The limiting behavior of residuals from measurement error regressions

Stephen M. Miller
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
Miller, Stephen M., "The limiting behavior of residuals from measurement error regressions" (1986). Retrospective Theses and Dissertations. 8274.
https://lib.dr.iastate.edu/rtd/8274
INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Miller, Stephen M.

THE LIMITING BEHAVIOR OF RESIDUALS FROM MEASUREMENT ERROR REGRESSIONS

Iowa State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Ph.D. 1986
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages 
2. Colored illustrations, paper or print 
3. Photographs with dark background 
4. Illustrations are poor copy 
5. Pages with black marks, not original copy 
6. Print shows through as there is text on both sides of page 
7. Indistinct, broken or small print on several pages √
8. Print exceeds margin requirements 
9. Tightly bound copy with print lost in spine 
10. Computer printout pages with indistinct print 
11. Page(s) _________ lacking when material received, and not available from school or author.
12. Page(s) _________ seem to be missing in numbering only as text follows.
13. Two pages numbered _______. Text follows.
14. Curling and wrinkled pages 
15. Dissertation contains pages with print at a slant, filmed as received √
16. Other

__________________________________________________________________________
__________________________________________________________________________

University
Microfilms
International
The limiting behavior of residuals from measurement error regressions

by

Stephen M. Miller

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1986
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION AND LITERATURE REVIEW</strong></td>
<td></td>
</tr>
<tr>
<td>A. Definition of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>B. Review of Errors-in-Variables Regression Models</td>
<td>4</td>
</tr>
<tr>
<td>1. General linear model</td>
<td>6</td>
</tr>
<tr>
<td>2. Linear regression model</td>
<td>6</td>
</tr>
<tr>
<td>3. Multivariate general linear model</td>
<td>6</td>
</tr>
<tr>
<td>4. Multivariate linear regression model</td>
<td>7</td>
</tr>
<tr>
<td>5. Univariate structural model</td>
<td>7</td>
</tr>
<tr>
<td>6. Univariate functional model</td>
<td>10</td>
</tr>
<tr>
<td>7. Univariate ultrastructural model</td>
<td>11</td>
</tr>
<tr>
<td>8. Multivariate structural model</td>
<td>12</td>
</tr>
<tr>
<td>9. Multivariate functional model</td>
<td>13</td>
</tr>
<tr>
<td>C. Review of Weak Convergence of Empirical Processes Based on Residuals</td>
<td>14</td>
</tr>
<tr>
<td><strong>II. REPRESENTATIONS FOR RESIDUALS FROM REGRESSION MODELS</strong></td>
<td>22</td>
</tr>
<tr>
<td>A. Univariate Regression Models</td>
<td>22</td>
</tr>
<tr>
<td>B. Multivariate Regression Models</td>
<td>31</td>
</tr>
<tr>
<td><strong>III. WEAK CONVERGENCE OF SOME MODIFIED EMPIRICAL PROCESSES</strong></td>
<td>41</td>
</tr>
<tr>
<td>A. Univariate Results</td>
<td>41</td>
</tr>
<tr>
<td>B. Multivariate Results</td>
<td>65</td>
</tr>
</tbody>
</table>
# Table of Contents

IV. GOODNESS-OF-FIT STATISTICS BASED ON REGRESSION RESIDUALS 72
   A. EDF Goodness-of-Fit Statistics 72
      1. Kolmogorov-Smirnov type statistics 73
      2. Cramer-von Mises statistic 76
      3. Watson's statistic 76
      4. Anderson-Darling statistic 77
      5. Applications to multivariate data 78
   B. Chi-Squared Tests of Fit 81
   C. Tests Based on Moments 88
   D. Tests Based on L-Statistics 94

V. TESTS FOR AUTOCORRELATION, OUTLIERS, AND HOMOGENEITY OF VARIANCE 105
   A. Tests for Autocorrelation 105
   B. Tests for Outliers 109
   C. Tests for Homogeneity of Variance 116

VI. MONTE CARLO SIMULATIONS AND EXAMPLES 130
   A. Role Performance Data 130
   B. Lung Vital Capacity Data 147

VII. REFERENCES 165

VIII. ACKNOWLEDGEMENTS 172

IX. APPENDIX A: REVIEW OF WEAK CONVERGENCE 173

X. APPENDIX B: PERCENTAGE POINTS FOR SOME DISTRIBUTIONS 180
I. INTRODUCTION AND LITERATURE REVIEW

The statistical problem of regression models with Gaussian errors in potentially both the dependent and independent variables has a long history. The problem of assessing the goodness-of-fit of distributional normality to a set of data has an even longer history. Only recently has theoretical work been done to determine the limiting distribution of goodness-of-fit statistics based on residuals from classical univariate regression models - one dependent variable with fixed independent variables measured without error. We consider the problem of testing for distributional normality based on residuals from multivariate errors-in-variables regression models - one or more dependent variables with possibly random independent variables measured with error. We also consider the problem of how to test for outliers, autocorrelation, and homogeneity of variance based on errors-in-variables regression residuals. We begin by giving a mathematical description of the problem.

A. Definition of the Problem

The \textit{multivariate errors-in-variables regression model} is defined by,

\[ Y_t = X_t \beta + e_t, \quad X_t = x_t + u_t, \quad (1.1) \]
for $t = 1, \ldots, n$, where $Y_t, 1 \times r$, and $X_t, 1 \times k$, are both observable random vectors based on a sample of size $n$. The matrix $\beta, k \times r$, contains the unknown regression coefficients. The sequence of $n$ random $1 \times k$-vectors $(x_1, \ldots, x_n)$ is usually not directly observable because of measurement errors vectors $u_1, \ldots, u_n$ which are also not observable. The random vectors $e_1, \ldots, e_n$ are also not directly observable, and can be thought of as being composed of measurement errors and equation errors. Equation errors represent the fact that the linear relationship in (1.1) may not hold exactly.

Note that the model of (1.1) can also be written,

$$Y_t = X_t \beta + v_t \quad \text{for } t = 1, \ldots, n,$$

(1.2)

where

$$v_t = e_t - u_t \beta.$$

In this formulation the model looks like the usual multivariate regression model, except that now $X_t$ and $v_t$ are in general not independent (or uncorrelated) due to the measurement error in $X_t$. As in the usual multivariate regression model, if $\beta$ is known the vector $v_t$ becomes observable since,

$$v_t = Y_t - X_t \hat{\beta}.$$

(1.3)

In practice we can only obtain an estimate $\hat{\beta}$ of $\beta$. Then we define,
\[ \hat{v}_t = y_t - x_t \hat{\beta} \quad \text{for } t = 1, \ldots, n \] (1.4)

and call these the residual vectors from the model of (1.1) and (1.2).

We will restrict attention to models for which \((e_t, u_t)'\) for \(t = 1, \ldots, n\) are an iid sequence of random vectors from a \(N_p(0, \Sigma)\) distribution, where \(p = r + k\). We give more details about this assumption and others in the next section, but for now notice that this implies that \(v_t'\) for \(t = 1, \ldots, n\) is an iid sequence from a \(N_r(0, \Sigma_{vv})\) distribution. Now consider the following assumptions.

**Assumption 1.1.** As the sample size \(n\) increases the sequence of random \(1 \times k\)-vectors \((x_1, \ldots, x_n)\) in model (1.1) has the property that,

\[
\frac{1}{n} \sum_{t=1}^{n} x'_t x_t + L_2 \quad \text{a.s.} \quad \frac{1}{n} \sum_{t=1}^{n} x'_t + L_1 \quad \text{a.s.}
\]

where \(L_2\) is a finite \(k \times k\) matrix, and \(L_1\) is a finite \(1 \times k\) vector. This implies that \(m_{xx} + \bar{m}_{xx}\) a.s. where \(\bar{m}_{xx}\) is a finite \(k \times k\) matrix and,

\[
m_{xx} = (n - 1)^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(x_t - \bar{x})', \quad \bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t \quad \text{.} \quad (1.5)
\]

**Assumption 1.2.** The estimator \(\hat{\beta}\) is consistent and \(\hat{\beta} - \beta = O_p(n^{-1/2})\). This is called \(n^{1/2}\)-consistency.
We show that these two assumptions are sufficient conditions to justify the use of the residual vectors defined in (1.4) in goodness-of-fit tests of distributional normality of $v_t$'s defined in (1.2). We do this by examining the weak convergence of empirical processes associated with the residuals defined in (1.4). We also show that, in large samples, these are sufficient conditions to use the regression "t-tests" and "F-tests" from the regression of the residuals on lagged values of the residuals as a test of autocorrelation. We also show that assumptions 1.1 and 1.2 are sufficient conditions for the maximum standardized errors-in-variables regression residual to be used in tests for outliers. We also show that assumptions 1.1 and 1.2 are sufficient conditions to use the regression "t-tests" and "F-tests" from the regression of the squares of the residuals on certain predicted values of the unobservable $x_t$'s.

In the remainder of this chapter we review errors-in-variables regression models and review the subject of weak convergence of empirical processes based on residuals.

B. Review of Errors-in-Variables Regression Models

The errors-in-variables regression model described in (1.1) is quite general. The model states that,

$$Y_t = x_t \delta + e_t, \quad X_t = x_t + u_t,$$

for $t = 1, \ldots, n$, which can also be written,
\[ Y_t = X_t \beta + v_t \quad \text{for } t = 1, \ldots, n, \quad (1.7) \]

where
\[ v_t = e_t - u_t \beta. \]

The vector \((Y_t, X_t)\) for \(t = 1, \ldots, n\) is observable, where \(Y_t\) is \(1 \times r\), and \(X_t\) is \(1 \times k\). The matrix \(\beta\), \(k \times r\), contains the unknown regression coefficients. In addition we assume that
\[ (e_t, u_t)' \text{ are iid } N_p(0, \Sigma), \quad (1.8) \]

where \(p = k + r\), and are jointly independent of the sequence of random \(1 \times k\)-vectors \((x_1, \ldots, x_n)\). We allow the covariance matrix \(\Sigma\) to be singular, but require \(\Sigma_{vv} = V(v_t)\) to be positive definite, where
\[ \Sigma_{vv} = (I_r - \beta') \Sigma (I_r - \beta'). \quad (1.9) \]

We also assume that the sequence of random \(1 \times k\)-vectors \((x_1, \ldots, x_n)\) has the property that,
\[ n^{-1} \sum_{t=1}^{n} x_t' x_t + L_2 \text{ a.s.}, \quad n^{-1} \sum_{t=1}^{n} x_t + L_1 \text{ a.s.}, \quad (1.10) \]

where \(L_2\) is a finite \(k \times k\) matrix, and \(L_1\) is a finite \(1 \times k\) vector. This implies that \(m_{xx} + \tilde{m}_{xx}\) a.s. where \(\tilde{m}_{xx}\) is a finite matrix and \(m_{xx}\) is defined in (1.5). The explanation from (1.6) to (1.10) completely specifies the model.
For notational convenience we write \( \Sigma \) in partitioned form as,

\[
V((e_t, u_t)') = \Sigma = \begin{pmatrix}
\Sigma_{ee} & \Sigma_{eu} \\
\Sigma_{ue} & \Sigma_{uu}
\end{pmatrix}
\]

(1.11)

where \( \Sigma_{ee} \) is \( r \times r \), \( \Sigma_{eu} = \Sigma_{ue}' \) is \( r \times k \), and \( \Sigma_{uu} \) is \( k \times k \).

We next review some special cases of the model.

1. **General linear model**

   In this case \( r = 1 \), \( \Sigma_{uu} = 0 \), \( \Sigma_{eu} = 0 \), \( \sigma_{ee} > 0 \) and \( x_1, \ldots, x_n \) are the first \( n \) elements of an infinite sequence of fixed \( 1 \times k \)-dimensional vectors which satisfy (1.10). This model has a long history, and is widely used, particularly in experimental designs. A good reference is Searle (1971).

2. **Linear regression model**

   In this case \( r = 1 \), \( \Sigma_{uu} = 0 \), \( \Sigma_{eu} = 0 \), \( \sigma_{ee} > 0 \) and \( x_1, \ldots, x_n \) are \( 1 \times k \)-vectors of the form \( x_t = (1, x_{t2}) \) where \( x_{t2} \)

   \( t = 1, \ldots, n \) is a sample from a distribution on \( \mathbb{R}^{k-1} \) with finite second moments. A good reference is Graybill (1976).

3. **Multivariate general linear model**

   In this case \( r > 1 \), \( \Sigma_{uu} = 0 \), \( \Sigma_{eu} = 0 \), \( \Sigma_{ee} \) is positive definite, and the \( x_t \)'s satisfy the same conditions as in the general linear model. This model is widely used in MANOVA as well as in econometrics for systems of linear equations. Good references are Mardia, Kent and Bibby (1979), and Johnson and Wichern (1982).
4. **Multivariate linear regression model**

In this case \( r > 1 \), \( \Sigma_{uu} = 0 \), \( \Sigma_{eu} = 0 \), \( \Sigma_{ee} \) is positive definite, and the \( x_t \)'s satisfy the same conditions as in the linear regression model. In addition to the references given for the multivariate general linear model, also see Anderson (1984a).

5. **Univariate structural model**

We first examine the simple univariate model. In this case \( r = 1 \), \( k = 2 \), with the \( x_t \)'s of the form \( x_t = (1, x_{t2}) \) where \( x_{t2} \) for \( t = 1, \ldots, n \) is a sample a distribution on \( \mathbb{R} \) with finite second moment. In addition \( \sigma_{ee} > 0 \) and,

\[
\begin{pmatrix}
0 & 0 \\
0 & \sigma_{uu}
\end{pmatrix}, \quad \Sigma_{eu} = (0, \sigma_{eu}),
\]

where \( \sigma_{uu} > 0 \) and \( \sigma_{eu}^2 < \sigma_{uu} \sigma_{ee} \). Notice that if \( \sigma_{uu} = 0 \) this reduces to the simple linear regression model described above.

This model has a long history, and has received an extensive amount of research. We briefly review the subject next and refer the reader to other sources for more details.

Primary interest has centered on estimating the slope parameter \( \beta_1 \), where we write \( \beta' = (\beta_0, \beta_1) \) and \( \beta_0 \) is the intercept. Reiersol (1950) proved that if no other information is available about the model, then a necessary and sufficient condition for \( \beta_1 \) to be identified is that the distribution of the \( x_{t2} \)'s not be normal. Under the
assumption of nonnormality, consistent estimators have been proposed based on higher order cumulants, see Geary (1942, 1943), and higher order moments, see Drion (1951) and Scott (1950). These methods have not been popular since they are messy to compute and can have large variability due to the use of higher order moments. Bickel and Ritov (1986) take a semiparametric approach by treating the nonnormal distribution of the $x_{t2}$'s as an unknown parameter. They describe how to obtain efficient estimates of $\beta_1$ in the sense of Koshenik and Levitt (1976). This semiparametric approach of treating the distribution of the $x_{t2}$'s as an unknown parameter was first investigated by Wolfowitz (1952, 1953, 1954, 1957) and Kiefer and Wolfowitz (1956).

The assumption of nonnormality described above is just one of many ways to impose additional information into the model, and each way has its own particular advantages and disadvantages. A possible drawback of the estimators based on nonnormality described above is that it is unclear how effective they are when the distribution of the $x_{t2}$'s is close to normal, since by Reiersol's result the situation is close to being nonidentified.

Another way to introduce additional information is to assume another vector $W_t$ for $t = 1, \ldots, n$, correlated with $x_{t2}$, is observed which can be used as an instrument to estimate $\beta_1$. For more information on instrumental variable estimators see Carter (1976), Carter and Fuller (1980), and Fuller (1986). It is interesting to note that the "gouping method" of estimation proposed by Wald (1940) is
actually an instrumental variable estimator with the ability to form
groups acting as the instrument. Unfortunately this method is not very
practical since, as pointed out by Fuller (1986) and Bekker (1986), the
conditions needed to ensure the consistency of the estimator are usually
not met in practice.

In some situations it is possible to observe repeated replications,
and this information can be used to construct consistent estimators.
For more information see Tukey (1951).

By far the most common way of imposing additional information has
been to make assumptions about the variances and covariances $\sigma_{ee}$,
$\sigma_{eu}$, $\sigma_{uu}$ and $\sigma_{xx}$ where $\sigma_{xx} = V(x_{2t})$. Some of these include,

a) $\sigma_{uu}$ known, $\sigma_{eu} = 0$, $\sigma_{ee} > 0$, and $\sigma_{xx} > 0$.

b) $\delta = \sigma_{uu}^{-1}\sigma_{ee}$ known, with $\sigma_{eu} = 0$.

c) $(\sigma_{xx} + \sigma_{uu})^{-1}\sigma_{xx}$ known.

d) $\sigma_{ee}$, $\sigma_{eu}$, $\sigma_{uu}$ all known.

e) $\sigma_{ee}$, $\sigma_{eu}$, $\sigma_{uu}$ all known up to a common multiple.

Consistent estimators for these cases, and others, exist and more
information can be found in Fuller (1986) and Moran (1971) and the
references contained in them. The method of maximum likelihood is
commonly used to construct estimators, by assuming that $(x_{t2}, e_t, u_t)$
are jointly normally distributed for $t = 1, \ldots, n$. 
The exact distributions of the estimators we have discussed are usually not known in finite samples, so it is common to rely on large sample theory to approximate the distributions. The distributional assumption of normality for \((e_t, u_t)\) often yields convenient expressions for the covariance matrix of the limiting distributions of the estimates.

We have confined out discussion to the simple univariate structural model with scalar \(x_{t2}\), but similar comments hold for the general univariate structural model with vector \(x_{t2}\) from a distribution on \(\mathbb{R}^m\) where \(m > 1\). See Fuller (1986).

6. Univariate functional model

We first examine the simple univariate functional model. In this case \(r = 1\), \(k = 2\), and the \(x_t's\) are of the form \(x_t = (1, x_{t2})\) where \(x_{12}, \ldots, x_{n2}\) are the first \(n\) elements of an infinite sequence of real numbers which satisfy (1.10). As in the structural model we assume \(\sigma_{ee} > 0\) and,

\[
\Sigma_{uu} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{uu} \end{pmatrix}, \quad \Sigma_{eu} = (0, \sigma_{eu}),
\]

where \(\sigma_{uu} > 0\) and \(\sigma_{eu}^2 < \sigma_{uu} \sigma_{ee}\). Note that if \(\sigma_{uu} = 0\) this reduces to the simple linear model.

The functional model is similar to the structural model described before, except that now the \(x_{t2}'s\) are unknown parameters. Imagine an infinite sequence of observations for our model \((Y_t, X_t)\) for
t = 1, ..., n. Neyman and Scott (1948) referred to parameters entering the distribution of the observations for finitely many t as incidental parameters and those entering for infinitely many t as structural parameters. Generally, we can only hope to estimate the structural parameters consistently. In our model the $x_{t2}$'s are incidental parameters and $\beta$, $\sigma_{uu}$, $\sigma_{eu}$, and $\sigma_{ee}$ are structural parameters.

As in the structural model, additional information is needed in order to estimate the parameters of the model. As before, the methods of instrumental variables and of repeated observations can be applied. It is most common to make assumptions about $\sigma_{ee}$, $\sigma_{eu}$ and $\sigma_{uu}$. We will not pursue the matter any further except to say that consistent estimators of $\beta$ can be constructed. For more on the subject see Fuller (1986). For a Bayesian approach see Lindley and El-Sayyad (1968). We have confined our discussion to the simple univariate functional model with scalar $x_{t2}$ but similar comments hold for the general univariate functional model with vector $x_{t2}$.

7. Univariate ultrastructural model

We first examine the simple univariate unreplicated ultrastructural model of Dolby (1976). In this case $r = 1$, $k = 2$, and the $x_t$'s are of the form $x_t = (1, x_{t2})$ where

\[
\begin{pmatrix}
  x_{2t} \\
  e_t \\
  u_t
\end{pmatrix}
\sim N\left( \begin{pmatrix}
  \mu_t \\
  0 \\
  0
\end{pmatrix}, \begin{pmatrix}
  \sigma_{xx} & 0 & 0 \\
  0 & \sigma_{ee} & 0 \\
  0 & 0 & \sigma_{uu}
\end{pmatrix} \right),
\]

(1.12)
for $t = 1, \ldots, n$. Dolby (1976) called the model generated by (1.12) **ultrastructural** because it generalizes the usual structural model by allowing the $x_{t2}$'s to have different means $\mu_t$. If $\mu_1 = \ldots = \mu_n$ then (1.12) reduces to a structural model, while if $\sigma_{xx} = 0$ it reduces to the usual functional model. Gleser (1985) has derived the maximum likelihood estimators for the model under the assumption that the ratio $\delta = \sigma_{ee} \sigma_{uu}^{-1}$ is known, and derived the form of the limiting distribution of the estimators under the assumption that,

$$\bar{\mu} + \bar{m} < \infty, \quad m_{\mu \mu} + \bar{m}_{\mu \mu} < \infty$$

as $n \to \infty$.

The assumption of trivariate normality in (1.12) is not a necessary condition in order to obtain a limiting distribution for the estimators above. In general all we need is $(x_{t2} - \mu_t)$ for $t = 1, \ldots, n$ to be a sample from a distribution with finite second moment. For more on this model as well as the case of vector $x_{t2}$ see Fuller (1986).

8. **Multivariate structural model**

In this case $r > 1$, $k > 2$ with the $x_t$'s of the form $x_t = (1, x_{t2})$ where $x_{2t}$ for $t = 1, \ldots, n$ is a sample from a distribution on $\mathbb{R}^{k-1}$ with finite second moment. A special case is the factor analysis model which assumes $\Sigma_{eu} = 0$, with $\Sigma_{ee}$ and $\Sigma_{uu}$ both diagonal matrices. See Anderson and Rubin (1956), and Lawley and Maxwell (1971) for more discussion of the factor analysis model.
A common method of estimation for the multivariate structural model is to assume the $x_{t2}$'s are a sample from a normal distribution, and use the method of maximum likelihood. This is the approach taken by Joreskog (1970, 1973), and implemented in the computer program LISREL VI by Joreskog and Sorbom. The limiting distribution of the estimators can be derived under the weaker condition that the $x_{t2}$'s are a sample from a distribution on $\mathbb{R}^{k-1}$ with finite second moments. This is the approach of Fuller, Amemiya, and Pantula (1985), and implemented in the program ISU FACTOR written by Sastry Pantula.

Anderson (1973), Browne (1974) and others have applied the method of generalized least squares to multivariate structural models. In many cases the method of maximum likelihood and generalized least squares yield asymptotically equivalent estimators. For more on multivariate structural models see Amemiya (1982), Amemiya and Fuller (1984), Anderson (1984b), Amemiya (1986), Fuller (1986) and the references contained in them.

9. Multivariate functional model

In this case $r > 1$, $k > 2$ with the $x_t$'s of the form

$$x_t = (1, x_{t2})$$

where $x_{12}, \ldots, x_{n2}$ are the first n elements of an infinite sequence of real valued vectors in $\mathbb{R}^{k-1}$ which satisfy (1.10).

Gleser and Watson (1973) considered maximum likelihood estimation for the multivariate functional model under the assumption that $r = k$, $\Sigma_{eu} = 0$, $\Sigma_{ee} = \sigma^2 \xi$, $\Sigma_{uu} = \sigma^2 \phi$ where $\phi$ is a known $k \times k$ matrix, and $\sigma^2$ is an unknown positive constant. Bhargava (1977, 1979) and Gleser (1981) generalized these results. Dahm and Fuller (1986) applied
the method of generalized least squares to the multivariate functional model and derived the limiting distribution of the estimators. For more on the functional model see Dahm (1979), Amemiya (1982), Fuller (1986) and the references contained in them.

C. Review of Weak Convergence of Empirical Processes Based on Residuals

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution $F$ on the real line. It is well known that $U_i = F(X_i)$ for $i = 1, \ldots, n$ are iid uniform random variables on the unit interval. It is a classic result that the stochastic process

$$\{W_n(\omega): \omega \in [0, 1]\}$$

defined by,

$$W_n(\omega) = n^{-1/2} \sum_{i=1}^{n} [1(U_i < \omega) - \omega], \quad (1.13)$$

converges in law to the Brownian bridge process on $[0, 1]$. The Brownian bridge is a Gaussian process with mean and covariance function given by,

$$E(W(\omega)) = 0 \quad \text{for all } \omega \in [0, 1], \quad (1.14)$$

$$E(W(\omega_1)W(\omega_2)) = \min(\omega_1, \omega_2) - \omega_1 \omega_2 \quad \text{for } \omega_1, \omega_2 \in [0, 1],$$

where $W$ is a Brownian bridge process. In general $W_n$ is called an empirical process because it is of the form,
\[ n^{1/2} \left[ G_n(x) - G(x) \right] \text{ for } x \text{ in the support of } G, \quad (1.15) \]

where \( G \) is a distribution function, and \( G_n \) is the empirical distribution function from a sample of size \( n \). We also can define the process \( \{ W_n'(x): x \in \mathbb{R} \} \) by,

\[ W_n'(x) = n^{-1/2} \sum_{t=1}^{n} \left[ 1(X_t < x) - F(x) \right], \quad (1.16) \]

which converges to a limiting gaussian process \( \{ W'(x): x \in \mathbb{R} \} \) with mean and covariance function given by,

\[ E[W'(x)] = 0 \text{ for all } x \in \mathbb{R}, \quad (1.17) \]

\[ E[W'(x_1)W'(x_2)] = \min[F(x_1), F(x_2)] - F(x_1)F(x_2) \text{ for } x_1, x_2 \in \mathbb{R}. \]

For more on these results see Billingsley (1968), Doob (1949), Donsker (1951), and Shorack and Wellner (1986).

In general, the distribution function \( F \) may be a function of a vector of parameters \( \theta \) in some space \( \Theta \). Durbin (1973a) proved the weak convergence of the empirical process based on \( \hat{U}_i = F(X_i; \hat{\theta}) \) for \( i = 1, \ldots, n \) where \( \hat{\theta} \) is an estimator of \( \theta \). Kac, Kiefer and Wolfowitz (1955) had earlier shown that all finite dimensional distributions converged to a multivariate normal distribution. Neuhaus (1976) extended Durbin's results to a wider class of models. Both Durbin and Neuhaus require \( \hat{\theta} \) to be of the form,
where the random vector $Z(X; \theta)$ has a finite covariance matrix. Wood (1978) looked at the special case of location-scale families of distributions, and obtained the results of Durbin and Neuhaus under more easily verified conditions. Rao and Sethuraman (1975) had earlier looked at the case of scale families, and random estimates of scale (which they also call perturbations).

Loynes (1980) looked at the weak convergence of empirical processes based on generalized residuals defined in Cox and Snell (1968). Under that model, data $X_1, \ldots, X_n$ are generated by $X_i = g_i(\epsilon_i, \theta)$ for $i = 1, \ldots, n$, where $\theta$ is an unknown parameter; $g_1, \ldots, g_n$ are a sequence of known functions; and $\epsilon_1, \ldots, \epsilon_n$ are iid random variables. Given an estimate $\hat{\theta}$ the generalized residuals are defined by,

$$\hat{\epsilon}_i = h_i(X_i, \hat{\theta}) \text{ for } i = 1, \ldots, n,$$  

(1.19)

where $h_i$ is uniquely defined (by hypothesis) by the requirement that $X_i = g_i(\epsilon_i, \theta)$ if and only if $\hat{\epsilon}_i = h_i(X_i, \hat{\theta})$. Loynes' results are quite general, but the sufficient conditions are difficult to verify. As an application Loynes gives the model,

$$X_i = \alpha + u_i \beta + \epsilon_i \text{ for } i = 1, \ldots, n,$$  

(1.20)
where $\varepsilon_1, \ldots, \varepsilon_n$ are iid $N(0, \sigma^2)$ and $\{u_i\}_{i=1}^{\infty}$ is a sequence of fixed numbers such that $n^{-1} \sum_{i=1}^{n} u_i + b < \infty$ as $n \to \infty$. Letting $\hat{\alpha}$ and $\hat{\beta}$ denote the least squares estimators of $\alpha$ and $\beta$, and defining

$$\hat{\varepsilon}_i = X_i - \hat{\alpha} - u_i \hat{\beta} \quad \text{for} \quad i = 1, \ldots, n, \quad (1.21)$$

and

$$s^2 = (n - 2)^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2,$$

Loynes showed that the empirical process $W_n$ defined by,

$$W_n(\omega) = n^{-1/2} \sum_{i=1}^{n} \{1[\Phi(s^{-1} \hat{\varepsilon}_i) < \omega] - \omega\} \quad \text{for} \quad \omega \in [0, 1], \quad (1.22)$$

converges in law to the Gaussian process $W$ with mean and covariance function given by

$$E(W(\omega)) = 0 \quad \text{for} \quad \omega \in [0, 1], \quad (1.23)$$

$$E(W(\omega_1)W(\omega_2)) = \min(\omega_1, \omega_2) - \omega_1 \omega_2$$

$$- [1 + \frac{1}{2} \Phi^{-1}(\omega_1) \Phi^{-1}(\omega_2)] \Phi^{-1}(\omega_1) \Phi^{-1}(\omega_2) \quad \text{for} \quad \omega_1, \omega_2 \in [0, 1].$$
We have used the notation that \( \phi \) is the standard normal distribution function and \( \phi \) is its density. Result (1.23) had earlier been obtained by Pierce and Kopecky (1979). In both of these cases the authors needed \( \hat{\alpha} \) and \( \hat{\beta} \) to be asymptotically normal, while Pierce and Kopecky also needed the fact that \( \hat{\alpha} \) and \( \hat{\beta} \) are efficient. Mukantseva (1977) earlier obtained the same results as Loynes, and Pierce and Kopecky, but under weaker conditions on the \( u_i \)'s. In particular, he did not use the limiting normality and efficiency of \( (\hat{\alpha}, \hat{\beta}) \) in his proof. Wood (1981b) extended Mukantseva's method to the case of ordinary least squares with vector valued \( u_i \).

Wood (1984) considered the following model. Let \( Y \) be an \( n \times 1 \) vector of observations from the multiple regression model \( Y = X\beta + \varepsilon \) where \( X = (X_1, X_2, \ldots, X_k) \) is an \( n \times k \) non-stochastic matrix of constants; \( X_1 \) is a column of ones; \( \beta \) is a \( k \times 1 \) vector of unknown regression coefficients; and \( \varepsilon \) is an \( n \times 1 \) vector of iid \( N(0, \sigma^2) \) random variables with \( \sigma^2 > 0 \). She considered the ridge regression estimator,

\[
\hat{b} = (X'X + \theta_n D)^{-1}X'Y,
\]

where \( \{\theta_n\}_{i=1}^{\infty} \) is a sequence of positive constants converging to zero; \( D_2 = \text{diag}(1^{X_2}_{12}, ..., 1^{X_k}_{12}) \); \( D = \text{diag}(0, D_2) \); \( |X_i|^2 = \sum_{j=1}^{n} x_{ij}^2 \) and \( X_i \) is centered about zero. She showed that the empirical process based on the transformed variables \( \phi(s^{-1}e_i) \) for \( i = 1, \ldots, n \) where,
\[ e = Y - Xb, \quad s^2 = (n - k)^{-1}e'e, \quad (1.25) \]

converges to the limiting Gaussian process given in (1.23). She required that,

\[ \max_{2 < k \leq n} D_{i}^{-1/2} (X'X_{2}) D_{i}^{-1/2} + C \quad (1.26) \]

as \( n \to \infty \), where \( C \) is a positive definite matrix and \( X = (X_1' X_2') \). Under the same model set up, Shorack (1985) proved that if \( b \) is any estimator such that \( b = \beta + O_p(n^{-1/2}) \) and if \( n^{-1}X'X + C \) a positive definite matrix, then the Wood's result still holds.

These results are important for the following reason. Let \( Z_1, \ldots, Z_n \) be iid from \( N(\mu, \sigma^2) \) where \( \sigma^2 > 0 \). Then the empirical process based on the transformed data \( \phi[s^{-1}(Z_i - \bar{Z})] \) for \( i = 1, \ldots, n \), where \( \bar{Z} \) and \( s^2 \) are the sample mean and variance, converges weakly to the limiting Gaussian process in (1.23). Thus in terms of weak convergence of the associated empirical processes, standardized regression residuals behave in large samples like a set of standardized iid observations. It has been pointed out that this has important applications to goodness-of-fit statistics since a large class of tests known as EDF statistics are functions of the empirical process, and the limiting distributions of these statistics are determined by the weak convergence of the empirical process. Thus in large samples, composite tests of normality (with estimated mean and variance) such as
Kolmogorov-Smirnov and Cramer-von Mises when based on regression residuals have the same limiting distribution as in the usual one sample tests of composite normality based on iid observations. For more on limiting distributions for functions of empirical processes see Durbin (1973b). See Stephens (1974) for tables of percentage points of the limiting distributions of many EDF statistics.

Wood (1981a) considered empirical processes based on random vectors. Let the $p \times 1$ vectors $Y_1, \ldots, Y_n$ be iid from $N_p(\mu, \Sigma)$ where $\Sigma$ is positive definite. Let $\bar{\mu}$ and $\Sigma$ be the sample mean vector and covariance matrix, and define,

$$Z_i = S^{-1/2}(Y_i - \bar{\mu}) \quad \text{for } i = 1, \ldots, n,$$

where $S^{-1/2}$ is the symmetric square root of $S^{-1}$, and

$$Z'_i = (Z_{i1}, \ldots, Z_{ip})^T.$$  

Wood showed that the process $W_n$ defined by,

$$W_n(\omega) = (pn)^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ I[\Phi(Z_{ij}) < \omega] - \omega \right\},$$

for $\omega \in [0, 1]$, converges for fixed $p$ and $n \to \infty$ to the limiting Gaussian process defined in (1.23). Thus tests of multivariate normality can be constructed by applying the composite EDF tests of univariate normality to the collection of $np$ variables $\{Z_{ij}\}$. For more see Wood (1981a).

We have concentrated on empirical processes based on residuals from regression models. Recently work has also been done on the weak
convergence of empirical processes based on residuals from p-th order autoregressive time series. See Pierce (1985) and Wood (1985).
II. REPRESENTATIONS FOR RESIDUALS FROM REGRESSION MODELS

The residuals from measurement error regression models are complicated functions of the data. In general, their exact behavior will depend on many factors, including the behavior of the unobservable \( x_t \)'s and the method of estimation of the parameters of the model. Nevertheless, we show below that under our model assumptions given in Chapter One we can represent the residuals in relatively simple theoretical forms. While not useful in practice, these theoretical representations sufficiently characterize their behavior to allow us to prove general theorems in later chapters. We begin by considering the univariate model, and then generalize the results to multivariate models.

A. Univariate Regression Models

Recall that the model can be written,

\[
Y_t = X_t \beta + v_t, \quad x_t = x_t + u_t, \tag{2.1}
\]

for \( t = 1, \ldots, n \), where \( v_t = e_t - u_t \beta \). By introducing the random variable \( x_t \), we can also write the model as,

\[
Y_t = x_t \beta + (1 + \sigma_{vv}^{-1} \Sigma_{vu}) v_t \quad \text{for} \quad t = 1, \ldots, n, \tag{2.2}
\]

\[
x_t = x_t + v_t \sigma_{vv}^{-1} \Sigma_{vu},
\]
where \( \bar{x}_t \) is defined by the relation,

\[
\bar{x}_t = x_t - v_t \sigma_{vv}^{-1} \Sigma_{vu} \quad \text{for } t = 1, \ldots, n, \tag{2.3}
\]

and \( \Sigma_{vu} = \Sigma(v_t, u_t) = \Sigma_{uu} - \sigma_{vu}^2 \). The random variable \( \bar{x}_t \) is a predictor of \( x_t \). It is constructed by subtracting from \( x_t \) the best predictor of \( u_t \) given \( v_t \) under the normal model. Note that we can write,

\[
\bar{x}_t = x_t + u_t - v_t \sigma_{vv}^{-1} \Sigma_{vu},
\]

so under our model \( \bar{x}_t \) and \( v_t \) are independent since \( x_t \) and \( v_t \) are independent by assumption while,

\[
v_t \quad \text{and} \quad u_t - v_t \sigma_{vv}^{-1} \Sigma_{vu},
\]

are independent by the assumption of normality. It is also true that \( v_1, \ldots, v_n \) are iid \( N(0, \sigma_{vv}) \) and jointly independent of the sequence of \( 1 \times k \) vectors \( (\bar{x}_1, \ldots, \bar{x}_n) \). In addition we also have the following result.

