REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 103–118

LIKELIHOOD RATIO TESTS IN LINEAR MODELS WITH LINEAR INEQUALITY RESTRICTIONS ON REGRESSION COEFFICIENTS

Authors: Miguel Fonseca
– Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
fmig@fct.unl.pt

João Tiago Mexia
– Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
jtm@fct.unl.pt

Bimal K. Sinha
– Department of Mathematics and Statistics, University of Maryland Baltimore County, U.S.A.
sinha@umbc.edu

Roman Zmyślony
– Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Poland
r.zmyslony@wmie.uz.zgora.pl

Received: August 2012 Revised: October 2013 Accepted: November 2013

Abstract:

- This paper develops statistical inference in linear models, dealing with the theory of maximum likelihood estimates and likelihood ratio tests under some linear inequality restrictions on the regression coefficients. The results are widely applicable to models used in environmental risk analysis and econometrics.

Key-Words:

- *likelihood ratio test; linear constrains; regression models.*

AMS Subject Classification:

- 62J05, 62F30.

1. INTRODUCTION

Inference about the regression coefficients in a standard linear regression model under the usual assumptions of normality, independence and homoscedasticity of errors and without any constraints on the regression parameters is quite old. A good amount of research has also been done in this set up under some linear inequality constraints on the regression coefficients (see Liew (1976), Gourieroux *et al.* (1982), Self and Liang (1985), Mukerjee and Tu (1995), Andrews (1999), Andrews (2001), Meyer (2003) and Kopylev and Sinha (2010)). Most of the discussions in these papers are asymptotic in nature, and also under the assumption that the underlying dispersion matrix of errors is either completely known or asymptotically estimated and hence used as if it were known. It turns out that, under linear inequality constraints on the regression coefficients, quite often the null distribution of the likelihood ratio test statistic for the nullity of a regression coefficient is a linear combination of several independent chisquares rather than being just one chisquare. We add that the proofs in some papers are geometric in nature while in others it is algebraic in nature, but they are quite involved in both the cases due to the very nature of the model and the testing problem.

A brief literature review is in order. The first paper on this topic seems to be due to Gourieroux *et al.* (1982), followed by the celebrated paper by Self and Liang (1985). The emphasis in both the papers is the derivation of the asymptotic properties of the maximum likelihood estimates and the associated LRT when some parameters lie on their boundaries. In an excellent paper by Mukerjee and Tu (1995), the exact small sample LRT is derived and its properties have been studied in the special case of a simple linear regression model with the nonnegativity restriction on both the intercept and the slope parameters, and inference being on an arbitrary linear function of the two parameters. The paper by Meyer (2003) discusses a test for linear regression versus convex regression while Kopylev and Sinha (2010), primarily motivated by Self and Liang (1985), develop explicit and useful expressions of the MLEs and LRTs in dimensions two and three, the entire treatment being asymptotic in nature.

In this paper we revisit this important inference problem in the case of a standard linear regression model with some linear inequality constraints on the regression coefficients and develop the LRT for the nullity of just one linear function when the variance is unknown. Our treatment is exact, and we offer two solutions. This is in the same spirit as in Mukerjee and Tu (1995). The paper is organized as follows. In Section 2 we consider the linear regression problem with two regression coefficients, both being nonnegative, and derive the LRT for the nullity of one of them. In Section 3 we consider the case of a linear regression with three regression coefficients, all of which are nonnegative, and describe the LRT for the nullity of one of them. In both the settings, normality and independence of errors with an unknown variance are assumed. In each case we derive the likelihood ratio test and discuss some aspects of the corresponding null distribution of the LRT. Results of some simulation studies are reported in Section 4 in the case of two regression coefficients, comparing the Type I errors of the usual LRT (without taking into account any correction due to nonnegativity of regression coefficients) and the proposed LRT, clearly showing the benefit of the corrections. Such benefits have also been observed and reported in Mukerjee and Tu (1995).

