sigma_\varepsilon^2}$
Reliability ratio $\kappa = \dfrac{\sigma^2}{\sigma^2 + \sigma_\delta^2}$ known	$\tilde{\beta}_4 = \dfrac{s_{xy}}{\kappa s_{xx}}$	None
Variance ratio $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ known	$\tilde{\beta}_5 = \dfrac{(s_{yy} - \lambda s_{xx}) + \left\{(s_{yy} - \lambda s_{xx})^2 + 4\lambda (s_{xy})^2\right\}^{1/2}}{2 s_{xy}}$	None
$\nu = \dfrac{\lambda}{\beta^2}$ known	$\tilde{\beta}_6 = \dfrac{(\nu - 1) s_{xy} + \mathrm{sign}(s_{xy}) \left\{(\nu - 1)^2 (s_{xy})^2 + 4\nu s_{xx} s_{yy}\right\}^{1/2}}{2\nu s_{xx}}$	$s_{xx} \neq 0$
Both variances σ_δ^2 and σ_ε^2 known.	$\tilde{\beta}_7 = \mathrm{sign}(s_{xy}) \left\{\dfrac{s_{yy} - \sigma_\varepsilon^2}{s_{xx} - \sigma_\delta^2}\right\}^{1/2}$	$s_{xx} > \sigma_\delta^2$ $s_{yy} > \sigma_\varepsilon^2$

There is an ambiguity in the sign to be used in the equations for $\tilde{\beta}_6$ and $\tilde{\beta}_7$. This is resolved by assuming that the slope estimator always has the same sign as s_{xy}, as mentioned above to ensure that equation (11)* gives a non negative estimate of the variance σ^2. A discussion of these estimators is given in Section 9.

It may seem that the restriction leading to the estimator $\tilde{\beta}_6$ is not one that would often be made on the basis of a priori evidence. The reason for the inclusion of this estimator, which seems not to have been previously suggested, is that it is a generalisation of an estimator that has been widely recommended, the geometric mean estimator. This is the geometric mean of the slopes of the regression of y on x and the reciprocal of the regression of x on y. Section 9 contains further discussion. Asymptotic variances concerning this estimator will not be included in this report.

The assumption that both error variances σ_δ^2 and σ_ε^2 are known is somewhat different from the other cases. By assuming that two parameters are known there are only four remaining unknown parameters, but five first and second moment equations that could be used to estimate them. One possibility of obtaining a solution is to use only four of the five equations (4) to (8) inclusive, or a simple combination of these. If equation (6) is excluded, the estimator for the slope β is $\tilde{\beta}_3$, but then the assumed value of σ_δ^2 will almost certainly not agree exactly with the value that would be obtained from equation (12)*. If equation (7) is excluded, the estimator for the slope is $\tilde{\beta}_2$, but then it is most unlikely that the assumed value of σ_ε^2 will agree exactly with the value obtained from equation (13)*. If equations (6) and (7) are combined, using the known ratio $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_\delta^2}$, the estimator $\tilde{\beta}_5$ is obtained, and then neither of equations (12)* and (13)* will be satisfied by the a priori values assumed for σ_δ^2 and σ_ε^2. Another possibility that leads to a simple estimator for the slope β is to exclude equation (8), and it is this that leads to the estimator $\tilde{\beta}_7$ in Table 1.

6. Estimates Making Use of the Third Moments

The third order moments are written as follows.

$$s_{xxx} = \frac{\sum (x_i - \bar{x})^3}{n}$$

$$s_{xxy} = \frac{\sum (x_i - \bar{x})^2 (y_i - \bar{y})}{n}$$

$$s_{xyy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})^2}{n}$$

$$s_{yyy} = \frac{\sum (y_i - \bar{y})^3}{n} .$$

The four third moment equations take a simple form. Some details on the derivation of these expressions is given in Appendix 1.

$$s_{xxx} = \tilde{\mu}_{\xi 3} + \tilde{\mu}_{\delta 3} \tag{15}$$

$$s_{xxy} = \tilde{\beta}\tilde{\mu}_{\xi 3} \tag{16}$$

$$s_{xyy} = \tilde{\beta}^2 \tilde{\mu}_{\xi 3} \tag{17}$$

$$s_{yyy} = \tilde{\beta}^3 \tilde{\mu}_{\xi 3} + \tilde{\mu}_{\varepsilon 3} \tag{18}$$

Together with the first and second moment equations, equations (4) - (8) inclusive, there are now nine equations in nine unknown parameters. The additional parameters introduced here are the third moments $\mu_{\xi 3}$, $\mu_{\delta 3}$ and $\mu_{\varepsilon 3}$. There are therefore unique estimators for all nine parameters. However, it is unlikely in practice that there is as much interest in these third moments as there is in the first and second moments, more especially, the slope and intercept of the line. Thus a simpler way of proceeding is probably of more general value.

