INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations 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 through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

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

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

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

ESTIMATORS FOR THE ERRORS-IN-VARIABLES MODEL

Iowa State University

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

Ph.D. 1982
Estimators for the errors-in-variables model

by

Yasuo Amemiya

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
1982
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION AND LITERATURE REVIEW</strong></td>
<td>1</td>
</tr>
<tr>
<td>A. Definition of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>B. Review of the Multivariate Linear Errors-in-Variables Model</td>
<td>5</td>
</tr>
<tr>
<td>1. Structural relationships</td>
<td>5</td>
</tr>
<tr>
<td>2. Functional relationships</td>
<td>8</td>
</tr>
<tr>
<td>C. Review of the Nonlinear Errors-in-Variables Model</td>
<td>14</td>
</tr>
<tr>
<td><strong>II. DEFINITIONS AND THEOREMS</strong></td>
<td>23</td>
</tr>
<tr>
<td>A. Definitions</td>
<td>23</td>
</tr>
<tr>
<td>1. Matrix-vector operations</td>
<td>23</td>
</tr>
<tr>
<td>2. Order in probability</td>
<td>25</td>
</tr>
<tr>
<td>B. Theorems</td>
<td>26</td>
</tr>
<tr>
<td>1. Matrix theorems</td>
<td>26</td>
</tr>
<tr>
<td>2. Central limit theorems</td>
<td>38</td>
</tr>
<tr>
<td><strong>III. THE MULTIVARIATE LINEAR MODEL WITH ESTIMATED ERROR COVARIANCE MATRIX</strong></td>
<td>55</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>55</td>
</tr>
<tr>
<td>B. Maximum Likelihood Estimators</td>
<td>58</td>
</tr>
<tr>
<td>1. Functional relationship</td>
<td>62</td>
</tr>
<tr>
<td>2. Structural relationship</td>
<td>77</td>
</tr>
<tr>
<td>3. Maximum likelihood estimators adjusted for degrees of freedom</td>
<td>94</td>
</tr>
<tr>
<td>C. Asymptotic Properties</td>
<td>98</td>
</tr>
<tr>
<td>1. Strong consistency</td>
<td>98</td>
</tr>
<tr>
<td>2. Limiting distribution</td>
<td>113</td>
</tr>
<tr>
<td>D. Comments</td>
<td>133</td>
</tr>
</tbody>
</table>
IV. THE NONLINEAR MODEL WITH KNOWN ERROR COVARIANCE MATRIX 135

A. Introduction 135
B. Results of Wolter and Fuller 139
C. Bias of the Pseudo-Maximum Likelihood Estimator 146
D. Bias of the Modified Maximum Likelihood Estimator 162
E. A Class of Estimators Adjusted for Bias 165
F. Examples of the Adjusted Estimators 175
   1. Adjustment based on the local quadratic approximation 176
   2. Adjustment based on the curvature relative to the tangent 180
   3. Adjustment based on the expansion terms 187
   4. Comparison among the adjusted estimators 189
G. A Monte Carlo Study 191
   1. Generating the data 192
   2. Estimators compared 193
   3. Results and conclusions 196

V. INSTRUMENTAL VARIABLE ESTIMATION OF THE NONLINEAR MODEL 199

A. Introduction 199
B. The Instrumental Variable Estimator 208
   1. The estimator 208
   2. Consistency 210
   3. Limiting distribution 219
   4. Bias 242
C. An Estimator of the Error Covariance Matrix 247
D. An Estimator of the True Value 256
| E. The Modified Instrumental Variable Estimator | 264 |
| 1. Introduction | 264 |
| 2. The estimator | 270 |
| 3. Preliminary lemmas | 272 |
| 4. Consistency | 277 |
| 5. Limiting distribution | 280 |
| 6. Bias | 293 |
| F. Modified Estimators of the Error Covariance Matrix and the True Value | 294 |
| G. An Iterative Estimation Procedure | 298 |
| VI. BIBLIOGRAPHY | 305 |
| VII. ACKNOWLEDGEMENTS | 311 |
I. INTRODUCTION AND LITERATURE REVIEW

The statistical consideration of models containing errors of measurement began as early as the 1870s. An extensive literature deals with estimation for the univariate linear errors-in-variables model. Less is known concerning estimation for the model with multiple linear relationships or for the model with a nonlinear relationship. This thesis considers estimation for the multivariate linear errors-in-variables model and for the nonlinear errors-in-variables model. We begin by giving a mathematical definition of the errors-in-variables model.

A. Definition of the Problem

Let \( \{X_t\}_{t=1}^{\infty} \) be a sequence of, possibly constant, \( r \)-dimensional random row vectors. Let \( \{x_t\}_{t=1}^{\infty} \) be a sequence of, possibly constant, \( q \)-dimensional random row vectors. Let \( \beta \) be a \( kr \times 1 \) vector of parameters belonging to \( \mathcal{G} \) where \( \mathcal{G} \) is a subset of \( kr \) dimensional Euclidean space \( \mathbb{R}^{kr} \). Suppose that

\[
X_t = \xi(x_t; \beta), \quad t = 1, 2, \ldots, \tag{1.1}
\]

where the components of the \( r \)-vector \( \xi \) are real valued Borel measurable functions mapping \( \mathbb{R}^{q} \) into \( \mathbb{R}^{1} \).
Let

\[ Y_t = Y_t + \xi_t , \quad t = 1, 2, \ldots, n , \quad (1.2) \]

\[ \xi_t = \xi_t + \eta_t , \quad t = 1, 2, \ldots, n , \quad (1.3) \]

where \( Y_t \) and \( \xi_t \) are observable random row vectors of dimensions \( r \)
and \( q \), respectively, \( \xi_t \) and \( \eta_t \) are unobservable error vectors of
dimensions \( r \) and \( q \), and \( \xi_t = (\xi_t, \eta_t) \) are independently and
identically distributed with mean zero and covariance matrix \( \Sigma_{\xi\xi} \). The
model (1.1), (1.2), and (1.3) defines the **errors-in-variables model**.

If the relationship (1.1) is linear in both \( \xi_t \) and \( \xi_t \), that is, if

\[ Y_t = \xi_t \xi , \quad t = 1, 2, \ldots, \quad (1.4) \]

where \( q = k \), and \( \xi \) is a \( k \times r \) matrix of parameters formed from
\( kr \times 1 \) vector \( \xi \), then the model (1.1), (1.2), and (1.3) is called the
**linear errors-in-variables model**. The model is the **univariate linear
errors-in-variables model** if \( r = 1 \), and is the **multivariate linear
errors-in-variables model** if \( r > 1 \). When the relationship (1.1) is
not of the form (1.4), the model (1.1), (1.2), and (1.3) is called the
**nonlinear errors-in-variables model**.

The distinction between \( \xi_t \) being fixed or random is important in
the estimation of the errors-in-variables model. If the \( \xi_t \),
t = 1,2,..., are constant vectors, the errors-in-variables model is called the functional relationship model, while if the \( \chi_t \), t = 1,2,..., are nonconstant random vectors, the model is called the structural relationship model. In the functional relationship model, the \( \theta \), the parameters specifying the distribution of \( \varepsilon_t \), and the \( \chi_t \), t = 1,2,..., constitute the set of all parameters. For the structural relationship model, the set of all parameters consists of \( \theta \) and the parameters specifying the joint distribution of \( \varepsilon_t \) and \( \chi_t \). In most research on the structural relationship model, the \( \chi_t \), t = 1,2,..., n, are assumed to be independently and identically distributed with finite second moments and the \( \chi_t \) and \( \varepsilon_t \) are assumed independent.

Neyman and Scott (1948) introduced terms to distinguish between two important types of parameters. They referred to parameters entering the distribution of the observable random variables for finitely many t as incidental parameters, and those entering for infinitely many t as structural parameters. If we consider the consistency of estimators as the number of observations tends to infinity, then it is generally only the structural parameters which we can hope to estimate consistently.

The error variable \( \varepsilon_t \) in (1.2) may sometimes be decomposed into two parts,

\[
\varepsilon_t = \varepsilon_t^0 + \varepsilon_t^* ,
\]

where \( \varepsilon_t^* \) is a vector of measurement errors, and \( \varepsilon_t^0 \) is a vector of
errors in the equation (1.1). The term $\varepsilon_t^0$ represents the failure for the relationship to hold perfectly. Generally, it is easier to obtain information on $\varepsilon_t^*$ than on $\varepsilon_t^0$. Most research on the errors-in-variables model has been on models with some information on $\varepsilon_t^*$ and with no error in the equation. A classical errors-in-variables model assumes no error in the equation and known measurement error covariance matrix. For applying theories developed for such a model in practice, it is critical that the assumption of perfect relationship be examined.

The concept of identifiability is important in estimation of models such as the errors-in-variables model. If no pair of distinct values of a parameter in the parameter space gives an identical distribution function of the observable random variables, then the parameter is called identifiable. Otherwise, the parameter is called nonidentifiable. We say a model is identified, if all parameters in the model are identifiable, and is not identified, otherwise. In the literature, we find various types of additional information used to identify the errors-in-variables model. Generally, distinct types of information lead to distinct methods of estimation.


In the rest of this chapter, we review the literature on the multivariate linear errors-in-variables model and the nonlinear errors-in-variables model.

B. Review of the Multivariate Linear Errors-in-Variables Model

1. Structural relationships

An extensive literature exists for the following factor analysis model. Let the $1 \times p$ random vectors $\xi_t$, $t = 1,2,\ldots,n$, be independently and identically distributed with mean $\mu_Z$ and covariance matrix

$$\Sigma_{\xi\xi} = A \Sigma_{XX} A' + \Sigma_{\xi\varepsilon}, \quad (1.5)$$

where $A$ is a $p \times k$ matrix of unknown factor loadings, $\Sigma_{XX}$ is a $k \times k$ positive definite covariance matrix of unobservable common factors, and $\Sigma_{\xi\varepsilon}$ is the error covariance matrix. The multivariate linear structural relationship with the assumption that $\xi_t$ are independently and identically distributed with finite second moments and independent of $\xi_t$ is a special case of the factor analysis model (1.5), where
A' = (\theta', \Sigma_{zz}) .

Most of the work on the factor analysis model concentrates on inferences for the parameters in (1.5) based on a sample covariance matrix \Sigma_{zz} under the assumption of a diagonal \Sigma_{ee}.

The maximum likelihood estimators and their limiting properties are discussed in Lawley (1940, 1941, 1943, 1967, 1976), Anderson and Rubin (1956), Jöreskog (1967), and Jennrich and Thayer (1973). Detailed discussions of the problem of factor analysis are given in Anderson and Rubin (1956) and Lawley and Maxwell (1971).

Jöreskog and Goldberger (1972) considered another method of estimation for the factor analysis model. As in most works on factor analysis, their parameterization assumes \Sigma_{ee} = \Psi is diagonal and \Sigma_{xx} = I . They defined the generalized least squares estimators of \Theta and \Psi to be the values of \Theta and \Psi that minimize

\[
\frac{1}{2} \text{tr}(I - \Sigma_{zz}^{-1} \Sigma_{zz}^{-1}) ,
\]

where

\[
\Sigma_{zz} = \Theta \Theta' + \Psi^2 .
\]

Jöreskog and Goldberger showed that the generalized least squares estimators of \Theta and \Psi have the same limiting distribution as the
maximum likelihood estimators. They also showed that the minimum value of (1.6) is asymptotically distributed as a chi-square random variable with \( \frac{1}{2} [(p - k)^2 - (p + k)] \) degrees of freedom.

Anderson (1973) applied the generalized least squares method to the general linear covariance structures. His model includes the multivariate linear structural relationship with known \( \beta \) as a special case. Anderson showed that the generalized least squares estimator is asymptotically equivalent to the maximum likelihood estimator.

Jöreskog (1970, 1973a, 1973b) considered the general covariance structure model where observations satisfy certain linear structural equations. His model is general enough to include the multivariate linear structural relationship, the factor analysis model, and the simultaneous equation model with errors in the exogenous variables. Jöreskog proposed maximum likelihood estimation based on the assumption that all random variables included in the model are normally distributed. Under the normality assumption, the limiting covariance matrix of the maximum likelihood estimator is given by the inverse of the information matrix. He suggested a certain iterative procedure to obtain the estimates and an estimate of the limiting covariance matrix.

Browne (1974) discussed in detail the generalized least squares estimation of a vector of unknown parameters \( \beta \) in the covariance matrix \( \Sigma_{zz}(\beta) \). Under the assumption that the sample covariance matrix \( \hat{\Sigma}_{zz} \) has the Wishart distribution, he found the optimal choice of the weight matrix, and called the estimator obtained by such a weight matrix the best generalized least squares estimator. Browne showed that
the best generalized least squares estimator is asymptotically equivalent to the maximum likelihood estimator based on the Wishart distribution. He also showed that the minimum value of the generalized sum of squares function is asymptotically distributed as chi-square with \[
\frac{1}{2} p(p + 1) - q \]
degrees of freedom, where \( \Sigma_{ZZ} \) is \( p \times p \) and \( z \) is \( q \times 1 \).

Others have written about the multivariate linear structural relationship, including Grubbs (1948), Bock and Bargmann (1966), Robinson (1977), and Carter and Fuller (1980).

2. Functional relationships

The first work on the maximum likelihood estimation for the multivariate linear functional relationship was that of Anderson (1951). He considered the problem of estimating linear restrictions on coefficients for multivariate regression analysis. Assuming the residual vector to have a multivariate normal distribution, he obtained the maximum likelihood estimators of the linear restrictions, the coefficients, and the covariance matrix of residuals. He also gave confidence regions for the linear restrictions and discussed various testing problems associated with the model. Anderson stated that the limiting distribution of the maximum likelihood estimator of the restrictions can be obtained from the asymptotic results for certain eigenvectors. He derived limiting results by letting the number of observations increase holding the dimension of regression coefficients fixed. In terms of the errors-in-variables model, this is equivalent to
letting the number of replications increase while the number of the true values remains fixed. If we consider the means over replications as observations in the functional relationship model, then Anderson's model is equivalent to the model (1.2), (1.3), and (1.4) where \( \Sigma_{ee} \) decreases at rate \( T^{-1} \), the number of observations, \( n \), is fixed, and there is an unbiased estimate of \( \Sigma_{ee} \) whose distribution is a multiple of Wishart. A discussion of the different approaches to the derivation of limiting results is given in Anderson (1976).

Another work on the multivariate linear functional relationship which uses a different parameterization is that of Anderson and Rubin (1956). They showed that the maximum likelihood estimator does not exist for the factor analysis model with fixed factors. In our parameterization, this model is the functional relationship model with a diagonal error covariance matrix. For this model, Anderson and Rubin suggested estimation by maximizing the likelihood of the sample covariance matrix. They showed the asymptotic normality of the estimator.

The first explicit treatment of the multivariate linear functional relationship was given by Gleser and Watson (1973). They considered the functional relationship model (1.2), (1.3), and (1.4) where \( k = r \), and for a known \( k \times k \) positive definite matrix \( \Phi \) and an unknown positive constant \( \sigma^2 \).
They derived the maximum likelihood estimators $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{\Sigma}_t$ of the parameters $\beta$, $\sigma^2$, and $\Sigma_t$. Gleser and Watson showed that $\hat{\beta}$ and $2\hat{\sigma}^2$ are consistent for $\beta$ and $\sigma^2$ respectively, as the number of observations $(Y_t, X_t)$ increases. They were unable to obtain the asymptotic distribution of the estimators.

Bhargava (1977) considered the model (1.2), (1.3), and (1.4) where $k = r$,

$$
\Sigma_{\epsilon \epsilon} = \begin{pmatrix}
\sigma^2 & 0 \\
0 & \sigma^2
\end{pmatrix},
$$

and $\Sigma$ is an unknown $k \times k$ positive definite matrix. While able to prove the existence of a solution to the maximum likelihood equations, Bhargava could express the solution in closed form only if $\Sigma$ and $\beta$ had the same known eigenvectors.

Bhargava (1979), following the work of Gleser and Watson, obtained the maximum likelihood estimators of the model (1.2), (1.3), and (1.4) where for an unknown positive constant $\sigma^2$.
He showed that the maximum likelihood estimators $\hat{\theta}$ and $\hat{\sigma}^2$ converge in probability to $\theta$ and $(k + r)^{-1} \sigma^2$ as the number of observations increases.

Healy (1980) reproduced Anderson's maximum likelihood estimator for the multivariate linear functional relationship with replicate observations. Healy used matrix algebra and regression results in the derivation of the maximum likelihood estimator, while Anderson used the differentiation technique. Healy considered the consistency of the maximum likelihood estimators under different assumptions than those of Anderson. Let \( n \) be the number of true values, and let \( N \) be the total number of observations. Under the assumption

$$1 > t = \lim_{N \to \infty} N^{-1} n,$$

Healy showed that the maximum likelihood estimator of the coefficient matrix is strongly consistent and the maximum likelihood estimator of $\Sigma$ converges almost surely to a matrix which is different from $\Sigma$. He did not derive either the limiting distribution nor the covariance matrix of the estimator.

Gleser (1981) considered in detail the multivariate linear functional relationships with and without the intercept, where the error covariance matrix has the form

$$\Sigma_{\epsilon\epsilon} = \sigma^2 \mathbf{I}.$$
\[ \Sigma_{\epsilon \epsilon} = \sigma^2 \mathbf{I} , \]

and \( \sigma^2 \) is unknown. He defined the ordinary least squares estimators of \( \beta \) and \( \xi_t \) relative to an orthogonally invariant norm \( \| \cdot \| \) to be the values of \( \beta \) and \( \xi_t \) that minimize \( \| R_1(\beta, \xi_t) \| \), where \( R_1(\beta, \xi_t) \) is a \( n \times (k + r) \) matrix with \( t \)-th row being

\[ (\xi_t', \xi_t') - \xi_t (\xi_t', \xi_t') . \]

He also defined the generalized least squares estimator of \( \beta \) relative to an orthogonally invariant norm \( \| \cdot \| \) to be the value of \( \beta \) that minimizes \( \| R_2(\beta) \| \), where \( R_2(\beta) \) is a \( n \times r \) matrix with \( t \)-th row being

\[ (\xi_t - \xi_t \beta) (\xi + \beta' \beta)^{-1/2} . \]

Gleser showed that the maximum likelihood estimator of \( \beta \) under normal errors is an ordinary least squares estimator relative to any orthogonally invariant norm and is a generalized least squares estimator relative to any orthogonally invariant norm. He also showed that the maximum likelihood estimator of \( \xi_t \) under normal errors is an ordinary least squares estimator relative to any orthogonally invariant norm. The almost sure limits of the maximum likelihood estimators \( \hat{\beta} \) and \( \hat{\xi} \) were obtained under a weak assumption on the error terms. Gleser
derived the limiting distributions of $\hat{\theta}$ and $\hat{\sigma}^2$. He also presented the large sample confidence regions for $\hat{\theta}$ and $\hat{\sigma}^2$. He pointed out that if the model is such that

$$\Sigma_{\varepsilon \varepsilon} = \sigma^2 \Phi,$$  \hspace{1cm} (1.7)

where $\Phi$ is a known positive definite matrix, then the observations can be transformed so that the error covariance matrix is $\sigma^2 I$. Thus, theoretically speaking, the limiting covariance matrix of the maximum likelihood estimator for the model with the assumption (1.7) can be obtained from Gleser's results. However, Gleser failed to obtain a simple closed form for the limiting covariance matrix for the model with covariance matrix given by (1.7).

Dahm and Fuller (1981) applied the generalized least squares method to the sample covariance matrix to estimate $\hat{\theta}$ and $\hat{\sigma}^2$ of the multivariate linear functional relationship. They showed that the generalized least squares estimators of $\hat{\theta}$ and $\hat{\sigma}^2$ constructed under the assumption of the structural model have the same limiting distribution as the best generalized least squares estimators. They also showed that an estimator of the covariance of the limiting distribution may be obtained by proceeding as if the model is a structural relationship.

Villegas (1982) obtained, through geometric arguments, the maximum likelihood estimator of the multivariate linear functional relationship with or without intercept, when there is available an independent
Wishart matrix as an estimate of a multiple of the error covariance matrix. He also obtained the estimators which minimize certain sums of squares. He gave no discussion on the properties of his estimators.

Other literature on the multivariate linear functional relationship includes Tintner (1945), Geary (1948), Whittle (1952), HÖschel (1978a, 1978b), Nussbaum (1978), and Chan (1980).

Dahm (1979) reviewed the multivariate linear errors-in-variables model for both the structural and functional cases.

C. Review of the Nonlinear Errors-in-Variables Model

The literature on the nonlinear errors-in-variables models has been limited, compared with that for the linear models. Only the univariate model containing a single nonlinear relationship has been studied. Most research has concentrated on analysis of functional relationship models. Estimators have been constructed on either the least squares principle or the maximum likelihood principle.

Early works on the nonlinear functional relationship model include Deming (1931, 1943) and Cook (1931). They suggested an estimation procedure which applies the least squares principle to the linear portion of the Taylor expansion of the functional relationship. These authors assumed the error covariance matrix known.

Using heuristic arguments, Clutton-Brock (1967) suggested that a pseudo-likelihood function be used for the nonlinear model with the error covariance matrix known. His pseudo-likelihood function is an analogue of that for the model without measurement errors.
Villegas (1969) considered the nonlinear functional relationship with \( n \) replicate observations on each of \( k \) fixed true values \((y_{ti}, \xi_{ti})\), where \( \xi_{ti} \) is \( 1 \times m \). He used the \( k \) mean vectors \((\bar{y}_t, \bar{\xi}_t)\) as data points for the functional relationship and used the within mean square matrix as a consistent estimator of the error covariance matrix. Thus, in his model, the error variances of the data points are decreasing at the rate \( n^{-1} \). He proposed an iterative estimation procedure. Assuming the existence of a preliminary estimator with error of \( O_p(n^{-1/2}) \) and normal errors, he showed that his estimator has the normal distribution in the limit. Villegas also showed that the unweighted least squares estimator has error of \( O_p(n^{-1/2}) \) and, thus, can be used as the preliminary estimator.

Dolby and Lipton (1972) derived the maximum likelihood equation for the nonlinear functional relationship with scalar \( \chi_t \), a general covariance structure, normal errors, and replications. They also obtained the inverse of the information matrix, and suggested its use for 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. The statement is not necessarily applicable in the presence of infinitely many incidental parameters.

Dolby (1972) proposed the estimator obtained by applying the Newton-Raphson technique to the likelihood equation for the nonlinear functional relationship with known general covariance structure. He obtained the information matrix and made the same statement with regard
to the asymptotic covariance matrix as made in his earlier paper with Lipton. Dolby also discussed an estimation procedure which is a generalization of the generalized least squares estimation of Sprent (1966).

Egerton and Laycock (1979) considered the nonlinear functional relationship where the $\mathbf{x}$ is a vector,

$$
\mathbf{z}_c = \sigma^2 \mathbf{g},
$$

and $\mathbf{g}$ is known. The likelihood equation was derived. They proposed an adjustment to the maximum likelihood estimator of $\sigma^2$, and called, without justification, the resulting estimator a suitable consistent estimator of $\sigma^2$. They also obtained the inverse of the information matrix, and commented that the portion of the inverse matrix for the structural parameters underrate the true variability of the estimators. However, they did not give conditions which guarantee the existence of an asymptotic distribution for the estimators.

The theoretical development for estimation of the nonlinear functional relationship reviewed to this point is largely unsatisfactory. Except for Villegas (1969), there has been no rigorous investigation of the conditions guaranteeing the consistency or the asymptotic normality of the maximum likelihood estimators. Recall that Villegas assumed that the number of unknown true values is fixed and that the error variances decrease at the rate $n^{-1}$. The first consistent and asymptotically normal estimator for the nonlinear
functional relationship without replications or decreasing error variances was given by Wolter and Fuller (1982a). They proposed a method of moment type estimator for polynomial functional relationships with normally distributed errors. Their estimator is analogous to the maximum likelihood estimator for the linear model, but is not the maximum likelihood estimator. As in Wolter and Fuller (1982a), we review the estimator in terms of the quadratic functional relationship. The model is

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

\[
Y_t = y_t + e_t,
\]

\[
X_t = x_t + u_t, \quad t = 1, 2, \ldots, n,
\]  

(1.8)

where the \(x_t\) are fixed, and \((e_t, u_t)\) are independent normal random variables with mean zero and known covariance matrix

\[
\Sigma = \begin{bmatrix}
\sigma_e^2 & \sigma_{eu} \\
\sigma_{ue} & \sigma_u^2
\end{bmatrix}.
\]

Wolter and Fuller observed that the model (1.8) can be written as

\[
y_t = y_t e,
\]

\[
Y_t = y_t + e_t,
\]
\[ \hat{\xi}_t = \xi_t + \xi_t', \]

where

\[ \xi = (\beta_0, \beta_1, \beta_2)', \]

\[ \xi_t = (1, x_t, x_t^2), \]

\[ \xi_t = (1, x_t, x_t^2 - \sigma_u^2), \]

\[ \xi_t = (0, u_t, 2x_t u_t + u_t^2 - \sigma_u^2). \]

Note that \((e_t, \xi_t)\) has zero mean and covariance matrix

\[ \begin{bmatrix}
\sigma_e^2 & \xi_{ef}(t) \\
\xi_{fe}(t) & \xi_{ff}(t)
\end{bmatrix} \]

\[ = \begin{bmatrix}
\sigma_e^2 & 0 & 2x_t \sigma_e \\
0 & \sigma_e & \sigma_e \\
2x_t \sigma_e & \sigma_e & 2x_t \sigma_u \\
sym. & 4x_t^2 \sigma_u^2 + 2\sigma_u^4
\end{bmatrix}. \]

An unbiased estimator of \(\hat{\xi}_t\) is
\[ \hat{\Sigma}_t = \begin{bmatrix} \sigma_e^2 & 0 & \sigma_{eu} & 2X_t \sigma_{eu} \\ 0 & 0 & 0 & 0 \\ \sigma_u^2 & 2X_t \sigma_u^2 & \text{sym.} \\ 4\sigma_u^2 (X_t^2 - \sigma_u^2) + 2\sigma_u^4 \end{bmatrix}, \]

and an unbiased estimator of \( \hat{\xi} = n^{-1} \sum_{t=1}^n \hat{\xi}_t \) is

\[ \hat{\xi} = n^{-1} \sum_{t=1}^n \hat{\xi}_t = \begin{bmatrix} \sigma_{ee} & \hat{\Sigma}_{ef} \\ \hat{\xi}_{fe} & \hat{\xi}_{ff} \end{bmatrix}. \]

Note that the polynomial structure of the model and the normality of errors have been utilized. Let

\[ M = n^{-1} \sum_{t=1}^n \begin{pmatrix} Y_t \\ W_t \end{pmatrix} (Y_t, W_t) = \begin{pmatrix} M_{YY} & M_{YW} \\ M_{YW} & M_{WW} \end{pmatrix}, \]

Wolter and Fuller defined the estimator of \( \hat{\xi} \) as

\[ \hat{\xi} = [M_{WW} - \hat{\alpha} \hat{\xi}_{ff}]^{-1} [M_{YW} - \hat{\alpha} \hat{\xi}_{fe}], \] (1.9)
where $\hat{a}$ is the smallest root of

$$| y - \hat{a} \hat{x} | = 0.$$ 

They showed that the estimator $\hat{g}$ in (1.9) is consistent and is asymptotically normally distributed. They also presented a consistent estimator of the asymptotic covariance matrix of $\hat{g}$. A normalized statistic based on $\hat{a}$ was proposed for testing the goodness of fit of the model. Wolter and Fuller suggested that the properties of the estimator can be improved by making a small-order modification in the estimator. Based on their Monte-Carlo study, they suggest that the modified estimator may be recommended over the maximum likelihood estimator in samples as small as 30 with ratio of $\sigma^2_u$ to $n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2$ as large as 0.6. The estimated variance and associated t-statistic can be used for inferential statements, with only a small departure from nominal significance levels.

Fuller (1982) extended the use of the method of moments type estimator of Wolter and Fuller to a wide class of models. He considered a class of models which have a linear representations. In his model, the true values of $x_t$ may be fixed or random and errors in the equation may exist. He assumed that conditional on the $x_t$ the deviations of observations $(y_t, x_t)$ from the true values have mean zero and finite second moments, and that there is available for each $t$ an estimator of $\xi_t$ which is unbiased given $x_t$ and can be expressed as
\[
\sum_{i=1}^{m} \psi^*_i \psi_{ti},
\]

where \( m \) is a fixed number, \( m \leq k + 1 \), and \( \psi_{ti} \) are observable vectors. Fuller defined the method of moments type estimator, and derived the limiting distribution of the estimator under a wide range of assumptions. He demonstrated that his class covers many useful models including the polynomial errors-in-variables model. His estimation procedure is applicable to the nonlinear errors-in-variables model, provided that the observation can be expressed as the sum of a linear systematic part and an error with zero mean, and that there is an estimator of error covariance matrix for each observation having the form (1.10).

Wolter and Fuller (1982b) rigorously discussed the maximum likelihood type estimators for the general nonlinear functional relationship with known error covariance matrix. Their work will be reviewed in Chapter IV.

Estimation of specific nonlinear models has been considered by Hey and Hey (1960), Chan (1965), O'Neill et al. (1969), and Anderson (1981).


The remainder of this thesis deals with the estimation of three types of errors-in-variables models. These are the multivariate linear
structural model, the multivariate linear functional model, both with estimated error covariance matrix, and the nonlinear model with one functional relationship. Chapter II is devoted to presenting background notation, definitions, and theorems to be used in later chapters. Chapter III contains the derivations and properties of the maximum likelihood estimators for the multivariate linear errors-in-variables models. In Chapter IV, we review the work of Wolter and Fuller (1982b), discuss the properties of their estimators, and present alternative estimators. The instrumental variable estimation of the nonlinear functional relationship is considered in Chapter V.
II. DEFINITIONS AND THEOREMS

This chapter is devoted to presenting definitions and theorems to be used in later chapters. The proofs of several theorems are omitted, but references to available proofs are given.

A. Definitions

1. Matrix-vector operations

In a number of situations where we are interested in functions of elements of a matrix, it is more convenient to arrange the elements of the matrix as a vector.

Definition 2.1. Let \( \mathbf{A} = (a_{ij}) \) be a \( p \times q \) matrix, and let \( \mathbf{A}_j \) denote the \( j \)-th column of \( \mathbf{A} \). Then,

\[
\text{vec} \ \mathbf{A} = (a_{11}, a_{21}, \ldots, a_{p1}, a_{12}, a_{22}, \ldots, a_{p2}, \ldots, a_{1q}, a_{2q}, \ldots, a_{pq})
\]

The above vector can be written succinctly as

\[
(\mathbf{A}_1', \mathbf{A}_2', \ldots, \mathbf{A}_q')'.
\]

Note that \( \text{vec} \ \mathbf{A} \) is the vector obtained by listing the columns of \( \mathbf{A} \) one beneath the other in a single column vector. If the matrix \( \mathbf{A} \) is a symmetric \( p \times p \) matrix, \( \text{vec} \ \mathbf{A} \) will contain \( \frac{1}{2}p(p-1) \) pairs of identical elements. In some situations, it is convenient to retain only one element of each pair. This can be accomplished by listing the elements in each column that are on or below the diagonal.

Definitions 2.2. Let \( \mathbf{A} = (a_{ij}) \) be a \( p \times p \) matrix. Then,
vech \( \mathbf{A} = (a_{11}, a_{21}, \ldots, a_{p1}, a_{22}, a_{32}, \ldots, a_{p2}, a_{33}, a_{43}, \ldots, a_{p3}, \ldots, a_{pp})' \).

For symmetric \( \mathbf{A} \), vech \( \mathbf{A} \) contains the unique elements of \( \mathbf{A} \). Therefore, it is possible to recreate vec \( \mathbf{A} \) from vech \( \mathbf{A} \).

**Definition 2.3.** Let \( \mathbf{A} = (a_{ij}) \) be a \( p \times p \) symmetric matrix. Let \( \mathcal{P}_p \) be the \( p^2 \times \frac{1}{2} p(p+1) \) matrix such that

\[
\text{vec } \mathbf{A} = \mathcal{P}_p \text{ vech } \mathbf{A},
\]

and define \( \mathcal{P}_p \) by

\[
\mathcal{P}_p = (\mathcal{P}_p' \mathcal{P}_p)^{-1} \mathcal{P}_p'.
\]

Note that \( \mathcal{P}_p \) is unique and of full column rank and that

\[
\text{vech } \mathbf{A} = \mathcal{P}_p \text{ vec } \mathbf{A} = (\mathcal{P}_p' \mathcal{P}_p)^{-1} \mathcal{P}_p' \mathcal{P}_p \text{ vech } \mathbf{A}.
\]

There are many linear transformations of vec \( \mathbf{A} \) into vech \( \mathbf{A} \), but the transformation \( \mathcal{P}_p \), which is the Moore-Penrose generalized inverse of \( \mathcal{P}_p \), is particularly useful.

The direct, or Kronecker product of two matrices will also be used frequently.

**Definition 2.4.** The Kronecker product of a \( p \times q \) matrix \( \mathbf{A} = (a_{ij}) \) and an \( m \times n \) matrix \( \mathbf{B} \), denoted by \( \mathbf{A} \otimes \mathbf{B} \), is the \( pm \times qn \) matrix
2. Order in probability

The notion of order in probability will provide a powerful tool for obtaining the large sample results in later chapters. For sequences of random variables, definitions of order in probability were introduced by Mann and Wald (1943). Let \( \{X_n\} \) be a sequence of \( k \)-dimensional random variables and \( \{g_n\} \) a sequence of positive real numbers. Also, let \( X_{jn} \) be the \( j \)-th element of \( X_n \).

**Definition 2.5.** We say \( X_n \) is of smaller order in probability than \( g_n \) and write

\[
X_n = o_p(g_n),
\]

if, for every \( \varepsilon > 0 \) and \( \delta > 0 \) there exists an \( N \) such that for all \( N, \)

\[
P \left( |X_{jn}| > \varepsilon g_n \right) < \delta, \quad j = 1, 2, \ldots, k.
\]

**Definition 2.6.** We say \( X_n \) is at most of order in probability \( g_n \) and write

\[
a_n \cdot \begin{bmatrix}
a_{11} b & a_{12} b & \cdots & a_{1q} b \\
a_{21} b & a_{22} b & \cdots & a_{2q} b \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1} b & a_{p2} b & \cdots & a_{pq} b
\end{bmatrix}.
\]
if, for every $\epsilon > 0$, there exists a positive real number $M_\epsilon$ such that

$$P \{ |X_{jn}| > M_\epsilon g_n \} \leq \epsilon, \quad j=1,2,\ldots, k,$$

for all $n$.

We do not present various properties of orders in probability which will be used in later chapters. Many useful properties of orders in probability are given in Chapter 5 of Fuller (1976).

B. Theorems

1. Matrix theorems

We now present some theorems to be used in succeeding chapters.

We begin by presenting some useful results in matrix algebra.

**Theorem 2.1.** Let

$$\Sigma_{\xi\xi} = \begin{pmatrix} \Sigma_{ee} & \Sigma_{eu} \\ \Sigma_{ue} & \Sigma_{uu} \end{pmatrix}$$

be a $p \times p$ symmetric positive definite matrix, where $\Sigma_{ee}$ is $r \times r$.

