COMPARING TRANSFORMATIONS USING TESTS OF SEPARATE FAMILIES

by

Lloyd J. Edwards and Ronald W. Helms

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 1894

March 1992

Comparing Transformations Using Tests of Separate Families

By Lloyd J. Edwards and Ronald W. Helms
Department of Biostatistics, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27599

Summary

Cox's method for testing separate hypotheses is applied to derive a decision rule for choosing between the models $f(\underset{\sim}{Y}) \sim N_n(\underset{\sim}{X}\underset{\sim}{\beta}_1, \sigma_1^2\underset{\sim}{I}_n)$ and $g(\underset{\sim}{Y}) \sim N_n(\underset{\sim}{Z}\underset{\sim}{\beta}_2, \sigma_2^2\underset{\sim}{I}_n)$. The asymptotic distributions of the decision statistics are derived, subject to suitable regularity conditions.

Some key words: Regression; Separate families; Transformation.

1. Introduction

A number of situations arise in which one must choose between the following two hypothesized models for a set of data:

$$H_1:\quad f(\underset{\sim}{y}) \sim N_n(\underset{\sim}{X}\underset{\sim}{\beta}_1, \sigma_1^2\underset{\sim}{I}_n) \tag{1.1a}$$

$$H_2:\quad g(\underset{\sim}{y}) \sim N_n(\underset{\sim}{Z}\underset{\sim}{\beta}_2, \sigma_2^2\underset{\sim}{I}_n) \tag{1.2a}$$

or written in standard regression notation as:

$$H_1:\quad f(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}_1 + \underset{\sim}{\epsilon} \tag{1.1b}$$

$$H_2:\quad g(\underset{\sim}{y}) = \underset{\sim}{Z}\underset{\sim}{\beta}_2 + \underset{\sim}{u} \tag{1.2b}$$

where $f(\underset{\sim}{y}) = [f(y_i)]$ is an n x 1 vector of transformed values of a response variable, $g(\underset{\sim}{y}) = [g(y_i)]$, $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ are n x p and n x q matrices of values of independent variables, $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ are unknown primary parameter vectors (p x 1 and q x 1, respectively), σ_1^2 and σ_2^2 are unknown variance parameters, and $\underset{\sim}{\epsilon} \sim N_n(\underset{\sim}{0}, \sigma_1^2\underset{\sim}{I}_n)$, $\underset{\sim}{u} \sim N_n(\underset{\sim}{0}, \sigma_2^2\underset{\sim}{I}_n)$. A diagnostic procedure may lead to a negative conclusion that one or both of these models does not fit the data well. Since diagnostic procedures do not combine information from both models they do not lead to a positive decision rule for selecting one model instead of the other.

Cox (1961, 1962) defined two composite hypotheses as separate if a simple hypothesis in one cannot be reached as a limit of simple hypotheses in the other. In particular, if one hypothesis is a subhypothesis of another the two hypotheses are not separate. Cox proposed a general methodology, based on likelihood ratios, for testing separate hypotheses after recognizing that ad hoc methods certainly existed for handling these types of problems. Cox left open many details and examples for subsequent investigation.

In the ensuing years, several authors investigated testing separate families of hypotheses either by Cox's method or by proposing an alternative method (see Atkinson (1970), Sawyer (1983), Mizon (1984), Loh (1985)). Pesaran (1974) derived Cox's test for testing separate linear models with nonstochastic explanatory variables and later (Pesaran, 1982) studied power calculcations for these tests. Jackson (1968) derived the Cox test for comparing several families, including lognormal versus gamma, exponential versus lognormal, and others. Pereira (1977) studied consistency of the Cox test compared to Atkinson's (1970) test and Pereira (1978) applied Cox's method to derive a procedure for choosing between separate log-linear models. Numerous other references can be found in Mizon and Richard (1986) and Godfrey (1988). To our knowledge no one has applied Cox's method to derive a procedure for choosing between two different transformed regression models.

In this paper we apply Cox's method to derive a decision procedure for choosing between two different transformed regression models with possibly different design matrices, along the lines of Pesaran (1974).

2. A Procedure Based On Cox's Method

The hypotheses to be tested are given by (1.1) and (1.2). The y_i are assumed positive. Unlike typical likelihood ratio test situations, in this context neither H_1 nor H_2 is a "null hypothesis" with a "preferred status". At some points in the procedure H_1 will be the "null hypothesis" and at other points H_2 will be the null.

