SIMEX and TLS: An equivalence result

Polzehl, Jörg
Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin, Germany
polzehl@wias-berlin.de

Zwanzig, Silvelyn Uppsala University, Department of Mathematics,
Box 480, SE-751 06 Uppsala, Sweden
20@math.uu.se

Keywords: Errors-in-variables, SIMEX, Moment estimator, Total Least Squares

AMS 2000 Subject Classification: 62F12, 62J05

Abstract

SIMEX was introduced by Cook and Stefanski [6] as a simulation type estimator in errors-in-variables models. The idea of the SIMEX procedure is to compensate for the effect of the measurement errors while still using naive regression estimators. Polzehl and Zwanzig [13] defined a symmetrized version of this estimator. In this paper we establish some results relating these two simulation-extrapolation-type estimators to well known consistent estimators like the total least squares estimator (TLS) and the moment estimator (MME) in the context of errors-in-variables models. We further introduce an adaptive SIMEX (ASIMEX), which is calculated like SIMEX, but based on an estimated variance. The main result of this paper is that SYMEX, ASIMEX are equivalent to TLS. Additionally we see that SIMEX is equivalent to the moment estimator.

1 Introduction

In [6] Cook and Stefanski introduced a new estimator in errors-in-variables models, see also [11]. This estimator attained broad interest in the statistical literature and has been extended for use in a wide range of measurement error models, see e.g. [3], [1], [4], [5], [16], [10], [8], [12], [9], [15] and [2]. The idea behind this simulation-extrapolation estimator (SIMEX) is to compensate for the effect of the measurement errors while still using naive regression estimators. This is achieved by studying the naive regression estimate as a function of error variance in the predictor variables by simulation. Extrapolation to the unobservable case of zero measurement variance is then used to construct the SIMEX estimator.

In Polzehl and Zwanzig [13] a symmetrized modification (SYMEX) of this idea is proposed for the case of a multiple linear errors-in-variables model.

In this paper we establish some results relating both SIMEX and SYMEX to well known consistent estimators like the moment estimator (MME) and the total least squares estimator (TLS) in the context of linear functional relations.

We further introduce an adaptive SIMEX (ASIMEX), which is calculated like SIMEX, but based on an estimated variance.

Our main result is that SYMEX, ASIMEX are equivalent to TLS. Additionally we see that SIMEX is equivalent to the moment estimator. Both results essentially depend on the use of an appropriate model for extrapolation. We explicitely derive this model, see (3.1).

The comparison with TLS and the moment estimator implies, in the linear case, the consistency of all SIMEX estimators using (3.1) for $n \to \infty$ and fixed variances.

The paper is organized as follows. Section 2 introduces the model and the moment and total least squares estimators. In Section 3 we explain the construction of simulation-extrapolation-type estimators. Theoretical results are established in Section 4 while Section 5 is used to illustrate the results by a small simulation study. Proofs of the theorems in Section 4 are deferred to an Appendix.

2 The model and estimators

Suppose that we have observations:

$$\mathbf{Z} = \zeta + \varepsilon \tag{2.1}$$

where

$$\mathbf{Z} = (Z_{il})_{i=1...n, l=0...p}, \zeta = (\zeta_{il})_{i=1...n, l=0...p}, \;\; \varepsilon = (\varepsilon_{il})_{i=1...n, l=0...p}.$$

The measurement errors ε_{il} are *i.i.d.* with

$$E\varepsilon_{il} = 0\,,\; Var(\varepsilon_{il}) = \sigma^2. \tag{2.2}$$

Further we suppose a linear relationship between the expectations of $\mathbf{Z}$

$$0 = \zeta\alpha, \quad \alpha = (\alpha_0, ..., \alpha_p)\,,\; \min_l |\alpha_l| > 0 \tag{2.3}$$

where the parameter α consists of one known and p unknown components. With out loss of generality we set -1 for the known component of α. We require

$$M_{\zeta\zeta} = \frac{1}{n}\zeta^T\zeta \text{ is positive definite for } n > n_0. \tag{2.4}$$

There are $p+1$ different variants for the parameterization in (2.3), depending which of the components is -1. Set

$$\alpha^{(0)} = (-1, \beta_1, ..., \beta_p)^T \tag{2.5}$$

and for $l = 1, ..., p$

$$\alpha^{(l)} = \left(\alpha_0^{(l)}, ..., \alpha_{l-1}^{(l)}, -1, \alpha_{l+1}^{(l)}, ..., \alpha_p^{(l)}\right)^T, \ \alpha^{(l)} = -\frac{1}{\beta_l}\alpha^{(0)}. \tag{2.6}$$

The parameter of interest is

$$\beta = (\beta_1, ..., \beta_p)^T.$$

Note each parameterization in (2.5) and (2.6) corresponds to an explicite errors-in-variables model, where $\mathbf{Z}_l$, the $l'th$ column of $\mathbf{Z}$, is the "response" variable and the other rows of $\mathbf{Z}$ correspond to the "predictor" variables, which are observed with measurement errors. In the same way we can also consider naive regression models

$$\mathbf{Z}_l = \mathbf{Z}_{(-l)}\alpha_{(-l)}^l + \varepsilon_l, \quad \text{for}\, l = 0, \ldots, p, \tag{2.7}$$

where $\alpha_{(-l)}^l$ is the parameter vector formed from $\alpha^{(l)}$ in (2.6) by deleting $\alpha_l^{(l)} = -1$ and where $\mathbf{Z}_{(-l)}$ is the submatrix formed from $\mathbf{Z}$ by deleting the column $\mathbf{Z}_l$. In (2.7) the $l'th$ variable is considered as "response" and the errors in the other "predictor" variables are ignored.

