Estimators for a nonlinear functional relationship

Kirk Marcus Wolter
Iowa State University

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

Part of the Statistics and Probability Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
WOLTER, Kirk Marcus, 1948-
ESTIMATORS FOR A NONLINEAR FUNCTIONAL RELATIONSHIP.

Iowa State University, Ph.D., 1974
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106
Estimators for a nonlinear functional relationship

by

Kirk Marcus Wolter

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

1974
# TABLE OF CONTENTS

I. INTRODUCTION TO THE ERRORS-IN-VARIABLES PROBLEM

   A. Definition of the Problem .......................... 1

   B. The Linear Model Reviewed
      1. Maximum likelihood and related methods ........ 9
      2. Wald's method or the method of grouping ........ 14
      3. Instrumental variables ................................ 15
      4. Estimation via replicated observations .......... 16
      5. The method of cumulants ............................ 17
      6. The method of ranks ................................ 18
      7. Bayesian methods .................................. 18
      8. Controlled observations ............................. 18
      9. Interval estimation ................................ 19

   C. The Nonlinear Model Reviewed ..................... 20

II. BACKGROUND DEFINITIONS AND THEOREMS ............ 33

   A. Definitions .................................... 33

   B. Theorems ........................................ 36

III. A PRELIMINARY ESTIMATOR FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP ... 44

   A. Notation and Assumptions ........................ 44

   B. The Estimator ................................... 50

   C. Asymptotic Properties of the Estimator ......... 53

   D. An Estimator of the Covariance Matrix of the Estimator .... 70
### IV. ESTIMATORS FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP WHEN THE ERROR VARIANCES ARE DECREASING

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. The Model and Assumptions</td>
<td>75</td>
</tr>
<tr>
<td>B. The Ordinary Least Squares Estimator</td>
<td>80</td>
</tr>
<tr>
<td>C. The Preliminary Estimator ( \hat{\beta} )</td>
<td>85</td>
</tr>
<tr>
<td>D. The Likelihood Function and Iterative Estimators</td>
<td>97</td>
</tr>
<tr>
<td>1. The maximum likelihood estimator</td>
<td>97</td>
</tr>
<tr>
<td>2. A pseudo-maximum likelihood estimator</td>
<td>102</td>
</tr>
<tr>
<td>3. Asymptotic properties</td>
<td>111</td>
</tr>
<tr>
<td>4. A revised estimator</td>
<td>134</td>
</tr>
<tr>
<td>5. Estimating the variance of the estimators</td>
<td>144</td>
</tr>
<tr>
<td>6. Preliminary estimator versus revised estimator</td>
<td>145</td>
</tr>
<tr>
<td>7. Pseudo estimators vis-à-vis maximum likelihood estimator</td>
<td>147</td>
</tr>
</tbody>
</table>

### V. ESTIMATORS FOR THE GENERAL NONLINEAR FUNCTIONAL RELATIONSHIP

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Introduction</td>
<td>151</td>
</tr>
<tr>
<td>B. The Estimator</td>
<td>154</td>
</tr>
<tr>
<td>C. The Asymptotic Properties</td>
<td>162</td>
</tr>
<tr>
<td>D. A Revised Estimator</td>
<td>184</td>
</tr>
<tr>
<td>E. Miscellaneous Remarks</td>
<td>195</td>
</tr>
</tbody>
</table>

### VI. NUMERICAL RESULTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. A Monte Carlo Study</td>
<td>198</td>
</tr>
<tr>
<td>1. Generating the data</td>
<td>198</td>
</tr>
</tbody>
</table>
2. Adjusting the estimators 202
3. Results and conclusions 207

B. An Example 220

VII. SUMMARY 239
VIII. REFERENCES 249
IX. ACKNOWLEDGEMENTS 256

X. APPENDIX A: ADDITIONAL ESTIMATORS FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP 257
A. A Method of Moments Estimator 257
B. A Pseudo Instrumental Variable Estimator 260
C. A Revised Root Type Estimator 262

XI. APPENDIX B: THE THEORY OF CONTROLLED VARIABLES FOR A POLYNOMIAL RELATIONSHIP 265

XII. APPENDIX C: THE LINEAR ERRORS-IN-VARIABLES MODEL WITH LINEAR PARAMETRIC CONSTRAINTS 271
I. INTRODUCTION TO THE ERRORS-IN-VARIABLES PROBLEM

The statistical problem of fitting a curve through a set of \( N \) points, all of whose coordinates are subject to error, was first considered as early as the 1870's. This is the classical errors-in-variables problem.

Throughout the history of the errors-in-variables problem, statisticians have devoted their major research efforts towards, what we shall define to be, the linear model, while relatively unsatisfactory efforts have been directed towards nonlinear models. It is the purpose of this thesis to provide a theory of estimation for the nonlinear model, and to consider in depth a specific nonlinear model: the quadratic model. We begin by giving a mathematical definition of the errors-in-variables model.

A. Definition of the Problem

Let \((\Omega, \mathcal{F}, \mathcal{G})\) be a probability space, let \(\Theta \subset \mathbb{R}^p\), and let \(N\) be an integer greater than \(p\). Let \(\{z_t\}_{t=1}^{\infty}\) and \(\{\epsilon_t\}_{t=1}^{\infty}\) be sequences of \(q\)-dimensional random variables defined on the space \((\Omega, \mathcal{F}, \mathcal{G})\). It will be convenient to consider \(\epsilon_t\) and \(z_t\) \((1 \times q)\) vectors and to consider elements of \(\Theta\), which we denote by \(\theta\), \((p \times 1)\) vectors.

Next, let \(G: \mathbb{R}^q \times \Theta \to \mathbb{R}^1\) be a real valued, Lebesgue measurable function. Then, for some \(\theta \in \Theta\) suppose the relationship
\[ G(z_t; \theta) = 0 \]  

holds for \( t = 1, 2, \ldots \). If we are able to observe \( z_t \) for \( t = 1, \ldots, N \), then Model 1.1 is not an errors-in-variables model. The errors-in-variables nature of Equation 1.1 manifests itself only when we are not able to observe \( z_t \) directly, but when we are able to observe the \((1 \times q)\) random variable 

\[ Z_t = z_t + \epsilon_t \]  

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

The errors-in-variables model, as just defined in Equation 1.1 and 1.2, may now be categorized according to several criteria. The form of the function \( G \) determines whether the model is linear or nonlinear. If \( p = q \) and \( G \) is of the form

\[ G(z_t; \theta) = z_t \theta \]  

then we shall refer to the model as linear. If \( G \) is not linear, then we will say the model is nonlinear.

The assumptions made concerning the vectors \( z_t \) also distinguish different categories of the errors-in-variables model. If \( z_t \) is a nonobservable fixed constant, or constant random variable, then Model 1.1 will be called a functional relationship. If \( z_t \) is a nonobservable, nonconstant random variable for each \( t \), then Model 1.1 will be called
a structural relationship. This terminology is due to Kendall (1951) who introduced it with respect to the linear model. Thus we are merely extending Kendall's terminology to include the nonlinear model. As we shall see, the way one goes about estimating $\theta$ from the observations $Z_t$, $t = 1, \ldots, N$, depends critically upon whether one assumes Equation 1.1 is a structural or functional relationship.

Before proceeding further, it is necessary for us to consider the distributions of the random variables $\epsilon_t$ and $z_t$. In this thesis we will assume

$$E(\epsilon_t) = \int_{\Omega} \epsilon_t \, d\theta = 0 \quad (1.4)$$

and

$$E(\epsilon_t^t \epsilon_s^s) = \int_{\Omega} \epsilon_t^t \epsilon_s^s \, d\theta$$

$$= \frac{1}{ts} < \infty \quad (1.5)$$

for $t = 1, 2, \ldots$ and $s = 1, 2, \ldots$. In addition, we will assume

$\{z_t\}_{t=1}^{\infty}$ is a sequence of independent random variables and that the sequences $\{z_t\}_{t=1}^{\infty}$ and $\{\epsilon_t\}_{t=1}^{\infty}$ are independent. In specific situations we may also assume the $\epsilon_t$'s are normally distributed. However, in the present general context we shall not make that assumption.

We can now see that $\theta$, together with all parameters specified in the distributions of $\{z_t\}_{t=1}^{\infty}$ and $\{\epsilon_t\}_{t=1}^{\infty}$ constitute the set of all
parameters involved in the errors-in-variables model. Neyman and Scott (1948) have introduced two terms which distinguish between the two important types of parameters in this set. \( \theta \) together with those parameters which occur in the distribution of \( Z_t = z_t + \epsilon_t \) for infinitely many \( t \), are called structural parameters. Those parameters which occur in the distribution of \( Z_t = z_t + \epsilon_t \) for only finitely many \( t \) are called incidental parameters. For example, suppose Model 1.1 is a linear, functional relationship and the \( \epsilon_t \) are independently distributed as \( N(0, \sigma^2) \). In this example, the structural parameters are \( \theta \) and \( \sigma^2 \), while the incidental parameters are the \( z_t \)'s themselves. Generally, it is only the structural parameters which we can hope to estimate consistently. We shall return to this idea when we discuss the work of Kiefer and Wolfowitz (1956).

The basic goal of our study of the errors-in-variables model will be the estimation of \( \theta \) from the observations \( \{ Z_t \}_{t=1}^N \). Naturally, we expect our method of estimation to depend on our knowledge, or lack of knowledge, of the other structural and incidental parameters in the model. But, in fact, our ability to estimate \( \theta \) often depends on our knowledge of the other parameters. This is the important concept of identifiability to which we next turn.

All the parameters in an errors-in-variables model together with the distributions of \( z_t \) and \( \epsilon_t \), for each \( t \), generate one and only one
distribution of the observed variables, \( \{Z_t\}_{t=1}^N \). However, a particular distribution of the observed variables may be generated by differing sets of parameters as well as different distributions of the unobserved variables, \( \{z_t\}_{t=1}^\infty \) and \( \{\varepsilon_t\}_{t=1}^\infty \). This led Reiersøl (1950) to call the Model 1.1, 1.2 identifiable if \( \theta \) could be uniquely determined from knowledge of the distribution function of the observed variables, \( \{Z_t\}_{t=1}^N \). Contrariwise, he called the model nonidentifiable if \( \theta \) could not be uniquely determined. It is clear that we can hope to estimate the parameters in an errors-in-variables model if and only if the model is identifiable. Thus we will use these terms also to describe the parameters in the model.

Thus far we have introduced the errors-in-variables model, stated some general assumptions, and defined a number of terms which we will need in our dealings with the model. In the remainder of this chapter, we consider the linear model, the nonlinear model, and methods of estimation appropriate to each.

B. The Linear Model Reviewed

To facilitate our discussion of the linear model, we introduce an alternative notation to that used in Equations 1.1 and 1.2. Define the sequences of random variables \( \{y_t\}_{t=1}^\infty \), \( \{Y_t\}_{t=1}^N \), \( \{x_t\}_{t=1}^\infty \), \( \{X_t\}_{t=1}^N \).
\[ \{ e_t \}_{t=1}^\infty \text{, and } \{ u_t \}_{t=1}^\infty \text{ such that the following hold:} \]

\[ z_t = (y_t, \ x_t), \]
\[ Z_t = (Y_t, \ X_t), \]

and

\[ \epsilon_t = (e_t, \ u_t), \]

where \( y_t, \ Y_t, \) and \( e_t \) are scalars and \( x_t, \ X_t, \) and \( u_t \) are \((1 \times q-1)\).

Now define the \((N \times q)\) matrices \( Z, \ z, \) and \( \epsilon; \) the \((N \times q-1)\) matrices \( X, x, \) and \( u; \) the \((N \times 1)\) matrices \( Y, y, \) and \( e; \) and the \((p \times 1)\) vector \( \theta \) as

\[
Z = \begin{bmatrix}
Z_1 \\
Z_2 \\
\vdots \\
Z_N
\end{bmatrix}, \quad z = \begin{bmatrix}
z_1 \\
z_2 \\
\vdots \\
z_N
\end{bmatrix}, \quad \epsilon = \begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_N
\end{bmatrix},
\]

\[
X = \begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_N
\end{bmatrix}, \quad x = \begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_N
\end{bmatrix}, \quad u = \begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_N
\end{bmatrix},
\]

\[
Y = \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_N
\end{bmatrix}, \quad y = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_N
\end{bmatrix}, \quad e = \begin{bmatrix}
e_1 \\
e_2 \\
\vdots \\
e_N
\end{bmatrix}.
\]
and 

\[ q = \begin{bmatrix} 1 \\ -\beta \end{bmatrix} \, . \]

Letting \( p = q \), we can now rewrite the linear model, as given in Equation 1.3, as

\[ y = x\beta \quad (1.6) \]

where we can observe

\[ Y = y + \epsilon \]

and

\[ X = x + u \, . \quad (1.7) \]

In accord with this new notation, we shall partition the error covariance structure of Equation 1.5 as follows:

\[ \Phi_{ts} = \begin{bmatrix} \sigma_{e_t e_s} & \Phi_{e_t u_s} \\ \Phi_{u_t e_s} & \Phi_{u_t u_s} \end{bmatrix} \, . \quad (1.8) \]

Then by Equations 1.4 and 1.5 we see that

\[ E(e_t) = 0 \, , \]
\[ E(u_t) = 0 \, , \]
\[ E(e_t e_s) = \sigma_{e_t e_s} \, , \]
\[ E(e_t u_s) = \Phi_{e_t u_s} \, , \]
\[ E(u_t e_s) = \Phi_{u_t e_s} \, , \]
\[ E(u_t u_s) = \Phi_{u_t u_s} \, . \]
and
\[ E(u_t'u_s) = \frac{1}{t} u_t'u_s \]

for \( t = 1, 2, \ldots \) and \( s = 1, 2, \ldots \), where \( e_t'e_s \) is \((1 \times 1)\), \( e_t'u_s \) is \((1 \times q-1)\) and \( u_t'u_s \) is \((q-1 \times q-1)\).

Although no distributional assumption regarding the error vectors, \( \varepsilon_t \), has yet been made, it is most common to assume \( \varepsilon_t = (e_t', u_t') \) is distributed as a \( q \)-variate normal random variable. Reiersøl (1950) considered this case and proved a result which is of fundamental importance to the estimation of \( \beta \) in Equation 1.6. Under the additional assumption that

\[ \frac{1}{t} u_t'u_s = \frac{1}{t} \quad \text{for} \quad t = s \\
= 0 \quad \text{for} \quad t \neq s , \]

with \( \frac{1}{t} \) unknown, Reiersøl showed \( \beta \) is identifiable if and only if the \( z_t \)'s are not normally distributed, or are not fixed constants. The implication of this is depressingly clear: in the simplest, most familiar situations, we can not hope to estimate \( \beta \) without knowledge of the error covariance matrix.

In view of this result let us now look at the various methods that have been used to estimate \( \beta \). We identify seven such methods:
1. maximum likelihood, least squares, and related procedures;

2. Wald's method or the method of grouping;

3. instrumental variables;

4. methods using replicated observations;

5. the method of cumulants;

6. the method of ranks; and

7. Bayesian methods.

Let us consider each of these methods in turn.

1. **Maximum likelihood and related methods**

   As would be expected, the first work on the linear model was for the simple case where \( q = p = 2 \). Here, \( x \) is an \((N \times 1)\) vector and \( \beta \) is a scalar parameter to be estimated.

   Assuming both \( u \) and \( e \) are nonzero, R.J. Adcock (1877, 1878) suggested \( \beta \) be estimated by minimizing the sum of squares of the perpendicular distances from the observed points to the fitted line. K. Pearson (1901) independently suggested this estimator in 1901. A fundamental objection to this procedure is that it is not invariant under transformations of the coordinate system.

   In 1879 C.H. Kummell (1879) suggested a procedure which is invariant under transformations of the coordinate system, provided some information concerning the error variances is available. His
procedure is to minimize the sum of squares of the weighted distances from the observed points to the fitted line, where the weights are proportional to the inverses of the variances of $e_t$ and $u_t$.

Gini (1921) considered the structural relationship with

$$
\Phi_{ts} = \Phi = \begin{bmatrix}
\sigma_e^2 & 0 \\
0 & \sigma_u^2 \\
\end{bmatrix}, \quad t = s
$$

$$
= 0 \quad \text{, } t \neq s
$$

and showed $\beta$ is identifiable if we know the variance of $e_t$, the variance of $u_t$, or the ratio of the two variances. He also showed his estimate to be bounded by the estimate from the least squares regression of $Y$ on $X$ and by the reciprocal of the estimate from the least squares regression of $X$ on $Y$.

Under the assumption that the $\xi_t$ are independently distributed as $N(0, \Phi)$ and $\Phi = \text{diag}$(\(\sigma_e^2, \sigma_u^2\)) is unknown, Dent (1935) purported to have found the maximum likelihood estimates of the structural parameters, say $\hat{\beta}$, $\hat{\sigma}_e^2$, and $\hat{\sigma}_u^2$, for the functional model. However, in view of Reiersøl's (1950) result on identifiability we know Dent's estimates are necessarily unsatisfactory. This point has been thoroughly elucidated by Lindley (1947) and Solari (1969). Lindley showed that Dent's estimates satisfied the relation $\hat{\sigma}_e^2 = \hat{\beta}^2 \hat{\sigma}_u^2$, and argued that since this relation was not specified in the original model, the
estimates could not be consistent. Solari has recently shown that Dent's estimates did not occur at the maximum of the likelihood, but at a saddle point.

Let us consider the general case where $q = p$ is not necessarily 2. Following the work of M. J. van Uven (1930), Koopmans (1937) found the maximum likelihood estimator of $\beta$ for the functional model when the $\epsilon_t$ are independently distributed as $N(0, \delta)$, $\delta$ known up to a multiple. He showed the estimate to be consistent. Malinvaud (1966), following Lawley (1953), presented an asymptotic covariance matrix for Koopmans' estimate.

Assuming $\delta$ is unknown, but an estimate $S$ of $\delta$ is available, F. S. Acton (1959) suggested the estimator of $\beta$ obtained by replacing $\delta$ by $S$ in Koopmans' maximum likelihood estimator. Villegas (1961) showed that the estimator thus obtained is the maximum likelihood estimate when $\delta$ is unknown, provided $S$ is distributed as a Wishart with mean $\delta$ and is independent of $Z_t$ for $t = 1, \ldots, N$.

Villegas (1966) considered the problem of estimating the linear functional relationship when the number of distinct $x_t$ is finite, but where there is an increasing number of replicate observations for each $x_t$. He showed simple least squares is asymptotically efficient within a certain class of estimators which he called "ordinary estimators."
In 1971, Fuller (1971) studied several cases of the functional relationship, each with differing knowledge of the homogeneous error covariance matrix, $\Omega$. He modified the maximum likelihood estimators in such a way as to guarantee the existence of finite moments. Also, he derived the asymptotic distribution of each estimator and showed his estimators to have smaller mean square error than the usual, maximum likelihood estimators.

DeGracie and Fuller (1972) considered the problem of estimating the slope in the analysis of covariance when the concomitant variable is subject to error.

Sprent (1966) presented a very general estimator which does not require

$$
\Omega_{ts} = \Omega, \quad t = s \\
= 0, \quad t \neq s;
$$

does not require the errors $\epsilon_t$ be normally distributed; but does require the matrices $\Omega_{ts}$ be known for all $t$ and $s$. He called his estimator the generalized least squares estimator. However, Sprent did not develop the statistical properties of this estimator.

In a similar, general setting, Booth (1973) considered estimation for the functional model where

$$
\Omega_{ts} = \Omega_{tt}, \quad t = s \\
= 0, \quad t \neq s$$
and $\theta_t$ is known for each $t$. He presented both a preliminary and a revised estimator of $\beta$, each of which is consistent and satisfies the relation $\frac{\hat{\beta} - \beta}{N^{-1/2}} = O_p(N^{-1/2})$. The asymptotic distribution of each estimator was also derived.

We conclude this section by discussing two relevant papers on the maximum likelihood method. In considering the maximum likelihood estimator of $\beta$ for the functional relationship, we are confronted by the unknown incidental parameters $x_t$. Neyman and Scott (1948) were the first to demonstrate that the maximum likelihood method is not necessarily consistent when incidental parameters are present. Indeed, as Solari (1969) pointed out, the maximum likelihood estimates of the structural parameters for the functional model do not even exist unless $x$, or parts of $x$, are known. This result was implicit in the work of Reiersøl (1950).

For the structural relationship, maximum likelihood may not be in as much trouble. Kiefer and Wolfowitz (1956) have shown that if the incidental parameters, $x_t$, are independent, identically distributed and if the structural parameters are identifiable, then the maximum likelihood estimators of the structural parameters are consistent under regularity conditions. They illustrate this result with an example where $x_t$ has a nonnormal distribution.
2. Wald's method or the method of grouping

In 1940 Wald (1940) presented an estimator of $\beta$ which has drawn considerable attention. To be precise, Wald considered the functional model

$$y = \beta_0 + \beta_1 x,$$

where $x$ is $(N \times 1)$ and where the $\epsilon_t = (e_t, u_t)$ are uncorrelated with zero mean and

$$\Sigma_{ts} = \begin{bmatrix} \sigma^2_e & 0 \\ 0 & \sigma^2_u \end{bmatrix}, t = s,$$

$$= 0, \quad t \neq s.$$

He obtained estimators of $\beta_0$, $\beta_1$, $\sigma^2_e$, and $\sigma^2_u$ by dividing the $N$ observations into two equal groups. Further, he demonstrated consistency for these estimators under the assumptions that the grouping is independent of the errors, $\epsilon_t$, and that

$$\lim \inf_{N \to \infty} \left| \frac{\Sigma x_t}{N} - \frac{\Sigma x_t}{N} \right| > 0.$$

Unfortunately, these assumptions are rarely met, though Wald did present a pathological example for which the assumptions hold.
Since Wald's original publication, numerous statisticians have been attracted by and have published on this same general method. M.S. Bartlett (1949) presented a modification of Wald's method in which he divided the observations into three groups, ignoring the observations in the middle group. He showed this method to have greater efficiency than Wald's original method, provided \( u = 0 \).

Nair and Banerjee (1942) discussed the efficiency of Wald type estimators and demonstrated via the Monte Carlo technique that Bartlett's modification was more efficient than Wald's original method when \( u \neq 0 \). Dorff and Gurland (1961a, 1961b) examined the small sample bias and mean square error of Wald's estimators.

Gibson and Jowett (1957a, 1957b) analyzed the efficiency of Wald's method and extended it to the case of multiple regression with two independent variables, i.e. \( x \) is \((Nx2)\). Hooper and Thiel (1956) proposed an alternative multivariate generalization and studied the efficiency of their estimator vis-à-vis Gibson and Jowett's estimator.

3. **Instrumental variables**

An instrumental variable, say \( w \), is a random variable which is uncorrelated with the errors, \( \epsilon \), but is correlated with the true values, \( x \). Extensive consideration has been given to constructing consistent estimates of \( \beta \) through use of instrumental variables. They are of interest in that they provide an alternative to knowledge of the error.
covariance matrix as a means of identifying the model.

Reiersøl (1945) was the first to employ instrumental variables in the estimation of a structural relation. He assumed the existence of two linearly related instrumental variables, $w_1$ and $w_2$, where the parameters of the linear relation are known, and where $w_1$ and $w_2$ are observed with error. Then, using observations on the instrumental variables, say $W_1$ and $W_2$, he was able to consistently estimate the simple linear model, i.e. $p = q = 2$.

Geary (1949) considered a situation in which observations on only one instrumental variable, say $W$, are available, and where $W$ is observed without error. For the structural model with $q = p = 2$, he showed his estimator of $\beta$ to be consistent, provided $\text{Cov}(W, X) \neq 0$.

4. Estimation via replicated observations

Another kind of additional information which enables us to identify the linear model is the presence of replicated observations. That is, we suppose the availability of observations

$$X_{tj} = x_t + u_{tj}$$
and
$$Y_{tj} = y_t + \epsilon_{tj}$$
for $t = 1, \ldots, N$ and $j = 1, \ldots, M_t$.

Assuming $\{(y_t, x_t)\}_{t=1}^{\infty}$ is a sequence of independent random variables; $(\epsilon_{tj}, u_{tj})$ is independent of $(\epsilon_{si}, u_{si})$ for $t \neq s$ or $j \neq i$; and
(y_t, x_t) is independent of (e_{s,j}, u_{s,j}) for all t, s, and j, we can perform a one-way analysis of variance on the X's and Y's, and thus obtain estimators of \( \beta \).

This method has been described by Tukey (1951). He gave several estimators of \( \beta \), applicable to both the functional and structural model, and showed them to be consistent as \( N \to \infty \) and as \( M_t \to \infty \), for at least one \( t \).

Housner and Brennan (1948) presented another estimator of \( \beta \) for this situation. Their estimator is consistent as \( M_t \to \infty \), for at least two distinct \( t \)'s, independently of \( N \).

The aforementioned work of Villegas (1961, 1966) is also applicable to the case of replicated observations.

5. **The method of cumulants**

For the structural model, Geary (1942, 1943) observed that the cumulants of \( Z_t \) are equal to those of \( z_t \), provided the elements of \( \epsilon_t \) are independent. He was thus able to construct linear equations in \( \beta \) which could be solved to yield consistent estimators of \( \beta \).

However, when the \( z_t \)'s are normally distributed all cumulants of degree greater than two are zero. Thus, as Geary noted, the method fails for this, the most common, structural relationship.

For functional relationships, there are no cumulants of the \( z_t \)'s. Thus the method fails again. These failures of the method are, of
course, consistent with Reiersøl's identifiability result.

6. **The method of ranks**

For the functional relationship where $x$ is $(N \times 1)$, Theil (1950) has constructed an estimator of $\beta$ with the use of ranks. His method, however, hinges on the unrealistic assumption that the observations, $X_t$, are in the same order as the true values, $x_t$.

7. **Bayesian methods**

Very recently, Lindley and El-Sayyad (1968) have attacked the functional relationship with a Bayesian analysis. They developed a Bayesian approach to the general case of incidental and structural parameters, and then specialized this approach to the linear functional relationship. For this relationship, they derived the joint posterior density of the structural parameters, and for a variety of prior distributions looked at approximations to the marginal posterior distribution of the slope.

8. **Controlled observations**

Before leaving point estimation of $\beta$, let us consider the important contribution of Berkson (1950). Berkson added two new words to the terminology of errors-in-variables. He defined an uncontrolled experiment to be an experiment in which the experimenter is not attempting to measure $X$ according to preassigned schedule, but is
merely observing $X$ and the associated $Y$. A controlled experiment, according to Berkson, is an experiment in which the experimenter is observing $X$ and the associated $Y$, where $X$ is fixed. That is, $X$ is adjusted to a series of preassigned values, while the true, unknown values, $X$, are fluctuating. In Berkson's terminology, we have thus far been considering uncontrolled experiments. For controlled experiments, Berkson demonstrates that ordinary least squares yields a consistent estimate of $\beta$ in the linear model. See also Lindley (1953) and Scheffe (1958) on this matter.

9. Interval Estimation

Several authors have addressed the problem of interval estimation of $\beta$. Wald (1940) derived confidence limits for $\beta_0$ and $\beta_1$ in Equation 1.9 through use of his grouping technique. Durbin (1954), Williams (1955, 1959), and Bartlett (1957) have pointed out that the instrumental variable technique enables one to obtain confidence regions for $\beta$. The problem of finding confidence limits for the slope in Equation 1.9 when the ratio of error variances is known has been solved by Creasy (1956), while R. L. Brown (1957) has found a confidence region for the line when both error variances are known.

More recently, Villegas (1964) has considered interval estimation for the functional relationship with replicate observations and independent, identically and normally distributed errors, $\epsilon_t$. He gives a test
for testing whether the unknown relation is a given relation, and de-
defines a confidence region for \( \beta \) to be the set of all points which lie on hyperplanes not rejected by the test. However, he is not able to give the exact confidence coefficient of the corresponding region.

Before considering nonlinear models, mention must be made of the works of Kendall and Stuart (1961) and Madansky (1959). Both give excellent reviews of the linear model and go into far more detail than was possible here. See also Moran (1971) for a very recent review of the linear model.

C. The Nonlinear Model Reviewed

To review the nonlinear model, it will be helpful to present an alternative model specification to that given in Equations 1.1 and 1.2. Define the sequences of random variables \( \{y_t\}_{t=1}^{\infty} \), \( \{Y_t\}_{t=1}^{N} \), \( \{x_t\}_{t=1}^{\infty} \), \( \{X_t\}_{t=1}^{\infty} \), \( \{e_t\}_{t=1}^{\infty} \), and \( \{u_t\}_{t=1}^{\infty} \) as in Section B; and define the random matrices \( Z, z, X, x, Y, y, \epsilon, e, \) and \( u \) also as in Section B. Using these definitions, the nonlinear model given by Equations 1.1 and 1.2 may be rewritten as

\[
G(y_t, x_t; \theta) = 0
\]  

where we can observe
\[ Y_t = y_t^* + \epsilon_t \]  
\[ X_t = x_t^* + u_t \]  
(1.11)

for \( t = 1, 2, \ldots, N \) and where \( p \) does not necessarily equal \( q \).

In considering Model 1.10 we will recall Equations 1.4 and 1.5 and will further assume

\[ \delta_{ts} = 0, \quad t \neq s, \]  
(1.12)

unless otherwise stated. We will partition \( \delta_{tt} \) as in Equation 1.8.

In addition, to avoid the nonerrors-in-variables aspects of Model 1.10, 1.11, we will assume \( e_s \neq 0 \), for some \( s \), and \( u_t \neq 0 \), for some \( t \).

While no identifiability theorem, i.e. Reiers\'s, has been proved for Model 1.10, it is nevertheless clear that \( \theta \) is not, in general, identifiable. Additional information of some kind will generally be needed to estimate \( \theta \).

It is possible to extend Geary's method of cumulants (1942, 1943, 1949), Wald's method of grouping (1940) (cf. Nair and Shrivastava (1942)), and Theil's method of ranks (1950) to polynomial models, i.e. \( G \) is of the form

\[ y_t = \beta_0 + \beta_1 x_t + \ldots + \beta_{q-2} x_t^{q-2} \]  
(1.13)
where \( q > 4 \). However, each of these methods has the pitfalls noted in Section B. For the single best reference to this material, consult Kendall and Stuart (1961).

The analysis of controlled variables is not necessarily extendable. Whereas Berkson (1950) demonstrated ordinary least squares to be appropriate in the estimation of the linear model with controlled observations, Geary (1953) has shown that ordinary least squares may not be appropriate in the estimation of polynomial models with controlled observations. He considered the cubic model, i.e.

\[
y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3,
\]

with independent, identically and normally distributed errors. For this case he was able to show ordinary least squares consistently estimates

\[
(\beta_0 + \beta_2 \sigma_u^2), (\beta_1 + 3\beta_3 \sigma_u^2), \beta_2, \beta_3,
\]

where \( \sigma_u^2 = \text{Var}(u_1) \). Thus, without knowledge of \( \sigma_u^2 \), only \( \beta_2 \) and \( \beta_3 \) are identifiable.

Very recently, Fedorov (1974) considered the general nonlinear errors-in-variables model with controlled variables. He assumed known error variances and suggested an iterative estimation procedure
based on the least squares principle. Fedorov established some convergence properties for this estimator.

We shall now review estimation procedures based either on the maximum likelihood principle or on the least squares principle. Each of these methods relies on knowledge of the error variances or on replicated observations for model identification, and each deals with the functional relationship.

Possibly the earliest attempt to estimate $\theta$ in Model 1.10 was by Deming (1931, 1943). His method was based on the least squares principle. That is, to estimate $\theta$ he suggested minimizing the sum of squares

$$ Q = \sum_{t=1}^{N} (Z_t - z_t)^2 + \sum_{t} \lambda_t (Z_t - z_t) $$

subject to the $N$ conditions

$$ G(z_t; \theta) = 0. $$

The minimization was accomplished by defining the function

$$ L(z, \theta, \lambda) = Q + \sum_{t=1}^{N} \lambda_t G(z_t; \theta), $$

where $\lambda' = (\lambda_1 \ldots \lambda_N)$ are LaGrange multipliers. However, instead of minimizing Equation 1.15 with respect to $(z, \theta, \lambda)$, Deming minimized
\[ L_0(z, \theta, \lambda) = Q + \sum_{t=1}^{N} \lambda_t G_0(z_t; \theta) \]  
(1.16)

where \( G_0(z_t; \theta) \) is the linear portion of the Taylor expansion of \( G(z_t; \theta) \) about the point \((Z_t; \theta_0)\), and \( \theta_0 \) is an "initial guess" at \( \theta \).

Differentiating Equation 1.16 with respect to \((z, \theta, \lambda)\) and setting the derivatives to zero yielded a system of linear equations in \((z, \theta, \lambda)\) which Deming was able to solve, the resulting estimators being denoted by

\[ (\hat{z}, \hat{\theta}, \hat{\lambda}) \] .

Deming gave an approximation to the covariance matrix of \( \hat{\theta} \) but did not investigate the conditions under which the approximation would be valid.

Also in 1931, Cook (1931) expanded upon Deming's work. Cook pointed out that the ratio of the error variances is all that need be known to apply Deming's method, not the variances themselves. In addition, he suggested an iterative estimation scheme where the functions \( G(z_t; \theta) \) are relinearized about the most recent estimates, \((\hat{z}, \hat{\theta})\), and new estimates are computed via Deming's technique. He concludes his paper with a rather archaic discussion of the errors of estimation.

Recently, several authors have revived the general notion of model linearization as proposed in the methods of Deming and Cook. Dolby
and Lipton (1972) considered a specific form of the Model 1.10, namely

\[ y_t - g(x_t; \theta) = 0 \quad (1.17) \]

where \( g(x; \theta) \) is a function with finite, continuous first derivatives with respect to \( x \) and \( \theta \). Replicate observations were used to identify the model and to estimate the covariance matrices \( \Sigma_{tt} \). Assuming normally distributed errors, \( \epsilon_t \), they substituted Equation 1.17 directly into the log likelihood function, and proposed solution of the likelihood equations by the Newton-Raphson technique. Dolby and Lipton, without giving conditions, stated that the inverse of the information matrix is the asymptotic covariance matrix of the estimates of the structural parameters. The statement is not necessarily applicable in the presence of infinitely many incidental parameters (cf. Neyman and Scott (1948); Kiefer and Wolfowitz (1956).

Also in 1972, Dolby (1972) considered the Model 1.17 with normal errors. In contrast to his earlier paper with Lipton, Dolby now assumed no replication, but a known, general covariance matrix with \( \Sigma_{ts} \) not necessarily 0 for \( t \neq s \). For this model, he generalized the Newton-Raphson solution of the likelihood equations given in Dolby and Lipton (1972) and made the same statement with regard to the information matrix. In addition, he proposed an alternative estimation procedure which he called generalized least squares. He showed this
procedure to be equivalent to maximum likelihood, and to Sprent's (1966) generalized least squares provided the model is linear.

Britt and Luecke (1973) have extended Cook's modification of Deming's method to cover the general error covariance structure, i.e. where $\Sigma_{ts}$ is not necessarily 0. This involved no additional principle. They also presented an approximation to the covariance matrix of their estimator without specifying rigorous conditions under which the approximation would be valid.

O'Neill et al. (1969) independently derived the Dolby method for polynomial models with normal errors.

Clutton-Brock (1967) has taken a somewhat different approach to estimating $\theta$ in Equation 1.17. For the case where $x$ is $(N \times 1)$; he advocated minimizing the pseudo-likelihood

$$\left[ \sum_{t=1}^{N} \frac{1}{\sigma_{v_t}^2} \right] \exp \left[ -\frac{1}{2} \sum_{t=1}^{N} \sigma_{v_t}^{-2} (y_t - g(x_t; \theta))^2 \right],$$

where $\Psi_{ts} = \Psi = \text{diag. } (\sigma_e^2, \sigma_u^2), t = s$

$= 0, \quad t \neq s$;

$$\sigma_{v_t}^2 = \left[ 1, -\frac{\partial}{\partial x_t} g(x_t; \theta) \right] \Psi \left[ 1, -\frac{\partial}{\partial x_t} g(x_t; \theta) \right]'$$

and the derivatives, $\frac{\partial}{\partial x_t} g(x_t; \theta)$, are evaluated at "initial guesses"
of \((x; \theta)\). Heuristic arguments were given in support of the resulting estimator.

It is important for us to note that each of the above methods, from Deming's through Clutton-Brock's, suffers from at least two important deficiencies. First, each of these methods relies upon linear approximations of one kind or another, and thus upon initial estimators of the parameters \((x; \theta)\). Not one of the above papers devotes attention to the way in which the preliminary estimators should be chosen, other than to suggest making an "initial guess" based on a data plot. Second, no rigorous, mathematically defensible examination of the statistical properties of the final estimators has been given. None of the authors considered the consistency or other asymptotics of their estimators. Only a few authors considered the variances of their estimators, and then only in terms of vague approximations.

A notable exception to this criticism is the work of Villegas (1969). In considering Model 1.17, however, Villegas assumed a different observational scheme from that given by Equation 1.11: he considered the situation where the error variances are decreasing at the rate \(1/N\). To be precise, he assumed \(N\) replicate observations on each of \(K\) (fixed) true values:

\[
\begin{align*}
Y_{tj} &= y_t + e_{tj} \\
X_{tj} &= x_t + u_{tj}
\end{align*}
\]
for $t = 1, \ldots, K$ and $j = 1, \ldots, N$.

The estimator Villegas presented requires a preliminary estimator of $\beta$, say $\hat{\beta}$, satisfying $(\hat{\beta} - \beta) = \mathcal{O}(N^{-1/2})$, and a consistent estimator $S = \text{diag. } (s_{v_1}^2, \ldots, s_{v_k}^2)$ of $V = \text{diag. } (\sigma^2_{v_1}, \ldots, \sigma^2_{v_k})$, where

$$
\sigma^2_{v_t} = (1, -g_{x_t}(\beta; \beta))^\top (1, -g_{x_t}(\beta; \beta))' ;
$$

$$
\Sigma_{ts} = \Sigma \quad \text{for } t = s
$$
$$
= 0 \quad \text{for } t \neq s ;
$$

and $g_{x_t}(\beta; \beta)$ denotes the first derivative of $g$ with respect to $x$ evaluated at the true value $(x_t; \beta)$. The estimator is defined to be that $\beta$ which minimizes the following function:

$$
\frac{K}{\sum_{t=1}^{T} s_{v_t}^2} \left[ \overline{Y}_t - g(\overline{X}_t; \hat{\beta}) - g_{\beta}(\overline{X}_t; \hat{\beta})(\beta - \hat{\beta}) \right]^2
$$

where

$$
\overline{Y}_t = N^{-1} \sum_{j=1}^{N} Y_{tj} ,
$$

$$
\overline{X}_t = N^{-1} \sum_{j=1}^{N} X_{tj} ,
$$

and $g_{\beta}(\overline{X}_t; \hat{\beta})$ denotes the first derivative of $g$ with respect to $\beta$ evaluated at $(\overline{X}_t; \hat{\beta})$. Villegas showed that the resulting estimator, say $\hat{\beta}_v$, satisfies
\( \sqrt{N(\hat{\theta}_v - \theta)} \xrightarrow{d} N(0, \{ [G_{\theta}(x; \theta)]' V^{-1}[G_{\theta}(x; \theta)] \}^{-1}) \),

where

\[
G_{\theta} = \begin{bmatrix}
g_{\theta}(x_1; \theta) \\
g_{\theta}(x_2; \theta) \\
\vdots \\
g_{\theta}(x_k; \theta)
\end{bmatrix}.
\]

Four authors have looked at specific nonlinear models. Hey and Hey (1960) considered the estimation of \( \theta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \) where the functional relationship was the rectangular hyperbola

\[
(x_t - \beta_1)(y_t - \beta_2) = \beta_0.
\]

Their estimator was based on a variation of model linearization and the least squares principle. As with the earlier authors, they did not consider the statistical properties of their estimator.

In 1965, Chan (1965) considered the circular functional relationship in parametric representation; i.e.

\[
X_t = \xi + \rho \cos x_t + u_t,
\]
\[
Y_t = \gamma + \rho \sin x_t + e_t.
\]
\[ \theta = \begin{bmatrix} \xi \\ \eta \\ \rho \end{bmatrix}, \]

where the circle is of radius \( \rho \) centered at \((\xi, \eta)\). He assumed

\[ \Phi_{ts} = \Phi = \sigma^2 I, \quad t = s \]
\[ = 0 \quad t \neq s . \]

Chan's estimation procedure was to minimize the sum of squares of the perpendicular distances from the observed points to the fitted curve. For a certain class of error distributions, necessary and sufficient conditions for the consistency and asymptotic normality of his estimates were given. A result on maximum likelihood estimation due to Neyman and Scott (1948) provided the underlying theory for most of Chan's proofs. Unfortunately, he did not provide proof that the regularity conditions of Neyman and Scott's were met.

Kendall and Stuart (1961) briefly considered maximum likelihood estimation for the quadratic functional model

\[ y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 \quad (1.18) \]

with covariance structure
and normal errors. They found iterative methods were needed to solve the likelihood equations. No properties of the resulting estimators were given.

Griliches and Ringstad (1970) considered the bias in the ordinary least squares estimator of $\beta' = (\beta_0, \beta_1, \beta_2)$ for the quadratic structural relationship. Assuming $x_t$ was independent, identically distributed (i.i.d.) as $N(0, \sigma_x^2)$; $u_t$ was i.i.d. as $N(0, \sigma_u^2)$; and $\{x_t\}$ was independent of $\{u_t\}$, they established that

$$\text{plim } \hat{\beta}_1 = \beta_1 (1 - \lambda)$$

and

$$\text{plim } \hat{\beta}_2 = \beta_2 (1 - \lambda)^2,$$

where $\lambda = \sigma_u^2 / (\sigma_x^2 + \sigma_u^2)$ and where $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ is the ordinary least squares estimator of $\beta'$. That is, in the presence of measurement error, the estimated linear coefficient is biased towards zero by the factor $(1 - \lambda)$, while the estimated quadratic coefficient is biased towards zero by the square of the bias factor of the linear coefficient.

The remainder of this thesis is primarily concerned with the quadratic functional relationship as specified by Equation 1.18. Throughout, we assume a known error covariance matrix for model
identification.

Chapters 3 and 4 are devoted to the presentation of estimators of \( \beta' = (\beta_0, \beta_1, \beta_2) \), and to the establishment of the asymptotic properties of the estimators. Chapter 6 reports on the small sample behavior of the estimators as observed in a Monte Carlo study. Also, the estimators are illustrated with an example from the earth sciences. Then, Appendix A gives three additional estimators for the quadratic functional relationship.

A general nonlinear functional relationship is considered in Chapter 5. Estimators of the unknown parameters are presented which generalize the methods of Chapter 4. Additionally, the asymptotic properties of the estimators are studied.

Finally, Appendices B and C are devoted to two miscellaneous topics. While not in the mainstream of the thesis, these topics are intimately related to the overall problem of errors-in-variables.
II. BACKGROUND DEFINITIONS AND THEOREMS

This chapter is devoted to a number of fundamental definitions and theorems which underlie the work of the succeeding chapters. In the case of several of the theorems proofs are omitted, but references to available proofs are given.

A. Definitions

We begin by defining the concept of order as used in analysis. The following definitions are given by Fuller (1972).

Let \( \{a_n\} \) be a sequence of real numbers and \( \{g_n\} \) a sequence of positive real numbers.

**Definition 2.1**

We say \( a_n \) is of smaller order than \( g_n \) and write

\[
a_n = o(g_n)
\]

if

\[
\lim_{n \to \infty} \frac{a_n}{g_n} = 0.
\]

**Definition 2.2**

We say \( a_n \) is at most of order \( g_n \) and write

\[
a_n = O(g_n)
\]
if for a real number $M > 0$, and all $n$ greater than some finite $N_0$,

$$\left| \frac{a_n}{g_n} \right| \leq M.$$  

The concept of convergence in probability of a sequence of random variables will be indispensable to the arguments of this text. This definition is given by Chung (1968).

**Definition 2.3**

The sequence $\{X_n\}$ of random variables is said to converge in probability to $X$ if and only if for every $\epsilon > 0$ we have

$$\lim_{n \to \infty} P\{|X_n - X| > \epsilon\} = 0.$$  

This is denoted by

$$\text{plim}_{n \to \infty} X_n = X$$

or by

$$X_n \overset{p}{\to} X.$$  

Closely related to the concepts of order and convergence in probability is the notion of order in probability. This concept was first introduced by Mann and Wald (1943).

