ERRORS IN VARIABLES AND PROPERTIES OF STATISTICAL INFERENCE

by

Lloyd J. Edwards

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1883T
Errors in Variables and Properties of Statistical Inference

by

Lloyd J. Edwards

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1990

Approved by:

Advisor

Reader

Reader

Reader
Abstract

Lloyd J. Edwards. Errors in Variables and Properties of Statistical Inference (under the direction of Dr. P. K. Sen.)

The objective here is to investigate the effects errors in variables have on tests of separate families of linear hypotheses (i.e., non-nested linear hypotheses) using the methodology introduced by D. R. Cox (1961, 1962). Two hypotheses are called separate if an arbitrary simple hypothesis in one cannot be obtained as a limit of simple hypotheses in the other.

We derive Cox’s test of non-nested linear models for stochastic explanatory variables measured with error under the assumption of normality (normal structural models). These previously underived results reflect the effect the error covariance of the stochastic explanatory variables has on Cox’s test. The effect of the error covariance is shown to appear in the test statistic itself and its variance.

We also derive Cox’s test of non-nested linear models for fixed explanatory variables measured with error under the assumption of normality (normal functional models). These previously underived results are shown to differ in form from that of Pesaran (1974) for the case of fixed explanatory variables not measured with error. As in the case of the structural model, the effect of the error covariance is shown to appear in the test statistic itself and its variance. Since the “true” explanatory variables are treated as nuisance parameters, the number of parameters in the models increase with the sample size n. We propose a method of reducing the number of parameters in the likelihood function and investigate conditions (practical and theoretical) for selecting an appropriate reduced likelihood function.

We apply our results to real data, discuss the effects of applying our results, and discuss plans for future research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>LIST OF TABLES</strong></td>
<td>v</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter I</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Introduction and Literature Review</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1 Tests of Separate Families of Hypotheses</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1.2 Methodology</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.1.3 Further Developments</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1.1.4 Regularity conditions for the Cox test of separate families of</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>hypotheses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2 Errors in Variables for Linear Models</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>1.2.1 Introduction</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>1.2.2 Maximum Likelihood Estimation</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>1.2.3 Remarks</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter II</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testing Separate Normal Structural Models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.1 Introduction</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2.1.1 Hypotheses and Assumptions</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2.1.2 Derivation of Cox’s Test</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>2.2 Consistency of Test</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>2.2.1 A Sufficient Condition</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>2.3 Interchanging Hypotheses</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter III</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testing Separate Normal Functional Models</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.1 Introduction</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>3.1.1 Hypotheses and Assumptions</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>3.1.2 Derivation of Cox’s Test</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3.1.3 A Note on the Selection of Quantile Design Matrix</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>3.1.4 Discussion</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>3.2 Interchanging Hypotheses</td>
<td>58</td>
</tr>
</tbody>
</table>
Chapter IV
Applications

4.1 Introduction ............................................................................................................. 60
    4.1.1 Assuming Normal Structural Models ............................................................. 62
    4.1.2 Results from Testing Normal Structural Models ............................................ 65
4.2 Assuming Normal Functional Models ..................................................................... 72
    4.2.1 Results from Testing Normal Functional Models ............................................. 80
4.3 Results from Standard Testing Procedures ............................................................. 81

Chapter V
Discussion and Plans for Further Research

5.1 Introduction ............................................................................................................. 82
    5.1.1 Effect of Measurement Errors ....................................................................... 83
5.2 Plans for Future Research ....................................................................................... 84

Appendix: Proofs of Selected Results .......................................................................... 86
References ....................................................................................................................... 91
# LIST OF TABLES

Table 4.1: Surgical Unit Example  
Descriptive Statistics ........................................................................ 66

Table 4.2: Surgical Unit Example  
Correlation Matrix ........................................................................... 67

Table 4.3: Surgical Unit Example  
\( H_0: \) variables \( X_1 \) and \( X_2 \) vs \( H_1: \) variables \( X_3 \) and \( X_4 \)  
Normal Structural Models ..................................................................... 68

Table 4.4: Surgical Unit Example  
\( H_0: \) variables \( X_2 \) and \( X_4 \) vs \( H_1: \) variables \( X_1 \) and \( X_3 \)  
Normal Structural Models ..................................................................... 69

Table 4.5: Surgical Unit Example  
\( H_0: \) variables \( X_1 \) and \( X_4 \) vs \( H_1: \) variables \( X_2 \) and \( X_3 \)  
Normal Structural Models ..................................................................... 70

Table 4.6: Surgical Unit Example  
\( H_0: \) variable \( X_4 \) vs \( H_1: \) variables \( X_1, X_2 \) and \( X_3 \)  
Normal Structural Models ..................................................................... 71

Table 4.7: Surgical Unit Example  
\( H_0: \) variables \( X_1 \) and \( X_2 \) vs \( H_1: \) variables \( X_3 \) and \( X_4 \)  
Normal Functional Models .................................................................... 76

Table 4.8: Surgical Unit Example  
\( H_0: \) variables \( X_2 \) and \( X_4 \) vs \( H_1: \) variables \( X_1 \) and \( X_3 \)  
Normal Functional Models .................................................................... 77

Table 4.9: Surgical Unit Example  
\( H_0: \) variables \( X_1 \) and \( X_4 \) vs \( H_1: \) variables \( X_2 \) and \( X_3 \)  
Normal Functional Models .................................................................... 78

Table 4.10: Surgical Unit Example  
\( H_0: \) variable \( X_4 \) vs \( H_1: \) variables \( X_1, X_2 \) and \( X_3 \)  
Normal Functional Models .................................................................... 79
CHAPTER 1
Introduction and Literature Review

OBJECTIVE: Errors in variables models (or measurement error models) have been explored since the latter part of the 19th century. Much activity and results on estimation and statistical inference have been forthcoming in linear and nonlinear errors in variables models, particularly in the past 10 to 15 years. The objective here is to investigate the effects errors in variables have on tests of separate families of linear hypotheses (i.e., non-nested linear hypotheses) using the methodology introduced by Cox (1961, 1962).

1.1 Tests of Separate Families of Hypotheses

1.1.1 Introduction

Cox (1961, 1962) recognized a need to find a general method for handling testing of separate families of hypotheses, a class of problems which until that time had not received much attention in the literature. Two hypotheses are called separate if an arbitrary simple hypothesis in one cannot be obtained as a limit of simple hypotheses in the other. Prior to Cox’s 1961 paper, ad hoc methods had been employed to test separate (non-nested) hypotheses.

The idea of two hypotheses being separate is a function of both the parameters of the hypotheses and the definitions of the hypotheses. For instance, both the null and
alternative hypotheses could have the same number of parameters in their parameter vectors, say a location and scale parameter, but the hypotheses separate due to the definitions of the hypotheses. As an example, consider testing the null hypothesis, \( H_0 \), that the random variable \( Y \) has a normal distribution with unknown mean and variance against \( H_1 \), that \( \log(Y) \) has a normal distribution with unknown mean and variance (\( Y \) positive). The two hypotheses have the same number of parameters in their parameter vectors. However, the hypotheses are separate since a simple hypothesis in \( H_0 \) could not be reached as a limit of simple hypotheses in \( H_1 \).

An example of nested (non-separate) hypotheses is as follows: Let \( Y_1, \ldots, Y_n \) be independently normally distributed and let \( H_0 \) be the hypothesis that

\[
E(Y_i) = \alpha, \quad i = 1, \ldots, n,
\]

and \( H_1 \) be the hypothesis that

\[
E(Y_i) = \beta_1, \quad i = 1, \ldots, \gamma,
\]

\[
\beta_2, \quad i = \gamma+1, \ldots, n,
\]

where \( \alpha, \beta_1, \beta_2, \) and \( \gamma \) are unknown. The two hypotheses are not separate since any hypothesis in \( H_0 \) can be achieved as a sequence of simple hypotheses in \( H_1 \).

As an another example of separate hypotheses, consider testing the null hypothesis, \( H_0 \), that the probability density function (pdf) is log-normal with unknown parameter values against \( H_1 \), that the pdf is exponential with unknown parameter value. Irwin (1942) discusses the difficulty in distinguishing between these distributions. Cox presents several other examples in his two papers.

Cox’s method is a large sample procedure based on a modification of the
Neyman-Pearson likelihood ratio.

1.1.2 Methodology

Let $Y=(Y_1, Y_2, \ldots, Y_n)$ be an observed random vector and suppose we are interested in testing the composite null hypothesis, $H_0$, that the probability density function (pdf) is $f(y, \alpha)$, against the composite alternative, $H_1$, that the pdf is $g(y, \beta)$, where $\alpha \in \Omega_0 \subset \mathbb{R}^{k_1}$ and $\beta \in \Omega_1 \subset \mathbb{R}^{k_2}$, where $k_1, k_2 > 1$. The following assumptions are made:

(i) $f(y, \alpha)$ and $g(y, \beta)$ represent separate families in the sense that for an arbitrary value of $\alpha$, say $\alpha_0$, $f(y, \alpha_0)$ cannot be approximated arbitrarily closely by $g(y, \beta)$;

(ii) the parameters $\alpha$ and $\beta$ may be treated as varying continuously even when a component of say $\beta$ is the serial number of the observation at which a discontinuity occurs;

(iii) the values of $\alpha$, or $\beta$, are interior to $\Omega_0$, or $\Omega_1$, so that the type of distribution problem discussed by Chernoff (1954) is excluded.

Let $L_0(\hat{\alpha})$ be the maximized log-likelihood under $H_0$ and $L_1(\hat{\beta})$ be the maximized log-likelihood under $H_1$ where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimates of $\alpha$ and $\beta$, respectively. Cox proposed using the following test statistic:

$$T_0 = \{ L_0(\hat{\alpha}) - L_1(\hat{\beta}) \} - E_{\hat{\alpha}} \{ L_0(\hat{\alpha}) - L_1(\hat{\beta}) \},$$

which compares the observed difference of log-likelihoods with an estimate of the expected log-likelihoods. Expectation is taken under $H_0$ and $\alpha$ is replaced with its maximum likelihood estimate, $\hat{\alpha}$, under $H_0$. Under $H_0$, $T_0$ should be nearly zero, but under $H_1$, $T_0$ should be negative. Thus, a large negative value of $T_0$ would lead to the rejection of the null hypothesis, $H_0$. 
Let $L_{01} = L_0(\hat{\alpha}) - L_1(\hat{\beta})$. A few general remarks are in order.

(i) Usually in likelihood ratio applications $\Omega_0 \subset \Omega_1$, so that $L_{01} < 0$. Under the assumption of separate families, this inequality may not hold.

(ii) If the components of $Y$ are independent, $L_{01}$ is the sum of $n$ independent terms and an application of the central limit theorem will usually prove the asymptotic normality of $L_{01}$. Approximations to the percentage points of $L_{01}$ can then be obtained under both $H_0$ and $H_1$.

(iii) The problem considered is one of significance testing and not discrimination, which is to say $H_0$ and $H_1$ are considered unsymmetrically, where $H_1$ serves only to indicate the type of alternative for which high power is required. In discrimination, one is concerned with the degree for which each of the models provides a reasonable description of the data. $H_0$ and $H_1$ are not the only possible hypotheses.

(iv) The roles of $H_0$ and $H_1$ can be interchanged, yielding the test statistic $T_1$, where

$$T_1 = \left[ L_1(\hat{\beta}) - L_0(\hat{\alpha}) \right] - E_\beta \left[ L_1(\hat{\beta}) - L_0(\hat{\alpha}) \right]$$

In general, $T_0$ and $T_1$ will be different functions of the observations. Both $T_0$ and $T_1$ should be investigated. Expectation is taken under $H_1$ and $\beta$ is replaced by its maximum likelihood estimate, $\hat{\beta}$, under $H_1$.

Let the elements of $Y$ be independent. The log-likelihoods are sums with one item per observation. Using the following structures the covariance matrix can be expressed in notational form:

$$F = \log(f(y,\alpha)), \quad F_{\alpha_i} = \frac{\partial \log(f(y,\alpha))}{\partial \alpha_i}, \quad F_{\alpha_i \alpha_j} = \frac{\partial^2 \log(f(y,\alpha))}{\partial \alpha_i \partial \alpha_j},$$
\[ G = \log(g(y, \beta)), \quad G_{\beta_i} = \frac{\partial \log(g(y, \beta))}{\partial \beta_i}, \quad G_{\beta_i \beta_j} = \frac{\partial^2 \log(g(y, \beta))}{\partial \beta_i \partial \beta_j}. \]

Also, let \( F_k, G_k, \) etc. refer to the \( k^{th} \) term in the summation representation of \( F, G, \) etc. Thus, \( T_0 \) and the variance of \( T_0 \) can be written asymptotically as:

\[ T_0 = \sum_{k=1}^{n} \{ F_k - G_k - E_{\alpha}(F - G) - \sum_{j,l} E_{\alpha}(F, F_j) C_{\alpha}(F - G, F_{\alpha j}) F_{k,\alpha l} \}, \text{ and} \]

\[ V_{\alpha}(T_0) = n \{ V_{\alpha}(F - G) - \sum_{j,l} E_{\alpha}(F, F_j) C_{\alpha}(F - G, F_{\alpha j}) C_{\alpha}(F - G, F_{\alpha l}) \}. \]

where \( V_{\alpha}(\cdot), C_{\alpha}(\cdot) \) denote variance and covariance under \( H_0 \) and \( E_{\alpha}(F, F_j) \) is an element of the matrix inverse to \( E_{\alpha}(F_{\alpha j} F_{\alpha l}) \). Similarly, one can write the covariance for the statistic \( T_1 \) by interchanging \( H_0 \) and \( H_1 \).

For scalar \( \alpha \) and \( \beta \), \( T_0 \) can be written asymptotically as:

\[ T_0 = \sum_{k} \{ F_k - G_k - E_{\alpha}(F - G) - \frac{C_{\alpha}(F - G, F_{\alpha})}{V_{\alpha}(F_{\alpha})} F_{k,\alpha} \}, \]

with asymptotic variance

\[ V_{\alpha}(T_0) = n \{ V_{\alpha}(F - G) - \frac{C_{\alpha}^2(F - G, F_{\alpha})}{V_{\alpha}(F_{\alpha})} \}. \]

Cox has shown that by using simplifying assumptions, under \( H_0 \), \( T_0 \) is asymptotically normally distributed with mean zero and covariance \( V_{\alpha}(T_0) \). Under \( H_0 \), \( \hat{\beta} \) converges to \( \beta_\alpha \), i.e., the MLE of \( \beta \), \( \hat{\beta} \), is computed under \( H_1 \) and its limiting behavior is investigated under \( H_0 \).

As an example of the latter concept, suppose \( H_0 \) is the hypothesis that the pdf, \( f(y, \alpha) \), is normal with mean \( \alpha \) and variance 1, and that \( H_1 \) is the hypothesis that the
pdf, \(g(y, \beta)\), is Cauchy with location parameter \(\beta\) and scale parameter 1. It can be shown that \(\beta_\alpha\) exists and is equal to \(\alpha\). However, \(\alpha_\beta\) does not exist since \(\hat{\alpha}\) is the sample mean and it does not converge in probability under \(H_1\).

In constructing the test statistic for \(H_0\) vs \(H_1\), basically four steps are required:

(i) the maximum likelihood estimates of \(\alpha\) and \(\beta\);

(ii) the log-likelihood ratio \(L_{01}\);

(iii) the expected value \(E_{\hat{\alpha}}\{L_{01}\}\);

(iv) the asymptotic variance of \(T_0\).

If another estimation procedure is used which is asymptotically equivalent to the maximum likelihood estimates, then the asymptotic distribution theory will still hold. In constructing the test statistic, \(T_0\), points (iii) and (iv) above pose the most difficulties.

Cox presented several examples in detail including the log-normal distribution versus the exponential distribution, the Poisson distribution versus the geometric distribution, and applications to quantal response models.

1.1.3 Further Developments

Cox (1962) left several issues unresolved for further development. The most relevant to this discussion were:

(i) the examination of the adequacy of the limiting normal approximation to the distribution of \(T_0\), in particular the inclusion of the terms of order 1 in probability in some of the expansions used in the development of the procedure;

(ii) the calculation of power functions and the comparison with alternative tests, for which the distribution of \(T_0\) under \(H_1\) would be required;

(iii) the replacement of the test statistics by asymptotically equivalent statistics that are easier to compute;

(iv) the investigation of possible asymptotic optimum properties;
(v) the extensions to deal with problems about time series.

Walker (1967) applied Cox's test to various problems in time series analysis. He considered two main types of problems: (1) time series generated by an autoregression of specified order \( p \) under \( H_0 \) and by a moving-average of specified order \( q \) under \( H_1 \); (2) the time series is generated by a simple harmonic oscillation with superimposed random error under \( H_0 \) and by an autoregression under \( H_1 \). Since time series analysis isn't the focus of this presentation, no further elaboration will be given.

Jackson (1968) investigated the adequacy of the asymptotic results of testing the log normal distribution versus the exponential distribution (a test explored by Cox) and the log normal distribution versus the gamma distribution and presented power functions of the tests of log normal versus exponential. He also compared the performance of Cox's test with that of other tests for departure from the exponential distribution.

Jackson found that \( T_0 \) (null: exponential, alternative: log normal) had high power and that \( T_1 \) had low power for small \( n \) but rose steadily as \( n \) increased, both tests using a two-sided significance level of 0.10. The asymptotic results for \( T_0 \) seemed adequate for \( n \) as low as 50. For \( T_1 \) the correction to the mean was relatively large but the asymptotic null variance seemed adequate for all \( n \).

Jackson compared \( T_0 \) with three other test statistics: (1) Jackson's \( T \) statistic (Jackson, 1967) which is based on the sum of products of the ordered values and the expected order statistics for an exponential, (2) Lewis's \( S' \) statistic (Lewis, 1965) which is a weighted combination of order statistics, and (3) Moran's \( M \) statistic (Moran, 1951) which is essentially a comparison of the geometric and arithmetic means. The comparisons were made by using the power of the tests at the 0.05 one-sided significance level. The power of the tests depended on the shape parameter of the log normal distribution and Jackson found that when this shape parameter was 1, \( T_0 \) was much
more powerful than the other tests. For other values of the shape parameter, \( T_0 \) did not do that well.

Atkinson (1970) examined an alternative idea proposed by Cox (but not expounded upon) of discriminating between models. The idea is to combine the two hypotheses into a general model of which they both would be special cases. The general form of the combined pdf (for two hypotheses) is written as

\[
f_{\lambda}(y) = \frac{\{ f(y, \alpha) \}^{\lambda_1} \{ g(y, \beta) \}^{\lambda_2}}{\int \{ f(z, \alpha) \}^{\lambda_1} \{ g(z, \beta) \}^{\lambda_2} dz},
\]

where three distinct situations can occur:

1. Both \( \lambda_1 \) and \( \lambda_2 \) can be estimated.
2. Only one of the \( \lambda \)'s can be estimated.
3. Neither \( \lambda \) can be estimated. The combined pdf in this case has only one general form and two special cases corresponding to the component distributions.

