L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

L_1-tests in linear models: Tests with maximum relative power

Christine Müller

Free University of Berlin, Germany

Abstract: In a linear model $Y_{nN} = x_{nN}^T\beta + Z_{nN}$, $n = 1, ..., N$, the problem of testing the hypothesis $H_0 : L\beta = l$ versus $H_1 : L\beta \neq l$ is considered. As tests Wald-type tests based on asymptotically linear estimators are used. For such tests the asymptotic efficiency at the ideal model and the asymptotic bias caused by outliers or other deviations from the ideal model depend only on the influence function of the underlying estimator. As for estimation most efficient robust tests can be found by maximizing the efficiency under the side condition that the bias is bounded by some bias bound b. But this has the disadvantage that the solutions depend on the bias bound b. To determine b one can regard measures which are composed by the efficiency and the bias. For estimation such measure is the mean squared error while for testing the power relative to the bias is used. It is shown that the L_1-tests, i.e. Wald-type tests based on the L_1-estimator, maximize this relative power. This result is in opposition to that for estimation where the L_1-estimators do not maximize the mean squared error.

Key words: Linear model, L_1-test, bias of the level, relative power.

AMS subject classification: Primary 62F35 ; secondary 62J05, 62J10, 62K05.

1 Introduction

A general linear model

$$Y_N = X_N\beta + Z_N$$

is considered, where $Y_N = (Y_{1N}, \ldots, Y_{NN})^T$ is the vector of observations, $\beta \in I\!R^r$ an unknown parameter vector, $X_N = (x_{1N}, \ldots, x_{NN})^T \in I\!R^{N\times r}$ the known design matrix with regressors $x_{1N}, \ldots, x_{NN} \in I\!R^r$ and $Z_N =$

$(Z_{1N},\ldots,Z_{NN})^T$ the vector of errors. A realization of the random vector Y_N is denoted by $y_N = (y_{1N},\ldots,y_{NN})^T$. In this linear model a hypothesis of the general form

$$H_0 : L\beta = l$$

shall be tested versus the alternative

$$H_1 : L\beta \neq l,$$

where $l \in I\!R^s$ and $\varphi(\beta) = L\beta$ is a linear aspect of the unknown parameter vector β with given matrix $L \in I\!R^{s\times r}$ of rank s.

A large class for testing the hypothesis $H_0 : L\beta = l$ is the class of Wald-type tests based on asymptotically linear estimators, briefly called ALE-tests (see Müller, 1992a,b, 1995a,b; Rieder, 1994, p.153). To define these tests we assume that the ideal distribution of the standardized errors Z_{nN}/σ is P and that the design $x_{1N},\ldots,x_{NN}$ is converging to an asymptotic design measure δ in the following sense:

$$\lim_{N\to\infty} \frac{1}{N} \sum e_{x_{nN}}(\{x\}) = \delta(\{x\})$$

for all $x \in \text{supp}(\delta)$, where $\text{supp}(\delta)$ is the support of δ and e_x is the Dirac measure on $x \in I\!R^r$. Then the ALE-tests have a test statistic of the form

$$\tau_N(y_N, X_N) = N\,(\widehat{\varphi}_N(y_N, X_N) - l)^T C_N(y_N, X_N)^{-1}(\widehat{\varphi}_N(y_N, X_N) - l),$$

where $\widehat{\varphi}_N$ is an asymptotically linear estimator for $\varphi(\beta) = L\beta$ with influence function ψ and $C(y_N, X_N)$ is a consistent estimator for the asymptotic covariance matrix of $\widehat{\varphi}_N$, i.e. of $\sigma^2 C(\psi,\delta)$ with