**Theorem 2.1.** Under the assumptions of our model,

\[
n^{-1} \sum_{t=1}^{n} x_t' x_t + L_2 + \Sigma_{\eta\eta} \quad \text{a.s.},
\]

\[
n^{-1} \sum_{t=1}^{n} x_t + L_1 \quad \text{a.s.},
\]
where \( \V_{xx} \) is a finite matrix and,

\[
\V_{xx} = (n - 1)^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(x_t - \bar{x})', \quad \bar{x} = n^{-1} \sum_{t=1}^{n} x_t. 
\]  

(2.4)

Proof. We can write,

\[
x_t = x_1 + \eta_t \quad \text{for} \quad t = 1, \ldots, n,
\]

(2.5)

where \( \eta_1, \ldots, \eta_n \) are iid \( N_k(0, \Sigma_{\eta\eta}) \) and independent of \( (x_1, \ldots, x_n) \) with,

\[
\Sigma_{\eta\eta} = \Sigma_{uu} - \sigma_{vv}^{-1} \Sigma_{uv} \Sigma_{vu}.
\]

(2.6)

It follows immediately that \( \bar{x} \rightarrow L_1 \). We next show that \( \V_{xx} \rightarrow \V_{xx} \) a.s. The other convergence result follows by an analogous proof. Note that,

\[
\V_{xx} = \V_{xx} + \M_{x\eta} + \M_{\eta x} + \M_{\eta\eta},
\]

(2.7)

with notation analogous to (2.4). Since \( \V_{xx} \rightarrow \V_{xx} \) a.s. by assumption, and \( \M_{\eta\eta} \rightarrow \Sigma_{\eta\eta} \) a.s. by the strong law of large numbers it is enough to show that \( \M_{x\eta} \rightarrow 0 \) a.s. We prove this in the scalar case only. The
matrix result follows by applying the argument below to each element of the matrix. In terms of scalars we can write,

\[ m_{x\eta} = (n - 1)^{-1} \sum_{t=1}^{n} (x_t - \bar{x}) \eta_t. \]

Define \( z_t = (x_t - \bar{x}) \eta_t \) for \( t = 1, \ldots, n \). Then conditional on \((x_1, \ldots, x_n)\) the \( z_t \)'s are independent with conditional mean and variance,

\[ E(z_t | x_1, \ldots, x_n) = 0, \quad V(z_t | x_1, \ldots, x_n) = (x_t - \bar{x})^2 \sigma_{\eta \eta}. \]

Let \( \{x_i\}_{i=1}^\infty \) be a sequence such that \( m_{xx} + m_{xx} \). By assumption we can find such a sequence with probability one. By the Lemma 2.1 which follows this proof, we know that \( m_{xx} + m_{xx} \) implies that,

\[ \sum_{t=1}^{\infty} t^{-2} (x_t - \bar{x})^2 < \infty. \]

Thus by Kolmogorov's criterion for the strong law of large numbers \( m_{x\eta} + 0 \) a.s. conditional on the sequence \( \{x_i\}_{i=1}^\infty \). But since such a sequence occurs with probability one, it follows that \( m_{x\eta} + 0 \) a.s. unconditionally as well.

Lemma 2.1. Let \( \{w_i\}_{i=1}^\infty \) be a sequence of real numbers such that \( n^{-1} \sum_{k=1}^{n} w_k^2 + L < \infty \) as \( n \to \infty \). Then \( \sum_{k=1}^{\infty} k^{-2} w_k^2 < \infty. \)
Proof. Let \( \{a_i\}_{i=1}^{\infty}, \{b_i\}_{i=1}^{\infty} \) be two sequences of real numbers, and define \( s_n = \sum_{i=1}^{n} b_i \) for \( n > 1 \). Abel's partial summation formula says that,

\[
\sum_{i=1}^{n} a_i b_i = a_n s_n + \sum_{i=1}^{n} (a_i - a_{i+1}) s_i.
\]

Now let \( a_k = k^{-2} \) and \( b_k = w_k^2 \). Then,

\[
\sum_{k=1}^{n} k^{-2} w_k^2 = (n + 1)^{-2} \left( \sum_{j=1}^{n} w_j^2 \right) + \sum_{j=1}^{n} \left[ \frac{2j+1}{j^2(1+j)^2} \right] \left( \sum_{i=1}^{j} w_i^2 \right).
\]

By assumption there exists a \( K < \infty \) such that \( n^{-1} \sum_{k=1}^{n} w_k^2 < K \) for all \( n \). Thus,

\[
\sum_{k=1}^{n} k^{-2} w_k^2 < n^{-1} K + 2 \sum_{j=1}^{n} \frac{j(1+j)}{j^2} ,
\]

\[
< K(n^{-1} + \pi^2/3) .
\]

Thus, \( \sum_{k=1}^{\infty} k^{-2} w_k^2 < \infty \). \( \square \)

Let \( \hat{\beta} \) be an estimator of \( \beta \) such that \( \hat{\beta} - \beta = O_p(\sqrt{n}) \). The residual \( \hat{v}_t \) for \( t = 1, \ldots, n \) is given by,

\[
\hat{v}_t = Y_t - X_t \hat{\beta} , \tag{2.8}
\]

\[
= v_t - X_t (\hat{\beta} - \beta) ,
\]
In the notation of Fuller (1986) we define,

\[ s_{vv} = (n - k)^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2, \quad \bar{v} = \frac{1}{n} \sum_{t=1}^{n} v_t. \]  

(2.9)

By using (2.4) we can write \( s_{vv} \) as,

\[ s_{vv} = \{m_{vv}[1 - \sigma_{vv}^{-1} \xi_{vu}(\hat{g} - \bar{g})]^2 - 2[1 - \sigma_{vv}^{-1} \xi_{vu}(\hat{g} - \bar{g})]m_{vx}(\hat{g} - \bar{g}) \]

\[ + (\hat{g} - \bar{g})m_{xx}(\hat{g} - \bar{g})(n - k)^{-1}(n - 1). \]

Under our assumptions, \( m_{xx}^n \to 0 \) a.s. which follows by the same proof used to show \( m_{xx} \to 0 \) a.s. in Theorem 2.1 above. Thus we can write,

\[ s_{vv} = m_{vv}[1 - \sigma_{vv}^{-1} \xi_{vu}(\hat{g} - \bar{g})]^2 + o_p(n^{-1/2}). \]  

(2.10)

Now returning to expression (2.8) we can write for \( t = 1, \ldots, n \),

\[ s_{vv}^{-1/2}(\hat{v}_t - \bar{v}) = s_{vv}^{-1/2}(\hat{v}_t - \bar{v})(1 - \sigma_{vv}^{-1} \xi_{vu}(\hat{g} - \bar{g})) - s_{vv}^{-1/2}(\bar{x}_t - \bar{x})(\hat{g} - \bar{g}), \]

\[ = \tilde{m}_{vv}^{-1/2}(\hat{v}_t - \bar{v})(1 + a_n) + (\bar{x}_t - \bar{z})B_n, \]  

(2.11)

where,
\[ a_n = m_{m_n} e^{1/2} \left[ 1 - \sigma_{m_n}^{-1} \left( \hat{\theta} - \theta \right) \right] - 1 \]  \hspace{1cm} (2.12)

\[ b_n = -s_{m_n}^{-1/2} \left( \hat{\theta} - \theta \right) . \]

We write it this way in terms of \( a_n \) and \( b_n \) for convenience later on. Notice that \( b_n = O_p(n^{-1/2}) \), from the fact that 
\[ \hat{\theta} - \theta = O_p(n^{-1/2}) \] by assumption. Also note that,

\[ s_{m_n}^{-1/2} = \left| \frac{1}{m_{m_n}} \left[ 1 - \sigma_{m_n}^{-1} \left( \hat{\theta} - \theta \right) \right] \right|^{-1} + o_p(n^{-1/2}) , \]  \hspace{1cm} (2.13)

so

\[ a_n = \text{sgn}[1 - \sigma_{m_n}^{-1} \left( \hat{\theta} - \theta \right)] - 1 + o_p(n^{-1/2}) \]

where the \( \text{sgn}(\cdot) \) function is defined by

\[ \text{sgn}(x) = \begin{cases} +1 & \text{if } x > 0 , \\ 0 & \text{if } x = 0 , \\ -1 & \text{if } x < 0 . \end{cases} \]

Now let \( \{c_n\}_{n=1}^{\infty} \) be a sequence of positive real numbers decreasing to zero, and let \( \varepsilon > 0 \) be given. Then

\[ P(c_n^{-1} | \text{sgn}[1 - \sigma_{m_n}^{-1} \left( \hat{\theta} - \theta \right)] - 1| > \varepsilon) < P(\sigma_{m_n}^{-1} \left( \hat{\theta} - \theta \right) > 1} . \]

The right hand side converges to zero since \( \hat{\theta} - \theta = O_p(n^{-1/2}) \). Thus,
If we let \( c_n = n^{-\frac{1}{2}} \) then we get \( a_n = o_p(n^{-\frac{1}{2}}) \). The expression on the left hand side of (2.11) is called the **standardized residual** associated with \( \hat{v}_t \). We state our result as a theorem so we can refer to it later.

**Theorem 2.2.** Let the model assumptions hold, then the standardized residuals can be represented as,

\[
\hat{v}_t \sim \frac{\sqrt{n}}{\sigma_{\hat{v}\hat{v}}} (\hat{v}_t - \bar{v}) = \frac{\sqrt{n}}{\sigma_{\hat{v}\hat{v}}} (\hat{v}_t - \bar{v})(1 + a_n) + (\hat{x}_t - \bar{x}_t)B_n,
\]

for \( t = 1, \ldots, n \), where \( a_n = o_p(n^{-\frac{1}{2}}) \) and \( B_n \) is a random \( k \times 1 \) vector such that \( B_n = O_p(n^{-\frac{1}{2}}) \).

We have concentrated on representations for the standardized residuals. In practice, we are also interested in predicting the true \( X_t \) values. Following Fuller (1986) we form the predictor \( \hat{X}_t \) by replacing the population values in the definition of \( \hat{X}_t \) by sample values. In particular, we define \( \hat{X}_t \) by,

\[
\hat{X}_t = X_t - \bar{v}_t \frac{\sqrt{n}}{\sigma_{\hat{v}\hat{v}}} \hat{v}_t,
\]

or

\[
= X_t - \bar{v}_t \hat{v}_t,
\]

(2.14)
for $t = 1, \ldots, n$, where $\hat{\sigma} = \sigma \varepsilon \sigma^{-1} \varepsilon$ and call (2.14) the estimated true values. We assume that,

\[
\hat{\sigma} = \sigma = O_p(n^{-1/2}) , \tag{2.15}
\]

where $\hat{\sigma} = \sigma \varepsilon \sigma^{-1} \varepsilon$. This is generally equivalent to the assumption that $\hat{\sigma} = \sigma = O_p(n^{-1/2})$. From the definition of $\bar{x}_t$ and expression (2.4) we can write,

\[
\bar{x}_t = x_t - [v_t - x_t(\hat{\sigma} - \sigma)] \hat{\sigma} , \tag{2.16}
\]

\[
= x_t[I_k + (\hat{\sigma} - \sigma) \hat{\sigma}] - v_t \hat{\sigma} ,
\]

\[
= (\bar{x}_t + v_t \hat{\sigma})[I_k + (\hat{\sigma} - \sigma) \hat{\sigma}] - v_t \hat{\sigma} ,
\]

\[
= \bar{x}_t[I_k + (\hat{\sigma} - \sigma) \hat{\sigma}] + v_t[(\hat{\sigma} - \hat{\sigma}) + \hat{\sigma}(\hat{\sigma} - \sigma) \hat{\sigma}] .
\]

We state this as a theorem for later reference.

**Theorem 2.3.** Let the model assumptions hold, then we can write,

\[
(\bar{x}_t - \bar{x}) = (\bar{x}_t - \bar{x})(I_k + C_n) + (v_t - \bar{v})d_n ,
\]

where $C_n$ is a random $k \times k$ matrix such that $C_n = O_p(n^{-1/2})$, and $d_n$ is a random $1 \times k$ vector such that $d_n = O_p(n^{-1/2})$. 

B. Multivariate Regression Models

Recall that the model can be written,

\[ Y_t = X_t \beta + v_t, \quad X_t = x_t + u_t, \]  

(2.17)

for \( t = 1, \ldots, n \), where \( v_t = e_t - u_t \beta \). As was done in the univariate case we can write the model as,

\[ Y_t = X_t \beta + v_t (I_t + \Sigma_{vv}^{-1} \Sigma_{vu} \beta) \]  

for \( t = 1, \ldots, n \),  

(2.18)

\[ X_t = x_t + v_t \Sigma_{vv}^{-1} \Sigma_{vu}, \]

where \( x_t \) is defined by the relation,

\[ x_t = X_t - v_t \Sigma_{vv}^{-1} \Sigma_{vu} \]  

for \( t = 1, \ldots, n \).  

(2.19)

By analogy with the univariate model, it follows that \( v_1, \ldots, v_n \) are iid \( N_r(0, \Sigma_{vv}) \) and jointly independent of the sequence of \( 1 \times k \) vectors \( (x_1, \ldots, x_n) \). Theorem 2.1 still holds.

Let \( \hat{\beta} \) be an estimator of \( \beta \) such that \( \hat{\beta} - \beta = O_p(n^{-1/2}) \). The residual vector \( \hat{v}_t \) for \( t = 1, \ldots, n \) is given by,

\[ \hat{v}_t = Y_t - X_t \hat{\beta}, \]  

(2.20)

\[ = v_t - x_t (\hat{\beta} - \beta), \]
We also define,

\[ S_{vv} = (n - k)^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})'(\hat{v}_t - \bar{v}) , \quad \bar{v} = n^{-1} \sum_{t=1}^{n} \hat{v}_t . \quad (2.21) \]

And write,

\[ m_{vv} = (n - 1)^{-1} \sum_{t=1}^{n} (v_t - \bar{v})'(v_t - \bar{v}) , \quad \bar{v} = n^{-1} \sum_{t=1}^{n} v_t , \quad (2.22) \]

and define \((S_{vv})_{ij} = s_{vij} \), \((m_{vv})_{ij} = m_{vij} \). Also define \( \beta_i \) for \( i = 1, \ldots, r \) to be the \( i \)-th column of \( \beta \), and write \( v_t = (v_{t1}, \ldots, v_{tr}) \) and similarly for \( \hat{v}_t \).

By analogy with the univariate case we seek a representation for the \textbf{standardized residual vector},

\[ (\hat{v}_t - \bar{v})S_{vv}^{-1/2} \quad \text{for} \quad t = 1, \ldots, n . \quad (2.23) \]

First notice that,

\[ S_{vv} = (I_r - E_{vv} E_{vu}(\hat{\beta} - \beta))'m_{vv}(I_r - E_{vv} E_{vu}(\hat{\beta} - \beta)) \quad (2.24) \]

\[ -(\hat{\beta} - \beta)'m_{x\bar{v}}(I_r - E_{vv} E_{vu}(\hat{\beta} - \beta)) \]

\[ - (I_r - E_{vv} E_{vu}(\hat{\beta} - \beta))'m_{\bar{v}x}(\hat{\beta} - \beta) \]
\[ + (\hat{\beta} - \beta) ' \mathbf{m} \mathbf{m} ' (\hat{\beta} - \beta) (n - k)^{-1} (n - 1) . \]

As in the univariate case \( \mathbf{m}^{+} = 0 \) a.s., so we can write
\[
S_{vv} = \{(I_{r} - \Sigma_{vv}^{-1} \Sigma_{vu} (\hat{\beta} - \beta))' \mathbf{m}_{vv} [I_{r} - \Sigma_{vv}^{-1} \Sigma_{vu} (\hat{\beta} - \beta)]\} + o_{p}(n^{-1/2}) .
\]

(2.25)

Now returning to expression (2.20) we can write for \( t = 1, \ldots, n \)
\[
(\hat{\psi}_{t} - \overline{\psi}) S_{vv}^{1/2} = (\psi_{t} - \overline{\psi}) [I_{r} - \Sigma_{vv}^{-1} \Sigma_{vu} (\hat{\beta} - \beta)] S_{vv}^{1/2} - (\overline{\psi}_{t} - \overline{\psi}) (\hat{\beta} - \beta) S_{vv}^{1/2}
\]
\[
= (\psi_{t} - \overline{\psi}) S_{vv}^{1/2} (I_{r} + A_{n}) + (\overline{\psi}_{t} - \overline{\psi}) B_{n} ,
\]

(2.26)

where
\[
A_{n} = \mathbf{m}_{vv}^{1/2} [I_{r} - \Sigma_{vv}^{-1} \Sigma_{vu} (\hat{\beta} - \beta)] S_{vv}^{1/2} - I_{r} ,
\]

(2.27)

\[
B_{n} = - (\hat{\beta} - \beta) S_{vv}^{-1/2} .
\]

We write it in the form in terms of \( A_{n} \) and \( B_{n} \) for convenience later on. Notice that \( B_{n} = 0_{p}(n^{-1/2}) \), from the fact that \( \hat{\beta} - \beta = 0_{p}(n^{-1/2}) \) by assumption. To determine the behavior of \( A_{n} \) we need the following lemma.
Lemma 2.2. Let $Q_n$ be a random $r \times r$ matrix such that

$$Q_n = I_r + o_p(a_n), \quad \text{and} \quad Q_n'Q_n = I_r + o_p(a_n)$$

for a sequence of positive numbers $\{a_n\}_{n=1}^{\infty}$ decreasing to zero. Then $\text{diag}(Q_n - I_r) = o_p(a_n)$.

Proof. Let $Q_n = I_r + A_n$ where $A_n = o_p(a_n)$. Then,

$$Q_n'Q_n = I_r + A_n + A_n' + A_n'A_n,$$

which implies that,

$$A_n + A_n' + A_n'A_n = o_p(a_n),$$

which implies that,

$$A_n + A_n' = o_p(a_n) \quad \text{since} \quad A_n'A_n = o_p(a_n^2).$$

Thus we get,

$$2 \text{diag}(A_n) = \text{diag}(A_n + A_n') = o_p(a_n),$$

so,

$$\text{diag}(Q_n - I_r) = \text{diag}(A_n) = o_p(a_n). \quad \square$$

Returning to $A_n$, we can write $A_n = Q_n - I_r$ where,
From expression (2.25) we know that,

\[ S_{vv} = \sigma_{vv} + o_p(n^{-1/2}) . \]

This implies that,

\[ S_{vv}^{-1/2} = \sigma_{vv}^{-1/2} + o_p(n^{-1/2}) , \]

which implies that,

\[ Q_n = I_r + o_p(n^{-1/2}) . \]

If we premultiply and postmultiply expression (2.25) by \( S_{vv}^{-1/2} \) we get,

\[ Q_n' Q_n = I_r + o_p(n^{-1/2}) . \]

Thus applying Lemma 2.2 gives,

\[ \text{diag}(Q_n - I_r) = o_p(n^{-1/2}) . \]

We now state our results as a theorem for later reference.

**Theorem 2.4.** Let the model assumptions hold, then the standardized residual vectors can be represented as,
for $t = 1, \ldots, n$, where $B_n$ is a random $k \times r$ matrix such that $B_n = o_p(n^{-1/2})$, and $A_n$ is a random $r \times r$ matrix such that $\text{diag}(A_n) = o_p(n^{-1/2})$ with off-diagonal elements which are $o_p(n^{-1/2})$.

Notice that Theorem 2.4 reduces to Theorem 2.2 in the univariate case when $r = 1$. While the representation in Theorem 2.4 will be useful later, it will also be useful to have representations for $s_{v_{ti}}^{-1/2}(v_{ti} - v_i)$ for $t = 1, \ldots, n$, (2.29)

and $i \in \{1, \ldots, r\}$. These are called the standardized components of the residual vector, to contrast them with the components of the standardized residual vector in (2.23). For an $i \in \{1, \ldots, r\}$ we can write,

$$v_{ti} = v_{ti} - x_t \hat{\beta}_i,$$  (2.30)

$$= v_{ti} - x_t (\hat{\beta}_i - \beta_i),$$

$$= v_{ti} [1 - \sigma_{v_{ti}}^{-1} \xi_{v}(i)u(\hat{\beta}_i - \beta_i)] - x_t(i) (\hat{\beta}_i - \beta_i),$$

where $\xi_{v}(i)u = C(v_{ti}, u_t)$ and,

$$x_t(i) = x_t - v_{ti} \sigma_{v_{ti}}^{-1} \xi_{v}(i)u.$$  (2.31)
The random variable $\mathbf{x}_{t(i)}$ is a predictor of $\mathbf{x}_t$ and is formed by subtracting from $\mathbf{x}_t$ the best predictor of $\mathbf{u}_t$ given $\mathbf{v}_{ti}$ under normality. We can write $\mathbf{x}_{t(i)}$ as,

$$\mathbf{x}_{t(i)} = \mathbf{x}_t + \mathbf{v}_{ti}$$

for $t = 1, \ldots, n$, \hspace{1cm} (2.32)

where

$$\mathbf{v}_{ti} = \mathbf{u}_t - \mathbf{v}_t \sigma_{vii}^{-1} \mathbf{v}_t \mathbf{v}(i) \mathbf{u}.$$

Thus $\mathbf{v}_{1(i)}$, $\ldots$, $\mathbf{v}_{n(i)}$ are iid $N_r(0, \Sigma_{\mathbf{v}(i)})$ where

$$\Sigma_{\mathbf{v}(i)} = \Sigma_{\mathbf{u}} - \sigma_{vii}^{-1} \mathbf{v}(i) \mathbf{v}(i) \mathbf{u}.$$

By construction $\mathbf{v}_{ti}$ and $\mathbf{v}_{ti}$ are independent which implies that $\mathbf{v}_{ti}$ and $\mathbf{x}_{t(i)}$ are independent. In the univariate case $\mathbf{x}_{t(i)}$ and $\mathbf{x}_t$, defined in (2.19), reduce to $\mathbf{x}_t$, defined in (2.3). In the multivariate case $\mathbf{x}_{t(i)}$ and $\mathbf{x}_t$ are different. Under the normal model $\mathbf{x}_{t(i)}$ is the best predictor of $\mathbf{x}_t$ given $\mathbf{v}_{ti}$, whereas $\mathbf{x}_t$ is the best predictor of $\mathbf{x}_t$ given the entire vector $\mathbf{v}_t$.

By using (2.30) we can write $s_{vv11}$ as,

$$s_{vv11} = \left[ 1 - \sigma_{vii}^{-1} \mathbf{v}(i) \mathbf{v}(i) \mathbf{u} (\hat{\beta}_1 - \beta_1) \right]^2 s_{vv11}$$

$$- 2 \left[ 1 - \sigma_{vii}^{-1} \mathbf{v}(i) \mathbf{v}(i) \mathbf{u} (\hat{\beta}_1 - \beta_1) \right] s_{vv11} (\hat{\beta}_1 - \beta_1)$$

\hspace{1cm} (2.33)
\[ + (\hat{\beta}_1 - \beta_1)' \bar{m}_{xx}(i)(\hat{\beta}_1 - \beta_1) (n-k)^{-1} (n-1), \]

where,

\[ \bar{m}_{xx}(i) = (n-1)^{-1} \sum_{t=1}^{n} (x_t - \bar{x})'[x_t(i) - \bar{x}(i)], \quad (2.34) \]

\[ \bar{m}_{xx}(i) = (n-1)^{-1} \sum_{t=1}^{n} [x_t(i) - \bar{x}(i)]'[x_t(i) - \bar{x}(i)]. \]

Under our assumptions \( \bar{m}_{xx}(i) \rightarrow 0 \) a.s. which follows by the same proof used to show \( \bar{m}_{xx}(i) \rightarrow 0 \) a.s. in Theorem 2.1. Thus we can write,

\[ s_{vvii} = \bar{m}_{vvii}[1 - \bar{\sigma}_{vvii}^2]u(\hat{\beta}_1 - \beta_1)]^2 + o_p(n^{-1/2}). \quad (2.35) \]

We have used the fact that \( \bar{m}_{xx}(i) + \bar{m}_{xx}(i) \) a.s., where \( \bar{m}_{xx}(i) \) is a finite matrix, which follows by the same proof used in Theorem 2.1. Now returning to (2.30) we can write for \( t = 1, ..., n \)

\[ s_{vvii}^{-1/2}(v_t - \bar{v}) = (v_t - \bar{v})[1 - \bar{\sigma}_{vvii}^2]u(\hat{\beta}_1 - \beta_1)]s_{vvii}^{-1/2} \]

\[ - [x_t(i) - \bar{x}(i)](\hat{\beta}_1 - \beta_1)s_{vvii}^{-1/2}, \]

\[ = s_{vvii}^{-1/2}(v_t - \bar{v})(1 + a_n) + [x_t(i) - \bar{x}(i)]B_n, \]

where,
By analogy with the univariate result, we state this as a theorem for later reference.

**Theorem 2.5.** Let the model assumption hold, then we can write,

\[
a_{ni} = s_{vii}^{-1/2} \left[ 1 - c_{vii}^{-1} \hat{v}(1) u \left( \hat{b}_1 - b_1 \right) \right] - 1,
\]

\[
B_{ni} = - s_{vii}^{-1/2} (\hat{b}_1 - b_1).
\]

for \( t = 1, \ldots, n \), where \( a_{ni} = o_p(n^{-1/2}) \) and \( B_{ni} \) is a random vector such that \( B_{ni} = o_p(n^{-1/2}) \).

The order in probability result for \( B_{ni} \) follows because \( \hat{b}_1 - b_1 = o_p(n^{-1/2}) \) by assumption, and \( a_{ni} = o_p(n^{-1/2}) \) follows by the same proof used to show \( a_n = o_p(n^{-1/2}) \) is Theorem 2.2.

We have concentrated on representations for the standardized residual vectors. In practice we are also interested in predicting the true \( X_t \) values. As was done in the univariate case, we define the predictor \( \hat{X}_t \) by,

\[
\hat{X}_t = X_t - \hat{v}_{t} \hat{\delta},
\]

\[
= X_t - \hat{v}_{t} \delta,
\]
where $\hat{\delta} = \frac{1}{\sum_{vu}^2}$. We assume that,

$$\hat{\delta} - \delta = O_p(n^{-1/2}) ,$$

where $\delta = \frac{1}{\sum_{vu}^2}$. This is generally equivalent to the assumption that $\hat{\delta} - \delta = O_p(n^{-1/2})$. From the definition of $\hat{x}_t$ and expression (2.20) we can write,

$$\hat{x}_t = x_t - [v_t - x_t(\hat{\delta} - \delta)]\hat{\delta} ,$$

$$= x_t[I_k + (\hat{\delta} - \delta)\hat{\delta}] - v_t\hat{\delta} ,$$

$$= (\hat{x}_t + v_t\delta)[I_k + (\hat{\delta} - \delta)\hat{\delta}] - v_t\hat{\delta} ,$$

$$= \hat{x}[I_k + (\hat{\delta} - \delta)\hat{\delta}] + v_t[(\hat{\delta} - \delta) + \delta(\hat{\delta} - \delta)] .$$

We state this as a theorem for later reference.

**Theorem 2.6.** Let the model assumptions hold, then we can write,

$$(\hat{x}_t - \bar{x}) = (\hat{x}_t - \bar{x})(I_k + C_n) + (v_t - \bar{v})D_n ,$$

where $C_n$ is a random $k \times k$ matrix such that $C_n = O_p(n^{-1/2})$, and $D_n$ is a random $r \times k$ matrix such that $D_n = O_p(n^{-1/2})$. 


III. WEAK CONVERGENCE OF SOME MODIFIED EMPIRICAL PROCESSES

In the previous chapter we derived representations for standardized residuals from regression models. In this chapter we prove the weak convergence of empirical processes based on a class of random variables which include the representations given above as a special case. The results we give are presented in a very general framework, so the notation used in this chapter differs from the notation used elsewhere in this thesis. One thing that remains the same is that $Z$ and $\mathbb{m}_{ZZ}$ will always denote the sample mean vector and covariance matrix from the sample of random vectors $Z_1, \ldots, Z_n$.

A. Univariate Results

The results of this section generalize the following classic result from the theory of weak convergence of empirical processes.

**Theorem 3.1.** Let $Y_1, Y_2, \ldots, Y_n$ be iid $N(\mu_Y, \sigma_{YY})$ with $\sigma_{YY} > 0$. Define for $\omega \in [0, 1]$,

$$W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\Phi(Z_t) < \omega] - \omega\},$$

and,

$$Z_t = \mathbb{m}_{YY}^{-1/2}(Y_t - \bar{Y}) \text{ for } t = 1, \ldots, n.$$
Then \( W_n \xrightarrow{L} W \) as a stochastic process on \([0, 1]\), where \( W \) is a Gaussian process with mean and covariance function given by,

\[
E(W(\omega)) = 0 \quad \text{for } \omega \in [0, 1]
\]

and

\[
E(W(\omega_1)W(\omega_2)) = \min(\omega_1, \omega_2) - \omega_1\omega_2
\]

\[
- [1 + \frac{1}{2} \phi^{-1}(\omega_1)\phi^{-1}(\omega_2)]\phi[\phi^{-1}(\omega_1)]\phi[\phi^{-1}(\omega_2)]
\]

for \( \omega_1 \) and \( \omega_2 \in [0, 1] \).

**Proof.** See Kac, Keifer and Wolfowitz (1955).

Theorem 3.1 is useful if the random variables \( Y_1, \ldots, Y_n \) are directly observable. In the case of regression residuals the theorem does not apply since the \( v_t \)'s are not directly observable. Furthermore, the observable \( \hat{v}_t \)'s are not iid normal random variables. We next present our main theorem which generalizes Theorem 3.1 to include empirical processes based on certain non-iid sequences of random variables. The class of non-iid sequences for which the theorem holds contains the class of standardized regression residuals presented in Chapter Two by virtue of the representation theorems given there.

Following Theorem 3.2 we present a series of lemmas used in its proof. Our method of proof is closely related to the techniques of Rao and Sethuraman (1975), and Loynes (1980).
Theorem 3.2. Let \( Y_1, Y_2, \ldots, Y_n \) be iid \( N(\mu_Y, \sigma_{YY}) \) with \( \sigma_{YY} > 0 \), and independent of the random \( k \)-vectors \( (X_1, X_2, \ldots, X_n) \) where \( \mu_{XX} + \bar{m}_{XX} \) a.s. Let \( B_n \) be a random \( 1 \times k \) vector such that \( B_n = o_p(n^{-1/2}) \), and let \( a_n \) be a random variable such that \( a_n = o_p(n^{-1/2}) \). Define

\[
Z_t = (1 + a_n)\mu_{YY}^{-1/2} (Y_t - \bar{Y}) + \frac{B_n}{n} (X_t - \bar{X}) ,
\]

\[
Z^*_t = \mu_{YY}^{-1/2} (Y_t - \bar{Y}) \text{ for } t = 1, 2, \ldots, n,
\]

and,

\[
W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\phi(Z_t) < \omega] - \omega\} ,
\]

\[
W^*_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\phi(Z^*_t) < \omega] - \omega\} \text{ for } \omega \in [0, 1] .
\]

Then \( \sup_{\omega \in [0, 1]} |W^*_n(\omega) - W_n(\omega)| = o_p(1) \).

**Proof.** First note that we can write for \( \omega \in [0, 1] \).

\[
W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{1[\mu_{YY}^{-1/2} (Y_t - \bar{Y}) < (1 + c_n)\phi^{-1}(\omega) + D_n(X_t - \bar{X})] - \omega\} + o_p(1) ,
\]

where \( c_n = -(1 + a_n)^{-1}a_n = o_p(n^{-1/2}) \) and,
\( D_n = -(1 + a_n)^{-1} E_n = o_p(n^{-1/2}) \).

The remainder is \( o_p(1) \) uniformly over \( \omega \in [0, 1] \), because \( P(1 + a_n) < 0 \) → 0 as \( n \to \infty \). After some manipulations we can write,

\[
W_n(\omega) = W_0(\omega) + W_{1,n}(\omega) + W_{2,n}(\omega) - W_{3,n}(\omega) + o_p(1), \tag{3.2}
\]
where, again, the remainder is \( o_p(1) \) uniformly over \( \omega \in [0, 1] \), and the other terms are given by

\[
W_{1,n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{ \Phi[z_{tn}(\omega)] + y_n(\omega)] - \Phi[y_n(\omega)] \}, \tag{3.3}
\]
\[
W_{2,n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{ 1[U_t < \Phi[z_{tn}(\omega)]] - 1[U_t < \omega] - \Phi[z_{tn}(\omega)] + \omega \},
\]
\[
W_{3,n}(\omega) = n^{-1/2} \sum_{t=1}^{n} \{ 1[U_t < \Phi[y_n(\omega)]] - 1[U_t < \omega] - \Phi[y_n(\omega)] + \omega \},
\]

where,

\[
y_n(\omega) = \mu_Y - \mu_Y + \left( \sigma_{XY} \sigma_{XX} \right)^{1/2} \phi^{-1}(\omega), \tag{3.4}
\]
\[
z_{tn}(\omega) = y_n(\omega) + \left[ \phi^{-1}(\omega)c_n + D_n(X_t - \bar{X}) \right] \left( \sigma_{XY} \sigma_{XX} \right)^{1/2},
\]
\[
U_t = \phi[\alpha_{XY}(X_t - \mu_X)].
\]
By Lemma 3.3 we know that

$$\sup_{\omega \in [0,1]} |W_{1,n}(\omega)| = o_p(1).$$

By Lemma 3.4 we know that

$$\sup_{\omega \in [0,1]} |W_{2,n}(\omega)| = o_p(1),$$

and also

$$\sup_{\omega \in [0,1]} |W_{3,n}(\omega)| = o_p(1)$$

by taking $C_n = 0$ in the notation of Lemma 3.4. Thus,

$$\sup_{\omega \in [0,1]} |W_n(\omega) - W_n^*(\omega)| = o_p(1).$$

**Lemma 3.1.** Let $a, b \in \mathbb{R}$ with $b \neq 0$, and let $p$ be nonnegative integer. Then,

$$\sup_{\omega \in [0,1]} \left| [\phi^{-1}(\omega)]^p \phi[a + b\phi^{-1}(\omega)] \right| < |b|^{-p} \left| |a| + 4p \right|^p,$$

where the right hand side is defined to be one when $p = 0$.

**Proof.** By definition we know,

$$x^p \phi(a + bx) = (2\pi)^{-1/2} x^p \exp[-\frac{1}{2} (a + bx)^2].$$
Therefore,

\[ \frac{d}{dx}[x^p \phi(a + bx)] = px^{p-1} \phi(a + bx) + bx^p \phi'(a + bx). \]

But \( \phi'(x) = -x \phi(x) \). Thus the derivative above is equal to,

\[ x^{p-1}[p - bx(a + bx)] \phi(a + bx). \]

Thus the critical points are derived from solving the equation,

\[ x^{p-1}[p - bx(a + bx)] = 0. \]

The nonzero solutions to this are given by,

\[ x = (2b^2)^{-1}[-ab \pm (a^2b^2 + 4pb^2)^{1/2}]. \]

Thus we have,

\[ |x^p \phi(a + bx)| < (2a)^{-1/2} |2b|^{-p}[|a| + (a^2 + 4p)^{1/2}]^p, \]

\[ < |2b|^{-p}[2(a^2 + 4p)^{1/2}]^p, \]

\[ < |b|^{-p}[a^2 + (4p)^2]^{p/2}, \]

\[ < |b|^{-p}[|a| + 4p]^p, \]
where the last step follows from Minkowski's inequality.