Consider the data matrix:

$$\underset{(p+1)\times(p+1)}{M_{ZZ}} = \frac{1}{n}\mathbf{Z}^T\mathbf{Z} \tag{2.8}$$

The notation M_{ZZ} coincides with Fuller (1986) , [7]. We can compose the data matrix M_{ZZ} by the different adjunct matrices

$$\underset{p\times p}{M_{(-l)(-l)}} = \mathbf{Z}_{(-l)}^T\mathbf{Z}_{(-l)} \tag{2.9}$$

$$\underset{1\times 1}{M_{ll}} = \mathbf{Z}_l^T\mathbf{Z}_l, \ \underset{p\times 1}{M_{(-l)l}} = \mathbf{Z}_{(-l)}^T\mathbf{Z}_l \ \ \underset{1\times p}{M_{l(-l)}} = \mathbf{Z}_l^T\mathbf{Z}_{(-l)}. \tag{2.10}$$

For $l = 0$ we get, again using the notation of Fuller (1986) , [7],

$$M_{00} = M_{YY}, \ M_{(-0)(-0)} = M_{XX}, \ M_{0(-0)} = M_{YX}, \ M_{(-0)0} = M_{XY}. \tag{2.11}$$

The naive least squares estimator $\widehat{\alpha}^{(l)}$ related to the "wrong" regression model in (2.7) is given by

$$\widehat{\alpha}_{(-l)}^l = M_{(-l)(-l)}^{-1}M_{(-l)l}, \ \ \widehat{\alpha}_l^l = -1. \tag{2.12}$$

Thus we get $(p+1)$ different naive estimators for β

$$\widehat{\beta}_{naive,l} = \frac{\widehat{\alpha}_{(-0)}^l}{\widehat{\alpha}_0^l}, \quad l = 0, ..., p. \tag{2.13}$$

Note that for $l = 0$ we have the "traditional" naive least squares estimate

$$\widehat{\beta}_{naive,0} = M_{XX}^{-1} M_{XY}. \tag{2.14}$$

The total least squares (TLS) estimator $\widehat{\beta}_{TLS}$ is the maximum likelihood estimator under Gaussian error distribution in (2.2). Following [7] we get

$$\widehat{\beta}_{TLS} = \left(M_{XX} - \lambda_{\min}\left(M_{ZZ}\right) I\right)^{-1} M_{XY} \tag{2.15}$$

for the TLS estimator,

$$\widehat{\sigma}^2 = \frac{n}{n-p} \lambda_{\min}\left(M_{ZZ}\right) \tag{2.16}$$

for a variance estimator and

$$\widehat{\beta}_{MME} = \left(M_{XX} - \sigma^2 I\right)^{-1} M_{XY} \tag{2.17}$$

for the moment estimator.

3 SIMEX Methods

Each SIMEX procedure consists of three main steps:

1. **Simulation:** Generate new samples with successive higher measurement error variances by adding pseudo errors to the original observations.

2. **Model fit:** Adapt a parametric model which describes the relationship between estimates and error variances.

3. **Extrapolation:** Determine the length of the backwards step relating to a theoretic sample with measurement error variance zero.

In the original paper of Cook Stefanski the SIMEX procedure is introduced for every component of the parameter β_l separately. In step 1 they add pseudo errors only on the corresponding variable X_l. In step 2 they consider a linear, a quadratic and a nonlinear model. For justifying the backwards step in 3 they require the knowledge of the variance of the measurement error in X_l.

In difference to this we propose a simulatenous estimation of all parameters components in all SIMEX methods.

3.1 SIMEX and ASIMEX

We perform the steps 1-3 for improving the naive estimator $\widehat{\beta}_{naive,0}$ given (2.14) related to model (2.7) with $l = 0$.

1. **Simulation**

   This step is carried out for every λ out of a set $\{\lambda_k; k = 0, ..., N_s\}$, with $\lambda_0 = 0$.

   (a) Generate independent pseudo measurement errors $\sqrt{\lambda}\varepsilon^*_{b,i,j}$ with $\varepsilon^*_{b,i,j} \sim N(0,1)$, $b = 1, ..., B$, $i = 1, ..., n$, $j = 1, ..., p$, and add them to X_{ij} :

   $$X_{b,i,j}(\lambda) = X_{ij} + \lambda^{1/2}\varepsilon^*_{b,i,j}$$

   (b) Define the new observation matrices $\mathbf{Z}_{bj}(\lambda)$ using $X_{b,i,j}(\lambda)$ instead of X_{ij} . Thus the new data matrix is

   $$M_{ZZ}(\lambda) = \frac{1}{n\ B}\sum_{i,b} Z_{ib}(\lambda) Z_{ib}(\lambda)^t$$

   and calculate the naive estimator $\widehat{\beta}_{naive,0}(\lambda) = M^{-1}_{XX}(\lambda)\, M_{XY}(\lambda)$.