$$C(\psi,\delta) := \int \psi(z,x)\,\psi(z,x)^T\,P(dz)\,\delta(dx).$$

Thereby an estimator $\widehat{\varphi}_N$ for $\varphi(\beta) = L\beta$ is called asymptotically linear with influence function $\psi : I\!R \times I\!R^r \to I\!R^s$ if $\int |\psi(z,x)|^2\,P(dz)\,\delta(dx) < \infty$, $\int \psi(z,x)\,P(dz) = 0$ for all $x \in \text{supp}(\delta)$, $\int \psi(z,x)\,x^T\,z\,P(dz)\,\delta(dx) = L$ and

$$\lim_{N\to\infty} P^N_{\beta_N}$$

$$-\frac{\sigma}{N}\sum\nolimits_{n=1}^{N} \psi\left(\frac{y_{nN} - x_{nN}^T\beta_N}{\sigma}, x_{nN}\right)\Bigg| > \epsilon\Bigg\}\Bigg) = 0$$

for all $\epsilon > 0$, $\sigma \in I\!R^+$ and $\beta_N = \beta + N^{-1/2}\overline{\beta}$ with $\beta, \overline{\beta} \in I\!R^r$ and $L\beta = l$. The set of all influence functions is denoted by Ψ. Many wellknown estimators as M-estimators and R-estimators are asymptotically linear.

In particular the classical F-test is an ALE-tests, where $\widehat{\varphi}_N$ is the Gauss-Markov estimator and $\psi(z,x) = L\,I(\delta)^{-} x\, z$. Thereby M^{-} denotes the generalized inverse of the matrix M, i.e. $M M^{-} M = M$, and $I(\delta)$ the information matrix, i.e.

$$I(\delta) = \int x\, x^T \delta(dx).$$

If $\psi(z,x) = \psi^1(z,x) := L\,I(\delta)^{-} x \operatorname{sgn}(z) \sqrt{\frac{\pi}{2}}$ then it is the influence function of the L_1-estimator and the corresponding ALE-test is called L_1-test.

If the standardized errors $Z_{1N}/\sigma, \ldots, Z_{NN}/\sigma$ are independent and identically distributed according to the ideal distribution P, then under H_0 the ALE-test statistic τ_N has asymptotically a central chi-squared distribution with s degrees of freedom. Hence, the critical value of an asymptotic level α ALE-test can be determined as the $(1-\alpha)$ quantile of the chi-squared distribution. Under contiguous alternatives of the form $\beta_N = \beta + N^{-1/2}\overline{\beta} \in \mathbb{R}^r$ with $L\beta_N = l + N^{-1/2}\gamma$ the ALE-test statistic has asymptotically a chi-squared distribution with noncentrality parameter $\gamma^T[\sigma^2\, C(\psi,\delta)]^{-1}\gamma$ so that the power of the test is an increasing function of $\gamma^T C(\psi,\delta)^{-1}\gamma$.

If the errors $Z_{1N}/\sigma, \ldots, Z_{NN}/\sigma$ have instead of the ideal distribution P distributions which are contaminated by outliers and other deviations, then under H_0 the asymptotic error probability can exceed the level α. The maximum bias of the level, which is possible under contaminated distributions, is an increasing function of

$$\|\psi^T C(\psi,\delta)^{-1}\psi\|_\delta := \max_{(z,x)\in \mathbb{R}\times \mathrm{supp}(\delta)} \psi(z,x)^T C(\psi,\delta)^{-1}\psi(z,x) \qquad (1)$$

See Müller (1992a,b, 1995a,b), Rieder (1994), Heritier and Ronchetti (1994).

