

Regression with errors in variables: estimators based on third order moments

K. van Montfort

A. Mooijaart

*Department of Psychology
Leiden University, Hooigracht 15
2312 KM Leiden, The Netherlands*

J. de Leeuw

*Department of Data Theory FSW
Leiden University, Middelstegegracht 4
2312 TW Leiden, The Netherlands*

In this paper consistent and, in a well-defined sense, optimal moment-estimators of the regression coefficient in a simple regression model with errors in variables are derived. The asymptotic variance and other asymptotic properties of these estimators are given. As is known for a long time, serious estimation problems exist in this model. There are two ways out of this problem: using either *additional assumptions* or *additional information* in the data. A lot of attention has been paid to the use of additional assumptions. However, quite often this leads to rather unrealistic models. In this paper we use additional information in the data. That means here that, besides first and second order moments, third order moments are formulated as functions of the model parameters. Besides theoretical derivations a small study with generated data is discussed. This study shows that for samples larger than 50 the estimates we consider behave nicely.

Key Words & Phrases: estimation by higher order moments, BAN estimates.

1. INTRODUCTION

In this paper we discuss regression with errors in variables. This somewhat peculiar terminology refers to the following problem. We deal with a number of random variables, which we call *observed*. The basic notion in the models discussed in this paper is that there exists, *approximately*, exactly one linear relation between these observed variables. The notion of an approximate linear relationship is modelled by assuming that the observed random variables are error-perturbed versions of unobserved or *latent* variables, between which there exists one exact linear relationship. Thus relations between unobservable variables lead to inexact relationships between observables.

We shall treat in this paper the bivariate case only, although generalizations to the multivariate case are obvious from our treatment. Related work is

reported in BEKKER, WANSBEEK, and KAPTEYN (1985), and BEKKER and DE LEEUW (1986).

First of all we shall, briefly, discuss the model and show the kind of problems that arise in this model. This model was discussed in the literature for the first time by GINI (1921), but it did become well known from the work of FRISCH (1934). Thus we refer to it as the Frisch-model. This part of the paper doesn't give anything new, but merely serves as an introduction to error in variable models. See for an overview of the theoretical properties of the Frisch model MADANSKY (1959), MORAN (1971), AIGNER et al. (1984), T.W. ANDERSON (1984) and DEISTLER (1986). For the multivariable case, not discussed in this paper, we refer to SCHNEEWEISS (1976), KALMAN (1982) and KLEPPER and LEAMER (1984).

2. THE FRISCH MODEL

Consider the following simple regression model. We have two observed random variables x and y , which have a representation in terms of unobserved random variables of the form

$$x = \xi + \delta \quad (1a)$$

$$y = \alpha + \beta\xi + \epsilon \quad (1b)$$

Thus error is additive. The unobserved random variables δ and ϵ all have expectation zero, and they are independent of each other and of the random variable ξ . We write μ_x, μ_y and λ_1 for the expectations of x, y and ξ , respectively. Further, we write μ_{st} for the expectation $E((x - \mu_x)^s (y - \mu_y)^t)$ if $s + t \geq 2$, λ_s for $E((\xi - \lambda_1)^s)$ if $s \geq 2$, ω_s for $E(\delta^s)$, and θ_s for $E(\epsilon^s)$. We assume that all moments that occur in our formulas exist. For the moments around the origin of order one we find

$$\mu_x = \lambda_1 \quad (2a)$$

$$\mu_y = \alpha + \beta\lambda_1 \quad (2b)$$

Clearly these equations do not suffice to identify the three unknowns. Thus we must also look at higher order moments. The moments around the mean of order two are

$$\mu_{20} = \lambda_2 + \omega_2 \quad (2c)$$

$$\mu_{02} = \beta^2\lambda_2 + \theta_2 \quad (2d)$$

$$\mu_{11} = \beta\lambda_2 \quad (2e)$$

Now (2) defines 5 equations with 6 unknowns. Again there is no identification, but because variances are nonnegative, we must have

$$0 \leq \omega_2 \leq \mu_{20} \quad (3a)$$

$$0 \leq \theta_2 \leq \mu_{02} \quad (3b)$$

and it also follows from (2) that admissible pairs (ω_2, θ_2) are in the intersection of the rectangle defined by (3) and the hyperbola

$$(\mu_{20} - \omega_2)(\mu_{02} - \theta_2) = (\mu_{11})^2. \tag{4}$$

This is illustrated in figure 1, with $\mu_{20} = .4, \mu_{02} = .6$, and $\mu_{11} = \sqrt{.02}$.