The form of the distributions being tested and whether the values of the parameters are specified or are estimated will dictate which of the three situations will occur.

It is noted that more than two competing hypotheses could be examined by this exponential combination procedure by using, in general, a pdf proportional to

\[
\prod_{i=1}^{P} \{ f_i(y, \theta_i) \}^{\lambda_i},
\]

but we will limit our discussion to only two hypotheses.
If the log-likelihood ratio is used as a test statistic in which both component distributions contain a vector of parameters requiring estimation, then the calculation of the MLE's of $\lambda_1$ and $\lambda_2$ would involve a multivariable function maximization which would usually necessitate a complicated numerical integration at each point at which the function is evaluated. This could be a monumental task even with a powerful computing facility.

Atkinson proposed an alternative to the maximum likelihood ratio test so as to reduce the necessary computations. Upon adding the constraint $\lambda_1 + \lambda_2 = 1$, it is desirable to test the hypothesis $\lambda = 1$ (where $\lambda_1 = \lambda$ and $\lambda_2 = 1 - \lambda$), which implies testing for departures from the first model in the direction of the second model; $\alpha$ and $\beta$ are nuisance parameters. An asymptotically normal statistic for testing hypotheses about the value of a single parameter in the presence of nuisance parameters was suggested by Bartlett (1953) and, independently, by Neyman (1959).

Under Atkinson's formulation, for simplification, assume that the log-likelihood, $L$, contains only the scalar nuisance parameter $\theta$. Let

$$I_{11} = - E \left( \frac{\partial^2 L}{\partial \lambda^2} \right), \quad I_{12} = - E \left( \frac{\partial^2 L}{\partial \theta \partial \lambda} \right), \quad I_{22} = - E \left( \frac{\partial^2 L}{\partial \theta^2} \right),$$

where the derivatives are evaluated and expectations calculated under the null hypothesis. Atkinson proposed the following test statistic:

$$T = \frac{(\partial L/\partial \lambda) - (I_{12}/I_{22})(\partial L/\partial \theta)}{\{I_{11} - (I_{12}^2/I_{22})\}^{1/2}}.$$

This is the derivative of the log-likelihood with respect to the parameter of interest adjusted for the regression on the partial derivative with respect to the nuisance
parameter, divided by the appropriate standard error. The adjustment for regression ensures that the statistic is asymptotically unbiased for all allowable estimates of $\theta$. Under the null hypothesis the test statistic will be asymptotically normally distributed. Extensions to a vector of $p$ nuisance parameters is straightforward.

Atkinson showed that the test statistic, $T$, is asymptotically equivalent to the Cox statistic when testing separate hypotheses. Using the notation in Section 1.1.2, the numerator of Atkinson's $T$ statistic for testing two separate hypotheses when $\alpha$ and $\beta$ are scalar quantities is written as:

$$T(\beta_{\hat{\alpha}}) = \sum_{k=1}^{n} \{ F_k(\hat{\alpha}) - G_k(\beta_{\hat{\alpha}}) - E_{\hat{\alpha}} \{ F(\hat{\alpha}) - G(\beta_{\hat{\alpha}}) \} \}.$$ 

As in Cox’s methodology, expectation is taken under $H_0$ and $\alpha$ is replaced with its maximum likelihood estimate, $\hat{\alpha}$. Recall that $\beta_{\hat{\alpha}}$ is the limit of $\hat{\beta}$ under $H_0$. Since $\alpha$ is estimated, $\beta_{\hat{\alpha}}$ is replaced by $\beta_{\hat{\alpha}}$ in calculating the statistic.

When $\lambda=1$, the likelihood is independent of $\beta$ so that the estimate of $\beta$ is not determined by the theory of Atkinson. Atkinson noted, however, that substitution of any preassigned value of $\beta$ yields a statistic which is asymptotically standard normal.

Since the purpose of the statistic is to test for departures from the first distribution in the direction of the second, an estimate of $\beta$ which best describes the data under the null hypothesis is used, namely $\beta_{\hat{\alpha}}$.

In comparison, Cox's test statistic for scalar $\alpha$ and $\beta$ is written as:

$$T(\hat{\beta}) = \sum_{k=1}^{n} \{ F_k(\hat{\alpha}) - G_k(\hat{\beta}) - E_{\hat{\alpha}} \{ F(\hat{\alpha}) - G(\beta_{\hat{\alpha}}) \} \}.$$ 

Observe that Cox uses $G_k(\hat{\beta})$ and Atkinson uses $G_k(\beta_{\hat{\alpha}})$. Under the null hypothesis,
the numerator of Atkinson's statistic would have zero expectation if the true value of $\alpha$ were known whereas the numerator of Cox's statistic has this property only asymptotically as $\tilde{\beta} \rightarrow \beta_{\alpha}$.

For vector valued parameters $\alpha$ and $\beta$ under Atkinson's formulation, let

$$F_{i,a_j} = \frac{\partial}{\partial a_j} \{ \log(f(y_i, \alpha)) \}$$

be the $j^{th}$ elements of the $1 \times p$ vector $X_i$, where $p$ is the number of nuisance parameters. Now write

$$X'X = \{ C_{\alpha}(F_{a_j}, F_{a_k}) \}$$

and

$$X'Y = \{ C_{\alpha}(F - G, F_{a_j}) \}.$$ 

The numerator of Atkinson's statistic is now written as

$$\sum_{i=1}^{n} \{ F_i - G_i - E_{\alpha}(F - G) - X_i(X'X)^{-1}X'Y \}$$

with variance

$$n \{ V_{\alpha}(F - G) - Y'X(X'X)^{-1}X'Y \}.$$ 

Atkinson goes on to apply the test statistic based on the exponential combination of pdf's to examples such as testing the exponential distribution against the log-normal, applications to quantal responses and other examples.

Dyer (1972, 1973, 1974) discussed discrimination procedures and hypothesis testing procedures for separate families of hypotheses. Dyer restricted his attention to procedures which were invariant under location and scale transformation. Dyer assumes
the null and alternative hypothesis are both from location-scale families that have distributions of the form:

$$\mathcal{F} = \{ \sigma^{-1} f((x-\mu)/\sigma) \mid \mu \in \mathbb{R}, \sigma > 0 \} ,$$

or to the case where $\mu$, the location parameter, is known and the distributions under the null and alternative hypotheses are scale families ($\mu$ and $\sigma$ are scalars). For hypothesis testing, Dyer explicitly states that he did not consider the methods suggested by Cox, Jackson, and Atkinson due to the difficulties involved in the computation of the test statistics. His interest centered mainly on statistics which were relatively easy to compute.

Pesaran (1974) applied Cox's method to the testing of separate univariate linear hypotheses, concentrating primarily on models for which the explanatory variables are fixed under repeated sampling. Pesaran also briefly discussed the exponential combination procedure, which he calls the comprehensive procedure, elaborated on by Atkinson. He then makes small sample comparisons of Cox's method and Atkinson's method by way of Monte Carlo simulation. The exact F-test statistic was used for the comprehensive model.

Pesaran observed that on average Cox's method overstated the assumed significance level of 0.05 of the test by about 3 percent for samples of size 40 or less; the F-test for the comprehensive model did not differ significantly from the assumed significance level. Cox's method resulted in smaller type II errors in every instance as compared to that of the F-test for the comprehensive procedure. Under the assumption that type I and type II errors are equally important, then Cox's method was seen to be better than the comprehensive procedure when the sample size was 40 or less and the correlation between the competing set of explanatory variables was large.
Pesaran concluded that when dealing with small samples and highly collinear sets of competing explanatory variables, Cox’s method was preferable to the comprehensive (exponential combination) procedure for testing non-nested univariate linear hypotheses. In addition, Pesaran emphasizes that the application of Cox’s tests of separate families of hypotheses to linear hypotheses can be framed as a method of selecting appropriate explanatory variables for a given response variable.

Quandt (1974) compared three procedures for testing separate (non-nested) linear hypotheses which included Pesaran’s formulation of Cox’s method, Atkinson’s exponential combination method, and a linear combination method. For the cases which Quandt considered, he concluded that there was no best method based on theoretical or empirical grounds.

Quesenberry and Starbuck (1976) explored optimal tests for separate families of hypotheses from continuous distributions. They show that in many cases a uniformly most powerful similar (UMPS) test exists and that the test is equivalent to a uniformly most powerful invariant (UMPI) test. In addition, the optimal tests are shown to have power that is a monotone function of the null hypothesis family of distributions.

Pereira (1977a) compiled a bibliography listing material related to the problem of choosing between models. He gives a very brief abstract of each of the listings and indicates whether the main approach taken was classical, bayesian, or likelihood.

Pereira (1977b) discussed the consistency of the tests used by Cox and Atkinson by considering the probability limits of the tests under the alternative hypothesis (which should be negative). An investigation into the finite sample properties of the tests is also discussed. Pereira shows that Cox’s method always converges in probability limit to a negative value under the alternative hypothesis. However, Atkinson’s method doesn’t
necessarily converge in probability limit to a negative value and in that case Atkinson's method will not be consistent. An illustration is given by testing the exponential distribution against the log normal distribution as the alternative. Finite sample comparisons were made for a few specific null and alternative hypotheses and the results indicated that Atkinson's method showed better agreement for the first two moments and Cox's method showed better agreement for skewness and kurtosis.

Pereira (1978) investigated the application of Cox's test in testing log linear regression models. The consequences of using one model when another is true was assessed. Asymptotic efficiencies of the estimators of the regression coefficients from the false model under the null hypothesis were analyzed. All log linear models investigated were univariate models with fixed regressor variables.

Pesaran and Deaton (1978) extended Pesaran's 1974 results from the univariate linear tests of non-nested hypotheses to tests of multivariate non-nested nonlinear regression models under the assumption that full information maximum likelihood estimation is available. Whereas the univariate linear case provided a method for selecting between competing explanatory variables, the multivariate nonlinear case, in addition, provides a method of selecting between alternative functional forms. The authors' results are shown to be recognizable as generalizations of the univariate single-equation case. An application of the derived results is given for the analysis of the relationship between consumption and income using U.S. quarterly data.

A couple of alternative methods of testing separate families of hypotheses were developed in 1982 and 1983. Epps, Singleton, and Pulley (1982) proposed using the difference between the empirical moment generating function and its theoretical counterpart for the testing of separate families of hypotheses. K. R. Sawyer (1983) proposed using an information criterion, an asymmetric test statistic for discriminating between models based on the information theory developed by Kullback and Leibler, for
the testing of separate families of hypotheses. Sawyer provided an application of the information criteria to discriminating between linear regression models. Both methods result in asymptotically normal test statistics after suitable normalization. Overall, the large and small sample applications of these alternative methods were comparable to Cox’s test statistic.

Pesaran (1982) derived and compared the local power of three different methods of testing non-nested regression models. The three methods considered were the F test, Cox’s test, and a linearized version of Cox’s test statistic (referred to as the J test) suggested by Davidson and MacKinnon (1981). All three test are shown to have the correct size asymptotically. The asymptotic power of the F and J test against local alternatives are shown to be less than that of Cox’s test, unless the number of non-overlapping variables of the alternative hypothesis over the null hypothesis is unity. Monte Carlo experiments were also used to check the validity of the theoretical results.

Aguirre-Torres and Gallant (1983) derived the asymptotic distribution of the generalized Cox test for testing multivariate nonlinear regression models. The asymptotic distribution is derived for the null and non-null models. In addition, a new test statistic was introduced which uses a bootstrap estimate for estimating the expectation of the Cox difference. The statistic does not require an explicit specification of the error distribution of the null model.

Loh (1985) observed that there was no guarantee that Cox’s test would be asymptotically level $\alpha$ in some of the applications, even for very smooth problems. Loh introduced a new test statistic for testing separate families of hypotheses which is based on repeated application of the parametric bootstrap idea over slowly shrinking confidence regions. A number of examples were provided in examining the finite-sample properties of the new method, in addition to the derivation of the theoretical properties of the new method. Pesaran (1982), however, showed that Cox’s test is asymptotically level $\alpha$ for testing linear regression models.
1.1.4 Regularity conditions for Cox's test of separate families of hypotheses

Cox (1962) left unresolved the adequacy of the limiting normal approximation to his test statistic. In the years following his introduction of the problem of testing separate families of hypotheses, several authors have applied his test statistic but no general regularity conditions and rigorous proof of the asymptotic normality of his test statistic had been given, until White (1982) provided general regularity conditions and a rigorous proof of the asymptotic normality of Cox's test statistic. White provided a demonstration of the applicability of the regularity conditions in the case of discriminating between linear regression models. In what follows, a presentation of the regularity conditions and a statement of the theorem regarding asymptotic normality is given.

Using the notation of White, let \( U_t, t=1,\ldots,n \) be iid. Suppose \( H_f: f(u,\theta) \) is the null hypothesis that the density function of \( U_t \) is \( f \), where \( \theta \in \Theta \) and \( H_h: h(u,\gamma) \) is the alternative hypothesis that the density function of \( U_t \) is \( h \), where \( \gamma \in \Gamma \). Let \( \hat{\theta}_n \) be the maximum likelihood estimator for \( \theta_0 \), the 'true' value of \( \theta \) under the null hypothesis. Let \( \hat{\gamma}_n \) be the quasi-maximum likelihood estimator, which is the estimator that maximizes \( n^{-1} \sum_t \log(h(u_t,\gamma)) \) over \( \Gamma \), and let \( \gamma_* \) denote the value to which \( \hat{\gamma}_n \) converges under the null hypothesis, \( H_f \).

Cox's test statistic can then be written as:

\[
T_f = n^{-1} \sum_{t=1}^{n} \log \left( \frac{f(U_t,\hat{\theta}_n)/h(U_t,\hat{\gamma}_n)}{f(u,\hat{\theta}_n)/h(u,\hat{\gamma}_n)} \right) - \int \log \left( \frac{f(u,\hat{\theta}_n)/h(u,\hat{\gamma}_n)}{f(u,\theta_0)/h(u,\gamma_*)} \right) f(u,\hat{\theta}_n) \, d\nu,
\]

where it is based on the difference

\[
n^{-1} \sum_{t=1}^{n} \log \left( \frac{f(U_t,\hat{\theta}_n)/h(U_t,\hat{\gamma}_n)}{f(u,\theta_0)/h(u,\gamma_*)} \right) - \int \log \left( \frac{f(u,\theta_0)/h(u,\gamma_*)}{f(u,\theta_0)} \right) f(u,\theta_0) \, d\nu,
\]

with \( \theta_0 \) and \( \gamma_* \) replaced by their estimates, \( \hat{\theta}_n \) and \( \hat{\gamma}_n \).


Assumptions 1, 2, and 3 [White (1982)] give general conditions under which the quasi-maximum likelihood estimator (QMLE) will be well behaved. Since the choice of $H_f$ or $H_h$ as the null hypothesis is arbitrary, the following assumptions are stated in terms of $f$ with the form for $h$ in parentheses:

**Assumption 1:** The independent random $1 \times M$ vectors $U_t$, $t=1,...,n$, have common joint distribution function $G$ on $\Omega$, a measurable Euclidean space, with measurable Radon-Nikodým density $g=dG/d\nu$ (Note: $g$ may correspond to $f$ or $h$, depending on whether $H_f$ or $H_h$ is the null hypothesis).

**Assumption 2:** The family of distribution functions $F(u,\theta)$ ($H(u,\gamma)$) has Radon-Nikodým densities $f(u,\theta)=dF(u,\theta)/d\nu$ ($h(u,\gamma)=dH(u,\gamma)/d\nu$) which are measurable in $u$ for every $\theta$ in $\Theta$ ($\gamma$ in $\Gamma$), a compact subset of a $p$-dimensional ($q$-dimensional) Euclidean space, and continuous in $\theta$ ($\gamma$) for every $u$ in $\Omega$. When $H_f$ ($H_h$) is the null hypothesis, the minimal support of $f$ ($h$) does not depend on $\theta$ ($\gamma$).

**Assumption 3:** (a) $|\log(f(u,\theta))| \leq m(u)$ for all $\theta$ in $\Theta$, where $m$ is integrable with respect to $G$; and (b) $E(\log(f(u,\theta)))$ has a unique maximum at $\theta_*$ in $\Theta$.

The existence of the QMLE's $\hat{\theta}_n$ and $\hat{\gamma}_n$ are guaranteed by Assumptions 1-3 and they also ensure the estimates converge almost surely to $\theta_*$ and $\gamma_*$, the parameter values which maximize the expected log-likelihood [White (1982, theorems 2.1, 2.2)]. Observe that under the hypothesis $H_f$, $\theta_*=\theta_0$. 
When the partial derivatives, expectations, and appropriate inverses exist, the following matrices are defined for $f$ with analogs for $h$, where the analogs for $h$ are not written explicitly:

$$A_n(\theta) = \{ n^{-1} \sum_{t=1}^{n} \partial^2 \log(f(U_t, \theta))/\partial \theta_i \partial \theta_j \},$$

$$B_n(\theta) = \{ n^{-1} \sum_{t=1}^{n} \partial \log(f(U_t, \theta))/\partial \theta_i \cdot \partial \log(f(U_t, \theta))/\partial \theta_j \},$$

$$A(\theta) = \{ E(\partial^2 \log(f(U_t, \theta))/\partial \theta_i \partial \theta_j) \},$$

$$B(\theta) = \{ E(\partial \log(f(U_t, \theta))/\partial \theta_i \cdot \partial \log(f(U_t, \theta))/\partial \theta_j) \},$$

$$C_n(\theta) = A_n(\theta)^{-1} B_n(\theta) A_n(\theta)^{-1},$$

$$C(\theta) = A(\theta)^{-1} B(\theta) A(\theta)^{-1}.$$

**Assumption 4:** $\partial \log(f(u, \theta))/\partial \theta_i$, $i=1,...,p$, are measurable functions of $u$ for each $\theta$ in $\Theta$ and continuously differentiable functions of $\theta$ for each $u$ in $\Omega$.

**Assumption 5:** $|\partial^2 \log(f(u, \theta))/\partial \theta_i \partial \theta_j|$ and $|\partial \log(f(u, \theta))/\partial \theta_i \cdot \partial \log(f(u, \theta))/\partial \theta_j|$, $i,j=1,...,p$, are dominated by functions integrable with respect to $G$ for all $u$ in $\Omega$ and $\theta$ in $\Theta$.

**Assumption 6:** (a) $\theta_*$ is interior to $\Theta$, and (b) $A(\theta_*)$ and $B(\theta_*)$ are non-singular.

**Assumption 7:** $(\log(f(u, \theta)/h(u, \gamma)))^2$ is dominated by a measurable function integrable with respect to $G$ for all $\theta, \gamma$ in $\Theta \times \Gamma$.

Assumptions 1-6 guarantee that $\sqrt{n}(\hat{\theta}_n - \theta_*)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_*)$ are
asymptotically normal and assumption 7 will allow the Lindeberg-Levy central limit theorem to be applicable.