We end this section with a general observation that in the context of a linear model

$$\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{W}) \,, \tag{1.1}$$

if there are known linear inequality constraints on the regression coefficients $\boldsymbol{\beta}$, and the inference problem is to test the equality of such a linear constraint versus it is bigger (or smaller), under a suitable (known) matrix transformation we can always assume without any loss of generality that the inequality constraints as well as the testing problem depend solely on the regression coefficients themelves. This is precisely the formulation we adopt in the remainder of the paper. We also observe an important point from Self and Liang (1985) and Kopylev and Sinha (2010). Under normality and independence of errors, the maximization of the likelihood with respect to the entire regression coefficients $\boldsymbol{\beta}$, which is is equivalent to the minimization of the familiar normal quadratic form $(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\,\mathbf{V}\,(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})$ with respect to the regression coefficients $\boldsymbol{\beta}$, where $\mathbf{V}$ is the estimates covariance matrix, can be safely carried out only with respect to the subset of the regression coefficients which satisfy the inequality constraints, thus completely ignoring the minimization aspect with respect to the unrestricted regression coefficients. Hence, although our proposed solutions in this paper are derived for linear regression models with two nonnegative regression coefficients, this formulation can be adapted for any number of unrestricted regression coefficients!

These results can be very useful in econometrics, extending, for example, the results of Andrews (1999) and Andrews (2001). Similar methodologies can also be applied in environmental risk analysis, as it can be seen in Sinha, Kopylev and Fox (2012).

2. TWO REGRESSION COEFFICIENTS ON THE BOUNDARY

2.1. Model

Consider the linear model

$$\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{W}) \,, \tag{2.1}$$

with unknown parameters $\boldsymbol{\beta}$ and σ^2, and known matrices $\mathbf{X}$ and $\mathbf{W}$. Then the usual maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and σ^2 are given by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}y \,, \tag{2.2}$$

$$\tilde{\sigma}^2 = \frac{S}{n} \,, \qquad S = (\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{W}^{-1}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}) \,. \tag{2.3}$$

We assume without any loss of generality that $\boldsymbol{\beta} = (\beta_1, \beta_2)$, and derive below the likelihood ratio test (LRT) statistic for the hypothesis

$$H_0\colon \beta_1 = 0 \;\text{ vs. }\; H_1\colon \beta_1 > 0 \,; \qquad \beta_1, \beta_2 \geq 0 \,. \tag{2.4}$$

As mentioned earlier, we point out that any linear regression model with a linear inequality constraint on the original regression coefficients and a linear hypothesis on another linear function of them can be reduced to the above setup by suitable linear transformations of $\underset{\sim}{y}$. Let $\mathbf{V}_{\boldsymbol{\psi}}$ be the inverse of the Fisher information matrix for $\boldsymbol{\psi} = [\beta_1 \; \beta_2 \; \sigma^2]$, which will have the form

$$\mathbf{V}_{\boldsymbol{\psi}} = \begin{bmatrix} \sigma^2(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{n} \end{bmatrix} . \tag{2.5}$$

All throughout we assume that σ^2 is unknown, and we proceed in two ways to develop a test for H_0. Our first approach is based on taking σ^2 to be known and deriving an LRT for H_0, and then replacing σ^2 by its natural estimate, namely, the sample residual variance, and checking what kind of properties the resultant test statistic would possess. This is done by extensive simulation carried out in Section 4. The second approach is to derive the *genuine* LRT when σ^2 is unknown. Although the latter test statistic has an explicit form, its null distribution is rather complicated. We study its properties again by simulation in Section 4. A point of caution is in order here. Unlike the asymptotic treatments in Self and Liang (1985) and Kopylev and Sinha (2010), the null distributions of the test statistics in both the above cases depend on the nuisance parameter β_2 (in fact, via β_2/σ). This is in sharp contrast with the contents of *all* the previous papers!

2.2. σ^2 known

The derivation of the LRT in this case essentially follows from Kopylev and Sinha (2010) who derived it algebraically. We provide below an alternative proof using some geometrical arguments. Following the results presented in [5], the LRT statistic for known σ will have the exact form:

$$L = \min_{\boldsymbol{\beta}\in C}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}')(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \min_{\boldsymbol{\beta}\in C_0}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}')(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) \,, \tag{2.6}$$

where C is the cone represented by $\mathbb{R}_0^+ = \prod_{i=1}^2 [0,+\infty[$ and C_0 is the cone represented by $\{0\} \times [0,+\infty[$.