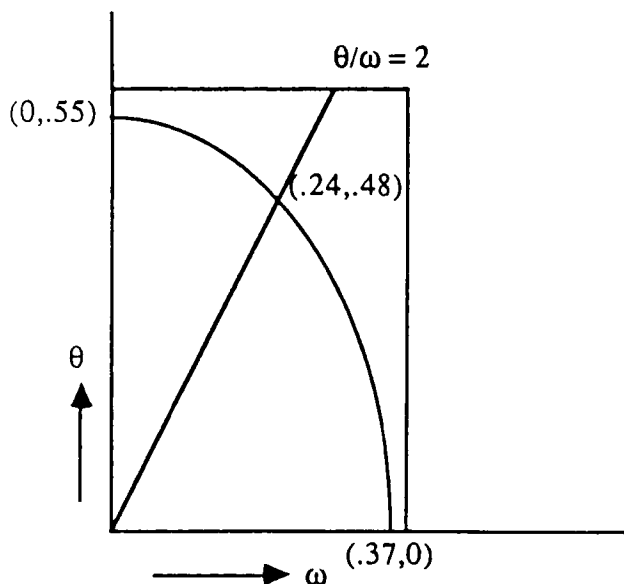


FIGURE 1. Frisch hyperbola $(.4 - \omega)(.6 - \theta) = .02$

Without additional information the only result we have about β so far is simply that

$$|\mu_{11} / \mu_{20}| \leq |\beta| \leq |\mu_{02} / \mu_{11}|, \tag{5a}$$

$$\text{sgn } \beta = \text{sgn } \mu_{11} \tag{5b}$$

(μ_{11} / μ_{20} is the classical regression coefficient for the regression of y on x and μ_{02} / μ_{11} is the reciprocal of the regression coefficient for the regression of x on y in the usual regression model without errors in variables.)

We can easily estimate the bounds on the left and the right by using sample variances and covariances, but more precision than the inequality (5) is not possible. In this paper we shall investigate if increased precision is possible by using either additional assumptions or additional information in the data.

3. ADDITIONAL ASSUMPTIONS AND PREJUDICES

There is no reason to choose any particular point in the hyperbola. KALMAN (1982, 1983) discusses this in terms of *prejudice*. In our context a prejudice is an additional assumption which makes it possible to select a point on the hyperbola, but which is not forced on us by the nature of the problem. We simply assume something in order to get rid of the uncertainty. This does not imply that each additional assumption we make is an example of prejudice. Some assumptions could be based on prior knowledge, or on physical considerations of symmetry.

The best known of these additional assumptions is $\omega_2 = 0$, i.e. x is measured without error. This may be a good approximation in the context of designed experiments. It follows from (2) that $\lambda_2 = \mu_{20}$, $\beta = \mu_{11}/\mu_{20}$, $\theta_2 = \mu_{02} - (\mu_{11})^2/\mu_{20}$, $\lambda_1 = \mu_x$ and $\alpha = \mu_y - (\mu_{11}/\mu_{20})\mu_x$. Kalman calls this the *Least squares prejudice*. Similarly we can also assume that $\theta_2 = 0$. Then $\omega_2 = \mu_{20} - (\mu_{11})^2/\mu_{02}$, $\lambda_2 = (\mu_{11})^2/\mu_{02}$, $\beta = \mu_{02}/\mu_{11}$, $\lambda_1 = \mu_x$ and $\alpha = \mu_y - (\mu_{02}/\mu_{11})\mu_x$. This is another least squares prejudice, far less common than the first one.

Another common additional assumption is that the ratio θ_2/ω_2 is known. The intersection of the line and the hyperbola in figure 1 then gives the desired answer. We can also assume that the two error variances are equal. This defines orthogonal regression, first described by ADCOCK (1878) and KUMMEL (1879), but most well-known from the work of PEARSON (1901).

The important thing is, of course, that these methods are just prejudices if they are only intended to force identification. If they incorporate true prior knowledge, the situation is different.

4. ADDITIONAL INFORMATION IN THE DATA

Besides an additional assumption which often is some kind of prejudice, as discussed above, one can use additional information in the data. Two well-known methods in the case of normally distributed variables are the method of *instrumental variables* and the *grouping method*.