2. **Model fit**

   Define the parametric model that relates the expectation of the naive estimates to λ as

   $$\beta(\lambda, \theta) \ = (M_{XX} + \lambda I_p)^{-1}\theta, \tag{3.1}$$

   with parameter $\theta \in \mathbb{R}^p$. Fit this model by

   $$\widehat{\theta} = \arg\min \sum_{k=0}^{N_s} \left\| \widehat{\beta}_{naive,0}(\lambda_k) - \beta(\lambda_k, \theta) \right\|^2 . \tag{3.2}$$

3. **Extrapolation**

   The backwards step for the SIMEX estimator is given by the known measurement error variance σ^2 . Define the SIMEX estimator as

   $$\hat{\beta}_{SIMEX} = \beta(-\sigma^2, \widehat{\theta}). \tag{3.3}$$

   The ASIMEX estimator applies the variance estimator $\lambda_{\min}(M_{ZZ})$, see (2.16), instead of σ^2 :

   $$\hat{\beta}_{ASIMEX} = \beta(-\lambda_{\min}(M_{ZZ}), \widehat{\theta}). \tag{3.4}$$

3.2 SYMEX

The main idea is to apply the first two steps of SIMEX to all naive estimators given in (2.13). Then the backward step is determined by the λ^*, where all estimators deliver the same result.

1. **Simulation step**

   In contrast to he SIMEX procedure in subsection 3.1 the simulation step is carried out is carried out for all $(p+1)$ models in (2.7). Note, the pseudo errors are only added to the "predictor" variables in the respective model. Define the projection matrix P_l,

$$P_l = \begin{pmatrix} 0 & & & .0 \\ & 1 & & \\ & & \ddots & \\ 0.. & & & ..0 \end{pmatrix}. \tag{3.5}$$

   with all except the l'th diagonal element being zero.

   (a) For every $\lambda \in \{\lambda_k; k = 0, ..., K\}$ generate new samples by adding pseudo measurement errors

$$Z_{ib}(\lambda) = Z_i + (I - P_l)\lambda^{1/2}\varepsilon^*_{ib},\ i = 1, ..., n,\ b = 1, ..., B, \tag{3.6}$$

   with independent pseudo errors $\varepsilon^*_{ij} \sim P^* = N_{p+1}(0, I)$.

   (b) Calculate the $(p+1) \times (p+1)$ data matrix

$$M_{ZZ}(\lambda) = \frac{1}{n\ B}\sum_{i,b} Z_{ib}(\lambda)Z_{ib}(\lambda)^T \tag{3.7}$$

   and the naive estimators related to (2.12)

$$\widehat{\alpha}^l_{naive,(-l)}(\lambda) = M_{(-l)(l)}(\lambda)^{-1} M_{(-l)l}(\lambda),\ \widehat{\alpha}^l_{naive,l}(\lambda) = -1,$$

   where the decomposition of $M_{ZZ}(\lambda)$ into $M_{ll}(\lambda)$, $M_{(l)(-l)}(\lambda)$, $M_{(-l)(l)}(\lambda)$ and $M_{(-l)(-l)}(\lambda)$ is defined analogously to (2.9) and (2.10).

2. **Model fit**

   In difference to the SIMEX procedure we now have two tasks: model fit and reparametrization.

(a) Apply the model fit step of subsection 3.1 to all naive estimators. Introduce $\alpha^l(\lambda,\theta)$ by

$$\alpha^l_{(-l)}(\lambda,\theta) = \left(M_{(-l)(-l)} + \lambda I_p\right)^{-1}\theta \text{ and } \alpha^l_l(\lambda,\theta) = -1, \quad (3.8)$$

with parameter $\theta \in R^p$. Fit this model by

$$\widehat{\theta}_{(l)} = \arg\min \sum_{k=0}^{N_s} \left\| \widehat{\alpha}^l(\lambda_k) - \alpha^l(\lambda_k,\theta) \right\|^2 . \quad (3.9)$$

(b) Retransform the parameterization analogously to (2.13) and define

$$\widehat{\beta}^l(\lambda) = \frac{-1}{\alpha^l_0(\lambda,\widehat{\theta}_{(l)})}\alpha^l_{(-0)}(\lambda,\widehat{\theta}_{(l)}),\ l = 0,...,p. \quad (3.10)$$

3. **Extraploation**

Determine, for $\lambda_{min} = \min_l \lambda_{min}(M_{(-l)(-l)})$, an optimal value λ^* for the extrapolation step by

$$\lambda^* = \arg \min_{\lambda\in(-\lambda_{\min},0]} \sum_{l=0}^{p} \sum_{m=l+1}^{p} ||\widehat{\beta}^l(\lambda) - \widehat{\beta}^m(\lambda)||^2. \quad (3.11)$$

Define the SYMEX estimator by

$$\hat{\beta}_{SYMEX} = \frac{1}{p+1}\sum_{l=0}^{p}\widehat{\beta}^l(\lambda^*). \quad (3.12)$$

4 Theoretical Results

Let us summarize the model assumptions (2.1), (2.2), (2.3), (2.4), under **M.** The following condition **C** for the original observations **Z** is needed.
Condition C: The observations **Z** fulfill the condition **C** iff M_{ZZ} is positive definite,

$$\lambda_{\min}(M_{ZZ}) > 0$$

and all components of the total least squares estimator in (2.15) are not zero,

$$\hat{\beta}_{TLS,j} \neq 0\ , \text{ for all } j = 1,...,p.$$

Note that under **M** the total least squares estimator is consistent and the probability, that the observations fulfill the condition **C**, is tending to one for increasing sample size.

