In Haagen K., Bartholomev D.J. & Deistler M. (Eds.):
*Statistical Modeling and Latent Variables*, pp. 341-355, 1993.

# General Least Squares Regression in Linear Errors-in-Variables Models with Correlated Errors

Liqun Wang
Institute for Statistics and Econometrics, University of Basel
Petersgraben 51, CH-4051 Basel, Switzerland

**Abstract**

The class of the so-called general least squares procedures in the linear errors-in-variables models with correlated noise components is considered. The word "general" means that the data are projected onto the hyperplanes along an arbitrarily given direction, not necessarily parallel to the axes of coordinates. A full description of the structure of this class is given. The asymptotic properties of the estimators are investigated. Monte Carlo simulations of the asymptotic biases of the general least squares, the total least squares and the maximum likelihood estimators are presented.

## 1 Introduction

Errors-in-variables models have received more and more attention in recent years. Aigner et al (1984), Anderson (1984) and Fuller (1987) summarize various aspects and results, mainly for static models. Whereas a recent, rather thorough reference of the literature is given in Deistler and Anderson (1989), with emphasis on dynamic systems. Another direction of research of the problem is using non-statistical framework. In a series of lectures Kalman (1990) investigates the problem in such a setting and provides some very deep results, one of which is the general least squares scheme for estimating the regression coefficients for given number of equations. This paper is largely motivated by Kalman (1990). However here we are mainly concerned with statistical aspects and applications of the general least squares procedures.

Consider the following linear errors-in-variables model

$$y_t = \alpha + \beta'\xi_t + u_t, \quad x_t = \xi_t + v_t, \tag{1.1}$$

where $\xi_t \in \mathbb{R}^m$ are the vectors of unobserved variables, $y_t \in \mathbb{R}$ and $x_t \in \mathbb{R}^m$ are the observed variables, $u_t$ and $v_t$ are the corresponding measurement errors and $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^m$ are the parameters.

The problem of estimating the slope parameters in this model has a long history. It is well-known that the ordinary least squares (OLS) estimator is no longer optimal in the usual sense and is biased toward the origin. A straightforward generalization of the OLS procedure is the total least squares (TLS) (sometimes also called orthogonal regression), in which the perpendicular distance of the hyperplane to the data are minimized.

However, we observe that in some situations, especially when the errors $u_t$ and $v_t$ are highly correlated, the observed data $(y_t, x_t')$ tend to be deviate from the true values $(\alpha + \beta'\xi_t, \xi_t')$ in a common direction. This suggests another possibility to generalize the OLS procedure: first, the data are projected onto the hyperplanes along the fixed "true" direction, not necessarily parallel to the axes of coordinates, and then, the sum of squared residuals is minimized. For notational convenience we call this procedure the directed least squares (DLS) procedure.

From the statistical point of view it is well-known that the model (1.1) suffers from the problem of identification (Hsiao 1983). Consequently in practical applications usually a certain kind of a priori information must be used to reduce the number of unknown parameters. One typical example of such cases is the maximum likelihood (ML) estimation, presuming that certain information about the error covariances be available (Fuller 1987). In this paper we show that in such cases this information may also be used to calculate the DLS estimate which is equally good as and sometimes even better than the ML and the TLS estimates.

In section 2 we derive the general form of the DLS estimate for an arbitrarily given direction. Section 3 deals with the statistical properties of the estimator. In section 4 we show how to use the information in the error covariances to choose a direction to apply the DLS estimator for two typical models which are used very often in practical applications. Results of Monte Carlo simulations of the asymptotic biases of the DLS, the TLS and the ML estimators are also presented.

## 2 The directed least squares estimates

In this section we first derive the general form of the DLS estimate for an arbitrarily given direction and then discuss some geometric and algebraic properties of the class of all such estimates. Without loss of generality we assume that the data $z_t = (y_t, x_t')' \in I\!R^n$, $n = m + 1$, $t = 1, 2, \ldots, T$ are generated from the model

$$y_t = \beta'\xi_t + u_t, \quad x_t = \xi_t + v_t \tag{2.1}$$

and the data moment matrix is nonsingular: $M_z = \frac{1}{T}\sum_{t=1}^{T} z_t z_t' > 0$. In this case the model may be written as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}, \quad A' \begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix} = 0, \quad A' = (1, -\beta') \in I\!R^n.$$

Given any vector $D = (1, d')' \in \mathbb{R}^n$ (that the first element is taken to be one is not a restriction of generality), the projection in $\mathbb{R}^n$ along the direction D onto the hyperplane $\Pi : A'z = 0$, where $A'D \neq 0$, is $P_d = I - D(A'D)^{-1}A'$. Thus the sum of squared residuals (divided by the sample size T) is

$$\begin{aligned} L(d,\beta) &= \frac{1}{T}\sum_{t=1}^{T}\|z_t - P_d z_t\|^2 \\ &= \frac{1}{T}\sum_{t=1}^{T} z_t'(I-P_d)'(I-P_d)z_t \\ &= \operatorname{tr}\big((I-P_d)'(I-P_d)M_z\big) \\ &= D'D(A'D)^{-2}A'M_zA. \end{aligned} \tag{2.2}$$

The least squares solution to the problem is then to find the $\hat{\beta}$ which minimizes (2.2). Applying the well-known Cauchy-Schwarz inequality we have, for every $\beta \in \mathbb{R}^m$,

$$L(d,\beta) \geq D'D(D'M_z^{-1}D)^{-1}$$

and the minimum is attained if and only if

$$M_z\hat{A} = cD, \tag{2.3}$$

where $\hat{A} = (1, -\hat{\beta}')'$ and $c$ is a constant. Writing (2.3) as

$$\begin{pmatrix} M_y & M_{xy}' \\ M_{xy} & M_x \end{pmatrix}\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix} = c\begin{pmatrix} 1 \\ d \end{pmatrix}$$

and solving this matrix equation we have

$$\hat{\beta}_d = (M_x - dM_{xy}')^{-1}(M_{xy} - dM_y), \tag{2.4}$$

provided $\det(M_x - dM_{xy}') \neq 0$.

