Journal of Statistical Planning and Inference 81 (1999) 293–309

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Nonlinear unbiased estimation in the linear regression model with nonnormal disturbances

Henning Knautz

Institute for Statistics and Econometrics, University of Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany

Received 23 July 1996; accepted 25 January 1999

Abstract

In the application of the linear regression model there continues to be wide-spread use of the Least Squares Estimator (LSE) due to its theoretical optimality. For example, it is well known that the LSE is the best unbiased estimator under normality while it remains best linear unbiased estimator (BLUE) when the normality assumption is dropped. In this paper we extend an approach given in Knautz (1993) that allows improvement of the LSE in the context of nonnormal and nonsymmetric error distributions. It will be shown that there exist linear plus quadratic (LPQ) estimators, consisting of linear and quadratic terms in the dependent variable, which dominate the LS estimator, depending on second, third and fourth moments of the error distribution. A simulation study illustrates that this remains true if the moments have to be estimated from the data. Computation of confidence intervals using bootstrap methods reveal significant improvement compared with inference based on the LS especially for nonsymmetric distributions of the error term. © 1999 Elsevier Science B.V. All rights reserved.

MSC: 62J05; 62G15

Keywords: Linear regression model; Linear plus quadratic (LPQ) estimator; Bootstrap confidence intervals

1. Introduction

Let us be given the standard linear regression model

$$y = X\beta + \epsilon, \tag{1}$$

where the model matrix X of dimension $n \times p$ is assumed to have full column rank. This implies $X^{-}X = I$ for any generalized inverse X^{-} of X. The unknown parameter vector β is of length p and y represents the $n \times 1$ vector of observations of the endogeneous variable. Following Ullah et al. (1983), Tracy and Srivastava (1988) among others,

PII: S0378-3758(99)00023-3

we impose the following assumptions on the components ϵ_i of the error term ϵ. Let $\epsilon_i, i=1,\dots,n$, be independently distributed with

$$E(\epsilon_i)=0 \quad \text{and} \quad E(\epsilon_i^2)=\sigma^2, \tag{2}$$

and

$$E(\epsilon_i^3)=\sigma^3 m_{3i} \quad \text{and} \quad E(\epsilon_i^4)=\sigma^4 m_{4i}. \tag{3}$$

In the following let us denote $\{a_i\}=(a_1,\dots a_p)'$ for a column vector a.

We are interested in the estimation of the parameter vector β. For this we will consider the following class of estimators:

$$\mathscr{C}:=\left\{\tilde\beta|\tilde\beta=X^- y+\frac{1}{\sigma}\{y'H_i y\}-\sigma\{\operatorname{tr} H_i\}, H_iX=0, i=1,\dots,p\right\}, \tag{4}$$

where X^- is any generalized inverse of X, $H_i, i=1,\dots,p$ are $n\times n$ matrices which are assumed to be symmetric without loss of generality.

Parameter σ, m_{3i} and $m_{4i}, i=1,\dots,n$ are assumed to be known for the moment.

Observe that $\tilde\beta\in\mathscr{C}$ consists of a linear part $X^- y$ which estimates β unbiasedly, due to $X^-X=I$, and of quadratic terms $(1/\sigma)y'H_i y-\sigma\operatorname{tr} H_i$ for each component which have expectation zero. This follows from Lemma 1 which is stated below.

Motivation for this type of estimators is twofold: Firstly it can be derived from a characterization of all unbiased estimators in the linear regression model (cf. Koopmann 1982, p. 37) which are restricted to estimators which are equivariant under translations $\Pi_\alpha(y)=y+X\alpha, \alpha\in\mathbb{R}^p$. This implies the condition $H_iX=0$ in (4).

Secondly, similar estimators proved to be useful in the context of variance components estimation (cf. Kleffe, 1978; Rao and Kleffe, 1988, p. 32). These authors also derived the necessary formulas to deal with this type of linear plus quadratic functions. We state the necessary results in the following lemma:

Lemma 1. *Let us have $y=g+\epsilon$ where g is an arbitrary but fixed nonstochastic vector and ϵ is distributed according to* (2) *and* (3). *Further let $a,b\in\mathbb{R}^n$, and $A,B\in\mathbb{R}^{n\times n}$ be symmetric matrices for which $Ag=0$ and $Bg=0$ hold. Expectation and covariances of linear plus quadratic functions in y are given by*

$$E\left(a'y+\frac{1}{\sigma}y'Ay\right)=a'g+\sigma\operatorname{tr} A, \tag{5}$$

$$\begin{aligned}&\operatorname{Cov}\left(a'y+\frac{1}{\sigma}y'Ay, b'y+\frac{1}{\sigma}y'By\right)\\ &\quad=\sigma^2\operatorname{tr}\{B(2A+[\operatorname{Diag}A]\Delta+[\operatorname{Diag}a\mu'])\}+\sigma^2 b'\{a+[\operatorname{Diag}A]\mu\},\end{aligned} \tag{6}$$

where $[\operatorname{Diag}A]$ represents the diagonal matrix formed by the diagonal elements of A, $\mu=(m_{31},\dots,m_{3n})'$ is the vector of third moments and $\Delta=\operatorname{Diag}(m_{4i}-3)$ is the diagonal matrix of excess parameters.

To apply (5) to the components of $\tilde{\beta} \in \mathscr{C}$ we observe that $g = X\beta$ in the linear regression model. The condition $H_i X = 0$ is equivalent to $H_i = MH_iM$, where $M = I - XX^+$ which implies $H_i g = 0$.

It follows that this class consists of all unbiased equivariant estimators of β. By choosing $H_i = 0$ we see that the least squares estimator $b = X^+ y$ also belongs to $\mathscr{C}$.

2. Derivation of the optimal estimator and its properties

In the following we derive the optimal LPQ estimator in class $\mathscr{C}$ (cf. Knautz, 1993, Chapter 4). The different estimators in $\mathscr{C}$ are compared by the matrix MSE criterion:

$$\mathrm{MSE}(\tilde{\beta}, \beta) = E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'.$$

The optimal estimator is found with the help of the following lemma.