All theoretical results of this section are obtained, conditionally on the observations Z, with respect to the distribution of the pseudo errors P^*. We denote a sequence $r(Z(\lambda))$ (possibly vector valued) by $o_{P^*}(1)$ iff it converges in probability P^* for $B \to \infty$ to zero for all fixed observations $\mathbf{Z}$.

Theorem 4.1 *Under the assumptions* ***M*** *and* ***C*** *it holds that*

1. $\lambda^* = -\lambda_{\min}(M_{ZZ}) + o_{P^*}(1)$
2. *the estimators* $\widehat{\beta}_{SYMEX}$, $\widehat{\beta}_{ASIMEX}$, $\widehat{\beta}_{TLS}$ *are the same up to a term of order* $o_{P^*}(1)$.

Theorem 4.2 *Under the assumptions* ***M*** *and* ***C*** *it holds*

$$\widehat{\beta}_{SIMEX} = \widehat{\beta}_{MME} + o_{P^*}(1).$$

The proofs are deferred to the Appendix.

5 Numerical results

We illustrate the results from Theorem 4.1 and 4.2 using a small simulation. We generate data according to equations (2.3) and (2.5) with $p = 2$, $\beta = (2, .5)$ and $\zeta_{(-0)} \sim N_{p,50}(0_p, I_p)$. Errors in equation (2.1) are generated as i.i.d. Gaussian random variables, i.e. $\varepsilon_{il} \sim N(0, 0.5)$. Results are based on 250 simulated examples.

Figure 1 shows pairwise plots of the estimates of the second component of the parameter β obtained by SIMEX (3.3), MME (2.17), SYMEX (3.12), ASIMEX (3.4) and TLS (2.15). The upper triangular plots show estimates obtained with $B = 50$ while the plots below the diagonal reflect results for $B = 2$.

Figure 1 clearly illustrates the relations between SIMEX and MME and between ASIMEX, SYMEX and TLS. Additionally, due to its symmetric construction, SYMEX seems to provide a good approximation of TLS already for $B = 2$ while SIMEX requires larger values of B for a good approximation of MME. Note that ASIMEX uses extrapolation to approximate TLS while SYMEX can be viewed as an approximation to ASIMEX by estimating the length of the extrapolation

Figure 1: Pairwise plots of SIMEX, MME, SYMEX, ASIMEX and TLS estimates of $\beta_2 = .5$. The upper triangular plots show estimates obtained with $B = 50$ while the lower triangular plots provide results for $B = 2$

step λ^*. This is reflected by a slightly weaker correlation between SYMEX and TLS compared to the correlations between SYMEX and ASIMEX and between ASIMEX and TLS.

Figure 2 provides scatterplots of λ^*, obtained for $B = 2, 5$ and 50 against $-\lambda_{min}(M_{ZZ})$. The plots clearly indicate that λ^* is a simulative approximation to $-\lambda_{min}(M_{ZZ})$ with improving quality as B grows.

Table 1 provides results of our simulation in terms of Mean Squared Error (MSE) for the five estimates depending on the size B of the simulation experiment as observed in our simulations. We also give the mean squared distance of SIMEX, ASIMEX and SYMEX to both the MME and TLS estimates in order to illustrate

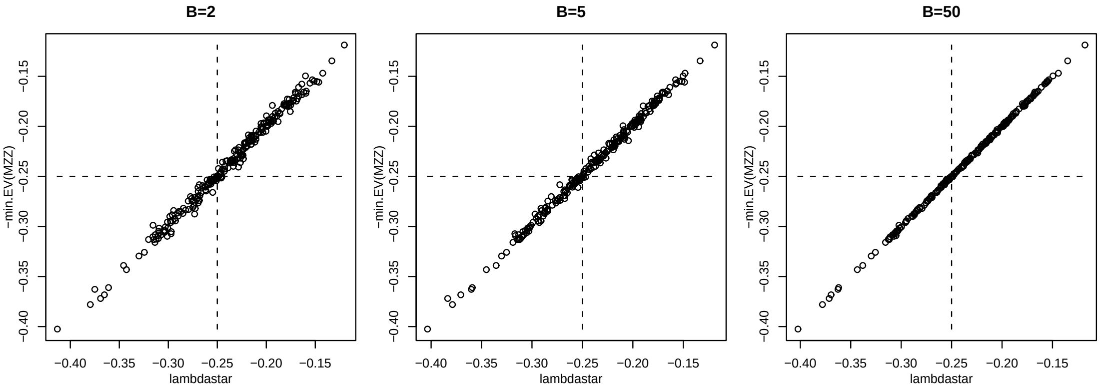

Figure 2: Scatterplots of $(\lambda^*, -\lambda_{min}(M_{ZZ}))$ ($B = 2, 5$ and 50 from left to right).

the effect of B. The results are in complete agreement with our Theorems 4.1 and 4.2.