Thus we have obtained the least squares solution to the problem of estimation of slope parameters in model (2.1). We call $\hat{\beta}_d$ in (2.4) the DLS estimate of $\beta$. Note that this estimate depends on the direction d, which is unknown in general. In fact, every vector $d \in \mathbb{R}^m$ satisfying $\det(M_x - dM_{xy}') \neq 0$ may serve as a direction. On the other hand, it is easy to verify that, for the given data $M_z > 0$, any $\beta \in \mathbb{R}^m$ satisfying $M_y - M_{xy}'\beta \neq 0$ is a least squares solution if we choose the direction

$$d = (M_y - M_{xy}'\beta)^{-1}(M_{xy} - M_x\beta). \tag{2.5}$$

Let $\mathcal{D} = \{d \in \mathbb{R}^m | \det(M_x - dM_{xy}') \neq 0\}$ and $\mathcal{B} = \{\beta \in \mathbb{R}^m | M_y - M_{xy}'\beta \neq 0\}$, then the class of all DLS estimates $\{\hat{\beta}_d \mid d \in \mathcal{D}\}$ is exactly $\mathcal{B}$ and, viewed as a mapping, $\hat{\beta}_d$ defined by (2.4) is one to one between $\mathcal{D}$ and $\mathcal{B}$. Furthermore for $n = 2$, it is easily seen that $\hat{\beta}_d$ is strictly decreasing in the intervals $(-\infty, m_x/m_{xy})$ and $(m_x/m_{xy}, +\infty)$. We summarize the above discussion in the following theorem.

**Theorem 2.1.** *Given the data* $M_z$ *satisfying* $\det M_z \neq 0$, *then*

*(1) Given any direction* $d \in \mathcal{D}$, *the loss function* $L(d, \beta)$ *in (2.2) is minimized at* $\hat{\beta}_d$ *which is given by* (2.4).

*(2) Conversely, for any* $\beta \in \mathcal{B}$, *there is one direction* $d \in \mathbb{R}^m$ *(given by (2.5)), such that* $\hat{\beta}_d$ *is the corresponding least squares solution.*

*(3) The mapping* $d \to \hat{\beta}_d$ *is a homeomorphism between* $\mathcal{D}$ *and* $\mathcal{B}$, *which are the open and dense subsets of* $\mathbb{R}^m$.

*(4) For* $n = 2$, $\hat{\beta}_d$ *is piecewise strictly decreasing on* $\mathcal{D}$.

**Remark 2.1.** If we do not have any a priori information about the direction d, then it seems natural to minimize the loss

$$L(d, \hat{\beta}_d) = D'D(D'M_z^{-1}D)^{-1} \tag{2.6}$$

with respect to $d$ further to get an overall optimal solution. It is easily seen that this will lead to the total least squares (TLS) estimate. Indeed, the right hand side of (2.6) attains the minimum if and only if $D$ is the eigenvector of $M_z$ associated with the smallest eigenvalue $\lambda = \lambda_{min}$, i.e., if and only if D satisfies $M_z D = \lambda_{min} D$, which gives $d_{TLS} = (\lambda_{min} I - M_x)^{-1} M_{xy}$. Now if $\hat{A}$ is determined by (2.3), then $M_z \hat{A} = \lambda_{min} \hat{A}$. Solving this equation we obtain the familiar TLS estimate

$$\hat{\beta}_{TLS} = (M_x - \lambda_{min} I)^{-1} M_{xy}.$$

However, the derivation of the estimates here gives an interpretation other than the traditional ones, e.g. that the TLS has the minimum perpendicular distance to the data. As has been mentioned earlier, the TLS procedure seems to be rational if we do not have any information about the errors in the data. As is easily seen however, the TLS ignores such kind of information if they are available, which is more often than not the case in controlled experiments. As will be shown later, in such cases TLS is not the optimal solution in the class of all DLS estimates.

**Remark 2.2.** Finally we note that in applications usually the intercept is explicitly expressed in the model, as in (1.1). Thus it is convenient to give an explicit expression for the estimates. The calculations are straightforward if the model (1.1) is viewed as a special case of model (2.1) in which one component of $\xi_t$ takes the constant one. The only issue to be careful is that, since the constant variable one is error free, the corresponding direction adjustment should be zero, i.e., the direction vector should be taken as $(1, 0, d')'$. Thus the corresponding estimates of the parameters in model (1.1) are given by

$$\hat{\beta}_d = (S_x - dS'_{xy})^{-1}(S_{xy} - dS_y), \quad \hat{\alpha}_d = \bar{y} - \bar{x}'\hat{\beta}_d,$$

provided $\det(S_x - dS'_{xy}) \neq 0$, where

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t, \quad S_x = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(x_t - \bar{x})', \quad S_{xy} = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})$$

and $\bar{y}$ and $S_y$ are similarly defined.

# 3 Statistical properties of the DLS estimator

In this section we are concerned with the statistical aspect of the DLS estimator. Usually the model (1.1) has two different forms, namely the structural form and the functional form, depending on whether the $\xi_t$ are stochastic variables or fixed parameters. In this paper we make the following assumptions for the structural form

**(S1)** $\xi_t$ has finite first and second moments $\mu_\xi = E\xi_t$ and $\Sigma_\xi = E(\xi_t - \mu_\xi)(\xi_t - \mu_\xi)' > 0$.