Using the method of instrumental variables we need an *instrumental variable* (IV). An IV is a random variable z , which is correlated with x and is not correlated with $v = \epsilon - \beta\delta(y \equiv \beta x + v)$. The estimator $\hat{\beta}_{IV} = (S_{zx})^{-1} S_{zy}$, where S_{zx} and S_{zy} are sample covariances, is called the IV-estimator of β and is consistent under weak assumptions.

If z is uncorrelated with x the sampling variance of the IV-estimator is infinity large. Even a small correlation between z and x causes a very large sampling variance. So, the variable z has to be fairly strongly correlated with x .

In practice, there is no means for checking whether IV's are really uncorrelated with v . An instrumental variable, based on x 's, is likely to be correlated with δ 's and v 's also. The corresponding IV-estimator may not be safe even for consistent estimation of parameters. Obtaining "good" IV's is very difficult. SARGAN (1958), PAL (1981) and BEKKER et al. (1985) pay more attention to

instrumental variables.

The method of grouping divides the data into two groups - those with above and those with below median observations on x - and then fits a line through the group means. Sometimes the method of grouping divides the data into three groups, throwing out the middle one, and fits a line between the means in the upper and the lower groups. If the grouping of the observations on the basis of x could be guaranteed to be the same as the grouping of the observations on the basis of ξ , this estimator is consistent for β . This condition does not hold under most plausible distributional assumptions, for instance under normality of ξ and δ .

PAKES (1982) concludes that under non-normality we can generally do better than grouping. Further he notes that under normality the inconsistencies of the grouping estimator and the OLS-estimator are the same.

REIERSØL (1950) derives conditions for the identifiability of β . He shows that β is identifiable if there exists a nonzero (finite or infinite) cumulant κ_{rs} of the joint distribution function of x and y , with $r \geq 1, s \geq 1$, and either r or s but not both equal to 1. So, we have the result that β is identifiable if ξ is not normally distributed. When ξ is normally distributed, a necessary and sufficient condition for the identifiability of β is that neither the distribution of ϵ nor the distribution of δ is divisible by a normal distribution.

(If three variables a, b and d are such that for every $t \in R$ $c_a(t) = c_b(t) * c_d(t)$, where $c_a(\cdot), c_b(\cdot)$ and $c_d(\cdot)$ are the characteristic functions of a, b and d , we say that the distribution of a is divisible by the distribution of b and divisible by the distribution of d .)

BEKKER (1986a) formulates necessary and sufficient identification conditions for the multiple regression model with errors in the variables. Assuming normality of the errors he shows that the $p \times 1$ vector β of regression coefficients is identified if and only if there does exist a non-singular $p \times p$ matrix $A = (A_1; A_2)$ such that $\xi^T A_1$ is distributed normally and independently of $\xi^T A_2$ (ξ is here a p -dimensional random vector). So non-normality of ξ keeps identification and so consistent estimators of β based on second and higher-order moments, for example, can be constructed. The same argument was used by MOOJAART (1985) for finding a unique solution in factor analysis.

In the next part of this paper we assume ξ to be not normally distributed and we deal with the problem of estimating the regression coefficient β . This idea is closely related to the work of GEARY (1942) and PAL (1980). Their approach yields consistent and reasonably efficient estimators of regression coefficients based on uni- and bi-variate moments of third or higher order.

Using third order sample moments we derive in the next section a consistent estimator $\hat{\beta}$ of β . $\hat{\beta}$ has asymptotically minimal variance in the class of consistent β -estimators which are functions of the sample moments of x and y up to order three.

After that we assume the errors δ and ϵ to be symmetrically distributed. In that special case we also derive an "optimal" third-order-estimator $\hat{\beta}^{opt}$ of β .

5. THE USE OF THIRD ORDER MOMENTS

We shall now take third order moments into account. Thus we have the five equations (2a) - (2e) and in addition

$$\mu_{03} = \beta^3 \lambda_3 + \theta_3 \quad (6a)$$

$$\mu_{12} = \beta^2 \lambda_3 \quad (6b)$$

$$\mu_{21} = \beta \lambda_3 \quad (6c)$$

$$\mu_{30} = \lambda_3 + \omega_3 \quad (6d)$$

(If ξ is symmetrically distributed then $\lambda_3 = 0$. In that case instead of system (6) we use equations of fourth order moments.)

