

AD-A092 535

NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/6 12/1
TESTS FOR REGRESSION PARAMETERS IN POWER TRANSFORMATION MODELS. (U)
1980 R J CARROLL AFOSR-80-0080

UNCLASSIFIED

AFOSR-TR-80-1053

NL

1 1/2
AD
ADDRESS



END
DATE
FILMED
181
DTIC

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
18 AFOSR-TR-80-1053	AD-A092535		
4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED		
6 TESTS FOR REGRESSION PARAMETERS IN POWER TRANSFORMATION MODELS	9 Interim rept.		
7. AUTHOR(s)	6. PERFORMING ORG. REPORT NUMBER		
10 R. J./Carroll			
8. CONTRACT OR GRANT NUMBER(s)	15 AFOSR-80-0080		
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS		
University of North Carolina Department of Statistics Chapel Hill, N. C. 27514	16 6110ZF 2304/A5		
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE		
Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	11/1980		
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES		
12 13	12		
15. SECURITY CLASS. (of this report)	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
UNCLASSIFIED			
16. DISTRIBUTION STATEMENT (of this Report)			
Approved for public release; distribution unlimited			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
E			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)			
Power Transformation, Box-Cox Family, Testing Hypotheses, Likelihood Ratio Tests			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)			
We study tests of hypotheses for regression parameters in the power transformation model. In this model, a simple test consists of estimating the correct scale and then performing the usual linear model F-test in this estimated scale. We explore situations in which this test has the correct level asymptotically as well as better power than Wald's test or the likelihood ratio test. ←			

AD A092535

DDC FILE COPY

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

182 850

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR- 80 - 1053

Tests for Regression Parameters
in Power Transformation Models
by

R.J. Carroll*

University of North Carolina at Chapel Hill

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/ _____	
Availability Codes	
Dist.	Avail and/or special
A	

Key Words and Phrases: Power Transformations, Box-Cox Family, Testing
Hypotheses, Likelihood Ratio Tests.

AMS 1970 Subject Classifications: Primary 62 E 20; Secondary 62 G 35.

Research supported by the US Air Force Office of Scientific Research
under contract AFOSR-80-0080.

80 12 01 021

Approved for public release;
Distribution unlimited.

Abstract

We study tests of hypotheses for regression parameters in the power transformation model. In this model, a simple test consists of estimating the correct scale and then performing the usual linear model F-test in this estimated scale. We explore situations in which this test has the correct level asymptotically as well as better power than Wald's test or the likelihood ratio test.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TECHNICAL TO DDC
THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS
APPROVED FOR PUBLIC RELEASE IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

1. Introduction

We study tests of hypotheses for a regression parameter within the context of the Box and Cox (1964) power transformation family. The model is given by

$$\begin{aligned}
 (1) \quad y_i^{(\lambda)} &= x_i \beta + \sigma \epsilon_i, \quad i=1, \dots, N, \\
 x_i \beta &= (v_i \ w_i) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad (w_i), \gamma_2 \text{ are scalars} \\
 y^{(\lambda)} &= (Y^\lambda - 1)/\lambda \quad \text{if } \lambda \neq 0 \\
 &= \log Y \quad \text{if } \lambda = 0.
 \end{aligned}$$

Here σ is the standard deviation and $\epsilon_1, \dots, \epsilon_N$ are independently and identically distributed with mean zero and variance one. We are interested in testing the hypothesis

$$(2) \quad H_0: \gamma_2 = 0.$$

In what follows, λ^* will be an estimate of λ , $\beta^* = (\gamma_1^*, \gamma_2^*)$ will be the least squares estimates in the estimated scale λ^* and $\hat{\beta} = (\hat{\gamma}_1, \hat{\gamma}_2)$ will be the least squares estimates calculated in the true but unknown scale λ . A substantial literature now exists, although there has been no real emphasis on the hypothesis testing problem (2); see Andrews (1971), Atkinson (1973), Hinkley (1975), Carroll (1980), Bickel and Doksum (1980), and Carroll and Ruppert (1980) as a subset of this literature.

Of course, if the errors are normally distributed, the obvious method for testing (2) is the likelihood ratio test or LRT. Equivalent to this in large samples is the Wald test WT which is based on γ_2^* divided by an appropriate estimate of its standard error. In practice what is most often done is to select the scale λ^* and then do the usual analysis in this scale, i.e., divide γ_2^* by the usual formula for its estimated standard error. We denote such an analysis by CT, for conditional test based on the estimated scale λ^* . Hinkley (1975) was apparently the first to recognize that these tests are not all equivalent.