**(S2)** $\xi_t$ and $\epsilon_t = (u_t, v_t')'$ are independent.

Whereas for the functional form we assume

**(F1)** For $T \to \infty$, $\frac{1}{T}\sum_{t=1}^{T} \xi_t \to \mu_\xi \in I\!R^m$ exists.

**(F2)** For $T \to \infty$, $\frac{1}{T}\sum_{t=1}^{T} (\xi_t - \mu_\xi)(\xi_t - \mu_\xi)' \to \Sigma_\xi > 0$.

Under either of the two groups of assumptions and notations the following moments equations hold

$$\mu_y = \alpha + \beta'\mu_\xi, \quad \mu_x = \mu_\xi \tag{3.1}$$

$$\Sigma_y = \beta'\Sigma_\xi\beta + \Sigma_u, \quad \Sigma_{xy} = \Sigma_\xi\beta + \Sigma_{vu}, \quad \Sigma_x = \Sigma_\xi + \Sigma_v. \tag{3.2}$$

For a proof see e.g. Schneeweiss and Mittag (1986). However, since our derivations of the asymptotic properties of the DLS estimator are based solely on these moment equations, it is not necessary to distinguish between the two forms of the model. Given the data, the equations (3.1) are used to determine the intercept $\alpha$ and (3.2) the slope parameter $\beta$. However, we are mainly interested in the estimator of the slope parameter $\hat\beta_d$, since the calculation of $\hat\alpha_d$ and the derivation of its properties are mathematically straightforward, e.g. the consistency of $\hat\alpha_d$ follows immediately from the consistency of $\hat\beta_d$.

First let us consider the consistency of the DLS estimator. We say that the estimator $\hat\beta_d$ is consistent for $\beta$, if $\text{plim}_{T\to\infty}\hat\beta_d = \beta$ holds for all $\beta \in I\!R^m$. The following theorem shows the consistency of the DLS estimator.

**Theorem 3.1.** *If* $\det(\Sigma_x - d\Sigma_{xy}') \neq 0$, *then*

*(1)* $\text{plim}_{T\to\infty}\hat\beta_d = \beta$ *holds if and only if* $d = (\Sigma_u - \Sigma_{vu}'\beta)^{-1}(\Sigma_{vu} - \Sigma_v\beta)$.

*(2)* $\text{plim}_{T\to\infty}\hat\beta_d = \beta$ *for all* $\beta \in I\!R^m$ *if and only if* $\Sigma_u\Sigma_v = \Sigma_{vu}\Sigma_{vu}'$ *and* $d = \Sigma_{vu}\Sigma_u^{-1} = \Sigma_v\Sigma_{vu}(\Sigma_{vu}'\Sigma_{vu})^{-1}$.

**Proof.** (1) From the definition of the DLS (2.4) and the moment equations (3.2) we have

$$\begin{aligned}\operatorname*{plim}_{T\to\infty}\hat{\beta}_d &= (\Sigma_x - d\Sigma'_{xy})^{-1}(\Sigma_{xy} - d\Sigma_y)\\ &= \big((I-d\beta')\Sigma_\xi + \Sigma_v - d\Sigma'_{vu}\big)^{-1}\big((I-d\beta')\Sigma_\xi\beta + \Sigma_{vu} - d\Sigma_u\big)\\ &= \beta + \big((I-d\beta')\Sigma_\xi + \Sigma_v - d\Sigma'_{vu}\big)^{-1}(\Sigma_{vu} - d\Sigma_u - (\Sigma_v - d\Sigma'_{vu})\beta). \end{aligned} \tag{3.3}$$

Thus $\operatorname{plim}_{T\to\infty}\hat{\beta}_d = \beta$ if and only if

$$\Sigma_{vu} - d\Sigma_u - (\Sigma_v - d\Sigma'_{vu})\beta = 0 \tag{3.4}$$

or equivalently $d = (\Sigma_u - \Sigma'_{vu}\beta)^{-1}(\Sigma_{vu} - \Sigma_v\beta)$, provided $\Sigma_u - \Sigma'_{vu}\beta \neq 0$.

(2) From (3.4) $\operatorname{plim}_{T\to\infty}\hat{\beta}_d = \beta$ for all $\beta$ if and only if $\Sigma_{vu} - d\Sigma_u = 0$ and $\Sigma_v - d\Sigma'_{vu} = 0$, which is equivalent to the conditions in the theorem. □

From the above theorem we see that the consistency can be achieved only in the case where the error covariance matrix is singular. Consequently it is necessary to consider another criterion of the estimation error when the DLS estimator is applied. One natural criterion is the asymptotic absolute bias (AAB) of the estimator, which is defined as

$$\mathrm{AAB}(\hat{\beta}_d, \beta) = \|\operatorname*{plim}_{T\to\infty}\hat{\beta}_d - \beta\| = \|\beta_d - \beta\|. \tag{3.5}$$