We assume both $f(\cdot)$ and $g(\cdot)$ are one-to-one, continuously differentiable, real-valued functions. For convenience, define:

$\underset{\sim}{x}_i = \text{Row}_i(\underset{\sim}{X})$, i=1,...,n and $\underset{\sim}{h} = [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))]$ is an n x 1 vector of the function g composed with the function f^{-1}, (2.1)

$[\cdot]$ will denote an n x 1 vector throughout the remainder of this paper,

$$\underset{\sim}{M}_x = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T \text{ and } m_{ij}^{(x)} \text{ is the ij}^{th} \text{ element of the matrix } \underset{\sim}{M}_x, \tag{2.2a}$$

$$\underset{\sim}{M}_z = \underset{\sim}{I} - \underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T \text{ and } m_{ij}^{(z)} \text{ is the ij}^{th} \text{ element of the matrix } \underset{\sim}{M}_z. \tag{2.2b}$$

We shall consider the asymptotic properties of test procedures as the number of observations, n, goes to infinity. We make the following regularity assumptions:

$$\lim_{n\to\infty}(\frac{1}{n}\underset{\sim}{X}^T\underset{\sim}{X}) = \underset{\sim}{\Sigma}_{x'x} \text{ (non-singular)}, \tag{2.3a}$$

$$\lim_{n\to\infty}(\frac{1}{n}\underset{\sim}{Z}^T\underset{\sim}{Z}) = \underset{\sim}{\Sigma}_{z'z} \text{ (non-singular)}, \tag{2.3b}$$

$$\lim_{n\to\infty}(\frac{1}{n}\underset{\sim}{Z}^T\underset{\sim}{h}) = \underset{\sim}{\Sigma}_{z'z} \,(\neq \underset{\sim}{0}) \,. \tag{2.3c}$$

The initial phase of Cox's method treats H_1 as a "null hypothesis" and leads to the statistic (Cox, 1961, 1962)

$$T_1^* = T_1^*(\hat{\underset{\sim}{\theta}}_1,\hat{\underset{\sim}{\theta}}_2) = \hat{L}_{21} - n\{\underset{n\to\infty}{\mathrm{Plim}}(\hat{L}_{21}/n)\}_{\underset{\sim}{\theta}_1=\hat{\underset{\sim}{\theta}}_1}, \tag{2.4}$$

where the log-likelihoods for H_1 and H_2 are given by $L_1(\underset{\sim}{\theta}_1)$ and $L_2(\underset{\sim}{\theta}_2)$; $\hat{L}_{21}=L_1(\hat{\underset{\sim}{\theta}}_1) - L_2(\hat{\underset{\sim}{\theta}}_2)$ is the maximized log-likelihood ratio; $\underset{\sim}{\theta}_1 = (\sigma_1^2, \underset{\sim}{\beta}_1^T)^T$, $\underset{\sim}{\theta}_2 = (\sigma_2^2, \underset{\sim}{\beta}_2^T)^T$ with $\hat{\underset{\sim}{\theta}}_1$ and $\hat{\underset{\sim}{\theta}}_2$ denoting the maximum likelihood estimators of $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_2$ under H_1 and H_2. Plim($\cdot$) denotes the probability limit under the null hypothesis H_1, where the limit is a function of the unknown parameter $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_1$ is replaced by $\hat{\underset{\sim}{\theta}}_1$.

Cox's procedure would compare the maximized log-likelihood ratio with its expected value under the null hypothesis if the expected value were available. Here, the expected value is not available and has been replaced by the rightmost term in (2.4). Cox showed that when H_1 is true, the test statistic T_1^* is asymptotically normally distributed with mean zero and variance $V_1(T_1^*)$, under certain regularity conditions (see White (1982)). Cox showed that the variance, which is a function of the unknown parameter $\underset{\sim}{\theta}_1$, can be given by:

$$V_1(T_1^*) = V_1(L_{21}) - \frac{1}{n}\underset{\sim}{\eta}^T\underset{n\to\infty}{\mathrm{Plim}}(n\underset{\sim}{I}_1^{-1})\underset{\sim}{\eta}\,, \tag{2.5}$$