Upon simplification, we get an equivalent expression for L:

$$\begin{aligned} L_0 \;=\; & \min_{\beta_1,\beta_2\geq 0}\Big[v_{11}(\beta_1-\tilde\beta_1)^2 - 2\,v_{12}(\beta_1-\tilde\beta_1)(\beta_2-\tilde\beta_2) + v_{22}(\beta_2-\tilde\beta_2)^2\Big] \\ & - \min_{\beta_2\geq 0}\Big[v_{11}\tilde\beta_1^2 - 2\,v_{12}\tilde\beta_1(\beta_2-\tilde\beta_2) + v_{22}(\beta_2-\tilde\beta_2)^2\Big] \end{aligned} \tag{2.7}$$

where v_{11}, v_{12} and v_{22} come from

$$\mathbf{V} \;=\; (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \;=\; \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}. \tag{2.8}$$

Note that dividing the estimators $\tilde\beta_1$ and $\tilde\beta_2$ as well as the parameters β_1 and β_2 by $\sqrt{v_{11}}$ and $\sqrt{v_{22}}$, respectively, we can rewrite L_0 as

$$\begin{aligned} \ell \;=\; & \min_{\theta_1,\theta_2\geq 0}\Big[(\theta_1-\tilde\theta_1)^2 - 2\,\rho(\theta_1-\tilde\theta_1)(\theta_2-\tilde\theta_2) + (\theta_2-\tilde\theta_2)^2\Big] \\ & - \min_{\theta_2\geq 0}\Big[(1-\rho^2)\,\tilde\theta_1^2 + (\theta_2-\tilde\theta_{2.1})^2\Big] \end{aligned} \tag{2.9}$$

where $\boldsymbol\theta = \mathrm{diag}\Big(\frac{1}{\sqrt{v_{11}}}, \frac{1}{\sqrt{v_{22}}}\Big)\boldsymbol\beta$ and $\tilde{\boldsymbol\theta} = \mathrm{diag}\Big(\frac{1}{\sqrt{v_{11}}}, \frac{1}{\sqrt{v_{22}}}\Big)\tilde{\boldsymbol\beta}$, and $\tilde\theta_{2.1} = \tilde\theta_2 - \rho\tilde\theta_1$. It is easy to see that the hypotheses H_0 and H_1 remain invariant under this transformation.

2.3. Minimization

Let us assume that $\rho > 0$ and start with the minimization of $Q(\theta_1,\theta_2) = (\theta_1-\tilde\theta_1)^2 - 2\,\rho(\theta_1-\tilde\theta_1)(\theta_2-\tilde\theta_2) + (\theta_2-\tilde\theta_2)^2$. When $\min\{\tilde\theta_1,\tilde\theta_2\} < 0$ and $\tilde\theta_2 \geq \rho\tilde\theta_1$, putting

$$\begin{cases} x \;=\; \theta_1 - \tilde\theta_1 \\ y \;=\; \theta_2 - \tilde\theta_2 \end{cases}, \tag{2.10}$$

the level curves of the ellipsoid for the d-level curve are given by

$$x^2 - 2\,\rho\,x\,y + y^2 \;=\; d^2 \tag{2.11}$$

and, choosing the positive value,

$$x \;=\; \rho y + \sqrt{d^2 - (1-\rho^2)y^2}\;, \tag{2.12}$$

we get

$$\frac{dx}{dy} \;=\; \rho - \frac{(1-\rho^2)y}{\sqrt{d^2-(1-\rho^2)y^2}}\;. \tag{2.13}$$

To have $\frac{dx}{dy} = 0$, one must have $y = \pm\frac{\rho d}{\sqrt{1-\rho^2}}$. Since the solution we seek is positive, so

$$x = \frac{d}{\sqrt{1-\rho^2}} \; . \tag{2.14}$$

In order that the vertical tangent coincides with the vertical axis one must have $x = -\tilde{\theta}_1$, and so $d = -\tilde{\theta}_1\sqrt{1-\rho^2}$ and $y = -\tilde{\theta}_1\rho$. Thus,

$$\begin{cases} \theta_1 = 0 \\ \theta_2 = \tilde{\theta}_2 - \rho\tilde{\theta}_1 = \tilde{\theta}_{2.1} \end{cases} . \tag{2.15}$$

We also have that $Q(0, \tilde{\theta}_{2.1}) = (1-\rho^2)\tilde{\theta}_1^2$. Under these assumptions, suppose we can attain a value smaller than $(1-\rho^2)\tilde{\theta}_1^2$, say $\left(-\tilde{\theta}_1\sqrt{1-\rho^2}-\epsilon\right)^2$, with $\epsilon > 0$. In that case, the largest value possible for x would be

$$x \;=\; -\tilde{\theta}_1 - \frac{\epsilon}{\sqrt{1-\rho^2}} \; , \tag{2.16}$$

which implies that

$$\theta_1 \;=\; -\frac{\epsilon}{\sqrt{1-\rho^2}} \;<\; 0 \; , \tag{2.17}$$

which is *not* a valid solution for θ_1.