Now (2) and (6) are nine equations in nine unknowns, and they can be solved uniquely for the parameters. The solution is

$$\beta = \mu_{12} / \mu_{21} \quad (7a)$$

$$\lambda_1 = \mu_x \quad (7b)$$

$$\alpha = \mu_y - \mu_x \mu_{12} / \mu_{21} \quad (7c)$$

$$\lambda_3 = (\mu_{21})^2 / \mu_{12} \quad (7d)$$

$$\lambda_2 = (\mu_{11} \mu_{21}) / \mu_{12} \quad (7e)$$

$$\omega_3 = \mu_{30} - (\mu_{21})^2 / \mu_{12} \quad (7f)$$

$$\omega_2 = \mu_{20} - (\mu_{11} \mu_{21}) / \mu_{12} \quad (7g)$$

$$\theta_3 = \mu_{03} - (\mu_{12})^2 / \mu_{21} \quad (7h)$$

$$\theta_2 = \mu_{02} - (\mu_{11} \mu_{12}) / \mu_{21} \quad (7i)$$

Recall that up to now we have assumed nothing about the distribution of the variables, except that it is assumed that the means of the error variables are zero, while ξ is not normally distributed and ξ, ϵ and δ are independent of each other. These assumptions are enough to estimate the parameter β , the parameter of our main interest, and the other parameters. These estimates are given simply by substituting the sample moments m_{st} for the μ_{st} in (7). Note that it is no longer guaranteed that the estimates of λ_2, ω_2 and θ_2 are non-negative.

It follows from these considerations that $\hat{\beta} = m_{12} / m_{21}$ is optimal in the sense of minimum variance, in the class of all consistent estimators which are functions of the moments up to order three. A proof can be found in Appendix 2.

6. THE CASE OF SYMMETRIC ERRORS

We now assume, with PAL (1980), that the error variables are symmetrically distributed. This 'prejudice' makes the junction of system (2) and (6) overidentified, which means that we can look more closely into the problem of *optimal estimates*. Use of additional assumptions also makes it possible to

improve our previous 'best' estimate of β .

If $\omega_3 = \theta_3 = 0$ then (6a) - (6d) is equivalent to

$$\beta = \mu_{03} / \mu_{12} = \mu_{12} / \mu_{21} = \mu_{21} / \mu_{30}. \tag{8}$$

Expression (8) gives us three estimators of β

$$\hat{\beta} = \mathbf{m}_{03} / \mathbf{m}_{12} \tag{9a}$$

$$\hat{\beta}_2 = \mathbf{m}_{12} / \mathbf{m}_{21} \tag{9b}$$

$$\hat{\beta}_3 = \mathbf{m}_{21} / \mathbf{m}_{30}, \tag{9c}$$

where \mathbf{m}_{rs} is a consistent estimator of $\mu_{rs} (0 \leq r + s \leq 3)$. Observe that (9b) is the same as the estimator from (7a) and is consequently our earlier optimal estimator. Further (9a), (9b) and (9c) are three estimators of PAL (1980).

It is shown in Appendix 2 that an optimal estimator in the class of consistent estimators which are functions of the moments up to order three, is

$$\hat{\beta}^{opt} = (u' V^{-1} \mathbf{t}) / (u' V^{-1} u), \tag{10}$$

where $\mathbf{t}' = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$, $u' = (1, 1, 1)$ and V the asymptotic covariance matrix of \mathbf{t} . Because V is unknown, in expression (10) we replace V by a consistent estimator \hat{V} of V . The resulting estimator is no longer a function of moments up to order three only, but is asymptotically equivalent to $\hat{\beta}^{opt}$ given by (10).

REMARK 1. If ξ is also symmetrically distributed then $\mu_{30} = \mu_{21} = \mu_{12} = \mu_{03} = 0$. In that special case consistent third-order-estimators of β do not exist. Then we could instead of deriving an optimal third-order-estimator of β , deduce an optimal fourth-order-estimator of β . This estimator has asymptotically minimal variance in the class of consistent β -estimators which are functions of the moments up to order four. In this paper we do not pay attention to this symmetrical case.

REMARK 2. In this paper we assume the errors ϵ and δ to be independent of each other. However, we can relax this assumption. See Appendix 1.