Lemma 3.2. Let $X_1, X_2, \ldots, X_n$ be a sequence of random $k$-vectors such that $m_{XX} \xrightarrow{p} m_{XX}$, and let $B_n$ be a random $1 \times k$ vector such that $B_n = O_p(n^{-1/2})$. Then,

$$n^{-1} \sum_{t=1}^{n} |B_n(X_t - \bar{X})|^{1+\delta} = O_p(n^{-1/2}) \quad \text{for} \quad \delta \in [0, 1)$$

$$= O_p(n^{-1}) \quad \text{for} \quad \delta > 1.$$

Proof. For $\delta \in [0, 1)$ we can write,

$$n^{-1} \sum_{t=1}^{n} |B_n(X_t - \bar{X})|^{1+\delta} \leq \max_{1 \leq t \leq n} |B_n(X_t - \bar{X})|^{1+\delta} n^{-1} \sum_{t=1}^{n} |B_n(X_t - \bar{X})|$$

$$\leq [(nB_nB_n^t)^{\delta/2}E_{n}m_{XX} n^{1/2}] = O_p(n^{-1/2}),$$

where we repeatedly used Hölder's inequality. For $\delta > 1$ we can write,

$$n^{-1} \sum_{t=1}^{n} |B_n(X_t - \bar{X})|^{1+\delta} \leq \max_{1 \leq t \leq n} |B_n(X_t - \bar{X})|^{1+\delta} n^{-1} \sum_{t=1}^{n} |B_n(X_t - \bar{X})|^2$$

$$\leq [(nB_nB_n^t)^{(\delta-1)/2}E_{n}m_{XX} n^{-1/2}]$$

$$= O_p(n^{-1}) \quad \square$$

Lemma 3.3. Let $(X_1, X_2, \ldots, X_n)$ be a sequence of random $k$-vectors such that $m_{XX} \xrightarrow{p} m_{XX}$. Let $a_n, b_n, c_n$ be random
variables such that \( a_n = o_p(1) \), \( b_n = o_p(1) \), and \( c_n = o_p(n^{-1/2}) \), and let \( D_n \) be a random \( 1 \times k \) vector such that \( D_n = o_p(n^{-1/2}) \).

Define for \( \omega \in [0, 1] \),

\[
y_n(\omega) = a_n + (1 + b_n)\phi^{-1}(\omega),
\]

\[
z_{tn}(\omega) = \phi^{-1}(\omega)c_n + D_n(X_t - \bar{X}),
\]

and,

\[
W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{\phi[z_{tn}(\omega) + y_n(\omega)] - \phi[y_n(\omega)]\}.
\]

Then \( \sup_{\omega \in [0,1]} |W_n(\omega)| = o_p(1). \)

**Proof.** We can write by Taylor's theorem,

\[
W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \{\phi[y_n(\omega)]z_{tn}(\omega) + \phi'(y^*)z_{tn}(\omega)\}, \tag{3.5}
\]

where \( y^* \) is on the line segment joining \( y_n(\omega) \) and \( y_n(\omega) + z_{tn}(\omega) \).

Thus we have,

\[
W_n(\omega) = n^{1/2} c_n \phi[y_n(\omega)]\phi^{-1}(\omega) + n^{-1/2} \sum_{t=1}^{n} \phi'(y^*)z_{tn}^2(\omega). \tag{3.6}
\]

Now by expanding \( z_{tn}^2(\omega) \) and using Lemma 3.1 we can write,
\[
\sup_{\omega \in [0,1]} |W_n(\omega)| < n^{1/2} |c_n||1 + b_n|^{-1}(|a_n| + 4) + n^{1/2} 4D_n \max_{\omega \in [0,1]} D_n' \tag{3.7}
\]

\[
+ n^{-1/2} c_n^2 \sum_{t=1}^{n} \sup_{\omega \in [0,1]} |\phi(\gamma_t^*)[\phi^{-1}(\omega)]^2| 
\]

\[
+ 2 n^{-1/2} |c_n| \sum_{t=1}^{n} |D_n(\bar{X}_t - \bar{X}) \sup_{\omega \in [0,1]} |\phi(\gamma_t^*)\phi^{-1}(\omega)| + o_p(1) ,
\]

where the last \( o_p(1) \) term comes from the fact that,

\[
P[|1 + b_n| = 0] = 0 \text{ as } n \to \infty.
\]

We also used the fact that \( \phi'(x) = -x\phi(x) \), so \( |\phi'(x)| < 4 \) follows from Lemma 3.1. The first two terms of (3.7) are \( o_p(1) \). To deal with the third and fourth terms of (3.7) we need to write \( \gamma_t^* \) as,

\[
\gamma_t^* = y_n(\omega) + \lambda_{tn}(\omega)z_{tn}(\omega) , \tag{3.8}
\]

where \( \lambda_{tn}(\omega) \) is a random variable between zero and one. We can write this as

\[
\gamma_t^* = e_{tn} + f_{tn} \phi^{-1}(\omega) , \tag{3.9}
\]

where,

\[
e_{tn} = a_n + \lambda_{tn}(\omega)D_n(\bar{X}_t - \bar{X}) ,
\]
Now we can write,

\[ |\phi'(\gamma_T)\phi^{-1}(\omega)| = |\phi^{-1}(\omega)\gamma_T \phi(\gamma_T)|, \tag{3.10} \]

\[ < |e_{tn}| |\phi(\gamma_T)\phi^{-1}(\omega)| + |e_{tn}| |\phi(\gamma_T)[\phi^{-1}(\omega)]^2|, \]

and,

\[ |\phi'(\gamma_T)[\phi^{-1}(\omega)]^2| = |[\phi^{-1}(\omega)]^2 \gamma_T \phi(\gamma_T)|, \tag{3.11} \]

\[ < |e_{tn}| [\phi^{-1}(\omega)]^2 |\phi(\gamma_T)| + |e_{tn}| [\phi^{-1}(\omega)]^3 |\phi(\gamma_T)|. \]

Now note that,

\[ |f_{tn}| > 1 - |b_n + \lambda_{tn}(\omega)c_n| > 1 - |b_n| - |c_n|. \]

Thus given \( \varepsilon > 0 \) we have,

\[ P(\inf |f_{tn}|: \omega \in [0, 1], 1 < t < n) < 1 - \varepsilon \to 0 \text{ as } n \to \infty, \]

or we can write this as,

\[ P(\sup |f_{tn}|^{-1}: \omega \in [0, 1], 1 < t < n) > (1 - \varepsilon)^{-1} \to 0 \text{ as } n \to \infty. \]
Thus, if we let $\varepsilon$ in $(0, 1)$ be given and use Lemma 3.1 repeatedly we can write (3.7) as,

$$
\sup_{\omega \in [0,1]} |w''(u)| < n^{-1/2} c_2(1 - \varepsilon)^{-2} \sum_{t=1}^{n} \{2 g_{tn}^3 + 52 g_{tn}^2 + 396 g_{tn} + 1728\}
+ 2 |c_n| (1 - \varepsilon)^{-1/2} \sum_{t=1}^{n} |D_n(x_t - \bar{x})| \{2 g_{tn}^2 + 20 g_{tn} + 64\}
+ o_p(1). \tag{3.12}
$$

where $g_{tn} = |a_n| + |D_n(x_t - \bar{x})|$, so $|e_{tn}| < g_{tn}$ for all $\omega \in [0, 1]$.

The lemma now follows from Lemma 3.2. □

**Lemma 3.4.** Let $U_1, U_2, \ldots, U_n$ be iid Uniform $(0, 1)$ random variables independent of the random k-vectors $(X_1, X_2, \ldots, X_n)$, where $\mathbb{E}X = \mathbb{E}X$ a.s. Let $a_n$ and $b_n$ be random variables such that

$a_n = o_p(n^{-1/2})$ and $b_n = o_p(n^{-1/2})$, and let $C_n$ be a random $1 \times k$ vector such that $C_n = O_p(n^{-1/2})$. Define for $\omega \in [0, 1]$.

$$
x_{tn}(\omega) = a_n + (1 + b_n) \phi^{-1}(\omega) + C_n(x_t - \bar{x}),
$$

and,

$$
W_n(\omega) = n^{-1/2} \sum_{t=1}^{n} \left[ 1\{U_t < \phi[z_{tn}(\omega)]\} - 1\{U_t < \omega\} - \phi[z_{tn}(\omega)] + \omega \right].
$$

Then $\sup_{\omega \in [0,1]} |W_n(\omega)| = o_p(1)$. 

Proof. We can define

$$z_{tn}(w; \xi) = \xi_1 + (1 + \xi_2)\phi^{-1}(w) + \xi_3(x_t - \bar{x}),$$

(3.13)

where \( \xi = (\xi_1, \xi_2, \xi_3) \). Thus, \( z_{tn}(w) = z_{tn}(w; a_n, b_n, C_n) \). We also define \( W_n(w; \xi) \) in terms of \( z_{tn}(w; \xi) \) above. Let \( \varepsilon > 0 \) be given. We can find an \( L_{\varepsilon} \in (0, \infty) \) such that,

$$\left| z_{tn}(w; \xi) \right| > 1 - \varepsilon,$$

where \( C_{n, \varepsilon} \) is the hypercube in \( \mathbb{R}^{k+2} \) centered at \( 0 \), with sides of length \( 2L_{\varepsilon}n^{-1/2} \). Thus we can write,

$$P\{\sup_{w \in [0,1]} |W_n(w)| > \varepsilon\},$$

(3.14)

$$< E\{P\{\sup_{w \in [0,1]} \sup_{\xi \in C_{n, \varepsilon}} |W_n(w; \xi)| > \varepsilon | (X_1, X_2, \ldots, X_n)\} \} + \varepsilon.$$  

Thus, if we can show that the conditional probability within the expectation operator converges to zero for almost every sequence \( \{X_1, X_2, \ldots, X_n\}_{n=1}^{\infty} \), then by the Bounded Convergence Theorem we will have shown that,

$$\lim_{n \to \infty} P\{\sup_{w \in [0,1]} |W_n(w)| > \varepsilon\} < \varepsilon.$$

But since \( \varepsilon > 0 \) was arbitrary, we will have shown that,
\[ \sup_{\omega \in [0,1]} \left| W_n(\omega) \right| = \alpha_p(1). \]

But further notice that,

\[ P \left( \sup_{\omega \in [0,1]} \sup_{\xi \in C_n, \epsilon} |W_n(\omega)| > \epsilon \left| (X_1, X_2, \ldots, X_n) \right| \right) \]

\[ < \sup_{\omega \in [0,1]} \sup_{\xi \in C_n, \epsilon} W_n(\omega) > \epsilon \left| (X_1, X_2, \ldots, X_n) \right| \]

\[ \leq \sup_{\omega \in [0,1]} \sup_{\xi \in C_n, \epsilon} -W_n(\omega) > \epsilon \left| (X_1, X_2, \ldots, X_n) \right| . \]

The steps that follow show that the first term on the right hand side converges to zero. It will be obvious how to modify the procedure to show that the second term also converges to zero.

Let \( \epsilon > 0 \) be given, and define \( L_{\epsilon} \) and \( C_n, \epsilon \) as indicated above. Now subdivide the hypercube into (approximately) \( (2L_{\epsilon}^{-1})^{k+2} \) hypercubes with sides of length \( \epsilon_1 n^{-1/2} \), where \( \epsilon_1 \) is a positive real number to be chosen later. Let the \( k \)-th such cube be \( C_{k, \epsilon} \). Let \( \xi_{kn}^1(\omega), \xi_{kn}^2(\omega) \) be the values of \( \xi \in C_{k, \epsilon} \) at which \( z_{kn}(\omega; \xi) \) takes its maximum and minimum values, respectively. Note that the first two components of \( \xi_{kn}^1(\omega) \) and \( \xi_{kn}^2(\omega) \) do not depend on \( t \). Then,

\[ z_{kn}(\omega; \xi_{kn}^1(\omega)) - z_{kn}(\omega; \xi_{kn}^2(\omega)) < q_{kn}(\omega, \epsilon_1), \]  

where,
\[ q_{tn}(\omega, \varepsilon_1) = \sup \{ \phi[z_{tn}(\omega; \xi)] - \phi[z_{tn}(\omega; \xi)] : \| \xi - \xi \| < \varepsilon_1 n^{-1/2}, \xi \in c_n, \varepsilon \leq c_n, \varepsilon \} \].

For convenience, we define \( \| \xi - \xi \| < \varepsilon_1 n^{-1/2} \) to mean that each component of the vector \( (\xi - \xi) \) in absolute value satisfies the inequality. Thus we can write,

\[ W_n(\omega) < n^{-1/2} \sum_{t=1}^{n} [1\{U_t < \phi[z_{t,tn}(\omega)]\} - 1\{U_t < \omega\} + n^{-1/2} \sum_{t=1}^{n} q_{tn}(\omega, \varepsilon_1) , \]  

where \( z_{t,tn}(\omega) = z_{tn}[\omega, \xi_{t,tn}] \).

Now we need the following result.

**Result 3.1.** \( n^{-1/2} \sum_{t=1}^{n} q_{tn}(\omega, \varepsilon_1) \) can be made arbitrarily small by choice of \( \varepsilon_1 > 0 \) (almost surely), uniformly in \( n, \omega, \) and \( t \).

**Proof.** First note that,

\[ \frac{\partial}{\partial \xi_1} z_{tn}(\omega; \xi) = 1 , \]

\[ \frac{\partial}{\partial \xi_2} z_{tn}(\omega; \xi) = \phi^{-1}(\omega) , \]

\[ \frac{\partial}{\partial \xi_3} z_{tn}(\omega; \xi) = (x_t - \bar{x})' . \]
Thus we can write,

\[ |q^*_{tn}(\omega, \varepsilon_1)| \leq n^{-\frac{1}{2}} \varepsilon_1 \left( 1 + |\phi^{-1}(\omega)| + \frac{1}{2} |x_t - \bar{x}| \right) \]

\[ \times \left| \phi[\varphi_{t1}^{*} + (1 + \xi_{2t}^{*})\phi^{-1}(\omega) + \varphi_{3t}^{*} (x_t - \bar{x})] \right|, \]

where \( \mathbf{1} \) is a \( k \)-vector of ones, and \( \varphi_{t}^{*} = (\varphi_{1t}^{*}, \varphi_{2t}^{*}, \varphi_{3t}^{*}) \) maximizes the density above evaluated at \( \omega \). Now by using Lemma 3.1, we can write

\[ |q^*_{tn}(\omega, \varepsilon_1)| \leq n^{-\frac{1}{2}} \varepsilon_1 \left( 1 + \mathbf{1}^T |x_t - \bar{x}| \right) \]

\[ + \left| 1 + \xi_{2t}^{*} \right|^{-1} \left( |\varphi_{t1}^{*}| + |\varphi_{3t}^{*} (x_t - \bar{x})| + 4 \right), \]

which is guaranteed to be well defined for \( n > L^2 \). We can further write,

\[ |q^*_{tn}(\omega, \varepsilon_1)| \leq n^{-\frac{1}{2}} \varepsilon_1 \left( 1 + \mathbf{1}^T |x_t - \bar{x}| \right) \]

\[ + (1 - n^{-\frac{1}{2}} L_{\varepsilon})^{-1} \left( 4 + n^{-\frac{1}{2}} L_{\varepsilon} + k \frac{1}{2} L_{\varepsilon} [\text{tr}(m_{XX})]^{1/2} \right). \]

Thus we have,

\[ \left| \sum_{t=1}^{n} q^*_{tn}(\omega, \varepsilon_1) \right| \leq \varepsilon_1 \left[ 1 + (\frac{1}{2} L_{\varepsilon} m_{XX})^{1/2} \right] \]

\[ + (1 - n^{-\frac{1}{2}} L_{\varepsilon})^{-1} \left( 4 + n^{-\frac{1}{2}} L_{\varepsilon} + k \frac{1}{2} L_{\varepsilon} [\text{tr}(m_{XX})]^{1/2} \right). \]
Thus the result follows by the almost sure convergence of $m_{XX}$.

Now divide $[0, 1]$ into (approximately) \( n^{1/2} \lambda^{-1} \) intervals of length \( \lambda n^{-1/2} \) by points \( \omega_s \), where \( \omega_s < \omega_{s+1} \) and write

\[ G_s^\varepsilon = \{ \omega: \omega_s < \omega < \omega_{s+1} \} \].

Now note that if we restrict attention to \( n > L_\varepsilon \) then \( \Phi[z_t, n(\omega; \varepsilon)] \), for \( \varepsilon \in C_n, \varepsilon \), is increasing in \( \omega \) since \( 1 + \varepsilon_2 > 0 \). Therefore the first term of the right hand side of (3.17) can be written, if \( \omega \in G_s^\varepsilon \)

\[
-\frac{1}{2} n \sum_{t=1}^{n} \left[ I\{U_t < \Phi[z_t, n(\omega)]\} - I\{U_t < \omega\} - \Phi[z_t, n(\omega)] + \omega\right] ,
\]

\[
\leq -\frac{1}{2} n \sum_{t=1}^{n} \left[ I\{U_t < \Phi[z_t, n(\omega_{s+1})]\} - I\{U_t < \omega_s\} + \Phi[z_t, n(\omega_{s+1})] + \omega_s\right]
\]

\[
+ n -\frac{1}{2} n \sum_{t=1}^{n} \left( \omega_{s+1} - \omega_s + \Phi[z_t, n(\omega_{s+1})] - \Phi[z_t, n(\omega_s)]\right) ,
\]

Now we can write the second term as equal to

\[
\lambda + n -\frac{1}{2} n \sum_{t=1}^{n} \left( \Phi[z_t, n(\omega_{s+1}; \xi_{t, n}(\omega_{s+1}))] - \Phi[z_t, n(\omega_{s+1}; \xi_{t, n}(\omega_s))]\right)
\]

\[
+ n -\frac{1}{2} n \sum_{t=1}^{n} \left( \Phi[z_t, n(\omega_{s+1}; \xi_{t, n}(\omega_s))] - \Phi[z_t, n(\omega_s; \xi_{t, n}(\omega_s))]\right) ,
\]

\[
< \lambda + n -\frac{1}{2} n \sum_{t=1}^{n} \Phi[q^*_t(\omega, \varepsilon_1)] + n -\frac{1}{2} n \sum_{t=1}^{n} h_n(\lambda, \omega),
\]

where,
Now we need the following result.

**Result 3.2.** \( n^{-1/2} \sum_{t=1}^{n} h_{tn}(\lambda, \omega) \) can be made arbitrarily small by choice of \( \lambda > 0 \) (almost surely), uniformly in \( n, \omega, \) and \( \xi \).

**Proof.** For \( \xi \in C_{n,e}, \omega \in [0,1], \) and \( \lambda > 0 \) small, \( n > L_{e}^{2} \), we can write,

\[
\phi[\xi_{1} + (1 + \xi_{2})\phi^{-1}(\omega + \lambda n^{-1/2}) + \xi_{3}(\underline{x} - \overline{x}) + \phi[\xi_{1} + (1 + \xi_{2})\phi^{-1}(\omega) + \xi_{3}(\underline{x} - \overline{x})],
\]

\[
< \phi[L_{\epsilon} n^{-1/2} \left[ 1 + \lambda_{\epsilon} |\underline{x} - \overline{x}| \right] + (1 - n^{-1/2} L_{\epsilon}) \phi^{-1}(\lambda n^{-1/2})],
\]

which equals,

\[
1 - \phi[- L_{\epsilon} n^{-1/2} \left[ 1 + \lambda_{\epsilon} |\underline{x} - \overline{x}| \right] + (1 - n^{-1/2} L_{\epsilon}) \phi^{-1}(1 - \lambda n^{-1/2})].
\]

This just points out the fact that the greatest change occurs at the endpoints. Now we can write,

\[
\phi[L_{\epsilon} n^{-1/2} \left[ 1 + \lambda_{\epsilon} |\underline{x} - \overline{x}| \right] + (1 - L_{\epsilon} n^{-1/2}) \phi^{-1}(\lambda n^{-1/2})].
\]
where $\phi$ is on the line segment joining,

$$\phi^{-1}(n^{-1/2}) \text{ and } \phi^{-1}(\lambda n^{-1/2}) + n^{-1/2} L \epsilon [1 - \phi^{-1}(\lambda n^{-1/2}) + \frac{1}{\epsilon} |X_t - \bar{x}|] .$$

We can represent $\phi$ as,

$$\phi = \phi^{-1}(n^{-1/2}) + \alpha [n^{-1/2} L \epsilon [1 + \frac{1}{\epsilon} |X_t - \bar{x}| - \phi^{-1}(\lambda n^{-1/2})] ,$$

where $\alpha \in (0, 1)$. But note that for $t = 1, 2, \ldots, n$,

$$n^{-1/2} \frac{1}{\epsilon} |X_t - \bar{x}| < k \frac{1}{\epsilon} [\text{tr}(\omegaXX)]^{1/2} .$$

Thus by the almost sure convergence of $\omegaXX$ we can choose $\lambda$ small enough so that for all $n$ and $t$, $\phi(\phi^*) < \phi(\phi^*)$ where,

$$\phi^* = (1 - n^{-1/2} L \epsilon) \phi^{-1}(\lambda n^{-1/2}) + n^{-1/2} L \epsilon + L \epsilon \frac{1}{\epsilon} [\text{tr}(\omegaXX)]^{1/2} .$$

Thus we can write,

$$\sup_{\omega \in [0,1]} \left| n^{-1/2} \sum_{t=1}^{n} h_t(\lambda, \omega) \right| < \lambda + L \epsilon \phi(\phi^*)[1 - \phi^{-1}(\lambda n^{-1/2})] .$$
\[ n^{-1} \sum_{t=1}^{n} \frac{1}{2} |x_t - \bar{x}|, \]

\[ < \lambda + L_{\epsilon} \phi(\theta^*) \left[ 1 + \left( \frac{1}{\lambda_{\infty}} \right)^{1/2} - \phi^{-1}(\lambda - \frac{1}{2}) \right]. \]

Now the result follows from the almost sure convergence of \( \lambda_{\infty} \) and the use of Lemma 3.1.

Thus the only term we are left with from (3.7) is,

\[ n^{-1/2} \sum_{t=1}^{n} \left[ 1 \{ U_t < \phi[z_{1n}(\omega_{s+1})] \} - 1 \{ U_t < \omega_s \} - \phi[z_{1n}(\omega_{s+1})] + \omega_s \right], \]

since \( n^{-1/2} \sum_{t=1}^{n} q_{tns}^{*}(\omega_{s+1}, \epsilon_1) \) can be made arbitrarily small by choice of \( \epsilon_1 \) as was shown in Result 3.1. We can write the above expression as,

\[ n^{-1/2} \sum_{t=1}^{n} (B_{tns} - p_{tns}) \text{sgn}(tn+s) = \sum_{t=1}^{n} W_{tns}, \]

where \( p_{tns} = |\phi[z_{1n}(\omega_{s+1})] - \omega_s| \), \( B_{tns} \) are Bernoulli random variables with expectation \( p_{tns} \), independent over \( t \), and \( \text{sgn}(tn+s) = +1 \) or \(-1\) according to the sign of \( \phi[z_{1n}(\omega_{s+1})] - \omega_s \).

We will be done if we can show that,

\[ p(\max_{t, s} W_{tns} > \epsilon) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty, \]

where the maximum is taken over

\[ 1 \leq s \leq \left\lfloor \frac{\left( 2L_{\epsilon_{\infty}}^{-1} \right)^{1/2}}{1} + 1 \right\rfloor + 2, \quad \text{and} \quad 1 \leq s \leq \left\lfloor \frac{1}{L_{\epsilon_{\infty}}} \right\rfloor + 1; \]
where \( \lfloor \cdot \rfloor \) stands for the least integer function. Now we can write,

\[
\sum_{t=1}^{n} W_{tns} = n^{-1/2} \left( \sum_{t \in T_+} (B_{tns} - p_{tns}) - \sum_{t \in T_-} (B_{tns} - p_{tns}) \right),
\]

where, \( T_+ = \{t: \text{sgn}(tns) = +1\} \), and \( T_- = \{t: \text{sgn}(tns) = -1\} \).

Thus,

\[
P[ \sum_{t=1}^{n} W_{tns} > \varepsilon] < P[ \sum_{t \in T_+} (B_{tns} - p_{tns}) > n^{1/2} \varepsilon/2] \quad (3.21)
\]

\[
+ P[ \sum_{t \in T_-} (B_{tns} - p_{tns}) < -n^{1/2} \varepsilon/2].
\]

Using the argument of Rao and Sethuraman (1975), and the modification of Loynes (1980), we can write for \( u > 0 \),

\[
P( \sum_{t \in T_+} (B_{tns} - p_{tns}) > n^{1/2} \varepsilon/2),
\]

\[
< \exp(-u n^{1/2} \varepsilon/2) \sum_{t \in T_+} \exp[u(B_{tns} - p_{tns})],
\]

\[
= \exp(-u n^{1/2} \varepsilon/2) \sum_{t \in T_+} [1 - p_{tns}(1 - e^u)] \exp(-u p_{tns}),
\]

\[
< \exp(-u n^{1/2} \varepsilon/2) \sum_{t=1}^{n} [1 - p_{tns}(1 - e^u)] \exp(-u p_{tns}),
\]

\[
\equiv a(n, u, \varepsilon) \quad \text{(say)},
\]

where we have used the fact that for \( u > 0 \),
\[(1 - a(1 - e^u)) \exp(-u\alpha) > 1 \text{ for } a \in [0, 1].\]

Now if we examine \(n^{-1} \log[a(n, u, \varepsilon)]\) we see that,

\[
n^{-1} \log[a(n, u, \varepsilon)] = -u n^{-1/2} \varepsilon/2 - u n^{-1} \sum_{t=1}^{n} p_{t,n} \alpha \]

\[+ n^{-1} \sum_{t=1}^{n} \log[1 - p_{t,n} \alpha (1 - e^u)],\]

\[< -u n^{-1/2} \varepsilon/2 - [u + (1 - e^u)] n^{-1} \sum_{t=1}^{n} p_{t,n} \alpha ,\]

since \(\log(1 - \delta) < \delta\) for \(\delta \in [0, 1]\). Thus,

\[\log[a(n, u, \varepsilon)] < n^{1/2} \{u\varepsilon/2 + [u + (1 - e^u)] n^{-1/2} \sum_{t=1}^{n} p_{t,n} \alpha \} .\]

But note that,

\[n^{-1/2} \sum_{t=1}^{n} p_{t,n} \alpha = n^{-1/2} \sum_{t=1}^{n} \left| \phi[z_{t,n} (\omega_{t+1})] - \omega_{t+1} \right| ,\]

\[< n^{-1/2} \sum_{t=1}^{n} \left| \phi[z_{t,n} (\omega_{t+1}; \xi_{t,n} (\omega_{t+1}))] - \phi[z_{t,n} (\omega_{t+1}; \theta)] \right| + \lambda ,\]

\[< \left( (e^{-1} 2L_{\varepsilon}) + 1 \right) \{n^{-1/2} \sum_{t=1}^{n} q_{t,n} (\omega_{t+1}, \varepsilon_1) \} + \lambda ,\]

\[< (e^{-1} 2L_{\varepsilon} + 1) K(\varepsilon_1) + \lambda ,\]

where,
\[ K(\epsilon_1) = \sup\{ |n^{-1/2} \sum_{t=1}^{n} q_t^*(w, \epsilon_1)| : w \in [0, 1], n \} \]

We know that \( K(\epsilon_1) < \infty \) by Result 3.1. Thus we can write,

\[ \log[a(n, u, \epsilon)] < n^{1/2} \left\{ \frac{u \epsilon}{2} + [u + 1 - \epsilon] \right\} \left[ \frac{(1 + 2 \epsilon^{-1} L) K(\epsilon_1) + \lambda} {1 + \epsilon L} \right] \]

and the right hand side does not depend on \( s \) or \( k \). Thus by an appropriate choice of \( u^* \in (0, \infty) \) we can write,

\[ a(n, u^*, \epsilon) < \exp(- c_2 n^{-1/2}) \]

where \( c_2 \in (0, \infty) \), and \( c_2 \) does not depend on \( n \), \( s \), or \( k \). We can do the same procedure for the second term of (3.21), and combining these results together we obtain,

\[ P\{ \max_{l,s t=1}^{n} W_{ntns} > \epsilon \} < 2 (2 L \epsilon^{-1} + 1)^{k+2} (1 + n^{-1/2} \lambda) \exp(- c_2 n^{1/2}) \]

This expression goes to zero as \( n \to \infty \). Thus we have shown that,

\[ \sup_{\omega \in [0, 1]} |W_n(\omega)| = o_p(1) \]

Theorem 3.2 has several applications. We state these as corollaries.

**Corollary 3.1.** Let the assumptions and definitions of Theorem 3.2 hold, then \( W_n \overset{L}{\to} W \) where \( W \) is the limiting Gaussian process given in Theorem 3.1.
Proof. By Theorem 3.1 \( W_n \xrightarrow{L} W \), so by Theorem 3.2 \( W_n \) has the same limiting distribution as \( W^* \). \( \square \)

**Corollary 3.2.** Let the assumptions of Theorem 3.2 hold, and define for \( x \in \mathbb{R} \)

\[
F_n(x) = n^{-1} \sum_{t=1}^{n} 1(Z_t < x),
\]

\[
F_n^*(x) = n^{-1} \sum_{t=1}^{n} 1(Z^*_t < x),
\]

where \( Z_t \) and \( Z^*_t \) were defined in Theorem 3.2. Then,

\[
\sup_{x \in \mathbb{R}} |F_n^*(x) - F_n(x)| = o_p(n^{-1/2}).
\]

**Proof.** Let \( \omega \) be defined in terms of \( x \) by the mapping \( \omega = \phi(x) \) for \( x \in \mathbb{R} \), where \( \phi \) is the standard normal distribution function. Remember that \( \phi \) is a one-to-one mapping from the real numbers to the unit interval. Note that,

\[
n^{1/2} |F_n^*(x) - F_n(x)| = |W_n^*(\omega) - W_n(\omega)|,
\]

under the mapping \( \phi \). The result now follows from Theorem 3.2. \( \square \)

The proof of Theorem 3.2 is quite technical, but the result is actually quite useful. To apply the theorem, and its corollaries, to the case of standardized residuals in the univariate case we do the following.
1. Let the \( v_t \)'s of Chapter Two take the place of the \( Y_t \)'s of Theorem 3.2.

2. Let the \( x_t \)'s of Chapter Two take the place of the \( X_t \)'s of Theorem 3.2.

3. Let the standardized residuals of expression (2.11) take the place of the \( Z_t \)'s in Theorem 3.2.

Thus under the conditions given in Chapter Two, the conditions of Theorem 3.2 hold. Thus the empirical process based on the transformed standardized regression residuals converges weakly to the same limit law of Theorem 3.1. Also Corollary 3.2 gives the order of the difference between \( F_n \) and \( F_n^* \) in terms of the sup-norm metric. This norm is also called the Kolmogorov metric for distribution functions. These results will be used extensively in Chapter Four.

Theorem 3.2 and its corollaries can also be applied to the standardized components of the residual vector defined in (2.29) for the multivariate model. For an \( i \in \{1, \ldots, r\} \) we do the following.

1. Let the \( v_{ti} \) for \( t = 1, \ldots, n \) of Chapter Two take the place of the \( Y_t \)'s of Theorem 3.2.

2. Let the \( x_{t(i)} \) for \( t = 1, \ldots, n \) of (2.31) take the place of the \( X_t \)'s of Theorem 3.2.

3. Let the standardized components of the residual vector defined in (2.29) take the place of the \( Z_t \)'s in Theorem 3.2.
Thus under the conditions given in Chapter Two, Theorem 3.2 and its corollaries hold. We will use these ideas in the next chapter when we deal with goodness-of-fit tests for normality.

B. Multivariate Results

Our results in this section are an extension of an idea of Wood (1981a), which was discussed in Section C of Chapter One. We begin with Theorem 3.3 which is the vector analog of Theorem 3.2.

**Theorem 3.3.** Let \( Y_1, Y_2, \ldots, Y_n \) be iid \( N_p(\mu_Y, \Sigma_Y) \) with \( \Sigma_Y \) positive definite, and independent of the random \( k \)-vectors \( (X_1, X_2, \ldots, X_n) \). Assume \( \Sigma_{XX} = \Sigma_{XX} \) a.s. Let \( B_n \) be a random \( p \times k \) matrix such that \( B_n = O_p(n^{-1/2}) \). Let \( A_n \) be a random \( p \times p \) matrix such that \( \text{diag}(A_n) = o_p(n^{-1/2}) \), and the off diagonal elements of \( A_n \) are \( o_p(n^{-1/2}) \). Define, for \( n > p + 1 \),

\[
Z_t = (I_p + A_n)^{-1/2}(Y_t - \bar{Y}) + B_n(X_t - \bar{X}),
\]

\[
Z_t^* = (Z_{t1}^*, Z_{t2}^*, \ldots, Z_{tp}^*)',
\]

\[
Z_{ti}^* = n^{-1/2}(e_{ti} - \bar{e}_i),
\]

where, \( e_{ti} = \Sigma_{XY}^{-1/2}(Y_t - \mu_Y) \) for \( t = 1, 2, \ldots, n \). For \( \omega \in [0, 1] \) and \( i = 1, 2, \ldots, p \), define

\[
W_{n,i}(\omega) = n^{-1/2} \sum_{t=1}^{n} 1[\theta(Z_{ti}^*) < \omega] - \omega,
\]
\[
W_{n,i}^*(\omega) = n^{-\frac{1}{2}} \sum_{t=1}^{n} \{1[\theta(Z_{t,i}^*) < \omega] - \omega\}.
\]

Then \[\sup_{\omega \in [0,1]} |W_{n,i}^*(\omega) - W_{n,i}(\omega)| = o_p(1).\]

**Proof.** First note that we can write,

\[
Z_t = (I_p + A_n)Q_n \varepsilon_{t,\omega}^{1/2} (\xi_t - \bar{x}) + B_n (X_t - \bar{x}),
\]

where

\[
Q_n = \sqrt{\frac{1}{Z_{YY}}} \varepsilon_{t,\omega}^{1/2} = (\sqrt{\frac{1}{Z_{YY}}} \varepsilon_{t,\omega}^{1/2}) (\sqrt{\frac{1}{Z_{YY}}} \varepsilon_{t,\omega}^{1/2})^{1/2}.
\]

Notice that \(Q_n\) is an orthogonal matrix which is converging in probability to \(I_p\) at the rate \(O_p(n^{-1/2})\). Next we examine the \((1, \ell)\)-th element of the matrix \((I_p + A_n)Q_n \varepsilon_{t,\omega}^{1/2}\). For notational convenience, let us represent the \((1, \ell)\)-th element by

\[
[(I_p + A_n)Q_n \varepsilon_{t,\omega}^{1/2}]_{1,\ell} = \sum_{j=1}^{p} \sum_{k=1}^{p} ( \delta_{ij} + a_{ij}) q_{jk} m_{k\ell},
\]

where \(a_{ij}\), \(q_{jk}\), and \(m_{k\ell}\) are elements of the matrices \(A_n\), \(Q_n\), and \(m_{1/2}\) respectively; and \(\delta_{ij} = 1\) if \(i = j\), and equals zero if \(i \neq j\). First we examine the case where \(i = \ell\). Then,

\[
[(I_p + A_n)Q_n \varepsilon_{t,\omega}^{1/2}]_{ii} = \sum_{j=1}^{p} \sum_{k=1}^{p} ( \delta_{ij} + a_{ij}) q_{jk} m_{ki},
\]

\[
= \sum_{k=1}^{p} (1 + a_{ii}) q_{ik} m_{ki} + \sum_{j \neq i}^{p} a_{ij} q_{jk} m_{ki}.
\]
From the definition of $A_n$ we know that \( a_{ii} = o_p(n^{-1/2}) \), and \( a_{ij} = o_p(n^{-1/2}) \) for \( i \neq j \). From the definition of \( m_{\varepsilon \varepsilon} \), we know that \( m_{kk} = o_p(n^{-1/2}) \) for \( k \neq i \), and \( m_{ii} = m_{\varepsilon \varepsilon i i} \). From the definition of \( Q_n \) and Lemma 2.2 we know that \( q_{ii} = 1 + o_p(n^{-1/2}) \) and \( q_{ij} = o_p(n^{-1/2}) \) for \( i \neq j \). Combining these results we obtain

\[
[(r_p + A_n)Q_n m_{\varepsilon \varepsilon}^{-1/2}]_{ii} = (1 + c_{ni})m_{\varepsilon \varepsilon i i}^{-1/2},
\]

where \( c_{ni} = o_p(n^{-1/2}) \) for \( i = 1, 2, \ldots, p \).