**Assumption 8:** $|\partial \log [f(u,\theta)/h(u,\gamma)]f(u,\theta)/\partial \theta_i|$, $|\partial \log [f(u,\theta)/h(u,\gamma)]f(u,\theta)/\partial \gamma_j|$, $i=1,\ldots,p$, $j=1,\ldots,q$, are dominated for all $\theta$, $\gamma$ in $\Theta \times \Gamma$ by functions integrable with respect to $\nu$.

Assumption 8 allows one to take derivatives and limits inside the integral

$$\int \log [f(u,\theta)/h(u,\gamma)]f(u,\theta)\,d\nu.$$ 

Assumptions 1-8 are enough to establish the asymptotic normality of $\sqrt{n}T_f$.

Let $\nabla_\theta$ denote the gradient operator with respect to $\theta$ (with $\nabla_\gamma$ defined similarly). Define the following two functions:

(i) $\psi(\theta,\gamma) = \int \nabla_\theta (\log(f(u,\theta))) \cdot \log [f(u,\theta)/h(u,\gamma)]f(u,\theta)\,d\nu$, a $1 \times p$ vector function,

(ii) $w(\theta,\gamma)^2 = \int \left(\log [f(u,\theta)/h(u,\gamma)]\right)^2 f(u,\theta)\,d\nu - (\int \log [f(u,\theta)/h(u,\gamma)]f(u,\theta)\,d\nu)^2$

$$+ 2\psi(\theta,\gamma)A(\theta)^{-1}\psi(\theta,\gamma)' + 2\psi(\theta,\gamma)C(\theta)\psi(\theta,\gamma)'.$$ 

White shows that $w(\theta_0,\gamma_0)^2$ is the asymptotic variance of $\sqrt{n}T_f$. For a consistent estimate of $w(\theta_0,\gamma_0)^2$ let

$$\hat{w}_n^2 = n^{-1}\sum_{t=1}^n \left(\log [f(U_t,\hat{\theta}_n)/h(U_t,\hat{\gamma}_n)]\right)^2 - (\int \log [f(U_t,\hat{\theta}_n)/h(U_t,\hat{\gamma}_n)]f(U_t,\hat{\theta}_n)\,d\nu)^2$$ 

$$+ 2\psi(\hat{\theta}_n,\hat{\gamma}_n)A_n(\hat{\theta}_n)^{-1}\psi(\hat{\theta}_n,\hat{\gamma}_n)'.$$
To avoid having to evaluate integrals, let

\[ \hat{w}_n^2 = n^{-1} \sum_{t=1}^{n} \left( \log \left( f(U_t, \hat{\theta}_n) / h(U_t, \hat{\gamma}_n) \right) \right)^2 - \left( n^{-1} \sum_{t=1}^{n} \log \left( f(U_t, \hat{\theta}_n) / h(U_t, \hat{\gamma}_n) \right) \right)^2 \]

\[ + 2 \psi_n(\hat{\theta}_n, \hat{\gamma}_n) A_n(\hat{\theta}_n)^{-1} \psi_n(\hat{\theta}_n, \hat{\gamma}_n)', \]

where

\[ \psi_n(\theta, \gamma) = n^{-1} \sum_{t=1}^{n} \nabla_{\theta} (\log(f(U_t, \theta))) \cdot \log(f(U_t, \theta)) / h(U_t, \gamma). \]

Both of the above estimators were proposed by Cox (1962).

**Assumption 9:** \( A(\theta_0) = -B(\theta_0) \) under \( H_f \).

**Assumption 10:** \( |\partial \log(f(u, \theta)) / \partial \theta_i \cdot \log(f(u, \theta)) / h(u, \gamma)|, i = 1, \ldots, p, \) are dominated by functions integrable with respect to \( G \).

Assumption 9 allows \( w(\theta_0, \gamma) \) to be simplified and Assumption 10 ensures the consistency of \( \hat{w}_n^2 \).

Assumptions 9 and 10 can be dropped by considering the following estimator of the asymptotic variance of \( \sqrt{n}T_f \), which is consistent whether \( H_f \) or \( H_h \) holds or not:

\[ \hat{w}_n^2 = n^{-1} \sum_{t=1}^{n} \left( \log \left( f(U_t, \hat{\theta}_n) / h(U_t, \hat{\gamma}_n) \right) \right)^2 - \left( n^{-1} \sum_{t=1}^{n} \log \left( f(U_t, \hat{\theta}_n) / h(U_t, \hat{\gamma}_n) \right) \right)^2 \]

\[ + 2 \psi_n(\hat{\theta}_n, \hat{\gamma}_n) A_n(\hat{\theta}_n)^{-1} \psi_n(\hat{\theta}_n, \hat{\gamma}_n)' + \psi_n(\hat{\theta}_n, \hat{\gamma}_n) C_n(\hat{\theta}_n) \psi_n(\hat{\theta}_n, \hat{\gamma}_n)'. \]
The main result can now be given.

Theorem (White, 1982):

(i) Given Assumption 1, Assumptions 2-6 for $f$ and $h$, and Assumptions 7-9, $\hat{w}_n^{-1} \sqrt{n} T_f$ is distributed asymptotically normal as $N(0,1)$ under $H_f$, provided $w(\theta_0, \gamma_*)^2 > 0$.

(ii) Given Assumption 1, Assumptions 2-6 for $f$ and $h$, and Assumptions 7-10, $\hat{w}_n^{-1} \sqrt{n} T_f$ is distributed asymptotically normal as $N(0,1)$ under $H_f$, provided $w(\theta_0, \gamma_*)^2 > 0$.

(iii) Given Assumption 1, Assumptions 2-6 for $f$ and $h$, and Assumptions 7 and 8, $\hat{w}_n^{-1} \sqrt{n} T_f$ is distributed asymptotically normal as $N(0,1)$ under $H_f$, provided $w(\theta_0, \gamma_*)^2 > 0$. 
1.2 Errors in Variables for Linear Models

1.2.1 Introduction

Errors in variables models (measurement error models) have been explored since the latter part of the 19th century when Adcock (1877, 1878) investigated estimation properties under somewhat restrictive but realistic assumptions in simple linear regression models. Since then much has been accomplished in the way of estimation and hypothesis testing in error in variables models, especially in the past 10 to 15 years.

Fuller's book, Measurement Error Models (1987), represents the most comprehensive single source of information on errors in variables models here to date. Fuller's book covers the topic of errors in variables in simple linear regression models to multivariate linear regression models to nonlinear regression models. The book's emphasis is placed on estimation techniques, which includes estimating true values for the fixed model and predicting true values for the random model.

Fuller's book is used as the main source of reference for errors in variables models here. The following section presents errors in variables in univariate linear models for vector explanatory variables using the notation of Fuller. The approach here is simply to present the basic ideas governing errors in variables models.

1.2.2 Maximum Likelihood Estimation

Let the errors in variables model be defined as follows:

\[ y_t = x_t \beta, \]

\[ (Y_t, X_t) = (y_t, x_t) + (e_t, u_t), \]

for \( t = 1, 2, \ldots, n \), where \( \{x_t\} \) is a sequence of \( k \)-dimensional row vectors of true values and
\( \epsilon_t = (e_t, u_t)' \) is the vector of measurement errors.

The idea behind errors in variables models is that instead of measuring the true value \( x_t \), which is unobservable, one observes the sum

\[ X_t = x_t + u_t, \]

where \( u_t \) is a random variable. The observed variable \( X_t \) is sometimes called the \textit{manifest} variable or the \textit{indicator} variable. The unobserved variable \( x_t \) is sometimes called the \textit{latent} variable. If the \( x_t \) are considered fixed, then the model is referred to as a \textit{functional} model; if the \( x_t \) are considered stochastic, then the model is referred to as a \textit{structural} model.

Rewrite the defining errors in variables model as

\[ z_t \alpha = 0, \quad Z_t = z_t + \epsilon_t, \]

where \( z_t = (y_t, x_t) \), \( Z_t = (Y_t, X_t) \), and

\[ \alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) = (1, -\beta'). \]

The following theorem provides maximum likelihood estimation for the functional model when the error covariance structure is known up to a scalar multiple.

Theorem (Fuller, Theorem 2.3.1, pp 124-125): Let the above model hold and let \( \{x_t\} \) be fixed. Let \( \epsilon_t \sim N(0, \Sigma_{\epsilon \epsilon}) \) (independent normal), where \( \Sigma_{\epsilon \epsilon} = \Gamma_{\epsilon \epsilon} \sigma^2 \) and \( \Gamma_{\epsilon \epsilon} \) is known. Then the maximum likelihood estimators of \( \beta \) and \( \sigma^2 \) are
\[ \hat{\beta} = (M_{XX} - \hat{\lambda}T_{uu})^{-1}(M_{XY} - \hat{\lambda}T_{ue}), \]

\[ \hat{\sigma}^2_m = (k + 1)^{-1}\hat{\lambda}, \]

where \( M_{ZZ} = n^{-1} \sum_{t=1}^{n} Z_t'Z_t \), with \( M_{XX} \) and \( M_{XY} \) defined similarly, and \( \hat{\lambda} \) is the smallest root of

\[ |M_{ZZ} - \lambda T_{ee}| = 0. \]

The maximum likelihood estimator of \( z_t, t=1,2,\ldots,n \), is

\[ \hat{z}_t = Z_t - (Y_t - X_t\hat{\beta})[(1, -\hat{\beta}')(\lambda T_{ee}(1, -\hat{\beta}'))^{-1}(1, -\hat{\beta}')]'. \]

For the functional error in variables model with vector explanatory variables, maximum likelihood estimation fails to produce consistent estimators of all parameters (see Fuller, Section 2.2). The problem can be viewed as essentially a “degrees of freedom” problem. This problem does not exist for the structural model.

Fuller gives the limiting distribution of the estimators for the functional model when an estimator of \( \Sigma_{ee} \) is proposed (Theorem 2.3.2, pp 127-129). For the structural model, the majority of the results developed for the estimator of \( \beta \) for the functional model are also appropriate.

1.2.3 Remarks

The model in Section 1.2.2 is considered a model with no error in the equation, i.e., the true value \( y_t \) has exactly a linear relationship with the true value \( x_t \). If the covariance matrices for the measurement error are estimated, then one would need to specify a model with an error in the equation since the random variable \( e_t \) may be
composed of two parts.

Fuller devotes a considerable amount of coverage to the simple linear case. Most, but not all, the results for simple linear errors in variables models can be extended to models with vector explanatory variables.

To reiterate, Section 1.2.2 presented a very brief discussion of the errors in variables model and was not meant as a comprehensive review. Many issues in errors in variables models exist, such as instrumental variable estimation, factor analysis, models with unequal error variances, and results when the assumption of normality of error terms is violated. Additional results for errors in variables models will be presented as needed.
CHAPTER 2
Testing Separate Normal Structural Models

2.1 Introduction

This section presents the derivation of Cox's test statistic for testing two normal structural models. Pesaran (1974) derived Cox's test statistic for the normal functional model with no errors in the explanatory variables. White (1982) presented Cox's test statistic for the normal structural model with no errors in the explanatory variables.

In the following derivation, we follow along the lines of Pesaran. In brief, Pesaran used an asymptotically unbiased estimate of the true expectations under the null hypothesis. The estimates required an appropriate use of partial derivatives and the information matrix. In contrast, White computed Cox's test statistic by rewriting the joint density as the conditional density multiplied by the marginal density, then he used the true expectations under the null hypothesis.

2.1.1 Hypotheses and Assumptions

Define the null and alternative hypotheses as follows:

\[ H_0 : Y = \beta_0 + x\beta_1 + \epsilon \]
\[ X = x + \delta_x \]

\[ H_1 : Y = \gamma_0 + z\gamma_1 + u \]
\[ Z = z + \delta_z \]
where x and z are random nxp and nxq matrices of unobservable explanatory variables.

Y is an nx1 column vector of observed, continuous response measurements, \( \beta_0 \) and \( \beta_1 \) are n and p dimensional column vectors, \( \gamma_0 \) and \( \gamma_1 \) are n and q dimensional column vectors, and \( \epsilon \) and \( u \) are nx1 column vectors of measurement errors in the continuous response vector Y. Let \( \beta_0 = \beta_0 1_n \) and \( \gamma_0 = \gamma_0 1_n \), where \( 1_n \) is an nx1 column vector of 1's and \( \beta_0 \) and \( \gamma_0 \) are scalars.

The null and alternative hypotheses are assumed to be non-nested, i.e., all the columns in X cannot be obtained from those of Z and vice-versa. It should be noted, however, that X and Z may contain a few common explanatory variables, say \( 0 \leq k < \min(p,q) \). Let the common explanatory variables represent the first k corresponding columns of X and Z.

Under the null hypothesis, \( H_0 \), we observe \( X = x + \delta_x \), where \( \delta_x \) is the nxp matrix of measurement errors in the explanatory variables. Analogously, under the alternative hypothesis, \( H_1 \), we observe \( Z = z + \delta_z \), where \( \delta_z \) is the nxq matrix of measurement errors in the explanatory variables.

It is preferable to define the null and alternative models using the i-th response measurement, \( Y_i \), and the appropriate rows of the explanatory variables. Define the following:

\[
Y = (Y_1, \ldots, Y_n)', \quad \text{where the } Y_i \text{'s are independent,}
\]

\[
X_i = \text{Row}_i(X), x_i = \text{Row}_i(x), \quad \text{and } \delta_{z_i} = \text{Row}_i(\delta_x), \quad \text{where } \text{Row}_i(\cdot) \text{ stands for } i \text{-th row,}
\]

\[
\epsilon = (\epsilon_1, \ldots, \epsilon_n)', \quad \text{where the } \epsilon_i \text{'s are iid,}
\]

\[
Z_i = \text{Row}_i(Z), z_i = \text{Row}_i(z), \quad \text{and } \delta_{z_i} = \text{Row}_i(\delta_z),
\]

\[
u = (u_1, \ldots, u_n)', \quad \text{where the } u_i \text{'s are iid.}
\]
Thus, the null and alternative hypotheses can be expressed as:

\[ H_0: Y_i = \beta_0 + x_i \beta_1 + \epsilon_i \]
\[ X_i = x_i + \delta x_i \]

\[ H_1: Y_i = \gamma_0 + z_i \gamma_1 + u_i \]
\[ Z_i = z_i + \delta z_i \]

for \( i = 1, \ldots, n \). For the normal structural models, assume the following for \( i = 1, \ldots, n \):

\[
\begin{bmatrix}
 x'_{i} \\
 \epsilon_i \\
 \delta'_{x_i}
\end{bmatrix}
\sim
\text{NI}
\begin{bmatrix}
 \mu_x' \\
 0 \\
 0
\end{bmatrix},
\begin{bmatrix}
 \Sigma_{x'x} & 0 & 0 \\
 0 & \sigma^2_\epsilon & 0 \\
 0 & 0 & \Sigma_{\delta'_x \delta_x}
\end{bmatrix}
\]

\[
\begin{bmatrix}
 z'_{i} \\
 u_i \\
 \delta'_{z_i}
\end{bmatrix}
\sim
\text{NI}
\begin{bmatrix}
 \mu_z' \\
 0 \\
 0
\end{bmatrix},
\begin{bmatrix}
 \Sigma_{z'z} & 0 & 0 \\
 0 & \sigma^2_u & 0 \\
 0 & 0 & \Sigma_{\delta'_z \delta_z}
\end{bmatrix}
\]

Also assume that \( \Sigma_{\delta'_x \delta_x} \) and \( \Sigma_{\delta'_z \delta_z} \) are known. Observe that the above assumptions indicate that the measurement errors are independent and also, are independent of the unobservable explanatory variables. In addition, the \( Y_i \)'s are identically distributed. NI denotes Normal Independent.
Fuller (1987, p. 105) shows that the maximum likelihood estimates of the parameters for the null hypothesis are given by

$$(\hat{\mu}_z, \hat{\beta}_0) = (\bar{X}, \bar{Y} - \bar{X}\hat{\beta}_1),$$

$$\hat{\beta}_1 = (m_{X'y} - \Sigma_{\delta_z'\delta_z})^{-1}m_{X'Y},$$

$$\hat{\sigma}^2 = m_{YY} - 2m_{YX}\hat{\beta}_1 + \hat{\beta}_1'm_{X'X}\hat{\beta}_1 - \hat{\beta}_1'\Sigma_{\delta_z'\delta_z}\hat{\beta}_1,$$

where $\hat{\Sigma}_{x'x} = m_{X'X} - \Sigma_{\delta_z'\delta_z}$ provided $\hat{\Sigma}_{x'x}$ is positive definite and $\hat{\sigma}^2 \geq \Sigma_{\epsilon'\epsilon} - \Sigma_{\delta_z'\delta_z} \Sigma_{\delta_z'\epsilon} = 0$, where $\Sigma_{\delta_z'\delta_z}$ is the Moore-Penrose generalized inverse of $\Sigma_{\delta_z'\delta_z}$ and $\Sigma_{\delta_z'\epsilon} = 0$ by assumption.

The maximum likelihood estimates of the parameters for the alternative hypothesis are similarly defined:

$$(\hat{\mu}_z, \hat{\gamma}_0) = (\bar{Z}, \bar{Y} - \bar{Z}\hat{\gamma}_1),$$

$$\hat{\gamma}_1 = (m_{Z'Z} - \Sigma_{\delta_z'\delta_z})^{-1}m_{Z'Y},$$

$$\hat{\sigma}_u^2 = m_{YY} - 2m_{YZ}\hat{\gamma}_1 + \hat{\gamma}_1'm_{Z'Z}\hat{\gamma}_1 - \hat{\gamma}_1'\Sigma_{\delta_z'\delta_z}\hat{\gamma}_1,$$

where $\hat{\Sigma}_{x'x} = m_{Z'Z} - \Sigma_{\delta_z'\delta_z}$ provided $\hat{\Sigma}_{x'x}$ is positive definite and $\hat{\sigma}_u^2 \geq \Sigma_{\epsilon'\epsilon} - \Sigma_{\delta_z'\delta_z} \Sigma_{\delta_z'u} = 0$, where $\Sigma_{\delta_z'\delta_z}$ is the Moore-Penrose generalized inverse of $\Sigma_{\delta_z'\delta_z}$ and $\Sigma_{\delta_z'u} = 0$ by assumption.
For notation, lowercase letter $m$, appropriately subscripted, is used for the sample covariance matrix:

$$m_{Z'Z} = (n-1)^{-1} \sum_{i=1}^{n} (Z_i - \bar{Z})(Z_i - \bar{Z})',$$

where $\bar{Z} = n^{-1} \sum_{i=1}^{n} Z_i$.