Table 1: Mean Squared Error of MME, TLS, SIMEX, ASIMEX and SYMEX and mean squared difference of SIMEX, ASIMEX and SYMEX to MME and TLS

β_l	Criterion	B	MME	TLS	SIMEX	ASIMEX	SYMEX
$\beta_1 = 2$	MSE	2	0.0688	0.0398	0.0672	0.0398	0.0401
		5			0.0682	0.0399	0.0401
		20			0.0687	0.0398	0.0398
		50			0.0686	0.0397	0.0397
	$\|\hat\beta - \hat\beta_{MME}\|^2$	2	0	0.0180	2.0e-04	0.0185	0.0185
		5			5.3e-05	0.0181	0.0180
		20			1.0e-05	0.0182	0.0182
		50			3.4e-06	0.0181	0.0181
	$\|\hat\beta - \hat\beta_{TLS}\|^2$	2	0.0180	0	0.0173	9.4e-05	1.8e-04
		5			0.0179	4.2e-05	8.7e-05
		20			0.0181	8.3e-06	1.3e-05
		50			0.0180	2.8e-06	6.1e-06
$\beta_2 = 0.5$	MSE	2	0.0442	0.0390	0.0484	0.0406	0.0419
		5			0.0429	0.0380	0.0379
		20			0.0436	0.0384	0.0384
		50			0.0448	0.0391	0.0392
	$\|\hat\beta - \hat\beta_{MME}\|^2$	2	0	0.0022	0.0050	0.0028	0.0036
		5			1.4e-03	0.0027	0.0031
		20			3.7e-04	0.0023	0.0024
		50			1.1e-04	0.0021	0.0022
	$\|\hat\beta - \hat\beta_{TLS}\|^2$	2	0.00215	0	0.0066	6.9e-04	0.0015
		5			0.0032	3.0e-04	6.8e-04
		20			0.0025	8.6e-05	1.6e-04
		50			0.0024	2.8e-05	5.8e-05

6 Appendix

We now provide the proofs of Theorem 4.1 and 4.2. Similar results, obtained using a different argumentation can be found in [14].

First let us recall a general result for minimizing quadratic forms. Let $A = (A_{lk})_{l,k=0\ldots p}$ be a nonnegative definite $(p+1) \times (p+1)$ matrix composed by the different adjunct matrices

$$\underset{p\times p}{A_{(-l)(-l)}} = (A_{hk})_{h,k=0\ldots p,h\neq l,k\neq l}\,,\ \underset{1\times 1}{A_{ll}},$$
$$\underset{p\times 1}{A_{(-l)l}} = (A_{hl})_{h=0\ldots p,h\neq l}\,,\ \underset{1\times p}{A_{l(-l)}} = (A_{lk})_{k=0\ldots p,k\neq l}\,.$$

Further let $\gamma_{(0)},...,\gamma_{(p)}$ be the eigenvectors to the eigenvalues $\lambda_0,...,\lambda_p$ of A, where $\lambda_0 \leq \lambda_1 \leq ... \leq \lambda_p$. Under $\gamma_{(0)l} \neq 0$ for normalized quadratic forms it holds:

$$\widehat{x}_{norm} = \arg \min_{x,x_l=-1} \frac{x^T A x}{\|x\|^2} = -\frac{1}{\gamma_{(0)l}} \gamma_{(0)}.$$

Consider now the nonnormalized quadratic form.

$$\widehat{x} = \arg \min_{x,x_l=-1} x^T A x.$$

Lemma 6.1 *1. Assume $A_{(-l)(-l)}$ is positive definite. Then*

$$\widehat{x}_{(-l)} = A^{-1}_{(-l)(-l)} A_{(-l)l}.$$

2. Assume $\lambda_0 > 0$. Then

$$\widehat{x} = -\frac{1}{c_{ll}} \sum_{i=0}^{p} \frac{\gamma_{(i)l}}{\lambda_i} \gamma_{(i)}, \ \text{with } c_{ll} = \sum_{i=0}^{p} \frac{1}{\lambda_i} \left(\gamma_{(i)l}\right)^2$$

3. Assume $\lambda_0 = 0$ and $\gamma_{(0)l} \neq 0$. Then

$$\widehat{x} = -\frac{1}{\gamma_{(0)l}} \gamma_{(0)}.$$

Proof:

1. For all x with $x_l = -1$ we have

$$x^t A x = A_{ll} - 2A_{l(-l)} x_{(-l)} + x^T_{(-l)} A_{(-l)(-l)} x_{(-l)}.$$

Thus $\widehat{x}_{(-l)} = A^{-1}_{(-l)(-l)} A_{(-l)l}$.