Lemma 2. *Let the components of $\tilde{\beta} \in \mathscr{C}$ be uncorrelated with all functions $f = b'y + (1/\sigma)y'By - \sigma \operatorname{tr} B$, where $b'X = 0$ and $BX = 0$. Then $\tilde{\beta}$ is the optimal estimator in $\mathscr{C}$ with respect to the matrix MSE criterion.*

Proof. For arbitrary $c \in \mathbb{R}^p$ we have $\mathrm{Cov}(c'\tilde{\beta}, f) = 0$. Due to a result from Rao (1973, p. 317) it follows that $c'\tilde{\beta}$ has minimal variance among all unbiased and invariant estimators for $c'\beta$ given by $a'y + (1/\sigma)y'Ay - \sigma \operatorname{tr} A$, where $AX = 0$ and $a'X = c'$. This implies

$$\mathrm{Var}(c'\tilde{\beta}) \leqslant \mathrm{Var}(c'\widetilde{\beta}^*) \tag{7}$$

for any other estimator $\widetilde{\beta}^* \in \mathscr{C}$, or equivalently since (7) holds for an arbitrary vector c

$$\mathrm{Cov}(\tilde{\beta}) \leqslant \mathrm{Cov}(\widetilde{\beta}^*) \tag{8}$$

in the sense of the Loewner partial ordering. □

Now we are in the position to determine an explicit formula for the optimal estimator. Let '$*$' denote the Hadamard product, i.e. the elementwise multiplication of two matrices of equal size and $\operatorname{diag} A$ represents the column vector formed by the diagonal elements of the $n \times n$ matrix A.

Theorem 1. *Let Δ and μ be defined as in Lemma 1 and for $M = I - XX^+$ let $M * M > 0, X^+$ denoting the Moore Penrose inverse of X. The optimal estimator $\tilde{\beta}_{\mathrm{LPQ}} \in \mathscr{C}$ is given by*

$$\tilde{\beta}_{\mathrm{LPQ}} = Ly + \frac{1}{\sigma}\{y'H_i y\} + \sigma d \tag{9}$$

with

$$L = X^+ + (D\tilde{K} * 1_p\mu')M, \tag{10}$$

$$d = D\tilde{K}1_n, \tag{11}$$

where 1_p *is a p-dimensional vector of ones, and*

$$D = X^+ * 1_p\mu' \tag{12}$$

$$\tilde{K} = \{2(M * M)^{-1} + \Delta - M * \mu\mu'\}^{-1}. \tag{13}$$

The vectors $\operatorname{diag} H_i$ *containing the diagonal elements of* $H_i, i = 1,\ldots, p$, *are computed from*

$$(\operatorname{diag} H_1 | \cdots | \operatorname{diag} H_p) = -\tilde{K}D'. \tag{14}$$

After transformation of $\operatorname{diag} H_i$ *into the diagonal matrix* $[\operatorname{Diag} H_i]$ *we finally compute*

$$H_i = (1/2)M\{[\operatorname{Diag} M[\operatorname{Diag} H_i]\mu\mu'] - [\operatorname{Diag} H_i]\Delta - [\operatorname{Diag}(X^+)'e_i\mu']\}M, \tag{15}$$

where e_i *is the ith unit vector of dimension p.*

Proof. Let us denote $\tilde{\beta}_i = l_i'y + (1/\sigma)y'H_iy - \sigma \operatorname{tr} H_i$ as the ith component of $\tilde{\beta} \in \mathscr{C}$, where $l_i' = e_i'X^-$. We have $H_iX = 0$ or equivalently $H_i = MH_iM$. Then it follows from Lemmas 1 and 2 that

$$b'(l_i + [\operatorname{Diag} H_i]\mu) + \operatorname{tr} B(2H_i + [\operatorname{Diag} H_i]\Delta) + [\operatorname{Diag} l_i\mu'] = 0 \tag{16}$$

for all b, B with $b'X = 0$ and $BX = 0$. These conditions are equivalent to $b = Me$ and $B = MEM$, where e, E are arbitrary, E is symmetric. Hence

$$M\{[\operatorname{Diag} H_i]\mu + l_i\} = 0 \tag{17}$$

and

$$M\{2H_i + [\operatorname{Diag} H_i]\Delta + [\operatorname{Diag} l_i\mu']\}M = 0. \tag{18}$$

Since it holds for any g-inverse X^- that $X^- = X^+ + UM$, where $M = I - XX^+$ and U arbitrary, it follows from (17) that

$$l_i = (X^+)'e_i - M[\operatorname{Diag} H_i]\mu. \tag{19}$$

Using $H_i = MH_iM$ and inserting (19) in (18) gives (15). Applying the diag-operator on (15) and observing that for any conformable matrices A, C and diagonal matrix F we have the relation

$$\operatorname{diag} AFC' = (A * C) \operatorname{diag} F,$$

we obtain

$$[2I + (M * M)(\Delta - M * \mu\mu')] \operatorname{diag} H_i = -(M * M) \operatorname{diag}((X^+)'e_i\mu'). \tag{20}$$

From this and (13) which is shown to exist in Knautz (1993, p. 26) we arrive at (14), where

$$
\begin{aligned}
D &= (\operatorname{diag}((X^+)'e_1\mu')|\ldots|\operatorname{diag}((X^+)'e_p\mu'))' \\
&= X^+ * 1_p\mu'.
\end{aligned}
\tag{21}
$$

Since $[\operatorname{Diag} H_i]\mu = (\operatorname{diag} H_i) * \mu$ we have

$$
L = \begin{pmatrix} l_1' \\ \vdots \\ l_p' \end{pmatrix} = X^+ - \begin{pmatrix} \mu' * (\operatorname{diag} H_1)' \\ \vdots \\ \mu' * (\operatorname{diag} H_p)' \end{pmatrix} M \tag{22}
$$

$$
= X^+ - ((\operatorname{diag} H_1|\cdots|\operatorname{diag} H_p)' * 1_p\mu')M \tag{23}
$$

which by applying (14) gives (10). Finally we obtain

$$
d = \{-\operatorname{tr} H_i\} = \{-1_n' \operatorname{diag} H_i\} = D\tilde{K}1_n \tag{24}
$$

by again using (14). □

The following corollary gives a more concise representation of the optimal estimator.