Example #1. Consider the location model for log-normal data with $\lambda = 0$, $\sigma = 1$ and normal errors following

$$\log Y_i = \mu + \epsilon_i.$$

Estimate λ by the normal theory MLE. Then $N^{1/2}(\mu^* - \mu)$ is asymptotically normally distributed with mean zero and variance $1 + (1 + \mu^2)^2/6$. Thus, in testing $H_0: \mu = 0$, the test CT which rejects when $N^{1/2}|\mu^*/\sigma^*|$ exceeds the normal $1-\alpha/2$ percentile always has a higher level than the desired level α , at least asymptotically. □

However, in some cases the test CT is quite good. Bickel and Doksum (1980) recognized this for the case of simple linear regression through the origin. We present an illustrative example.

Example #2. Consider log-normal simple linear regression with $\lambda = 0$, $\sigma = 1$:

$$\log Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

$$N^{-1} \sum_{i=1}^N X_i^k = \mu_k, \quad k = 1, 2, 3, 4$$

$$\mu_1 = 0, \mu_2 = 1.$$

In this important special case, it is possible to show by very detailed likelihood calculations that $N^{1/2}(\beta_1^* - \beta_1)$ is asymptotically normally distributed with mean zero and variance

$$(3) \quad s^2(\beta_1) = 1 + 4\beta_1^2 (\beta_0 + \beta_1 \mu_3/2)^2 (6 + 8\beta_1^2 + \beta_1^4 (\mu_4 - 1 - \mu_3))^{-1}.$$

Of course, $N^{1/2}(\hat{\beta}_1 - \beta_1)$ has limit variance equal to one. The test CT rejects $H_0: \beta_1 = 0$ whenever

$$N^{1/2}|\beta_1^*|/\sigma^* > t(\alpha, N-2),$$

where $t(\alpha, N-2)$ is the two-sided t -percentile. The Wald test WT

rejects when $N^{1/2} |\beta_1^*| / \sigma^* s(\beta_1^*)$ is large. What equation (3) says is that asymptotically

(a) the test CT has the correct level, since $s^2(0) = 1$,

(b) the test CT has higher power than the WT or LRT. $\square$

The theory developed by Bickel and Doksum (1980) and outlined in the next section suggests that when the naive test CT has the correct level, it will also have better power than the WT or LRT. That such a phenomenon is possible is not too surprising in view of the fact that when $\beta=0$ and $\sigma=1$ are known, the statistical curvature (Efron (1975)) for estimating λ is $\gamma_0^2 = 10^{2/3}$ when $\lambda=0$; Efron suggests $\gamma_0^2 \geq 1/8$ is large!

The purpose of this article is to describe situations in which the test CT has the correct level; this we do in Section 3. Basically, we require only that the $\{w_i\}$ be independent of the $\{v_i\}$, a situation that obtains in many important models including simple linear regression and balanced factorial designs.

2. Small σ Asymptotics

Assume that the errors $\{e_i\}$ are symmetrically distributed. Bickel and Doksum (1980) find major technical simplifications by letting $\sigma \rightarrow 0$ as $N \rightarrow \infty$. Define

$$A = (x_1' \dots x_N')' \quad P = A(A'A)^{-1}A'$$

$$Q = (A'A)^{-1}A'd' \quad d = (d_1 \dots d_N)$$

$$d_i = (\lambda^{-2} \{ (v_i - 1) - v_i \log(v_i) \})$$

$$v_i = 1 + \lambda x_i' \beta$$

$$N^{-1}e_N' = N^{-1}(dd' - dP d') \rightarrow E > 0.$$

From standard regression theory, we know that $N^{1/2}(\hat{\beta} - \beta)/\sigma$ is asymptotically normally distributed with mean zero and covariance

$$(4) \quad \sum_0 = \lim_{N \rightarrow \infty} (N^{-1}A'A)^{-1}.$$

Bickel and Doksum show that $N^{1/2}(\hat{\beta}^* - \beta)/\sigma$ has asymptotic covariance

$$(5) \quad \sum_1 = \sum_0 + \lim_{N \rightarrow \infty} N^{-1}QQ'/e_N.$$

Suppose that $x_i = (v_i, w_i)$ with w_i scalar and v_i a $(1 \times p)$ vector. Define $\eta = (0 \dots 0, 1)$. Then asymptotically the test CT rejects $H_0: \gamma_2 = 0$ if

$$N^{1/2} | \eta \hat{\beta}^* | / (\eta \sum_0 \eta')^{1/2} > t(\alpha, N-p-1).$$

Equations (4) and (5) tell us that CT has the correct level if

$$(6) \quad \eta(\sum_1 - \sum_0)\eta' = 0 \quad \text{when } H_0: \gamma_2 = 0 \text{ obtains.}$$

Further, since QQ' is positive semidefinite, CT will have power at least as large as WT or LRT when (6) obtains. Another example illustrates this.

