Kukush, Alexander; Schneeweiss, Hans

Working Paper

Asymptotic optimality of the quasi-score estimator in a class of linear score estimators

Discussion Paper, No. 477

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Kukush, Alexander; Schneeweiss, Hans (2006) : Asymptotic optimality of the quasi-score estimator in a class of linear score estimators, Discussion Paper, No. 477, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München,
https://doi.org/10.5282/ubm/epub.1845

This Version is available at:
https://hdl.handle.net/10419/31040

Asymptotic optimality of the quasi-score estimator in a class of linear score estimators

Alexander Kukush[1] and Hans Schneeweiss[2]

[1] Kiev National Taras Shevchenko University,
Vladimirskaya st. 60, 01033, Kiev, Ukraine,
e-mail: alexander_kukush@univ.kiev.ua

[2] University of Munich,
Akademiestr.1, 80799 Munich, Germany,
e-mail: schneew@stat.uni-muenchen.de

Abstract

We prove that the quasi-score estimator in a mean-variance model is optimal in the class of (unbiased) linear score estimators, in the sense that the difference of the asymptotic covariance matrices of the linear score and quasi-score estimator is positive semi-definite. We also give conditions under which this difference is zero or under which it is positive definite. This result can be applied to measurement error models where it implies that the quasi-score estimator is asymptotically more efficient than the corrected score estimator.

1 Introduction

Assume that the relation between a dependent (continuous or discrete) variable y and a (possibly vector-valued) continuous variable x is given by a pair of conditional mean and variance functions

$$\mathbb{E}(y|x) =: m(x,\beta) \quad (1)$$
$$\mathbb{V}(y|x) =: v(x,\beta), \quad (2)$$

which are known to the statistician except for an unknown k-dimensional parameter vector β to be estimated from an i.i.d. sample $(x_i, y_i), i = 1, \ldots, n$. The functions m and v are assumed to be sufficiently smooth with respect to x and β, and $v > 0$. We also assume that the distribution of x does not

depend on β (i.e., β is a regression parameter vector and does not contain parameters describing the distribution of x). Such a model is called a *mean-variance model*, cf. Carroll *et al.* (1995).

To estimate this model we consider the class of all unbiased estimating functions that are linear in y:

$$S_L := S_L(x, y, \beta) := g(x, \beta)y - h(x, \beta), \tag{3}$$

where g and h are column vectors of the same dimension, k, as β. As functions of x and β, they are assumed to be sufficiently smooth. We call S_L a *linear score function*. By definition, S_L is unbiased, which means that

$$\mathbb{E}S_L(x, y, \beta) = 0 \tag{4}$$

for all β, where the expectation is taken under the same β as the β in the argument of S_L.

The estimator of β based on S_L is called *linear score* (LS) *estimator* $\widehat{\beta}_L$ and is given as the solution to the equation

$$\sum_1^n S_L(x_i, y_i, \widehat{\beta}_L) = 0.$$

Under general conditions, similar to those studied in Kukush and Schneeweiss (2005), $\widehat{\beta}_L$ is consistent and asymptotically normal.

Within this class of estimators, the *quasi-score* (QS) *estimator* $\widehat{\beta}_Q$ stands out, cf. Wedderburn (1974) for the related concept of quasi-likelihood. It is based on the *quasi-score function*

$$S_Q := S_Q(x, y, \beta) := [y - m(x, \beta)]v(x, \beta)^{-1}\frac{\partial m(x, \beta)}{\partial \beta}, \tag{5}$$

which is obviously a member of the class of linear score functions.

We want to prove that $\widehat{\beta}_Q$ is optimal within the class of linear score estimators in the sense that its asymptotic covariance matrix (ACM) is less or equal (in the Loewner sense) to the ACM of any other estimator of this class. We also give conditions under which the ACMs are equal and under which the $\geq$ sign can be replaced with the $>$ sign.

This kind of problem shows up in the context of measurement error models. In Kukush *et al.* (2005), a measurement error model based on an error-free regression model from an exponential family was considered. The likelihood

score function of the error-free model can be transformed into a so-called corrected score function, which has exactly the form (3). In addition, the quasi-score function (5) can be constructed from a derived mean-variance model if the distributions of the latent regressor and the measurement error are given.