Analogously, when $\min\{\tilde{\theta}_1, \tilde{\theta}_2\} < 0$ and $\tilde{\theta}_2 \le \rho^{-1}\tilde{\theta}_1$ we have $(\theta_1, \theta_2) = (\tilde{\theta}_{1,2}, 0)$ and $Q(\tilde{\theta}_{1.2}, 0) = (1-\rho^2)\tilde{\theta}_2^2$.

On the other hand, when $\rho^{-1}\tilde{\theta}_1 < \tilde{\theta}_2 < \rho\tilde{\theta}_1$, we take $(\theta_1, \theta_2) = (0,0)$ and get $Q(0,0) = \tilde{\theta}_1^2 - 2\,\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2$. If we take any other valid solution, say (ϵ_1, ϵ_2), with $\epsilon_1, \epsilon_2 > 0$, it is easy to see that

$$Q(\epsilon_1, \epsilon_2) - Q(0,0) \;=\; -2\,\tilde{\theta}_{2.1}\epsilon_1 - 2\,\tilde{\theta}_{1.2}\epsilon_2 + \epsilon_1^2 + \epsilon_2^2 - 2\,\rho\epsilon_1\epsilon_2 \;>\; 0 \; , \tag{2.18}$$

and so the optimal solution is in fact $(0,0)$. Summing up the various cases, the first term of ℓ, written as Q_1, simplifies to

$$\frac{Q_1}{\sigma^2(1-\rho^2)} = \begin{cases} 0\,; & \tilde{\theta}_1 > 0, \;\; \tilde{\theta}_2 > 0 \\ \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0, \;\; \tilde{\theta}_1 < 0 \\ \tilde{\theta}_2^2\,; & \tilde{\theta}_{1.2} > 0, \;\; \tilde{\theta}_2 < 0 \\ \dfrac{\tilde{\theta}_1^2 - 2\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}\,; & \tilde{\theta}_{1.2} < 0, \;\; \tilde{\theta}_{2.1} < 0 \end{cases} . \tag{2.19}$$

Under the null hypothesis, *i.e.*, when $\theta_1 = 0$, one gets the likelihood

$$\ell_0 \;=\; (1-\rho^2)\,\tilde{\theta}_1^2 + (\theta_2 - \tilde{\theta}_{2.1})^2 \; . \tag{2.20}$$

It is easy to see that the minimum for this ℓ_0 is achieved when $\theta_2 = \tilde{\theta}_{2.1}$ for $\tilde{\theta}_{2.1} \geq 0$, and $\theta_2 = 0$ for $\tilde{\theta}_{2.1} < 0$. So, the second term of ℓ define Q_0 as

$$\frac{Q_0}{\sigma^2(1-\rho^2)} = \begin{cases} \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0 \\ \dfrac{\tilde{\theta}_1^2 - 2\,\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}\,; & \tilde{\theta}_{2.1} < 0 \end{cases} . \tag{2.21}$$

2.4. Likelihood ratio with known σ^2

Combining the above results, it follows that the LRT rejects H_0 for large values of λ given by

$$\lambda = \begin{cases} \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0, \;\; \tilde{\theta}_1 > 0 \\ \dfrac{\tilde{\theta}_{2.1}^2 + (1-\rho^2)\tilde{\theta}_1^2}{2(1-\rho^2)}\,; & \tilde{\theta}_{2.1} < 0, \;\; \tilde{\theta}_2 > 0 \\ 0\,; & \tilde{\theta}_{2.1} > 0, \;\; \tilde{\theta}_1 < 0 \\ \dfrac{\tilde{\theta}_{1.2}^2}{1-\rho^2}\,; & \tilde{\theta}_{1.2} > 0, \;\; \tilde{\theta}_2 < 0 \\ 0\,; & \tilde{\theta}_{2.1} < 0, \;\; \tilde{\theta}_{1.2} < 0 \end{cases} . \tag{2.22}$$