In order to calculate this bias, let us first discuss more about the direction $d$. For notational simplicity in the remaining part of this section we restrict ourselves to the case $n = 2$. For this case we use the lower case letters to denote variables and moments. Thus the error covariance matrix is

$$\Sigma_\epsilon = \begin{pmatrix}\sigma_u & \sigma_{vu}\\ \sigma_{vu} & \sigma_v\end{pmatrix} = \sigma_u\begin{pmatrix}1 & \rho\sqrt{\delta}\\ \rho\sqrt{\delta} & \delta\end{pmatrix},$$

where $\rho = \sigma_{vu}/\sqrt{\sigma_v\sigma_u}$, $\delta = \sigma_v/\sigma_u$ are the error correlation coefficient and the error variance ratio respectively. In the case where the error covariance matrix $\Sigma_\epsilon$ is nonsingular, the two directions

$$d_1 = \frac{\sigma_{vu}}{\sigma_u} = \rho\sqrt{\delta}, \quad d_2 = \frac{\sigma_v}{\sigma_{vu}} = \frac{\sqrt{\delta}}{\rho} \tag{3.6}$$

given in Theorem 3.1.(2) are not identical. There is another intuitive reason that $d_1$ and $d_2$ play a special role, namely they are the two extreme bounds of all possible directions if the average direction in the real errors is viewed as the regression coefficient in the linear relation $v_t = du_t$. Indeed, by (3.6), the intervals between $d_1$ and $d_2$ will cover the whole real line as $\rho \to 0$ and they will concentrate at one point $\sqrt{\delta}$ and $-\sqrt{\delta}$ as $\rho \to 1$ and $\rho \to -1$ respectively.

Next we look at in more detail the function

$$\beta_d = \operatorname*{plim}_{T\to\infty} \hat{\beta}_d = \frac{\sigma_{xy} - d\sigma_y}{\sigma_x - d\sigma_{xy}}.$$

It is easily seen that $\beta_d$ is strictly decreasing in $d$ in each interval of $(-\infty, \sigma_x/\sigma_{xy})$ and $(\sigma_x/\sigma_{xy}, +\infty)$. Thus if $d_1$ and $d_2$ lie at the same side of $\sigma_x/\sigma_{xy}$, then for any $d$ between $d_1$ and $d_2$ the AAB of $\hat{\beta}_d$ in (3.5) does not exceed that of $\hat{\beta}_{d1}$ and $\hat{\beta}_{d2}$.

On the other hand, it is also easily seen that the function $\beta_d$ changes very fast around its odd pole $\sigma_x/\sigma_{xy}$, i.e., any small change in $d$ may lead to a very large change in $\beta_d$. This sensitivity might lead to large estimation bias if the true direction lies in this region and is not known exactly. However this will not occur in the case $|\rho| = 1$. Because of the one to one correspondence between $d$ and $\beta_d$, it is possible and also more convenient to describe this "sensitive region" of the DLS procedure in terms of $\beta$.

**Lemma 3.1.** *$\sigma_x/\sigma_{xy}$ lies between $d_1$ and $d_2$ if and only if $\beta$ lies between*

$$b_1 = \frac{\rho}{\sqrt{\delta}} \quad \text{and} \quad b_2 = \frac{1 + (1-\rho^2) r_{v\xi}}{\rho\sqrt{\delta}}, \tag{3.7}$$

*where $r_{v\xi} = \sigma_v/\sigma_\xi$ is the so-called noise-to-signal ratio. The directions of inequalities depend on the sign of the correlation coefficient $\rho$.*

**Proof.** We consider the case where $\sigma_{vu} > 0$. From (3.2) it follows that $d_1 < \sigma_x/\sigma_{xy} < d_2$ if and only if

$$\frac{\sigma_{vu}}{\sigma_u} < \frac{\sigma_x}{\sigma_{xy}} = \frac{\sigma_\xi + \sigma_v}{\sigma_\xi \beta + \sigma_{vu}} < \frac{\sigma_v}{\sigma_{vu}},$$

which is easily shown to be equivalent to $b_1 < \beta < b_2$. The case $\sigma_{vu} < 0$ may be treated in the same way with all inequalities reversed. □

It is easily seen that the sensitive interval corresponds to a certain neighbourhood of the odd pole of the function $\beta_d$ and the intervals for positive $\rho$ are symmetric about zero to the intervals for negative $\rho$. Table 3.1 shows some sensitive intervals for $0 < \rho < 1$, in which case the sensitive interval is $(b_1, b_2)$. We observe that most of these sensitive intervals are around 1 (except for $\delta = 0.5$ and $\rho = 0.9$) and have a width no more than 2.5.

As has been mentioned earlier, in practical applications the DLS procedure might have relatively large asymptotic bias if the unknown true parameter lies in the sensitive interval. Thus the DLS procedure is recommended only if there is strong a priori information that the true parameter $\beta$ lies outside the sensitive interval. In this case it is possible to give the bound for the asymptotic bias for the DLS estimator.

Table 3.1: Some sensitive intervals for $\beta$.

| | | | |
|---|---|---|---|
| | $\delta = 1$ | | |
| | $r_{v\xi} = 0.1$ | $r_{v\xi} = 0.01$ | $r_{v\xi} = 0.001$ |
| $\rho = 0.9$ | (0.9, 1.13) | (0.9, 1.11) | (0.9, 1.11) |
| $\rho = 0.7$ | (0.7, 1.50) | (0.7, 1.44) | (0.7, 1.43) |
| $\rho = 0.5$ | (0.5, 2.15) | (0.5, 2.02) | (0.5, 2.00) |
| | $\delta = 0.5$ | | |
| | $r_{v\xi} = 0.1$ | $r_{v\xi} = 0.01$ | $r_{v\xi} = 0.001$ |
| $\rho = 0.9$ | (1.27, 1.60) | (1.27, 1.57) | (1.27, 1.57) |
| $\rho = 0.7$ | (0.99, 2.12) | (0.99, 2.03) | (0.99, 2.02) |
| $\rho = 0.5$ | (0.71, 3.04) | (0.71, 2.85) | (0.71, 2.83) |

**Theorem 3.2.** *Suppose $d_1$ and $d_2$ lie at the same side of $\sigma_x/\sigma_{xy}$. Let $b_1, b_2$ be as in (3.7) and $c > 0$ be any given constant. Then for any $d$ between $d_1$ and $d_2$ and for all $\beta$ in the region*

$$|\beta - b_1| \geq \frac{1}{c\delta}, \quad |1 - \frac{b_2}{\beta}| \geq \frac{1}{c|\rho|\sqrt{\delta}} \tag{3.8}$$