It can be shown that the corrected score estimator is (asymptotically) less efficient than the quasi-score estimator. This was proved in Kukush *et al.* (2005) with the help of an intermediate estimator and by going back to the underlying likelihood score function of the error-free model. However, it turns out that one need not resort to the original error-free model and, indeed, the proof can be greatly simplified if one just stays with the derived mean-variance model in the manifest variables x and y. The optimality statement we are going to prove is also more general as it does not only apply to measurement error models but to any mean-variance model.

If the class of linear score functions is restricted to the class of *conditionally unbiased* linear score functions, i.e., to score functions S_L^* with the property $\mathbb{E}(S_L^*|x) = 0$, then the optimality of QS within this class is almost an immediate consequence of Theorem 2.3 in Heyde (1997), which deals with score functions of the form $[y-m(x,\beta)]g(\frac{\partial m(x,\beta)}{\partial \beta})$. Here, however, we want to prove the optimality of QS within the wider class of linear score functions that are unconditionally unbiased, and for that we need some additional arguments. We can prove this more general result by applying Heyde's (1997) very general optimality criterion, Theorem 2.1. But we go beyond this optimality criterion when we give conditions for strict optimality of QS.

In Section 2 we give our main result: a proof of the optimality of QS and a condition when LS and QS have the same efficiency. Section 3 gives conditions under which QS is strictly better than LS. Some concluding remarks are found in Section 4.

2 Optimality of QS

In the sequel, we often omit the arguments in the various functions, g, h, m, S_L etc. E.g., we abbreviate (3) by writing $S_L = gy - h$. We also denote the derivative of a function with respect to β by the subscript β. Thus

$$m_\beta := \frac{\partial m(x,\beta)}{\partial \beta},$$

where m_β is a column vector of the same dimension, k, as β. By convention, the derivative of a column vector-valued function $l(\beta)$ with respect to β is always meant to be the matrix $l_\beta := \frac{\partial l}{\partial \beta^\top}$ with (i,j)-element $\frac{\partial l_i}{\partial \beta_j}$.

The ACM of any linear score estimator $\widehat{\beta}_L$ is given, under general conditions, by the sandwich formula

$$\Sigma_L = A_L^{-1} B_L A_L^{-\top}, \tag{6}$$

where

$$\begin{aligned} A_L &:= -\mathbb{E} S_{L\beta} \\ B_L &:= \mathbb{E} S_L S_L^\top. \end{aligned}$$

We implicitly assume that A_L is non-singular.

The ACM of the quasi-score estimator $\widehat{\beta}_Q$ is given by a similar sandwich formula, which, however, simplifies to

$$\Sigma_Q = (\mathbb{E} v^{-1} m_\beta m_\beta^\top)^{-1}. \tag{7}$$

The following theorem states the optimality of QS within the class of linear score estimators.

Theorem 1

a) In a mean-variance model

$$\Sigma_Q \leq \Sigma_L \tag{8}$$

for any linear score estimator $\widehat{\beta}_L$.

b) Moreover, $\Sigma_Q = \Sigma_L$ for a specific β if, and only if,

$$\begin{aligned} h(x,\beta) &= m(x,\beta) g(x,\beta) && (9) \\ g(x,\beta) &= K(\beta) v(x,\beta)^{-1} m_\beta(x,\beta) && (10) \end{aligned}$$

for some non-stochastic non-singular matrix $K(\beta)$.

c) Finally, $\Sigma_Q = \Sigma_L$ for all β if, and only if, $\widehat{\beta}_Q = \widehat{\beta}_L$ for all samples.