The above representation of the difference of the minimum of the two quadratic forms is exactly similar to what appears in Kopylev and Sinha (2010). At this point two things need to be settled. First, the null distribution of λ, and then the fact that σ^2 is unknown and it needs to be replaced by an estimate. Since under $H_0\colon \beta_1 = 0$ and $\beta_2 \geq 0$ is unknown, it is obvious that the exact null distribution of our LRT λ will depend on β_2! This is indeed a major difference between our result and that of Kopylev and Sinha (2010) where the argument is asymptotic in nature, resulting in the null distribution of LRT being independent of σ as well as any nuisance parameter. Below we assume that $\beta_2 = 0$ and derive the null distribution of LRT still assuming that σ^2 is known, and then *rescale* λ to take care of unknown σ^2. We will call this the modified LRT. Simulation studies carried out in Section 4 about the Type I error of the modified LRT for unknown β_2 and unknown σ^2 reveal that the performance of the modified LRT is quite good.

Write $V_1 = \tilde{\theta}_1$, $V_2 = \tilde{\theta}_2$, $W_1 = \frac{\tilde{\theta}_{1.2}}{\sqrt{1-\rho^2}}$ and $W_2 = \frac{\tilde{\theta}_{2.1}}{\sqrt{1-\rho^2}}$, and note that under H_0, $V_1 \sim N(0,1)$, $\text{cov}[V_1, W_2] = \text{cov}[V_2, W_1] = 0$, and $V_2 \sim N[\delta, 1]$ with $\delta > 0$.

We now express

$$
\begin{aligned}
\mathbb{P}\big[\lambda < x\big] \;&=\; \mathbb{P}\big[V_1^2 < x \wedge V_1 > 0 \wedge W_2 > 0\big] \\
&+\; \mathbb{P}\left[V_1^2 + W_2^2 < x \wedge W_2 > -\frac{\rho}{1-\rho} V_1 \wedge W_2 > 0\right] \\
&+\; \mathbb{P}\big[W_1^2 < x \wedge V_2 < 0 \wedge W_1 > 0\big] \\
&+\; \mathbb{P}\big[V_1 < 0 \wedge W_1 > 0 \wedge W_2 < 0\big] \,.
\end{aligned} \tag{2.23}
$$

The computation of the above probabilities can be somewhat complicated using Cartesian coordinates. Below we use the familiar polar coordinates.

It is well known that a random two dimensional vector whose components are two independent normal vectors with null mean value and variance σ^2 has the same distribution as the vector

$$
\begin{pmatrix} R \cos U \\ R \sin U \end{pmatrix} \tag{2.24}
$$

where

$$
R \sim \sqrt{\sigma^2 \chi_2^2}\ , \tag{2.25}
$$

$$
U \sim \text{Unif}(0, 2\pi)\ , \tag{2.26}
$$

these variables being independent. In fact, the following equality can be obtained:

$$
\iint\limits_{A} \frac{e^{\frac{x^2+y^2}{2\sigma^2}}}{2\pi\sigma^2}\, dx\, dy \;=\; \iint\limits_{\Lambda} r\, \frac{e^{\frac{r^2}{2\sigma^2}}}{2\pi\sigma^2}\, du\, dr\ , \tag{2.27}
$$

where A is a subset of $\mathbb{R}^2$ and Λ is a subset of $\Omega = [0, \infty[\times[0, 2\pi[$. The polar coordinate transformation

$$
\mathbf{p}(r, u) \;=\; \begin{cases} x = r \cos u \\ y = r \sin u \end{cases} \tag{2.28}
$$

guarantees a bijective function between $\mathbb{R}^2$ and Ω (for $(0, 0)$, take $r = u = 0$).