Let the ideal distribution P be the standard normal distribution. In Müller (1995a,b) the question was considered which ALE-test has maximum power under the side condition that the maximum bias is bounded by some bias bound. Expressed by influence functions this means: Which $\psi \in \Psi$ maximizes $\gamma^T C(\psi,\delta)^{-1}\gamma$ for all $\gamma \in \mathbb{R}^s$ under the side condition

$$\|\psi^T C(\psi,\delta)^{-1}\psi\|_\delta \leq b, \qquad (2)$$

where b is some given bias bound. But this means that the matrix $C(\psi,\delta)^{-1}$ should be maximized in the positive definite sense under the side condition (2). In general this optimization problem has no solution (see Krasker and Welsch, 1982). To find solutions one can regard instead of the whole matrix $C(\psi,\delta)^{-1}$ functionals of the matrix. An appropriate functional for testing is the determinant of the matrix. In Müller (1995a,b) it was shown that for maximizing $\det(C(\psi,\delta)^{-1})$ under the side condition (2) solutions can be derived if the design δ is D-optimal, i.e. $\delta \in \arg\min\{\det(L\,I(\tilde{\delta})^{-}L^T);\ \tilde{\delta} \in \Delta\}$. But the solutions depend on the bias bound b so that the question exists how to choose b. Here in Section 2 we propose a criterion for choosing b based on a relative power value. We show that the best ALE-tests with respect to this criterion is the L_1-test. In Section 3 we compare this result with a corresponding result for estimation, and in Section 4 we give an example.

2 Tests with maximum relative power

Müller (1995a,b) showed that the influence function ψ_b given by

$$\psi_b(z,x) := \begin{cases} L\,I(\delta)^{-}x\,\mathrm{sgn}(z)\,\sqrt{\frac{\pi}{2}}, & \text{for } b = s, \\ L\,I(\delta)^{-}x\,\mathrm{sgn}(z)\,\frac{\min\{|z|,\sqrt{b}\,y_b\}}{2\Phi(\sqrt{b}\,y_b)-1}, & \text{for } b > s, \end{cases}$$

with

$$y_b^2 = \frac{1}{s}\,g(\sqrt{b}\,y_b) > 0$$

is a solution of the problem of maximizing the power criterion $\det(C(\psi,\delta)^{-1})$ under the bias side condition (2) if δ is a D-optimal design. Thereby Φ denotes the distribution function of the standard normal distribution and g is given by

$$g(y) := \int \min\{|z|, y\}^2 P(dz).$$

Note that ψ_b with $b = s$ is the influence function ψ^1 of the L_1-test and that $s = b_{\min} := \min\{\|\psi^T C(\psi,\delta)^{-1}\psi\|_\delta;\ \psi \in \Psi\}$. Recall that s is the rank of the matrix $L \in I\!R^{s\times r}$.

For every solution ψ_b we have that the quantity (1) providing the maximum bias satisfies

$$\|\psi_b^T C(\psi_b,\delta)^{-1}\psi_b\|_\delta = b$$

and that the power criterion satisfies

$$\det(C(\psi_b,\delta)^{-1}) = \frac{\left(u\left(\frac{b}{s}\right)\right)^s}{\det(L\,I(\delta)^{-}L^T)}. \tag{3}$$

Thereby $u : [1,\infty) \to (0,\infty)$ is defined by

$$u(a) = \begin{cases} \frac{(2\Phi(\tilde{w}(a))-1)^2}{g(\tilde{w}(a))} & \text{for } a > 1, \\ \frac{2}{\pi} & \text{for } a = 1, \end{cases}$$

where $\tilde{w}(a) > 0$ is implicitly given by $a\,g(\tilde{w}(a)) - \tilde{w}(a)^2 = 0$. In particular we have $\tilde{w}(\frac{b}{s}) = \sqrt{b}\,y_b$. With the implicit function theorem it can be shown that u is an increasing function (see Müller, 1995a). This means that the power of the ALE test based on ψ_b increases when the maximum bias value b increases, and vice versa.