Let $k = p - r$, and let $g$ be a $k \times r$ matrix. Then,

$$[(g, I_r) \Sigma_{\xi\xi}^{-1} (g, I_r)]^{-1} = \Sigma_{uu} - \Sigma_{uv} \Sigma_{vv}^{-1} \Sigma_{vu},$$
\[
((\mathbf{e}, \mathbf{I}) \Sigma_{eE}^{-1} (\mathbf{e}, \mathbf{I})')^{-1} (\mathbf{e}, \mathbf{I}) \Sigma_{eE}^{-1} (\mathbf{e}, \mathbf{I})' = -\Sigma_{uv} \Sigma_{vv}^{-1},
\]

\[
\mathbf{e}' [((\mathbf{e}, \mathbf{I}) \Sigma_{eE}^{-1} (\mathbf{e}, \mathbf{I})')^{-1} (\mathbf{e}, \mathbf{I}) \Sigma_{eE}^{-1} (\mathbf{I}, \mathbf{Q})'] - \mathbf{I} = -\Sigma_{ev} \Sigma_{vv}^{-1},
\]

where

\[
\Sigma_{uv} = \Sigma_{vu} = \Sigma_{ue} - \Sigma_{uu} \mathbf{g},
\]

\[
\Sigma_{ev} = \Sigma_{ee} - \Sigma_{eu} \mathbf{g}.
\]

Proof. By the formula for the inverse of a partitioned symmetric positive definite matrix,

\[
((\mathbf{e}, \mathbf{I}) \Sigma_{eE}^{-1} (\mathbf{e}, \mathbf{I})')
\]

\[
= (\mathbf{e}, \mathbf{I})
\]

\[
= \left(\begin{array}{c}
\Sigma_{uu}^{-1} - \Sigma_{uu}^{-1} \Sigma_{ue} \Sigma_{uu}^{-1} \\
- \Sigma_{uu}^{-1} \Sigma_{ue} \Sigma_{uu}^{-1}
\end{array}\right) \left(\begin{array}{c}
\Sigma_{uu}^{-1} + \Sigma_{uu}^{-1} \Sigma_{ue} \Sigma_{uu}^{-1}
\end{array}\right)
\]

\[
= \Sigma_{uu}^{-1} + (\mathbf{g} - \Sigma_{uu}^{-1} \Sigma_{ue}) \Sigma_{uu}^{-1} (\mathbf{g}' - \Sigma_{ev} \Sigma_{vv}^{-1}),
\]

where

\[
\mathbf{g} = \Sigma_{ee} - \Sigma_{eu} \Sigma_{uu}^{-1} \Sigma_{ue}.
\]
By the formula for the inverse of a sum of matrices,

\[
[(\xi, \iota) \Sigma_{\xi \iota}^{-1}(\xi, \iota)']^{-1}
\]

\[
= \Sigma_{\xi \xi} - \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\xi \iota}) [\xi + (\xi' - \xi_{\xi \iota} \Sigma_{\xi \xi}^{-1})]
\]

\[
= \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\iota \iota})^{-1} (\xi' - \xi_{\xi \iota} \Sigma_{\xi \xi}^{-1}) \Sigma_{\xi \xi}
\]

\[
= \Sigma_{\xi \xi} - (\Sigma_{\xi \xi} (\xi - \Sigma_{\iota \iota} (\xi_{\iota \iota} - \xi_{\xi \iota} \xi_{\xi \iota}^{-1} - \xi_{\iota \iota} + \xi_{\iota \iota} \xi_{\xi \iota}^{-1})))
\]

\[
= \Sigma_{\xi \xi} - \Sigma_{\xi \iota} \Sigma_{\iota \iota}^{-1} \xi_{\xi \iota}
\]

Also,

\[
[(\xi, \iota) \Sigma_{\xi \iota}^{-1}(\xi, \iota)']^{-1} (\xi, \iota) \Sigma_{\xi \iota}^{-1} (\iota, \xi)'
\]

\[
= \{ \Sigma_{\xi \xi} - \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\iota \iota}) [\xi + (\xi' - \xi_{\xi \iota} \Sigma_{\xi \xi}^{-1})]
\]

\[
= \{ \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\iota \iota})^{-1} (\xi' - \xi_{\xi \iota} \Sigma_{\xi \xi}^{-1}) \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\iota \iota}) \xi_{\xi \iota}^{-1}
\]

\[
= (\Sigma_{\xi \iota} (\xi - \Sigma_{\iota \iota} \xi_{\iota \iota}^{-1}) [\xi + (\xi' - \xi_{\xi \iota} \Sigma_{\xi \iota}^{-1})]
\]

\[
= \Sigma_{\xi \xi} (\xi - \Sigma_{\xi \xi}^{-1} \xi_{\iota \iota})^{-1} \xi_{\xi \iota}^{-1}
\]
\[= - (\Sigma_{ee} - \Sigma_{eu} \xi) (\Sigma_{ee} - \xi' \Sigma_{ee} \xi + \xi' \Sigma_{uu} \xi)\]
\[= - \xi' \Sigma_{uu} \xi^{-1} \cdot\]

Finally,

\[\xi'(\xi, \iota) \Sigma_{ee}^{-1} (\xi, \iota)'^{-1}(\xi, \iota) \Sigma_{ee}^{-1} (\iota, \emptyset)' - \iota\]
\[= - \xi' \Sigma_{uv} \xi_{vv}^{-1} - \iota\]
\[= - (\xi' \Sigma_{uv} + \Sigma_{vv}) \xi_{vv}^{-1}\]
\[= - (\Sigma_{ee} - \Sigma_{eu} \xi) \xi_{vv}^{-1} \]
\[= - \xi_{ev} \xi_{vv}^{-1} \cdot\]

The next two theorems provide useful relationships involving vec operators. The proofs are elementary and thus omitted. Dahm (1979), for example, gives proofs of both the theorems.

Theorem 2.2. Let \(A, B, \) and \(C\) be \(p \times q, q \times m,\) and \(m \times n\) matrices, respectively. Then,

\[\text{vec} (A \otimes B) = (C' \otimes A) \text{vec} \ B.\]
Theorem 2.3. Let $A$, $B$, $C$, and $D$ be $p \times q$, $q \times m$, $p \times n$, and $n \times m$ matrices, respectively. Then,

$$\text{tr } (A \ B \ D' \ C') = (\text{vec } A)' (B \otimes C) (\text{vec } D).$$

The next two theorems give formulae for derivatives of matrix functions. For the proofs, see, for example, Dahm (1979).

Theorem 2.4. Let $A = \xi_i(\gamma)$ be a $p \times p$ symmetric positive definite matrix, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)'$. Then, for $i = 1, 2, \ldots, r$,

$$\frac{\partial \log |A|}{\partial \gamma_i} = \text{tr } (A^{-1} \frac{\partial A}{\partial \gamma_i}) .$$

Theorem 2.5. Let $A = \xi_i(\gamma)$ be a $p \times p$ nonsingular matrix, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)'$. Then, for $i = 1, 2, \ldots, r$,

$$\frac{\partial A^{-1}}{\partial \gamma_i} = -A^{-1} \frac{\partial \xi_i}{\partial \gamma_i} A^{-1} .$$

The proof of the next theorem is given in Bellman (1970, p. 115).

Theorem 2.6. (Courant-Fischer min-max theorem)

Let $A$ be a $n \times n$ symmetric matrix. Let

$\lambda_1 > \lambda_2 > \ldots > \lambda_n$ be eigenvalues of $A$. Then

$$\lambda_1 = \max \frac{\mathbf{x}' A \mathbf{x}}{\mathbf{x}' \mathbf{x}},$$

$$\frac{\mathbf{x}_1}{\mathbf{x}_1} \neq \mathbf{x}_2 \neq \frac{\mathbf{x}_2}{\mathbf{x}_2} \neq \ldots \neq \frac{\mathbf{x}_n}{\mathbf{x}_n}$$
\[
\lambda_2 = \min_{x \neq 0} \max_{x'^{\top}y = 0} \frac{x'^{\top}Ax}{x'^{\top}x}, \\
\lambda_k = \min_{x_i \neq 0} \max_{x'^{\top}y_i = 0} \frac{x'^{\top}Ax}{x'^{\top}x}, \quad i = 1, \ldots, (k-1) \\
\lambda_n = \min_{x_i \neq 0} \max_{x'^{\top}y_i = 0} \frac{x'^{\top}Ax}{x'^{\top}x}, \quad i = 1, \ldots, (n-1)
\]

Note that the Courant-Fischer min-max theorem is typically stated with an extra condition

\[
y_i^t y_i = 1, \quad i = 1, 2, \ldots, (k-1),
\]

on the vectors over which the minimum is evaluated. Obviously, such a condition does not change the minimum value.

**Theorem 2.7.** Let \( A \) be a \( n \times n \) symmetric positive semidefinite matrix. Let \( X \) be any \( n \times p \) (\( p < n \)) matrix satisfying

\[
X'^{\top}X = I_{p \times p}.
\]  
(2.1)

Then,
\[ \lambda_k [P'AP] > \lambda_{n-p+k} [A], \quad k = 1, 2, \ldots, p, \tag{2.2} \]

where \( \lambda_k [\cdot] \) denotes the \( k \)-th largest eigenvalue of a symmetric matrix. The equality in (2.2) holds for all \( k = 1, 2, \ldots, p \), when

\[ P = Q_2 G, \tag{2.3} \]

where \( Q_2 \) is the matrix of orthonormal eigenvectors of \( A \) corresponding to \( p \) smallest eigenvalues, and \( G \) is a \( p \times p \) orthogonal matrix. If \( \lambda_{n-p} [A] > \lambda_{n-p+1} [A] \), then the equality in (2.2) holds for all \( k = 1, 2, \ldots, p \), if and only if \( P \) is of the form (2.3).

**Proof.** For \( k = 2, 3, \ldots, p \), by Theorem 2.6 and (2.1),

\[ \lambda_k [P'AP] = \min_{X_i \neq 0} \max_{i=1, \ldots, (k-1)} \frac{X_i'P'APX_i}{X_i'X_i}, \tag{2.4} \]

where

\[ C(P) = \{ \text{column space of } P \}, \]

and \( C(P) \) is the column space of \( P \). By (2.1),
Let \( \xi_j \), \( j = 1,2,..., (n-p) \) be a basis of the orthogonal complement \( C^-(\mathcal{P}) \) of \( C(\mathcal{P}) \). Then,

\[
\mathcal{S}_{k,\mathcal{Y}} = \{ z \in \mathbb{R}^n; z'\xi_i = 0, i=1,...,(k-1), z'\xi_j = 0, j=1,2,...,(n-p) \},
\]

(2.5)

where

\[
\xi_i = \mathcal{P} \xi_i, \quad i = 1,2,...,(k-1).
\]

(2.6)

Therefore, by (2.4), (2.5), and (2.6),

\[
\lambda_k[\mathcal{P}' \mathcal{A} \mathcal{P}] = \min_{\xi_i \neq 0} \max_{\xi_j = 0} \frac{z'Ax}{z'z}
\]

\[
= \begin{cases} 
\xi_i \in \mathcal{C}(\mathcal{P}) & i=1,...,(k-1) \\
\xi_i \in \mathcal{C}^-(\mathcal{P}) & j=1,...,(n-p)
\end{cases}
\]

(2.4)
where the last equality follows from Theorem 2.6. For \( k = 1 \), by Theorem 2.6 and (2.1),

\[
\lambda_1 [P' A P] = \max_{\xi} \frac{\xi' P' A P \xi}{\xi' \xi} = \max_{\xi \in C(\xi)} \frac{\xi' A \xi}{\xi' \xi}
\]

\[
= \max_{\xi' \xi_j = 0} \frac{\xi' A \xi}{\xi' \xi}
\]

\[
j = 1, \ldots, (n-p)
\]
\[ \min_{u_j \neq 0} \frac{x'Az}{x'x} \quad \max_{x'z_j = 0} \frac{x'Az}{x'x} \quad j=1,\ldots,(n-p) \quad j=1,\ldots,(n-p) \]

= \lambda_{n-p+1}[A].

If \( P = Q_2G \) for some \( p \times p \) orthogonal matrix, then for \( k=1,2,\ldots, p \),

\[ \lambda_k[P'AP] = \lambda_k[G'Q_2AQ_2G] \]

= \lambda_k[G'Q_2AQ_2]

= \lambda_k[diag \{ \lambda_{n-p+1}[A], \ldots, \lambda_n[A] \}] 

= \lambda_{n-p+k}[A].

Suppose that \( \lambda_{n-p}[A] > \lambda_{n-p+1}[A] \). Then, the eigenspace \( E_2 \) of \( A \) corresponding to \( p \) smallest eigenvalues is unique. Let \( H \) be any \( n \times p \) matrix satisfying

\[ H'AH = diag \{ \lambda_{n-p+1}[A], \ldots, \lambda_n[A] \}, \quad (2.7) \]

and

\[ H'H = I_{p \times p}. \quad (2.8) \]
Suppose there exists an \( n \times 1 \) vector \( \mathbf{h} \) belonging to the column space of \( \mathbf{H} \) but not belonging to \( \mathbf{E}_2 \). Then, \( \mathbf{h} \) must belong to the eigenspace of \( \mathbf{A} \) corresponding to the largest \( (n-p) \) eigenvalues. Thus, there exists a \( p \times 1 \) vector \( \xi \) and a \( (n-p) \times 1 \) vector \( \xi' \) such that

\[
\mathbf{h} = \mathbf{H} \xi, \quad (2.9)
\]

\[
\mathbf{h} = \mathbf{Q}_1 \xi', \quad (2.10)
\]

where \( \mathbf{Q}_1 \) is the matrix of orthonormal eigenvectors of \( \mathbf{A} \) corresponding to the largest \( (n-p) \) eigenvalues. Now, by (2.7) and (2.9),

\[
\mathbf{h}' \mathbf{A} \mathbf{h} = \xi' \mathbf{H}' \mathbf{A} \mathbf{H} \xi
\]

\[
= \xi' \text{diag} \{ \lambda_{n-p+1} [\mathbf{A}], \ldots, \lambda_n [\mathbf{A}] \} \xi
\]

\[
< \lambda_{n-p+1} [\mathbf{A}] \xi' \xi.
\]

By (2.10),

\[
\mathbf{h}' \mathbf{A} \mathbf{h} = \xi' \mathbf{Q}_1' \mathbf{A} \mathbf{Q}_1 \xi
\]

\[
= \xi' \text{diag} \{ \lambda_1 [\mathbf{A}], \ldots, \lambda_{n-p} [\mathbf{A}] \} \xi
\]

\[
> \lambda_{n-p} [\mathbf{A}] \xi' \xi.
\]

Since \( \lambda_{n-p} [\mathbf{A}] > \lambda_{n-p+1} [\mathbf{A}] \), and since \( \xi' \xi = \xi' \xi \) by (2.8), (2.9),
and (2.10), we have arrived at a contradiction. Hence, every vector in the column space of \( \mathbf{H} \) is also in \( E_2 \). Therefore, any matrix \( \mathbf{H} \) satisfying (2.7) and (2.8) must have the form

\[
\mathbf{H} = \mathbf{Q}_2 \mathbf{G},
\]

where \( \mathbf{G} \) is some \( p \times p \) orthogonal matrix. Now, suppose that

\[
\lambda_k \{ \mathbf{P}'\mathbf{A}\mathbf{P} \} = \lambda_{n-p+k} \{ \mathbf{A} \}, \quad k = 1, 2, \ldots, p.
\]

Then,

\[
\mathbf{P}'\mathbf{P}'\mathbf{A}\mathbf{P} = \text{diag} \{ \lambda_{n-p+1} \{ \mathbf{A} \}, \ldots, \lambda_n \{ \mathbf{A} \} \},
\]

where \( \mathbf{P} \) is the matrix of orthogonal eigenvectors of \( \mathbf{P}'\mathbf{A}\mathbf{P} \). Thus, by the above result (2.11), there exists a \( p \times p \) orthogonal matrix \( \mathbf{Q} \) such that

\[
\mathbf{P}\mathbf{Q} = \mathbf{Q}_2 \mathbf{G}.
\]

Therefore,

\[
\mathbf{P} = \mathbf{Q}_2 \mathbf{F},
\]

where
This concludes the proof of the necessity part for the equality and thus the proof of Theorem 2.7.

The following theorem provides a useful representation of the fourth moments of a normal random vector. The proof is given in Dahm (1979).

**Theorem 2.8.** Let $\mathbf{X}$ be a $p \times 1$ multivariate normal random vector with mean zero and covariance matrix $\Sigma$. Then,

$$V\{\text{vech}(XX')\} = 2 \psi_p (\Sigma \otimes \Sigma) \psi_p' .$$

2. Central limit theorems

In later chapters, we derive the limiting distributions of estimators. Hence, it is convenient to summarize various forms of the central limit theorems in this section.

**Theorem 2.9.** Let $(X_t, Y_t)$ be $(p+q)$-dimensional independently and identically distributed random row vectors. Assume that $X_t$ and $Y_t$ are independent, and that $(X_t, Y_t)$ has finite fourth moments. Let
\[ u_X = E\{\xi_t\} , \]
\[ u_Y = E\{\eta_t\} , \]
\[ \xi_{XX} = V\{\xi_t\} , \]
\[ \xi_{YY} = V\{\xi_t\} , \]
\[ \xi_X = V\{\text{vech}[((\xi_t - u_X)'(\xi_t - u_X)])\} , \]
\[ \xi_Y = V\{\text{vech}[((\eta_t - u_Y)'(\eta_t - u_Y)])\} , \]
\[
\begin{pmatrix}
\psi_{XX} & \psi_{XY} \\
\psi_{YX} & \psi_{YY}
\end{pmatrix}
\]
\[ = (n-1)^{-1} \sum_{t=1}^{n} [(\xi_t, \xi_t) - (\overline{\xi}, \overline{\xi})]'[(\xi_t, \xi_t) - (\overline{\xi}, \overline{\xi})] , \]
\[ (\overline{\xi}, \overline{\eta}) = n^{-1} \sum_{t=1}^{n} (\xi_t, \xi_t) , \]
\[
\xi = \begin{pmatrix}
\xi_{XX} & 0 \\
0 & \xi_{YY}
\end{pmatrix}
\].

Then,
\[
\begin{pmatrix}
\text{vech}(\psi_{XX} - \xi_{XX}) \\
\text{vec} \psi_{XY} \\
\text{vech}(\psi_{YY} - \xi_{YY})
\end{pmatrix}
\xrightarrow{L} N \begin{pmatrix}
Q , \\
\xi_X & 0 & 0 \\
0 & \xi_{XX} + \xi_{YY} & 0 \\
0 & 0 & \xi_Y
\end{pmatrix} .
\]
Proof. We have

\[ m = (n-1)^{-1} \sum_{t=1}^{n} [(\xi_t, \xi_t) - (\xi_X, \xi_Y)]'(\xi_t, \xi_t) - (\xi_X, \xi_Y)] 

- (n-1)^{-1} n[(\bar{\xi}, \bar{\xi}) - (\xi_X, \xi_Y)]'(\bar{\xi}, \bar{\xi}) - (\xi_X, \xi_Y)] 

= n^{-1} \sum_{t=1}^{n} [(\xi_t - \xi) - (\xi_X, \xi_Y)]'(\xi_t - \xi) - (\xi_X, \xi_Y)] + O_p(n^{-1}).

Hence,

\[
\begin{bmatrix}
\text{vech } m_{XX} \\
\text{vec } m_{YX} \\
\text{vech } m_{YX}
\end{bmatrix}
= n^{-1} \sum_{t=1}^{n} \bar{W}_t + O_p(n^{-1}),
\]

where

\[
\bar{W}_t = \begin{bmatrix}
\xi_{1t} \\
\xi_{2t} \\
\xi_{3t}
\end{bmatrix}
= \begin{bmatrix}
\text{vech}(\xi_t - \xi)(\xi_t - \xi) \\
\text{vec } (\xi_t - \xi)'(\xi_t - \xi) \\
\text{vech}(\xi_t - \xi)'(\xi_t - \xi)
\end{bmatrix}.
\]

Now, \( \bar{W}_t \) are independently and identically distributed random variables such that

\[
\mathbb{E}[\bar{W}_t] = \begin{bmatrix}
\text{vech } \xi_{XX} \\
0 \\
\text{vech } \xi_{YY}
\end{bmatrix},
\]

\[
\text{Var}[\bar{W}_t] = \xi_X.
\]
where we have used the independence of $\zeta_t$ and $\eta_t$. Thus, the result follows from the multivariate central limit theorem. \hfill \Box

**Theorem 2.10.** Let $(\zeta_t, \eta_t)$ be $(p+q)$-dimensional independently and identically distributed random row vectors. Assume that $\zeta_t$ and $\eta_t$ are independent, that $\eta_t$ has finite fourth moments, and that $\zeta_t$ has finite second moments. Then,
\[ \frac{1}{n} \left[ \begin{array}{c} \text{vec } \Sigma_{XX} \\ \text{vec}(\Sigma_{YY} - \Sigma_{XY}) \end{array} \right] \xrightarrow{L} N \left( \begin{array}{ccc} 0 \\ 0 \\ 0 \end{array} \right), \]

where the notations are the same as in Theorem 2.9.

**Proof.** Since the independence and the existence of second moments of \((X_t, Y_t)\) imply the existence of second moments of \(\text{vec } \Sigma_{XX}\), the result follows by the argument used in the proof of Theorem 2.9. \(\Box\)

The next lemma will be used to prove some central limit theorems.

**Lemma 2.1.** Let \(\{A_t\}\) be a sequence of real numbers such that

\[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} A_t = A. \]

Then,

\[ \lim_{n \to \infty} n^{-1} A_n = 0, \]

and

\[ \lim_{n \to \infty} n^{-1} \sup_{1 \leq t \leq n} |A_t| = 0. \]

**Proof.** We observe that

\[ n^{-1} A_n = n^{-1} \sum_{t=1}^{n} A_t - n^{-1} (n-1)^{-1} \sum_{t=1}^{n} A_t \]

\[ \to 0, \text{ as } n \to \infty. \]
Thus, for every $\varepsilon > 0$, there exists a $T$ such that for all $t > T$ \[ \left| \frac{1}{t} A_t \right| < \varepsilon. \]

Hence, for all $n > \max \{ T, \varepsilon^{-1} \sup_{1 \leq t < T-1} |A_t| \},$ \[ n^{-1} \sup_{1 \leq t \leq n} |A_t| = \max \{ n^{-1} \sup_{1 \leq t < T-1} |A_t|, n^{-1} \sup_{T \leq t \leq n} |A_t| \} \]

\[ < \max \{ \varepsilon, \sup_{T \leq t \leq n} \left| \frac{1}{t} A_t \right| \} \]

\[ < \varepsilon. \]

\[ \square \]

The following three central limit theorems will be useful in later chapters.

**Theorem 2.11.** Let $\{\xi_t\}$ be a sequence of independently and identically distributed $k$-dimensional random column vectors. Assume $E[\xi_t] = 0$ and $\text{V} \{\xi_t\} = \xi$. Let $\{A_t\}$ be a sequence of $k$-dimensional real column vectors such that \[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} A_t A_t' = \xi. \] (2.12)

Then, \[ n^{-1/2} \sum_{t=1}^{n} A_t' \xi_t \overset{L}{\longrightarrow} N(Q, \text{tr}[\xi \xi']), \] (2.13)

and \[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \text{V} \{A_t' \xi_t\} = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} A_t' \xi A_t. \]
= \text{tr}[\xi G] .

Proof. Let
\[
V_n = V\left\{ \sum_{t=1}^{n} A_t^T X_t \right\} = \sum_{t=1}^{n} A_t^T \xi A_t .
\]

Then, by (2.12),
\[
\lim_{n \to \infty} n^{-1} V_n = \lim_{n \to \infty} \text{tr}[\xi n^{-1} \sum_{t=1}^{n} A_t^T A_t'] = \text{tr}[\xi G] . \tag{2.14}
\]

If \( \text{tr}[\xi G] = 0 \), then for every \( \epsilon > 0 \),
\[
P\left\{ n^{-1/2} \sum_{t=1}^{n} A_t^T X_t > \epsilon \right\} < \epsilon^{-2} n^{-1} V_n \longrightarrow 0 , \text{ as } n \to \infty ,
\]
and thus (2.13) holds with the limit given by the degenerate distribution. Now we assume \( \text{tr}[\xi G] > 0 \). Let \( \epsilon > 0 \) be given, and let
\[
S_{nt\epsilon} = \{ \xi \in R : (A_t^T \xi)^2 > \epsilon^2 V_n \} ,
\]
\[
S_{nt\epsilon}^* = \{ \xi \in R : (\xi' \xi) > \epsilon^2 V_n [\sup_{1 \leq t \leq n} A_t^T A_t]^{-1} \} .
\]

Then, by Hölder's inequality, \( S_{nt\epsilon} \) is a subset of \( S_{nt\epsilon}^* \). Hence,
\[
\frac{1}{n} \sum_{t=1}^{n} \int_{S_{nt \in \mathbb{E}}} (\xi_t \xi_t^2) \ d \xi (\xi) \leq \frac{1}{n} \sum_{t=1}^{n} \int_{S_{nt \in \mathbb{E}}} (\xi_t \xi_t)^2 \ d \xi (\xi)
\]

\[
\leq \frac{1}{n} \left( \sum_{t=1}^{n} \xi_t \xi_t \right) \int_{S_{nt \in \mathbb{E}}} (\xi' \xi) \ d \xi (\xi).
\]

Since,

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \xi_t \xi_t = \lim_{n \to \infty} \text{tr}[n^{-1} \sum_{t=1}^{n} \xi_t \xi_t'] = \text{tr} G, \quad (2.16)
\]

by Lemma 2.1,

\[
\lim_{n \to \infty} n^{-1} \sup_{1 \leq t \leq n} (\xi_t' \xi_t') = 0. \quad (2.17)
\]

By (2.14) and (2.17),

\[
\lim_{n \to \infty} \epsilon^2 V_n \left[ \sup_{1 \leq t \leq n} (\xi_t' \xi_t') \right]^{-1} = \lim_{n \to \infty} \epsilon^2 n^{-1} V_n \left[ n^{-1} \sup_{1 \leq t \leq n} (\xi_t' \xi_t') \right]^{-1}
\]

\[
= \infty. \quad (2.18)
\]

Since \( \xi_t \) has second moments,

\[
\sigma^2 = \int (\xi' \xi) \ d \xi (\xi) < \infty. \quad (2.19)
\]
By (2.18) and Chebyshev's inequality,

$$P\{X_t \in S_{ne}\} < \varepsilon^2 \inf_{1 \leq t \leq n} \left( (A_t' A_t)^{-1} \right)^{-1} \to 0 \text{, as } n \to \infty$$ (2.20)

By (2.19) and (2.20),

$$\lim n \sup_{S_{ne}} \left( (x'x) d F (x) \right) = 0 .$$ (2.21)

Therefore, by (2.14), (2.15), (2.16), and (2.21),

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} (A_t' x)^2 d F (x) = 0 .$$

Thus, by the Lindeberg central limit theorem and (2.14),

$$n^{-1/2} \sum_{t=1}^{n} A_t' X_t = \left( n^{-1/2} \sum_{t=1}^{n} A_t' X_t \right) \to N (0, \text{tr}[\xi G]) .$$

Theorem 2.12. Let \( \{X_t\} \) be a sequence of independently and identically distributed k-dimensional random column vectors. Assume \( E[X_t] = 0 \) and \( V[X_t] = \xi \). Let \( \{A_t\} \) be a sequence of \( k \times p \) real matrices such that

$$\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \text{vec} A_t (\text{vec} A_t)' = \mathbb{I}$$

...
\[
\begin{pmatrix}
\mathbb{H}_{11} & \cdots & \mathbb{H}_{1p} \\
\vdots & \ddots & \vdots \\
\mathbb{H}_{p1} & \cdots & \mathbb{H}_{pp}
\end{pmatrix},
\]
where each \( \mathbb{A}_{ij} \) is a \( k \times k \) matrix. Then,

\[
n^{-\frac{1}{2}} \sum_{t=1}^{n} \mathbb{A}_t^\prime \mathbb{X}_t \xrightarrow{L} N(\mathbf{Q}, \mathbf{V}),
\]

where

\[
\mathbf{V} = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \mathbb{A}_t^\prime \mathbb{X}_t \mathbb{A}_t
\]

\[
\begin{pmatrix}
\text{tr}(\mathbb{X} \mathbb{H}_{11}) & \cdots & \text{tr}(\mathbb{X} \mathbb{H}_{1p}) \\
\vdots & \ddots & \vdots \\
\text{tr}(\mathbb{X} \mathbb{H}_{p1}) & \cdots & \text{tr}(\mathbb{X} \mathbb{H}_{pp})
\end{pmatrix},
\]

**Proof.** Let \( \lambda = (\lambda_1, \ldots, \lambda_p)' \) be an arbitrary nonnull \( p \times 1 \) real vector, and consider

\[
\lambda^\prime n^{-\frac{1}{2}} \sum_{t=1}^{n} \mathbb{A}_t^\prime \mathbb{X}_t = n^{-\frac{1}{2}} \sum_{t=1}^{n} \lambda^\prime \mathbb{A}_t^\prime \mathbb{X}_t.
\]
Let $A_t^{(i)}$ be the $i$-th column of $A_t$. Then,

$$n^{-1} \sum_{t=1}^{n} A_t A_t' = n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{p} \lambda_i A_t^{(i)} \sum_{j=1}^{p} \lambda_j A_t^{(j)'}$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} \left\{ \lambda_i \lambda_j n^{-1} \sum_{t=1}^{n} A_t^{(i)} A_t^{(j)'} \right\}$$

$$\longrightarrow \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_i \lambda_j R_{ij}, \text{ as } n \to \infty.$$

Thus, by Theorem 2.11,

$$\lambda' n^{-1/2} \sum_{t=1}^{n} A_t A_t' \overset{L}{\longrightarrow} N(0, \text{tr} \left[ \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_i \lambda_j R_{ij} \right]).$$

We observe that

$$\text{tr} \left[ \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_i \lambda_j R_{ij} \right] = \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_i \lambda_j \text{tr} \left[ R_{ij} \right]$$

$$= \lambda' \Lambda \lambda.$$

Hence, the result follows from the multivariate central limit theorem.

\[\blacksquare\]

Theorem 2.13. Let $X_t$ be a sequence of $p$-dimensional real row vectors such that

$$\lim_{n \to \infty} \frac{X_t}{n} = \frac{\mu}{n},$$

where
Let $\xi_t$ be $q$-dimensional independently and identically distributed random column vectors with mean $\mu$, covariance matrix $\Sigma_{\xi\xi}$, and finite fourth moments. Then,

$$m_{\xi\xi} = (n-1)^{-1} \sum_{t=1}^{n} (\xi_t - \bar{\xi}')(\xi_t - \bar{\xi})',$$

$$\bar{\xi} = n^{-1} \sum_{t=1}^{n} \xi_t.$$

Proof. Let $\lambda' = (\lambda_1', \lambda_2')$ be an arbitrary $[pq+2^{-1}q(q+1)]$-dimensional nonnull real vector. Now, using the double subscript notation for
\[ \lambda_1 \text{ and } \lambda_2, \text{ we have} \]

\[
\begin{bmatrix}
\text{vec } \Sigma_{XX} \\
\text{vech}(\Sigma_{XX} - \Sigma_{XX})
\end{bmatrix}
\]

\[ = \lambda_1' \text{ vec } \Sigma_{XX} + \lambda_2' \text{ vech}(\Sigma_{XX} - \Sigma_{XX}) \]

\[ = \lambda_1'(n-1)^{-1} \sum_{t=1}^{n} \text{ vec}[(\xi_t - \bar{\xi})'\xi_t'] + \lambda_2'(n-1)^{-1} \sum_{t=1}^{n} \text{ vech}(\xi_t\xi_t') \]

\[ - \lambda_2' \text{ vech } \Sigma_{XX} + O_p(n^{-1}) \]

\[ = (n-1)^{-1} \sum_{t=1}^{n} \frac{p}{q} \sum_{i=1}^{\frac{q}{p}} \lambda_{ij} (x_{it} - \bar{x}_i) \]

\[ + (n-1)^{-1} \sum_{t=1}^{n} \sum_{m=1}^{q} \sum_{k=1}^{q} \lambda_{2km} (e_{kt} e_{mt} - \sigma) + O_p(n^{-1}) \]

\[ = (n-1)^{-1} \sum_{t=1}^{n} \left\{ \sum_{j=1}^{q} w_{jt} e_{jt} + \sum_{m=1}^{q} \sum_{k=1}^{q} \lambda_{2km} (e_{kt} e_{mt} - \sigma) \right\} \]

\[ + O_p(n^{-1}) \]

\[ = n^{-1} \sum_{t=1}^{n} \xi_t'y_t + O_p(n^{-1}) , \quad (2.22) \]

where

\[ w_{jt} = \frac{p}{q} \lambda_{ij} (x_{it} - \bar{x}_i) , \]

\[ z_t' = (w_{1t}, \ldots, w_{qt}, \lambda_{211}, \ldots, \lambda_{2qq}) , \]
\[ Y_t = (\xi_t', \text{vech}(\xi_t' \xi_t - \xi_{ee})')' \]

By the assumptions, \( Y_t \) are independently and identically distributed, and satisfy

\[ E\{a_t'Y_t\} = 0 \]
\[ V\{a_t'Y_t\} = a_t'\Sigma a_t \]

where

\[ \Sigma = \begin{pmatrix}
\xi_{ee} & \xi_3 \\
\xi_3' & \Gamma_{ee}
\end{pmatrix} \]

\[ \xi_3 = E\{\xi_t' \text{vech}(\xi_t' \xi_t - \xi_{ee})\} \]

Now,

\[ n^{-1} \sum_{t=1}^{n} w_{it} w_{jt} = P \sum_{k=1}^{p} \sum_{w=1}^{p} \lambda_{ki} \lambda_{mj} n^{-1} \sum_{t=1}^{n} (x_{kt} - \bar{x}_k)(x_{mt} - \bar{x}_m) \]
Thus, by (2.23) and Theorem 2.11,
\[ n^{-1/2} \sum_{t=1}^{n} \sum_{t} \xi_{t} \xi_{t} \xrightarrow{L} N(0, \text{tr}(\xi \Lambda)) \]  \hspace{1cm} (2.24)

Using Theorem 2.3 and that

\[ \text{vec} \Lambda_1 = \lambda_1 \]

we have

\[
\text{tr}(\xi \Lambda) = \text{tr} \begin{bmatrix} \xi_{ee} & \bar{\xi}_3 \\ \bar{\xi}_3' & \xi_{ee} \end{bmatrix} \begin{bmatrix} \Lambda_1' \bar{\Xi}_{xx} \Lambda_1 & 0 \\ 0 & \lambda_2 \lambda_2' \end{bmatrix} = \text{tr}(\Lambda_1 \Sigma_{ee} + \bar{\Xi}_{xx}) + \text{tr} \lambda_1' \Sigma_{ee} \lambda_2
\]

\[
= (\text{vec} \Lambda_1)' (\Sigma_{ee} \bar{\Xi}_{xx}) \text{vec} \Lambda_1 + \lambda_1' \Sigma_{ee} \lambda_2
\]

\[
= \lambda_1' (\Sigma_{ee} \bar{\Xi}_{xx}) + \lambda_2' \Sigma_{ee} \lambda_2
\]

\[
\begin{bmatrix} \Sigma_{ee} \bar{\Xi}_{xx} & Q \\ Q & \xi_{ee} \end{bmatrix} \lambda = \lambda
\]

Hence, by (2.22), (2.24), and (2.25),
Thus, the result follows from the multivariate central limit theorem.

\[ \lambda, n^{-1/2} \begin{pmatrix} \text{vec } \mathbf{m}_{EE} \\ \text{vech}(\hat{\Sigma}_{EE} - \hat{\Sigma}_{EE}) \end{pmatrix} \xrightarrow{L} N \left( \begin{array}{c} 0, \lambda' \\ \begin{pmatrix} \hat{\Sigma}_{EE} & \hat{\mathbf{m}}_{XX} \\ \hat{\mathbf{m}}_{XX}^\top & 0 \\ 0 & \hat{\Sigma}_{EE} \end{pmatrix} \end{array} \right) \lambda \]
III. THE MULTIVARIATE LINEAR MODEL WITH ESTIMATED ERROR COVARIANCE MATRIX

In this chapter, we consider the maximum likelihood estimation of the multivariate linear error-in-variables model when there is available an independent estimate of the error covariance matrix. Both structural and functional relationships are considered. The properties of the estimators are derived under relatively weak assumptions.

A. Introduction

One of the most commonly discussed models in the errors-in-variables problem is the model with two variables, no errors in the equation, and error covariance matrix either known or known up to a multiple. We extend this simple model to the case where there are multiple relationships and an independent estimator of the error covariance matrix is available. Let

\[
\begin{align*}
Z_t &= X_t + \varepsilon_t, \\
X_t &= X_t + \tilde{e}_t, \\
\bar{X}_t &= \bar{X}_t + u_t,
\end{align*}
\]  

(3.1)

where \(\bar{X}_t\) is an unobservable \(1 \times k\) vector, \(X_t\) is the \(1 \times k\) vector of observed values of \(X_t\), \(\bar{X}_t\) is the \(1 \times r\) vector of observations on the dependent variables \(\bar{X}_t\), \(p = k+r\), \(\beta_o\) is a \(1 \times r\) vector of unknown parameters, \(\tilde{e}_t\) is a \(k \times r\) matrix of unknown coefficients, and
\( \varepsilon_t = (\varepsilon_t, \eta_t) \) are independently and identically distributed with mean zero and covariance matrix \( \Sigma_{\varepsilon \varepsilon} \). In addition, there is available an estimator \( S_{\varepsilon \varepsilon} \) of \( \Sigma_{\varepsilon \varepsilon} \). We assume \( S_{\varepsilon \varepsilon} \) is distributed independently of \( (X_t, Y_t) \) for all \( t \).