Hence, applying the polar transformation on the pair $(V_1, W_2) \mapsto (R, U)$ and noting that (V_1, W_2) and (V_2, W_1) are pairs of independent standard normal variables, one can rewrite (2.23) as

$$
\begin{aligned}
\mathbb{P}\left[\frac{\ell}{\sigma^2} < x\right] \;&=\; \mathbb{P}\big[V_1 < 0 \wedge W_1 > 0 \wedge W_2 < 0\big] + \frac{1}{2}\,\mathbb{P}\big[\chi_1^2 < x\big] \\
&+\; \mathbb{P}\big[R^2 < x \wedge 2\pi - \arcsin\rho < U < 2\pi\big]\,, \\
\mathbb{P}\left[\frac{\ell}{\sigma^2} < x\right] \;&=\; \frac{1}{2} - \frac{\arcsin\rho}{2\pi} + \frac{1}{2}\,\mathbb{P}\big[\chi_1^2 < x\big] + \frac{\arcsin\rho}{2\pi}\,\mathbb{P}\big[\chi_2^2 < x\big]\,,
\end{aligned} \tag{2.29}
$$

noting that $\arctan \frac{-\rho}{1-\rho} = \arcsin\rho$. It is then established the following.

Theorem 2.1. *The exact distribution of the LRT statistic under* $H_0 : \theta_1 = 0$ *vs* $H_1 : \theta_1 > 0$, *for* $\theta_2 = 0$ *and known* σ^2 *is a mixture of* χ_0^2, χ_1^2 *and* χ_2^2 *with coefficients* $\frac{1}{2} - p$, $\frac{1}{2}$ *and* p, *respectively, with*

$$p = \frac{\arcsin(\rho)}{2\pi} \ ,$$

where ρ *the correlation coefficient between* $\tilde{\theta}_1$ *and* $\tilde{\theta}_2$.

Up until now, it was assumed that the variance component σ^2 was known, which in practice is rarely the case. Take

$$S_0 = \frac{nS}{n-p} \ , \tag{2.30}$$

where S, defined in (2.3) is the maximum likelihood estimator for σ^2, which is independent of $\tilde{\theta}_1$ and $\tilde{\theta}_2$, and observe that

$$(n-p)S \ \sim \ \sigma^2 \chi^2_{n-p} \ . \tag{2.31}$$

Hence, one can use S_0 to enable the computation of the distribution of ℓ. Thus, taking the expression in (30) and multiplying ℓ by S_0, one gets

$$\begin{aligned} \mathbb{P}\big[\ell < (1-\rho^2)x\big] \ &= \mathbb{P}\big[V_1 < 0 \wedge W_1 > 0 \wedge W_2 < 0\big] + \frac{1}{2}\,\mathbb{P}\big[F_{1,n-p} < (1-\rho^2)x\big] \\ &\quad + \mathbb{P}\left[\frac{R^2}{S_0} < (1-\rho^2)x \ \wedge \ 2\pi - \arcsin\rho < U < 2\pi\right] \ , \\ \mathbb{P}\big[\ell < x\big] \ &= \frac{1}{2} - \frac{\arcsin\rho}{2\pi} + \frac{1}{2}\,\mathbb{P}\big[F_{1,n-p} < (1-\rho^2)x\big] \\ &\quad + \frac{\arcsin\rho}{2\pi}\ \mathbb{P}\big[F_{2,n-p} < (1-\rho^2)x\big] \ . \end{aligned} \tag{2.32}$$

So, another version of Theorem 2.1 for the rescaled or modified LRT is given by

Theorem 2.2. *The exact distribution of the LRT under* $H_0 : \theta_1 = 0$ *vs* $H_1 : \theta_1 > 0$, *for* $\theta_2 = 0$, *is a mixture of* F_0, $F_{1,n-p}$ *and* $F_{2,n-p}$ *with coefficients* $\frac{1}{2} - p$, $\frac{1}{2}$ *and* p, *respectively, with*

$$p = \frac{\arcsin(\rho)}{2\pi} \ ,$$

where ρ *the correlation coefficient between* $\tilde{\theta}_1$ *and* $\tilde{\theta}_2$.

Tables 1, 2 and 3 represent the rejection probability for some values of β_1, taking $\alpha = 0.05$.

