Schneeweiss, Nittner:

Estimating A Polynomial Regression With Measurement Errors In The Structural And In The Functional Case - A Comparison

Sonderforschungsbereich 386, Paper 197 (2000)

ESTIMATING A POLYNOMIAL REGRESSION WITH MEASUREMENT ERRORS IN THE STRUCTURAL AND IN THE FUNCTIONAL CASE – A COMPARISON.

HANS SCHNEEWEISS AND THOMAS NITTNER
UNIVERSITY OF MUNICH

SUMMARY

Two methods of estimating the parameters of a polynomial regression with measurement errors in the regressor variable are compared to each other with respect to their relative efficiency and robustness. One of the two estimators (SLS) is valid for the structural variant of the model and uses the assumption that the true regressor variable is normally distributed, while the other one (ALS and also its small sample modification MALS) does not need any assumption on the regressor distribution. SLS turns out to react rather strongly on violations of the normality assumption as far as its bias is concerned but is quite robust with respect to its MSE. It is more efficient than ALS or MALS whenever the normality assumption holds true.

Keywords: Polynomial regression, measurement errors, efficiency, robustness

1. INTRODUCTION

It is well known that if a regressor variable of a linear regression is measured with errors, the ordinary least squares (OLS), or naive, estimator of the corresponding slope parameter will be biased, the bias usually being such that it will attenuate the true value of the slope parameter, Cheng and Van Ness (1999), Fuller (1987), Schneeweiss and Mittag (1986). Assuming that the error is a random variable with expectation zero and independent of the true regressor variable and that the value of its variance is known, one can use this known error variance to construct an adjusted least squares (ALS) estimator, which does not have any bias, but is in fact consistent. The construction principle involved is that of the corrected score function method, which can be applied not only to the linear but also to a large class of non-linear models, Nakamura (1990), Buonaccorsi (1996). In this paper it will be applied to the polynomial regression model, just as in Cheng and Schneeweiss (1998), see also Chan and Mak (1985), and Stefanski (1989).

The ALS estimator does not use any further information beyond the knowledge of the error variance. Suppose however that, even though the time regressor variable cannot be observed, its distribution were known, then this additional knowledge could be used to construct a possibly superior estimator by using another principle than that for the ALS. The idea is to start from a mean function in the true regressor variables as the original model, supplemented by a variance function, and to transform it into a mean function model in the observable regressor by taking conditional expectations. Again this principle can be applied to a large class of models including the polynomial regression model, Thamerus (1998), Carrol et al. (1995). As the case of a known regressor distribution corresponds to what is usually called the structural variant of a measurement error model, this new estimator will be denoted by the same name: a structural least squares estimator (SLS). It can be most easily constructed if a normal distribution is assumed for the regressor variable.

In such a case, SLS will presumably be better than ALS in the sense of having a smaller (asymptotic) variance, both estimators being consistent. However, if the normality assumption is not valid, the SLS estimator may loose its superiority. Indeed, the ALS estimator is more robust than the SLS estimator owing to the fact that it does not depend on any particular distribution of the regressor variable. In fact, the true regressor can even be thought of as being nonstochastic, i. e., just an unknown constant for each observation, a case which is called the functional variant of a measurement error model. Thus ALS is a method good for the functional variant but can also be used in the structural variant case, whereas SLS explicitly makes use of the distributional assumption of the structural variant of the measurement error model and does depend on the validity of this assumption.

In the present paper we want to compare these two estimation methods, firstly when the distribution of the true regressor variable is correctly specified as Gaussian and secondly when it is non-Gaussian but incorrectly assumed to be Gaussian. One may expect the ALS estimator not to be effected very much by the shape of the regressor distribution and thus it will behave similarly whether the distribution is correctly specified or not, but the SLS estimator will clearly depend on the correct specification of the regressor distribution. The question is to what extent does the SLS estimator react to a misspecification of the regressor distribution. When will its properties deteriorate so much that it will become inferior to the more robust ALS estimator?