*it holds*

$$\mathrm{AAB}(\hat{\beta}_d, \beta) = |\operatorname*{plim}_{T\to\infty} \hat{\beta}_d - \beta| \leq c(1-\rho^2) r_{v\xi}. \tag{3.9}$$

**Proof.** From (3.3) we have, for any $d$,

$$\beta_d - \beta = \frac{\sigma_{vu} - d\sigma_u - (\sigma_v - d\sigma_{vu})\beta}{(1-d\beta)\sigma_\xi + \sigma_v - d\sigma_{vu}}.$$

Thus by (3.6)

$$\begin{aligned}\beta_{d1} - \beta &= \frac{-(\sigma_v - d_1\sigma_{vu})\beta}{(1-d_1\beta)\sigma_\xi + \sigma_v - d_1\sigma_{vu}} \\ &= \frac{-(1-\rho^2) r_{v\xi}\beta}{1 - d_1\beta + (1-\rho^2) r_{v\xi}} \\ &= \frac{(1-\rho^2) r_{v\xi}}{(1 - b_2/\beta)\rho\sqrt{\delta}}\end{aligned}$$

and

$$\begin{aligned}\beta_{d2} - \beta &= \frac{\sigma_{vu} - d_2\sigma_u}{(1-d_2\beta)\sigma_\xi} \\ &= \frac{(1-\rho^2) r_{v\xi}}{(\beta - b_1)\delta}.\end{aligned}$$

Since $d_1$ and $d_2$ lie at the same side of $\sigma_x/\sigma_{xy}$, the function $\beta_d$ is strictly decreasing in the interval between $d_1$ and $d_2$. It follows that for any $d$ between $d_1$ and $d_2$

$$\begin{aligned}|\beta_d - \beta| &\le \max\{|\beta_{d1} - \beta|,\ |\beta_{d2} - \beta|\}\\ &= (1-\rho^2) r_{v\xi} \max\{\frac{1}{|\beta - b_1|\delta}\ \frac{1}{|1 - b_2/\beta||\rho|\sqrt{\delta}}\}\\ &\le (1-\rho^2) r_{v\xi} c,\end{aligned}$$

provided $\beta$ lies in the region (3.8). □

From (3.8) and (3.9) we see that the choice of the constant $c$ is a trade-off between the estimation precision and the range of admissible $\beta$'s. It is also easy to calculate that for any $c$ satisfying

$$0 < c < \frac{1}{\sqrt{\delta}(\sqrt{1 + (1-\rho^2) r_{v\xi}} - |\rho|)} \tag{3.10}$$

the region (3.8) will exclude the sensitive interval between $b_1$ and $b_2$. Thus by Lemma 3.1 and Theorem 3.2 we have the following result.

**Corollary 3.1.** *Let $c$ be any constant satisfying (3.10). Then (3.9) holds for any $d$ between $d_1$ and $d_2$ and all $\beta$ in the region (3.8).*

**Example 3.1.** Let $\delta = 1$, $r_{v\xi} = 0.01$, $c = 2$ and $\rho = 0.7$, then by (3.6) $d_1 = 0.7$ and $d_2 = 1.43$. From Table 3.1 we see that the sensitive interval is $(b_1, b_2) = (0.7, 1.44)$. Now it is easy to verify that the region (3.8) is the union of the intervals $(-\infty, 0.2)$ and $(5.03, +\infty)$ and this region does not contain the sensitive interval. Thus by Theorem 3.2, for any $0.7 < d < 1.43$ and all $\beta$ being outside the interval $(0.2, 5.03)$, it holds $|\operatorname{plim}_{T\to\infty} \hat\beta_d - \beta| \le c(1-\rho^2) r_{v\xi} = 0.0102$.

In the remaining part of this section we consider a kind of weaker consistency in the sense that $\hat\beta_d$ converges to the true $\beta$ in probability as the sample size T is large and as the noise-to-signal ratio is small. Formally, we consider the convergence

$$\hat\beta_d \xrightarrow{P} \beta, \quad \text{as} \quad T \to \infty \quad \text{and} \quad \frac{tr\Sigma_\epsilon}{tr\Sigma_\xi} \to 0, \tag{3.11}$$

where $\Sigma_\epsilon$ is the variance-covariance matrix of the errors $\epsilon_t = (u_t, v_t')'$. For this kind of weaker consistency we have the following result.

**Theorem 3.3.** *If $n = 2$, then (3.11) holds for any $d$, $\beta$ satisfying $d\beta \ne 1$.*

**Proof.** From (3.3) we have

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \hat{\beta}_d &= \beta + \frac{\sigma_{vu} - d\sigma_u - (\sigma_v - d\sigma_{vu})\beta}{(1-d\beta)\sigma_\xi + \sigma_v - d\sigma_{vu}} \\
&= \beta + \frac{(1+d\beta)\rho\sqrt{\sigma_v\sigma_u} - \beta\sigma_v d\sigma_u}{(1-d\beta)\sigma_\xi + \sigma_v - d\rho\sqrt{\sigma_v\sigma_u}} \\
&= \beta + \frac{(1+d\beta)\rho\sqrt{(\sigma_v/\sigma_\xi)(\sigma_u/\sigma_\xi)} - \beta(\sigma_v/\sigma_\xi) - d(\sigma_u/\sigma_\xi)}{(1-d\beta) + (\sigma_v/\sigma_\xi) - d\rho\sqrt{(\sigma_v/\sigma_\xi)(\sigma_u/\sigma_\xi)}} \\
&= \beta + O\left(\frac{\sigma_v + \sigma_u}{\sigma_\xi}\right)
\end{aligned}$$

and the result follows immediately. □

# 4 Applications of the DLS estimator

In this section we explain how to use the a priori information in the error covariance matrix to choose a suitable direction for the DLS estimator. We also compare it with the usual TLS and the ML estimators under the AAB criterion

$$\mathrm{AAB}(\hat{\beta}, \beta) = |\operatorname*{plim}_{T\to\infty} \hat{\beta} - \beta|.$$