Corollary 1. *The estimator* $\tilde{\beta}_{\mathrm{LPQ}}$ *given in Theorem 1 can be rewritten as*

$$
\tilde{\beta}_{\mathrm{LPQ}} = X^+y + D\tilde{K}\{z * \mu - (1/\sigma)(M * M)^{-1}(z * z) + \sigma 1_n\}, \tag{25}
$$

where $z = My$ *is the vector of least squares residuals and* D, $\tilde{K}$ *are given in* (12) *and* (13).

Proof. Observe for the linear term of $\tilde{\beta}_{\mathrm{LPQ}}$ that

$$
(D\tilde{K} * 1_p\mu')My = (D\tilde{K} * 1_p\mu')z \tag{26}
$$

$$
= D\tilde{K}(z * \mu). \tag{27}
$$

For the quadratic term we have

$$
y'H_iy = z'G_iz = (\operatorname{diag} G_i)'(z * z), \tag{28}
$$

where the diagonal matrix G_i is given by

$$
G_i = (1/2)\{[\operatorname{Diag} M[\operatorname{Diag} H_i]\mu\mu'] - [\operatorname{Diag} H_i]\Delta - [\operatorname{Diag}(X^+)e_i\mu']\}.
$$

Now it holds that $\operatorname{diag} H_i = (M * M)\operatorname{diag} G_i$. Hence

$$
(\operatorname{diag} G_1|\cdots|\operatorname{diag} G_p) = -(M * M)^{-1}\tilde{K}D'. \tag{29}
$$

Combining (29) with (28) gives the desired result. □

Remark 1. The positive definiteness of $M*M$ which implies the existence of $(M*M)^{-1}$ in general is no problem. Precise conditions have been formulated in the context of MINQU estimation (cf. Mallela, 1972 or Rao, 1970).

Remark 2. Since σ^2, m_{3i} and m_{4i}, $i=1,\dots,n$ are generally unknown the estimator $\tilde{\beta}_{\mathrm{LPQ}}$ as such is infeasible. However, it can be seen in the sequel that these quantities can be replaced by appropriate estimators leading to a feasible linear plus quadratic estimator of β.

To conclude this section we present the dispersion matrix of the estimator $\widetilde{\beta}_{\mathrm{LPQ}}$ given in Theorem 1:

Corollary 2. *The dispersion matrix of* $\tilde{\beta}_{\mathrm{LPQ}}$ *in Theorem* 1 *is given by*

$$\mathrm{Cov}(\tilde{\beta}_{\mathrm{LPQ}}) = \sigma^2\{(X'X)^{-1} - D\tilde{K}D'\}. \tag{30}$$

Proof. Using formula (6) we compute the covariance between two components of the estimator as

$$\mathrm{Cov}(\tilde{\beta}_{\mathrm{LPQ},i}, \tilde{\beta}_{\mathrm{LPQ},j}) = \sigma^2\{e_i'(X'X)^{-1}e_j + e_j'X^+[\mathrm{Diag}\, H_i]\mu\}. \tag{31}$$

Since

$$e_j'X^+[\mathrm{Diag}\, H_i]\mu = (\mathrm{diag}\, H_i)'\,\mathrm{diag}\,(X^+)'e_j\mu' \tag{32}$$

$$= -e_i'D\tilde{K}D'e_j, \tag{33}$$

by (14) and (21) we obtain the desired result. □

In Knautz (1993, p. 26) it is also shown that $\tilde{K}$ is a positive-definite matrix. This implies that the LPQ estimator $\tilde{\beta}_{\mathrm{LPQ}}$ dominates the least-squares estimator in the sense of the MSE criterion whenever $D = X^+ * 1_p\mu'$ is different from zero which is the case when $\mu \neq 0$, i.e. there is at least one component of the error vector which is not symmetrically distributed.

3. Feasible estimation and consistency

As it was already mentioned in Section 2, the quantities σ, m_{3i}, m_{4i}, $i = 1,\dots,n$ in general will be unknown in practical data situations. Hence the optimal estimator is infeasible in this case. For the derivation of a feasible estimator it is necessary to estimate higher moments of the error distribution. For this purpose we confine ourselves to the situation of identically distributed errors, i.e. Eq. (3) is replaced by

$$E(\epsilon_i^3) = \sigma^3 m_3 \quad \text{and} \quad E(\epsilon_i^4) = \sigma^4 m_4, \tag{34}$$

for $i=1,\dots,n$. Replacing the unknown quantities σ^2, m_3 and $m_4 - 3$ in $\tilde{\beta}_{\mathrm{LPQ}}$ by consistent estimators $\hat{\sigma}^2, \hat{m}_3$ and $\hat{m}_{43}$ yields the feasible estimator

$$\tilde{\beta}^*_{\mathrm{LPQ}} = X^+y + \hat{m}_3X^+\tilde{K}^*\{\hat{m}_3z - (1/\hat{\sigma})(M*M)^{-1}(z*z) + \hat{\sigma}1_n\}, \tag{35}$$

where

$$\tilde{K}^* = \{2(M*M)^{-1} + \hat{m}_{43}I_n - \hat{m}_3^2M\}^{-1}. \tag{36}$$

In this section we will propose appropriate estimators for σ^2, m_3, and $m_4 - 3$ and give conditions under which the resulting estimator $\tilde{\beta}^*_{\mathrm{LPQ}}$ will be consistent.

Estimation of the unknown moments is made on the basis of the vector of LS residuals $z = My$. For σ^2 we utilize the unbiased estimator

$$\hat{\sigma}^2_{\mathrm{LS}} = z'z/v, \tag{37}$$

where $v = n - p$. In Knautz (1993) estimators for m_3 and $m_4 - 3$ were derived using results by Anscombe (1961, 1981), Pukelsheim (1980) and by McCullagh and Pregibon (1987). They give estimators for higher cumulants in the context of the linear regression model. The resulting estimators are consistent and dominate the naive estimators for m_3 and $m_4 - 3$. The latter are derived from the empirical distribution of the LS residuals $z_1, \ldots, z_n$. They are given by

$$\tilde{m}_3 = \frac{(1/n)\sum_{i=1}^n (z_i - \bar{z})^3}{s_z^3}, \tag{38}$$

$$\tilde{m}_{43} = \frac{(1/n)\sum_{i=1}^n (z_i - \bar{z})^4}{s_z^4} - 3, \tag{39}$$

where $s_z^2 = (1/n)\sum_{i=1}^n (z_i - \bar{z})^2$. The quantities $\tilde{m}_3$ and $\tilde{m}_{43}$ represent skewness and excess of the empirical distribution of the z_i. In Knautz (1993) a method was proposed to restrict estimators for the unknown moments such that

$$\tilde{m}_3^2 \leqslant \tilde{m}_{43} + 2. \tag{40}$$

This restriction reflects the fact that for an arbitrary distribution we have the relation (cf. Rohatgi and Szekely, 1989)

$$m_3^2 \leqslant m_4 - 1, \tag{41}$$

which can be sharpened for special classes of distributions. Eq. (40) turned out to be rather important for obtaining feasible LPQ estimators with reasonable properties. Note that this equation automatically holds for $\tilde{m}_3$ and $\tilde{m}_{43}$ since these are the moments of the empirical distribution.