Now for \( i \neq \ell \) we obtain,

\[
[(r_p + A_n)Q_n m_{\varepsilon \varepsilon}^{-1/2}]_{i,\ell} = d_{i,\ell} = o_p(n^{-1/2}).
\]

If we let \( b_{ni} \) denote the \( i \)-th row of \( B_n \), we can represent \( Z_{t1} \) as

\[
Z_{t1} = (1 + c_{ni})Z_{t1}^* + \sum_{j \neq i} d_{ij}(\varepsilon_{tj} - \bar{\varepsilon}_j) + b_{ni}(x_i - \bar{x}).
\]

The theorem now follows from Theorem 3.2.

Notice that the \( W_{n,i}^*, \ i = 1, \ldots, p \) are independent processes. Thus the theorem says that the processes \( W_{n,i} \) for \( i = 1, \ldots, p \) are independent in the limit. This will be used when we deal with the distribution of goodness-of-fit statistics for multivariate data.

Theorem 3.3 deals with \( p \) separate empirical processes based on \( n \) observations each. In other words, the total number of random variables, \( np \), is broken up into \( p \) groups and an empirical process is
created for each group. The next corollary is for empirical processes based on more general groupings.

**Corollary 3.3.** Let the conditions and notation of Theorem 3.3 hold. Let $S$ be a subset of distinct elements of the set $\{1, 2, \ldots, p\}$ chosen independently of the data, and let $1 < s < p$ be the number of elements in $S$. Define for $\omega \in [0, 1]$,

$$W_n(\omega) = s^{-1/2} \sum_{i \in S} W_{n,i}(\omega),$$

$$W^*(\omega) = s^{-1/2} \sum_{i \in S} W^*_{n,i}(\omega).$$

Then

$$\sup_{\omega \in [0,1]} |W_n(\omega) - W^*_n(\omega)| = o_p(1).$$

**Proof.** Note that,

$$\sup_{\omega \in [0,1]} |W_n(\omega) - W^*_n(\omega)| \leq s^{-1/2} \sum_{i \in S} \sup_{\omega \in [0,1]} |W^*_{n,i}(\omega) - W_{n,i}(\omega)|.$$

Thus the result follows from Theorem 3.3.

The following theorem is a multivariate version of Theorem 3.1. It is the tool used to obtain limiting distributions for empirical processes based on standardized vectors.

**Theorem 3.4.** Let $Y_1, Y_2, \ldots, Y_n$ be iid $N_p(0, I_p)$, and define for $\omega \in [0, 1]$,
Then \( W_n \overset{L}{\to} W \) as a stochastic process on \([0, 1]\), where \( W \) is the Brownian bridge described in Theorem 3.1.

**Proof.** Note that for each \( n \) all the \( W_{n,i} \) are independent stochastic processes on \([0, 1]\). Since they all converge to the same Gaussian process, their mean multiplied by \( p^{1/2} \) converges to the same process. \( \square \)

The next corollary is a simple application of these results.

**Corollary 3.4.** Let the conditions and notation of Theorem 3.3 and Corollary 3.3 hold. Then \( W_n \overset{L}{\to} W \) as a stochastic process on \([0, 1]\), where \( W \) is the Brownian bridge described in Theorem 3.1.

**Proof.** The result follows immediately from Corollary 3.3 and Theorem 3.4. \( \square \)

A multivariate extension of Corollary 3.2 is given next.
Corollary 3.5. Let the assumptions of Theorem 3.3 hold, and define for \( x \in \mathbb{R} \) and \( i = 1, \ldots, p \),

\[
F_{n,i}(x) = n^{-1} \sum_{t=1}^{n} I(Z_{ti} < x),
\]

\[
F^*_n(x) = n^{-1} \sum_{t=1}^{n} I(Z^*_{ti} < x),
\]

where \( Z_{ti} \) and \( Z^*_{ti} \) are defined in Theorem 3.3. Then, for \( i = 1, \ldots, p \),

\[
\sup_{x \in \mathbb{R}} |F^*_n(x) - F_{n,i}(x)| = o_p(n^{-1/2}).
\]

Proof. The proof is the same one used in Corollary 3.2, applied now for each \( i \).

We also have the following.

Corollary 3.6. Let the conditions and notation of Theorem 3.3 hold. Let \( S \) be a subset of distinct elements of the set \( \{1, 2, \ldots, p\} \), and let \( 1 < s < p \) be the number of elements in \( S \). Define for \( x \in \mathbb{R} \),

\[
F_n(x) = (ns)^{-1} \sum_{t=1}^{n} \sum_{i \in S} I(Z_{ti} < x),
\]

\[
F^*_n(x) = (ns)^{-1} \sum_{t=1}^{n} \sum_{i \in S} I(Z^*_{ti} < x).
\]
Then,

$$\sup_{x \in \mathbb{R}} |F^*(x) - F_n(x)| = o_p(n^{-1/2}).$$

**Proof.** Follows immediately from Corollary 3.5. □

Notice that $F_{n,i}^*$ for $i = 1, ..., p$ are independent random processes. Thus Corollary 3.5 says that the processes $F_{n,i}^*$ for $i = 1, ..., p$ are independent in the limit. This will be used when we deal with the distribution of goodness-of-fit statistics for multivariate data.

To apply Theorem 3.3 and its corollaries to the case of standardized residual vectors we do the following.

1. Let the $v_i$'s of Chapter Two take the place of the $Y_i$'s of Theorem 3.3.

2. Let the $x_i$'s of Chapter Two take the place of the $z_i$'s of Theorem 3.3.

3. Let the standardized residual vectors of expression (2.23) take the place of the $Z_i$'s in Theorem 3.3.

Thus under the conditions given in Chapter Two, the conditions of Theorem 3.3 hold. In particular, the empirical process based on any subset of the components of the vector of transformed standardized regression residuals converges weakly to the limit law of Theorem 3.1.
IV. GOODNESS-OF-FIT STATISTICS BASED ON REGRESSION RESIDUALS

In the previous chapter we examined the large sample behavior of the empirical distribution function based on standardized regression residuals. In this chapter we use those results to examine the limiting behavior of goodness-of-fit statistics based on residuals.

A. EDF Goodness-of-Fit Statistics

We begin by considering the univariate model. Let \( \hat{F}_n(\omega) \) for \( \omega \in [0, 1] \) be the empirical distribution function based on \( \hat{Z}_1, \ldots, \hat{Z}_n \) where,

\[
\hat{Z}_t = \Phi \left( \frac{s_{vv}^{-1/2} (v_t - \bar{v})}{s_{vv}} \right),
\]

and \( s_{vv} \) and \( \bar{v} \) were defined in (2.9). Let \( F_n(\omega) \) for \( \omega \in [0, 1] \) be the empirical distribution function based on \( Z_1, \ldots, Z_n \) where,

\[
Z_t = \Phi \left( \frac{m_{vv}^{-1/2} (v_t - \bar{v})}{m_{vv}} \right),
\]

where \( m_{vv} \) and \( \bar{v} \) are the sample variance and mean of the \( v_t \)'s .

Also, let

\[
\hat{W}_n(\omega) = n^{1/2} [\hat{F}_n(\omega) - \omega], \quad W_n(\omega) = n^{1/2} [F_n(\omega) - \omega],
\]

for \( \omega \in [0, 1] \).
Goodness-of-fit statistics which are functions of $W_n$ or $\hat{W}_n$ are called **EDF goodness-of-fit statistics**, where EDF stands for empirical distribution function. In the previous chapter we found conditions under which $\hat{W}_n$ and $W_n$ converge to the same limiting Gaussian process $W$ defined in Theorem 3.1. It is intuitive that nice functions of $\hat{W}_n$ or $W_n$ will converge in distribution in $D[0, 1]$ to the corresponding function of $W$. This idea was originally proposed by Doob (1949) and provides a way to describe the limiting distribution of EDF statistics. These ideas are special cases of applications of the general theory of weak convergence in metric spaces. Some fundamental concepts of weak convergence are reviewed in Appendix A. We next give some examples.

1. **Kolmogorov-Smirnov type statistics**

The one-sided Kolmogorov-Smirnov statistics are defined by

$$
\hat{D}^+_{n} = n^{-1/2} \sup_{\omega \in [0,1]} \hat{W}_n(\omega),
$$

and

$$
\hat{D}^-_{n} = n^{-1/2} \inf_{\omega \in [0,1]} \hat{W}_n(\omega).
$$

In terms of the $\hat{Z}_t$'s, these become

$$
\hat{D}^+_{n} = \max_{1 \leq t \leq n} (t/n - \hat{Z}_{t:n}),
$$

and

$$
\hat{D}^-_{n} = \min_{1 \leq t \leq n} (t/n - \hat{Z}_{t:n}).
$$
and,

$$ \hat{D}_n^- = \max_{1 \leq t \leq n} \left[ Z_{t:n} - (t - 1)/n \right] , $$

where $Z_{t:n}$ stands for the $t$-th order statistic from a sample of size $n$.

The functionals $\sup x(t)$ and $\inf x(t)$ are continuous with respect to the Skorohod metric on $D[0, 1]$. See Appendix A, and Billingsley (1968). Hence under the conditions which ensure that $\hat{W}_n \xrightarrow{L} W$, the limit laws of $n^{1/2} \hat{D}_n^+$ and $n^{1/2} \hat{D}_n^-$ are given by the laws of the random variables,

$$ D^+ = \sup_{\omega \in [0,1]} W(\omega) , \tag{4.6} $$

and,

$$ D^- = \inf_{\omega \in [0,1]} W(\omega) . $$

The distributions of $D^+$ and $D^-$ have apparently not been tabulated.

The one-sided Kolmogorov-Smirnov statistics have not been widely used. Instead functions of these statistics have been used. The Kolmogorov-Smirnov statistic is given by

$$ \hat{D}_n = \max(\hat{D}_n^+, -\hat{D}_n^-) , \tag{4.7} $$
The Kulper statistic is given by

\[ \hat{V}_n = D_n^+ - D_n^- , \]

(4.8)

\[ = n^{-1/2} \sup_{\omega \in [0,1]} \hat{W}_n(\omega) - n^{-1/2} \inf_{\omega \in [0,1]} \hat{W}_n(\omega) . \]

The functionals \( \sup_{t} |x(t)| \) and \( \sup_{t} x(t) - \inf_{t} x(t) \) are continuous with respect to the Skorohod metric on \( D[0, 1] \). Thus under the conditions which ensure \( \hat{W}_n \overset{L}{\to} W \), the limit laws of \( n^{1/2} \hat{D}_n \) and \( n^{1/2} \hat{V}_n \) are given by the laws of the random variables,

\[ D = \sup_{\omega \in [0,1]} |W(\omega)| , \]

(4.9)

and,

\[ V = \sup_{\omega \in [0,1]} W(\omega) - \inf_{\omega \in [0,1]} W(\omega) . \]

The distributions of \( D \) and \( V \) have been tabled by Stephens (1974), along with small sample modifications. These are presented in Table 10.1 of Appendix B.
2. **Cramer-von Mises statistic**

The Cramer-von Mises statistic is defined by

\[ C_n^2 = \frac{1}{n} \int_0 W_n^2(\omega) \, d\omega. \quad (4.10) \]

In terms of the \( Z_t \)'s this can be written as

\[ C_n^2 = (12n)^{-1} + \sum_{t=1}^{n} \frac{1}{(2n)^{-1}(2t-1)} \left[ Z_{t:n} - (2n)^{-1}(2t-1) \right]^2. \quad (4.11) \]

Durbin (1973b) showed that the function defined by the integral in (4.10) is continuous with respect to the Skorohod metric on \( D[0, 1] \).

Thus under the conditions which ensure \( \hat{W}_n \overset{L}{\to} W \), the limit law of \( C_n^2 \) is given by the law of the random variable,

\[ C^2 = \frac{1}{n} \int_0 W^2(\omega) \, d\omega. \quad (4.12) \]

Percentage points for the distribution of \( C^2 \) have been tabled by Stephens (1974), along with small sample modifications. We reproduce these in Table 10.1 of Appendix B.

3. **Watson's statistic**

The Watson statistic is defined by

\[ \hat{U}_n^2 = \frac{1}{n} \int_0 W_n^2(\omega) \, d\omega - \left[ \frac{1}{n} \int_0 W_n(\omega) \, d\omega \right]^2. \quad (4.13) \]

In terms of the \( Z_t \)'s this can be written as
\[ \hat{U}_n^2 = \frac{C_n^2}{n} - n(Z - \frac{1}{2})^2, \quad (4.14) \]

where \( \frac{C_n^2}{n} \) is the Cramer-von Mises statistic, and \( Z = n^{-1} \sum_{t=1}^{n} \hat{Z}_t \).

Durbin (1973b) showed that the functions in (4.13) are continuous in the Skorohod metric on \( D[0, 1] \). Thus under the conditions which ensure that \( \hat{W}_n \xrightarrow{L} W \), the limit law of \( \hat{U}_n^2 \) is given by the law of the random variable,

\[ U^2 = \int_0^1 W^2(\omega) \, d\omega - \left[ \int_0^1 W(\omega) \, d\omega \right]^2. \quad (4.15) \]

Percentage points for the distribution of \( U^2 \) have been tabled by Stephens (1974), along with small sample modifications. We reproduce these in Table 10.1 of Appendix B.

4. **Anderson-Darling statistic**

The Anderson-Darling statistic is defined by

\[ \hat{A}_n^2 = \int_0^1 [\omega(1 - \omega)]^{-2} \hat{W}_n^2(\omega) \, d\omega. \quad (4.16) \]

In terms of the \( \hat{Z}_t \)'s this can be written as

\[ \hat{A}_n^2 = -n - n^{-1} \sum_{t=1}^{n} (2t - 1)[\log \hat{Z}_{t:n} + \log(1 - \hat{Z}_{n+1-t:n})]. \quad (4.17) \]

This statistic was originally studied by Anderson and Darling (1952). Durbin (1973b) and Rosenblatt (1952) showed that the function defined by the integral in (4.16) is continuous in the Skorohod metric.
on $D[0, 1]$ except at a set of paths of measure zero under the
distribution of the limiting process $W$. Thus under the conditions
which ensure that $\hat{W}_n \overset{L}{\to} W$, the limit law of $\hat{A}_n^2$ is given by the law
of the random variable,

$$A^2 = \frac{1}{\omega(1 - \omega)[1 - \omega]^{-2} \omega^2(\omega)} \, d\omega. \quad (4.18)$$

Percentage points for the distribution of $A^2$ have been tabled by
Stephens (1974), along with small sample modifications. We reproduce
these in Table 10.1 of Appendix B.

5. Applications to multivariate data

Of the EDF statistics we have listed, the Kolmogorov-Smirnov $D_n$
statistic is the most widely used. It is programmed into many software
packages including SAS, which uses it for sample sizes greater than 50
and uses the percentage points from Stephens (1974).

We have concentrated on regression residuals from univariate
models, but the EDF statistics are easily applied to multivariate
regression models as well. Our presentation is an extension of an idea
of Wood (1981a). Let the $(1 \times r)$-vector $\hat{Z}_t^*$ be defined for
t $= 1, ..., n$ by,

$$\hat{Z}_t^* = (\hat{r}_t - \bar{r}) S_{\bar{r} \bar{r}}^{-1/2}, \quad (4.19)$$

where $S_{\bar{r} \bar{r}}$ and $\bar{r}$ were defined in (2.21). We called $\hat{Z}_t^*$ the
standardized residual vector in Chapter Two. Now define for
\[ \hat{Z}_{ti} = \Phi(\hat{Z}_{ti}^*) , \] (4.20)

where \( \hat{Z}_{ti}^* \) is the \( i \)-th component of \( \hat{Z}_t^* \). Let \( \hat{F}_{nr}(\omega) \) for \( \omega \in [0, 1] \) be the empirical distribution function based on the collection of \( nr \) variables \( \hat{Z}_{ti}^* \) for \( t = 1, \ldots, n \) and \( i = 1, \ldots, r \). Also define for \( \omega \in [0, 1] \),

\[ \hat{W}_{nr}(\omega) = n^{1/2} [\hat{F}_{nr}(\omega) - \omega] . \] (4.21)

We gave conditions in Chapter Three under which \( \hat{W}_{nr} \) converged to the Gaussian process \( W \). Thus the EDF statistics described above, when based on the \( nr \) random variables in (4.20), have the same limit laws described above. This provides a test of multivariate normality based on readily available univariate statistics. It should be noted that we could have defined (4.20) for only a subset of the components of \( \hat{Z}_t^* \).

In particular, let \( S \) be a subset of distinct elements of the set \( \{1, \ldots, r\} \), and let \( s \) be the number of elements in \( S \). Now define for \( t = 1, \ldots, n \) and \( i \in S \),

\[ \hat{Z}_{ti} = \Phi(\hat{Z}_{ti}^*) . \] (4.22)

Let \( \hat{F}_{ns}(\omega) \) for \( \omega \in [0, 1] \) be the empirical distribution function based on the collection of \( sn \) variables \( \hat{Z}_{ti}^* \) for \( t = 1, \ldots, n \) and
\[ \hat{W}_{nS}(\omega) = n^{1/2} [\hat{W}_{nS}(\omega) - \omega] \quad (4.23) \]

The convergence of \( \hat{W}_{nS} \) to the Gaussian process \( \hat{W} \) was proven in Corollary 3.4. Thus we can also test for joint normality based on subsets of the standardized residual vector.

We have concentrated on testing for multivariate normality by forming a large group of the components of the standardized residuals and calculating a single EDF statistic based on that group. Another method which is reasonable is the following. Let \( T \) represent any of the EDF statistics discussed above, and let \( \hat{T}_{n,i} \) for \( i = 1, \ldots, r \) be the statistic calculated from the \( n \) values of the \( \hat{Z}_{t}^{i} \)'s for each of the \( r \) components of the standardized residual vector \( \hat{Z}_{t}^{i} \) in (4.19). From Chapter Three we know that the statistics \( \hat{T}_{n,i} \) for \( i = 1, \ldots, r \), properly normalized, converge in law to random variables which are independent in the limit. Thus it makes sense to consider a random variable such as

\[ \hat{T}_{n} = \max_{1 \leq i \leq r} \hat{T}_{n,i} \quad (4.24) \]

The weak convergence results tell us that,

\[ P(\hat{T}_{n,i} < t) \rightarrow F_{T}(t) \quad \text{as} \quad n \rightarrow \infty, \quad (4.25) \]
for $i = 1, \ldots, r$ where $F_n(\cdot)$ is some distribution function, and the convergence in (4.25) is for all continuity points of $F_n$. This implies, by independence in the limit, that

$$P(T_n > t) = 1 - [F_n(t)]^R$$

(4.26)

Based on this convergence result it is easy to construct significance tests based on the random variable $T_n$. We construct the $r$ separate significance tests based on $T_{n,1}, \ldots, T_{n,r}$ and calculate the significance levels $\alpha_1, \ldots, \alpha_r$. Then the overall significance level is given by,

$$\hat{\alpha} = 1 - \prod_{i=1}^{r} (1 - \alpha_i)$$

(4.27)

which is justified in large samples by (4.25).

B. Chi-Squared Tests of Fit

We begin by considering the univariate model. Let the notation of Section A with $F_n$, $F_n$, and (4.1)-(4.3) hold. It will be clear that the chi-squared statistic actually can be considered an EDF statistic, but we are considering it separately because of some of its special features.

Many authors have considered the problem of applying the Pearson chi-squared test for goodness-of-fit based on an iid sample from a continuous distribution. In the case of testing for normality with
unknown mean and variance, the behavior of the chi-squared statistic is dependent upon how \( \mu \) and \( \sigma^2 \) are estimated as well as how the data are grouped. Our approach follows that of Dahiya and Gurland (1972) who use the sample mean and variance to estimate \( \mu \) and \( \sigma^2 \), and group the data into random equal probability intervals.

To fix ideas, imagine that we can actually observe \( \bar{v}_1, \ldots, \bar{v}_n \) and we want to calculate a chi-squared test based on \( k \) categories. The method of Dahiya and Gurland (1972) prescribes the \( k \) random intervals,

\[
I_j = \left( \bar{v} + \frac{1}{\sqrt{v}} c_{k,j-1}, \bar{v} + \frac{1}{\sqrt{v}} c_{k,j} \right),
\]

for \( j = 1, \ldots, k \), where

\[
c_{k,j} = \phi^{-1}(k^{-1}j) \quad \text{for} \ j = 0, \ldots, k,
\]

and \( \phi^{-1} \) is the inverse function corresponding to the standard normal distribution function \( \phi \). The intervals are called random equal probability intervals because if, conditional on \( \bar{v} \) and \( m_{vv} \), \( Z \) were a random variable such that \( Z \sim N(\bar{v}, m_{vv}) \) then,

\[
P(Z \in I_j) = k^{-1} \quad \text{for} \ j = 1, \ldots, k.
\]

The chi-squared statistic based on these intervals is
Dahlya and Gurland (1972) have shown that for fixed \( k \) and \( n \rightarrow \infty \), the random variable in (4.30) converges in law to a random variable with distribution defined by

\[
\chi^2_{k-3} + \lambda_1 Z_1^2 + \lambda_2 Z_2^2 ,
\]

(4.31)

for \( k = 3, 4, \ldots \), where \( \chi^2_{k-3} \) is a chi-squared random variable with \( k - 3 \) degrees of freedom and independent of \((Z_1, Z_2)' \sim N_2(0, I_2)\). The constants \( \lambda_1 \) and \( \lambda_2 \) have been given by Watson (1957) and are

\[
\lambda_1 = 1 - k \sum_{i=1}^{k} \psi^2(i) , \quad \lambda_2 = 1 - (0.5)k \sum_{i=1}^{k} \psi^2(i) ,
\]

(4.32)

where,

\[
\psi_r(i) = c_{k,i}^r \psi(c_{k,i}) - c_{k,i-1}^r \psi(c_{k,i-1}) ,
\]
where the $c_{k,i}$'s were defined in (4.28), and $\phi$ is the standard normal density. Percentage points for the random variable in (4.31) are reproduced for $3 < k < 15$ in Table 10.3 of Appendix B.

The chi-squared statistic in (4.30) is easy to calculate in practice, which perhaps makes it more attractive than the EDF statistics described earlier. In order to calculate the chi-squared statistic based on regression residuals, we replace the $v_t$'s by the $v_t^*$'s and $(\bar{v}, m_{vv})$ by $(\bar{v}, s_{vv})$. We show next that the limiting distribution of the chi-squared statistic is the same as that of the variable in (4.31).

In order to more closely examine the limiting behavior of the chi-squared statistic in (4.30) we need to write it as a function of the empirical processes we have developed. Note that we can write (4.30) as,

\[ \sum_{i=1}^{k} E_i (O_i - E_i) = (n^{-1}) \sum_{i=1}^{k} \left\{ \sum_{i=1}^{n} 1(v_t \in I_i) - k^{-1}n \right\}^2, \quad (4.33) \]

\[ = (nk) \sum_{i=1}^{k} \left\{ (n^{-1}) \sum_{t=1}^{n} 1(v_t < \bar{v} + \frac{1}{2} m_{vv} c_{k,i}) - n^{-1} \sum_{t=1}^{n} 1(v_t < \bar{v} + \frac{1}{2} m_{vv} c_{k,i-1}) - k^{-1} \right\}^2, \]

\[ = (nk) \sum_{i=1}^{k} \left\{ F_n(k^{-1}i) - F_n[k^{-1}(i-1)] - k^{-1} \right\}^2, \]

\[ = k \sum_{i=1}^{k} \left\{ W_n(k^{-1}i) - W_n[k^{-1}(i-1)] \right\}^2. \]
Thus we can write the chi-squared statistic as a functional of the empirical process $W_n$. Because of this, we denote the chi-squared statistic by $\chi_k^2(W_n)$ and further simplify (4.33) to,

$$\chi_k^2(W_n) = kW_n^2(k^{-1}) + k\{1 - W_n[k^{-1}(k - 1)]\}^2$$ \hspace{1cm} (4.34)

$$+ k \sum_{i=2}^{k-1} (W_n[k^{-1}i] - W_n[k^{-1}(i - 1)])^2 .$$

The real valued function defined on $D[0, 1]$ is, for fixed $k$, continuous with respect to the Skorohod metric on $D[0, 1]$. Since $W_n \overset{L}{\longrightarrow} W$, we know that

$$\chi_k^2(W_n) \overset{L}{\longrightarrow} \chi_k^2(W) \text{ as } n \to \infty ,$$ \hspace{1cm} (4.35)

where

$$\chi_k^2(W) = kW^2(k^{-1}) + k\{1 - W[k^{-1}(k - 1)]\}^2$$

$$+ k \sum_{i=2}^{k-1} (W(k^{-1}i) - W[k^{-1}(i - 1)])^2 .$$

Under the conditions which ensure $\hat{W_n} \overset{L}{\longrightarrow} W$,

$$\chi_k^2(\hat{W_n}) \overset{L}{\longrightarrow} \chi_k^2(W) .$$ \hspace{1cm} (4.36)
The limit random variable in (4.35) may look complicated, but by the uniqueness of the weak limit we know that its distribution is precisely that of (4.31) as calculated by Dahiya and Gurland (1972). Thus, the chi-squared statistic of (4.30) based on regression residuals has a known limiting distribution, and it is exactly the one described by (4.31).

We have dealt only with the univariate model, but the chi-squared test can also be applied to multivariate data. As in (4.19) let \( \hat{\mathbf{z}}_t^* \) denote the standardized residual vector, where

\[
\hat{\mathbf{z}}_t^* = (\hat{\mathbf{y}}_t - \mathbf{\bar{y}}) \mathbf{S}_{vv}^{-1/2} \quad \text{for} \quad t = 1, \ldots, n. \tag{4.37}
\]

Now for fixed \( k \) construct the intervals,

\[
I_i^* = (c_{k,i-1}^*, c_{k,i}^*) \quad \text{for} \quad i = 1, \ldots, k. \tag{4.38}
\]

Also define

\[
0_i = \sum_{t=1}^{n} \sum_{j=1}^{r} I(\hat{\mathbf{z}}_t^* \in I_i^*),
\tag{4.39}
\]

\[
E_i = k^{-1}(\alpha_i).
\]

Now, as was shown in (4.33), it follows that

\[
\sum_{i=1}^{k} E_i^{-1} (0_i - E_i)^2 = \chi^2_k(\mathbf{w}_{nr}), \tag{4.40}
\]
where \( \hat{W}_{nr} \) was defined (4.21), and \( X^2(\cdot) \) was defined in (4.33) and (4.34). Thus under the conditions which ensure that \( \hat{W}_{nr} \xrightarrow{L} W \), it follows

\[
X_k^2(\hat{W}_{nr}) \xrightarrow{L} X_k^2(W) \quad \text{as} \quad n \to \infty,
\]

(4.41)

with limiting distribution given by (4.31).

The statistic defined by (4.39) and (4.40) is particularly easy to calculate since the intervals \( I_i^j \) for \( i = 1, \ldots, k \) are fixed and not random. This happens because we first put the residual vector through a scale and location transformation. This can also be done in the univariate case if one is working with the standardized residuals.

The above chi-squared test used all \( r \)-components of the standardized residual vector. As was done earlier we can also construct a test based on only a subset of the components. As was done in Section A.5 of Chapter Four, let \( S \) be a subset of distinct elements of the set \( \{1, \ldots, r\} \), and let \( s \) be the number of elements in \( S \). Now define

\[
0_i = \sum_{t=1}^{n} \sum_{j \in S} 1(Z^*_t \in I_j^i),
\]

(4.42)

\[
E_i = k^{-1}(ns).
\]

By applying Corollary 3.4, it follows that the chi-squared statistic has a limiting distribution given by (4.31).
As a final note, we mention that it is also possible to calculate \( r \) separate chi-squared statistics based on the \( r \) components of the standardized residual vector, then combine them as was done in (4.24). The comments following (4.24) also hold with no additional modifications.

C. Tests Based on Moments

Tests based on moments do not appear to be widely used, but they are easily handled by our weak convergence theory. Their convenience is that they are reasonably easy to calculate, and can provide a quick information about the data. Our attention will be restricted to the usual statistic of skewness, which we call \( g_1 \), and the usual statistic of kurtosis, which we call \( g_2 \).

We begin by considering the univariate model. To fix ideas, imagine that we can actually observe \( v_1, \ldots, v_n \). Now define,

\[
m_r = n^{-1} \sum_{t=1}^{n} (v_t - \overline{v})^r ,
\]

(4.43)

for \( r = 2, 3, 4 \). Then \( g_1 \) and \( g_2 \) are defined by

\[
g_1 = m_2^{-3/2} m_3 , \quad g_2 = m_2^{-2} m_4 - 3 .
\]

(4.44)

Under normality of the \( v_t \)'s, it is well known that

\[
n^{1/2} g_1 \xrightarrow{L} N(0, 6) , \quad n^{1/2} g_2 \xrightarrow{L} N(0, 24) .
\]

(4.45)
We show next that the same limiting distributions result when \( g_1 \) and \( g_2 \) are calculated with the \( \hat{v}_t \)'s instead of with the actual \( v_t \)'s.

In order to examine the limiting behavior of the skewness and kurtosis statistics of (4.44) we need to write them as functions of the empirical processes we have developed. To do so, we first define the functional for \( r \in \{1, 2, \ldots \} \) by

\[
\mu_r(F) = \int_{-\infty}^{\infty} x^r dF(x),
\]

where \( F \) is a distribution function with finite \( r \)-th moment. We can rewrite \( \mu_r(F) \) as follows,

\[
\mu_r(F) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} ru^{r-1} du \ dF(x),
\]

\[
= \int_{-\infty}^{\infty} ru^{r-1} du \ dF(x),
\]

\[
= r \int_{0}^{\infty} \left[ F(u) - \Phi(u) \right] u^{r-1} du - r \int_{0}^{\infty} \Phi(u) u^{r-1} du.
\]

Now letting \( \Phi \) denote the standard normal distribution function we can write

\[
\mu_r(F) - \mu_r(\Phi) = - r \int_{0}^{\infty} \left[ F(u) - \Phi(u) \right] u^{r-1} du,
\]

\[
= - r \int_{-\infty}^{0} \left[ F(u) - \Phi(u) \right] u^{r-1} du,
\]
where,

\[ \psi_x(\omega) = - r \phi^{-1}(\omega)r^{-1} \phi^{-1}(\omega) \text{ for } \omega \in (0, 1), \]

and \( \phi \) is the standard normal density.

Now let \( G_n(x) \) for \( x \in \mathbb{R} \) denote the empirical distribution function based on standardized random variables,

\[ m_{vv}^{-1/2} (v_t - \bar{v}) \text{ for } t = 1, \ldots, n. \quad (4.50) \]

Then \( g_1 \) and \( g_2 \) of (4.44) are

\[ g_1 = [(n - 1)^{-1}n]^{3/2} \{ \mu_3(G_n) - \mu_3(\phi) \} + O(n^{-1}), \quad (4.51) \]

\[ g_2 = [(n - 1)^{-1}n]^{2} \{ \mu_4(G_n) - \mu_4(\phi) \}, \]

where we have used the fact that \( \mu_3(\phi) = 0 \) and \( \mu_4(\phi) = 3 \). The factor \( [(n - 1)^{-1}n]^{3/2} \) arises because of the divisor in the definition of the moments \( m_{vv} \) and \( m_r \) for \( r = 2, 3, 4 \). Now combining (4.51) and (4.48) we can write \( g_1 \) and \( g_2 \) as functions of the empirical process \( W_n \) defined in (4.3).

\[ n^{1/2} g_1(W_n) = [(n - 1)^{-1}n]^{3/2} \int_0^1 \psi_x(\omega) W_n(\omega) d\omega + O(n^{-1/2}), \quad (4.52) \]
Because $W_n \xrightarrow{L} W$ by Theorem 3.1, the results of Shorack and Wellner (1986) imply that

$$n^{1/2} g_2(W_n) = \left[(n - 1)^{-1} n\right]^{3/2} \int_0^1 \psi_4(\omega)W_n(\omega) d\omega.$$  

The limiting random variables in (4.53) may look complicated, but by uniqueness of the weak limit, their distributions are given by (4.45).

We now return to the skewness and kurtosis statistics based on the $\hat{v}_t$'s instead of the $v_t$'s. If we call these statistics $\hat{g}_1$ and $\hat{g}_2$, then just as in (4.52), we have

$$n^{1/2} \hat{g}_1 = \left[(n - k)^{-1} n\right]^{3/2} \int_0^{1/2} \psi_3(\omega)W_n(\omega) d\omega + O(n^{-1/2}),$$

$$n^{1/2} \hat{g}_2 = \left[(n - k)^{-1} n\right]^{3/2} \int_0^{1/2} \psi_4(\omega)W_n(\omega) d\omega.$$  

Under conditions which ensure $\hat{W}_n \xrightarrow{L} W$,

$$n^{1/2} \hat{g}_1 \xrightarrow{L} \int_0^{1/2} \psi_3(\omega)W(\omega) d\omega,$$

$$n^{1/2} \hat{g}_2 \xrightarrow{L} \int_0^{1/2} \psi_4(\omega)W(\omega) d\omega.$$
Thus, the skewness and kurtosis statistics based on regression residuals have known limiting distributions, and the distributions are precisely the ones given in (4.45).

One method of extending our results to the multivariate model is to apply the skewness and kurtosis statistics to the standardized components of the residual vector to obtain \( r \) separate tests of marginal normality. Another way is to apply the statistics to the components of the standardized residual vector to obtain \( r \) separate tests of marginal normality which would have the nice property that they are independent in the limit. This last fact follows by Theorem 3.3. A third way would be to apply the statistics to the collection of \( rn \) random variables obtained by combining the \( r \) components of the \( n \) standardized residual vectors. The statistics have the correct limiting distributions by virtue of expression (4.53) and Corollary 3.4.

A fourth method to measure multivariate skewness and kurtosis has been proposed by Mardia (1970, 1974). In terms of the \( \bar{v}_t \)'s, these statistics are

\[
b_{1,r} = n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} h_{ts}^3,
\]

\[
b_{2,r} = n^{-1} \sum_{t=1}^{n} h_{tt}^2,
\]

where

\[
h_{ts} = (\bar{v}_t - \bar{\bar{v}}) (\bar{v}_s - \bar{\bar{v}})',
\]
for $t, s = 1, \ldots, n$. It has been shown by Mardia (1970) that

$$6^{-1}nb_{1,r} \xrightarrow{L} x_f^2,$$

where $f = 6^{-1}r(r + 1)(r + 2)$, and

$$n^{1/2} [b_{2,r} - r(r + 2)] \xrightarrow{L} N[0, 8r(r + 2)],$$

as $n \to \infty$. Now define

$$b_{1,r} = n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} \hat{h}_{ts},$$

$$b_{2,r} = n^{-1} \sum_{t=1}^{n} \hat{h}_{tt},$$

where

$$\hat{h}_{ts} = (\hat{\nu}_t - \bar{\nu}) S^{-1}_{\nu\nu}(\hat{\nu}_s - \bar{\nu})',$$

for $t, s = 1, \ldots, n$. By the representation given in Theorem 2.4 and after a bit of algebra it follows that

$$6^{-1}nb_{1,r} \xrightarrow{L} x_f^2,$$

where $f = 6^{-1}r(r + 1)(r + 2)$, and

$$n^{1/2} [b_{2,r} - r(r + 2)] \xrightarrow{L} N[0, 8r(r + 2)],$$
as \( n \to \infty \).