The simplest way of making use of these equations is to make a single further assumption, namely that $\mu_{\xi 3}$ is non zero. There is a practical requirement associated with this assumption, and this is that the sample third moments should be significantly different from 0. It is this requirement that has probably led to the use of third moment estimators receiving relatively little attention in recent literature. It is not always the case that the observed values of x and y are sufficiently skewed to allow these equations to be used with any degree of confidence. Moreover sample sizes needed to identify third order moments with a practically useful degree of precision are somewhat larger than is the case for first and second order moments. However, if the assumption can be justified from the data then a straightforward estimator for the slope parameter is obtained without assuming anything known a priori about the values taken by any of the parameters. This estimator is obtained by dividing equation (17) by equation (16).

$$\tilde{\beta}_8 = \frac{s_{xyy}}{s_{xxy}} \tag{19}*$$

The value for β obtained from this equation can then be substituted in equations (11)* - (14)* to estimate the intercept α and all three variances σ^2, σ_δ^2 and σ_ε^2. The third moment $\mu_{\xi 3}$ can be estimated from equation (16).

$$\tilde{\mu}_{\xi 3} = \frac{s_{xxy}}{\tilde{\beta}_8} \tag{20}*$$

Estimators for $\mu_{\delta 3}$ and $\mu_{\varepsilon 3}$ may be obtained from equations (15) and (18) respectfully.

Other simple ways of estimating the slope are obtained if the additional assumptions $\mu_{\delta 3} = 0$ and $\mu_{\varepsilon 3} = 0$ are made. These would be appropriate assumptions to make if the distributions of the error terms δ and ε are symmetric. Note, however, that this does not imply that the distribution of ξ is symmetric. The observations have to be skewed to allow the use of estimators based on the third moments. With these assumptions the slope β could be estimated by dividing equation (16) by (15) or by dividing equations (18) and (17).

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

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

We do not investigate these estimators further in this report, since we feel that estimators that make fewest assumptions are likely to be of the most practical value.

7. Estimates Making Use of the Fourth Moments

The fourth order moments are written as

$$s_{xxxx} = \frac{\sum (x_i - \bar{x})^4}{n}$$

$$s_{xxxy} = \frac{\sum (x_i - \bar{x})^3 (y_i - \bar{y})}{n}$$

$$s_{xxyy} = \frac{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}{n}$$

$$s_{xyyy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})^3}{n}$$

$$s_{yyyy} = \frac{\sum (y_i - \bar{y})^4}{n}$$

By using a similar approach to the one adopted in deriving the third moment estimating equations, the fourth moment equations can be derived.

$$s_{xxxx} = \tilde{\mu}_{\xi 4} + 6\tilde{\sigma}^2 \tilde{\sigma}_\delta^2 + \tilde{\mu}_{\delta 4} \qquad (21)$$

$$s_{xxxy} = \tilde{\beta}\tilde{\mu}_{\xi 4} + 3\tilde{\beta}\tilde{\sigma}^2 \tilde{\sigma}_\delta^2 \qquad (22)$$

$$s_{xxyy} = \tilde{\beta}^2 \tilde{\mu}_{\xi 4} + \tilde{\beta}^2 \tilde{\sigma}^2 \tilde{\sigma}_\delta^2 + \tilde{\sigma}^2 \tilde{\sigma}_\varepsilon^2 + \tilde{\sigma}_\delta^2 \tilde{\sigma}_\varepsilon^2 \qquad (23)$$

$$s_{xyyy} = \tilde{\beta}^3 \tilde{\mu}_{\xi 4} + 3\tilde{\beta}\tilde{\sigma}^2 \tilde{\sigma}_\varepsilon^2 \qquad (24)$$

$$s_{yyyy} = \tilde{\beta}^4 \tilde{\mu}_{\xi 4} + 6\tilde{\beta}^2 \tilde{\sigma}^2 \tilde{\sigma}_\varepsilon^2 + \tilde{\mu}_{\varepsilon 4} \qquad (25)$$

Together with the first and second moment equations these form a set of ten equations, but there are only nine unknown parameters. The fourth moment equations have introduced three additional parameters $\mu_{\xi 4}$ $\mu_{\delta 4}$ and $\mu_{\varepsilon 4}$, but four new equations. One of the equations is therefore not needed. The easiest practical way of estimating the parameters is to use equations (22) and (24), together with equations (6), (7) and (8).

Equation (22) is multiplied by $\tilde{\beta}^2$ and subtracted from equation (24).

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

Equation (6) is multiplied by $\tilde{\beta}^2$ and subtracted from equation (7).

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

Thus, making use also of equation (8) an estimating equation is obtained for the slope parameter β.

$$\tilde{\beta}_9 = \left\{\frac{s_{xyyy} - 3s_{xy}s_{yy}}{s_{xxxy} - 3s_{xy}s_{xx}}\right\}^{1/2} \qquad (26)^*$$

There may be a practical difficulty associated with the use of equation (26)* if the random variable ξ is Normally distributed. In this case the fourth moment is equal to 3 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 kurtosis is given by the following expression

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \qquad (27)$$

If the distribution of ξ has zero measure of kurtosis the average values of the five sample moments used in equation (26)* are as follows.

$$E[s_{xyyy}] = 3\beta^3\sigma^4 + 3\beta\sigma^2\sigma_\varepsilon^2$$

$$E[s_{xxxy}] = 3\beta\sigma^4 + 3\beta\sigma^2\sigma_\delta^2$$

$$E[s_{xx}] = \sigma^2 + \sigma_\delta^2$$

$$E[s_{yy}] = \beta^2\sigma^2 + \sigma_\varepsilon^2$$

$$E[s_{xy}] = \beta\sigma^2$$

Then it can be seen that the average value of the numerator of equation (26)* is approximately equal to zero, as is the average value of the denominator. Thus there is an additional assumption that has to be made for this equation to be reliable as an estimator, and that is that equation (27) does not hold, $\mu_{\xi 4}$ must be different from $3\sigma^4$. In practical terms, both the numerator and the denominator of the right hand side of equation (26)* must be significantly different from zero.