7. THE ASYMPTOTIC VARIANCE OF $\hat{\beta}^{opt}$

The asymptotic variance of $\hat{\beta}^{opt}$ has the form $(u' V^{-1} u)^{-1}$. If we have expressions for the asymptotic covariances and variances of $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ then we are able to compute the asymptotic variance of $\hat{\beta}^{opt}$. These variances and covariances can be derived by using the delta-method. See also e.g. KENDALL and STUART, vol I, chapter 10 (1963) for a derivation of variances and covariances of ratio's of random variables. Below we give the asymptotic variance of $\hat{\beta}_1$ and the asymptotic covariance of $\hat{\beta}_1$ and $\hat{\beta}_2$. The other asymptotic vari-

ances and covariances are analogous.

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= (1/\mu_{12}^2)(\text{var}(\mathbf{m}_{03}) - 2\beta\text{cov}(\mathbf{m}_{03}, \mathbf{m}_{12}) + \beta^2\text{var}(\mathbf{m}_{12})) \\ \text{cov}(\hat{\beta}_1, \hat{\beta}_2) &= -(1/\mu_{21}^2)\text{cov}(\mathbf{m}_{03}, \mathbf{m}_{21}) + (\beta/\mu_{21}^2)\text{cov}(\mathbf{m}_{12}, \mathbf{m}_{21}) \\ &\quad - (\beta/\mu_{21}\mu_{12})\text{var}(\mathbf{m}_{12}) + (1/\mu_{21}\mu_{12})\text{cov}(\mathbf{m}_{03}, \mathbf{m}_{12}). \end{aligned}$$

We can estimate the variances and covariances of $\mathbf{m}_{30}, \mathbf{m}_{21}, \mathbf{m}_{12}$ and \mathbf{m}_{03} in the sample consistently with formulas given in Kendall and Stuart, Vol. 1. So we have consistent estimators of the asymptotic covariances and variances of $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$, and we can estimate the asymptotic variance of $\hat{\beta}^{opt}$ consistently.

Instead of using sample statistics for estimating the asymptotic variances and covariances of the three $\hat{\beta}$'s, we also can express them in terms of the model parameters. This makes it more easy to investigate special cases. For a detailed discussion and (awkward) formulas we refer to VAN MONTFORT (1986). Some of his asymptotic results will be given here:

If $\delta=0$ or $\beta \cong 0$ then $\hat{\beta}^{opt}$ and $\hat{\beta}_3$ are identical.

If $\epsilon=0$ or $\beta \cong \infty$ then $\hat{\beta}^{opt}$ and $\hat{\beta}_1$ are identical.

8. EXAMPLE WITH GENERATED DATA

In this example scores on ϵ, δ and ξ are drawn independently from a standard normal distribution for ϵ and δ and from a chi-square distribution with one degree of freedom for ξ . β is set equal to 1. From the x and y scores the parameter β is estimated by the method discussed before. The number of replications in this study is set equal to 1000. We give the results for two different sample sizes: $N=50$ and $N=200$. In table 1 the estimates $\hat{\beta}_0$ (the ordinary least squares estimate, i.e. assuming $\omega_2=0$), $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}^{opt}$ and the Pal-estimates $\hat{\beta}_4 = (\mathbf{m}_{03}/\mathbf{m}_{30})^{1/3}$, $\hat{\beta}_5 = (\mathbf{m}_{03}/\mathbf{m}_{21})^{1/2}$ and $\hat{\beta}_6 = (\mathbf{m}_{12}/\mathbf{m}_{30})^{1/2}$ are given.

TABLE 1.
Different estimates of β

parameter	Value	Theor. Asympt. Stand.Dev. (200)	Means (50)	Stand.dev. (50)	Means (200)	Stand.dev. (200)
β_0	.67	.076	.639	.151	.656	.077
β_1	1	.106	1.004	.570	.998	.114
β_2	1	.098	1.116	1.426	1.002	.108
β_3	1	.106	1.323	4.775	1.009	.125
β^{opt}	1	.097	.996	.294	.995	.106
β_4	1	.097	1.064	.383	1.002	.103
β_5	1	.099	1.043	.481	0.999	.104
β_6	1	.099	1.123	.594	1.009	.110
μ_{20}	3	.566	2.941	1.123	2.979	.565
μ_{11}	2	.552	1.966	1.099	1.980	.558
μ_{03}	8	4.186	7.482	10.197	7.705	5.864
μ_{12}	8	4.124	7.439	9.810	7.710	5.784
μ_{21}	8	4.143	7.385	9.675	7.739	5.800
μ_{30}	8	4.675	7.374	9.778	7.791	5.909