The comparison will be done by way of a simulation study and will thus cover small sample properties of the estimators. As with small samples ALS does not behave very nicely, the estimates becoming very unstable, ALS has to be modified so that its small sample variations become more stable and it can be compared more easily with SLS, which apparently needs no modification. The modified method is called MALS in this paper. The idea of modification stems from Fuller (1987) for a linear model and has been adapted to the polynomial measurement error model by Cheng et al. (1998).

In a recent paper, Kuha and Temple (1999) carried out a similar study trying to answer the same question as in this paper. They do, however, not assume the error variance to be known –the usual assumption– but rather the "noise-to-signal-ratio". Asymptotically these two approaches do not differ, but in small samples there may be differences. Kuha and Temple also do not go beyond the quadratic model. On the other hand, they study some other estimation methods as well.

In the next section a brief exposition is given of the estimation methods, ALS and SLS, involved. Section 3 then describes the simulation study and presents its results. The final section has some concluding remarks.

2. ESTIMATION METHODS

2.1 Adjusted least squares (ALS and MALS)

The model that is investigated is a polynomial regression in a latent variable ξ that can only be measured with a measurement error δ.

$$y_i = \beta_0 + \beta_1\xi_i + \ldots + \beta_k\xi_i^k + \epsilon_i \quad (2.1)$$

$$x_i = \xi_i + \delta_i\,, \quad i = 1, \ldots, n \quad (2.2)$$

x being the observed regressor variable. We assume the errors (ϵ_i, δ_i) to be iid Gaussian, independent of the ξ_i's, with variances σ_ϵ^2 and σ_δ^2 and covariance $\sigma_{\epsilon\delta} = 0$. It is then possible to construct polynomials $t_r(x)$ of degree r such that $\mathrm{E}t_r(x_i) = \xi_i^r$. Let H_i be a $(k+1)\times(k+1)$ matrix with elements $(H_i)_{rs} = t_{r+s}(x_i)$, $r, s = 0, \ldots, k$ and h_i a $(k+1)\times 1$ vector with elements $(h_i)_r = y_i t_r(x_i)$, $r = 0, \ldots, k$, then the unmodified ALS estimator $\hat\beta_{ALS}$ of $\beta = (\beta_0, \ldots, \beta_k)'$ is given as the solution of

$$\overline{H}\hat\beta_{ALS} = \overline{h}\ , \quad (2.3)$$

where the bar denotes averages, e. g., $\overline{H} = \frac{1}{n}\sum H_i$; for details see Cheng and Schneeweiss (1998).

This estimator is consistent and asymptotically normal. For small samples, however, it can give rise to large estimation errors, at least occasionally, and in particular if the noise-to-signal-ratio $\sigma_\delta^2/\sigma_\xi^2$ is large, say, larger than 0.1. A modification of ALS is available which reduces the estimator's variance considerably without introducing any conceivable bias. For this, define the vector $t_i = (t_0(x_i), \ldots, t_k(x_i))'$ and the matrix $V_i = t_i t_i' - H_i$. Then the MALS estimator of β is given as the solution of

$$(\overline{tt'} - a\overline{V})\hat\beta_{MALS} = \overline{h}\ , \quad (2.4)$$

where

$$a = \begin{cases} 1 - \frac{\alpha}{n} & \text{if} \quad \rho > 1 + \frac{1}{n} \\ \rho\frac{n-\alpha}{n+1} & \text{if} \quad \rho \le 1 + \frac{1}{n} \end{cases}\ ,$$

ρ being the smallest positive root (which always exists) of

$$\det \left[\begin{pmatrix} \overline{y^2} & \overline{yt'} \\ \overline{ty} & \overline{tt'} \end{pmatrix} - \rho \begin{pmatrix} 0 & 0' \\ 0 & \overline{V} \end{pmatrix} \right] = 0 \ , \tag{2.5}$$

where $\alpha = k + 4$. The estimator is an adaptation of Fuller's (1987) small sample improvement of the parameter estimates in a linear model with measurement errors. For details see Cheng et al. (1998).