If a reliable estimate of the slope β can be obtained from equation (26)*, equations (10)* - (13)* enable the intercept α and the variances σ^2, σ_δ^2 and σ_ε^2 to be estimated. The fourth moment $\mu_{\xi 4}$ of ξ can then be estimated from equation (22), and the fourth moments $\mu_{\delta 4}$ and $\mu_{\varepsilon 4}$ of the error terms δ and ε can be estimated from equations (20) and (24) respectively, though estimates of these higher moments of the error terms are less likely to be of practical value.

Although $\tilde{\beta}_9$ has a compact closed form, its variance is rather cumbersome. Indeed, the variance of $\tilde{\beta}_9$ depends on the sixth central moments of ξ. Since it is impractical to estimate this moment with any degree of accuracy, there will be no discussion of the asymptotic variance of this estimator.

8. Variances and Covariances of the Estimators

In order to derive formulas for the asymptotic variances and covariances of the estimators derived in previous sections, the variances and covariances of the sample moments are needed. Further details on variances and covariances of the estimators are included in the technical paper by Gillard and Iles (2006). However, a brief exposition is given here, and in Appendix 2.

Since most of the estimators described in this report are non linear functions of the sample moments, the problem of finding exact formulas for the variances and covariances is not a straightforward one. However, an approximate method, called the delta method, or method of statistical differentials, gives simple formulas for the estimators quoted here. These have proved in simulation studies to be highly reliable even for moderate (n = 50) sample sizes. The method is sometimes described in statistics texts, for example DeGroot (1989) and is often used in linear models to derive a variance stabilisation transformation (see Draper and Smith, 1998). For further details see Kotz and Johnson (1988) and Bishop et al (1975). The 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.

For each of the restricted cases discussed earlier, (apart from the restriction $\upsilon = \frac{\lambda}{\beta^2}$) the variance covariance matrices can be partitioned into a sum of three matrices, A, B and C. This is reported in Gillard and Iles (2006).

The matrix A alone is needed if the assumptions are made that ξ, δ and ε all have zero third moments and zero measures of kurtosis, as given by equation (27). These assumptions would be valid if all three of these variables are Normally distributed.

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

The C matrices are additional terms that are needed if the third moments and measures of kurtosis are non zero for the error terms δ and ε. We believe it to be likely that these C matrices will prove of less value to practitioners than the A and B matrices.

For estimators based on higher order moments, the algebra is more cumbersome, and the expressions are not as concise as for the restricted cases. In Gillard and Iles (2006), the tools needs to construct the variance covariance matrices for $\tilde{\beta}_8$ are included.

The expressions that are of most practical use in the application of regression methods are the variances and covariances of the estimators of the intercept and slope parameters α and β. These enable approximate tests and confidence intervals to be calculated for these parameters, and also approximate confidence bands to be found for the line. We give here the formulas derived from the A and B matrices, that is assuming that the error variables have Normal like third and fourth moments. The variances for the slope estimators in Table 1 based on the first and second moments are first given in Table 2. The formulas for the variance of $\tilde{\beta}_8$ is not as simple, and is given later.

Estimates of a combination of the parameters is needed in some of the variance and covariance formulas given below. This estimator is derived here. With the assumptions that have been made up to this point, 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..

$$\Sigma = \begin{pmatrix} \sigma_\xi^2 + \sigma_\delta^2 & \beta\sigma_\xi^2 \\ \beta\sigma_\xi^2 & \beta^2\sigma_\xi^2 + \sigma_\varepsilon^2 \end{pmatrix} \tag{28}$$

This variance covariance matrix is estimated by the matrix S.

$$S = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{pmatrix} \tag{29}$$

The determinant of the matrix Σ is $|\Sigma| = \beta^2\sigma^2\sigma_\delta^2 + \sigma^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2$. This is therefore estimated by the determinant of S.

$$|\tilde{\Sigma}| = |S| = s_{xx}s_{yy} - (s_{xy})^2 \qquad (30)^*$$

Estimators of the variances $\tilde{\sigma}^2$, $\tilde{\sigma}_\delta^2$ and $\tilde{\sigma}_\varepsilon^2$ are given in equations (11)*, (12)* and (13)* respectively, and the fourth moment $\mu_{\xi 4}$ is estimated from equation (22) .

Table 2:

Variances and Estimators of the Variances for the Slope Parameter Estimators given in Table 1.