It is unclear which of these procedures one should use in practice. This question is explored by a Monte Carlo study in Section B of Chapter Six.

D. Tests Based on L-Statistics

In the first three sections of this chapter we have shown how the weak convergence results of Chapter Three imply that many goodness-of-fit statistics based on residuals behave, in large samples, like statistics based on an iid sample with estimated mean and variance. The procedure is clear. We simply show that a statistic can be written as a real valued function on \( D[0, 1] \) which is continuous with respect to the Skorohod metric, except at (possibly) a set of probability zero with respect to the distribution of the limiting Gaussian process \( W \) given in Theorem 3.1.

Unfortunately not all statistics can be handled in this way. A particular example is the class of linear combinations of order statistics, also called L-statistics. There does exist a theory about the differentiability of L-statistics with respect to empirical distribution functions from iid samples, based on the theory of statistical functionals. This statistical functional approach has been used to great advantage in the theory of robustness and asymptotic normal theory for complex statistical functionals. See Fernholz (1983) and Huber (1981). Unfortunately this theory does not directly apply to L-statistics based on regression residuals. The main problem is that the
regression residuals are not iid random variables, and the results of Chapters Two and Three give us no immediate information about the convergence of the ordered $v_1$'s in terms of the ordered $v_t$'s.

We next develop a theory of convergence of $L$-statistics based on regression residuals. The topic has been investigated by Bickel (1973) in the context of linear models, but his motivation and results are different from ours. Our results enable us to examine the limiting behavior of the Shapiro-Wilk statistic and related statistics, based on regression residuals.

We begin by considering the univariate model. Let $F_n(x)$ for $x \in \mathbb{R}$ denote the empirical distribution function based on

$$m^{-1/2}(v_t - \bar{v}) \text{ for } t = 1, \ldots, n,$$  \hspace{1cm} (4.56)

and let $\hat{F}_n(x)$ for $x \in \mathbb{R}$ denote the empirical distribution function based on

$$s^{-1/2}(v_t - \bar{v}) \text{ for } t = 1, \ldots, n.$$  \hspace{1cm} (4.57)

Notice that this notation differs from that of (4.1)-(4.2) that was used in earlier sections of this chapter. This is because we need to look at distributions on the entire real line instead of on $[0, 1]$ as was done earlier.

We consider $L$-statistics of the form,
where $\hat{v}_{n;i}$ for $i = 1, \ldots, n$, is the $i$-th order statistic based on a sample of size $n$. The $a_{n;i}$'s are fixed constants which are allowed to depend on $i$ and $n$. The expression in (4.58) is inconvenient to work with, since we don't know much about the $\hat{v}_{n;i}$'s. We would prefer to write (4.58) as a function of $\hat{F}_n$ since we know more about the limiting behavior of $\hat{F}_n$. Thus, define the function $J_n(u)$ for $u \in [0, 1]$ as follows.

$$J_n(u) = a_{n;i} \text{ for } n^{-1}(i - 1) < u < n^{-1}i, \quad (4.59)$$

for $i = 1, \ldots, n$ and let $J_n(1) = a_{n,n}$. Thus we can write (4.58) as

$$T(\hat{F}_n; J_n) = \int_0^1 \hat{F}_n^{-1}(u)J_n(u)du, \quad (4.60)$$

where we define $\hat{F}_n^{-1}$ for an arbitrary distribution function $F$ on the real line by,

$$\hat{F}_n^{-1}(u) = \inf\{x: F(x) > u\}, \quad (4.61)$$

for $u \in (0, 1)$. We introduce the notation $T(\hat{F}_n; J_n)$ for convenience and to emphasize that the L-statistic of (4.58) is completely determined by the function $J_n$, and is a functional of the distribution function.
We are now ready to prove the following theorem about the limiting behavior of \( T(\hat{F}_n; J_n) \).

**Theorem 4.1.** Let the sequence of functions \( \{J_n\}_{n=1}^{\infty} \) satisfy the following two conditions.

1. \( \{J_n\}_{n=1}^{\infty} \) are uniformly integrable on \([0, 1]\),
2. For \( \delta \in (0, \frac{1}{2}) \), \( \sup_n \sup_{\delta < u < 1-\delta} |J_n(u)| < \infty \).

Then \( T(\hat{F}_n; J_n) \) and \( T(F_n; J_n) \) are well defined for all \( n \) and,

\[
|T(\hat{F}_n; J_n) - T(F_n; J_n)| = o_p(1),
\]
as \( n \to \infty \).

**Proof.** First note that by Hölder's inequality,

\[
\int_0^1 |F_n^{-1}(u)J_n(u)| \, du \leq \left\{ \int_0^1 [F_n^{-1}(u)]^2 \, du \right\}^{1/2} \left\{ \int_0^1 J_n^2(u) \, du \right\}^{1/2}.
\]

But,

\[
\int_0^1 [F_n^{-1}(u)]^2 \, du = \int_{-\infty}^{\infty} x^2 \, dF_n(x) < 1 \text{ for all } n,
\]

since \( F_n \) is based on the standardized variables in (4.56). Also, uniformly integrability gives
Thus $T(F_n; J_n)$ is well defined for all $n$. Now

$$\frac{1}{n} \int [\hat{F}_{n}(u)]^2 du = \int x^2 d\hat{F}_{n}(x) < 1 \text{ for all } n,$$  \hfill (4.65)

because $\hat{F}_{n}$ is based on (4.57). It follows that $T(F_n; J_n)$ is also well defined for all $n$.

Let $\varepsilon > 0$ be given, and let $\delta \in (0, \frac{1}{2})$ be a number to be chosen later. Then

$$|T(F_n; J_n) - T(F_n; J'_n)| < \frac{1}{n} \int |J_n(u)||\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du,$$

$$= \int_{0}^{1-\delta} |J_n(u)||\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du$$

$$+ \int_{\delta}^{1} |J_n(u)||\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du$$

$$+ \int_{1-\delta}^{1} |J_n(u)||\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du. \hfill (4.66)$$

We consider the three terms to the right of the equality. Note that,

$$\int_{\delta}^{1-\delta} |J_n(u)||\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du \leq \{\sup_{\delta < u < 1-\delta} |J_n(u)|\} \int_{\delta}^{1-\delta} |\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du$$

$$< M_6 \int_{0}^{1} |\hat{F}_{n}^{-1}(u) - F_{n}^{-1}(u)| du, \hfill (4.67)$$
where,

$$M_\delta = \sup_n \sup_{\delta u \leq 1-\delta} |J_n(u)|$$

and $M_\delta$ is finite by assumption. Furthermore,

$$\int_{0}^{1} |\hat{F}_n^{-1}(u) - F_n^{-1}(u)| \, du = \int_{-\infty}^{\infty} |\hat{F}_n(x) - F_n(x)| \, dx,$$

since both expressions are the area between $F_n$ and $\hat{F}_n$. Also

$$\int_{-\infty}^{\infty} |F_n(x) - \hat{F}_n(x)| \, dx = \int_{-\infty}^{\infty} |\hat{F}_n(x) - F_n(x)| \, dx,$$

since $\hat{F}_n$ and $F_n$ are based on the standardized variables of (4.56) and (4.57). We can also write,

$$\int_{-\infty}^{\infty} |\hat{F}_n(x) - F_n(x)| \, dx < 2n^{1/2} ||\hat{F}_n - F_n||,$$

where,

$$||\hat{F}_n - F_n|| = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_n(x)|.$$

Thus,

$$\int_{\delta}^{1-\delta} |J_n(u)||\hat{F}_n^{-1}(u) - F_n^{-1}(u)| \, du < 2M_\delta n^{1/2} ||\hat{F}_n - F_n||.$$
The second term of (4.66) satisfies

\[ \delta \int_0^\infty |J_n(u)| \left| \mathcal{F}^{-1}_n(u) - F_n(u) \right| du < \left\{ \delta \int_0^\infty J_n^2(u) du \right\}^{1/2} \left\{ \delta \int_0^\infty \left( \mathcal{F}^{-1}_n(u) - F_n(u) \right)^2 du \right\}^{1/2}. \]  

(4.72)

Define,

\[ L_\delta = \sup_n \delta \int_0^\infty J_n^2(u) du, \]  

(4.73)

and note that

\[ \delta \int_0^\infty \left( \mathcal{F}^{-1}_n(u) - F_n(u) \right)^2 du < \int_0^\infty \left( \mathcal{F}^{-1}_n(u) - F_n(u) \right)^2 du, \]  

(4.74)

\[ < 2 \int_0^\infty \left( F_n(u) \right)^2 du + 2 \int_0^\infty \left( \mathcal{F}^{-1}_n(u) \right)^2 du, \]

\[ < 4, \]

where the second step follows from the C1-inequality, and the last step follows from (4.63) and (4.65). Thus we can bound the left hand side of (4.72) by

\[ \delta \int_0^\infty |J_n(u)| \left| \mathcal{F}^{-1}_n(u) - F_n(u) \right| du < 2L_\delta^{1/2}. \]  

(4.75)

Likewise we can bound the third term of (4.66) by
\[ \frac{1}{1-\delta} \int_{1-\delta} |J_n(u)| \left| \hat{F}_n^{-1}(u) - F_n^{-1}(u) \right| du < 2U_{\delta}^{1/2}, \quad (4.76) \]

where

\[ U_{\delta} = \sup_n \frac{1}{1-\delta} \int_{1-\delta} J_n^2(u) du. \quad (4.77) \]

Combining (4.66)-(4.77) we obtain,

\[ |T(F_n; J_n) - T(F_n; J)\] \[ \leq 2(M_n^{1/2} \| F \| - F_n \| + L_{\delta}^{1/2} + U_{\delta}^{1/2}) \]. \quad (4.78) \]

By the uniform integrability assumption we can choose \( \delta \in (0, \frac{1}{2}) \) such that,

\[ 2(U_{\delta}^{1/2} + U_{\delta}^{1/2}) < \epsilon, \quad (4.79) \]

and we have \( M_{\delta} < \infty \). Thus,

\[ |T(F_n; J_n) - T(F_n; J)\] \[ < 2M_{\delta}^{1/2} \| F \| - F_n \| + \epsilon. \quad (4.80) \]

Now by Corollary 3.2 we have that,

\[ \lim_{n \to \infty} |T(F_n; J_n) - T(F_n; J)\] \[ < \epsilon. \quad (4.81) \]

But since \( \epsilon > 0 \) was arbitrary, this implies that
We next look at some specific examples of goodness-of-fit tests which are functions of L-statistics of the form (4.58). The first example is D'Agostino's statistic, which is defined by (4.58) with a\_n,i's given by,

\[ a_{n,i} = n^{-1} \left[ i - (0.5)(n + 1) \right] , \]  

(4.82)

for \( i = 1, \ldots, n \). D'Agostino (1971) proposed this statistic, and showed that it compared favorably with the Shapiro-Wilk statistic (to be described later). D'Agostino (1971) also showed that with \( J_n \) defined by (4.59) in terms of the \( a_{n,i} \)'s of (4.82), then

\[ T(F_n; J_n) \xrightarrow{F} (2\pi^{1/2})^{-1} . \]  

(4.83)

The \( \{J_n\}_{n=1}^{\infty} \) sequence clearly satisfy the conditions of Theorem 4.1, so the theorem holds.

We next present the statistic of Shapiro and Wilk (1965), and statistics closely related to it. The Shapiro-Wilk statistic is \( T^2(F_n; J_n) \) where the \( a_{n,i} \)'s, which define the \( \{J_n\}_{n=1}^{\infty} \) sequence, are the normalized "best linear unbiased" coefficients tabulated in Sarhan and Greenberg (1956). Shapiro and Francia (1972) modified the Shapiro-Wilk statistic by using \( a_{n,i} \)'s which are the normalized expected values of the standard normal order statistics based on a sample of
size $n$. In practice the $a_{n,i}$'s are often replaced by approximations to the expected order statistics. The Blom estimates are,

$$a_{n,i} = \phi[(n + \frac{1}{4})^{-1}(i - 3/8)] . \quad (4.84)$$

The Tukey estimates are,

$$a_{n,i} = \phi[(n + 1/3)^{-1}(i - 1/3)] . \quad (4.85)$$

The van der Waerden estimates are,

$$a_{n,i} = \phi[(n + 1)^{-1}i] . \quad (4.86)$$

All of the definitions for the $a_{n,i}$'s given above produce sequences which satisfy the conditions of Theorem 4.1, so the theorem holds. In particular it is well known that,

$$T^2(F_0; J_0) \overset{P}{\underset{n\to\infty}{\to}} 1 , \quad (4.87)$$

under any of the definitions of the $a_{n,i}$'s above. Thus we know that,

$$T^2(\hat{F}_n; J_n) \overset{P}{\underset{n\to\infty}{\to}} 1 . \quad (4.88)$$

The Shapiro-Wilk statistic is a powerful and popular test of normality and its distribution, based on iid samples from a normal
distribution, has been tabled for various values of \( n \). Even though (4.88) is of interest, it does not tell us how well the actual distribution of \( T^2(F_n; J_n) \) is approximated by those tabled values. This becomes even more interesting when we note that,

\[
n^{1/2} [T^2(F_n; J_n) - 1] \overset{P}{\to} 0 ,
\]

while,

\[
n[T^2(F_n; J_n) - c_n] \overset{L}{\to} \sum_{k=3}^{\infty} k^{-1}(z_k^2 - 1) ,
\]

where \( \{z_k\}_{k=3}^{\infty} \) is an iid sequence of \( N(0, 1) \) random variables, and \( \{c_n\}_{n=1}^{\infty} \) are a sequence of positive constants such that \( c_n = o(\log n) \).

See Gregory (1977), de Wet and Venter (1972), and van der Watt (1969).

We will examine this question more closely in Chapter Six where we examine the results of some Monte Carlo experiments.

We have concentrated on the univariate model, but the results generalize immediately to the multivariate model by applying Corollaries 3.5 and 3.6.
V. TESTS FOR AUTOCORRELATION, OUTLIERS, AND HOMOGENEITY OF VARIANCE

In the previous chapter we looked at ways of assessing the assumption of distributional normality based on measurement error regression residuals. In this chapter we examine other ways of using the regression residuals to check the assumptions of the model.

A. Tests for Autocorrelation

We begin by considering the univariate model,

\[ Y_t = \beta x_t + \epsilon_t, \quad x_t = x_t + \eta_t, \]

for \( t = 1, \ldots, n \), under the assumptions given in Section 1.B. Fuller (1986) recommends plotting the \( \hat{\epsilon}_t \)'s across time to assess the model assumptions. In particular, such a plot should show no autocorrelation. To test the hypothesis that the \( \hat{\epsilon}_t \)'s have zero autocorrelation, we show below that we can regress the \( \hat{\epsilon}_t \)'s against lagged values of the \( \hat{\epsilon}_t \)'s and test for zero coefficients in the regression.

For convenience we examine the case of regressing \( \hat{\epsilon}_t \) against the first \( p \)-lags. It will be clear that the procedure generalizes to other situations including non-consecutive lags. First, define

\[ \hat{\eta}_t = (\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2}, \ldots, \hat{\epsilon}_{t-p})', \quad (5.1) \]
for \( t = p+1, \ldots, n \). Then the regression coefficient of \( \hat{v}_t \) on \( \hat{W}_t \), with an intercept in the model is given by

\[
\hat{\gamma} = [n \sum_{t=p+1}^{n} (\hat{W}_t - \bar{W})(\hat{W}_t - \bar{W})']^{-1} \sum_{t=p+1}^{n} (\hat{W}_t - \bar{W})\hat{v}_t ,
\]

(5.2)

where

\[
\bar{W} = (n - p)^{-1} \sum_{t=p}^{n} \hat{W}_t .
\]

From Chapter Two we can write,

\[
(v_t - \bar{v}) = (v_t - \bar{v})(1 + a_n) + (x_t - x)B_n ,
\]

(5.3)

where \( a_n = O_p(n^{-1/2}) \), and \( B_n \) is a \( 1 \times k \) random vector such that \( B_n = O_p(n^{-1/2}) \). The expression holds for the lagged values and, also, for deviations from the mean when the mean \( \bar{v} \) is calculated over only \( (n - p) \) of the observations for \( p \) fixed. For example, expression (5.3) holds for

\[
[v_{t-j} - \bar{v}_{(-j)}] = [v_{t-j} - \bar{v}_{(-j)}](1 + a_n) + [x_{t-j} - x_{(-j)}]B_n ,
\]

(5.4)

where

\[
\bar{v}_{(-j)} = (n - p)^{-1} \sum_{t=p+1}^{n} v_{t-j} \quad \text{for} \quad j = 1, \ldots, p ,
\]
with the analogous definition for \( \tilde{v}(-j) \) and \( \tilde{x}(-j) \). First we examine expressions of the form,

\[
\sum_{t=p+1}^{n-1} \frac{1}{n} \prod_{j=1}^{p} \left[ \hat{v}_{t-j} - \tilde{v}(-j) \right] (\hat{v}_t - \tilde{v}) \quad \text{for} \quad j = 1, \ldots, p, \quad (5.5)
\]

where

\[
\tilde{v} = (n-p)^{-1} \sum_{t=p+1}^{n} \hat{v}_t.
\]

We can write,

\[
\sum_{t=p+1}^{n-1} \frac{1}{n} \prod_{j=1}^{p} \left[ \hat{v}_{t-j} - \tilde{v}(-j) \right] (\hat{v}_t - \tilde{v}) = \sum_{i=1}^{4} \hat{M}_i, \quad (5.6)
\]

where the \( \hat{M}_i \)'s are given below.

\[
\hat{M}_1 = (1 + a_n)^2 n^{-1} \sum_{t=p+1}^{n} \prod_{j=1}^{p} \left[ \hat{v}_{t-j} - \tilde{v}(-j) \right] (\hat{v}_t - \tilde{v}), \quad (5.7)
\]

\[
\hat{M}_2 = (1 + a_n) n^{-1} \sum_{t=p+1}^{n} \prod_{j=1}^{p} \left[ \hat{v}_{t-j} - \tilde{v}(-j) \right] (\hat{x}_t - \tilde{x}) B_n,
\]

\[
\hat{M}_3 = (1 + a_n) n^{-1} \sum_{t=p+1}^{n} \prod_{j=1}^{p} (\hat{v}_t - \tilde{v}) (\hat{x}_{t-j} - \tilde{x}(-j)) B_n,
\]

\[
\hat{M}_4 = n^{-1} \sum_{t=p+1}^{n} \prod_{j=1}^{p} [(\hat{x}_t - \tilde{x}) B_n] (\hat{x}_{t-j} - \tilde{x}(-j)) B_n.
\]
By Lemma 3.2 and Holder's inequality we have $M_4 = O_p(n^{-1/2})$, and the independence of $v_t$ and the $x_t$'s gives $M_2 = o_p(n^{-1/2})$, $M_3 = o_p(n^{-1/2})$. Also,

$$M_1 = n^{-1} \sum_{t=p+1}^{n} v_{t-j} v_t + o_p(n^{-1/2}). \quad (5.8)$$

Combining this together we have

$$\hat{x} = \sigma_{vv} n^{-1} \sum_{t=p+1}^{n} W_t v_t + o_p(n^{-1/2}), \quad (5.9)$$

where,

$$W_t = (v_{t-1}, \ldots, v_{t-p})',$$

and we have used

$$(n-p)^{-1} \sum_{t=p+1}^{n} (\hat{W}_t - \bar{W})(\hat{W}_t - \bar{W})' = \sigma_{vv} I_p + o_p(n^{-1/2}). \quad (5.10)$$

Thus, it follows from Theorem 8.2.1 of Fuller (1976) that

$$n^{1/2} \hat{x} \xrightarrow{L} N(0, I_p). \quad (5.11)$$

Since $\hat{x} = O_p(n^{-1/2})$, and because of expression (5.10), the residual mean square from the regression of $v_t$ on $\hat{W}_t$ equals
\[(n - p)^{-1} \sum_{t=p+1}^{n} (v_t - \bar{v})^2 + o_p(1), \quad (5.12)\]

which converges to \( \sigma_{vv} \).

Expressions (5.10), (5.11), and (5.12) imply that the regression "t-tests" and "F-tests" for the elements of \( \hat{\chi} \) converge to \( N(0, 1) \) and chi-squared random variables respectively. Thus, in large samples, such tests can be used to test the hypothesis of zero autocorrelation of the \( v_t \)'s.

B. Tests for Outliers

We consider the univariate model,

\[ Y_t = \beta_t + e_t, \quad X_t = x_t + u_t, \]

for \( t = 1, \ldots, n \), under the assumptions given in Section B of Chapter One. In practice, some of the regression residuals may be of large absolute value. This may indicate a defect in the model, or that an error was made in recording an observation. Thus, it is useful to have a test on residuals for outliers. We construct an outlier test based on the regression residual with the largest absolute deviation.

From Chapter Two we know that the standardized residuals can be written as

\[ s_{vv}^{-1/2} (v_t - \bar{v}) = \frac{1}{n} (v_t - \bar{v})(1 + a_n) + (x_t - x)B_n, \quad (5.13) \]
for \( t = 1, \ldots, n \), where \( a_n = o_p(n^{-1/2}) \) and \( B_n \) is a \( 1 \times k \) random vector such that \( B_n = O_p(n^{-1/2}) \). Thus we can write,

\[
\left| s_{vv}^{-1/2} (v_t - \Bar{v}) - m_{vv}^{-1/2} (v_t - \Bar{v}) \right| < \left| m_{vv}^{-1/2} (v_t - \Bar{v}) a_n \right| + \left| (x_t - \bar{x}) B_n \right|, \tag{5.14}
\]

and further write,

\[
\max_{1 \leq t \leq n} \left| s_{vv}^{-1/2} (v_t - \Bar{v}) - m_{vv}^{-1/2} (v_t - \Bar{v}) \right| < \max_{1 \leq t \leq n} \left| m_{vv}^{-1/2} (v_t - \Bar{v}) a_n \right| + \max_{1 \leq t \leq n} \left| (x_t - \bar{x}) B_n \right|. \tag{5.15}
\]

Since the \( v_t \)'s are iid \( N(0, \sigma_{vv}) \) with \( \sigma_{vv} > 0 \) we know that

\[
\max_{1 \leq t \leq n} |v_t| = o_p((\log n)^{1/2}) \tag{5.16}
\]

Therefore

\[
\max_{1 \leq t \leq n} \left| m_{vv}^{-1/2} (v_t - \Bar{v}) a_n \right| = o_p((n^{-1/2} \log n)^{1/2}) \tag{5.17}
\]

To examine the behavior of

\[
\max_{1 \leq t \leq n} \left| (x_t - \bar{x}) B_n \right|
\]
we need the following two lemmas.

**Lemma 5.1.** Let \( \{a_i\}_{i=1}^{\infty} \) be a sequence of real numbers such that \( n^{-1} \sum_{i=1}^{n} a_i + A < \infty \). Then

\[
\max_{1 \leq i \leq n} |a_i| = o(n).
\]

**Proof.** Define \( S_n = \sum_{i=1}^{n} a_i \). By assumption

\[
n^{-1}S_n - (n-1)^{-1}S_{n-1} + 0 \text{ as } n \to \infty.
\]

But,

\[
n^{-1}S_n - (n-1)^{-1}S_{n-1} = n^{-1}[a_n - (n-1)^{-1}S_{n-1}],
\]

which implies that \( n^{-1}a_n + 0 \). This also implies that \( n^{-1}|a_n| + 0 \).

Let \( \varepsilon > 0 \) be given, then there exists an \( N_0 \) such that \( n^{-1}|a_n| < \varepsilon \) for all \( n > N_0 \). This implies that \( m^{-1}|a_n| < \varepsilon \) for all \( m > n > N_0 \). Now choose \( N_1 \) such that \( n^{-1} \max_{1 \leq i \leq N_0} |a_i| < \varepsilon \). Thus \( n^{-1} \max_{1 \leq i \leq n} |a_i| < \varepsilon \) for all \( n > N_1 \). Since \( \varepsilon > 0 \) was arbitrary, the proof is complete. \( \square \)

**Lemma 5.2.** Let \( \{b_i\}_{i=1}^{\infty} \) be a sequence of real numbers such that \( n^{-1} \sum_{i=1}^{n} b_i^2 + B < \infty \). Then

\[
\max_{1 \leq i \leq n} |b_i - \overline{b}_n| = o(n^{1/2}),
\]
where \( \bar{b}_n = n^{-1} \sum_{i=1}^{n} b_i \).

**Proof.** By Lemma 5.1 we know that \( \max_{1 \leq i \leq n} b_i^2 = o(n) \), which implies that \( \max_{1 \leq i \leq n} |b_i| = o(n^{1/2}) \). Note that,

\[
\max_{1 \leq i \leq n} |b_i - \bar{b}_n| < \max_{1 \leq i \leq n} |b_i| + |\bar{b}_n|.
\]

The result now follows since

\[
|\bar{b}_n| < (n^{-1} \sum_{i=1}^{n} b_i^2)^{1/2} = o(1).
\]

By Theorem 2.1 we know that

\[
n^{-1} \sum_{t=1}^{n} x_t^2 - \sum_{t=1}^{n} x_t^2 + L_2 + \varepsilon_{nn} \text{ a.s.}
\]

By applying Lemma 5.2 we get

\[
\max_{1 \leq t \leq n} |\bar{x}_{ti} - x_i| = o(n^{1/2}) \text{ a.s.,}
\]

for \( i = 1, \ldots, k \). This implies that

\[
\max_{1 \leq t \leq n} |(x_t - \bar{x})\varepsilon_n| = o_p(1).
\]

Thus,
Now we need the following lemma.

**Lemma 5.3.** Let \( \{a_i\}_{i=1}^n \) and \( \{b_i\}_{i=1}^n \) be two sequences of real numbers. Then,

\[
\max_{1 \leq i \leq n} |a_i| - \max_{1 \leq i \leq n} |b_i| < \max_{1 \leq i \leq n} |a_i - b_i|.
\]

**Proof.** We can write,

\[
|a_i| = |b_i + (a_i - b_i)| < |b_i| + |a_i - b_i|,
\]

which gives,

\[
\max_{1 \leq i \leq n} |a_i| < \max_{1 \leq i \leq n} |b_i| + \max_{1 \leq i \leq n} |a_i - b_i|.
\]

Reversing the roles of \( a_i \) and \( b_i \) yields the desired result. \( \square \)

Now we can apply the order in probability result equivalent to Lemma 5.3 to expression (5.14) to obtain,

\[
\max_{1 \leq t \leq n} \left| s_{vv}^{-1/2} (\hat{v}_t - \bar{v}) - m_{vv}^{-1/2} (v_t - \bar{v}) \right| = o_p(1). \quad (5.18)
\]

From the theory of order statistics it is known that the random variable,
converges to a particular extreme value distribution where,

\[ c_n = (2 \log n)^{1/2} - (0.5)(2 \log n)^{1/2} (\log \log n + \log 4\pi) . \]

See David (1981). The result in (5.18) is not strong enough to imply that,

\[ (2 \log n)^{1/2} \left[ \max_{1 \leq t \leq n} \left| s_{\nu \nu}^{-1/2} (\hat{\nu}_t - \bar{\nu}) \right| - c_n \right] , \tag{5.20} \]

has the same limiting distribution as in (5.19). Expression (5.20) would hold if the \( x_t \)'s satisfied a slightly stronger moment condition namely,

\[ n^{-1} \sum_{t=1}^{n} |x_{t1}|^{2+\delta} = O(1) \text{ a.s. for some } \delta > 0 \]

for \( i = 1, \ldots, k \). In this case the \( o_p(1) \) remainder in (5.18) could be replaced by \( o_p(n^{-\lambda}) \) where \( \lambda = \delta(4 + 2\delta)^{-1} \). We summarize these results in the following theorem.

**Theorem 5.1.** Let the univariate model of Chapter One hold. Then expression (5.18) holds. If in addition

\[ n^{-1} \sum_{t=1}^{n} |x_{t1}|^{2+\delta} = O(1) \text{ a.s. for some } \delta > 0 \]
for \( i = 1, \ldots, k \), then the random variable of (5.20) has the same limiting distribution as the random variable of (5.19).

Percentage points for the distribution of the random variable,

\[
(n - 1)^{-1/2} \max_{1 \leq t \leq n} \left| m^{-1/2} (v^*_t - \bar{v}) \right|, \tag{5.21}
\]

often called the maximum normed residual, have been tabulated by Stefansky (1972). We reproduce a version of these tables in Table 10.4 of Appendix B. Our results above indicated that in large samples we can use these tables to approximate the distribution of the random variable,

\[
(n - k)^{-1/2} \max_{1 \leq t \leq n} \left| s^{-1/2} (v^*_t - \bar{v}) \right|. \tag{5.22}
\]

This gives us an approximate outlier test for residuals from measurement error regressions, for use in large samples.

The adequacy of the approximation to the distribution of the random variable in (5.22) depends on many factors, but clearly one of the most important is the behavior of the \( x^*_t \)'s. If the \( x^*_t \)'s vary greatly about their mean then we would expect the rate of convergence in (5.18) to be very slow. On the other hand, if the \( x^*_t \)'s are bounded then the remainder in (5.18) is \( o_p \left( (n^{-1} \log n)^{1/2} \right) \). This should be considered when judging the adequacy of the distributional approximation for (5.22).
We begin by considering the univariate model,

\[ Y_t = \beta_0 + x_t \beta + e_t, \quad x_t = x_t + \epsilon_t, \]

for \( t = 1, \ldots, n \), where we have separated the intercept term from the other regression variables in \( x_t \) and \( x_t \) is a 1 \( \times (k - 1) \) vector. We assume that \( m_{xx} \rightarrow m_{xx} \) a.s., where \( m_{xx} \) is a positive definite matrix. The other model assumptions remain the same.

Fuller (1986) recommends plotting \( \hat{\nu}_t \) against the elements of the vector \( \hat{x}_t \). This is analogous to plotting the residuals against independent variables in the usual regression setting. Since the \( \hat{v}_t \)'s are converging to the \( v_t \)'s, and the \( \hat{\nu}_t \)'s are converging to the \( \nu_t \)'s, the plots should behave approximately like the theoretical plots of \( v_t \) against \( x_t \). In particular, the independence of \( v_t \) and \( \nu_t \) implies that the \( v_t \)'s should have zero mean and common variance for all \( x_t \).

In the usual regression setting, a test for homogeneity of variance can be constructed by regressing the squared residuals against the independent variables and an intercept, and testing for zero coefficients. This is closely related to a statistic commonly attributed to Anscombe (1961). The limiting distribution of Anscombe's statistic was investigated by Bickel (1978) under a general linear model with bounded mean function. Bickel noted that Anscombe's statistic is closely related to the locally best test of homogeneity of variance, under the
normal model, versus a particular class of alternatives. We show below
that the regression of \((\hat{v}_t - \bar{v})^2\) on \(\hat{x}_t\) yields a test of homogeneity
of variance for the \(v_t\)'s.

From Chapter Two we know that we can write,

\[(\hat{v}_t - \bar{v}) = (\hat{v}_t - \bar{v})(1 + a_n) + (\hat{x}_t - \bar{x})B_n, \quad (5.23)\]

for \(t = 1, \ldots, n\), where \(a_n = O_p(n^{-1/2})\) and \(B_n\) is a \((k - 1) \times 1\)
random vector such that \(B_n = O_p(n^{-1/2})\). We can also write,

\[(\hat{x}_t - \bar{x}) = (\hat{x}_t - \bar{x})(I_{k-1} + C_n) + (\hat{v}_t - \bar{v})d_n, \quad (5.24)\]

for \(t = 1, \ldots, n\) where \(C_n\) is a random \((k - 1) \times (k - 1)\) matrix,
\(C_n = O_p(n^{-1/2})\), \(d_n\) is a \(1 \times (k - 1)\) random vector, and
\(d_n = O_p(n^{-1/2})\).

Let \(\hat{\gamma}\) be the regression coefficient vector from the regression of
\((\hat{v}_t - \bar{v})^2\) on \(\hat{x}_t\) with an intercept. Then,

\[\hat{\gamma} = m_{xx}^{-1}(n - 1)^{-1} \sum_{t=1}^{n} (\hat{x}_t - \bar{x})'(\hat{v}_t - \bar{v})^2. \quad (5.25)\]

Noting that

\[(\hat{v}_t - \bar{v})^2 = (\hat{v}_t - \bar{v})^2(1 + a_n)^2 + [(\hat{x}_t - \bar{x})B_n]^2 + 2(1 + a_n)(\hat{v}_t - \bar{v})(\hat{x}_t - \bar{x})B_n, \quad (5.26)\]
we can write,

\[(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{x}_t - \mathbf{\bar{x}})'(\mathbf{v}_t - \mathbf{\bar{v}})^2 = 6 \sum_{i=1}^{6} M_i, \quad (5.27)\]

where

\[
M_1 = (1 + a_n)^2 d_n'(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{v}_t - \mathbf{\bar{v}})^3, \quad (5.28) \]

\[
M_2 = 2(1 + a_n) d_n'(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{v}_t - \mathbf{\bar{v}})^2 (\mathbf{x}_t - \mathbf{\bar{x}}) B_n, \]

\[
M_3 = d_n'(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{v}_t - \mathbf{\bar{v}})[(\mathbf{x}_t - \mathbf{\bar{x}}) B_n]^2, \]

\[
M_4 = (1 + a_n)^2 (I_{k-1} + C_n)(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{v}_t - \mathbf{\bar{v}})^2 (\mathbf{x}_t - \mathbf{\bar{x}})' , \]

\[
M_5 = (I_{k-1} + C_n)(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{x}_t - \mathbf{\bar{x}})' [(\mathbf{x}_t - \mathbf{\bar{x}}) B_n]^2 , \]

\[
M_6 = 2(1 + a_n) (I_{k-1} + C_n)(n - 1)^{-1} \sum_{t=1}^{n} (\mathbf{v}_t - \mathbf{\bar{v}})(\mathbf{x}_t - \mathbf{\bar{x}})'(\mathbf{x}_t - \mathbf{\bar{x}}) B_n. \]

The behavior of the first five of the $M_i$ is described by the following lemma.

**Lemma 5.4.** Under our assumptions $M_1$, $M_2$, $M_3$, and $M_5$ are all $o_p(n^{-1/2})$ and

\[
M_4 = n^{-1} \sum_{t=1}^{n} (\mathbf{x}_t - \mathbf{\bar{x}})' v_t^2 + o_p(n^{-1/2}).
\]
Proof. That \( M_1 = o_p(n^{-1/2}) \) follows from the fact that 
\( \mathbb{E}(v_i^2) = 0 \) under normality. For \( M_2 \), note that

\[
|M_2| < 2|1 + a_n||d_n^t| \max_{1 \leq t \leq n} (v - \bar{v})^2(n - 1)^{-1} \sum_{t=1}^{n} |(x^*_t - \bar{x})b_n|,
\]

where \(|\cdot|\) of a vector is defined to be elementwise absolute value in comparison. By normality,

\[
\max_{1 \leq t \leq n} (v - \bar{v})^2 = o_p(\log n),
\]

and, by Lemma 3.2,

\[
(n - 1)^{-1} \sum_{t=1}^{n} |(x^*_t - \bar{x})b_n| = o_p(n^{-1/2}).
\]

Thus, \( M_2 = o_p(n^{-1/2} \log n) \) and the result follows. For \( M_3 \), note that,

\[
|M_3| < |d_n^t| \max_{1 \leq t \leq n} (v - \bar{v})(n - 1)^{-1} \sum_{t=1}^{n} |(x^*_t - \bar{x})b_n|^2.
\]

By normality,

\[
\max_{1 \leq t \leq n} |v - \bar{v}| = o_p([\log n]^{1/2}),
\]

and by Lemma 3.2,
Thus, $M_3 = O_p [(n^{-3} \log n)^{1/2}]$ and the result follows. For $M_3$, we can write

$$|M_3| < I_{k-1} + C_n \max_{1 \leq t \leq n} |(\bar{x}_t - \bar{x})| (n - 1)^{-1} \sum_{t=1}^{n} |(\bar{x}_t - \bar{x})_n|^{2}, \quad (5.35)$$

where the "max" expression is meant to mean the elementwise maximum. By expression (5.32) and Lemma 5.2 it follows that $M_3 = o_p(n^{-1/2})$.