Let $\{X_n\}$ be a sequence of random variables and $\{g_n\}$ a sequence of positive real numbers.
Definition 2.4

The random variable \( X_n \) is of probability order \( o_p (g_n) \) if and only if

\[
\frac{X_n}{g_n} \xrightarrow{p} 0.
\]

In this case we write

\[
X_n = o_p (g_n).
\]

Definition 2.5

The random variable \( X_n \) is of probability order \( O_p (g_n) \) if and only if for every \( \varepsilon > 0 \) there exists a positive real number, \( M_\varepsilon \), and an \( N_\varepsilon \) such that

\[
P\{|X_n| > M_\varepsilon g_n\} \leq \varepsilon
\]

for all \( n > N_\varepsilon \). In this case we write

\[
X_n = O_p (g_n).
\]

A concept which is essential to the work of this thesis is the convergence of sequences distribution functions.

Definition 2.6

A sequence \( \{F_n\} \) of distribution functions is said to converge weakly (vaguely) to the function \( F \) if and only if there exists a dense subset \( D \) of \( \mathbb{R}^1 \) such that
for all $x \in D$ and $F$ is right continuous and monotone nondecreasing.

In this case we write

$$F_n \xrightarrow{w} F.$$ 

**Definition 2.7**

A sequence of random variables $\{X_n\}$ is said to converge in distribution to $X$ if and only if the sequence $\{F_n\}$ of corresponding distribution functions converges weakly to $F$, the distribution function of $X$. This is denoted by

$$X_n \xrightarrow{d} X$$

or by

$$X_n \xrightarrow{d} F.$$ 

**B. Theorems**

The first four theorems given below establish important results with regard to order in probability and convergence in probability.

**Theorem 2.1 (Chebyshev's Inequality)**

If $\varphi$ is a strictly positive and increasing function on $(0, \infty)$, $\varphi(\varepsilon) = \varphi(-\varepsilon)$, and $X$ is a random variable such that $E[\varphi(X)] < \infty$, then for each $\varepsilon > 0$
\[ P[|X| \geq \epsilon] \leq \frac{E[\varphi(X)]}{\varphi(\epsilon)}. \]

Proof:

(See Chung (1968), page 46.)

**Theorem 2.2**

If the sequence of random variables \( \{X_n\} \) satisfy

\[ E(X_n^2) = O(a_n^2) \]

then

\[ X_n = O_p(a_n). \]

Proof:

By Definition 2.2 there exists an \( N_0 \) and an \( M_1 \) such that

\[ E(X_n^2) < M_1 a_n^2 \]

for all \( n > N_0 \). By Chebyshev's inequality, for \( M_2 > 0 \)

\[ P[|X_n| > M_2 a_n] \leq \frac{E(X_n^2)}{M_2^2 a_n^2}. \]

Thus, given \( \epsilon > 0 \), we choose \( M_2 > M_1 \epsilon^{-1/2} \), and the result follows.

Q.E.D.

**Theorem 2.3**

If \( \{X_n\} \) is a sequence of random variables such that
\[ X_n = O_p(n^{-r}) \]

for \( r > 0 \), then

\[ X_n \xrightarrow{p} O. \]

**Proof:**

By Definition 2.5, given \( \epsilon > 0 \) there exists an \( M_\epsilon > 0 \) and an \( N_\epsilon \) such that

\[ P(\{|X_n| > M_\epsilon n^{-r}\}) \leq \epsilon \]

for all \( n > N_\epsilon \). Given \( \delta > 0 \), we define

\[ N_{\epsilon, \delta} = (M_\epsilon / \delta)^{1/r}. \]

Now we have

\[ P(\{|X_n| > \delta N_{\epsilon, \delta} n^{-r}\}) \leq \epsilon \]

for all \( n > N_\epsilon \). Thus

\[ P(\{|X_n| > \delta\}) \leq \epsilon \]

for all \( n > \max\{N_\epsilon, N_{\epsilon, \delta}\} \). **Q.E.D.**

**Theorem 2.4**

If \( \{X_n\} \) and \( \{Y_n\} \) are sequences of random variables such that

\[ X_n = O_p(f_n) \quad \text{and} \quad Y_n = O_p(g_n), \]

for \( \epsilon > 0 \), then

\[ X_n \xrightarrow{p} O. \]

**Proof:**

By Definition 2.5, given \( \epsilon > 0 \) there exists an \( M_\epsilon > 0 \) and an \( N_\epsilon \) such that

\[ P(\{|X_n| > M_\epsilon n^{-r}\}) \leq \epsilon \]

for all \( n > N_\epsilon \). Given \( \delta > 0 \), we define

\[ N_{\epsilon, \delta} = (M_\epsilon / \delta)^{1/r}. \]

Now we have

\[ P(\{|X_n| > \delta N_{\epsilon, \delta} n^{-r}\}) \leq \epsilon \]

for all \( n > N_\epsilon \). Thus

\[ P(\{|X_n| > \delta\}) \leq \epsilon \]

for all \( n > \max\{N_\epsilon, N_{\epsilon, \delta}\} \). **Q.E.D.**
where \( \{f_n\} \) and \( \{g_n\} \) are sequences of positive real numbers, then

\[
X_n Y_n = O_p(f_n g_n)
\]

and

\[
X_n + Y_n = O_p(\max\{f_n, g_n\}).
\]

If \( X_n = o_p(f_n) \) and \( Y_n = O_p(g_n) \), then

\[
X_n Y_n = o_p(f_n g_n).
\]

Proof:

(See Mann and Wald (1943).)

The next three lemmas establish some important relationships between convergence in probability and convergence in distribution.

**Theorem 2.5**

Let \( \{F_n\} \), \( F \) be the distribution functions of the random variables \( \{X_n\} \), \( X \). If

\[
X_n \overset{P}{\to} X
\]

then

\[
F_n \overset{w}{\to} F.
\]

More briefly stated, convergence in probability implies convergence in distribution.

Proof:

(See Chung (1968), page 84.)
Theorem 2.6

Let \( \{X_n, Y_n\} \) be a sequence of pairs of random variables. If

\[
|X_n - Y_n| \stackrel{P}{\to} 0 \quad \text{and} \quad Y_n \xrightarrow{d} F,
\]

then

\[
X_n \xrightarrow{d} F.
\]

Proof:

(See Rao (1965), page 101.)

Theorem 2.7

Let \( \{X_n, Y_n\} \) be a sequence of pairs of random variables. If

\[
X_n \xrightarrow{d} X \quad \text{and} \quad Y_n \xrightarrow{P} C
\]

then

\[
X_n Y_n \xrightarrow{d} CX.
\]

Proof:

(See Rao (1965), page 102.)

The next two theorems are the two forms of the central limit theorem which will be employed in this thesis.
Theorem 2.8 (Liapounov Central Limit Theorem)

For each \( n > 1 \) let there be \( b^*_n \) independent random variables \( \{X_{nj}, 1 \leq j \leq b^*_n\} \), where \( b^*_n \to \infty \) as \( n \to \infty \). Furthermore, let

\[
\begin{align*}
E(X_{nj}) &= \mu_{nj}, & \var(X_{nj}) &= \sigma^2_{nj} \\
E(S_n) &= \sum_{j=1}^{b^*_n} \mu_{nj} = \mu_n, & \var(S_n) &= \sum_{j=1}^{b^*_n} \sigma^2_{nj} = \sigma^2_n, \\
E(|X_{nj}|^{2+\delta}) &= \gamma_{nj} < \infty, & \Gamma_n &= \sum_{j=1}^{b^*_n} \gamma_{nj},
\end{align*}
\]

where \( S_n = \sum_{j=1}^{b^*_n} X_{nj} \) and \( \delta > 0 \). If

\[
\frac{1}{(\sigma_n)^{2+\delta}} \Gamma_n \to 0
\]

as \( n \to \infty \), then the random variable

\[
\frac{S_n - \mu_n}{\sigma_n}
\]

converges in distribution to \( N(0, 1) \).

Proof:

(See Chung (1968), page 185.)
Theorem 2.9

Let \( F_n \) denote the joint distribution function of the \( k \) dimensional random variable \( X_n, n = 1, 2, \ldots \), let \( F_{\lambda n} \) denote the distribution function of \( \lambda' X_n \), and let \( F \) denote the joint distribution function of a \( k \) dimensional random variable \( X \). If for each nonzero vector \( \lambda \),

\[
F_{\lambda n} \xrightarrow{w} F_{\lambda},
\]

the distribution function of \( \lambda' X \), then

\[
F_n \xrightarrow{w} F.
\]

Proof:

(See Rao (1965), page 108.)

We conclude this chapter with two theorems which incorporate order in probability concepts with the classical Taylor series.

Theorem 2.10

If \( \{X_n\} \) is a sequence of scalar random variables with

\[
X_n = a + O_p (r_n),
\]

where \( r_n \xrightarrow{n} 0 \), and if \( g(x) \) is a function with \( s \) continuous derivatives at \( x = a \), then
\[ g(X_n) = g(a) + g^{(1)}(a)(X_n - a) + \ldots \]

\[ + \frac{1}{(s-1)!} g^{(s)}(a)(X_n - a)^{s-1} + O_p(r_n^s) , \]

where \( g^{(s)}(a) \) is the \( s \)th derivative of \( g(x) \) evaluated at \( x = a \).

The analogous result holds for vector random variables.

**Proof:**

(See Fuller (1972).)

---

**Theorem 2.11**

If \( A_n \) is a \((p \times p)\) nonsingular matrix, \( B_n \) is a \((p \times p)\) matrix, the elements of \( B_n \) are \( O_p(n^{-r}) \), \( r > 0 \), and \((A_n + B_n)^{-1}\) exists, then

\[
(A_n + B_n)^{-1} = A_n^{-1} - A_n^{-1}B_nA_n^{-1} + O_p(n^{-2r})
\]

\[
= A_n^{-1} + O_p(n^{-r}) .
\]

**Proof:**

(See Fuller (1972).)
III. A PRELIMINARY ESTIMATOR FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP

Our general approach to estimating the unknown structural parameters in the quadratic errors-in-variables model will be as follows: first, we shall construct a preliminary estimator whose error will be \( O(N^{-1/2}) \); and second, we shall construct a revised estimator whose definition requires the preliminary estimator. In this chapter we shall introduce one preliminary estimator and investigate its asymptotic properties. We first restate the quadratic model and introduce the notation needed in dealing with our estimator.

A. Notation and Assumptions

The quadratic functional relationship is given by

\[
y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2
\]

where

\[
Y_t = y_t + e_t
\]

\[
X_t = x_t + u_t
\]

can be observed for \( t = 1, 2, \ldots, N \), and \( \{x_t\}_{t=1}^{\infty} \) is a sequence of fixed constants, i.e. constant random variables. For the errors of measurement we assume
45

\[ \text{E}(e_t) = 0 , \quad (3.2a) \]
\[ \text{E}(u_t) = 0 , \]

and

\[ E \left\{ \begin{bmatrix} e_t \\ u_t \end{bmatrix} \begin{bmatrix} e_s \\ u_s \end{bmatrix} \right\} = \begin{bmatrix} \sigma_{e(t)}^2 & \sigma_{eu(t)}^2 \\ \sigma_{ue(t)}^2 & \sigma_{u(t)}^2 \end{bmatrix} < \infty , \quad t = s \]
\[ = 0 , \quad t \neq s . \]

To define our estimator, we shall require additional notation.

Define the (1 x 3) vectors \( w_t, f_t, \) and \( W_t \) by setting

\[ w_t = [1, x_t, x_t^2] ; \]
\[ f_t = [0, u_t, 2x_t u_t + (u_t^2 - \sigma_{u(t)})] ; \quad (3.3) \]

and

\[ W_t = w_t + f_t \]
\[ = [1, X_t, X_t^2 - \sigma_{u(t)}] . \]

Define the (1 x 4) vectors \( z_t, \epsilon_t, \) and \( Z_t \) by setting

\[ z_t = (y_t, w_t) , \]
\[ \epsilon_t = (e_t, f_t) , \quad (3.4) \]

and

\[ Z_t = z_t + \epsilon_t = (Y_t, W_t) . \]
And, define

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (3.5)$$

and

$$\theta = \begin{bmatrix} 1 \\ -\beta \end{bmatrix} .$$

Note that Model 3.1a may now be expressed as

$$z_t \theta = 0 \quad (3.6a)$$

where we can observe

$$Z_t = z_t + \epsilon_t \quad (3.6b)$$

for \( t = 1, 2, \ldots, N. \)

It will also be convenient to introduce the following notation:

$$M_t = Z_t^t Z_t = \begin{bmatrix} M_{YY}(t) & M_{YW}(t) \\ M_{YW}(t) & M_{WW}(t) \end{bmatrix} ; \quad (3.7a)$$

$$\overline{M} = \frac{1}{N} \sum_{t=1}^{N} M_t = \begin{bmatrix} \overline{M}_{YY} & \overline{M}_{YW} \\ \overline{M}_{YW} & \overline{M}_{WW} \end{bmatrix} ; \quad (3.7b)$$

$$m_t = z_t^t z_t = \begin{bmatrix} m_{yy}(t) & m_{yw}(t) \\ m_{wy}(t) & m_{ww}(t) \end{bmatrix} ; \quad (3.7c)$$
and, 
\[
\bar{m} = \frac{1}{N} \sum_{t=1}^{N} m_t = \begin{bmatrix}
\bar{m}_{yy} & \bar{m}_{yw} \\
\bar{m}_{wy} & \bar{m}_{ww}
\end{bmatrix}
\] (3.7d)

The properties of our estimator of \( \beta \) rest on the following assumptions:

**Assumption 3.1**

The random variables \((e_t, u_t)\) are independent, satisfy Equations 3.2a, 3.2b, and have bounded \(8+\delta\) moments for \(t = 1, 2, \ldots\) with \(\delta > 0\).

This assumption allows us to define

\[
E(e_t) = 0 \quad ,
\] (3.8a)

\[
\Sigma_t = E(e_t^t e_t) = \begin{bmatrix}
\sigma^2_e(t) & \mathbb{H}_{ef}(t) \\
\mathbb{H}_{ef}(t) & \sigma^2_f(t)
\end{bmatrix} < \infty
\] (3.8b)

for \(t = 1, 2, \ldots\), and

\[
\Sigma = \frac{1}{N} \sum_{t=1}^{N} \Sigma_t = \begin{bmatrix}
\sigma^2_e & \mathbb{H}_{ef} \\
\mathbb{H}_{fe} & \sigma^2_f
\end{bmatrix}
\] . (3.9)

**Assumption 3.2**

The error covariance matrices
are known. Also, for any $\xi$ in an open sphere containing the true parameter, $\beta$,

$$0 < K < (1, -\xi') \Psi^{1/2} \Psi^{1/2}$$

for any $N > 3$, where $K$ is fixed.

**Assumption 3.3a**

$m_{ww}$ is a positive definite matrix for all $N > 3$.

**Assumption 3.3b**

$$\lim_{N \to \infty} m_{ww} = \overline{m}_{ww}$$

exists and is positive definite. Also,

$$N^{-1} \sum_{t=1}^{N} x_t^5, \quad N^{-1} \sum_{t=1}^{N} x_t^6, \quad \text{and} \quad N^{-1} \sum_{t=1}^{N} |x_t|^{6+\delta}$$

converge for $\delta > 0$.

**Assumption 3.4**

$$\lim_{N \to \infty} N^{-1} \sum_{t=1}^{N} \text{abs}(\Phi_t) = H < \infty$$

exists, where $\text{abs}(\Phi_t)$ denotes a $(4 \times 4)$ matrix whose elements are the
absolute values of the elements of $\mathbf{d}_t$.

**Assumption 3.5**

An estimator of $\mathbf{d}_t$, say $\hat{\mathbf{d}}_t$, is available for $t = 1, 2, \ldots, N$ such that $\hat{\mathbf{d}}_t$ and $\hat{\mathbf{d}}_s$ are independent for $t \neq s$, $E(\hat{\mathbf{d}}_t) = \mathbf{d}_t$, and

$$\hat{\mathbf{d}} = \frac{1}{N} \sum_{t=1}^{N} \hat{\mathbf{d}}_t = \mathbf{d} + O(N^{-1/2}).$$

It is worth noting that the above distributional assumptions are satisfied by the normal distribution. If the measurement errors $(e^t, u_t)$ are distributed as a bivariate normal for $t = 1, 2, \ldots$, then

$$\frac{\mathbf{d}_t}{\mathbf{d}_t} = \begin{bmatrix}
\sigma_e^2 & 0 & 2X_t \sigma_{eu(t)} \\
0 & 0 & 0 \\
\sigma_{u(t)}^2 & 2X_t \sigma_{u(t)}^2 & \text{sym.}
\end{bmatrix}
\begin{bmatrix}
\sigma_e^2 & 0 & 2X_t \sigma_{eu(t)} \\
0 & 0 & 0 \\
\sigma_{u(t)}^2 & 2X_t \sigma_{u(t)}^2 & \text{sym.}
\end{bmatrix}
$$

and

$$\frac{\hat{\mathbf{d}}_t}{\mathbf{d}_t} = \begin{bmatrix}
\sigma_e^2 & 0 & 2X_t \sigma_{eu(t)} \\
0 & 0 & 0 \\
\sigma_{u(t)}^2 & 2X_t \sigma_{u(t)}^2 & \text{sym.}
\end{bmatrix}
\begin{bmatrix}
\sigma_e^2 & 0 & 2X_t \sigma_{eu(t)} \\
0 & 0 & 0 \\
\sigma_{u(t)}^2 & 2X_t \sigma_{u(t)}^2 & \text{sym.}
\end{bmatrix}
$$

satisfies Assumption 3.5.
B. The Estimator

We define an estimator of $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ to be that $\beta$ which minimizes the function

$$h(\theta) = \frac{\theta' \hat{M} \theta}{\theta' \hat{E} \theta}, \quad (3.10)$$

where $\theta = \begin{bmatrix} 1 \\ -\beta \end{bmatrix}$. This type of estimator was used in a different setting by Booth (1973), and our developments are similar.

From Equation 3.10 we see that if $\hat{\theta}$ minimizes $h(\theta)$ with respect to $\theta$, then so will $c \hat{\theta}$, where $c$ is any nonzero constant. Consequently, we can uniquely determine only the ratios between the elements of $\hat{\theta}$. If we let

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \quad \text{and} \quad \hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \end{bmatrix},$$

then, for example, $\hat{\theta}_2 / \hat{\theta}_1$, $\hat{\theta}_3 / \hat{\theta}_1$, $\hat{\theta}_4 / \hat{\theta}_1$ are uniquely determined, and the vector
minimizes \( h(\theta) \), provided \( \hat{\theta}_1 = 0 \). But given Model 3.1a, 3.1b and Assumptions 3.1 and 3.4, \( \theta_1 = 0 \). Thus it is apparent that minimizing \( h(\theta) \) with respect to \( \theta \) is equivalent to minimizing \( h(\theta) \) with respect to \( \beta \). If \( \hat{\beta} \) minimizes \( h(\theta) \) with respect to \( \beta \), then this means

\[
\hat{\theta} = \begin{bmatrix}
1 \\
\hat{\theta}_2 / \hat{\theta}_1 \\
\hat{\theta}_3 / \hat{\theta}_1 \\
\hat{\theta}_4 / \hat{\theta}_1 
\end{bmatrix}
\]

with probability one.

Since \( h(\theta) \) is a homogeneous function of degree 0 we see that the value of \( \theta \) which minimizes Equation 3.10 also minimizes the LaGrangian function

\[
\mathcal{L} = \theta^\top \hat{M} \theta - \alpha (\theta^\top \hat{F}_\theta \theta - k)
\]

where \( \alpha \) is the LaGrangian multiplier and \( k \) is an arbitrary nonzero constant. Differentiating \( \mathcal{L} \) with respect to \( \theta \) and setting the derivative to zero yields the expression

\[
\frac{\partial \mathcal{L}}{\partial \theta} = 2\hat{M}\theta - 2\alpha \hat{F}_\theta \theta = 0
\]
By rearranging terms and using $\hat{\theta}$ to denote the minimizing value of $\theta$, we obtain

$$[\bar{M} - \alpha \hat{\delta}] \hat{\theta} = 0.$$ 

But this system of equations has a nontrivial solution for $\hat{\theta}$ only if

$$|\bar{M} - \alpha \hat{\delta}| = 0.$$ 

To minimize $\xi$ we must take $\alpha$ to be the smallest root of this determinental equation, say $\hat{\alpha}$.

By these arguments, we see that our estimator of $\beta$, say $\hat{\beta}$, satisfies

$$\hat{\beta} = \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}$$

and

$$(\bar{M} - \hat{\alpha} \hat{\delta}) \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix} = 0 \quad (3.11)$$

where $\hat{\alpha}$ is the smallest root of

$$|\bar{M} - \alpha \hat{\delta}| = 0 \quad (3.12)$$

Using the partitioned form of $\bar{M}$ and $\hat{\delta}$ and Equation 3.11, we see that $\hat{\beta}$ satisfies

$$\bar{M}_{WY} - \bar{M}_{WW} \hat{\beta} - \hat{\alpha}_{fe} \hat{\delta}_{fe} + \hat{\alpha}_{ff} \hat{\delta}_{ff} \hat{\beta} = 0.$$
Rearranging terms yields the expression

\[
[\overline{M}_{WW} - \hat{\alpha} \hat{\Sigma}_{fff}] \hat{\beta} = [\overline{M}_{WY} - \hat{\alpha} \hat{\Sigma}_{fe}] 
\]  \hspace{1cm} (3.13)

and finally

\[
\hat{\beta} = [\overline{M}_{WW} - \hat{\alpha} \hat{\Sigma}_{fff}]^{-1}[\overline{M}_{WY} - \hat{\alpha} \hat{\Sigma}_{fe}] . 
\]  \hspace{1cm} (3.14)

Let us consider the asymptotic properties of the estimator given by Equation 3.14.

C. Asymptotic Properties of the Estimator

It is convenient to introduce the following notation:

\[
\Delta \overline{M} = \overline{M} - E(\overline{M}) 
\]

\[
= \overline{M} - E\left[ \frac{1}{N} \sum_{t=1}^{N} z_t^t z_t + z_t^t \epsilon_t + \epsilon_t^t z_t + \epsilon_t^t \epsilon_t \right] 
\]

\[
= \overline{M} - \overline{m} - \overline{\delta} ; 
\]  \hspace{1cm} (3.15a)

\[
\Delta \theta = \begin{bmatrix} 0 \\ -\Delta \beta \end{bmatrix} = \begin{bmatrix} 1 \\ \hat{\beta} \end{bmatrix} - \begin{bmatrix} 1 \\ \beta \end{bmatrix} = \hat{\theta} - \theta ; 
\]  \hspace{1cm} (3.15b)

\[
\Delta \alpha = \hat{\alpha} - 1 ; 
\]  \hspace{1cm} (3.15c)

and

\[
\Delta \hat{\delta} = \frac{1}{N} \sum_{t=1}^{N} (\Delta \hat{\delta}_t) = \frac{1}{N} \sum_{t=1}^{N} (\hat{\delta}_t - \delta_t) 
\]

\[
= \hat{\delta} - \delta. 
\]  \hspace{1cm} (3.15d)
We now prove a series of lemmas which lead to the main theorem of this section.

Lemma 3.1

Given Model 3.1a, 3.1b with Assumptions 3.1 through 3.5,

\[ \Delta \alpha = o_p(1) \]

and

\[ \Delta \beta = o_p(1) \]

Proof:

Consider the definition of \( \hat{\alpha} \) given by Equation 3.12, i.e. \( \hat{\alpha} \) is the smallest root of

\[ |\bar{M} - \alpha \frac{\hat{\delta}}{\hat{\epsilon}}| = 0 \]

By definition

\[ \bar{M} = \frac{1}{N} \sum_{t=1}^{N} Z_t' Z_t = \frac{1}{N} \sum_{t=1}^{N} (z_t' \epsilon_t + \epsilon_t' z_t + \epsilon_t' \epsilon_t) \]

Thus, utilizing the assumptions of the lemma, we obtain

\[ \text{plim } \bar{M} = \lim_{N \to \infty} \left[ \frac{\bar{m}}{1} + \frac{\overline{\delta}}{1} \right] = \frac{\bar{m}}{1} + \frac{\overline{\delta}}{1} \]

Also, by Assumptions 3.4 and 3.5 we see that

\[ \text{plim } \hat{\delta} = \lim_{N \to \infty} \hat{\delta} = \overline{\delta} \]
Since $\hat{\alpha}$ is a continuous function of the elements of $\overline{M}$ and $\overline{F}$, the probability limit of $\hat{\alpha}$, say $\overline{\alpha}$, is the smallest root of the determinental equation

$$\left| \overline{m} + \overline{F} - \alpha \overline{F} \right| = 0 .$$

Consequently, if $\xi$ is an arbitrary nonzero $(4 \times 1)$ vector, then

$$\overline{\alpha} = \min_{\xi} \frac{\xi' \overline{m} \xi}{\xi' \overline{F} \xi}$$

$$= \min_{\xi} \frac{\xi' \overline{m} \xi}{\xi' \overline{F} \xi} + 1 .$$

But $\overline{m}$ and $\overline{F}$ are positive semi-definite by assumption;

$$\xi' \overline{m} \xi = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} (z_t \xi)'(z_t \xi) ;$$

and $z_t \theta = 0$ for $t = 1, 2, \ldots$. Hence,

$$\min_{\xi} \frac{\xi' \overline{m} \xi}{\xi' \overline{F} \xi} = 0 ,$$

$$\operatorname{plim} \hat{\alpha} = \overline{\alpha} = 1 ,$$

and the first portion of the lemma is established.
Consider the expression for \( \hat{\beta} \) given by Equation 3.14. Let us investigate the probability limits of the two factors, \( \bar{M}_{WW} - \hat{\alpha} \hat{\mathbb{F}}_{ff} \) and \( \bar{M}_{WY} - \hat{\alpha} \hat{\mathbb{F}}_{fe} \), which appear in that expression.

By definition,

\[
\bar{M}_{WW} - \hat{\alpha} \hat{\mathbb{F}}_{ff} = \frac{1}{N} \sum_{t=1}^{N} \left( w'_t w_t + w'_t f_t + f'_t w_t + f'_t f_t \right) - (1 + \Delta \alpha) \hat{\mathbb{F}}_{ff} - \hat{\alpha} (\Delta \mathbb{F}_{ff}) .
\]  

(3.16)

Since \( \Delta \mathbb{F}_{ff} = O_p(N^{-1/2}) \), \( \Delta \alpha = o_p(1) \), \( \mathbb{E}( f'_t ) = 0 \), and \( \mathbb{E}( f'_t f_t ) = \frac{1}{2} \mathbb{E}( f'_t ) \), it follows by our assumptions that

\[
\text{plim}_{N \to \infty} \left[ \bar{M}_{WW} - \hat{\alpha} \hat{\mathbb{F}}_{ff} \right] = \lim_{N \to \infty} \left[ \bar{m}_{ww} + \hat{\mathbb{F}}_{ff} - \mathbb{F}_{ff} \right] = \bar{m}_{ww}
\]

Similarly,

\[
\bar{M}_{WY} - \hat{\alpha} \hat{\mathbb{F}}_{fe} = \frac{1}{N} \sum_{t=1}^{N} \left( w'_t y_t + w'_t e_t + f'_t y_t + f'_t e_t \right) - (1 + \Delta \alpha) \hat{\mathbb{F}}_{fe} - \hat{\alpha} (\Delta \mathbb{F}_{fe})
\]

(3.17)

and by the assumptions of the lemma

\[
\text{plim}_{N \to \infty} \left[ \bar{M}_{WY} - \hat{\alpha} \hat{\mathbb{F}}_{fe} \right] = \lim_{N \to \infty} \left[ \bar{m}_{wy} + \hat{\mathbb{F}}_{fe} - \mathbb{F}_{fe} \right] = \bar{m}_{wy}
\]
If we apply these results to Equation 3.14 we obtain

\[ \text{plim } \hat{\beta} = \frac{m^{-1}}{m_{ww} m_{wy}} = \beta, \]

and the second portion of the lemma is established. Q.E.D.

By the definition of \( \hat{\theta} \) and \( \hat{\alpha} \), \( \hat{\alpha} \) is the smallest root of \( |M - \alpha S| \theta = 0 \)

and \( \hat{\theta} \) satisfies

\[ (M - \alpha S) \hat{\theta} = 0. \] \hspace{1cm} (3.18)

Introducing the notation given in Equations 3.15a, 3.15b, 3.15c, and 3.15d into Equation 3.18 yields the expression

\[ 0 = \left[ \Delta M + m + \frac{\theta}{S} - (1 + \Delta \alpha) \frac{\theta}{S} - (1 + \Delta \alpha)(\Delta S) \right] \hat{\theta} \]

\[ = \left[ \Delta M - (\Delta \alpha) \frac{\theta}{S} - (1 + \Delta \alpha)(\Delta S) \right] \hat{\theta}. \] \hspace{1cm} (3.19)

And premultiplying this by \( \theta' \) gives

\[ 0 = \theta' \left[ \Delta M - (\Delta \alpha) \frac{\theta}{S} - (1 + \Delta \alpha)(\Delta S) \right] \hat{\theta}, \] \hspace{1cm} (3.20)

since \( \theta' m = \frac{1}{N} \sum_{t=1}^{N} (z_t \theta)' z_t = 0. \) This expression puts us in position to prove our next lemma.

**Lemma 3.2**

Given Model 3.1a, 3.1b and Assumptions 3.1 through 3.5,
\[ \Delta\alpha = O_p\left(N^{-1/2}\right). \]

Proof:

By rearranging the terms in Equation 3.20 we obtain the following expression for \( \Delta\alpha \):

\[ (\Delta\alpha) = \frac{\theta'(\Delta M)\hat{\theta} - \theta'(\Delta \overline{F})\hat{\theta}}{\theta'(\overline{M}\hat{\theta}) + \theta'(\Delta \overline{F})\hat{\theta}}. \]  \hspace{1cm} (3.21)

By the assumptions of the lemma, \( \Delta M = O_p\left(N^{-1/2}\right) \) and \( \Delta \overline{F} = O_p\left(N^{-1/2}\right) \), and by Lemma 3.1, \( \Delta \theta = \begin{bmatrix} 1 \\ -\Delta\beta \end{bmatrix} = o_p(1) \). Thus we have

\[ \theta'(\Delta M)\hat{\theta} = \theta'(\Delta M)\theta + \theta'(\Delta M)(\Delta \theta) \]
\[ = \theta'(\Delta M)\theta + o_p\left(N^{-1/2}\right) \]
\[ = O_p\left(N^{-1/2}\right) \];

\[ \theta'(\Delta \overline{F})\hat{\theta} = \theta'(\Delta \overline{F})\theta + \theta'(\Delta \overline{F})(\Delta \theta) \]
\[ = \theta'(\Delta \overline{F})\theta + o_p\left(N^{-1/2}\right) \]
\[ = O_p\left(N^{-1/2}\right) \];

and

\[ \theta'(\overline{F})\hat{\theta} = \theta'(\overline{F})\theta + \theta'(\overline{F})(\Delta \theta) \]
\[ = \theta'(\overline{F})\theta + o_p(1) \].
By combining these results into Equation 3.21 we obtain
\[
(\Delta \alpha) = \frac{91(\Delta M)q - 91(\Delta \hat{f})q + o(N^{-1/2})}{p} + o(1) + o
\]
\[
= O_p(N^{-1/2}). \quad Q.E.D.
\]

With this lemma in hand, we are able to establish the order of the error in \( \hat{\beta} \).

**Lemma 3.3**

Given Model 3.1a, 3.1b and Assumptions 3.1 through 3.5,
\[
\Delta \beta = O_p(N^{-1/2}).
\]

**Proof:**

Since \( W_t = w_t + f_t \) we can write
\[
\bar{M}_{WW} = \frac{1}{N} \sum_{t=1}^{N} W_t W_t = \bar{m}_{ww} + \frac{1}{N} \sum_{t=1}^{N} (w_t f_t + f_t w_t + f_t f_t).
\]

Thus we can write the inverse of Equation 3.16 as
\[
[\bar{M}_{WW} - \alpha \hat{f}_f]^{-1} = \left[ \bar{m}_{ww} + \frac{1}{N} \sum_{t=1}^{N} (w_t f_t + f_t w_t) \right]^{-1}.
\]
\[
\begin{align*}
&+ \frac{1}{N} \sum_{t=1}^{N} (f \cdot f - \Phi_{gg}(t)) \\
&- (\Delta \alpha) \Phi_{ff} - (1 + \Delta \alpha)(\Delta \Phi_{gg})^{-1} \\
&= \left[ \overline{m}_{ww} + \hat{a} \right]^{-1},
\end{align*}
\]

where
\[
\hat{a} = \frac{1}{N} \sum_{t=1}^{N} (w^t f + f^t w) + \frac{1}{N} \sum_{t=1}^{N} (f^t f - \Phi_{ff}(t)) - (\Delta \alpha) \Phi_{ff} - (1 + \Delta \alpha)(\Delta \Phi_{ff})
\]

By the assumptions of the lemma, together with Lemma 3.2, each term of \( \hat{a} \) is \( O_p (N^{-1/2}) \). Thus \( \hat{a} = O_p (N^{-1/2}) \).

Similarly, if we define \( \hat{b} \) to be
\[
\hat{b} = \frac{1}{N} \sum_{t=1}^{N} (w^t e + f^t y) + \frac{1}{N} \sum_{t=1}^{N} (f^t e - \Phi_{fe}(t)) \\
- (\Delta \alpha) \Phi_{fe} - (1 + \Delta \alpha)(\Delta \Phi_{fe})
\]

then Equation 3.17 becomes
\[
\left[ \overline{m}_{wy} - \alpha \Phi_{fe} \right] = \left[ \overline{m}_{wy} + \hat{b} \right].
\]
Note that \( \hat{b} = O_p(N^{-1/2}) \) follows from our assumptions and from Lemma 3.2.

Utilizing a Taylor series expansion (cf. Theorem 2.11), we may now express Equation 3.14 as follows:

\[
\hat{\beta} = [\overline{M}_{WW} - \hat{\alpha} \hat{\delta}_{ff}]^{-1}[\overline{M}_{WY} - \hat{\alpha} \hat{\delta}_{fe}]
\]

\[
= [\overline{m}_{ww} + \hat{a}]^{-1}[\overline{m}_{wy} + \hat{b}]
\]

\[
= [\overline{m}_{ww}^{-1} - \overline{m}_{ww}^{-1} \hat{a} \overline{m}_{ww}^{-1}][\overline{m}_{wy} + \hat{b}] + O_p(N^{-1})
\]

\[
= \overline{m}_{ww}^{-1} \overline{m}_{wy} + \overline{m}_{wy} \hat{b} - \overline{m}_{ww}^{-1} \hat{a} \overline{m}_{ww}^{-1} \overline{m}_{wy} + O_p(N^{-1})
\]

\[
= \beta + \overline{m}_{ww}^{-1} \hat{b} - \overline{m}_{ww}^{-1} \hat{a} \beta + O_p(N^{-1})
\]

Thus

\[
\Delta \beta = \hat{\beta} - \beta = \overline{m}_{ww}^{-1} [\hat{b} - \hat{a} \beta] + O_p(N^{-1}) . \tag{3.24}
\]

And since both \( \hat{a} \) and \( \hat{b} \) are \( O_p(N^{-1/2}) \), we have our conclusion, namely

\[
\Delta \beta = O_p(N^{-1/2}) . \quad \text{Q.E.D.}
\]

Before presenting the main theorem of this section, it will be convenient to define the residual error, say \( v_t \), as
\( v_t = e_t - f_t \beta \)
\[ = Y_t - W_t \beta \]
\[ = Z_t \theta \]
\[ = \epsilon_t \theta \]

for \( t = 1, 2, \ldots \). By the moment properties of \( e_t \) and \( u_t \) we have at once that

\[ E(v_t) = 0 \]

and

\[ \operatorname{Var}(v_t) = E(v_t^2) = \theta^t E(\epsilon_t^t \epsilon_t) \theta \]
\[ = \theta^t \frac{1}{v_t} \theta < \infty \]  

for \( t = 1, 2, \ldots \). Define

\[ \sigma^2_v = \theta^t \frac{1}{v_t} \theta \]
\[ \tilde{\sigma}^2_v = \theta^t \frac{\hat{\theta}}{v_t} \theta \]
\[ \frac{\sum}{\sigma}^2_v = \theta^t \frac{\hat{\theta}}{v_t} \theta = \frac{1}{N} \sum_{t=1}^{N} \sigma^2_v \]  

and

Also, define

\[ E(f_t^t v_t) = \frac{1}{v_t} \]
for $t = 1, 2, \ldots$ and

$$E\left(\frac{1}{N} \sum_{t=1}^{N} f'_t v_t\right) = \frac{1}{N} \sum_{t=1}^{N} \hat{f}'_{fv}(t) = \hat{f}'_{fv} . \quad (3.26)$$

By the definition of $v_t$, we have

$$\hat{f}'_{fv}(t) = \hat{f}'_{fe}(t) - \hat{f}'_{ff}(t) \beta$$

and

$$\hat{f}'_{fv} = \hat{f}'_{fe} - \hat{f}'_{ff} \beta . \quad (3.27)$$

In an obvious notation let

$$\hat{f}'_{fv}(t) = \hat{f}'_{fe}(t) - \hat{f}'_{ff}(t) \beta ,$$

$$\hat{f}'_{fv} = \frac{1}{N} \sum_{t=1}^{N} \hat{f}'_{fv}(t) = \hat{f}'_{fe} - \hat{f}'_{ff} \beta , \quad (3.28)$$

and

$$\Delta \hat{f}'_{fv} = \hat{f}'_{fv} - \hat{f}'_{fv} = \frac{1}{N} \sum_{t=1}^{N} \Delta \hat{f}'_{fv}(t)$$

$$= \Delta \hat{f}'_{fe} - \Delta \hat{f}'_{ff} \beta .$$

With this new notation in hand we may write $(\hat{b} - \hat{a} \beta)$ as follows:

$$(\hat{b} - \hat{a} \beta) = \frac{1}{N} \sum_{t=1}^{N} \left[ w'_t (e_t - f_t \beta) + f'_t (y_t - w_t \beta) \right]$$

$$+ \frac{1}{N} \sum_{t=1}^{N} \left[ f'_t (e_t - f_t \beta) - (\hat{f}'_{fe}(t) - \hat{f}'_{ff}(t) \beta) \right]$$
\[- (\Delta \alpha) \left[ \hat{\beta}_{fe} - \hat{\beta}_{ff} \beta \right] \]
\[- (1 + \Delta \alpha) \left[ \Delta \hat{\beta}_{fe} - \Delta \hat{\beta}_{ff} \beta \right] \]
\[= \frac{1}{N} \sum_{t=1}^{N} w_t v_t + 0 \]
\[+ \left( \frac{1}{N} \sum_{t=1}^{N} \left[ f^t v_t - \hat{\beta}_{fv(t)} \right] \right) \]
\[- (\Delta \alpha) \hat{\beta}_{fv} \]
\[ - (1 + \Delta \alpha) (\Delta \hat{\beta}_{fv}) \quad (3.29) \]

since \( y_t - w_t \beta = 0 \) for \( t = 1, 2, \ldots \). This expression will be useful in proving Theorem 3.1.

**Theorem 3.1**

Given Model 3.1a, 3.1b and Assumptions 3.1 through 3.5,

\[ \sqrt{N} (\hat{\beta} - \beta) \xrightarrow{d} N[0, \frac{m^{-1}}{\mu_{ww}} \lim_{N \to \infty} E\left( \frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t' \right) m^{-1}] \]

where

\[ \phi_t = W_t v_t - \left( \frac{v_t - \bar{v}_t}{\bar{v}} \right) \hat{\beta}_{fv} - \hat{\beta}_{fv(t)} \]

for \( t = 1, 2, \ldots \).
Proof:

From Equations 3.24 and 3.29 we have

\[ (\hat{\beta} - \beta) = \frac{-1}{m w^2 w} \left[ \frac{1}{N} \sum_{t=1}^{N} w^t v^t + \frac{1}{N} \sum_{t=1}^{N} (f^t v^t - \Phi_{fv}(t)) \right] \]

\[ - (\Delta \alpha) \Phi_{fv} - (1 + \Delta \alpha)(\Delta \Phi_{fv}) + O_p(N^{-1}). \]

But by assumption, \( \Delta \Phi_{fv} = O_p(N^{-1/2}) \) and by Lemma 3.2

\( \Delta \alpha = O_p(N^{-1/2}). \) Thus

\[ (\hat{\beta} - \beta) = \frac{-1}{m w^2 w} \left[ \frac{1}{N} \sum_{t=1}^{N} w^t v^t + \frac{1}{N} \sum_{t=1}^{N} (f^t v^t - \Phi_{fv}(t)) \right] \]

\[ - (\Delta \alpha) \Phi_{fv} - (\Delta \Phi_{fv}) + O_p(N^{-1}) . \]

Consider \( (\Delta \alpha) \). From Lemma 3.3, Theorem 2.11, Equation 3.21, and from the relations \( \Delta M = O_p(N^{-1/2}) \), \( \Delta \Phi = O_p(N^{-1/2}) \) we obtain

\[ (\Delta \alpha) = \frac{\theta^t(\Delta M) \hat{\alpha} - \theta^t(\Delta \Phi) \hat{\theta}}{\theta^t \Phi \hat{\alpha} + \theta^t(\Delta \Phi) \hat{\theta}} \]

\[ = \frac{\theta^t(\Delta M) \hat{\alpha} - \theta^t(\Delta \Phi) \hat{\theta} + O_p(N^{-1})}{\theta^t \Phi \hat{\alpha} + O_p(N^{-1/2})} \]

\[ = \frac{\theta^t(\Delta M) \hat{\alpha} - \theta^t(\Delta \Phi) \hat{\theta}}{\theta^t \Phi \hat{\alpha}} + O_p(N^{-1}) . \]
And by definition of $\Delta M$ and $\Delta \Phi$

$$\theta'(\Delta M) \theta = \theta'(M \theta - m \theta - \Phi \theta)$$

and

$$\theta'(\Delta \Phi) \theta = \theta'(\Phi \theta - \Phi \theta)$$

Therefore, since $m \theta = 0$,

$$(\Delta \alpha) = \frac{\theta'(M \theta - \Phi \theta)}{\sigma_{\Phi \theta}} + O(N^{-1})$$

(3.31)

$$= \frac{1}{N} \sum_{t=1}^{N} \left( \frac{v_{t}^2}{\sigma_{v}} \right) - \frac{1}{N} \sum_{t=1}^{N} \sigma_{v}^2 + O(N^{-1}).$$

Substituting Equation 3.31 into Equation 3.30 yields

$$\hat{\beta} - \beta = \overline{m}^{-1}_{ww} \left\{ \frac{1}{N} \sum_{t=1}^{N} (w_{t} + f_{t}) v_{t} - \frac{\Phi_{fv}}{\sigma_{fv}} \right\}$$

$$- \left[ \frac{1}{N} \sum_{t=1}^{N} \frac{v_{t}^2}{\sigma_{v}} - \frac{1}{N} \sum_{t=1}^{N} \sigma_{v}^2 \right] \frac{\Phi_{fv} - (\Phi_{fv} - \Phi_{fv})}{\sigma_{fv}}$$

$$+ O \left( \frac{1}{N} \right)$$
\[\begin{align*}
\frac{m^{-1}}{\text{ww}} \left\{ \frac{1}{N} \sum_{t=1}^{N} \left[ \frac{v_{t}^{2} - \hat{v}_{t}^{2}}{\sigma_{v}^{2}} \right] \right\} \\
+ \mathcal{O}_{p}(N^{-1}) \\
= \frac{m^{-1}}{\text{ww}} \frac{1}{N} \sum_{t=1}^{N} \phi_{t} + \mathcal{O}_{p}(N^{-1}), \tag{3.32}
\end{align*}\]

where
\[\phi_{t} = \frac{W_{t}' \nu_{t}}{\left( \frac{v_{t}^{2} - \hat{v}_{t}^{2}}{\sigma_{v}^{2}} \right)} \hat{f}_{v} - \hat{f}_{v(t)} \cdot \]

From Equation 3.32 and Theorem 2.6, it follows that the limiting distribution of \(\sqrt{N}(\hat{\beta} - \beta)\) is the same as the limiting distribution of

\[m^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \phi_{t}. \]

Thus let us investigate the limiting distribution of

\[m^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \phi_{t}. \]

Let \(\lambda\) be an arbitrary nonzero \((3 \times 1)\) vector, and consider

\[\frac{\lambda'}{\sqrt{N}} \sum_{t=1}^{N} \phi_{t} = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \lambda' \phi_{t}. \]

First, note that the random variables \(\phi_{t}\) are independent by

Assumptions 3.1 and 3.5. Second, note that
\begin{align*}
E(\lambda' \phi_t) &= \lambda' E(\phi_t) = \lambda' E \left[ W_t v_t - \left( \frac{v_t^2 - \theta_t^* \theta_t^*}{\sigma_v^2} \right) \bar{v}_t - \bar{v}_t(t) \right] \\
&= \lambda' E \left[ w_t^* v_t + f_t v_t - \left( \frac{v_t^2 - \theta_t^* \theta_t^*}{\sigma_v^2} \right) \bar{v}_t - \bar{v}_t(t) \right] \\
&= \lambda' \left[ 0 + \bar{v}_t(t) - \left( \frac{\sigma_v^2}{\sigma_v^2} \right) \bar{v}_t - \bar{v}_t(t) \right] \\
&= 0
\end{align*}

for \( t = 1, 2, \ldots \), again by Assumptions 3.1 and 3.5. Third, note that

\[ E[(\lambda' \phi_t)^2] = E[\lambda' \phi_t \phi_t^*] = \lambda' E(\phi_t \phi_t^*), \]

\( \lambda < \infty \)

and

\[ E\left[ |\lambda' \phi_t|^{2+\delta}\right] < \infty, \]

where \( \delta > 0 \). This follows since the random variables \((e_t, u_t)\) possesses finite \(8+\delta\) moments for \( t = 1, 2, \ldots \).