Estimator	Variance of slope b	Estimator of variance of slope b
$\tilde{\beta}_1$	$\dfrac{\beta^2\sigma_\delta^2 + \sigma_\varepsilon^2}{n\mu^2}$	$\dfrac{s_{yy}}{n\bar{x}^2}$
$\tilde{\beta}_2$	$\dfrac{1}{n\sigma^4}\left[\lvert\Sigma\rvert + 2\beta^2\sigma_\delta^4\right]$	$\dfrac{1}{n(\tilde{\sigma}^2)^2}\left[\lvert S\rvert + 2(\tilde{\beta}_2\tilde{\sigma}_\delta^2)^2\right]$
$\tilde{\beta}_3$	$\dfrac{1}{n\sigma^4}\left[\lvert\Sigma\rvert + \dfrac{2\sigma_\varepsilon^4}{\beta^2}\right]$	$\dfrac{1}{n(\tilde{\sigma}^2)^2}\left[\lvert S\rvert + 2\left(\dfrac{\tilde{\sigma}_\varepsilon^2}{\tilde{\beta}_3}\right)^2\right]$
$\tilde{\beta}_4$	$\dfrac{1}{n\sigma^4}\left[\lvert\Sigma\rvert + (1-\kappa)^2\beta^2(\mu_{\xi 4} - 3\sigma^4)\right]$	$\dfrac{1}{n(\tilde{\sigma}^2)^2}\left[\lvert S\rvert + (1-\kappa)^2\tilde{\beta}_4^2(\tilde{\mu}_{\xi 4} - 3(\tilde{\sigma}^2)^2)\right]$
$\tilde{\beta}_5$	$\dfrac{\lvert\Sigma\rvert}{n\sigma^4}$	$\dfrac{\lvert S\rvert}{n(\tilde{\sigma}^2)^2}$
$\tilde{\beta}_7$	$\dfrac{1}{n\sigma^4}\left[\lvert\Sigma\rvert + \dfrac{(\beta^2\sigma_\delta^2 - \sigma_\varepsilon^2)^2}{2\beta^2}\right]$	$\dfrac{1}{n(\tilde{\sigma}^2)^2}\left[\lvert S\rvert + \dfrac{(s_{yy} - \tilde{\beta}_7^2 s_{xx})^2}{2\tilde{\beta}_7^2}\right]$

Notice that in the special case that $\mu_{\xi 4} = 3\sigma^2$ the formula for the variance of $\tilde{\beta}_4$ is identical to that for $\tilde{\beta}_5$. This would hold if ξ is Normally distributed.

The above variance estimators assume that both δ and ε are Normally distributed (or have Normal like third and fourth moments). Details on the corrections when δ and ε are not Normally distributed are offered in Gillard and Iles (2006).

We now give formulas for the variance of the slope estimator $\tilde{\beta}_8$ based on the third and fourth moments where it is assumed that the error terms δ and ε are assumed to be Normally distributed. The expression Var[$\tilde{\beta}_8$] has been calculated when δ and ε are assumed not to be Normally distributed by Gillard and Iles (2006).

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

Notice that the formula involves the third and fourth moments of ξ, but not higher moments.

To estimate this variance, all three parameters, σ^2, σ_δ^2 and σ_ε^2 have to be estimated using equations (11)*, (12)* and (13)* respectively. The third moment $\mu_{\xi 3}$ is estimated from equation (22)*. The fourth moment $\mu_{\xi 4}$ can be estimated from one of equations (22), (23) or (24). The combinations $(\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2)$, $(\sigma^2 + \sigma_\delta^2)$ and $(\beta^2\sigma^2 + \sigma_\varepsilon^2)$ are estimated by $(\tilde{\beta}^2 s_{xx} + s_{yy} - 2\tilde{\beta}s_{xy})$, s_{xx} and s_{yy} respectively.

9. Discussion

It was not our intention in writing this report to advocate the complete abandoment of simple linear regression methods. If there are no measurement errors associated with the x variable in linear regression, then $\sigma_\delta^2 = 0$ and the optimum way of fitting the line if the rest of the assumptions made in this report hold true is to use the least squares regression line of y on x. The intercept and slope estimators of this line are both unbiased. Conversely, if there are no measurement errors associated with the variable y then the regression of x on y gives the best line. However, in the presence of measurement error in both x and y, neither the x on y nor the y on x regression give unbiased estimators of the slope and intercept. In fact the true line lies between these two extremes. (see, for example, Casella and Berger, 1990).

In this report a number of possible solutions have been presented to the problem of identifying an appropriate relationship between the (unmeasured) variables ξ and η, based on the observations x and y. In the following section a procedure is suggested for selecting the most appropriate line for a particular purpose. Some observations on the inter-relationships between these different estimators is discussed here.

The estimators that are given in Table 1 are, with one exception, maximum likelihood estimators for the case where the variable ξ and the errors δ and ε are all assumed to be Normally distributed. The exception is $\tilde{\beta}_7$, the case where both error variances are known, where the maximum likelihood estimator for the slope is $\tilde{\beta}_5$. Most of these cases were derived by Hood (1998) and Kendall and Stuart (1979). The derivation of the estimators in this report is based on the method of moments and no assumption of an exact form for these distributions is necessary.

The case where the intercept α is assumed to be known leads to the estimator $\tilde{\beta}_1$, which is a ratio of means. In a sense, therefore, this estimator is related to the ratio of means estimators suggested by Wald (1940) and Bartlett (1949) but based on grouped data.

The estimator $\tilde{\beta}_2$, which should be used if prior knowledge gives the value for the error variance σ_δ^2 of the error in the x variable, is a simple modification of the slope of the regression of y on x. The modification is to subtract the known variance σ_δ^2 from the sum of squares of x, s_{xx}, in the denominator of the expression. The effects of equation error have led some authors, notably Dunn (2004), 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 on the right hand side of equation (7), because 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, is more likely to be reliable than those that assume prior knowledge of σ_ε^2

Where it is believed that prior knowledge gives a reliable value for the error variance σ_ε^2 of the error in the y variable the estimator $\tilde{\beta}_3$ should be used. This is a modification of the reciprocal of the slope of the regression of x on y. The modification is to subtract the known variance σ_ε^2 from the sum of squares of y, s_{yy}, in the numerator of the expression.

Knowledge of the reliability ratio implicitly is knowledge of the bias of the y on x regression slope where there are errors in both variables. The unbiased estimator $\tilde{\beta}_4$ in this case is obtained from the slope of the regression line of y on x simply by dividing by the reliability ratio.