The model just defined arises frequently in practical situations. Physical experiments often have models with no errors in the equations. The model with intercept term \( \beta_0 \) was chosen because of its wider use in practice. Theoretical development for the model without intercept follows with minor modifications from that for the intercept model. Also, it is more realistic to assume the existence of an estimator \( S_{\varepsilon \varepsilon} \) of \( \Sigma_{\varepsilon \varepsilon} \) than to assume that \( \Sigma_{\varepsilon \varepsilon} \) is known or known up to a multiple. We chose to assume that \( S_{\varepsilon \varepsilon} \) is an estimator of \( \Sigma_{\varepsilon \varepsilon} \) rather than of a multiple of \( \Sigma_{\varepsilon \varepsilon} \) because we consider this to be the usual case. There are two common sources for \( S_{\varepsilon \varepsilon} \). Independent experiments in the past often provide such estimators. Also, when replicated observations are measured at some of the true values \( (X_t, Y_t) \), a multiple of the within replication sum of squares can be used as an estimator of \( \Sigma_{\varepsilon \varepsilon} \). Under normality of the errors, the means over replicates used as the data points are independent of the estimator of \( \Sigma_{\varepsilon \varepsilon} \) based on the within sum of squares.

In this chapter, we will call upon one or more of the following assumptions.

**Assumption 3.1.** The \( \beta_t \) are fixed constants satisfying

\[
\lim_{n \to \infty} \bar{X} = \mu_X ,
\]

\[
\lim_{n \to \infty} \bar{Y} = \mu_Y .
\]
\[
\lim_{n \to \infty} \mu_{xx} = \Sigma_{xx},
\]

where

\[
\bar{x} = n^{-1} \sum_{t=1}^{n} x_t,
\]

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

and \( \Sigma_{xx} \) is a \( k \times k \) positive definite matrix. Also,

\( \Sigma_{vv} = (I, -\bar{\beta}') \Sigma_{ee} (I, -\bar{\beta})' \) is positive definite.

**Assumption 3.2.** The \( x_t \) are independently and identically distributed with mean \( \mu_x \) and covariance matrix \( \Sigma_{xx} \). The \( x_t \) are independent of \( \varepsilon_t \), for all \( t \) and \( t' \). The covariance matrices \( \Sigma_{xx} \) and

\[
\Sigma_{zz} = (\bar{z}, \bar{\ell})' \Sigma_{xx} (\bar{z}, \bar{\ell}) + \Sigma_{ee}
\]

are positive definite.

**Assumption 3.3.** The fourth moments of \( x_t \) exist.

**Assumption 3.4.** The \( x_t \) are normally distributed.

**Assumption 3.5.** The \( \varepsilon_t \) are independently and identically distributed with mean \( 0 \) and nonsingular covariance matrix \( \Sigma_{ee} \).

**Assumption 3.6.** The \( \varepsilon_t \) are normally distributed.

**Assumption 3.7.** The \( \varepsilon_{ee} \) is independent of \( Z_t \) for all \( t \) and is based on \( d \) degrees of freedom. Also
Assumption 3.8. The distribution of $d S_{\varepsilon \varepsilon}$ is the Wishart distribution with covariance matrix $S_{\varepsilon \varepsilon}$ and degrees of freedom $d$.

B. Maximum Likelihood Estimators

We derive the maximum likelihood estimators of the parameters of the model (3.1), assuming that $S_{\varepsilon \varepsilon}$ is a multiple of a Wishart matrix, and that all other random variables in the model have a multivariate normal distribution. The functional case, where $X_t$ are fixed, and the structural case, where $X_t$ are random, are treated separately. First, we introduce some notations. Let

$$Z_t = (Y_t, X_t),$$

$$z_t = (\bar{Y}, \bar{X}) = \frac{1}{n} \sum_{t=1}^{n} Z_t = (\bar{Y}, \bar{X}),$$

$$E_{ZZ} = (n-1)^{-1} \sum_{t=1}^{n} (Z_t - \bar{Z})'(Z_t - \bar{Z}).$$
Also, let \( \hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_p \geq 0 \) be the eigenvalues of 
\( S_{ee}^{-1/2} \mathbf{m}_{ZZ} S_{ee}^{-1/2} \), and let \( \hat{q}_i, i = 1, 2, \ldots, p, \) be the corresponding 
eigenvectors such that

\[
S_{ee}^{-1/2} \mathbf{m}_{ZZ} S_{ee}^{-1/2} \hat{q}_i = \hat{\lambda}_i \hat{q}_i , \tag{3.2}
\]

\[
\hat{q}_Q' = \hat{\Theta}_Q \hat{Q} = \mathbb{I} ,
\]

where

\[
\hat{q}_Q = (\hat{q}_1, \ldots, \hat{q}_k, \hat{q}_{k+1}, \ldots, \hat{q}_p) = (\hat{q}_{(1)}, \hat{q}_{(2)}) ,
\]

and \( p = k+r \). If

\[
\hat{\lambda}_i > 1 , \quad i = 1, 2, \ldots, k ,
\]

then define

\[
\hat{\xi} = (\hat{A}_{rk} \hat{A}_{kk}^{-1})' , \tag{3.3}
\]

\[
\hat{\xi}_o = \hat{\xi} - \bar{\xi} \hat{\xi} , \tag{3.4}
\]
where

\[ \hat{A} = (\hat{A}_{rk}', \hat{A}_{kk}')' = \frac{1}{2} \{ \hat{q}_1 (\lambda_1 - 1)^{1/2}, \hat{q}_2 (\lambda_2 - 1)^{1/2}, \ldots, \hat{q}_k (\lambda_k - 1)^{1/2} \} , \]

and let \( \hat{\Sigma}_{xx} \) be the lower right \( k \times k \) submatrix of \( \hat{A}A' \). To obtain an alternative representation for the estimators, we let

\[ \hat{\Sigma}_{(1)}^{1/2} = \hat{S}_{xx} \hat{Q}_{(1)} = (\hat{\Sigma}_{rr}^{1/2}, \hat{\Sigma}_{kr}^{1/2})' , \]

\[ \hat{\Sigma}_{(2)}^{1/2} = \hat{S}_{xx} \hat{Q}_{(2)} = (\hat{\Sigma}_{rr}^{1/2}, \hat{\Sigma}_{kr}^{1/2})' , \]

\[ \hat{\Sigma}_{(2)}^{1/2} = \text{diag}\{ (\lambda_1 - 1)^{1/2}, \ldots, (\lambda_k - 1)^{1/2} \} . \]

Then,

\[ \hat{A} = \frac{1}{2} \hat{S}_{xx} \hat{Q}_{(1)} \hat{P}_{rk}^{1/2} = \begin{bmatrix} \hat{\Sigma}_{nk}^{1/2} \\ \hat{\Sigma}_{kk}^{1/2} \end{bmatrix} , \quad (3.5) \]

and

\[ \hat{\Sigma} = (\hat{\Sigma}_{kk}^{-1})^{-1} (\hat{\Sigma}')^{-1/2} (\hat{\Sigma}')^{1/2} \hat{\Sigma}_{rk} \]

\[ = (\hat{\Sigma}_{kk}^{-1})^{-1} \hat{\Sigma}_{rk} . \quad (3.6) \]
Since

\[ \hat{E}_{rk}^T \hat{E}_{rr} + \hat{E}_{kk}^T \hat{E}_{kr} = \hat{E}_{r}^T \hat{E}_{r} = \hat{Q}(1) \hat{Q}(2) = 0 \]

we have

\[ \hat{\beta} = -\hat{T}_{kr}^T \hat{T}_{rr}^{-1} \quad \text{(3.7)} \]

Note that expressions (3.6) or (3.7) are defined for \( \hat{\lambda}_k < 1 \).

Also,

\[ \hat{\Sigma} = \hat{P}_{(1)} (\hat{A}_1 - I) \hat{E}_{(1)}^T \quad \text{(3.8)} \]

\[ \hat{\Sigma}_{xx} = \hat{P}_{kk} (\hat{A}_1 - I) \hat{E}_{kk}^T \quad \text{(3.9)} \]

where

\[ \hat{A}_1 = \text{diag}\{\hat{\lambda}_1, \ldots, \hat{\lambda}_k\} \]

We also define
1. Functional relationship

When the $\xi_t$ are fixed, they are unknown parameters. Thus, for the functional model, the parameters to be estimated are $\xi_0$, $\xi$, $\Sigma_{\xi\xi}$, and $\xi_t$, $t = 1, 2, \ldots, n$. The following theorem gives the maximum likelihood estimators for the functional relationship, where the $\xi_t$ are normal and the $\Sigma_{\xi\xi}$ is a multiple of a Wishart matrix.

**Theorem 3.1.** Let the model (3.1) hold, and let Assumptions 3.1 and 3.5 through 3.8 hold. Then, the maximum likelihood estimators of $\xi$ and $\xi_0$ are $\hat{\xi}$ in (3.7) and $\hat{\xi}_0$ in (3.4), respectively. The maximum likelihood estimators of $\Sigma_{\xi\xi}$ and $\xi_t$ are

$$
\hat{\Sigma}_{\xi\xi} = (d+n)^{-1}[(n-1) \hat{\Sigma}_{\xi\xi} + \hat{\Sigma}_{\xi\xi}^{-1} \hat{\Sigma}_{\xi\xi} \hat{\Sigma}_{\xi\xi}^{-1} + d \Sigma_{\xi\xi}], \quad (3.10)
$$

and

$$
\hat{\xi}_t = \hat{\xi}_t - \hat{\xi}_t \hat{\Sigma}_{\xi\xi}^{-1} \hat{\Sigma}_{\xi\xi}, \quad (3.11)
$$

where

$$
\hat{\Sigma}_{\xi\xi} = (\mathbb{I} - \hat{\xi}^\prime) \Sigma_{\xi\xi} (\mathbb{I} - \hat{\xi}^\prime)^\prime, \quad
$$

$$
\hat{\Sigma}_{\xi\xi} = \hat{\Sigma}_{\xi\xi} = \Sigma_{\xi\xi} (\mathbb{I} - \hat{\xi}^\prime)^\prime. \quad
$$
\[ \hat{\psi}_{vv} = (I, -\hat{\xi}'), \hat{\psi}_{zz} (I, -\hat{\xi})', \]

\[ \hat{\psi}_t = (\hat{\phi}_t - \hat{\xi}_o) - \hat{\phi}_t \hat{\xi} = \hat{y}_t - \hat{\xi} - (\hat{\phi}_t - \hat{\xi}) \hat{\xi}, \]

\[ \hat{\xi}_{vv} = (I, -\hat{\xi})' \hat{\xi}_{ee} (I, -\hat{\xi})', \]

\[ \hat{\xi}_{vu} = (I, -\hat{\xi})' \hat{\xi}_{ee} (0, I)', \]

**Proof.** The log likelihood function is

\[
\log L = C_o - \frac{n}{2} \log |\Sigma_{ee}| - \frac{1}{2} \sum_{t=1}^{n} (Z_t - Z_t) \Sigma_{ee}^{-1} (Z_t - Z_t)',
\]

\[- \frac{d}{2} \log |\Sigma_{ee}| - \frac{d}{2} \text{tr}(\Sigma_{ee}^{-1}),
\]

where

\[ Z_t = (y_t, x_t), \]

\[ Z_t = (\phi_o, 0) + x_t (\xi, I), \]

and \( C_o \) is a constant. First we fix \( \xi \) and \( \Sigma_{ee} \) and obtain the maximum likelihood estimators of \( \phi_o \) and \( x_t \) in terms of \( \xi \) and \( \Sigma_{ee} \). Since \( \phi_o \) and \( x_t \) appear only in the part for \( (y_t, x_t) \) of the likelihood, by the standard argument for normal maximum
likelihood estimation, we have

\[ \hat{\beta}_o = \bar{X} - \bar{X} \bar{g}, \quad (3.12) \]

\[ \hat{X}_t = (X_t - \hat{\beta}_o, X_t) \Sigma_{ee}^{\frac{1}{2}} (g, \xi)' [(g, \xi) \Sigma_{ee}^{\frac{1}{2}} (g, \xi)']^{-1}. \quad (3.13) \]

Also, it follows from (3.13) and Theorem 2.1,

\[ (\hat{X}_t, \hat{X}_t) = (\hat{\beta}_o, \xi) + (X_t - \hat{\beta}_o, X_t)[\xi - (g, \xi)' \Sigma_{vv}^{\frac{1}{2}} \Sigma_{ve}] \]

\[ = (\hat{X}_t, \hat{X}_t) - \hat{X}_t \Sigma_{vv}^{\frac{1}{2}} \Sigma_{ve}, \quad (3.14) \]

where

\[ \hat{X}_t = X_t - \hat{\beta}_o - \bar{X} \bar{g}, \]

\[ \Sigma_{vv} = (\xi, - g') \Sigma_{ee} (\xi, - g'), \]

\[ \Sigma_{ve} = (\xi, - g') \Sigma_{ee}. \]

Substituting (3.12) and (3.14) into the likelihood, we obtain
\[
\log L(\hat{\beta}_o, \hat{x}_t) = C_o - \frac{n}{2} \log |\Sigma_{ee}| =
- \frac{1}{2} \sum_{t=1}^{n} \frac{\hat{x}_t \Sigma_{ee}^{-1} \hat{x}_t}{\Sigma_{ee} \hat{v}_t} - \frac{d}{2} \log |\Sigma_{ee}| - \frac{d}{2} \text{tr}(\Sigma_{ee} \Sigma_{ee}^{-1})
\]

\[
= C_o - \frac{1}{2} (d+n) \log |\Sigma_{ee}| - \frac{1}{2} (n-1) \text{tr}(\Sigma_{ee} \Sigma_{ee}^{-1})
\]

\[
- \frac{d}{2} \text{tr}(\Sigma_{ee} \Sigma_{ee}^{-1}) , \\
\] (3.15)

where

\[
\Sigma_{vv} = (n-1)^{-1} \sum_{t=1}^{n} \hat{x}_t \hat{v}_t
\]

\[
= (\Sigma, -\beta') \Sigma_{zz} (\Sigma, -\beta')'.
\]

Next we fix \( \hat{\beta} \) and maximize (3.15) with respect to \( \Sigma_{ee} \).

Given \( \hat{\beta} \), we reparameterize \( \Sigma_{ee} \) in the following way. Let

\[
\Sigma = (\Sigma_1, \Sigma_2)
\]

\[
= \begin{bmatrix}
\Sigma \\
-\Sigma
\end{bmatrix},
\]

\[
\begin{bmatrix}
\beta' \\
-\Sigma
\end{bmatrix} (\Sigma + \Sigma \beta')^{-1/2}.
\]
Then,

\[ \Gamma' \Sigma_{ee} \Gamma = \Sigma^* \]
\[ = \begin{bmatrix}
\Sigma_{vv} & \Sigma_{vw} \\
\Sigma_{vw} & \Sigma_{ww}
\end{bmatrix}, \]

where

\[ \Sigma_{vv} = \Sigma_{vv}' \]
\[ = \Gamma_1' \Sigma_{ee} \Gamma_2, \]
\[ \Sigma_{ww} = \Gamma_2' \Sigma_{ee} \Gamma_2. \]

Since \( \Gamma \) is nonsingular, this reparameterization is one-to-one. The reparameterized form of (3.15) is

\[ \log L (\hat{\beta}_o, \hat{x}_t) = C_o - \frac{1}{2} (d+n) \log |\Gamma' \Sigma_{ee}^{-1} \Gamma| \]
\[ - \frac{1}{2} (n-1) \text{tr}(\Sigma_{vv}^{-1} \Gamma' \Sigma_{ee} \Gamma) \]
\[ - \frac{d}{2} \text{tr}(\Gamma' \Sigma_{ee} \Gamma' \Sigma_{ee}^{-1}) \]
\[ - \frac{1}{2} \text{tr}(\Sigma_{vv} \Sigma_{vv}^{-1}). \]
\[ = c_0 + (d+n) \log |\Gamma| - \frac{1}{2} (d+n) \log |\Sigma^*| \]
\[ - \frac{1}{2} (n-1) \text{tr}(m_{vv} \Sigma_{vv}^{-1}) - \frac{d}{2} \text{tr}(\Sigma_{vv} \Sigma_{vv}^{-1}) \]
\[(3.16)\]

where

\[ \Sigma^* = \Gamma' \Sigma_{ww} \Gamma \]

We further reparameterize \( \Sigma^* \) by letting

\[ \Sigma_{ww,v} = \Sigma_{ww} - \chi \Sigma_{vv} \chi' \],
\[ \chi = \Sigma_{ww,v}^{-1} \]

Then, the log likelihood (3.16) becomes

\[ \log L(\hat{\Sigma}_0, \Sigma_t) \]
\[ = c_0 + (d+n) \log |\Gamma| - \frac{1}{2} (d+n) \log(|\Sigma_{vv}|, |\Sigma_{ww,v}|) \]
\[ - \frac{1}{2} (n-1) \text{tr}(m_{vv} \Sigma_{vv}^{-1}) - \frac{d}{2} \text{tr}[\Sigma_{vv} (\Sigma_{vv}^{-1} + \chi' \Sigma_{ww,v}^{-1} \chi)] \]
\[ + \frac{d}{2} \text{tr}[- \chi \Sigma_{ww,v} \Sigma_{ww,v}^{-1} - \Sigma_{ww,v}^{-1} \chi' \Sigma_{ww,v} + \Sigma_{ww,v}^{-1}] \]
\[ = C_0 + (d+n) \log |S| - \frac{1}{2} (d+n) \log |S_{vv}| \]
\[ - \frac{1}{2} \text{tr} \left\{ \left( (n-1)S_{vv} + d S_{vv} \right) S_{vv}^{-1} \right\} - \frac{1}{2} (d+n) \log S_{ww,v} \]
\[ - \frac{d}{2} \text{tr} \left\{ (\chi S_{vv} \gamma' - \chi S_{vv} - S_{ww,v} \gamma' + S_{ww}) S_{ww,v}^{-1} \right\}, \quad (3.17) \]

where

\[ S^* = \begin{pmatrix} S_{vv} & S_{vw} \\ S_{vw} & S_{ww} \end{pmatrix}, \]

and we have used

\[ |S^*| = |S_{vv}| |S_{ww,v}|, \]

\[ (S^*)^{-1} = \begin{pmatrix} S_{vv}^{-1} + \chi' S_{ww,v}^{-1} \gamma & - \chi' S_{ww,v}^{-1} \\ - S_{ww,v} \gamma & S_{ww,v}^{-1} \end{pmatrix}. \]

By Lemma 3.2.2 of Anderson (1958, p. 46), for fixed \( \gamma \), the expression in (3.17) is maximized with

\[ \hat{S}_{vv} = (d+n)^{-1} \left\{ (n-1)S_{vv} + d S_{vv} \right\}, \quad (3.18) \]
\[ \hat{\Sigma}_{WW, V} = (d+n)^{-1} d(\chi S_{VV} \gamma' - \chi S_{VW} - S_{WW} \gamma' + S_{WW}), \quad (3.19) \]

and the maximum is given by

\[
\log L(\hat{\theta}_0; \hat{\chi}; \hat{\Sigma}_{VV}, \hat{\Sigma}_{WW, V}) = C_1 + (d+n) \log |\Sigma| \\
- \frac{1}{2} (d+n) \log |(d+n)^{-1}[d(S_{VV}) + dS_{VV}]| \\
- \frac{1}{2} (d+n) \log |(d+n)^{-1} d(\chi S_{VV} \gamma' - \chi S_{VW} - S_{WW} \gamma' + S_{WW})|, \quad (3.20)
\]

where \( C_1 \) is a constant. Now, the determinant

\[
|\chi S_{VV} \gamma' - \chi S_{VW} - S_{WW} \gamma' + S_{WW}| \\
= |(\chi - S_{VV} S^{-1}_{VV}) S_{VV} (\chi - S_{VV} S^{-1}_{VV})' + S_{WW} - S_{WW} S^{-1}_{VV} S_{VV}|
\]

is minimized when

\[ \hat{\chi} = S_{VV} S^{-1}_{WW} \quad (3.21) \]
Such \( \hat{\chi} \) gives the maximum of the log likelihood (3.20)

\[
\log L(\hat{\eta}_0, \hat{\xi}_t, \hat{\xi}_{vv}, \hat{\xi}_{ww,v}, \hat{\chi})
= C_2 + \frac{1}{2}(d+n)\log|\Gamma|^2
- \frac{1}{2}(d+n)\log|(n-1)m_{vv} + d S_{vv}|
- \frac{1}{2}(d+n)\log|S_{ww} - S_{vv} S_{vv}^{-1} S_{vv}|
= C_2 + \frac{1}{2}(d+n)\log|\Gamma' \Gamma|
- \frac{1}{2}(d+n)\log|\{(n-1)m_{vv} + d S_{vv}, (S_{ww} - S_{vv} S_{vv}^{-1} S_{vv})\}|
\]

where \( C_2 \) is a constant. Therefore,

\[
\left\{ L(\hat{\eta}_0, \hat{\xi}_t, \hat{\xi}_{vv}, \hat{\xi}_{ww,v}, \hat{\chi}) \right\} \frac{2}{d+n}
= C_3 |\Gamma'^{\top} \Gamma|/\{(n-1)m_{vv} + d S_{vv}, |S_{vv} - S_{vv} S_{vv}^{-1} S_{vv}|\}
= C_3 |\Gamma'^{\top} \Gamma|/\{|(n-1)m_{vv} + d S_{vv}, |S_{vv} S_{vv}^{-1} S_{vv}|\}
= C_3 |\Gamma'^{\top} \Gamma|/\{|(n-1)m_{vv} + d S_{vv}, |S_{vv}|^{-1} |S_{vv}|\}
= C_3 |\Gamma'^{\top} \Gamma|/\{|(n-1)m_{vv} + d S_{vv}, |S_{vv}|^{-1} |\Gamma' S_{ee} \Gamma|\}
\[ \begin{align*}
&= c_3 \left| S_{vv} \right| \left/ \left| (n-1)S_{vv} + dS_{vv} \right| \left| S_{ee} \right| \right. \\
&= c_3 \left| S_{ee} \right|^{-1} \left| (I, - e') \left( I, - e' \right)' \right| \left| (I, - e') (n-1)S_{zzz} \right. \\
&\quad \quad + dS_{ee} (I, - e')' \left|^{-1} \right. , \quad (3.22)
\end{align*} \]

where \( c_3 \) is a constant. Hence, the maximum likelihood estimator of \( \beta \) is the value of \( \hat{\beta} \) which minimizes the ratio

\[ R = \left| (I, - e') S_{zzz} + (n-1)^{-1} dS_{ee} (I, - e') \right| \left| (I, - e') S_{ee} (I, - e')' \right|^{-1} . \quad (3.23) \]

Let

\[ \theta(2) = \frac{1}{2} S_{ee} (I, - e')' . \]

Then,

\[ R(\theta(2)) = R = \left| \theta(2) \frac{1}{2} S_{ee} \frac{1}{2} S_{zzz} \theta(2) + (n-1)^{-1} d\theta(2) \theta(2) \right| \left| \theta(2) \theta(2) \right|^{-1} . \quad (3.24) \]
Let \( G \) be an arbitrary \( r \times r \) nonsingular matrix, and for a given \( Q(2) \), let

\[
Q^*(2) = Q(2)G.
\]

Then,

\[
R(Q^*(2)) = R(Q(2)).
\]

Also,

\[
\hat{E}^* = - \hat{E}^* G^{-1} T_{rr}^{-1} \\
= - \hat{E}^* T_{rr} G G^{-1} T_{rr}^{-1} \\
= - \hat{E}^* T_{rr} T_{rr}^{-1} \\
= \hat{E}^*
\]

where

\[
\hat{S}^{-1/2} Q(2) = (T_{rr}', T_{kr}')',
\]

\[
\hat{S}^{-1/2} Q^*(2) = \hat{S}^{-1/2} Q(2) G
\]

\[
= (T_{rr}', T_{kr}')'.
\]
Hence, to obtain the maximum likelihood estimator \( \hat{\theta} \) of \( \theta \), we minimize \( R(\Omega(2)) \) in (3.24) with respect to \( \Omega(2) \) subject to the constraint

\[
\Omega'(2)\Omega(2) = I, \tag{3.26}
\]

and use the relation (3.25) to determine \( \hat{\theta} \). Under the condition (3.26)

\[
R(\Omega(2)) = |\Omega'(2)[S^{-1/2}_e - \frac{m}{S} S^{-1/2}_e + (n-1)^{-1} d I] \Omega(2)|. \tag{3.27}
\]

Note that the roots \( \hat{\lambda}_i \) defined in (3.2) are distinct with probability one. The eigenvalues of \( [S^{-1/2}_e - \frac{m}{S} S^{-1/2}_e + (n-1)^{-1} d I] \) are

\[
\hat{\lambda}_i + (n-1)^{-1} d, \ i = 1, 2, \ldots, p,
\]

and \( \hat{\Omega} \) is the matrix of the corresponding orthonormal eigenvectors. Therefore, by Theorem 2.7, the minimum of (3.27) is

\[
\sum_{i=k+1}^{p} \lambda_i + r(n-1)^{-1} d,
\]

where
and is attained if and only if

\[ \varrho(2) = \hat{\varrho}(2) \xi , \]

for some \( r \times r \) orthogonal matrix \( \xi \). Hence, the maximum likelihood estimator \( \hat{\xi} \) is uniquely determined by

\[ \hat{\xi} = -T^{T}r^{T}r \xi . \]

Now, let

\[ \hat{L}_{1} = (I, -\hat{\xi}')' , \]

\[ \hat{L}_{2} = (\hat{\xi}, I)' (I + \hat{\xi} \hat{\xi}')^{-1/2} , \]

\[ S^{*} = \begin{bmatrix} \hat{S}_{vv} & \hat{S}_{vw} \\ \hat{S}_{vw} & \hat{S}_{ww} \end{bmatrix} \]

\[ = (\hat{L}_{1}', \hat{L}_{2}') S_{ee} (\hat{L}_{1}, \hat{L}_{2}) , \]

\[ M_{vv} = \hat{L}_{1}' M_{zz} \hat{L}_{1} . \]

Substituting (3.28) into (3.21), we obtain the maximum likelihood estimator of \( \gamma \) as
\[ \hat{Y} = \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}}^{-1} \quad (3.29) \]

Substituting (3.28) and (3.29) into (3.18) and (3.19), we have the maximum likelihood estimators

\[ \tilde{E}_{\mathbf{vv}} = (d+n)^{-1} \left[ (n-1) \mathbf{m}_{\mathbf{vv}} + d \mathbf{S}_{\mathbf{vv}} \right] , \quad (3.30) \]

\[ \tilde{E}_{\mathbf{ww}, \mathbf{v}} = (d+n)^{-1} d(\mathbf{S}_{\mathbf{ww}} - \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{ww}}) . \quad (3.31) \]

By the invariance property of maximum likelihood estimation, the maximum likelihood estimators of \( \mathbf{E}_{\mathbf{vv}} \) and \( \mathbf{E}_{\mathbf{ww}} \) are

\[ \tilde{E}_{\mathbf{vv}} = \mathbf{Y} \tilde{E}_{\mathbf{vv}} \]

\[ = (d+n)^{-1} \mathbf{S}_{\mathbf{vv}} \left[ (n-1) \mathbf{m}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}} + d \mathbf{I} \right] \]

\[ = (d+n)^{-1} \left[ (n-1) \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}} \mathbf{m}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}} + d \mathbf{S}_{\mathbf{vv}} \right] , \quad (3.32) \]

and

\[ \tilde{E}_{\mathbf{ww}} = \tilde{E}_{\mathbf{ww}, \mathbf{v}} + \mathbf{Y} \tilde{E}_{\mathbf{vv}} \mathbf{Y} \]

\[ = (d+n)^{-1} d(\mathbf{S}_{\mathbf{ww}} - \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{vv}} \mathbf{S}_{\mathbf{ww}}) \]
Observe that

\[
\mathbf{L}^{-1} = \left( \mathbf{l}_1, \mathbf{l}_2 \right)^{-1}
\]

\[
= \begin{bmatrix}
\mathbf{I} \\
\hat{E}'
\end{bmatrix}
\begin{bmatrix}
\mathbf{I} + \hat{E} \hat{E}'^{-1}
\end{bmatrix}^{-1/2}
\]

\[
= \begin{bmatrix}
(\mathbf{I} + \hat{E} \hat{E}')^{-1} (\mathbf{I}, - \hat{E}') \\
(\mathbf{I} + \hat{E} \hat{E}')^{-1/2} (\hat{E}, \mathbf{I})
\end{bmatrix}
\]  \hspace{1cm} (3.34)

By (3.34) and the invariance property of maximum likelihood estimation, we transform (3.31), (3.32), and (3.33) to the original parameterization to obtain the maximum likelihood estimator of \( \mathbf{L}_{\xi \xi} \) as

\[
\mathbf{L}_{\xi \xi} = \mathbf{L}^{-1} \begin{bmatrix}
\hat{\mathbf{L}}_{\mathbf{V} \mathbf{V}} & \hat{\mathbf{L}}_{\mathbf{V} \mathbf{W}} \\
\hat{\mathbf{L}}_{\mathbf{W} \mathbf{V}} & \hat{\mathbf{L}}_{\mathbf{W} \mathbf{W}}
\end{bmatrix} \mathbf{L}^{-1}
\]
Finally, substituting $\hat{v}$ and $\tilde{v}$ into (3.12) and (3.14) gives the maximum likelihood estimators of $S_o$ and $\tilde{S}_t$.

2. Structural relationship

For the multivariate linear structural relationship, we assume that the $X_t$ and $\mathcal{E}_t$ are normally distributed and that the $S_{ee}$ is a multiple of a Wishart matrix. Thus, the unknown parameters are $\beta_o$, $\beta$, $\mu_X$, $\Sigma_{XX}$, and $S_{ee}$.
Theorem 3.2. Let the model (3.1) hold, and let Assumptions 3.2 and 3.4 through 3.8 hold. If \( \lambda_k < (n-1)^{-1}n \), the maximum likelihood estimator does not exist. If \( \lambda_k > (n-1)^{-1}n \), the maximum likelihood estimators \( \hat{\theta} \) and \( \hat{\theta}_o \) are \( \hat{\theta} \) in (3.7) and \( \hat{\theta}_o \) in (3.4), respectively, and the maximum likelihood estimators of \( \mu_X \), \( \Sigma_{XX} \), and \( \Sigma_{\varepsilon \varepsilon} \) are, respectively,

\[
\hat{\mu}_X = \bar{X},
\]

\[
\hat{\Sigma}_{XX} = \hat{\Sigma}_{kk} \left\{ n^{-1}(n-1) A_1 - I \right\} \hat{\Sigma}_{kk},
\]

and

\[
\hat{\Sigma}_{\varepsilon \varepsilon} = (n+d)^{-1} \left\{ d \hat{\Sigma}_{\varepsilon \varepsilon} + n \left[ n^{-1}(n-1) \hat{\Sigma}_{ZZ} - \hat{\Sigma}_{zz} \right] \right\}
\]

where

\[
\hat{\Sigma}_{zz} = (\hat{\theta}, \bar{X})' \hat{\Sigma}_{kk} (\hat{\theta}, \bar{X})
\]

\[= \frac{1}{2} \frac{1}{S_{\varepsilon \varepsilon}} Q(1) \left\{ n^{-1}(n-1) A_1 - I \right\} Q(1) \frac{1}{S_{\varepsilon \varepsilon}}. \]

Proof. The log likelihood function is
\[ \log L = c_0 - \frac{n}{2} \log |\Sigma_{zz}| - \frac{1}{2} \sum_{t=1}^{n} (Z_t - \mu_Z) \Sigma_{zz}^{-1} (Z_t - \mu_Z)' \]

\[ - \frac{d}{2} \log |\Sigma_{\epsilon\epsilon}| - \frac{d}{2} \text{tr}(S_{\epsilon\epsilon}^{-1}) , \]

where

\[ Z_t = (Y_t', X_t') , \]

\[ \mu_Z = (\beta_0, \theta) + \mu_{X}(\beta, \lambda) . \]

Since

\[ \hat{\mu}_Z = \bar{Z} = n^{-1} \sum_{t=1}^{n} (Y_t', X_t') \]

maximizes the likelihood with respect to \( \mu_Z \), the results for \( \hat{\beta}_0 \) and \( \hat{\mu}_X \) follow. Let

\[ T = S_{\epsilon\epsilon}^{-1/2} \hat{Z} . \]

Then,

\[ T' S_{\epsilon\epsilon} T = I , \]
Define new parameters

\[ E_T = T' E_Z Z T = \hat{A} \cdot \]

and let

\[ E_{TZZ} = T' E_{ZZ} T \cdot \]

Then, the log likelihood function in terms of the transformed variables and new parameters is

\[
\log L (\hat{\xi}) = C_0 - \frac{1}{2} (n+d) \log |T^{-1}T^{-1}| \\
- \frac{1}{2} n \log |E_{TZZ}| - \frac{1}{2} (n-1) \text{tr} (\hat{A} E_{TZZ}^{-1}) \\
- \frac{1}{2} d \log |E_{Tee}| - \frac{1}{2} d \text{tr} (E_{Tee}^{-1}) ,
\]
where $\theta$ contains the elements of vech $\Sigma_{TEE}$ and $\eta$. Therefore, we can minimize

$$f(\theta) = n \log |\Sigma_{TZZ}| + (n-1) \text{tr} (A \Sigma_{TZZ}^{-1})$$

$$+ d [\log |\Sigma_{TEE}| + \text{tr} (\Sigma_{TEE}^{-1})] . \quad (3.38)$$

Since only $kr + \frac{1}{2} k(k+1)$ elements in $\eta$ are free, we impose a restriction

$$\eta^T \hat{A}^{-1} \eta = D , \quad (3.39)$$

where $D$ is a diagonal matrix with free diagonal elements. Observe that

$$\frac{\partial \Sigma_{TZZ}}{\partial \eta_{ij}} = \eta J_{kp,ji} + J_{pk,ij} \eta'$$

$$= (\theta_{p,i-1}', \theta_{p,j}', \theta_{p,p-1}) + (\theta_{p,i-1}', \theta_{p,j}', \theta_{p,p-1})',$$

$$\frac{\partial \Sigma_{TEE}}{\partial \sigma_{TEE ij}} = \frac{\partial \Sigma_{TZZ}}{\partial \sigma_{TZZ ij}}$$

$$= \begin{cases} 
J_{pp,ij} + J_{pp,ij} , & i \neq j , \\
J_{pp,ii} , & i = j , 
\end{cases}$$
where \( J_{ab,cd} \) is the \( a \times b \) matrix of zero elements except for the \((c,d)\)-th element being unity, and \( Q_{ab} \) is the \( a \times b \) matrix of zeros. Hence, the derivatives of \( f(\theta) \) with respect to the elements of \( \theta \) are for \( i = 1,2,\ldots, p \), and \( j = 1,2,\ldots, k \),

\[
\frac{\partial f(\theta)}{\partial \eta_{ij}} = \text{tr} \left\{ E_{TZZ}^{-1} \left[ (n-1) \hat{A} - n E_{TZZ} \right] \frac{\partial^2 E_{TZZ}}{\partial \eta_{ij}} \right\}
\]

\[
= 2 \left\{ E_{TZZ}^{-1} \left[ (n-1) \hat{A} - n E_{TZZ} \right] E_{TZZ} \right\}_{ij},
\]

\[
\frac{\partial f(\theta)}{\partial T_{T\epsilon ij}} = \text{tr} \left\{ E_{TZZ}^{-1} \left[ (n-1) \hat{A} - n E_{TZZ} \right] \frac{\partial^2 E_{TZZ}}{\partial T_{T\epsilon ij}} \right\}
\]

\[
+ d \text{tr} \left\{ E_{T\epsilon}^{-1} \left( I - E_{T\epsilon} \right) E_{T\epsilon}^{-1} \frac{\partial^2 E_{T\epsilon}}{\partial T_{T\epsilon ij}} \right\}.
\]

\[
\frac{\partial f(\theta)}{\partial T_{T\epsilon ii}} = \left\{ E_{TZZ}^{-1} \left[ (n-1) \hat{A} - n E_{TZZ} \right] E_{TZZ} \right\}_{ii},
\]

and for \( i > j = 1,2,\ldots, p-1 \),

\[
\frac{\partial f(\theta)}{\partial T_{T\epsilon ij}} = 2 \left\{ E_{TZZ}^{-1} \left[ (n-1) \hat{A} - n E_{TZZ} \right] E_{TZZ} \right\}_{ij}.
\]
Setting the derivatives equal to zero, we obtain the necessary conditions for critical points of \( f(\theta) \):

\[
\Sigma_{Tzz}^{-1} [A^* - \Sigma_{Tzz}] \Sigma_{Tzz}^{-1} \eta = 0 ,
\]

or

\[
\Sigma_{Tzz}^{-1} [A^* - \Sigma_{Tzz}] \Sigma_{Tzz}^{-1} + d \Sigma_{Tee}^{-1} (I - \Sigma_{Tee}) \Sigma_{Tee}^{-1} = 0 ,
\]

where

\[
A^* = n^{-1}(n-1) \hat{A}
\]

\[
= \text{diag} \{ \lambda_1^*, \lambda_2^*, \ldots, \lambda_p^* \} .
\]