DISCUSSION

From Table 1 we see that the estimates of β behave reasonably well, even for a sample size of 50. We see that $\hat{\beta}_0$ is biased, as it should be. Further we see that the six estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$ and $\hat{\beta}_6$ are different with different standard deviations. For a sample size of 50 the estimate $\hat{\beta}^{opt}$ is better than the other six estimates, because it has smaller standard deviation.

For the sample size of 200 we see that all the third-order-estimates behave very well. Their standard deviations are near to 0.100. Further the standard deviation of the estimate $\hat{\beta}^{opt}$ is smaller than the standard deviations of the estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\hat{\beta}_6$, and bigger than the standard deviations of the estimates $\hat{\beta}_4$ and $\hat{\beta}_5$. From the second and the sixth column of Table 1 follows that for $N = 200$ our third-order-estimates behave not asymptotically yet.

The same study was done for a sample size of 25. In this case our third-order-estimates were very bad: negative estimates of β were frequently found and the standard deviations of the estimates were extremely large, e.g. larger than 20. Often the estimates $\hat{\beta}_5$ and $\hat{\beta}_6$ were not defined because the sample-estimates of μ_{03}/μ_{21} and μ_{12}/μ_{30} had negative values.

Also we estimated β by an IV-estimator, with instrumental variable $z = \cos \xi$. For the sample sizes of 25, 50 and 200 the means of the estimates of β were 1.108, 1.035 and 1.006, and the corresponding standard deviations were 0.876, 0.245 and 0.106. However, in practice there is no obvious procedure to find

instrumental variables.

The conclusion from this small study is that our procedure of estimating β behaves very well, in particular for samples with size larger than 50.

APPENDIX 1

In this paper we assume the errors ϵ and δ to be independent of each other. However, we can relax this assumption. We can postulate:

$$\epsilon = \lambda\delta + \omega, \quad (A1)$$

where λ is a scalar, δ and ω are independent and ω is symmetrically distributed. $\lambda=0$ corresponds with the case that ϵ and δ are independent.

Replacing the independence of ϵ and δ by (A1), $\hat{\beta}^{opt}$ is still optimal for all consistent estimators of β which are functions of the moments up to order three. Further the asymptotic properties of $\hat{\beta}^{opt}$ don't change.

APPENDIX 2

In this Appendix we derive some general results on best asymptotic normal (or BAN) estimation from which the specific results in the text are easy consequences. We deal with the situation in which we have a sequence \mathbf{x}_n of random m -dimensional vectors, which are asymptotically normal in the sense that there exist a vector μ_0 and a positive semidefinite matrix Σ_0 such that $n^{\frac{1}{2}}(\mathbf{x}_n - \mu_0)$ converges in law to $N(0, \Sigma_0)$. In general μ_0 and Σ_0 are unknown. The problem that interests us is the optimal estimation μ_0 .

The problem becomes interesting if we have prior information of the form $\mu_0 \in \Omega$, with Ω a differentiable manifold of dimension p . In fact we shall assume that Ω is an open submanifold of R^p . This means that $\Omega = \eta(\Theta)$, with Θ an open subset of some R^p , and with η differentiable (η is defined on R^p). We suppose there is a unique θ_0 in Θ such that $\mu_0 = \eta(\theta_0)$. We study estimators of the form $\Phi(x)$, mapping R^m into Θ , where Φ is differentiable, and Fisher-consistent for θ , which means that $\Phi(\eta(\theta)) = \theta$ for all $\theta_0 \in \Theta$. Under these conditions it follows immediately from the general delta method type of argument (RAO, 1973, section 6a.2) that $n^{\frac{1}{2}}(\Phi(\mathbf{x}_n) - \theta_0)$ is asymptotically $N(0, G_0' \Sigma_0 G_0)$, where G_0 is $\delta\Phi/\delta x$ evaluated at μ_0 . But, more importantly, it also follows that $G_0 \Sigma_0 G_0' \geq (H_0' \Sigma_0^{-1} H_0)^{-1}$, where H_0 is the matrix of partial derivatives $\delta\eta/\delta\theta$ at θ_0 . This is a general result on BAN-estimation, given for example by WIJSMAN (1959a, 1959b). Estimates for which the above inequality is satisfied as an equality are called BAN. If η is one-to-one, then Fisher-consistency already implies that $G_0 = H_0^{-1}$, and all Fisher-consistent estimates are automatically BAN.