For an appropriate choice of the bias bound b we can set the power value $\det(C(\psi_b,\delta)^{-1})$ in relation to the maximum bias given by (1). There are in principle two possibilities: b should be chosen so that the difference between the power value and the bias value is maximized, or b should be chosen so that the ratio of the the power value and the bias value is maximal. Maximizing the difference

$$\det(C(\psi_b,\delta)^{-1}) - \|\psi_b^T C(\psi_b,\delta)^{-1}\psi_b\|_\delta = \frac{\left(u\left(\frac{b}{s}\right)\right)^s}{\det(L\,I(\delta)^- L^T)} - b \tag{4}$$

has the disadvantage that the solution would depend on the formulation of the hypotheses. Namely, if we use instead of the hypothese $H_0 : L\beta = l$ the equivalent hypotheses $H_0 : \lambda\, L\beta = \lambda\, l$ with $\lambda \neq 1$, then we have to use $\tilde{\psi}_b := \lambda\,\psi_b$ instead of ψ_b. While the bias value (1) is invariant with respect to λ, this is not the case for the power value so that a solution of maximizing a difference like (4) would depend on λ. This problem does not appear if we use the ratio of the power value and the bias value. Hence, b should be chosen so that the relative power value

$$\frac{\det(C(\psi_b,\delta)^{-1})^{1/s}}{\|\psi_b^T C(\psi_b,\delta)^{-1}\psi_b\|_\delta} = \frac{1}{b\,\cdot\,\det(C(\psi_b,\delta))^{1/s}} \tag{5}$$

is maximized. Thereby we take the sth root of the determinant of the covariance matrix to ensure that an improvement of the covariance matrix by a factor c provides also an improvement of the relative power value by the factor c. Note also that the sth root of the determinant is often used as a measure for the entropy and that it is the geometric mean of the diagonal elements if the covariance matrix is a diagonal matrix.

The following theorem shows that the L1-test, i.e. the ALE-test based on ψ_b with $b = s$, has the maximum relative power.

Theorem 1 *$b = s$ maximizes the relative power value (5) with respect to b, i.e. the L_1-test has maximum relative power.*

Proof: Using (3) maximizing of (5) is equivalent to minimizing

$$b\,\cdot\,\det(C(\psi_b,\delta))^{1/s} = b\,\cdot\,\frac{1}{u\left(\frac{b}{s}\right)}\,\det(L\,I(\delta)^- L^T)^{1/s}$$

$$= t\left(\frac{b}{s}\right)\cdot s\,\cdot\,\det(L\,I(\delta)^- L^T)^{1/s}$$

where

$$t(a) := a \, \frac{1}{u\,(a)} \text{ for } a \geq 1.$$

By calculating the first and second derivative of $t : [1, \infty) \to (0, \infty)$ it can be shown that t is a convex function with $\lim_{a \downarrow 1} t'(a) = 0$, where t' denotes the first derivative of t. This means that $t\left(\frac{b}{s}\right)$ is minimized by $b = s$. To calculate the first and the second derivative of t it is helpful to calculate the derivatives of u. At first note that the implicit function theorem provides

$$\tilde{w}'(a) = \frac{\tilde{w}(a)\; g(\tilde{w}(a))}{2\,a\,[2\Phi(\tilde{w}(a)) - 1 - 2\,\tilde{w}(a)\,\Phi'(\tilde{w}(a))]} > 0.$$

Then, with $h(y) := y\,\Phi(-y) - \Phi'(y)$, it can be shown that

$$\begin{aligned} u'(a) &= \frac{-(2\Phi(\tilde{w}(a)) - 1)\,2\,h(\tilde{w}(a))}{\tilde{w}(a)} > 0, \\ u''(a) &< 0 \quad \text{and} \\ u''(a)\,a + 2\,u'(a) &\leq 0 \end{aligned}$$

is satisfied for all $a > 1$ (see Müller, 1995a, Lemma 12.7). The rule of L'Hospital provides $\lim_{a \downarrow 1} \tilde{w}(a) = 0$ and $\lim_{a \downarrow 1} u(a) = \frac{2}{\pi}$ so that

$$t'(a) = \frac{1}{u(a)} \left(1 + \frac{2\,h(\tilde{w}(a))\,\tilde{w}(a)}{2\Phi(\tilde{w}(a)) - 1}\right)$$

is converging to 0 for $a \downarrow 1$ (see Müller 1995a, Lemma 12.8). □

3 Comparison with estimation problems

An asymptotically linear estimator $\widehat{\varphi}_N$ for estimating $\varphi(\beta) = L\beta$ with influence function ψ is under contaminated distributions asymptotically normally distributed. The asymptotic covariance matrix is $C(\psi, \delta)$ and the maximum asymptotic bias is given by

$$\|\psi\|_\delta := \max_{(z,x) \in I\!R \times \operatorname{supp}(\delta)} |\psi(z, x)| \tag{6}$$