Proof: First note that (3) and (4) imply

$$\mathbb{E}(gm - h) = 0$$

and furthermore, since the expectation is taken with respect to x and the distribution of x does not depend on β,

$$\mathbb{E}(g_\beta m + g m_\beta^\top - h_\beta) = 0.$$

It follows that

$$\begin{aligned} A_L &= -\mathbb{E}(g_\beta y - h_\beta) \\ &= -\mathbb{E}(g_\beta m - h_\beta) \\ &= \mathbb{E} g m_\beta^\top. \end{aligned} \tag{11}$$

In addition, by (3),

$$\begin{aligned} B_L &= \mathbb{E}[g(y-m) - (h - mg)][g(y-m) - (h-mg)]^\top \\ &= \mathbb{E}[v g g^\top + (h-mg)(h-mg)^\top] \end{aligned} \tag{12}$$

From (6), (11) and (12),

$$(\mathbb{E} g m_\beta^\top)^{-1} \mathbb{E} v g g^\top (\mathbb{E} g m_\beta^\top)^{-\top} \leq \Sigma_L \tag{13}$$

In order to prove (8), we need only to show that

$$(\mathbb{E} v^{-1} m_\beta m_\beta^\top)^{-1} \leq (\mathbb{E} g m_\beta^\top)^{-1} \mathbb{E} v g g^\top (\mathbb{E} g m_\beta^\top)^{-\top} \tag{14}$$

or equivalently

$$\mathbb{E} v^{-1} m_\beta m_\beta^\top \geq (\mathbb{E} g m_\beta^\top)^\top (\mathbb{E} v g g^\top)^{-1} \mathbb{E} g m_\beta^\top. \tag{15}$$

Here we assume that $\mathbb{E} v g g^\top > 0$. Denote

$$v^{\frac{1}{2}} (\mathbb{E} v g g^\top)^{-\frac{1}{2}} g =: p, \quad v^{-\frac{1}{2}} m_\beta =: q. \tag{16}$$

Then (15) is equivalent to

$$\mathbb{E} q q^\top \geq \mathbb{E} q p^\top \mathbb{E} p q^\top, \tag{17}$$

where $\mathbb{E} p p^\top = I$. Now, since

$$\begin{pmatrix} p \\ q \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}^\top = \begin{pmatrix} p p^\top & p q^\top \\ q p^\top & q q^\top \end{pmatrix} \geq 0,$$

therefore also

$$\begin{pmatrix} I & \mathbb{E} p q^\top \\ \mathbb{E} q p^\top & \mathbb{E} q q^\top \end{pmatrix} \geq 0. \tag{18}$$

The desired result will then follow from the following purely algebraic Lemma, which is proved in the appendix.

Lemma: Let Q be a symmetric $(m \times m)$-matrix and let P be $m \times k$ and I the $k \times k$ identity matrix. Then

$$\begin{pmatrix} I & P^\top \\ P & Q \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad Q \geq PP^\top \tag{19}$$

and

$$\begin{pmatrix} I & P^\top \\ P & Q \end{pmatrix} > 0 \quad \Rightarrow \quad Q > PP^\top. \tag{20}$$

Inequality (17) now follows from (18) and (19) with $Q = \mathbb{E}qq^\top$ and $P = \mathbb{E}qp^\top$, $(m = k)$. This proves (14) and part a) of Theorem 1.

In order to prove part b), we note that $\Sigma_Q = \Sigma_L$ for any specific β iff there is equality both in (13) and (15). We have equality in (13) iff

$$\mathbb{E}(h - mg)(h - mg)^\top = 0,$$

which is equivalent to $h = mg$, a.s. But as m, g, and h are continuous functions in x and x is a continuous variable, this is equivalent to

$$h = mg \tag{21}$$

as function of x (given β).