Throughout this section we consider the univariate model ($n = 2$)

$$y_t = \beta\xi_t + u_t, \quad x_t = \xi_t + v_t,$$

where $\xi_t \sim N(0,1)$, $\epsilon_t = (u_t, v_t)' \sim N(0, \Sigma_\epsilon)$ and $\xi_t$ and $\epsilon_t$ are independent. Especially we assume that $\sigma_{vu} \neq 0$. We will consider two typical models in this form, namely the model with an error in the equation and the model with no error in the equation. For more details about these models as well as the corresponding TLS and ML estimators, we refer the reader to Fuller (1987).

## 4.1 The model with an error in the equation

In this model it is assumed that in the covariance matrix

$$\Sigma_\epsilon = \begin{pmatrix} \sigma_u & \sigma_{vu} \\ \sigma_{vu} & \sigma_v \end{pmatrix}$$

$\sigma_v, \sigma_{vu}$ are known, whereas $\sigma_u$ is not. In this case we may choose $d = \sigma_v/\sigma_{vu} = \sqrt{\delta}/\rho$ for the DLS estimator. Now the smallest eigenvalue of $M_z$ is

$$\lambda_{min} = \frac{m_y + m_x - \sqrt{(m_y - m_x)^2 + 4m_{xy}^2}}{2}$$

Table 4.1: The ML, DLS, TLS and OLS estimates for the models with an error in the equation: $\rho = 1$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 57.22 | 19.12 | 9.57 | 4.81 | 2.91 | 0.52 | 0.05 |
| DLS | 60.03 | 20.03 | 10.00 | 5.00 | 3.00 | 0.50 | 0.00 |
| TLS | 59.98 | 19.98 | 9.96 | 4.96 | 2.96 | 0.53 | 0.05 |
| OLS | 57.20 | 19.11 | 9.57 | 4.81 | 2.90 | 0.52 | 0.05 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | -0.43 | -0.91 | -2.81 | -4.72 | - 9.49 | -19.00 | -57.12 |
| DLS | -0.50 | -1.00 | -3.00 | -5.00 | -10.02 | -20.00 | -60.01 |
| TLS | -0.47 | -1.00 | -3.04 | -5.05 | -10.07 | -20.05 | -60.06 |
| OLS | -0.43 | -0.91 | -2.81 | -4.71 | - 9.49 | -18.99 | -57.09 |

Table 4.2: AAB's of the ML, DLS, TLS and OLS estimators for the models with an error in the equation: $\rho = 1$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 2.792 | 0.886 | 0.429 | 0.190 | 0.095 | 0.024 | 0.046 |
| DLS | 1.027 | 0.350 | 0.167 | 0.073 | 0.036 | 0.009 | 0.018 |
| TLS | 1.027 | 0.348 | 0.171 | 0.081 | 0.049 | 0.028 | 0.049 |
| OLS | 2.818 | 0.894 | 0.433 | 0.191 | 0.096 | 0.024 | 0.047 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | 0.069 | 0.094 | 0.191 | 0.286 | 0.509 | 1.002 | 2.893 |
| DLS | 0.026 | 0.036 | 0.072 | 0.110 | 0.198 | 0.367 | 1.080 |
| TLS | 0.037 | 0.038 | 0.079 | 0.118 | 0.205 | 0.369 | 1.082 |
| OLS | 0.070 | 0.095 | 0.193 | 0.288 | 0.514 | 1.012 | 2.921 |

and hence the TLS estimator is given by

$$\begin{aligned}\hat{\beta}_{TLS} &= \frac{m_{xy}}{m_x - \lambda_{min}} \\ &= \frac{m_y - m_x + \sqrt{(m_y - m_x)^2 + 4m_{xy}^2}}{2m_{xy}}.\end{aligned}$$

Note that now the ML estimate does not always exist and is given by

$$\hat{\beta}_{ML} = \frac{m_{xy} - \sigma_{vu}}{m_x - \sigma_v},$$

provided $\sigma_v < m_x$ and $(m_{xy} - \sigma_{vu})^2 \leq (m_x - \sigma_v)(m_y - \sigma_{vu}^2/\sigma_v)$. However, these conditions are practically always fulfilled in Monte Carlo simulations.

Table 4.3: The ML, DLS, TLS and OLS estimates for the models with an error in the equation: $\rho = 0.8$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 57.16 | 19.08 | 9.56 | 4.80 | 2.90 | 0.51 | 0.04 |
| DLS | 60.01 | 19.99 | 10.01 | 5.00 | 3.01 | 0.44 | -0.02 |
| TLS | 59.96 | 19.95 | 9.96 | 4.96 | 2.97 | 0.52 | 0.04 |
| OLS | 57.14 | 19.07 | 9.56 | 4.80 | 2.89 | 0.51 | 0.04 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | -0.44 | -0.91 | -2.82 | -4.73 | - 9.47 | -19.01 | -57.08 |
| DLS | -0.51 | -1.01 | -3.00 | -5.01 | - 9.99 | -20.01 | -59.98 |
| TLS | -0.47 | -1.00 | -3.03 | -5.04 | -10.03 | -20.05 | -60.02 |
| OLS | -0.44 | -0.91 | -2.82 | -4.73 | - 9.47 | -19.00 | -57.05 |