Example 3. We consider the small σ asymptotics for the two group analysis of covariance model when $\lambda = 0$:

$$\log Y_i = \mu + \rho s_i + \beta X_i + \sigma \epsilon_i, \quad i=1, \dots, 2N$$

$$s_1 = 1, s_2 = -1, s_3 = 1, s_4 = -1, \dots, s_{2N} = -1.$$

The parameter ρ is the treatment effect, and we are interested in testing

$$H_0: \rho = 0.$$

Let E be as above and set

$$\begin{aligned} \sum_1^{2N} X_i &= 0, & \sum_1^{2N} X_i^2 &= 2N \\ \sum_1^{2N} s_i X_i &= 2Na, & \sum_1^{2N} s_i X_i^2 &= 2Nb \\ \sum_1^{2N} X_i^3 &= 2Nc. \end{aligned}$$

We will show that the test CT has the correct level when the design is balanced over the two treatments in the first two moments of $\{X_i\}$, i.e., $a=b=0$. When $\lambda = 0$ is known, $N^{1/2}(\hat{\rho}-\rho)/\sigma$ has asymptotic variance $(1-a^2)^{-1}$. Estimating λ by maximum likelihood, we find that $N^{1/2}(\hat{\rho}^*-\rho)/\sigma$ has asymptotic variance

$$s^2 = (1-a^2)^{-1} + (4E(1-a^2)^2)^{-1} \left((\beta^2 - 2a\rho\beta)b - \beta^2 a c + 2\mu\rho(1-a^2) \right)^2.$$

Hence the test CT has the correct level if $a = b = 0$. □

4. The level of the test CT

In Example #3 we saw that when $\lambda = 0$ and the covariates are independent of the treatment assignment, then the test CT has the correct level. A generalization of this to the model (1) and hypothesis(2) might then need $\{v_i\}$ to be independent of $\{w_i\}$. In the following we state conditions which formalize this notion and are at the same time not restricted to the MLE λ^* as an estimate of λ . Our assumptions are stated in such a fashion as to allow the design $\{x_i\}$ to also be randomly generated.

Assumption #1. λ^* is root-N consistent, i.e., $N^{1/2}(\lambda^* - \lambda) = O_p(1)$.

Assumption #2. $N^{1/2}(\hat{\beta} - \beta)$ is asymptotically normally distributed.

Assumption #3. There exists σ^* with $\sigma^*/\sigma \xrightarrow{P} 1$.

Assumption #4. For the sequence $\{w_i\}$,

$$N^{-1} \sum_{i=1}^N w_i \rightarrow 0$$

$$N^{-1} \sum_{i=1}^N w_i^2 \rightarrow 1,$$

almost surely if the $\{w_i\}$ are random.

Assumption #5. Let $F_N(v, w, e)$ be the empirical distribution function of $\{(x_i, e_i) = (v_i, w_i, e_i)\}$. Suppose there exists distribution functions F_1, F_2 and F_3 such that

$$F_N(v, w, e) \xrightarrow{w} F_1(v)F_2(w)F_3(e)$$

almost surely. Further, $\{(x_i, e_i)\}$ are uniformly bounded.

Theorem. Make the Assumptions #1 - #5. Then under $H_0: \gamma_2 = 0$ the statistics $N^{1/2} \hat{\gamma}_2 / \hat{\sigma}$ and $N^{1/2} \gamma_2^* / \sigma^*$ are each asymptotically normally distributed with mean zero and variance one. Hence, the test CT has the correct asymptotic level. $\square$

Note that the theorem is stated only in terms of the design, as long as we have appropriate estimates of λ and σ . The value of λ itself is not important, in contrast to Examples #1 - #3.

Some comment on the assumptions is in order. Assumption #1 is crucial, but obviously λ^* need not be the normal theory MLE; see Hinkley (1975), Carroll (1980) and Bickel and Doksum (1980) for other suggestions. Assumption #4 will hold if there is an intercept; otherwise it seems necessary, as the work of Bickel and Doksum indicates for simple regression through the origin. Assumptions #2 - #3 are hardly onerous.