On the other hand, equality in (15) means equality in (17) and this means

$$\mathbb{E}\begin{pmatrix} p \\ q \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix}^\top = \begin{pmatrix} I & \mathbb{E}pq^\top \\ \mathbb{E}qp^\top & \mathbb{E}qp^\top \mathbb{E}pq^\top \end{pmatrix} = \begin{pmatrix} I \\ \mathbb{E}qp^\top \end{pmatrix}\begin{pmatrix} I \\ \mathbb{E}qp^\top \end{pmatrix}^\top.$$

Thus the $(2k \times 2k)$-matrix $\mathbb{E}(p^\top, q^\top)^\top(p^\top, q^\top)$ has rank k. But then

$$K_1 p + K_2 q = 0, \text{a.s.}$$

with some non-stochastic $(k \times k)$-matrices K_1 and K_2 such that rank $(K_1, K_2) = k$. The matrices K_1 and K_2 are functions of β. Since $\mathbb{E}qq^\top = \mathbb{E}v^{-1}m_\beta m_\beta^T > 0$ and $\mathbb{E}pp^\top = I > 0$, K_1 and K_2 are both non-singular. Hence $p = -K_1^{-1}K_2 q$, a.s. From the definition of p and q in (16) it follows that $g = Kv^{-1}m_\beta$, a.s., with some non-stochastic non-singular matrix $K = K(\beta)$. By the argument that led to (21), then also

$$g = Kv^{-1}m_\beta \tag{22}$$

as functions of x (given β). Conversely, (22) implies equality in (15), as one can see directly by substituting g from (22) in (15) his proves part b) of Theorem 1.

Part c) now follows immediately: If (21) and (22) hold true for all β, then

$$S_L = (y - m)g = KS_Q. \tag{23}$$

Thus if $\Sigma_Q = \Sigma_L$ for all β, S_L and S_Q generate the same estimator, i.e., $\widehat{\beta}_L = \widehat{\beta}_Q$. The converse is obvious. ♦

Remark 1: The linear score function in (23)

$$S_S := (y - m)g$$

gives rise to the special linear score estimator $\widehat{\beta}_S$, which we may call a simple (linear) score (SS) estimator. With regard to asymptotic efficiency it stands between $\widehat{\beta}_L$ and $\widehat{\beta}_Q$ in so far as

$$\Sigma_Q \leq \Sigma_S \leq \Sigma_L,$$

where Σ_S is the ACM of $\widehat{\beta}_S$. Inequality $\Sigma_S \leq \Sigma_L$ is just a restatement of (13), as Σ_S is equal to the l.h.s. of (13), and $\Sigma_Q \leq \Sigma_S$ is a restatement of (14).

Remark 2: The SS estimator corresponds to the SS estimator introduced in Kukush et al. (2005), except that there it was related to the CS estimator of a measurement error model, whereas here it belongs to a general mean-variance model.

Remark 3: The simple score function S_S has the property that not only $\mathbb{E}S_S = 0$ but even $\mathbb{E}(S_S|x) = 0$. It is of a form which was also considered in Heyde's (1997) Theorem 2.3, except that there the function g was a function of β only via m_β, whereas here it is a general function of x and β. Apart from this slight difference, one can immediately deduce $\Sigma_Q \leq \Sigma_S$ from Heyde (1997, Theorem 2.3). We preferred to give an independent proof via the lemma because that enabled us to study also the case when $\Sigma_Q = \Sigma_S = \Sigma_L$ (part b) of Theorem 1).

An example for part b) of Theorem 1 is the following. Consider the special linear score estimator based on the score function

$$S_L = (y - m)m_\beta.$$

Thus S_L differs from S_Q just by the omission of the factor v^{-1} and would be equal to S_Q if v were constant (i.e., homoscedastic). Clearly $\Sigma_Q \leq \Sigma_L$

for this estimator. According to Theorem 1 b) (10) $\Sigma_Q = \Sigma_L$ if, and only if, $g := m_\beta = Kv^{-1}m_\beta$ (The other condition (9) is satisfied anyway). Since the component functions $m_{\beta i}(x), i = 1, ..., k$, of $m_\beta(x)$ are linearly independent, there exist values $x_1, ..., x_k$ such that the matrix M with elements $M_{ij} = m_{\beta i}(x_j)$ is non-singular. We then have $vM = KM$ and consequently $vI = K$. As K is a constant matrix for given β, v is also constant given β and does not depend on x. Thus $\Sigma_Q = \Sigma_L$ if, and only if, $v = v(\beta) = const.$ If this is true for every β, then $\hat{\beta}_Q = \hat{\beta}_L$.