Now let us transform \mathbf{x}_n by the one-to-one differentiable transformation Γ to $\mathbf{y}_n = \Gamma(\mathbf{x}_n)$. Consider estimates of the form $\Psi(y)$, which are Fisher-consistent for θ in the sense that $\Psi(\Gamma(\eta(\theta))) = \theta$ for all θ in Θ . Suppose $T_0 = \delta\Gamma/\delta x$, evaluated in μ_0 . It follows that $n^{\frac{1}{2}}(\Psi(\mathbf{y}_n) - \theta_0)$ is asymptotically $N(0, W_0)$, with $W_0 \geq (H_0' T_0' (T_0 \Sigma_0 T_0')^{-1} T_0 H_0)^{-1} = (H_0' \Sigma_0^{-1} H_0)^{-1}$. The lower bound on the variance thus remains the same. A BAN estimate $\Psi(\mathbf{y}_n)$, for the model $\Gamma(\eta(\Theta))$, is at the same time a BAN estimate $\Phi(\mathbf{x}_n)$, with $\Phi = \Psi(\Gamma)$, for the model $\eta(\Theta)$.

Let us now apply the general results discussed above to the example treated in this paper. Take μ_0 equal to the nine moments of order less than or equal to three, \mathbf{x}_n equal to the corresponding nine sample moments \mathbf{m} , θ_0 equal to the vector with nine parameters, and define η by equations (2) and (6). It follows that $\eta^{-1}(\mathbf{m})$ is BAN for θ_0 , and these are exactly the estimates computed by using sample moments in (7). This implies that $\mathbf{m}_{12}/\mathbf{m}_{21}$ is BAN for β .

The second result again uses the one-to-one transformation η^{-1} . We have seen that the model is that the two coordinates (7f) and (7h) of $\eta^{-1}(\mu_0)$, corresponding with ω_3 and θ_3 , are zero. This result can be used quite simply to derive BAN estimates of the remaining parameters. Suppose \mathbf{y}_1 estimates the seven nonzero parameters of the system, and \mathbf{y}_2 estimates the two parameters (7f) and (7h) which are supposed to be zero. Partition the dispersion matrix

W_0 of $\eta^{-1}(\mathbf{m})$, and its inverse, correspondingly. Then $\mathbf{y}_1 + (\mathbf{W}_{11})^{-1} \mathbf{W}_{12} \mathbf{y}_2$, where \mathbf{W} estimates W_0 consistently, is BAN for the remaining seven parameters. This follows from the general minimum chi square theory for computing BAN estimates discussed by FERGUSON (1958), CHIANG (1952), and WIJSMAN (1959).

A third method in the symmetric error case transforms (6) to $\lambda_3 = \mu_{30}$ together with the three determinations of β in (8). This defines the one-to-one transformation Γ , and now we require that three coordinates of Γ are equal. As in VAN DER POL and DE LEEUW (1987) it follows that the linear combination, with coefficients that add up to one, of the three estimates of β given by (9), which has the smallest possible variance among such combinations, gives the BAN estimate of β . This is (10). In general the two BAN estimates we discussed for the symmetric error case will be different, because the one-one transformations $\bar{\Gamma}$ and η^{-1} are not related linearly.