where we define $L_{21} = L_1(\underset{\sim}{\theta}_1) - L_2(\underset{\sim}{\theta}_{21})$ and $\underset{\sim}{\theta}_{21}$ is the asymptotic expectation of $\underset{\sim}{\theta}_2$ under H_1 ; $\underset{\sim}{I}_1 = \underset{\sim}{I}_1(\underset{\sim}{\theta}_1)$ is the information matrix of $\underset{\sim}{\theta}_1$ under H_1;

$$\underset{\sim}{\eta} = \underset{\sim}{\eta}(\underset{\sim}{\theta}_1) = n\,\frac{\partial\{\underset{n\to\infty}{\mathrm{Plim}}(\hat{L}_{21}/n)\}}{\partial\underset{\sim}{\theta}_1}\,. \tag{2.6}$$

Using the log-likelihood equations for H_1 and H_2, it is straightforward to show that

$$\hat{L}_{21} = \frac{n}{2}\log\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}\,, \tag{2.7}$$

where

$$L_1(\underset{\sim}{\theta}_1) = -\frac{n}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}(f(\underset{\sim}{y}) - \underset{\sim}{X}\underset{\sim}{\beta}_1)^T(f(\underset{\sim}{y}) - \underset{\sim}{X}\underset{\sim}{\beta}_1)$$

and

$$L_2(\underset{\sim}{\theta}_2) = -\frac{n}{2}\log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2}(g(\underset{\sim}{y}) - \underset{\sim}{Z}\underset{\sim}{\beta}_2)^T(g(\underset{\sim}{y}) - \underset{\sim}{Z}\underset{\sim}{\beta}_2) .$$

The maximum likelihood estimates of the parameters under H_1 and H_2 are given by

$$\hat{\underset{\sim}{\beta}}_1 = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T f(\underset{\sim}{y}) \ , \quad \hat{\sigma}_1^2 = \frac{1}{n}f(\underset{\sim}{y})^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)f(\underset{\sim}{y}) = \frac{1}{n}f(\underset{\sim}{y})^T\underset{\sim}{M}_x f(\underset{\sim}{y})$$

and

$$\hat{\underset{\sim}{\beta}}_2 = (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T g(\underset{\sim}{y}) \ , \quad \hat{\sigma}_2^2 = \frac{1}{n}g(\underset{\sim}{y})^T(\underset{\sim}{I} - \underset{\sim}{Z}(\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T)g(\underset{\sim}{y}) = \frac{1}{n}g(\underset{\sim}{y})^T\underset{\sim}{M}_z g(\underset{\sim}{y}).$$

Under H_1, we find that

$$n\hat{\sigma}_2^2 = [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))]^T\underset{\sim}{M}_z[g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))] \tag{2.9}$$

We use a first order Taylor expansion on the function $h(\epsilon_i) = g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))$ about the point $\epsilon_i = 0$ for i=1,...,n to obtain the following:

$$h(\epsilon_i) = g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i)) \ \approx \ g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i \ , \tag{2.10}$$

where $g'(\cdot)$ and $g''(\cdot)$ will denote the first and second derivatives of the function $g(\cdot)$.

Thus, under H_1 we have

$$n\hat{\sigma}_2^2 \ \approx \ [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i]^T\underset{\sim}{M}_z[g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i]. \tag{2.11}$$

So, under suitable conditions (see Appendix A), $\hat{\sigma}_2^2 - \frac{1}{n}\tilde{\sigma}_{21}^2 \to 0$ in probability as $n\to\infty$, where

$$\tilde{\sigma}_{21}^2 = nE_{\underset{\sim}{\theta}_1}(\hat{\sigma}_2^2) \ \approx \ \sigma_1^2\text{trace}(\underset{\sim}{M}_z\underset{\sim}{V}) + [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))]^T\underset{\sim}{M}_z[g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))] \tag{2.12}$$

where $\underset{\sim}{V} = \text{diag}\{[g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))^2]\}$.