(see Bickel, 1981, 1984; Rieder, 1985, 1987, 1994). Solutions ψ_b^e which minimize the trace of the covariance matrix $C(\psi, \delta)$ under the bias side condition $\|\psi\|_\delta \leq b$ can be characterized explicitly if the design δ is based on linearly independent regressors or if the design δ is A-optimal, i.e. $\delta \in \arg\min\{\operatorname{tr}(L\,I(\tilde{\delta})^- L^T);\ \tilde{\delta} \in \Delta\}$ (see Kurotschka and Müller, 1992; Müller, 1994a). As for testing the optimal influence function ψ_b^e depends heavily on the bias bound b. Moreover, only the solution ψ_b^e with $b = b_{\min}^e := \min\{\|\psi\|_\delta;\ \psi \in \Psi\}$ corresponds to the L$_1$-estimator, i.e. satisfies $\psi_b^e = \psi^1$.

Similarly to testing an optimal bias bound can be found by combining the efficiency criterion $\mathrm{tr}(C(\psi_b^e, \delta))$ with the bias value $\|\psi_b^e\|_\delta$. A natural combined criterion is the asymptotic mean squared error

$$M_\delta(b) := \|\psi_b^e\|_\delta^2 + \mathrm{tr}(C(\psi_b^e, \delta)) = b^2 + \mathrm{tr}(C(\psi_b^e, \delta)). \tag{7}$$

It can be shown that as for testing M_δ is convex. But in opposition to testing M_δ attains its minimum for a value $b > b_{\min}^e$ so that the L$_1$-estimator does not minimize the mean squared error. See Müller (1994b,c).

Hence, we have the following situation: The approaches for estimation and testing look very similar and leads to similar constrained optimization problems of maximizing the efficiency under a bias bound b. Nevertheless the problem of finding an optimal bias bound by using a natural criterion which combines efficiency and bias leads to qualitative different results. For testing the best bias bound is $b = b_{\min} = s$ so that the L$_1$-test is optimal while for estimation the best bias bound is $b > b_{\min}^e$ so that the L$_1$-estimator is not optimal. Moreover, for estimation the optimal bias bound depends strongly on the model, the design and the aspect $L\beta$ and it can be calculated only per computer. For testing the optimal bias bound is simply s, the rank of L.

4 Example

Consider a one-way lay-out model with four levels. i.e. we have four samples with unknown means β_1, β_2, β_3, β_4 so that the observations are given by

$$Y_{nN} = \beta_i + Z_{nN},$$

if the observation Y_{nN} belongs to the sample i, $i = 1, 2, 3, 4$. This model can be expressed as a linear model with $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T \in I\!R^4$ and

$$x_{nN} = x(i) := (1_1(i), 1_2(i), 1_3(i), 1_4(i))^T.$$

Often one sample, say sample 1, is a controll group. Then an interesting aspect of β is the linear aspect $\varphi(\beta) = (\beta_2 - \beta_1, \beta_3 - \beta_1, \beta_4 - \beta_1)^T \in I\!R^3$. Then testing $H_0 : \varphi(\beta) = 0$ against $H_1 : \varphi(\beta) \neq 0$ is equivalent with testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$. A D-optimal design for $\varphi(\beta)$ is $\delta = \frac{1}{4}(e_{x(1)} + e_{x(2)} + e_{x(3)} + e_{x(4)})$ which means that the four samples are of equal size. At this design the influence function of the L$_1$-test for H_0 and the L$_1$-estimator for $\varphi(\beta)$ has the form

$$\psi^1(z, x(i)) = \begin{cases} (-1, -1, -1)^T \operatorname{sgn}(z)\, 4\sqrt{\frac{\pi}{2}}, & \text{for } i = 1, \\ (1_2(i), 1_3(i), 1_4(i))^T \operatorname{sgn}(z)\, 4\sqrt{\frac{\pi}{2}}, & \text{for } i \neq 1. \end{cases}$$