Table 4.4: AAB's of the ML, DLS, TLS and OLS estimators for the models with an error in the equation: $\rho = 0.8$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 2.856 | 0.926 | 0.437 | 0.202 | 0.104 | 0.017 | 0.038 |
| DLS | 1.075 | 0.341 | 0.166 | 0.074 | 0.040 | 0.061 | 0.027 |
| TLS | 1.075 | 0.342 | 0.166 | 0.080 | 0.046 | 0.024 | 0.041 |
| OLS | 2.882 | 0.935 | 0.441 | 0.204 | 0.105 | 0.017 | 0.039 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | 0.063 | 0.086 | 0.182 | 0.272 | 0.527 | 0.989 | 2.933 |
| DLS | 0.029 | 0.036 | 0.070 | 0.106 | 0.199 | 0.380 | 1.092 |
| TLS | 0.035 | 0.036 | 0.075 | 0.113 | 0.201 | 0.383 | 1.093 |
| OLS | 0.064 | 0.087 | 0.184 | 0.274 | 0.532 | 0.998 | 2.961 |

The Monte Carlo simulations of the ML, the DLS, the TLS and the OLS estimates and the corresponding AAB's are carried out for the case $\delta = 1$ and $r_{v\xi} = 0.05$. The results in Table 4.1 – 4.4 are based on 1000 replications and 100 observations in each replication. These results show that the DLS estimator has the smaller asymptotic absolute bias than the other estimators except when $\beta = 0.5$ (for $\rho = 0.8$), which lies in the sensitive interval. It is also worth noting that the DLS estimator performs clearly better than other estimators for $-20 < \beta < 20$.

Table 4.5: The ML, DLS, TLS and OLS estimates for the models with no error in the equation: $\rho = 0.9$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 60.01 | 19.99 | 10.01 | 5.00 | 3.00 | 0.50 | 0.00 |
| DLS1 | 60.02 | 20.00 | 10.02 | 5.01 | 3.02 | 0.49 | 0.00 |
| DLS2 | 60.01 | 19.99 | 10.01 | 5.00 | 3.01 | 0.48 | -0.01 |
| TLS | 59.96 | 19.95 | 9.96 | 4.96 | 2.96 | 0.53 | 0.04 |
| OLS | 57.14 | 19.07 | 9.57 | 4.80 | 2.90 | 0.52 | 0.04 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | -0.50 | -1.00 | -3.00 | -5.00 | -10.02 | -20.00 | -60.11 |
| DLS1 | -0.50 | -0.99 | -2.99 | -4.99 | -10.01 | -19.99 | -60.10 |
| DLS2 | -0.51 | -1.00 | -3.01 | -5.00 | -10.02 | -20.00 | -60.11 |
| TLS | -0.47 | -1.00 | -3.04 | -5.04 | -10.06 | -20.04 | -60.15 |
| OLS | -0.43 | -0.91 | -2.82 | -4.72 | - 9.49 | -18.99 | -57.19 |

Table 4.6: AAB's of the ML, DLS, TLS and OLS estimators for the models with no error in the equation: $\rho = 0.9$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 1.064 | 0.340 | 0.165 | 0.073 | 0.037 | 0.011 | 0.018 |
| DLS1 | 1.064 | 0.340 | 0.166 | 0.074 | 0.040 | 0.013 | 0.018 |
| DLS2 | 1.064 | 0.340 | 0.165 | 0.073 | 0.037 | 0.024 | 0.020 |
| TLS | 1.066 | 0.341 | 0.167 | 0.081 | 0.046 | 0.026 | 0.045 |
| OLS | 2.880 | 0.930 | 0.436 | 0.200 | 0.100 | 0.020 | 0.043 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | 0.027 | 0.035 | 0.069 | 0.105 | 0.198 | 0.367 | 1.084 |
| DLS1 | 0.027 | 0.036 | 0.069 | 0.105 | 0.198 | 0.367 | 1.083 |
| DLS2 | 0.028 | 0.036 | 0.069 | 0.105 | 0.198 | 0.367 | 1.084 |
| TLS | 0.038 | 0.037 | 0.077 | 0.112 | 0.206 | 0.371 | 1.089 |
| OLS | 0.069 | 0.092 | 0.184 | 0.282 | 0.517 | 1.018 | 2.828 |

## 4.2 The model with no error in the equation

In this model it is assumed that $\Sigma_\epsilon = \sigma^2\Gamma$ where

$$\Gamma = \begin{pmatrix} 1 & \sigma_{vu}/\sigma_u \\ \sigma_{vu}/\sigma_u & \sigma_v/\sigma_u \end{pmatrix} = \begin{pmatrix} 1 & \rho\sqrt{\delta} \\ \rho\sqrt{\delta} & \delta \end{pmatrix}$$

is known whereas $\sigma^2$ is unknown. In this case we may choose either $d_1 = \rho\sqrt{\delta}$ or $d_2 = \sqrt{\delta}/\rho$ and apply the corresponding DLS estimators. Since now the error