REFERENCES

- ADCOCK, R.J. (1878), A problem in least squares, *The Analyst* 5, 53-54.
- AIGNER, D.J., C. HSIAO, A. KAPTEYN and T. WANSBEEK (1984), Latent variable models in econometrics, in: Z. Griliches and M.D. Intrilligator (Eds.), *Handbook of Econometrics*, North Holland P.C., Amsterdam.
- ANDERSON, T.W. (1984), Estimating linear statistical relationships, *Annals of Statistics* 12, 1-45.
- BEKKER, P.A. (1986a), Comment on identification in the linear errors in variables model, *Econometrica* 54, 215-217.
- BEKKER, P.A. (1986b), *Essays on identification in linear models with latent variables*, Dissertation, Wibo, Helmond.
- BEKKER, P.A., T. WANSBEEK and A. KAPTEYN (1985), Errors in variables in econometrics: new developments and recurrent themes, *Statistica Neerlandica* 39, 129-141.
- BEKKER, P.A. and J. DE LEEUW (1986), Rank of reduced dispersion matrices, *Psychometrika*, in press.
- CHIANG, C.L. (1952), On regular best asymptotically normal estimates with an application to a stochastic process, Doctoral Dissertation, University of California.
- DEISTLER, M. (1986), Linear errors-in-variables systems, to appear in: S. Bitanti (editor), *Springer Lecture Notes*, Berlin.
- FERGUSON, T.S. (1958), A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities, *Annals of Mathematical Statistics* 29, 1046-1062.
- FRISCH, R. (1934), Statistical confluence analysis by means of complete regression systems, Publication No. 5, University of Oslo, Economic Institute.
- GEARY, R.C. (1942), Inherent relations between random variables, *Proceedings of the Royal Irish Academy, Sec. A* 47, 63-76.
- GEARY, R.C. (1943), Relations between statistics: the general and the

- sampling problem when the samples are large, *Proceedings of the Royal Irish Academy, Sec. A* 49, 177-196.
- GINI, C. (1921), Sull'interpolazione di una retta quando i valori della variable indipendente sono affetti da errori accidentali, *Metron* 1, 63-82.
- KALMAN, R.E. (1982), System identification from noisy data, in: A. Bednarek and L. Cesari (Eds.), *Dynamical System II*, a University of Florida International Symposium, Academic press, New York.
- KALMAN, R.E. (1983), Identifiability and modeling in Econometrics, in: P.R. Krishnaiah (Ed.), *Developments in Statistics*, Vol. 4, Academic Press, New York.
- KENDALL, M.G. and A. STUART (1963), *The Advanced Theory of Statistics*, Vol. 1, Griffin, London.
- KENDALL, M.G. and A. STUART (1967), *The Advanced Theory of Statistics*, Vol. 2, Griffin, London.
- KLEPPER, S. and E. LEAMER (1984), Consistent sets of estimates for regressions with errors in all variables, *Econometrica* 52, 163-183.
- KUMMEL, C.H. (1879), Reduction of observed equation which contain more than one observed quantity, *The Analyst*, 6, 97-105.
- MADANSKY, A. (1959), The fitting of straight lines when both variables are subject to error, *Journal of the American Statistical Association* 54, 173-205.
- MOOIJAART, A. (1985), Factor analysis for non-normal variables, *Psychometrika* 50, 323-342.
- MORAN, P.A.P. (1971), Estimating structural and functional relationships, *Journal of Multivariable Analysis* 1, 232-255.
- PAKES, A. (1982), On the asymptotic bias fo the Wald-type estimator of a straight line when both variables are subject to error, *International Economic Review* 23, 491-497.
- PAL, M. (1980), Consistent moment estimators of regression coefficients in the presence of errors in variables, *Journal of Econometrics* 14, 349-364.
- PAL, M. (1981), Estimation in errors in variables models, Unpublished dissertation, Indian Statistical Institute.
- PEARSON, K. (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine and Journal of Science* 2, 559-572.
- REIERSØL, O. (1950), Identifiability of a linear relation between variables which are subject to error, *Econometrica* 18, 375-389.
- RAO, C.R. (1973), *Linear statistical inference and its application*, Wiley, New York.
- SARGAN, J.D. (1958), The estimation of economic relationships using instrumental variables, *Econometrica* 26, 393-415.
- SCHNEEWEISS, H. (1976), Consistent estimation of a regression with errors in the variables, *Metrika* 23, 101-115.
- VAN DER POL, J. and J. DE LEEUW (1987), Aspects of estimation of the polychoric correlation coefficient, *Psychometrika*, in press.
- VAN MONTFORT, K. (1986), Third order moments in regression with errors in variables, Research Note, Department of Psychology, University of Leiden.
- WIJSMAN, R.A. (1959a), On the theory of BAN estimates, *Annals of*

Mathematical Statistic 30, 185-191.

WIJSMAN, R.A. (1959b), Correction to "On the theory of BAN estimates",
Annals of Mathematical Statistics 31, 1269-1270.

Received November 1986, Revised March 1987