2.2 Structural least squares (SLS)

In order to introduce SLS assume $\xi_i \sim \text{i i d } N(\mu_\xi, \sigma_\xi^2)$, write the regression model (2.1) as a conditional mean-variance model (see Caroll et al. 1995):

$$\mathrm{E}(y \mid \xi) = \beta_0 + \beta_1 \xi + \ldots + \beta_k \xi^k \tag{2.6}$$

$$\mathrm{V}(y \mid \xi) = \sigma_\epsilon^2 \ , \tag{2.7}$$

and find a new mean-variance model in the observable variable x by taking conditional expectations given x:

$$\mathrm{E}(y \mid x) = \sum_{j=0}^{k} \beta_j \mu_j(x) \tag{2.8}$$

$$\mathrm{V}(y \mid x) = \sigma_\epsilon^2 + \sum_{j=1}^{k} \sum_{l=1}^{k} \beta_j \beta_l \{\mu_{j+l}(x) - \mu_j(x)\mu_l(x)\} \ , \tag{2.9}$$

where $\mu_r(x) = \mathrm{E}(\xi^r \mid x)$. The conditional moments are easily computed using the fact that the conditional distribution of ξ given x is $N(\mu(x), \tau^2)$ with

$$\mu(x) = \mu_x + (1 - \sigma_\delta^2/\sigma_x^2)(x - \mu_x) \tag{2.10}$$

$$\tau^2 = \sigma_\delta^2(1 - \sigma_\delta^2/\sigma_x^2) \ . \tag{2.11}$$

Let $\mu_r^* = \mathrm{E}[\{\xi - \mu(x)\}^r \mid x]$ be the r-th conditional central moment of ξ given x, then

$$\mu_r^* = \begin{cases} 0 & \text{if } \ r \text{ is odd} \\ 1 \cdot 3 \cdot 5 \cdots (r-1)\tau^r & \text{if } \ r \text{ is even} \end{cases}$$

and $\mu_r(x)$ is given by

$$\mu_r(x) = \sum_{j=0}^{r} \begin{pmatrix} r \\ j \end{pmatrix} \mu_j^* \mu(x)^{r-j} \ . \tag{2.12}$$

μ_x and σ_x^2 can be estimated by their empirical counterparts.

$$\hat{\mu}_x = \overline{x} \ , \quad \hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

If these are substituted for μ_x and σ_x^2 in (2.10) to (2.12), estimates of τ^2 and $\mu(x)$ and finally of $\mu_r(x)$ arise. Replacing the $\mu_r(x)$ in (2.8) and (2.9) by their estimates and substituting the observable values x_i for the variable x we finally get a mean-variance model for the observable data with mean and variance functions

$$\hat{\mathrm{E}}(y \mid x = x_i) = \sum_{j=0}^{k} \beta_j \hat{\mu}_j(x_i) \tag{2.13}$$

$$\hat{\mathrm{V}}(y \mid x = x_i) = \sigma_\epsilon^2 + \sum_{j=1}^{k}\sum_{l=1}^{k} \beta_j\beta_l\{\hat{\mu}_{j+l}(x_i) - \hat{\mu}_j(x_i)\hat{\mu}_l(x_i)\} \tag{2.14}$$

One can derive estimates for the β's for this model by an iteratively reweighted least squares method, where in each step, s, an estimate for σ_ϵ^2 has to be updated using the residuals of the previous step, $s-1$:

$$\hat{\sigma}_\epsilon^{2\ (s)} = (n-k-1)^{-1} \sum_{i=1}^{n} [y_i - \{\sum_{j=0}^{k} \hat{\beta}_j^{(s-1)} \hat{\mu}_j(x_i)\}]^2 \tag{2.15}$$

For details see Thamerus (1998) and for the general method Carroll et al. (1995). It might be mentioned that an approximate method exists, where $\mathrm{E}(y|x)$ is approximated by replacing ξ in (2.6) with $\mathrm{E}(\xi|x)$. This is the regression calibration method, Carrol et al. (1995), which however can only reduce the measurement error bias, not remove it. A more elaborate expanded regression calibration method is also available and is, in fact, used by Kuha and Temple (1999) in their simulation study. In the quadratic model it coincides with the method used here, but not in higher order polynomials, where it is only approximately unbiased.