By (3.40),

\[
\eta = \Sigma_{Tzz} A^* \eta
\]

\[
= (\eta_{\theta'} + \Sigma_{Tee}) A^* \eta ,
\]

and thus,
\[
\mathbf{A}^{*-1/2} \mathbf{H} \Gamma = \mathbf{A}^{*-1/2} \mathbf{X}_{\text{Te}} \mathbf{A}^{*-1/2} \mathbf{H}^{1/2},
\]

where

\[
\Gamma = \mathbf{I} - \mathbf{H}' \mathbf{A}^{*-1} \mathbf{H}
\]

\[
= \mathbf{I} - \mathbf{H}^2
\]

\[
= \text{diag} \{ \gamma_1, \gamma_2, \ldots, \gamma_p \}
\]

is a diagonal matrix by the restriction (3.39). Therefore, \( \mathbf{A}^{*-1/2} \mathbf{H} \Gamma \) are necessarily the eigenvectors of \( \mathbf{A}^{*-1/2} \mathbf{X}_{\text{Te}} \mathbf{A}^{*-1/2} \) corresponding to \( \gamma_i \). Which \( k \) roots are used to construct \( \mathbf{H} \) is undetermined. Since \( \mathbf{H} \) has a full column rank by Assumption 3.2, we need to choose exactly \( k \) roots. Let

\[
\Gamma = \text{block diag} \{ \Gamma_1, \Gamma_2 \},
\]

where \( \Gamma_1 \) contains the \( k \) roots corresponding to \( \mathbf{H} \). Also, let

\[
\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2)
\]

be the matrix of orthonormal eigenvectors corresponding to \( \Gamma_1 \) and \( \Gamma_2 \). Then,
\[ \lambda^{-1/2} \eta_i = c_i \xi_i \quad , \quad i = 1,2, \ldots, k. \]

for some constants \( c_i \). Since

\[ 1 - \gamma_i = D_{ii} = \eta_i \lambda^{* -1} \eta_i \]

\[ = c_i^2 \xi'_i \xi_i = c_i^2 \]

we have

\[ \hat{\eta} = \lambda^{1/2} \xi_1 (I - \xi_1^{1/2}), \quad (3.42) \]

provided the \( k \) roots \( \gamma_i \) chosen are all less than unity. Observe that

\[ \Sigma_{\text{Tee}} = \lambda^{1/2} \xi \xi' \lambda^{1/2}. \quad (3.43) \]

Thus,

\[ \Sigma_{\text{Tee}}^{-1} (I - \Sigma_{\text{Tee}}) \Sigma_{\text{Tee}}^{-1} = \lambda^{1/2} \xi \xi'^{-1} \xi' (\lambda^{* -1} - \xi \xi' \xi') \xi \xi'^{-1} \xi' \lambda^{* -1/2}. \quad (3.44) \]

By (3.42) and (3.43),
\[ \hat{E}_{TZZ} = \hat{n} \hat{n}' + \Sigma_{TEC} \]

\[ = A^{\frac{1}{2}} E_1 (I - E_1) E_1' A^{\frac{1}{2}} + A^{\frac{1}{2}} E_1 E_1' A^{\frac{1}{2}} \]

\[ = A^{\frac{1}{2}} (E_1 E_1' + E_2 E_2' E_1') A^{\frac{1}{2}} \]

\[ = A^* + A^{\frac{1}{2}} E_2 (I - E_2) E_2' A^{\frac{1}{2}} \quad (3.45) \]

Hence,

\[ \hat{E}_{TZZ}^{-1} (A^* - \hat{E}_{TZZ}) \hat{E}_{TZZ}^{-1} \]

\[ = A^{-\frac{1}{2}} (E_1 E_1' + E_2 E_2') \]

\[ = A^{\frac{1}{2}} E_1 (I - E_2) E_2' E_2 (E_1 E_1' + E_2 E_2') \]

\[ = A^{\frac{1}{2}} E_2 E_2' E_2 (I - E_2) E_2' E_2 E_2' A^{\frac{1}{2}} \]

\[ = A^{\frac{1}{2}} E_2 E_2' E_2 (I - E_2) E_2' E_2 (I - E_2) E_2' E_2 A^{\frac{1}{2}} \quad (3.46) \]

where we have used

\[ E_1 E_1^{-1} E_1' E_2 = E_1 E_1^{-1} E_1' E_2 + E_2 E_2^{-1} E_1' E_2 \]

\[ = E_2 E_2^{-1} E_2' E_2 \quad . \]

By (3.44) and (3.46), the necessary condition (3.41) with (3.43)
\[
A_2^0 = \text{diag} \{ \lambda_{k+1}^0, \lambda_{k+2}^0, \ldots, \lambda_p^0 \}.
\]

Then,

\[
\hat{P}_1 = A_1^{0-1},
\]

\[
\hat{P}_2 = (n+d)^{-1} [d A_2^{0-1} + n \mathbb{I}] .
\]  

(3.48)

By (3.43) and (3.48),

\[
\hat{\Sigma}_{Tee} = \Lambda^{1/2} \hat{P} \left[
\begin{pmatrix}
A_1^{0-1} & 0 \\
0 & (n+d)^{-1}[d A_2^{0-1} + n \mathbb{I}]
\end{pmatrix}
\right] \hat{P}^t \Lambda^{1/2} .
\]  

(3.49)

Note that the matrix \( \hat{P} \) is the permutation matrix such that

\[
\hat{P} \text{ diag } \{ \lambda_1^0, \lambda_2^0, \ldots, \lambda_p^0 \} \hat{P}^t = \Lambda^* .
\]  

(3.50)

Hence, by (3.49) and (3.50),

\[
\hat{\Sigma}_{Tee} = \text{diag} \{ \hat{\delta}_1, \ldots, \hat{\delta}_p \} ,
\]  

(3.51)

where \( \hat{\delta}_1 = 1 \) if \( \lambda_1^* \) is in the chosen set of \( \lambda_1^0, \lambda_2^0, \ldots, \lambda_p^0 \),

and \( \hat{\delta}_i = (n+d)^{-1}(d+n\lambda_i^0) \) if \( \lambda_i^* \) is not in the set. By (3.42), (3.48),
and (3.51) the critical points of the function \( f(\theta) \) satisfying (3.39) have the form:

\[
\hat{\theta} = \int^{1/2} P_1 (I + A_1^{-1} \hat{\theta})^{1/2},
\]

\[
\hat{\Sigma}_{zz} = \text{diag} \{ \hat{\delta}_1, \ldots, \hat{\delta}_p \},
\]

(3.52)

where \( \lambda_1^0, \lambda_2^0, \ldots, \lambda_k^0 \) are any \( k \) diagonal elements of \( A^* \) which are greater than unity. If \( \lambda_k^* < 1 \), i.e. \( \lambda_k < (n-1)^{-1} n \), then the maximum likelihood estimator does not exist. Now assume \( \lambda_k^* > 1 \) so that \( \lambda_i^0 > 1 \) for \( i = 1, \ldots, k \). To determine which of the critical points (3.52) gives the minimum, using the relation (3.45), we substitute (3.52) into (3.38). For the critical point (3.52), we have, by the relation (3.45) and the nature of \( P \),

\[
\hat{\Sigma}_{zz} = \int^{1/2} P \begin{pmatrix} I & 0 \\ 0 & \hat{\Lambda} \end{pmatrix} \begin{pmatrix} P' \hat{\Lambda}^{1/2} & 0 \\ 0 & \hat{\Lambda}^{1/2} \end{pmatrix} = \text{diag} \{ \hat{\nu}_1, \ldots, \hat{\nu}_p \},
\]

(3.53)

where \( \hat{\nu}_i = \lambda_i^* \) if \( \lambda_i^* \) is among \( \lambda_1^0, \lambda_2^0, \ldots, \lambda_k^0 \) and \( \hat{\nu}_i = \delta_i \) if \( \lambda_i^* \) is not among \( \lambda_1^0, \lambda_2^0, \ldots, \lambda_k^0 \). Thus, the function \( f(\theta) \) evaluated at the critical point (3.52) is
\[ f(\theta) = n \left[ \sum_{i=1}^{P} \log \hat{v}_i + k + \text{tr} \left( \hat{\Sigma}^{-1} \right) \right] + d \left[ \sum_{i=1}^{P} \log \hat{\delta}_i + \sum_{i=1}^{P} \hat{\delta}_i^{-1} \right] \]

\[ = n \sum_{i=1}^{k} \log \hat{\lambda}_i^0 + (n+d) \sum_{i=k+1}^{P} \log \left\{ (n+d)^{-1}(d+n\lambda_i^0) \right\} + (n+d) k \]

\[ + n \sum_{i=k+1}^{P} (d+n\lambda_i^0)^{-1} \lambda_i^0 (n+d) + d \sum_{i=k+1}^{P} (d+n\lambda_i^0)^{-1}(n+d) \]

\[ = n \sum_{i=1}^{k} \log \lambda_i^0 + (n+d) \left[ \log d \delta + n A^* \right] - \sum_{i=1}^{k} \log (d+n\lambda_i^0) \]

\[ - (n+d) r \log(n+d) + (n+d) p \]

\[ = C_0 + \sum_{i=1}^{k} g(\lambda_i^0), \quad (3.54) \]

where \( C_0 \) is free of the choice of \( \lambda_i^0 \) and

\[ g(w) = n \log w - (n+d) \log(d+nw). \]

Now, for all \( w > 1, d > 0, \) and \( n > 0 \),

\[ \frac{\partial g(w)}{\partial w} = w^{-1}(d+nw)^{-1}nd(1-w) < 0. \]

Hence, \( g(w) \) is monotone decreasing for \( w > 1 \). Therefore, the expression \( (3.54) \) is minimized when the \( k \) largest roots, \( \lambda_1^*, \lambda_2^*, \ldots, \lambda_k^* \), are chosen as \( \lambda_1^0, \lambda_2^0, \ldots, \lambda_k^0 \). Thus,
\[ \hat{\delta}_i = 1, \quad i=1,2,\ldots,k, \]
\[ = (n+d)^{-1} (d+n\lambda_i^*)^2, \quad i=k+1, k+2,\ldots,p, \]
\[ \hat{\nu}_i = \lambda_i^*, \quad i=1,2,\ldots,k, \]
\[ = (n+d)^{-1} (d+n\lambda_i^*)^2, \quad i=k+1, k+2,\ldots,p. \]

Hence, by (3.51) and (3.53), the maximum likelihood estimators of \( \hat{\Sigma}_{TZZ} \) and \( \hat{\Sigma}_{Tee} \) are

\[ \hat{\Sigma}_{TZZ} = \text{block diag} \{ \Lambda_1^*, (n+d)^{-1} (n\Lambda_2^* + d I) \}, \]

\[ \hat{\Sigma}_{Tee} = \text{block diag} \{ I, (n+d)^{-1} (n\Lambda_2^* + d I) \}, \]

where

\[ \Lambda_1^* = \text{diag} \{ \lambda_1^*, \lambda_2^*, \ldots, \lambda_k^* \} = n^{-1} (n-1) \hat{A}_1, \]

\[ \Lambda_2^* = \text{diag} \{ \lambda_{k+1}^*, \lambda_{k+2}^*, \ldots, \lambda_p^* \} = n^{-1} (n-1) \hat{A}_2. \]

By the invariance property of maximum likelihood estimation,

\[ \hat{\Sigma}_{zz} = \hat{n}' \hat{\Sigma}_{zz} \hat{n} \]
\[ = \hat{\Sigma}_{zz}^{-1} \text{block diag} \{ \Lambda_1^* - I, Q \} \hat{\Sigma}_{zz}^{-1} \]
\[ = \frac{1}{2 \sqrt{p}} \hat{Q}_{(1)} \left( A_1^* - I \right) \hat{Q}_{(1)}' \frac{1}{2 \sqrt{p}} = 0, \hspace{1cm} (3.55) \]

and

\[ \sum_{22} = T^{-1} \sum_{22} T^{-1} \]

\[ = S_{22} + (n+d)^{-1}n \frac{1}{2 \sqrt{p}} \hat{Q}_{(1)} \left( A_1^* - I \right) \hat{Q}_{(1)}' S_{22} \]

\[ = S_{22} + (n+d)^{-1}n[n^{-1}(n-1)\mu_{zz} - S_{22}] \]

\[ - \frac{1}{2 \sqrt{p}} \hat{Q}_{(1)} \left( A_1^* - I \right) \hat{Q}_{(1)}' S_{22} \]

\[ = (n+d)^{-1}d S_{22} + (n+d)^{-1}n[n^{-1}(n-1)\mu_{zz} - \hat{x}_{zz}] \]

Also, by (3.55),

\[ \hat{S} = \left( \hat{E}_{kk} \right)^{-1} \hat{E}_{rk} = - \hat{x}_{kr} \hat{x}_{rr} \]

\[ \hat{Z}_{xx} = \hat{P}_{kk} \left( A_1^* - I \right) \hat{P}_{kk}' \]

Under the assumptions of Theorem 3.2, we can perform a goodness of fit test. The alternative to the model (3.1) is the unrestricted normal model where \( Z_t \) are independently and identically distributed with
unrestricted mean and covariance matrix and $\Sigma_{\xi\xi}$ is a multiple of Wishart matrix independent of $Z_t$.

**Theorem 3.3.** Let the model (3.1) hold, and let Assumptions 3.2 and 3.4 through 3.8 hold. Then, the likelihood ratio statistic of the model (3.1) against the unrestricted model is

$$
\chi^2 = (n+d) \sum_{i=k+1}^{p} \log\{(n+d)^{-1}[\hat{\lambda}_i + d]\} - n \sum_{i=k+1}^{p} \log[n^{-1}(n-1)\hat{\lambda}_i].
$$

If the model specification is correct, then the test statistic is asymptotically distributed as a chi-square random variable with $2^{-1}r(r+1)$ degrees of freedom.

**Proof.** Using the notation in the proof of Theorem 3.2, we have for the unrestricted model

$$
\hat{\Sigma}_{ZZ} = \hat{A},
$$
$$
\hat{\Sigma}_{\xi\xi} = \hat{I}.
$$

The expression for the test statistic follows by evaluating the function $f(\hat{\theta})$ in (3.38) and taking the difference. The distribution of observations is a product of a normal distribution and a Wishart distribution. Also, the mean and the covariance matrices are twice differentiable functions of the parameters. Thus, by the standard likelihood ratio theory, the test statistic converges in law to a chi-square random variable. The degrees of freedom are given by the
difference of the number of parameters for the two models for $\Sigma_{ZZ}$. The model (3.1) has $kr + 2^{-1} k(k+1)$ parameters in $\Sigma_{ZZ}$, and the unrestricted model has $2^{-1} p(p+1)$ parameters. Therefore, the difference is $2^{-1} r(r+1)$. □

3. Maximum likelihood estimators adjusted for degrees of freedom

The maximum likelihood estimator does not take into account the number of parameters to be estimated. As in regression analysis, we often use an estimator which is obtained from the maximum likelihood estimator by making an adjustment for degrees of freedom. Usually, such an adjustment does not change the asymptotic properties of an estimator. However, more care is required when there exist the incidental parameters appearing in the distributions of finitely many observations. Consider the maximum likelihood estimator $\tilde{\Sigma}_{\epsilon \epsilon}$ in (3.10) for the functional relationship model. The parameter $\Sigma_{\epsilon \epsilon}$ is a structural parameter which appears in the distributions of infinitely many observations as $n$ and $d$ increase. But, the presence of the incidental parameters $\theta_t$ for the functional model makes the maximum likelihood estimator $\tilde{\Sigma}_{\epsilon \epsilon}$ inconsistent. In the expression (3.10) for the $\tilde{\Sigma}_{\epsilon \epsilon}$, the term

$$\tilde{S}_{\epsilon \epsilon} \tilde{S}_{\epsilon \epsilon}^{-1} \tilde{S}_{vv} \tilde{S}_{vv}^{-1} \tilde{S}_{\epsilon \epsilon}$$

is estimating $S_{\epsilon \epsilon} S_{\epsilon \epsilon}^{-1} S_{vv} S_{vv}^{-1} S_{\epsilon \epsilon}$, where
\[ \Sigma_{EE} = \Sigma_{VE} = \Sigma_{EE} (I, -\xi')' \]

\[ \Sigma_{VV} = (I, -\xi') \Sigma_{EE} (I, -\xi')' . \]

Hence, even when both \( n \) and \( d \) tend to infinity, we do not expect the \( \tilde{\Sigma}_{EE} \) for the functional case to be a consistent estimator of \( \Sigma_{EE} \).

However, it follows from (3.29), (3.30), and (3.31) that using the transformation used in the proof of Theorem 3.1 we can concentrate the inconsistency of \( \tilde{\Sigma}_{EE} \) in \( \tilde{\Sigma}_{WW,V} \) and obtain the consistent maximum likelihood estimators \( \tilde{\Sigma}_{VV} \) and \( \hat{\gamma} \) of \( \Sigma_{VV} \) and \( \Sigma_{VV} \Sigma_{VV}^{-1} \).

Furthermore, the inconsistency of \( \Sigma_{WW,V} \) is due only to a proportion factor. Thus, making an obvious adjustment, we consider \( d^{-1}(d+n)\tilde{\Sigma}_{WW,V} \) as an estimator of \( \Sigma_{WW,V} \). Therefore, transforming to the original parameterization \( \Sigma_{EE} \), we obtain the maximum likelihood estimator of \( \Sigma_{EE} \) adjusted for the presence of the incidental parameters as follows. By (3.34),

\[
\begin{bmatrix}
\tilde{\Sigma}_{VV} & \tilde{\Sigma}_{VW} \\
\tilde{\Sigma}_{WV} & \tilde{\Sigma}_{WW}
\end{bmatrix}
= \frac{1}{L} \begin{bmatrix}
\tilde{\Sigma}_{VV} & d^{-1}(d+n)\tilde{\Sigma}_{WW,V} + \tilde{\Sigma}_{VV} \tilde{\Sigma}_{VV}^{-1} \tilde{\Sigma}_{WW} \\
\tilde{\Sigma}_{WV} & \tilde{\Sigma}_{WW}
\end{bmatrix}
\]
\[ \begin{pmatrix}
0 & 0 \\
0 & S_{\omega \nu} - S_{\omega \nu} S_{\nu \nu} S_{\nu \nu} ^{-1} S_{\nu \nu}
\end{pmatrix} ^{-1} \]

\[ = \xi_{ee} + (d+n)^{-1} \left[ n(\xi,1)'(I+\hat{\xi}'' \hat{\xi}''^{-1}) - \frac{1}{2} (S_{\omega \nu} - S_{\omega \nu} S_{\nu \nu} S_{\nu \nu} ^{-1} S_{\nu \nu})(I+\hat{\xi}'' \hat{\xi}''^{-1}) - \frac{1}{2} (\xi,1) \right] \]

\[ = (d+n)^{-1} \left[ d \xi_{ee} + (n-1) S_{ee} \left( S_{ee} S_{ee} S_{ee} S_{ee} ^{-1} S_{ee} \right) \right. \\

\[ + n(\xi,1)'(I+\hat{\xi}'' \hat{\xi}''^{-1})(\xi,1)(S_{ee} - S_{ee} S_{ee} S_{ee} S_{ee} ^{-1} S_{ee})(\xi,1)'(I+\hat{\xi}'' \hat{\xi}''^{-1})(\xi,1) \left. \right] \]

\[ = (d+n)^{-1} \left[ d \xi_{ee} + (n-1) S_{ee} \hat{T}(2) \hat{A}(2) \hat{T}(2) \hat{A}(2) \right. \\

\[ + n \hat{P}_{(1)} \left( \hat{P}_{(1)} \hat{P}_{(1)} \right) ^{-1} \hat{P}_{(1)} (S_{ee} - S_{ee} S_{ee} S_{ee} S_{ee} ^{-1} S_{ee}) \hat{P}_{(1)} \left( \hat{P}_{(1)} \hat{P}_{(1)} \right) ^{-1} \hat{P}_{(1)} \left. \right] \]

\[ = (d+n)^{-1} \left[ d \xi_{ee} + (n-1) S_{ee} \hat{Q}(2) \hat{A}(2) \hat{Q}(2) \hat{A}(2) \right. \\

\[ + n \frac{1}{2} S_{ee} \hat{Q}(1) \hat{Q}(1) \left. \right] \]

(3.56)

where

\[ \hat{\xi}_{zz} = \frac{1}{2} \hat{\xi}_1 \{ n^{-1} (n-1) \hat{A}_1 - I \} \hat{\xi}_1 \frac{1}{2} \]

\[ \hat{\xi}_{zz} = \frac{1}{2} \hat{\xi}_1 \{ n^{-1} (n-1) \hat{A}_1 - I \} \hat{\xi}_1 \frac{1}{2} \]
Thus, changing a divisor for a portion of the estimator (3.10) for the functional model, we have obtained \( \hat{\Sigma}_{EE} \) which is identical to \( \tilde{\Sigma}_{EE} \) in (3.37), the maximum likelihood estimator for the structural model. By adjusting for degrees of freedom, we obtain the same maximum likelihood estimators of the parameters \( \beta_0, \beta, \) and \( \Sigma_{EE} \) for both the functional relationship and the structural relationship.

Since the model (3.1) has the intercept term, we have used a divisor \( (n-1) \) for \( \Sigma_{ZZ} \), the matrix of sums of squares and cross products. This choice of divisor led to the presence of a factor \( n^{-1}(n-1) \) in the maximum likelihood estimators \( \tilde{\Sigma}_{XX} \) in (3.36), \( \tilde{\Sigma}_{EE} \) in (3.37), and \( \hat{\Sigma}_{EE} \) in (3.56). By analogy to usual regression analysis, we adjust the estimators \( \tilde{\Sigma}_{XX} \) and \( \tilde{\Sigma}_{EE} \), and define the maximum likelihood estimators of \( \Sigma_{XX} \) and \( \Sigma_{EE} \) adjusted for degrees of freedom as

\[
\hat{\Sigma}_{XX} = \hat{\Sigma}_{kk} (\hat{A}_1 - \hat{I}) \hat{P}_k', \quad (3.57)
\]

\[
\hat{\Sigma}_{EE} = (n-1+d)^{-1} \left\{ d \hat{\Sigma}_{EE} + (n-1) \left( \hat{\Sigma}_{ZZ} - \hat{\Sigma}_{zz} \right) \right\}, \quad (3.58)
\]

where

\[
\hat{\Sigma}_{ZZ} = \hat{\Sigma}_{(1)} (\hat{A}_1 - \hat{I}) \hat{P}_1'.
\]

In the following section, we consider the properties of the maximum likelihood estimators adjusted for degrees of freedom. The estimators
are \( \hat{\theta}_o \) in (3.4), \( \hat{\theta} \) in (3.7), and \( \hat{\Sigma}_{ee} \) in (3.58) for both the functional and structural models, and \( \hat{\Sigma}_{xx} \) in (3.57) for the structural model. Since we consider only the asymptotic properties, the adjustments used in (3.57) and (3.58) do not change the results.

C. Asymptotic Properties

1. Strong consistency

In this section, we show the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom under a broad class of assumptions. We begin by presenting two lemmas which will be used to show the strong consistency.

Lemma 3.1. Let the model (3.1) hold, and let Assumptions 3.5 and 3.7 hold. In addition, let either (a) Assumption 3.1, or (b) Assumption 3.2 hold. Then, for both the cases (a) and (b), as \( n \to \infty \),

\[
\Sigma_{ee}^{1/2} \Sigma_{zz}^{1/2} \to \Sigma_{ee}^{1/2} \Sigma_{zz}^{1/2} \Sigma_{ee}^{1/2}, \text{ a.s.},
\]

where

\[
\Sigma_{zz} = (\Theta, 0)' \Sigma_{xx} (\Theta, 0) + \Sigma_{ee}.
\]

Proof. For case (b), the result follows from Assumption 3.7 and the strong law of large numbers. For case (a), we observe that

\[
\Sigma_{zz} = \Sigma_{zz} + \Sigma_{ze} + \Sigma_{ez} + \Sigma_{ee}, \quad (3.59)
\]
where

\[ B_{zz} = (\beta, \beta)' B_{xx} (\beta, \beta) , \]

\[ B_{ze} = (\beta, \beta)' B_{xe} \]

\[ = B_{ez} , \]

\[ B_{xe} = (n-1)^{-1} \sum_{t=1}^{n} (z_t - \bar{z}) x_t' . \]

By Assumption 3.1,

\[ \lim_{n \to \infty} B_{zz} = (\beta, \beta)' B_{xx} (\beta, \beta) . \quad (3.60) \]

By the strong law of large numbers,

\[ B_{ee} \longrightarrow B_{ee} \quad \text{a.s.}. \quad (3.61) \]

The \((i,j)\)-th element of \( B_{ze} \) is

\[ (n-1)^{-1} \sum_{t=1}^{n} (z_{it} - z_i) x_{jt} = (n-1)^{-1} \sum_{t=1}^{n} z_{it} x_{jt} , \]

where \( z_{it} \) is the \( i \)-th element of \( z_t \),
\[-z_i = n^{-1} \sum_{t=1}^{n} z_{it},\]

and

\[z_{it}^* = z_{it} - \bar{z}_i.\]

It follows from Assumption 3.1 that there exist constants $K_i$ satisfying

\[\lim_{n \to \infty} (n-1)^{-1} \sum_{t=1}^{n} z_{it}^2 = K_i < \infty.\]

Hence, there exists a finite number $K_0$ such that for any $n$

\[(n-1)^{-1} \sum_{t=1}^{n} z_{it}^2 < K_0.\]  \hspace{1cm} (3.62)

By Abel's partial summation formula, for any two series \(\{a_t\}\) and \(\{b_t\}\) and any $n > M > 0$, 
\[
\sum_{t=M}^{n} a_t (b_{t+1} - b_t) = a_n b_{n+1} - a_M b_M - \sum_{t=M+1}^{n} b_t (a_t - a_{t-1}) 
\]  
\[(3.63)\]

Letting

\[ a_t = t^{-2} , \]

\[ b_t = \sum_{j=1}^{t-1} z_{i,j}^2 , \]

\[ M = 2 , \]

and using (3.62) we have for any \( n > 2 \)

\[
\sum_{t=2}^{n} \frac{1}{t^2} z_{i,t}^2 = \frac{1}{n^2} \sum_{j=1}^{n} z_{i,j}^2 - \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{t-1} z_{i,j}^2 \left[ \sum_{j=1}^{t-1} \frac{1}{t^2} \left( \frac{1}{(t-1)^2} \right) \right] 
\]

\[
= \frac{1}{n(n-1)} \sum_{j=1}^{n} z_{i,j}^2 - \frac{1}{4} \sum_{i=1}^{n} \frac{2t-1}{t^2(t-1)} \sum_{j=1}^{t-1} z_{i,j}^2 
\]

\[
\frac{1}{n} K_0 + 2 \sum_{t=3}^{n} \frac{1}{t(t-1)} \frac{1}{(t-1)} \sum_{j=1}^{t-1} z_{i,j}^2 
\]
Since \( \sum_{t=1}^{n} t^{-2} z_{it}^{2} \) is a series of nonnegative terms, (3.64) implies that for some \( V_{o} \)

\[
\lim_{n \to \infty} \sum_{t=1}^{n} \frac{1}{t^2} z_{it}^{2} = V_{o} < \infty.
\]

Thus, as \( n \to \infty \),

\[
\sum_{t=1}^{n} \frac{1}{t^2} V_{i} z_{it}^{2} \epsilon_{jt} = \sum_{t=1}^{n} \frac{1}{t^2} z_{it}^{2} \sigma_{\epsilon jj} \epsilon_{jj} \to V_{o} \sigma_{\epsilon jj} < \infty.
\]

By (3.65), the independence of \( z_{it}^{*} \epsilon_{jt} \), and the fact that

\[
E[z_{it}^{*} \epsilon_{jt}] = 0,
\]
it follows from the standard probability theorem (See, for example, Chung (1974), p. 125.) that

\[(n-1)^{-1} \sum_{t=1}^{n} z_{it}^* e_{jt} \rightarrow 0 \text{ a.s. .}\]

Therefore,

\[m_{z\epsilon} = m_{\epsilon z} \]

\[\rightarrow 0 \text{ a.s. .} \quad (3.66)\]

By (3.59), (3.60), (3.61), and (3.66),

\[m_{zz} \rightarrow (\bar{g}, \bar{I})' \Sigma_{XX} (\bar{g}, \bar{I}) + \Sigma_{\epsilon \epsilon} \text{ a.s. .} \quad (3.67)\]

Also, by Assumption 3.7, as \( n \rightarrow \infty \),

\[S_{\epsilon \epsilon} \rightarrow \Sigma_{\epsilon \epsilon} \text{ a.s. .} \quad (3.68)\]

By (3.67) and (3.68), the result for the case (a) follows. \(\Box\)

We introduce further notations. Let

\[E_{pp} = [(\bar{g}, \bar{I}) \Sigma_{\epsilon \epsilon}^{-1} (\bar{g}, \bar{I})']^{-1} .\]
Also, let \( \nu_1 > \nu_2 > \ldots > \nu_k \) be the eigenvalues of \( \Sigma_{pp}^{-1/2} \Sigma_{xx} \Sigma_{pp}^{-1/2} \), and let \( R \) be the matrix of corresponding orthonormal eigenvectors.

Note that by Theorem 2.1

\[
\Sigma_{pp} = \Sigma_{uu} - \Sigma_{uv} \Sigma_{vv}^{-1} \Sigma_{vu}, \tag{3.69}
\]

where

\[
\Sigma_{vv} = (I - \beta') \Sigma_{ee} (I - \beta')',
\]

\[
\Sigma_{uv} = \Sigma_{vu}
\]

\[
= \Sigma_{ue} - \Sigma_{uu} \beta.
\]

Note also that \( \nu_k > 0 \) since \( \Sigma_{pp}^{-1/2} \Sigma_{xx} \Sigma_{pp}^{-1/2} \) is positive definite.

**Lemma 3.2.** Let \( \lambda_1 > \lambda_2 > \ldots > \lambda_p \) be the eigenvalues of the matrix

\[
\Sigma_{ee}^{-1/2} \Sigma_{zz} \Sigma_{ee}^{-1/2},
\]

and denote the corresponding orthonormal eigenvectors by

\[
Q = (Q^{(1)}, Q^{(2)}),
\]

where \( Q^{(1)} \) has \( k \) columns. Then,
\[ \lambda_i = 1 + \nu_i, \quad i = 1, 2, \ldots, k, \]
\[ = 1, \quad i = k+1, k+2, \ldots, p, \]
\[ \Omega_1 = \Sigma_{ee}^{-1/2} \langle \xi, \xi \rangle', \Sigma_{pp}^{1/2} \xi, \quad \text{(3.70)} \]
\[ \Omega_2 = \frac{1}{2} \Sigma_{ee}^{-1/2} (I - \xi \xi') \Sigma_{vv}^{-1/2}. \quad \text{(3.71)} \]

**Proof.** We observe that the columns of \( \Omega \) given by (3.71) are orthonormal. Also,

\[ \Sigma_{xx}^{-1/2} \xi_{xx}^{pp} \xi_{pp}^{1/2} = \Sigma_{xx}^{-1/2} \xi_{pp}^{1/2} \xi_{pp}^{1/2} \]
\[ = \Sigma_{pp}^{1/2} \xi_{pp}^{1/2} \Omega \]

where

\[ \Omega = \text{diag} \{ \nu_1, \nu_2, \ldots, \nu_k \}. \]

Thus,

\[ \Sigma_{ee}^{-1/2} \xi_{ee}^{-1/2} \Omega_1 = \Sigma_{ee}^{-1/2} \langle \xi, \xi \rangle' \xi_{xx} \xi_{pp}^{-1/2} + \Omega_1 \]
\[ = \Omega_1 (\Omega + I). \quad \text{(3.72)} \]
We also observe that

$$
\Sigma_{ee}^{-1/2} \Sigma_{zz} \Sigma_{ee}^{-1/2} \Omega(2) = \Sigma_{ee}^{-1/2} (\bar{\beta}, \bar{\gamma})' \Sigma_{xx} (\bar{\beta}, \bar{\gamma})(\bar{\beta}' = \Omega(2) + \Omega(2)'.
$$

(3.73)

The results follow from (3.72) and (3.73).