2.1.2 Derivation of Cox's Test

To derive Cox's test statistic, we follow along the lines of Pesaran (1974). Let the log-likelihood functions for $H_0$ and $H_1$ be denoted by $L_0(\theta_0)$ and $L_1(\theta_1)$, where $\theta_0 = [\mu_x, \sigma_x^2, \beta_0, \beta'_1, (\text{vech } \Sigma_{x'}x)]'$ and $\theta_1 = [\mu_x, \sigma_x^2, \gamma_0, \gamma'_1, (\text{vech } \Sigma_{x'}x)]'$. Also, let $\hat{L}_{10}$ denote the maximum likelihood ratio of $L_{10}$. Cox's test statistic for testing $H_0$ against $H_1$ can be written as

$$T_0 = T_0(\hat{\theta}_0) = \hat{L}_{10} - (n-1)[P\lim_{n \to \infty} (\hat{L}_{10}/(n-1))]_{\theta = \hat{\theta}_0},$$

where $P\lim(\cdot)$ denotes the probability limit and the probability limit is taken under $H_0$, with the true $\theta$ replaced by its maximum likelihood estimate, $\hat{\theta}_0$, under $H_0$. Here, $\hat{L}_{10} = L_0(\hat{\theta}_0) - L_1(\hat{\theta}_1)$, where $\hat{\theta}_0$ and $\hat{\theta}_1$ are the maximum likelihood estimators of $\theta_0$ and $\theta_1$. We require that $P\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i'z_j = \Sigma_{x'x} = \text{Cov}(x,z)$ and $P\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \delta_{x_i} = \Sigma_{\delta'x}$ exist and not be equal to zero. We also require that $\Sigma_{x'x}$, $\Sigma_{\delta'x}$, $\Sigma_{x'\delta}$, $\Sigma_{\delta'\delta}$ all be non-singular matrices.

Cox's idea is to compare the maximum likelihood ratio with its expected value under the null hypothesis, $H_0$. The true expectation of the likelihood is taken and then the parameter, $\theta$, is replaced by its maximum likelihood estimate. In the above expression for $T_0$, the true expectation of the maximum likelihood ratio is replaced by
the probability limit (under $H_0$) of the maximum likelihood ratio divided by $n-1$. The probability limit under $H_0$ can be shown to converge to the true expectation using Khintchine's Weak Law of Large Numbers (WLLN).

Under Cox's formulation, $T_0$ is asymptotically normally distributed with mean zero and variance given by $V_0(T_0)$:

$$V_0(T_0) = V_0(L_{10}) - \frac{1}{n} \eta' \lim_{n \to \infty} (nI_0^{-1}) \eta,$$

where we define $L_{10} = L_0(\theta_0) - L_0(\theta_{10})$ and $\theta_{10}$ is the asymptotic expectation of $\hat{\theta}_1$ under $H_0$.

$I_0 = I_0(\theta_0)$ is the information matrix of $\theta_0$ and $\eta$ is given by

$$\eta = \eta(\theta_0) = (n-1) \frac{\partial [\lim_{n \to \infty} (L_{10}/(n-1))]}{\partial \theta_0}.$$

The log-likelihood functions of the independent and identically distributed normal random vectors $(Y_i, X_i)$ and $(Y_i, Z_i)$ under $H_0$ and $H_1$ for $i=1,...,n$ can be written as follows:

$$L_0(\theta_0) = -\frac{n}{2}(p+1)\log 2\pi - \frac{(n-1)}{2} \log |\Sigma_0|$$

$$- \frac{(n-1)}{2} \sum_{i=1}^{n} (Y_i - \beta_0 - \mu_x \beta_1, (X_i - \mu_x) \Sigma_0^{-1} (Y_i - \beta_0 - \mu_x \beta_1, (X_i - \mu_x)')$$
\[
L_1(\theta_1) = -\frac{n}{2}(q+1)\log 2\pi - \frac{n-1}{2} \log |\Sigma_1| \\
- \frac{n-1}{2} \sum_{i=1}^{n} (Y_i - \gamma_0 - \mu_z \gamma_1, (Z_i - \mu_z)) \Sigma_1^{-1} (Y_i - \gamma_0 - \mu_z \gamma_1, (Z_i - \mu_z))',
\]

where

\[
\Sigma_0 = \begin{bmatrix}
\beta_1' \Sigma_{x'x} \beta_1 + \sigma^2 & \beta_1' \Sigma_{x'x} \\
\Sigma_{x'x} \beta_1 & \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x}
\end{bmatrix}
\]

and

\[
\Sigma_1 = \begin{bmatrix}
\gamma_1' \Sigma_{z'z} \gamma_1 + \sigma^2 & \gamma_1' \Sigma_{z'z} \\
\Sigma_{z'z} \gamma_1 & \Sigma_{z'z} + \Sigma_{\delta_z'\delta_z}
\end{bmatrix}
\]
The log-likelihood functions of \((Y_i, X_i)\) and \((Y_i, Z_i)\) for \(i=1,\ldots, n\) evaluated at \(\hat{\theta}_0\) and \(\hat{\theta}_1\) are given by

\[
L_0(\hat{\theta}_0) = -\frac{n}{2}(p+1)\log 2\pi - \frac{(n-1)}{2}(p+1) - \frac{(n-1)}{2}\log |\Sigma_0|
\]

and

\[
L_1(\hat{\theta}_1) = -\frac{n}{2}(q+1)\log 2\pi - \frac{(n-1)}{2}(q+1) - \frac{(n-1)}{2}\log |\Sigma_1|.
\]

We can then derive

\[
L_{10} = L_0(\hat{\theta}_0) - L_1(\hat{\theta}_1) = -\frac{n}{2}(p-q)\log 2\pi - \frac{(n-1)}{2}(p-q) + \frac{(n-1)}{2}\log \frac{|\Sigma_1|}{|\Sigma_0|}.
\]

Now, under \(H_0\) it can be shown that \(\hat{\sigma}_u^2\) converges in probability to \(\sigma_{u0}^2\), where

\[
\sigma_{u0}^2 = \sigma_z^2 + \beta_1'(\Sigma_Z'Z - \Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\beta_1)\beta_1.
\]

by repeated use of Khintchine's Weak Law of Large Numbers. The asymptotic expectation of \(\hat{\gamma}_0\) and \(\hat{\gamma}_1\) under \(H_0\) can be shown to be given by

\[
\gamma_{00} = \beta_0 + \mu_x\beta_1 - \mu_z\gamma_{10},
\]

\[
\gamma_{10} = \Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\beta_1,
\]

where \(\Sigma_{Z'Z} = \Sigma_{Z'Z} - \Sigma_{Z'Z}\delta_z\delta_z'\) and \(\Sigma_{Z'X} = \Sigma_{Z'X} + \Sigma_{Z'X}\delta_x\delta_x'\). We note that \(\gamma_{10}\) is the "specification bias" discussed by Theil (1957) for the case of stochastic explanatory variables measured with error.
Now observe that

\[ \lim_{n \to \infty} \left( \hat{L}_{10}/(n-1) \right) = -\frac{1}{2}(p-q) \log 2\pi - \frac{1}{2}(p-q) - \frac{1}{2} \log \frac{|\Sigma_{10}|}{|\Sigma_0|}, \]

where

\[ \Sigma_{10} = \begin{bmatrix} \gamma_{10}' \Sigma_{zz} \gamma_{10} + \sigma^2_{u0} & \gamma_{10}' \Sigma_{xz} \\ \Sigma_{xz} \gamma_{10} & \Sigma_{xz} + \Sigma_{\delta_{z} \delta_{z}} \end{bmatrix} \]

\[ = \begin{bmatrix} \beta_1' \Sigma_{xz} \beta_1 + \sigma^2_{e} & \beta_1' \Sigma_{xz} \\ \Sigma_{xz} \beta_1 & \Sigma_{xz} + \Sigma_{\delta_{z} \delta_{z}} \end{bmatrix} \]

is the probability limit of \( \hat{\Sigma}_1 \) under \( H_0 \).

Cox's test statistic, ignoring constant term, can then be written as

\[ T_0 = \hat{L}_{10} - (n-1)\left[ \lim_{n \to \infty} \left( \hat{L}_{10}/(n-1) \right) \right]_{\theta = \hat{\theta}_0} = \frac{(n-1)}{2} \log \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_{10}|}, \]

where \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_{10} \) are obtained by replacing the unknown parameters with their maximum likelihood estimates.
To derive $V_0(T_0)$, we first consider $L_{10} = L_0(\theta_0) - L_1(\theta_{10})$:

$$L_{10} = -\frac{n}{2}(p+q)\log 2\pi - \frac{(n-1)}{2} \log \frac{\|\Sigma_{10}\|}{\|\Sigma_0\|}$$

$$= -\frac{n}{2}(p+q)\log 2\pi - \frac{(n-1)}{2} \log \frac{\|\Sigma_{10}\|}{\|\Sigma_0\|}$$

$$- \frac{(n-1)}{2} \sum_{i=1}^{n} (Y_i - \beta_0 - \mu_x \beta_1, (X_i - \mu_x)') \Sigma_0^{-1} (Y_i - \beta_0 - \mu_x \beta_1, (X_i - \mu_x)')'$$

$$+ \frac{(n-1)}{2} \sum_{i=1}^{n} (Y_i - \gamma_0 - \mu_z \gamma_1, (Z_i - \mu_z)') \Sigma_1^{-1} (Y_i - \gamma_0 - \mu_z \gamma_1, (Z_i - \mu_z)')'$$

Thus, we have

$$V_0(L_{10}) = V_0[- \frac{(n-1)}{2} \sum_{i=1}^{n} ((x_i - \mu_x)\beta_1 + \epsilon_i, (X_i - \mu_x)) \Sigma_0^{-1} ((x_i - \mu_x)\beta_1 + \epsilon_i, (X_i - \mu_x))'$$

$$+ \frac{(n-1)}{2} \sum_{i=1}^{n} (z_i - \mu_z)\gamma_{10} + u_i, (Z_i - \mu_z) \Sigma_1^{-1} (z_i - \mu_z)\gamma_{10} + u_i, (Z_i - \mu_z)')']$$

We observe that the variance expression above is that of the sum of two quadratic forms which are themselves sums of independent and identically distributed quadratic forms by assumption.

We can then derive the following expression:

$$V_0(L_{10}) = \frac{n(n-1)^2}{4}(p + q + 2) + \frac{n(n-1)^2}{2} \text{tr}(\Sigma_0^{-1} \Sigma_1^{-1} \Sigma_*')$$
where

\[
\Sigma_* = \begin{bmatrix}
\gamma_1^0 \Sigma'_{x'x} \beta_1 + \sigma_\xi^2 & \beta_1' \Sigma_{x'} \\
\Sigma_{x'x} \gamma_{10} & \Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}}
\end{bmatrix}.
\]

Now we must derive \( \eta \) and \( I_{0} \) defined previously. First observe that

\[
|\Sigma_{10}| = |\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}}| \cdot |\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 - \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1|.
\]

and

\[
|\Sigma_{0}| = |\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}}| \cdot |\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 - \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1|,
\]

where the following two conditions hold since the matrices are variance-covariance matrices:

\[
\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 > \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1
\]

and

\[
\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 > \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1.
\]

Thus, we can write

\[
P_{\lim_{n \to \infty}} (L_{10}/(n-1)) = -\frac{1}{2}(p-q) \log 2\pi - \frac{1}{2}(p-q) - \frac{1}{2} \log \frac{|\Sigma_{10}|}{|\Sigma_{0}|}
\]

\[
= -\frac{1}{2}(p-q) \log 2\pi - \frac{1}{2}(p-q) - \frac{1}{2} \log \frac{|\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}}|}{|\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}}|}
\]

\[
- \frac{1}{2} \log (\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 - \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1)
\]

\[
+ \frac{1}{2} \log (\beta_1' \Sigma_{x'x} \beta_1 + \sigma_\xi^2 - \beta_1' \Sigma_{x'x} (\Sigma_{x'x} + \Sigma_{\delta'_{x} \delta_{x}})^{-1} \Sigma_{x'x} \beta_1).
\]
Now, we have

\[
\frac{\partial \text{Plim}_{n \to \infty} (\hat{L}_{10}/(n-1))}{\partial \mu_x} = 0 ,
\]

\[
\frac{\partial \text{Plim}_{n \to \infty} (\hat{L}_{10}/(n-1))}{\partial \sigma_{\epsilon}^2} = - \frac{1}{2} \frac{1}{\beta_1' \Sigma_{\epsilon' \epsilon} \beta_1 + \sigma_{\epsilon}^2 - \beta_1' \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1}
\]

\[
+ \frac{1}{2} \frac{1}{\beta_1' \Sigma_{\epsilon' \epsilon} \beta_1 + \sigma_{\epsilon}^2 - \beta_1' \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1},
\]

\[
\frac{\partial \text{Plim}_{n \to \infty} (\hat{L}_{10}/(n-1))}{\partial \beta_0} = 0 ,
\]

\[
\frac{\partial \text{Plim}_{n \to \infty} (\hat{L}_{10}/(n-1))}{\partial \beta_1} = - \left[ \frac{1}{\beta_1' \Sigma_{\epsilon' \epsilon} \beta_1 + \sigma_{\epsilon}^2 - \beta_1' \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1} \right]
\]

\[
\cdot [ \Sigma_{\epsilon' \epsilon} \beta_1 - \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1 ]
\]

\[
+ \left[ \frac{1}{\beta_1' \Sigma_{\epsilon' \epsilon} \beta_1 + \sigma_{\epsilon}^2 - \beta_1' \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1} \right]
\]

\[
\cdot [ \Sigma_{\epsilon' \epsilon} \beta_1 - \Sigma_{\epsilon' \epsilon} (\Sigma_{\epsilon' \epsilon} + \Sigma_{\delta' \delta})^{-1} \Sigma_{\epsilon' \epsilon} \beta_1 ] ,
\]
\[
\frac{\partial \text{Plim}_{n \to \infty} \left( \hat{L}_{10}/(n-1) \right)}{\partial \text{vech} \Sigma_{x'x}} = \text{vech} \left\{ \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} - \frac{1}{2} \text{diag} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \right\} \\
- \frac{1}{2} \left[ \frac{\text{vech} A_*}{2 \beta_1^* \Sigma_{x'x} \beta_1 + \sigma_1^2 - \beta_1^* \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \Sigma_{x'x} \beta_1} \right] \\
+ \frac{1}{2} \left[ \frac{\text{vech} B_*}{2 \beta_1^* \Sigma_{x'x} \beta_1 + \sigma_1^2 - \beta_1^* \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \Sigma_{x'x} \beta_1} \right]
\]

where

\[
A_* = \left( 2\beta_1 \beta_1' - \text{diag} \left( \beta_1 \beta_1' \right) \right) A_{**},
\]

\[
A_{**} = I - \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} I_{q \times p} (k) - I_{p \times q} (k) \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \Sigma_{x'x}
\]

\[
+ \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} I_{q \times q} (k) \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \Sigma_{x'x}
\]

and

\[
B_* = \left( 2\beta_1 \beta_1' - \text{diag} \left( \beta_1 \beta_1' \right) \right) B_{**},
\]

\[
B_{**} = I - \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} - \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-1} \Sigma_{x'x}
\]

\[
+ \Sigma_{x'x} \left( \Sigma_{x'x} + \Sigma_{\delta_x'\delta_x} \right)^{-2} \Sigma_{x'x}
\]

\(I_{\text{sxt}}^{(k)}\) is an sxt matrix which has a kxk identity matrix in its upper left block and zeros everywhere else. If k=0 then \(I_{\text{sxt}}^{(k)} = 0\). Recall that k is the number of common explanatory variables in the models of \(H_0\) and \(H_1\).

The above expressions provide the components of

\[
\eta = \eta(\theta_0) = (n-1) \frac{\partial \text{Plim}_{n \to \infty} \left( \hat{L}_{10}/(n-1) \right)}{\partial \theta_0}.
\]
Likewise, the components of the information matrix under $H_0$, $I_0$, can be computed and displayed. Due to the complexity of computing the information matrix with the log-likelihood for $(Y_i, X_i)'$, $i=1,...,n$, we instead compute $I_0$ using the log-likelihood function for $(x_i, \epsilon_i, \delta_{x_i})'$, $i=1,...,n$.

The information matrix is thus given by

$$
I_0 = I_0(\theta_0) = \left[
\begin{array}{cccccc}
n\Sigma_{x'x} & 0 & 0 & 0 & 0 \\
0 & \frac{n+1}{2\sigma_\epsilon^4} & 0 & 0 & 0 \\
0 & 0 & \frac{n}{\sigma_\epsilon^2} & \frac{n\mu_x}{\sigma_\epsilon^2} & 0 \\
0 & 0 & \frac{n\mu_x}{\sigma_\epsilon^2} & \frac{n}{\sigma_\epsilon^2} \Sigma_{x'x}^{-1} & 0 \\
0 & 0 & 0 & 0 & C_*
\end{array}
\right]
$$

where

$$
C_* = \frac{(n-1)^2}{2} \{\text{vech}(2\Sigma_{x'x}^{-1} - \text{diag}(\Sigma_{x'x}^{-1}))\} \cdot \{\text{vech}(2\Sigma_{x'x}^{-1} - \text{diag}(\Sigma_{x'x}^{-1}))\}'
$$

$$
- \frac{n(n-1)}{2} \{\text{vech}(2\Sigma_{x'x}^{-1} - \text{diag}(\Sigma_{x'x}^{-1}))\} \cdot \{\text{vech}(\Sigma_{x'x}^{-1})\}'
$$

$$
+ \frac{n^2}{4} \{\text{vech}(\Sigma_{x'x}^{-1})\} \cdot \{\text{vech}(\Sigma_{x'x}^{-1})\}'.
$$
Since the information matrix is in block diagonal form, it is straight-forward to obtain its inverse, \( R_0^{-1}(\theta_0) \).

Finally, we have all the components for \( V_0(T_0) \) and we can obtain \( \hat{V}_0(T_0) \) by replacing the parameters with their estimates, assuming \( H_0 \) is true.

Now, by Cox's formulation, \( T_0/[V_0(T_0)]^{1/2} \) has an asymptotically standard normal distribution under \( H_0 \).

To summarize, we have derived Cox's test of non-nested linear models for the case of stochastic explanatory variables measured with error by following along the lines of Pesaran (1974). Pesaran derived Cox's test under the assumptions of no measurement errors in the explanatory variables and fixed (non-stochastic) explanatory variables.

2.2 Consistency of Test

By definition (see Lehmann (1986) ), a test is said to be consistent if for any member of the alternative hypothesis, \( H_1 \), the probability of rejecting the null hypothesis, \( H_0 \), tends to one as the sample size goes to infinity.

The exact distribution of Cox's test derived in section 2.1 cannot be obtained since it depends on unknown parameters. To investigate the consistency of Cox's test, we must consider asymptotic behavior of the test.