2. Under $\lambda_0 > 0$ the inverse matrix A^{-1} exists. Denote the respective adjuncts of A^{-1} by $A^{(-l)(-l)},\ A^{(-l)l},\ A^{l(-l)},\ A^{ll}$. It holds

$$A^{(-l)l} = -A^{-1}_{(-l)(-l)}A_{(-l)l}A^{ll}.$$

Hence $\widehat{x}_{(-l)} = A^{-1}_{(-l)(-l)}A_{(-l)l} = -\frac{1}{A^{ll}}A^{(-l)l}$. Further

$$A^{(-l)l} = \sum_{i=0}^{p}\frac{\gamma_{(i)l}}{\lambda_i}\gamma_{(i)(-l)},\ A^{ll} = \sum_{i=0}^{p}\frac{1}{\lambda_i}\left(\gamma_{(i)l}\right)^2 = c_{ll},$$

so

$$\widehat{x}_{(-l)} = -\frac{1}{c_{ll}}\sum_{i=0}^{p}\frac{\gamma_{(i)l}}{\lambda_i}\gamma_{(i)(-l)},\ \ \widehat{x}_l = -1,$$

which gives the statement.

3. Under $\lambda_0 = 0$ we have for $\gamma_{(0)l} \neq 0$

$$0 \le \min_{x, x_l=-1} x^T A x \le \frac{1}{\gamma_{(0)l}}\gamma_{(0)}^T A \gamma_{(0)} = \lambda_0 = 0.$$

□

Applying Lemma 6.1 , if all necessary inverses exist, we get

$$\widehat{\alpha}^l_{naive,(-l)}\left(\lambda\right) = \arg\min_{\alpha_{(-l)},\alpha_l=-1}\alpha^t M_{ZZ}\left(\lambda\right)\alpha\ = M_{(-l)(-l)}\left(\lambda\right)^{-1}M_{(-l)l}\left(\lambda\right) \qquad (6.1)$$

$$\widehat{\beta}_{TLS} = \arg\min_{\alpha_{(-0)},\alpha_0=-1}\frac{\alpha^t M_{ZZ}\alpha}{\alpha^t\alpha} = -\frac{1}{\gamma_{(0)0}}\gamma_{(0)(-0)}. \qquad (6.2)$$

Now we study

$$\alpha^l_{(-l)}(\lambda) = \arg\min_{\alpha_{(-l)},\alpha_l=-1}\alpha^t\left(M_{ZZ} + \lambda I\right)\alpha\ = \left(M_{(-l)(-l)} + \lambda I_p\right)^{-1}M_{(-l)l} \qquad (6.3)$$

Lemma 6.2 *Under the assumptions* $\boldsymbol{M}$ *and* $\boldsymbol{C}$ *for all* $\lambda \in [0, C]$ *holds*

$$\widehat{\alpha}^l_{naive}\left(\lambda\right) = \alpha^l(\lambda) + o_{P^*}(1)\,.$$

Proof: First we show that

$$M_{ZZ}\left(\lambda\right) = M_{ZZ} + \lambda(I - P_l) + o_{P^*}(1). \qquad (6.4)$$

From (3.7) with (3.6) we obtain

$$\begin{aligned} M_{ZZ}(\lambda) &= \frac{1}{n\,B}\sum_{i,b}\left(Z_i+(I-P_l)\lambda^{1/2}\varepsilon^*_{ib}\right)\left(Z_i+(I-P_l)\lambda^{1/2}\varepsilon^*_{ib}\right)^t \\ &= M_{ZZ}+\lambda^{1/2}(I-P_l)\frac{1}{n}\sum_i\left(\overline{\varepsilon}^*_i Z_i^t+Z_i\overline{\varepsilon}^{*t}_i\right) \\ &\quad +\lambda(I-P_l)\frac{1}{n}\sum_i\left(\frac{1}{B}\sum_b\varepsilon^*_{ib}\varepsilon^{*t}_{ib}\right)(I-P_l). \end{aligned}$$

Because of

$$\overline{\varepsilon}^*_i=\frac{1}{B}\sum_b\varepsilon^*_{ib}=o_{P^*}(1)$$

and

$$\frac{1}{B}\sum_b\varepsilon^*_{ib}\varepsilon^{*t}_{ib}=I+o_{P^*}(1),$$

we get for idempotent $(I-P_l)$ the result (6.4). From (6.4) we obtain that

$$M_{(-l)(-l)}(\lambda)=M_{(-l)(-l)}+\lambda I_p+o_{P^*}(1)$$

and

$$M_{(-l)l}(\lambda)=M_{(-l)l}+o_{P^*}(1).$$

Under **C** for $\lambda\in[0,C]$ the matrix $M_{(-l)(-l)}+\lambda I_p$ is positive definite. Then for sufficient large B the matrix $M_{(-l)(-l)}(\lambda)$ is positive definite too. The statement follows from (6.1) and (6.3). $\square$

Lemma 6.3 *Under the assumptions **M**, **C** and for all $\lambda>-\lambda_{\min}(M_{(-l)(-l)})$ it holds that*

$$\alpha^l(\lambda,\widehat{\theta}_{(l)})=\alpha^l(\lambda)+o_{P^*}(1). \tag{6.5}$$

Proof: We have in (3.9):

$$\widehat{\theta}_{(l)}=\arg\min\sum_{k=0}^{N_s}\left\|\widehat{\alpha}^l(\lambda_k)-(M_{(-l)(-l)}+\lambda_k I_p)^{-1}\theta\right\|^2.$$