For $M_4$, we note that

$$(n - 1)^{-1} \sum_{t=1}^{n} (\bar{x}_t - \bar{x})^t (v_t - v)^2 = (n - 1)^{-1} \sum_{t=1}^{n} (\bar{x}_t - \bar{x})^t v_t^2 - 2\bar{x}_v x_v. \quad (5.36)$$

The result for $M_4$ now follows from the fact that $m_{xv} = o_p(1)$. □

In order to deal with the $M_6$ term of expression (5.28) we need the following lemma.

**Lemma 5.4.** Let $\{Z_i\}_{i=1}^{\infty}$ be a sequence of uncorrelated $(0, \sigma^2)$ random variables independent of the sequence of random variables $\{w_i\}_{i=1}^{\infty}$. Let $n^{-1} \sum_{i=1}^{n} w_i + L \leq a.s.$ and $n^{-1} \sum_{i=1}^{n} |w_i| = O(1)$ a.s. Then

$$n^{-1} \sum_{i=1}^{n} w_i Z_i = o_p(1).$$
Proof. Fix a sequence \( \{w_i\}_{i=1}^{\infty} \) such that \( n^{-1} \sum_{i=1}^{n} w_i = L \) and \( n^{-1} \sum_{i=1}^{n} |w_i| = O(1) \). Then conditional on that sequence

\[ E(n^{-1} \sum_{i=1}^{n} w_i Z_i) = 0, \quad \text{and} \quad V(n^{-1} \sum_{i=1}^{n} w_i Z_i) = \sigma^2 n^{-2} \sum_{i=1}^{n} w_i^2, \]

for all \( n \). Note that,

\[ n^{-2} \sum_{i=1}^{n} w_i^2 < \{n^{-1} \sum_{i=1}^{n} |w_i|\} n^{-1} \max_{1 \leq i \leq n} |w_i|. \]

Thus by Lemma 5.1

\[ V(n^{-1} \sum_{i=1}^{n} w_i Z_i) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \]

Since the sequence \( \{w_i\}_{i=1}^{\infty} \) occurs with probability one, the result follows.

The expression for \( M_6 \) can be written

\[ M_6 = 2(1 + a_n) (I_{k-1} + c_n) M_{vxx} B_n, \quad (5.37) \]

where,

\[ (M_{vxx})_{ij} = (n - 1)^{-1} \sum_{t=1}^{n} (v_t - \bar{v})(x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j). \quad (5.38) \]

Our goal is to show that \( M_{vxx} = o_p(1) \), which would imply that

\[ M_6 = o_p(n^{-1/2}). \]

We will show that \( (M_{vxx})_{ij} = o_p(1) \) for each \((i, j)\).
pair. For a given \((i, j)\) pair define,

\[
  w_{tij} = (\bar{x}_t - \bar{x}_i)(\bar{x}_t - \bar{x}_j),
\]

for \(t = 1, \ldots, n\). The \(\{w_{tij}\}_{t=1}^n\) satisfy the conditions of Lemma 5.4 since \(\overrightarrow{\text{xx}} + \overrightarrow{\text{xx}} = \overrightarrow{\text{xx}} + \sum_{\text{nn}}\) a.s. and

\[
  n^{-1} \sum_{t=1}^n |(\bar{x}_t - \bar{x}_i)(\bar{x}_t - \bar{x})| < [(\overrightarrow{\text{xx}}_{11})^{1/2} [(\overrightarrow{\text{xx}}_{jj})^{1/2},
\]

which is bounded almost surely. It follows that

\[
  (n - 1)^{-1} \sum_{t=1}^n w_{tij} = o_p(1),
\]

Since

\[
  \overrightarrow{\nu}(n - 1)^{-1} \sum_{t=1}^n (\bar{x}_t - \bar{x}_i)(\bar{x}_t - \bar{x}_j) = o_p(n^{-1/2}),
\]

it follows that \((\overrightarrow{\text{xx}}_{ij}) = o_p(1)\) and \(\overrightarrow{\eta} = o_p(n^{-1/2})\).

Now we return to the regression coefficient \(\widehat{\chi}\) of expression (5.25). Using Lemma 5.4 we can write,

\[
  \widehat{\chi} = \overrightarrow{\text{xx}}^{-1} \sum_{t=1}^n (\bar{x}_t - \bar{x})' \nu_t^2 + o_p(n^{-1/2}),
\]

where we have used the fact that,
Fix a sequence \( \{x_t\}_{t=1}^\infty \) such that \( \mu_{m} + \mu_{m} \). Such a sequence occurs with probability one. Let \( \lambda \in \mathbb{R}^{k-1} \), where \( \lambda \neq 0 \), be given. Let

\[
T_n(\lambda) = \lambda^\top \mu_{m}^{-1} \sum_{t=1}^n (x_t - x)'v_t = \sum_{t=1}^n c_{n,t}(v_t' - v) ,
\]

where

\[
c_{n,t} = \lambda^\top \mu_{m}^{-1}(x_t - x)' ,
\]

and we have defined \( m_{ij} = (\mu_{m}^{-1})_{ij} \) for notational convenience. Now

\[
n^{-1} \sum_{t=1}^n c_{n,t}^2 = \lambda^\top \mu_{m}^{-1}(n^{-1} \sum_{t=1}^n (x_t - x)'(x_t - x))\mu_{m}^{-1} + \lambda^\top \mu_{m}^{-1} .
\]

Also

\[
\max_{1 \leq t \leq n} |c_{n,t}| < \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \lambda_i m_{ij} \max_{1 \leq t \leq n} \left| x_{t,j} - x_j \right| .
\]
Thus, by Lemma 5.2,

$$\max_{1 \leq t \leq n} |c_{n,t}| = o(n^{1/2}).$$

Now we need the following lemma.

**Lemma 5.5.** Let \( \{e_t\}_{t=1}^{\infty} \) be an iid sequence of random variables with \( \mathbb{E}\{e_1\} = 0 \) and \( \mathbb{V}\{e_1\} = \sigma^2 \). Let \( \{c_{n,t}: 1 \leq t < n, n > 1\} \) be a triangular array of real numbers such that \( n^{-1} \sum_{t=1}^{n} c_{n,t}^2 + C^2 < \infty \) where \( C^2 > 0 \), and \( \max_{1 \leq t \leq n} |c_{n,t}| = o(n^{1/2}) \). Then,

$$n^{-1/2} \sum_{t=1}^{n} c_{n,t} e_t \xrightarrow{L} N(0, \sigma^2 C^2).$$

**Proof.** Let \( \varepsilon > 0 \) be given, and define \( s_n^2 = \sigma^2 n^{-1} \sum_{t=1}^{n} c_{n,t}^2 \). Then

$$s_n^{-2} \sum_{t=1}^{n} c_{n,t}^2 \mathbb{E}\{e^2: |e| > \varepsilon s_n n^{1/2} |c_{n,t}|\} \leq \mathbb{E}\{e^2: |e| > \varepsilon s_n (n^{-1/2} \max_{1 \leq t \leq n} |c_{n,t}|)^{-1}\}.$$

By assumption,

$$\varepsilon s_n (n^{-1/2} \max_{1 \leq t \leq n} |c_{n,t}|)^{-1} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Because the \( e_t \)'s are iid with finite variance, the right hand side of (5.44) converges to zero as \( n \rightarrow \infty \). Thus the triangular array
\[ \{n^{-1/2} c_{n,t} e_t: 1 < t < n, n \to \infty\} \text{satisfies the conditions of the Lindeberg central limit theorem, which implies that} \]

\[ s_n^{-1/2} \sum_{t=1}^{n} c_{n,t} e_t \xrightarrow{L} N(0, 1), \]

which implies that

\[ n^{-1/2} \sum_{t=1}^{n} c_{n,t} e_t \xrightarrow{L} N(0, \sigma^2 C^2). \]

Since the \( c_{n,t} \)'s given in (5.43) satisfy the conditions of Lemma 5.5, and since \( V(v_v^2) = 2\sigma^2 v_v \) it follows that

\[ T_n(\lambda) \xrightarrow{L} N(0, 2\sigma^2 \lambda \lambda^\top_{xx}). \quad (5.45) \]

Since (5.45) holds for an arbitrary non-null \( \lambda \), we know, by the Cramer-Wald device that conditional on \( \{x_t\}_{t=1}^{\infty} \),

\[ n^{1/2} x \xrightarrow{L} N(0, 2\sigma^2 \lambda \lambda^\top_{xx}). \quad (5.46) \]

But since this limiting distribution is attained for almost all sequences \( \{x_t\}_{t=1}^{\infty} \), this implies that,

\[ n^{1/2} x \xrightarrow{L} N(0, 2\sigma^2 \lambda \lambda^\top_{xx}), \quad (5.47) \]

unconditionally as well.
Our goal is to show that the regression "t-tests" and "F-tests" for the elements of \( \hat{\chi} \) converge to \( N(0, 1) \) and chi-square random variables, respectively. The proof will be complete if we can show that the residual mean square from the regression of \((\bar{\tau}_t - \bar{\tau})^2\) on \( \bar{x}_t \) and an intercept converges to \( 2\sigma^2_{Vv} \). Since \( \hat{\chi} \xrightarrow{P} 0 \), it is enough to show that,

\[
n^{-1} \sum_{t=1}^{n} \frac{[(\hat{\tau}_t - \bar{\tau})^2 - s_{VV}]^2}{n} \xrightarrow{P} 2\sigma^2_{Vv}, \tag{5.48}
\]

or equivalently,

\[
n^{-1} \sum_{t=1}^{n} \frac{[s_{VV}^{-1}(\hat{\tau}_t - \bar{\tau})^2 - 1]^2}{n} \xrightarrow{P} 2, \tag{5.49}
\]

since \( s_{VV} \xrightarrow{P} \sigma_{VV} \). By the results of Chapter Two we can write,

\[
s_{VV}^{-1}(\hat{\tau}_t - \bar{\tau})^2 = n^{-1}_v\hat{\tau}_t(1 + a_n)^2 + [(\bar{x}_t - \bar{\tau})B_n]^2 + 2(1 + a_n)m^{-1/2}_v(\tau - \bar{\tau})(\bar{x}_t - \bar{\tau})B_n, \tag{5.50}
\]

where \( a_n = o_p(n^{-1/2}) \), and \( B_n \) is a \((k - 1)\times 1\) random vector such that \( B_n = o_p(n^{-1/2}) \). Now we can write,

\[
n^{-1} \sum_{t=1}^{n} \frac{[s_{VV}^{-1}(\hat{\tau}_t - \bar{\tau})^2 - 1]^2}{n} = \frac{6}{\Sigma \Sigma_i M_i - n^{-1}k}, \tag{5.50}
\]

where,
\[ M_1' = (1 + \alpha_n)^{m_{n-1}} \sum_{t=1}^{n} (v_t - \bar{v})^4, \]  
\[ M_2' = 2(1 + \alpha_n)^{2m_{n-1}} \sum_{t=1}^{n} (v_t - \bar{v})^2[(x_t - \bar{x})B_n]^2, \]  
\[ M_3' = 4(1 + \alpha_n)^{3m_{n-1}} \sum_{t=1}^{n} (v_t - \bar{v})^3[(x_t - \bar{x})B_n], \]  
\[ M_4' = n^{-1} \sum_{t=1}^{n} \left[(x_t - \bar{x})B_n\right]^4, \]  
\[ M_5' = 4(1 + \alpha_n)^{m_{n-1}} \sum_{t=1}^{n} (v_t - \bar{v})[(x_t - \bar{x})B_n]^3, \]  
\[ M_6' = 4(1 + \alpha_n)^{2m_{n-1}} \sum_{t=1}^{n} (v_t - \bar{v})^2[(x_t - \bar{x})B_n]^2. \]

By Lemma 3.2, and the fact that
\[ \max_{1 \leq t \leq n} |v_t - \bar{v}|^v = o_p[(\log n)^{v/2}], \]  
for \( v > 1 \) we have that \( M_2' \), \( M_3' \), \( M_4' \), \( M_5' \), and \( M_6' \) are all \( o_p(1) \). By the properties of the normal distribution,
\[ m_{n-1} \sum_{t=1}^{n} (v_t - \bar{v})^4 \xrightarrow{P} 2. \]

Thus expression (5.48) is true. We summarize our results in the following theorem.

**Theorem 5.2.** Let the univariate model of Chapter One hold. Then the regression "t-statistics" and "F-statistics" from the regression of
\[(v_t - \bar{v})^2 \text{ on } \hat{x}_t \text{ and an intercept converge to } N(0, 1) \text{ and chi-square random variables, respectively.} \]

We have concentrated on the univariate model, but we can also test for homogeneity of variance in the multivariate case. We write the model as,

\[ Y_t = \beta_0 + \beta_1 x_t + e_t, \quad \hat{x}_t = \hat{x}_t + a_t, \]

for \( t = 1, \ldots, n \), where the intercept term is separated from the other regression variables as was done above in the univariate case.

The residual vector \( \hat{v}_t \) is defined by

\[ \hat{v}_t = Y_t - \hat{\beta}_0 - \hat{x}_t \hat{\beta}_1, \quad (5.53) \]

for \( t = 1, \ldots, n \), where \( \hat{v}_t = (\hat{v}_{t1}, \ldots, \hat{v}_{tn}) \). The same proof used in the univariate case can be used to show that for \( i = 1, \ldots, r \) we can regress \( (\hat{v}_{t1} - \bar{v}_i)^2 \) on \( \hat{x}_t \) and the "t-tests" and "F-tests" will be valid in large samples. It may also be of interest to regress the "cross-product" terms \( (\hat{v}_{t1} - \bar{v}_i)(\hat{v}_{tj} - \bar{v}_j) \) on \( \hat{x}_t \). If we call the regression vector \( \chi_{di} \), then we have,

\[ n^{1/2} \chi_{di} \overset{L}{\to} N(0, (\sigma^2_{v_1i} + \sigma^2_{v_1j} + \sigma_{v_{ij}})^{-1} \sigma_{xx}) \]

where we have used the notation,
\[ \sigma_{vij}^2 = E[v_{ti}v_{tj}] \]  

(5.55)

for \( i, j \in \{1, \ldots, r\} \). It is also true that the residual mean square from the regression converges to,

\[ \sigma_{vij}^2 + \sigma_{vii}\sigma_{vjj} \]  

(5.56)

Thus the regression "t-tests" and "F-tests" work in this case as well.
VI. MONTE CARLO SIMULATIONS AND EXAMPLES

In the previous chapters we described several diagnostic procedures that can be used to check the assumptions of the multivariate errors-in-variables regression model. We have examined the limiting behavior of these procedures to justify their use in large samples. In practice it is important to know how well these asymptotic results hold in moderate or small samples. In this chapter we explore this question by analyzing two actual data sets, and by conducting some Monte Carlo studies based on the data sets.

A. Role Performance Data

In this section we consider some data studied by Warren, White and Fuller (1974) and also analyzed in Fuller (1986). In the original study, the responses of 98 managers of Iowa farmer cooperatives were analyzed, but, following Example 2.2.1 of Fuller (1986), we will analyze a subset of the original data containing 55 observations. The postulated model is

\[ y_t = \beta_0 + \sum_{i=1}^{4} x_{ti} \beta_i + q_t, \]

\[ Y_t = y_t + w_t, \]

\[ X_{ti} = x_{ti} + u_{ti} \text{ for } i = 1, \ldots, 4, \]
where $y_t$ is the role performance of the $t$-th manager, $x_{t1}$ is knowledge of the economic phases of management, $x_{t2}$ is value orientation, $x_{t3}$ is role satisfaction and $x_{t4}$ is past training. For a more detailed description of the variables see the references listed above. The data are presented in Table 6.1. The random variable $q_t$ represents a random equation error brought about by the failure of the linear relationship to hold exactly. It is assumed that $q_t \sim N(0, \sigma_{qq})$. The random vector $a_t = (w_t, u_{t1}, ..., u_{t4})'$ of measurement errors is assumed to satisfy $a_t \sim NI(0, \Sigma_{aa})$ and be independent of $q_t$. For this example, we consider $\{x_t\}_{t=1}^{55}$ to be a sequence of fixed vectors.

Using replicated determinations on the same individuals an estimate $S_{aa}$ of $\Sigma_{aa}$ was calculated to be,

$$S_{aa} = \text{diag}(0.0037, 0.0203, 0.0438, 0.0180, 0.0). \quad (6.2)$$

We assume that past training, which is defined to be the total years of formal schooling divided by six, is measured without error. The other variables contain errors in measurement. Depending on the particular variable, these measurement errors account for between 20% and 40% of the total variation in the observed responses.

Estimation procedures for the parameters of this model are given in Fuller (1980) and the references given above. We follow the procedure of Example 2.2.1 of Fuller (1986), but do not go into the details here. We define $v_t$ to be
### Table 6.1: Data from role performance study (Fuller, 1986)

<table>
<thead>
<tr>
<th>Observation</th>
<th>Knowledge orientation</th>
<th>Role satisfaction</th>
<th>Past training</th>
<th>Role performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.193</td>
<td>2.656</td>
<td>2.333</td>
<td>2.000</td>
</tr>
<tr>
<td>2</td>
<td>1.654</td>
<td>3.300</td>
<td>2.320</td>
<td>2.000</td>
</tr>
<tr>
<td>3</td>
<td>1.193</td>
<td>2.489</td>
<td>2.737</td>
<td>2.000</td>
</tr>
<tr>
<td>4</td>
<td>1.077</td>
<td>2.478</td>
<td>2.203</td>
<td>2.000</td>
</tr>
<tr>
<td>5</td>
<td>1.539</td>
<td>2.822</td>
<td>2.840</td>
<td>2.000</td>
</tr>
<tr>
<td>6</td>
<td>1.385</td>
<td>3.000</td>
<td>2.373</td>
<td>2.000</td>
</tr>
<tr>
<td>7</td>
<td>1.462</td>
<td>3.111</td>
<td>2.497</td>
<td>2.667</td>
</tr>
<tr>
<td>8</td>
<td>1.385</td>
<td>2.545</td>
<td>2.617</td>
<td>2.167</td>
</tr>
<tr>
<td>9</td>
<td>1.539</td>
<td>2.556</td>
<td>2.997</td>
<td>2.000</td>
</tr>
<tr>
<td>10</td>
<td>1.654</td>
<td>2.945</td>
<td>2.150</td>
<td>2.167</td>
</tr>
<tr>
<td>11</td>
<td>1.462</td>
<td>2.778</td>
<td>2.227</td>
<td>1.833</td>
</tr>
<tr>
<td>12</td>
<td>1.154</td>
<td>2.545</td>
<td>2.017</td>
<td>2.167</td>
</tr>
<tr>
<td>13</td>
<td>1.424</td>
<td>3.611</td>
<td>2.303</td>
<td>2.333</td>
</tr>
<tr>
<td>14</td>
<td>1.116</td>
<td>2.956</td>
<td>2.517</td>
<td>2.333</td>
</tr>
<tr>
<td>15</td>
<td>1.270</td>
<td>2.856</td>
<td>1.770</td>
<td>2.333</td>
</tr>
<tr>
<td>16</td>
<td>1.347</td>
<td>2.956</td>
<td>2.430</td>
<td>2.000</td>
</tr>
<tr>
<td>17</td>
<td>1.116</td>
<td>2.545</td>
<td>2.043</td>
<td>2.000</td>
</tr>
<tr>
<td>18</td>
<td>1.077</td>
<td>3.356</td>
<td>2.410</td>
<td>2.000</td>
</tr>
<tr>
<td>19</td>
<td>1.423</td>
<td>3.211</td>
<td>2.150</td>
<td>2.000</td>
</tr>
<tr>
<td>20</td>
<td>0.923</td>
<td>2.556</td>
<td>2.180</td>
<td>2.000</td>
</tr>
<tr>
<td>21</td>
<td>1.385</td>
<td>2.589</td>
<td>2.490</td>
<td>2.000</td>
</tr>
<tr>
<td>22</td>
<td>1.270</td>
<td>2.900</td>
<td>1.920</td>
<td>2.333</td>
</tr>
<tr>
<td>23</td>
<td>1.116</td>
<td>2.167</td>
<td>2.663</td>
<td>1.333</td>
</tr>
<tr>
<td>24</td>
<td>1.346</td>
<td>2.922</td>
<td>2.520</td>
<td>2.000</td>
</tr>
<tr>
<td>25</td>
<td>0.846</td>
<td>1.711</td>
<td>3.150</td>
<td>1.500</td>
</tr>
<tr>
<td>26</td>
<td>1.077</td>
<td>2.556</td>
<td>2.297</td>
<td>2.000</td>
</tr>
<tr>
<td>27</td>
<td>1.231</td>
<td>3.567</td>
<td>2.307</td>
<td>2.167</td>
</tr>
<tr>
<td>28</td>
<td>0.962</td>
<td>2.689</td>
<td>2.830</td>
<td>1.333</td>
</tr>
<tr>
<td>29</td>
<td>1.500</td>
<td>2.978</td>
<td>2.737</td>
<td>2.000</td>
</tr>
<tr>
<td>30</td>
<td>1.577</td>
<td>2.945</td>
<td>3.117</td>
<td>2.167</td>
</tr>
</tbody>
</table>
Table 6.1. (continued)

<table>
<thead>
<tr>
<th>Observation</th>
<th>Knowledge</th>
<th>Value orientation</th>
<th>Role satisfaction</th>
<th>Past training</th>
<th>Role performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>1.885</td>
<td>3.256</td>
<td>2.647</td>
<td>2.667</td>
<td>0.334</td>
</tr>
<tr>
<td>32</td>
<td>1.231</td>
<td>2.956</td>
<td>2.217</td>
<td>1.667</td>
<td>-0.076</td>
</tr>
<tr>
<td>33</td>
<td>1.808</td>
<td>2.811</td>
<td>2.327</td>
<td>2.000</td>
<td>-0.043</td>
</tr>
<tr>
<td>34</td>
<td>1.039</td>
<td>2.733</td>
<td>2.447</td>
<td>2.000</td>
<td>-0.126</td>
</tr>
<tr>
<td>35</td>
<td>1.385</td>
<td>2.400</td>
<td>2.347</td>
<td>2.167</td>
<td>-0.056</td>
</tr>
<tr>
<td>36</td>
<td>1.846</td>
<td>2.944</td>
<td>2.410</td>
<td>2.000</td>
<td>0.203</td>
</tr>
<tr>
<td>37</td>
<td>1.731</td>
<td>3.200</td>
<td>2.277</td>
<td>2.000</td>
<td>0.023</td>
</tr>
<tr>
<td>38</td>
<td>1.500</td>
<td>2.911</td>
<td>2.577</td>
<td>2.000</td>
<td>0.047</td>
</tr>
<tr>
<td>39</td>
<td>1.231</td>
<td>3.167</td>
<td>2.507</td>
<td>2.333</td>
<td>0.011</td>
</tr>
<tr>
<td>40</td>
<td>1.346</td>
<td>3.322</td>
<td>2.653</td>
<td>2.500</td>
<td>0.153</td>
</tr>
<tr>
<td>41</td>
<td>1.347</td>
<td>2.833</td>
<td>2.587</td>
<td>2.667</td>
<td>0.100</td>
</tr>
<tr>
<td>42</td>
<td>1.154</td>
<td>2.967</td>
<td>3.140</td>
<td>2.000</td>
<td>-0.089</td>
</tr>
<tr>
<td>43</td>
<td>0.923</td>
<td>2.700</td>
<td>2.557</td>
<td>1.833</td>
<td>0.007</td>
</tr>
<tr>
<td>44</td>
<td>1.731</td>
<td>3.033</td>
<td>2.423</td>
<td>2.000</td>
<td>0.089</td>
</tr>
<tr>
<td>45</td>
<td>1.808</td>
<td>2.911</td>
<td>2.793</td>
<td>2.000</td>
<td>0.182</td>
</tr>
<tr>
<td>46</td>
<td>1.193</td>
<td>3.311</td>
<td>2.283</td>
<td>2.333</td>
<td>0.259</td>
</tr>
<tr>
<td>47</td>
<td>1.308</td>
<td>2.245</td>
<td>2.210</td>
<td>2.000</td>
<td>0.007</td>
</tr>
<tr>
<td>48</td>
<td>1.424</td>
<td>2.422</td>
<td>2.350</td>
<td>2.000</td>
<td>-0.015</td>
</tr>
<tr>
<td>49</td>
<td>1.385</td>
<td>2.744</td>
<td>2.330</td>
<td>2.000</td>
<td>-0.023</td>
</tr>
<tr>
<td>50</td>
<td>1.385</td>
<td>2.956</td>
<td>2.130</td>
<td>2.000</td>
<td>-0.150</td>
</tr>
<tr>
<td>51</td>
<td>1.347</td>
<td>2.933</td>
<td>2.837</td>
<td>2.167</td>
<td>0.152</td>
</tr>
<tr>
<td>52</td>
<td>1.539</td>
<td>3.411</td>
<td>2.600</td>
<td>2.167</td>
<td>0.377</td>
</tr>
<tr>
<td>53</td>
<td>1.385</td>
<td>1.856</td>
<td>2.790</td>
<td>2.000</td>
<td>0.043</td>
</tr>
<tr>
<td>54</td>
<td>1.654</td>
<td>3.089</td>
<td>2.500</td>
<td>2.000</td>
<td>0.184</td>
</tr>
<tr>
<td>55</td>
<td>1.308</td>
<td>2.967</td>
<td>2.813</td>
<td>2.667</td>
<td>0.127</td>
</tr>
</tbody>
</table>
for \( t = 1, \ldots, 55 \). By assumption, the \( v_t \)'s are iid normal random variables. We also define \( \hat{v}_t \), as usual, by

\[
\hat{v}_t = Y_t - \hat{\beta}_0 - \sum_{i=1}^{4} X_{ti} \hat{\beta}_i
\]

for \( t = 1, \ldots, 55 \), and \( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_4 \) are \( n^{1/2} \)-consistent estimators of \( \beta_0, \beta_1, \ldots, \beta_4 \). The definition of \( \hat{x}_t \) is given in Section 2.2 of Fuller (1986). The standardized \( \hat{v}_t \)'s and the \( \hat{x}_t \)'s are presented in Table 6.2. We next demonstrate our diagnostic procedures using these \( \hat{v}_t \)'s and \( \hat{x}_t \)'s.

Table 6.3 displays nine goodness-of-fit test statistics for normality based on the 55 values of \( \hat{v}_t \). As noted in the table, the EDF statistics were all calculated with the modification of Stephens (1974), and the significance levels are obtained from Table 10.1 of Appendix B. The choice of 11 intervals for the chi-squared statistic was made to conform to the rule of thumb of forming intervals with expected cell counts of at least 5, and the significance level was obtained from Table 10.2 of Appendix B. These first six tests are all in close agreement. All tests are close to the five percent level, but do not indicate a large deviation from normality. The last three statistics in Table 6.3 are in close agreement, and also do not indicate
Table 6.2. Residuals and expected true values from the role performance data

<table>
<thead>
<tr>
<th>Observation</th>
<th>$\frac{-1}{2} \hat{s}_{vv}$</th>
<th>$\hat{x}_1$</th>
<th>$\hat{x}_2$</th>
<th>$\hat{x}_3$</th>
<th>$\hat{x}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.096</td>
<td>1.198</td>
<td>2.660</td>
<td>2.335</td>
<td>2.000</td>
</tr>
<tr>
<td>2</td>
<td>1.679</td>
<td>1.755</td>
<td>3.391</td>
<td>2.349</td>
<td>2.000</td>
</tr>
<tr>
<td>3</td>
<td>1.052</td>
<td>1.256</td>
<td>2.546</td>
<td>2.755</td>
<td>2.000</td>
</tr>
<tr>
<td>4</td>
<td>-0.024</td>
<td>1.075</td>
<td>2.476</td>
<td>2.202</td>
<td>2.000</td>
</tr>
<tr>
<td>5</td>
<td>0.255</td>
<td>1.554</td>
<td>2.836</td>
<td>2.844</td>
<td>2.000</td>
</tr>
<tr>
<td>6</td>
<td>-0.167</td>
<td>1.374</td>
<td>2.990</td>
<td>2.370</td>
<td>2.000</td>
</tr>
<tr>
<td>7</td>
<td>0.397</td>
<td>1.485</td>
<td>3.132</td>
<td>2.503</td>
<td>2.666</td>
</tr>
<tr>
<td>8</td>
<td>-0.735</td>
<td>1.339</td>
<td>2.504</td>
<td>2.603</td>
<td>2.166</td>
</tr>
<tr>
<td>9</td>
<td>1.737</td>
<td>1.644</td>
<td>2.650</td>
<td>3.027</td>
<td>2.000</td>
</tr>
<tr>
<td>10</td>
<td>0.173</td>
<td>1.664</td>
<td>2.954</td>
<td>2.153</td>
<td>2.166</td>
</tr>
<tr>
<td>11</td>
<td>-0.237</td>
<td>1.447</td>
<td>2.764</td>
<td>2.222</td>
<td>1.833</td>
</tr>
<tr>
<td>12</td>
<td>0.344</td>
<td>1.174</td>
<td>2.563</td>
<td>2.022</td>
<td>2.166</td>
</tr>
<tr>
<td>13</td>
<td>-1.164</td>
<td>1.352</td>
<td>3.547</td>
<td>2.282</td>
<td>2.333</td>
</tr>
<tr>
<td>14</td>
<td>-0.559</td>
<td>1.081</td>
<td>2.924</td>
<td>2.506</td>
<td>2.333</td>
</tr>
<tr>
<td>15</td>
<td>2.627</td>
<td>1.429</td>
<td>2.999</td>
<td>1.816</td>
<td>2.333</td>
</tr>
<tr>
<td>16</td>
<td>0.522</td>
<td>1.378</td>
<td>2.984</td>
<td>2.439</td>
<td>2.000</td>
</tr>
<tr>
<td>17</td>
<td>0.318</td>
<td>1.134</td>
<td>2.561</td>
<td>2.048</td>
<td>2.000</td>
</tr>
<tr>
<td>18</td>
<td>-0.245</td>
<td>1.062</td>
<td>3.342</td>
<td>2.405</td>
<td>2.000</td>
</tr>
<tr>
<td>19</td>
<td>-1.304</td>
<td>1.343</td>
<td>3.139</td>
<td>2.126</td>
<td>2.000</td>
</tr>
<tr>
<td>20</td>
<td>-0.421</td>
<td>0.897</td>
<td>2.532</td>
<td>2.172</td>
<td>2.000</td>
</tr>
<tr>
<td>21</td>
<td>-1.587</td>
<td>1.287</td>
<td>2.501</td>
<td>2.461</td>
<td>2.000</td>
</tr>
<tr>
<td>22</td>
<td>0.969</td>
<td>1.328</td>
<td>2.953</td>
<td>1.937</td>
<td>2.333</td>
</tr>
<tr>
<td>23</td>
<td>0.317</td>
<td>1.134</td>
<td>2.184</td>
<td>2.668</td>
<td>1.333</td>
</tr>
<tr>
<td>24</td>
<td>0.287</td>
<td>1.363</td>
<td>2.937</td>
<td>2.525</td>
<td>2.000</td>
</tr>
<tr>
<td>25</td>
<td>-0.023</td>
<td>0.844</td>
<td>1.709</td>
<td>3.149</td>
<td>1.500</td>
</tr>
<tr>
<td>26</td>
<td>-0.316</td>
<td>1.057</td>
<td>2.538</td>
<td>2.291</td>
<td>2.000</td>
</tr>
<tr>
<td>27</td>
<td>-0.687</td>
<td>1.189</td>
<td>3.529</td>
<td>2.294</td>
<td>2.166</td>
</tr>
</tbody>
</table>
Table 6.2. (continued)

<table>
<thead>
<tr>
<th>Observation</th>
<th>$s_{vv}^{-1/2}$</th>
<th>$\hat{x}_1$</th>
<th>$\hat{x}_2$</th>
<th>$\hat{x}_3$</th>
<th>$\hat{x}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>1.961</td>
<td>1.080</td>
<td>2.796</td>
<td>2.864</td>
<td>1.333</td>
</tr>
<tr>
<td>29</td>
<td>-0.292</td>
<td>1.482</td>
<td>2.961</td>
<td>2.731</td>
<td>2.000</td>
</tr>
<tr>
<td>30</td>
<td>-1.928</td>
<td>1.459</td>
<td>2.838</td>
<td>3.082</td>
<td>2.166</td>
</tr>
<tr>
<td>31</td>
<td>-0.010</td>
<td>1.883</td>
<td>3.255</td>
<td>2.646</td>
<td>2.666</td>
</tr>
<tr>
<td>32</td>
<td>-0.397</td>
<td>1.206</td>
<td>2.933</td>
<td>2.209</td>
<td>1.666</td>
</tr>
<tr>
<td>33</td>
<td>-2.039</td>
<td>1.683</td>
<td>2.699</td>
<td>2.290</td>
<td>2.000</td>
</tr>
<tr>
<td>34</td>
<td>-0.291</td>
<td>1.020</td>
<td>2.717</td>
<td>2.441</td>
<td>2.000</td>
</tr>
<tr>
<td>35</td>
<td>-0.282</td>
<td>1.367</td>
<td>2.384</td>
<td>2.341</td>
<td>2.166</td>
</tr>
<tr>
<td>36</td>
<td>-0.164</td>
<td>1.835</td>
<td>2.935</td>
<td>2.407</td>
<td>2.000</td>
</tr>
<tr>
<td>37</td>
<td>-1.655</td>
<td>1.629</td>
<td>3.109</td>
<td>2.247</td>
<td>2.000</td>
</tr>
<tr>
<td>38</td>
<td>-0.600</td>
<td>1.463</td>
<td>2.878</td>
<td>2.566</td>
<td>2.000</td>
</tr>
<tr>
<td>39</td>
<td>-0.443</td>
<td>1.203</td>
<td>3.142</td>
<td>2.498</td>
<td>2.333</td>
</tr>
<tr>
<td>40</td>
<td>0.055</td>
<td>1.349</td>
<td>3.325</td>
<td>2.654</td>
<td>2.500</td>
</tr>
<tr>
<td>41</td>
<td>0.247</td>
<td>1.361</td>
<td>2.846</td>
<td>2.591</td>
<td>2.666</td>
</tr>
<tr>
<td>42</td>
<td>-1.404</td>
<td>1.068</td>
<td>2.889</td>
<td>3.115</td>
<td>2.000</td>
</tr>
<tr>
<td>43</td>
<td>1.307</td>
<td>1.002</td>
<td>2.771</td>
<td>2.579</td>
<td>1.833</td>
</tr>
<tr>
<td>44</td>
<td>-0.976</td>
<td>1.671</td>
<td>2.979</td>
<td>2.405</td>
<td>2.000</td>
</tr>
<tr>
<td>45</td>
<td>-0.602</td>
<td>1.770</td>
<td>2.878</td>
<td>2.782</td>
<td>2.000</td>
</tr>
<tr>
<td>46</td>
<td>2.021</td>
<td>1.315</td>
<td>3.421</td>
<td>2.319</td>
<td>2.333</td>
</tr>
<tr>
<td>47</td>
<td>0.978</td>
<td>1.367</td>
<td>2.298</td>
<td>2.227</td>
<td>2.000</td>
</tr>
<tr>
<td>48</td>
<td>-0.003</td>
<td>1.423</td>
<td>2.422</td>
<td>2.349</td>
<td>2.000</td>
</tr>
<tr>
<td>49</td>
<td>-0.371</td>
<td>1.362</td>
<td>2.724</td>
<td>2.323</td>
<td>2.000</td>
</tr>
<tr>
<td>50</td>
<td>-1.619</td>
<td>1.286</td>
<td>2.866</td>
<td>2.101</td>
<td>2.000</td>
</tr>
<tr>
<td>51</td>
<td>0.502</td>
<td>1.377</td>
<td>2.960</td>
<td>2.845</td>
<td>2.166</td>
</tr>
<tr>
<td>52</td>
<td>1.550</td>
<td>1.632</td>
<td>3.496</td>
<td>2.627</td>
<td>2.166</td>
</tr>
<tr>
<td>53</td>
<td>0.963</td>
<td>1.443</td>
<td>1.908</td>
<td>2.807</td>
<td>2.000</td>
</tr>
<tr>
<td>54</td>
<td>-0.001</td>
<td>1.653</td>
<td>3.088</td>
<td>2.499</td>
<td>2.000</td>
</tr>
<tr>
<td>55</td>
<td>0.194</td>
<td>1.319</td>
<td>2.977</td>
<td>2.816</td>
<td>2.666</td>
</tr>
</tbody>
</table>
Table 6.3. Goodness-of-fit tests based on residuals from the role performance data

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>Value</th>
<th>Significance level&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.879&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.05 &lt; p &lt; .10</td>
</tr>
<tr>
<td>Kuiper</td>
<td>1.514&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.025 &lt; p &lt; .05</td>
</tr>
<tr>
<td>Cramer von-Mises</td>
<td>0.129&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.025 &lt; p &lt; .05</td>
</tr>
<tr>
<td>Watson</td>
<td>0.126&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.025 &lt; p &lt; .05</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>0.713&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.05 &lt; p &lt; .10</td>
</tr>
<tr>
<td>Chi-squared (11 intervals)</td>
<td>14</td>
<td>.05 &lt; p &lt; .10</td>
</tr>
<tr>
<td>Shapiro-Wilk (Blom estimates)</td>
<td>0.973</td>
<td>p &gt; .10</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.302</td>
<td>p &gt; .10</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.117</td>
<td>p &gt; .10</td>
</tr>
</tbody>
</table>

<sup>a</sup>Calculated with the modification of Stephens (1974).