Recent simulation results applying $\tilde{m}_3$ and $\tilde{m}_{43}$ for insertion in the feasible estimator $\tilde{\beta}^*_{\mathrm{LPQ}}$ surprisingly revealed no significant difference to the application of the more sophisticated estimators derived from Pukelsheim (1980) and McCullagh and Pregibon (1987). As (40) holds automatically, no further improvement of these naive estimators is necessary which also saves some amount of computation time. Consequently, we will apply these simpler moment estimators to compute the feasible LPQ estimator $\tilde{\beta}^*_{\mathrm{LPQ}}$.

It is interesting to observe that the equivariance property of the LPQ estimator which was used in its construction is not destroyed by the insertion of the estimators $\tilde{m}_3$ and $\tilde{m}_{43}$. If we consider translations of $\beta \mapsto \beta + \alpha$, $\alpha \in \mathbb{R}^p$ for the parameter vector or equivalently, $y \mapsto \pi_\alpha(y) = y + X\alpha$ we obtain

$$\tilde{\beta}^*_{\mathrm{LPQ}}(\pi_\alpha(y)) = \alpha + \tilde{\beta}^*_{\mathrm{LPQ}}(y) \tag{42}$$

since we have $\tilde{\beta}^*_{\mathrm{LPQ}} = \hat{\beta}_{\mathrm{LS}} + g_1(z)$, where $g_1(z)$ is a function depending on y only through the vector of LS residuals z. This implies that the bias of the feasible LPQ estimator is independent of the true parameter vector β.

Additionally a second equivariance property may be observed. Transforming the error terms by $\epsilon \mapsto \delta\epsilon$ or equivalently $y \mapsto \pi_\delta(y) = X\beta + \delta\epsilon$, it follows by a similar reasoning

$$\mathrm{Bias}(\tilde{\beta}^*_{\mathrm{LPQ}}(\pi_\delta(y))) = \delta\,\mathrm{Bias}(\tilde{\beta}^*_{\mathrm{LPQ}}(y)), \tag{43}$$

$$\mathrm{MSE}(\tilde{\beta}^*_{\mathrm{LPQ}}(\pi_\delta(y))) = \delta^2\,\mathrm{MSE}(\tilde{\beta}^*_{\mathrm{LPQ}}(y)). \tag{44}$$

This means that bias and mean squared error of the LPQ estimator increase proportionally to the error variance.

In the rest of this section we will state conditions under which $\tilde{\beta}^*_{\mathrm{LPQ}}$ based on consistent estimators of the moments is consistent for the parameter vector β. This will extend results given in Knautz (1993) and shows that the necessary assumptions are only slightly stronger than those usually applied for consistency of the LS estimator.

Theorem 2. *Let us be given the model* (1), (2) *and* (34). *Let* $\tilde{m}_3$ *and* $\tilde{m}_{43}$ *be the naive estimators of* m_3 *and* $m_4 - 3$ *given by* (38) *and* (39). *Further assume*

(A1) $\lim_{n\to\infty}(1/n)X'X = Q$, *$Q$ positive definite,*

(A2) $E(|\epsilon_i|^{4+\delta}) < \infty$, *for some* $\delta > 0$ *and for* $i = 1,\dots,n$,

(A3) $\lim_{n\to\infty}\lambda_{\min}(M * M) = c > 0$.

Under assumptions (A1)–(A3) *the feasible LPQ estimator* $\tilde{\beta}^*_{\mathrm{LPQ}}$ *is consistent for parameter vector* β.

Proof. Firstly assumption (A1) is sufficient for the consistency of the LS estimator $\hat{\beta} = (X'X)^{-1}X'y$ and implies $\mathrm{plim}(1/n)X'\epsilon = 0$, where 'plim' denotes the usual probability limit.

Assumption (A2) ensures the consistency of $\tilde{m}_3$ and $\tilde{m}_{43}$ (cf. White and MacDonald, 1980) but also implies the consistency of the estimators derived from Pukelsheim (1980) and McCullagh and Pregibon (1987), cf. Knautz (1993, p. 70). There it was also shown that consistency of the feasible LPQ estimator holds under the following two assumptions (see Knautz, 1993, p. 78):

$$\mathrm{plim}\,(1/n)X'\tilde{K}^*\epsilon = 0, \tag{45}$$

$$\mathrm{plim}\,(1/n)X'\tilde{K}^*\xi = 0, \tag{46}$$

where

$$\xi = (M * M)^{-1}(z * z) - \sigma^2 1_n, \tag{47}$$

and z represents the vector of LS residuals.