The case where the ratio λ of error variances is known, $\tilde{\beta}_5$, is related to the orthogonal regression line. If $\lambda = 1$ these distances are in a direction that is orthogonal to the line. Casella and Berger (1990), amongst other authors, gave a proof of this result. If $\lambda \neq 1$ the line still has a geometrical interpretation, but the sum of squares of distances from the line that is minimised is in a direction different from perpendicular. Riggs et al (1978), based on their simulation studies, recommended the use of this estimator but emphasised the importance of having a reliable prior knowledge of the ratio λ of error variances. Use of this line has been criticised on the grounds that if the scale of measurement of the line is changed then a different line would be fitted (Bland 2000, p187). In fact, this criticism cannot be substantiated as long as it is kept in mind that it is knowledge of the ratio of the error variances λ that is used in fitting in line. If the data are rescaled in any way there is an exactly corresponding rescaling of λ that leads to the same line being fitted. The presence of equation error discussed above might, however, make it difficult to obtain a reliable a priori estimate of the ratio λ of error variances.

The estimator $\tilde{\beta}_6$ has been included in this report because it is linked with other estimators. The ratio $\nu = \frac{\lambda}{\beta^2}$ is a dimensionless quantity. If $\nu = 0$ the estimator reduces to the regression of x on y. If $\nu = \infty$ it is the regression of y on x.

If $\nu = 1$, the estimator reduces to a simple estimator.

$$\tilde{\beta} = \left\{\frac{s_{yy}}{s_{xx}}\right\}^{1/2}$$

This is the geometric mean of the slope of the regression of y on x, $\frac{s_{xy}}{s_{xx}}$ and the reciprocal of the slope of the regression of x on y, $\frac{s_{yy}}{s_{xy}}$. Probably because it is a compromise solution to the errors in variables regression problem and apparently makes no use of prior assumptions about the parameters, it has been recommended by several authors, for example Draper and Smith (1998, p 92). They gave a geometric interpretation of this estimator. The line with this slope has the minimum sum of products of the horizontal and vertical distances of the observations from the line. However, unless the restriction $\beta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ is true, the estimator is biased. If however this restriction is true the variance of estimator takes a very simple form.

$$\mathrm{Var}[\tilde{\beta}] = \frac{|\Sigma|}{n(\sigma_\xi^2 + \sigma_\delta^2)}$$

This is estimated by $$\frac{|S|}{n(s_{xx})^2} = \frac{s_{xx}s_{yy} - (s_{xy})^2}{n(s_{xx})^2}$$

A technical criticism of the use of this estimator is that it may have infinite variance (Creasy, 1956). As Creasy pointed out, however, this occurs when the scatter of the observations in the (x, y) plane is so large in both directions that it is visually impossible to determine if one line or a different line at right angles should be used to describe the relationship between y and y. Thus the criticism applies in cases where a linear relationship between x and y is not strongly indicated by the observations. The geometric mean estimator is also related to the orthogonal regression estimator $\tilde{\beta}_5$. If the ratio λ is taken to be $\frac{s_{yy}}{s_{xx}}$, the two estimators are identical.

The estimator $\tilde{\beta}_7$ can clearly be seen to be a modification of the geometric mean estimator described above in that both numerator and denominator are modified by subtraction of the known error variance. If $(s_{xx} - \sigma_\delta^2)$ is substituted from equation (9), the Frisch hyperbola, into the formula for $\tilde{\beta}_7$ the estimator $\tilde{\beta}_3$ is obtained. Similarly if $(s_{yy} - \sigma_\varepsilon^2)$ is substituted from equation (9), the estimator $\tilde{\beta}_2$ is obtained.

In the case where no knowledge is available that enables a restriction on the parameter space to be assumed, none of the estimators based on the first and second moments alone can be used, although as discussed above the geometric mean estimator has been suggested as an ad hoc compromise. However, if the third order moments s_{xyy} and s_{xxy} are both significantly different from 0, the estimator $\tilde{\beta}_8$ can be used, and is a reliable estimator if the sample size is sufficiently large. Another possible solution that could be used is $\tilde{\beta}_9$, but here the data must be such that both the numerator and denominator of the estimator are significantly different from zero.

Doubtless there will be cases where there is insufficient detailed prior knowledge of the parameters of the model to assign precise values to parameters, but nevertheless there may be some range of values for these parameters that are believed to be more likely to be true than others. In such cases there are two possible types of plot that may be of practical value in the identification of an appropriate range of values of the slope parameter β and hence, by using equations (9)* to (13)* ranges of values for the other parameters. The first of these plots is the Frisch hyperbola given in equation (9).

An example of a Frisch hyperbola plot of $\tilde{\sigma}_{\delta}^2$ against $\tilde{\sigma}_{\varepsilon}^2$ is given in Figure 1. The use of this plot can be illustrated by assuming that it is believed that the error variance σ_{δ}^2 for x is about equal to 1.0, but is possibly between 0.8 and 1.2. This gives an error variance for y, σ_{ε}^2 that is approximately between 1.2 and 1.6, with 1.4 as the most likely value. This range of indicated values for σ_{ε}^2 could then be appraised to determine whether this range of values is plausible. The possibility of equation error must always be borne in mind, however. The indicated values of σ_{ε}^2 indicated by this plot may at first sight be larger than expected from measurement error alone. It could be the equation error contribution that has inflated the estimates.