Another example in the context of measurement error models is found in Kukush *et al.* (2005), Section 7.1.

We can give another, very short, proof of Theorem 1a) by employing a criterion of Heyde (1997, Theorem 2.1). According to this criterion $\widehat{\beta}_Q$ is more efficient than $\widehat{\beta}_L$ if

$$(\mathbb{E}S_{L\beta})^{-1}\mathbb{E}S_L S_Q^\top$$

does not depend on β.

Indeed, by (11),

$$\mathbb{E}S_{L\beta} = -A_L = -\mathbb{E}gm_\beta^\top$$

and

$$\begin{aligned} \mathbb{E}S_L S_Q^\top &= \mathbb{E}[g(y-m) - (h-gm)]v^{-1}(y-m)m_\beta^\top \\ &= \mathbb{E}gm_\beta^\top. \end{aligned}$$

Thus

$$(\mathbb{E}S_{L\beta})^{-1}\mathbb{E}S_L S_Q^\top = -I$$

is obviously independent of β, and it follows that $\Sigma_Q \leq \Sigma_L$ for any linear score estimator. ♦

This argument, however, does not allow us to give conditions under which $\Sigma_Q = \Sigma_L$ (Theorem 1b)) or $\Sigma_Q < \Sigma_L$ (Theorem 2).

3 Strict optimality of QS

We want to give sufficient conditions such that $\Sigma_Q < \Sigma_L$, (i.e., $\Sigma_L - \Sigma_Q$ is positive definite).

Theorem 2 Suppose that for any specific β one of the following two conditions holds, where $g = (g_1, \ldots, g_k)^\top$, $h = (h_1, \ldots, h_k)^\top$ and $\beta = (\beta_1, \ldots, \beta_k)^\top$, g and h being the vectors of a linear score function (3).

a) The functions $h_i - mg_i, i = 1, \ldots, k$, are linearly independent as functions of x.

b) $\text{span}\{g_1, \ldots, g_k\} \cap \text{span}\{v^{-1}m_{\beta_1}, \ldots, v^{-1}m_{\beta_k}\} = \{0\}$.

Then

$$\Sigma_Q < \Sigma_L.$$

Proof: Under condition a),

$$\mathbb{E}(h - mg)(h - mg)^\top > 0.$$

This implies, see (12), that $\mathbb{E}vgg^T < B_L$ and hence that we have strict inequality in (13). Together with (14) this implies $\Sigma_Q < \Sigma_L$.

Under condition b), the components p_i and $q_i, i = 1, \ldots, k$, of the vectors p and q defined in (16) satisfy

$$\text{span}\{p_1, \ldots, p_k\} \cap \text{span}\{q_1, \ldots, q_k\} = \{0\}.$$

As $\{p_1, \ldots, p_k\}$ and $\{q_1, \ldots, q_k\}$ are both linearly independent sets of functions of x, see the proof of Theorem 1, it follows that the $2k$ functions $p_1, \ldots, p_k$, $q_1, \ldots, q_k$ are linearly independent. Therefore, we can replace the $\geq$ sign in (18) by the $>$ sign. From (20) of the lemma it now follows that

$$\mathbb{E}qq^\top > \mathbb{E}qp^\top \mathbb{E}pq^\top,$$

i.e., we have strict inequality in (15) and consequently also in (14). Together with (13) this implies $\Sigma_Q < \Sigma_L$. ♦

Remark 4: We may supplement this result by stating that condition a) implies $\Sigma_S < \Sigma_L$ and condition b) implies $\Sigma_Q < \Sigma_S$, and hence both imply $\Sigma_Q < \Sigma_L$.

Applications of Theorem 2 in the context of measurement error models are found in Kukush *et al.* (2005), Section 7.2.