<sup>b</sup>With respect to the theoretical distributions.

rejection of normality. The significance levels for the skewness and kurtosis statistics were obtained by comparing their values to percentage points given in Table 34B and Table 34 of *Tables for Statisticians and Biometricians* by Pearson and Hartley (1954). The significance levels for the Shapiro-Wilk statistic were obtained from tables contained in Shapiro and Francia (1972).

The outlier test based on the v<sub>t</sub>'s yields,

\[
\max_{1 < t < 55} \left| s^{-1/2} v_t \right| = 2.627. \tag{6.5}
\]
The largest residual is associated with observation 15. The critical values in Table 10.4 of Appendix B are only given for \( n < 30 \), but since critical values increase with \( n \) it is clear that the significance level associated with the value in (6.5) is greater than .05. Thus, the data conform to the behavior predicted under the normal model.

We tested for first order autocorrelation based on the \( \hat{v}_t \)'s by regressing the \( \hat{v}_t \)'s on one lagged value and an intercept. The F-statistic for testing for a zero regression coefficient on the lagged value is given by,

\[
F = 0.147 , \quad (6.6)
\]

and is approximately distributed as \( F_{52}^1 \). This value is clearly not significant so there appears to be little autocorrelation in the \( \hat{v}_t \)'s. This test was computed only for illustration because there is no natural ordering for the observations.

We tested for homogeneity of variance by regressing \( \hat{v}_t^2 \) on \( \hat{x}_{t1}, \hat{x}_{t2}, \hat{x}_{t3}, \hat{x}_{t4} \) and an intercept. The F-statistic for testing that the regression coefficients on \( \hat{x}_{t1}, \hat{x}_{t2}, \hat{x}_{t3}, \) and \( \hat{x}_{t4} \) are all zero is given by,

\[
F = 0.950 , \quad (6.7)
\]

and is approximately distributed as \( F_{50}^4 \). This value is clearly not significant so the model assumption of homogeneous error variance can be
retained. If the F-statistic in (6.7) had been larger we would have looked at the t-statistics for the individual regression coefficients, but there is no need to examine them here.

We have demonstrated the use of the diagnostic procedures on the role performance data, and found no reason to reject the assumptions of the model. We next present the results of a Monte Carlo study based on these data.

The purpose of our Monte Carlo study is to determine how well the asymptotic theory of earlier chapters performs in moderate to small samples. In particular we will examine the performance of the procedures under the null hypothesis that model (1.6) of Chapter One holds with Gaussian errors.

Our Monte Carlo study was conducted in order to simulate "an actual data set" as much as possible. By "actual data set" we mean a data set we might encounter in practice. In order to do this we simulated data in accordance with mode (6.1) and sample size $n = 55$ to match the role performance data. To be specific, at each stage of Monte Carlo replication we generated,

$$y_{t}^* = \hat{\beta}_0 + \sum_{i=1}^{4} x_{ti} \hat{\beta}_i + q_{t}^*,$$

$$Y_{t}^* = y_{t}^* + w_{t}^*,$$

$$X_{t}^* = x_{ti} + u_{ti}^* \quad \text{for } i = 1, \ldots, 4,$$
for \( t = 1, \ldots, 55 \), where \( (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_4) \) were the estimates calculated from the original data, and \( \hat{x}_{t_i} \) for \( i = 1, \ldots, 4 \) are the estimated true-\( x \) values given in Table 6.2 also based on the original data. In addition, \( q^*_t \) for \( t = 1, \ldots, 55 \) were \( \text{NI}(0, \sigma_{qq}^2) \) random variables where \( \sigma_{qq}^2 \) is an estimate of \( \sigma_{qq}^2 \) based on the original data, and \( a^*_t = (w^*_t, u^*_{t1}, \ldots, u^*_{t4})' \) for \( t = 1, \ldots, 55 \) were \( \text{NI}(0, S_{aa}) \) random vectors independent of the \( q^*_t \)'s, where \( S_{aa} \) is given in (6.2). At each stage we calculated,

\[
\hat{v}^*_t = v^*_t - \hat{\beta}^*_0 - \sum_{i=1}^{4} x^*_t \hat{\beta}^*_i
\]

for \( t = 1, \ldots, 55 \) where \( (\hat{\beta}^*_0, \hat{\beta}^*_1, \ldots, \hat{\beta}^*_4) \) are the estimates based on the simulated data, and we also calculated \( \hat{x}^*_t \). At each stage we calculated our test procedures based on \( (\hat{v}^*_t, \hat{x}^*_t) \) for \( t = 1, \ldots, 55 \). We also recorded

\[
v^*_t = q^*_t + w^*_t - \sum_{i=1}^{4} u^*_t \hat{\beta}^*_i
\]

and calculated some of the test procedures based on the \( v^*_t \)'s in order to examine how closely the simulated residuals \( \hat{v}^*_t \) approximated the behavior of the simulated errors \( v^*_t \). We repeated this procedure for a total of 5000 Monte Carlo replications.

Table 6.4 summarizes the results of the Monte Carlo study. The middle section of the table gives the percent of Monte Carlo samples for which a given test statistic fell beyond the .01, .05, .10, .90, .95 and
<table>
<thead>
<tr>
<th>Statistic</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
<th>Corr. a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>NC b</td>
<td>NC</td>
<td>NC</td>
<td>.0998</td>
<td>.0456</td>
<td>.0106</td>
<td>.350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Kuiper</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1090</td>
<td>.0572</td>
<td>.0140</td>
<td>.339</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0966</td>
<td>.0482</td>
<td>.0096</td>
<td>.468</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Watson</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1016</td>
<td>.0526</td>
<td>.0112</td>
<td>.441</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1076</td>
<td>.0522</td>
<td>.0100</td>
<td>.501</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Chi-squared (11 int.)</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1082</td>
<td>.0584</td>
<td>.0130</td>
<td>.146</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Shapiro-Wilk (Blom est.)</td>
<td>.0114</td>
<td>.0524</td>
<td>.0996</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.6202</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0031)</td>
<td>(0.0042)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0087)</td>
</tr>
<tr>
<td>Skewness</td>
<td>.0086</td>
<td>.0458</td>
<td>NC</td>
<td>NC</td>
<td>.0506</td>
<td>.0110</td>
<td>.8276</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0031)</td>
<td></td>
<td></td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.0082</td>
<td>.0514</td>
<td>NC</td>
<td>NC</td>
<td>.0508</td>
<td>.0106</td>
<td>.7703</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0031)</td>
<td></td>
<td></td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.0058)</td>
</tr>
<tr>
<td>Outlier</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0946</td>
<td>.0512</td>
<td>.0106</td>
<td>.7136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0902</td>
<td>.0438</td>
<td>.0092</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td></td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0866</td>
<td>.0406</td>
<td>.0080</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0042)</td>
<td>(0.0031)</td>
<td>(0.0014)</td>
<td></td>
</tr>
</tbody>
</table>

*The sample correlation between the statistic based on \( v^* \) and the statistic based on \( v^*_t \) from 5000 Monte Carlo samples.*

b*Not calculated.*
critical points. These critical points are for the theoretical
distributions described earlier. The numbers in parentheses are
standard errors which were calculated from the usual binomial variance
formula. The right-hand column of Table 6.4 is the sample correlation
between the statistic based on $\hat{v}^*$ and the same statistic based on $v^*_c$
for the 5000 Monte Carlo samples. The numbers in parentheses under the
correlations are standard errors calculated by the formula

$$(5000)^{-1/2} (1 - \rho^2).$$

The formula was derived under the assumption that
the statistics based on $\hat{v}^*$ and $v^*_c$ are distributed as a bivariate
normal random variable under the randomization imposed by the Monte
Carlo sampling. While the assumption of bivariate normality is clearly
not valid, the formula does provide a useful approximation of the
standard error. The idea of calculating the correlation was taken from
White and MacDonald (1980), who studied tests of normality based on
ordinary least squares regression residuals.

Of the goodness-of-fit statistics, only the Kolmogorov-Smirnov,
Kuiper and chi-squared statistics have percentages which fall more than
two standard errors away from the theoretical significance levels. Not
surprisingly, these are also the statistics with the lowest correla-
tions. It is interesting that of the EDF statistics, the ones with the
highest correlation (i.e., Cramer-von Mises, Watson, and Anderson-
Darling) are all based on integral functionals of the empirical
process. It is known from Monte Carlo studies on iid samples that the
Anderson-Darling statistic seems to approach its limiting distribution
faster than do the other EDF statistics. The correlations indicate that this seems to hold for residuals also.

The extremely low correlation for the chi-squared statistic is surprising. Despite the low correlation, the percentiles of the statistic are in good agreement with the theoretical percentiles. One possible explanation for the low correlation is that our large sample theory was based on a fixed number of intervals, \( k \), with large sample sizes, \( n \). In this example \( k = O(n) \) because \( 11 = 55/5 \). If we let \( n \to \infty \) and \( k \to \infty \) then in the notation of Chapter Four,

\[
\hat{\chi}_k^2(W_n) = O_p(k), \tag{6.11}
\]

and by using expression (4.34) we can show that

\[
|\chi_k^2(W_n) - \chi_k^2(W_{\hat{W}})| = o_p(n^{-1/2} k^{3/2}). \tag{6.12}
\]

Thus if \( k = O(n) \) then the remainder in (6.12) is \( o_p(n) \) which says that \( \chi_k^2(W_n) \) and \( \chi_k^2(W_{\hat{W}}) \) are not guaranteed to be close, even in large samples. Notice that if we choose \( k = O(n^{1/3}) \) then the remainder term in (6.12) is \( o_p(1) \) so the two statistics will be close in large samples. In the next section, where we examine data from a medical study, we try choosing the number of intervals by the rule of thumb,

\[
k = \max\{3, [n^{1/3}]\}, \tag{6.13}
\]
where \([\cdot]\) is the rounding function.

The other goodness-of-fit statistics in Table 6.4 perform quite well. The Shapiro-Wilk statistic has a higher correlation than any of the EDF statistics, and its Monte Carlo distribution matches its theoretical distribution quite well. Interestingly, the two highest correlations in the table are given by the skewness and kurtosis statistics. These statistics are not as widely used as many of the other statistics in the table, but they seem to be the least affected by the use of residuals in place of the actual errors. The difference between the skewness statistic with a correlation of .8276, and the Kolmogorov-Smirnov statistic with a correlation of .350 is most dramatic. From a practical point of view because the skewness and kurtosis are the easiest statistics in the table to calculate.

The results for the last three test procedures in Table 6.4 are mixed. The outlier test has a high correlation, and its Monte Carlo distribution matches its theoretical distribution fairly well. The tests for autocorrelation, and homogeneity of variance both appear to be conservative since all the observed significance levels are smaller than their theoretical values. This may be due in part to the fact that the distributions are compared to an F-distribution instead of to a \(\chi^2\)-distribution. The \(\chi^2\)-distribution has shorter tails.

A second Monte Carlo study was conducted with the role performance data to determine the effect of smaller sample sizes on the test procedures. In the second study a sample size of \(n = 25\) was used, where the sample was a fixed subset of 25 of the \(\hat{x}_t\)'s given in Table
6.2. The other aspects of the Monte Carlo process were unchanged. The results of the second study are summarized in Table 6.5. The correlations in the righthand side of Table 6.5 are uniformly lower than the correlations in Table 6.4. The correlations in the two tables are in the same relative order, with skewness and kurtosis having the largest correlations and the Shapiro-Wilk statistic having a greater correlation than any of the EDF statistics. Even with relatively low correlations, most of the statistics approximate their theoretical distributions reasonably well. This is particularly true for the Kolmogorov-Smirnov statistic. Note that the chi-squared statistic was calculated using 5 intervals to conform to the "five per cell" rule of thumb. The correlation for the chi-squared statistic is not that much smaller in Table 6.5 than in Table 6.4. This is roughly what we would expect from expression (6.12) because the ratio of \( k/n \) is equal to 1/5 for the two tables. As in Table 6.4 the outlier test performs reasonably well, while the test for autocorrelation and the test for homogeneity of variance are again conservative.

The overall results from the two studies indicate that the distributions of all test statistics were well approximated by the limit distributions. Of the goodness-of-fit statistics, the Shapiro-Wilk, skewness and kurtosis statistics are least affected by using residuals instead of the actual errors to test for normality. Of the EDF statistics, the ones based on integral functionals of the empirical process seem least affected. While the "five per cell" chi-squared statistics based on true and estimated residuals have small correlation, the percentiles of the statistic for the residuals are similar to the
Table 6.5. Test procedures based on residuals from 5000 simulated data sets of size $n = 25$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
<th>Corr.(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>NC(^b)</td>
<td>NC</td>
<td>NC</td>
<td>.1018</td>
<td>.0478</td>
<td>.0094</td>
<td>.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Kuiper</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1260</td>
<td>.0656</td>
<td>.0146</td>
<td>.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Cramer-von-Mises</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0940</td>
<td>.0472</td>
<td>.0108</td>
<td>.269</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Watson</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1096</td>
<td>.0572</td>
<td>.0140</td>
<td>.245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1288</td>
<td>.0624</td>
<td>.0148</td>
<td>.306</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Chi-Squared (11 int.)</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1162</td>
<td>.0576</td>
<td>.0110</td>
<td>.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Shapiro-Wilk (Blom est.)</td>
<td>.0136</td>
<td>.051</td>
<td>.0918</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.410</td>
</tr>
<tr>
<td></td>
<td>(.0014)</td>
<td>(.0031)</td>
<td>(.0042)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>.0118</td>
<td>.0528</td>
<td>NC</td>
<td>NC</td>
<td>.0526</td>
<td>.0106</td>
<td>.633</td>
</tr>
<tr>
<td></td>
<td>(.0014)</td>
<td>(.0031)</td>
<td></td>
<td></td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.011)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.0138</td>
<td>.0584</td>
<td>NC</td>
<td>NC</td>
<td>.0544</td>
<td>.0142</td>
<td>.541</td>
</tr>
<tr>
<td></td>
<td>(.0014)</td>
<td>(.0031)</td>
<td></td>
<td></td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.012)</td>
</tr>
<tr>
<td>Outlier</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0954</td>
<td>.0500</td>
<td>.0120</td>
<td>.541</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td>(.0014)</td>
<td>(.012)</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0818</td>
<td>.0368</td>
<td>.0080</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorr.</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0754</td>
<td>.0322</td>
<td>.0038</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0042)</td>
<td>(.0031)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The sample correlation between the statistic based on $v*$ and the statistic based on $v*$ from 5000 Monte Carlo samples.

\(^b\)Not calculated.
theoretical percentiles. The outlier test performed well, even with only 20 degrees of freedom as in Table 6.5. The test of autocorrelation and the test of homogeneity of variance appear to be conservative in small samples.

A non-null Monte Carlo study to investigate the power properties of our test procedures would be of interest, but we choose not to examine that question here because of limited resources and the enormous number of possible alternatives to the null hypothesis. A power study would need to concentrate on a subset of our procedures, and to examine a restricted number of specific alternatives of interest.

B. Lung Vital Capacity Data

In this section we consider some data analyzed by Barnett (1969). The data arose from a medical study to determine the relative merits of two instruments used to measure human lung characteristics. The original study involved two instruments, a standard instrument and an experimental instrument, and two operators, a skilled operator and an unskilled operator. Four measurements of lung vital capacity, corresponding to the four instrument-operator combinations, were made on each of 72 patients. Following Barnett (1969) we will concentrate on the following three variables,

$Y_1$ = The experimental instrument operated by the skilled operator.

$Y_2$ = The experimental instrument operated by the unskilled operator.

$X$ = The standard instrument operated by the skilled operator.
The data are presented in Table 6.6. The postulated model is,

\[ Y_{t1} = \beta_{11} + x_t \beta_{11} + e_{t1}, \]  

\[ Y_{t2} = \beta_{02} + x_t \beta_{12} + e_t, \]  

\[ X_t = x_t + u_t, \]

for \( t = 1, \ldots, 72 \). It is assumed that the \( \varepsilon_t 's \) are \( \text{NI}(0, \Sigma_{\varepsilon\varepsilon}) \) where \( \varepsilon_t = (e_{t1}, e_{t2}, u_t)' \) and,

\[ \Sigma_{\varepsilon\varepsilon} = \text{diag}(\sigma_{e11}, \sigma_{ee2}, \sigma_{uu}). \]  

Barnett (1969) further assumed that the \( x_t 's \) are iid \( \text{N}(\mu_x, \sigma_{xx}) \) and independent of the \( \varepsilon_t 's \). All of the parameters of the model are assumed to be unknown, and need to be estimated.

The model described in (6.14) is commonly known as a factor analysis model, with three variables and one factor. The parameterization we have chosen seems natural for the problem of comparing the two instruments, by using the standard instrument and skilled operator combination as a base for comparison. The variable \( x_t \) represents the true lung vital capacity as measured by the standard instrument on the \( t \)-th patient, and \( u_t \) represents a measurement error incurred by the skilled operator. The \( x_t 's \) are assumed random because we envision the 72
Table 6.6. Readings of lung vital capacity for 72 patients on three instrument-operator combinations (Barnett, 1969)

<table>
<thead>
<tr>
<th>X</th>
<th>Y₁</th>
<th>Y₂</th>
<th>X</th>
<th>Y₁</th>
<th>Y₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>345</td>
<td>403</td>
<td>372</td>
<td>106</td>
<td>85</td>
<td>60</td>
</tr>
<tr>
<td>131</td>
<td>161</td>
<td>160</td>
<td>200</td>
<td>127</td>
<td>170</td>
</tr>
<tr>
<td>382</td>
<td>415</td>
<td>370</td>
<td>228</td>
<td>238</td>
<td>235</td>
</tr>
<tr>
<td>211</td>
<td>274</td>
<td>252</td>
<td>194</td>
<td>167</td>
<td>158</td>
</tr>
<tr>
<td>186</td>
<td>154</td>
<td>169</td>
<td>258</td>
<td>285</td>
<td>211</td>
</tr>
<tr>
<td>194</td>
<td>202</td>
<td>180</td>
<td>140</td>
<td>168</td>
<td>148</td>
</tr>
<tr>
<td>236</td>
<td>243</td>
<td>235</td>
<td>126</td>
<td>100</td>
<td>103</td>
</tr>
<tr>
<td>288</td>
<td>265</td>
<td>286</td>
<td>232</td>
<td>236</td>
<td>236</td>
</tr>
<tr>
<td>198</td>
<td>180</td>
<td>166</td>
<td>200</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>312</td>
<td>325</td>
<td>304</td>
<td>240</td>
<td>147</td>
<td>174</td>
</tr>
<tr>
<td>176</td>
<td>139</td>
<td>120</td>
<td>288</td>
<td>324</td>
<td>314</td>
</tr>
<tr>
<td>148</td>
<td>170</td>
<td>164</td>
<td>342</td>
<td>320</td>
<td>320</td>
</tr>
<tr>
<td>184</td>
<td>140</td>
<td>165</td>
<td>100</td>
<td>65</td>
<td>84</td>
</tr>
<tr>
<td>358</td>
<td>368</td>
<td>396</td>
<td>140</td>
<td>135</td>
<td>138</td>
</tr>
<tr>
<td>188</td>
<td>209</td>
<td>207</td>
<td>188</td>
<td>160</td>
<td>135</td>
</tr>
<tr>
<td>240</td>
<td>255</td>
<td>248</td>
<td>128</td>
<td>116</td>
<td>133</td>
</tr>
<tr>
<td>222</td>
<td>229</td>
<td>227</td>
<td>312</td>
<td>311</td>
<td>325</td>
</tr>
<tr>
<td>254</td>
<td>262</td>
<td>196</td>
<td>377</td>
<td>390</td>
<td>370</td>
</tr>
<tr>
<td>92</td>
<td>64</td>
<td>103</td>
<td>342</td>
<td>312</td>
<td>329</td>
</tr>
<tr>
<td>224</td>
<td>230</td>
<td>230</td>
<td>274</td>
<td>285</td>
<td>288</td>
</tr>
<tr>
<td>224</td>
<td>203</td>
<td>214</td>
<td>284</td>
<td>271</td>
<td>275</td>
</tr>
<tr>
<td>226</td>
<td>240</td>
<td>245</td>
<td>380</td>
<td>344</td>
<td>340</td>
</tr>
<tr>
<td>386</td>
<td>398</td>
<td>368</td>
<td>210</td>
<td>165</td>
<td>193</td>
</tr>
<tr>
<td>278</td>
<td>189</td>
<td>200</td>
<td>182</td>
<td>106</td>
<td>105</td>
</tr>
<tr>
<td>222</td>
<td>184</td>
<td>136</td>
<td>140</td>
<td>135</td>
<td>110</td>
</tr>
<tr>
<td>188</td>
<td>190</td>
<td>184</td>
<td>220</td>
<td>164</td>
<td>111</td>
</tr>
<tr>
<td>94</td>
<td>106</td>
<td>100</td>
<td>194</td>
<td>182</td>
<td>127</td>
</tr>
<tr>
<td>248</td>
<td>215</td>
<td>215</td>
<td>326</td>
<td>325</td>
<td>327</td>
</tr>
<tr>
<td>166</td>
<td>176</td>
<td>180</td>
<td>196</td>
<td>189</td>
<td>192</td>
</tr>
<tr>
<td>404</td>
<td>400</td>
<td>377</td>
<td>132</td>
<td>114</td>
<td>100</td>
</tr>
<tr>
<td>254</td>
<td>208</td>
<td>225</td>
<td>284</td>
<td>365</td>
<td>351</td>
</tr>
<tr>
<td>178</td>
<td>139</td>
<td>120</td>
<td>206</td>
<td>172</td>
<td>178</td>
</tr>
<tr>
<td>128</td>
<td>80</td>
<td>113</td>
<td>220</td>
<td>190</td>
<td>227</td>
</tr>
<tr>
<td>194</td>
<td>203</td>
<td>188</td>
<td>126</td>
<td>86</td>
<td>115</td>
</tr>
<tr>
<td>176</td>
<td>186</td>
<td>186</td>
<td>304</td>
<td>285</td>
<td>267</td>
</tr>
<tr>
<td>204</td>
<td>147</td>
<td>116</td>
<td>214</td>
<td>256</td>
<td>272</td>
</tr>
</tbody>
</table>
patients in the study as a sample from a population of people who could have been included in the study.

We use the estimation procedure of Barnett (1969), which is the method of maximum likelihood under the assumption of normality for $x_t$ and $\varepsilon_t$. Carter and Fuller (1980) and Fuller (1986) give a more detailed description of this approach. We define

$$v_{t1} = Y_{t1} - \beta_{01} - X_t \beta_{11},$$

$$= e_{t1} - u_t \beta_{11},$$

and

$$v_{t2} = Y_{t2} - \beta_{02} - X_t \beta_{12},$$

$$= e_{t2} - u_t \beta_{12},$$

for $t = 1, \ldots, 72$. By assumption, the vector $(v_{t1}, v_{t2})'$ has a bivariate normal distribution. We define $\hat{v}_{t1}$ and $\hat{v}_{t2}$ by

$$\hat{v}_{t1} = Y_{t1} - \hat{\beta}_{01} - X_t \hat{\beta}_{11},$$

$$\hat{v}_{t2} = Y_{t2} - \hat{\beta}_{02} - X_t \hat{\beta}_{12},$$

for $t = 1, \ldots, 72$, where $(\hat{\beta}_{01}, \hat{\beta}_{02}, \hat{\beta}_{11}, \hat{\beta}_{12})$ are $n^{1/2}$-consistent estimators of $(\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12})$. We define $\hat{x}_t$ as the estimator
given in Section 1.5 of Fuller (1986). The $\hat{v}_{t1}$'s and $\hat{v}_{t2}$'s standardized by their individual sample variances, the standardized residual vectors, and the $\hat{x}_t$'s are presented in Table 6.7. We next demonstrate our diagnostic procedures using the $\hat{v}_{t1}$'s, $\hat{v}_{t2}$'s and $\hat{x}_t$'s.

Table 6.8 contains nine goodness-of-fit test statistics for normality based on the two standardized components of the residual vector, and the estimated true values. We have concentrated on testing for normality based on the residuals, but it is easy to show that we can also test for normality based on the $\hat{x}_t$'s if the true $x_t$'s are iid normal random variables. In the notation of Theorem 2.6,

$$(\hat{x}_t - \bar{x}) = (x_t - \bar{x})(1 + c_n) + (v_t - \bar{v})D_n$$  (6.18)

where $c_n = o_p(n^{-1/2})$ and $D_n$ is a 2 x 1 random vector such that $D_n = o_p(n^{-1/2})$. It follows by the method used to prove Theorem 2.2 that,

$$\frac{m_{xx}^{1/2}}{\sqrt{n}}(\hat{x}_t - \bar{x}) = \frac{m_{xx}^{1/2}}{\sqrt{n}}(x_t - \bar{x})(1 + a_n) + (v_t - \bar{v})B_n ,$$  (6.19)

where $a_n = o_p(n^{-1/2})$ and $B_n$ is a 2 x 1 random vector such that $B_n = o_p(n^{-1/2})$. Thus the empirical process results of Chapter Three hold for empirical processes based on the standardized estimated true
Table 6.7. Residuals and expected true values from the lung vital capacity data

<table>
<thead>
<tr>
<th>Observation</th>
<th>$s_{\hat{v}1v1}^{-1/2}$</th>
<th>$s_{\hat{v}2v2}^{-1/2}$</th>
<th>$(\hat{v} - \hat{w})s_{vv}^{-1/2}$</th>
<th>$\hat{v}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.200</td>
<td>0.656</td>
<td>1.266</td>
<td>0.133</td>
</tr>
<tr>
<td>2</td>
<td>1.734</td>
<td>1.626</td>
<td>1.369</td>
<td>1.156</td>
</tr>
<tr>
<td>3</td>
<td>0.246</td>
<td>-0.604</td>
<td>0.762</td>
<td>-1.022</td>
</tr>
<tr>
<td>4</td>
<td>2.185</td>
<td>1.624</td>
<td>2.014</td>
<td>0.853</td>
</tr>
<tr>
<td>5</td>
<td>-0.400</td>
<td>0.090</td>
<td>-0.632</td>
<td>0.395</td>
</tr>
<tr>
<td>6</td>
<td>0.704</td>
<td>0.141</td>
<td>0.908</td>
<td>-0.267</td>
</tr>
<tr>
<td>7</td>
<td>0.412</td>
<td>0.330</td>
<td>0.363</td>
<td>0.194</td>
</tr>
<tr>
<td>8</td>
<td>-0.779</td>
<td>0.079</td>
<td>-1.167</td>
<td>0.632</td>
</tr>
<tr>
<td>9</td>
<td>-0.071</td>
<td>-0.384</td>
<td>0.159</td>
<td>-0.498</td>
</tr>
<tr>
<td>10</td>
<td>0.108</td>
<td>-0.192</td>
<td>0.285</td>
<td>-0.345</td>
</tr>
<tr>
<td>11</td>
<td>-0.481</td>
<td>-0.969</td>
<td>-0.028</td>
<td>-1.056</td>
</tr>
<tr>
<td>12</td>
<td>1.396</td>
<td>1.186</td>
<td>1.187</td>
<td>0.755</td>
</tr>
<tr>
<td>13</td>
<td>-0.733</td>
<td>0.042</td>
<td>-1.076</td>
<td>0.549</td>
</tr>
<tr>
<td>14</td>
<td>-0.266</td>
<td>0.912</td>
<td>-1.000</td>
<td>1.473</td>
</tr>
<tr>
<td>15</td>
<td>1.117</td>
<td>1.100</td>
<td>0.846</td>
<td>0.819</td>
</tr>
<tr>
<td>16</td>
<td>0.617</td>
<td>0.567</td>
<td>0.495</td>
<td>0.394</td>
</tr>
<tr>
<td>17</td>
<td>0.499</td>
<td>0.559</td>
<td>0.333</td>
<td>0.462</td>
</tr>
<tr>
<td>18</td>
<td>0.327</td>
<td>-1.359</td>
<td>1.391</td>
<td>-2.149</td>
</tr>
<tr>
<td>19</td>
<td>0.304</td>
<td>1.284</td>
<td>-0.437</td>
<td>1.621</td>
</tr>
<tr>
<td>20</td>
<td>0.458</td>
<td>0.579</td>
<td>0.260</td>
<td>0.517</td>
</tr>
<tr>
<td>21</td>
<td>-0.320</td>
<td>0.126</td>
<td>-0.543</td>
<td>0.393</td>
</tr>
<tr>
<td>22</td>
<td>0.676</td>
<td>0.938</td>
<td>0.328</td>
<td>0.882</td>
</tr>
<tr>
<td>23</td>
<td>-0.384</td>
<td>-0.791</td>
<td>-0.011</td>
<td>-0.867</td>
</tr>
<tr>
<td>24</td>
<td>-2.622</td>
<td>-2.027</td>
<td>-2.364</td>
<td>-1.133</td>
</tr>
<tr>
<td>25</td>
<td>-0.798</td>
<td>-2.014</td>
<td>0.228</td>
<td>-2.329</td>
</tr>
<tr>
<td>26</td>
<td>0.568</td>
<td>0.450</td>
<td>0.505</td>
<td>0.260</td>
</tr>
<tr>
<td>27</td>
<td>1.446</td>
<td>1.134</td>
<td>1.294</td>
<td>0.647</td>
</tr>
<tr>
<td>28</td>
<td>-0.817</td>
<td>-0.626</td>
<td>-0.741</td>
<td>-0.345</td>
</tr>
<tr>
<td>29</td>
<td>0.937</td>
<td>1.053</td>
<td>0.622</td>
<td>0.871</td>
</tr>
<tr>
<td>30</td>
<td>-0.959</td>
<td>-1.122</td>
<td>-0.605</td>
<td>-0.955</td>
</tr>
<tr>
<td>31</td>
<td>1.230</td>
<td>-0.538</td>
<td>-1.398</td>
<td>0.054</td>
</tr>
<tr>
<td>32</td>
<td>-0.551</td>
<td>-1.034</td>
<td>-0.084</td>
<td>-1.102</td>
</tr>
<tr>
<td>33</td>
<td>-0.498</td>
<td>0.395</td>
<td>-0.979</td>
<td>0.892</td>
</tr>
<tr>
<td>34</td>
<td>0.733</td>
<td>0.367</td>
<td>0.796</td>
<td>0.034</td>
</tr>
<tr>
<td>35</td>
<td>0.874</td>
<td>0.897</td>
<td>0.638</td>
<td>0.692</td>
</tr>
<tr>
<td>36</td>
<td>-1.234</td>
<td>-1.994</td>
<td>-0.406</td>
<td>-2.011</td>
</tr>
<tr>
<td>37</td>
<td>0.419</td>
<td>-0.387</td>
<td>0.861</td>
<td>-0.830</td>
</tr>
<tr>
<td>38</td>
<td>-1.671</td>
<td>-0.336</td>
<td>-2.156</td>
<td>0.634</td>
</tr>
<tr>
<td>39</td>
<td>0.548</td>
<td>0.590</td>
<td>0.381</td>
<td>0.473</td>
</tr>
<tr>
<td>40</td>
<td>-0.305</td>
<td>-0.480</td>
<td>-0.109</td>
<td>-0.479</td>
</tr>
<tr>
<td>41</td>
<td>0.851</td>
<td>1.065</td>
<td>1.938</td>
<td>2.079</td>
</tr>
<tr>
<td>42</td>
<td>1.620</td>
<td>0.994</td>
<td>1.636</td>
<td>0.333</td>
</tr>
</tbody>
</table>
Table 6.7. (continued)

<table>
<thead>
<tr>
<th>Observation</th>
<th>$s_{v_{v_{1}1}}^{-1/2}$</th>
<th>$s_{v_{v_{2}2}}^{-1/2}$</th>
<th>$(\hat{v} - \overline{v})s_{v_{v}}^{-1/2}$</th>
<th>$\hat{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>0.149</td>
<td>0.177</td>
<td>0.092</td>
<td>0.152</td>
</tr>
<tr>
<td>44</td>
<td>0.350</td>
<td>0.488</td>
<td>0.168</td>
<td>0.460</td>
</tr>
<tr>
<td>45</td>
<td>0.378</td>
<td>0.455</td>
<td>0.229</td>
<td>0.395</td>
</tr>
<tr>
<td>46</td>
<td>-2.499</td>
<td>-1.525</td>
<td>-2.529</td>
<td>-0.503</td>
</tr>
<tr>
<td>47</td>
<td>0.922</td>
<td>0.871</td>
<td>0.724</td>
<td>0.624</td>
</tr>
<tr>
<td>48</td>
<td>-1.089</td>
<td>-0.716</td>
<td>-1.068</td>
<td>-0.292</td>
</tr>
<tr>
<td>49</td>
<td>0.052</td>
<td>0.486</td>
<td>-0.255</td>
<td>0.655</td>
</tr>
<tr>
<td>50</td>
<td>0.667</td>
<td>0.711</td>
<td>0.469</td>
<td>0.566</td>
</tr>
<tr>
<td>51</td>
<td>-0.297</td>
<td>-0.935</td>
<td>0.211</td>
<td>-1.131</td>
</tr>
<tr>
<td>52</td>
<td>0.540</td>
<td>0.960</td>
<td>0.119</td>
<td>1.004</td>
</tr>
<tr>
<td>53</td>
<td>-0.295</td>
<td>0.401</td>
<td>-0.694</td>
<td>0.767</td>
</tr>
<tr>
<td>54</td>
<td>-0.299</td>
<td>-0.441</td>
<td>-0.126</td>
<td>-0.427</td>
</tr>
<tr>
<td>55</td>
<td>-1.320</td>
<td>-0.461</td>
<td>-1.570</td>
<td>0.223</td>
</tr>
<tr>
<td>56</td>
<td>0.289</td>
<td>0.592</td>
<td>0.010</td>
<td>0.648</td>
</tr>
<tr>
<td>57</td>
<td>-0.466</td>
<td>-0.101</td>
<td>-0.596</td>
<td>0.166</td>
</tr>
<tr>
<td>58</td>
<td>-1.732</td>
<td>-1.387</td>
<td>-1.528</td>
<td>-0.818</td>
</tr>
<tr>
<td>59</td>
<td>-0.925</td>
<td>-0.011</td>
<td>-1.313</td>
<td>0.500</td>
</tr>
<tr>
<td>60</td>
<td>-1.644</td>
<td>-1.589</td>
<td>-1.267</td>
<td>-1.162</td>
</tr>
<tr>
<td>61</td>
<td>0.667</td>
<td>-0.080</td>
<td>1.007</td>
<td>-0.558</td>
</tr>
<tr>
<td>62</td>
<td>-1.305</td>
<td>-2.656</td>
<td>-0.058</td>
<td>-2.904</td>
</tr>
<tr>
<td>63</td>
<td>0.127</td>
<td>-1.357</td>
<td>1.103</td>
<td>-2.013</td>
</tr>
<tr>
<td>64</td>
<td>-0.383</td>
<td>0.002</td>
<td>-0.548</td>
<td>0.258</td>
</tr>
<tr>
<td>65</td>
<td>0.258</td>
<td>0.417</td>
<td>0.086</td>
<td>0.418</td>
</tr>
<tr>
<td>66</td>
<td>0.342</td>
<td>-0.102</td>
<td>0.558</td>
<td>-0.374</td>
</tr>
<tr>
<td>67</td>
<td>2.246</td>
<td>2.048</td>
<td>1.814</td>
<td>1.414</td>
</tr>
<tr>
<td>68</td>
<td>-0.583</td>
<td>-0.305</td>
<td>-0.624</td>
<td>-0.045</td>
</tr>
<tr>
<td>69</td>
<td>-0.555</td>
<td>0.624</td>
<td>-1.217</td>
<td>1.257</td>
</tr>
<tr>
<td>70</td>
<td>-0.254</td>
<td>0.516</td>
<td>-0.714</td>
<td>0.903</td>
</tr>
<tr>
<td>71</td>
<td>-0.764</td>
<td>-0.978</td>
<td>-0.426</td>
<td>-0.880</td>
</tr>
<tr>
<td>72</td>
<td>1.560</td>
<td>2.093</td>
<td>0.804</td>
<td>1.934</td>
</tr>
</tbody>
</table>
Table 6.8. Goodness-of-fit tests based on the standardized components of the residual vectors and estimated true values from the lung vital capacity data

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$v_{t1}$</th>
<th>$v_{t2}$</th>
<th>$x_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>.597</td>
<td>.755</td>
<td>.845</td>
</tr>
<tr>
<td>Kuiper</td>
<td>.15 &lt; p</td>
<td>.15 &lt; p</td>
<td>.05 &lt; p &lt; .10</td>
</tr>
<tr>
<td>Carmer-von Mises</td>
<td>.0468</td>
<td>.081</td>
<td>.153</td>
</tr>
<tr>
<td>Watson</td>
<td>.0458</td>
<td>.0724</td>
<td>.131</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>.295</td>
<td>.488</td>
<td>1.026</td>
</tr>
<tr>
<td>Chi-squared (4 intervals)</td>
<td>1.222</td>
<td>2.778</td>
<td>2.111</td>
</tr>
<tr>
<td>Shaprio-Wilk (Blom estimates)</td>
<td>.989</td>
<td>.984</td>
<td>.960</td>
</tr>
<tr>
<td>Skewness</td>
<td>-.188</td>
<td>-.364</td>
<td>.524</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.127</td>
<td>2.866</td>
<td>2.444</td>
</tr>
</tbody>
</table>

*Calculated with respect to the theoretical distributions.*
values. It follows that the results of Chapter Four hold as well. The outlier test and test of autocorrelation can also be applied to the \( \hat{x}_t \)'s.