The second type of plot that may be useful in such circumstances is a sensitivity plot. Suppose that the preferred estimator for the slope is the case where the variance ratio λ is assumed to be known, but the precise value of λ is not known with certainty. The estimator of the slope in this case is $\tilde{\beta}_5$.

$$\tilde{\beta}_5 = \frac{s_{yy} - \lambda s_{xx} + \left\{(s_{yy} - \lambda s_{xx})^2 + 4\lambda (s_{xy})^2\right\}}{2s_{xy}} \tag{33}$$

A value of β can then be calculated for each plausible value of λ. A plot of such values is given in Figure 2. Suppose a priori evidence suggests that λ is equal to 1.4, but there is some doubt about the exact value, and it is possible that λ might be between 1.33 and 1.5. The corresponding values of β are between approximately 1.145 and 1.155, with 1.15 as the most likely value. Similar sensitivity plots are readily devised for other estimators given in Table 1.

In some cases the estimators given in Table 1 will give similar numerical values. One such case would be where the slope β is small. In that case the numerical values for $\tilde{\beta}_2$ and $\tilde{\beta}_5$ are highly likely to be numerically similar. This is because the correlation between these two estimators is approximately equal to 1 if β is small. Using the delta method described above an equation for this correlation coefficient can be worked out.

$$\mathrm{Corr}[\tilde{\beta}_2, \tilde{\beta}_5] \approx \left\{1 + \frac{2\beta^2\sigma_\delta^4}{|\Sigma|}\right\}$$

Thus if $\tilde{\beta}_7$ is small compared with $|\Sigma|$, or if σ_δ^2 is small compared with $|\Sigma|$ this correlation coefficient is approximately equal to 1, and in practice values of $\tilde{\beta}_2$ and $\tilde{\beta}_5$ will be numerically similar. In similar fashion it can be shown that $\tilde{\beta}_3$ and $\tilde{\beta}_5$ will be numerically similar if the slope β is large.

10. A Systematic Approach for Fitting Lines where there are Errors in Both Variables.

A systematic procedure for estimating the slope of a straight line relationship between y and x can now be presented, making use of the theory presented in this report.

- If the measurement error in x is small in comparison with that of y, use the simple linear regression of y on x.

- If the measurement error in y is small in comparison with that of x, use the reciprocal of regression of x on y.

In the following cases it is assumed that there are believed to be significant measurement errors in both x and y. Equations for the estimators are given in Table 1, except as noted.

Care has to be taken for some of these estimators that admissibility conditions are satisfied. In practice, if a variance estimate is obtained that is negative, the assumptions that have been made are incorrect, and a reappraisal of a priori information is needed. One may also check that the slope estimate lies between the y on x regression and x on y regression estimates respectively.

- If the intercept α is known a priori, use estimator $\tilde{\beta}_1$.

- If the error variance σ_δ^2 is known, use estimator $\tilde{\beta}_2$.

- If the error variance σ_ε^2 is known, use estimator $\tilde{\beta}_3$.

- If the reliability ratio $\kappa = \dfrac{\sigma^2}{\sigma^2 + \sigma_\delta^2}$ is known, use estimator $\tilde{\beta}_4$.

- If the ratio of error variances $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ is known, use estimator $\tilde{\beta}_5$.

- If no a priori knowledge is available, but the sample third moments are significantly different from zero, use estimator $\tilde{\beta}_8$. For this estimator to be reliable a sample size of at least 50 is needed.

- If no a priori knowledge is available, but the coefficients of kurtosis are significantly different from zero, use estimator $\tilde{\beta}_9$. For this estimator to be reliable a sample size of at least 100 is needed.

If a single most appropriate estimate for the slope is obtained, estimates of the variances, the intercept and the mean μ are obtained by using equations (10)* to (14)* inclusive.

- If imprecise prior information is available, and the conditions for the use of one of $\tilde{\beta}_8$ or $\tilde{\beta}_9$ are not satisfied, use the Frisch hyperbola and sensitivity plot to identify a range of possible values for β that accord with the prior knowledge.

- If no a priori knowledge is available, and it is decided to make the ad hoc assumption that $\beta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$, or equivalently that the ratio of error variances λ is equal to $\frac{s_{yy}}{s_{xx}}$, use estimator $\tilde{\beta}_6$. This estimator should be used with some care, however. It is unlikely that these conditions will be met in practice, and the estimator is then biased. Alternatively, one might use the orthogonal regression estimator, $\tilde{\beta}_6$ with $\lambda = 1$. This is equivalent to minimising the sums of squares of the orthogonal projections from the data point to the regression line.

11. References

Adcock, R.J. (1878). A problem in least squares. *Analyst*, **5**, 53-54.

Anderson, T.W. (1984). Estimating linear statistical relationships. *Ann. Statist.*, **12**, 1-45.

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

Bishop, Y.M.M., Fienberg, S., Holland, P. (1975). *Discrete Multivariate Analysis.* MIT Press, Cambridge.

Bland, M. (2000). *An Introduction to Medical Statistics (3rd edition).* Oxford University Press, Oxford.

Bowman, K.O. and Shenton, L.R. (1985). Method of moments. *Encyclopedia of Statistical Sciences (Volume 5)*, 467-453, John Wiley & Sons, Canada.

Carroll, R.J. and Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *Amer. Statist.*, **50**, 1-6.

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

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

Cheng, C-L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*. Arnold, London.

Cox, N.R. (1976). The linear structural relation for several groups of data. *Biometrika,* **63**, 231-237.

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

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ.

DeGroot, M.H. (1989). *Probability and Statistics.* Addison-Wesley Publishing Company, USA.