Recall that Cox's test for testing separate normal structural models as stated in section 2.1 is given by

\[
T_0 = \hat{L}_{10} - (n-1)[P_{\theta=\theta_0} \lim_{n \to \infty} (\hat{L}_{10}/(n-1))] = \frac{(n-1)}{2} \log \frac{|\hat{\Sigma}_1|}{|\Sigma_{10}|}
\]

Under \( H_1 \), \( T_0 \) should have a negative mean asymptotically. So if the test, \( T_0 \), converges in probability to a negative quantity under \( H_1 \) then the test will be consistent.
2.2.1 A Sufficient Condition

It can be shown that under $H_1$, \( \frac{1}{(n-1)} T_0 \) converges in probability to $T_{01}$, where

\[
T_{01} = \frac{1}{2} \log \frac{|\Sigma_1|}{|\Sigma_{10}|}
\]

and

\[
\Sigma_1 = \begin{bmatrix}
\gamma_1' \Sigma' x z \gamma_1 + \sigma_u^2 & \gamma_1' \Sigma' x z \\
\Sigma' x z \gamma_1 & \Sigma' x z + \Sigma' \delta' z \delta z
\end{bmatrix}
\]

\[
\Sigma_{10} = \begin{bmatrix}
\gamma_1' \Sigma' x z \gamma_1 + \sigma_u^2 & \gamma_1' \Sigma' x z \Sigma' \Sigma' x z \\
\Sigma' x z \Sigma' \Sigma' x z \gamma_1 & \Sigma' x z + \Sigma' \delta' z \delta z
\end{bmatrix}
\]

Observe that $T_{01}$ is negative if and only if $|\Sigma_{10}^*| > |\Sigma_1|$. It can be shown, via matrix theory, that

\[
|\Sigma_1| = |\Sigma' x z + \Sigma' \delta' z \delta z| \cdot \gamma_1' \Sigma' x z \gamma_1 + \sigma_u^2 - \gamma_1' \Sigma' x z (\Sigma' x z + \Sigma' \delta' z \delta z)^{-1} \Sigma' x z \gamma_1|
\]

and

\[
|\Sigma_{10}^*| = |\Sigma' x z + \Sigma' \delta' z \delta z| \cdot \gamma_1' \Sigma' x z \gamma_1 + \sigma_u^2
\]

\[
- \gamma_1' \Sigma' x z \Sigma' \Sigma' x z (\Sigma' x z + \Sigma' \delta' z \delta z)^{-1} \Sigma' x z \Sigma' \Sigma' x z \Sigma' x z \gamma_1|.
\]
Thus, $|\Sigma_{10}^*| > |\Sigma_1|$ if and only if

$$
\gamma_1' \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z} \gamma_1 > 
\gamma_1' \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} \gamma_1.
$$

If we let

$$
U = \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z}
$$

and

$$
V = \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z},
$$

then $|\Sigma_{10}^*| > |\Sigma_1|$ if and only if $\gamma_1' (U - V) \gamma_1 > 0$. By definition, the latter statement, $\gamma_1' (U - V) \gamma_1 > 0$, is achieved if and only if $U - V$ is a positive definite matrix (since $U - V$ is symmetric).

From the previous discussion, a sufficient condition for Cox’s test to be consistent for testing separate normal structural models is now given by the following theorem.

Theorem 2.2.1 Assume the conditions stated in section 2.1 for testing separate normal structural models. A sufficient condition for Cox’s test to be consistent is that $U - V$ is positive definite, where

$$
U = \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z}
$$

and

$$
V = \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} (\Sigma_{z'z} + \Sigma_{z'\delta_z} \Sigma_{z'z})^{-1} \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z}.
$$

A further simplification of the above result can be given. Suppose we let

$$
\Sigma_{z'z'z'z} = \Sigma_{z'z} - \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z}.
$$
Thus, we have
\[
U - V = \Sigma_{x'z} (\Sigma_{x'x} + \Sigma_{\delta'z\delta_z})^{-1} \Sigma_{x'z} \\
- (\Sigma_{x'z} - \Sigma_{x'z'x'}) (\Sigma_{x'x} + \Sigma_{\delta'z\delta_z})^{-1} (\Sigma_{x'z} - \Sigma_{x'z'x'}) \\
= 2 \Sigma_{x'z'x'} (\Sigma_{x'x} + \Sigma_{\delta'z\delta_z})^{-1} (\Sigma_{x'z} - \Sigma_{x'z'x'}) \\
+ \Sigma_{x'z'x'} (\Sigma_{x'x} + \Sigma_{\delta'z\delta_z})^{-1} \Sigma_{x'z'x'}
\]
This latter expression is positive definite if \( \Sigma_{x'z'x'} \) is positive definite.

Corollary 2.2.1 A sufficient condition for Cox’s test to be consistent is that \( \Sigma_{x'z'x'} \) is positive definite, where
\[
\Sigma_{x'z'x'} = \Sigma_{x'z'} - \Sigma_{x'x} \Sigma_{x'x}^{-1} \Sigma_{x'z'}.
\]

It should be noted, however, that the converse of Theorem 2.2.1 isn’t necessarily true, i.e., consistency of the test does not imply that \( U - V \) is positive definite in general. Observe that under \( H_1 \) the asymptotic expectation of \( T_0 \) should be a negative quantity. Hence, \( T_0 \) is of order \( n \) in probability (\( O_p(n) \)) so that \( U - V \) positive definite is a sufficient condition for the test to be consistent. However, suppose we know the test is consistent. Then \( T_0 \) is of order \( n \) in probability still, but \( U - V \) could be positive semidefinite, i.e., \( \gamma_1(U - V) \gamma_1 \geq 0 \).

2.3 Interchanging Hypotheses

As discussed in Chapter 1 (page 4), the hypotheses \( H_0 \) and \( H_1 \) are to be considered asymmetrically. Thus, it is instructive to investigate Cox’s test with \( H_1 \) as the null hypothesis and \( H_0 \) as the alternative hypothesis.

A test statistic \( T_1 \) and variance \( V_1(T_1) \) can be derived using the Cox formulation for testing \( H_1 \) versus \( H_0 \) just as before.
CHAPTER 3
Testing Separate Normal Functional Models

3.1 Introduction

This section presents the derivation of Cox's test statistic for testing two normal functional models. As mentioned in section 3, Pesaran (1974) derived Cox's test statistic for the normal functional model with no errors in the explanatory variables.

Again, we follow along the lines of Pesaran (1974). Estimation in the normal functional model requires that we estimate the fixed but unobservable "true" explanatory variables, treating them as nuisance parameters. Maximum likelihood estimation will be employed to estimate all parameters. However, maximum likelihood estimation fails to yield consistent estimators of all parameters in the models since the number of parameters increases with the sample size, n (see Fuller, p 104).

Fuller (1987) gives the maximum likelihood estimates for the normal functional model under the assumption that the entire error covariance matrix is known up to a scalar multiple. However, we will assume that the error covariance matrix of the explanatory variables is known and the within-subject error variance is unknown. This assumption leads to analogous, but slightly different estimates of some parameters in the models than what is given by Fuller.
3.1.1 Hypotheses and Assumptions

Define the null and alternative hypotheses as follows:

$$H_0: Y = x\beta_1 + \epsilon$$
$$X = x + \delta_x$$

$$H_1: Y = z\gamma_1 + u,$$
$$Z = z + \delta_z$$

where $x$ and $z$ are fixed nxp and nxq matrices of unobservable explanatory variables.

Again, the null and alternative hypotheses are assumed to be non-nested, i.e., all the columns in $X$ cannot be obtained from those of $Z$ and vice-versa. It should be noted, however, that $X$ and $Z$ may contain a few common explanatory variables, say $0 \leq k < \min(p,q)$. Let the common explanatory variables represent the first $k$ corresponding columns of $X$ and $Z$.

We observe $X = x + \delta_x$, where $\delta_x$ is the nxp matrix of measurement errors in the explanatory variables under the null hypothesis, $H_0$. Analogously we observe $Z = z + \delta_z$, where $\delta_z$ is the nxq matrix of measurement errors in the explanatory variables under the alternative hypothesis, $H_1$.

$Y$ is an nx1 column vector of observed, continuous response measurements, $\beta_1$ is a p dimensional column vector, $\gamma_1$ is a q dimensional column vector, and $\epsilon$ and $u$ are nx1 column vectors of measurement errors in the continuous response vector $Y$.

It is preferable to define the null and alternative models using the $i$-th response measurement, $Y_i$, and the appropriate rows of the explanatory variables. Define the following:

$$Y = (Y_1, \ldots, Y_n)'$$, where the $Y_i$'s are independent,
\( X_i = \text{Row}_i(X) \), \( x_i = \text{Row}_i(x) \), and \( \delta x_i = \text{Row}_i(\delta x) \), where \( \text{Row}_i(\cdot) \) stands for \( i \)-th row,

\[
\epsilon = (\epsilon_1, \ldots, \epsilon_n)', \text{ where the } \epsilon_i \text{ are iid,}
\]

\( Z_i = \text{Row}_i(Z) \), \( z_i = \text{Row}_i(z) \), and \( \delta z_i = \text{Row}_i(\delta z) \),

\[
u = (u_1, \ldots, u_n)', \text{ where the } u_i \text{ are iid.}
\]

Thus, the null and alternative hypotheses can be expressed as:

\[
H_0: Y_i = x_i \beta_1 + \epsilon_i \\
X_i = x_i + \delta x_i
\]

\[
H_1: Y_i = z_i \gamma_1 + u_i, \\
Z_i = z_i + \delta z_i
\]

for \( i = 1, \ldots, n \). For the normal functional models, assume the following for \( i = 1, \ldots, n \):

\[
\begin{bmatrix}
\epsilon_i \\
\delta x_i
\end{bmatrix} \sim \text{NI}\left[
\begin{bmatrix}
0 \\
0
\end{bmatrix}, \begin{bmatrix}
\sigma_\epsilon^2 & 0 \\
0 & \Sigma_{\delta x, \delta x}
\end{bmatrix}\right],
\]

\[
\begin{bmatrix}
u_i \\
\delta z_i
\end{bmatrix} \sim \text{NI}\left[
\begin{bmatrix}
0 \\
0
\end{bmatrix}, \begin{bmatrix}
\sigma_u^2 & 0 \\
0 & \Sigma_{\delta z, \delta z}
\end{bmatrix}\right].
\]

Also assume that \( \Sigma_{\delta x, \delta x} \) and \( \Sigma_{\delta z, \delta z} \) are known. Observe that the above assumptions
indicate that the measurement errors are independent and independent of the unobservable explanatory variables.

The maximum likelihood estimates of the parameters for the null hypothesis can be shown to be given as

\[ \hat{x}_i = (\hat{\sigma}_\epsilon^2 \hat{\beta}_1 \hat{\beta}_1 + \Sigma_{\delta_x \delta_x}^{-1})^{-1}(\hat{\sigma}_\epsilon^2 Y_i \hat{\beta}_1 + \Sigma_{\delta_x \delta_x}^{-1} X_i), \]

\[ \hat{\beta}_1 = (M_{X'X} - \Sigma_{\delta_x \delta_x}^{-1})^{-1}M_{X'Y}, \]

\[ \hat{\sigma}_\epsilon^2 \text{ is the smallest positive root of } \sigma_\epsilon^4 + \sigma_\epsilon^2(2B_0 - A_0) + B_0^2 = 0, \]

where

\[ A_0 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i \hat{\beta}_1)^2, \]

\[ B_0 = \hat{\beta}_1 \Sigma_{\delta_x \delta_x} \hat{\beta}_1. \]

Using the quadratic formula, we can show that \( \hat{\sigma}_\epsilon^2 \) is the smallest positive quantity given by:

\[ \hat{\sigma}_\epsilon^2 = \frac{(A_0 - 2B_0) \pm \sqrt{A_0^2 - 4A_0B_0}}{2}. \]

The maximum likelihood estimates of the parameters for the alternative hypothesis are similarly defined:

\[ \hat{\gamma}_i = (\hat{\sigma}_u^2 \hat{\gamma}_1 \hat{\gamma}_1 + \Sigma_{\delta_x \delta_x}^{-1})^{-1}(\hat{\sigma}_u^2 Y_i \hat{\gamma}_1 + \Sigma_{\delta_x \delta_x}^{-1} Z_i), \]

\[ \hat{\gamma}_1 = (M_{Z'Z} - \Sigma_{\delta_x \delta_x})^{-1}M_{Z'Y}, \]
\( \hat{\sigma}_u^2 \) is the smallest positive root of \( \sigma_u^4 + \sigma_u^2 (2B_1 - A_1) + B_1^2 = 0 \), where

\[
A_1 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - Z_i \hat{\gamma}_1)^2,
\]

\[
B_1 = \hat{\gamma}_1 \Sigma_{\hat{\delta}_Z \hat{\delta}_Z} \hat{\gamma}_1.
\]

Again, using the quadratic formula, we can show that \( \hat{\sigma}_u^2 \) is the smallest positive quantity given by:

\[
\hat{\sigma}_u^2 = \frac{(A_1 - 2B_1) + \sqrt{A_1^2 - 4A_1B_1}}{2}.
\]

For notation, uppercase letter \( M \), appropriately subscripted, is used for the sample covariance matrix:

\[
M_{Z'Z} = n^{-1} \sum_{i=1}^{n} Z_i Z_i'.
\]

3.1.2 Derivation of Cox's Test

To derive Cox's test statistic, we again follow along the lines of Pesaran (1974). Let the log-likelihood functions for \( H_0 \) and \( H_1 \) be denoted by \( L_0(\theta_0) \) and \( L_1(\theta_1) \), where \( \theta_0 = (\sigma_u^2, \text{vec}(x)', \beta_1')' \) and \( \theta_1 = (\sigma_u^2, \text{vec}(z)', \gamma_1')' \). Also, let \( \hat{L}_{10} \) denote the maximum likelihood ratio of \( L_{10} \).
The log-likelihood functions of the independent normal random vectors \((Y_i, X_i)\) and \((Y_i, Z_i)\) under \(H_0\) and \(H_1\) for \(i=1,\ldots, n\) can be written as follows:

\[
L_0(\theta_0) = -\frac{n}{2} \log |2\pi \Sigma_{\delta_x,\delta_x'}| - \frac{n(p+1)}{2} \log \sigma_\delta^2
- \frac{1}{2\sigma_\delta^2} \sum_{i=1}^{n} (Y_i - x_i \beta_1)(Y_i - x_i \beta_1)'
- \frac{1}{2} \sum_{i=1}^{n} (X_i - x_i) \Sigma_{\delta_x,\delta_x'}^{-1} (X_i - x_i)'
\]

and

\[
L_1(\theta_1) = -\frac{n}{2} \log |2\pi \Sigma_{\delta_z,\delta_z'}| - \frac{n(q+1)}{2} \log \sigma_\delta^2
- \frac{1}{2\sigma_\delta^2} \sum_{i=1}^{n} (Y_i - z_i \gamma_1)(Y_i - z_i \gamma_1)'
- \frac{1}{2} \sum_{i=1}^{n} (Z_i - z_i) \Sigma_{\delta_z,\delta_z'}^{-1} (Z_i - z_i)'
\]

The log-likelihood functions of \((Y_i, X_i)\) and \((Y_i, Z_i)\) for \(i=1,\ldots, n\) evaluated at \(\hat{\theta}_0\) and \(\hat{\theta}_1\) are given by

\[
L_0(\hat{\theta}_0) = -\frac{n}{2} \log |2\pi \Sigma_{\delta_x,\delta_x'}| - \frac{np+1}{2} \log \sigma_\delta^2
\]

and

\[
L_1(\hat{\theta}_1) = -\frac{n}{2} \log |2\pi \Sigma_{\delta_z,\delta_z'}| - \frac{nq+1}{2} \log \sigma_\delta^2
\]
We can then derive \( \hat{L}_{10} = L_0(\hat{\theta}_0) - L_1(\hat{\theta}_1) \):

\[
\hat{L}_{10} = \frac{\text{tr}(\Sigma_{Z'Z})}{2} + \frac{n(q-p)}{2} \log \det \Sigma_{Z'Z} + \frac{n}{2} \log \frac{\hat{\sigma}_u^2}{\sigma^2(\mu + 1)}.
\]

Now, under \( H_0 \) it can be shown that \( \hat{\sigma}_u^2 \) converges in probability to \( \sigma^2 u_0 \), where

\[
\sigma^2 u_0(\theta_0) = \frac{(A_{10} - 2B_{10})}{2} \pm \sqrt{\frac{A_{10}^2 - 4A_{10}B_{10}}{2}},
\]

which is a function of the unknown parameter vector, \( \theta_0 \), where under \( H_0 \) we have

\[
A_{10} = \lim_{n \to \infty} A_1 = \sigma_2^2 + \beta'_1 (\Sigma_{X'X} - \Sigma_{\delta'\delta}) \beta_1 - 2 \beta'_1 \Sigma_{X'Z} (\Sigma_{Z'Z} - \Sigma_{\delta'\delta})^{-1} \Sigma_{Z'X} \beta_1
\]

\[
+ \beta'_1 \Sigma_{X'Z} (\Sigma_{Z'Z} - \Sigma_{\delta'\delta})^{-1} \Sigma_{Z'Z} (\Sigma_{Z'Z} - \Sigma_{\delta'\delta})^{-1} \Sigma_{Z'X} \beta_1
\]

and

\[
B_{10} = \lim_{n \to \infty} B_1 = \beta'_1 \Sigma_{X'Z} (\Sigma_{Z'Z} - \Sigma_{\delta'\delta})^{-1} \Sigma_{\delta'\delta} (\Sigma_{Z'Z} - \Sigma_{\delta'\delta})^{-1} \Sigma_{Z'X} \beta_1
\]

where the following regularity assumptions have been made:

\[
\lim_{n \to \infty} M_{X'X} = \Sigma_{X'X} \quad \text{(nonsingular)},
\]

\[
\lim_{n \to \infty} M_{Z'Z} = \Sigma_{Z'Z} \quad \text{(nonsingular)},
\]

\[
\lim_{n \to \infty} M_{Z'X} = \Sigma_{Z'X} \quad (\neq 0).
\]

The asymptotic expectation of \( \hat{\gamma}_1 \) under \( H_0 \) can be shown to be given by

\[
\hat{\gamma}_{10} = \Sigma_{Z'Z}^{-1} \Sigma_{Z'X} \beta_1.
\]
where $\Sigma_{x,z} = \Sigma_{z,z} - \Sigma_{x,z}^{\delta_z \delta_z}$ and $\Sigma_{x,z} = \Sigma_{z,z} - \Sigma_{x,z}^{\delta_z \delta_x}$. As in Chapter 2, we note that $\gamma_{10}$ is the "specification bias" discussed by Theil (1957) for the case of fixed explanatory variables measured with error.

Now observe that

$$\text{P} \lim_{n \to \infty} (L_{10}/n) = \frac{(q-p)}{2} + \frac{1}{2} \log \frac{|\Sigma_{x,z}^{\delta_z \delta_z}|}{|\Sigma_{x,z}^{\delta_z \delta_x}|} + \frac{1}{2} \log \frac{\sigma^2_{u0}}{\sigma^2_{\varepsilon}} + \frac{2(q+1)}{2(p+1)}.$$ 

where it is assumed that the estimates of $\sigma^2_\varepsilon$ and $\beta_1$ are consistent under $H_0$.