In the Appendix it is shown that (45) and (46) can be derived from assumptions (A1) to (A3). Assumption (A3) ensures the positive definiteness of $M * M$ for all sample sizes n. □

4. Simulation

In this section we present the results of a Monte Carlo simulation on the feasible LPQ estimator for which the asymptotic properties have been given in the last section. Our main purpose is the examination of its small sample properties under the model assumptions (1), (2) and (34). The simulation design was chosen in the following way:

- *Design matrix*: Design matrix X was obtained from a data set on fuel consumption (FUEL) for $n=48$ US federal states described in Weisberg (1985, p. 35). It consists a constant term and of four regressors which are TAX (tax rate), DLIC (percentage of population with driver's license), INC (average income) and ROAD (length of federal highways). The following equation was estimated from the original dataset fit by LS
$$\widehat{\text{FUEL}}_t = 377.29 - 34.79\text{TAX} + 13.36\text{DLIC} - 66.6\text{INC} - 2.43\text{ROAD}$$
with $\hat{\sigma}^2 = 4396.5$.
- *Parameter vector*: Three different values were chosen for parameter vector β (see Table 1): Vector β_1 is close to the OLS estimates while $\beta_2 = 100\,v_{\max}$ and $\beta_3 = 100\,v_{\min}$ where $v_{\min}, v_{\max}$ are the eigenvectors of $X'X$ corresponding to the minimal and maximal eigenvalue respectively.
- *Error variance*: For the variance σ^2 of the error term ϵ there were chosen the following three values
$$\sigma_1^2 = 3.600, \qquad \sigma_2^2 = 14.400, \qquad \sigma_3^2 = 900.$$
- *Distribution of the error term*: The components ϵ_i of the error term were generated according to well-known distributions with different skewness m_3 and excess $m_4 - 3$ and appropriately standardized to obtain expectation zero and variance σ^2 (see Table 2).

Some details on the distributions are given in the appendix. The simulation is intended to give an impression of the possible improvement that can be achieved by feasible LPQ estimation especially under skewed error distributions.

An additional consideration of error distributions which are not so well behaved (e.g. having bimodal densities or mixture distributions) as was suggested by one of the referees seems to be useful under robustness aspects especially in the estimation of the moments m_3 and m_4 but is beyond of the scope of this article.

For each combination of parameters and error distribution 500 iterations were conducted generating the dependent variable y and computing the least squares and the feasible LPQ estimate. The performance of these point estimates was examined by computing the following quantities for each component $i = 1, \ldots, 5$.

- *Ratio of the average absolute bias of LPQ and LS estimates*:
$$\frac{\frac{1}{500}\sum_{j=1}^{500} |\tilde{\beta}_{\text{LPQ},i}^{(j)} - \beta_i|}{\frac{1}{500}\sum_{j=1}^{500} |\hat{\beta}_{\text{LS},i}^{(j)} - \beta_i|}.$$

Table 1

Vector	Components				
β_1	400	−40	15	−60	0
β_2	1.71	13.07	98.40	7.27	9.51
β_3	98.83	−5.44	−0.78	−1.91	−0.96

Table 2

Distribution	m_3	m_4-3
Lognormal (Log N)	6.19	110.94
Exponential (Exp)	2.00	6.00
Pareto ($a=5$, Par1)	4.65	70.80
Pareto ($a=10$, Par2)	2.81	14.83
Beta (Beta)	−2.66	6.57
Gamma (Gam)	2.83	12.00
Normal (Norm)	0.00	0.00
Laplace (Lap)	0.00	3.00

- *Ratio of Mean Squared Error (MSE) of LPQ and LS estimates*:

$$\frac{\frac{1}{500}\sum_{j=1}^{500}(\tilde{\beta}^{(j)}_{\mathrm{LPQ},i}-\beta_i)^2}{\frac{1}{500}\sum_{j=1}^{500}(\hat{\beta}^{(j)}_{\mathrm{LS},i}-\beta_i)^2}.$$

- Percentage of iterations where LPQ estimate is closer to the true parameter than LS (‘Prob(LPQ better)’)
- *Ratio of Ranges of LPQ and LS estimates*:

$$\frac{\max_j(\tilde{\beta}^{(j)}_{\mathrm{LPQ},i})-\min_j(\tilde{\beta}^{(j)}_{\mathrm{LPQ},i})}{\max_j(\hat{\beta}^{(j)}_{\mathrm{LS},i})-\min_j(\hat{\beta}^{(j)}_{\mathrm{LS},i})}.$$

Additionally interval estimates were computed by applying percentile bootstrap methods (cf. Jeong and Maddala, 1993) resampling 1000 times from the LPQ or LS residuals

$$\tilde{z}=y-X\tilde{\beta}^*_{\mathrm{LPQ}}, \tag{48}$$

$$z=y-X\hat{\beta}_{\mathrm{LS}}. \tag{49}$$

in each of the 500 iterations. For the bootstrap intervals generated in this way the following quantities were computed.

- Coverage Probability (LPQ intervals).
- Average of the ratios of interval length (LPQ/LS).
- Percentage of the 500 iterations yielding a LPQ interval that is shorter than the corresponding LS interval.

Due to the equivariance properties of LPQ and LS estimator which were discussed in the last section the above measures should not depend on the special choice for the parameter vector β and error variance σ^2 but on the choice of the error distribution. This was confirmed by the simulation results. For this reason results are given in the appendix only for the parameter combination β_1 and σ_1^2. The following conclusions can be drawn from our results:

4.1. Point estimates

The bias measure indicates that the empirical bias of the LPQ estimator is even smaller than that of the LS estimator for nonsymmetric distributions and approximately equal for the two symmetric distributions. This shows that the feasible LPQ estimator seems to be unbiased as its theoretical counterpart.

The MSE measure gives an indication how much improvement compared with LS can be achieved for different error distributions. In the case of nonsymmetric distributions the empirical MSE of the LPQ estimator approximately ranges between 25% (beta distribution) and 50% (exponential distribution) of the MSE of the LS estimator. For the symmetric distributions the ratio is approximately one.

The two remaining measures also indicate that the LPQ estimates are closer concentrated around the true parameters with a high probability to dominate the LS estimates.

4.2. Interval estimates

The coverage probability of the computed LPQ bootstrap intervals is in general acceptable being somewhat below the nominal level of 95% for the symmetric distributions. The interval length is significantly shorter than that of the corresponding LS intervals ranging from 50% (beta distribution) to 75% (exponential distribution) of the LS interval length for the nonsymmetric distributions. With a probability close to one the LPQ interval can be expected to be shorter than its LS counterpart. The results for the two symmetric distributions indicate that no improvement can be achieved in this situation.

5. Concluding remarks

In this article we presented an unbiased linear plus quadratic estimator which dominates the least-squares estimator in the case of nonsymmetric error distributions with finite fourth moments. This emphasizes that the LS estimator is best *linear* unbiased under general error distributions, i.e. it can be dominated by unbiased *nonlinear* estimators. This seems to be an interesting illustration of the theory.