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

Draper, N. and Smith, H. (1998). *Applied Regression Analysis*. Wiley, New York.

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 Mathematicae*, **13**, 256-260.

Dolby, G.R. (1976). The ultrastructural relation: a synthesis of the functional and structural relations. *Biometrika*, **63**, 39-50.

Dolby, G.R. and Lipton, S. (1972). Maximum likelihood estimation of the general nonlinear functional relationship with replicated observations and correlated errors. *Biometrika*, **59(1)**, 121-129.

Dunn, G. (2004). *Statistical Evaluation of Measurement Errors (2nd Edition).* Arnold, London.

Dunn, G. and Roberts, C. (1999). Modelling method comparison data. *Statistical Methods in Medical Research*, **8**, 161-179.

Durbin, J. (1954). Errors in variables. *Int. Statist. Rev.*, **22**, 23-32.

Fuller, W.A. (1987). *Measurement Error Models.* Wiley, New York.

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

Gibson, W.M. and Jowett, G.H. (1957). Three-group regression analysis. Part 1: Simple regression analysis. *Appl. Statist.*, **6**, 114-122.

Gillard, J.W. and Iles, T.C. (2006). Variance covariance matrices for linear regression with errors in both variables. *Cardiff School of Mathematics Technical Report.*

Golub, G.H. and Van Loan, C.F. (1981). An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, **17**, 883-893.

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

Hood, K. (1998). Some statistical aspects of method comparison studies. *Cardiff University Ph.D Thesis.*

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.

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

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

Karni, E. and Weissman, I. (1974). A consistent estimator of the slope in a regression model with errors in the variables. *Journal of the American Statistical Association.* **65**, 211-213.

Kendall, M.G. (1951). Regression, structure, and functional relationship, I. *Biometrika*, **38**, 11-25.

Kendall, M.G. and Stuart, A. (1979). *The Advanced Theory of Statistics,* **Vol.2** (4th Edition). Griffin, London.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. and Nizam, A (1997). *Applied Regression Analysis and Other Multivariable Methods (3rd Edition).* Duxbury Press, CA.

Kotz, S. and Johnson, N.L. (1988). *Encyclopedia of Statistical Sciences.* Wiley, New York.

Kummel, C.H. (1879). Reduction of observed equations which contain more than one observed quantity. *Analyst.* **6**, 97-105.

Kukush, A., Markovsky, I., and Van Huffel, S. (2003). Consistent estimation in the bilinear multivariate errors-in-variables model. *Metrika,* **57**, 253-285.

Lawley, D. and Maxwell, A. (1971). *Factor Analysis as a Statistical Method (2nd Edition).* Elsevier Publishing, New York.

Lindley, D.V. (1947). Regression lines and the linear functional relationship. *J. R. Statist. Soc. Suppl.,* **9**, 218-244.

Lindley, D.V. and El Sayyad, G.M. (1968). The Bayesian estimation of a linear functional relationship. *J. R. Statist. Soc. B*, **30**, 190-202.

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

Markovsky I., Kukush A., Van Huffel S., (2002). Consistent least squares fitting of ellipsoids. *Numerische Mathematik.* **98(1)**, 177-194.

Markovsky I., Van Huffel S., Kukush A., (2003). On the computation of the multivariate structured total least squares estimator. *Numer. Linear Algebra Appl.* **11**, 591-608.

Moran, P.A.P. (1971). Estimating structural and functional relationships. *J. Multivariate Anal.,* **1**, 232-255.

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

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

Rabe-Hesketh, S., Pickles, A., and Taylor, C. (2000). Generalised linear latent and mixed models. *Stata Technical Bulletin*, **53**, 47-57.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2001). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter,* **13,** 17-23.

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.

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

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalised Latent Variable Modelling.* Chapman and Hall/CRC, Florida.

Sprent, P. (1969). *Models in Regression and Related Topics*, Matheun & Co Ltd, London.

Van Huffel, S. and Lemmerling, P. (Eds) (2002). *Total Least Squares and Errors-in-Variables Modelling: Analysis, Algorithms and Applications*, Kluwer, Dordrecht.

Van Huffel, S. (1997) and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia.

Van Montfort, K. (1988). *Estimating in Structural Models with Non-Normal Distributed Variables: Some Alternative Approaches.* M & T Series 12. DSWO Press, Leiden.

Van Montfort, K., Mooijaart, A., and de Leeuw, J. (1987). Regression with errors in variables. *Statist. Neerlandica*, **41**, 223-239.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.*, **11**, 284-300.

Zellner, A. (1980). *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.

12. Appendices

Appendix 1

The moment equations based on the third and fourth moments are slightly more difficult to derive than for the first and second order moment equations. One example illustrates the general approach.

$$
\begin{aligned}
E\left[ns_{xxy}\right] &= E\left[\sum (x_i - \bar{x})^2 (y_i - \bar{y})\right] \\
&= E\left[\sum \left\{(\xi_i - \bar{\xi}) + (\delta_i - \bar{\delta})\right\}^2 \left\{\beta(\xi_i - \bar{\xi}) + (\varepsilon_i - \bar{\varepsilon})\right\}\right] \\
&= E\Big[\sum \beta(\xi_i - \bar{\xi})^3 + \sum (\xi_i - \bar{\xi})^2 (\varepsilon_i - \bar{\varepsilon}) + 2\sum \beta(\xi_i - \bar{\xi})^2 (\delta_i - \bar{\delta}) \\
&\quad + 2\sum \beta(\xi_i - \bar{\xi})(\delta_i - \bar{\delta})(\varepsilon_i - \bar{\varepsilon}) + \sum \beta(\xi_i - \bar{\xi})(\delta_i - \bar{\delta})^2 + \sum (\delta_i - \bar{\delta})^2 (\varepsilon_i - \bar{\varepsilon})\Big]
\end{aligned}
$$

Terms of order n^{-1} are neglected, so the expectations of all the cross products in this expression are zero, because of the assumptions that ξ, δ and ε are mutually uncorrelated and to order n^{-1} terms such as $E\left[(\xi_i - \bar{\xi})\right]$ are zero. Hence

$E\left[ns_{xxxy}\right] = n\beta\mu_{\xi 3}$.