The results in Table 6.8 indicate that the assumption of marginal normality is reasonable for both \( v_{t1} \) and \( v_{t2} \), but the \( \hat{x}_t \)'s appear to have a positively skewed distribution. This is not surprising since there is no reason to expect a human population characteristic such as lung vital capacity to be a normally distributed random variable, but it is reasonable to hope that the errors in measurement would be approximately normal. It is interesting that Barnett (1969) attempted to check the model assumptions using the raw data \((Y_{t1}, Y_{t2}, X_t)\) for \( t = 1, \ldots, 72 \). He looked at scatter plots of \( Y_1 \) against \( X \) and of \( Y_2 \) against \( X \). He noted that apart from a slight positive skewness in \( X \) the assumption of trivariate normality for \((Y_{t1}, Y_{t2}, X_t)\) was reasonable. Our procedures appear to be more sensitive for detecting departures from normality of the true \( x_t \)'s.

This is not surprising since we expect \( \hat{x}_t \) to be superior to the actual observation \( X_t \) as a predictor of \( x_t \). Therefore, tests based on the \( \hat{x}_t \)'s should be more effective than tests based on the \( X_t \)'s.

We used \( k = 4 \) intervals in calculating the chi-squared statistic in Table 6.8. We chose \( k = 4 \) to comply with the experimental rule-of-thumb devised in expression (6.13) which yields \( [(72)^{1/3}] = 4 \). In Tables 6.9 and 6.11 we use \( k = 5 \) intervals for the combined samples since \( [(144)^{1/3}] = 5 \).
Table 6.9. Goodness-of-fit tests based on the components of the standardized residual vectors from the lung vital capacity data

<table>
<thead>
<tr>
<th>Statistic</th>
<th>( \hat{v}_{t1} )</th>
<th>( \hat{v}_{t2} )</th>
<th>Combined(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>.639</td>
<td>1.171</td>
<td>1.086</td>
</tr>
<tr>
<td></td>
<td>.15 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Kuiper</td>
<td>.908</td>
<td>1.92</td>
<td>1.779</td>
</tr>
<tr>
<td></td>
<td>.15 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>.0395</td>
<td>.262</td>
<td>.2186</td>
</tr>
<tr>
<td></td>
<td>.15 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Watson</td>
<td>.0356</td>
<td>.227</td>
<td>.187</td>
</tr>
<tr>
<td></td>
<td>.15 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>.252</td>
<td>1.554</td>
<td>1.246</td>
</tr>
<tr>
<td></td>
<td>.15 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Chi-Squared (4 &amp; 5 intervals)</td>
<td>.778</td>
<td>8.556</td>
<td>7.111</td>
</tr>
<tr>
<td></td>
<td>.10 &lt; p</td>
<td>p &lt; .01</td>
<td>.01 &lt; p &lt; .05</td>
</tr>
<tr>
<td>Shapiro-Wilk (Blom estimates)</td>
<td>.990</td>
<td>.945</td>
<td>.975</td>
</tr>
<tr>
<td></td>
<td>.10 &lt; p</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>Skewness</td>
<td>-.287</td>
<td>-.786</td>
<td>-.537</td>
</tr>
<tr>
<td></td>
<td>.10 &lt; p</td>
<td>p &lt; .02</td>
<td>p &lt; .02</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.858</td>
<td>3.331</td>
<td>3.09</td>
</tr>
<tr>
<td></td>
<td>.10 &lt; p</td>
<td>.10 &lt; p</td>
<td>.10 &lt; p</td>
</tr>
</tbody>
</table>

\(^a\)Calculated with respect to the theoretical tabled distributions.

\(^b\)Calculated from \((2)(72) = 144\) combined variables.
In Table 6.9 we assess bivariate normality of \((v_{t1}, v_{t2})\) based on the standardized residual vectors \((\hat{v}_t - \bar{v})S_{vv}^{-1/2}\) for \(t = 1, \ldots, 72\). The first two columns of Table 6.9 give nine test statistics based on the two individual components of the standardized residual vector, and the last column gives the tests based on the 144 variables obtained by combining the two columns of the standardized residual vector. The first component appears to be normally distributed, but the second component is significantly nonnormal and negatively skewed. As a result, the tests based on the 144 combined variables also indicates nonnormality. It is interesting that the tests of normality of \(v_{t1}\) and \(v_{t2}\) in Table 6.8 and Table 6.9 lead to different conclusions. This suggests that if we are really interested in the joint behavior of \((v_{t1}, v_{t2})\) then we should look at tests based on the standardized residual vectors \((\hat{v}_t - \bar{v})S_{vv}^{-1/2}\) for \(t = 1, \ldots, n\) instead of the individual standardized components of the residual vector

\[s_{vvl1}(\hat{v}_{t1} - \hat{v}_1), s_{vvl2}(\hat{v}_{t2} - \hat{v}_2)\] for \(t = 1, \ldots, n\).

We also calculated the tests for outliers, autocorrelation, and homogeneity of variance based on the variables of Table 6.7. Because we found no anomalies, we do not present the results here.

We next present the results of a Monte Carlo study based on the Lung Vital Capacity Data. Our purpose is to determine how well the test procedures based on the standardized residual vector perform under the null hypothesis of normally distributed errors. In order to simulate an "actual data set", we generated data in accordance with model (6.14) and
a sample size of \( n = 72 \) to match the Lung Vital Capacity data. To be specific, at each stage of Monte Carlo replication we generated

\[
Y_{t1}^* = \beta_{01} + X_t^* \hat{\beta}_{11} + e_{t1}^* ,
\]

\[
Y_{t2}^* = \beta_{02} + X_t^* \hat{\beta}_{12} + e_{t2}^* ,
\]

\[
X_t^* = x_t^* + u_t^* ,
\]

for \( t = 1, \ldots, 72 \), where \( (\hat{\beta}_{01}, \hat{\beta}_{02}, \hat{\beta}_{11}, \hat{\beta}_{12}) \) were the estimates calculated from the original data, and \( x_t^* \) for \( t = 1, \ldots, 72 \) are iid random variables according to the distribution function,

\[
\hat{F}(x) = (72)^{-1} \sum_{t=1}^{72} I(x_t^* < x) ,
\]

which is the empirical distribution function of the estimated true values \( \hat{x}_t \) for \( t = 1, \ldots, 72 \). We generated the \( x_t^* \) according to (6.20) because the model for the original data (6.14) specified that the true \( x_t \)'s were iid random variables, and our analysis of the data indicated that the distribution of the \( x_t \)'s was nonnormal. In addition \( e_t^* = (e_{t1}^*, e_{t2}^*, u_t^*) \) for \( t = 1, \ldots, 72 \) were iid \( N(0, \Sigma_{ee}) \) random vectors, where \( \Sigma_{ee} = \text{diag}(\hat{\sigma}_{e11}, \hat{\sigma}_{e22}, \hat{\sigma}_{uu}) \) was estimated from the original data. At each stage we calculated

\[
\hat{e}_t^* = Y_t^* - \hat{\beta}_0 - X_t^* \hat{\beta}_1^* ,
\]
for \( t = 1, \ldots, 72 \), where \((\hat{\theta}_0^*, \hat{\theta}_1^*)\) are the estimates based on the simulated data, and we also calculated \( \hat{x}_t^* \). At each stage we calculated our test procedures based on \((\hat{v}_t^*, \hat{x}_t^*)\) for \( t = 1, \ldots, 72 \). The true errors are

\[
\begin{align*}
v_{t1}^* &= \epsilon_{t1}^* - u_{t11}^* \\
v_{t2}^* &= \epsilon_{t2}^* - u_{t12}^*
\end{align*}
\]  

(6.22)

and we calculated some of the test procedures based on the \( v_t^* \)'s in order to examine how closely the simulated residuals \( \hat{v}_t^* \) approximated the behavior of the simulated actual errors \( v_t^* \). We repeated this procedure for a total of 1000 Monte Carlo replications.

Table 6.10 summarizes the results of the Monte Carlo study dealing with the first component of the standardized residual vector. We present only the results for the first component because the results for the second component were nearly identical. The structure of Table 6.10 is the same as that of Tables 6.4 and 6.5. The numbers in parentheses are standard errors and were calculated by the same methods described for Tables 6.4 and 6.5.

All the statistics in Table 6.10 approximate their theoretical distribution well, and none have percentage points which fall more than two standard errors away from the theoretical significance levels. The correlations are uniformly greater than they were in Tables 6.4 and 6.5. This can be attributed to the greater sample size and to the fact that there is only one independent variable in the model of (6.14). It is interesting that, while the correlations are all larger, they still
Table 6.10. Test procedures based on the first component of the standardized residual vector from 1000 simulated data sets of size \( n = 72 \)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
<th>Corr. (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>NC (^b)</td>
<td>NC</td>
<td>NC</td>
<td>.1084</td>
<td>.0490</td>
<td>.0151</td>
<td>.753</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.021)</td>
</tr>
<tr>
<td>Kuiper</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1062</td>
<td>.0433</td>
<td>.0086</td>
<td>.751</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.021)</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0921</td>
<td>.0387</td>
<td>.0098</td>
<td>.884</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.015)</td>
</tr>
<tr>
<td>Watson</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0942</td>
<td>.0465</td>
<td>.0084</td>
<td>.869</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.016)</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0973</td>
<td>.0496</td>
<td>.0125</td>
<td>.892</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Chi-squared (^4) int.</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0821</td>
<td>.0416</td>
<td>.0118</td>
<td>.690</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.023)</td>
</tr>
<tr>
<td>Shapiro-Wilk (^\text{Blom est.})</td>
<td>.0132</td>
<td>.0574</td>
<td>.1050</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.930</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
<td>(.0095)</td>
<td></td>
<td></td>
<td></td>
<td>(.012)</td>
</tr>
<tr>
<td>Skewness</td>
<td>.0121</td>
<td>.0625</td>
<td>NC</td>
<td>NC</td>
<td>.0472</td>
<td>.0164</td>
<td>.975</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
<td></td>
<td></td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.007)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.0076</td>
<td>.0475</td>
<td>NC</td>
<td>NC</td>
<td>.0463</td>
<td>.0082</td>
<td>.968</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
<td></td>
<td></td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.008)</td>
</tr>
<tr>
<td>Outlier</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0991</td>
<td>.0471</td>
<td>.0123</td>
<td>.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td>(.010)</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.1010</td>
<td>.048</td>
<td>.0044</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td></td>
</tr>
<tr>
<td>Auto-correlation</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>.0996</td>
<td>.0525</td>
<td>.0082</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.0095)</td>
<td>(.0069)</td>
<td>(.0031)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The sample correlation between the statistic based on \( \hat{\omega}_t \) and the statistic based on \( \hat{\omega}_t^* \) from 1000 Monte Carlo samples.

\(^b\) Not calculated.
occur in the same relative order. The skewness and kurtosis have the greatest correlations, and the Shapiro–Wilk statistic has a greater correlation than any of the EDF statistics. The chi-squared statistic appears to perform better under the $k = \max\{3, \lfloor n^{1/3} \rfloor \}$ rule-of-thumb used in Table 6.10 than the $k = \lfloor n/5 \rfloor$ rule used in Tables 6.4 and 6.5, but the chi-square statistic still performs worst overall. The outlier test performs well. The autocorrelation test and test for homogeneity of variance are no longer conservative as in Tables 6.4 and 6.5 and now appear to perform quite well.

Table 6.11 summarizes the results of the Monte Carlo study dealing with the combined components of the standardized residual vector. The only organizational change in Table 6.10 from the previous tables is that we removed the autocorrelation test and test for homogeneity of variance and replaced them with Mardia's test of multivariate skewness and kurtosis described in Section C of Chapter Four. The cut-off points for the theoretical distribution of these statistics were taken from Mardia (1974).

All the statistics in Table 6.11 approximate their theoretical distributions well, with none of the percentage points falling more than two standard errors away from their theoretical significance levels. Interestingly, the correlations are all about the same as they were in Table 6.10, except for the chi-squared statistic which is significantly smaller. We may have expected the correlations in Table 6.11 to be larger than in Table 6.10 since the statistics are based on a combined sample of 144 observations instead of 72 observations. This larger
Table 6.11. Test procedures based on the combined component of the standardized residual vector from 1000 simulated data sets of size $n = 72$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Percent beyond theoretical quantiles</th>
<th>Corr.$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuiper</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-squared (5 int.)</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapiro-Wilk (Blom est.)</td>
<td>.0112</td>
<td>.0606</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
</tr>
<tr>
<td>Skewness</td>
<td>.0147</td>
<td>.0564</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.0089</td>
<td>.0532</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
</tr>
<tr>
<td>Outlier</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mardia's Skewness</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mardia's Kurtosis</td>
<td>.0087</td>
<td>.0436</td>
</tr>
<tr>
<td></td>
<td>(.0031)</td>
<td>(.0069)</td>
</tr>
</tbody>
</table>

$^a$The sample correlation between the statistic based on $\hat{\psi}_t$ and the statistic based on $\hat{\psi}_t^*$ from 1000 Monte Carlo samples.

$^b$Not calculated.
combined sample size is taken into account, for instance, when calculating the significance levels, so the statistics do act distributionally like they are based on a larger sample. The reason for the similar correlations is that while the statistics in Table 6.11 are based on a larger combined sample, the error in approximating $v^*_c$ by $\hat{v}^*_c$ remains the same for both tables. The order of approximation may explain why the chi-square statistic has a lower correlation in Table 6.11 than in Table 6.10. We found from the Monte Carlo studies associated with the Role Performance Data of Section A that using too many intervals in the chi-square statistic results in a low correlation. Apparently, we should use the same numbers of intervals in the combined samples as we use in the uncombined samples. In our example this would mean using $k = \lfloor (72)^{1/3} \rfloor = 4$ intervals instead of $k = \lfloor (144)^{1/3} \rfloor = 5$. We recalculated the chi-square statistic based on the combined sample with $k = 4$ intervals and the correlation was about the same as in Table 6.10.

The outlier statistic in Table 6.11 performs well, and our reasoning above explains why its correlation is nearly the same as in Table 6.10. The multivariate skewness and kurtosis statistics of Mardia perform well and have large correlations which are about the same as the other skewness and kurtosis statistics in Table 6.11.

Our overall results from the Monte Carlo study indicate that the distributions of the test procedures based on the separate components of the standardized residual vectors are well approximated by the theoretical null distributions for the sample size and degrees of freedom.
considered. The relative error in the approximations of the different procedures was the same as in the Monte Carlo studies associated with the role performance data. The closeness of the approximations was uniformly better for the larger sample size of the lung vital capacity data. Our results also indicate that the distributions of the test procedures based on the combined components of the standardized residual vectors are well approximated by the theoretical distributions. The quality of the approximations is about the same for the combined as for the separate components of the standardized residual vector.


Wald, A. (1940). Fitting straight lines when both variables are subject to error and ranks of the means are known. Ann. Math. Statist. 11:284-300.


VIII. ACKNOWLEDGEMENTS

My sincere thanks go to Wayne Fuller for making my four years of graduate study so rewarding. I could not have chosen a better major professor.

I also wish to thank the members of my graduate committee: Professors Yasuo Amemiya, David Harville, Elgin Johnston, and Glen Meeden.

Finally, I wish to thank Chris Olson for her efficient and skillful typing of my dissertation. Thanks also go to the Survey Section for the use of its word processor.
IX. APPENDIX A: REVIEW OF WEAK CONVERGENCE

In this appendix we briefly review some topics from the theory of weak convergence on metric spaces, and give some special results for the space $D[0, 1]$. Most of these results are taken from Billingsley (1968).

We begin with the following definitions.

**Definition 9.1.** A metric space $(S, \rho)$ is a nonempty set of $S$ of elements (called points) together with real-valued function $\rho$ (called a metric) defined on $S \times S$ such that for all $x, y, z$ in $S$:

i) $\rho(x, y) > 0$,

ii) $\rho(x, y) = 0$ if and only if $x = y$,

iii) $\rho(x, y) = \rho(y, x)$,

iv) $\rho(x, y) < \rho(x, z) + \rho(z, y)$.

**Definition 9.2.** A metric space $(S, \rho)$ is called separable if it contains a countable dense subset.

**Definition 9.3.** A set $A$ is called open in the metric space $(S, \rho)$ if for every $x \in A$, there exists $\delta > 0$ such that all $y \in S$ with $\rho(x, y) < \delta$ belong to $A$. 
Definition 9.4. A topology on a set $S$ is a collection $T$ of subsets (called open sets) of $S$ having the following properties:

1) $S$ and the empty set are in $T$,

2) The union of the elements of any subcollection of $T$ is in $T$.

3) The intersection of the elements of any finite subcollection of $T$ is in $T$.

Notice that the properties of Definition 9.4 are satisfied by the open sets of Definition 9.3 associated with the metric space. Thus we say that a metric $\rho$ induces a topology on the set $S$, and the resulting topological space $(S, T)$ is called metrizable. It is important to make the distinction between a metric space $(S, \rho)$ and its associated topological space $(S, T)$, since two different metric spaces can give rise to the same topological space.

In order to do probability theory on the metric space $(S, \rho)$ we need to make the following additional definitions.

Definition 9.5. Let $(S, \rho)$ be a metric space with associated topological space $(S, T)$. Then the class of Borel subsets $\mathcal{B}(S)$ of $(S, \rho)$ and $(S, T)$ is the $\sigma$-algebra generated by the open sets of $T$.

Definition 9.6. Let $A$ be a subset of $S$, where $(S, \rho)$ is a metric space. Then the boundary of $A$, denoted by $\partial A$, consists of
those points in $S$ which are limits of sequences of points in $A$ and are also limits of sequences of points outside $A$.

Now consider a sequence of probability measures $\{P_n\}_{n=1}^{\infty}$, and $P$ defined on the Borel sets $\mathcal{B}(S)$ of a metric space $(S, \rho)$ with topology $T$. We say that $P_n$ converges weakly to $P$ if $P_n(A) \to P(A)$ for all $A \in \mathcal{B}(S)$ such that $P(\exists A) = 0$. This concept of weak convergence is closely related to the concept of convergence in law which is defined in terms of the sequence of random elements $(X_n)_{n=1}^{\infty}$, and $X$. We define this below.

**Definition 9.7.** Let $(X_n)_{n=1}^{\infty}$, and $X$ be a sequence of random elements taking on values in a metric space $(S, \rho)$. Then we say that the sequence of random elements converges in law to the random element $X$, denoted by $X_n \xrightarrow{L} X$, if the distributions $P_n$ of $X_n$ converge weakly to the distribution $P$ of $X$.

We have not formally defined random elements, but they are just measurable mappings from some underlying probability space to the metric space $(S, \rho)$. Notice that the definition of convergence in law does not require that the random elements be defined on the same underlying probability space. A case where this distinction about the underlying probability space is important is given in the following theorem.

**Theorem 9.1.** Let $X$, and $Y$ be random elements of the separable metric space $(S, \rho)$, and defined on a common probability space. Then $\rho(X, Y)$ is a random variable.
It turns out that the conclusion of Theorem 9.1 may not hold if the metric space is not separable. See Billingsley (1968).

The concept of convergence in law is a useful one in statistics. The following theorem is a practical tool for proving convergence in law results for a sequence of random elements, in terms of the known behavior of another sequence of random elements.

**Theorem 9.2.** Let \( \{X_n\}_{n=1}^{\infty} \), \( \{Y_n\}_{n=1}^{\infty} \), \( X \), and \( Y \) be random elements in the separable metric space \((S, \rho)\), all defined on a common probability space. If \( X_n \xrightarrow{L} X \) and \( \rho(X_n, Y_n) \xrightarrow{P} 0 \), then \( Y_n \xrightarrow{L} Y \).

Notice that separability was needed in Theorem 9.2 to ensure that \( \rho(X_n, Y_n) \) is a well defined random variable. See Billingsley (1968).

The last general theorem we are going to present has to do with convergence in law of functions of random elements. To set up the problem, let \( h \) be a measurable mapping from the metric space \((S, \rho)\) to another metric space \((S', \rho')\), and let \( D_h \) be the set of discontinuities of \( h \) (so \( D_h \) is a subset of \( S \)). It turns out that \( D_h \in B(S) \) even if \( h \) is not measurable. See Billingsley (1968). We are now ready to state the theorem. For a proof see Billingsley (1968).

**Theorem 9.3.** (In terms of notation above.) Let \( \{X_n\}_{n=1}^{\infty} \), and \( X \) be random elements of the metric space \((S, \rho)\). If \( X_n \xrightarrow{L} X \) and \( P(X \in D_h) = 0 \), then \( h(X_n) \xrightarrow{L} h(X) \).
Theorem 9.2 and 9.3 are very useful, and we have used them repeatedly in earlier chapters. In those chapters, we dealt with two metric spaces. The first space is $\mathbb{R}^n$ for $n > 1$ with the Euclidean metric, which is a separable metric space. The second space is $D[0, 1]$, which is defined to be the space of functions on $[0, 1]$ that are right-continuous and have left-hand limits. This is the natural space in which to examine an empirical process on $[0, 1]$ since cumulative distribution functions are elements of $D[0, 1]$.

The most natural metric to put on $D[0, 1]$ would seem to be the uniform metric which is defined by

$$\rho_u(x, y) = \sup_{\omega \in [0, 1]} |x(\omega) - y(\omega)|,$$  \hspace{1cm} (9.1)

where $x$ and $y$ are elements of $D[0, 1]$. The topology induced by $\rho_u$ is called the uniform topology on $D[0, 1]$, and is denoted by $T_u$. It turns out that $\{D[0, 1], \rho_u\}$ is not a separable metric space. See Billingsley (1968). Because of this, Skorohod introduced another metric on $D[0, 1]$. Let $\Lambda$ be the class of strictly increasing continuous functions on $[0, 1]$ such that $\lambda(0) = 0$ and $\lambda(1) = 1$. The Skorohod metric is defined by

$$\rho_S(x, y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{0 < \omega < 1} |x(\omega) - y(\lambda(\omega))| + \sup_{0 < \omega < 1} \left| \omega - \lambda(\omega) \right| \right\},$$  \hspace{1cm} (9.2)

where $x$ and $y$ are elements of $D[0, 1]$. The topology induced by $\rho_S$ is called the Skorohod topology on $D[0, 1]$, and is denoted by...
It turns out that \( \{D[0, 1], \rho_S \} \) is a separable metric space. See Billingsley (1968).

It is easy to see that by letting \( \lambda(\omega) = \omega \), we get the inequality,

\[
\rho_S(x, y) < \rho_u(x, y),
\]

for \( x \) and \( y \) elements of \( D[0,1] \). This implies that \( T_u \) contains \( T_S \), and it can be further shown that \( T_S \) is a proper subset of \( T_u \).

The theoretical significance of this has to do with problems of measurability when one tries to interpret weak convergence in \( D[0, 1] \) in terms of the uniform topology. See Billingsley (1968). We have avoided the measurability problem in this thesis by always considering weak convergence in \( D[0, 1] \) in terms of the Skorohod topology.

The practical significance of (9.3) is that convergence of a sequence of functions in \( D[0, 1] \) in the uniform metric implies convergence in the Skorohod metric. This is helpful because the uniform metric is easier to work with. Let \( \{X_n\}_{n=1}^\infty \) and \( \{Y_n\}_{n=1}^\infty \) be sequences of random functions in \( D[0, 1] \), defined on the same underlying probability space. If we can find a sequence of random variables \( \{a_n\}_{n=1}^\infty \) such that,

\[
\rho_u(X_n, Y_n) < a_n,
\]

then we know that,
Thus if \( a_n \xrightarrow{p} 0 \) we can write,

\[
\rho_S(X_n, Y_n) \xrightarrow{p} 0 ,
\]

since \( \rho_S(X_n, Y_n) \) is a well defined sequence of random variables under the Skorohod topology on \( D[0, 1] \), by Theorem 9.1.

Interestingly, the theoretical solution outlined in (9.4)-(9.6) is ignored by many people. People sometimes construct the sequence of random \( a_n \)'s which converge to zero, and then say

\[
\rho_u(X_n, Y_n) \xrightarrow{p} 0 .
\]

Technically, such a procedure is not correct, since the \( \rho_u(X_n, Y_n) \)'s may not be well defined random variables under the uniform metric, since the uniform metric does not give rise to a separable metric space.

Nevertheless, the \( a_n \)'s would imply that (9.6) holds, so "in spirit" (9.7) does hold. Since it is common practice to make statements in the form of (9.7), we have continued to make them throughout this thesis, but the reader should remember that it is (9.6) which is actually being implied.
In this appendix we give percentage points for some of the distributions discussed in earlier chapters.

We begin with the EDF statistics of Chapter Four. As in Chapter Four, we let \( D_n \) denote the Kolmogorov-Smirnov statistic, \( V_n \) the Kuiper statistic, \( C^2_n \) the Cramer-von Mises statistic, \( U^2_n \) Watson's statistic, and \( A^2_n \) the Anderson-Darling statistic. These statistics are based on an iid sample from a normal distribution with unknown mean and variance. The sample mean and variance are used as estimators of the unknown parameters. Stephens (1974) gave percentage points for the limiting distributions of these statistics (properly normalized). We reproduce these in Table 10.1 below.

Table 10.1 Upper percentage points for the limiting distribution of EDF statistics (Stephens, 1974)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>15.0%</th>
<th>10.0%</th>
<th>5.0%</th>
<th>2.5%</th>
<th>1.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n^{1/2} D_n )</td>
<td>0.775</td>
<td>0.819</td>
<td>0.895</td>
<td>0.955</td>
<td>1.035</td>
</tr>
<tr>
<td>( n^{1/2} V_n )</td>
<td>1.320</td>
<td>1.386</td>
<td>1.489</td>
<td>1.585</td>
<td>1.693</td>
</tr>
<tr>
<td>( C^2_n )</td>
<td>0.091</td>
<td>0.104</td>
<td>0.126</td>
<td>0.148</td>
<td>0.178</td>
</tr>
<tr>
<td>( U^2_n )</td>
<td>0.085</td>
<td>0.096</td>
<td>0.116</td>
<td>0.136</td>
<td>0.163</td>
</tr>
<tr>
<td>( A^2_n )</td>
<td>0.576</td>
<td>0.656</td>
<td>0.787</td>
<td>0.918</td>
<td>1.092</td>
</tr>
</tbody>
</table>
Stephens (1974) also described the results of Monte Carlo
determinations of the sampling distributions of the EDF statistics in
Table 10.1 for finite samples. On the basis of his studies, he
recommends some small sample modifications to the EDF statistics. We
present these modifications in Table 10.2 below.

Table 10.2 Small sample modifications to the EDF statistics
(Stephens, 1974)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^{1/2}D_n$</td>
<td>$[n^{1/2} - 0.01 + (0.85)n^{-1/2}]D_n$</td>
</tr>
<tr>
<td>$n^{1/2}V_n$</td>
<td>$[n^{1/2} + 0.05 + (0.82)n^{-1/2}]V_n$</td>
</tr>
<tr>
<td>$C_n^2$</td>
<td>$[1 + (2n)^{-1}]C_n^2$</td>
</tr>
<tr>
<td>$U_n^2$</td>
<td>$[1 + (2n)^{-1}]U_n^2$</td>
</tr>
<tr>
<td>$A_n^2$</td>
<td>$[1 + (4)n^{-1} - (25)n^{-2}]A_n^2$</td>
</tr>
</tbody>
</table>

The modified statistics have distributions which are better
approximated, in finite samples, by the asymptotic values of Table 10.1
than the unmodified statistics. It is recommended that the modified
statistics be used in tests of significance for goodness-of-fit. The
statistical package SAS uses the modification to calculate the
significance level of the Kolmogorov-Smirnov statistic.

We next consider the limit distribution of the chi-square statistic
described in Section B of Chapter Four. The limiting distribution is
that of the random variable,
\[ x_{k-3}^2 + \lambda_1 Z_1^2 + \lambda_2 Z_2^2 \text{ for } k = 3, 4, \ldots, \quad (10.1) \]

where \( x_{k-3}^2 \) is a chi-square random variable, with \( k - 3 \) degrees of freedom, independent of \((Z_1, Z_2)' \sim N_2(0, I_2)\). The constants \( \lambda_1 \), \( \lambda_2 \) were given by Watson (1957), and were discussed in Chapter Four. Dahiya and Gurland (1972) presented percentage points for the distribution of the random variable in (10.1) for \( 3 < k < 15 \). We reproduce Dahiya and Gurland's table in Table 10.3 below.

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Percentage points</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>10.0%</td>
</tr>
<tr>
<td>3</td>
<td>2.371</td>
</tr>
<tr>
<td>4</td>
<td>3.928</td>
</tr>
<tr>
<td>5</td>
<td>5.442</td>
</tr>
<tr>
<td>6</td>
<td>6.905</td>
</tr>
<tr>
<td>7</td>
<td>8.322</td>
</tr>
<tr>
<td>8</td>
<td>9.703</td>
</tr>
<tr>
<td>9</td>
<td>11.055</td>
</tr>
<tr>
<td>10</td>
<td>12.384</td>
</tr>
<tr>
<td>11</td>
<td>13.694</td>
</tr>
<tr>
<td>12</td>
<td>14.988</td>
</tr>
<tr>
<td>13</td>
<td>16.267</td>
</tr>
<tr>
<td>14</td>
<td>17.535</td>
</tr>
<tr>
<td>15</td>
<td>18.792</td>
</tr>
</tbody>
</table>
When the number of classes $k$ is large, the distribution of the random variable in (10.1) is very close to a chi-square distribution with $k - 3$ degrees of freedom. This is because the $\lambda_i$'s are positive decreasing functions of $k$ bounded above by one. For smaller values of $k$ the distribution in Table 10.3 differs markedly from that of a $X_{k-3}^2$ random variable. It is recommended that Table 10.3 be used in tests of significance for goodness-of-fit tests based on the chi-squared tested described in Chapter Four.

We next consider a distribution associated with the outlier test described in Chapter Five. Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ random variables. The distribution of the random variable,

$$\max_{1 \leq t \leq n} \left| \frac{1}{\sigma^2} (X_t - \bar{X}) \right|,$$

called the maximum normed residual, has been tabulated for various values of $n$ by Stefansky (1972). We use these results to tabulate the distribution of the random variable,

$$\max_{1 \leq t \leq n} \left| \frac{1}{\sigma^2} (X_t - \bar{X}) \right|,$$

in Table 10.4 below.

We present the distribution of the random variable in (10.3) because it is easier to use than the random variable in (10.2). In practice we can directly compare the largest standardized regression residual to the values in Table 10.4.
0.4. Upper percentage points for the distribution of the random variable $\max_{1 \leq t \leq l} \left| m_{XX}^{-1/2} (X_t - \bar{X}) \right|$ for normal samples (Stefansky, 1972)

<table>
<thead>
<tr>
<th>Size of sample</th>
<th>Percentage points</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>5.0%</td>
</tr>
<tr>
<td>5</td>
<td>1.716</td>
</tr>
<tr>
<td>6</td>
<td>1.887</td>
</tr>
<tr>
<td>7</td>
<td>2.021</td>
</tr>
<tr>
<td>8</td>
<td>2.127</td>
</tr>
<tr>
<td>9</td>
<td>2.215</td>
</tr>
<tr>
<td>10</td>
<td>2.289</td>
</tr>
<tr>
<td>11</td>
<td>2.356</td>
</tr>
<tr>
<td>12</td>
<td>2.411</td>
</tr>
<tr>
<td>13</td>
<td>2.463</td>
</tr>
<tr>
<td>14</td>
<td>2.506</td>
</tr>
<tr>
<td>15</td>
<td>2.548</td>
</tr>
<tr>
<td>16</td>
<td>2.587</td>
</tr>
<tr>
<td>17</td>
<td>2.620</td>
</tr>
<tr>
<td>18</td>
<td>2.651</td>
</tr>
<tr>
<td>19</td>
<td>2.681</td>
</tr>
<tr>
<td>20</td>
<td>2.707</td>
</tr>
<tr>
<td>22</td>
<td>2.759</td>
</tr>
<tr>
<td>24</td>
<td>2.801</td>
</tr>
<tr>
<td>26</td>
<td>2.840</td>
</tr>
<tr>
<td>28</td>
<td>2.879</td>
</tr>
<tr>
<td>30</td>
<td>2.908</td>
</tr>
</tbody>
</table>