Insertion of simple estimates of the third and fourth moment yielded a feasible estimator which was shown to be consistent under certain reasonable assumptions. This estimator showed good performance in a simulation study with moderate sample size, improving least-squares based inference for nonsymmetric distributions being only slighty worse in the ideal situation of normal distribution.

In the simulation the LPQ estimator was only compared with the least-squares estimator. This is due to illustrative purposes. It is well known that there exist other estimators (e.g. Stein estimators) which also dominate least squares. Comparison with these estimators might also be interesting.

Some work remains to be done in the examination of the LPQ approach when the underlying error distributions violate the assumptions (e.g. correlated or heteroskedastic errors, nonexistence of higher moments).

Appendix A

A.1. Proof of consistency

In the following we will demonstrate that (45) and (46) can be derived from Assumptions (A1)–(A3) which completes the proof of Theorem 2.

To derive $\operatorname{plim}(1/n)X'\tilde{K}^{*}\epsilon = 0$ we write (cf. Knautz, 1993, p. 103)

$$\frac{1}{n}X'\tilde{K}^{*}\epsilon = \frac{X'X}{n}\left(\frac{X'X}{n} + \widetilde{m}_3^2\frac{X'\hat{A}^{-1}X}{n}\right)^{-1}\frac{X'\hat{A}^{-1}\epsilon}{n}, \tag{A.1}$$

where

$$\hat{A} = 2(M*M)^{-1} + (\widetilde{m}_{43} - \widetilde{m}_3^2)I_n. \tag{A.2}$$

(1) In the first step we prove $\operatorname{plim}(1/n)X'A^{-1}\epsilon = 0$, where

$$A = 2(M*M)^{-1} + (m_4 - 3 - m_3^2)I_n.$$

Now

$$E\left(\frac{X'A^{-1}\epsilon}{n}\right) = 0, \tag{A.3}$$

$$\operatorname{Cov}\left(\frac{X'A^{-1}\epsilon}{n}\right) = \frac{\sigma^2}{n^2}X'A^{-2}X, \tag{A.4}$$

where

$$A^{-2} = V[2\Delta_\lambda^{-1} + (m_4 - 3 - m_3^2)I_n]^{-2}V' \tag{A.5}$$

using the spectral decomposition $M*M = V\Delta_\lambda V'$, where Δ_λ is the diagonal matrix of eigenvalues for which we have $\lambda_i \leqslant 1$, $i = 1,\dots,n$ due to $M*M \leqslant I_n$ and $V'V = I_n$. Hence it follows

$$A^{-2} = V\,\mathrm{Diag}[1/(2/\lambda_i + m_4 - 3 - m_3^2)^2]V', \tag{A.6}$$

$$A^{-2} \leqslant V\,\mathrm{Diag}[1/(m_4 - 1 - m_3^2)^2]V', \tag{A.7}$$

$$A^{-2} \leqslant 1/(m_4 - 1 - m_3^2)^2 I_n. \tag{A.8}$$

This implies

$$\lim_{n\to\infty} \frac{\sigma^2}{n^2} X'A^{-2}X \leqslant \lim_{n\to\infty} \frac{\sigma^2}{n^2(m_4-1-m_3^2)^2} X'X = 0, \tag{A.9}$$

which completes the first step.

(2) Now we will point out that it does not make a difference asymptotically when the true moments in A are replaced by their consistent estimates thus yielding $\hat{A}$, i.e.

$$\operatorname{plim} \frac{X'A^{-1}\epsilon}{n} = \operatorname{plim} \frac{X'\hat{A}^{-1}\epsilon}{n}. \tag{A.10}$$

We write

$$\frac{X'(A^{-1}-\hat{A}^{-1})\epsilon}{n} = \frac{1}{n} X'VDV'\epsilon \tag{A.11}$$

where D is the diagonal matrix with elements

$$\delta_i = \frac{1}{2/\lambda_i + m_4 - 3 - m_3^2} - \frac{1}{2/\lambda_i + \widetilde{m}_{43} - \widetilde{m}_3^2}. \tag{A.12}$$

For δ_i, $i=1,\dots,n$, we can make use of Assumption (A2) which gives

$$\operatorname{plim} \widetilde{m}_{43} = m_4 - 3 \quad \text{and} \quad \operatorname{plim} \widetilde{m}_3 = m_3,$$

i.e. $\operatorname{plim} \delta_i = 0$ for $i=1,\dots,n$. Denote a_i' with elements a_{ij} for the ith row of $X'V$, $i=1,\dots,p$, and $\tilde{\epsilon} = V\epsilon$. Then the ith row of $(1/n)X'VDV'\epsilon$ is given by

$$a_i'D\tilde{\epsilon} = \frac{1}{n} \sum_{j=1}^{n} a_{ij}\delta_j\tilde{\epsilon}_j. \tag{A.13}$$

Applying the Cauchy–Schwarz inequality we obtain

$$\begin{aligned} \frac{1}{n}\left|\sum_{j=1}^{n} a_{ij}\delta_j\tilde{\epsilon}_j\right| &\leqslant \frac{1}{n}\left(\sum_{j=1}^{n}(a_{ij}\delta_j)^2 \sum_{j=1}^{n}\tilde{\epsilon}_j^2\right)^{1/2} \\ &\leqslant \left(\max_{1\leqslant j\leqslant n} \delta_j^2 \frac{1}{n}\sum_{j=1}^{n} a_{ij}^2 \frac{1}{n}\sum_{j=1}^{n}\tilde{\epsilon}_j^2\right)^{1/2}. \end{aligned} \tag{A.14}$$

Note that $\operatorname{plim} \max_{1\leqslant j\leqslant n} \delta_j^2 = 0$ since $\operatorname{plim} \delta_j = 0$ for all j and

$$\sum_{j=1}^{n} a_{ij}^2 = e_i'X'V'VXe_i = e_i'X'Xe_i = (X'X)_{ii}$$

where e_i denotes the ith unit vector. Then it follows that

$$\operatorname{plim} \frac{1}{n}\sum_{j=1}^{n} a_{ij}^2 = (Q)_{ii} > 0 \tag{A.15}$$

due to (A1) and finally

$$\begin{aligned} \operatorname{plim} \frac{1}{n}\sum_{j=1}^{n}\tilde{\epsilon}_j^2 &= \operatorname{plim} \frac{1}{n}\tilde{\epsilon}'\tilde{\epsilon} \\ &= \operatorname{plim} \frac{1}{n}\epsilon'\epsilon = \sigma^2. \end{aligned} \tag{A.16}$$

Hence the probability limit of the upper bound in (A.14) is zero which gives the desired result.