3. SIMULATIONS

In order to compare the performance of ALS, both unmodified and modified, with that of SLS a simulation study was run. Several polynomial models were studied, which differed in the degree of the polynomial, $k = 2$ or 3, and in the distribution of ξ. The parameter values were $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = -0.5$, and (for $k = 3$) $\beta_3 = 0.5$. In all cases the error variances were taken to be $\sigma_\epsilon^2 = \sigma_\delta^2 = 0.1$. (A much smaller error variance $\sigma_\delta^2 = 0.01$ was also experimented with, but in this case the results of the various estimation methods did not differ very much. For larger error variances like $\sigma_\delta^2 = 0.5$ the results became rather unstable.) The sample size was fixed at $n = 201$. Three distributions for ξ were chosen: the Gaussian distribution $N(0, \frac{1}{3})$, the uniform distribution, where the ξ_i were fixed at the 201 equidistant points $-1 + \frac{i}{100}$, $i = 0, \ldots, 200$, and the exponential distribution $\mathrm{Exp}(\lambda)$ shifted to the left by the amount $\frac{1}{\lambda}$ with $\lambda = \sqrt{3}$. In all three cases $\mathrm{E}(\xi) = 0$ and $\mathrm{Var}(\xi) = \frac{1}{3}$. (Clearly for the uniform distribution, which here is taken to be nonstochastic, $\mathrm{E}(\xi)$ and $\mathrm{Var}(\xi)$

have to be replaced with $\overline{\xi}$ and $s_\xi^2 = \frac{1}{n}\sum(\xi_i - \overline{\xi})^2$, and the equality $s_\xi^2 = \frac{1}{3}$ holds only approximately).

For these, altogether 6, models artificial samples were generated and the three estimation methods ALS, MALS and SLS were applied. The simulations were run with $N = 100$ replications. Bias and standard error of the estimates were computed and can be compared between models and between estimation methods. The naive OLS estimator was also computed but is not reproduced here. It clearly showed a significant bias in almost all cases, as was to be expected.

In table 4.1 the simulation results are presented for $k = 2$ and in table 4.2 for $k = 3$. Note that the noise-to-signal ratio $\sigma_\delta^2/\text{Var}(\xi) = 0.3$ is rather high so that a noticeable bias for the naive estimator (not shown here) results. The SLS estimator is consistent if ξ is actually normally distributed. The other estimators (ALS and MALS) are always consistent, whatever the distribution of ξ. Nevertheless they may show some bias in small or medium sized samples, where $n = 201$ may be considered medium sized. It is for this reason that the bias is shown in Tables 4.1 and 4.2, even though it turns out to be rather small and often insignificant for all the consistent estimators.

From table 4.1 it is seen that in the quadratic case ALS and MALS estimators hardly differ at all so that one might think the small sample modification of ALS was not necessary. There are however, albeit rare, cases where the ALS estimate has an extremely high estimation error, which is then greatly reduced by the MALS method. Apparently such a case did not come up in the present simulation study. Nevertheless the MALS method should always be used, if only for precautionary reasons.

The necessity of using MALS instead of ALS is seen most clearly in Table 4.2. For the cubic regression, the MALS estimator has always a conspicuously smaller standard deviation. While the standard deviation of MALS is rather modest, that of ALS is often extremely large, rendering the ALS method almost useless in this case.

Let us now compare MALS to SLS. The standard deviations of the SLS estimators are always smaller than those of MALS, regardless of the distribution of ξ. However, if we consider the bias, it is seen that on the whole, though not always, SLS has a smaller bias than MALS if ξ is normally distributed, but a significantly higher bias if the distribution of ξ deviates from the normal one, the difference being most prominent in the case of the exponential distribution.

It should be noted that the values for the bias are only estimated values. A rough rule-of-thumb 95%-confidence interval for the true bias is given by $\hat{B}_{\hat{\beta}} \pm 2\hat{\sigma}_{\hat{\beta}}/\sqrt{N}$, where $\hat{B}_{\hat{\beta}}$ is the estimated bias of the estimated regression coefficient $\hat{\beta}$, as shown in the tables, $\hat{\sigma}_{\hat{\beta}}$ is the corresponding estimated standard deviation, and $N = 100$. Bias values with one asterisk differ from zero by $2\hat{\sigma}_{\hat{\beta}}/\sqrt{N}$ and with two asterisks by $3\hat{\sigma}_{\hat{\beta}}/\sqrt{N}$.