Cox's test statistic can then be written as

$$T_0 = T_0(\theta_0) = L_{10} - n \text{P} \lim_{n \to \infty} (L_{10}/n) \theta_0 = \frac{n(q+1)}{2} \log \frac{\hat{\sigma}^2_{u0}}{\sigma^2_{u0}}.$$ 

To derive $V_0(T_0)$ (in actuality we will compute an approximation to $V_0(T_0)$), we first consider $L_{10} = L_0(\theta_0) - L_1(\theta_{10})$:

$$L_{10} = \frac{n(q-p)}{2} + \frac{1}{2} \log \frac{|\Sigma_{x,z}^{\delta_z \delta_z}|}{|\Sigma_{x,z}^{\delta_z \delta_x}|} + \frac{1}{2} \log \frac{\sigma^2_{u0}}{\sigma^2_{\varepsilon}} - \frac{1}{2} \hat{\sigma}^2_{\varepsilon} \sum_{i=1}^{n} \epsilon^2$$

$$+ \frac{1}{2} \sum_{i=1}^{n} (Z_i - z_i) \Sigma_{x,z}^{\delta_z \delta_z} (Z_i - z_i) + \frac{1}{2} \sum_{i=1}^{n} (X_i - x_i) \Sigma_{x,z}^{\delta_z \delta_x} (X_i - x_i)$$

$$+ \frac{1}{2\sigma^2_{u0}} \sum_{i=1}^{n} (x_i \beta_1 - x_i \gamma_{10} + \epsilon_i)(x_i \beta_1 - x_i \gamma_{10} + \epsilon_i)^{'}$$

$$= \frac{n(q-p)}{2} + \frac{1}{2} \log \frac{|\Sigma_{x,z}^{\delta_z \delta_z}|}{|\Sigma_{x,z}^{\delta_z \delta_x}|} + \frac{1}{2} \log \frac{\sigma^2_{u0}}{\sigma^2_{\varepsilon}} + \frac{1}{2} \left( \frac{1}{\sigma^2_{u0}} - \frac{1}{\sigma^2_{\varepsilon}} \right) \sum_{i=1}^{n} \epsilon^2$$

$$+ \frac{1}{2} \sum_{i=1}^{n} (Z_i - z_i) \Sigma_{x,z}^{\delta_z \delta_z} (Z_i - z_i) + \frac{1}{2} \sum_{i=1}^{n} (X_i - x_i) \Sigma_{x,z}^{\delta_z \delta_x} (X_i - x_i)$$

$$+ \frac{1}{2\sigma^2_{u0}} \sum_{i=1}^{n} (x_i \beta_1 - x_i \gamma_{10})(x_i \beta_1 - x_i \gamma_{10})^{'} + \frac{1}{\sigma^2_{u0}} \sum_{i=1}^{n} (x_i \beta_1 - x_i \gamma_{10})\epsilon_i.$$
Thus, we have

\[ V_0(\frac{p}{10}) = V_0\left[ \frac{1}{2}\left(\frac{1}{\sigma^2_{u_0}} - \frac{1}{\sigma^2_{v}}\right)\sum_{i=1}^{n} \varepsilon^2 + \frac{1}{\sigma^2_{u_0}}\sum_{i=1}^{n} (x_i \beta_1 - z_i \gamma_{10}) \varepsilon_i \right. \]

\[ + \left. \frac{1}{2}\left\{ \sum_{i=1}^{n} (Z_i - z_i) \Sigma^{-1}_{\delta^2_z} (Z_i - z_i)' - \sum_{i=1}^{n} (X_i - x_i) \Sigma^{-1}_{\delta^2_x} (X_i - x_i)' \right\} \right] \]

We can then derive the following expression:

\[ V_0(\frac{p}{10}) = \frac{\left(p+q\right)}{2} + \frac{\left(1 - \frac{1}{\sigma^2_{u_0}}\right)}{2} \sigma^2_{\varepsilon} + \frac{\sigma^2_{z}}{\sigma^2_{u_0}}\sum_{i=1}^{n} (x_i \beta_1 - z_i \gamma_{10}) (x_i \beta_1 - z_i \gamma_{10}) \]

\[ + \frac{\sigma^2_{z}}{2} \text{tr}(\Sigma^{-1}_{\delta^2_z} \Sigma_{\delta^2_x} \Sigma_{\delta^2_x} \Sigma_{\delta^2_z} \Sigma_{\delta^2_z}). \]

Unfortunately, we cannot evaluate the above quantity since \( x_i \) and \( z_i \) are unobservable.

However, observe the following

\[ \sum_{i=1}^{n} (x_i \beta_1 - z_i \gamma_{10}) (x_i \beta_1 - z_i \gamma_{10}) = \sum_{i=1}^{n} (a^i x_i \beta_1 - b^i z_i \gamma_{10} + c^i z_i \gamma_{10}). \]

We propose using the following asymptotic expression to approximate the above expression:

\[ \frac{n}{2} (\beta^T \Sigma_{x^2} \beta) - 2 \beta^T \Sigma_{x^2} \gamma_{10} + \gamma_{10} \Sigma_{x^2} \gamma_{10} \]

Substituting the latter expression into \( V_0(\frac{p}{10}) \) we get

\[ V_0(\frac{p}{10}) \approx \frac{\left(p+q\right)}{2} + \frac{\left(1 - \frac{1}{\sigma^2_{u_0}}\right)}{2} \sigma^2_{\varepsilon} + \frac{\sigma^2_{z}}{\sigma^2_{u_0}} (\beta^T \Sigma_{x^2} \beta) - 2 \beta^T \Sigma_{x^2} \gamma_{10} + \gamma_{10} \Sigma_{x^2} \gamma_{10}. \]
Now we must derive $\eta$ and $I_0$ (information matrix) as defined in Chapter 2 (section 2.1.2). However, since the unobservable design matrix, $x$, represents an $n \times p$ matrix of nuisance parameters, then as $n$ goes to infinity the number of parameters in $\eta$ and $I_0$ goes to infinity also.

We propose to remedy this problem by using a reduced likelihood function. The likelihood function is reduced by approximating the original design matrix with a design matrix based on the quantiles of the original explanatory variables. So instead of a matrix of nuisance parameters which increases with $n$, we will have a matrix with a fixed number of nuisance parameters. Thus, we compute $\eta^*$ and $I_0^*$ in place of $\eta$ and $I_0$.

Now, $\eta^*$ and $I_0^*$ will contain the unknown parameters $\beta_1^*$ and $\sigma_\epsilon^2$ (and the assumed known error covariance matrix, $\Sigma_{\delta x \delta x}^*$) which are a result of using the quantile design matrix. We propose replacing $\beta_1^*$, $\sigma_\epsilon^2$, and $\Sigma_{\delta x \delta x}^*$ with $\beta_1$, $\sigma_\epsilon^2$, and $\Sigma_{\delta x \delta x}$ (the unknown and known parameters based on the full likelihood function) since in any applications, estimates of $\beta_1^*$ and $\sigma_\epsilon^2$ will be used and it seems reasonable to make use of the knowledge of the obtainable estimates of $\beta_1$ and $\sigma_\epsilon^2$ based on the full likelihood function.

Formally, we replace the original design matrix, $x$, by computing a predetermined set of quantiles (for example minimum, 25th, median, 75th, maximum) for each explanatory variable (column of $x$). Each value of $x$ is then replaced by an appropriate quantile by categorizing the particular value. For example, suppose we use the previously mentioned quantiles for a particular explanatory variable (column of $x$) and compute half the distance between each quantile. Let these halfway points be denoted by $h_1$, $h_2$, $h_3$, and $h_4$. We can then replace each value of that variable which lies in the "right open, left closed" interval $[\text{min}, h_1)$ with the minimum, each value which lies in $[h_1, h_2)$ with the 25th, each value which lies in $[h_2, h_3)$ with the median, and so forth. Let the resulting quantile design matrix be denoted by $x^*$ and hence the observed design matrix by $X^*$, with the $i^{th}$ row of $x^*$ and $X^*$ denoted by $x_i^*$ and $X_i^*$. 
The reduced likelihood function under the null hypothesis, $H_0$, can then be written as

\[ L_0(\theta^*_0) = -\frac{n}{2} \log|2\pi \Sigma^*_x\delta_x| - \frac{n(p+1)}{2} \log \sigma^2 \]

\[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - x^*_i \beta^*_1)(Y_i - x^*_i \beta^*_1)' \]

\[-\frac{1}{2} \sum_{i=1}^{n} (X_i^* - x^*_i) \Sigma^{* -1}_{x} (X_i^* - x^*_i)' \]

and similarly for the likelihood function under $H_1$.

Under $H_0$, there are $p$ explanatory variables. Let $\omega_{0i}, i=1,...,p$, be mx1 column vectors where $m$ is the chosen number of quantiles for each explanatory variable and $\omega_{0i}$ contains the $m$ chosen quantiles for the $i$th explanatory variable in ascending order. Thus, $\theta^*_0= (\sigma^2, \omega, \beta^*_1)'$, where $\omega = (\omega_{01},...,\omega_{0p})'$. Observe that we have reduced the number of nuisance parameters from a sample size dependent $np$ to a fixed $mp$.

Now we are ready to compute $\eta^*$. First of all we notice that $\text{Plim}_{n \to \infty} (\hat{L}^*_0/n)$ is completely analogous to $\text{Plim}_{n \to \infty} (\hat{L}^*_{10}/n)$. Secondly,

\[ \frac{\partial \text{Plim}_{n \to \infty} (\hat{L}^*_{10}/n)}{\partial \omega} = 0. \]

This latter expression indicates that $\omega$ contributes nothing to $\eta^*$. Thus, it is appropriate to compute the other components of $\eta^*$ using the full likelihood functions (see discussion on previous page):

\[ \frac{\partial \text{Plim}_{n \to \infty} (\hat{L}^*_{10}/n)}{\partial \sigma^2_{\epsilon}} = \frac{(q+1)}{2\sigma_{u0}^2} \frac{\partial \sigma^2_{u0}}{\partial \sigma^2_{\epsilon}} - \frac{(p+1)}{2\sigma_{\epsilon}^2} \]

\[ = \frac{(q+1)}{2\sigma_{u0}^2} \left[ \frac{1}{2} (A_{10} - 2B_{10}) (A_{10}^2 - 4A_{10}B_{10})^{-1/2} \right] - \frac{(p+1)}{2\sigma_{\epsilon}^2}, \]
\[
\frac{\partial \lim_{n \to \infty} \left( \hat{L}_{10}/n \right)}{\partial \beta_1} = \left( \frac{\partial A_{10}}{\partial \beta_1} - 2 \frac{\partial B_{10}}{\partial \beta_1} \right) \frac{(q+1)}{2\sigma^2_{u0}} \times \pm \left( A_{10}^2 - 4A_{10}B_{10} \right)^{-1/2} \left[ (A_{10} - 2B_{10}) \frac{\partial A_{10}}{\partial \beta_1} - 4A_{10} \frac{\partial B_{10}}{\partial \beta_1} \right],
\]

where

\[
\frac{\partial A_{10}}{\partial \beta_1} = 2 \left( (\Sigma_{x'x} - \Sigma_{\delta_z\delta_z}) - 2\Sigma_{x'z}(\Sigma_{z'z} - \Sigma_{\delta_z\delta_z})^{-1}\Sigma_{z'x} \right) + \Sigma_{x'z}(\Sigma_{z'z} - \Sigma_{\delta_z\delta_z})^{-1}\Sigma_{z'z}(\Sigma_{z'z} - \Sigma_{\delta_z\delta_z})^{-1}\Sigma_{z'x} \beta_1
\]

and

\[
\frac{\partial B_{10}}{\partial \beta_1} = 2\Sigma_{x'z}(\Sigma_{z'z} - \Sigma_{\delta_z\delta_z})^{-1}\Sigma_{\delta_z\delta_z}(\Sigma_{z'z} - \Sigma_{\delta_z\delta_z})^{-1}\Sigma_{z'x} \beta_1.
\]

The above expressions provide the components of

\[
\eta^* = \eta^*(\theta_0^*) = n \frac{\partial \left[ \lim_{n \to \infty} \left( \hat{L}_{10}^*/n \right) \right]}{\partial \theta_0^*},
\]

where \(\eta^*(\theta_0^*)\) indicates \(\eta^*\) is a function of the unknown parameter vector \(\theta_0^*\).

To express the information matrix, \(I_0^*\), it is necessary to utilize additional notation involving the derivative of the likelihood function under \(H_0^*\) with respect to \(\omega\). Consider taking the derivative of \(x_i^* \beta_1^*\) with respect to \(\omega\), i.e. \(\frac{\partial x_i^* \beta_1^*}{\partial \omega}\). Due to the construction of \(x^*\) a possible example of \(\frac{\partial x_i^* \beta_1^*}{\partial \omega}\) assuming the chosen quantiles are the minimum, 25th, median, 75th, and maximum is given by

\[
\frac{\partial x_i^* \beta_1^*}{\partial \omega} = [\beta_{11}^*, 0, 0, 0, 0, 0, \beta_{12}^*, 0, 0, ..., 0, 0, 0, 0, \beta_1^*].
\]

The interpretation of the above is that for the \(i^{th}\) row of the original design matrix \(x\),
the first explanatory variable's value has been categorized as the minimum, the second
explanatory variable's value has been categorized as the median, etc., and the last
explanatory variable's value has been categorized as the maximum.

Denote \( \frac{\partial x^*_i}{\partial \omega} \beta_1^* \) by \( Q_{0i}(\beta_1^*) \), which is in general an mp×1 vector for \( i=1,...,p \). For
the above example, \( Q_{0i}(\beta_1^*) \) is a 5p×1 vector for \( i=1,...,p \). Similarly, denote the mp×p
matrices of 0's and 1's given by \( \frac{\partial x^*}{\partial \omega} \) by \( R_{0i} \) for \( i=1,...,p \).

Finally, we can write an expression for \( I_0^*(\theta_0, x^*) \), which is a function of the
parameters from the full model and the reduced model (see the discussion on page 53):

\[
I_0^* = I_0^*(\theta_0, x^*) = \begin{bmatrix}
\frac{n(p+1)}{2\sigma^4_{\xi}} & 0 & 0 \\
0 & \Sigma_{22} & \Sigma_{23} \\
0 & \Sigma_{32} & \frac{n}{\sigma^2_{\xi}} \Sigma x'x
\end{bmatrix},
\]

where

\[
\Sigma_{22} = \sum_{i=1}^{n} \left( \frac{1}{\sigma^2_{\xi}} Q_{0i}(\beta_1) Q_{0i}(\beta_1)' + R_{0i} \Sigma_{\delta_x \delta_x} R_{0i}' \right),
\]

\( Q_{0i}(\beta_1) \) represents \( Q_{0i}(\beta_1^*) \) evaluated at \( \beta_1^* = \beta_1 \)

and

\[
\Sigma_{23} = \frac{1}{\sigma^2_{\xi}} \sum_{i=1}^{n} Q_{0i}(\beta_1) x^*_i \text{ with } \Sigma_{32} = \Sigma_{23}.
\]
We now have all the necessary components of $V_0^*(T_0)$, the variance of $T_0$ under $H_0$ using the substituted components $\eta^*$ and $I_0^*$. Since we have chosen to use a reduced likelihood function for the computation of $\eta$ and $I_0$ and also substituted appropriate approximations where necessary, it is expected that $V_0^*(T_0)$ will be larger than the true variance of $T_0$ under $H_0$, $V_0(T_0)$.

To obtain $\hat{V}_0^*(T_0)$, we replace all the parameters involved with their maximum likelihood estimates, where

$$\hat{s}_1^* = \left(\hat{\sigma}_i^2 \hat{\beta}_1 + \Sigma_{x}\hat{\delta}_x \right)^{-1}(\hat{\sigma}_i^2 \hat{Y}_i \hat{\beta}_1 + \Sigma_{x} \hat{X}_i) .$$

### 3.1.3 A Note on the Selection of Quantile Design Matrix

It can be shown that as the number of quantiles approaches the sample size, $n$, then the quantile design matrix will approach the original design matrix (see Serfling 1980).

An 'optimal' selection of quantiles will depend upon a number of factors. 'Optimal' in this procedure will loosely mean that the information contained in the quantile design matrix will be nearly as good as that in the original design matrix under such constraints as the sample size being much larger than the number of parameters and maybe a different number of quantiles chosen for each explanatory variable. The selection of quantiles will most certainly depend upon the distribution of the explanatory variables.

A thorough numerical investigation is required to develop a procedure which will select an optimal quantile design matrix. It is expected that this task will be undertaken not in this dissertation but in some future research.

Presently, we use an ad hoc method of choosing the quantile design matrix.
3.1.4 Discussion

The fact that maximum likelihood estimation does not yield consistent estimates of all parameters in the model due to the increasing number of parameters associated with the design matrix \( x \) required us to use a reduction method to achieve our desired results. Clearly, \( \hat{x} \) is only as good as the fitted model, i.e., if the model does not fit well, then we should not expect a good estimate of \( x \).

Several authors are beginning to take a different approach to looking at estimating \( x \) and the associated regression parameters. Whittemore (1989) considers replacing the true unobserved explanatory variables by their James-Stein estimates, based on the observed values. The remaining parameters are then estimated by the more common techniques. Whittemore and Keller (1988) proposes using modified estimates of the regression parameters computed from the naive estimates. The modified estimates involve only the mean and variance of the error distribution conditional on the observed explanatory variables.

Pierce, et al (1989a, 1989b) in two unpublished manuscripts proposes looking at the conditional expectation of the unobservable explanatory variables conditioned on the observed values. This particular procedure has a definitional drawback for the functional model but a way of circumventing it is proposed by the authors. Schafer (1989) in another unpublished manuscript also proposes investigating the idea of the conditional expectation of the unobservable explanatory variables conditioned on the observed values using iteratively weighted least squares.

Once the unobserved explanatory variables are replaced using any of the methods mentioned previously, then the problem of testing separate hypotheses can be investigated as discussed by Pesaran (1974).

3.2 Interchanging Hypotheses

As stated at the end of Chapter 2, since the hypotheses are considered
asymmetrically, $H_0$ and $H_1$ should be interchanged with $H_1$ as the null hypothesis and $H_0$ as the alternative. Cox’s test of separate hypotheses then has test statistic $T_1$ and variance $V_1(T_1)$. 
CHAPTER 4
Applications

4.1 Introduction

This chapter presents our applications of Cox's test statistic for testing two normal structural models and for testing two normal functional models. The theoretical aspects of our derivation of Cox's test statistic for the structural and functional models are given in Chapters 2 and 3, respectively.

To date, there is a large and ever growing body of literature on measurement error models and similarly, there is a large and ever growing body of literature on testing separate families of hypotheses. To our knowledge no one has attempted to integrate these two existing theories. We were successful in integrating the theory of measurement error models and Cox's test of separate (nonnested) linear models by making various simplifying assumptions and by using large sample approximations. We apply the results of our derivations to real data in this chapter.