In the following theorem, we prove the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom. In the proof of Theorem 3.4, let $\omega$ be a point in the probability space of all sequences of observations, and let the superscript $(n)$ indicate that the quantity is calculated from the first $n$ elements of $\omega$. This notation is discontinued once Theorem 3.4 is proven. Let

$$
\Sigma_{zz} = (\bar{\beta}, \bar{\gamma})' \Sigma_{xx} (\bar{\beta}, \bar{\gamma}),
$$

$$
\hat{D}_{kk} = \hat{P}_{kk} A_1 \hat{P}_{kk}',
$$

$$
\hat{D}_{rr} = \hat{T}_{rr} A_2 \hat{T}_{rr}'.
$$

Theorem 3.4. Let the model (3.1) hold, and let Assumptions 3.5 and 3.7 hold. In addition, let either (a) Assumption 3.1, or (b) Assumption 3.2 hold. Let the estimators $\hat{\xi}_o(n), \hat{\gamma}(n), \hat{\xi}_{ee},$ and $\hat{\xi}_{xx}$ be defined by (3.4), (3.7), (3.58), and (3.57), respectively. Then, for both the cases (a) and (b), as $n \to \infty$, 
\[ \hat{\mathcal{E}}_{\hat{\mathcal{E}}_{\hat{E}}} \xrightarrow{\text{a.s.}} \mathcal{E}_{\mathcal{E}} \]

\[ \hat{\mathcal{E}}_{\hat{E}} \xrightarrow{\text{a.s.}} \mathcal{E} \]

\[ \hat{\mathcal{E}}_{\hat{\mathcal{E}}} \xrightarrow{\text{a.s.}} \mathcal{E}_{\mathcal{E}} \]

\[ \hat{\mathcal{E}}_{\hat{\mathcal{E}}_{\hat{X}}} \xrightarrow{\text{a.s.}} \mathcal{E}_{\mathcal{X}} \]

\[ \hat{\mathcal{E}}_{\hat{\mathcal{E}}_{\hat{K}}} \xrightarrow{\text{a.s.}} \mathcal{E}_{\mathcal{K}} \]

\[ \hat{\mathcal{E}}_{\hat{\mathcal{E}}_{\hat{R}}} \xrightarrow{\text{a.s.}} \mathcal{E}_{\mathcal{R}} \]

\[ \mathcal{E}_{\mathcal{E}} = \mathcal{E}_{\mathcal{X}} + \mathcal{E}_{\mathcal{K}} \]

\[ = \mathcal{E}_{\mathcal{X}} - \mathcal{E}_{\mathcal{U}} \mathcal{E}_{\mathcal{V}}^{-1} \mathcal{E}_{\mathcal{V}} \]

**Proof.** Since the eigenvectors \( \lambda_{i}^{(n)} \) are locally continuous functions of the elements in \( \frac{1}{2} \mathcal{S}_{\mathcal{E}}^{1/2} \mathcal{S}_{\mathcal{Z}}^{1/2} \mathcal{S}_{\mathcal{E}}^{-1/2} \), it follows from Lemma 3.1 that for \( i = 1, 2, \ldots, p \),
\[ \hat{\lambda}_i^{(n)} \longrightarrow \lambda_i, \text{ a.s.} \]

Fix \( \omega \) such that

\[ \hat{\lambda}_i^{(n)} (\omega) \longrightarrow \lambda_i, \quad (3.74) \]

\[ \mathbb{E}_{ZZ}^{(n)} (\omega) \longrightarrow \mathbb{E}_{ZZ}, \quad (3.75) \]

\[ \mathbb{E}_{EE}^{(n)} (\omega) \longrightarrow \mathbb{E}_{EE}, \quad \text{as } n \to \infty. \quad (3.76) \]

The set of \( \omega \) satisfying (3.74), (3.75), and (3.76) has probability one. Since \( \hat{Q}^{(n)} (\omega) \) is orthogonal for all \( n \), each element of \( \hat{Q}^{(n)} (\omega) \) is bounded. Thus, for every subsequence \( \{ \hat{Q}^{(n_i)} (\omega) \} \), there exists a convergent sub-subsequence \( \{ \hat{Q}^{(n_{ij})} (\omega) \} \). The limit of such a convergent sub-subsequence depends on the sub-subsequence, and we denote the limit by

\[ \bar{b}_{ij} (\omega) = (\bar{b}_{ij}^{(1)} (\omega), \bar{b}_{ij}^{(2)} (\omega)). \]

Since each \( \hat{Q}^{(n_{ij})} (\omega) \) is orthogonal, so is \( \bar{b}_{ij} (\omega) \). For all \( j \),

\[ s_{EE}^{(n_{ij})} - \frac{1}{2} \mathbb{E}_{ZZ}^{(n_{ij})} \mathbb{E}_{EE}^{(n_{ij})} - \frac{1}{2} \mathbb{E}_{EE}^{(n_{ij})} \hat{Q}^{(n_{ij})} (\omega) \]
\[
\mathcal{Q}_i^{(n_i, j)}(\mathbf{w}) \mathbf{diag}\{\lambda_1^{(n_i, j)}(\mathbf{w}), \ldots, \lambda_p^{(n_i, j)}(\mathbf{w})\}.
\]

Taking limits over \( j \) on both sides of the equality in (3.77), and using (3.74), (3.75), and (3.76), we obtain

\[
\mathcal{Q}_i^{(n_i, j)}(\mathbf{w}) \mathbf{diag}\{\lambda_1^{(n_i, j)}(\mathbf{w}), \ldots, \lambda_p^{(n_i, j)}(\mathbf{w})\}.
\]

where \( \lambda_1 \) are defined in (3.70). Hence, the columns of \( \mathcal{Q}_i^{(n_i, j)}(\mathbf{w}) \) are orthogonal eigenvectors of \( \Sigma_{\mathbf{e}_j}^{-1/2} \Sigma_{\mathbf{z}} \mathbf{z}^{\top} \). Since by Lemma 3.2

\[
\lambda_k = 1 + \nu_k > 1 = \lambda_i, \quad i = k+1, \ldots, p,
\]

the two eigensubspaces of \( \Sigma_{\mathbf{e}_j}^{-1/2} \Sigma_{\mathbf{z}} \mathbf{z}^{\top} \) corresponding to the first \( k \) eigenvalues and the last \( r \) eigenvalues, respectively, are unique. Thus, there exist orthogonal matrices \( \mathcal{Q}_i^{(1)}; k \times k \), and \( \mathcal{Q}_i^{(2)}; r \times r \), such that

\[
\mathcal{E}_i^{(1)}(\mathbf{w}) = \mathcal{Q}_i^{(1)} \mathcal{Q}_i^{(2)}
\]

and

\[
= \mathcal{Q}_i^{(1)} \text{ block diag}\{\mathcal{Q}_i^{(11)}, \mathcal{Q}_i^{(12)}, \ldots, \mathcal{Q}_i^{(1s)}\},
\]
\[ B_{ij}^{(2)}(\omega) = \mathcal{Q}_1^{(2)} \mathcal{Q}_1^{(2)} \]

where \( s \) is the number of distinct roots \( \nu_u \), \( \mathcal{Q}_1^{(1)} \) is orthogonal with dimension corresponding to the multiplicity of \( \nu_u \), and \( \mathcal{Q}_1^{(2)} \) and \( \mathcal{Q}_2^{(2)} \) are defined by (3.71). Therefore, by (3.70) and (3.71), as \( j \to \infty \),

\[ \begin{align*}
\mathcal{Q}_1^{(1)}(\omega)[\mathcal{Q}_1^{(1)}(\omega)]' & \to B_{ij}^{(1)}(\omega) (\mathcal{Q}_1^{(1)}(\omega))' \\
& = \mathcal{Q}_1^{(1)} B_{ij}^{(1)} \mathcal{Q}_1^{(1)} \\
& = \mathcal{Q}_1^{(1)} Q_1^{(1)} \\
& = \mathcal{E}_{ee}^{-\frac{1}{2}}(\mathcal{E}_{pp} \mathcal{E}_{pp}^{1/2} \mathcal{E}_{pp}^{1/2} (\mathcal{E}_{pp} \mathcal{E}_{pp}) \mathcal{E}_{ee}^{-1/2} \\
& = \mathcal{E}_{ee}^{-\frac{1}{2}}(\mathcal{E}_{pp} \mathcal{E}_{pp}^{1/2} \mathcal{E}_{pp}^{1/2} (\mathcal{E}_{pp} \mathcal{E}_{pp}) \mathcal{E}_{ee}^{-1/2} , \quad (3.78)
\end{align*} \]

\[ \begin{align*}
\mathcal{Q}_2^{(2)}(\omega)[\mathcal{Q}_2^{(2)}(\omega)]' & \to \mathcal{Q}_2^{(2)} Q_2^{(2)} \\
& = \mathcal{E}_{ee}^{1/2} (\mathcal{E}_{pp}^{-1} (\mathcal{E}_{pp}^{-1}) \mathcal{E}_{ee}^{1/2} , \quad (3.79)
\end{align*} \]
\[
\begin{align*}
\longrightarrow \mathcal{Q}(1) \text{ block diag}\{\mathcal{Q}_{ij}^{(1)}, \ldots, \mathcal{Q}_{ij}^{(1s)}\} \text{ block diag}\{\lambda_1 I, \ldots, \lambda_s I\} \\
\text{ block diag}\{\mathcal{Q}_{ij}^{(1)}', \ldots, \mathcal{Q}_{ij}^{(1s)}'\} \mathcal{Q}'(1)
\end{align*}
\]

\[
\mathcal{Q}(1) = \text{diag}\{\lambda_1, \ldots, \lambda_k\} \mathcal{Q}'(1)
\]

\[
= \mathcal{E}_{ee}^{-1/2} (\mathcal{E}, I)' \mathcal{E}_{pp}^{-1/2} (\mathcal{E} + \mathcal{D}_v)' \mathcal{E}_{ee}^{-1/2}
\]

\[
= \mathcal{E}_{ee}^{-1/2} (\mathcal{E}, I)' \mathcal{E}_{pp}^{-1/2} (\mathcal{E}, I) \mathcal{E}_{ee}^{-1/2} + \mathcal{E}_{ee}^{-1/2} \mathcal{E}_{zz} \mathcal{E}_{ee}^{-1/2}
\]

\[
(3.80)
\]

\[
\mathcal{Q}_{(2)}^{(n_{ij})} (\mathcal{E}) A_2^{(n_{ij})} (\mathcal{E}) (\mathcal{Q}_{(2)}^{(n_{ij})} (\mathcal{E}))' \longrightarrow \mathcal{Q}_{(2)} \mathcal{Q}_{ij}^{(2)} \mathcal{Q}_{ij}^{(2)} \mathcal{Q}'(2)
\]

\[
= \mathcal{Q}(2) \mathcal{Q}'(2)
\]

\[
= \mathcal{E}_{ee}^{-1/2} (\mathcal{E}, -\mathcal{E})' \mathcal{E}_{vv}^{-1/2} (\mathcal{E}, -\mathcal{E}) \mathcal{E}_{ee}^{-1/2}
\]

\[
(3.81)
\]

Thus, by (3.68) and (3.80), as \(j \rightarrow \infty\),

\[
\mathcal{P}_{kk}^{(n_{ij})} (\mathcal{E}) \longrightarrow \mathcal{E}_{\eta \eta}
\]

\[
(3.82)
\]

By (3.68) and (3.81), as \(j \rightarrow \infty\),
By (3.68) and (3.79),
\[ \hat{D}_{\text{rr}}^{(n_{ij})}(\omega) \longrightarrow \Sigma^{-1}_{v v}. \]  
(3.83)

Also, by (3.68), (3.78), and (3.80), as \( j \rightarrow \infty \),
\[ \hat{L}_{\text{zz}}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{zz}. \]  
(3.85)

and thus
\[ \hat{L}_{\text{xx}}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{xx}. \]  
(3.87)

By (3.67), (3.68), (3.85), and Assumption 3.7, as \( j \rightarrow \infty \),
\[ \hat{L}_{\text{ee}}^{(n_{ij})}(\omega) \longrightarrow \Sigma_{ee}. \]  
(3.88)

If every subsequence has a sub-subsequence converging to the same limit, 
then the sequence converges to the limit for such an \( \omega \). Since the 
set of such \( \omega \) has probability one, the results for \( \hat{D}_{\text{kk}}^{(n)} \), \( \hat{D}_{\text{rr}}^{(n)} \), 
\( \hat{L}_{\text{rr}}^{(n)} \), \( \hat{L}_{\text{xx}}^{(n)} \), \( \hat{L}_{\text{ee}}^{(n)} \), and \( \Sigma_{ee} \) follows from (3.82), (3.83),
(3.84), (3.86), (3.87), and (3.88). Finally, for either the case (a) or (b), as \( n \to \infty \),

\[
\bar{X} \to \mu_X, \text{ a.s.,}
\]

\[
\bar{Y} \to \beta_o + \mu_X \beta, \text{ a.s.,}
\]

and thus the result for \( \hat{\beta}^{(n)}_o \) follows from that for \( \hat{\beta}^{(n)} \).

2. Limiting distribution

The asymptotic normality of the maximum likelihood estimators adjusted for degrees of freedom holds without the normality assumption on \( \xi_t \) and \( \xi_t' \). But, to write down an explicit form of the limiting covariance matrix, we assume that \( \xi_t \) are normal and that \( \Sigma_{\epsilon\epsilon} \) is a multiple of a Wishart matrix. The following theorem shows that inferences about \( \hat{\beta}_o \), \( \hat{\beta} \), and \( \Sigma_{\epsilon\epsilon} \) based on the limiting results are the same for a wide range of assumptions on \( \xi_t \).

**Theorem 3.5.** Let the model (3.1) hold, and let Assumptions 3.5 through 3.8 hold.

(a) If either Assumption 3.1 or Assumption 3.2 holds, then

\[
\frac{1}{n^{1/2}} \begin{bmatrix}
(\hat{\beta}_o - \beta_o)' \\
\text{vec}(\hat{\beta} - \beta) \\
\text{vech}(\hat{\Sigma}_{\epsilon\epsilon} - \Sigma_{\epsilon\epsilon})
\end{bmatrix} \xrightarrow{L} \mathcal{N} \left( \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}, \\
\begin{bmatrix}
V_{\beta\beta} & V_{\beta\epsilon} & V_{\epsilon\epsilon} \\
V_{\beta\epsilon}' & V_{\epsilon\epsilon} & V_{\epsilon\epsilon}' \\
V_{\epsilon\epsilon}' & V_{\epsilon\epsilon} & V_{\epsilon\epsilon}'
\end{bmatrix} \right),
\]

where
\[
V_{\infty} = \Sigma_{\nu\nu} + (I_{rrx} \otimes \mu_x)\Sigma_{\beta\beta} (I_{rrx} \otimes \mu'_{x}),
\]

\[
V_{0\delta} = - (I_{rrx} \otimes \mu_x)\Sigma_{\beta\beta},
\]

\[
V_{0\epsilon} = - (I_{rrx} \otimes \mu_x)\Sigma_{\beta\epsilon},
\]

\[
\Sigma_{\beta\beta} = \Sigma_{\nu\nu} \{\Sigma_{xx}^{-1}(\Sigma_{xx} + (1+c)\Sigma_{\rho\rho})\Sigma_{xx}^{-1}\},
\]

\[
\Sigma_{\beta\epsilon} = -2c[\Sigma_{\nu\nu} \{\Sigma_{xx}^{-1} \Sigma_{\rho\rho} (\xi, I)\}]\psi_p',
\]

\[
\Sigma_{\epsilon\epsilon} = 2c\psi_p [(\Sigma_{\epsilon\epsilon} \otimes \Sigma_{\epsilon\epsilon}) - (1+c^{-1})^{-1}(\Sigma_{\epsilon\nu} \Sigma_{\nu\nu}^{-1}\Sigma_{\epsilon\nu})(\Sigma_{\epsilon\nu} \Sigma_{\nu\nu}^{-1}\Sigma_{\epsilon\nu})]\psi_p,
\]

\[
\Sigma_{\rho\rho} = \Sigma_{uu} - \Sigma_{uv}^1 \Sigma_{\epsilon\nu}^{-1}\Sigma_{uv},
\]

\[
\Sigma_{\nu\nu} = (I, - \beta') \Sigma_{\epsilon\epsilon} (I, - \beta')',
\]

\[
\Sigma_{\epsilon\nu} = \Sigma_{\nu\nu}^1 = \Sigma_{ue} - \Sigma_{uu} \beta,
\]

\[
\Sigma_{\nu\epsilon} = \Sigma_{\epsilon\epsilon}^1 = (I, - \beta') \Sigma_{\epsilon\epsilon},
\]

\[
c = \lim_{n \to \infty} d^{-1} n.
\]

(b) If Assumption 3.2 and Assumption 3.3 hold, then
\[
\frac{1}{\sqrt{2}} \begin{pmatrix}
(\beta_0 - \beta_0)' \\
\text{vec}(\beta - \beta) \\
\text{vech}(\xi_{xx} - \xi_{xx}) \\
\text{vech}(\xi_{xx} - \xi_{xx})
\end{pmatrix}
\xrightarrow{L \rightarrow N}
\begin{pmatrix}
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0
\end{pmatrix} \\
\begin{pmatrix} v_{oo} \\ v_{o\beta} \\ v_{o\varepsilon} \\ v_{ox}
\end{pmatrix} \\
\begin{pmatrix} v'_{o\beta} \\ v'_{o\varepsilon} \\ v'_{o\varepsilon} \\ v'_{ox}
\end{pmatrix} \\
\begin{pmatrix} v'_{o\beta} \\ v'_{o\varepsilon} \\ v'_{o\varepsilon} \\ v'_{xx}
\end{pmatrix}
\end{pmatrix},
\]

where

\[V_{ox} = -(A_{rr} \otimes \mu_x) V_{\beta x},\]

\[V_{\beta x} = 2[\Sigma_{\psi u} \otimes \Sigma_{xx}^{-1} (\Sigma_{xx} + (1+c)\Sigma_{pp})] \psi_k',\]

\[V_{\varepsilon x} = 2\psi_p [(\Sigma_{\varepsilon u} \otimes \Sigma_{\varepsilon u}) - (\Sigma_{\varepsilon u} \otimes \Sigma_{\varepsilon u})] \psi_k',\]

\[V_{xx} = \Gamma_x + 2\psi_k [\Sigma_{xx} \otimes \Sigma_{uu} + \Sigma_{uu} \otimes \Sigma_{xx} + (1+c)(\Sigma_{uu} \otimes \Sigma_{uu} - \Sigma_{uu} \otimes \Sigma_{uu})] \psi_k',\]

\[\Gamma_x = V[\text{vech}(x_t' - \mu_x)'(x_t' - \mu_x)],\]

\[\Sigma_{uu} = \Sigma_{uu}^{-1} \Sigma_{uu},\]

\[\Sigma_{u\varepsilon} = \Sigma_{uu}^{-1} \Sigma_{u\varepsilon} \cdot\]
Proof. (a) We observe that

\[
S_{\varepsilon \varepsilon}^{-1} \mathbb{W}_{ZZ} = S_{\varepsilon \varepsilon}^{-1/2} S_{\varepsilon \varepsilon}^{-1/2} \mathbb{W}_{ZZ} S_{\varepsilon \varepsilon}^{-1/2} S_{\varepsilon \varepsilon}^{-1/2}
\]

\[
= S_{\varepsilon \varepsilon}^{-1/2} (\Omega_{(1)} \hat{A}_{1} \Omega'_{(1)} + \Omega_{(2)} \hat{A}_{2} \Omega'_{(2)} ) S_{\varepsilon \varepsilon}^{1/2}
\]

\[
= S_{\varepsilon \varepsilon}^{-1} \hat{\mathbb{P}}_{(1)} \hat{A}_{1} \hat{\mathbb{P}}'_{(1)} + \hat{\mathbb{P}}_{(2)} \hat{A}_{2} \hat{\mathbb{P}}'_{(2)} S_{\varepsilon \varepsilon}
\]

\[
= S_{\varepsilon \varepsilon}^{-1} (\hat{\mathbb{P}}_{(1)} \hat{\mathbb{P}}'_{(1)} + \hat{\mathbb{P}}_{(2)} \hat{\mathbb{P}}'_{(2)} ) S_{\varepsilon \varepsilon}
\]

\[
(\hat{\varepsilon}, \hat{I}) S_{\varepsilon \varepsilon}^{-1} S_{\varepsilon \varepsilon} \mathbb{W}_{ZZ} (\hat{I}, -\hat{\varepsilon}')' = (\hat{\varepsilon}, \hat{I}) S_{\varepsilon \varepsilon}^{-1} (\hat{\varepsilon}, \hat{I})' \hat{P}_{kk} (\hat{\varepsilon} - \hat{\varepsilon})
\]

\[- (\hat{\varepsilon} - \hat{\varepsilon}) \hat{P}_{rr} (\hat{I}, -\hat{\varepsilon}') S_{\varepsilon \varepsilon}(\hat{I}, -\hat{\varepsilon}')'
\]

\[= U_{kk} (\hat{\varepsilon} - \hat{\varepsilon}) + (\hat{\varepsilon} - \hat{\varepsilon}) U_{rr}
\]

where

\[U_{kk} = (\hat{\varepsilon}, \hat{I}) S_{\varepsilon \varepsilon}^{-1} (\hat{\varepsilon}, \hat{I})' \hat{P}_{kk}
\]

\[U_{rr} = - \hat{P}_{rr}(\hat{I}, -\hat{\varepsilon}') S_{\varepsilon \varepsilon}(\hat{I}, -\hat{\varepsilon})
\]

Hence, using Theorem 2.2, we obtain
\[ \vec{\text{vec}}(\xi, \bar{\xi}) = \vec{\text{vec}}[\xi, \bar{\xi}]^{-1} \text{vec}(\xi, \bar{\xi}) \]

By Assumption 3.7, Theorem 2.1, and Theorem 3.4,

\[
(I_{rr} \oplus I_{kk}) \text{vec}(\pi) + (I_{rr} \oplus I_{kk}) \text{vec}(\pi')
\]

and the limit is nonsingular. Thus,

\[
(I_{rr} \oplus I_{kk}) \text{vec}(\pi) + (I_{rr} \oplus I_{kk}) \text{vec}(\pi')
\]

is nonsingular for large \( n \). Hence, by (3.90),

\[
\text{vec}(\pi) = (I_{rr} \oplus I_{kk})^{-1} \text{vec}(\xi, \bar{\xi})^{-1} \text{vec}(\xi, \bar{\xi}) \]

\[
= (I_{rr} \oplus I_{kk})^{-1} \text{vec}(\xi, \bar{\xi})^{-1} \text{vec}(\xi, \bar{\xi}) \]

Using Theorem 2.13 if Assumption 3.1 holds, and using Theorem 2.10 if Assumption 3.2 holds, we obtain

\[(\xi, I)' \mathbb{M}_{\varepsilon \varepsilon} + \mathbb{M}_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon} = 0_p(n^{-1/2})\]  \hspace{1cm} (3.93)

By Assumption 3.8,

\[S_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon} = 0_p(n^{-1/2})\]  \hspace{1cm} (3.94)

Using the Taylor expansion of a matrix, we have

\[S_{\varepsilon \varepsilon}^{-1} = \Sigma_{\varepsilon \varepsilon}^{-1} - \Sigma_{\varepsilon \varepsilon}^{-1}(S_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon})\Sigma_{\varepsilon \varepsilon}^{-1} + 0_p(n^{-1})\]

Hence,

\[(\xi, I)S_{\varepsilon \varepsilon}^{-1}[(\xi, I)' \mathbb{M}_{\varepsilon \varepsilon} + \mathbb{M}_{\varepsilon \varepsilon}](I, \xi)^)' + 0_p(n^{-1})\]

\[= (\xi, I)\Sigma_{\varepsilon \varepsilon}^{-1} - \Sigma_{\varepsilon \varepsilon}^{-1}(S_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon})\Sigma_{\varepsilon \varepsilon}^{-1}[(\xi, I)' \mathbb{M}_{\varepsilon \varepsilon} + \mathbb{M}_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon}](I, \xi)^)' + 0_p(n^{-1})\]

\[= (\xi, I)\Sigma_{\varepsilon \varepsilon}^{-1}[(\xi, I)' \mathbb{M}_{\varepsilon \varepsilon} + \mathbb{M}_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon}](I, \xi)^)' - (\xi, I)\Sigma_{\varepsilon \varepsilon}^{-1}(S_{\varepsilon \varepsilon} - \Sigma_{\varepsilon \varepsilon})(I, \xi)^)' + 0_p(n^{-1})\]
\[ + o_p(n^{-1}) \]

\[ = \sum_{\mathcal{P}}^{-1}(m_{xx} + m_{\xi} - S_{\xi}) + o_p(n^{-1}) \]

\[ = o_p(n^{-1/2}) \quad (3.95) \]

where

\[ m_{xx} = m_{xx}(I, -g')' \]

\[ m_{\xi} = \Sigma_{pp}(g, I) \Sigma_{ee}^{-1} m_{ee}(I, -g')' \]

\[ S_{\xi} = \Sigma_{pp}(g, I) \Sigma_{ee}^{-1} S_{ee}(I, -g')' \]

and we have used Theorem 2.1. By (3.91), (3.92), and (3.95),

\[ \text{vec}(\hat{g} - g) = [I_{xx} \otimes (\Sigma_{pp}^{-1})^{-1}\text{vec}(\Sigma_{pp}^{-1}(m_{xx} + m_{\xi} - S_{\xi})) + o_p(n^{-1/2}) \]

\[ = (I_{xx} \otimes \Sigma_{pp})^{-1}(I_{xx} \otimes \Sigma_{pp})\text{vec}(m_{xx} + m_{\xi} - S_{\xi}) + o_p(n^{-1/2}) \]

\[ = (I_{xx} \otimes \Sigma_{pp})\text{vec}(m_{xx} + m_{\xi} - S_{\xi}) + o_p(n^{-1/2}) \quad (3.96) \]

Observing that
\[ -\frac{1}{2} \mathbb{P}_{zz} S_{ee} \frac{1}{2} \hat{\Omega}(2) = \hat{\Omega}(2) \hat{A}_2 \]

we have

\[ \mathbb{P}_{zz}(I, -\hat{e}')' \hat{T}_{rr} = \mathbb{P}_{zz} \hat{T}(2) \]

\[ = \mathbb{P}_{zz} S_{ee} \frac{1}{2} \hat{\Omega}(2) \]

\[ = \frac{1}{2} \hat{\Omega}(2) \hat{A}_2 \]

\[ = S_{ee} (I, -\hat{e}')' \hat{T}_{rr} \hat{A}_2 \]

Multiplying \((I, -\hat{e}')'\) from the left, we obtain

\[ (I, -\hat{e}') \mathbb{P}_{zz}(I, -\hat{e}')' \hat{T}_{rr} = (I, -\hat{e}) S_{ee} (I, -\hat{e}')' \hat{T}_{rr} \hat{A}_2 \]

\[ = S_{ee} (I, -\hat{e}')' \hat{T}_{rr} \hat{A}_2 \]

Hence,

\[ \hat{T}_{rr} (\hat{A}_2 - I) \hat{T}_{rr}' = [((I, -\hat{e}') S_{ee} (I, -\hat{e}')')^{-1} (I, -\hat{e}') \mathbb{P}_{zz} (I, -\hat{e}')]' \]

\[ - I] \hat{T}_{rr} \hat{T}_{rr}' \]

\[ = [((I, -\hat{e}') S_{ee} (I, -\hat{e}')')^{-1} (I, -\hat{e}') (\mathbb{P}_{zz} - S_{ee})]^{-1} \]

\[ = [\mathbb{P}_{zz} - S_{ee}]^{-1} (I, -\hat{e}') S_{ee} (I, -\hat{e}')' \]

\[ = [\mathbb{P}_{zz} - S_{ee}]^{-1} (I, -\hat{e}') \]

\[ = [\mathbb{P}_{zz} - S_{ee}]^{-1} \hat{A}_2 \]
\[(\mathbb{I}, \hat{\mathcal{E}}')' \hat{\mathcal{T}}_{tr} \hat{T}_{tr}' \]  \hspace{1cm} (3.98)

By (3.96),
\[\hat{\mathcal{E}} = \mathcal{E} + o_p(n^{-1/2}) \]  \hspace{1cm} (3.99)

Thus,
\[(\mathbb{I}, \hat{\mathcal{E}}')S_{ee} (\mathbb{I}, \hat{\mathcal{E}}')' = \mathcal{E}_{vv} + o_p(n^{-1/2}) \]  \hspace{1cm} (3.100)

Also, by Theorem 3.4,
\[\hat{T}_{tr} \hat{T}_{tr}' = \mathcal{E}_{vv}^{-1} + o_p(1) \]  \hspace{1cm} (3.101)

Using (3.93), (3.94), and (3.99), we have
\[(\mathbb{I}, \hat{\mathcal{E}}')(m_{zz} - S_{ee})(\mathbb{I}, \hat{\mathcal{E}}')' = (\mathbb{I}, \hat{\mathcal{E}}')(m_{ee} + m_{ee} - S_{ee} + S_{ee} - S_{ee})\]
\[\left[(\mathbb{I}, \hat{\mathcal{E}}')' - (\mathcal{E}_v, \hat{\mathcal{E}}' - \mathcal{E}_v)\right] \]
\[= (\mathbb{I}, \hat{\mathcal{E}}')(m_{ee} - S_{ee})(\mathbb{I}, \hat{\mathcal{E}}')' + o_p(n^{-1})\]
\[= \mathcal{E}_{vv} - S_{vv} + o_p(n^{-1})\]
\[= o_p(n^{-1/2}) \]  \hspace{1cm} (3.102)
where

\[ \eta_{vv} = (n-1)^{-1} \sum_{t=1}^{n} (z_{t} - \bar{z})(z_{t} - \bar{z}) , \]

\[ \bar{z} = \bar{v} - u_{t} \hat{\theta} , \]

\[ \bar{y} = \bar{y} = n^{-1} \sum_{t=1}^{n} y_{t} , \]

\[ S_{vv} = (I_{n} - \hat{\theta} \hat{\theta}) S_{ee} (I_{n} - \hat{\theta} \hat{\theta})' . \]

By (3.98), (3.100), (3.101), and (3.102),

\[ \hat{T}_{rr} (\hat{\theta}_{2} - I) \hat{T}_{rr}' = S_{vv}^{-1} (\eta_{vv} - S_{vv}) S_{vv}^{-1} + o_{p}(n^{-1/2}) \]

\[ = o_{p}(n^{-1/2}) . \quad (3.103) \]

We have, from (3.9) and (3.89),

\[ \mu_{zz} = (\hat{z}, I)'(\hat{\Sigma}_{xx} + P_{kk} \hat{\Sigma}'_{kk})(\hat{z}, I) + S_{ee} (I_{n} - \hat{\theta} \hat{\theta})' D_{rr} (I_{n} - \hat{\theta} \hat{\theta}) S_{ee} \]

\[ = (\hat{z}, I)' \hat{\Sigma}_{xx} (\hat{z}, I) + P_{(1)} \hat{P}_{(1)} + S_{ee} \hat{T}_{(2)} \hat{A}_{2} \hat{T}_{(2)} S_{ee} \]

\[ = (\hat{z}, I)' \hat{\Sigma}_{xx} (\hat{z}, I) + 1/2 \hat{S}_{ee} (I_{n} - \hat{Q}_{(2)} \hat{Q}_{(2)}') S_{ee}^{1/2} \]

\[ + 1/2 \hat{S}_{ee} \hat{Q}_{(2)} \hat{A}_{2} \hat{Q}_{(2)} S_{ee}^{1/2} . \]
Therefore,

\[
\hat{\Sigma}_{ee} = (n-1+d)^{-1}\left\{ d \Sigma_{ee} + (n-1)\Sigma_{ee}(\mathbf{I}, \mathbf{\hat{\Sigma}}')^{\prime}\mathbf{T}_{rr}(\mathbf{\hat{\Sigma}} - \mathbf{I})\mathbf{T}_{rr}'(\mathbf{I}, \mathbf{\hat{\Sigma}})\Sigma_{ee}\right\}.
\]

Thus, using (3.94), (3.99), and (3.103), we have

\[
\hat{\Sigma}_{ee} = \Sigma_{ee} + (n-1+d)^{-1}(n-1)\Sigma_{ee}\Sigma_{vv}^{-1}(\mathbf{\Pi}_{vv} - \Sigma_{vv})\Sigma_{vv}^{-1}\Sigma_{ee} + o_p(n^{-1/2}).
\]  

(3.105)

Since under either Assumption 3.1 or Assumption 3.2

\[
\bar{\mathbf{x}} = \mu_x + o_p(1),
\]

it follows from (3.99) that

\[
\mathbf{\hat{s}_o} = \bar{\mathbf{x}} - \bar{\mathbf{x}} \mathbf{\hat{s}} - \bar{\mathbf{x}} (\mathbf{\hat{s}} - \mathbf{\hat{s}})
\]

\[
= \bar{\mathbf{y}} - \mu_x (\mathbf{\hat{s}} - \mathbf{\hat{s}}) + o_p(n^{-1/2})
\]

\[
= \bar{\mathbf{y}} - \{\text{vec}(\mathbf{\hat{s}} - \mathbf{\hat{s}})\}'(\mathbf{I}_{txr} = \mathbf{\hat{y}}') + o_p(n^{-1/2}).
\]  

(3.106)
Using Theorem 2.13 if Assumption 3.1 holds, and using Theorem 2.9 if Assumption 3.2 holds, we have

\[
\begin{bmatrix}
X' \\
\text{vec } \xi \\
\text{vec } \mu \\
\text{vech}(\Sigma - \Sigma) \\
\end{bmatrix}
\xrightarrow{L} N(0, \Sigma_1),
\]

(3.107)

where

\[
\Sigma_1 = \text{block diag}\{\Sigma, \Sigma \odot \Sigma, \Sigma, \Sigma \odot \Sigma, 2\Psi_T(\Sigma \odot \Sigma)\}.
\]

Also, by Assumption 3.8,

\[
\begin{bmatrix}
\text{vech}(\Sigma - \Sigma) \\
\text{vec } \xi \\
\text{vech}(\Sigma - \Sigma) \\
\end{bmatrix}
= (nd^{-1})^{1/2} \frac{1}{d^{1/2}}
\begin{bmatrix}
\text{vech}(\Sigma - \Sigma) \\
\text{vec } \xi \\
\text{vech}(\Sigma - \Sigma) \\
\end{bmatrix}
\xrightarrow{L} N(0, \Sigma_1),
\]

(3.108)

where

\[
F_1 = c \begin{bmatrix}
2\Psi_p(\Sigma \odot \Sigma) \Psi_p & \Psi & 2\Psi_p(\Sigma \odot \Sigma) \Psi_T \\
\Sigma & \Sigma & 0 \\
2\Psi_T(\Sigma \odot \Sigma) \Psi_p & 0 & 2\Psi_T(\Sigma \odot \Sigma) \Psi_T \\
\end{bmatrix}
\]
\[ E_{12} = 2\psi_p (E_{\epsilon\epsilon} \equiv E_{\epsilon\epsilon}) \psi_p' \{(I - \hat{\Sigma}^{-1}) \equiv (E_{\epsilon\epsilon}^{-1} (\Sigma_{\epsilon\epsilon} \equiv \Sigma_{\epsilon\epsilon}) \psi_p \} \]

\[ = 2 (E_{\epsilon\epsilon} \equiv E_{\epsilon\epsilon}) \psi_p' \]

By (3.96), (3.105), (3.106), (3.107), (3.108), and the independence of \( E_{\epsilon\epsilon} \) and \( Z_t \), \( n^{1/2} \{ (E_{\epsilon\epsilon}^{-1} (\Sigma_{\epsilon\epsilon} \equiv \Sigma_{\epsilon\epsilon}) \psi_p \} \) converges in law to a multivariate normal random vector, and also

\[ V_{\beta \beta} = (I \equiv E_{xx}^{-1}) [E_{vv} \equiv E_{xx} + (1+c)(E_{vv} \equiv E_{pp})] (I \equiv E_{xx}^{-1}) \]

\[ = E_{vv} \equiv E_{xx}^{-1} [E_{xx} + (1+c)E_{pp}] E_{xx}^{-1} \]

\[ V_{\beta \epsilon} = -(I \equiv E_{xx}^{-1}) 2c (E_{ve} \equiv E_{ee}) \psi_p' \]

\[ = -2c (E_{ve} \equiv E_{xx}^{-1} E_{ee}) \psi_p' \]

\[ V_{\epsilon \epsilon} = 2c \psi_p (E_{\epsilon\epsilon} \equiv E_{\epsilon\epsilon}) \psi_p' + (1+c)^{-1} 2\psi_p [ (E_{\epsilon\epsilon}^{-1} E_{\epsilon\epsilon}) \psi_p' \]

\[ \equiv (E_{\epsilon\epsilon}^{-1} E_{\epsilon\epsilon}) \psi_p' 2(1+c) \psi_p' (E_{vv} \equiv E_{vv}) \psi_p' E_{\epsilon\epsilon}^{-1} [ (E_{vv}^{-1} E_{vv}) \equiv (E_{vv}^{-1} E_{vv}) \psi_p' \]

\[ - 2c (1+c)^{-1} \psi_p [ (E_{\epsilon\epsilon}^{-1} E_{\epsilon\epsilon}) \equiv (E_{\epsilon\epsilon}^{-1} E_{\epsilon\epsilon}) \psi_p' \]

\[ - 2c (1+c)^{-1} \psi_p (E_{vv} \equiv E_{vv}) \psi_p' [ (E_{vv}^{-1} E_{vv}) \equiv (E_{vv}^{-1} E_{vv}) \psi_p' \]
= 2c\psi_p(Σ_{εe} m_{εe})\psi_p' + 2(1+c)(1+c^{-1})^{-2}\psi_p[(Σ_{εe} \Sigma^{-1}_{εe}) m(Σ_{εe} \Sigma^{-1}_{εe})]\psi_p'

- 4c(1+c^{-1})^{-1}\psi_p[(Σ_{εe} \Sigma^{-1}_{εe}) m(Σ_{εe} \Sigma^{-1}_{εe})]\psi_p'

= 2c\psi_p[(Σ_{εe} m_{εe}) - (1+c^{-1})^{-1}(Σ_{εe} \Sigma^{-1}_{εe}) m(Σ_{εe} \Sigma^{-1}_{εe})]\psi_p'

where

\[ Σ_{εe} = Σ_{pp}(ξ, \bar{J}) \]

and we have used for any \((a \times b)\) matrix \(A\)

\[ ψ_a(A \oplus A)ψ_b = ψ_a(Α \oplus A) \]

\[ = ψ_a(Α \oplus A) \cdot \]

(b) Observe that

\[ Σ_{εe}(ξ, \widehat{ξ})' = Σ_{εe}(ξ, \widehat{ξ})' + o_p(n^{-1/2}) \]

\[ = Σ_{εe} + o_p(n^{-1/2}) \cdot \]

(3.109)

Using (3.103) and (3.109), we write the lower \(k \times k\) corner of (3.104) as
\[ \hat{\Sigma}_{XX} = m_{XX} - S_{uu} - \Sigma_{uv} \Sigma_{vv}^{-1}(\bar{m}_{vv} - S_{vv})\Sigma_{vv}^{-1}\Sigma_{vu} + o_p(n^{-1/2}) \, . \]

Therefore,
\[ \hat{\Sigma}_{XX} - \Sigma_{XX} = (m_{XX} - \Sigma_{XX}) - (\bar{m}_{uu} - \Sigma_{uu}) - (S_{uu} - \Sigma_{uu}) \]
\[ + (\bar{S}_{uu} - \Sigma_{uu}) + o_p(n^{-1/2}) \, , \] (3.110)

where
\[ \bar{m}_{uu} = \Sigma_{uv} \Sigma_{vv}^{-1} \Sigma_{vu} \, , \]
\[ \bar{S}_{uu} = \Sigma_{uv} \Sigma_{vv}^{-1}\Sigma_{vu} \, , \]
\[ \bar{\Sigma}_{uu} = \Sigma_{uv} \Sigma_{vv}^{-1}\Sigma_{vu} \, , \]
\[ \bar{\Sigma}_{XX} = \Sigma_{XX} + \Sigma_{uu} \, . \]

Under Assumption 3.2 and Assumption 3.3,
$$\sqrt{n} \begin{bmatrix} \vec{X}' \\ \text{vec } \mathbb{E}_{XX} \\ \text{vec } \mathbb{E}_{XX} \\ \text{vech}(\mathbb{E}_{vv} - \mathbb{E}_{vv}) \\ \text{vech}(\mathbb{E}_{xx} - \mathbb{E}_{xx}) \\ \text{vech}(\mathbb{E}_{uu} - \mathbb{E}_{uu}) \end{bmatrix} \xrightarrow{L} \mathbb{N} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} G(1) \\ G(12) \end{bmatrix} \begin{bmatrix} 0 \\ G(12) \end{bmatrix} \begin{bmatrix} G(12) \\ G(2) \end{bmatrix}, \quad (3.111)$$

where

$$G(12) = \begin{bmatrix} 0 & 2\psi_k (\Sigma_{uv} \otimes \Sigma_{xx}) & G_{53} & 2\psi_r (\Sigma_{vu} \otimes \Sigma_{uu}) \psi_k' \\ 0 & 0 & 0 & 2\psi_r (\Sigma_{vu} \otimes \Sigma_{uu}) \psi_k' \end{bmatrix},$$

$$G(2) = \begin{bmatrix} I_x + 2\psi_k (\Sigma_{xx} \otimes \Sigma_{uu} + \Sigma_{uu} \otimes \Sigma_{xx} + \Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' & 2\psi_k (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' \\ 2\psi_k (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' & 2\psi_k (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' \end{bmatrix},$$

$$G_{53} = 2\psi_k [\Sigma_{uv} \otimes [\Sigma, \Sigma]^{-1} (\Sigma, \Sigma) \otimes \Sigma_{pp}].$$

Also, by Assumption 3.8,
\[
\begin{align*}
\begin{pmatrix}
\text{vech}(\Sigma_{ee} - \Sigma_{ee}) \\
\text{vec } \Sigma_{ev} \\
\text{vech}(\Sigma_{vv} - \Sigma_{vv}) \\
\text{vech}(\Sigma_{uu} - \Sigma_{uu}) \\
\text{vech}(\Sigma_{uu}^{**} - \Sigma_{uu}^{**})
\end{pmatrix}
\end{align*}
\xrightarrow{\frac{1}{n^2}}
\mathcal{N}
\left[
\begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
E_{(1)} & E_{(12)} \\
E_{(12)} & E_{(2)}
\end{pmatrix}
\right],
(3.112)
\]

where

\[
E_{(12)} = c
\begin{pmatrix}
2\Psi_p (\Sigma_{eu} \otimes \Sigma_{eu}) \psi_k' \\
2\Psi_p (\Sigma_{eu}^{**} \otimes \Sigma_{eu}^{**}) \psi_k' \\
0 \\
0 \\
2\Psi_r (\Sigma_{vu} \otimes \Sigma_{vu}) \psi_k' \\
2\Psi_r (\Sigma_{vu}^{**} \otimes \Sigma_{vu}^{**}) \psi_k'
\end{pmatrix},
\]

\[
E_{(2)} = c
\begin{pmatrix}
2\Psi_k (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' \\
2\Psi_k (\Sigma_{uu}^{**} \otimes \Sigma_{uu}^{**}) \psi_k' \\
2\Psi_k (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_k' \\
2\Psi_k (\Sigma_{uu}^{**} \otimes \Sigma_{uu}^{**}) \psi_k'
\end{pmatrix}.
\]

Hence, by (3.96), (3.105), (3.106), (3.110), (3.111), (3.112), and the independence of \( \Sigma_{ee} \) and \( Z_t \), the result follows. \( \square \)

Note that the existence of the fourth moment of \( X_t \) is necessary only for the asymptotic normality of \( \hat{\Sigma}_{XX} \). If Assumption 3.4 holds, i.e., if \( X_t \) is normal, then
\[
V_{XX} = 2\psi_k [(\Sigma_{XX} \circ \Sigma_{XX}) - (\Sigma_{uu} \circ \Sigma_{uu})] + c(\Sigma_{uu} \circ \Sigma_{uu}) \\
- c(\Sigma_{uu} \circ \Sigma_{uu})' \psi_k',
\] (3.113)

where

\[
\Sigma_{XX} = \Sigma_{XX} + \Sigma_{uu}
\]

Except for the part associated with \( V_{XX} \), the limiting covariance matrix of the estimators is a function of the unknown parameters. Thus, a consistent estimator of the limiting covariance matrix can be obtained by replacing \( \hat{a} \), \( \Sigma_{ee} \), \( \Sigma_{XX} \), \( u_X \), and \( c \) with \( \hat{a} \), \( \hat{\Sigma}_{ee} \), \( \hat{\Sigma}_{XX} \), \( \hat{u}_X \), and \( d^{-1}n \), respectively. A consistent estimator of \( V_{XX} \) can be obtained in the same way if, in addition, the \( X_t \) are assumed to be normally distributed. In the next theorem, we choose slightly different degrees of freedom for the estimated covariance matrices by analogy to usual regression analysis.