Because Lemma 6.2 and (6.3):

$$\begin{aligned} &\sum_{k=0}^{N_s}\left\|\widehat{\alpha}^l(\lambda_k)-(M_{(-l)(-l)}+\lambda_k I_p)^{-1}\theta\right\|^2 \\ =&\sum_{k=0}^{N_s}\left\|\alpha^l(\lambda_k)-(M_{(-l)(-l)}+\lambda_k I_p)^{-1}\theta+o_{P^*}(1)\right\|^2 \\ =&\left(\theta-M_{(-l)(l)}+o_{P^*}(1)\right)^t\sum_{k=0}^{N_s}(M_{(-l)(-l)}+\lambda_k I_p)^{-2}\left(\theta-M_{(-l)(l)}+o_{P^*}(1)\right). \end{aligned}$$

Hence

$$\widehat{\theta}_{(l)} = M_{(-l)(l)} + o_{P^*}(1)$$

and (6.5) follows from (3.8). □

Proof of Theorem 4.1:

First we show that $\lambda^* = -\lambda_{\min}(M_{ZZ}) + o_{P^*}(1)$. λ^* is defined in (3.11) as fixpoint, where all renormalized estimators are equal:

$$\frac{\alpha^l_{(-0)}(\lambda^*, \widehat{\theta}_{(l)})}{\alpha^l_0(\lambda^*, \widehat{\theta}_{(l)})} = \frac{\alpha^k_{(-0)}(\lambda^*, \widehat{\theta}_{(k)})}{\alpha^k_0(\lambda^*, \widehat{\theta}_{(k)})}, \quad l, k, = 0, ..., p$$

Use Lemma 6.3 we get

$$\frac{\alpha^l_{(-0)}(\lambda, \widehat{\theta}_{(l)})}{\alpha^l_0(\lambda, \widehat{\theta}_{(l)})} = \frac{\alpha^l_{(-0)}(\lambda)}{\alpha^l_0(\lambda)} + o_{P^*}(1). \tag{6.6}$$

Compare now the leading terms in (6.6).

First assume $\lambda_{\min}(M_{ZZ} + \lambda I) > 0$.

That is $\lambda_{\min}(M_{ZZ} + \lambda I) = \lambda_{\min}(M_{ZZ}) + \lambda$ and $\lambda > -\lambda_{\min}(M_{ZZ})$. From Lemma 6.1 follows that

$$\frac{\alpha^0(\lambda)}{\alpha^l_l(\lambda)} = -\frac{1}{c_{0l}} \sum_{i=0}^{p} \frac{\gamma_{(i)l}}{\lambda_i} \gamma_{(i)}, \quad c_{0l} = \sum_{i=0}^{p} \frac{\gamma_{(i)0}\gamma_{(i)l}}{\lambda_i},$$

where $\gamma_{(0)}, ..., \gamma_{(p)}$ are the eigenvectors to the eigenvalues $\lambda_0, ..., \lambda_p$ of $M_{ZZ} + \lambda I$. Then the fixpoint equation for the leading terms is

$$-\frac{1}{c_{0l}} \sum_{i=0}^{p} \frac{\gamma_{(i)l}}{\lambda_i} \gamma_{(i)} = -\frac{1}{c_{0k}} \sum_{i=0}^{p} \frac{\gamma_{(i)k}}{\lambda_i} \gamma_{(i)} \text{ for all } l, k = 0, ..., p. \tag{6.7}$$

That is

$$\frac{\gamma_{(i)l}}{c_{0l}} = \frac{\gamma_{(i)k}}{c_{0k}} \text{ for all } i, l, k = 0, ..., p \ .$$

Specifically we have

$$\sum_{l=0}^{p} \gamma^2_{(i)l} = \sum_{l=0}^{p} c^2_{0l} \left(\frac{\gamma_{(i)k}}{c_{0k}} \right)^2 \text{ for all } i, k = 0, ..., p.$$

The eigenvectors are orthonormal such that $\sum_{l=0}^{p} \gamma^2_{(i)l} = 1$ and therefore

$$\gamma_{(i)k} = \frac{c_{0k}}{\sqrt{\sum_{l=0}^{p} c^2_{0l}}} \text{ for all } i, k = 0, ..., p.$$

This implies a linear dependence between the eigenvectors, which is a contradiction. There is no solution of (6.7) for $\lambda > -\lambda_{\min}(M_{ZZ})$.