(3) To complete the proof of (45) we demonstrate that $X'\hat{A}^{-1}X = O_p(n)$. This follows from the inequality

$$\hat{A} > (\tilde{m}_{43} - \tilde{m}_3^2 + 2)I_n \tag{A.17}$$

which follows from $M * M < I_n$. Hence

$$\frac{1}{n}X'\hat{A}^{-1}X < \frac{1}{\tilde{m}_{43} - \tilde{m}_3^2 + 2}\frac{1}{n}X'X, \tag{A.18}$$

where the probability limit of the upper bound is given by $1/(m_4 - 1 - m_3^2)Q$. Since $(1/n)X'\hat{A}^{-1}X$ is nonnegative definite for all n we have $(1/n)X'\hat{A}^{-1}X = O_p(1)$.

Denoting $\operatorname{plim}(1/n)X'\hat{A}^{-1}X = \tilde{Q}$ from steps (1)–(3) and (A.1) it follows that

$$\operatorname{plim}\frac{1}{n}X'\tilde{K}^*\epsilon = Q(Q + \tilde{Q})^{-1} \cdot 0 = 0. \tag{A.19}$$

To show $\operatorname{plim}(1/n)X'\tilde{K}^*\xi = 0$, where $\xi = (M * M)^{-1}(z * z) - \sigma^2 1_n$, let us write

$$\frac{1}{n}X'\tilde{K}^*\xi = \frac{X'X}{n}\left(\frac{X'X}{n} + \hat{m}_3^2\frac{X'\hat{A}^{-1}X}{n}\right)^{-1}\frac{X'\hat{A}^{-1}\xi}{n}. \tag{A.20}$$

(4) It remains to examine the last factor of the right-hand side. For $(1/n)X'A^{-1}\xi$ we have expectation zero due to $E(\xi) = 0$ (cf. Knautz, 1993, p. 80) and

$$\begin{aligned}\operatorname{Cov}\frac{1}{n}X'A^{-1}\xi &= \frac{\sigma^4}{n^2}X'A^{-1}\left(2(M * M)^{-1} + (m_4 - 3)I_n\right)A^{-1}X \\ &\leqslant \frac{\sigma^4}{n^2}\left(m_4 - 3 + \frac{2}{\lambda_{\min}}\right)X'A^{-2}X \\ &\leqslant \frac{\sigma^4}{n^2}\frac{m_4 - 3 + (2/\lambda_{\min})}{(m_4 - 1 - m_3^2)^2}X'X,\end{aligned} \tag{A.21}$$

which follows from $(M * M)^{-1} < (1/\lambda_{\min})I_n$, $\lambda_{\min}$ denoting the smallest eigenvalue of $M * M$, and Eq. (A.6). Consequently we have

$$\lim_{n\to\infty}\operatorname{Cov}\left(\frac{1}{n}X'A^{-1}\xi\right) = 0, \tag{A.22}$$

which gives $\operatorname{plim}(1/n)X'A^{-1}\xi = 0$.

(5) As in step (2) it remains to show that

$$\operatorname{plim}\frac{X'A^{-1}\xi}{n} = \operatorname{plim}\frac{X'\hat{A}^{-1}\xi}{n}. \tag{A.23}$$

Writing

$$\frac{X'(A^{-1}-\hat{A}^{-1})\xi}{n}=\frac{1}{n}X'VDV'\xi, \tag{A.24}$$

where a_i, D as before and $\tilde{\xi}=V'\xi$, we obtain for the ith row of $(1/n)X'VDV'\xi$ (cf. step (2))

$$\left|\frac{1}{n}\sum_{j=1}^{n}a_{ij}\delta_j\tilde{\xi}_j\right|\leqslant\left(\max_{1\leqslant j\leqslant n}\delta_j^2\frac{1}{n}\sum_{j=1}^{n}a_{ij}^2\frac{1}{n}\sum_{j=1}^{n}\tilde{\xi}_j^2\right)^{1/2}. \tag{A.25}$$

It remains to show that $(1/n)\xi'\xi=\mathrm{O_p}(1)$ which would imply that the probability limit of the upper bound in (A.25) is zero. Now

$$\xi'\xi=(z*z)'(M*M)^{-2}(z*z)+\sigma^4 1_n'1_n-2\sigma^2 1_n'(M*M)^{-1}(z*z), \tag{A.26}$$

where $1_n'1_n=n$, and

$$(z*z)'(M*M)^{-2}(z*z)\leqslant\frac{1}{\lambda_{\min}^2}(z*z)'(z*z). \tag{A.27}$$

For $(z*z)'(z*z)=1_n'z^{*4}$ where z^{*4} is the vector of the LS residuals with elements raised to the fourth power, we obtain $\operatorname{plim}(1/n)1_n'z^{*4}=\sigma^4 m_4$. Hence the first term in (A.26) is $\mathrm{O_p}(n)$ which trivially holds also for the second term. For the last term this property follows along the same lines of the proof for the identity $\operatorname{plim}(1/n)X'\xi=0$ (cf. Knautz, 1993, p. 81).

Collecting the results we see that (A.23) is shown and applying (A1) and steps (3)–(5) we obtain the desired result $\operatorname{plim}(1/n)X'\tilde{K}^*\xi=0$ which completes the proof.

A.2. Distributions in the simulation

The following distributions considered in the simulation are given in Table 3.