4 Conclusion

When one wants to estimate a parametric regression of y on x given by a conditional mean function $\mathbb{E}(y|x) = m(x, \beta)$ and supplemented by a conditional variance function $\mathbb{V}(y|x) = v(x, \beta)$, then the quasi-score (QS) estimator is often the estimator of ones choice. It is known, see Heyde (1997, Theorem 2.3) that this estimator is asymptotically most efficient within the class of all estimators that are based on a linear-in-y conditionally unbiased estimating function

$$S_L^*(x, y, \beta) = g(x, \beta)y - h(x, y),$$

where conditional unbiasedness means that $\mathbb{E}(S_L^*|x) = 0$. We can prove that this optimality result holds within the wider class of so-called linear score estimators, which are based on unconditionally unbiased linear-in-y estimating functions, i.e., on linear functions $S_L(x, y, \beta)$ of the same form as S_L^* but with $\mathbb{E}S_L = 0$ and not necessarily $\mathbb{E}(S_L|x) = 0$. The special class of conditionally unbiased estimating functions is characterized by the property that $pm = q$, which gives the function S_L^* the special form

$$S_L^* = (y - m)g.$$

We called such a function a simple (linear) score function. By proving that a linear score estimator is less efficient than a simple score estimator, we can extend Heyde's optimality result (for the quasi-score estimator) to the more general class of linear score estimators (Theorem 1a)). In addition we show that QS has the same efficiency as LS if and only if the two estimators coincide. We also give sufficient conditions under which QS is strictly more efficient than LS (Theorem 2).

Mean-variance models sometimes have an additional (dispersion) parameter φ in the variance function, which then is to be denoted by $v(x, \beta, \varphi)$. To estimate φ, one can use various unbiased estimating functions $S_\varphi(x, y, \beta, \varphi)$, e.g., $S_\varphi = (y - m)^2 - v$. Formula (7) for Σ_Q remains, however, unchanged and all the results of this paper hold true even in this more general case.

Linear score estimators appear naturally in the context of measurement error models. The so-called *corrected score* (CS) *estimator* is a linear score estimator. Thus we have as a Corollary to Theorem 1 that CS is less efficient that QS.

Appendix

Proof of the Lemma:

Let

$$\begin{pmatrix} I & P^\top \\ P & Q \end{pmatrix} \geq 0. \qquad (A1)$$

This implies that for any k-vector x_1 and m-vector x_2

$$x_1^\top x_1 + x_1^\top P^\top x_2 + x_2^\top P x_1 + x_2^\top Q x_2 \geq 0. \qquad (A2)$$

Set $x_1 = -P^\top x_2$, then the l.h.s. of (A2) becomes

$$x_2^\top P P^\top x_2 - 2x_2^\top P P^\top x_2 + x_2^\top Q x_2 = x_2^\top (Q - P P^\top) x_2 \geq 0$$

for all x_2. Hence

$$Q - P P^\top \geq 0.$$

Conversely, $Q \geq P P^T$ implies

$$\begin{aligned} & x_1^\top x_1 + x_1^\top P^\top x_2 + x_2^\top P x_1 + x_2^\top Q x_2 \\ \geq\ & x_1^\top x_1 + x_1^\top P^\top x_2 + x_2^\top p x_1 + x_2^\top P P^\top x_2 \\ =\ & (x_1^\top + x_2^\top P)(x_1^\top + x_2^\top P)^\top \geq 0, \end{aligned}$$

which implies (A1).
Now suppose the $\geq$ sign in (A1) is replaced with the $>$ sign. Then by the same argument for any $x_2 \neq 0$,

$$x_2^\top (Q - P P^\top) x_2 > 0$$

and hence

$$Q - P P^\top > 0.$$

5 References

- Carroll, R.J., Ruppert, D., and Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models.* Chapman and Hall, London.
- Heyde, C.C. (1997), *Quasi-Likelihood And Its Application.* Springer, New York.

- Kukush, A. and Schneeweiss, H. (2005), Comparing different estimators in a nonlinear measurement error model, I. *Mathematical Methods of Statistics* **14**, 53-79.

- Kukush, A., Schneeweiss H., and Shklyar, S. (2005), *Quasi score is more efficient than Corrected Score in a general nonlinear measurement error model* Discussion Paper 451, Sonderforschungsbereich 386 University of Munich.

- Wedderburn, R.W.M. (1974), Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrica* **61**, 439-447.