The only restrictive assumption is #5. The boundedness is needed only in a technical sense to make the proof fairly easy. Also, one of the most common assumptions in regression is that the errors $\{e_i\}$ are independent of the design $\{x_i = (v_i, w_i)\}$. The heart of Assumption #5 is thus the requirement that $\{v_i\}$ be unrelated to $\{w_i\}$.

When is this requirement satisfied? It certainly holds for simple linear regression under Assumption #4, either with or without an origin. More importantly, it holds for balanced $\kappa \times 2$ designs where the test is for the treatment effect. Another important example in which Assumption #5 will hold is general two-group analysis of covariance in which the covariates are equally distributed across the treatments, as might occur with random or blocked allocation.

It should also be noted that the CT test will be of the correct level even when $\{w_i\}$ and γ_2 are vectors rather than scalars, as in general factorial designs or many-group analysis of covariance. The requirement remains that $\{w_i\}$ should be unrelated to $\{v_i\}$.

5. Conclusion

We have shown that the rather naive test CT, which picks a scale λ^* and then performs the usual F-test, has the correct asymptotic level in many important statistical problems. Generally speaking, this level obtains in balanced designs. When the test CT has the correct asymptotic level, it generally outperforms Wald's test and the likelihood ratio test.

6. Proof of the Theorem

Write $A = (x_1' \dots x_N')'$. By Assumption #4 - #5 we have

$$(8) \quad N^{-1} A' A \longrightarrow \sum_* = \begin{pmatrix} \sigma^{**} & 0 \\ 0 & 1 \end{pmatrix},$$

the convergence being with probability one in case of randomness. By (8) and Assumptions #2 - #4, $N^{1/2}(\hat{\gamma}_2 - \gamma_2)/\sigma^*$ is asymptotically normally distributed with mean zero and variance one, so it suffices to show that under $H_0: \gamma_2 = 0$,

$$(9) \quad N^{1/2}(\gamma_2^* - \hat{\gamma}_2)/\sigma \xrightarrow{P} 0.$$

Because of (8), (9) will follow by proving

$$(10) \quad N^{-1/2} \sum_1^N w_i (\gamma_i^{(\lambda^*)} - \gamma_i^{(\lambda)}) / \sigma \xrightarrow{P} 0.$$

By a Taylor expansion, the left hand side of (10) becomes

$$N^{1/2} \left(\frac{\lambda^* - \lambda}{\sigma} \right) N^{-1} \sum_1^N w_i \frac{\partial}{\partial \lambda} \gamma_i^{(\lambda)} \xrightarrow{P} 0,$$

so that by Assumption #1 we need only show

$$(11) \quad N^{-1} \sum_{i=1}^N w_i \frac{\partial}{\partial \lambda} \gamma_i^{(\lambda)} \xrightarrow{P} 0.$$

Now, under $H_0: \gamma_2 = 0$ there exists a bounded function G with

$$\frac{\partial}{\partial \lambda} Y_i^{(\lambda)} = G(v_i, \gamma_1, \sigma \epsilon_i),$$

so that (11) becomes

$$(12) \quad N^{-1} \sum_{i=1}^N w_i G(v_i, \gamma_1, \sigma \epsilon_i) \xrightarrow{P} 0.$$

But from Assumption #5 the left side of (12) converges to

$$\int w G(v, \gamma, \sigma \epsilon) dF_1(w) dF_2(v) dF_3(\epsilon) = 0,$$

the last following from Assumption #5.

References

Andrews, D.F. (1971): A note on the selection of data transformations.
Biometrika 58, 249-254.

Atkinson, A.C. (1973): Testing transformations to normality. J. Royal Statist. Soc. Ser. B, 35, 473-479.

Bickel, P.J. and Doksum, K.A. (1980): An analysis of transformations revisited. To appear in J. Amer. Statist. Assoc.

Box, G.E.P. and Cox, D.R. (1964): An analysis of transformations.
J. Royal Statist. Soc. Ser. B, 26, 211-252.

Carroll, R.J. (1980): A robust method for testing transformations to achieve approximate normality. J. Royal Statist. Soc. Series B, 42, 71-78.

Carroll, R.J. and Ruppert, D. (1980): Prediction and the power transformation family. Tentatively accepted by Biometrika.

Efron, B. (1975): Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Statist. 3, 1189-1242.

Hinkley, D.V. (1975): On power transformations to symmetry, Biometrika 62, 101-111.

FILMED
— 8