Table 3

Distribution	Density $f(x)$	Parameter choice
Lognormal	$1/(\sqrt{2\pi}\sigma x)\exp\{-(\ln x-\mu)^2/2\sigma^2\}$	$\mu=0, \sigma^2=1$
Exponential	$f(x)=\lambda\exp(-\lambda x),\ x\geqslant 0$	$\lambda=1$
Pareto	$(a/x_0)(x/x_0)^{a+1},\ x>x_0$	$x_0=1,\ a=5, a=10$
Laplace	$(1/2\lambda)\exp(-\lvert x-\mu\rvert/\lambda)$	$\mu=0, \lambda=1$
Normal	$1/(\sqrt{2\pi}\sigma)\exp\{-(x-\mu)^2/2\sigma^2\}$	$\mu=0,\ \sigma^2=1$
Gamma	$b^p/\Gamma(p)\exp(-bx)x^{p-1},\ x>0$	$b=1,\ p=0.5$
Beta	$1/B(p,q)x^{p-1}(1-x)^{q-1},\ x\in(0,1)$	$p=1,\ q=0.1$

Table 4

	Relative bias					MSE-ratio (LPQ/LS)				
Log N	0.56	0.51	0.57	0.52	0.53	0.29	0.24	0.28	0.22	0.26
Exp	0.69	0.72	0.70	0.68	0.72	0.49	0.49	0.50	0.47	0.53
Par1	0.59	0.60	0.60	0.57	0.60	0.32	0.33	0.35	0.32	0.36
Par2	0.64	0.66	0.65	0.63	0.66	0.40	0.41	0.43	0.40	0.44
Lap	0.95	0.98	0.95	0.95	0.98	0.91	0.95	0.90	0.93	0.98
Norm	1.00	1.01	1.02	1.01	1.02	1.00	1.02	1.02	1.03	1.03
Gam	0.56	0.59	0.56	0.60	0.61	0.32	0.34	0.32	0.35	0.35
Beta	0.48	0.47	0.45	0.46	0.47	0.24	0.24	0.23	0.24	0.23
	Prob(LPQ better)					Ratio of range(LPQ/LS)				
Log N	71.0	69.4	66.0	72.0	69.8	0.54	0.44	0.37	0.31	0.49
Exp	68.0	62.6	66.8	66.0	66.8	0.71	0.61	0.66	0.79	0.81
Par1	68.2	66.4	69.6	69.2	68.8	0.42	0.45	0.41	0.58	0.63
Par2	68.6	64.2	68.6	68.2	68.6	0.55	0.52	0.53	0.70	0.73
Lap	52.0	52.0	53.2	57.4	49.8	1.01	0.82	0.99	1.04	0.97
Norm	48.2	50.2	45.8	48.8	50.2	0.97	0.97	0.99	0.97	1.04
Gam	74.0	70.8	73.0	69.4	68.6	0.50	0.58	0.53	0.60	0.62
Beta	75.4	72.6	77.0	77.2	75.8	0.76	0.56	0.62	0.72	0.64

Table 5

	Coverage probability					Ratio of interval length					Prob(LPQ interval shorter)				
Log N	93.4	94.6	94.2	94.8	94.0	0.60	0.61	0.61	0.60	0.61	100.0	100.0	100.0	100.0	100.0
Exp	92.4	91.8	93.4	93.2	92.8	0.73	0.73	0.74	0.74	0.74	99.8	99.8	99.4	99.8	99.6
Par1	92.8	91.8	93.4	93.4	93.4	0.63	0.63	0.64	0.64	0.64	100.0	100.0	100.0	100.0	100.0
Par2	92.2	91.2	93.6	93.4	92.6	0.68	0.68	0.69	0.69	0.69	100.0	99.8	99.8	99.8	100.0
Lap	91.4	90.0	92.2	90.8	91.6	0.95	0.95	0.95	0.95	0.95	74.6	76.2	73.4	77.6	71.4
Norm	90.0	92.2	91.4	93.6	91.0	1.00	1.00	1.00	1.00	1.00	37.2	38.8	40.8	39.6	39.8
Gam	94.8	95.0	92.4	95.6	95.0	0.63	0.63	0.63	0.63	0.64	100.0	100.0	100.0	100.0	100.0
Beta	96.4	96.4	93.8	94.6	95.6	0.50	0.50	0.50	0.50	0.51	100.0	100.0	100.0	100.0	100.0

A.3. Simulation results

Details of point estimation with parameters β_1 and σ_1^2 are given in Table 4. Details of interval estimation with parameters β_1 and σ_1^2 are given in Table 5.

References

Anscombe, F.J., 1961. Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability, vol. 1, pp. 1–36.

Anscombe, F.J., 1981. Computing in Statistical Science Through APL. Springer, Heidelberg.

Jeong, J., Maddala, G.S., 1993. A perspective on application of bootstrap in econometrics. In: Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, vol. 11. Elsevier Science B.V., Amsterdam, pp. 573–609.

Kleffe, J., 1978. Simultaneous estimation of expectation and covariance matrix in linear models. Math. Oper. Statist., Ser. Statist. 443–478.

Knautz, H., 1993. Nichtlineare Schätzung des Parametervektors im linearen Regressionsmodell. Mathematical Systems in Economics, vol. 133. Anton Hain, Frankfurt a.M.

Koopmann, R., 1982. Parameterschätzung bei a priori Information. Vandenhoek & Ruprecht, Göttingen.
Mallela, P., 1972. Necessary and sufficient conditions for MINQU-estimation of heteroskedastic variances in linear models. J. Amer. Statist. Assoc. 67, 486–487.
McCullagh, P., Pregibon, D., 1987. k-statistics and dispersion effects in regression. Ann. Statist. 15, 202–219.
Pukelsheim, F., 1980. Multilinear estimation of skewness and kurtosis in linear models. Metrika 27, 103–113.
Rao, C.R., 1970. Estimation of heteroskedastic variances in linear models. J. Amer. Statist. Assoc. 65, 161–172.
Rao, C.R., 1973. Linear Statistical Inference and Its Applications. Wiley, New York.
Rao, C.R., Kleffe, J., 1988. Estimation of Variance Components and Applications. North-Holland, Amsterdam.
Rohatgi, V.K., Szekely, G.J., 1989. Sharp inequalities Between skewness and kurtosis. Statist. Probab. Lett. 8, 297–299.
Tracy, D.S., Srivastava, A.K., 1988. A synthesis of Stein-rule and mixed regression procedures in linear regression models under nonnormality. Commun. Statist. Theory Methods 17, 691–704.
Ullah, A., Srivastava, V.K., Chandra, R., 1983. Properties of shrinkage estimators in linear regression when disturbances are not normal. J. Econometrics 21, 389–402.
Weisberg, S., 1985. Applied Linear Regression, second ed. Wiley, New York.
White, H., MacDonald, G.M., 1980. Some large-sample tests for nonnormality in the linear regression model. J. Amer. Statist. Assoc. 75, 16–31.