Now we have

\[
\lim_{N \to \infty} \frac{\sum_{t=1}^{N} E[|\lambda' \phi_t|^{2+\delta}]}{\left( \sum_{t=1}^{N} E[(\lambda' \phi_t)^2] \right)^{2+\delta/2}} = \]
by Assumptions 3.3b and 3.4. By the Liapounov central limit theorem (cf. Theorem 2.8) this gives

\[
\lim_{N \to \infty} \frac{N^{-1} \sum_{t=1}^{N} \mathbb{E} \left[ |\lambda_t \phi_t|^{2+\delta} \right]}{N^{\delta/2} \left( \sum_{t=1}^{N} \mathbb{E} (\phi_t \phi_t') \lambda \right)^{(2+\delta)/2}} = 0
\]

Furthermore, by our assumptions

\[
\mathbb{E} \left[ \frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t' \right] = O(1)
\]

and thus by Theorem 2.7

\[
\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \lambda_t \phi_t \xrightarrow{d} N(0, \lambda_t \lim_{N \to \infty} \mathbb{E} \left[ \frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t' \right] \lambda).
\]

By the multivariate central limit theorem we now have
\[
\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \phi_t \longrightarrow N(0, \lim_{N \to \infty} \text{E}[\frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t^t]).
\]

Finally, since \( \lim_{N \to \infty} \bar{m}_{ww} = \bar{m}_{ww} \), we obtain

\[
\bar{m}_{ww}^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \phi_t \overset{d}{\longrightarrow} N(0, \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \text{E}[\phi_t \phi_t^t] \bar{m}_{ww}^{-1}).
\]

and the conclusion of the theorem follows. \( \text{Q.E.D.} \)

D. An Estimator of the Covariance Matrix of the Estimator

In this section we shall be concerned with estimating the covariance matrix of the preliminary estimator \( \hat{\beta} \). For this purpose, we shall rely on the asymptotic properties of \( \hat{\beta} \) developed in the last section.

To be specific, we shall estimate the matrix

\[
\bar{m}_{ww}^{-1} \text{E} [\frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t^t] \bar{m}_{ww}^{-1}.
\]  
\[ (3.33) \]

With the estimation of Equation 3.33 in mind, recall that

\[
\phi_t = W_t v_t \begin{bmatrix} v^2_t - \sigma^2 v_t^t \\ v^2_t - \sigma^2 v_t \end{bmatrix} \hat{f}_{v} - \hat{f}_{v(t)}.
\]
where $v_t = Y_t - W_t \beta$, 

$$\frac{\sigma^2}{\sigma_{v_t}} = \hat{\theta} \frac{\sigma^2}{\sigma_{v_t}} \hat{\theta},$$

$$\frac{\sigma^2}{\sigma_v} = \theta \frac{\sigma^2}{\sigma_v} \theta,$$

$$\frac{\sigma_{fv}}{\sigma_{fe}} \frac{\sigma_{ff}}{\sigma_{fe}} \beta,$$

and $\frac{\sigma_{fv(t)}}{\sigma_{fe(t)}} \frac{\sigma_{ff(t)}}{\sigma_{fe(t)}} \beta$.

From Lemma 3.3, $\hat{\beta} = \beta + O_p(N^{-1/2})$, and by Assumption 3.5, 

$$\frac{\sigma_{fv}}{\sigma_{fe}} + O_p(N^{-1/2}).$$

Thus we have

$$\hat{v}_t = Y_t - W_t \hat{\beta} = v_t + O_p(N^{-1/2}),$$

$$\frac{\sigma^2}{\sigma_{v_t}} = \hat{\theta} \frac{\sigma^2}{\sigma_{v_t}} \hat{\theta} = O_p(N^{-1/2}),$$

$$\frac{\sigma^2}{\sigma_v} = \theta \frac{\sigma^2}{\sigma_v} \theta = O_p(N^{-1/2}),$$

$$\frac{\sigma_{fv}}{\sigma_{fe}} \frac{\sigma_{ff}}{\sigma_{fe}} = O_p(N^{-1/2}),$$

(3.34)

$$\frac{\sigma_{fv(t)}}{\sigma_{fe(t)}} \frac{\sigma_{ff(t)}}{\sigma_{fe(t)}} = O_p(N^{-1/2});$$

and $\frac{\sigma_{fv(t)}}{\sigma_{fe(t)}} \frac{\sigma_{ff(t)}}{\sigma_{fe(t)}} \beta = O_p(N^{-1/2});$

Now define $\hat{\phi}_t$ as follows:

$$\hat{\phi}_t = W_t \hat{v}_t = \left[ \begin{array}{c} \frac{\sigma^2}{\sigma_{v_t}} \hat{\theta} \\ \frac{\sigma^2}{\sigma_v} \theta \end{array} \right] \hat{\sigma}_{fv(t)} \frac{\sigma_{ff(t)}}{\sigma_{fe(t)}} \beta.$$
By Equation 3.34 it is clear that

$$\hat{\phi}_t = \phi_t + O_p(N^{-1/2})$$

Thus

$$\frac{1}{N} \sum_{t=1}^{N} \hat{\phi}_t^i = \frac{1}{N} \sum_{t=1}^{N} \phi_t^i + O_p(N^{-1/2})$$

If in addition to Assumptions 3.1 through 3.5, we assume the random variables \((e_t, u_t)\) have bounded 16th moments, and the constants \(x_t\) are uniformly bounded, then clearly

$$\text{Var}\{N^{-1} \sum_{t=1}^{N} \phi_t^i \phi_t^j\} = N^{-2} \sum_{t=1}^{N} \text{Var}\{\phi_t^i \phi_t^j\}$$

$$= O(N^{-1})$$

Under these circumstances, this gives

$$N^{-1} \sum_{t=1}^{N} \phi_t^i \phi_t^j = \mathbb{E}\{N^{-1} \sum_{t=1}^{N} \phi_t^i \phi_t^j\} + O_p(N^{-1/2})$$

and from Equation 3.35

$$N^{-1} \sum_{t=1}^{N} \hat{\phi}_t^i \hat{\phi}_t^j = \mathbb{E}\{N^{-1} \sum_{t=1}^{N} \phi_t^i \phi_t^j\} + O_p(N^{-1/2})$$

To estimate \(\frac{1}{N} \sum_{t=1}^{N} \phi_t^i \phi_t^j\) let us recall Equation 3.22. There we saw
\[
[\overline{M}_{WW} - \hat{\alpha} \hat{F}_{ff}]^{-1} = [\overline{m}_{ww} + \hat{a}]^{-1}
\]

where \( \hat{a} = O_p(N^{-1/2}) \). Therefore,

\[
[\overline{M}_{WW} - \hat{\alpha} \hat{F}_{ff}]^{-1} = \overline{m}_{ww}^{-1} + O_p(N^{-1/2}).
\]

Taken together, the above results in the following theorem:

**Theorem 3.2**

Given Model 3.1a, 3.1b, Assumptions 3.1 through 3.5, and the additional assumptions that the random variables \((e_t, u_t)\) have bounded sixteenth moments and the constants \(x_t\) are uniformly bounded\(^1\), then

\[
\hat{H}^{-1} \hat{D} \hat{H}^{-1} = \overline{m}_{ww}^{-1} E\left[ \frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t' \right] \overline{m}_{ww}^{-1} + O_p(N^{-1/2})
\]

where

\[
\hat{H} = (\overline{M}_{WW} - \hat{\alpha} \hat{F}_{ff})
\]

and

\[
\hat{D} = \frac{1}{N} \sum_{t=1}^{N} \hat{\phi}_t \hat{\phi}_t'.
\]

\(^1\) These assumptions are probably stronger than needed to insure the error in the estimator is \(O_p(N^{-1/2})\), and are certainly stronger than needed to establish the consistency of the estimator.
Proof:

This result follows immediately from Equations 3.37 and 3.38.

Q.E.D.

As a consequence of this theorem, we will use $\hat{H}^{-1} \hat{D} \hat{H}^{-1}$ to estimate the covariance matrix of $\hat{\beta}$. 
IV. ESTIMATORS FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP WHEN THE ERROR VARIANCES ARE DECREASING

In this chapter we consider estimation for the quadratic functional relationship under the additional, more restrictive assumption that the error variances decrease with increasing sample size. First, we restate the model and state new assumptions, employing a slightly different notation. Second, under the new assumptions we establish the asymptotic properties of the ordinary least squares estimator and of the preliminary estimator $\hat{\beta}$ as defined in Chapter 3. Third, the maximum likelihood equations are presented. Finally, the asymptotic properties of two iterative type pseudo-maximum likelihood estimators are considered.

A. The Model and Assumptions

In this chapter we let \( \{b_N\}_{N=1}^{\infty} \) and \( \{a_N\}_{N=1}^{\infty} \) be sequences of positive real numbers such that \( N = \frac{b_N}{a_N} \) for \( N = 1, 2, \ldots \); we suppose the existence of a sequence of experiments indexed by \( N \); and we let \( b_N \) denote the number of observations in the \( N \)th experiment.

The quadratic functional relationship for the \( N \)th experiment is specified by the exact mathematical relationship

\[
\begin{align*}
 y_{Nt} = \beta_0 + \beta_1 x_{Nt} + \beta_2 x_{Nt}^2
\end{align*}
\]
where we observe

\[ Y_{Nt} = y_{Nt} + e_{Nt}, \]

\[ X_{Nt} = x_{Nt} + u_{Nt}, \]

(4.1b)

for \( t = 1, \ldots, b_N \), and where \( \{x_{Nt}\}_{t=1}^{b_N} \) is a sequence of fixed constants, i.e. constant random variables.

To shorten this notation, henceforth we will suppress the subscript \( N \) from all variables. Thus the model becomes

\[ y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2, \]

where we observe

\[ Y_t = y_t + e_t \]

and

\[ X_t = x_t + u_t \]

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

Throughout the work of this chapter we will have occasion to call upon one or more of the following assumptions:

Assumption 4.1a

The errors of measurement \((e_t, u_t)\) are independent, have zero means, and bounded \(2+\delta\) moments, \(\delta > 0\), for \( t = 1, \ldots, b_N \).
Assumption 4.1b

The random variables \((e_t, u_t)\) are independent, have zero means, and bounded \(4+\delta\) moments, \(\delta > 0\), for \(t = 1, \ldots, b_N\).

Assumption 4.2

The error covariance matrix

\[
E\left(\begin{pmatrix} e_t \\ u_t \end{pmatrix} \right) \left(\begin{pmatrix} e_t \\ u_t \end{pmatrix} \right) = \Phi = \begin{pmatrix} \sigma_e^2 & \sigma_{eu} \\ \sigma_{ue} & \sigma_u^2 \end{pmatrix}
\]

is known and positive definite.

Assumption 4.3

The matrix \(\overline{m}_{ww}\) defined by

\[
\overline{m}_{ww} = b_N^{-1} \sum_{t=1}^{b_N} w_t^t w_t^t
\]

is positive definite for all \(b_N > 3\), where \(w_t = (1, x_t, x_t^2)\).

Assumption 4.4

\[
\lim_{N \to \infty} \overline{m}_{ww} = \overline{m}_{ww}
\]

exists and is positive definite. Furthermore,

\[
b_N^{-1} \sum_{t=1}^{b_N} |x_t|^5, \ b_N^{-1} \sum_{t=1}^{b_N} |x_t|^6, \ b_N^{-1} \sum_{t=1}^{b_N} |x_t|^{6+\delta}
\]
converge for some $\delta > 0$.

**Assumption 4.5**

The sequence $\{b_N\}_{N=1}^\infty$ is monotonically increasing.

**Assumption 4.6**

The elements of the error covariance matrix satisfy $\Phi = O(a_N)$.

**Assumption 4.7a**

The elements of the sequence $\{a_N\}_{N=1}^\infty$ satisfy $a_N = o(N^{-1/2})$.

**Assumption 4.7b**

The elements of the sequence $\{a_N\}_{N=1}^\infty$ satisfy $a_N^{3/2} = o(N^{-1/2})$.

**Assumption 4.8**

The matrix $\overline{m}^*$ given by

$$\overline{m}^* = b^{-1} \frac{b_N}{N} \sum_{t=1}^N (a_N^{-1} \frac{2}{\sigma_{v_t}^2})^{-1} \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} [1 \ 1 \ x_t^2]$$

is positive definite for all $b_N > 3$, where $\sigma_{v_t}^2 = \sigma_e^2 - 2\gamma_t \sigma_{e_u} + \gamma^2_t \sigma_u^2$ and $\gamma_t = \beta_1 + 2x_t \beta_2$.

**Assumption 4.9**

$$\lim_{N \to \infty} \overline{m}^*_{xx} = \overline{m}^*_{xx}$$
exists and is positive definite. Also,

\[ b^{-1} \sum_{t=1}^{bN} (1 + |x_t|^r)^{-1} \]

converges for \( r = 2, 3, \ldots, 10 \).

**Assumption 4.10**

A preliminary estimator of \( \beta \), say \( \bar{\beta} \), exists and satisfies

\[ \bar{\beta} - \beta = O_p(N^{-1/2}) . \]

**Assumption 4.11**

An estimator of

\[ E \begin{pmatrix} e_t \\ f_t \end{pmatrix} = \mathbf{f}_t , \]

say \( \mathbf{f}_t \), is available for \( t = 1, \ldots, b_N \) such that \( \mathbf{f}_t \) and \( \mathbf{f}_s \) are independent for \( t \neq s \), \( E(\mathbf{f}_t) = \mathbf{f}_t \), and

\[ \mathbf{f} = b^{-1} \sum_{t=1}^{bN} \mathbf{f}_t = b^{-1} \sum_{t=1}^{bN} \mathbf{f}_t + O_p(a_N N^{-1/2}) . \]

Here, \( f_t = [(0, u_t, 2x_t u_t + (u_t^2 - \sigma_u^2)] \) is defined as in Chapter 3.
B. The Ordinary Least Squares Estimator

The ordinary least squares estimator for Model 4.1a, 4.1b, denoted by $\hat{\beta}_{OLS}$, is given by

$$
\hat{\beta}_{OLS} = \left( \frac{1}{b^{1}} \sum_{N}^{bN} \left[ \begin{array}{c} 1 \\ X_t \\ X_t^2 \end{array} \right] \right)^{-1} \left( \frac{1}{b^{1}} \sum_{N}^{bN} \left[ \begin{array}{c} 1 \\ X_t \\ X_t^2 \end{array} \right] \right) Y_t
$$

(4.2)

Under the error assumptions of Section A it can be shown that $\hat{\beta}_{OLS}$ is a consistent estimator of $\beta$. The next two lemmas summarize this fact.

Lemma 4.1

Under Assumptions 4.1a, 4.3, 4.4, 4.6, and 4.7a,

$$
\hat{\beta}_{OLS} - \beta = O_p(N^{-1/2}).
$$

Proof:

Define

$$
v_t^* = e_t - \gamma_t u_t
$$

(4.3)

where $\gamma_t = \beta_1 + 2x_t \beta_2$. Note that

$$
Y_t - \beta_0 - \beta_1 X_t - \beta_2 X_t^2 = v_t^* - \beta_2 u_t^2.
$$

(4.4)
Substituting Equation 4.4 into Equation 4.2 yields

\[
\hat{\beta}_{OLS} = \beta + \left( b^{-1} \sum_{t=1}^{b_N} X_t \right) \begin{bmatrix} 1 \\ X_t \end{bmatrix} \begin{bmatrix} 1 \\ X_t^{2} \\ X_t \end{bmatrix}^{-1} \left( b^{-1} \sum_{t=1}^{b_N} X_t \right)^{-1} \begin{bmatrix} 1 \\ X_t \end{bmatrix} \begin{bmatrix} v_t^* - \beta u_t^2 \end{bmatrix}.
\]

(4.5)

Consider

\[
\left\{ b^{-1} \sum_{t=1}^{b_N} \begin{bmatrix} 1 \\ X_t \end{bmatrix} \begin{bmatrix} 1 \\ X_t^{2} \\ X_t \end{bmatrix}^{-1} \right\}.
\]

By Assumptions 4.1a, 4.6, 4.7a, and 4.4 we have

\[
E(u_t) = 0 ;
\]

\[
\text{Var}\left\{ b^{-1} \sum_{t=1}^{b_N} u_t \right\} = b^{-1} \sigma_u^2 = O(b^{-1} a_N) = O(N^{-1}) ;
\]

\[
\text{Var}\left\{ b^{-1} \sum_{t=1}^{b_N} x_t u_t \right\} = b^{-2} \sum_{t=1}^{b_N} x_t \sigma_u^2 = O(b^{-1} a_N) = O(N^{-1})
\]

for \( r = 1, 2, 3 \);

\[
u_t^2 = O_p\left( a_N \right) = o_p\left( N^{-1/2} \right) ;
\]

\[
u_t^3 = O_p\left( a_N^{1.5} \right) = o_p\left( N^{-1/2} \right) ;
\]
and
\[ u_t^4 = O_p\left( a_{N_0}^2 \right) = O_p\left( N^{-1/2} \right). \]

Thus
\[
\begin{align*}
\frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} X_t &= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t + \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} u_t \\
&= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t + O_p\left( N^{-1/2} \right);
\end{align*}
\]
\[
\begin{align*}
\frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^2 &= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} (x_t^2 + 2x_t u_t) + o_p\left( N^{-1/2} \right) \\
&= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^2 + O_p\left( N^{-1/2} \right);
\end{align*}
\]
\[
\begin{align*}
\frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^3 &= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} (x_t^3 + 3x_t^2 u_t) + o_p\left( N^{-1/2} \right) \\
&= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^3 + O_p\left( N^{-1/2} \right);
\end{align*}
\]

and
\[
\frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^4 = \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} (x_t^4 + 4x_t^3 u_t) + o_p\left( N^{-1/2} \right) \\
= \frac{b_N}{b_{N^*}} \sum_{t=1}^{b_N} x_t^4 + O_p\left( N^{-1/2} \right). 
\]

Consequently,
and by Assumption 4.3 and Theorem 2.11

\[
\left\{ b^{-1} \sum_{t=1}^{bN} \begin{pmatrix} 1 \\ X_t \\ X_t \end{pmatrix} \begin{bmatrix} 1 & X_t & X_t^2 \end{bmatrix} \right\}^{-1} = \left( \frac{m_{ww}}{N} \right)^{-1} + O_p \left( N^{-1/2} \right). \tag{4.6}
\]

Next consider

\[
\left\{ b^{-1} \sum_{t=1}^{bN} \begin{pmatrix} 1 \\ X_t \\ X_t \end{pmatrix} \begin{bmatrix} v_t^* - \beta_2 u_t^2 \end{bmatrix} \right\}.
\]

By Assumptions 4.1a, 4.4, 4.6, and 4.7a we obtain

\[
E(v_t^*) = 0 ;
\]

\[
u_t^2 = O_p \left( a_N \right) = o_p \left( N^{-1/2} \right) ;
\]

\[
u_t v_t^* = O_p \left( a_N \right) = o_p \left( N^{-1/2} \right) ;
\]

\[
u_t^2 v_t = O_p \left( a_N 1.5 \right) = o_p \left( N^{-1/2} \right) ;
\]

\[
\text{Var} \left\{ b^{-1} \sum_{t=1}^{bN} x_t^* \right\} = b^{-2} \sum_{t=1}^{bN} x_t^2 \sigma^2 v_t = O(N^{-1})
\]
for \( r = 0, 1, 2 \), where \( \sigma^2_{v_t} = \sigma^2_e - 2\gamma_t \sigma_{eu} + \gamma^2_t \sigma^2_u \). This gives

\[
b^{-1} \sum_{N t=1}^{b_N} \begin{pmatrix} 1 \\ x_t^2 \\ X_t \end{pmatrix} (v_t^* - \beta_2 u_t^2) = b^{-1} \sum_{N t=1}^{b_N} \begin{pmatrix} 1 \\ x_t^2 \\ x_t \end{pmatrix} v_t^* + o_p(N^{-1/2})
\]

\[
= O_p(N^{-1/2}) .
\] (4.7)

Finally, from Equations 4.7, 4.6, and 4.5 we have

\[
\hat{\beta}_{OLS} - \beta = (\bar{m}_{ww})^{-1} \left\{ b^{-1} \sum_{N t=1}^{b_N} \begin{pmatrix} 1 \\ x_t \\ x_t^2 \end{pmatrix} v_t^* \right\} + o_p(N^{-1/2})
\]

\[
= O_p(N^{-1/2}) . \quad \text{Q.E.D.} \] (4.8)

**Lemma 4.2**

Under Assumptions 4.1a, 4.3, 4.4, 4.6, and 4.7b,

\[
\hat{\beta}_{OLS} - \beta = O_p(\max[ a_N, N^{-1/2} ]).
\]

**Proof:**

This proof follows the proof of Lemma 4.1 exactly. Here, however, it is found that
By the results established in these lemmas, it is clear that \( \hat{\beta}_{OLS} \) satisfies the requirements of Assumption 4.10. Consequently, \( \hat{\beta}_{OLS} \) may be used as a preliminary estimator in the iterative procedures to be discussed later in this chapter.

We now consider another candidate for preliminary estimator.

C. The Preliminary Estimator \( \hat{\beta} \)

In this section we reconsider the estimator

\[
\hat{\beta} = \left[ \overline{M}_{WW} - \hat{\alpha}_f \hat{f}_{ff} \right]^{-1} \left[ \overline{M}_{WY} - \hat{\alpha}_f \hat{f}_{fe} \right] \tag{4.9}
\]
where \( \hat{\alpha} \) is the smallest root of the equation

\[
|M - \alpha \hat{x}| = 0.
\]  

(4.10)

This estimator was considered in Chapter 3, and we employ the same notation here as there, with the exception that all sums are taken over \( t = 1, \ldots, b_N \) in this section. The aim of this work will be to establish the asymptotic properties of \( \hat{\beta} \) under the new assumptions set forth in Section A.

We now state and prove several lemmas which lead to the main theorems of this section.

**Lemma 4.3**

Under Assumptions 4.1b, 4.2, 4.3, 4.4, 4.6, 4.7b, and 4.11, \( \hat{\alpha} \), defined by Equation 4.10, is bounded in probability.

**Proof:**

Since \( \hat{\alpha} \) is obtained by choosing that \( \xi \) which minimizes the ratio

\[
\frac{\xi' \bar{M} \xi}{\xi' \hat{F} \xi}
\]

it follows that

\[
\alpha = \min_{\xi} \frac{\xi' \bar{M} \xi}{\xi' \hat{F} \xi} \leq \frac{\hat{\xi}' \bar{M} \hat{\xi}}{\hat{\xi}' \hat{F} \hat{\xi}} = \frac{b^{-1} \sum_{t=1}^{b} \frac{N}{N} v_t^2}{\frac{2\sigma^2}{\nu}},
\]  

(4.11)
where \( \theta' = (1, -\beta') \),
\[
\nu_t = y_t - W_t \beta = e_t - f' \beta,
\]
and \( \frac{\hat{\sigma}^2}{\nu} = \theta' \hat{\Sigma} \theta \). By Theorem 2.2 and Assumptions 4.1b, 4.4, 4.6, 4.7b, and 4.11,
\[
b^{-1} \sum_{t=1}^{bN} v_t^2 = E \{ b^{-1} \sum_{t=1}^{bN} v_t^2 \} + o_p(N^{-1/2})
\]
\[
= \theta' \hat{\Sigma} \theta + o_p(N^{-1/2})
\]
and
\[
\theta' \hat{\Sigma} \theta = \theta' \hat{\Sigma} \theta + o_p(N^{-1/2}).
\]
Thus we have
\[
0 \leq \alpha \leq \frac{a^{-1} N \sum v_t^2}{a^{-1} N} = 1 + o_p(1).
\]
(4.13)
\(\text{Q.E.D.}\)

**Lemma 4.4**

Under Assumptions 4.1b, 4.2, 4.3, 4.4, 4.6, 4.7b, and 4.11,
\[
\hat{\beta} - \beta = O_p(\max[ a_N, N^{-1/2} ]).
\]
Proof:

By Lemma 4.3, Assumptions 4.6 and 4.11 we have $\hat{\alpha} \hat{\delta} = O_p (a_N)$.

Thus

$$M_{WW} - \hat{\alpha} \hat{\delta} = M_{WW} + O_p (a_N)$$

and

$$M_{WY} - \hat{\alpha} \hat{\delta} = M_{WY} + O_p (a_N).$$

It then follows from Equations 4.9 and 4.2 that

$$\hat{\beta} = \hat{\beta}_{OLS} + O_p (a_N).$$

From Lemma 4.2 we obtain the result

$$\hat{\beta} - \beta = O_p (\max[ a_N, N^{-1/2}]). \quad (4.14)$$

Q.E.D.

Now recall the following notation:

$$\Delta \bar{M} = \bar{M} - E(\bar{M}) = b^{-1} \sum_{t=1}^{bN} (z_t' \varepsilon_t + \varepsilon_t' \bar{z}_t)$$

$$+ b^{-1} \sum_{t=1}^{bN} (\varepsilon_t' \varepsilon_t - \frac{1}{b} \varepsilon_t' \varepsilon_t).$$
\[\Delta \alpha = \hat{\alpha} - 1,\]
\[\Delta \hat{\lambda} = \hat{\lambda} - \lambda,\]
\[\hat{\theta}' = (1, -\beta'),\]
and \[\Delta \theta = \hat{\theta} - \theta\] Also recall Equation 3.21

\[
\Delta \alpha = \frac{\theta'(\Delta \hat{\lambda}) \theta + \theta'(\Delta \hat{\lambda})(\Delta \theta) - \theta'(\Delta \hat{\lambda}) \theta}{\theta'(\Delta \hat{\lambda}) \theta + \theta'(\Delta \hat{\lambda}) \theta}.
\] (3.21)

Since \[z_t \theta = 0\]
and
\[
b^{-1} \sum_{t=1}^{N} (\epsilon_t' \epsilon_t - \frac{1}{b} \epsilon_t') = O(a^{1/2} N^{-1/2})
\]
we obtain
\[
\theta'(\Delta \hat{\lambda}) \theta = \theta'[b^{-1} \sum_{t=1}^{N} (\epsilon_t' \epsilon_t - \frac{1}{b} \epsilon_t')] \theta
\]
\[= O(a^{-1/2} N^{-1/2}).\] (4.15)

By Assumptions 4.1b, 4.4, 4.6, and 4.11

\[\Delta \hat{\lambda} = O_p(N^{-1/2})\] and \[\Delta \hat{\lambda} = O_p(a^{-1} N^{-1/2}).\]

and by Lemma 4.4 \[\Delta \theta = o_p(1).\] Thus from Equations 3.21 and 4.15 we obtain
\[ a_N(\Delta \alpha) = \frac{\theta'(\Delta \hat{M}) \theta + \theta'(\Delta \hat{M})(\Delta \hat{\theta}) - \theta'(\Delta \hat{\theta})\hat{\theta}}{\theta'(\hat{a}^{-1}_N \hat{\theta}) + \theta'(\hat{a}^{-1}_N \Delta \hat{\theta}) \hat{\theta}} \]

\[ = \frac{\theta'(\Delta \hat{M}) \theta + o_p(N^{-1/2})}{\theta'(\hat{a}^{-1}_N \hat{\theta}) + o_p(N^{-1/2})} \]

\[ = O_p(a^{-1/2}_N N^{-1/2}). \quad (4.16) \]

We have just proved Lemma 4.5.

**Lemma 4.5**

Under Assumptions 4.1b, 4.2, 4.3, 4.4, 4.6, 4.7b, and 4.11, it follows that

\[ \Delta \alpha = O_p(a^{-1/2}_N N^{-1/2}) = O_p(b^{-1/2}_N). \]

As in Chapter 3, we can write

\[ (\overline{M}_{WW} - \hat{\alpha}_{\hat{f}_f}) = \overline{m}_{ww} + \hat{a} \]

\[ (\overline{M}_{WX} - \hat{\alpha}_{\hat{f}_f}) = \overline{m}_{wx} + \hat{b}, \quad (4.17) \]

and

where \[ \hat{a} = b^{-1}_N \sum_{t=1}^{N} (w'_t f'_t + f'_t w'_t) + \]

\[ \hat{b} \]
\[ b^{-1} \sum_{t=1}^{N} \left( f_t^t f_t - \frac{x}{2} f_t f(t) \right) \]

\[ \Delta \alpha \frac{\bar{x}}{f} - (1 + \Delta \alpha)(\Delta \frac{\bar{x}}{f}) \]

and

\[ \hat{b} = b^{-1} \sum_{t=1}^{N} \left( w_t e_t + f_t y_t \right) \]

\[ + b^{-1} \sum_{t=1}^{N} \left( f_t e_t - \frac{x}{2} f_t f(t) \right) \]

\[ = (\Delta \alpha) \frac{\bar{x}}{f} - (1 + (\Delta \alpha))(\Delta \frac{\bar{x}}{f}) \]

Under the Assumptions of Lemma 4.5, \( \hat{a} \) may be written as

\[ \hat{a} = b^{-1} \sum_{t=1}^{N} \left( w_t^t f_t + f_t^t w_t \right) + o_p(N^{-1/2}) \]

\[ = O_p(N^{-1/2}) \] \hspace{1cm} \text{(4.18)}

and \( \hat{b} \) may be written as

\[ \hat{b} = b^{-1} \sum_{t=1}^{N} \left( w_t e_t + f_t y_t \right) + o_p(N^{-1/2}) \]

\[ = O_p(N^{-1/2}) \] \hspace{1cm} \text{(4.19)}
Utilizing Theorem 2.11 and Equations 4.9, 4.17, 4.18, and 4.19 we obtain

\[
\hat{\beta} = [\left(\overline{M}_{WW} - \hat{\alpha}\hat{F}_{ff}\right)^{-1}\left[\overline{M}_{WY} - \hat{\alpha}\hat{F}_{fe}\right] = \left[\overline{m}_{ww} + a\right]^{-1}\left[\overline{m}_{wy} + \hat{b}\right]
= \left[\overline{m}_{ww}^{-1} - \overline{m}_{ww}^{-1} \hat{a} \overline{m}_{ww}^{-1}\right]\left[\overline{m}_{wy} + \hat{b}\right] + O_p(N^{-1})
= \beta + \overline{m}_{ww}^{-1}\left[\hat{b} - \hat{\alpha}\beta\right] + O_p(N^{-1}) .
\]

(4.20)

Lemma 4.6 then follows from Equations 4.18, 4.19, and 4.20.

**Lemma 4.6**

Under Assumptions 4.1b, 4.2, 4.3, 4.4, 4.6, 4.7b, and 4.11, it follows that

\[
\Delta\beta = \hat{\beta} - \beta = O_p(N^{-1/2}) .
\]

We are now able to prove the main theorems of this section.

**Theorem 4.1**

Given Assumptions 4.1b, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7b, and 4.11, it follows that
\[
\sqrt{N} (\hat{\beta} - \beta) \xrightarrow{d} N(0, \lim_{N \to \infty} m_{\mathbf{WW}} - 1 \left\{ b N \sum_{t=1}^{b N} a^{-1} \sigma_{w_t w_t}^{2} \right\} m_{\mathbf{WW}} - 1)
\]

where \( \sigma_{v_t}^{2} = \sigma_{e}^{2} - 2 \gamma \sigma_{e u} + \gamma^{2} \sigma_{u}^{2} \) and \( \gamma = \beta_1 + 2 x_2 \beta_2 \).

Proof:

Define \( v_t = e_t - f_t \beta \). Then from Equations 4.18 and 4.19 we obtain

\[
\hat{v}_t = b_{-1} \sum_{t=1}^{b N} w_t' (e_t - f_t \beta) \\
+ b_{-1} \sum_{t=1}^{b N} f_t (v_t - w_t \beta) + o_p (N^{-1/2}) \\
= b_{-1} \sum_{t=1}^{b N} w_t' v_t + o_p (N^{-1/2}) \quad (4.21)
\]

since \( v_t - w_t \beta = 0 \).

From Equations 4.21 and 4.20, we have

\[
\sqrt{N} (\hat{\beta} - \beta) = \frac{-1}{m_{\mathbf{WW}}} \left\{ \sqrt{N} b^{-1} \sum_{t=1}^{b N} w_t' v_t \right\} + o (1).
\]

It then follows that the limiting distribution of \( \sqrt{N} (\hat{\beta} - \beta) \) is the limiting distribution of \( \frac{-1}{m_{\mathbf{WW}}} \left\{ \sqrt{N} b^{-1} \sum_{t=1}^{b N} w_t' v_t \right\} \) (cf. Theorem 2.6).
Consider the limiting distribution of

\[ \sqrt{N} b^{-1} \sum_{N}^{b} \lambda' w^t v_t \]

where \( \lambda \) is an arbitrary, nonzero (3 x 1) vector. By Assumption 4.1b the random variables \((\lambda' w^t v_t)\) are independent and have zero mean. Also,

\[ \text{Var}(\lambda' w^t v_t) = \lambda' w^t E(v_t^2) w_t \lambda < \infty \]

and

\[ E[|\lambda' w^t v_t|^{2+\delta}] < \infty \]

for \( \delta > 0 \), since \((e_t, u_t)\) possesses finite \(4+\delta\) moments for \( t = 1, 2, \ldots \). To see this recall that \( v_t \) is a function of \( u_t^2 \). Consequently, from Assumption 4.4 we obtain

\[ \lim_{N \to \infty} \frac{b_N \sum_{t=1}^{\left[ \frac{2+\delta}{2} \right]} E[|\lambda' w^t v_t|^{2+\delta}]}{b_N \left[ \sum_{t=1}^{\left[ \frac{2+\delta}{2} \right]} E[(\lambda' w^t v_t)^2] \right]} = 0 \]

and by Assumption 4.5 and the Liapounov central limit theorem we
must have

\[ \sqrt{N} b^{-1} \sum_{t=1}^{b_N} \lambda' w_t v_t \]

\[ \{ \lambda' b^{-1} \sum_{t=1}^{b_N} w_t w_t E(N b^{-1} v_t^2) \lambda \}^{1/2} \]

\[ \xrightarrow{d} N(0, 1) . \]

Furthermore, by our assumptions,

\[ E(N b^{-1} v_t^2) = a_N^{-1} E(v_t^2) \]

\[ = a_N^{-1} \sigma_{v_t}^2 + o(1) = O(1) \]

and \( b^{-1} \sum_{t=1}^{b_N} a_{N v_t}^{-1} \sigma_{w_t} w_t w_t \) converges. Thus, by Theorem 2.7,

\[ \sqrt{N} b^{-1} \sum_{t=1}^{b_N} \lambda' w_t v_t \xrightarrow{d} N(0, \lambda' \lim_{N \to \infty} b^{-1} \sum_{t=1}^{b_N} a_{N v_t}^{-1} \sigma_{w_t} w_t w_t \lambda) , \]

and by the multivariate central limit theorem

\[ \sqrt{N} b^{-1} \sum_{t=1}^{b_N} w_t v_t \xrightarrow{d} N(0, \lim_{N \to \infty} b^{-1} \sum_{t=1}^{b_N} a_{N v_t}^{-1} \sigma_{w_t} w_t w_t) . \]
To obtain the limiting distribution of \( \sqrt{N} (\hat{\beta} - \beta) \), we note that

\[
\lim_{N \to \infty} \frac{\bar{m}}{\text{ww}} = \frac{m}{\text{ww}} \text{ exists. Therefore, by Theorem 2.7 and Assumption 4.4, we have}
\]

\[
\frac{\bar{m}^{-1} \sqrt{N} b^{-1} \sum_{t=1}^{bN} w_t v_t}{N} \xrightarrow{d} -1 \\
N(0, \lim_{N \to \infty} \frac{\bar{m}^{-1} \left[ b^{-1} \sum_{t=1}^{bN} a^{-1} \sigma^2 \sigma \right] \bar{m}^{-1}}{N})
\]

and the conclusion of the theorem follows. \( \text{Q.E.D.} \)

In each of Lemmas 4.3 through 4.6, as well as in Theorem 4.1, we specified Assumption 4.7b, i.e. \( a_N^{1.5} = o(N^{-1/2}) \). By these results, we are able to use \( \hat{\beta} \) as a preliminary estimator in each of the iterative procedures to be given which specify Assumption 4.7b and which require a preliminary estimator whose error is \( o_p(N^{-1/2}) \).

The important results of Theorem 4.1 can also be established for \( \hat{\beta} \) when Assumption 4.7a replaces 4.7b, i.e. \( a_N = o(N^{-1/2}) \). This fact is summarized by Theorem 4.2. The proof is similar to the proof of Theorem 4.1 and is thus deleted.

**Theorem 4.2**

Given Assumptions 4.1a, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7a, and 4.11, it follows that
By this result, we are also able to use $\hat{\beta}$ as a preliminary estimator in an iterative scheme which specifies Assumption 4.7a and which requires a preliminary estimator whose error is $O_p(N^{-1/2})$.

Now, let us consider some iterative procedures which are derived from the likelihood function.

D. The Likelihood Function and Iterative Estimators

1. The maximum likelihood estimator

Assuming the random variables $(e_t, u_t)$ are normally distributed, the maximum likelihood estimators (MLE) of $\beta$ and $x_t$, $t = 1, \ldots, b_N$, for Model 4.1a, 4.1b, say $\tilde{\beta}$ and $\tilde{x}_t$, are those values of $\beta$ and $x_t$ which maximize the likelihood function

$$L(\beta, x_1, x_2, \ldots, x_{b_N}) =$$

$$\frac{-b_N}{2} - \frac{b_N}{2} \exp \left\{ \frac{1}{2} \sum_{t=1}^{b_N} \left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \\ X_t - x_t \end{array} \right] \right\}^{-1}$$

$$\left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \\ X_t - x_t \end{array} \right].$$
Since the logarithm function is monotonically increasing,

\[ L(\beta, x_1, \ldots, x_b) \] and \( \log L(\beta, x_1, \ldots, x_b) \) are maximized by

the same value of \((\beta, x_1, \ldots, x_b)\). Consequently, \( \tilde{\beta} \) and \( \tilde{x}_t \) are

those values of \( \beta \) and \( x_t \) which maximize

\[
\log L(\beta, x_1, \ldots, x_b) = -b_N \log (2\pi) - b_N/2 \log |\Phi|
\]

\[
-\frac{1}{2} \sum_{t=1}^{b_N} \left[ Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \right] \frac{1}{2} \left[ X_t - x_t \right]
\]

Clearly then, \( \tilde{\beta} \) and \( \tilde{x}_t \) are those values of \( \beta \) and \( x_t \) which minimize

the function

\[
Q(\beta, x_1, \ldots, x_b) =
\]

\[
\sum_{t=1}^{b_N} \left[ Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \right] \frac{1}{2} \left[ Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \right]
\]

\[
\frac{1}{2} \left[ X_t - x_t \right] \frac{1}{2} \left[ X_t - x_t \right]
\]

(4.22)

where \( \Phi^{-1} = \begin{bmatrix} \sigma_{ee} & \sigma_{eu} \\ \sigma_{ue} & \sigma_{uu} \end{bmatrix} \).
Differentiating $Q$ with respect to $\beta' = (\beta_0, \beta_1, \beta_2)$ and $x_t$, $t = 1, \ldots, b_N$, and setting the resulting derivatives to zero yields the following system of likelihood equations:

$$0 = \sigma_{ee}(Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2)(\beta_1 + 2 x_t \beta_2)$$

$$+ \sigma_{eu} [(Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2) - (\beta_1 + 2 x_t \beta_2)(X_t - x_t)]$$

$$+ \sigma_{uu}(X_t - x_t)$$

for $t = 1, 2, \ldots, b_N$; and

$$\left\{ \begin{array}{c}
\sigma_{ee} \sum_{t=1}^{b_N} \left( \begin{array}{c}
1 \\
x_t^1 \\
x_t^2
\end{array} \right) \left[ \begin{array}{c}
1 \\
x_t \\
x_t^2
\end{array} \right] \beta = \\
\sigma_{eu} \sum_{t=1}^{b_N} (X_t - x_t) \left( \begin{array}{c}
1 \\
x_t \\
x_t^2
\end{array} \right)
\end{array} \right\} .$$

$$\left( \begin{array}{c}
\sigma_{ee} \\
\sigma_{eu} \\
\sigma_{uu}
\end{array} \right) = (\sigma_{ee}^2 - \sigma_{eu}^2)^{-1} \left( \begin{array}{c}
\sigma_{ee}^2 \\
-\sigma_{eu}^2 \\
-\sigma_{ue} \sigma_{eu}
\end{array} \right)$$
Equation 4.24 may alternatively be expressed as

\[
\left\{ \frac{b_N}{\sum_{t=1}^{b_N}} \begin{pmatrix} \sim x_t \\ \sim^2 x_t \end{pmatrix} \begin{bmatrix} 1 & \sim x_t & \sim^2 x_t \end{bmatrix} \right\} \sim \beta =
\]

\[
\left\{ \frac{b_N}{\sum_{t=1}^{b_N}} Y_t \begin{pmatrix} \sim x_t \\ \sim^2 x_t \end{pmatrix} - \frac{\sigma_{eu}^2}{\sigma_u^2} \frac{b_N}{\sum_{t=1}^{b_N}} (X_t - \sim x_t) \begin{bmatrix} 1 \\ \sim x_t \\ \sim^2 x_t \end{bmatrix} \right\}, \quad (4.25)
\]

and Equation 4.23 may be expressed as

\[
0 = \sigma_u^2 (Y_t - \sim \beta_0 - \sim \beta_1 x_t - \sim \beta_2 x_t^2)(\beta_1 + 2x_t \beta_2)
\]

\[
- \sigma_{eu} \left[ (Y_t - \sim \beta_0 - \sim \beta_1 x_t - \sim \beta_2 x_t^2) - (\sim \beta_1 + 2x_t \beta_2)(X_t - \sim x_t) \right] \quad (4.26)
\]

\[
+ \sigma_e^2 (X_t - \sim x_t).
\]

The maximum likelihood estimators of \( \beta \) and \( x_t \), \( t = 1, \ldots, b_N \), are thus given by simultaneous solution of Equations 4.25 and 4.26. These likelihood equations, however, do not necessarily provide an intuitive grasp of the situation. Nevertheless, the geometric interpretation of maximum likelihood is clear. From Equation 4.22 it is
seen that \( \tilde{\beta} \) and \( \tilde{x}_t \) are chosen such that the sum of squares of the of the weighted distances between the fitted parabola and the observed data points, \((X_t, Y_t), t = 1, \ldots, b_N\), is minimized. This is graphically displayed in Figure 4.1 for the case \( \frac{1}{\psi} = \sigma^2 I \).

\[
y = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\beta}_2 x^2
\]

**Figure 4.1.** Geometric interpretation of \( \tilde{\beta} \) and \( \tilde{x}_t \) where

\[
\frac{1}{\psi} = \sigma^2 I
\]
Returning to the likelihood equations, it is seen that the expression defining $\hat{\beta}$, Equation 4.25, and the expression defining $x_t$, Equation 4.26, are nonlinear. As a result, an explicit expression for the maximum likelihood estimator of $\beta$ has not been obtained. Consequently, two important questions arise. First, how does one compute the maximum likelihood estimators. Second, what properties do the resulting estimators possess. These questions are dealt with in the next section.