**Theorem 3.6.** Let the model (3.1) hold, and let Assumptions 3.5 through 3.8 hold.

(a) If either Assumption 3.1 or Assumption 3.2 holds, then the covariance matrix of the limiting distribution of

\[
(\hat{\Sigma}_O, [\text{vec } \hat{\Sigma}_O]', [\text{vech } \hat{\Sigma}_{ee}]')'
\]

is consistently estimated by

\[
\frac{1}{n} \hat{V}_{(a)}
\]

where
\[ \hat{V}(a) = \begin{pmatrix}
\hat{V}_{\omega \omega} & \hat{V}_{\omega \beta} & \hat{V}_{\omega \epsilon} \\
\hat{V}_{\omega \beta} & \hat{V}_{\beta \beta} & \hat{V}_{\beta \epsilon} \\
\hat{V}_{\omega \epsilon} & \hat{V}_{\beta \epsilon} & \hat{V}_{\epsilon \epsilon}
\end{pmatrix}, \]

\[ \hat{V}_{\omega \omega} = n^{-1} \hat{\Sigma}_{vv} + (I_{rxr} \otimes \bar{x}) \hat{V}_{\beta \beta} (I_{rxr} \otimes \bar{x}), \]

\[ \hat{V}_{\omega \beta} = - (I_{rxr} \otimes \bar{x}) \hat{V}_{\beta \beta}, \]

\[ \hat{V}_{\omega \epsilon} = - (I_{rxr} \otimes \bar{x}) \hat{V}_{\beta \epsilon}, \]

\[ \hat{V}_{\beta \beta} = (n-1)^{-1} \hat{\Sigma}_{vv} \otimes \hat{\Sigma}_{xx}^{-1} + [(n-1)^{-1} + d^{-1}] \hat{\Sigma}_{vv} \otimes (\hat{\Sigma}_{xx}^{-1} \hat{v}^p \hat{v}^{-1} \hat{v}_{xx}), \]

\[ \hat{V}_{\beta \epsilon} = - 2 d^{-1} \{ \hat{\Sigma}_{ve} \otimes [\hat{\Sigma}_{xx}^{-1} \hat{v}^p \hat{v}_{pp}(\bar{v}, \bar{v})] \} \psi'_p, \]

\[ \hat{V}_{\epsilon \epsilon} = 2 \psi'_p \{ d^{-1} (\hat{\Sigma}_{ve} \otimes \hat{\Sigma}_{ee}) - [d^{-1} - (n-1+d)^{-1}] (\hat{\Sigma}_{ve} \otimes \hat{\Sigma}_{vv} \hat{\Sigma}_{ve}) \}
\]

\[ \otimes (\hat{\Sigma}_{ve} \otimes \hat{\Sigma}_{vv} \hat{\Sigma}_{ve}) \} \psi'_p, \]

\[ \hat{\Sigma}_{pp} = \hat{\Sigma}_{uu} - \hat{\Sigma}_{uv} \hat{\Sigma}_{vv} \]

\[ \hat{\Sigma}_{vv} = (n-k-1+d)^{-1} (n-1+d) (I,-\hat{\phi}') \hat{\Sigma}_{ee} (I,-\hat{\phi}')', \]
\[ \hat{E}_{VE} = \hat{E}_{EV} = (\hat{E}_{ve}, \hat{E}_{vu}) \]

\[ = (\hat{I}, -\hat{g}') \hat{E}_{EE}. \]

(b) If Assumption 3.2 and Assumption 3.4 hold, then the covariance matrix of the limiting of \((\hat{E}_O, \text{vec } \hat{g}'), [\text{vech } \hat{E}_{EE}]', [\text{vech } \hat{E}_{xx}]'\) is consistently estimated by \(n \hat{\nabla}(b)\), where

\[ \hat{\nabla}(b) = \begin{bmatrix} \hat{\nabla}(a) & \hat{\xi} \\ \hat{\xi}' & \hat{\nabla}_{xx} \end{bmatrix}, \]

\[ \hat{\xi} = (\hat{\xi}_{ox}, \hat{\xi}_{ox}, \hat{\xi}_{ex})', \]

\[ \hat{\xi}_{ox} = - (\hat{X}_{rx} \equiv \hat{X}) \hat{\xi}_{ox}, \]

\[ \hat{\xi}_{ox} = 2(n-1)^{-1} \hat{Z}_{vu} \equiv \hat{Z}_{xx} + [(n-1)^{-1} + d^{-1}] \hat{Z}_{vu} \equiv (\hat{Z}_{xx} \hat{Z}_{pp}) \hat{\psi}', \]

\[ \hat{\nabla}_{ex} = 2 d^{-1} \hat{\psi} [(\hat{E}_{ex} \equiv \hat{E}_{ex}) - (\hat{E}_{ex} \equiv \hat{E}_{ex})] \hat{\psi}', \]

\[ \hat{\nabla}_{xx} = 2 \psi k (n-1)^{-1} (\hat{Z}_{xx} + \hat{Z}_{uu}) \hat{Z}_{xx} + \hat{Z}_{uu} + d^{-1} (\hat{Z}_{uu} \equiv \hat{Z}_{uu}) \]

\[ - [(n-1)^{-1} + d^{-1}] (\hat{Z}_{uu} \equiv \hat{Z}_{uu}) \hat{\psi}', \]
\[ \hat{\Sigma}_{uu} = \hat{\Sigma}_{uv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vu}, \]
\[ \hat{\Sigma}_{ue} \hat{\Sigma}_{eu} = \hat{\Sigma}_{uv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve}. \]

**Proof.** The results follow from Theorem 3.4, Theorem 3.5, and (3.113).

\[
\]

**D. Comments**

In this section, we comment on the results in this chapter and the work of other authors. The maximum likelihood estimators for the functional model derived in Theorem 3.1 were first obtained by Anderson (1951), and were reproduced by Healy (1980). Our derivation, following that of Healy, clearly pointed out the nature of inconsistency in the maximum likelihood estimator of \( \hat{\Sigma}_{ee} \), and enabled us to obtain a simple adjustment to produce the consistent estimator \( \hat{\Sigma}_{ee}^\prime \).

The maximum likelihood estimation of the multivariate structural model considered in Theorem 3.2 and Theorem 3.3 has not been discussed in the literature.

In Theorem 3.4, we showed the strong consistency of the maximum likelihood estimators adjusted for degrees of freedom under the assumption that \( d^{-1}n \rightarrow c \). Under the same assumption on the rate of increase of \( d \), Healy (1980) obtained the almost sure limits of the maximum likelihood estimators without adjustment for degrees of freedom for the functional model with normal errors.
For the same model as that of Healy, Anderson (1951) stated that the limiting distribution of the eigenvectors of a certain matrix can be used to obtain the limiting distribution of $\hat{\beta}$ as $d$ increases holding $n$ fixed. But, he did not obtain an explicit form of the limiting distribution. In Theorem 3.5, we derived the limiting joint distribution of the maximum likelihood estimators adjusted for degrees of freedom of all the structural parameters. We showed in Theorem 3.5 and Theorem 3.6 that the limiting covariance matrix and its estimator have the same forms for a wide range of assumptions on the unobservable true values $\chi_\xi$. It can be shown that the limiting covariance matrix of the maximum likelihood estimators adjusted for degrees of freedom of $\hat{\beta}_0$, $\hat{\beta}$, and $\hat{\Sigma}_{XX}$ for the model with known $\Sigma_{\epsilon\epsilon}$ is given by setting $c = 0$ in the corresponding portions of the covariance matrices in Theorem 3.5. Gleser (1981) gave the covariance matrix of the estimators for $\Sigma_{\epsilon\epsilon} = I$ known.
IV. THE NONLINEAR MODEL WITH KNOWN ERROR COVARIANCE MATRIX

This chapter deals with estimation of the general nonlinear errors-in-variables model under the assumption that the error covariance matrix is known. We consider the functional model with a single relationship. Some results in this chapter are extensions of the work by Wolter and Fuller (1982b). Also, some of the techniques used in the proofs will be discussed thoroughly in Chapter V. Therefore, the presentation of the proofs in this chapter is not as detailed as those in other chapters.

A. Introduction

To define the model, let \( \{b_n\}_{n=1}^{\infty} \) and \( \{a_n\}_{n=1}^{\infty} \) be sequences of positive real numbers such that \( n = a_n b_n \) for \( n = 1, 2, \ldots, \infty \), \( a_n = o(n) \), and \( b_n = o(n) \). We assume the existence of a sequence of experiments indexed by \( n \). Let

\[
y_t^0 = f(x_t^0; \varepsilon^0),
\]

\[
Y_{nt} = y_t^0 + e_{nt},
\]

\[
X_{nt} = x_t^0 + u_{nt}, \quad t = 1, 2, \ldots, b_n,
\]

where \( x_t^0 \) are \( 1 \times q \) vectors of fixed constants belonging to a parameter space \( \Gamma \), a convex subset of \( q \)-dimensional Euclidean space,
$\mathbf{g}^0$ is a $k \times 1$ vector belonging to a parameter space $\mathbb{G}$, a convex compact subset of $k$-dimensional Euclidean space, $f(z; \theta)$ is continuous on $\mathbb{I} \subseteq \mathbb{G}$, $(Y_{nt}, X_{nt})$ are observed in the $n$-th experiment, and $\varepsilon_{nt} = (e_{nt}, y_{nt})$ denote errors of measurement. We assume that $\varepsilon_{nt}$ are independently and identically distributed with mean zero and covariance matrix $\Sigma_n$, and that $\Sigma_n$ is of order $a_n^{-1}$. One way of interpreting this assumption is to let $a_n$ denote the number of observations made at each point $(y^0_t, x^0_t)$, $t = 1, 2, \ldots, b_n$. Under this interpretation, the total number of observations is $b_n a_n = n$ and each of the vectors $(Y_{nt}, X_{nt})$ used in the analysis is the mean of $a_n$ observations. Another interpretation is that the asymptotic results obtained with $n$ tending to infinity hold if the ratios of measurement error variances to the sums of squares of $b_n$ true values are small. Thus, the asymptotic results can be used as approximations if either the error variances are small or the number of observations is large.

We assume $f(x; \mathbf{g})$ possesses continuous first and second derivatives with respect to both arguments on $\mathbb{I} \subseteq \mathbb{G}$. Let $\partial f(x; \mathbf{g})$ denote the $q$-dimensional row vector of partial derivatives of $f(x; \mathbf{g})$ with respect to the elements of $x$ evaluated at $(z; \theta)$; let $\partial f(x; \mathbf{g})$ denote the $k$-dimensional column vector of partial derivatives with respect to the elements of $\mathbf{g}$ evaluated at $(z; \theta)$; let $\partial^2 f(x; \mathbf{g})$ denote the $k \times q$ matrix of second partial derivatives with respect to the elements of $\mathbf{g}$ and $x$ evaluated at $(z; \theta)$; let $\partial^2 f(x; \mathbf{g})$ denote the $q \times q$ matrix of second partial derivatives with respect to the elements of $x$ evaluated at $(z; \theta)$; and let
$f_{\beta\beta}(z; \beta)$ be the $k \times k$ matrix of second partial derivatives with respect to the elements of $\beta$ evaluated at $(z; \beta)$. We also let

$$f_t^0 = f(x_t^0; \beta^0),$$

$$f_{xt}^0 = f_x(x_t^0; \beta^0),$$

$$f_{\beta t}^0 = f_\beta(x_t^0; \beta^0),$$

$$f_{xxt}^0 = f_{xx}(x_t^0; \beta^0),$$

$$f_{\beta xt}^0 = f_{\beta x}(x_t^0; \beta^0),$$

$$f_{\beta\beta t}^0 = f_{\beta\beta}(x_t^0; \beta^0).$$

We introduce assumptions associated with the model (4.1).

**Assumption 4.1.** The $x_t^0$ are fixed constants and are interior points of $\Gamma$, and $\Gamma$ is a convex subset of $q$-dimensional Euclidean space.

**Assumption 4.2.** The parameter vector $\beta^0$ is an interior point of $\Omega$, and $\Omega$ is a convex compact subset of $k$-dimensional Euclidean space.

**Assumption 4.3.** The random variables $\xi_{nt}$, $t = 1, 2, \ldots, b_n$, are independently distributed with mean zero and covariance matrix $\Sigma_n$, where $\Sigma_n$ are known positive definite matrices satisfying
for a nonsingular matrix $\Phi$.

**Assumption 4.4.** Let

$$m_{xx} = n^{-1} \sum_{t=1}^{h_n} \sigma_{vnt}^{-2} \xi_{vnt} \xi_{vnt}' ,$$

where

$$\sigma_{vnt}^2 = [1, -f_0^0] \Sigma_n [1, -f_0^0]' .$$

Then, for a positive definite matrix $\overline{m}_{xx}$,

$$\lim_{n \to \infty} \frac{m_{xx}}{\overline{m}_{xx}} = 1 .$$

**Assumption 4.5.** There is available a preliminary estimator $\overline{\xi}$ in $\mathcal{U}$ satisfying

$$\overline{\xi} - \xi^0 = o_p(\max[a_n^{-1}, n^{-1/2}]) .$$

**Assumption 4.6.** The partial derivatives through order two of $f(x; \xi)$ are continuous and bounded on $\mathcal{U} \subseteq \mathcal{U}$.

**Assumption 4.6a.** The partial derivatives through order three of $f(x; \xi)$ are continuous and bounded on $\mathcal{U} \subseteq \mathcal{U}$. 
Assumption 4.7. The $2 + \delta$ moments of $\frac{1}{2} a_n^{-1} \varepsilon_{nt}$ are bounded for some $\delta > 0$.

Assumption 4.7a. For some real $L$ and all $t$ and $n$,

$$E[|\varepsilon_{nt}|^L] < L a_n^{-2},$$

where the norm is the Euclidean norm.

Assumption 4.8. $a_n^{-1} = o(n^{-1/2})$.

Assumption 4.8a. $a_n^{-1} = o(n^{-1/3})$.

B. Results of Wolter and Fuller

In this section, we summarize the work of Wolter and Fuller (1982b). Assuming a preliminary estimator $\hat{\theta}$ satisfying Assumption 4.5 is available, they considered two estimators of $\theta^0$ for the model (4.1).

Under Assumptions 4.1, 4.2, and 4.3 and the assumption of the normality of $\varepsilon_{nt}$, the maximum likelihood estimators of $x_t^0$ and $\theta^0$ are those values of $x_t$ and $\theta$ contained in $F = \theta$ that minimize the sum of squares

$$\sum_{t=1}^{b_n} q(\theta, x_t; Y_{nt}, X_{nt}).$$
While an explicit expression for the maximum likelihood estimator of \( \theta^0 \) has not been obtained, Wolter and Fuller considered an iterative procedure leading to an estimator of \( \theta^0 \). Given a preliminary estimator \( \bar{\theta} \) of \( \theta^0 \), let \( \bar{X}_t \) be the value of \( X_t \) contained in \( \pi \) that minimizes \( q(\bar{\theta}, \bar{X}_t; Y_{nt}, \bar{X}_{nt}) \). The local approximation to the sum of squares (4.2) is

\[
\sum_{t=1}^{b_n} [\tilde{e}_{nt} - \Delta y_t, \tilde{u}_{nt} - \Delta x_t] S_n^{-1} [\tilde{e}_{nt} - \Delta y_t, \tilde{u}_{nt} - \Delta x_t]' ,
\]

(4.3)

where

\[
\tilde{e}_{nt} = Y_{nt} - f(\bar{X}_t; \bar{\theta}) ,
\]

\[
\tilde{u}_{nt} = X_{nt} - \bar{X}_t ,
\]

\[
\Delta y_t = f(x_t; \theta) - f(\bar{x}_t; \bar{\theta}) ,
\]

\[
\Delta x_t = x_t - \bar{x}_t .
\]
Retaining only linear terms in the Taylor expansion of \( f(x_t, \bar{g}) \), we obtain

\[
\Delta y_t = f'_\bar{g}(\bar{x}_t; \bar{g})(\Delta \bar{g}) + f'_x(\bar{x}_t; \bar{g})(\Delta x_t)' ,
\]

(4.4)

where

\[
\Delta \bar{g} = \bar{g} - \bar{\bar{g}} .
\]

Minimizing (4.3) with respect to \( \Delta y_t, \Delta x_t, \) and \( \Delta \bar{g} \) subject to (4.4), Wolter and Fuller obtained an improved estimator of \( \bar{g}^0 \) given by

\[
\hat{\bar{g}} = \bar{\bar{g}} + \hat{\Delta} \bar{g} ,
\]

(4.5)

where \( \hat{\Delta} \bar{g} \) satisfies

\[
\bar{\Sigma}_{xx}(\Delta \bar{g}) = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f'_\bar{g}(\bar{x}_t; \bar{g})[\tilde{e}_{nt} - \bar{y}_{nt} f'_x(\bar{x}_t; \bar{g})] ,
\]

\[
\bar{\Sigma}_{vnt}^2 = [1, - f'_x(\bar{x}_t; \bar{g})] \bar{\Sigma}_n [1, - f'_x(\bar{x}_t; \bar{g})]' ,
\]

\[
\bar{\Sigma}_{xx} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} f'_\bar{g}(\bar{x}_t; \bar{g}) f'_\bar{g}'(\bar{x}_t; \bar{g}) .
\]
Given the preliminary estimators \( \hat{\beta} \) and \( \bar{\beta}_t \), Wolter and Fuller introduced a modified estimator of \( \hat{\beta}^0 \) given by

\[
\tilde{\beta} = \hat{\beta} + \Delta \tilde{\beta} ,
\]

where \( \Delta \tilde{\beta} \) satisfies

\[
\bar{\Omega}_{xx}(\Delta \tilde{\beta}) = n^{-1} \sum_{t=1}^{b_n} \sigma^{-2}_{nt} \xi_{\beta}(\bar{\beta}_t; \hat{\beta}) \tilde{v}_{nt} ,
\]

\[
\tilde{v}_{nt} = \varepsilon_{nt} - \bar{v}_{nt} \xi^1_x(\bar{\beta}_t; \hat{\beta}) - \frac{1}{2} \text{tr}[\xi_{xx}(\bar{\beta}_t; \hat{\beta})(\bar{v}_{nt}^\prime \bar{v}_{nt} - \xi_{uun})] ,
\]

\[
\bar{\Sigma}_n = \begin{bmatrix}
\sigma_{een} & \xi_{eun} \\
\xi_{uen} & \xi_{uun}
\end{bmatrix} .
\]

Without proofs, we state the results and some extensions of Wolter and Fuller (1982b).

**Result 4.1.** Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.5, and 4.6 hold. Then,

\[
\bar{X}_t = X_t^0 + \bar{x}_{nt} + o_p(\max\{a^{-1}_n, n^{-1/2}\}) ,
\]

where

\[
\bar{x}_{nt} = X_{nt} \Sigma_n^{-1} [\xi_{xt}^0, \xi_1^\prime] \Lambda_{nt}^{-1} ,
\]
\[
A_{nt} = [\xi_{xt}, 1] \Sigma_n^{-1} [\xi_{xt}, 1]'.
\]

Also, there is a constant \( k_o \) such that for \( t = 1, 2, \ldots, b_n \),

\[
E[|\varepsilon_t - \xi_t^0|^2] < k_o a_n^{-1}.
\]

If, in addition, Assumption 4.7a holds, there is a constant \( K \) such that for \( t = 1, 2, \ldots, b_n \),

\[
E[|\varepsilon_t - \xi_t^0|^4] < K a_n^{-2}.
\]

Observe that the leading term in the error of estimation, denoted by \( \delta_{nt} \), is of the form of the error in the generalized least squares estimator with \([\xi_{xt}, 1]\) the "X-matrix" and \( \Sigma_n \) the covariance matrix of the error vector \( \varepsilon_{nt} \).

Result 4.2. Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.5, and 4.6 hold. Then,

\[
a_n^{-2} v_{nt} = a_n v_{nt} + a_n \Delta v_{nt} + O_p [\max[a_n^{-1}, n^{-1/2}]],
\]

where

\[
\Delta v_{nt} = 2 \delta_{nt} \xi_{xt} [\Sigma_{uun} \xi_{xt} - \Sigma_{uen}].
\]
Result 4.3. Let the model (4.1) hold, and let Assumptions 4.1 through 4.6 hold. Then,
\[ \tilde{\mu}_{\infty} = \mu_{\infty} + o_p \left( \max \{ a_n^{-1}, n^{-1/2} \} \right), \]
\[ \hat{\beta} - \beta^0 = \mu_{\infty}^{-1} (\delta_1 \hat{\beta}) + o_p (a_n^{-1}) \]
\[ = o_p \left( \max \{ a_n^{-1}, n^{-1/2} \} \right), \]
where
\[ \delta_1 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{st}^{0*} \nu^*_{nt} \]
\[ \nu^*_{nt} = \nu_{nt} - u_{nt} \xi_{xt}^{0*}. \]

If, in addition, Assumptions 4.7 and 4.8 hold, then
\[ n^{1/2} (\hat{\beta} - \beta^0) \xrightarrow{L} N \left( 0, \mu_{\infty}^{-1} \right). \]

Result 4.4. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,
\[ \tilde{\xi} - \xi^0 = \mu_{\infty}^{-1} (\delta_1 \hat{\xi}) + o_p \left( \max \{ a_n^{-3/2}, a_n^{-1/2} n^{-1/2} \} \right) \]
If, in addition, Assumption 4.8a holds, then

$$\frac{1}{\sqrt{n}} (\tilde{g} - g^0) \xrightarrow{L} N(0, \tilde{\sigma}_{xx}^{-1}) .$$

The procedure for obtaining $\hat{\gamma}$ can be iterated using $\tilde{g}$ as the preliminary estimator in the second round of calculation. Since the $\hat{g}$ was obtained through the linearization of the sum of squares (4.2), the maximum likelihood estimators of $x_t$ and $g^0$ can be obtained as the limits of the iteration by including a modification in the procedure to guarantee convergence. By Result 4.3, the $\hat{g}$ satisfies the same property as the $\tilde{g}$;

$$\hat{g} - g^0 = O_p(\max \{a_n, n^{-1/2}\}) .$$

Thus, Result 4.3 holds for the final estimator obtained by any finite number of iterations. The procedure for obtaining $\tilde{g}$ can also be iterated, and for any finite number of iterations the asymptotic properties of the final estimator are given in Result 4.4.

We call $\hat{g}$ the **pseudo-maximum likelihood estimator**, and $\tilde{g}$ the **modified maximum likelihood estimator**.

Recall that we assumed $a_n = o(n)$ and $b_n = o(n)$. If $b_n$ is a constant and $a_n = o(n)$, then $\hat{g}$ and $\tilde{g}$ are still consistent. But,
if \( a_n = o(n) \), the limiting distributions of \( \frac{1}{n^2} (\hat{\theta} - \theta^0) \) and \( \frac{1}{n} (\hat{\theta} - \theta^0) \) are not normal unless \( \varepsilon_{nt} \) are normally distributed.

Wolter and Fuller suggested certain modifications to improve the small sample properties of \( \tilde{\theta} \). Their Monte Carlo study showed that such modified estimators were superior to \( \tilde{\theta} \). They also showed that under some assumptions the ordinary least squares estimator \( \hat{\theta}_L \) of \( \theta^0 \) obtained by minimizing

\[
\min_{\theta} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \theta)]^2
\]

satisfies Assumption 4.5, and thus can be used as the preliminary estimator \( \bar{\theta} \).

C. Bias of the Pseudo-Maximum Likelihood Estimator

In this section, we investigate further the properties of \( \hat{\theta} \) and of an estimator of \( \theta^0 \) based on \( \hat{\theta} \). By Result 4.3, the error in \( \hat{\theta} \) as an estimator of \( \theta^0 \) is \( O_p(\max[a_n^{-1}, n^{-1/2}]) \). The term of \( O_p(a_n^{-1}) \) is due to the nonlinearity of the function \( f(z; \theta) \). That is, the nonzero second partial derivative \( f_{zz}(z; \theta) \) introduces a bias in the estimator \( \hat{\theta} \). We observe that \( \bar{X}_t \) was obtained by projecting the point \((Y_{nt}, X_{nt})\) onto a nonlinear surface \( f(z; \bar{\theta}) \) in the metric \( \varepsilon_n^{-1} \). Thus, \( \bar{X}_t \) is biased, even if \( \bar{\theta} \) is close to \( \theta^0 \). Also, \( \hat{\theta} \)
minimizes the sum of distances between \((y_{nt}, \bar{x}_{nt})\) and \((y_t, \bar{x}_t)\) by a linearization based on \([f(\bar{x}_t; \bar{\beta}), \bar{x}_t]\). Hence, the bias in \(\bar{x}_t\) leads to a bias in \(\hat{\beta}\). In this section, we obtain an exact expression for the portion of the bias in \(\hat{\beta}\) that is 0 \(\left(a_n^{-1}\right)\). In order to simplify our presentation in the following theorem, we use a stronger set of assumptions than that in Result 4.3.

**Theorem 4.1.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

\[
\hat{\beta} - \beta^0 = \frac{1}{n} \sum_{t=1}^{b_n} \sigma^{-2}_{vnt} f^0 v^t y^t = o_p(\max[a_n^{-3/2}, n^{-1/2} a_n^{-1/2}]),
\]

where

\[
\delta_1 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma^{-2}_{vnt} f^0 v^t y^t = o_p(n^{-1/2}),
\]

\[
\delta_2 \hat{\beta} = n^{-1} \sum_{t=1}^{b_n} \sigma^{-2}_{vnt} f^0 v^t y^t f^0 x^t x^t \sigma^{-1}_{nt} \hat{\delta}_{nt} - u^t_{nt} = o_p(\sigma^{-1}_n),
\]

\[
\delta_{nt} = \xi_{nt} \sigma^{-1}_{nt} \xi_{xt} \xi_{nt}^{-1} [\xi_{nt}^0, \xi_{nt}^0] \Lambda_{nt}^{-1},
\]

\[
\Lambda_{nt} = [\xi_{nt}^0, \xi_{nt}^0] \xi_{nt}^{-1} [\xi_{nt}^0, \xi_{nt}^0]',
\]
and \( f_0^t, x_t, f_0^0 \), and \( x_{xxt}^0 \) were defined following (4.1). Also,

\[
E[\delta_{1}\hat{g}] = 0 ,
\]

\[
E[\delta_{2}\hat{g}] = -\frac{1}{2} b_n^{-1} \sum_{t=1}^{b_n} (a_n c_n^2)^{-1} x_{g}^0 \text{ tr}(x_{xxt}^0 A_{nt}^{-1})
\]

\[
= O(a_n^{-1}) .
\]

**Proof.** By Result 4.3,

\[
\bar{N}_{xx} = N_{xx} + O_p(\text{max}[a_n^{-1}, n^{-1/2}]) .
\] (4.7)

where \( \bar{N}_{xx} \) is defined following (4.5). By Result 4.1, for

\( t = 1, 2, \ldots, b_n \),

\[
E[|\bar{z}_t - z_{t}^0|] < K a_n^{-2} ,
\]

and it follows that for a constant \( K_1 \) and for \( t = 1, 2, \ldots, b_n \),

\[
E[|\bar{z}_t - z_{t}^0|] < K_1 a_n^{-1/2} ,
\] (4.8)

\[
E[|\bar{z}_t - z_{t}^0|^2] < K_1 a_n^{-1} ,
\] (4.9)
Also, it can be shown that

\[
\begin{align*}
\frac{b_n}{n} \sum_{t=1}^{b_n} \left[ (a_n \sigma_{vnt}^2)^{-1} - (a_n \sigma_{vnt}^2)^{-1} + (a_n \sigma_{vnt}^2)^{-2} (a_n \Delta_{vnt}) \right]^2 &= o_p(a_n^{-2}) .
\end{align*}
\]

Using Assumptions 4.4 and 4.6a, (4.8), (4.9), (4.10), and (4.11), we obtain

\[
\frac{b_n}{n} \sum_{t=1}^{b_n} \left[ (\sigma_{vnt}^2)^{-1} \epsilon_{nt} (\tilde{x}_t; \tilde{\theta}) [\tilde{e}_{nt} - \tilde{u}_{nt}^x \tilde{x}_t; \tilde{\theta}] \right] = \frac{b_n}{n} \sum_{t=1}^{b_n} \left[ (a_n \sigma_{vnt}^2)^{-1} - (a_n \sigma_{vnt}^2)^{-2} a_n \Delta_{vnt} \right] \\
\left[ \frac{\tilde{\theta}^0 + \frac{1}{2} \tilde{\theta}^t - \tilde{\gamma}_{nt}}{v_{nt} - \tilde{\gamma}_{nt} \tilde{\gamma}_{nt} - \tilde{\gamma}_{nt}} \right] + o_p(\max\{a_n^{-3/2}, n^{-1/2} a_n^{-1/2}\}) .
\]

Therefore, by (4.7) and (4.12),
\[
\lim_{n \to \infty} (\hat{\beta} - \beta^0) = \delta_1 \hat{\beta} + \delta_2 \hat{\beta} + b_n \sum_{t=1}^{b_n} (a_n \sigma_n^2)^{-1} [\varepsilon_0 \hat{\delta}^t - \sigma_n^2 \varepsilon_0 \Delta \varepsilon_0]^{nt} \\
+ \mathcal{O}_p(\max[\frac{a_n^{-3/2}}{n}, \frac{a_n^{-1/2}}{a_n}]) \quad (4.13)
\]

We observe that

\[
E[\hat{\varepsilon}_n = 0] = \mathcal{A}^{-1} \mathcal{E}_n \mathcal{E}^{-1}_n [1, -\varepsilon_0] = 0 \quad (4.14)
\]

Also, by Assumption 4.7a,

\[
E[\delta_n, \delta^2] = 0 (a_n^{-2}) \quad (4.15)
\]

Since \( \Delta \varepsilon_0 \) is a linear combination of \( \delta_n \), it follows from (4.14) and (4.15) that the third term in the right hand side of (4.13) is \( \mathcal{O}_p(\frac{n^{-1/2}a_n^{-1/2}}{a_n}) \). Since \( \delta_1 \hat{\beta} \) is a linear combination of \( \varepsilon_0 \), by Assumption 4.3 and 4.4,

\[
E[\delta_1 \hat{\beta}] = 0 \\
V[\delta_1 \hat{\beta}] = 0(b_n^{-1}a_n^{-1}) = 0(n^{-1})
\]
and thus,
\[ \delta_1 \hat{\theta} = O_p(n^{-1/2}) . \]

We observe that by Assumptions 4.3, 4.4, and 4.6a, for some constant \( L_1 \),
\[
E[|\delta_2 \hat{\theta}|] = L_1 b_n^{-1} \sum_{t=1}^{b_n} E[|\delta_{nt}|^4 + |\hat{\theta}_{nt}|^2 |y_{nt}|^2] = O(a_n^{-1}) .
\]
Hence,
\[ \delta_2 \hat{\theta} = O_p(a_n^{-1}) . \]
Finally,
\[
E[\text{tr}\{\hat{\theta}_{nt}^t \hat{\theta}_{nt}\}] = A_{nt}^{-1} ,
\]
\[
E[\text{tr}\{y_{nt}^t \hat{\theta}_{nt}\}] = A_{nt}^{-1} = O(a_n^{-1}) .
\]
Therefore, the results follow. □

Hence, taking the expectation through terms of \( O_p(a_n^{-1}) \), the bias
in \( \hat{\theta} \) as an estimator of \( \theta^0 \) is

\[
- \frac{1}{2} \mathbf{m}^{-1} \mathbf{b}^{-1} \sum_{t=1}^{n} (a \sigma^2_{\text{vnt}})^{-1} \mathbf{f}_t^0 \text{tr}[\mathbf{f}_t^{0\prime} \mathbf{A}_{nt}^{-1}]
\]

which is of \( O(a_n^{-1}) \). The bias (4.16) is due to the nonlinearity of the function \( f(x; \theta) \), and vanishes if the model is linear in \( x_t \), i.e., if \( f_{xxt}^0 = 0 \) for all \( t \). Note that by Theorem 2.1

\[
\mathbf{A}_{nt}^{-1} = \mathbf{\Sigma}_{uun} - \mathbf{\Sigma}_{uvnt} \sigma_{vnt}^2 \mathbf{\Sigma}_{vunt}
\]

where

\[
\mathbf{\Sigma}_{uvnt} = \mathbf{\Sigma}_{vunt}'
\]

\[
= \mathbf{\Sigma}_{uen} - \mathbf{\Sigma}_{uun} f_{xt}^0
\]

We observe that the bias (4.16) is a weighted average of

\[
- \frac{1}{2} \text{tr}[\mathbf{f}_t^{0\prime} \mathbf{A}_{nt}^{-1}]
\]

where the weights are

\[
\mathbf{m}^{-1} \mathbf{b}^{-1} (a_n \sigma_{\text{vnt}}^2)^{-1} \mathbf{f}_t^0
\]

for \( t = 1, 2, \ldots, b_n \).
Thus, the quantity in (4.17) can be considered as a contribution of each true value to the bias of ~. Such a contribution is small for an ~ at which either f' or A' is small. Note that f' is the nonlinearity at the point ~, and that A' is the approximate covariance matrix of ~ as an estimator of ~.