A simple and comprehensive measure of precision is the overall MSE which here is defined as the sum of the MSE's for $\hat{\beta}_0$ to $\hat{\beta}_k$, $k = 2$ or 3. This measure is shown in Table 4.3 for the six models.

It is seen that the SLS estimators have smaller overall MSE than the MALS estimators in models with a normal and uniform distribution of ξ. For the exponential distribution, however, the overall MSE of the SLS estimators is larger in the quadratic regression ($k = 2$), but still smaller in the cubic regression ($k = 3$), although only slightly so.

4. CONCLUSION

Several conclusions can be drawn from the results of this simulation experiment.

1. The estimators considered in this paper are rather stable and do not differ too much in the quadratic model. On the other hand, due to the high multicollinearity, all the estimators become rather unstable in the cubic case and differ considerably with regard to their variances.

2. In particular the ALS estimator, a simple adjustment of the naive estimator, although being consistent, has very bad small sample properties for the cubic regression. Here a modification of ALS, viz. MALS, greatly reduces the instability of the estimator giving rise to reasonable standard errors. In the quadratic model, ALS and MALS hardly differ. This changes, however, when σ_δ^2 increases, e. g., to 0.5 . Then ALS becomes unstable also for the quadratic case.

3. While MALS (just like ALS) is a consistent method whatever the distribution of ξ, another estimation procedure developed for the structural variant of the measurement error model, viz. SLS, depends heavily on the assumption of normally distributed ξ-variables. As long as this assumption is true, SLS is superior to MALS, both with regard to bias and to the standard error.

4. When the distribution of ξ deviates from the normal distribution, SLS becomes strongly biased, the more so the farther away the distribution of ξ gets from normality. However the standard error of the SLS estimator is still rather small, indeed so small that the overall MSE of SLS is smaller than that of MALS in most cases except for the cubic regression with an exponential distribution of ξ.

5. This MSE behavior of SLS will certainly change when either the sample size is increased or the error variance σ_δ^2 becomes smaller. In these cases the overall MSE will typically be always larger for the SLS estimators whenever the distribution of ξ is non-normal. This is testified by the results of table 4.4. They show that for $\sigma_\delta^2 = 0.01$ the overall MSE of SLS is always considerably larger than the MSE of MALS, except for the case of normal ξ .

6. To sum up, SLS is always superior to MALS when the assumption of normality of ξ is valid. Whenever ξ deviates from normality, SLS becomes biased and, as

far as one is solely concerned with the bias, MALS should be prefered to SLS. The picture is not so clear when one takes the MSE as a precision criterion. With respect to this measure, SLS is rather robust, at least for not too large samples and if σ_δ^2 is large enough. For small σ_δ^2 and for large sample size, SLS deteriorates with respect to its overall MSE as compared to MALS.

ACKNOWLEDGEMENT

We thank T. Augustin for some useful comments on an earlier draft of the paper.

ξ Normal				
	ALS	MALS	SLS	
β_0	0.01	0.01	0.00	Bias
β_1	0.02	0.02	-0.09 **	
β_2	-0.04	-0.04	-0.09 **	
β_0	0.06	0.06	0.04	Standard deviation
β_1	0.10	0.10	0.07	
β_2	0.19	0.19	0.08	

ξ Uniform				
	ALS	MALS	SLS	
β_0	0.04	0.04 **	-0.05 **	Bias
β_1	0.02	0.02	-0.07 **	
β_2	-0.12	-0.11 **	-0.19 **	
β_0	0.11	0.11	0.04	Standard deviation
β_1	0.10	0.10	0.06	
β_2	0.36	0.34	0.07	

ξ Exponential				
	ALS	MALS	SLS	
β_0	0.01	0.01	-0.06 **	Bias
β_1	0.01	0.01	-0.34 **	
β_2	-0.02	-0.02	0.24 **	
β_0	0.05	0.05	0.04	Standard deviation
β_1	0.16	0.16	0.06	
β_2	0.15	0.15	0.10	