2. A pseudo-maximum likelihood estimator

In this section an estimator of $\beta$ will be constructed by minimization of a functional approximation to the $Q$ of Equation 4.22, say $Q'$. To define the function $Q'$ we require a preliminary estimator of $\beta$, say $\bar{\beta}$, satisfying

$$\bar{\beta} - \beta = O_p(N^{-1/2}). \tag{4.27}$$

We also require an initial estimator, say $\hat{x}_t$, of $x_t$ for $t = 1, \ldots, b_N$.

For each $t = 1, \ldots, b_N$ we define the estimator $\hat{x}_t$ to be the real valued root of the polynomial
\[ P_t(x) = \sigma_u^2 (Y_t - \beta_0 - \beta_1 x - \beta_2 x^2)(\beta_1 + 2x\beta_2) \]

\[ -\sigma_{eu} [(Y_t - \beta_0 - \beta_1 x - \beta_2 x^2) - (\beta_1 + 2x\beta_2)(X_t - x)] \]

\[ + \sigma_e^2 (X_t - x) \]

\[ = 0 \quad \text{(4.28)} \]

which minimizes the expression

\[ \begin{bmatrix} Y_t - \beta_0 - \beta_1 x - \beta_2 x^2 \\ X_t - x \end{bmatrix}^t \Phi^{-1} \begin{bmatrix} Y_t - \beta_0 - \beta_1 x - \beta_2 x^2 \\ X_t - x \end{bmatrix} \]

with respect to the set of \( x \) which are roots of Equation 4.28. Since \( P_t(x) \) is a cubic equation in \( x \), one or three real roots are possible. A graphical demonstration of these possibilities is given in Figures 4.2a and 4.2b.

Expression 4.28 is derived from likelihood Equation 4.26. The maximum likelihood estimator \( \tilde{\beta} \) in Equation 4.26 is replaced by the preliminary estimator \( \bar{\beta} \) in Equation 4.28.

To define \( Q' \), we expand \( y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 \) in a Taylor series about the point \((\bar{\beta}, \hat{x}_1, \ldots, \hat{x}_N)\), displaying only the linear terms. This yields
Figure 4.2a. Three real roots

Figure 4.2b. One real root
\[ y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 \]

\[ = (\beta_0 + \beta_1 \hat{x}_t + \beta_2 \hat{x}_t^2) + (\beta_0 - \bar{\beta}_0) \]

\[ + \hat{x}_t (\beta_1 - \bar{\beta}_1) + \hat{x}_t^2 (\beta_2 - \bar{\beta}_2) + (\bar{\beta}_1 + 2\hat{x}_t \bar{\beta}_2)(x_t - \hat{x}_t) + R_t \]

\[ = (\beta_0 + \beta_1 \hat{x}_t + \beta_2 \hat{x}_t^2) + (\Delta \beta_0) \]

\[ + \hat{x}_t (\Delta \beta_1) + \hat{x}_t^2 (\Delta \beta_2) + d_t (\Delta x_t) + R_t \quad (4.29) \]

where \[ \Delta \beta = \begin{bmatrix} \Delta \beta_0 \\ \Delta \beta_1 \\ \Delta \beta_2 \end{bmatrix} = \beta - \bar{\beta} = \begin{bmatrix} \beta_0 - \bar{\beta}_0 \\ \beta_1 - \bar{\beta}_1 \\ \beta_2 - \bar{\beta}_2 \end{bmatrix} \]

\[ \Delta x_t = x_t - \hat{x}_t \]

\[ d_t = \bar{\beta}_1 + 2\hat{x}_t \bar{\beta}_2 \]

and \[ R_t = \text{a remainder} \]. Substituting the linear portion of Equation 4.29 into Equation 4.22 and calling the result \( Q' \) yields the following definition:

\[ Q'(\Delta \beta, \Delta x_1, \ldots, \Delta x_{b_N}) = \]

\[ \sum_{t=1}^{b_N} \left[ (Y_t - \bar{\beta}_0 - \bar{\beta}_1 x_t - \bar{\beta}_2 x_t^2) - (\Delta \beta_0) - \hat{x}_t (\Delta \beta_1) - \hat{x}_t (\Delta \beta_2) - d_t (\Delta x_t) \right] \]

\[ \left( X_t - \hat{x}_t \right) - (\Delta x_t) \]
where $\hat{e}_t = Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2$ and $u_t = X_t - \hat{x}_t$.

We seek to minimize the function $Q'$ with respect to $(\Delta \beta, \Delta x_1, \ldots, \Delta x_N)$.

Differentiating Equation 4.30 with respect to $(\Delta x_t)$ and setting the derivative equal to zero yields
\[
\frac{\partial Q'}{\partial (\Delta x_t)} = \left(\sigma_u^2 \sigma_e^2 - \sigma_{eu}^2\right)^{-1}
\]

\[
\left[\sigma_u^2 (-2d_t) \left[\hat{e}_t - (\Delta \beta_0) - \hat{x}_t (\Delta \beta_1) - \hat{x}_t^2 (\Delta \beta_2) - d_t (\Delta x_t)\right] - 2\sigma_{eu} \left[-e_t - (\Delta \beta_0) - x_t (\Delta \beta_1) - x_t^2 (\Delta \beta_2) - d(\Delta x_t)\right] - d_t (\Delta x_t)\right]
\]

\[
-2\sigma_e^2 [\hat{u}_t - (\Delta x_t)]
\]

\[= 0\]

Then rearranging terms and naming the resulting solution \((\Delta \hat{x}_t)\) and \((\Delta \hat{\beta})\) yields

\[
(\Delta \hat{x}_t) = \frac{\hat{u}_t (\sigma_e^2 - d_t \sigma_{eu}) + (\sigma_u^2 \sigma_e - \sigma_{eu}) \left[\hat{e}_t - (\Delta \beta_0) - x(\Delta \beta_0) - x(\Delta \beta_1) - x^2 (\Delta \beta_2)\right]}{d^2 \sigma_u - 2d \sigma_{eu} + \sigma_e^2}
\]

(4.31)

Differentiating Equation 4.30 with respect to \((\Delta \beta)\) and setting the derivative to zero yields
Rearranging terms and naming the resulting solution $(\hat{\Delta} \beta)$ and $(\hat{\Delta} x_t)$ yields

$$\frac{\partial Q'}{\partial (\Delta \beta)} = 2 \sum_{t=1}^{b_N} (\sigma_e^2 \sigma_u^2 - \sigma_{eu}^2)^{-1} \left[ \begin{array}{c} 1 \\ \hat{e}_t \end{array} \right] - \hat{x}_t (\Delta \beta_1) - \hat{x}_t^2 (\Delta \beta_2) - d_t (\Delta x_t)^T$$

$$= 2 \sum_{t=1}^{b_N} (\sigma_e^2 \sigma_u^2 - \sigma_{eu}^2)^{-1} \left[ \begin{array}{c} 1 \\ -\sigma_{eu} \hat{u}_t - (\Delta x_t) \end{array} \right]$$

where $g_t = \hat{e}_t - (\hat{\Delta} \beta_0) - \hat{x}_t (\Delta \beta_1) - \hat{x}_t^2 (\Delta \beta_2)$.
If we substitute Equation 4.31 into Equation 4.32, we obtain

$$\sum_{t=1}^{N} \sigma_{u}^{2} \left( g_{t} - d_{t} \left\{ \frac{\hat{u}_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) + (\sigma_{d_{t}}^{2} - \sigma_{d_{t}})g_{t}}{d_{t}\sigma_{u}^{2} - 2d_{t}\sigma_{e} + \sigma_{e}^{2}} \right\} \right)$$

$$-\sigma_{e}u_{t} \left\{ \frac{\hat{u}_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) + (\sigma_{d_{t}}^{2} - \sigma_{d_{t}})g_{t}}{d_{t}\sigma_{u}^{2} - 2d_{t}\sigma_{e} + \sigma_{e}^{2}} \right\} \right\} \left[ \begin{array}{c} 1 \\ x_{t} \\ \bar{x}_{t}^{2} \end{array} \right]$$

$$= 0 . \quad (4.33)$$

Letting

$$\sigma_{v_{t}}^{2} = d_{t}\sigma_{e}^{2} - 2d_{t}\sigma_{e} + \sigma_{e}^{2},$$

Equation 4.33 may be rewritten as

$$\sum_{t=1}^{N} \sigma_{u}^{2} \left( g_{t} - d_{t} \left\{ \frac{\hat{u}_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) + (\sigma_{d_{t}}^{2} - \sigma_{d_{t}})g_{t}}{d_{t}\sigma_{u}^{2} - 2d_{t}\sigma_{e} + \sigma_{e}^{2}} \right\} \right)$$

$$-\sigma_{e}u_{t} \left\{ \frac{\hat{u}_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) + (\sigma_{d_{t}}^{2} - \sigma_{d_{t}})g_{t}}{d_{t}\sigma_{u}^{2} - 2d_{t}\sigma_{e} + \sigma_{e}^{2}} \right\} \right\} \left[ \begin{array}{c} 1 \\ x_{t} \\ \bar{x}_{t}^{2} \end{array} \right] =$$

$$\frac{b_{N}}{\sum_{t=1}^{N} \sigma_{v_{t}}^{2}} \left\{ \sigma_{u}^{2} \left[ g_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) - d_{t}\hat{u}_{t}(\sigma_{e}^{2} - d_{t}\sigma_{e}) \right] \right\}$$
\[-\sigma_{eu} \left( d \hat{u}_t \left( \sigma_{u}^2 - \sigma_{eu} \right) \right) \sigma_{t} \hat{u}_t = 0. \]

Finally, from the definition of \( g_t \), an explicit expression for \( (\Delta \hat{\beta}) \) is obtained:

\[
b^{-1} \sum_{N}^{bN} a_{N}^{-1} \left[ \begin{array}{c}
1
\end{array} \right] \left[ \begin{array}{c}
\hat{x}_{t}
\end{array} \right] \left( \hat{x}_{t}^{2} \right) (\Delta \hat{\beta}) = \]

\[
b^{-1} \sum_{N}^{bN} a_{N}^{-1} \left[ \begin{array}{c}
(\hat{e}_t - e_t \hat{u}_t)
\end{array} \right] \left[ \begin{array}{c}
\hat{x}_{t}
\end{array} \right] \left( \hat{x}_{t}^{2} \right). \tag{4.34} \]

We have now minimized \( Q^1 \) and have determined that \( (\Delta \hat{\beta}) \), given by Equation 4.34, is the minimizing value of \( (\Delta \hat{\beta}) \). A revised estimator of \( \beta \), say \( \hat{\beta} \), may now be constructed by defining

\[
\hat{\beta} = \overline{\beta} + (\hat{v})(\Delta \hat{\beta}) \tag{4.35} \]

where \( \hat{v} \) is that \( v \in [0, 1] \) which minimizes

\[
Q(\overline{\beta} + v(\Delta \hat{\beta}), \hat{x}_1, \ldots, \hat{x}_{bN})
\]
with respect to $v$.

In the next section we will investigate the asymptotic properties of $\hat{x}_t$, given by Equation 4.28, and of $(\hat{A}\beta)$ given by Equation 4.34. But first, we summarize the estimation procedure suggested by this section:

1. Compute a preliminary estimator of $\beta$, say $\hat{\beta}$, satisfying
   
   $$\hat{\beta} - \beta = O_p(N^{-1/2}).$$
   
   (Recall that both $\hat{\beta}$ and $\hat{\beta}_{OLS}$ satisfy this requirement.)

2. For each $t = 1, 2, \ldots, b_N$, compute $x_t$ defined by Equation 4.28.

3. Compute $(\hat{A}\beta)$ as defined by Equation 4.34.

4. Compute $\hat{\beta}$ as defined by Equation 4.35.

5. Return to step one and iterate if desired.

3. Asymptotic properties

This section is concerned with the asymptotic properties of the estimators defined in the previous section. We begin by proving three lemmas which establish the asymptotic properties of $\hat{x}_t$.

Lemma 4.7

Under Assumptions 4.6, 4.7a, and 4.10,

$$\hat{x}_t - x_t = O_p(a_N^{1/2})$$
for each \( t = 1, \ldots, b_N \).

Proof:

By definition, \( \hat{x}_t \) is obtained as that value of \( x \) which minimizes the function

\[
D^2_t(x) = \left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x - \beta_2 x^2 \\ X_t - x \end{array} \right] a_N^{-1} \left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x - \beta_2 x^2 \\ X_t - x \end{array} \right].
\]

Thus \( \sqrt{D^2_t(\hat{x}_t)} \leq \sqrt{D^2_t(x_t)} \). But

\[
D^2_t(x_t) = \left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \\ X_t - x_t \end{array} \right] a_N^{-1} \left[ \begin{array}{c} Y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2 \\ X_t - x_t \end{array} \right].
\]

\[= O_p(a_N)\]

since \( e_t = O_p(a_N^{1/2}) \), \( u_t = O_p(a_N^{1/2}) \), and \( \bar{\beta} - \beta = O_p(N^{-1/2}) \). This gives

\[0 \leq \sqrt{D^2_t(\hat{x}_t)} \leq \sqrt{D^2_t(x_t)} = O_p(a_N^{1/2}). \quad (4.36)\]
Observe that \( D_t^2(\hat{x}_t) \) may be written as

\[
D_t^2(\hat{x}_t) = a_N \sigma_{u}^{-2} (X_t - \hat{x}_t)^2 + a_N \left[ \sigma_e^2 - \frac{\sigma_{eu}}{2} \right]^{-1} \left\{ (Y_t - \beta_0 - \beta_1 X_t - \beta_2 \hat{x}_t)^2 - \frac{\sigma_{eu}}{\sigma_u^2} (X_t - \hat{x}_t)^2 \right\}.
\]

Thus, since \( \left| \frac{\sigma_{eu}}{\sigma_u^2} \right| = \frac{\sigma_e^2}{\sigma_u^2} - \frac{\sigma_{eu}^2}{\sigma_u^4} > 0 \), we obtain

\[
D_t^2(\hat{x}_t) \geq (a_N \cdot \sigma_{u}^{-2})(X_t - \hat{x}_t)^2.
\] (4.37)

From Equations 4.36 and 4.37 it follows that \( (X_t - \hat{x}_t) = O_p \left( \frac{1}{a_N^{1/2}} \right) \).

But \( X_t - \hat{x}_t = u_t = O_p \left( \frac{1}{a_N^{1/2}} \right) \). Consequently,

\[
(\hat{x}_t - x_t) = O_p \left( \frac{1}{a_N^{1/2}} \right).
\]  Q.E.D.

**Lemma 4.8**

Under Assumptions 4.6, 4.7a, and 4.10,

\[
\hat{x}_t = x_t + \delta_t + O_p (N^{-1/2})
\]

for \( t = 1, \ldots, b_N \), where
\[
\delta_t = (a_{\text{N}}^{-1} \sigma^2)_{v_t}^{-1} \left[ u_t (\sigma^2_{e} - \gamma_t \sigma_{eu}) a_{\text{N}}^{-1} e_t (\gamma_t \sigma^2_{u} - \sigma_{eu}) a_{\text{N}}^{-1} \right]
\]

and \( \gamma_t = \beta_1 + 2x_t \beta_2 \).

Proof:

From Equation 4.28, \( \hat{x}_t \) satisfies the cubic equation

\[
0 = \sigma_u^2 (Y_t - \bar{\beta}_0 - \bar{\beta}_1 \hat{x}_t - \bar{\beta}_2 \hat{x}_t^2) (\bar{\beta}_1 + 2\hat{x}_t \hat{\beta}_2) - \sigma_{eu} \left[ (Y_t - \bar{\beta}_0 - \bar{\beta}_1 \hat{x}_t - \bar{\beta}_2 \hat{x}_t^2) - (\bar{\beta}_1 + 2\hat{x}_t \hat{\beta}_2) (X_t - \hat{x}_t) \right] + \sigma_e^2 (X_t - \hat{x}_t) .
\]

By rearranging terms we may rewrite this as

\[
\hat{x}_t = \frac{a_{\text{N}}^{-1} \left\{ X_t (\sigma^2_{e} - \sigma_{eu} d_t) + (\sigma^2_{u} d_{t} - \sigma_{eu}) (Y_t - \bar{\beta}_0 + \bar{\beta}_2 \hat{x}_t^2) \right\}}{a_{\text{N}}^{-1} \sigma^2_{v_t}}
\]

(4.38)

where \( d_t = \bar{\beta}_1 + 2\hat{x}_t \hat{\beta}_2 \) and \( \hat{\sigma}^2_{v_t} = \sigma^2_{e} - 2d_t \sigma_{eu} + d_t^2 \sigma^2_{u} \).

From Lemma 4.7 and our assumptions regarding \( \bar{\beta} \) we can write
\[ d_t - \gamma_t = [(\bar{\beta}_1 + 2\hat{x}_t \beta_2) - (\beta_1 + 2\hat{x}_t \beta_2)] + \]
\[ = 2(\hat{x}_t - x_t)\beta_2 + O_p(N^{-1/2}) \]
\[ = O_p(a_N^{1/2}). \quad (4.39) \]

Note then that \( (d_t - \gamma_t)^2 = O_p(a_N) = o_p(N^{-1/2}). \)

By Equation 3.39, \( \sigma^2_{\nu_t} \) may be expressed as

\[ a_{-1}^{-1} \sigma_{\nu_t}^2 = \sigma_e^2 - 2d_t \sigma_{eu} + d_t^2 a_{-1}^{-1} \]
\[ = [(\sigma_e^2 - 2 \gamma_t \sigma_{eu} + \gamma_t^2 \sigma_{eu}) - 2(d_t - \gamma_t)\sigma_{eu} \]
\[ + 2 \gamma_t (d_t - \gamma_t)\sigma_{eu}^2 + (d_t - \gamma_t)^2 \sigma_{eu}^2] a_{-1}^{-1} \]
\[ = a_{-1}^{-1} \sigma_{\nu_t}^2 + 2(d_t - \gamma_t)(\gamma_t \sigma_{eu} - \sigma_{eu}) a_{-1}^{-1} + O_p(N^{-1/2}) \]

where \( \sigma_{\nu_t}^2 = \sigma_e^2 - 2 \gamma_t \sigma_{eu} + \gamma_t^2 \sigma_{eu} \). Then, by Theorem 2.11,

\( (a_{-1}^{-1} \sigma_{\nu_t}^2)^{-1} \) may be expanded in a Taylor series about \( (a_{-1}^{-1} \sigma_{\nu_t}^2)^{-1} \):
Substituting Equation 4.39 for \( d_t - \gamma_t \) finally yields

\[
(a_{N - 1} \sigma_v^2)^{-1} = \left[ a_{N - 1} \sigma_v^2 + 2(d_t - \gamma_t)(\gamma_t \sigma_u^2 - \sigma_{eu}) a_{N - 1} \right]^{-1} + o_p(N^{-1/2})
\]

\[
= (a_{N - 1} \sigma_v^2)^{-1} - 2(d_t - \gamma_t)(\gamma_t \sigma_u^2 - \sigma_{eu}) a_{N - 1} (a_{N - 1} \sigma_v^2)^{-2} + o_p(N^{-1/2}).
\]

Next, we consider \( Y_t - \hat{\beta}_0 + \hat{\beta}_2 \hat{x}_t^2 \). By Lemma 4.7 and by our assumptions we have

\[
Y_t - \hat{\beta}_0 + \hat{\beta}_2 \hat{x}_t^2 = (y_t + e_t) - [\beta_0 + (\hat{\beta}_0 - \beta_0)]
\]

\[
+ [\beta_2 + (\hat{\beta}_2 - \beta_2)] [x_t + (\hat{x}_t - x_t)]^2
\]

\[
= y_t + e_t - \beta_0 + \beta_2 x_t^2 + 2\beta_2 x_t (\hat{x}_t - x_t) + o_p(N^{-1/2})
\]

\[
= (\beta_0 + \beta_1 x_t + \beta_2 x_t^2) - \beta_0 + \beta_2 x_t^2
\]

\[
+ e_t + 2\beta_2 x_t (\hat{x}_t - x_t) + o_p(N^{-1/2})
\]

\[
= x_t \gamma_t + 2\beta_2 x_t (\hat{x}_t - x_t) + e_t + o_p(N^{-1/2}).
\]
From Equations 4.41 and 4.39 the numerator of Equation 4.38 may be expressed as

\[ a_{-1}^N \left[ x_t (\sigma^2_{e} - \sigma_{eu} d_t) + (\sigma^2_u d_t - \sigma_{eu}^2) (Y_t - \beta_0 + \beta_2 \hat{x}_t) \right] = \]

\[ (x_t + u_t) \{ \sigma^2_{e} - \sigma_{eu} [\gamma_t + (d_t - \gamma_t)]\} a_{-1}^N + \]

\[ a_{-1}^N \{ \sigma^2_{u} [\gamma_t + (d_t - \gamma_t)] - \sigma_{eu} \} \cdot [x_t \gamma_t + 2\beta_2 x_t (\hat{x}_t - x_t) + e_t + O_p (N^{-1/2})] \]

\[ = (x_t + u_t) \{ \sigma^2_{e} - \sigma_{eu} [\gamma_t + 2\beta_2 (\hat{x}_t - x_t)] \} a_{-1}^N + \]

\[ a_{-1}^N \{ \sigma^2_{u} [\gamma_t + 2\beta_2 (\hat{x}_t - x_t)] - \sigma_{eu} \} \cdot [x_t \gamma_t + 2\beta_2 x_t (\hat{x}_t - x_t) + e_t + O_p (N^{-1/2})] \]

Finally, by Equations, 4.42, 4.40, and 4.38 we can express \( \hat{x}_t \) as

\[ \hat{x}_t = \{ x_t (a_{-1}^N \sigma^2_{v_t}) + u_t (\sigma^2_{e} - \gamma_t \sigma_{eu}) a_{-1}^N + e_t (\gamma_t \sigma^2_{u} - \sigma_{eu}) a_{-1}^N \]

\[ + [2\beta_2 (\hat{x}_t - x_t)] 2x_t (\gamma_t \sigma^2_{u} - \sigma_{eu}) a_{-1}^N + O_p (N^{-1/2}) \} \]
\[
\begin{align*}
\{ (a_{-1,2}^N)^{-1} - 4\beta_2 (\hat{x}_t - x_t)(\gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} (a_{-1,2}^N)^{-2} + O_p(N^{-1/2}) & \\
= x_t + (a_{-1,2}^N)^{-1} \{ u_t (\sigma_t^2 - \gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} + e_t (\gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} & \\
+ 4\beta_2 x_t (\hat{x}_t - x_t)(\gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} \} & \\
- x_t (a_{-1,2}^N)^{-1} [4\beta_2 (\hat{x}_t - x_t)(\gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} (a_{-1,2}^N)^{-2}] + O_p(N^{-1/2}) & \\
= x_t + (a_{-1,2}^N)^{-1} \{ u_t (\sigma_t^2 - \gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} + e_t (\gamma_t^2 - \sigma_{eu}) a_{-1}^{-1} & \\
+ O_p(N^{-1/2}). & 
\end{align*}
\]

The conclusion of the lemma follows. Q.E.D.

To establish the next lemma we need to recall some previous notation. In Chapter 3 \(v_t\) was defined by

\[
v_t = Y_t - \beta_0 - \beta_1 x_t - \beta_2 (x_t^2 - u_t^2)
\]

\[
= e_t - \beta_1 u_t - \beta_2 [(2x_t u_t + (u_t^2 - \sigma_u^2)]
\]

\[
= e_t - \gamma u_t - \beta_2 (u_t^2 - \sigma_u^2).
\]

Letting \(v_t^* = v_t + \beta_2 (u_t^2 - \sigma_u^2)\)

\[
(4.43)
\]

we state and prove the next lemma. In effect, this lemma establishes
that $x_t$ and the "residual" error $v_t^*$ are "nearly" uncorrelated.

Lemma 4.9

Under Model 4.1a, 4.1b and Assumption 4.1a,

$$\text{Cov}(v_t^*, \delta_t) = 0.$$ 

Proof:

By the definition of $v_t^*$ and $\delta_t$ we have

$$\text{Cov}(v_t^*, \delta_t) = (a N \sigma_v^2)^{-1} \text{Cov}(e_t - \gamma u_t, u_t(\sigma_e^2 - \gamma_t \sigma_{eu}) + e_t(\gamma_t \sigma_u^2 - \sigma_{eu}))$$

$$= (a N \sigma_v^2)^{-1} \{ \sigma_{eu}(\sigma_e^2 - \gamma_t \sigma_{eu}) + \sigma_e^2(\gamma_t \sigma_u^2 - \sigma_{eu})$$

$$- \gamma_t \sigma_u^2(\sigma_e^2 - \gamma_t \sigma_{eu}) - \gamma_t \sigma_{eu}(\gamma_t \sigma_u^2 - \sigma_{eu}) \}$$

$$= 0. \quad \text{Q.E.D.}$$

Having established the properties of $\hat{x}_t$, we now turn our attention to the asymptotic properties of $\hat{\Delta}$. First, we establish the order of the error in $\hat{\Delta}$. 
Lemma 4.10

Given Model 4.1a, 4.1b, Assumptions 4.1a, 4.2, 4.4, 4.6, 4.7a, 4.8, 4.9 and 4.10, then

\[(\hat{\Delta}\beta) - (\Delta\beta) = o_p(N^{-1/2})\]

where \((\Delta\beta) = \beta - \bar{\beta}\).

Proof:

Define

\[
\bar{M}_{\hat{x}\hat{x}} = b^{-1} \sum_{t=1}^{N} (a_{-1})^{-2} \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \\ \hat{x}_t \end{bmatrix}
\]

Then \((\hat{\Delta}\beta)\) may be written as

\[
(\hat{\Delta}\beta) = (\bar{M}_{\hat{x}\hat{x}})^{-1} \left\{ b^{-1} \sum_{t=1}^{N} (a_{-1})^{-2} \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \\ \hat{x}_t \end{bmatrix} \right\} \quad (4.45)
\]

By definition

\[
(\hat{e}_t - d_t \hat{u}_t) = \left( Y_t - \beta_0 - \beta_1 \hat{x}_t - \beta_2 \hat{x}_t^2 \right)
\]
and \( Y_t - \beta_0 - \beta_1 X_t - \beta_2 (X_t^2 - \sigma_u^2) = v_t \). Thus \( (e_t - d_t u_t) = \)

\[
[\beta_0 + \beta_1 X_t + \beta_2 X_t^2 + (v_t - \beta_2 \sigma_u^2) - \beta_0 - \beta_1 \hat{x}_t - \beta_2 \hat{x}_t^2] \\
- (\bar{\beta}_1 + 2\bar{x}_t \bar{\beta}_2)(X_t - \hat{x}_t). \tag{4.47}
\]

If we expand \( (\beta_0 + \beta_1 X_t + \beta_2 X_t^2) \) in a Taylor series about the point \((\hat{x}_t, \beta)\), then

\[
\beta_0 + \beta_1 \hat{x}_t + \beta_2 \hat{x}_t^2 = \\
\beta_0 + \beta_1 \hat{x}_t + \beta_2 \hat{x}_t^2 + (\beta_1 + 2\hat{x}_t \beta_2)(X_t - \hat{x}_t) + o_p(N^{-1/2}). \tag{4.48}
\]

by Assumptions 4.6, 4.7a, and Lemma 4.7. Substituting Equation 4.48 into Equation 4.47 yields

\[
(e_t - d_t u_t) = \\
(\Delta \beta_0) + (\Delta \beta_1) \hat{x}_t + (\Delta \beta_2) \hat{x}_t^2 + (v_t - \beta_2 \sigma_u^2) + o_p(N^{-1/2}).
\]

But \( v_t - \beta_2 \sigma_u^2 = e_t - \gamma u_t - \beta_2 u_t^2 \)
since $u_t^2 = O_p(a_N) = o_p(N^{-1/2})$ by Assumptions 4.6, 4.7a. Hence

$$(e_t - d\hat{u}_t) = (\Delta\beta_0) + (\Delta\beta_1)t + (\Delta\beta_2)t^2 + v_t^* + o_p(N^{-1/2}) \quad (4.49)$$

From Equation 4.49 we obtain

$$b_{-1}^N \sum_{t=1}^N (a_{-12}^{-1})^{-1} \left( e_t - d\hat{u}_t \right) \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} = \begin{bmatrix} \hat{r}_{1,t} \\ \hat{r}_{2,t} \end{bmatrix}$$

$$\bar{M}_{XX}^* (\Delta\beta) + b_{-1}^N \sum_{t=1}^N (a_{-12}^{-1})^{-1} v_t^* \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} + o_p(N^{-1/2}) \quad (4.50)$$

Then combining Equations 4.50 and 4.45 yields

$$(\hat{\Delta}\beta) = (\bar{M}_{XX}^*)^{-1} \{ \bar{M}_{XX}^* (\Delta\beta) +$$
\[
\begin{align*}
\frac{1}{b} \sum_{t=1}^{bN} \left( a \frac{-1}{N} \hat{x}_t \right) - 1 \cdot \hat{v}_t \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{2}x_t \end{bmatrix} + o_p(N^{-1/2}) \\
= (\Delta \beta) + (M_{xx}^*)^{-1} \left\{ \frac{1}{b} \sum_{t=1}^{bN} \left( a \frac{-1}{N} \hat{x}_t \right) - 1 \cdot \hat{v}_t \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{2}x_t \end{bmatrix} + o_p(N^{-1/2}) \right\}
\end{align*}
\]

Hence
\[
(\Delta \beta) - (\Delta \beta) = (M_{xx}^*)^{-1} \left\{ \frac{1}{b} \sum_{t=1}^{bN} \left( a \frac{-1}{N} \hat{x}_t \right) - 1 \cdot \hat{v}_t \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{2}x_t \end{bmatrix} + o_p(N^{-1/2}) \right\}
\]

We now show that $\tilde{M}_{xx}^* = \tilde{m}_{xx}^* + O_p(N^{-1/2})$. From Equation 4.39

\[
d_t - \gamma_t = 2(\hat{x}_t - x_t)\beta_2 + O_p(N^{-1/2})
\]

Thus
\[
(\Delta \gamma_t) = d_t - \gamma_t = 2 \delta_t \beta_2 + O_p(N^{-1/2}).
\]
Substituting Equation 4.52 for \((\Delta \gamma_t)^2\) yields

\[
\sum_{t}^{-1/2} = \sum_{t}^{1/2} - 4 \delta_t \beta_2 \left( \sum_{t}^{1/2} \right)
\]

since \(\delta_t = O_p \left( \sum_{N}^{1/2} \right)\). Then by Theorem 2.11

\[
(\sum_{t}^{1/2})^{-1} = (\sum_{t}^{1/2})^{-1} - \left[ 4 \delta_t \gamma \beta_2 \left( \sum_{t}^{1/2} \right) - 4 \delta_t \beta_2 \left( \sum_{t}^{1/2} \right) \right] (\sum_{t}^{1/2})^{-2}
\]

\[
+ O_p (N^{-1/2}). \tag{4.53}
\]

Note that

\[
E(\delta_t) = 0 \tag{4.54}
\]

and that \(\sigma_{\delta}^2 = \text{Var}(\delta_t) = O(\sum_{N}^{1/2})\). Thus, from Equations 4.53, 4.54, Assumption 4.9, and Theorem 2.2
\[ b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-1} = b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-1} + O(N^{-1/2}) \]  

(4.55)

since

\[ \text{Var}\{b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-2} \delta_t \} = O(b^{-1}_N a_N^2) = O(N^{-1}) \]

and

\[ \text{Var}\{b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-2} x_t \delta_t \} = O(b^{-1}_N a_N^2) = O(N^{-1}). \]

Also, from Equations 4.53, 4.54, Assumption 4.9, Lemma 4.8, and Theorem 2.2

\[ b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-1} \tilde{x}_t = \]

\[ = b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-1} x_t - b^{-1}_N \sum_{t=1}^{b_N} [4 \delta_t \gamma_t \beta_t (a^{-1} \sigma_v^2 u) + 4 \delta_t \beta_t (a^{-1} \sigma_v^2 v)] \]

\[ + O_p(N^{-1/2})\{x_t + \delta_t + O_p(N^{-1/2})\} \]

\[ = b^{-1}_N \sum_{t=1}^{b_N} (a^{-1} \sigma_v^2)^{-1} x_t - b^{-1}_N \sum_{t=1}^{b_N} [4 \delta_t \gamma_t \beta_t (a^{-1} \sigma_v^2 u) + 4 \delta_t \beta_t (a^{-1} \sigma_v^2 v)]. \]

(4.56)
since \( \text{Var} \{ b N \sum_{t=1}^{b N} \left[ 4 \delta \gamma t \beta u (a^{-1} N \sigma v) - 4 \delta \nu t (a^{-1} N \sigma u) \right] x_t (a^{-1} N \nu v^2) \} \\ = O\left( b N \sigma N \right) = O(N^{-1}). \)

Similarly, for \( r = 2, 3, 4, \)

\[
\sum_{t=1}^{b N} (a^{-1} N \sigma^2 v_t)^{-1} x_t = \sum_{t=1}^{b N} \left\{ (a^{-1} N \sigma^2 v_t)^{-1} - \left[ 4 \delta \gamma t \beta u (a^{-1} N \sigma v) - 4 \delta \nu t (a^{-1} N \sigma u) \right] (a^{-1} N \sigma^2 v_t)^{-2} \right. \\
+ O\left( N^{-1/2} \right) \left\{ x_t + \delta_t + O\left( N^{-1/2} \right) \right\}^r \\
= b N \sum_{t=1}^{b N} (a^{-1} N \sigma^2 v_t)^{-1} x_t + O\left( N^{-1/2} \right). \tag{4.57}
\]

It then follows, by Equations 4.57, 4.56, and 4.55 that

\[
\overline{M}_{xx}^* = \overline{m}_{xx}^* + O\left( N^{-1/2} \right). \tag{4.58}
\]

Next, we consider \( b N \sum_{t=1}^{b N} (a^{-1} N \sigma^2 v_t)^{-1} x_t \begin{bmatrix} 1 \\ x_t \\ x^2_t \end{bmatrix} \) term by term.
By Equation 4.53 and by Lemma 4.8 we have the following:

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} (a_{N, \sigma_{\nu_t}}^{-1})^{-1} v_t = \]

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} \left\{ (a_{N, \sigma_{\nu_t}}^{-1})^{-1} - \left[ 4 \delta t \beta_2 (a_{N, \sigma_{u_t}}^{-1}) - 4 \delta t \beta_2 (a_{N, \sigma_{eu}}^{-1}) \right] (a_{N, \sigma_{\nu_t}}^{-1})^{-2} \right\} v_t^* + o_p(\sqrt{N}) \]  

(4.59)

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} (a_{N, \sigma_{\nu_t}}^{-1})^{-1} x_t = \]

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} \left\{ (a_{N, \sigma_{\nu_t}}^{-1})^{-1} - \left[ 4 \delta t \gamma_t (a_{N, \sigma_{u_t}}^{-1}) - 4 \delta t \beta_2 (a_{N, \sigma_{eu}}^{-1}) \right] (a_{N, \sigma_{\nu_t}}^{-1})^{-2} \right\} \]

\[ v_t^* (x_t + \delta_t) + o_p(\sqrt{N}) \]  

(4.60)

and

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} (a_{N, \sigma_{\nu_t}}^{-1})^{-2} v_t x_t = \]

\[ b^{-1} \frac{b_N}{N} \sum_{t=1}^{N} \left\{ (a_{N, \sigma_{\nu_t}}^{-1})^{-1} - \left[ 4 \delta t \gamma_t (a_{N, \sigma_{u_t}}^{-1}) - 4 \delta t \beta_2 (a_{N, \sigma_{eu}}^{-1}) \right] (a_{N, \sigma_{\nu_t}}^{-1})^{-2} \right\} \]

\[ v_t^* (x_t^2 + 2x_t \delta_t + \delta_t^2) + o_p(\sqrt{N}) \]  

(4.61)
But \( v_t^* = O_p(a_{N^{1/2}}) \) and \( \delta_t = O_p(a_{N^{1/2}}) \) imply that

\[
\begin{align*}
  v_t^* \delta_t &= O_p(a_{N^{1/2}}) = o_p(N^{-1/2}).
\end{align*}
\]

Thus it follows that

\[
\begin{align*}
  b N^{-1} \sum_{t=1}^{N} (a_{N\sigma_u^2})^{-1} v_t^* \begin{bmatrix}
    x_t \cr
    x_t^2
  \end{bmatrix} = \begin{bmatrix}
    1 \cr
    0
  \end{bmatrix}
\end{align*}
\]

\[\text{(4.62)}\]

If we substitute Equations 4.62 and 4.58 into Equation 4.51, then we have

\[
\begin{align*}
  (\Delta \beta) - (\Delta \beta) &= (m_{xx})^{-1} \left\{ b N^{-1} \sum_{t=1}^{N} (a_{N\sigma_u^2})^{-1} v_t^* \begin{bmatrix}
    x_t \cr
    x_t^2
  \end{bmatrix} \right\} + o_p(N^{-1/2}).
\end{align*}
\]

\[\text{(4.63)}\]

But \( E(v_t^*) = E(e_t - \gamma u_t) = 0 \)
and
\[ \text{Var}(v^*_t) = \sigma_e^2 - 2 \gamma_t \sigma_e u + \sigma_u^2 \]

\[ = \sigma_{v_t}^2 = O(a_N). \]

Thus, by Theorem 2.2,

\[ b^{-1} \frac{b}{N} \sum_{t=1}^{bN} \left( a^{-1} \sigma_{v_t}^2 \right)^{-1} v^*_t \begin{bmatrix} 1 \\ x_t \\ \frac{1}{2} x_t^2 \end{bmatrix} = O_p(N^{-1/2}) \]

(4.64)

and the conclusion of the lemma follows. Q.E.D.

We now state and prove the major theorem of this section. This theorem establishes the limiting distribution of \( \hat{\beta} \).

**Theorem 4.3**

Given Model 4.1a, 4.1b, Assumptions 4.1a, 4.2, 4.4, 4.5, 4.6, 4.7a, 4.8, 4.9, and 4.10, then

\[ \sqrt{N} \left[ \hat{\beta} - \beta \right] \xrightarrow{d} N(0, \lim_{N \to \infty} (\bar{m}^{*})^{-1}) \]

where \( \hat{\beta} = \bar{\beta} + (\Delta \beta) \).
Proof:

From Lemma 4.10 and Equation 4.63

$$\sqrt{N} \{ \hat{\beta} - \beta \} = \sqrt{N} \{ (\Delta \hat{\beta}) - (\Delta \beta) \} = \sqrt{N} (m_{xx}^{-} - 1) \left\{ b \sum_{t=1}^{N} \left( a^{-1} \sigma_{v}^{2} \right)^{-1} v_{t}^{*} \begin{pmatrix} 1 \\ x_{t} \\ x_{t}^{2} \end{pmatrix} \right\} + o_{p} (1).$$

Thus, by Theorem 2.6, the limiting distribution of $\sqrt{N} \{ \hat{\beta} - \beta \}$ is the same as the limiting distribution of

$$\sqrt{N} \{ m_{xx}^{-} - 1 \left\{ b \sum_{t=1}^{N} \left( a^{-1} \sigma_{v}^{2} \right)^{-1} v_{t}^{*} \begin{pmatrix} 1 \\ x_{t} \\ x_{t}^{2} \end{pmatrix} \right\} \}.$$

We now investigate the limiting distribution of

$$\sqrt{N} \lambda^{'} b^{-1} \sum_{t=1}^{N} \left( a^{-1} \sigma_{v}^{2} \right)^{-1} v_{t}^{*} \begin{pmatrix} 1 \\ x_{t} \\ x_{t}^{2} \end{pmatrix},$$

where $\lambda$ is an arbitrary, nonzero $(3 \times 1)$ vector. First, note that the
are mutually independent by Assumption 4.1a. Second, note that

\[ E\left( (a^{-1/2}_N v_t)^{-1} (a^{-1/2}_N v_t^*)^\prime \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \right) = 0 \]

for \( t = 1, \ldots, b_N \) since \( E(v_t^*) = 0 \), also by Assumption 4.1a. Third, note that

\[ E\left( \left( (a^{-1/2}_N v_t)^{-1} (a^{-1/2}_N v_t^*)^\prime \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \right)^2 \right) = \]

\[ \left( a^{-1}_N \sigma_v^2 \right)^{-2} \lambda^\prime \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \left[ 1 \begin{bmatrix} x_t \\ x_t^2 \end{bmatrix} \right] \lambda E(a^{-1}_N v_t^* v_t) \]

\[ = \left( a^{-1}_N \sigma_v^2 \right)^{-1} \lambda^\prime \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \left[ 1 \begin{bmatrix} x_t \\ x_t^2 \end{bmatrix} \right] \lambda < \infty \]
and

$$E\left\{ \left| \left( a_{N}^{-1} \right)^{2} \left( a_{N}^{-1/2} v_{t}^{*} \right)^{\lambda^t} \right|^{2+\delta} \right\} < \infty,$$

where $\delta > 0$, since the random variables $(e_t, u_t)$ have bounded $2+\delta$ moments.

From Assumptions 4.4, 4.9 and 4.5 we obtain

$$\lim_{N \to \infty} \frac{b_N}{\sum_{t=1}^{N} E\left\{ \left| \left( a_{N}^{-1} \right)^{2} \left( a_{N}^{-1/2} v_{t}^{*} \right)^{\lambda^t} \right|^{2+\delta} \right\}} = \frac{b_N}{\sum_{t=1}^{N} E\left\{ \left[ \left( a_{N}^{-1} \right)^{2} \left( a_{N}^{-1/2} v_{t}^{*} \right)^{\lambda^t} \right]^{2} \right\}}^{2+\delta}$$

$$\lim_{N \to \infty} \frac{b_N^{-1}}{\sum_{t=1}^{N} E\left\{ \left| \left( a_{N}^{-1} \right)^{2} \left( a_{N}^{-1/2} v_{t}^{*} \right)^{\lambda^t} \right|^{2+\delta} \right\}} = \frac{b_N^{-1}}{\sum_{t=1}^{N} E\left\{ \left[ \left( a_{N}^{-1} \right)^{2} \left( a_{N}^{-1/2} v_{t}^{*} \right)^{\lambda^t} \right]^{2} \right\}}^{2+\delta}$$
Thus, by the Liapounov central limit theorem (cf. Theorem 2.8) we must have

\[
\frac{b_N}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \left( a_N^{-1/2} \sigma_{v_t}^{-1/2} \right) \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \left( \lambda' \sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \right)^{-1/2} \left( \lambda' \begin{bmatrix} 1 \\ \overline{m}_{ww} \end{bmatrix} \right)^{1/2}
\]

\[
\frac{\sqrt{N \lambda'} \sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{ww}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{ww}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

Furthermore, by Assumption 4.9,

\[
\lim_{N \to \infty} \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]

\[
\frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}} = \frac{\overline{m}_{xx}}{\sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t}^{-2} \right)^{-1} \lambda' \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix}}
\]
exists, and thus, by Theorem 2.7,

$$\sqrt{N} \lambda' b^{-1} \sum_{t=1}^{b_N} (a^{-1}_N \sigma^2 v_t^{-1} v_t^* \begin{bmatrix} 1 \\ x_t \\ 2 \\ x_t^2 \end{bmatrix} \rightarrow^d N(0, \lim_{N \to \infty} \lambda' \frac{m^*}{\lambda}).$$

By the multivariate central limit theorem this gives

$$\sqrt{N} b^{-1} \sum_{t=1}^{b_N} (a^{-1}_N \sigma^2 v_t^{-1} v_t^* \begin{bmatrix} 1 \\ x_t \\ 2 \\ x_t^2 \end{bmatrix} \rightarrow^d N(0, \lim_{N \to \infty} \frac{m^*}{\lambda}).$$

The result then follows by noting that

$$\frac{m^*}{\lambda} \rightarrow^d N(0, \lim_{N \to \infty} \frac{m^*}{\lambda}).$$

Q.E.D.