Assume $\frac{1}{n}\tilde{\sigma}^2_{21} \to \sigma^2_{21}(\underset{\sim}{\theta}_1)$ as $n \to \infty$. Since $\hat{\sigma}^2_1$ is a consistent estimate of σ^2_1 assuming H_1 is true, we have the following result:

$$\underset{n\to\infty}{\text{Plim}}(\hat{L}_{21}/n) \approx \frac{1}{2}\log\frac{\sigma^2_{21}(\underset{\sim}{\theta}_1)}{\sigma^2_1}. \tag{2.13}$$

Thus, we have a test statistic expression for T_1, which is an approximation of T^*_1:

$$T_1 = T_1(\hat{\underset{\sim}{\theta}}_1, \hat{\underset{\sim}{\theta}}_2) = \frac{n}{2}\log\frac{\hat{\sigma}^2_2}{\sigma^2_{21}(\hat{\underset{\sim}{\theta}}_1)}, \tag{2.14}$$

where $\sigma^2_{21}(\hat{\underset{\sim}{\theta}}_1) = \frac{1}{n}\{\hat{\sigma}^2_1 \text{trace}(\underset{\sim}{M}_z\hat{\underset{\sim}{V}}) + [g(f^{-1}(\underset{\sim}{x}_i\hat{\underset{\sim}{\beta}}_1))]^T\underset{\sim}{M}_z[g(f^{-1}(\underset{\sim}{x}_i\hat{\underset{\sim}{\beta}}_1))]\}$.

To derive the asymptotic variance of T^*_1 we need the asymptotic expectation of $\hat{\underset{\sim}{\beta}}_2$ under H_1. Under H_1, we can write

$$\begin{aligned}\hat{\underset{\sim}{\beta}}_2 &= (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T g(\underset{\sim}{y}) = (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T[g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))] \\ &\approx (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\underset{\sim}{Z}^T[g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i]\end{aligned}$$

again by using a first order Taylor expansion. We then have, by using the above expression, that $\hat{\underset{\sim}{\beta}}_2 \to \underset{\sim}{\beta}_{21}$ in probability as $n \to \infty$ under H_1, where

$$\underset{\sim}{\beta}_{21} = \Sigma^{-1}_{zz}\Sigma_{zh}. \tag{2.15}$$

Assuming H_1 is true, we can write

$$\begin{aligned}L_{21} &= \frac{n}{2}\log\frac{\sigma^2_{21}}{\sigma^2_1} - \frac{1}{2\sigma^2_1}(f(\underset{\sim}{y}) - \underset{\sim}{X}\underset{\sim}{\beta}_1)^T(f(\underset{\sim}{y}) - \underset{\sim}{X}\underset{\sim}{\beta}_1) + \frac{1}{2\sigma^2_{21}}(g(\underset{\sim}{y}) - \underset{\sim}{Z}\underset{\sim}{\beta}_{21})^T(g(\underset{\sim}{y}) - \underset{\sim}{Z}\underset{\sim}{\beta}_{21}) \\ &\approx \frac{n}{2}\log\frac{\sigma^2_{21}}{\sigma^2_1} - \frac{1}{2\sigma^2_1}\sum_{i=1}^{n}\epsilon^2_i + \frac{1}{2\sigma^2_{21}}\sum_{i=1}^{n}\{g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i - \underset{\sim}{z}_i\underset{\sim}{\beta}_{21}\}^2,\end{aligned}$$

again by using a first order Taylor expansion.

Now we have

$$\begin{aligned}V_1(L_{21}) \approx V_1\{&\frac{1}{2}\sum_{i=1}^{n}(\frac{1}{\sigma^2_{21}}(g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)))^2 - \frac{1}{\sigma^2_1})\epsilon^2_i \\ &+ \frac{1}{\sigma^2_{21}}\sum_{i=1}^{n}(g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) - \underset{\sim}{z}_i\underset{\sim}{\beta}_{21})g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i\,]\end{aligned}$$

$$= \tfrac{1}{2}\sigma_1^4 \sum_{i=1}^{n} \left(\frac{1}{\sigma_{21}^2}(g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)))^2 - \frac{1}{\sigma_1^2}\right)^2$$

$$+ \frac{\sigma_1^2}{\sigma_{21}^4} \sum_{i=1}^{n} \{(g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) - \underset{\sim}{z}_i\underset{\sim}{\beta}_{21})g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\}^2 \qquad (2.16)$$

since $\epsilon_i \sim$ independent identically distributed $N(0, \sigma_1^2)$ for $i=1,\ldots,n$.