2.5. LRT when σ^2 is unknown

In this section we derive the LRT for H_0 when the error variance $\sigma^2 > 0$ is unknown. Write

$$(2.33)\qquad \begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'W^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) &= (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{W}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) + Q(\boldsymbol{\beta}) \\ &= SS_{\text{res}} + Q(\boldsymbol{\beta})\end{aligned}$$

where $\hat{\boldsymbol{\beta}}$ is the usual weighted least squares estimate of $\boldsymbol{\beta}$ defined in (3), and $Q(\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}')(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$. Then it easily follows that the LRT statistic λ for H_0 defined as the ratio of null-restricted maximum of the likelihood of $\mathbf{y}$ to the unrestricted maximum of the same likelihood is given by

$$(2.34)\qquad \lambda = \left[\frac{SS_{\text{res}} + Q_1}{SS_{\text{res}} + Q_0}\right]^{n/2}$$

where Q_1 and Q_0 are the unrestricted and restricted values of the quadratic $Q(\boldsymbol{\beta})$ under the conditions $\beta_1, \beta_2 \geq 0$ and $\beta_1 = 0$, $\beta_2 \geq 0$, respectively.

Using the expressions for Q_1 and Q_0 from (20) and (22), respectively, and noting that our exact LRT rejects H_0 for large values of $\Delta = \frac{SS_{\text{res}}+Q_0}{SS_{\text{res}}+Q_1}$, we simplify Δ as

$$(2.35)\qquad \Delta = \begin{cases} \dfrac{SS_{\text{res}}}{SS_{\text{res}} + \tilde{\theta}_1^2}\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 > 0 \\ \dfrac{SS_{\text{res}}}{SS_{\text{res}} + \frac{\tilde{\theta}_1^2 - 2\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}}\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_2 > 0 \\ 1\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 < 0 \\ \dfrac{SS_{\text{res}} + \tilde{\theta}_2^2}{SS_{\text{res}} + \frac{\tilde{\theta}_1^2 - 2\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}}\,; & \tilde{\theta}_{1.2} > 0, \quad \tilde{\theta}_2 < 0 \\ 1\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_{1.2} < 0 \end{cases}.$$

The *crux* of the problem now is to derive the null distribution of Δ. It is obvious that although the null distribution of Δ is independent of σ^2, it does depend on the unknown second regression coefficient $\beta_2 \geq 0$ as in the previous case. Finding this null distribution even for a specified β_2 turns out to be extremely difficult, and we can present only some simulation results for this purpose.

Tables 7, 8 and 9 represent the rejection probability for some values of β_1 and β_2, taking $\alpha = 0.05$.

Based on these simulation results, we conclude that this test behaves very good, maintaining test size and gaining power compared to the usual F test and the *ad-hoc* test described earlier.

3. NUMERICAL RESULTS

3.1. Ad-hoc test

A set of simulations was performed to evaluate the power performance of the ad-hoc test when σ^2 is unknown. The model assumed was

$$y_j = \beta_0 + \beta_1 x_{1,j} + \beta_2 x_{2,j} + e_j \ , \tag{3.1}$$

with $j = 1, ..., 33$ and $\beta_1, \beta_2 \geq 0$. Considering that $e_j \sim N(0, \sigma^2)$, the procedure consists in generating values for $\tilde{\theta}_1$, $\tilde{\theta}_2$ and S_0, comparing the result of the usual F test with the derived test for $H_0\colon \beta_1 = 0$ vs $H_1\colon \beta_1 > 0$. The procedure was repeated 10000 times to obtain an empirical rejection probability. The chosen values for the parameters were:

$$\begin{aligned} \beta_1 &= 0, 1, 3, 10 \\ \beta_2 &= 0, 1, 3, 10, 30 \\ \rho &= 0, 0.1, 0.25, 0.5 \\ \sigma^2 &= 1 \ . \end{aligned} \tag{3.2}$$

The results appear in Tables 1 through 3.

Table 1: Rejection probability of ad-hoc test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.062	0.041	0.056	0.052	0.055
1	0.232	0.275	0.251	0.228	0.243
3	0.900	0.891	0.899	0.905	0.913
10	1.000	1.000	1.000	1.000	1.000

Table 2: Rejection probability of ad-hoc test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.053	0.045	0.044	0.053	0.045
1	0.218	0.229	0.232	0.266	0.238
3	0.879	0.853	0.904	0.900	0.860
10	1.000	1.000	1.000	1.000	1.000

Table 3: Rejection probability of ad-hoc test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.045	0.035	0.053	0.053	0.050
1	0.263	0.225	0.224	0.237	0.232
3	0.917	0.865	0.892	0.899	0.901
10	1.000	1.000	1.000	1.000	1.000

For comparison sake, we also present simulation results for the usual F test (Tables 4 to 6).