4. A revised estimator

In Subheading 2 of this section an estimation procedure for the quadratic functional model was given, and in Subheading 3 the
asymptotic properties of the given estimators were developed under Assumption 4.7a, i.e. $\frac{a}{N} = o(N^{-1/2})$. In this section, we consider the estimators of Subheading 2, with the exception that a high order correction is made to the right hand side of the equation defining $(\Delta \beta)$. Also, we now examine the estimators under the less restrictive Assumption 4.7b, i.e. $\frac{a^{1.5}}{N} = o(N^{-1/2})$.

The estimation procedure to be considered is given as follows:

1. Compute a preliminary estimator of $\beta$, say $\tilde{\beta}$, satisfying
   \[ \tilde{\beta} - \beta = O_p(N^{-1/2}). \]

2. For each $t = 1, 2, \ldots, b_N$, compute $\hat{x}_t$ defined by Equation 4.28.

3. Compute $(\Delta \beta)^*$ defined by

   \[ \bar{M}_{xx}^*(\Delta \beta)^* = \{b^{-1} \sum_{t=1}^{b_N} (a^{-1} \sigma_t^2)^{-1} \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} \begin{bmatrix} e_t - d \hat{u}_t - \tilde{\beta}_2 (\hat{u}_t - \hat{\sigma}_u^2) \\ \hat{u}_t - \hat{\sigma}_u^2 \end{bmatrix} \} \]

   where \[ e_t = d \hat{u}_t - \tilde{\beta} (\hat{u}_t - \hat{\sigma}_u^2) = Y_t - \tilde{\beta}_0 - \tilde{\beta}_1 X_t - \tilde{\beta}_2 (X_t - \hat{\sigma}_u^2) = \hat{v}_t. \quad (4.65) \]

4. Compute $\beta^*$ defined by
   \[ \beta^* = \tilde{\beta} + \nu^*(\Delta \beta)^* \quad (4.66) \]

   where $\nu^*$ is that $\nu \in [0, 1]$ which minimizes
\[ \Omega(\bar{\beta} + v(\Delta\beta)^*, \hat{x}_1, \ldots, \hat{x}_b_N) \]

with respect to \( v \).

5. Return to step one and iterate if desired.

As noted above, \((\Delta\beta)^*\) is identical to \((\hat{\Delta}\beta)\) with the exception that \((e_t - d_tu_t)\) is now replaced by \(\hat{v}_t\).

We now give three lemmas which describe the asymptotic properties of \(\hat{x}_t\). In the proofs of these lemmas we will have occasion to express the order of certain remainders as \(O_p(\max[a_N, N^{-1/2}])\). For notational convenience these will all be abbreviated by \(O_p(\max)\).

**Lemma 4.11**

Under Assumptions 4.6, 4.7b, and 4.10,

\[ \hat{x}_t = x_t + O_p(a_N^{1/2}) \]

for each \( t = 1, \ldots, b_N \).

**Proof:**

Follows immediately from proof of Lemma 4.7. \( \Box \).

**Lemma 4.12**

Under Assumptions 4.6, 4.7b, and 4.10,
\[ \hat{x}_t = x_t + \delta_t + O_p(\text{max}) \]

for \( t = 1, \ldots, b_N \), where

\[
\delta_t = (a_{-1}^T \sigma_{N}^{-2}) \left[ u_t (\sigma_t^2 - \gamma_t \sigma_t) a_{-1}^{-1} + e_t (\gamma_t \sigma_t^2 - \sigma_{eu}) a_{-1}^{-1} \right]
\]

and

\[
\gamma_t = \beta_1 + 2 x_t \beta_2.
\]

Proof:

We give an abbreviated proof which follows the proof of Lemma 4.8.

By Equation 4.52 we have

\[
a_{-1}^{-2} = \left[ (\sigma_t^2 - 2 \gamma_t \sigma_{eu} + \gamma_t^2 \sigma_{eu}) - 2(d_t - \gamma_t) \sigma_{eu} + 2 \gamma_t (d_t - \gamma_t) \sigma_{eu}^2 + (d_t - \gamma_t)^2 \sigma_{eu} \right] a_{-1}^{-1}
\]

\[
= a_{-1}^{-2} + 2(d_t - \gamma_t)^2 (\gamma_t \sigma_{eu}^2 - \sigma_{eu}) a_{-1}^{-1} + O_p(a_N)
\]

and then from Theorem 2.11 we obtain

\[
(a_{-1}^T \sigma_{N}^{-2})^{-1} = (a_{-1}^T \sigma_{N}^{-2})^{-1} - 4 \beta_2 (\hat{x}_t - x_t)(\gamma_t \sigma_t^2 - \sigma_{eu}) a_{-1}^{-1} (a_{-1}^T \sigma_{N}^{-2})^{-2}
\]

\[
+ O_p(a_N)
\]

(4.67)
From Lemma 4.7 we see

\[ Y_t - \beta_0 + \beta_2 x_t^2 = y_t + e_t - \beta_0 + \beta_2 x_t^2 + 2 \beta_2 x_t(\hat{x}_t - x_t) + O_p(\text{max}) \]

\[ = x_t y_t + 2 \beta_2 x_t(\hat{x}_t - x_t) + e_t + O_p(\text{max}) \]  \hspace{1cm} (4.68)

Then, by Equations 4.68 and 4.39, the numerator of Equation 4.38 may be expressed as

\[ a^{-1}_N [\frac{x_t (\sigma^2_e - \sigma_{eu} t) + (\sigma^2 u - \sigma_{eu}) (Y_t - \beta_0 + \beta_2 x_t^2)}{N}] \]

\[ = x_t (a^{-1}_N \sigma_v^2) + u_t (\sigma^2 - \gamma_t \sigma_{eu}) a^{-1}_N + e_t (\gamma_t \sigma^2 - \sigma_{eu}) a^{-1}_N \]

\[ + [2 \beta_2 (\hat{x}_t - x_t)] 2x_t (\gamma_{t} \sigma^2 - \sigma_{eu}) a^{-1}_N + O_p(\text{max}) \] \hspace{1cm} (4.69)

Finally, from Equations 4.38, 4.69, and 4.67 it follows that

\[ \hat{x}_t = x_t + (a^{-1}_N \sigma_v^{-1}) [u_t (\sigma^2 - \gamma_t \sigma_{eu}) a^{-1}_N + e_t (\gamma_t \sigma^2 - \sigma_{eu}) a^{-1}_N] \]

\[ + O_p(\text{max}) \] \hspace{1cm} Q.E.D.

We now restate Lemma 4.9 due to its importance for the remaining results of this section.
Lemma 4.9

Under Model 4.1a, 4.1b, and Assumption 4.1,

\[ \text{Cov}(v_t^*, \delta_t) = 0 , \]

where \( v_t^* = e_t - \gamma_t u_t' . \)

Next, we consider the asymptotic properties of \((\Delta \beta)^* \) under the Assumption 4.7b. First, we establish the order of the error in \((\Delta \beta)^* \).

Lemma 4.13

Given Model 4.1a, 4.1b, Assumptions 4.1b, 4.2, 4.4, 4.6, 4.7b, and 4.8-4.10,

\[ (\Delta \beta)^* - (\Delta \beta) = O_p(N^{-1/2}) \]

where \( \Delta \beta = \beta - \bar{\beta} . \)

Proof:

If we expand \((\beta_0 + \beta_1 x_t + \beta_2 x_t^2) \) in a Taylor series about the point \((\hat{x}_t, \beta) \) we obtain

\[ \beta_0 + \beta_1 \hat{x}_t + \beta_2 x_t^2 = \beta_0 + \beta_1 \hat{x}_t + \beta_2 \hat{x}_t^2 \]

\[ + (\beta_1 + 2\hat{x}_t \beta_2)(x_t - \hat{x}_t) + \beta_2(x_t - \hat{x}_t)^2 . \] (4.70)
Then from Equations 4.70 and 4.47 we have

\[
\hat{v}_t = (\hat{e}_t - d_t \hat{u}_t) - \beta_2 (u_t^2 - \sigma_u^2) = \\
[(\Delta \beta_0) + (\Delta \beta_1) \hat{x}_t + (\Delta \beta_2) \hat{x}_t^2 + \nu_t] + (\Delta \beta_2)[\hat{u}_t^2 - \sigma_u^2] \\
= (\Delta \beta_0) + (\Delta \beta_1) \hat{x}_t + (\Delta \beta_2) \hat{x}_t^2 + \nu_t + o_p(N^{-1/2}). \tag{4.71}
\]

where \( \nu_t = e_t - \gamma_t u_t - \beta_2 (u_t^2 - \sigma_u^2) \).

Substituting Equation 4.71 into Equation 4.66 yields

\[
(\Delta \beta)^* = (\bar{M}_{xx})^{-1} \left[ \bar{M}_{xx}(\Delta \beta) + b \sum_{i=1}^{b} \left( a_i \sigma_v \right) \left( \begin{array}{c}
1 \\
\bar{x}_t \\
\bar{x}_t^2
\end{array} \right) \hat{v}_t + o_p(N^{-1/2}) \right]
\]

and thus

\[
(\Delta \beta)^* - (\Delta \beta) = (\bar{M}_{xx})^{-1} \left[ b \sum_{i=1}^{b} \left( a_i \sigma_v \right) \left( \begin{array}{c}
1 \\
\bar{x}_t \\
\bar{x}_t^2
\end{array} \right) \hat{v}_t + o_p(N^{-1/2}) \right]. \tag{4.72}
\]

From Lemma 4.12 and the assumption \( \bar{\beta} - \beta = O_p(N^{-1/2}) \) we have that
\[ \Delta \gamma_t = d_t - \gamma_t = 2(\hat{x}_t - x_t)\beta_2 + \mathcal{O}_p(N^{-1/2}) \]
\[ = 2 \delta_t \beta_2 + \mathcal{O}_p(\text{max}) = \mathcal{O}_p(a_{1/2}^N). \] (4.73)

Thus
\[ a_{-1}^2 \sigma_{\nu_t} = a_{-1}^2 \sigma_{\nu_t}^2 - 2(\Delta \gamma_t)(a_{-1}^2 N \sigma_{e u}^2) + \]
\[ [2 \gamma_t(\Delta \gamma_t) + (\Delta \gamma_t)^2](a_{-1}^2 N \sigma_{u}^2) \]
\[ = a_{-1}^2 \sigma_{\nu_t}^2 - 4 \delta_t \beta_2(a_{-1}^2 N \sigma_{e u}^2) \]
\[ + 4 \delta_t \gamma_t \beta_2(a_{-1}^2 N \sigma_{u}^2) + \mathcal{O}_p(\text{max}) \]

and by Theorem 2.11

\[ (a_{-1}^2 \sigma_{\nu_t}^2)^{-1} = (a_{-1}^2 \sigma_{\nu_t}^2)^{-1} - [4 \delta_t \gamma_t \beta_2(a_{-1}^2 N \sigma_{u}^2)]^{-1} \]
\[ = 4 \delta_t \beta_2(a_{-1}^2 N \sigma_{e u}^2)(a_{-1}^2 N \sigma_{u}^2)^{-2} + \mathcal{O}_p(\text{max}) \] (4.74)

where \( \sigma_{\nu_t}^2 = \sigma_{e}^2 - 2 \gamma_t \sigma_{e u} + \gamma_t^2 \sigma_{u}^2 \).

Now by Equation 4.74, Lemma 4.12, and Theorem 2.2 it follows that
\[
\overline{M}_{\text{xx}} = \overline{m}_{\text{xx}} + O(p \text{ max}).
\] (4.75)

This is easily seen by following the proof of Lemma 4.10.

Next, we consider

\[
b^{-1} \sum_{t=1}^{bN} (a \sigma^2_v t) v_t \begin{bmatrix}
1 \\
\hat{x}_t \\
\hat{x}_t^2
\end{bmatrix}.
\]

By Equation 4.74 and Lemma 4.12 we can write

\[
b^{-1} \sum_{t=1}^{bN} (a \sigma^2_v t) v_t \begin{bmatrix}
1 \\
\hat{x}_t \\
\hat{x}_t^2
\end{bmatrix} = \begin{bmatrix}
1 \\
\hat{x}_t \\
\hat{x}_t^2
\end{bmatrix}
\]

\[
b^{-1} \sum_{t=1}^{bN} [(a \sigma^2_v t)^{-1} - 4 \delta \gamma v_t \beta \sigma_t^2 - 4 \delta \beta \sigma_t^2 (a \sigma^2_v t)^{-2}] v_t \cdot \begin{bmatrix}
1 \\
\hat{x}_t + \delta_t \\
\hat{x}_t^2 + 2x_t \delta_t
\end{bmatrix} + o_p(N^{-1/2})
\]

(4.76)

since \( v_t = O_p(a_N^{1/2}) \). Note that \( E(v_t) = 0 \), \( E(\delta_t) = 0 \), and
\[ \text{Cov}(v_t, \delta_t) = \text{Cov}[\beta^*_2(u_t^2 - \sigma^2_u), \delta_t] = O(a^{-1.5}_N) \quad (4.77) \]

by Assumptions 4.1b, 4.6 and Lemma 4.9. Consequently, from Equations 4.76, 4.77, Assumption 4.9, and Theorem 2.2, we obtain

\[ b^{-1} N \sum_{t=1}^{bN} (a^{-1} \sigma^2_v \sigma_t^2)^{-1} v_t \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} 
= b^{-1} N \sum_{t=1}^{bN} (a^{-1} \sigma^2_v \sigma_t^2)^{-1} v_t \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} + o_p(N^{-1/2}) \]

\[ = O_p(N^{-1/2}) \quad (4.78) \]

since

\[ \text{Var}\{b^{-1} N \sum_{t=1}^{bN} (a^{-1} \sigma^2_v \sigma_t^2)^{-1} v_t \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \} = O(b^{-1} a_{NN}) = O(N^{-1}). \]
The conclusion of the lemma follows from Equations 4.78, 4.75, and 4.72. Q.E.D.

To conclude this section, we state the following theorem which establishes the limiting distribution of \( \beta^* \). The proof follows the proof of Theorem 4.3 and is thus deleted.

**Theorem 4.4**

Given Model 4.1a, 4.1b, Assumptions 4.1b, 4.2, 4.4, 4.5, 4.6, 4.7b, and 4.8-4.10, then

\[
\sqrt{N} \{ \beta^* - \beta \} \xrightarrow{d} N(0, \lim_{N \to \infty} (m^*_{xx})^{-1})
\]

where \( \beta^* = \overline{\beta} + (\Delta \beta)^* \).

5. **Estimating the variance of the estimators**

   We now consider estimating the variance of the estimators \( \hat{\beta} \) and \( \beta^* \). For this purpose, we rely on the asymptotic properties of \( \hat{\beta} \) and \( \beta^* \) developed in Theorems 4.3 and 4.4 of this chapter. To be specific, we shall estimate the matrix \((m^*_{xx})^{-1}\), the asymptotic covariance matrix of \( \hat{\beta} \) and \( \beta^* \).

**Lemma 4.14**

Under Assumptions 4.1a, 4.6, 4.4, 4.7a, 4.8, 4.9, and 4.10, then a consistent estimator of \((m^*_{xx})^{-1}\) is given by \((\overline{M}^*_{xx})^{-1}\).
Proof:

By Equation 4.58 we have

$$\overline{M} = \overline{m} + O_p(N^{-1/2}) .$$

The result then follows from Theorem 2.11 by noting that

$$(\overline{M})^{-1} = (\overline{m})^{-1} + O_p(N^{-1/2}). \quad \text{Q.E.D.}$$

Lemma 4.15

Under Assumptions 4.1b, 4.6, 4.4, 4.7b, 4.8, 4.9, and 4.10, then a consistent estimator of $(\overline{m})^{-1}$ is given by $(\overline{M})^{-1}$.

Proof:

From (4.4.51) and Theorem 2.11

$$(\overline{M})^{-1} = (\overline{m})^{-1} + O_p(\max). \quad \text{Q.E.D.}$$

6. Preliminary estimator versus revised estimator

Having considered the estimators $\hat{\beta}$, $\tilde{\beta}$, and $\beta^*$, we now consider the question of whether, in some sense, it is worth computing the two-step estimator $\hat{\beta}$ or $\beta^*$, as opposed to merely computing a preliminary estimator such as $\tilde{\beta}$. The error in each of the estimators is
However, in terms of asymptotic variance, we can establish that \( \hat{\beta} \) or \( \beta^* \) are superior to \( \hat{\beta} \).

**Lemma 4.16**

The asymptotic variance of \( \hat{\beta} \) is greater than the asymptotic variance of either \( \hat{\beta} = \bar{\beta} + (\Delta \hat{\beta}) \) or \( \beta^* = \bar{\beta} + (\Delta \beta)^* \), provided the conditions under which each possesses a limiting distribution are met.

**Proof:**

It is sufficient to show that each of the diagonal elements of \( (\overline{m}_{\infty})^{-1} \) is less than or equal to the corresponding diagonal element of

\[
(\overline{m}_{\infty})^{-1} \left\{ b_N^{-1} \sum_{t=1}^{b_N} \left( a_N^{-1} \sigma_{v_t} \right) \begin{bmatrix} 1 \\ x_t^1 \\ x_t^2 \\ \vdots \\ x_t^b_N \end{bmatrix} \right\} (\overline{m}_{\infty})^{-1}.
\]

To see this we define the matrices \( V \) and \( w \) as follows:

\[
V = \text{diag} \left( a_N^{-1} \sigma_{v_1}^2, \ldots, a_N^{-1} \sigma_{v_b_N}^2 \right);
\]

\[
w = \begin{bmatrix}
1 & x_1^1 & x_1^2 \\
1 & x_2^1 & x_2^2 \\
& \vdots & \vdots \\
1 & x_{b_N}^1 & x_{b_N}^2
\end{bmatrix}.
\]
Now we obtain

\[(\overline{m}_{xx})^{-1} = (b \cdot w'w)^{-1}\]  

(4.79)

and

\[(\overline{m}_{ww})^{-1} \{b \cdot \sum_{t=1}^{N} \left( a^{-1} \sigma_{t}^{2} \right) \left[ \begin{array}{c} 1 \\ x_{t} \\ x_{t}^{2} \end{array} \right] \} (\overline{m}_{ww})^{-1} = (b \cdot w'w)^{-1} \{b \cdot w'w \} (b \cdot w'w)^{-1}. \]  

(4.80)

Note that Equation 4.79 is analogous to the covariance matrix of a generalized least squares estimator and Equation 4.80 is analogous to the covariance matrix of a simple least squares estimator. The conclusion thus follows by the Gauss Markov theorem. Q.E.D.

7. **Pseudo estimators vis-à-vis maximum likelihood estimator**

To conclude Chapter 4, we shall attempt to consolidate the results of Section D with our knowledge of the likelihood function and the maximum likelihood estimator. We begin by remarking that the classical results on the consistency and asymptotic normality of the maximum likelihood estimator (cf. Kendall and Stuart (1961)) do not hold for the present situation. The easiest way to see this is to note that here
the number of unknown parameters increases with the sample size. That is, as N increases, the number of unknown $x_t$'s increases. As a result, the regularity conditions of the classical theorems are not met. This fact is well documented by Neyman and Scott (1948) and others.

In view of this observation, a direct proof of consistency or asymptotic normality would be required for the quadratic functional relationship. However, such a proof would be very difficult indeed. We have considered likelihood equations 4.25, 4.26 at great length, yet the asymptotic properties of $\tilde{\beta}$ remain elusive. Nevertheless, we do have knowledge of the properties of several estimators which are intimately related to the maximum likelihood estimator.

In Chapter 1 we presented the work of Villegas (1969). Recall that he gave an iterative estimator, which depended upon preliminary estimators, and established the asymptotic properties of the estimator assuming the error variances decrease at the rate $1/N$. If the maximum likelihood estimator exists, and if the preliminary estimators are sufficiently close to their respective true values, and if the Villegas procedure converges, then we would expect this estimator to iterate to the maximum of the likelihood. Under these conditions the asymptotic properties of the maximum likelihood estimator would be given by the properties of the Villegas estimator. To attempt to prove this we would argue along the lines of similar proofs given for the Gauss-Newton
method, where the independent variables are measured without error. However, we emphasize that such a proof has not been given.

In Section D of the present chapter, we presented the iterative estimators \((\hat{\Delta}\beta)\) and \((\Delta\beta)^*\) which in turn defined the estimators \(\hat{\beta}\) and \(\beta^*\). The asymptotic properties of these estimators were established assuming the error variances decrease at the rate \(a_N\). Again, if the maximum likelihood estimator exists, if the preliminary estimator is sufficiently close to the true value, and if the iterative procedure converges, then both \(\hat{\beta}\) and \(\beta^*\) might converge to a local maximum of the likelihood. Under these conditions, the asymptotic properties of the maximum likelihood estimator would be given by the properties of \(\hat{\beta} = \beta + (\hat{\Delta}\beta)\) and \(\beta^* = \beta + (\Delta\beta)^*\). As before, we emphasize that this is an unproven conjecture.

Happily, some things remain unchanged. Under classical assumptions the inverse of the information matrix is the asymptotic covariance matrix of the maximum likelihood estimators. Dolby and Lipton (1972) have taken the expected value of the twice differentiated likelihood for the nonlinear functional relationship with normally distributed measurement errors. For the quadratic functional relationship, as defined in this chapter, this becomes
Thus, the inverse of the information matrix is the asymptotic co-
variance matrix of both \( \hat{\beta} = \bar{\beta} + (\hat{\Delta} \beta) \) and \( \beta^* = \bar{\beta} + (\Delta \beta)^* \) (cf. Theorems 4.3 and 4.4). Similar results hold for the Villegas situation. Note that in these three situations the error variance of \( X_t \) is decreasing as \( N \) increases.
V. ESTIMATORS FOR THE GENERAL NONLINEAR FUNCTIONAL RELATIONSHIP

A. Introduction

Chapters 2, 3, and 4 were concerned with a specific nonlinear errors-in-variables model: the quadratic functional relationship. We now extend the methods of those chapters to the general nonlinear functional relationship.

As in Chapter 4, we let \( \{ b_N \}_{N=1}^{\infty} \) and \( \{ a_N \}_{N=1}^{\infty} \) be sequences of positive real numbers such that \( N = b_N / a_N \) for \( N = 1, 2, \ldots \). Also, we suppose the existence of a sequence of experiments indexed by \( N \), and we let \( b_N \) denote the number of observations in the \( N \)th experiment.

The nonlinear errors-in-variables model to be considered is specified by the exact mathematical relationship

\[
y_{Nt} = g(x_{Nt}; \beta_0)
\]

(5.1)

where, in the \( N \)th experiment, we observe

\[
y_{Nt} = y_{Nt} + e_{Nt}
\]

\[
x_{Nt} = x_{Nt} + u_{Nt}
\]

(5.2)

for \( t = 1, \ldots, b_N \). \( x_{Nt}, u_{Nt}, \) and \( x_{Nt} \) are each \((1 \times q)\) vectors; \( y_{Nt}, \)
\( e_{Nt}, \) and \( Y_{Nt} \) are each scalars; and \( \beta_0 \) is a \((p \times 1)\) vector. The vector
\( \beta_0 \) is unknown and is to be estimated.

We suppose that \( \{x_{nt}^{bN} \}_{t=1}^{bN} \) is a sequence of fixed constants, that \( x_{nt} \in \mathbb{R}^q \) for each \( t \), and that \( \beta_0 \in \mathbb{R}^p \). Also, we suppose \( e_{nt} \) and \( u_{nt} \) represent measurement errors committed in attempting to observe the true values \( y_{nt} \) and \( x_{nt} \).

To construct estimators of \( \beta_0 \), it will be necessary to impose certain assumptions.

**Assumption 5.1**

\( g(x; \beta) \) is a continuous function from \( \mathbb{R}^q \times \mathbb{R}^p \) into \( \mathbb{R}^1 \).

**Assumption 5.2a**

The partial derivative of \( g(x; \beta) \) with respect to each argument exists and is continuous.

**Assumption 5.2b**

The second partial derivative of \( g(x; \beta) \) with respect to each pair of arguments exists and is continuous.

**Assumption 5.2c**

The third partial derivative of \( g(x; \beta) \) with respect to each trio of arguments exists and is continuous.

**Assumption 5.3**

A preliminary estimator of \( \beta_0 \), say \( \bar{\beta} \), is available and satisfies \( \bar{\beta} - \beta_0 = O_p(N^{-1/2}) \).
Assumption 5.4

The errors of measurement, \( \{ (e_{Nt}, u_{Nt}) \}_{t=1}^{b_N} \) are independently distributed random variables and possess the following moment properties:

\[
E(e_{Nt}) = 0 , \\
E(u_{Nt}) = \varphi ,
\]

and

\[
E \begin{pmatrix} e_{Nt} \\ u_{Nt} \end{pmatrix} e_{Ns} u_{Ns} = \begin{pmatrix} \mathbb{P} \\ \varphi \end{pmatrix} = \begin{pmatrix} \sigma^2 & \mathbb{P}_{eu} \\ \mathbb{P}_{ue} & \mathbb{P}_{uu} \end{pmatrix} \quad \text{for } t = s \\
\varphi \quad \text{for } t \neq s ,
\]

where \( \varphi \) denotes an appropriately dimensioned matrix of zeros.

Assumption 5.5

The error covariance matrix, \( \mathbb{P} \), is known and positive definite.\(^2\)

To shorten the notation, we suppress the \( N \) subscript.

---

\(^2\) This is an assumption of convenience. For example, if the errors are normally distributed this allows us to immediately write the likelihood. However, the methods to be given also apply to singular covariance matrices. If \( \mathbb{P} \) is singular, we first transform the data and then write the likelihood in lower dimension.
B. The Estimator

Following the approach of Chapter 4, an iterative estimator of $\beta_0$ will be constructed. Our strategy will be to define a preliminary estimator of $x_t$, say $\bar{x}_t$, and then to define a second-round estimator of $\beta_0$ in terms of $\bar{\beta}$ and $\bar{x}_t$. First, however, it will be convenient to establish some additional notation.

Let $g_{x}(x; \beta)$ denote the $(1 \times q)$ vector of first partial derivatives of the function $g$ with respect to the elements of $x$, evaluated at $(x; \beta)$. Let $g_{\beta}(x; \beta)$ denote the $(p \times 1)$ vector of first partial derivatives of $g$ with respect to the elements of $\beta$, evaluated at $(x; \beta)$. Let $g_{xx}(x; \beta)$ denote the $(q \times q)$ matrix of second mixed partial derivatives of $g$ with respect to the elements of $x$; evaluated at $(x; \beta)$. And let $g_{\beta x}(x; \beta)$ denote the $(p \times q)$ matrix of second mixed partial derivatives of $g$ with respect to $\beta$ and then $x$, evaluated at $(x; \beta)$.

In all cases, we shall use a prime to denote the transpose of the specified matrix.

If the errors of measurement, $(e_t, u_t)$, are normally distributed, then by Assumptions 5.4 and 5.5 the maximum likelihood estimators of $\beta_0$ and $x_t$, $t = 1, \ldots, b_N$, are those values of $\beta$ and $x_t$ which maximize the likelihood function.
\[ L(\beta, \xi_1, \xi_2, \ldots, \xi_{b_N}) = (\text{constant}) \cdot \exp\left\{ -\frac{1}{2} Q(\beta, \xi_1, \xi_2, \ldots, \xi_{b_N}) \right\}, \]

where

\[ Q(\beta, \xi_1, \xi_2, \ldots, \xi_{b_N}) = \sum_{t=1}^{b_N} \left[ Y_t - g(\xi_t; \beta), X_t - \xi_t \right]^{-1} \left[ Y_t - g(\xi_t; \beta), X_t - \xi_t \right]'. \]  

Since the logarithm function is monotonically increasing, \( L(\beta, \xi_1, \xi_2, \ldots, \xi_{b_N}) \) and \( \log L(\beta, \xi_1, \xi_2, \ldots, \xi_{b_N}) \) are maximized by the same value of \((\beta, \xi_1, \xi_2, \ldots, \xi_{b_N})\). Thus, as in Chapter 4, the maximum likelihood estimator is that value of \((\beta, \xi_1, \xi_2, \ldots, \xi_{b_N})\) which maximizes \( Q \) defined by Equation 5.3.

Unfortunately, an explicit expression for the maximum likelihood estimator will rarely be obtained. For example, an explicit expression was not obtained for the quadratic model. To circumvent this difficulty, we construct an iterative estimator, whose motivation rests with the likelihood function.

In this direction, we define a preliminary estimator of \( x_t \), say \( \bar{x}_t \), to be that value of \( \xi_t \) which minimizes
If we differentiate $Q'$ with respect to $\xi_t$, replace $\xi_t$ by $\bar{x}_t$, and set the result equal to zero, we obtain the following system of equations:

\[
\begin{align*}
\sigma_{ee} (Y_t - g(\bar{x}_t; \bar{\beta})) g_x(\bar{x}_t; \bar{\beta}) \\
+ [Y_t - g(\bar{x}_t; \bar{\beta})] \phi_{eu} \\
+ [X_t - \bar{x}_t] \phi_{ue} g_x(\bar{x}_t; \bar{\beta}) \\
+ [X_t - \bar{x}_t] \phi_{uu} = q
\end{align*}
\] (5.5)

where $\phi^{-1}$ is partitioned as

\[
\phi^{-1} = \begin{bmatrix}
\sigma_{ee} & \phi_{eu} \\
\phi_{ue} & \phi_{uu}
\end{bmatrix}
\]

The preliminary estimator $\bar{x}_t$ is obtained by solution of Equation 5.5. If they are multiple roots to this equation, then $\bar{x}_t$ is that root which minimizes $Q'$. Geometrically, from Equation 5.4 we see that $\bar{x}_t$ is defined to be the $x$ coordinate of the minimum weighted distance between $(Y_t, X_t)$ and the curve $y = g(x; \bar{\beta})$. Thus the root of Equation 5.5 provides a
global minimum of $Q'$. Moreover, by Assumption 5.1 that minimum
is unique with probability one.

For most functions $g$, solution of Equations 5.5 will be very diffi­
cult. Thus we define an approximate solution. Expanding Equation
5.5 in a Taylor series about the point $X_t$ yields

$$
\varphi = \sigma \epsilon [Y_t - g(X_t; \bar{\beta})] g_x(X_t; \bar{\beta}) + \left[ Y_t - g(X_t; \bar{\beta}) \right] g_{xx}(X_t; \bar{\beta})
$$

$$
+ (x_t - X_t) \left[ \sigma \epsilon [ -g'(X_t; \bar{\beta}) g_x(X_t; \bar{\beta}) + \left[ Y_t - g(X_t; \bar{\beta}) \right] g_{xx}(X_t; \bar{\beta}) \right]
$$

$$
- g_x'(X_t; \bar{\beta}) \frac{\epsilon}{\varphi} - \frac{\epsilon}{\varphi} g_x(X_t; \bar{\beta}) - \frac{\epsilon}{\varphi} uu \right] + R_t
$$

where $R_t$ is a remainder. Then, from Equation 5.6 we define $\hat{x}_t$ to be
the solution of

$$
\varphi = \sigma \epsilon [Y_t - g(X_t; \bar{\beta})] [g_x'(X_t; \bar{\beta}) + \frac{\epsilon}{\varphi}]
$$

$$
+ (x_t - X_t) \left[ \sigma \epsilon [ -g'(X_t; \bar{\beta}) g_x(X_t; \bar{\beta}) + \left[ Y_t - g(X_t; \bar{\beta}) \right] g_{xx}(X_t; \bar{\beta}) \right]
$$

$$
- g_x'(X_t; \bar{\beta}) \frac{\epsilon}{\varphi} - \frac{\epsilon}{\varphi} g_x(X_t; \bar{\beta}) - \frac{\epsilon}{\varphi} uu \right].
$$

In the next section, we will investigate the properties of $\bar{x}_t$; the
properties of $\hat{x}_t$; and the error committed in approximating $x_t$ by $\hat{x}_t$. 
But now, we proceed to define a second-round estimator of $\beta_0$.

We introduce the following notation:

\[
\begin{align*}
\bar{y}_t &= g(x_t; \beta), \\
\bar{e}_t &= y_t - \bar{y}_t, \\
\bar{u}_t &= x_t - \bar{x}_t, \\
\Delta x_t &= x_t - \bar{x}_t, \\
\Delta y_t &= y_t - \bar{y}_t, \\
\Delta \beta &= \beta_0 - \bar{\beta}, \\
d_t &= g(x_t; \bar{\beta}).
\end{align*}
\]

Then we define an estimator of $\Delta \beta$, say $\hat{\Delta \beta}$, to be that value of $\Delta \beta$ which minimizes the function

\[
L(\Delta \beta, \Delta y_1, \Delta x_1, \ldots, \Delta y_{b_N}, \Delta x_{b_N}, \lambda_1, \ldots, \lambda_{b_N})
\]

\[
= \sum_{t=1}^{b_N} (\bar{e}_t - \Delta y_t, \bar{u}_t - \Delta x_t) \frac{1}{\hat{\beta}} (\hat{e}_t - \Delta y_t, \hat{u}_t - \Delta x_t)' \\
= 2 \sum_{t=1}^{b_N} \lambda_t \left[ (\Delta y_t) - g'(x_t; \bar{\beta})(\Delta \beta) - d_t'(\Delta x_t) \right]
\] (5.8)
where the variables $\lambda_t$ are Lagrange multipliers.  

Necessary conditions for a stationary point of Equation 5.8 are

$$\frac{\partial L}{\partial (\Delta y_t, \Delta x_t)} = \varphi , \quad t = 1, \ldots, b_N,$$

$$\frac{\partial L}{\partial (\Delta \beta)} = \varphi ,$$

and

$$\frac{\partial L}{\partial \lambda_t} = 0 , \quad t = 1, \ldots, b_N,$$

where $\varphi$ is a null vector of appropriate dimension.

These give

$$(e_t - \Delta y_t, \bar{u}_t - \Delta x_t) \Phi^{-1} + \lambda_t (1, -d_t) = \varphi , \quad \text{for } t = 1, \ldots, b_N ;$$

$$\Sigma_{t=1}^{b_N} \lambda_t g_\beta (\bar{x}_t ; \bar{\beta}) = \varphi ; \quad \text{(5.10)}$$

---

\(^3\) Defining $\Delta \beta$ in this fashion is totally analogous to the manner in which the estimators of Chapter 4 were constructed.
and

\[(\Delta y_t) - g'_{\beta}(\bar{x}_t; \bar{\beta})(\Delta \beta) - d_t(\Delta x_t)' = 0, \quad (5.11)\]

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

Postmultiplying Equation 5.9 by \(\phi(1, -d_t)'\) yields

\[\binom{e_t - \Delta y_t}{u_t - \Delta x_t} \begin{pmatrix} 1 \\ -d_t' \end{pmatrix} + \lambda_t \sigma^2_{v_t} = 0 \quad (5.12)\]

where \(\sigma^2_{v_t} = (1, -d_t) \phi(1, -d_t)' \). If we rearrange the terms in Equation 5.12 we obtain

\[e_t - u_t d_t' + \lambda_t \sigma^2_{v_t} = \Delta y_t - \Delta x_t d_t' \quad (5.13)\]

From Equations 5.11 and 5.13 we have

\[g'_{\beta}(\bar{x}_t; \bar{\beta})(\Delta \beta) = \binom{e_t - u_t d_t'}{\Delta y_t - \Delta x_t d_t'} \quad (5.14)\]

And premultiplying Equation 5.14 by \(g_{\beta}(\bar{x}_t; \bar{\beta}) / \sigma^2_{v_t}\) gives
\[
\frac{g_\beta(x_t; \bar{\beta})g_\beta'(x_t; \bar{\beta})}{\sigma^2 v_t} (\Delta \beta) = \\
\frac{g_\beta(x_t; \bar{\beta})}{\sigma^2 v_t} \left[ e_t - \bar{u}_t d'_t + \lambda_t \frac{\sigma^2}{v_t} \right] \\
(5.15)
\]
for \( t = 1, \ldots, b_N \).

If we sum Equation 5.15 over \( t = 1, \ldots, b_N \), recall Equation 5.10, and replace \( \Delta \beta \) by \( \hat{\Delta} \beta \) then we obtain the following expression:

\[
\left[ \sum_{t=1}^{b_N} \frac{1}{\sigma^2 v_t} g_\beta(x_t; \bar{\beta})g_\beta'(x_t; \bar{\beta}) \right] (\hat{\Delta} \beta) = \\
\left[ \sum_{t=1}^{b_N} \frac{1}{\sigma^2 v_t} g_\beta(x_t; \bar{\beta})(e_t - \bar{u}_t d'_t) \right]. \\
(5.16)
\]

We have now minimized the function \( \mathcal{J} \) and have determined that \((\hat{\Delta} \beta)\), given by Equation 5.16, is the minimizing value of \((\Delta \beta)\). A second-round estimator of \( \beta_0 \), say \( \hat{\beta} \), may now be constructed by defining

\[
\hat{\beta} = \bar{\beta} + \hat{v}(\hat{\Delta} \beta) \\
(5.17)
\]
where \( \hat{v} \) is that \( v \in [0, 1] \) which minimizes

\[
Q(\hat{\beta} + v(\Delta \beta), \bar{x}_1, \ldots, \bar{x}_{N})
\]

with respect to \( v \).

In the next section we will investigate the asymptotic properties of the estimator \( \hat{\beta} \). But first, we summarize the estimation procedure suggested by this section:

1. Compute a preliminary estimator of \( \beta \), say \( \bar{\beta} \), satisfying
   \[
   \bar{\beta} - \beta = O_p(N^{-1/2}).
   \]
2. For each \( t = 1, \ldots, b_N \), compute \( \bar{x}_t \) defined by Equation 5.5.
3. Compute \( \Delta \hat{\beta} \) as defined by Equation 5.16.
4. Compute \( \hat{\beta} \) as defined by Equation 5.17.
5. Return to step one and iterate if desired.

C. The Asymptotic Properties

This section is concerned with the asymptotic properties of the estimators defined in Section B. We begin by giving some additional assumptions.
Assumption 5.6a

The random variables \((e_t, u_t)\) have bounded \(2+\delta\) moments, where \(\delta > 0\).

Assumption 5.6b

The random variables \((e_t, u_t)\) have bounded \(4+\delta\) moments, where \(\delta > 0\).

Assumption 5.7

The \((p \times p)\) matrix \(\mathbf{m}^*\) defined by

\[
\mathbf{m}^* = b^{-1} \sum_{N}^{b} \frac{1}{a} \frac{1}{N} \sum_{t=1}^{a} g_t^2(x_t; \beta_0) g_t^2(x_t; \beta_0)
\]

is positive definite for all \(b_N > p\), where

\[
\sigma^2 = (1, -g(x_t; \beta_0)) (1, -g(x_t; \beta_0))'.
\]

Assumption 5.8a

\[
\lim_{N \to \infty} \mathbf{m}^* = \mathbf{m}^*
\]

exists and is positive definite.

Assumption 5.8b

The constants \(x_t\) are uniformly bounded for \(t = 1, 2, \ldots, b_N\) and
$N = 1, 2, \ldots$.

**Assumption 5.9**

The sequence $\{b_N\}_{N=1}^\infty$ is monotonically increasing.

**Assumption 5.10**

The elements of the error covariance matrix satisfy $\hat{\psi} = O(aN)$.

**Assumption 5.11a**

The elements of the sequence $\{a_N\}_{N=1}^\infty$ satisfy $a_N = o(N^{-1/2})$.

**Assumption 5.11b**

The elements of the sequence $\{a_N\}_{N=1}^\infty$ satisfy $a_N^{3/2} = o(N^{-1/2})$.

With these assumptions in hand we are able to investigate the asymptotic properties of $\bar{x}_t$, $\hat{x}_t$, and $\Delta \hat{\beta}$ defined in the previous section. We begin by proving four lemmas which establish the properties of $\bar{x}_t$ and $\hat{x}_t$.

**Lemma 5.1**

Under Assumptions 5.10, 5.11a, 5.5, 5.3, and 5.2a,

$$\bar{x}_t - x_t = O_p(a_N^{1/2})$$

for each $t = 1, \ldots, b_N$.
Proof:

By definition, \( \bar{x}_t \) is obtained as that value of \( \xi \) which minimizes

\[
D^2_t(\xi) = [Y_t - g(\xi; \overline{\beta}), X_t - \xi](a_N^{-1})(Y_t - g(\xi; \overline{\beta}), X_t - \xi)'.
\]

(5.18)

Thus \( 0 \leq D^2_t(\bar{x}_t) \leq D^2_t(x_t) \).

By Theorem 2.10 and Assumptions 5.3 and 5.2a we have

\[
Y_t - g(x_t; \overline{\beta}) = Y_t - g(x_t; \beta_0) + O_p(N^{-1/2})
\]

\[
= e_t + O_p(N^{-1/2}).
\]

By Assumption 5.10, \( e_t = O_p(a_N^{1/2}) \) and \( u_t = O_p(a_N^{1/2}) \). Thus, it follows from Assumption 5.11a that

\[
Y_t - g(x_t; \overline{\beta}) = O_p(a_N^{1/2})
\]

and that

\[
X_t - x_t = O_p(a_N^{1/2}). \quad (5.19)
\]

From Equations 5.19 and 5.18 we have obtained

\[
0 \leq D^2_t(\bar{x}_t) \leq D^2_t(x_t) = O(a_N). \quad (5.20)
\]
Now we observe that $D^2_t(x_t)$ may be expressed as

$$D^2_t(x_t) = (X_t - \bar{x}_t)(a_N \hat{\phi}_u)^{-1}(X_t - \bar{x}_t)' + a_N(\sigma_e^2 - \hat{\phi}_e \hat{\phi}_u \hat{\phi}_u^{-1} \hat{\phi}_e)^{-1} \left( Y_t - g(x_t; \beta) - \hat{\phi}_e \hat{\phi}_u^{-1} (X_t - \bar{x}_t)' \right)^2$$

(5.21)

where $\hat{\phi}$ is partitioned as

$$\hat{\phi} = \begin{bmatrix} \sigma_e^2 & \hat{\phi}_e \\ \hat{\phi}_e & \hat{\phi}_u \\ \hat{\phi}_u & \hat{\phi}_u \end{bmatrix}.$$}

Clearly then, from Equations 5.20 and 5.21, we must have

$$0 \leq (X_t - \bar{x}_t)(a_N \hat{\phi}_u^{-1})(X_t - \bar{x}_t)' = D^2_t(x_t) = O_p(a_N).$$

Thus $X_t - \bar{x}_t = O_p(a_N^{1/2})$ and $\bar{x}_t - x_t = O_p(a_N^{1/2})$, since

$$X_t - x_t = u_t = O_p(a_N^{1/2}). \quad \text{Q.E.D.}$$

**Lemma 5.2**

Under Assumptions 5.10, 5.11a, 5.5, 5.3, and 5.2b,
\[ \bar{x}_t = x_t + \delta_t + o_p(N^{-1/2}) \]

for \( t = 1, \ldots, b_N \), where

\[
\delta_t = \{ u_t a_N \delta_{uu} + a_N \delta_{ue} \gamma_t \}
\]

\[ + e_t [a_N \delta_{eu} + a_N \sigma_{ee} \gamma_t] \Lambda_t^{-1} , \]

\[ \gamma_t = g_{x_t}(x_t; \beta_0) , \]

and

\[
\Lambda_t = a_N \delta_{uu} + a_N \delta_{ue} \gamma_t + \gamma_t' a_N \delta_{eu}
\]

\[ + \gamma_t' \gamma_t a_N \sigma_{ee} . \]

**Proof:**

Since \( g \) possesses continuous second partials, we can use a Taylor series expansion, Lemma 5.1, and Assumption 5.11a to obtain

\[
Y_t - g(\bar{x}_t; \bar{\beta}) = Y_t + e_t - g(x_t; \beta_0) - \gamma_t (\bar{x}_t - x_t)' + o_p(N^{-1/2})
\]

\[ = e_t - \gamma_t (\bar{x}_t - x_t)' + o_p(N^{-1/2}) \quad (5.22) \]

and

\[ d_t = g_{x_t}(x_t; \beta) = g_{x_t}(x_t; \beta_0) + o_p(a_N^{1/2}) \]
Substituting Equations 5.22 and 5.23 into Equation 5.5 and multiplying through by $a_N$ yields the following:

$$
\varphi = a_N \sigma_{ee} \left[ e_t - (\bar{y}_t - y_t) \gamma_t \right] \gamma_t
+ \left[ e_t - (\bar{y}_t - y_t) \gamma_t \right] a_N \frac{\partial}{\partial e} u
+ (X_t - \bar{x}_t) a_N \frac{\partial}{\partial x} u \gamma_t
+ (X_t - \bar{x}_t) a_N \frac{\partial}{\partial u} u + O_p(N^{-1/2}).
$$

If we recognize that $X_t = x_t + u_t$, then from Equation 5.24, $\bar{x}_t$ must satisfy

$$
(\bar{x}_t - x_t) \Lambda_t =
\left\{ u_t \left[ a_N \frac{\partial}{\partial u} u + a_N \frac{\partial}{\partial u} e \gamma_t \right] + e_t \left[ a_N \frac{\partial}{\partial e} u + \gamma_t a_N \sigma_{ee} \right] \right\}
+ O_p(N^{-1/2}).
$$

Since $\Lambda_t$ is order one and is nonsingular, the result of the lemma follows. Q.E.D.
To state the next lemma we need to establish some additional notation. Define

\[ v_t = Y_t - g(X_t; \beta_0) \]

and

\[ v_t^* = e_t - v_t \gamma_t' \]

The relation between \( v_t \) and \( v_t^* \) may be seen by expanding \( g(X_t; \beta_0) \).