According to Theorem 1 this influence function provide the maximum relative power within all ALE-tests for H_0. But this influence function does not provide the minimum mean squared error within all asymptotically linear estimators for $\varphi(\beta)$. The influence function providing the minimum mean squared error has the form

$$\psi_b^e(z, x(i)) = \begin{cases} (-1,-1,-1)^T \operatorname{sgn}(z) \frac{\min\{|z|, b\, w_b\}}{w_b \sqrt{3}}, & \text{for } i = 1, \\ (1_2(i), 1_3(i), 1_4(i))^T \operatorname{sgn}(z) \frac{\min\{|z|, b\, v_b\}}{v_b}, & \text{for } i \neq 1, \end{cases}$$

where $b \approx 8.7213$, $w_b \approx 0.0186$ and $v_b \approx 0.2411$ (see Müller, 1994c).

References

[1] Bickel, P.J. (1981). Quelque aspects de la statistique robuste. In *Ecole d'Eté de Probabilités de St. Flour*, Springer Lecture Notes in Math, Vol. 876, pp. 1-72.

[2] Bickel, P.J. (1984). Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12**, 1349-1368.

[3] Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Am. Statist. Assoc* **89**, 897-904.

[4] Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded-influence regression estimation. *J. Am. Statist. Assoc.* **77**, 595-604.

[5] Kurotschka, V. and Müller, Ch.H. (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Statist.* **20**, 331-350.

[6] Müller, Ch.H. (1992a). L_1-estimation and testing in conditionally contaminated linear models. In *L_1-Statistical Analysis and Related Methods*, Ed. Y. Dodge, pp. 69-76. Amsterdam: North-Holland.

[7] Müller, Ch.H. (1992b). Robust testing in conditionally contaminated linear models. Preprint No. A-92-12, Freie Universität Berlin, Fachbereich Mathematik und Informatik.

[8] Müller, Ch.H. (1994a). Optimal designs for robust estimation in conditionally contaminated linear models. *J. Statist. Plann. Infer.* **38**, 125-140.

[9] Müller, Ch.H. (1994b). Optimal bias bounds for robust estimation in linear models. In *Proc. of the International Conference on Linear Statistical Inference LINSTAT'93*, Eds. T. Caliński and R. Kala, pp. 97-102. Dordrecht: Kluwer Academic Publishers.

[10] Müller, Ch.H. (1994c). On the calculation of MSE minimizing robust estimators. In *COMPSTAT'94, Proceedings in Computational Statis-*

tics, Eds. R. Dutter and W. Grossmann, pp. 257-262. Heidelberg: Physica-Verlag.

[11] Müller, Ch.H. (1995a). Outlier robust inference for planned experiments. Habilitation thesis, Freie Universität Berlin. To appear as Springer Lecture Notes in Statistics.

[12] Müller, Ch.H. (1995b). Optimum robust testing in linear models. Preprint No. A-28-95, Freie Universität Berlin, Fachbereich Mathematik und Informatik. Submitted to *Ann. Statist.*

[13] Rieder, H. (1985). Robust estimation of functionals. Technical Report, Universität Bayreuth.

[14] Rieder, H. (1987). Robust regression estimators and their least favorable contamination curves. *Stat. Decis.* **5**, 307-336.

[15] Rieder, H. (1994). *Robust Asymptotic Statistics.* New York: Springer.