Table 4: Rejection probability of F test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0477	0.0471	0.0468	0.0481	0.0460
1	0.1646	0.1554	0.1653	0.1675	0.1592
3	0.8320	0.8293	0.8273	0.8276	0.8210
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5: Rejection probability of F test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0471	0.0512	0.0460	0.0474	0.0491
1	0.1634	0.1648	0.1624	0.1588	0.1577
3	0.8240	0.8296	0.8297	0.8258	0.8317
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6: Rejection probability of F test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0464	0.0504	0.0490	0.0494	0.0500
1	0.1544	0.1551	0.1694	0.1593	0.1630
3	0.8308	0.8275	0.8253	0.8280	0.8288
10	1.0000	1.0000	1.0000	1.0000	1.0000

The power increase for the ad-hoc test over the F test is evidently very significant for $\beta_1 > 0$.

3.2. Exact LR test with unknown σ^2

A batch of simulations was run for the exact test Δ. The set of parameters considered, the same as for the previous batch of simulations, was:

$$\begin{aligned} \beta_1 &= 0, 1, 3.10 \\ \beta_2 &= 0, 1, 3, 10, 30 \\ \rho &= 0, 0.25, 0.5 \\ \sigma^2 &= 1\ . \end{aligned} \tag{3.3}$$

The results appear in Tables 7 to 9.

Table 7: Rejection probability of exact LR test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0470	0.0526	0.0495	0.0495	0.0498
1	0.2476	0.2548	0.2485	0.2493	0.2466
3	0.9030	0.9015	0.9000	0.8961	0.8998
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 8: Rejection probability of exact LR test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0490	0.0499	0.0484	0.0521	0.0443
1	0.2702	0.2773	0.2512	0.2552	0.2514
3	0.9133	0.9063	0.8963	0.9012	0.8943
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 9: Rejection probability of exact LR test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0486	0.0592	0.0452	0.0459	0.0541
1	0.2769	0.2647	0.2511	0.2403	0.2564
3	0.9485	0.9324	0.8949	0.8934	0.9031
10	1.0000	1.0000	1.0000	1.0000	1.0000

Again, there is a clear gain of power over the usual F test for $\beta_1 > 0$.

3.2.1. Critical Values

The critical values for the exact likelihood ratio test were also obtained (Table 10).

Table 10: 5% quantiles.

$\beta_2 \backslash \rho$	0	0.25	0.5
0	0.9126540	0.9045819	0.8939074
1	0.9136219	0.9139177	0.9121665
3	0.9120875	0.9121392	0.9100027
10	0.9111222	0.9145002	0.9092603
30	0.9108427	0.9106236	0.9129284

It is easy to see that the critical values across the different values of β_2 are similar, and stabilize as this parameter increases. This leads us to believe that the use of these critical values would be valid for a wide range of unknown values of β_2.

ACKNOWLEDGMENTS

Our sincere thanks are due to an anonymous reviewer for some excellent comments and suggestions which led to improved presentation. This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

[1] Andrews, D. (1999). Estimation when a parameter is on a boundary, *Econometrica*, **67**, 1341–1383.

[2] Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica*, **69**, 683–734.

[3] Gourieroux, C.; Holly, A. textupand Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica*, **50**, 63–80.

[4] Liew C.K. (1976). Inequality Constrained Least-Squares Estimation, *Jour. Amer. Statist. Assoc.*, **71**, 746–751.

[5] Kopylev, L. and Sinha, B.K. (2010). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary, *Sankhya, Series B*, **73**, 20–41.

[6] Meyer, M.C. (2003). A test for linear versus convex regression function using shape-restricted regression, *Biometrika*, **90**, 223–232.

[7] Mukerjee, H. and Tu, R. (1995). Order-restricted inferences in linear regression, *Jour. Amer. Statist. Assoc.*, **90**, 717–728.

[8] Self, S.G. and Liang, K-Y. (1985). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Jour. Amer. Statist. Assoc.*, **82**, 605–610.

[9] Sinha, B.K.; Kopylev, L. and Fox, J. (2012). Some new aspects of statistical inference for multistage dose-response models with applications, *Technical report.*