The data is from an example in Neter, Wasserman, and Kutner (NWK 1985, pp 419-421). From NWK is the following description of the data: A hospital surgical unit was interested in predicting survival in patients undergoing a particular type of liver operation. A random selection of 54 patients was available for the analysis. From each patient record, the following information was extracted from the preoperational evaluation (units of measurement were not given):

$X_1$: blood clotting score

$X_2$: prognostic index, which includes the age of the patient

$X_3$: enzyme function test score
$X_4$: liver function blood clotting score

The dependent variable was survival time (assumed days since unit of measurement was not given), but due to the skewness of the survival time distribution, the logarithm (base 10) of the survival time was used as the dependent variable. The variables $X_1$, $X_2$, $X_3$, and $X_4$ were used as potential explanatory variables. Table 4.1 provides descriptive statistics in the form of sample means and variances for the explanatory variables $X_1$, $X_2$, $X_3$, $X_4$, and the dependent variable $Y$ (logarithm base 10 of survival time).

Observe that the data from NWK has two blood clotting scores. In illustrating the application of our results, we use various different combinations of the explanatory variables. Since two blood clotting scores are used, models of interest include one or the other blood clotting score in each hypothesis. Other combinations of the explanatory variables are also explored.
4.1.1 Assuming Normal Structural Models

In this section, we apply the results of Chapter 2 to the data described in section 4.1, where the underlying true explanatory variables are assumed to be stochastic (random). We first assume the null and alternative hypotheses are given by

\[ H_0: Y_i = \beta_0 + x_i \beta_1 + \epsilon_i \]
\[ X_i = x_i + \delta x_i \]

\[ H_1: Y_i = \gamma_0 + z_i \gamma_1 + u_i, \]
\[ Z_i = z_i + \delta z_i \]

for \( i = 1, \ldots, n (=54) \). For the normal structural models, we assume the following for \( i = 1, \ldots, n \):

\[
\begin{bmatrix}
    x_i' \\
    \epsilon_i \\
    \delta x_i
\end{bmatrix} \sim \text{NI} \begin{bmatrix}
    \mu_x' \\
    0 \\
    0
\end{bmatrix}, \begin{bmatrix}
    \Sigma_x x & 0 & 0 \\
    0 & \sigma_x^2 & 0 \\
    0 & 0 & \Sigma_{\delta x \delta x}
\end{bmatrix},
\]

\[
\begin{bmatrix}
    z_i' \\
    u_i \\
    \delta z_i
\end{bmatrix} \sim \text{NI} \begin{bmatrix}
    \mu_z \\
    0 \\
    0
\end{bmatrix}, \begin{bmatrix}
    \Sigma_z z & 0 & 0 \\
    0 & \sigma_u^2 & 0 \\
    0 & 0 & \Sigma_{\delta z \delta z}
\end{bmatrix},
\]

where \( X_i = (X_{1i}, X_{2i}) \) and \( Z_i = (X_{3i}, X_{4i}) \). As stated in the introduction, we make a series of comparisons in addition to the one given above. Even though the explanatory variables are correlated, with the two blood clotting variables, \( X_1 \) and \( X_4 \), showing the
highest correlation at 0.5 (see Table 4.2), for simplicity we assume that $\Sigma_{\delta_x\delta_x}$ and $\Sigma_{\delta_x'\delta_x'}$ are known and both are diagonal, i.e., we assume there is no correlation between the measurement errors in the explanatory variables.

A review of Table 4.2, which gives the correlation matrix of the variables used in the surgical unit example shows all the explanatory variables are positively correlated with the dependent variable. $X_4$ has the largest correlation (0.726) followed in descending order by $X_3$ (0.665), $X_2$ (0.593), and $X_1$ (0.346). The two blood clotting scores have the highest correlation amongst the explanatory variables (0.502) as mentioned previously. It can be seen that $X_4$ is also positively correlated with $X_2$ (0.369), and $X_3$ (0.416). The other correlations are rather small in magnitude and thus are not of particular note.

In order to evaluate the role of measurement error, we apply Cox's test under three different assumptions concerning the known measurement error covariance matrices. First we assume that the explanatory variables are measured with no error, i.e., $\Sigma_{\delta_x\delta_x}$ and $\Sigma_{\delta_x'\delta_x'}$ are both zero matrices. Secondly, we assume that the measurement error variances are 10\% of the observed sample variances. Thirdly, we assume the measurement error variances are 20\% of the observed sample variances.

As shown in Table 4.3, for testing $H_0$ versus $H_1$ with no error in the explanatory variables, Cox's test statistic, $T_0$, was evaluated as $-25.82$. The variance of $T_0$, $V_0(T_0)$, was determined to be $319693$. The normalized 2-sided test yielded a p-value of 0.9636. When reversing the hypotheses: $T_1 = -10.38$, $V_1(T_1) = 297060$, p-value = 0.9848.

For testing $H_0$ versus $H_1$ with error variances equal to 10\% of the observed sample variances in the explanatory variables, Cox's test statistic, $T_0$, was evaluated as $-25.28$. The variance of $T_0$, $V_0(T_0)$, was determined to be $314533$. The normalized 2-sided test yielded a p-value of 0.9641. When reversing the hypotheses: $T_1 = -9.9174$, $V_1(T_1) = 288115$, p-value = 0.9853.
For testing $H_0$ versus $H_1$ with error variances equal to 20% of the observed sample variances in the explanatory variables, Cox's test statistic, $T_0$, was evaluated as $-24.59$. The variance of $T_0$, $V_0(T_0)$, was determined to be 314487. The normalized 2-sided test yielded a p-value of 0.9650. When reversing the hypotheses: $T_1 = -9.3311$, $V_1(T_1) = 280395$, p-value = 0.9859.

For testing $H_0$ versus $H_1$, the result of Cox's test in each of the hypothesis testing scenarios given in Table 4.3 suggests there is not enough evidence to choose between the models. So whether there is no error in the explanatory variables or an assumed 10% or 20% of the sample variance is due to error, Cox's test indicates that there is not enough evidence in the data to choose between a model using blood clotting and prognostic index as explanatory variables versus a model using enzyme function test score and liver function blood clotting score as explanatory variables in relation to logarithmically (base 10) transformed survival time.

In addition, for testing $H_0$ versus $H_1$ in Table 4.3, it can be observed that the p-values consistently increased as the error variance assumption ranged from 0 to 20%. However, the increases in the p-values are not very large. Cox's test statistics consistently decreased in absolute value and the variance of the test statistics also decreased as the error variance assumption ranged from 0 to 20%.

Tables 4.4-4.6 present the results for testing $H_0$ versus $H_1$ with 0, 10%, and 20% (of sample variances) error variances for various combinations of the explanatory variables. The assumption of errors in the explanatory variables affected Cox's test statistic and its variance in various ways. A brief discussion regarding these effects is provided in the next section.
4.1.2 Results from Testing Normal Structural Models

The results shown in Tables 4.3-4.6 vary in the degree in which the assumption of errors in the explanatory variables affect Cox's test of separate (nonnested) linear models.

The following observations are derived from the results in Tables 4.3-4.6:

i) The p-values for testing the nonnested normal structural models increased with increasing error variances,

ii) Cox's test indicated that in each case there was insufficient evidence in the proposed models to choose between any tested pair of models,

iii) The assumption of errors in the explanatory variables affected Cox's test statistic and its variance in varying ways. For example, in Tables 4.3 and 4.4 the Cox test statistic consistently decreased in absolute value as the error variances increased whereas the variances of the test statistics usually decreased as the error variances increased. In Table 4.5 Cox's test statistic \( T_{0} \) first increased in absolute value and then decreased and the variances did likewise when testing \( H_{0} \) versus \( H_{1} \). When testing \( H_{1} \) versus \( H_{0} \) in Table 4.5, Cox's test statistic \( T_{1} \) consistently decreased in absolute value as the error variances increased while the variances of the test statistic consistently increased.
Table 4.1
Surgical Unit Example
Descriptive Statistics

Variables

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.78</td>
<td>63.19</td>
<td>77.11</td>
<td>2.74</td>
<td>2.21</td>
</tr>
<tr>
<td>Var</td>
<td>2.57</td>
<td>290.08</td>
<td>451.72</td>
<td>1.15</td>
<td>0.08</td>
</tr>
</tbody>
</table>

X1: blood clotting score
X2: prognostic index, which includes the age of the patient
X3: enzyme function test score
X4: liver function blood clotting score
Y : $\log_{10}(\text{survival time})$
Table 4.2
Surgical Unit Example

Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>X₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>1.000</td>
<td>0.346</td>
<td>0.593</td>
<td>0.665</td>
<td>0.726</td>
</tr>
<tr>
<td>X₁</td>
<td>1.000</td>
<td>0.090</td>
<td>-0.150</td>
<td>0.502</td>
<td></td>
</tr>
<tr>
<td>X₂</td>
<td>1.000</td>
<td>1.000</td>
<td>-0.024</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td>X₃</td>
<td>1.000</td>
<td>1.000</td>
<td>0.416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₄</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Y : $\log_{10}$(survival time)

X₁: blood clotting score

X₂: prognostic index, which includes the age of the patient

X₃: enzyme function test score

X₄: liver function blood clotting score
Table 4.3

Surgical Unit Example

$H_0$: variables $X_1$ and $X_2$ vs $H_1$: variables $X_3$ and $X_4$

Normal Structural Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -25.82$</td>
<td>319693</td>
<td>0.9636</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -10.38$</td>
<td>297060</td>
<td>0.9848</td>
</tr>
<tr>
<td>10%</td>
<td>$T_0 = -25.28$</td>
<td>314533</td>
<td>0.9641</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -9.92$</td>
<td>288115</td>
<td>0.9853</td>
</tr>
<tr>
<td>20%</td>
<td>$T_0 = -24.59$</td>
<td>314487</td>
<td>0.9650</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -9.33$</td>
<td>280395</td>
<td>0.9859</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.4
Surgical Unit Example

$H_0$: variables $X_2$ and $X_4$ vs $H_1$: variables $X_1$ and $X_3$

Normal Structural Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -17.16$</td>
<td>308712</td>
<td>0.9754</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -17.47$</td>
<td>309169</td>
<td>0.9747</td>
</tr>
<tr>
<td>10%</td>
<td>$T_0 = -15.98$</td>
<td>298256</td>
<td>0.9767</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -15.74$</td>
<td>298029</td>
<td>0.9770</td>
</tr>
<tr>
<td>20%</td>
<td>$T_0 = -14.43$</td>
<td>291208</td>
<td>0.9787</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -13.29$</td>
<td>300806</td>
<td>0.9807</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.5

Surgical Unit Example

$H_0$: variables $X_1$ and $X_4$ vs $H_1$: variables $X_2$ and $X_3$

Normal Structural Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -31.75$</td>
<td>322992</td>
<td>0.9554</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -7.43$</td>
<td>288572</td>
<td>0.9890</td>
</tr>
<tr>
<td>10%</td>
<td>$T_0 = -45.00$</td>
<td>476606</td>
<td>0.9480</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -6.08$</td>
<td>297143</td>
<td>0.9911</td>
</tr>
<tr>
<td>20%</td>
<td>$T_0 = -44.90$</td>
<td>466828</td>
<td>0.9476</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -3.29$</td>
<td>301763</td>
<td>0.9942</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.6
Surgical Unit Example

\(H_0: \text{variable } X_4 \text{ vs } H_1: \text{variables } X_1, X_2 \text{ and } X_3\)

Normal Structural Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(T_0 = -75.06)</td>
<td>349156</td>
<td>0.8989</td>
</tr>
<tr>
<td></td>
<td>(T_1 = -0.13)</td>
<td>278144</td>
<td>0.9998</td>
</tr>
<tr>
<td>10%</td>
<td>(T_0 = -71.75)</td>
<td>337388</td>
<td>0.9017</td>
</tr>
<tr>
<td></td>
<td>(T_1 = 3.68)</td>
<td>334648</td>
<td>0.9949</td>
</tr>
<tr>
<td>20%</td>
<td>(T_0 = -66.99)</td>
<td>334480</td>
<td>0.9078</td>
</tr>
<tr>
<td></td>
<td>(T_1 = 9.50)</td>
<td>633692</td>
<td>0.9905</td>
</tr>
</tbody>
</table>

* Percent of sample variance
4.2.1 Assuming Normal Functional Models

In this section, we apply the results of Chapter 3 to the data described in section 4.1, where the underlying true explanatory variables are assumed to be fixed. We first assume the null and alternative hypotheses are given by

\[ H_0: Y_i = x_i \beta_1 + \epsilon_i \]
\[ X_i = x_i + \delta x_i \]

\[ H_1: Y_i = z_i \gamma_1 + u_i , \]
\[ Z_i = z_i + \delta z_i \]

for \( i = 1, \ldots, n (=54) \). For the normal functional models, assume the following for \( i = 1, \ldots, n \):

\[
\begin{bmatrix}
\epsilon_i \\
\delta x_i
\end{bmatrix}
\sim \text{NI}
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\begin{bmatrix}
\sigma^2 & 0 \\
0 & \Sigma_{\delta x, \delta x}
\end{bmatrix},
\]

\[
\begin{bmatrix}
u_i \\
\delta z_i
\end{bmatrix}
\sim \text{NI}
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\begin{bmatrix}
\sigma^2 & 0 \\
0 & \Sigma_{\delta z, \delta z}
\end{bmatrix},
\]

where \( X_i = (X_{1i}, X_{2i}) \) and \( Z_i = (X_{3i}, X_{4i}) \) and it is assumed that \( \Sigma_{\delta x, \delta x} \) and \( \Sigma_{\delta z, \delta z} \) are known.

As we did in the previous section, we again apply Cox's test under as many as four different assumptions concerning the known measurement error covariance matrices. However, due to the fact that maximum likelihood does not yield consistent estimates for all parameters in the functional model, assuming too large an error
variance sometimes resulted in a complex-valued value of the estimated within subject error variance \( (\sigma_i^2 \text{ and/or } \sigma_d^2) \), i.e., we encountered in some cases the square root of a negative value in solving the quadratic equation for the estimated within subject error variance. Therefore, we often could only look at a limited range of error variances.

For the first example we assume that the explanatory variables are measured with no error, i.e., \( \Sigma_{\delta x} \Sigma_{\delta x} \) and \( \Sigma_{\delta x} \Sigma_{\delta x} \) are both zero matrices. Secondly, we assume that the measurement error variances are 5% of the observed sample variances. Thirdly, we assume the measurement error variances are 10% of the observed sample variances. Fourthly, we assume the measurement error variances are 15% of the observed sample variances. For other examples we were able to use error variances ranging from a maximum of as much as 7% to as little as 1%.

In addition, we use five quantiles (see Chapter 3 for details) for each explanatory variable in the reduced likelihood functions.

As shown in Table 4.7, for testing \( H_0 \) versus \( H_1 \) with no error in the explanatory variables, Cox's test statistic, \( T_0 \), was evaluated as \(-30.90\). The variance of \( T_0 \), \( V_0(T_0) \), was determined to be 8.998. The normalized 2-sided test yielded an approximate p-value of 0. When reversing the hypotheses: \( T_1 = -42.89 \), \( V_1(T_1) = 8.467 \), p-value \( \approx 0 \).

For testing \( H_0 \) versus \( H_1 \) with error variances equal to 5% of the observed sample variances in the explanatory variables, Cox's test statistic, \( T_0 \), was evaluated as 30.48. The variance of \( T_0 \), \( V_0(T_0) \), was determined to be 22968. The normalized 2-sided test yielded a p-value of 0.8406. When reversing the hypotheses: \( T_1 = 129.5 \), \( V_1(T_1) = 642748 \), p-value = 0.8717.

For testing \( H_0 \) versus \( H_1 \) with error variances equal to 10% of the observed sample variances in the explanatory variables, Cox's test statistic, \( T_0 \), was evaluated as 36.95. The variance of \( T_0 \), \( V_0(T_0) \), was determined to be 4814. The normalized 2-sided test yielded a p-value of 0.5944. When reversing the hypotheses: \( T_1 = 146.4 \), \( V_1(T_1) = \)
193381, p-value = 0.7392.

For testing $H_0$ versus $H_1$ with error variances equal to 15% of the observed sample variances in the explanatory variables, Cox's test statistic, $T_0$, was evaluated as 54.23. The variance of $T_0$, $V_0(T_0)$, was determined to be 471. The normalized 2-sided test yielded a p-value of 0.0125. When reversing the hypotheses: $T_1 = 196.8$, $V_1(T_1) = 124764$, p-value = 0.5774.

For testing $H_0$ versus $H_1$, the result of Cox's test indicates that both hypotheses should be rejected and thus a model previously unprescribed should be considered. This latter result contrasts with the result in Section 4.2.1 under the assumption of the normal structural model. Recall that under the assumption of no error in the normal structural model, there was not enough evidence to choose between the models.

When the explanatory variables were assumed measured with error and the error variances were assumed to be 5% and 10% of the sample variances, Cox's test suggests that there is not enough evidence in the data to choose between the models in each case. It is readily seen, however, that Cox's test statistics increase, their variances decrease, and the associated p-values decrease as the error variance ranges from 5% to 10%. Just the opposite occurred under the normal structural model presented in Section 4.2.1.

When the error variance was assumed to be 15% of the sample variance, Cox's test yielded an unexpected result. Cox's test indicates that we should reject the model using the explanatory variables $X_1$ (blood clotting score) and $X_2$ (prognostic index) in favor of the model using the explanatory variables $X_3$ (enzyme function test score) and $X_4$ (liver function blood clotting score) in relation to logarithmically (base 10) transformed survival time. Hence, in testing $H_0$ versus $H_1$, if the explanatory variables were assumed to be measured with no error, then based on Cox's test we would erroneously commence looking for a previously unformulated model instead of using a model with enzyme function test score and liver function blood clotting score as explanatory variables if in actuality the explanatory variables were measured with error.
where the error variances were 15% of the sample variance.

It should be noted that as the error variances ranged from 5% to 15% of the sample variance, Cox's test statistics consistently increased, their variances consistently decreased, and the associated p-values consistently decreased.