Under Assumptions 5.2b, 5.10, and 5.11a we have

\[
v_t = y_t + e_t - g(x_t; \beta_0) - g_x(x_t; \beta_0)(X_t - x_t)' + O_p(a_N)
\]

\[ = e_t - u_t \gamma_t' + O_p(a_N) \]

\[ = v_t^* + O_p(a_N). \tag{5.25} \]

Now we have the following:

**Lemma 5.3**

Under Assumption 5.4,

\[ \text{Cov} \{v_t^*, \delta_t\} = \phi. \]

**Proof:**

By the definition of \( v_t^* \) and \( \delta_t \) we have
\[
\text{Cov}\{\gamma^*_t, \delta_t\} = E\{\gamma^*_t \delta_t\} = \\
E\{(e_t - u_t \gamma'_t)[u_t (a_N \gamma_t^{uu} + a_N \gamma_t^{ue}) + e_t (a_N \gamma_t^{ue} + a_N \gamma_t^{ee})]\} \Lambda_t^{-1} \\
= \{(\gamma_t^{eu} - \gamma_t^{uu})(a_N \gamma_t^{uu} + a_N \gamma_t^{ue}) \} \Lambda_t^{-1} \\
+ (a_e^2 - \gamma_t^{ue})(a_N \gamma_t^{ee} + a_N \gamma_t^{ee}) \} \Lambda_t^{-1} \\
= \{\gamma_t^{eu} + \gamma_t^{ue} - \gamma_t^{uu} - \gamma_t^{ue} \} \Lambda_t^{-1} \\
+ \sigma_e^2 \gamma_t^{ee} - \gamma_t^{ue} \sigma_e^2 \gamma_t^{ee} - \gamma_t^{uu} \sigma_e^2 \gamma_t^{ee} \} \cdot a_N \Lambda_t^{-1} \\
= \varphi.
\]

The last equality follows from the following useful relations between a partitioned matrix and its inverse:

\[
\sigma_e^2 \gamma_t^{ee} = -\gamma_t^{eu} \gamma_t^{uu}, \\
\gamma_t^{ue} \sigma_t^{ee} = -\gamma_t^{uu} \gamma_t^{ue}, \\
\sigma_t^2 \sigma_t^{ee} + \gamma_t^{ee} = 1, \\
\text{and } \gamma_t^{uu} \gamma_t^{eu} + \gamma_t^{uu} \gamma_t^{uu} = 1.
\]

Q.E.D.
The import of this lemma is clear. It establishes that $\tilde{x}_t$ and "residual" error $v^*_t$ are "nearly" uncorrelated.

We next consider the relation of $\tilde{x}_t$ to its approximation $\hat{x}_t$.

Subtracting Equation 5.7 from 5.6 and multiplying through by $a_N$ we obtain

\[
\varphi = (\tilde{x}_t - \hat{x}_t) \left\{ a_N \sigma_{\epsilon \epsilon} \left[ -\frac{g'_x(X_t; \bar{\beta})}{g_x(X_t; \bar{\beta})} \right] + \right. \\
\left. \left[ Y_t - g(x_t; \bar{\beta}) \right] g_{xx}(X_t; \bar{\beta}) - g'_x(X_t; \bar{\beta}) a_N \frac{h_{\epsilon \epsilon}^u}{a_N R_t} \right. \\
\left. - a_N \frac{h_{\epsilon u}^u}{a_N} g_x(X_t; \beta) - a_N \frac{h_{\epsilon u}^u}{a_N} \right\} + a_N R_t
\]

where $R_t$ is a remainder. But under the conditions of Assumption 5.2c, $a_N R_t = O_p(a_N)$. Furthermore, the term within braces in Equation 5.26 is $O_p(1)$. Thus we have proved Lemma 5.4.

**Lemma 5.4**

Under Assumptions 5.10, 5.11a, 5.5, 5.3, and 5.2c,

\[
\tilde{x}_t - \hat{x}_t = O_p(a_N).
\]

for $t = 1, \ldots, b_N$.

By this lemma, it immediately follows that the results of Lemmas 5.1 and 5.2 hold also for $\hat{x}_t$. Thus, if Equations 5.5 are such that
they can not easily be solved for \( \bar{x}_t \), then we can solve Equations 5.6 for the asymptotically equivalent \( \hat{x}_t \).

Having established the properties of \( \bar{x}_t \) and \( \hat{x}_t \), we turn our attention to the asymptotic properties of \( \hat{\beta} \) defined by Equation 5.16. First, we establish the order of the error in \( \Delta \hat{\beta} \).

**Lemma 5.5**

Given Assumptions 5.2c, 5.3, 5.4, 5.5, 5.6a, 5.7, 5.8a, 5.8b, 5.10, and 5.11a, then

\[
\hat{\beta} - \beta_0 = \Delta \hat{\beta} = O_p(N^{-1/2}),
\]

where \( \Delta \beta = \beta_0 - \bar{\beta} \) and \( \hat{\beta} = \bar{\beta} + \Delta \hat{\beta} \).

**Proof:**

Define

\[
M_{xx}^{-1} = b^{-1} \sum_{N} (a^{-1/2})^{-1} g_{\beta}(x; \bar{\beta}) g_{\beta}(x; \bar{\beta}).
\]

Then \( \Delta \hat{\beta} \) may be written as

\[
\Delta \hat{\beta} = (M_{xx}^{-1})^{-1} \left\{ b^{-1} \sum_{N} (a^{-1/2})^{-1} g_{\beta}(x; \bar{\beta})(e_t - d u_t') \right\}. \tag{5.27}
\]

By definition

\[
e_t - d u_t' = [(\gamma_t - g(x_t; \bar{\beta})) - g_{x_t}(x_t; \bar{\beta})(X_t - x_t')]'.
\]
and
\[ Y_t - g(X_t; \beta_0) = v_t. \]

Thus
\[ \bar{e}_t - d_t \bar{u}_t = g(X_t; \beta_0) + v_t - g(X_t; \bar{\beta}) - g_x(X_t; \bar{\beta})(X_t - \bar{x}_t). \]  \hspace{1cm} (5.28)

If we expand \( g(X_t; \beta_0) \) in a Taylor series about the point \((x_t; \bar{\beta})\),
we obtain
\begin{align*}
g(X_t; \beta_0) &= g(x_t; \bar{\beta}) + g'_\beta (x_t; \bar{\beta}) (\Delta \beta) + g_x(x_t; \bar{\beta})(X_t - \bar{x}_t) \\
&\quad + o_p(N^{-1/2}). \hspace{1cm} (5.29)
\end{align*}

This follows since \( g \) has bounded second partials, since
\[ \Delta \beta = \beta_0 - \bar{\beta} = O_p(N^{-1/2}), \]
\[ X_t - \bar{x}_t = O_p(a_{N^{1/2}}), \] and since \( a_{N} = o(N^{-1/2}). \)

Substituting Equations 5.29 and 5.25 into Equation 5.28 yields
\begin{align*}
\bar{e}_t - d_t \bar{u}_t &= g'_\beta (x_t; \bar{\beta}) (\Delta \beta) + v_t + o_p(N^{-1/2}) \\
&= g'_\beta (x_t; \bar{\beta}) (\Delta \beta) + v^*_t + o_p(N^{-1/2}). \hspace{1cm} (5.30)
\end{align*}

Then from Equation 5.30 we obtain
If we combine Equations 5.31 and 5.27 we find that

\[
(\Delta \beta) - (\Delta \beta) = (\bar{M} \times x)_{-1} \left\{ b_{N} \left[ \sum_{t=1}^{b_{N}} a_{N}^{-1} \right] \mathbf{g}_{\beta}(\bar{x}_{t}; \bar{\beta}) \right\}
+ \mathcal{O}(N^{-1/2}).
\] (5.32)

We now consider \( \bar{M} \times x \) and its components. Let

\[
\Delta \gamma_{t} = d_{t} - \gamma_{t} = g_{x}(\bar{x}_{t}; \bar{\beta}) - g_{x}(x_{t}; \beta_{0}).
\]

By assumption, \( g \) has bounded third derivatives. Thus,

\[
g_{x}(\bar{x}_{t}; \bar{\beta}) = g_{x}(x_{t}; \beta_{0}) + (\bar{x}_{t} - x_{t}) g_{xx}(x_{t}; \beta_{0}) + \mathcal{O}_{p}(N^{-1/2})
\]
since \( \Delta \beta = \mathcal{O}_{p}(N^{-1/2}) \) and \( (\bar{x}_{t} - x_{t}) = \mathcal{O}_{p}(a_{N}^{1/2}) \). Consequently,

\[
\Delta \gamma_{t} = (\bar{x}_{t} - x_{t}) g_{xx}(x_{t}; \beta_{0}) + \mathcal{O}_{p}(N^{-1/2})
= \mathcal{O}_{p}(a_{N}^{1/2}).
\] (5.33)
From Equation 5.33 and the definitions of \( \sigma_{v_t}^{-2} \) and \( \sigma_{v_t}^2 \) we obtain the following useful expression:

\[
-a_{N}^{-1} \sigma_{v_t}^{-2} = a_{N}^{-1} (1, -d_t) \frac{1}{2} (1, -d_t)'
\]

\[
= a_{N}^{-1} \sigma_{v_t}^2
\]

\[
- (\Delta \gamma_t) a_{\nu}^{-1} \frac{1}{2} \nu - a_{N}^{-1} \frac{1}{2} \nu \nu \Delta (\gamma_t)'
\]

\[
+ (\Delta \gamma_t) a_{N}^{-1} \frac{1}{2} \nu \nu \gamma_t' + \gamma_t a_{N}^{-1} \Sigma_{\nu} \nu \Delta (\gamma_t)'
\]

\[
+ o_{p} (N^{-1/2})
\]

since \( a_{N} = o_{p} (N^{-1/2}) \). But \( \overline{x}_t - x_t = \delta_t + o_{p} (a_{N}) = \delta_t + o_{p} (N^{-1/2}) \) by Lemma 5.2. Thus, we have

\[
a_{N}^{-1} \frac{1}{2} \nu = a_{N}^{-1} \sigma_{v_t}^2 + 
\]

\[
a_{N}^{-1} [ - \delta_t g_{xx}(x_t; \beta_0) \frac{1}{2} \nu - \frac{1}{2} \nu \nu g_{xx}(x_t; \beta_0) \delta_t' 
\]

\[
+ \delta_t g_{xx}(x_t; \beta_0) \frac{1}{2} \nu \gamma_t' - \gamma_t \frac{1}{2} \nu \nu g_{xx}(x_t; \beta_0) \delta_t' ]
\]

\[
+ o_{p} (N^{-1/2})
\]

(5.34)
by Equation 5.33.

From Theorem 2.11 and Equation 5.34 we now have

\[
(a^{-1} - 2)_{v_t}^{-1} = (a^{-1} - 2)_{v_t}^{-1} - (a^{-1} \Delta \sigma_{v_t}^2)(a^{-1} \sigma_{v_t}^2)^{-2} + O_p(N^{-1/2}),
\]

(5.35)

where

\[
\Delta \sigma_{v_t}^2 = -\delta_t g_{xx}(x_t; \beta_0) \frac{\partial}{\partial x} - \frac{\partial}{\partial x} g_{xx}(x_t; \beta_0) \delta_t
\]

\[
+ \delta_t g_{xx}(x_t; \beta_0) \frac{\partial}{\partial u} \gamma_t - \gamma_t \frac{\partial}{\partial u} g_{xx}(x_t; \beta_0) \delta_t.
\]

(5.36)

Next, by our assumptions we can expand \( g_{\beta}(\bar{x}_t; \bar{\beta}) \) about \((x_t; \beta_0)\)

which gives

\[
g_{\beta}(\bar{x}_t; \bar{\beta}) = g_{\beta}(x_t; \beta_0) + g_{\beta x}(x_t; \beta_0)(\bar{x}_t - x_t) + O_p(N^{-1/2}),
\]

where \( g_{\beta x}(x_t; \beta_0) \) denotes the \((p \times q)\) matrix of second mixed partial derivatives of \( g \) with respect to \( \beta \) and then \( x \), evaluated at \((x_t; \beta_0)\). Then, by Lemma 5.2 we obtain
Using Equations 5.35 and 5.37 we can express \( \bar{M}_{\bar{\mathbf{x}}} \) as follows:

\[
\bar{M}_{\bar{\mathbf{x}}} = b^{-1} \sum_{t=1}^{bN} \left[ \left( a^{-1} \sigma_{v_t}^2 \right)^{-1} \left( a^{-1} \Delta_{\sigma_{v_t}^2} \right) \right] \left( a^{-1} \sigma_{v_t}^2 \right)^{-2} \cdot 
\]

\[
\left[ g_{\beta}(x_t; \beta_0) g_{\beta}(x_t; \beta_0) + g_{\beta}(x_t; \beta_0) \delta \cdot g_{\beta}(x_t; \beta_0) \right] + O_p(N^{-1/2}) 
\]

since \( \delta_t = O_p(a_{N}^{-1/2}) \) and \( a_N = o(N^{-1/2}) \). But \( a^{-1} \Delta_{\sigma_{v_t}^2} = O_p(a_{N}^{-1/2}) \) and \( \delta_t = O_p(a_{N}^{-1/2}) \) by Assumptions 5.2b and 5.8b. Thus, it follows from Equation 5.38 that

\[
\bar{M}_{\bar{\mathbf{x}}} = b^{-1} \sum_{t=1}^{bN} \left( a^{-1} \sigma_{v_t}^2 \right)^{-1} g_{\beta}(x_t; \beta_0) g_{\beta}(x_t; \beta_0) + O_p(a_{N}^{-1/2}) 
\]

\[
= \bar{m}_{\bar{\mathbf{x}}} + O_p(a_{N}^{-1/2}) .
\]

Now we consider the term...
By Equations 5.35 and 5.37 we can write

\[ b^{-1} \sum_{t=1}^{b_N} \left( \frac{1}{N} \sigma_t^2 \right) g_\beta (x_t; \beta) v_t^* \]

Since \( v_t^* = O_p(a_N^{-1/2}) \), \( \delta_t = O_p(a_N^{-1/2}) \), \( a_N^{-1} \Delta \sigma_t^2 = O_p(a_N^{-1/2}) \), and

\[ a_N = o(N^{-1/2}) \]

Combining Equations 5.40, 5.39, and 5.32 with Assumption 5.7 gives

\[ \hat{(\Delta \beta)} - (\Delta \beta) = \left( \bar{m}^*_x \right)^{-1} \left\{ b^{-1} \sum_{t=1}^{b_N} \left( \frac{1}{N} \sigma_t^2 \right) g_\beta (x_t; \beta_0) v_t^* \right\} + o_p(N^{-1/2}) \]

But, by Assumption 5.4

\[ E(v_t^*) = 0 \]
and 
\[ \text{Var}(\mathbf{v}^*_t) = (1, -\gamma_t)^T \Phi(1, -\gamma_t)' \]
\[ = \sigma^2_{\mathbf{v}_t} = O(a_N). \]

Thus, by Theorem 2.2 and Assumption 5.8a

\[ b^{-1} \sum_{t=1}^{bN} \left( a_{-1, 0} \sigma_{\mathbf{v}_t}^2 \right)^{-1} g_\beta(x_t; \beta_0) \mathbf{v}^*_t = O(p N^{-1/2}). \]

The conclusion of the lemma follows. Q. E. D.

We now state and prove the major theorem of this section. This theorem establishes the limiting distribution of \( \hat{\beta} \).

**Theorem 5.1**

Given the model 5.1, 5.2 with Assumptions 5.2c, 5.3, 5.4, 5.5, 5.6a, 5.7, 5.8a, 5.8b, 5.9, 5.10, and 5.11a, then

\[ \sqrt{N} [\hat{\beta} - \beta] \xrightarrow{d} N(0, \lim_{N \to \infty} (\overline{\mathbf{m}}^*)^{-1}) \]

where \( \hat{\beta} = \bar{\beta} + (\Delta\hat{\beta}) \)

**Proof:**

From Lemma 5.5 and Equation 5.41,
\begin{equation}
\sqrt{N} [\hat{\beta} - \beta] = (m)^{-1} \left[ \frac{b}{N} \sum_{t=1}^{b} \left( a^{-1} \frac{2}{N} \right) g_{\beta}(x_t; \beta_0) v_t^* \right] + o_p(1)
\end{equation}

And by Theorem 2.6, the limiting distribution of $\sqrt{N} [\hat{\beta} - \beta]$ is the same as the limiting distribution of

\begin{equation}
(m)^{-1} \left[ \frac{b}{N} \sum_{t=1}^{b} \left( a^{-1} \frac{2}{N} \right) g_{\beta}(x_t; \beta_0) v_t^* \right] .
\end{equation}

We now investigate the limiting distribution of

\begin{equation}
n_{\lambda}^t(a^{-1} \frac{2}{N} \sigma^2) g_{\beta}(x_t; \beta_0) v_t^* , \text{ where } \lambda \text{ is an arbitrary, nonzero (p x 1)}
\end{equation}

vector. First, note that the random variables

\begin{equation}
n_{\lambda}^t(a^{-1} \frac{2}{N} \sigma^2) g_{\beta}(x_t; \beta_0) v_t^*
\end{equation}

are mutually independent. This follows by Assumption 5.4 and the definition of $v_t^*$. Second, note that
\[ E\{\lambda'(a_n^{-1/2} - 1)^{-1} g_\beta(x_t; \beta_0) v_t^* \}\]

\[ = \lambda'(a_n^{-1/2} - 1)^{-1} g_\beta(x_t; \beta_0) E(v_t^*) \]

\[ = 0 \]

for \( t = 1, \ldots, b_n \), since \( E(v_t^*) = 0 \) by Assumption 5.4. Third, note that

\[ E\{[\lambda'(a_n^{-1/2} - 1)^{-1} g_\beta(x_t; \beta_0) v_t^*]^2 \} = \]

\[ (a_n^{-1/2} - 1)^{-2} \lambda^2 g_\beta(x_t; \beta_0) g'_{\beta}(x_t; \beta_0) \lambda E(v_t^*) \]

\[ a_n(a_n^{-1/2} - 1)^{-1} \lambda g_\beta(x_t; \beta_0) g'_{\beta}(x_t; \beta_0) \lambda \]

for \( t = 1, \ldots, b_n \), since \( E(v_t^*^2) = (1, -\gamma_t)^{\frac{1}{2}} (1, -\gamma_t)' = \sigma_{v_t}^2 \).

Finally, since \((e_t, u_t)\) has finite \(2+\delta\) moments, \(\delta > 0\), since the constants \(x_t\) are bounded, and since \(g\) has continuous partial derivatives

\[ E\{|\lambda'(a_n^{-1/2} - 1)^{-1} g_\beta(x_t; \beta_0) v_t^*|^{2+\delta}\} \]
is bounded for $t = 1, \ldots, b_N$ and $N = 1, 2, \ldots$.

By Assumptions 5.8a, 5.8b, and 5.9 we have

$$
\lim_{N \to \infty} \frac{b_N}{N} \sum_{t=1}^{b_N} \mathbb{E} \left\{ \left| \lambda' \left( a_N \sigma_N \right)^{-1} g_{\beta}(x_t; \beta_0) v_t^* \right|^{\frac{2+\delta}{2}} \right\}
$$

$$
\leq \lim_{N \to \infty} \frac{b_N}{N} \sum_{t=1}^{b_N} \mathbb{E} \left\{ \left| \lambda' \left( a_N \sigma_N \right)^{-1} g_{\beta}(x_t; \beta_0) v_t^* \right|^2 \right\}^{\frac{2+\delta}{2}}
$$

$$
= \lim_{N \to \infty} \frac{b_N}{N} \sum_{t=1}^{b_N} \mathbb{E} \left\{ \left| a_N^{-1/2} \lambda' \left( a_N \sigma_N \right)^{-1} g_{\beta}(x_t; \beta_0) v_t^* \right|^2 \right\}^{\frac{2+\delta}{2}}
$$

$$
= 0. \text{ Thus, by the Liapounov central limit theorem,}
$$

$$
\lim_{N \to \infty} \frac{b_N}{N} \sum_{t=1}^{b_N} \mathbb{E} \left\{ \lambda' \left( a_N \sigma_N \right)^{-1} g_{\beta}(x_t; \beta_0) v_t^* \right\}^{\frac{1}{2}}
$$

$$
\leq \lim_{N \to \infty} \frac{b_N}{N} \sum_{t=1}^{b_N} \mathbb{E} \left\{ \lambda' \left( a_N \sigma_N \right)^{-1} g_{\beta}(x_t; \beta_0) v_t^* \right\}^{1/2}
$$
Furthermore, by Assumption 5.8a
\[
\lim_{N \to \infty} \frac{-m^*}{\sigma_{xx}} = \frac{-m^*}{\sigma_{xx}}
\]
exists and thus by Theorem 2.7
\[
\sqrt{N} b_N^{-1} \sum_{t=1}^{b_N} \lambda'(a^{-1} \sigma^2_{x_t}^{-1})^{-1} g_{\beta}(x_t; \beta_0) v_t^*
\]

\[
\xrightarrow{d} N(0, \lim_{N \to \infty} \lambda' \frac{-m^*}{\sigma_{xx}} \lambda)
\]  \hspace{1cm} (5.42)

By the multivariate central limit theorem and Equation 5.42 we have
\[ \sqrt{N} b_{N}^{-1} \sum_{t=1}^{b_{N}} (a_{N}^{-1} \sigma_{N}^{-2})^{-1} g_{\beta}(x_{t}; \beta_{0}) v_{t} \]

\[ \xrightarrow{d} N\left(0, \lim_{N \to \infty} \frac{m^{*}}{m_{XX}} \right). \]

The result then follows by noting that

\[ \frac{m^{*}-1}{m_{XX}} \left( \sqrt{N} b_{N}^{-1} \sum_{t=1}^{b_{N}} (a_{N}^{-1} \sigma_{N}^{-2})^{-1} g_{\beta}(x_{t}; \beta_{0}) v_{t} \right) \]

\[ \xrightarrow{d} N\left(0, \lim_{N \to \infty} \frac{m^{*}-1}{m_{XX}} \right). \quad \text{Q.E.D.} \]

D. A Revised Estimator

In Section B of this chapter an estimation procedure was given, and in Section C the asymptotic properties of the given estimators were developed under Assumption 5.11a. In this section, we consider the estimators of Section B, with the exception that a high order correction is made to the right hand side of equation defining \( \Delta \beta \). Also, we examine these estimators under the less restrictive Assumption 5.11b, i.e. \( a_{N}^{1.5} = o(N^{-1/2}) \).

The estimation procedure to be considered is given as follows:
1. Compute a preliminary estimator of $\beta$, say $\beta$, satisfying
$$\beta - \beta = O_p(N^{-1/2}).$$

2. For each $t = 1, 2, \ldots, b_N$, compute $x_t$ defined by Equation 5.5.

3. Compute $(\Delta \beta)^*$ defined by

$$
(\Delta \beta)^* = (M_{xx}^*)^{-1} \left( b^{-1} \sum_{t=1}^{b_N} \left( a^{-1} \sigma_t^{-2} \right)^{-1} g_{\beta}(x_t; \beta)(e_t - d_{tt} u_t) - \frac{1}{2} \text{tr} \left[ g_{xx}(x_t; \beta)(u_t u_t - \mathbb{I}_{uu}) \right] \right),
$$

where $\text{tr}(\cdot)$ denotes the trace operator.

4. Compute $\beta^*$ defined by

$$
\beta = \beta + v^* (\Delta \beta)^*
$$

where $v^*$ is that $v \in [0, 1]$ which minimizes

$$Q(\beta + v(\Delta \beta)^*, x_1, \ldots, x_{b_N})$$

with respect to $v$.

5. Return to step one and iterate if desired.

We now give four lemmas which describe the properties of $x_t$ and $x_t^*$ under Assumption 5.11b. In the proofs of these lemmas we will
have occasion to express the order of certain remainders as
\[ O_p(\max[a_N, N^{-1/2}]). \] For notational convenience these will be
abbreviated by \( O_p(\max) \).

**Lemma 5.6**

Under Assumptions 5.10, 5.11b, 5.5, 5.3 and 5.2a, then

\[ \frac{-x_t - x_t}{t} = O_p(a_N^{1/2}) \]

for each \( t = 1, \ldots, b_N \).

**Proof:**

This lemma follows immediately from the proof of Lemma 5.1.

Q.E.D.

**Lemma 5.7**

Under Assumptions 5.10, 5.11b, 5.5, 5.3, and 5.2b,

\[ \frac{-x_t - x_t}{t} = O_p(\max) \]

for \( t = 1, \ldots, b_N \), where \( \delta_t \) is defined by Lemma 5.2.

**Proof:**

Since \( g \) possesses continuous second partials, we can use Taylor
series expansions and Lemma 5.6 to obtain
\[ Y_t - g(x_t; \beta) = y_t + e_t - g(x_t; \beta_0) - \gamma_t (x_t - x_t)' + O_p(\text{max}) \]  

(5.44)

\[ e_t - \gamma_t (x_t - x_t)' + O_p(\text{max}) \]

Substituting Equations 5.44 and 5.45 into Equation 5.5 and multiplying through by \( a_N \) yields the following

\[ \varphi = a_N \sigma_{ee} \left[ e_t - (x_t - x_t) \gamma_t \right] + \left[ e_t - (x_t - x_t) \gamma_t' \right] a_N^{1/2} \]

\[ + (X_t - x_t) a_N^{1/2} \rho_{ue} \gamma_t \]

\[ + (X_t - x_t) a_N^{1/2} \rho_{uu} + O_p(\text{max}) \]

The remainder of this proof follows the proof of Lemma 5.2.

Q.E.D.

We now restate Lemma 5.3 due to its importance for the remaining results of this section.
Lemma 5.3

Under Assumption 5.4,

\[ \text{Cov}(v^*_t, \delta_t) = 0, \]

where \( v^*_t = e_t - u_t \gamma' \).

Lemma 5.8 establishes the relationship between \( \bar{x}_t \) and \( \hat{x}_t \), and since the proof follows the proof of Lemma 5.4, it is deleted.

Lemma 5.8

Under Assumptions 5.10, 5.11b, 5.5, 5.3, and 5.2c,

\[ \bar{x}_t - x_t = O_p(a_N). \]

By this lemma, it immediately follows that the results of Lemmas 5.6 and 5.7 hold also for \( \hat{x}_t \).

Next, we consider the asymptotic properties of \((\Delta\beta)^*\) defined by Equation 5.43. We establish the order of the error in \((\Delta\beta)^*\).

Lemma 5.9

Given Assumptions 5.2c, 5.3, 5.4, 5.5, 5.6b, 5.7, 5.8a, 5.8b, 5.10, and 5.11b, then
\[(\beta^* - \beta) = (\Delta \beta)^* - (\Delta \beta) = \mathcal{O}_p(N^{-1/2}),\]

where \(\Delta \beta = \beta_0 - \bar{\beta}\) and \(\beta^* = \bar{\beta} + (\Delta \beta)^*\).

Proof:

Since \(Y_t - g(X_t; \beta_0) = v_t\) we can write

\[e_t - d_t u_t' - \frac{1}{2} u_t g_x(x_t; \bar{\beta}) u_t\]

\[= g(X_t; \beta_0) + v_t - g(x_t; \bar{\beta}) = g(x_t; \beta_0)(X_t - x_t) \quad (5.46)\]

\[-\frac{1}{2} (X_t - x_t) g_{xx}(x_t; \bar{\beta})(X_t - x_t)' .\]

If we expand \(g(X_t; \beta_0)\) in a Taylor series about \((x_t; \bar{\beta})\), we obtain

\[g(X_t; \beta_0) = g(x_t; \bar{\beta}) + g'_x(x_t; \bar{\beta})(\Delta \beta) + g''_x(x_t; \bar{\beta})(X_t - x_t)\]

\[+ \frac{1}{2} (X_t - x_t) g_{xxx}(x_t; \bar{\beta})(X_t - x_t)' + \mathcal{O}_p(N^{-1/2}) . \quad (5.47)\]

This follows since \(g\) has bounded third partials,

\[\Delta \beta = \beta_0 - \bar{\beta} = \mathcal{O}_p(N^{-1/2}),\quad X_t - \bar{x}_t = \mathcal{O}_p(a_N^{1/2}),\quad \text{and since } a_N = o(N^{-1/2}).\]

Substituting Equation 5.47 into Equation 5.46 yields...
Note that

\[ v_t = Y_t - g(x_t; \beta_0) \]

\[ = y_t + e_t - g(x_t; \beta_0) - g(x_t; \beta_0)u_t' \]

\[ = e_t - \gamma u_t' - \frac{1}{2} tr \left[ g(x_t; \beta_0)'u_t u_t' \right] + o \left( N^{-1/2} \right). \]  

(5.49)

Define

\[ \bar{v}_t = e_t - \gamma u_t' - \frac{1}{2} tr \left[ g(x_t; \beta_0)'u_t u_t' - \frac{\partial}{\partial u} \right]. \]  

(5.50)

Since \( g \) has bounded third derivatives and \( \frac{\partial}{\partial u} = O(A_N) \), clearly

\[ \text{tr} g_{xx}(\bar{x}_t; \bar{\beta}) \frac{\partial}{\partial u} \]

\[ = \text{tr} g_{xx}(x_t; \beta_0) \frac{\partial}{\partial u} + O \left( a_N^{1.5} \right) \]  

(5.51)
Then from Equations 5.48, 5.50, and 5.51, we have

\[ g'(x; \beta_0) \Delta_{uu} + o_p(N^{-1/2}). \]

If we substitute Equation 5.52 into 5.43, we obtain

\[ (\beta^* - \beta) = (M_{xx}^*)^{-1} \left\{ b^{-1} \sum_{t=1}^{N} \left( a_{t} - \sigma^2 \right)^{-1} g_x(x_t; \beta) \Delta_{vt} \right\} \]

\[ + o_p(N^{-1/2}). \tag{5.53} \]

We next consider \( M_{xx}^* \) and its components. Let \( \Delta = \gamma - \gamma_0 = g_x(x; \beta) - g_x(x_t; \beta_0) \). By assumption, \( g \) has bounded third derivatives and thus

\[ g_x(x_t; \beta) = g_x(x_t; \beta_0) + (x_t - x_t) g_{xx}(x_t; \beta_0) + O_p(\max) \]

since \( \Delta \beta = O_p(N^{-1/2}) \) and \( x_t - x_t = O_p(a_{N}^{1/2}) \).
Consequently,

\[ \Delta \gamma_t = (\bar{x}_t - x_t) g_{xx}(x_t; \beta_0) + O_p(max) \]

\[ = \delta_t g_{xx}(x_t; \beta_0) + O_p(max) \]  

\[ = O_p(a_N^{1/2}) \]  

From Equation 5.54 and the definitions of \( \sigma^2_v \) and \( \sigma^2_v \), we obtain

\[
\begin{align*}
\sigma^{-1/2}_v &= a^{-1}_N (1, -d_t) \phi(1, -d_t)' \\
&= a^{-1}_N \sigma^{-2}_v + a^{-1}_N \{ - (\Delta \gamma_t) \phi_{ue} - \phi_{ue} (\Delta \gamma_t)' \} + O_p(a_N) \\
&= a^{-1}_N \sigma^{-2}_v + a^{-1}_N (\Delta \sigma^2_v) + O_p(max),
\end{align*}
\]

where \( \Delta \sigma^2_v \) is defined by Equation 5.36.

By our assumptions and by Lemma 5.7, we can expand \( g_\beta(\bar{x}_t; \beta) \) about \( (x_t; \beta_0) \) which gives

\[
g_\beta(\bar{x}_t; \beta) = g_\beta(x_t; \beta_0) + g_{\beta x}(x_t; \beta_0)(\bar{x}_t - x_t)'
\]
\[ + O_p(\text{max}) \]
\[ = g_\beta(x_t; \beta_0) + g_{\beta x}(x_t; \beta_0) \delta_t + O_p(\text{max}) \quad (5.56) \]

Then, by Equations 5.56, 5.55, Assumptions 5.8a, 5.8b, 5.10, and 5.2c it follows that

\[ \overline{M}^*_m = \overline{m}_{xx}^* + O_p(a_{1/2}^N). \quad (5.57) \]

This is easily seen by following the proof of Lemma 5.5.

Next, we consider the term \[ b^{-1} \sum_{N=1}^{B_N} \left( a_{N \sigma_v}^{1/2} \right)^{-1} g_\beta(x_t; \beta_v) \bar{v}_t. \]

By Equations 5.56 and 5.55 we have

\[ b^{-1} \sum_{N=1}^{B_N} \left( a_{N \sigma_v}^{1/2} \right)^{-1} g_\beta(x_t; \beta_v) \bar{v}_t = \quad (5.58) \]

\[ b^{-1} \sum_{N=1}^{B_N} \left[ (a_{-1/2}^{N \sigma_v})^{-1} - (a_{-1/2}^{N \Delta_v}) (a_{-1/2}^{N \sigma_v})^{-2} \right] [g_\beta(x_t; \beta_0) + g_{\beta x}(x_t; \beta_0) \delta_t] \bar{v}_t \]

+ \[ O_p(N^{-1/2}) \],

since \( \bar{v}_t = O_p(a_{1/2}^N) \) and \( a_{1.5}^N = O_p(N^{-1/2}) \). But note that \( E(\bar{v}_t) = 0 \), ...
\[ E(\delta_t) = \varphi, \text{ and} \]

\[ \text{Cov}(\overline{v}_t, \delta_t) = \]

\[ \text{Cov}(v^*_t - \frac{1}{2} \text{tr}[g_{xx}(x_t; \beta_0)(u_t'u_t' - \frac{1}{N} uu')], \delta_t) \]

\[ = \varphi + O(a_{1.5}) \]

by Assumptions 5.4, 5.6b and Lemma 5.3. Consequently, from

Equation 5.58, Assumption 5.8a, 5.6b, 5.8b, 5.2c and Theorem 2.2

we obtain

\[ b^{-1} \sum_{t=1}^{bN} (a_{-1}^{-1} \sigma^{-2}_v) g_{\beta}(x_t; \beta_0) \overline{v}_t = \]

\[ = O\left(N^{-1/2}\right), \quad (5.59) \]

since

\[ \text{Var}\left[b^{-1} \sum_{t=1}^{bN} (a_{-1}^{-1} \sigma^{-2}_v) g_{\beta}(x_t; \beta_0) \overline{v}_t \right] = O(b^{-1} a_N) = O(N^{-1}) \]

The conclusion of the lemma follows from Equations 5.59, 5.57, and 5.53. Q.E.D.
To conclude this section, we state the following theorem which establishes the limiting distribution of $\beta^*$. The proof follows the proof of Theorem 5.1 and is thus deleted.

Theorem 5.2

Given Model 5.1, 5.2 with Assumptions 5.2c, 5.3, 5.4, 5.5, 5.6b, 5.7, 5.8a, 5.8b, 5.9, 5.10, and 5.11b, then

$$\sqrt{N} [\beta^* - \beta] \overset{d}{\longrightarrow} N(0, \lim_{N \to \infty} (\mathbf{m}^*_{xx})^{-1}) ,$$

where $\beta^* = \bar{\beta} + (\Delta \beta)^*$. 

E. Miscellaneous Remarks

This section is devoted to three additional and important topics. First, we consider estimating the variance of the estimators $\hat{\beta}$ and $\beta^*$. 

To estimate the variances, we rely on the asymptotic properties of $\hat{\beta}$ and $\beta^*$ as developed in Sections C and D. In particular, we seek to estimate $(\mathbf{m}^*_{xx})^{-1}$, the asymptotic covariance matrix of $\beta$ and $\beta^*$. The next two lemmas provide us with a consistent estimator of this matrix.

Lemma 5.10

Given the assumptions of Lemma 5.9, a consistent estimator of
$(-m_{xx})^{-1}$, and thus of the asymptotic covariance matrix of $\beta^*$, is given by $(-M_{xx})^{-1}$.

Proof:

This result follows immediately from Equation 5.57, i.e.

$$\bar{M}_{xx} = \bar{m}_{xx} + O(p^{1/2}).$$

Q.E.D.

Lemma 5.11

Given the assumptions of Lemma 5.5 a consistent estimator of $(-m_{xx})^{-1}$, and thus of the asymptotic covariance matrix of $\hat{\beta}$, is given by $(-M_{xx})^{-1}$.

Proof:

This result follows immediately from Equation 5.39, i.e.

$$\bar{M}_{xx} = \bar{m}_{xx} + O(p^{1/2}).$$

Q.E.D.

The second topic concerns the estimator $\hat{x}_t$ defined by Equation 5.7. As mentioned in Section B, $\hat{x}_t$ is an easily computed approximation to $\bar{x}_t$. Thus, if the complexity of the function $g$ prohibits computation of $\bar{x}_t$, then one may compute $\hat{x}_t$. 
Since we have $\hat{x}_t - \bar{x}_t = O_p(a_N)$, (cf. Lemma 5.4) all of the results of Section C remain valid if $\hat{\beta}$ is defined in terms of $\hat{x}_t$ rather than $\bar{x}_t$. Similarly, Lemma 5.8 establishes that $\hat{x}_t^* - \bar{x}_t^* = O_p(a^*_N)$ so that all of the results of Section D remain valid if $\beta^*$ is defined in terms of $\hat{x}_t^*$.

The third topic concerns the consolidation of the results in Sections C and D with our knowledge of the likelihood function and of the maximum likelihood estimator. Each of the remarks made in the final section of Chapter 4 is pertinent to this point.

We emphasize that neither the estimator $\hat{\beta}$ of Section C, nor the estimator $\beta^*$ of Section D is the maximum likelihood estimator for Model 5.1, 5.2. However, if the iterative procedure converges, then we conjecture that $\hat{\beta}$ and $\beta^*$ iterate to a local maximum of the likelihood. As in Chapter 4 though, this conjecture is unproven.

Finally, we note that the asymptotic covariance matrix of Sections C and D, namely $(\mathbf{M}_{xx}^*)^{-1}$, is the inverse of the information matrix as determined by Dolby and Lipton (1972).
VI. NUMERICAL RESULTS

In this chapter we consider the quadratic functional relationship defined by Equations 3.1a and 3.1b. Chapters 3 and 4 were concerned with constructing estimators of the unknown $\beta$ and with developing the asymptotic properties of the constructed estimators. Here we consider these estimators of $\beta$ from a numerical point of view.

Section A reports a Monte Carlo study which investigates the small sample behavior of the estimators. Use of the estimators is then illustrated in Section B with an example from the earth sciences.

A. A Monte Carlo Study

Each of the estimators presented in Chapters 3 and 4, as well as several other estimators, were included in a Monte Carlo study. Runs of 200 trials were computed for three different sets of true values for each estimator in the study. Various sample statistics, i.e. mean, variance, mean square error (M.S.E.), were then computed from the 200 trials. The small sample behavior of the estimators was evaluated on the basis of the sample statistics.

1. Generating the data

The model considered is

$$y_t = x_t^2 \beta_2.$$
where we observe

\[ Y_t = y_t + e_t \]

\[ X_t = x_t + u_t \]

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

For each run of 200 trials, the data was generated from one of the following three parameter sets:

**Parameter Set I**

The true value of \( \beta \) was

\[ \beta = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \]

and the error covariance matrix was of the form \( \sigma^2 I_2 \), where \( \sigma^2 = .0324 \). The sample size was \( N = 33 \) observations per trial with 3 replicate observations on each of the following values of \( x \): 0.0, .1, .2, \ldots, .9, 1.0.

**Parameter Set II**

The true value of \( \beta \) was

\[ \beta = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \]
and the error covariance matrix was of the form $\sigma^2 I_2$, where $\sigma^2 = .0144$. The sample size was $N = 66$ observations per trial with 6 replicate observations on each of the following values of $x$: $0.0, .1, .2, \ldots, .9, 1.0$.

Parameter Set III

The true value of $\beta$ was

$$
\beta = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
$$

and the error covariance matrix was of the form $\sigma^2 I_2$, where $\sigma^2 = .0081$. The sample size was $N = 66$ observations per trial with 6 replicate observations on each of the following values of $x$: $0.0, .1, .2, \ldots, .9, 1.0$.

In all cases, the errors of measurement, $(e_t, u_t)$, were generated as independent, identically and normally distributed random variables with zero mean and covariance matrix $\sigma^2 I_2$. For each trial within each run, a sample of $2N$ independent $U(0, 1)$ pseudo-random deviates was generated using a composite, multiplicative-congruential generator. Using a pair of these $U(0, 1)$ deviates, say $U_1, U_2$, a pair of independent, standard normal pseudo-random deviates $Z_1, Z_2$ was defined by setting
\[ Z_1 = (-2L \cos(2\pi U_2))^{1/2} \cos(2\pi U_2) \]

and

\[ Z_2 = (-2L \sin(2\pi U_2))^{1/2} \sin(2\pi U_2). \]

A vector \((e_t, u_t)\) with the necessary distributional properties was then defined as

\[ (e_t, u_t) = \sigma I_2 (Z_1, Z_2). \]

In this manner, 2N independent \(U(0, 1)\) pseudo-random deviates were converted into a sample of \(N\) independent \(N(0, \sigma^2 I_2)\) pseudo-random deviates, namely \((e_t, u_t)\) for \(t = 1, \ldots, N\).

The "observable" random variables \(Y_t, X_t\) were formed by adding \((e_t, u_t)\) to the corresponding \((y_t, x_t)\). That is,

\[ Y_t = \beta_2 X_t^2 + e_t \]

and

\[ X_t = x_t + u_t \]

for \(t = 1, \ldots, N\). The observations \(\{(Y_t, X_t)\}_{t=1}^{N}\) and the error covariance matrix, \(\sigma^2 I_2\), were then used to compute estimates of \(\beta\).
To compare different estimators on equal terms, the same set of pseudo-random deviates was employed for each estimator in each run of 200 trials.

2. Adjusting the estimators

During preliminary Monte Carlo work, we quickly determined that the sample distributions functions of $\hat{\beta}$, $\hat{\beta}$, and $\beta^*$ possessed heavy tails due to the presence of a small number of deviate observations. To eliminate this heavy tail property, these estimators were adjusted in a manner analogous to that given by Fuller (1971).

The estimator $\hat{\beta}$ was originally defined by Equation 3.14. The adjusted form of this estimator is given by

$$\hat{\beta}(k) = \left[ \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i} - (\alpha - k/N) \frac{\hat{\beta}}{\hat{\beta}_{fe}} \right]^{-1}.$$  \hspace{1cm} (6.1)

where $k > 0$ is a fixed number.