We consider a simple quadratic model as an example. Let

\[ y_0^* = \beta_0 + \beta_1 x_0^* + \beta_2 (x_0^*)^2, \]

\[ \Sigma_n = \sigma^2 I_{2\times 2}, \]

where

\[ \sigma^2 = o(a_n^{-1}). \]

Then, the quantity in (4.17) becomes

\[ -\sigma^2 \beta_2 [1 + (\beta_1 + 2 \beta_2 x_0^*)^2]^{-1} = -\sigma^2 \beta_2 + \sigma^2 \beta_2 [1 + (\beta_1 + 2 \beta_2 x_0^*)^2]^{-1} \]

\[ (\beta_1 + 2 \beta_2 x_0^*)^2. \]

Observing that

\[ \sum_{-1}^{b_{-1}} \sum_{-1}^{b_{-1}} \sum_{-1}^{b_{-1}} (a_n \sigma_2^2)^{-1}(1, x_t^0, x_t^0)^\prime = (1, 0, 0)^\prime, \]
the bias expression in (4.16) becomes

\[ (-\beta_2 \sigma^2, 0, 0)' + (\beta_2 \sigma^2) M_{xx}^{-1} \sum_{t=1}^{b_n} \left[ 1 + (\beta_1 + 2\beta_2 x_t^0)^2 \right]^{-2} (\beta_1 + 2\beta_2 x_t^0)^2 [1, x_t^0, (x_t^0)^2]' \]

where

\[ M_{xx} = \sum_{t=1}^{b_n} \left[ 1 + (\beta_1 + 2\beta_2 x_t^0)^2 \right]^{-1} [1, x_t^0, (x_t^0)^2]' [1, x_t^0, (x_t^0)^2] . \]

We observe that at the point \( x_t^0 = -2^{1-1} \beta_2^{-1} \beta_1 \) the bias contribution is concentrated in the intercept term and is given by \( -\beta_2 \sigma^2 \). Also, as \( x_t^0 \) moves away from \( -2^{1-1} \beta_2^{-1} \beta_1 \), the bias contribution decreases for \( \beta_0 \) and increases for \( \beta_1 \) and \( \epsilon_2 \).

By Theorem 4.1, the bias in \( \hat{\beta}_2 \) due to the curvature is given by

\[ \hat{\beta}_2 \hat{\epsilon}_2 \left( \beta_{xx}^{-1} \hat{\beta}_2 \right) , \]

which is a function of

\[ \delta_{nt} \mathbf{f}_0^0 \mathbf{x}_{nt} \mathbf{x}_{nt} \left( \frac{1}{2} \delta_{nt} - u_{nt} \right) = \frac{1}{2} (u_{nt} - \delta_{nt}) \mathbf{f}_0^0 \mathbf{x}_{nt} \mathbf{x}_{nt} (u_{nt} - \delta_{nt})' - \frac{1}{2} u_{nt} \mathbf{f}_0^0 \mathbf{x}_{nt} \mathbf{x}_{nt} . \]

Since the leading term for the error in \( \hat{x}_t \) is \( \delta_{nt} \),

\[ \hat{u}_{nt} = \hat{x}_{nt} - \hat{x}_t \]
estimates \( u_{nt} - \delta_{nt} \). Also,

\[
E[u_{nt}^0 u_{nt}'_t] = \text{tr}[g^0 \Sigma_{xxt-uu}] .
\]

Hence, the modification in \( \tilde{\xi} \) suggested by Wolter and Fuller (1982b) is the subtraction of an estimate of the bias contribution from each

\[
[\tilde{\xi}_{nt} - \bar{\xi}_{nt} f_{x_t} (\bar{\xi}_t; \bar{\xi})] .
\]

Using \( \bar{\xi} \) we can obtain an estimator of \( \xi_t^0 \) which is an improvement of the estimator \( \tilde{\xi}_t \). Let \( \hat{\xi}_t \) be the value of \( \xi_t \) contained in \( \bar{\xi} \) that minimizes \( q(\bar{\xi}_t; Y_{nt}, \xi_{nt}) \). Since

\[
\hat{\xi} - \xi^0 = O_p(\max[a^{-1}_n, n^{-1/2}]) ,
\]

Result 4.1 also holds for \( \hat{\xi}_t \). Now, we investigate the bias in \( \hat{\xi}_t \) as an estimator of \( \xi_t^0 \). The effect of the bias in \( \hat{\xi} \) on \( \hat{\xi}_t \) is given in the next theorem.

**Theorem 4.2.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

\[
\hat{\xi}_t - \xi^0 = \tilde{\delta}_{nt} + [\delta_1 \hat{\xi}_t + \delta_2 \hat{\xi}_t + \delta_3 \hat{\xi}_t](a_{nt}^{-1}) + O_p(\max[a^{-3/2}_n, a^{-1/2}_n n^{-1/2}]) ,
\]
where

\[ \hat{\delta}_{nt} = \varepsilon_{nt} \Sigma_n^{-1} [\mathbf{x}_t', \mathbf{I}]' \Delta_{nt}^{-1} \]

\[ = o_p(a_n^{-1}) , \]

\[ \delta_1 \hat{\varepsilon}_t = - [\mathbf{x}_t' \Sigma_{xx}^{-1} (\delta_1 \hat{\hat{\varepsilon}}), \mathbf{Q}] (a_n \Sigma_n)^{-1} [\mathbf{x}_t', \mathbf{I}]' \]

\[ = o_p(a_n^{-1}) , \]

\[ \delta_2 \hat{\varepsilon}_t = \varepsilon_{nt}^* (a_n \Sigma_n)^{-1} (\delta_1 \hat{\hat{\varepsilon}}), \mathbf{Q}] [\mathbf{x}_t'] \]

\[ = o_p(a_n^{-1}) , \]

\[ \delta_3 \hat{\varepsilon}_t = - \left[ \frac{1}{2} \varepsilon_{nt} \mathbf{x}_t' \Sigma_{nt} \delta_1 \hat{\varepsilon}_t + \mathbf{x}_t' \Sigma_{xx}^{-1} (\delta_2 \hat{\hat{\varepsilon}}), \mathbf{Q}] (a_n \Sigma_n)^{-1} [\mathbf{x}_t', \mathbf{I}]' \right] \]

\[ = o_p(a_n^{-1}) , \]

\[ \varepsilon_{nt}^* = \varepsilon_{nt} - \delta_{nt} (\mathbf{x}_t', \mathbf{I}) . \]

Also,

\[ E\{\delta_{nt}\} = E\{\delta_1 \hat{\varepsilon}_t\} = E\{\delta_2 \hat{\varepsilon}_t\} = 0 , \]
\[ E(\hat{\delta}_{nt}) = \frac{1}{2} \sum_{s=1}^{b_n} \sigma_{ns}^2 \sigma_{nt}^{-1} \sum_{s=1}^{b_n} \sigma_{ns}^2 \sigma_{nt}^{-1} \]

\[ - \frac{1}{2} \sum_{s=1}^{b_n} \sigma_{ns}^2 \sigma_{nt}^{-1} \sum_{s=1}^{b_n} \sigma_{ns}^2 \sigma_{nt}^{-1} \]

\[ = 0(a_n^{-1}) \]

where

\[ \Sigma_n^{-1} = \begin{bmatrix}
\sigma_{ee} & \sigma_{eu} \\
\sigma_{ne} & \Sigma_n
\end{bmatrix} \]

Proof. Since \( x_t^0 \) is an interior point of \( F \), and since \( \hat{x}_t \) is consistent for \( x_t^0 \) by Result 4.1, \( \hat{x}_t \) satisfies

\[ [y_{nt} = f(x_t^0; \hat{\theta}), x_{nt} - \hat{x}_t] (a_n \Sigma_n^{-1}) [x'_{nt}(x_t^0; \hat{\theta}), 1]' = o_p(\max[a_n^{-3/2}, n^{-1}]) . \]  

(4.18)

By Result 4.1,

\[ \hat{x}_t - x_t^0 = \delta_{nt} + o_p(\max[a_n^{-1}, n^{-1/2}]) . \]  

(4.19)

Using (4.19) and Theorem 4.1, we obtain that
\[
Y_{nt} - f(\hat{x}_t; \hat{g}) = e_{nt} - f^0_{xt}(\hat{x}_t - x^0_t)' - \frac{1}{2} \delta_{nt} \xi_{nt} \xi_{nt}'
- \xi^0_{bt} \xi_{xx}^{-1}(\delta_1 \hat{g} + \delta_2 \hat{g}) + o_p(\max[\alpha_n^{-3/2}, \alpha_n^{-1/2} - \frac{1}{2}])
\]
\[
= o_p(\alpha_n^{-1/2}) ,
\]
and that
\[
f_{\times}(\hat{x}_t; \hat{g}) = f^0_{xt} + \delta_{nt} f^0_{nt} + o_p(\max[\alpha_n^{-1}, \alpha_n^{-1/2}]) .
\]

Substituting (4.20) and (4.21) into (4.18), we have
\[
[e_{nt} - f^0_{xt}(\hat{x}_t - x^0_t)' - \frac{1}{2} \delta_{nt} f^0_{nt} \delta_{nt}' - \xi^0_{bt} \xi_{xx}^{-1}(\delta_1 \hat{g} + \delta_2 \hat{g}), u_{nt}(\hat{x}_t - x^0_t)]
\]
\[
(a_n \xi_n)^{-1}[f^0_{xt} + f^0_{nt} \delta_{nt}' - \xi^0_{bt} \xi_{xx}^{-1}(\delta_1 \hat{g} + \delta_2 \hat{g}), 0]
\]
\[
= o_p(\max[\alpha_n^{-3/2}, \alpha_n^{-1/2} - \frac{1}{2}]) .
\]

Hence, the expansion of \((\hat{x}_t - x^0_t)\) follows. We note that
\[
E[\delta_{nt}] = 0 ,
\]
\[
E[\delta_i \hat{g}] = 0 .
\]

Also,
\[
E[\delta_{nt} \xi_{nt}] = A_{nt}^{-1}(f^0_{xt}', I)[I - A_{nt}(f^0_{xt}', I)'A_{nt}(f^0_{xt}', I)]
\]
and thus, each element of $\delta_{nt}$ is uncorrelated with each element of $\varepsilon_{nt}^*$. Therefore, $E[\delta_{nt}^*] = 0$. Finally, using Theorem 4.1 and taking the expectation of $\delta_{nt}^*^\frac{1}{2}$, we obtain the results.

As in $\hat{g}$, there is a bias of $O(a_n^{-1})$ in $\hat{x}_t$ as an estimator of $x_t^0$. Also, the bias is again a function of the quantity in (4.17). But, in $\hat{x}_t$, the bias is due to the deviation of $\text{tr}[f_0^{-1}A_n^{-1}]$ from the weighted average

$$
\sum_{s=1}^{b_n} \sigma_n^{-2} f_0^s \sum_{s=1}^{b_n} \sigma_n^{-2} f_0^s \left[ f_0^s [f_0^{-1}]^{-1} f_0^s \right] \text{tr}[f_0^{-1}A_n^{-1}] .
$$

Hence, the bias in $\hat{x}_t$ is small for $x_t^0$ such that $\text{tr}[f_0^{-1}A_n^{-1}]$ is close to the weighted average (4.22). The term (4.22) is the contribution of the bias $\delta_{nt}^*\hat{g}$ in $\hat{g}$. Thus, the bias of $\hat{g}$ actually helps to reduce the bias of $\hat{x}_t$ for the values of $t$ for which $\text{tr}[f_0^{-1}A_n^{-1}]$ is close to (4.22). To investigate this point in more detail, we consider the maximum likelihood estimator $x_t^*$ under the assumption that $g_0^0$ is known. The estimator $x_t^*$ is the value of $x_t$ contained in $r$ that minimizes $q(g_0^0, x_t; Y_{nt}, X_{nt})$. The asymptotic properties of $x_t^*$ are given in the following theorem.

**Theorem 4.3.** Let the model (4.1) hold, and let Assumptions 4.1, 4.2, 4.3, 4.6a, and 4.7a hold. Then,
\begin{equation}
\hat{\delta}_{2t}^* - \hat{\delta}_{t}^0 = \delta_{nt}^* + [\delta_{nt}^* + \delta \hat{\delta}_{nt}^*] (a_n^{-1}) + O_p (\max [a_n^{-3/2}, a_n^{-1/2} - 1/2]) ,
\end{equation}

where \( \hat{\delta}_{2t}^* \) is given in Theorem 4.2, and

\begin{equation}
\delta x^*_t = -\frac{1}{2} \left[ \delta_{nt} x^0 \delta^t, O \right] (a_n^{-1})^{-1} [x^0, \xi] \right]
\end{equation}

\begin{equation}
= O_p (a_n^{-1}) .
\end{equation}

Also,

\begin{equation}
E[\delta x^*_t] = -\frac{1}{2} \text{tr}[x^0 A^{-1} a_n^{-1} [e e^0 + \Sigma e^0]]
\end{equation}

\begin{equation}
= O(a_n^{-1}) .
\end{equation}

**Proof.** It is easy to show that

\begin{equation}
\hat{\delta}_{t}^* - \hat{\delta}_{t}^0 = \delta_{nt}^* + O_p (a_n^{-1})
\end{equation}

\begin{equation}
= O_p (a_n^{-1/2}) .
\end{equation}

Then, \( \hat{\delta}_{t}^* \) satisfies

\begin{equation}
[y_{nt} - f(\hat{\delta}_{t}^*; \xi), \hat{\delta}_{nt} - \hat{\delta}_{t}^*] (a_n^{-1})^{-1} [f^0(\hat{\delta}_{t}^*; \xi^0), \xi] = O_p (a_n^{-3/2}) .
\end{equation}

(4.23)
Also,
\[ Y_{nt} - f(x^*; \hat{g}^0) = e_{nt} - f^0 (x^* - x^0)' - \frac{1}{2} \delta_{nt} f^0 \delta_{nt}' + O_p(a_n^{-3/2}) , \]
(4.24)

and
\[ f^*(x^*; \hat{g}^0) = f^0 + \delta_{nt} f^0_{xx} + O_p(a_n^{-1}) . \]
(4.25)

Substituting (4.24) and (4.25) into (4.23), we obtain the results.

Note that \( x^* \) was obtained by projecting \((Y_{nt}, x_{nt})\) onto the known nonlinear surface \( y = f(x; \hat{g}^0) \). The bias given in Theorem 4.3 is due to the nonlinearity of the function \( f(x; \hat{g}^0) \). Comparing the results in Theorem 4.2 and Theorem 4.3, we obtain an interesting result. The estimator \( \hat{x}^* \), the projection of \((Y_{nt}, x_{nt})\) onto the estimated surface \( y = f(x; \hat{g}) \), has a smaller bias than the estimator \( x^* \), the projection of \((Y_{nt}, x_{nt})\) onto the true surface \( y = f(x; \hat{g}^0) \), provided that \( \text{tr}[f^0 L_{xx}^{-1}] \) has the same sign as the weighted average (4.22). That is, the bias in \( \hat{g} \) due to the curvature of \( f(x; \hat{g}) \) partially cancels the bias due to the projection onto a nonlinear surface.

The iterative procedure for successively obtaining \( \hat{g} \) and \( \hat{x}_{nt} \) leads to the maximum likelihood estimator. The asymptotic properties of estimators of \( \hat{g}^0 \) and \( \hat{x}_{nt}^0 \) obtained by any finite number of iterations are given by Theorem 4.1 and Theorem 4.2. We conjecture that the same
asymptotic properties apply to the maximum likelihood estimator.

D. Bias of the Modified Maximum Likelihood Estimator

In this section, we investigate biases of the modified maximum likelihood estimator \( \tilde{\theta} \) and an estimator of \( \chi_t^0 \) based on \( \tilde{\theta} \), and compare the results to those in the previous section.

The following theorem is a restatement of Result 4.4.

**Theorem 4.4.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Let \( \tilde{\theta} \) be defined by (4.6). Then,

\[
\tilde{\theta} - \theta^0 = \mathbb{I}_{\text{XX}}^{-1}(\delta_1 \hat{\theta}) + O_p(\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}])
\]

where

\[
\delta_1 \hat{\theta} = n^{-1} \sum_{t=1}^{b_n} \sigma^{-2}_t v_t^* \sigma^{1/2}_t \eta_t^{1/2} = O_p(n^{-1/2})
\]

\[
v_t^* = e_t - x_t u_t^*.
\]

\[
\mathbb{E}[^{\delta_1 \hat{\theta}}] = 0.
\]

**Proof.** We observe that

\[
\tilde{v}_{nt} = e_{nt} + \xi_t^0 - f(X_{nt}; \tilde{\theta}) + \frac{1}{2} \text{tr}[\xi_t^0 (\tilde{L}_{u u})] + O_p(a_n^{-3/2})
\]
= f'(\tilde{\beta}; g_0 - \bar{g}) + \frac{1}{2} \text{tr}[f^0(x_t' u_t - \Sigma))]

+ o_p(\max[a_n^{-3/2}, a_n^{-1/2} - 1/2]) .

By Assumption 4.7a,

$$E[|\text{vech}(u_t' u_t - \Sigma)|^2] = O(a_n^{-2}) .$$

By (4.26) and Assumption 4.6a,

$$b_n \sum_{t=1}^{b_n} \text{tr}[f^0(x_t' u_t - \Sigma)] = O(a_n^{-1/2}) .$$

The results follow by (4.27) and the argument in the proof of Theorem 4.1.

Thus, the estimator \( \tilde{\beta} \) does not have the bias that is \( O(a_n^{-1}) \) and is due to the curvature of \( f(x; g) \). Therefore, \( \tilde{\beta} \) is superior to \( \hat{\beta} \) not only because the limiting distribution of \( n^{1/2}(\tilde{\beta} - g_0) \) can be obtained under weaker assumptions than those for \( n^{1/2}(\hat{\beta} - g_0) \), but also because \( \tilde{\beta} \) has bias of smaller order.

We define the estimator \( \tilde{x}_t \) of \( x_0 \) to be the value of \( x_t \) contained in \( \mathcal{F} \) that minimizes \( q(\tilde{\beta}, \tilde{x}_t; y_{nt}, X_{nt}) \). The next theorem presents the bias of \( \tilde{x}_t \).
Theorem 4.5. Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, and 4.7a hold. Then,

\[ \tilde{x}_t - x_t^0 = \delta_{nt} + \delta_{nt} + \delta'_{nt} (a_n^{-1}) + O_p(\max(a_n^{-3/2}, a_n^{-1/2} - 1/2}) \]

where \( \delta_{nt} \) and \( \delta'_{nt} \) are given in Theorem 4.2, and \( \delta'_{nt} \) is given in Theorem 4.3.

Proof. Since

\[ \tilde{\xi} - \xi^0 = O_p(\max(\max(a_n^{-3/2}, a_n^{-1/2})) \]

Result 4.1 holds for \( \tilde{x}_t \). Thus,

\[ \tilde{x}_t - x_t^0 = \delta_{nt} + O_p(\max(a_n^{-1}, a_n^{-1/2})) \]

\[ Y_{nt} - f(\tilde{x}_t; \tilde{\xi}) = e_{nt} - f_{xt}' (\tilde{x}_t - x_t^0)' - \frac{1}{2} \delta_{nt} f_{nt}^0 \delta'_{nt} \]

\[ - f_{nt}' m_{xx} (\delta_{nt} + \delta'_{nt}) + O_p(\max(a_n^{-3/2}, a_n^{-1/2} - 1/2}) \]

\[ \xi_{x}(\tilde{x}_t; \tilde{\xi}) = \xi_{xt} + \delta_{nt} \xi_{nt} + O_p(\max(a_n^{-1}, a_n^{-1/2})) \]
Hence, the results follow.

Taking the expectation through terms of $O_p(a^{-1})$, the bias of $\tilde{x}_t$ is the same as that of $\tilde{x}_t^*$, the estimator assuming known $g^0$. Since $\tilde{g}$ does not possess the bias due to the curvature, the cancellation of the nonlinearity biases in $\hat{x}_t$ does not occur for $\tilde{x}_t$. Therefore, for certain values of $t$ the estimator $\tilde{x}_t$ based on $\tilde{g}$ has a larger bias than the estimator $\hat{x}_t$ based on $\hat{g}$.

E. A Class of Estimators Adjusted for Bias

In this section, we introduce a class of estimators $(\tilde{g}_A, \tilde{x}_{tA})$ of $(\tilde{g}^0, \tilde{x}^0)$ for the model (4.1). For each member of the class, $\tilde{g}_A$ has the same asymptotic properties as the modified maximum likelihood estimator $\tilde{g}$. That is, each $\tilde{g}_A$ in the class has no bias due to the curvature, and has a limiting distribution under Assumption 4.8a. Also, we can choose certain members of the class for which the $\tilde{x}_{tA}$ have smaller biases than the estimators $\hat{x}_t$, $\tilde{x}_t$, and $\tilde{x}_t^*$.

To define the class of estimators, we assume that there are available preliminary estimators $\hat{g}$ satisfying Assumption 4.5 and $\hat{x}_t$ satisfying the conclusions of Result 4.1. For example, the maximum likelihood estimators of $g^0$ and $x^0_t$, i.e., the estimators $\hat{g}$ and $\hat{x}_t$ iterated to convergence, can be used as preliminary estimators. Let $dY_{nt}$ and $dX_{nt}$, $t = 1, 2, \ldots, b_n$, be functions of $\hat{g}$ and $\hat{x}_t$, where $dY_{nt}$ is a scalar and $dX_{nt}$ is a $1 \times q$ vector. The quantities $dY_{nt}$ and $dX_{nt}$ are modifications to observations $Y_{nt}$ and $X_{nt}$ and are introduced to reduce the bias due to the nonlinearity. Given $dY_{nt}$
and $dX_{nt}$, we define the adjusted estimators $\tilde{\theta}_A$ and $\tilde{\xi}_{tA}$ to be the maximum likelihood estimators of $\theta^0$ and $\xi_t^0$ based on the observations $(Y_{nt}^*, X_{nt}^*)$, where

$$Y_{nt}^* = Y_{nt} + dY_{nt},$$

$$X_{nt}^* = X_{nt} + dX_{nt}.$$  \hspace{1cm} (4.28)

That is, $\tilde{\theta}_A$ and $\tilde{\xi}_{tA}$ are the estimators $\hat{\theta}$ and $\hat{\xi}_t$ iterated to convergence using $(Y_{nt}^*, X_{nt}^*)$ as observations. Depending on the choice of $dY_{nt}$ and $dX_{nt}$, we have different adjusted estimators $\tilde{\theta}_A$ and $\tilde{\xi}_{tA}$. Of course, the adjustment term $(dY_{nt}, dX_{nt})$ has to satisfy some condition to reduce the biases of the resulting adjusted estimators $\tilde{\theta}_A$ and $\tilde{\xi}_{tA}$. The following Assumption 4.9 provides a condition for $(dY_{nt}, dX_{nt})$ to produce an adjusted estimator $\tilde{\theta}_A$ with a small bias. We later introduce Assumption 4.10 which provides a condition for $(dY_{nt}, dX_{nt})$ to produce an adjusted estimator $\tilde{\xi}_{tA}$ with a small bias. We define the class of adjusted estimators to be estimators $\tilde{\theta}_A$ and $\tilde{\xi}_{tA}$ which are obtained using $(dY_{nt}, dX_{nt})$ satisfying Assumptions 4.9 and 4.10. Specific examples of the adjustment term $(dY_{nt}, dX_{nt})$ are given in Section E. We first investigate theoretical properties.

Assumption 4.9. There exists a constant $K_2$ such that for all $t = 1, 2, \ldots, b_n$, and all $n$,
\[ E\left[ \left| \left( d_{nt}, d_{nt}^* \right) \right|^n \right] < K_2 a_n^{-4}. \]

Also,

\[ \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{0t}^{\mathbf{g}_0} \xi_{nt} \left\{ d_{nt} - d_{nt}^* \xi_{0t}^{\mathbf{g}_0} \xi_{nt} - \frac{1}{2} \text{tr}[\xi_{0t}^{\mathbf{g}_0} A_{nt}^{-1}] \right\} \]

\[ = o_p \left( \max \left[ a_n^{-3/2}, n^{-1/2} a_n^{-1/2} \right] \right). \quad (4.29) \]

Since under Assumptions 4.4, 4.6a, and 4.7a,

\[ \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{0t}^{\mathbf{g}_0} \xi_{nt} \left\{ \delta_{nt} \xi_{nt} \left( \frac{1}{2} \delta_{nt} - u_{nt} \right)' + \frac{1}{2} \text{tr}[\xi_{0t}^{\mathbf{g}_0} A_{nt}^{-1}] \right\} \]

\[ = o_p \left( \max \left[ a_n^{-3/2}, n^{-1/2} a_n^{-1/2} \right] \right), \]

the condition (4.29) is equivalent to the condition

\[ \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{0t}^{\mathbf{g}_0} \left\{ d_{nt} - d_{nt}^* + \delta_{nt} \xi_{0t}^{\mathbf{g}_0} \left( \frac{1}{2} \delta_{nt} - u_{nt} \right)' \right\} \]
Also, the condition (4.29) implies that \( dY_{nt} - d\xi_{nt} f_{\xi}^0 \) is estimating the bias contribution given in (4.17).

To investigate the asymptotic properties of \( \hat{\beta}_A \), we consider the following one-step estimator \( \hat{\beta}_A \) using a preliminary estimator \( \bar{\beta} \).

Let \( \bar{x}_{tA} \) be the value of \( x_t \) contained in \( \Gamma \) that minimizes

\[
q(\bar{\beta}, x_t; y_{nt}, \bar{x}_{nt}),
\]

and let

\[
\hat{\beta}_A = \bar{\beta} + (\hat{\Delta}_A),
\]

where \( \hat{\Delta}_A \) satisfies

\[
M_{xx}(\hat{\Delta}_A) = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}) [\hat{e}_{nt} - \hat{u}_{nt} \xi'_{x}(\bar{x}_{tA}; \bar{\beta})],
\]

\[
\hat{\beta} = \frac{1}{n} \sum_{t=1}^{b_n} \xi'_{\beta}(\bar{x}_{tA}; \bar{\beta}) \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}),
\]

\[
\hat{e}_{nt} = y_{nt} - f(\bar{x}_{tA}; \bar{\beta}),
\]

\[
\hat{\Delta}_A = \frac{1}{n} \sum_{t=1}^{b_n} \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}) \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}),
\]

\[
\sigma_{vnt}^{-2} = [1, - \xi_{x}(\bar{x}_{tA}; \bar{\beta})] \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}) [1, - \xi_{x}(\bar{x}_{tA}; \bar{\beta})]',
\]

\[
\hat{\Delta}_A = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}) \xi_{\beta}(\bar{x}_{tA}; \bar{\beta})
\]

\[
\hat{\Delta}_A = n^{-1} \sum_{t=1}^{b_n} \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}) \xi_{\beta}(\bar{x}_{tA}; \bar{\beta}).
\]
The properties of the one step estimator $\hat{\delta}_A$ are given in the following theorem.

**Theorem 4.6.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, 4.7a, and 4.9 hold. Then,

$$\hat{\delta}_A - \delta^0 = m_{XX}^{-1} (\delta_1 \hat{\delta}) + o_p (\max[a_n^{-3/2}, a_n^{-1/2} n^{-1/2}]) .$$

If, in addition, Assumption 4.8a holds, then

$$\frac{1}{n^{1/2}} (\hat{\delta}_A - \delta^0) \xrightarrow{L} N (0, m_{XX}^{-1}) .$$

**Proof.** By Assumption 4.9,

$$(dY_{nt}, dX_{nt}) = o_p (a_n^{-1}) .$$

Thus, it follows that

$$\frac{1}{n^{1/2}} (\xi_{tA} - \xi_t) = \delta_{nt} + o_p (\max[a_n^{-1}, n^{-1/2}]) .$$

(4.31)

We observe that

$$q(\tilde{\delta}, \xi_{tA}; Y^*_{nt}, X^*_{nt}) < q(\tilde{\delta}, \xi^0_{tA}; Y^*_{nt}, X^*_{nt}) .$$

(4.32)
By the compactness of \( \mathcal{G} \) and Assumptions 4.6a, 4.7a, and 4.9, it can be shown that for a constant \( K_3 \) and for \( t = 1,2,\ldots,b_n \),

\[
E\{q^2(\bar{g}, \bar{z}_t^0; \bar{y}_{nt}^*, \bar{z}_{nt}^*)\} < K_3.
\] (4.33)

It follows from (4.32) and (4.33) that for a constant \( K_4 \) and for \( t = 1,2,\ldots,b_n \),

\[
E[|\bar{x}_{nt}^* - \bar{x}_{tA}|^4] < K_4 a_n^{-2}.
\] (4.34)

By (4.34) and Assumptions 4.7a and 4.9, for a constant \( K_5 \) and for \( t = 1,2,\ldots,b_n \),

\[
E[|\bar{x}_{tA} - \bar{x}_t^0|^4] < K_5 a_n^{-2}.
\] (4.35)

By (4.35) and Assumptions 4.3 and 4.6a, it can be shown that

\[
b_n^{-1} \sum_{t=1}^{b_n} [(a_n^2)_{vnt}^{-1} - (a_n^2_{vnt})^{-1} + (a_n^2_{vnt})^{-2}(a_n A_{vnt})^2] = O_p(a_n^{-2}).
\] (4.36)

The expansion (4.35) implies that the order of an average over \( b_n \) values of \( (\bar{x}_{tM} - \bar{x}_t^0)^i \) is of \( O_p(a_n^{-i/2}) \) for \( i < 4 \). Thus, using this fact, Assumption 4.6a, and (4.36), we obtain
\[ M_{xx} = m_{xx} + O_p(\max[a_n^{-1}, n^{-1/2}]) . \] (4.37)

We observe that

\[ \hat{e}_{nt} - \hat{u}_{nt} \tilde{f}_x(x_tA; \tilde{g}) \]

\[ = f_{\beta t}^0(\tilde{g}^0 - \tilde{g}) + v_{nt}^* + dx_{nt} - dx_{nt} \tilde{x}_{nt}\frac{f_{\beta t}^0}{2} \tilde{\delta}_{nt}\tilde{\delta}_{nt}\frac{1}{2} \delta_{nt}^t - u_{nt}^t \]

\[ + O_p(\max[a_n^{-3/2}, n^{-1/2}a_n^{-1/2}]) , \] (4.38)

where the remainder is a function of the powers of \((\vec{x}_{tA} - \vec{x}_t^0)\), \((\vec{e} - \vec{g}_t^0)\), and \(dx_{nt}\). Thus, it follows from (4.14), (4.15), (4.30), (4.35), (4.36), (4.38), and Assumptions 4.5, 4.6a, and 4.9 that

\[ n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt} \tilde{f}_x(x_tA; \tilde{g})[\hat{e}_{nt} - \hat{u}_{nt} \tilde{f}_x(x_tA; \tilde{g})] \]

\[ = n^{-1} \sum_{t=1}^{b_n} \sigma_{vnt} \tilde{e}_g(\vec{x}_{tA}; \vec{g}) \tilde{f}_{\beta t}^0(\vec{g}^0 - \vec{g}) + (\hat{\delta}_{1} \hat{g}) \]

\[ + O_p(\max[a_n^{-3/2}, n^{-1/2}a_n^{-1/2}]) . \] (4.39)
Therefore, the result for the expansion of \((\hat{\xi}_A - \xi^0)\) follows from (4.37) and (4.39). The limiting distribution result holds by Result 4.4.

To consider the asymptotic properties of the adjusted estimator \(\hat{x}_{tA}\), we define \(\hat{x}_{tA}\) to be the value of \(x_t\) contained in \(F\) that minimizes \(q(\hat{\xi}_A, x_t; Y_{nt}^*, X_{nt}^*)\). We introduce a further assumption on \((dY_{nt}, dX_{nt})\).

**Assumption 4.10.** There exist \(\Delta Y_{nt}\) and \(\Delta X_{nt}\) such that

\[
dY_{nt} = \Delta Y_{nt} + o_p(\max[a_n^{-3/2}, n^{-1/2}a_n^{-1}])
\]

\[
dX_{nt} = \Delta X_{nt} + o_p(\max[a_n^{-3/2}, n^{-1/2}a_n^{-1}])
\]

\[
E[\Delta Y_{nt}] = \frac{1}{2} \text{tr}[\xi^{0,0}_{XX}^{-1}] + E[\Delta X_{nt}] \xi^{0,0}_{XX}^{-1}
\]

The next theorem gives the properties of \(\hat{x}_{tA}\).

**Theorem 4.7.** Let the model (4.1) hold, and let Assumptions 4.1 through 4.5, 4.6a, 4.7a, 4.9, and 4.10 hold. Then,

\[
\hat{x}_{tA} - x_t^0 = \Delta \hat{x}_{tA} + o_p(\max[a_n^{-3/2}, a_n^{-1/2}])
\]

where
\[ \Delta_{\hat{\mathbf{x}}_{tA}} = \hat{\delta}_{nt} + \left[ \delta_{1t} + \delta_{2t} + \delta_{3t} \right] (a_{n}^{-1}) \]

\[ + \left( \Delta Y_{nt}, \Delta X_{nt}, (a_{n}^{-1}) \right) \left( g_{0}^{0}, \xi \right) \left( a_{n}^{-1} \right). \]

Also,

\[ E[\Delta_{\hat{\mathbf{x}}_{tA}}] = E[\Delta X_{nt}]. \]

**Proof.** We observe that the modification \((4.28)\) is of \(O_{p}(a_{n}^{-1})\), and that

\[ \hat{\delta}_{A} - \delta_{0} = O_{p}(\max[a_{n}^{-1/2}, n^{-1/2}]). \]

Thus, by Result 4.1,

\[ \hat{\mathbf{x}}_{tA} - \mathbf{z}_{t} = \hat{\delta}_{nt} + O_{p}(\max[a_{n}^{-1}, n^{-1/2}]). \] (4.40)

Using \((4.40)\) and Theorem 4.6, we obtain that

\[ y_{nt}^{*} - f(\hat{\mathbf{x}}_{tA}; \hat{\mathbf{z}}_{nt}) = e_{nt} + \Delta Y_{nt} - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' - \frac{1}{2} \delta_{nt} \delta_{nt} - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' \]

\[ - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{z}_{t} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{z}_{t} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{z}_{t} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{x}_{tA} - \mathbf{z}_{t})' - \xi_{nt}(\mathbf{z}_{t} - \mathbf{z}_{t})'. \] (4.41)

and that
\[ f_{x}(\hat{x}_{tA}; \hat{g}_{A}) = f_{x}^{0} + \delta_{n}^{0} \xi_{xx} + 0_{p}(\max[\lambda_{n}^{-1}, n^{-1/2}]) . \]  
\hspace{1cm} \text{(4.42)}

It follows from (4.41) and (4.42) that

\[ (\delta_{nt} + \Delta Y_{nt} - f_{x}^{0} (\hat{x}_{tA} - x_{t}^{0}) - \frac{1}{2} \delta_{nt}^{0} \xi_{xx} - f_{l}^{0} \mu_{nt} \hat{g}_{l} - \delta_{nt} \hat{g}_{l} \mu_{nt} \Delta X_{nt} x_{t}^{0})] \]

\[ (\sigma_{n}^{-1}) [f_{x}^{0} + \delta_{nt}^{0} \xi_{xx}^{-1} \mu_{nt}] = 0_{p}(\max[\lambda_{n}^{-3/2}, \lambda_{n}^{-1/2} n^{-1/2}]) . \]

Therefore, the expansion for \((\hat{x}_{tA} - x_{t}^{0})\) follows. If Assumption 4.10 holds, then

\[ E \{ (\Delta Y_{nt}, \Delta X_{nt}) \Sigma_{n}^{-1} [f_{x}^{0}, \mu_{nt}] \} \]

\[ = \frac{1}{2} \text{tr}[f_{x}^{0} A_{nt}^{-1}], \Omega \right] \Sigma_{n}^{-1} [f_{x}^{0}, \mu_{nt}] A_{nt}^{-1} + E \{ \Delta X_{nt} \} . \]

Also, by Theorem 4.2 and Theorem 4.3,

\[ E \{ \delta_{nt} \} = E \{ \delta_{l} \xi_{l} \} = E \{ \delta_{l} \hat{g}_{l} \} = 0 , \]

\[ E \{ \delta_{x_{l}} \} = - (\frac{1}{2} \text{tr}[f_{x}^{0} A_{nt}^{-1}], \Omega \right) (\sigma_{n}^{-1} f_{x}^{0}, \mu_{nt}) \} . \]

Thus, under Assumption 4.10,

\[ E \{ \Delta \xi_{tA} \} = E \{ \Delta X_{nt} \} . \]
The iterative procedure for successively obtaining $\tilde{\varepsilon}_A$ and $\tilde{x}_{tA}$ leads to the adjusted estimators $\tilde{\varepsilon}_A$ and $\tilde{x}_{tA}$. The estimators of $\varepsilon^0$ and $x^0_t$ obtained by any finite number of such iterations have asymptotic properties given in Theorem 4.6 and Theorem 4.7. We might expect that the same asymptotic properties would apply to the adjusted estimators $\tilde{\varepsilon}_A$ and $\tilde{x}_{tA}$ which maximize the likelihood function of $(Y_{nt}^*, X_{nt}^*)$. But, we emphasize that this is an unproven conjecture.