Table 4.7: The ML, DLS, TLS and OLS estimates for the models with no error in the equation: $\rho = 0.7$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 59.95 | 20.01 | 9.99 | 5.01 | 3.00 | 0.50 | 0.00 |
| DLS1 | 59.98 | 20.05 | 10.04 | 5.06 | 3.07 | 0.48 | 0.00 |
| DLS2 | 59.95 | 20.01 | 10.00 | 5.01 | 3.01 | 0.37 | -0.04 |
| TLS | 59.91 | 19.97 | 9.96 | 4.97 | 2.98 | 0.52 | 0.03 |
| OLS | 57.12 | 19.08 | 9.55 | 4.80 | 2.89 | 0.51 | 0.03 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | -0.50 | -1.00 | -3.00 | -5.00 | -10.01 | -20.01 | -59.97 |
| DLS1 | -0.49 | -0.98 | -2.98 | -4.97 | - 9.97 | -19.98 | -59.93 |
| DLS2 | -0.52 | -1.01 | -3.01 | -5.00 | -10.01 | -20.01 | -59.97 |
| TLS | -0.48 | -1.00 | -3.03 | -5.03 | -10.04 | -20.05 | -60.00 |
| OLS | -0.44 | -0.92 | -2.82 | -4.73 | - 9.49 | -19.02 | -57.05 |

Table 4.8: AAB's of the ML, DLS, TLS and OLS estimators for the models with no error in the equation: $\rho = 0.7$.

| $\beta$ | 60 | 20 | 10 | 5 | 3 | 0.5 | 0 |
|---|---|---|---|---|---|---|---|
| ML | 1.074 | 0.351 | 0.167 | 0.075 | 0.043 | 0.013 | 0.018 |
| DLS1 | 1.075 | 0.353 | 0.172 | 0.090 | 0.079 | 0.021 | 0.018 |
| DLS2 | 1.074 | 0.351 | 0.167 | 0.076 | 0.045 | 0.129 | 0.040 |
| TLS | 1.075 | 0.351 | 0.170 | 0.078 | 0.046 | 0.022 | 0.035 |
| OLS | 2.890 | 0.924 | 0.449 | 0.204 | 0.107 | 0.015 | 0.034 |
| $\beta$ | -0.5 | -1 | -3 | -5 | -10 | -20 | -60 |
| ML | 0.025 | 0.032 | 0.068 | 0.101 | 0.188 | 0.342 | 1.073 |
| DLS1 | 0.026 | 0.034 | 0.070 | 0.102 | 0.188 | 0.342 | 1.074 |
| DLS2 | 0.031 | 0.035 | 0.068 | 0.101 | 0.188 | 0.342 | 1.073 |
| TLS | 0.031 | 0.033 | 0.073 | 0.105 | 0.192 | 0.343 | 1.073 |
| OLS | 0.058 | 0.082 | 0.176 | 0.274 | 0.514 | 0.989 | 2.958 |

variance ratio $\delta$ is known, we may use the "adjusted" TLS estimator which is given by

$$\hat{\beta}_{TLS} = \frac{\delta m_y - m_x + \sqrt{(\delta m_y - m_x)^2 + 4\delta m_{xy}^2}}{2\delta m_{xy}}.$$

The ML estimator is defined as

$$\hat{\beta}_{ML} = \frac{m_{xy} - \lambda\rho\sqrt{\delta}}{m_x - \lambda\delta},$$

where $\lambda$ is the smallest root of the equation

$$\det(M_z - \lambda\Gamma) = \delta(1-\rho^2)\lambda^2 - (\delta m_y + m_x - 2\rho\sqrt{\delta} m_{xy})\lambda + (m_x m_y - m_{xy}^2) = 0$$

and is given by

$$\frac{\delta m_y + m_x - 2\rho\sqrt{\delta} m_{xy} - \sqrt{(\delta m_y + m_x - 2\rho\sqrt{\delta} m_{xy})^2 - 4\delta(1-\rho^2)(m_x m_y - m_{xy}^2)}}{2\delta(1-\rho^2)}$$

for $|\rho| < 1$ and by $(m_x m_y - m_{xy}^2)/(\delta m_y + m_x - 2\rho\sqrt{\delta} m_{xy})$ for $|\rho| = 1$ respectively. In the later case the ML estimator is $\hat{\beta}_{ML} = (m_{xy} - \rho\sqrt{\delta} m_y)/(m_x - \rho\sqrt{\delta} m_{xy})$ and therefore coincides with the DLS estimator $\hat{\beta}_{d1}$.

The ML, the DLS, the TLS and the OLS estimates and the corresponding AAB's are shown in the Table 4.5 – 4.8. Again 1000 replications and 100 observations in each replication are made for the case $\delta = 1$ and $r_{v\xi} = 0.05$. These results show that for $\rho = 0.9$ the DLS estimator performs generally as well as the ML and better than the TLS and the OLS estimators. For $\rho = 0.7$ no one of the estimators dominates others uniquely. Note that $\hat{\beta}_{d1}$ performs better than $\hat{\beta}_{d2}$ for $-1 < \beta < 1$ whereas $\hat{\beta}_{d2}$ is better for other $\beta$'s.

# References

[1] Aigner, D.J., Hsiao, C., Kapteyn, A. and Wansbeek, T. (1984). Latent Variable Models in Econometrics. In Z. Griliches and M.D. Intriligator (Eds.): *Handbook of Econometrics*, Vol. II, North-Holland, Amsterdam.

[2] Anderson, T.W. (1984). Estimating Linear Statistical Relationships. *Ann. Statistics*, 12, 1–45.

[3] Deistler, M. and Anderson, B.D.O. (1989). Linear Dynamic Errors-in-Variables Models: Some Structure Theory. *Journal of Econometrics*, 41, 39–63.

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

[5] Hsiao, C. (1983). Identification. In Z. Griliches and M.D. Intriligator (Eds.): *Handbook of Econometrics*, Vol. I, North-Holland, Amsterdam.

[6] Kalman, R. (1990). *Nine Lectures on Identification*. Springer Lecture Notes on Economics and Mathematical Systems. To appear.

[7] Schneeweiss, H. and Mittag, H.-J. (1986). *Lineare Modelle mit fehlerbehafteten Daten*. Physica-Verlag, Heidelberg-Wien.