The estimators $(\Delta \beta)$ and $(\Delta \beta)^*$ were originally defined by Equations 4.34 and 4.65 respectively. These expressions involved the notation $a_N$ and $b_N$, where $N = b_N / a_N$. However, in this chapter, and in most real situations, we take $a_N = 1$ and $b_N = N$. 
Three adjusted forms of \((\hat{\Delta}\beta)\) and of \((\hat{\Delta}\beta)^*\) are defined as follows:

\[
1(\hat{\Delta}\beta)(k) = \left[ \frac{M^*_{xx}}{x^*} + \frac{k}{N} \hat{\beta}_{ff} \right]^{-1}.
\]

\[
\begin{bmatrix}
\hat{\epsilon}_t - d\hat{u}_t \\
\hat{x}_t^2
\end{bmatrix}
\begin{bmatrix}
1 \\
\hat{x}_t^2
\end{bmatrix}
\left[ \frac{N^{-1} \sum_{t=1}^{N} \hat{\sigma}_t^2}{N} \right]^{-1}
\]

\[
\]
\[
\left[ N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \begin{pmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{pmatrix} \right] v_t ; \quad (6.5)
\]

\[
\bar{\beta}^* (k) = \left[ \bar{M}_{w}^* - \Delta \right]^{-1} \bar{e}_t. \quad (6.6)
\]

\[
\left[ N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \begin{pmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{pmatrix} \right] [\hat{e}_t - d_t u_t] ;
\]

\[
\bar{\beta}^* (k) = \left[ \bar{M}_{w}^* - \Delta \right]^{-1} \bar{e}_t. \quad (6.7)
\]

\[
\left[ N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \begin{pmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{pmatrix} \right] \hat{v}_t.
\]

In the above equations

\[
k > 0 ;
\]

\[
\hat{\beta}^* = \begin{pmatrix} \hat{\beta}_{ee}^* & \hat{\beta}_{ef}^* \\ \hat{\beta}_{fe}^* & \hat{\beta}_{ff}^* \end{pmatrix} = N^{-1} \sum_{t=1}^{N} \sigma^{-2} \hat{\beta}_{tt}.
\]
\[ M_{XX}^* = N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \begin{bmatrix} 1 & X_t & X_t^2 \end{bmatrix} \]

\[ M_{\tilde{w} \tilde{w}}^* = N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \begin{bmatrix} 1 & \hat{x}_t & \hat{x}_t^2 \hat{x}_t^2 - \sigma_\delta^2 \end{bmatrix} \]

and

\[ \Delta = \begin{cases} (1 - \frac{k}{N}) N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \sigma_\delta^2 \begin{bmatrix} 1 & 2\hat{x}_t \\ 2\hat{x}_t & 4\hat{x}_t^2 \end{bmatrix} & \text{for } \lambda > 1 \\ (\lambda - \frac{k}{N}) N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \sigma_\delta^2 \begin{bmatrix} 1 & 2\hat{x}_t \\ 2\hat{x}_t & 4\hat{x}_t^2 \end{bmatrix} & \text{for } \lambda \leq 1 \end{cases} \]

where \( \sigma_\delta^2 = (\sigma_e^2 - \sigma_u^2 - \sigma_{eu}^2) / \sigma_v^2 \) and \( \hat{\lambda} \) is the smallest root of the determinant of the determinantal equation

\[ \left| M_{\tilde{w} \tilde{w}}^* - \lambda \left\{ N^{-1} \sum_{t=1}^{N} \sigma^{-2} v_t \sigma_\delta^2 \begin{bmatrix} 1 & 2\hat{x}_t \\ 2\hat{x}_t & 4\hat{x}_t^2 \end{bmatrix} \right\} \right| = 0. \]
Adjusted estimators of \( \hat{\beta} \) and \( \beta^* \), say \( \hat{\beta}(k), \hat{\beta^*}(k), \beta^*(k) \), are defined in the obvious way. Consider \( \beta^*(k) \) as an example:

\[
\beta^*(k) = \hat{\beta} + v^* \Delta \beta^*(k)
\]

where \( v^* \) is that \( v \in [0, 1] \) which minimizes \( Q(\beta + v^* \Delta \beta^*, \hat{x}_1, \ldots, \hat{x}_N) \) defined by Equation 4.22.

Estimators of the variances of the estimators are defined in a corresponding manner. Following Theorem 3.2 an adjusted estimator of the covariance matrix of \( \hat{\beta}(k) \) is given by

\[
\left[ \overline{M}_{WW} - (\hat{\alpha} - \frac{k}{N}) \overline{H}_{ff} \right]^{-1} \left[ N^{-1} \sum_{t=1}^{N} \phi_t \phi_t^t \right] \left[ \overline{M}_{WW} - (\hat{\alpha} - \frac{k}{N}) \overline{H}_{ff} \right]^{-1} \cdot (6.8)
\]

And, following Lemmas 4.14 and 4.15, adjusted estimators of the covariance matrices of \( \hat{\beta}(k), \beta^*(k) \), \( \beta^*(k), \beta^*(k) \), \( \beta^*(k), \beta^*(k) \) are given respectively by

\[
\left[ \overline{M}_{xx}^* + \frac{k}{N} \overline{H}_{ff}^* \right]^{-1},
\]

\[
\left[ \overline{M}_{xx}^* + \frac{k}{N} \overline{H}_{ff}^* \right]^{-1}.
\]
3. Results and conclusions

Tables 1 through 6 summarize the results of the most important Monte Carlo runs. We now consider these results and their implications.

Each of the adjusted estimators was defined in terms of a constant $k$. Following Fuller (1971) and Booth (1973), we determined that as $k$ increases, the bias in the adjusted estimators increases but the variance decreases. Seeking a small mean square error (M.S.E.) with a modest bias, $k = 4$ was found to be a good choice. This point is illustrated by runs 4, 5, 6, and 7 in Table 1. Note that for both parameter Set I and II $\hat{\beta}(4)$ performs much better than $\hat{\beta}(1)$ in terms of M.S.E. Also, the sample distribution function of $\hat{\beta}(4)$ was nearly that of a normal random variable, while the sample distribution function of $\hat{\beta}(1)$ possessed longer tails than under the normal hypothesis. For these reasons the remainder of the results of this chapter are for the
Table 1. Results of Monte Carlo study of $\hat{\beta}_{OLS}^\prime$, $\hat{\beta}(1)$, and $\hat{\beta}(4)$

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Parameter Set</th>
<th>Est.</th>
<th>Mean</th>
<th>Var.</th>
<th>M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>$\hat{\beta}_{OLS}(0)$</td>
<td>0.0095</td>
<td>0.0059</td>
<td>0.0060</td>
</tr>
<tr>
<td>1</td>
<td>I</td>
<td>$\hat{\beta}_{OLS}(1)$</td>
<td>0.3606</td>
<td>0.1045</td>
<td>0.2345</td>
</tr>
<tr>
<td>1</td>
<td>I</td>
<td>$\hat{\beta}_{OLS}(2)$</td>
<td>0.4040</td>
<td>0.0970</td>
<td>0.4522</td>
</tr>
<tr>
<td>2</td>
<td>II</td>
<td>$\hat{\beta}_{OLS}(0)$</td>
<td>0.0115</td>
<td>0.0014</td>
<td>0.0015</td>
</tr>
<tr>
<td>2</td>
<td>II</td>
<td>$\hat{\beta}_{OLS}(1)$</td>
<td>0.2930</td>
<td>0.0337</td>
<td>0.1195</td>
</tr>
<tr>
<td>2</td>
<td>II</td>
<td>$\hat{\beta}_{OLS}(2)$</td>
<td>0.5863</td>
<td>0.0394</td>
<td>0.2105</td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>$\hat{\beta}_{OLS}(0)$</td>
<td>0.0133</td>
<td>0.0009</td>
<td>0.0011</td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>$\hat{\beta}_{OLS}(1)$</td>
<td>0.2093</td>
<td>0.0238</td>
<td>0.0676</td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>$\hat{\beta}_{OLS}(2)$</td>
<td>0.7177</td>
<td>0.0301</td>
<td>0.1098</td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>0.0151</td>
<td>0.0604</td>
<td>0.0606</td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>-0.1247</td>
<td>1.8194</td>
<td>1.8350</td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>1.1400</td>
<td>1.8513</td>
<td>1.8709</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>$\hat{\beta}(1)$</td>
<td>0.0052</td>
<td>0.0046</td>
<td>0.0046</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>$\hat{\beta}(1)$</td>
<td>-0.0357</td>
<td>0.1556</td>
<td>0.1569</td>
</tr>
<tr>
<td>5</td>
<td>II</td>
<td>$\hat{\beta}(1)$</td>
<td>1.0367</td>
<td>1.1768</td>
<td>1.1781</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>-0.0007</td>
<td>0.0245</td>
<td>0.0245</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>0.0758</td>
<td>0.6530</td>
<td>0.6587</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>$\hat{\beta}(1)$</td>
<td>0.8871</td>
<td>0.6417</td>
<td>0.6544</td>
</tr>
</tbody>
</table>
Table 1 Continued

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Parameter Set</th>
<th>Est.</th>
<th>Mean</th>
<th>Var.</th>
<th>M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>II</td>
<td>$\hat{\beta}_0(4)$</td>
<td>.0020</td>
<td>.0041</td>
<td>.0041</td>
</tr>
<tr>
<td>7</td>
<td>II</td>
<td>$\hat{\beta}_1(4)$</td>
<td>-.0013</td>
<td>.1347</td>
<td>.1347</td>
</tr>
<tr>
<td>7</td>
<td>II</td>
<td>$\hat{\beta}_2(4)$</td>
<td>.9952</td>
<td>.1531</td>
<td>.1531</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>$\hat{\beta}_0(4)$</td>
<td>.0001</td>
<td>.0016</td>
<td>.0016</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>$\hat{\beta}_1(4)$</td>
<td>.0038</td>
<td>.0525</td>
<td>.0525</td>
</tr>
<tr>
<td>8</td>
<td>III</td>
<td>$\hat{\beta}_2(4)$</td>
<td>.9909</td>
<td>.0656</td>
<td>.0656</td>
</tr>
</tbody>
</table>
case \( k = 4 \).

Three preliminary type estimators of \( \beta \) were considered in the study: \( \hat{\beta}(k) \), \( \hat{\beta}_{\text{OLS}} \), and \( \hat{\beta}_M \) (see Appendix A). During preliminary work the properties of \( \hat{\beta}_M \) were found to be unsatisfactory relative to those of \( \hat{\beta}(4) \) and \( \hat{\beta}_{\text{OLS}} \). \( \hat{\beta}_M \) was thus dropped from consideration.

Sample statistics for \( \hat{\beta}(4) \) and \( \hat{\beta}_{\text{OLS}} \) are given in Table 1, runs 1, 2, 3, 6, 7, and 8. For each parameter set we see that the bias in \( \hat{\beta}_{\text{OLS}} \) is quite sizeable relative to the bias in \( \hat{\beta}(4) \). However, the variance of \( \hat{\beta}_{\text{OLS}} \) is less than the variance of \( \hat{\beta}(4) \) in every case. The M.S.E. results are mixed. For Parameter Set I \( \hat{\beta}_{\text{OLS}} \) has smaller M.S.E. than \( \hat{\beta}(4) \), while for Parameter Set III the reverse is true.

For Parameter Set II the M.S.E. of \( \hat{\beta}_2(4) \) is less than that of \( \hat{\beta}_{\text{OLS}}(2) \).

The sample distribution functions of \( \hat{\beta}(4) \) and \( \hat{\beta}_{\text{OLS}} \) were in reasonable agreement with the normal distribution for each parameter set studied.

Statistics analogous to Student's t, say \( \hat{t}(4) \), were computed for \( \hat{\beta}(4) \) using the estimator of its variance given by Equation 6.8. Sample percentiles for these statistics are given in Table 2. Note the reasonable agreement between the sample deviations and the theoretical deviations, with the agreement improving for larger sample sizes and smaller error variances. \( \hat{t}(4) \) performed best for Parameter Set III, while for Parameter Sets I and II the sample distribution functions of
Table 2. Comparison of theoretical t distribution with computed t's for $\hat{\beta}(4)$

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Theoretical Deviation $^b$</th>
<th>Parameter Set I</th>
<th>Parameter Set II</th>
<th>Parameter Set III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{t}_0(4)$</td>
<td>$\hat{t}_1(4)$</td>
<td>$\hat{t}_2(4)$</td>
</tr>
<tr>
<td>5%</td>
<td>-1.671</td>
<td>-1.780</td>
<td>-1.517</td>
<td>-3.082</td>
</tr>
<tr>
<td>10%</td>
<td>-1.296</td>
<td>-1.316</td>
<td>-1.050</td>
<td>-1.773</td>
</tr>
<tr>
<td>50%</td>
<td>0.0</td>
<td>0.194</td>
<td>0.015</td>
<td>-0.120</td>
</tr>
<tr>
<td>90%</td>
<td>1.296</td>
<td>1.505</td>
<td>1.010</td>
<td>1.550</td>
</tr>
<tr>
<td>95%</td>
<td>1.671</td>
<td>1.912</td>
<td>2.348</td>
<td>1.290</td>
</tr>
</tbody>
</table>

$^a$ Observed deviations for statistics analogous to Student's t computed using $\hat{\beta}(4)$ and the estimator of its variance.

$^b$ Theoretical deviations for Student's t distribution with 60 d.f.
\(\hat{t}_0(4)\) and \(\hat{t}_2(4)\) were skewed towards the negative scale and that of \(\hat{t}_1(4)\) was skewed towards the positive scale.

Many revised or iterative estimators of \(\beta\) were studied, among them the estimators of Chapter 4 and Appendix A. Due to their superior performance and intuitive appeal, only the results for the adjusted estimators of Chapter 4 are presented here.

Runs 9 and 10 offer a comparison between an adjusted estimator of \(\hat{\beta}\) and an adjusted estimator of \(\beta^*\). From a theoretical standpoint, the results of Chapter 4 indicate that \(\hat{\beta}\) should perform very nearly the same as \(\beta^*\). The Monte Carlo results confirm this, though \(\hat{\beta}^*(4)\) does perform almost 10% better than \(\hat{\beta}(4)\) in terms of M.S.E. On this basis we recommend adjusted estimators of \(\beta^*\) and suppress further consideration of \(\hat{\beta}\).

We can evaluate the effects of the first two adjustments to \(\beta^*\) by studying the results of runs 10 and 12. In terms of bias, variance, and M.S.E. both \(\hat{\beta}^*(4)\) and \(\hat{\beta}^*(4)\) perform identically.

Each of the iterative estimators discussed thus far has employed \(\hat{\beta}(4)\) as preliminary estimator. To evaluate \(\hat{\beta}_{OLS}\) as a preliminary estimator, run 11 may be compared to run 13 and run 12 may be compared to run 14. This evidence is startling. The bias, variance, and M.S.E. for the estimator which uses \(\hat{\beta}_{OLS}\) as preliminary estimator (see runs 13 and 14) are considerably smaller than the
Table 3. Results of Monte Carlo study of iterative estimators

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Parameter Set</th>
<th>Final Est.</th>
<th>Mean</th>
<th>Variance</th>
<th>M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>II</td>
<td>$\hat{\beta}_0^{(4)}$</td>
<td>-.0136</td>
<td>.0027</td>
<td>.0028</td>
</tr>
<tr>
<td>9</td>
<td>II</td>
<td>$\hat{\beta}_1^{(4)}$</td>
<td>-.0218</td>
<td>.1132</td>
<td>.1136</td>
</tr>
<tr>
<td>9</td>
<td>II</td>
<td>$\hat{\beta}_2^{(4)}$</td>
<td>1.0453</td>
<td>.1474</td>
<td>.1494</td>
</tr>
<tr>
<td>10</td>
<td>II</td>
<td>$\hat{\beta}_0^{*(4)}$</td>
<td>.0012</td>
<td>.0030</td>
<td>.0030</td>
</tr>
<tr>
<td>10</td>
<td>II</td>
<td>$\hat{\beta}_1^{*(4)}$</td>
<td>-.0326</td>
<td>.1073</td>
<td>.1084</td>
</tr>
<tr>
<td>10</td>
<td>II</td>
<td>$\hat{\beta}_2^{*(4)}$</td>
<td>1.0420</td>
<td>.1351</td>
<td>.1369</td>
</tr>
<tr>
<td>11</td>
<td>I</td>
<td>$\hat{\beta}_0^{*(4)}$</td>
<td>-.0050</td>
<td>.0524</td>
<td>.0524</td>
</tr>
<tr>
<td>11</td>
<td>I</td>
<td>$\hat{\beta}_1^{*(4)}$</td>
<td>-.0973</td>
<td>1.4456</td>
<td>1.4551</td>
</tr>
<tr>
<td>11</td>
<td>I</td>
<td>$\hat{\beta}_2^{*(4)}$</td>
<td>1.1502</td>
<td>1.4212</td>
<td>1.4438</td>
</tr>
<tr>
<td>12</td>
<td>II</td>
<td>$\hat{\beta}_0^{*(4)}$</td>
<td>.0012</td>
<td>.0030</td>
<td>.0030</td>
</tr>
<tr>
<td>12</td>
<td>II</td>
<td>$\hat{\beta}_1^{*(4)}$</td>
<td>-.0330</td>
<td>.1073</td>
<td>.1084</td>
</tr>
<tr>
<td>12</td>
<td>II</td>
<td>$\hat{\beta}_2^{*(4)}$</td>
<td>1.0423</td>
<td>.1350</td>
<td>.1368</td>
</tr>
</tbody>
</table>

---

*a* Each of the estimators represented in this table was computed using three iterations per trial.

*b* Experiments 9-12 used $\hat{\beta}(4)$ as a preliminary estimator. Experiments 13-17 used $\hat{\beta}_{OLS}$ as a preliminary estimator.

*c* These large M.S.E.'s were due in part to a small number of very "wild" observations.
Table 3 Continued

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Parameter Set</th>
<th>Final Est.</th>
<th>Mean</th>
<th>Var.</th>
<th>M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>I</td>
<td>$2\beta_0^*(4)$</td>
<td>- .0173</td>
<td>.0132</td>
<td>.0135</td>
</tr>
<tr>
<td>13</td>
<td>I</td>
<td>$2\beta_1^*(4)$</td>
<td>.0463</td>
<td>.3646</td>
<td>.3667</td>
</tr>
<tr>
<td>13</td>
<td>I</td>
<td>$2\beta_2^*(4)$</td>
<td>.9614</td>
<td>.3955</td>
<td>.3970</td>
</tr>
<tr>
<td>14</td>
<td>II</td>
<td>$2\beta_0^*(4)$</td>
<td>-.0080</td>
<td>.0025</td>
<td>.0025</td>
</tr>
<tr>
<td>14</td>
<td>II</td>
<td>$2\beta_1^*(4)$</td>
<td>.0306</td>
<td>.0782</td>
<td>.0791</td>
</tr>
<tr>
<td>14</td>
<td>II</td>
<td>$2\beta_2^*(4)$</td>
<td>.9737</td>
<td>.0956</td>
<td>.0953</td>
</tr>
<tr>
<td>15</td>
<td>I</td>
<td>$3\beta_0^*(4)$</td>
<td>.0031</td>
<td>.0195</td>
<td>.0195</td>
</tr>
<tr>
<td>15</td>
<td>I</td>
<td>$3\beta_1^*(4)$</td>
<td>-.0169</td>
<td>.5103</td>
<td>.5106</td>
</tr>
<tr>
<td>15</td>
<td>I</td>
<td>$3\beta_2^*(4)$</td>
<td>.9984</td>
<td>.4980</td>
<td>.4980</td>
</tr>
<tr>
<td>16</td>
<td>II</td>
<td>$3\beta_0^*(4)$</td>
<td>-.0019</td>
<td>.0028</td>
<td>.0028</td>
</tr>
<tr>
<td>16</td>
<td>II</td>
<td>$3\beta_1^*(4)$</td>
<td>-.0021</td>
<td>.0886</td>
<td>.0886</td>
</tr>
<tr>
<td>16</td>
<td>II</td>
<td>$3\beta_2^*(4)$</td>
<td>1.0048</td>
<td>1.070</td>
<td>1.070</td>
</tr>
<tr>
<td>17</td>
<td>III</td>
<td>$3\beta_0^*(4)$</td>
<td>-.0016</td>
<td>.0013</td>
<td>.0013</td>
</tr>
<tr>
<td>17</td>
<td>III</td>
<td>$3\beta_1^*(4)$</td>
<td>-.0012</td>
<td>.0406</td>
<td>.0406</td>
</tr>
<tr>
<td>17</td>
<td>III</td>
<td>$3\beta_2^*(4)$</td>
<td>1.0023</td>
<td>.0540</td>
<td>.0540</td>
</tr>
</tbody>
</table>
corresponding statistics of the estimator which uses $\hat{\beta}(4)$ (see runs 11 and 12). Though $\hat{\beta}_{OLS}$ is considerably biased, it is evident that its small variance (the variance of $\hat{\beta}_{OLS}$ is much less than the variance of $\hat{\beta}(4)$) must contribute to the superior performance of iterative estimators which use it as preliminary estimator. Of course, as noted below Table 3, the iterative estimators were computed using only 3 iterations per trial. Consequently, these results also indicate that more iterations would generally be required for convergence of estimators using $\hat{\beta}(4)$.

Concluding that $\hat{\beta}_{OLS}$ is preferable to $\hat{\beta}(4)$ as a preliminary estimator, we focus on the question of $\hat{\beta}^*(4)$ versus $\hat{\beta}^*(4)$. Run 13 may be compared to run 15 and run 14 to 16. Careful analysis of these figures indicates that the bias in $\hat{\beta}^*(4)$ is uniformly smaller than the bias in $\hat{\beta}^*(4)$. However, we note that the variance and M.S.E. of $\hat{\beta}^*(4)$ are uniformly smaller than the variance and M.S.E. of $\hat{\beta}^*(4)$.

The sample distribution functions of both $\hat{\beta}^*(4)$ and $\hat{\beta}^*(4)$ were in reasonable agreement with the distribution function of a normal random variable, with neither estimator superior to the other in this regard.

We are thus faced with a dilemma, which estimator to recommend, $\hat{\beta}^*(4)$ or $\hat{\beta}^*(4)$. If we consider their respective variance estimators though, the choice is clear.
Statistics analogous to Student's t were computed for $3\beta^*(4)$ and $2\beta^*(4)$. Sample percentiles for these statistics are given in Tables 4 and 5 respectively. In Table 4 we observe a very reasonable agreement between the observed deviations and the theoretical deviations. The agreement between the observed and theoretical deviations in Table 5 is very poor. We conclude that the estimator of the variance of $3\beta^*(4)$ is performing as the theoretical results of Chapter 4 indicate, while the estimator of the variance of $2\beta^*(4)$ is severely underestimating the variance.

As we observed for $t(4)$, we note that (see Table 4) the sample distribution functions of $3t^*_0(4)$ and $3t^*_2(4)$ are slightly skewed towards the negative scale. Conversely, the sample distributions function of $3t^*_2(4)$ is slightly skewed towards the positive scale.

Table 6 gives the asymptotic covariance matrix of $\sqrt{N}(\beta^* - \beta)$ for each parameter set studied. Comparing these figures with the sample variances for runs 15, 16, and 17 of Table 3, we see that the true asymptotic variances are uniformly smaller than the observed variances of $3\beta^*(4)$.

Finally, we conclude that a gain is made in computing an iterative estimator as opposed to a preliminary estimator only. Comparing runs 15, 16, and 17 with runs 1, 2, and 3 demonstrates that $3\beta^*(4)$ has smaller M.S.E. than $\hat{\beta}_{OLS}$ for Parameter Sets II and III. For
Table 4. Comparison of theoretical t distribution with computed t's for $3\hat{\beta}^*(4)^a$

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Theoretical Deviation</th>
<th>Observed Deviations b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Parameter Set I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_0^*(4)$</td>
</tr>
<tr>
<td>5%</td>
<td>-1.671</td>
<td>-1.607</td>
</tr>
<tr>
<td>10%</td>
<td>-1.296</td>
<td>-1.311</td>
</tr>
<tr>
<td>50%</td>
<td>0.0</td>
<td>-.002</td>
</tr>
<tr>
<td>90%</td>
<td>1.296</td>
<td>1.291</td>
</tr>
<tr>
<td>95%</td>
<td>1.671</td>
<td>1.564</td>
</tr>
</tbody>
</table>

a $3\hat{\beta}^*(4)$ is computed using three iterations per trial and using $\hat{\beta}_{OLS}$ as a preliminary estimator.

b Observed deviations for statistics analogous to Student's $t$ computed using $3\hat{\beta}^*(4)$ and the estimator of its variance.

c Theoretical deviations for Student's $t$ distribution with 60 d.f.
Table 5. Comparison of theoretical t distribution with computed $t'$s for $2\beta^*(4)^a$

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Theoretical Deviation $^c$</th>
<th>Observed Deviations $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter Set I</td>
<td>Parameter Set II</td>
</tr>
<tr>
<td></td>
<td>$t_0^*(4)$</td>
<td>$t_1^*(4)$</td>
</tr>
<tr>
<td>10%</td>
<td>-1.296</td>
<td>-1.951</td>
</tr>
<tr>
<td>50%</td>
<td>0.0</td>
<td>-.269</td>
</tr>
<tr>
<td>90%</td>
<td>1.296</td>
<td>1.747</td>
</tr>
<tr>
<td>95%</td>
<td>1.671</td>
<td>2.367</td>
</tr>
</tbody>
</table>

$^a$ $2\beta^*(4)$ is computed using three iterations per trial and using $\hat{\beta}_{OLS}$ is a preliminary estimator.

$^b$ Observed deviations for statistics analogous to Student's t computed using $2\beta^*(4)$ and the estimator of its variance.

$^c$ Theoretical deviations for Student's t distribution with 60 d.f.
Table 6. True asymptotic covariance matrix of 
\[ \sqrt{N}(\beta^* - \beta) \]

| Parameter Set | \[0.23194 \quad -1.0225 \quad 0.89059\] | \[7.4579 \quad -7.7327 \quad 8.8724\] |
|---------------|------------------------------------------|
| I             | \[0.10308 \quad 0.45442 \quad 0.39579\] | \[3.3144 \quad -3.4365 \quad 3.9430\] |
| II            | \[0.057982 \quad -0.25560 \quad 0.22263\] | \[1.8643 \quad -1.9330 \quad 2.2180\] |
| III           |                                           |                                   |
Parameter Set I $\hat{\beta}_{OLS}$ has smaller M.S.E. But $\hat{\beta}_{OLS}$ is so badly biased for this parameter set that $\delta_{\beta}(4)$ is still preferred.

Comparing runs 15, 16, and 17 with runs 6, 7, and 8 demonstrates that $\delta_{\beta}(4)$ has uniformly smaller M.S.E. and bias than $\hat{\beta}(4)$.

To summarize, these Monte Carlo results are in general agreement with the theoretical results of Chapters 3 and 4. The adjusted estimator $\delta_{\beta}(k)$, defined by Equation 6.7, is recommended as a final estimator, while $\hat{\beta}_{OLS}$, defined by Equation 4.2, is recommended as a preliminary estimator. The estimator of the variance of $\delta_{\beta}(k)$ (see Equation 6.9) is recommended. Furthermore, these results confirm that $k = 4$ is a good choice if the object is to minimize the M.S.E. while preserving a small bias.

B. An Example

In recent years, the various theories of global tectonics have received widespread attention by earth scientists (Sykes et al. (1969)): included are the hypotheses of continental drift and sea-floor spreading. To account for the movement of ocean floors and the drift of continents, earth scientists conjecture that the earth's surface is composed of a number of so-called "plates." They theorize that these plates are in constant drift about the earth's surface, floating on the earth's mantel.
The movements of plates offer an explanation of seismic mechanisms. At the interface of two plates, typically at an ocean trench, it is believed that one plate is subducted into the earth's mantle underneath the opposing plate. This movement of one plate beneath another places tremendous strain on the underlying bedrock. Eventually this strain is relieved by the occurrence of an earthquake.

Figure 6.1 locates the Tonga trench in the region of 19° S. latitude and 173° W. longitude. At this trench, the Pacific plate meets the Australian plate. This is an area of intense seismic activity.

Sykes et al. (1969) give the observed hypocentral locations of 89 earthquakes which occurred in the vicinity of the Tonga trench between January 1965 and January 1966. The location of each hypocenter is an observation on 3 variables: depth in kilometers, latitude in degrees, and longitude in degrees. All 89 of these earthquakes occurred within the boundary of the dark rectangle in Figure 6.1.

Figure 6.2 is a computer enlargement of the area within the rectangle in Figure 6.1. Each point in this figure represents the epicenter of one of the 89 earthquakes. The diagonal line in

---

4 The hypocenter of an earthquake locates the focus of the earthquake beneath the earth's surface.

5 The epicenter of an earthquake locates the point on the earth's surface of most intense seismic activity.
Figure 6.1. Location of Tonga trench and relocated epicenters of teleseismically recorded earthquakes in the region between the Tonga and New Hebrides trenches (Sykes, Isacks, and Oliver (1969))
Figure 6.2 is the approximate location of the Tonga trench.

The notable feature of this data is that the earthquake hypocenters are very shallow near the Tonga trench, while they become progressively deeper as one moves further away from the trench in a perpendicular direction. This leads to an interesting statistical question.

Figure 6.2. Epicenters of 89 earthquakes which occurred in the Tonga trench region between January 1965 and January 1966
Can a curve be fit through the earthquake hypocenters, and if so, what is the implication of the curve with respect to the fault plane of the earthquakes? Our attempt to answer these questions brings us into the realm of errors-in-variables, for each of the three variables which locate a hypocenter i.e. depth, latitude, and longitude, is measured with error.

We begin the analysis by transforming the data to a computationally convenient form. Translating the earth's coordinate system from the Greenwich-Equator orientation to the 173°W. longitude-19°S. latitude orientation, converting from units in degrees to units in kilometers, and then rotating the axis through a 62°51' angle leads to the transformed data given in Table 7. The 62°51' rotation was chosen so that the rotated longitude axis would be parallel to the Tonga trench and perpendicular to the quadrant of the earthquake fault plane. Standard tables for converting measurement units from degrees to kilometers exist in many books. In this case we multiplied degrees longitude by 100.175 km/deg. and degrees latitude by 110.72 km/deg.
Table 7. Hypocentral location of 89 earthquakes in the Tonga trench region

<table>
<thead>
<tr>
<th>Depth</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>541.00</td>
<td>439.728</td>
<td>-389.174</td>
</tr>
<tr>
<td>503.00</td>
<td>501.511</td>
<td>-366.497</td>
</tr>
<tr>
<td>479.00</td>
<td>339.862</td>
<td>-342.445</td>
</tr>
<tr>
<td>571.00</td>
<td>309.169</td>
<td>-394.210</td>
</tr>
<tr>
<td>488.00</td>
<td>431.373</td>
<td>-455.378</td>
</tr>
<tr>
<td>60.00</td>
<td>198.927</td>
<td>-72.480</td>
</tr>
<tr>
<td>567.00</td>
<td>282.081</td>
<td>-388.656</td>
</tr>
<tr>
<td>210.00</td>
<td>221.900</td>
<td>-207.051</td>
</tr>
<tr>
<td>620.00</td>
<td>333.114</td>
<td>-450.604</td>
</tr>
<tr>
<td>261.00</td>
<td>253.774</td>
<td>-210.971</td>
</tr>
<tr>
<td>120.00</td>
<td>178.673</td>
<td>-37.836</td>
</tr>
<tr>
<td>0.00</td>
<td>147.376</td>
<td>15.912</td>
</tr>
<tr>
<td>575.00</td>
<td>470.056</td>
<td>-380.376</td>
</tr>
<tr>
<td>584.00</td>
<td>535.953</td>
<td>-365.720</td>
</tr>
<tr>
<td>0.00</td>
<td>170.739</td>
<td>-39.528</td>
</tr>
<tr>
<td>27.00</td>
<td>162.326</td>
<td>13.447</td>
</tr>
<tr>
<td>69.00</td>
<td>391.126</td>
<td>-49.343</td>
</tr>
<tr>
<td>68.00</td>
<td>382.796</td>
<td>-20.967</td>
</tr>
<tr>
<td>35.00</td>
<td>158.275</td>
<td>-7.768</td>
</tr>
<tr>
<td>8.00</td>
<td>119.648</td>
<td>-2.809</td>
</tr>
<tr>
<td>598.00</td>
<td>419.700</td>
<td>-425.338</td>
</tr>
<tr>
<td>498.00</td>
<td>339.160</td>
<td>-333.799</td>
</tr>
<tr>
<td>17.00</td>
<td>133.353</td>
<td>-61.077</td>
</tr>
<tr>
<td>474.00</td>
<td>375.360</td>
<td>-338.875</td>
</tr>
<tr>
<td>81.00</td>
<td>294.018</td>
<td>-85.633</td>
</tr>
<tr>
<td>523.00</td>
<td>286.173</td>
<td>-354.218</td>
</tr>
</tbody>
</table>

\(^a\) Measured in kilometers.

\(^b\) Translated, rotated longitude in kilometers.

\(^c\) Translated, rotated latitude in kilometers.
Table 7 Continued

<table>
<thead>
<tr>
<th>Depth a</th>
<th>Longitude b</th>
<th>Latitude c</th>
</tr>
</thead>
<tbody>
<tr>
<td>64.00</td>
<td>220.823</td>
<td>-22.975</td>
</tr>
<tr>
<td>570.00</td>
<td>231.073</td>
<td>-380.223</td>
</tr>
<tr>
<td>306.00</td>
<td>265.784</td>
<td>-292.623</td>
</tr>
<tr>
<td>590.00</td>
<td>503.710</td>
<td>-436.294</td>
</tr>
<tr>
<td>221.00</td>
<td>242.479</td>
<td>-215.636</td>
</tr>
<tr>
<td>288.00</td>
<td>389.524</td>
<td>-257.311</td>
</tr>
<tr>
<td>572.00</td>
<td>234.768</td>
<td>-370.446</td>
</tr>
<tr>
<td>595.00</td>
<td>293.123</td>
<td>-411.445</td>
</tr>
<tr>
<td>553.00</td>
<td>442.887</td>
<td>-412.322</td>
</tr>
<tr>
<td>506.00</td>
<td>377.717</td>
<td>-341.045</td>
</tr>
<tr>
<td>556.00</td>
<td>526.621</td>
<td>-400.903</td>
</tr>
<tr>
<td>266.00</td>
<td>474.425</td>
<td>-223.903</td>
</tr>
<tr>
<td>247.00</td>
<td>203.670</td>
<td>-232.162</td>
</tr>
<tr>
<td>614.00</td>
<td>505.365</td>
<td>-429.816</td>
</tr>
<tr>
<td>258.00</td>
<td>196.987</td>
<td>-228.834</td>
</tr>
<tr>
<td>30.00</td>
<td>126.483</td>
<td>29.964</td>
</tr>
<tr>
<td>560.00</td>
<td>376.400</td>
<td>-399.135</td>
</tr>
<tr>
<td>557.00</td>
<td>465.019</td>
<td>-402.097</td>
</tr>
<tr>
<td>608.00</td>
<td>431.452</td>
<td>-421.563</td>
</tr>
<tr>
<td>604.00</td>
<td>415.759</td>
<td>-427.359</td>
</tr>
<tr>
<td>577.00</td>
<td>344.487</td>
<td>-412.124</td>
</tr>
<tr>
<td>595.00</td>
<td>448.130</td>
<td>-415.262</td>
</tr>
<tr>
<td>559.00</td>
<td>436.875</td>
<td>-403.022</td>
</tr>
<tr>
<td>612.00</td>
<td>462.269</td>
<td>-428.275</td>
</tr>
<tr>
<td>409.00</td>
<td>455.681</td>
<td>-328.081</td>
</tr>
<tr>
<td>0.00</td>
<td>350.940</td>
<td>46.004</td>
</tr>
<tr>
<td>72.00</td>
<td>385.579</td>
<td>-9.408</td>
</tr>
<tr>
<td>540.00</td>
<td>457.635</td>
<td>-390.123</td>
</tr>
<tr>
<td>516.00</td>
<td>368.140</td>
<td>-368.472</td>
</tr>
<tr>
<td>585.00</td>
<td>470.505</td>
<td>-412.794</td>
</tr>
<tr>
<td>196.00</td>
<td>453.082</td>
<td>-177.433</td>
</tr>
<tr>
<td>570.00</td>
<td>413.066</td>
<td>-373.577</td>
</tr>
<tr>
<td>7.00</td>
<td>142.654</td>
<td>-11.276</td>
</tr>
<tr>
<td>199.00</td>
<td>350.882</td>
<td>-174.680</td>
</tr>
<tr>
<td>25.00</td>
<td>288.162</td>
<td>-149.430</td>
</tr>
<tr>
<td>133.00</td>
<td>268.151</td>
<td>-122.541</td>
</tr>
<tr>
<td>0.00</td>
<td>81.029</td>
<td>33.674</td>
</tr>
</tbody>
</table>
Table 7 Continued

<table>
<thead>
<tr>
<th>Depth</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>622.00</td>
<td>496.569</td>
<td>-432.075</td>
</tr>
<tr>
<td>284.00</td>
<td>288.168</td>
<td>-258.627</td>
</tr>
<tr>
<td>209.00</td>
<td>463.991</td>
<td>-169.587</td>
</tr>
<tr>
<td>400.00</td>
<td>426.795</td>
<td>-337.265</td>
</tr>
<tr>
<td>607.00</td>
<td>500.865</td>
<td>-418.616</td>
</tr>
<tr>
<td>97.00</td>
<td>276.497</td>
<td>-87.864</td>
</tr>
<tr>
<td>610.00</td>
<td>402.353</td>
<td>-437.611</td>
</tr>
<tr>
<td>620.00</td>
<td>446.301</td>
<td>-411.696</td>
</tr>
<tr>
<td>58.00</td>
<td>201.513</td>
<td>-4.732</td>
</tr>
<tr>
<td>366.00</td>
<td>337.442</td>
<td>-311.037</td>
</tr>
<tr>
<td>589.00</td>
<td>420.237</td>
<td>-392.415</td>
</tr>
<tr>
<td>31.00</td>
<td>392.332</td>
<td>-10.448</td>
</tr>
<tr>
<td>571.00</td>
<td>231.498</td>
<td>-366.495</td>
</tr>
<tr>
<td>208.00</td>
<td>219.970</td>
<td>-191.155</td>
</tr>
<tr>
<td>497.00</td>
<td>336.977</td>
<td>-341.673</td>
</tr>
<tr>
<td>569.00</td>
<td>406.972</td>
<td>-398.093</td>
</tr>
<tr>
<td>0.00</td>
<td>396.220</td>
<td>52.337</td>
</tr>
<tr>
<td>42.00</td>
<td>330.983</td>
<td>12.127</td>
</tr>
<tr>
<td>39.00</td>
<td>321.313</td>
<td>28.559</td>
</tr>
<tr>
<td>0.00</td>
<td>299.820</td>
<td>38.927</td>
</tr>
<tr>
<td>50.00</td>
<td>413.644</td>
<td>-42.299</td>
</tr>
<tr>
<td>253.00</td>
<td>134.928</td>
<td>-229.138</td>
</tr>
<tr>
<td>570.00</td>
<td>312.064</td>
<td>-363.456</td>
</tr>
<tr>
<td>0.00</td>
<td>362.304</td>
<td>52.958</td>
</tr>
<tr>
<td>0.00</td>
<td>297.575</td>
<td>60.291</td>
</tr>
<tr>
<td>409.00</td>
<td>287.606</td>
<td>-386.130</td>
</tr>
</tbody>
</table>

Figure 6.3 gives a computer plot of depth versus the translated, rotated latitude. We seek to fit a curve through these points.

Let the variable $y$ denote depth and the variable $x$, latitude. Many models are possible for the relation between $y$ and $x$ in Figure 6.3.
Figure 6.3. Depth (km.) of 89 earthquakes versus translated, rotated latitude (km.)
To illustrate the methods of this thesis, we shall estimate the quadratic functional relationship, namely

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 \]

where \( \beta' = (\beta_0, \beta_1, \beta_2) \).

To identify the model we require the error covariance matrix be known. Sykes et al. (1969) state that the standard error in a hypocentral location is ±5 to ±10 kilometers. We take this to mean the error covariance matrix is of the form \( \sigma^2 I \), with \( \sigma \in [5, 10] \).

Table 8 summarizes the steps in computing \( \hat{\sigma} \), the smallest root of the equation.

Table 8. Calculation of eigenvalue to estimate standard error

| Obtain the residuals \( R_1, R_2, \) and \( R_3 \) from regressing \( Y, X, \) and \( (X^2 - 1.0) \) on a column of 1's. |
| Obtain the sum of squares and cross products matrix for the residuals: |

\[
\overline{M}_{RR} = (89)^{-1} \sum_{t=1}^{89} \begin{bmatrix} R_{1t} \\ R_{2t} \\ R_{3t} \end{bmatrix} \begin{bmatrix} R_{1t} & R_{2t} & R_{3t} \end{bmatrix}
\]
Table 8 Continued

\[
\begin{pmatrix}
.56339 \times 10^5 & - .39985 \times 10^5 & .16729 \times 10^8 \\
.29812 \times 10^5 & - .12044 \times 10^8 & \\
.52231 \times 10^6 & \\
\end{pmatrix}
\]

Obtain $\hat{F}_R$ by eliminating the 2nd row and column from $\hat{F}$:

\[
\hat{F}_R = \begin{bmatrix}
.1 \times 10^1 & 0 & 0 \\
.1 \times 10^1 & - .48246 \times 10^3 & \\
\text{sym.} & .35201 \times 10^6 & \\
\end{bmatrix}
\]

Solve the generalized eigenvalue problem

\[
|\overline{M}_{RR} - \alpha \hat{F}_R| = 0
\]

for the smallest root, $\hat{\alpha} = 487.88$. Calculate $\sqrt{\hat{\alpha}} = 22.088$.

\[
|\overline{M} - \alpha \hat{F}| = 0,
\]

where

\[
\overline{M} = (89)^{-1} \sum_{t=1}^{89} Z_t' Z_t,
\]

\[
Z_t = (Y_t, 1.0, X_t, X_t^2 - 1.0),
\]

and
By the results of Chapter 3, for normally distributed measurement errors with covariance matrix \( \sigma^2 I \), \( \sqrt{\hat{\sigma}} = 20.088 \) furnishes an estimate of \( \sigma \).

Since 20.88 is notably outside the range of standard errors given by Sykes, we conclude that either the model is incorrect or the standard error is much larger than commonly thought. To proceed with the example, we assume the model is correctly specified and treat \( \sigma = 20.088 \) as known.

The ordinary least squares estimator is easily computed:

\[
\hat{\beta}_{OLS} = \left( \sum_{t=1}^{89} X_t \right)^{-1} \left( \begin{array}{c} 1 \\ X_t \\ X_t^2 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ X_t \\ X_t^2 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ X_t \\ X_t^2 \end{array} \right)^{-1} \left( \begin{array}{c} Y_t \\ X_t \\ X_t^2 \end{array} \right)
\]

\[
= \left\{ \begin{array}{ccc}
.1010^{1} & -.241231.10^{3} & .880039.10^{5} \\
.880039.10^{5} & -.332730.10^{8} & .129679.10^{11} \\
\text{sym.} & \end{array} \right\}^{-1}
\]
This is used as a preliminary estimator in an iterative procedure.

Using \( \hat{\beta}_{OLS} \), we compute the iterative estimator \( 3\beta^*(4) \) defined by Equation 6.7. This estimator converged to 6 decimal places within 10 iterations. The computations involved in the final iteration are given in Table 9. We find that

\[
3\beta^*(4) = \begin{bmatrix} 26.256 \\ -0.53431 \\ 0.0020588 \end{bmatrix}.
\]

Note that \( 3\beta^*_2(4) > \hat{\beta}_{OLS(2)} \) and \( 3\beta^*_1(4) < \hat{\beta}_{OLS(1)} \).