Tables 4.8-4.10 present the results for testing $H_0$ versus $H_1$ with varying error variances for the same combination of explanatory variables given in Tables 4.4-4.6. Again, the assumption of errors in the explanatory variables affected Cox's test statistics and its variances in various ways. A brief discussion regarding these effects is provided in the next section.
Table 4.7

Surgical Unit Example

\( H_0 \): variables \( X_1 \) and \( X_2 \) vs \( H_1 \): variables \( X_3 \) and \( X_4 \)

Normal Functional Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( T_0 = -30.90 )</td>
<td>8.998</td>
<td>( \approx 0 )</td>
</tr>
<tr>
<td></td>
<td>( T_1 = -42.89 )</td>
<td>8.467</td>
<td>( \approx 0 )</td>
</tr>
<tr>
<td>5%</td>
<td>( T_0 = 30.48 )</td>
<td>22968</td>
<td>0.8406</td>
</tr>
<tr>
<td></td>
<td>( T_1 = 129.50 )</td>
<td>642748</td>
<td>0.8717</td>
</tr>
<tr>
<td>10%</td>
<td>( T_0 = 36.95 )</td>
<td>4814</td>
<td>0.5944</td>
</tr>
<tr>
<td></td>
<td>( T_1 = 146.40 )</td>
<td>193381</td>
<td>0.7392</td>
</tr>
<tr>
<td>15%</td>
<td>( T_0 = 54.23 )</td>
<td>471</td>
<td>0.0125</td>
</tr>
<tr>
<td></td>
<td>( T_1 = 196.80 )</td>
<td>124764</td>
<td>0.5774</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.8

Surgical Unit Example

$H_0$: variables $X_2$ and $X_4$ vs $H_1$: variables $X_1$ and $X_3$

Normal Functional Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -53.00$</td>
<td>10.207</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -20.87$</td>
<td>7.467</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>5%</td>
<td>$T_0 = 163.60$</td>
<td>150521</td>
<td>0.6733</td>
</tr>
<tr>
<td></td>
<td>$T_1 = 10.99$</td>
<td>33799</td>
<td>0.9523</td>
</tr>
<tr>
<td>7%</td>
<td>$T_0 = 189.20$</td>
<td>74936</td>
<td>0.4894</td>
</tr>
<tr>
<td></td>
<td>$T_1 = 10.96$</td>
<td>19514</td>
<td>0.9375</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.9

Surgical Unit Example

$H_0$: variables $X_1$ and $X_4$ vs $H_1$: variables $X_2$ and $X_3$

Normal Functional Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -70.78$</td>
<td>11.418</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -15.01$</td>
<td>4.747</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>5%</td>
<td>$T_0 = 238.20$</td>
<td>333708</td>
<td>0.6801</td>
</tr>
<tr>
<td></td>
<td>$T_1 = 1.09$</td>
<td>158511</td>
<td>0.9978</td>
</tr>
</tbody>
</table>

* Percent of sample variance
Table 4.10

Surgical Unit Example

$H_0$: variable $X_4$ vs $H_1$: variables $X_1, X_2$ and $X_3$

Normal Functional Models

<table>
<thead>
<tr>
<th>Error Variance*</th>
<th>Test Statistic</th>
<th>Variance</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T_0 = -123.60$</td>
<td>12.348</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td></td>
<td>$T_1 = 2.85$</td>
<td>0.938</td>
<td>0.0032</td>
</tr>
<tr>
<td>1%</td>
<td>$T_0 = 470.90$</td>
<td>58983959</td>
<td>0.9511</td>
</tr>
<tr>
<td></td>
<td>$T_1 = -8.14$</td>
<td>1022389</td>
<td>0.9936</td>
</tr>
</tbody>
</table>

* Percent of sample variance
4.2.2 Results from Testing Normal Functional Models

Similar to the results in section 4.1, the results shown in Tables 4.7-4.10 vary in the degree in which the assumption of errors in the explanatory variables affect Cox's test of separate (nonnested) linear models.

Investigation into the application of testing nonnested normal functional models highlights the need to quantify the conditions upon which the procedure is not applicable. The results in section 4.2.1 illustrates the fact that in some cases assuming even a 'small' error variance is enough for the procedure to be non-applicable. This could be a result of several things: sample size (since large sample approximations are used), quantile selections, model assumptions, etc.

The following observations are derived from the results in Tables 4.3-4.6:

i) The p-values for testing the nonnested normal functional models had a tendency to decrease with increasing error variances, unlike the results in section 4.1,

ii) Cox's test indicated that in each case when no error in the explanatory variables is assumed, both models are rejected. When error is assumed, most often Cox's test found their was insufficient evidence in the proposed models to choose between any tested pair of models. However, Table 4.7 shows when the error variance is assumed to be 15% of the observed variance, Cox's test rejected $H_0$ (using variables $X_1$ and $X_2$) in favor of $H_1$ (using variables $X_3$ and $X_4$).

iii) The assumption of errors in the explanatory variables affected Cox's test statistic and its variance in varying ways. In most cases Cox's test statistic is negative when no error is assumed in the explanatory variables while it is positive in most cases when the explanatory variables are assumed measured with error. Typically, the variance of Cox's test statistic decreased as the error variance grew larger.
4.3 Results from Standard Testing Procedures

In Neter, Wasserman, and Kutner (1985), results are given for selecting the independent variables assuming the variables are fixed and not measured with error. Three criteria were explored for an all possible regression procedure: $R_p^2$, $MSE_p$, and $C_p$ (see NWK 1985 pp 422-428 for definitions and details).

A model employing the independent variables $X_1$, $X_2$, and $X_3$ was the selected model using each of the criteria listed above. $R_p^2$ yielded a value of 0.972, $MSE_p$ a value of 0.0020, and $C_p$ a value of 3.1.

Table 4.10 gives the result of Cox's test for directly comparing a model using the fixed independent variables $X_1$, $X_2$, and $X_3$ not measured with error versus a model using the fixed independent variable $X_4$ not measured with error. The result of Cox's test indicates both models should be rejected. This implies that a model not previously considered should be formulated.
CHAPTER 5
Discussion and Plans for Further Research

5.1 Introduction

There exists a fairly extensive body of literature on measurement error (errors in variables) models and there also exists an ever growing body of literature on testing separate families of hypotheses. To our knowledge, however, no one has attempted to integrate the theory of measurement error models with tests of separate families of hypotheses.

In this dissertation we have successfully integrated the theories of measurement error models and Cox's test of separate families of (linear) hypotheses by making various simplifying assumptions and by using large sample approximations. The Cox test of separate (nonnested) linear hypotheses was derived for both the normal structural and normal functional models. These previously underrived results are given in Chapters 2 and 3 of this dissertation.

Recent advances in the last decade or so in the theory of measurement error models have made their applications more palatable to both statisticians and non-statisticians alike. However, there is still much room for improvement to make measurement error models a generally acceptable application tool. Until then, practitioners will continue to use statistical techniques designed for explanatory variables not measured with error even though it is known that their variables are measured with error.

Cox's test of separate families of hypotheses has typically been most applied in the area of economics as the recent literature attests. Coincidently, measurement error
models have been applied extensively to the field of economics also.

In the field of biostatistics, Cox's test of separate families of hypotheses has received a relatively small amount of attention. Measurement error models are only now beginning to see more and more applications in the field of biostatistics even though knowledge of measuring explanatory variables with error is certainly nothing new in the field of biostatistics.

It seems only natural to investigate what effect measurement errors in the explanatory variables have on the Cox test of separate (nonnested) linear models. In this dissertation we were able to partially address the effect errors in variables have on the Cox test of nonnested linear models. The results have been both revealing and surprising, particularly in the results for the normal functional model (see page 74). However, much work in the future is required to establish a firm foundation as to the full effect of errors in variables on the Cox test of nonnested linear models.

5.1.1 Effect of Measurement Errors

The assumption of measurement error in the explanatory variables varied from one hypothesis testing scenario to another depending on the extent of the error variance and whether the true explanatory variables were considered stochastic (structural model) or fixed (functional model). Joint normality of errors was a constant assumption.

When structural models were assumed, there was not enough evidence to detect a difference between the various combinations of hypotheses regardless of how much error variance was conjectured. The p-values consistently increased as the error variance increased. However, this type of consistency was not exhibited in the Cox test statistic and its variance.

The assumption of functional models painted a slightly different picture. When no error in the explanatory variables was assumed, the Cox test consistently rejected both models in each of the hypothesis testing scenarios. When measurement error was
assumed, the Cox test more often than not was not able to detect a difference between the given models. However, there was one instance where the assumption of an error variance of 15% of the sample variance resulted in the Cox test rejecting one model in favor of another. The latter result is a clear case of how measurement errors can bias the hypothesis testing outcome.

It is known that maximum likelihood estimation does not lead to consistent estimates of all parameters in the functional model (see Fuller 1987). The assumptions made in testing nonnested functional models revealed just how sensitive this technique is to the size of the error variance. Sometimes even small error variances (5% of sample variance) rendered the maximum likelihood estimate of the within subject error variance untenable. Under the assumptions of the model, it was expected that maximum likelihood estimation would cause problems for some error variances but not so small an error variance as 5% or 1% of the observed sample variance.

5.2 Plans for Future Research

In this dissertation we derived the Cox test of separate linear hypotheses for both the normal structural and normal functional model. The continuous response variable is assumed to be cross-sectional and the measurement error variances of the vector explanatory variables are assumed known. The models assumed independence and homogeneity of the within subject errors. There is no assumption of missing data, but case-wise deletion of missing observations would be the most natural technique to apply.

For future research, we plan on extending our results to the following:

i) Multivariate continuous data including both longitudinal and repeated measure data. We will investigate the Cox test under the assumption of no missing data and data missing at random,

ii) Assumption of non-normality of within subject errors and heterogeneity,
iii) Nonlinear models, nonlinear in both the parameters and the explanatory variables.

The application of our derived results revealed several areas which will require further investigation:

1) An extensive numerical study to provide an appropriate procedure to determine an 'optimal' selection of the quantile design matrix used in the functional model approach of Chapter 3,

2) A method of determining appropriate bounds on the size of the error variances for testing functional models,

3) 'Power' computations for the Cox test both the structural and functional models. Power is in quotes because the standard definition of power is not applicable for nonnested hypotheses.
APPENDIX
PROOFS OF SELECTED RESULTS

This appendix provides sketches of proofs for some selected results appearing in Chapters 2 and 3 of this dissertation. The page number of the result is first given, then the result, and then its proof.

The following theorem due to Slutsky (see Bickel and Doksum 1977) is used repeatedly:

Theorem: Let \( \{Z_n\}, \{U_n\} \) be sequences of random variables indexed by \( n \), \( Z \) a random variable, and \( u_0 \) a constant. If \( Z_n \) converges in distribution to \( Z \) (\( Z_n \overset{D}{\to} Z \)) and \( U_n \) converges in probability to \( u_0 \) (\( U_n \overset{P}{\to} u_0 \)) then:

(a) \( Z_n + U_n \overset{D}{\to} Z + u_0 \)

(b) \( U_n Z_n \overset{D}{\to} u_0 Z \)

Chapter 2

1) Page 33: \( \gamma_{10} = \Sigma_{\hat{x}'x}^{-1} \Sigma_{x'x} \beta_1 \)

proof: Under \( H_0 \), \( \hat{\gamma}_1 = (m_{Z'Z} - \Sigma_{\hat{\delta}'\hat{\delta}})^{-1} m_{Z'Y} \)

\[
= (m_{Z'Z} - \Sigma_{\hat{\delta}'\hat{\delta}})^{-1} (n-1) \sum_{i=1}^{n} (Z_i - \bar{Z})' (Y_i - \bar{Y})
\]

\[
= (m_{Z'Z} - \Sigma_{\hat{\delta}'\hat{\delta}})^{-1} (n-1) \sum_{i=1}^{n} (Z_i - \bar{Z})' (\beta_0 - \beta_0 + x_i \beta_1 - \bar{x} \beta_1 + \epsilon_i - \bar{\epsilon})
\]

\[
= (m_{Z'Z} - \Sigma_{\hat{\delta}'\hat{\delta}})^{-1} (n-1) \sum_{i=1}^{n} (x_i - \bar{x} + \delta z_i - \bar{\delta} z)' [(x_i - \bar{x}) \beta_1 + \epsilon_i - \bar{\epsilon}]
\]
\begin{align*}
= (\mathbf{m}_{Z'Z} - \Sigma_{\delta z'\delta z})^{-1} \{ (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})_{1} \\
\quad + (n-1)^{-1} \sum_{i=1}^{n} (\delta z_i - \bar{\delta z})(\delta z_i - \bar{\delta z})_{1} ((x_i - \bar{x})_{1} + \epsilon_i - \bar{\epsilon}) \}
\end{align*}

Under the assumption of independence and the combination of Slutsky's Theorem and the Weak Law of Large Numbers (WLLN), the latter expression converges in probability to:

\[ \gamma_{10} = \Sigma_{x'x}^{-1} \Sigma_{z'x} \beta_{1} \cdot \]

2) Page 33: \( \gamma_{00} = \beta_{0} + \mu_{x} \beta_{1} - \mu_{z} \gamma_{10} \)

Proof: Under \( H_0 \), \( \gamma_{00} = \bar{Y} - \bar{Z} \gamma_{1} = \beta_{0} + \bar{x} \beta_{1} + \bar{\tau} - \bar{\delta} \gamma_{1} \)

Again, using Slutsky's Theorem and WLLN, the above expression converges in probability to:

\[ \gamma_{00} = \beta_{0} + \mu_{x} \beta_{1} - \mu_{z} \gamma_{10} \cdot \]

3) Page 33: \( \sigma_{u0}^{2} = \sigma_{e}^{2} + \beta_{1}'(\Sigma_{x'x} - \Sigma_{z'x} \Sigma_{x'z} \Sigma_{z'x}) \beta_{1} \)

Proof: Under \( H_0 \), \( \sigma_{u}^{2} = m_{YY} - 2m_{YZ} \gamma_{1} + \gamma_{1}' m_{Z'Z} \gamma_{1} - \gamma_{1}' \Sigma_{\delta z'\delta z} \gamma_{1} \)

\begin{align*}
= (n-1)^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 - 2m_{YZ} \gamma_{1} + \gamma_{1}' m_{Z'Z} \gamma_{1} - \gamma_{1}' \Sigma_{\delta z'\delta z} \gamma_{1} \\
= (n-1)^{-1} \sum_{i=1}^{n} [ \beta_{1}'(x_i - \bar{x})(x_i - \bar{x})_{1} - 2\beta_{1}'(x_i - \bar{x})_{1}(\epsilon_i - \bar{\epsilon}) + (\epsilon_i - \bar{\epsilon})^2 ] \\
\quad - 2m_{YZ} \gamma_{1} + \gamma_{1}' m_{Z'Z} \gamma_{1} - \gamma_{1}' \Sigma_{\delta z'\delta z} \gamma_{1} 
\end{align*}
\[= (n-1)^{1/2} \sum_{i=1}^{n} \left[ \beta'_1 (x_i - \bar{x})'(x_i - \bar{x}) \beta_1 - 2\beta'_1 (x_i - \bar{x})'(\epsilon_i - \bar{\epsilon}) + (\epsilon_i - \bar{\epsilon})^2 \right] \]

\[- 2m_{YZ} \hat{\gamma}_1 + \hat{\gamma}_1' (m_{Z'Z} - \Sigma \delta' \delta) \hat{\gamma}_1 \]

Under the assumption of independence and the combination of Slutsky's Theorem and WLLN, the latter expression converges in probability to:

\[\sigma^2_{u0} = \beta'_1 \Sigma \delta' \delta \beta_1 + \sigma^2_{\epsilon} - 2\beta'_1 \Sigma' \Sigma \gamma_{10} + \gamma'_{10} \Sigma' \gamma_{10}.\]

This latter expression can be reduced to:

\[\sigma^2_{u0} = \sigma^2_{\epsilon} + \beta'_1 (\Sigma' \Sigma - \Sigma' \Sigma_{x'x}^{-1} \Sigma_{x'x}) \beta_1.\]

Chapter 3

4) Page 50: \( \gamma_{10} = \Sigma'_{x'z} \Sigma'_{x'x} \beta_1 \)

proof: Under \( H_0 \), \( \hat{\gamma}_1 = (M_{Z'Z} - \Sigma' \delta \delta^{-1} M_{Z'Y} \)

\[= (M_{Z'Z} - \Sigma' \delta \delta^{-1} n^{-1} \sum_{i=1}^{n} z_i' y_i) \]

\[= (M_{Z'Z} - \Sigma' \delta \delta^{-1} n^{-1} \sum_{i=1}^{n} (x_i + \delta_i)' (x_i \beta_1 + \epsilon_i) \]

\[= (M_{Z'Z} - \Sigma' \delta \delta^{-1} n^{-1} \sum_{i=1}^{n} [z_i' (x_i \beta_1 + \epsilon_i) + \delta_i' (x_i \beta_1 + \epsilon_i)] \]
Under the assumption of independence, the regularity conditions on page 50, and
the combination of Slutsky’s Theorem and WLLN, the latter expression converges
in probability to:

$$\gamma_{10} = \Sigma_{x'x}^{-1} \Sigma_{x'x} \beta_1,$$

where

$$\Sigma_{x'x} = \Sigma_{X'X} - \Sigma_{\delta_x' \delta_x} \quad \text{and} \quad \Sigma_{x'x} = \Sigma_{Z'Z} - \Sigma_{\delta_x' \delta_x}.$$

5) Page 50: $A_{10} = \sigma_\varepsilon^2 + \beta_1 (\Sigma_{X'X} - \Sigma_{\delta_x' \delta_x}) \beta_1$

$$- 2\beta_1 \Sigma_{X'Z} (\Sigma_{Z'Z} - \Sigma_{\delta_x' \delta_x})^{-1} \Sigma_{Z'X} \beta_1$$

$$+ \beta_1 \Sigma_{X'Z} (\Sigma_{Z'Z} - \Sigma_{\delta_x' \delta_x})^{-1} \Sigma_{Z'Z} (\Sigma_{Z'Z} - \Sigma_{\delta_x' \delta_x})^{-1} \Sigma_{Z'X} \beta_1$$

proof: Under $H_0$, $A_1 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - Z_i \hat{\gamma}_1)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i \beta_1 + \epsilon_i - Z_i \hat{\gamma}_1)^2$

$$= \frac{1}{n} \sum_{i=1}^{n} [\epsilon_i^2 + \beta_1 x_i \beta_1 + \hat{\gamma}_1 (x_i + \delta_1) (x_i + \delta_2) \hat{\gamma}_1 - 2 \beta_1 x_i (x_i + \delta_1) \hat{\gamma}_1$$

$$+ \epsilon_i (x_i \beta_1 + - Z_i \hat{\gamma}_1)]$$
Under the assumption of independence, the regularity conditions on page 50, and the combination of Slutsky’s Theorem and WLLN, the latter expression converges in probability to:

\[ A_{10} = \sigma_1^2 + \beta_1'(\Sigma X'X - \Sigma \delta'_z \delta_z)\beta_1 \]

\[ - 2\beta_1'\Sigma X'Z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma Z'X\beta_1 \]

\[ + \beta_1'\Sigma X'Z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma Z'Z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma Z'X\beta_1. \]

6) Page 50: \[ B_{10} = \beta_1'\Sigma X'Z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma \delta'_z \delta_z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma Z'X\beta_1 \]

proof: Under \( H_0 \), \[ B_1 = \gamma_1'\Sigma \delta'_z \delta_z \gamma_1 \]

Using Slutsky's Theorem, the latter expression converges in probability to:

\[ B_{10} = \gamma_{10}'\Sigma \delta'_z \delta_z \gamma_{10} \]

\[ = \beta_1'\Sigma X'Z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma \delta'_z \delta_z(\Sigma Z'Z - \Sigma \delta'_z \delta_z)^{-1}\Sigma Z'X\beta_1. \]
REFERENCES


Dyer, A. R. (1971), A comparison of classification and hypothesis testing procedures for choosing between competing families of distributions, including a survey of the goodness of fit tests. Technical Memorandum No. 18, Aberdeen Research and Development Center, Aberdeen Proving Group, Maryland.


REFERENCES


Irwin, J. O. (1942), The distribution of the logarithm of survival times when the true law is exponential. J. Hygiene 42, 328-333.


Kent, J. T. (1986), The underlying structure of nonnested hypothesis tests, Biometrika, 73, 333-343.


REFERENCES


REFERENCES