Appendix 2

Suppose estimators $\tilde{\theta}$ and $\tilde{\phi}$ of parameters θ and ϕ are calculated from two sample moments u and v. The formulas below can readily be generalised for cases where three or four sample moments are used in the estimator.

$$
\begin{aligned}
\tilde{\theta} &= f(u,v) \\
\tilde{\phi} &= g(u,v)
\end{aligned}
$$

Let $\overline{\dfrac{\partial f}{\partial u}} = \left.\dfrac{\partial f}{\partial u}\right|_{u=E[u]}$, the partial derivative evaluated at the expected values of the sample moment, u. Then,

$$
\operatorname{Var}\left[\tilde{\theta}\right] \approx \left\{\overline{\frac{\partial f}{\partial u}}\right\}^2 \operatorname{Var}[u] + \left\{\overline{\frac{\partial f}{\partial v}}\right\}^2 \operatorname{Var}[v] + 2\left\{\overline{\frac{\partial f}{\partial u}}\right\}\left\{\overline{\frac{\partial f}{\partial v}}\right\} \operatorname{Cov}[u,v]
$$

$$
\operatorname{Cov}\left[\tilde{\theta},\tilde{\phi}\right] \approx \overline{\frac{\partial f}{\partial u}}\,\overline{\frac{\partial g}{\partial u}} \operatorname{Var}[u] + \overline{\frac{\partial f}{\partial v}}\,\overline{\frac{\partial g}{\partial v}} \operatorname{Var}[v] + \left(\overline{\frac{\partial f}{\partial u}}\,\overline{\frac{\partial g}{\partial v}} + \overline{\frac{\partial f}{\partial v}}\,\overline{\frac{\partial g}{\partial u}}\right) \operatorname{Cov}[u,v].
$$

The algebra required to work out the variances and the covariances is quite lengthy. Nevertheless the resulting formulas are not difficult, and estimates of the parameters needed to estimate these variances and covariances are readily obtained from Sections 4, 5 and 6.

13. Figures

Figure 1

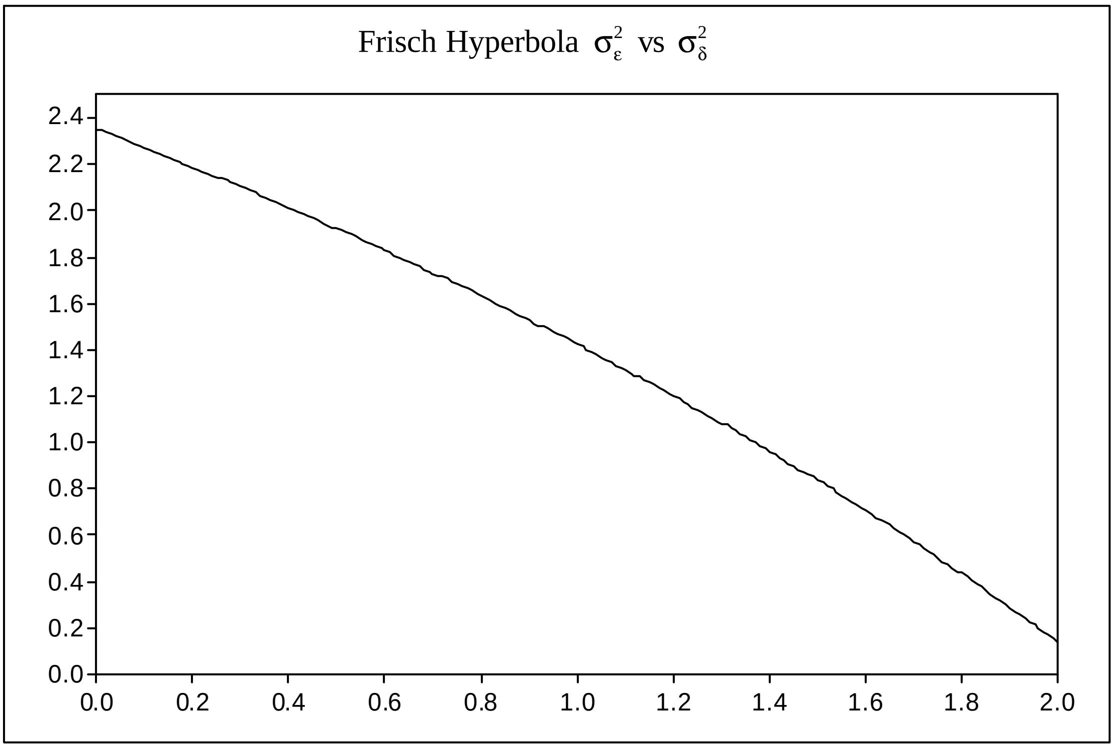

Figure 2