An estimator of the covariance matrix of \( \sqrt{N}(3\beta^*(4) - \beta) \) is given in Table 9 by \( (\bar{M}_{ww}^* - \Delta)^{-1} \). The estimated standard errors of \( 3\beta^*_0(4) \), \( 3\beta^*_1(4) \), and \( 3\beta^*_2(4) \) are 4.7670, 0.082637, and 0.00021799 respectively.
Table 9. Calculation of $\beta^* (4)$ for the tenth iteration

Obtain $\hat{x}_t$ by solving for the roots of

$$o = (Y_t - \beta_0 - \beta_1 x - \beta_2 x^2)(\beta_1 + 2x\beta_2) + (X_t - x).$$

Obtain $\hat{M}^*_{\text{ww}}$ and $\hat{M}^*_{\text{xx}}$:

$$\hat{M}^*_{\text{ww}} = (89)^{-1} \sum_{t=1}^{89} \sigma_{v_t}^{-2} \begin{bmatrix} \frac{1}{\hat{x}_t} \\ \hat{x}_t \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{x}_t} \\ \hat{x}_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_{v_t}^{-2}} \end{bmatrix}$$

$$= \begin{bmatrix} .78996 \cdot 10^{-3} \\ -.10035 \cdot 10^0 \\ .34739 \cdot 10^2 \\ .12455 \cdot 10^5 \\ .47036 \cdot 10^7 \\ \text{sym.} \end{bmatrix}$$

and

$$\hat{M}^*_{\text{xx}} = (89)^{-1} \sum_{t=1}^{89} \sigma_{v_t}^{-2} \begin{bmatrix} \frac{1}{\hat{x}_t} \\ \hat{x}_t \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{x}_t} \\ \hat{x}_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_{v_t}^{-2}} \end{bmatrix}$$

$$= \begin{bmatrix} .18772 \cdot 10^{-8} \\ .14529 \cdot 10^{-6} \\ -.91553 \cdot 10^{-4} \end{bmatrix}.$$

Obtain the residuals $R_{1t}$ and $R_{2t}$ by setting
Table 9 Continued

\[ R_{1t} = \hat{x}_t - \frac{89}{\sum t=1 \sigma_t^2} \hat{x}_t / \frac{89}{\sum t=1 \sigma_t^2} \]

and

\[ R_{2t} = \left( \hat{x}_t^2 - \sigma_t^2 \right) - \frac{89}{\sum t=1 \sigma_t^2} \left( \hat{x}_t^2 - \sigma_t^2 \right) / \frac{89}{\sum t=1 \sigma_t^2} \]

Obtain the sum of squares and cross products matrix for the residuals:

\[
\hat{M}_{RR}^* = (89)^{-1} \sum_{t=1}^{89} \begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix}
\]

\[
= \begin{bmatrix} .22209.10^2 & -.80422.10^4 \\ \text{sym.} & .31759.10^7 \end{bmatrix}
\]

Obtain \( \hat{\Sigma}_{R}^* \) defined by

\[
\hat{\Sigma}_{R}^* = (89)^{-1} \sum_{t=1}^{89} \sigma_t^2 \hat{\sigma}_t^2 \begin{bmatrix} 1 & 2\hat{x}_t \\ 2\hat{x}_t & 4\hat{x}_t^2 \end{bmatrix}
\]

\[
= \begin{bmatrix} .21664 & -.20281.10^2 \\ \text{sym.} & .14698.10^5 \end{bmatrix}
\]
Solve the generalized eigenvalue problem

\[ |\overline{M}_{RR}^* - \lambda \overline{C}_{RR}^*| = 0 \]

for the smallest root, \( \hat{\lambda} = 8.8243 \).

Compute

\[ (\overline{M}_{WW}^* - \Delta)^{-1} = \left( \overline{M}_{WW}^* - (1 - \frac{4}{89}) (89)^{-1} \sum_{t=1}^{89} \sigma_{v_t} \sigma_{t}^2 \right)^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 2\hat{\lambda}_t \\ \text{sym.} & 4\hat{\lambda}_t^2 \end{pmatrix} \]

\[ = \begin{pmatrix} 0.20225 \times 10^4 & 0.93869 \times 10^1 & 0.99099 \times 10^{-2} \\ 0.60777 \times 10^0 & 0.15421 \times 10^{-2} \\ \text{sym.} & 0.42292 \times 10^{-5} \end{pmatrix} \]

Calculate

\[ z(\Delta \beta)^* (4) = (\overline{M}_{WW}^* - \Delta)^{-1} \overline{M}_{xv}^* = \begin{pmatrix} 0.42531 \times 10^{-5} \\ -0.35264 \times 10^{-7} \\ -0.14454 \times 10^{-9} \end{pmatrix} \]

Consider the estimators \( \overline{\beta} + v z(\Delta \beta)^* (4) \), where \( v \) successively equals 1.0, .5, .25, .125, ... .
for \( \nu = 1.0, .5, .25, .125 \), but

\[
Q(\bar{\beta} + \nu_3 (\Delta \beta)^* (4), \hat{x}_1, \ldots, \hat{x}_{89}) < Q(\bar{\beta}, \hat{x}_1, \ldots, \hat{x}_{89})
\]

where \( \nu = .0625 \). Thus calculate

\[
3\beta^* (4) = \beta + (.0625) \Delta \beta^* (4) = \begin{bmatrix} 26.256 \\ -.53431 \\ .0020588 \end{bmatrix}
\]

\( a \) The function \( Q \) is defined by Equation 4.22; \( \hat{x}_t \) is formally
defined by Equation 4.28; and \( \beta \) is a preliminary estimator of \( \beta \),
in this case the value of \( 3\beta^* (4) \) after nine iterations:

\[
3\beta^* (4) = \begin{bmatrix} 26.256 \\ -.53431 \\ .0020588 \end{bmatrix}
\]
Figures 6.4 and 6.5 give graphical representations of the errors-in-variables fit and the ordinary least squares fit respectively. Again note that the errors-in-variables curve has the steeper slope.

A geological expert would be required to give this analysis definitive interpretation. However, some observations are possible.

Many functional models could be conceived to explain this data. The quadratic model is a reasonable choice, though perhaps not the best.

Ordinary least squares is not an appropriate estimation procedure in this case. An errors-in-variables analysis is preferred since each coordinate of a hypocentral location is observed with error.

The estimated standard error is 20.088 km., and from Figure 6.4 it is seen that at least 4 hypocenters deviate by several standard errors from the fitted curve. It is possible that these deviate observations are due to the existence of a more complex fault system than hypothesized by the quadratic model.
Figure 6.4. Estimated errors-in-variables fit through 89 earthquake hypocenters
Figure 6.5. Ordinary least squares fit through 89 earthquake hypocenters
The nonlinear errors-in-variables model was investigated with emphasis on the quadratic functional relationship. The quadratic model is defined by the exact mathematical relationship
\[ y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 \]
for \( t = 1, 2, \ldots \). The elements \( y_t \) and \( x_t \) are not directly observed. Rather, \( Y_t \) and \( X_t \) are observed for \( t = 1, \ldots, N \), where
\[ Y_t = y_t + e_t \]
\[ X_t = x_t + u_t \]
and where the random variables \((e_t, u_t)\) denote errors of measurement. It is assumed that the errors \( \{(e_t, u_t)\}_{t=1}^{\infty} \) are mutually independent; that \( E(e_t) = 0 \) and \( E(u_t) = 0 \); that the finite covariance matrices
\[
E\left( \begin{bmatrix} e_t \\ u_t \end{bmatrix} \begin{bmatrix} e_t \\ u_t \end{bmatrix} \right) = \begin{bmatrix} \sigma_e(t)^2 & \sigma_{eu}(t) \\ \sigma_{ue}(t) & \sigma_u(t)^2 \end{bmatrix}
\]
are known for each \( t = 1, 2, \ldots, N \); and that \( \{x_t\}_{t=1}^{\infty} \) is a sequence of fixed constants.

An estimator of \( \beta^t = (\beta_0, \beta_1, \beta_2) \), constructed by analogy to that for the linear functional relationship, is given by

\[
\hat{\beta} = \left\{ N^{-1} \sum_{t=1}^{N} W_t' W_t - \hat{\alpha} N^{-1} \sum_{t=1}^{N} \hat{\Phi}_{fe(t)} \right\}^{-1}.
\]

\[
\{ N^{-1} \sum_{t=1}^{N} W_t' Y_t - \hat{\alpha} N^{-1} \sum_{t=1}^{N} \hat{\Phi}_{fe(t)} \},
\]

where \( W_t = (1, X_t, X_t^2 - \sigma_u^2) \),

\[
\hat{\Phi}_t = \begin{bmatrix} \hat{\Phi}_{ee(t)} & \hat{\Phi}_{ef(t)} \\ \hat{\Phi}_{fe(t)} & \hat{\Phi}_{ff(t)} \end{bmatrix}
\]

is an estimator of the covariance matrix of \( (Y_t, W_t) \), and \( \hat{\alpha} \) is the smallest root of the equation

\[
| N^{-1} \sum_{t=1}^{N} \begin{bmatrix} Y_t \\ W_t' \end{bmatrix} - \hat{\alpha} N^{-1} \sum_{t=1}^{N} \hat{\Phi}_t | = 0.
\]
Under minimal assumptions \( \hat{\beta} \) is consistent with error \( O_p(N^{-1/2}) \). Furthermore, \( \sqrt{N}(\hat{\beta} - \beta) \) is asymptotically normal with zero mean and approximate covariance matrix

\[
\{N^{-1} \sum_{t=1}^{N} w_t' w_t\}^{-1} \cdot E\{N^{-1} \sum_{t=1}^{N} \phi_t \phi_t'\} \cdot \{N^{-1} \sum_{t=1}^{N} w_t' w_t\}^{-1},
\]

where

\[
w_t = (1, x_t, x_t^2),
\]

\[
\phi_t = W_t v_t = \left[ \frac{v_t^2 - \bar{\sigma}_v^2}{\bar{\sigma}_v^2} \right] \frac{\phi_{fv}}{\sigma_v} - \frac{\phi_{fv(t)}}{\sigma_v},
\]

and

\[
v_t = y_t - W_t \beta.
\]

A consistent estimator of this covariance matrix is constructed.

Under the more restrictive assumption that the error variances decrease with increasing sample size, the consistency of the ordinary least squares estimator,

\[
\hat{\beta}_{OLS} = \left\{b^{-1} \sum_{t=1}^{N} 1 \begin{bmatrix} X_t \\ X_t^2 \end{bmatrix} \begin{bmatrix} 1 & X_t & X_t^2 \end{bmatrix}^{-1} \right\},
\]

\[
\{b^{-1} \sum_{t=1}^{N} 1 \begin{bmatrix} X_t \\ X_t^2 \end{bmatrix} \begin{bmatrix} y_t \end{bmatrix} \right\},
\]
is established. In this expression, \( b_N \) denotes the number of observations in the \( N \)th of a sequence of experiments; the error variances are of order \( a_N \); and \( N = b_N/a_N \).

Given these assumptions, \( \beta \) has error \( O(N^{-1/2}) \) and \( \sqrt{N}(\hat{\beta} - \beta) \) is asymptotically normal with zero mean and covariance matrix

\[
\{ b_{-1}^{N} \Sigma \frac{w'_tw_t}{N} \}^{-1} \cdot \{ b_{-1}^{N} \Sigma \frac{a_{-1}^{2} \sigma_t^{2} w'_tw_t}{N} \cdot \{ b_{-1}^{N} \Sigma \frac{w'_tw_t}{N} \}^{-1}.
\]

Using \( \hat{\beta} \) or \( \hat{\beta}_{OLS} \) as a preliminary estimator two iterative estimators, \( \hat{\beta} \) and \( \hat{\beta}^* \), of \( \beta \) are constructed. Given the assumption of decreasing error variances, both \( \sqrt{N}(\hat{\beta} - \beta) \) and \( \sqrt{N}(\hat{\beta}^* - \beta) \) are asymptotically normal with zero mean and covariance matrix

\[
\{ N^{-1} \Sigma \frac{\sigma_{-2}^{2} w'_tw_t}{t} \}^{-1}
\]

Furthermore, the asymptotic variances of \( \hat{\beta} \) and \( \hat{\beta}^* \) are smaller than those of \( \hat{\beta} \) or \( \hat{\beta}_{OLS} \).

The iterative estimators are defined by \( \hat{\beta} = \bar{\beta} + (\Delta \hat{\beta}) \) and \( \hat{\beta}^* = \bar{\beta} + (\Delta \hat{\beta})^* \), where \( \bar{\beta} \) is a preliminary estimator,
\[(\Delta \beta) = \{b^{-1} \sum_{t=1}^{bN} \frac{\hat{\sigma}_t^{-2}}{v_t} \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} \}^{-1} \{1 \hat{x}_t \hat{x}_t^2 \} \].

\[(\Delta \beta)^* = \{b^{-1} \sum_{t=1}^{bN} \frac{\hat{\sigma}_t^{-2}}{v_t} \begin{bmatrix} 1 \\ \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} \}^{-1} \{1 \hat{x}_t \hat{x}_t^2 \} \].

\[\sigma_v^2 = \sigma^2 e - 2d_t \sigma e u + d_t^2 \sigma u^2,\]

\[d_t = \beta_1 + 2\hat{x}_t \beta_2,\]

\[\hat{e}_t = Y_t - \beta_0 - \beta_1 \hat{x}_t - \beta_2 \hat{x}_t^2,\]

\[\hat{u}_t = X_t - \hat{x}_t,\]

and

\[\hat{v}_t = Y_t - \beta_0 - \beta_1 X_t - \beta_2 X_t^2.\]
$\hat{x}_t$ is a preliminary estimator of $x_t$ and is defined by Equation 4.28. A consistent estimator of the asymptotic covariance matrix of $\beta^*$ and $\hat{\beta}$ is also given.

These estimators, among others, were considered in a Monte Carlo study. Preliminary work indicated a departure from normality in the sample distribution functions of the estimators. This was characterized by a small number of deviate observations.

To improve the small sample properties of the estimators, each was adjusted in the manner of Fuller (1971) and Booth (1973). Adjusted estimators of $\hat{\beta}$, $\hat{\beta}$, and $\beta^*$ and estimators of their covariance matrices were given by Equations 6.1-6.9.

Each of the adjusted estimators is asymptotically equivalent to its corresponding unadjusted estimator, provided the conditions under which the unadjusted estimator has a limiting distribution are met.

Based on the results of the Monte Carlo study, the estimator $\beta^*_o(4)$ is recommended as a final estimator, and $\hat{\beta}_{OLS}$ is recommended as a preliminary estimator. The estimator of the variance of $\beta^*_o(4)$, given by Equation 6.9, is recommended.

The general errors-in-variables model investigated is specified by the relationship

$$y_t = g(x_t; \beta)$$
for \( t = 1, 2, \ldots \), where \( x_t \) is a fixed \((1 \times q)\) vector for each \( t \) and \( \beta \) is an unknown \((p \times 1)\) vector. In the \( N \)th of a sequence of experiments \( Y_t \) and \( X_t \), defined by

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

are observed for \( t = 1, \ldots, b_N \).

By analogy with the quadratic model, two iterative estimators, \( \hat{\beta} \) and \( \beta^* \), of \( \beta \) are constructed. These are defined by Equations 5.17 and 5.43 respectively. Assuming the error variances decrease with increasing sample size, \( \sqrt{N}(\hat{\beta} - \beta) \) and \( \sqrt{N}(\beta^* - \beta) \) are asymptotically normally distributed with zero mean and covariance matrix

\[
\{N^{-1} \sum_{t=1}^{b_N} \sigma^{-2} \frac{g_\beta(x_t; \beta)}{\nabla g_\beta(x_t; \beta)} \}^{-1} ,
\]

where \( g_\beta(x_t; \beta) \) is the first derivative of \( g \) with respect to \( \beta \) evaluated at \((x_t; \beta)\); \( \sigma^{-2} = (1, -\gamma_t) \frac{1}{\sigma} (1, -\gamma_t)' \), and \( \gamma_t = g_x(x_t; \beta) \) is the first derivative of \( g \) with respect to \( x \) evaluated at \((x_t; \beta)\).

A consistent estimator of the covariance matrix of \( \hat{\beta} \) and \( \beta^* \) is given by
where \( \bar{\beta} \) is a preliminary estimator of \( \beta \) and \( \bar{x}_t \) is a preliminary estimator of \( x_t \) defined by Equation 5.5.

Following the Monte Carlo results for the quadratic model, an adjusted form of \( \beta^* \) is recommended for the general model. The adjusted estimator is defined by

\[
\beta^*(k) = \left[ b^{-1} \sum_{N} \sigma^{-2} \frac{g_{\beta}(\bar{x}_t; \bar{\beta}) g_{\beta}'(\bar{x}_t; \bar{\beta}) - \Delta}{v_t} \right]^{-1}.
\]

where \( k > 0 \),

\[
\Delta = \begin{cases} 
(1 - \frac{k}{N}) b^{-1} \sum_{N} \sigma^{-2} \frac{g_{\beta x}(\bar{x}_t; \bar{\beta}) E(\delta_1, \delta_2') g_{\beta x}'(\bar{x}_t; \bar{\beta})}{v_t} & \text{for } \lambda > 1 \\
(\lambda - \frac{k}{N}) b^{-1} \sum_{N} \sigma^{-2} \frac{g_{\beta x}(\bar{x}_t; \bar{\beta}) E(\delta_1, \delta_2') g_{\beta x}'(\bar{x}_t; \bar{\beta})}{v_t} & \text{for } \lambda \leq 1,
\end{cases}
\]

and \( \lambda \) is the smallest root of the determinantal equation
\[
\left( \begin{array}{c}
\Phi \\
\phi \\
\psi \\
\phi \\
\psi
\end{array} \right) = \mathcal{G} = \left( \begin{array}{c}
\mathcal{G} \\
\mathcal{G} \\
\mathcal{G} \\
\mathcal{G} \\
\mathcal{G}
\end{array} \right) \mathcal{S}
\]

\( (g : \mathcal{S}) \mathcal{S} = \mathcal{G} \)

\[ I - \{ \phi_0 \phi_1^2 \phi_2^3 + \phi_3^4 \phi_4^5 + \phi_5^6 + \phi_7^8 + \phi_9^9 \} \]

\[ \cdot (\phi_0 \phi_1 \phi_2 \phi_3 \phi_4 \phi_5 \phi_6 \phi_7 \phi_8 \phi_9) \]
A consistent estimator of the variance of $\beta^*(k)$ is given by

$$
\hat{\Phi}^{-1} = \begin{bmatrix}
\sigma_{ee} & \Sigma_{eu} \\
\Sigma_{ue} & \sigma_{uu}
\end{bmatrix}
$$

To minimize the mean square error of the estimator, while preserving a small bias, $k = 4$ is recommended.

Finally, under the assumption of decreasing error variances, $\beta^*(k)$ is asymptotically equivalent to $\beta^*$. 

VIII. REFERENCES


Bartlett, M.S. "The Fitting of Straight Lines if Both Variables Are Subject to Error." Biometrics 5 (1949): 207-212.


Fedorov, V.V. "Regression Problems with Controllable Variables Subject to Error." Biometrika 61 (1974): 49-56.


Nair, K. R., and Banerjee, K. S. "A Note on Fitting of Straight Lines if Both Variables Are Subject to Error." *Sankhyā* 6 (1942): 331.


IX. ACKNOWLEDGEMENTS

Several people have made major contributions to this thesis and deserve recognition.

My most sincere thanks are due to Professor Wayne A. Fuller. His expert guidance and genuine interest in this research provided a truly enjoyable, as well as an instructive, experience.

I am grateful to the members of my graduate committee: Professors Chien-Pai Han, Glen Meeden, H.A. David, and J.C. Mathews.

Also, I am indebted to Joe Reese for his help and interest in the earthquake example.

Finally, many thanks go to my wife and daughter. Their efforts, their patience, and their understanding through four years of graduate school can not be underestimated.
X. APPENDIX A: ADDITIONAL ESTIMATORS FOR THE QUADRATIC FUNCTIONAL RELATIONSHIP

A. A Method of Moments Estimator

In this section we give a method of moments estimator for the quadratic functional relationship as specified by Equations 3.1a, 3.1b.

Define the (3 x 3) matrix $\hat{H}$ and the (3 x 1) matrix $\hat{G}$ as follows:

\[
\hat{H} = \begin{bmatrix}
1 & N^{-1} \sum_{t=1}^{N} X_t & N^{-1} \sum_{t=1}^{N} X_t^2 - \sigma_u^2 \\
N^{-1} \sum_{t=1}^{N} X_t^2 - \sigma_u^2 & N^{-1} \sum_{t=1}^{N} (X_t^3 - 3\sigma_u^2 X_t) \\
\text{symmetric} & N^{-1} \sum_{t=1}^{N} (X_t^4 - 6\sigma_u^2 X_t^2 + 3\sigma_u^4)
\end{bmatrix}
\]

and

\[
\hat{G} = \begin{bmatrix}
N^{-1} \sum_{t=1}^{N} Y_t \\
N^{-1} \sum_{t=1}^{N} X_t Y_t - \sigma_{eu} \\
N^{-1} \sum_{t=1}^{N} (X_t^2 Y_t - \sigma_u^2 Y_t - 2\sigma_{eu} X_t)
\end{bmatrix}
\]

Consider the moment estimator

\[
\hat{\beta}_M = \hat{H}^{-1} \hat{G}
\]  
(10.1).
To establish the consistency of this estimator we make the following assumptions:

**Assumption 10.1**

The errors of measurement \((e_t, u_t)\) are independently, identically distributed as a \(N(0, \Psi)\).

**Assumption 10.2**

The error covariance matrix

\[
\Psi = \begin{bmatrix}
\sigma^2_e & \sigma_{eu} \\
\sigma_{ue} & \sigma^2_u
\end{bmatrix}
\]

is known.

**Assumption 10.3**

The matrix \(\overline{m}_{ww}\) as defined by Equation 3.7 is positive definite for \(N > 3\).

**Assumption 10.4**

The mean \(\bar{x}(t) = N^{-1} \sum_{t=1}^{N} x_t^r\) converges as \(N \rightarrow \infty\) for \(r = 1, 2, \ldots, 6\).

We are now able to state and prove the following lemma:

**Lemma 10.1**

Under Model 3.1a, 3.1b and Assumptions 10.1 through 10.4, the
error in $\hat{\beta}_M$ is $O_p(N^{-1/2})$, i.e. $\hat{\beta}_M - \beta = O_p(N^{-1/2})$.

Proof:

By Assumption 10.1 it is easily seen that

\[
E\{N^{-1} \sum_{t=1}^{N} X_t\} = N^{-1} \sum_{t=1}^{N} x_t;
\]

\[
E\{N^{-1} \sum_{t=1}^{N} X_t^2\} = N^{-1} \sum_{t=1}^{N} x_t^2 + \sigma_u^2;
\]

\[
E\{N^{-1} \sum_{t=1}^{N} X_t^3\} = N^{-1} \sum_{t=1}^{N} (x_t^3 + 3\sigma_u^2 x_t^2);
\]

\[
E\{N^{-1} \sum_{t=1}^{N} X_t^4\} = N^{-1} \sum_{t=1}^{N} (x_t^4 + 6\sigma_u^2 x_t^2 + 3\sigma_u^4).
\]

Since the normal distribution possesses finite moments, it follows immediately from Assumption 10.4, Theorem 2.2, and Equations 10.2 that

\[
\hat{H} = E(\hat{H}) + O_p(N^{-1/2})
\]

\[
= N^{-1} \sum_{t=1}^{N} \left[\begin{array}{c} \frac{1}{x_t} \\
\frac{x_t}{x_t^2} \\
\frac{x_t^2}{x_t^2} \end{array}\right] (1 x_t x_t^2) + O_p(N^{-1/2})
\]

\[
= \frac{m}{ww} + O_p(N^{-1/2}).
\]
Then, by Assumption 10.4, Theorem 2.2, and Equations 10.4 we can write

\[ \hat{G} = E(\hat{G}) + O_p(N^{-1/2}) \]
\[ = \bar{m}_{ww} \beta + O_p(N^{-1/2}). \] (10.5)

The lemma follows from Equations 10.3 and 10.5. \textit{Q. E. D.}

It is worth noting that this estimator may be used as a preliminary estimator in an iterative procedure. However, from an intuitive standpoint, we expect the estimator \( \hat{\beta} \) of Chapter 3 to perform better than \( \hat{\beta}_M \) since \( \hat{\beta} \) employs knowledge of \( \sigma^2 \) and \( \hat{\beta}_M \) does not.

B. A Pseudo Instrumental Variable Estimator

This section and the next are devoted to two additional iterative estimation procedures for the quadratic functional relationship given by Equations 4.1, 4.2. Under the assumptions of Chapter 4, each of these estimators is asymptotically equivalent to the estimators \( \hat{\beta} \) and \( \beta^* \) of Chapter 4. These estimators are presented here, apart from the main body of the thesis, due to their inferior small sample performance as indicated in the Monte Carlo study.
The first estimator, say $\hat{\beta}_{IV}$, is a pseudo-instrumental variable estimator and, in the notation of Chapter 4, is defined by

$$
\hat{\beta}_{IV} = \left\{ b \sum_{t=1}^{bN} \frac{1}{\sum_{t=1}^{bN} \frac{1}{\sigma^2_{\epsilon_t}}} \begin{bmatrix} \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} \right\}^{-1} \left( 1 \begin{bmatrix} X_t \\ X_t^2 \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \hat{x}_t^2 \end{bmatrix} \right) - 2 \frac{1}{\sum_{t=1}^{bN} \frac{1}{\sigma^2_{\epsilon_t}}} \left( X_t \hat{x}_t - \hat{\sigma}^2_{\epsilon_t} \right)
$$

(10.6)

where $\hat{x}_t$ is defined by Equation 4.28

$$
\hat{x}_t = \hat{x}_t^2 = \hat{x}_t^2 - 2d_t \hat{x}_t^2 + d_t^2 \hat{x}_t^2,
$$

and $\hat{\sigma}^2_{\epsilon_t}$ is a preliminary estimator of $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$. The asymptotic properties of $\hat{\beta}_{IV}$ are summarized in Theorem 10.1. The proof of this theorem is very similar to other proofs in this thesis, and is omitted for the sake of brevity.
Theorem 10.1

Given Assumptions 4.1b, 4.2, 4.4, 4.5, 4.6, 4.7b, 4.8, 4.9, and 4.10, then

\[ \sqrt{N}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, \lim_{N \to \infty} \left\{ b^{-1} \sum_{t=1}^{N} \frac{1}{a_{N}^{\sigma_{v_t}^2}} \begin{bmatrix} 1 \\ x_t \\ x_t^2 \end{bmatrix} \right\}^{-1}) \]

where \( \sigma^2_{v_t} = \sigma_e^2 - 2\gamma_t \sigma_{eu} + \gamma_t^2 \sigma_u^2 \) and \( \gamma_t = \beta_1 + 2x_t \beta_2 \).

C. A Revised Root Type Estimator

The estimator to be given here, say \( \hat{\beta}^* \), is a root type estimator and, in the notation of Chapter 4, is defined by

\[ \hat{\beta}^* = \left\{ b^{-1} \sum_{t=1}^{N} \frac{1}{a_{N}^{\sigma_{v_t}^2}} W_t W_t - \alpha^* b^{-1} \sum_{t=1}^{N} \frac{\hat{\eta}_t}{a_{N}^{\sigma_{v_t}^2}} \frac{\hat{\eta}_t}{\hat{\eta}_t} \right\}^{-1} \]

\[ = \left\{ b^{-1} \sum_{t=1}^{N} \frac{1}{a_{N}^{\sigma_{v_t}^2}} W_t Y_t - \alpha^* b^{-1} \sum_{t=1}^{N} \frac{\hat{\eta}_t}{a_{N}^{\sigma_{v_t}^2}} \frac{\hat{\eta}_t}{\hat{\eta}_t} \right\}^{-1} \]

where \( \alpha^* \) is the smallest root of
\[ b^{-1} \sum_{N_{t=1}}^{b_N} \frac{1}{a^{-1} \sum_{N_{v_{t}}}^{b_N}} Z_t Z_t - \alpha b^{-1} \sum_{N_{t=1}}^{b_N} \frac{\hat{\gamma}_t}{a^{-1} \sum_{N_{v_{t}}}^{b_N}} \Phi(t) = 0, \]

\[ d_t = \beta_1 + 2 \hat{x}_t \beta_2, \]

\[ \hat{\beta}_t = \sigma^2 - 2 d_t \sigma_e + d_t \sigma^2, \]

\[ \hat{x}_t \] is defined by Equation 4.28,

\[ \beta \] is a preliminary estimator of \( \beta, \)

\[ \Phi_t = \begin{bmatrix} \sigma_e^2 & \sigma_e \Phi(\tau) \\ \sigma_e \Phi(\tau) & \sigma(\Phi(\tau) \Phi(\tau)) \end{bmatrix} = \begin{bmatrix} \sigma_e^2 & 0 & 2X_t \sigma e \\ 0 & 0 & 0 \\ 2X_t \sigma^2 \\ 4\sigma^2 (X_t^2 - \sigma_u^2) + 2\sigma^4 \end{bmatrix}, \]

\[ W_t = [1, X_t, (X_t^2 - \sigma_u^2)], \]

\[ Z_t = [Y_t, W_t], \]

\[ \hat{\gamma}_t = \beta^t M_t \beta / \sigma^2 \]

\[ \sigma^2 \]

and \[ M_t = Z_t^t Z_t \]

was first suggested by Booth (1973) who worked with the linear functional relationship with known, but unequal error covariance matrices.
The asymptotic properties of \( \hat{\beta}^* \) are summarized in Theorem 10.2. The proof of this theorem is deleted for the sake of brevity.

**Theorem 10.2**

Given Assumptions 4.1b, 4.2, 4.4, 4.5, 4.6, 4.7b, 4.8, 4.9, and 4.10, then

\[
\sqrt{N}(\hat{\beta}^* - \beta) \xrightarrow{d} N(0, \lim_{N \to \infty} \{b^{-1} \sum_{t=1}^{bN} \frac{1}{aN} \sigma_{v_t}^{-1} w'_t w_t\}^{-1})
\]

where

\[
\sigma_{v_t}^2 = \sigma_e^2 - 2 \gamma_t \sigma_e u_{t} + \gamma_t^2 \sigma_u^2
\]

and

\[
\gamma_t = \beta_1 + 2x_t \beta_2.
\]
XI. APPENDIX B: THE THEORY OF CONTROLLED VARIABLES FOR A POLYNOMIAL RELATIONSHIP

This appendix is concerned with the polynomial errors-in-variables model

\[ y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \ldots + \beta_m x_t^m, \quad (11.1) \]

where we observe

\[ Y_t = y_t + e_t \]
\[ X_t = x_t + u_t \quad (11.2) \]

for \( t = 1, \ldots, N \) (\( N > m+1 \)), and where the observations \( \{X_t\}_{t=1}^N \) are fixed according to a preassigned schedule. Both the true value \( x_t \) and the error \( u_t \) are random variables, with the fixed observation \( X_t \) satisfying \( X_t = x_t + u_t \). That is,

\[ \text{Cov}(x_t, u_t) = -1. \]

\( X \) is called a controlled variable for this situation.

This model was first conceived by Berkson (1950) who considered the linear model, i.e. \( m = 1 \). Latter, Scheffe (1958) also studied the linear model. Each showed that simple least squares provides an
unbiased estimate of the structural parameters.

Geary (1953) considered the cubic model, i.e. \( m = 3 \), and showed that the Berkson results do not, in general, extend to polynomial models. Specifically, he demonstrated that the ordinary least squares estimators of the coefficients of \( x_3^3 \) and \( x_2^2 \) are unbiased, while the estimated coefficient of \( x_1^1 \) and the estimated intercept are biased. Here, we extend the results of Berkson and Geary to the general polynomial model given by Equation 11.1.

Throughout the work of this appendix we make the following assumptions:

Assumption 11.1

The errors of measurement \( (e_t, u_t) \) are independent, identically distributed with zero mean and finite \( m \)th moments.

Assumption 11.2

The distribution function of \( (e_1, u_1) \) is symmetric about the mean.

Under these assumptions, we consider the ordinary least squares estimator, namely
\[ \beta_{\text{OLS}} = \left\{ \frac{1}{N} \Sigma_{t=1}^{N} \begin{bmatrix} x_t \\ \vdots \\ x_{m_t} \end{bmatrix} \begin{bmatrix} 1 & x_t & \ldots & x_{m_t} \end{bmatrix} \right\} \left\{ \frac{1}{N} \Sigma_{t=1}^{N} \begin{bmatrix} 1 \\ x_t \\ \vdots \\ x_{m_t} \end{bmatrix} \right\} \]

(11.3)

where \( \beta' = (\beta_0, \beta_1, \ldots, \beta_m) \).

From Equation 11.2 we obtain

\[ x_t^p = (X_t - u_t)^p = \sum_{r=0}^{p} C_{p, r} x_t^{p-r} u_t^r \]  

(11.4)

for \( p = 0, 1, \ldots, m \), where

\[ C_{p, r} = \frac{p!}{r! (p-r)!} \]

Substituting Equation 11.4 into 11.1 then yields

\[ y_t = \beta_0 + \beta_1 x_t + \ldots + \beta_m x_{m_t} + e_t \]

\[ = \sum_{j=0}^{m} \beta_j x_t^j + e_t \]

\[ = \sum_{j=0}^{m} \beta_j \sum_{r=0}^{j} C_{j, r} x_t^{j-r} u_t^r + e_t \]
From Equation 11.5 we can write

\[
X^{p_y} = \sum_{r=0}^{m} u^r \sum_{j=r}^{m} \beta_j C_j \sum_{r=0}^{j-r} X^{j-r} + X^p e_t.
\]

where \( s = j-r \). Summing Equation 11.6 over \( t \) and taking the expected value yields

\[
E\{ \sum_{t=1}^{N} X^{p_y} \} = \sum_{s=0}^{m} \left\{ \sum_{t=1}^{N} X^{s+p} \right\} \sum_{j=s}^{m} \beta_j C_j \sum_{j-s}^{j} \mu^j_{j-s}.
\]

since \( X_t \) is fixed for \( t = 1, \ldots, N \). In Equation 11.7,

\[
K_s = \sum_{j=s}^{m} \beta_j C_j \sum_{j-s}^{j} \mu^j_{j-s}
\]

for \( s = 0, \ldots, m \), and \( \mu^j_{j-s} = E(u^j_{j-s}) \).
From Equations 11.7 and 11.3 we see that \( \hat{\beta}_{OLS} \) is an unbiased estimator of

\[
K = \begin{bmatrix}
K_0 \\
K_1 \\
\vdots \\
K_m
\end{bmatrix}
\]

Since all odd moments of \( u_1 \) are zero (by Assumption 11.2) simplified expressions may be obtained for certain elements of \( K \).

Specifically, we have

\[
K_m = \beta_m C_m, 0 \mu_0 = \beta_m ;
\]

\[
K_{m-1} = \beta_{m-1} C_{m-1}, 0 \mu_0 + \beta_m C_m, 1 \mu_1 = \beta_{m-1} ;
\]

\[
K_{m-2} = \beta_{m-2} C_{m-2}, 0 \mu_0 + \beta_{m-1} C_{m-1}, 1 \mu_1 + \beta_m C_m, 2 \mu_2
\]

\[
= \beta_{m-2} + \beta_m C_m, 2 \mu_2 ; \quad (11.9)
\]

\[
K_{m-3} = \beta_{m-3} C_{m-3}, 0 \mu_0 + \beta_{m-2} C_{m-2}, 1 \mu_1
\]

\[
+ \beta_{m-1} C_{m-1}, 2 \mu_2 + \beta_m C_m, 3 \mu_3
\]

\[
= \beta_{m-3} + \beta_{m-1} C_{m-1}, 2 \mu_2 ; \quad \text{etc.}
\]
We have now proved the following:

**Theorem 11.1**

Under the model and assumptions of this appendix, the ordinary least squares estimator provides unbiased estimates of the coefficients of the two highest powers of $x$, but biased estimates of the coefficients of the remaining powers of $x$. We note that this result is consistent with, but more general than the results of Berkson and Geary.

From Equations 11.8 and 11.9 we can see that $\beta_0, \beta_1, \ldots, \beta_{m-2}$ are not identified if the moments of $u_1$ are unknown. If, however, the moments of $u_1$ are known, i.e. $u_1 \sim N(0, \sigma_u^2)$ with $\sigma_u^2$ known, then we can use the estimator $\hat{\beta}_{OLS}$ and the relations specified by Equation 11.8 to construct unbiased estimators of $\beta_0, \beta_1, \ldots, \beta_{m-2}$. For example, if $\sigma_u^2 = \mu_2^2$ is known, then from Equation 11.9 we see that

$$(0, 0, \ldots, 1, 0, -\frac{\sigma_u^2}{m-2} \sigma_{OLS}^2)$$

is an unbiased estimator of $\beta_{m-2}$. 
In this appendix we consider the linear errors-in-variables model specified by

$$y_t = x_t \beta,$$  \hspace{1cm} (12.1)

where we observe

$$Y_t = y_t + e_t$$

and

$$X_t = x_t + u_t$$

for \( t = 1, \ldots, N \). We assume the errors of measurement

\( \{(e_t, u_t)\}_{t=1}^N \) are independently distributed with zero means and finite, known error covariance matrices given by \( \Psi_t \). Also, we assume \( x_t \) is \((1 \times p)\), \( \beta \) is \((p \times 1)\) and unknown, and \( \beta \in \mathbb{R}^p \).

The unique feature of this treatment is that we impose linear parametric constraints on the parameter space. That is, we suppose \( \beta \) satisfies \( A\beta = B \) where \( A \) is \((r \times p)\), \( B \) is \((r \times 1)\), \( r \leq p-1 \), and the row rank of \( A \) is \( r \).

The proposed method of solution is to transform the full model, involving the entire vector \( \beta \), to a reduced model involving only a \((p-r)\) dimensional portion of \( \beta \). Known estimation procedures,
i.e. Fuller (1971), Booth (1973), could then be used to estimate the reduced model, while the remaining $r$ elements of $\beta$ would be chosen to satisfy $A\beta = B$.

To deal effectively with this problem, we require some additional notation. Denote the $(i, j)^{th}$ element of $A$ by $a_{ij}$ and the $i$th element of $B$ by $b_i$. Partition $A$ by setting $A = [ C \mid D ]$ where $C$ is $(r \times r)$ and $D$ is $[ r \times (p-r) ]$.

Without loss of generality we suppose $C$ is nonsingular. Then we let $c_{ij}^{ij}$ denote the $(i, j)^{th}$ element of $C^{-1}$ and we let $(C^{-1})_{i.}$ denote the $i$th row of $C^{-1}$. Furthermore, we denote the $(i, j)^{th}$ element of $D$ by $d_{ij}$, we let $\beta^{(r)}$ denote the first $r$ elements of $\beta$, and we let $\beta^{(p-r)}$ denote the remaining $p-r$ elements of $\beta$.

Now the linear parametric constraints

\[ A\beta = [ C \mid D ]\beta = B \]

may be rewritten as

\[ [ I_r \mid C^{-1} D ]\beta = C^{-1} B \quad (12.3) \]

where $I_r$ is the identity matrix of order $r$. From Equation 12.3 we obtain
\[ \beta^{(r)} = C^{-1} B - C^{-1} D \beta^{(p-r)} \]

\[ = C^{-1} \left[ B - D \beta^{(p-r)} \right]. \quad (12.4) \]

Define

\[ \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \beta^{(r)} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}, \]

\[ \beta^{(p-r)} = \begin{bmatrix} \beta_{r+1} \\ \vdots \\ \beta_p \end{bmatrix}, \quad \text{and} \quad x_t = (x_{t1}, \ldots, x_{tp}) \quad \text{for} \quad t = 1, \ldots, N. \]

Then by Equation 12.4, Model 12.1 may be expressed by

\[ y_t = x_t \beta = \sum_{i=1}^{p} x_{ti} \beta_i \]

\[ = \sum_{i=1}^{r} x_{ti} \beta_i + \sum_{i=r+1}^{p} x_{ti} \beta_i \]

\[ = \sum_{i=1}^{r} x_{ti} [(C^{-1})_{ii} (B - D \beta^{(p-r)})] \]

\[ + \sum_{i=r+1}^{p} x_{ti} \beta_i. \]
Thus 
\[ y_t - \sum_{i=1}^{r} x_{t_i} [(C^{-1})i.B] = \quad (12.5) \]

\[ \sum_{i=r+1}^{p} x_{t_i} \beta_i - \sum_{i=1}^{r} x_{t_i} (C^{-1})i.D \beta^{(p-r)}. \]

Observe that 
\[ (C^{-1})i.B = \sum_{k=1}^{r} a_{ik} b_k \quad (12.6) \]

and that 
\[ \sum_{i=1}^{r} x_{t_i} (C^{-1})i.D = \quad (12.7) \]

\[ \left[ \sum_{i=1}^{r} x_{t_i} \sum_{k=1}^{r} c_{ik} d_{ki}, \ldots, \sum_{i=1}^{r} x_{t_i} \sum_{k=1}^{r} c_{ik} d_{k(p-r)} \right] \]

\[ = \left[ \sum_{i=1}^{r} x_{t_i} \sum_{k=1}^{r} c_{ik} a_{kr+1}, \ldots, \sum_{i=1}^{r} x_{t_i} \sum_{k=1}^{r} c_{ik} a_{kp} \right] \]

since \( d_{ki} = a_{kr+i} \). If we substitute Equations 12.6 and 12.7 into 12.5 then we obtain

\[ y_t - \sum_{j=1}^{r} x_{t_j} \sum_{k=1}^{r} c_{jk} b_k = \quad (12.8) \]

\[ \sum_{i=r+1}^{p} \left\{ x_{t_i} - \sum_{j=1}^{r} x_{t_j} \sum_{k=1}^{r} c_{jk} a_{ki} \right\} \beta_i. \]

Equation 12.8 is the reduced model of which we spoke earlier.

This may be considered a linear errors-in-variables model with no
restrictions on the parameter space. Here the model is

\[ y^*_t = x^*_t \beta^{(p-r)} \quad (12.9) \]

where we observe

\[ y^*_t = y^*_t + e^*_t \]
\[ x^*_t = x^*_t + u^*_t \]

for \( t = 1, \ldots, N \), and where

\[ y^*_t = y^*_t - \sum_{j=1}^{r} x^*_t \sum_{k=1}^{r} c^j_k b_k, \]
\[ e^*_t = e^*_t - \sum_{j=1}^{r} u^*_t \sum_{k=1}^{r} c^j_k b_k, \]
\[ x^*_t = \]
\[ [\{x^*_t r+1 - \sum_{j=1}^{r} x^*_t \sum_{k=1}^{r} c^j_k a^r_{k r+1}\}, \ldots, \{x^*_t p - \sum_{j=1}^{r} x^*_t \sum_{k=1}^{r} c^j_k a^r_p\}], \]
\[ u^*_t = \]
\[ [\{u^*_t r+1 - \sum_{j=1}^{r} u^*_t \sum_{k=1}^{r} c^j_k a^r_{k r+1}\}, \ldots, \{u^*_t p - \sum_{j=1}^{r} u^*_t \sum_{k=1}^{r} c^j_k a^r_p\}], \]
\[ u_t = (u_{t1}, \ldots, u_{tp}), \]

and \[ \beta^{(p-r)'} = (\beta_{r+1}, \ldots, \beta_p). \]
By our assumptions the errors of measurement \( \{ (e_t^*, u_t^*) \}_{t=1}^N \) are independently distributed with zero mean and with error covariance matrices

\[
\hat{\mathbf{x}}_t^* = \mathbf{T}_t \hat{\mathbf{p}}_t \mathbf{T}_t',
\]

\[
\mathbf{T} = \begin{bmatrix}
1 & \{ - \sum_{k=1}^r c_{1k} b_k \} & \ldots & \{ - \sum_{k=1}^r c_{rk} b_k \} & 0 & \ldots & 0 \\
0 & \{ - \sum_{k=1}^r c_{1k} a_{k+1} \} & \ldots & \{ - \sum_{k=1}^r c_{rk} a_{k+1} \} & 1 & \ldots & 0 \\
\vdots \\
0 & \{ - \sum_{k=1}^r c_{1k} a_{kp} \} & \ldots & \{ - \sum_{k=1}^r c_{rk} a_{kp} \} & 0 & \ldots & 1
\end{bmatrix}
\]

is \((p-r+1 \times p+1)\).

Summarizing, Equations 12.4 and 12.9 comprise our solution of the linear errors-in-variables model with linear parametric constraints. First, we obtain an estimator of \( \beta^{(p-r)} \) by use of an appropriate estimation procedure on the reduced model given by Equation 12.9. Second, an estimator of \( \beta^{(r)} \) is obtained from Equation 12.4 and the estimator of \( \beta^{(p-r)} \).