We now consider the case: $\lambda_{\min}(M_{ZZ}+\lambda I) = 0$. Because of **C** we get $\lambda_{\min}(M_{ZZ}) > 0$. Thus $\lambda_{\min}(M_{ZZ}+\lambda I) = \lambda_{\min}(M_{ZZ}) + \lambda = 0$, implies $\lambda = \lambda^{**} = -\lambda_{\min}(M_{ZZ})$. We conclude from Lemma 6.1 that

$$\frac{\alpha^l(\lambda^{**})}{\alpha_0^l(\lambda^{**})} = -\frac{1}{\gamma_{(0)0}}\gamma_{(0)} \tag{6.8}$$

where $\gamma_{(0)}$ is the eigenvector of $M_{ZZ}+\lambda^{**}I$ belonging to the zero eigenvalue. That means $\lambda^{**} = -\lambda_{\min}(M_{ZZ})$ is the solution of the fixpoint equation with $\lambda \in [-\lambda_{\min}(M_{ZZ}), 0]$. Hence

$$\lambda^* = -\lambda_{\min}(M_{ZZ}) + o_{P^*}(1). \tag{6.9}$$

Let us now compare the estimators: Comparing (6.3) with (2.15) it holds

$$\widehat{\beta}_{TLS} = \frac{\alpha^l_{(-0)}(\lambda^{**})}{\alpha_0^l(\lambda^{**})} \text{ for all } l = 0, ..., p.$$

Because of (3.4) and Lemma 6.2 we have

$$\begin{aligned}\widehat{\beta}_{ASIMEX} &= \frac{\alpha^0_{(-0)}(\lambda^{**}, \widehat{\theta}_{(l)})}{\alpha_0^0(\lambda^{**}, \widehat{\theta}_{(l)})} \\ &= \frac{\alpha^0_{(-0)}(\lambda^{**})}{\alpha_0^0(\lambda^{**})} + o_{P^*}(1) \\ &= \widehat{\beta}_{TLS} + o_{P^*}(1)\end{aligned}$$

From (3.12), (6.9) and Lemma 6.2 we obtain

$$\begin{aligned}\widehat{\beta}_{SYMEX} &= \frac{1}{p+1}\sum_{l=0}^{p}\frac{\alpha^l_{(-0)}(\lambda^{*}, \widehat{\theta}_{(l)})}{\alpha_0^l(\lambda^{*}, \widehat{\theta}_{(l)})} \\ &= \frac{1}{p+1}\sum_{l=0}^{p}\frac{\alpha^l_{(-0)}(\lambda^{*})}{\alpha_0^l(\lambda^{*})} + o_{P^*}(1) \\ &= \frac{1}{p+1}\sum_{l=0}^{p}\frac{\alpha^l_{(-0)}(\lambda^{**})}{\alpha_0^l(\lambda^{**})} + o_{P^*}(1) \\ &= \widehat{\beta}_{TLS} + o_{P^*}(1).\end{aligned}$$

Proof of Theorem 4.2:

The SIMEX estimator is defined in (3.8) by

$$\widehat{\beta}_{SIMEX} = \alpha^0_{(-0)}(-\sigma^2, \widehat{\theta}_{(0)}).$$

Thus by Lemma 6.3

$$\widehat{\beta}_{SIMEX} = \alpha^0_{(-0)}(-\sigma^2) + o_{P^*}(1),$$

where $\alpha^0_{(-0)}$ is given in (6.3). Using (2.17) and (2.11) for $l = 0$ we obtain

$$\alpha^0_{(-0)}(-\sigma^2) = \left(M_{XX} - \sigma^2 I\right)^{-1} M_{XY} = \widehat{\beta}_{MME}.$$

References

[1] Berry, S.M., Carroll, R.J. and Ruppert, D. (2002). Bayesian Smoothing and Regression Splines for Measurement Error Problems *Journal of the American Statistical Association*, **97**: 160-169.

[2] Carroll, R.J. and Hall, P. (2004). Low order approximations in deconvolution and regression with errors in variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**, 31

[3] Carroll, R.J., Ruppert,D. and Stefanski, L.A. (1996). *Measurement error in nonlinear models* , London: Chapman and Hall.

[4] Carroll, R.J., Küchenhoff, H., Lombard, F. and Stefanski, L.A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, **91**: 242-250.

[5] Carroll, R. J., Stefanski, L. A. (1997). Asymptotic theory for the Simex estimator in measurement error models (STMA V39 4250). *Advances in Statistical Decision Theory and Applications*, 151-164.

[6] Cook, J.R. and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, **89**: 1314-1328.

[7] Fuller, W.A. (1987). *Measurement error models* , New York: John Wiley.

[8] Holcomb, J.P. Jr. (1999). Regression with covariates and outcome calculated from a common set of variables measured with error: Estimation using the SIMEX method. *Statistics in Medicine*, **18**, 2847-2862.

[9] Li, Y. and Lin, X. (2003). Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association*, **98**: 191-203.

[10] Lin, X. and Carroll, R.J. (1999). SIMEX variance component tests in generalized linear mixed measurement error models. *Biometrics*, **55**, 613-619.

[11] Stefanski, L.A., Cook, J.R. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, **90**, 1247-1256.

[12] Lin, X. and Carroll, R.J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95**, 520-534.

[13] Polzehl, J. and Zwanzig S. (2003). On a symmetrized simulation extrapolation estimator in linear errors-in-variables models, *Computational Statistics and Data Analysis.*

[14] Polzehl, J. and Zwanzig S. (2003). On a comparison of different simulation extrapolation estimators in linear errors-in-variables models, Uppsala University, U.U.D.M. Report 2003:17.

[15] Staudenmayer, J. and Ruppert, D. (2004). Local polynomial regression and simulationextrapolation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 17-30.

[16] Wang, N., Lin, X., Gutierrez, R.G., Carroll, R.J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association*, **93**, 249-261.