Table 4.1: Bias and standard error of three estimators in three different models with $k = 2$

ξ Normal						
	ALS	MALS		SLS		
β_0	0.03	-0.00		-0.01		Bias
β_1	-0.69	-0.25	**	-0.08	**	
β_2	-0.23	-0.01		0.09	**	
β_3	1.14	0.35	**	-0.13	**	
β_0	0.50	0.08		0.04		Standard deviation
β_1	4.58	0.38		0.15		
β_2	2.92	0.38		0.13		
β_3	7.68	0.55		0.16		

ξ Uniform						
	ALS	MALS		SLS		
β_0	0.17	0.01		-0.06	**	Bias
β_1	0.10	-0.23	**	0.41	**	
β_2	-0.66	-0.04		0.20	**	
β_3	0.23	0.45	**	-0.73	**	
β_0	1.19	0.14		0.06		Standard deviation
β_1	7.49	0.60		0.13		
β_2	4.05	0.46		0.10		
β_3	15.98	1.02		0.13		

ξ Exponential						
	ALS	MALS		SLS		
β_0	-0.09	0.03	*	-0.19	**	Bias
β_1	0.34	-0.14	**	-0.29	**	
β_2	0.68	-0.37	**	0.85	**	
β_3	-0.85	0.28	**	-0.26	**	
β_0	1.79	0.13		0.05		Standard deviation
β_1	2.91	0.35		0.19		
β_2	10.85	0.74		0.16		
β_3	6.59	0.50		0.19		

Table 4.2: Bias and standard error of three estimators in three different models with $k = 3$

		distribution of ξ		
degree	estimator	normal	uniform	exponential
$k = 2$	MALS	0.05	0.15	0.05
	SLS	0.03	0.06	0.19
$k = 3$	MALS	0.78	1.89	1.17
	SLS	0.10	0.79	1.00

Table 4.3: Overall MSE of the estimators in six different models, $\sigma_\delta^2 = 0.1$

		distribution of ξ		
degree	estimator	normal	uniform	exponential
k=2	MALS	0.0016	0.0023	0.0014
	SLS	0.0015	0.0034	0.0062
k=3	MALS	0.0155	0.0342	0.0297
	SLS	0.0088	0.0803	0.0798

Table 4.4: Overall MSE of the estimators in six different models, $\sigma_\delta^2 = 0.01$

REFERENCES

Buonaccorsi, J.P. (1996): A Modified Estimating Equation Approach to Correcting for Measurement Error in Regression. Biometrika, **83**, 433 – 440.

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

Chan, L. K., Mak, T.K., (1985): On the Polynomial Functional Relationship. JRSSB, **47**, 510–518.

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

Cheng, C.-L., and Schneeweiss, H., (1998): The Polynomial Regression with Errors in the Variables. JRSSB, **60**, 189–199.

Cheng, C.-L., Schneeweiss, H., and Thamerus, M., (1998): A Small Sample Estimator for a Polynomial Regression with Errors in the Variables. Discussion Paper **113**, Sonderforschungsbereich 386, University of Munich.

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

Kuha, J.T., Temple, J. (1999) Covariate Measurement Error in Quadratic Regression. Paper presented at the 52nd Session of the International Statistical Institute, Helsinki.

Nakamura, T. (1990): Corrected Score Functions for Errors-in-Variables Models: Methodology and Application to Generalized Linear Models, Biometrika **77**, 127–137.

Schneeweiss, H. and Mittag, H. J., 1986: *Lineare Modelle mit fehlerbehafteten Daten.* Physica, Heidelberg.

Stefanski, L. A., (1989): Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models, Communications in Statistics, Part A – Theory and Methods, **18**, 4335–4358.

Thamerus, M. (1998): Different Nonlinear Regression Models with Incorrectly Observed Covariates. In R. Galata and H. Küchenhoff (Eds.): *Econometrics in Theory and Practice, Festschrift for Hans Schneeweiß.* Physica, Heidelberg, New York.