In addition, it can be shown that

$$\underset{\sim}{\eta} = n \frac{\partial\{\operatorname{Plim}_{n\to\infty}(\hat{L}_{21}/n)\}}{\partial \underset{\sim}{\theta}_1} \approx n(\eta_1, \underset{\sim}{\eta}_2^T)^T, \qquad (2.17a)$$

where

$$\eta_1 = \tfrac{1}{2}\{\frac{1}{n\sigma_{21}^2}\operatorname{trace}(\underset{\sim}{M}_z\underset{\sim}{V}) - \frac{1}{\sigma_1^2}\} \qquad (2.17b)$$

and

$$\underset{\sim}{\eta}_2 = \frac{1}{\sigma_{21}^2}\Big\{\frac{1}{n}\sum_{i=1}^{n} m_{ii}^{(z)}\{\sigma_1^2 g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))g''(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1)) + g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\}\underset{\sim}{x}_i^T\Big\}$$

$$+ \frac{1}{n}\sum_{i<j}^{n} m_{ij}^{(z)}\{\underset{\sim}{x}_i^T g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))g(f^{-1}(\underset{\sim}{x}_j\underset{\sim}{\beta}_1)) + \underset{\sim}{x}_j^T g'(f^{-1}(\underset{\sim}{x}_j\underset{\sim}{\beta}_1))g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\}. \qquad (2.17c)$$

It also can be shown that

$$\operatorname{Plim}_{n\to\infty}(n\underset{\sim}{\Gamma}_1^{-1}) = \begin{bmatrix} 2\sigma_1^4 & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma_1^2\Sigma_{x'x}^{-1} \end{bmatrix}. \qquad (2.18)$$

Thus, we have everything we need to evaluate

$$V_1(T_1) \approx V_1(L_{21}) - \tfrac{1}{n}\underset{\sim}{\eta}^T \operatorname{Plim}_{n\to\infty}(n\underset{\sim}{\Gamma}_1^{-1})\underset{\sim}{\eta}\,. \qquad (2.19)$$

To get $\hat{V}_1(T_1)$ we replace the unknown parameters by their consistent estimates.

Hence, by Cox's methodology we have $T_1/[V_1(T_1)]^{1/2}$ is asymptotically normally distributed with mean zero and variance one under H_1. This allows us to provide a large sample approximation of the associated p-value of the test.

In Cox's formulation the two hypotheses, H_1 and H_2, are considered asymmetrically, where H_2 serves as the type of alternative for which high power is desired. Hence, it is instructive to compute T_2, Cox's test statistic assuming the null and alternative hypotheses have been interchanged. Appendix B gives the main results for interchanging H_1 with H_2.

3. Discussion

We have derived the Cox statistic for tests of separate families of hypotheses for comparing two different transformed regression models. The testing procedure allows great flexibility in choosing possibly different design matrices for selecting between transformations. In the derivation we assumed that both $f(\cdot)$ and $g(\cdot)$ were continuously differentiable. However, as can be seen in the derivation, we only need the existence of the first and second derivatives.

The full utility of this method can only be assessed after further research into its asymptotic power and level of significance performance, small sample properties, and applications to particular testing scenarios by other researchers.

Although some of the equations may look formidable, the computations can be carried out rather straight-forward using any computer software which has matrix handling capabilities such as SAS© IML.

One common example of application would be to use the results of this paper to compare an untransformed linear regression model to that of a log-transformed model (see Edwards and Helms (1990)).

Acknowledgement

The authors would like to thank Dr. P. K. Sen for his helpful suggestions. This research was supported by NHLBI grant number P50HL19171.

REFERENCES

Atkinson, A. C. (1970), A method for discriminating between models. *J. R. Statist. Soc. B,* **32**, 323-353.

Box, G. E. P. and Cox, D. R. (1964), An analysis of transformations. *J. R. Statist. Soc. B* **26**, 211-243, discussion 244-252.

Cox, D. R. (1961), Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp.* 1, 105-123.

Cox, D. R. (1962), Further results on tests of separate families of hypothesis. *J. R. Statist. Soc. B* 24, 406-424.

Edwards, L. J. and Helms, R. W. (1990), On choosing between a linear or log-linear regression model. *Unpublished manuscript.* Department of Biostatistics, University of North Carolina at Chapel Hill.

Fisher, G. and McAleer, M. (1979), On the interpretation of the Cox test in econometrics. *Economics Letters* 4, 145-150.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches.* Cambridge University Press, New York.

Jackson, O. A. Y. (1968), Some results on tests of separate families of hypotheses. *Biometrika* **55**, 2, 355-363.

Loh, Wei-Yin (1985), A new method for testing separate families of hypotheses. *J. Amer. Statist. Assoc.*, **80**, 362-368.

Mizon, G. E. (1984), The encompassing principle in econometrics, *Econometrics and Quantitative Economics*, Ed. D. F. Hendry and K. F. Wallis, pp. 135-172. Oxford: Basil Blackwell.

Mizon, G. E. and Richard, J.-F. (1986), The encompassing principle and its application to testing non-nested hypotheses. *Econometrica* **54**, 3, pp 657-678.

Pereira, B., (1977), A note on the consistency and on the finite sample comparisons of some tests of separate families of hypotheses. *Biometrika* **64**, 1, 109-113.

Pereira, B., (1978), Tests and efficiencies of separate regression models. *Biometrika* **65**, 2, 319-327.

Pesaran, M. H. (1974), On the general problem of model selection. *Review of Economic Studies* **126**, 153-171.

Pesaran, M. H. and Deaton, A. S. (1978), Testing non-nested nonlinear regression models. *Econometrica*, **46**, 677-694.

Pesaran, M. H. (1982), Comparison of local power of alternative tests of non-nested regression models. *Econometrica*, **50**, 1287-1305.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications.* Wiley, New York.

Sawyer, K. R. (1983), Testing separate families of hypotheses: An information criterion. *J. R. Statist. Soc. B*, **45**, 89-99.

Searle, S. R. (1971), *Linear Models.* Wiley, New York.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics.* Wiley, New York.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

White, H. (1982), Regularity conditions for Cox's test of non-nested hypotheses. *J. Econometrics*, **19**, 301-318.

Appendix A

Sufficient condition for $\hat{\sigma}_2^2 - \frac{1}{n}\tilde{\sigma}_{21}^2 \to 0$ in probability as $n \to \infty$ (section 2, p 4):

Under H_1, upon using the first order Taylor expansion and simplifying the quadratic form, we find

$$n\hat{\sigma}_2^2 = [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))]^T \underset{\sim}{M}_z [g(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1 + \epsilon_i))] \tag{A.1}$$

$$\approx \underset{\sim}{h}^T \underset{\sim}{M}_z \underset{\sim}{h} + 2\underset{\sim}{h}^T \underset{\sim}{M}_z [g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i]$$

$$+ [g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i]^T \underset{\sim}{M}_z [g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i] \tag{A.2}$$

Since $g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\epsilon_i \sim$ independently distributed $N(0, \sigma_1^2\{g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))\}^2)$, for $i = 1,\ldots,n$, we have (see Searle (1971))

$$V_1(\hat{\sigma}_2^2) = \frac{2}{n^2}\Big\{2\sigma_1^2\{\underset{\sim}{h} + [g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))]\}^T \underset{\sim}{M}_z \underset{\sim}{V} \underset{\sim}{M}_z \{\underset{\sim}{h} + [g'(f^{-1}(\underset{\sim}{x}_i\underset{\sim}{\beta}_1))]\}$$

$$+ \operatorname{trace}(\underset{\sim}{M}_z \underset{\sim}{V})^2\Big\}. \tag{A.3}$$

Thus, by Chebychev's inequality, if $V_1(\hat{\sigma}_2^2) \to 0$ as $n \to \infty$, then $\hat{\sigma}_2^2 - \frac{1}{n}\tilde{\sigma}_{21}^2 \to 0$ in probability as $n \to \infty$, where $V_1(\hat{\sigma}_2^2)$ is given by the right-hand side of the above equation. So, the sufficient condition needed for the probability convergence is that the right-hand side of the above equation goes to zero as $n \to \infty$.

A completely analogous condition is obtained when interchanging the hypotheses.

Appendix B

Interchanging Hypotheses

In Cox's formulation the two hypotheses, H_1 and H_2, are considered asymmetrically, where H_2 serves as the type of alternative for which high power is desired. Hence, it is instructive to compute T_2, Cox's test statistic assuming the null and alternative hypotheses have been interchanged. Presenting T_2 will also help to illustrate why the hypotheses are considered asymmetrically.

Using notation analogous to that used in section 2 and corresponding expectations taken under H_2, we have the following:

$$\text{i) } T_2^* = \hat{L}_{12} - n\{\operatorname{Plim}_{n\to\infty}(\hat{L}_{12}/n)\}_{\underset{\sim}{\theta}_2=\hat{\underset{\sim}{\theta}}_2} \approx \frac{n}{2}\log\frac{\hat{\sigma}_1^2}{\sigma_{12}^2(\hat{\underset{\sim}{\theta}}_2)} = T_2\,, \tag{B.1}$$

where $\sigma_{12}^2(\hat{\underset{\sim}{\theta}}_2) = \frac{1}{n}\{\sigma_2^2\operatorname{trace}(\underset{\sim}{M}_x\underset{\sim}{W}) + [f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))]^T\, \underset{\sim}{M}_x\, [f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))]\}_{\underset{\sim}{\theta}_2=\hat{\underset{\sim}{\theta}}_2}$

and $\underset{\sim}{W} = \operatorname{diag}\{[f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))^2]\}$,

$$\text{ii) } V_2(T_2) \approx V_2(L_{12}) - \frac{1}{n}\underset{\sim}{\nu}^T\operatorname{Plim}_{n\to\infty}(n\underset{\sim}{I}_2^1)\underset{\sim}{\nu}\,, \tag{B.2a}$$

$$\text{where } V_2(L_{12}) \approx \frac{1}{2}\sigma_2^4\sum_{i=1}^{n}\Big(\frac{1}{\sigma_{12}^2}(f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2)))^2 - \frac{1}{\sigma_2^2}\Big)^2 + \frac{\sigma_2^2}{\sigma_{12}^4}\sum_{i=1}^{n}\{(f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2)) - \underset{\sim}{x}_i\underset{\sim}{\beta}_{12})f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))\}^2, \tag{B.2b}$$

$$\hat{\underset{\sim}{\beta}}_1 \to \underset{\sim}{\beta}_{12} = \underset{\sim}{\Sigma}_{x'x}^{-1}\underset{\sim}{\Sigma}_{x'd}\,,\ \text{where } \underset{\sim}{d} = [f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))] \tag{B.3}$$

$$\underset{\sim}{\nu} = n\,\frac{\partial\{\operatorname{Plim}_{n\to\infty}(\hat{L}_{12}/n)\}}{\partial\underset{\sim}{\theta}_2} \approx n(\nu_1, \underset{\sim}{\nu}_2^T)^T \text{ with} \tag{B.4a}$$

$$\nu_1 = \frac{1}{2}\Big\{\frac{1}{n\sigma_{12}^2}\operatorname{trace}(\underset{\sim}{M}_x\underset{\sim}{W}) - \frac{1}{\sigma_2^2}\Big\}, \tag{B.4b}$$

$$\underset{\sim}{\nu}_2 = \frac{1}{\sigma_{12}^2}\Big\{\frac{1}{n}\sum_{i=1}^{n}m_{ii}^{(x)}\{\sigma_2^2 f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))f''(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2)) + f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))\}\underset{\sim}{z}_i^T + \frac{1}{n}\sum_{i<j}^{n}m_{ij}^{(x)}\{\underset{\sim}{z}_i^T f'(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))f(g^{-1}(\underset{\sim}{z}_j\underset{\sim}{\beta}_2)) + \underset{\sim}{z}_j^T f'(g^{-1}(\underset{\sim}{z}_j\underset{\sim}{\beta}_2))f(g^{-1}(\underset{\sim}{z}_i\underset{\sim}{\beta}_2))\}\Big\}. \tag{B.4c}$$

and

$$\operatorname{Plim}_{n\to\infty}(n\underset{\sim}{I}_2^1) = \begin{bmatrix} 2\sigma_2^4 & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma_2^2\underset{\sim}{\Sigma}_{z'z}^{-1} \end{bmatrix}. \tag{B.5}$$

Again, by Cox's methodology we have $T_2/[V_2(T_2)]^{1/2}$ is asymptotically normally distributed with mean zero and variance one under H_2.