The adjusted estimators $\tilde{\varepsilon}_A$ and $\tilde{x}_{tA}$ can be improved by updating the function $(dY_{nt}^*, dx_{nt}^*)$ at each stage of iteration based on the most recent estimates of $\varepsilon^0$ and $x^0_t$, or using $\tilde{\varepsilon}_A$ and $\tilde{x}_{tA}$ to estimate a new adjustment term and repeating the maximization of the likelihood. Because of the computational simplicity we chose to define our class of adjusted estimators to be the maximum likelihood estimator based on the observations with fixed modifications that do not vary by iteration. If the maximum likelihood estimators without adjustment are used as the preliminary estimators $\tilde{\varepsilon}$ and $\tilde{x}$, then two runs of a standard nonlinear regression algorithm can produce any member of our class of adjusted estimators.

F. Examples of the Adjusted Estimators

As we have seen in the previous section, any adjusted estimator $\tilde{\varepsilon}_A$ satisfying Assumption 4.9 has no bias due to the curvature and has the same asymptotic properties as the estimator $\tilde{\varepsilon}$. Also, certain choices of $dY_{nt}^*$ and $dx_{nt}^*$ reduce errors in the adjusted estimator
\( \hat{X}_{tA} \). In this section, we present three examples of the adjusted estimators, and make comparisons among them. In Section G, we present the results from a Monte Carlo experiment carried out to study the small sample properties of various estimators of \( \hat{g}^0 \) including three estimators introduced in this section.

For the three adjusted estimators considered in this section, we use the maximum likelihood estimators \( \hat{g}_M \) and \( \hat{g}_{LM} \) as the preliminary estimators \( \hat{g} \) and \( \hat{g}_t \). In this section, we act as if \( \hat{g}_M, \hat{g}_{LM} \), and the three adjusted estimators have the same asymptotic properties as their corresponding one step estimators.

1. Adjustment based on the local quadratic approximation

A simple adjustment satisfying Assumptions 4.9 and 4.10 is obtained by setting

\[
\begin{align*}
\hat{X}_{nt} &= \frac{1}{2} \text{tr} \left[ \xi_{xx} (\hat{g}_{LM}; \hat{g}_M) \hat{A}_{nt}^{-1} \right] , \\
\hat{Y}_{nt} &= 0 ,
\end{align*}
\]

where

\[
\begin{align*}
\hat{A}_{nt}^{-1} &= \xi_{uun} - \xi_{uvnt} \xi_{vnt}^{-1} \xi_{vunt} , \\
\xi_{uvnt} &= \xi_{vunt} = \xi_{uun} - \xi_{uun} \xi_{x} (\hat{g}_{LM}; \hat{g}_M) , \\
\xi_{vunt} &= \left(1, - \xi_{x} (\hat{g}_{LM}; \hat{g}_M) \right) \xi_{uun} \left(1, - \xi_{x} (\hat{g}_{LM}; \hat{g}_M) \right)' .
\end{align*}
\]
Denote the resulting adjusted estimators by \( \tilde{\xi}_1 \) and \( \tilde{\xi}_{t1} \). Since \( \hat{\xi}_M \) and \( \hat{\xi}_M \) are the maximum likelihood estimators, under Assumptions 4.1 through 4.5, 4.6a, and 4.7a, \( (d_{1 \text{nt}}, d_{1 \text{nt}}) \) satisfies Assumptions 4.9 and 4.10 with

\[
\Delta_{1 \text{nt}} = \frac{1}{2} \text{tr}(\xi^0 \Sigma_{\text{nt}}^{-1}) ,
\]

\[
\Delta_{1 \text{nt}} = 0 .
\]

Thus, the results of Theorem 4.6 and Theorem 4.7 hold for \( \tilde{\xi}_1 \) and \( \tilde{\xi}_{t1} \). Since \( \Delta_{1 \text{nt}} = 0 \), the estimator \( \tilde{\xi}_{t1} \) is unbiased for \( \xi_t^0 \) through terms of \( O_p(a_n^{-1}) \). One interpretation of the adjusted estimators \( \tilde{\xi}_1 \) and \( \tilde{\xi}_{t1} \) can be obtained through the following observation.

Suppose the model (4.1) is quadratic in \( \xi_t^0 \) and is given by

\[
y_t^0 = \theta_o + \xi_t^0 \theta_1 + \frac{1}{2} \xi_t^0 \Sigma_2 \Sigma_t^{0'} . \tag{4.43}
\]

where \( \Sigma_2 \) is a symmetric \( q \times q \) matrix. The model (4.43) is linear in the parameters, and we can write

\[
y_t^0 = z_t^0 \Sigma_0 ,
\]

where
\[
\begin{align*}
\gamma_t^0 &= (1, \gamma_t^0, \frac{1}{2} [\gamma_t^0 \gamma_t^0]^T \theta_q), \\
\bar{\theta}^0' &= (\theta_0, \theta_1', \text{vech } \theta_2').
\end{align*}
\]

We observe that

\[
\begin{align*}
\gamma_{xt}^0 &= \theta_1' + \gamma_t^0 \theta_2', \\
\gamma_{xxt}^0 &= \theta_2', \\
\gamma_{\theta t}^0 &= \gamma_t^0'.
\end{align*}
\]

Thus, the term \(\mathbb{E}_{XX}^{-1}(\delta, \hat{\delta})\) in the expansion of the maximum likelihood estimator \(\hat{\theta}_M\) of \(\theta^0\) for this model becomes

\[
\begin{align*}
\mathbb{E}_{XX}^{-1}(\delta, \hat{\delta}) &= \left( \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 \right)^{-1} \left( \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 \right)^{-1} \left( \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \mathbb{E}_t \mathbb{E}_t \gamma_t^0 \gamma_t^0 \right)^{-1} \left( \frac{1}{2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 - \mathbb{E}_t \gamma_t^0 \right) \\
&= \left( \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 \right)^{-1} \left( \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 \right)^{-1} \left( \frac{1}{2} \mathbb{E}_t \gamma_t^0 \gamma_t^0 - \mathbb{E}_t \gamma_t^0 \right).
\end{align*}
\] (4.44)

where

\[
\begin{align*}
\sigma_{vnt}^2 &= \left[ 1, - \theta_1', - \mathbb{E}_t \theta_2' \right] \Sigma_n \left[ 1, - \theta_1', - \mathbb{E}_t \theta_2' \right]' \\
\delta_{nt} &= \mathbb{E}_n \mathbb{E}_n^{-1} \bar{\theta}_1 + \mathbb{E}_n \mathbb{E}_n^{-1} \bar{\theta}_2 \mathbb{E}_t \gamma_t^0, \xi_j A_n^{-1}
\end{align*}
\]
\[ A_{nt}^{-1} = \Sigma_{uun} - \Sigma_{uvn} \sigma_{vnt}^{-2} \Sigma_{vun}, \]
\[ \Sigma_{uvn} = \Sigma'_{vun} = \Sigma_{eun} - \Sigma_{uun} (\theta_1 + \theta_2 z^0_t). \]

Also,
\[ E[\delta_2 \hat{\theta}_2] = -\frac{1}{2} \left\{ \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} z_t^0 z_t^0 \right\}^{-1} \sum_{t=1}^{b_n} \sigma_{vnt}^{-2} z_t^0 \tr[\delta_2 A_{nt}^{-1}]. \]  

(4.45)

For a general nonlinear model (4.1), we approximate the function 
\[ f(x; \theta^0) \] around \( x = x^0_s \) by the first three terms of the Taylor expansion
\[ f(x; \theta^0) = f^0 + (x - x^0_s)f^0_{xs} + \frac{1}{2} (x - x^0_s)^2 f^0_{xxs} (x - x^0_s)', \]

(4.46)

Suppose that we estimate \( f^0_s, f^0_{xs}, \) and \( \text{vech} f^0_{xxs} \) instead of \( \theta^0 \) based on the observations on \( x^0_s, t = 1,2,...,b_n \), which are close to \( x^0_s \). Then, the bias of the maximum likelihood estimator is given by (4.45) with \( x^0_s, f^0_{xs}, \) and \( f^0_{xxs} \) replacing \( \theta_0, \theta_1, \) and \( \theta_2 \), respectively. If \( A_{nt} \) is approximated by \( A_{ns} \) for \( x^0_t \) close to \( x^0_s \), then the bias becomes
\[ -\frac{1}{2} \tr[\delta_0 \hat{\theta}_2 A_{ns}^{-1}] (1, 0, 0)', \]

(4.47)
Thus, the maximum likelihood estimators for $f_{xS}^0$ and \text{vech } $f_{xXS}^0$ are unbiased, while the maximum likelihood estimator of the intercept term has the bias

$$-\frac{1}{2} \text{tr}[f_{xXS}^0 A^{-1}_{ns}] .$$

Hence, the bias may be removed by subtracting (4.49) from $Y_{nt}$ corresponding to $x_0^0$ close to $x_s^0$. If we apply the same argument to each $x_s^0$, and if we estimate (4.48) by a natural estimator

$$-\frac{1}{2} \text{tr}[f_{XX} (\hat{x}_s^M; \hat{g}_M) A^{-1}_{ns}] ,$$

then we arrive at the adjustment given by $(dY_{ns}, dX_{ns})$. Therefore, the estimators $\tilde{x}_1$ and $\tilde{x}_t$ are derived based on the bias results for the local quadratic approximation to the function $f(x; g^0)$.

2. Adjustment based on the curvature relative to the tangent

The bias result (4.45) for the quadratic model (4.43) motivates another type of adjustment. Assume that

$$\tilde{E}_{\eta} = \sigma^2 I ,$$

(4.50)
where \( \sigma^2 = O(a_n^{-1}) \). Then, using (4.48), we rewrite (4.45) to obtain

\[
\mathbb{E}[\mathbf{w}_{xx}^{-1} \delta \mathbf{g}] = -\frac{1}{2} \sigma^2 \text{tr}[\mathbf{\theta}_2](1, 0, 0)' + \frac{1}{2} \sigma^2 \left\{ \sum_{t=1}^{b_n} \sigma_{vnt}^2 \mathbf{z}_t^0 \mathbf{z}_t^0 \right\}^{-1}
\]

Thus, the bias contribution from the point \( \mathbf{z}_t^0 = -\mathbf{\theta}_1 \mathbf{g}^{-1} \) is

\( (-\frac{1}{2} \sigma^2 \text{tr}[\mathbf{\theta}_2], 0, 0)' \). This suggests the following adjustment. First, for each \( \mathbf{z}_t^0 \), we rotate the coordinate system from \((y, \mathbf{x})\) to \((z, \mathbf{w})\) so that the function \( z = y(\mathbf{w}; \mathbf{g}_0) \) transformed from \( y = f(\mathbf{x}; \mathbf{g}_0) \) satisfies \( g_w(\mathbf{w}_t^0; \mathbf{g}_0) = 0 \), where \((\mathbf{z}_t^0, \mathbf{w}_t^0)\) is the rotated \((\mathbf{z}_t^0, \mathbf{x}_t^0)\). Then, the adjustment in the new coordinate system is obtained by adding \( \frac{1}{2} \sigma^2 \text{tr}[g_{ww}(\mathbf{w}_t^0; \mathbf{g}_0)] \) to the observation on \( \mathbf{z}_t^0 \). Then, we transform back to the original axes to obtain an adjustment to \((\mathbf{y}_n, \mathbf{x}_n)\). Note that \( g_{ww}(\mathbf{w}_t^0; \mathbf{g}_0) \) is the curvature of the function \( f(\mathbf{x}; \mathbf{g}_0) \) relative to the tangent at \( \mathbf{x} = \mathbf{x}_t^0 \).

The desired rotation of axes can be obtained by an orthogonal transformation

\[
(y, \mathbf{x}) = (z, \mathbf{w}) \mathbf{P} \quad ,
\]

where
\[ P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} , \]

\[ p_{11} = (1 + \xi^0 \xi_{xt}^0)^{-1/2} , \]

\[ p_{12} = -(1 + \xi^0 \xi_{xt}^0)^{-1/2} \xi^0 , \]

\[ p_{21} = (1 + \xi^0 \xi_{xt}^0)^{-1/2} \xi^0 , \]

\[ p_{22} = (1 + \xi^0 \xi_{xt}^0)^{-1/2} , \]

Since \( y = f(x; \xi^0) \), we have

\[ F(z, y) = z p_{11} + y p_{21} - f(z, p_{12} + y, p_{22}; \xi^0) = 0 . \]

(4.52)

We observe that

\[ F_z(z, y) = \frac{\partial}{\partial z} F(z, y) \]

\[ = p_{11} - \xi_x (z p_{12} + y p_{22}; \xi^0) P_{12} , \]

\[ F_w(z, y) = \frac{\partial}{\partial w} F(z, y) \]
Thus, by (4.52) and the implicit function theorem, for \((z, w)\) close to \((z^0, w^0)\), there exists a function \(z = g(w; \xi^0)\) such that

\[
g_w(z; \xi^0) = \left\{ f_z[g(w; \xi^0), w] \right\}^{-1} f_w[g(w; \xi^0), w]
\]

Hence,

\[
g_{ww}(w; \xi^0) = (f_z^{-1} f_w^{\text{F}} f_{12} - f_z f_{22}^{-1})(- f_{xx} f_{22}^{\text{F}} + g_w^{\text{F}} f_{12}^{\text{F}})
\]

where

\[
f_z = f_z[g(w; \xi^0), w]
\]

\[
f_w = f_w[g(w; \xi^0), w]
\]

\[
f_{xx} = f_{xx}[g(w; \xi^0) f_{12} + w f_{22}; \xi^0]
\]

\[
g_w = g_w(w; \xi^0)
\]

Note that

\[
f_x[g(w_t^0; \xi^0) f_{12} + w_t^0 f_{22}; \xi^0] = f_{xt}^0
\]
and thus,

\[ F_z [g(w_t^0; g_t^0), w_t^0] = (1 + \xi_{x,t}^0 f_0^0)^{1/2}, \]

\[ F_w [g(w_t^0; g_t^0), w_t^0] = 0, \]

\[ g_t^w (w_t^0; g_t^0) = 0. \]

Therefore,

\[ g_{ww} (w_t^0; g_t^0) = (1 + \xi_{x,t}^0 f_0^0)^{-1/2} [I + \xi_{x,t}^0 f_0^0]^{-1/2} \xi_{x,t}^0 [I + \xi_{x,t}^0 f_0^0]^{-1/2}. \]

The additive adjustment term for the observation on \((z_t^0, w_t^0)\) is \((dZ_{nt}^0, 0)\), where

\[ dZ_{nt}^0 = \frac{1}{2} \sigma^2 \text{tr}[g_{ww} (w_t^0; g_t^0)]. \]

Using (4.51), we transform back to the original coordinate \((y, x)\). Then, the additive adjustment term for \((y_{nt}, X_{nt})\) is

\[ dZ_{nt} (p_{11}, p_{12}) = \frac{1}{2} \sigma^2 (1 + \xi_{x,t}^0 f_0^0)^{-1} \text{tr}[\xi_{x,t}^0 [I + \xi_{x,t}^0 f_0^0]^{-1}] (1, - \xi_{x,t}^0). \]

Note that under the assumption (4.50)

\[ \sigma^2 (1 + \xi_{x,t}^0 f_0^0) = \sigma^2_{vnt}. \]
Thus, the adjustment (4.53) can be written as

$$
\frac{1}{2} \text{tr} \left( \xi^0 \Delta_{nt}^{-1} \right) \sigma_{\text{vnt}}^2 \Sigma_{\text{vnt}} = \Delta_{nt}^{-1} .
$$

If the error covariance matrix $\Sigma_n$ is not of the form (4.50), then we consider the tangent and the curvature relative to the tangent in the metric $\Sigma_{nt}^{-1}$. Instead of the orthogonal transformation (4.51), we obtain the desired rotation of axes by a transformation

$$(y, x) = (z, w)\Omega ,$$

where

$$
\Omega = \left( \xi_1^0, \xi_2^0 \right) ,
$$

$$
\xi_1 = \left( \sigma_{\text{vnt}}^2 \right)^{-1/2} (1, -\xi_{xt}^0) \Sigma_n ,
$$

$$
\xi_2 = \Delta_{nt}^{-1/2} [\xi_{xt}, \xi] .
$$

We observe that
Following the above steps with \( Q \) replacing \( P \), we find that the adjustment term for \((Y_{nt}, X_{nt})\) with a general \( \Sigma_n \) is also given by the expression (4.54). Replacing unknown parameters \( \gamma_t \) and \( \delta \) by \( \hat{\gamma}_{tM} \) and \( \hat{\delta}_M \), we define the adjustment based on the curvature relative to the tangent to be

\[
(d_{2Y_{nt}}, d_{2X_{nt}}) = \frac{1}{2} \text{tr} \left[ \Sigma_{XX}(\hat{\gamma}_{tM}; \hat{\delta}_M) \hat{\Sigma}_{nt}^{-1} \right] \hat{\sigma}^{-2}_{vnt} \Sigma_{vnt},
\]

(4.56)

where

\[
\hat{\Sigma}_{vnt} = (\hat{\sigma}^2_{vnt}, \hat{\sigma}^2_{vunt}),
\]

\[
\hat{\sigma}_{vnt} = \sigma_{een} - \Sigma_{XX}(\hat{\gamma}_{tM}; \hat{\delta}_M) \Sigma_{ussen}.
\]

By noting that

\[
\Sigma_{vnt} \left(1, - \frac{f^0}{\gamma_{nt}}\right)' = \sigma_{vnt}^2,
\]

it immediately follows that the adjustment (4.56) satisfies Assumptions 4.9 and 4.10 with \((\Delta_{2Y_{nt}}, \Delta_{2X_{nt}})\) being the quantity in (4.54). Therefore, the estimators \( \hat{\gamma}_2 \) and \( \hat{\gamma}_{t2} \) obtained using \((d_{2Y_{nt}}, d_{2X_t})\) have the properties stated in Theorem 4.6 and Theorem 4.7. Since
the expansion of \( \tilde{\xi}_{t2} \) is given by

\[
\tilde{\xi}_{t2} - \delta_t = \delta_{nt} + \left[ \delta_1 \hat{\xi}_t + \delta_2 \hat{x}_t + \delta \hat{x}^* \right] \left( a_n^{-1} \right) \\
+ O_p \left( \max \left[ a_n^{-3/2}, a_n^{-1/2}, a_n^{-1/4} \right] \right).
\]

Hence, up to the terms of \( O_p(a_n^{-1}) \), the adjusted estimators \( \tilde{\hat{\xi}}_2 \) and \( \tilde{\hat{\xi}}_{t2} \) are equivalent to the modified maximum likelihood estimators \( \tilde{\hat{\xi}}_1 \) and \( \tilde{\hat{\xi}}_t \). The relation that \( E(\delta \hat{x}^* | a_n^{-1}) = \Delta \hat{x}_{nt} \) could have been derived directly by Theorem 2.1.

3. Adjustment based on the expansion terms

The two previous adjusted estimators are obtained by subtracting estimates of the bias from the observations. Instead of estimating the bias, we can estimate a term in the expansion of \( \hat{\hat{\xi}}_M \) which introduces the bias in \( \hat{\hat{\xi}}_M \). From the expression (4.38), we observe that an adjustment can be obtained by letting \( \left( dX_{nt} - d\hat{x}_{nt} \hat{x}_{nt}^0 \right) \) estimate the term \( - \delta_{nt} \cdot \left( \frac{1}{2} \delta_{nt}^t u_{nt}^t - u_{nt} \right) \).

Let

\[
d_{3}X_{nt} = - \frac{1}{2} \text{tr} \left[ f_{\hat{\hat{\xi}}_{xs}} (\hat{\hat{\xi}}_M^t; \hat{\hat{\xi}}_M^t) (\hat{u}_{nt} \cdot \hat{u}_{nt} - \delta_{u_{nt}^t}) \right],
\]

\[
d_{3}X_{nt} = 0,
\]
where

\[ \hat{u}_{nt} = \hat{X}_{nt} - \hat{X}_t M. \]

We note that \( d_3 Y_{nt} \) has the same form as the modification used by Wolter and Fuller (1982b). It follows that

\[ d_3 Y_{nt} = \Delta_3 Y_{nt} + O_p(\max(a_n^{-3/2}, n^{-1/2} a_n^{-1/2})), \]

where

\[ \Delta_3 Y_{nt} = -\frac{1}{2} \text{tr}\{g_{xxt}^0[(u_{nt} - \delta_{nt})'(u_{nt} - \delta_{nt}) - \Sigma_{uun}]\}, \]

\[ E[\Delta_3 Y_{nt}] = \frac{1}{2} \text{tr}\{g_{xxt}^0 A_{nt}^{-1}\}. \]

Therefore, \((d_3 Y_{nt}, d_3 X_{nt})\) satisfies Assumption 4.10 with \( \Delta_3 X_{nt} = 0 \).

Since

\[ \delta_{nt} g_{xxt}^0 (\frac{1}{2} \delta_{nt}' - u_{nt}') = \frac{1}{2} \text{tr}\{g_{xxt}^0 [(u_{nt} - \delta_{nt})'(u_{nt} - \delta_{nt}) - u_{nt}' u_{nt}]\}, \]

we expect the adjustment \((d_3 Y_{nt}, d_3 X_{nt})\) to cancel the term \( \delta_2 \hat{q} \) in the expansion of \( \hat{g}_M \). However, to derive the asymptotic properties of the resulting estimators \( \tilde{g}_3 \) and \( \tilde{X}_{t3} \), we need to introduce an additional assumption.

**Assumption 4.11.** For some real \( L^0 \) and all \( t \) and \( n \),
Under Assumption 4.11 and the assumptions in Theorem 4.6, it follows from (4.29) that \((d_3Y_{nt}, d_3X_{nt})\) satisfies Assumption 4.9.

Thus, with Assumption 4.11, Theorem 4.6 and Theorem 4.7 hold for the adjusted estimators \(\tilde{\theta}_3\) and \(\tilde{\xi}_{t3}\). Hence, \(\tilde{\theta}_3\) has the same properties as \(\tilde{\theta}_1\) and \(\tilde{\theta}_2\), and \(\tilde{\xi}_{t3}\) is unbiased through terms of \(O\left(a_n^{-1}\right)\). Note that the adjustment \(d_3Y_{nt}\) is similar to the modification used in the modified maximum likelihood estimators \(\tilde{\theta}\) and \(\tilde{\xi}_t\). While \(\tilde{\theta}_3\) and \(\tilde{\theta}\) are asymptotically equivalent, the almost unbiased estimator \(\tilde{\xi}_{t3}\) is clearly superior to the estimator \(\tilde{\xi}_t\) which has the bias due to the curvature.

4. Comparison among the adjusted estimators

While all three adjusted estimators \(\tilde{\theta}_1\), \(\tilde{\theta}_2\), and \(\tilde{\theta}_3\) have the same expansion form, the estimators \(\tilde{\xi}_{t1}\), \(\tilde{\xi}_{t2}\), and \(\tilde{\xi}_{t3}\) have slightly different expansions. By Theorem 2.1 and Theorem 4.7, we obtain

\[
\tilde{\xi}_{t1} - \tilde{\xi}_t^0 = \Delta_{\tilde{\xi}}_t + \frac{1}{2} \text{tr}[\xi_{\tilde{\xi}t}^0 (\xi_{\tilde{\xi}t}^0 - \Lambda_{\tilde{\xi}t}^{-1})] \sigma^{-2}_{vnt} \xi_{vnt} + o_p \left( \max[a_n^{-3/2}, n^{-1/2}a_n^{-1/2}] \right),
\]

\[
\tilde{\xi}_{t2} - \tilde{\xi}_t^0 = \Delta_{\tilde{\xi}}_t + \frac{1}{2} \text{tr}[\xi_{\tilde{\xi}t}^0 (\xi_{\tilde{\xi}t}^0 - \Lambda_{\tilde{\xi}t}^{-1})] \sigma^{-2}_{vnt} \xi_{vnt} + o_p \left( \max[a_n^{-3/2}, n^{-1/2}a_n^{-1/2}] \right),
\]
\[ \tilde{\xi}_t^3 - \xi_0^3 = \Delta \tilde{\xi}_t^3 + \frac{1}{2} \text{tr} \left[ \xi_{xx}^0 (\delta'_{nt} \delta_{nt} - A_{nt}^{-1} + \tilde{v}_{nt} \tilde{v}_{nt} - \Sigma_{uun}) \right] \sigma_{vnt}^2 \Sigma_{vunt} \]

\[ + o_p \left( \max \left[ \frac{a_n^{3/2}}{n}, n^{-1/2} a_n^{-1/2} \right] \right), \quad (4.57) \]

where

\[ \Delta \tilde{\xi}_t^3 = \delta_{nt}^3 + [\delta'_{t} \delta_{nt} + \delta_{nt}^2] (a_{nt} A_{nt}^{-1}) \]

\[ \tilde{v}_{nt} = \tilde{u}_{nt} - \delta_{nt} \]

\[ \Sigma_{uun} = \Sigma_{uun} - A_{nt}^{-1} = \Sigma_{vnt} \sigma_{vnt}^2 \Sigma_{vunt} \]

It is obvious from (4.57) that \( \tilde{\xi}_t^3 \) is superior to \( \xi_0^3 \) as an estimator of \( \xi_0^3 \). If \( \xi_{nt} \) are normally distributed, then \( \delta_{nt}^3 \) is independent of \( v_{nt}^* \), \( \tilde{v}_{nt} \), and \( \xi_{nt}^* \). Also, the central third moments of the normal distribution are zero. Thus, it follows that the covariance between the first and second terms in each of the three expansions is zero. Hence, for the normal error model we need to compare the mean squared errors of the second terms in (4.57). Under the normality of \( \xi_{nt} \), we obtain

\[ \text{V} \{ \text{vech} (\delta'_{nt} \delta_{nt} - A_{nt}^{-1} + \tilde{v}_{nt} \tilde{v}_{nt} - \Sigma_{uun}) \} = 2 \psi_q (A_{nt}^{-1} + A_{nt}^{-1}) \psi_q \]

\[ + 2 \psi_q (\Sigma_{uun} - \Sigma_{uun}) \psi_q'. \]
Therefore, in terms of asymptotic mean squared error, $\tilde{x}_{t1}$ is preferred over $\tilde{x}_{t2}$ and $\tilde{x}_{t3}$ for the normal model.

For estimation of $\xi^0$, the choice of adjusted estimators is left undetermined. All three estimators $\tilde{\xi}_1$, $\tilde{\xi}_2$, and $\tilde{\xi}_3$ were derived based on the asymptotic results for the maximum likelihood type estimator of $\xi^0$. The estimator $\tilde{\xi}_1$ is associated with $\tilde{x}_1$ which is superior to the estimators of $\xi^0$ associated with $\tilde{x}_1$ and $\tilde{x}_2$. The estimator $\tilde{\xi}_2$ has an intuitive appeal in the sense that the adjustment is distributed over $Y_{nt}$ and $X_{nt}$ proportioned to $(\sigma_{vent}, \Sigma_{vunt})$. On the other hand, the adjustment in $\tilde{\xi}_3$ gives a cancellation of a term, which may be preferred over the adjustment in $\tilde{\xi}_1$ and $\tilde{\xi}_2$ using the expectation of the term. But, all three estimators have the same asymptotic properties up to the level of approximation we considered. In this section, we do not attempt to give any further comparison among the adjusted estimators of $\xi^0$. In the next section, using the Monte Carlo method, we investigate the small sample properties of various estimators of $\xi^0$ including the three adjusted estimators.

G. A Monte Carlo Study

In this section, we present the results from a Monte Carlo study carried out to compare the small sample behavior of the estimators of $\xi^0$. Due to an extensive amount of iterative computations required to obtain the estimates, we conducted a relatively small Monte Carlo experiment.
1. Generating the data

The model considered is

\[ y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} \]

\[ = (1, x_{1t}, x_{2t}) \xi^0 , \quad t = 1,2,\ldots,N, \]

where

\[ \xi^0 = (\beta_0, \beta_1, \beta_2)' , \]

and we observe

\[ y_t = y_t + e_t , \]

\[ X_t = x_t + u_t . \]

In this study, the sample size is \( N = 33 \), the vector of \( x \)-values is

\( (x_1, x_2, \ldots, x_{33}) = (1,1,1) \approx (-0.5, -0.4, -0.3, \ldots, 0.4,0.5) \),

the true parameter vector is \( \beta^0 = (0,0,1)' \), and the measurement errors \( (e_t, u_t) \) were generated as a random sample from a bivariate normal \( (0, 0.0324 I) \). Note that

\[ (33)^{-1} \sum_{t=1}^{33} (x_t - \bar{x})^2 = 0.1 . \]
Thus, the ratio of the standard error $\sigma_u$ to the root mean square for $x$ is 0.5692. For this parameter set, 100 samples were generated.

2. Estimators compared

For each 100 samples, we computed six estimators. These are the pseudo-maximum likelihood estimator $\hat{\beta}_M$, the modified maximum likelihood estimator $\tilde{\beta}_M$, three adjusted estimators $\hat{\beta}_{A1}$, $\hat{\beta}_{A2}$, and $\hat{\beta}_{A3}$, introduced in Section F, and the ordinary least squares estimator $\hat{\beta}_L$. Computations of the first five estimators require a preliminary estimator $\bar{\beta}$ of $\beta^0$. Wolter and Fuller (1982b) reported a Monte Carlo mean 1.37 for $\hat{\beta}_{M,2}$, where

$$\hat{\beta}_M = (\hat{\beta}_{M,0}, \hat{\beta}_{M,1}, \hat{\beta}_{M,2})'$$.

We used

$$\bar{\beta} = (0, 0, 1.37)$$

as the preliminary estimator of $\beta^0$ for $\hat{\beta}_M$, $\tilde{\beta}_M$, $\hat{\beta}_{A1}$, $\hat{\beta}_{A2}$, and $\hat{\beta}_{A3}$.

Wolter and Fuller stated that the small sample distribution function of the estimator $\hat{\beta}_M$ has thick tails, and introduced a modification to improve the small sample behavior of $\hat{\beta}_M$ and $\tilde{\beta}_M$. We applied their modification to $\hat{\beta}_M$, $\tilde{\beta}_M$, $\hat{\beta}_{A1}$, $\hat{\beta}_{A2}$, and $\hat{\beta}_{A3}$. Given the preliminary estimator $\bar{\beta} = (\bar{\beta}_0, \bar{\beta}_1, \bar{\beta}_2)$, let
\( \bar{x}_t = x_t - (1 - N^{-1}a)u_t , \) \hspace{1cm} (4.58)

where

\[
\bar{u}_t = \overline{\nu_t} \overline{\sigma_v} -2 \overline{\sigma_vu} , \\
\overline{\nu_t} = \overline{\nu_t} \overline{\sigma_v} -2 \overline{\sigma_vu} , \\
\overline{\sigma_v^2} = (1, - \bar{f}_{xt})\bar{\xi}(1, - \bar{f}_{xt})^t \\
= 0.0324 \left(1 + \bar{f}_{xt}^2\right)^t , \\
\overline{\sigma_vu} = (1, - \bar{f}_{xt})\bar{\xi}(0, 1)^t \\
= -0.0324 \bar{f}_{xt} , \\
\bar{f}_{xt} = \bar{\beta}_1 + 2 \bar{\beta}_2 x_t , \hspace{1cm} (4.59)
\]

\( N = 33 . \)

The modification associated with the \( \alpha \)-value is similar to that studied by Fuller (1980) for the linear errors-in-variables model. In our study, we used the value \( \alpha = 4 \).

Substituting (4.58) into the expression (4.5) defining the pseudo-maximum likelihood estimator, we obtained the first round value of
Then, replacing $X_t$ in (4.59) by $\bar{x}_t$ and $\bar{\theta}$ by the first round $\hat{\theta}_M$, a new estimate of $x^0_t$ was obtained by (4.58). The final estimate $\hat{\theta}_M$ was computed by four iterations of the procedure, successively applying (4.58) and (4.5).

The modified maximum likelihood estimator $\tilde{\theta}_M$ was computed by four iterations of the procedure, successively obtaining $\bar{x}_t$ and $\hat{\theta}_M$ by (4.58) and (4.6).

For the three adjusted estimators $\hat{\theta}_A_1$, $\hat{\theta}_A_2$, and $\hat{\theta}_A_3$, we let

$$d_1y_t = \bar{\beta}_2 \Lambda_t^{-1},$$

$$(d_2y_t, d_x^t) = \bar{\beta}_2 \Lambda_t^{-1} \sigma_v^{-2}(\sigma_v, \sigma_v),$$

$$d_3y_t = -\bar{\beta}_2 (u_t^2 - 0.0324),$$

where

$$\Lambda_t^{-1} = 0.0324 - \bar{\sigma}^2_v \sigma_v^{-2},$$

$$\sigma_v = (1, -\bar{x}_t)\xi(1, 0)' = 0.0324,$$

$$u_t = X_t - \bar{x}_t.$$
The estimate $\hat{\beta}_{A1}$ was computed by four iterations of the procedure, successively applying (4.58) and (4.5), where at each iteration observations $(y_t, x_t)$ were updated by adding $(d_t y_t, 0)$ based on the most recent $x_t$ and $\hat{\beta}_{A1}$.

The computational procedures for $\hat{\beta}_{A2}$ and $\hat{\beta}_{A3}$ are the same as that for $\hat{\beta}_{A1}$, except that the term added to observations $(y_t, x_t)$ at each iteration is $(d_{2t} y_t, d_{2t} x_t)$ for $\hat{\beta}_{A2}$ and is $(d_{3t} y_t, 0)$ for $\hat{\beta}_{A3}$.

Finally, the ordinary least squares estimator $\hat{\beta}_L$ was computed by regressing $y_t$ on $(1, x_t, x_t^2)$.

3. Results and conclusions

Our discussion concentrates on the results for estimators of $\beta_2^0$, the coefficient of the quadratic term. Table 1 contains the Monte Carlo properties of the six estimators of $\beta_2^0$.

<p>| Table 1. Monte Carlo Properties of Estimators of $\beta_2^0$ (100 samples) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean</th>
<th>Var.</th>
<th>MSE</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{M,2}$</td>
<td>1.152</td>
<td>0.633</td>
<td>0.706</td>
<td>0.506</td>
<td>0.951</td>
<td>1.561</td>
</tr>
<tr>
<td>$\hat{\beta}_{M,2}$</td>
<td>1.078</td>
<td>0.779</td>
<td>0.785</td>
<td>0.473</td>
<td>0.874</td>
<td>1.425</td>
</tr>
<tr>
<td>$\hat{\beta}_{A1,2}$</td>
<td>0.976</td>
<td>0.352</td>
<td>0.353</td>
<td>0.521</td>
<td>0.888</td>
<td>1.291</td>
</tr>
<tr>
<td>$\hat{\beta}_{A2,2}$</td>
<td>0.932</td>
<td>0.311</td>
<td>0.316</td>
<td>0.511</td>
<td>0.858</td>
<td>1.246</td>
</tr>
<tr>
<td>$\hat{\beta}_{A3,2}$</td>
<td>1.152</td>
<td>0.716</td>
<td>0.739</td>
<td>0.522</td>
<td>0.928</td>
<td>1.490</td>
</tr>
<tr>
<td>$\hat{\beta}_L$</td>
<td>0.386</td>
<td>0.063</td>
<td>0.440</td>
<td>0.231</td>
<td>0.367</td>
<td>0.529</td>
</tr>
</tbody>
</table>
The ordinary least squares estimator \( \hat{\beta}_{L,2} \) has the smallest variance, but severely underestimates \( \beta_2^0 = 1 \). The results for \( \hat{\beta}_{M,2} \) agree reasonably well with those reported for the same estimator in Wolter and Fuller (1982b).

The estimators \( \hat{\beta}_{A1,2} \) and \( \hat{\beta}_{A2,2} \) have smaller biases and smaller variances than the estimators \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), and \( \hat{\beta}_{A3,2} \). The mean squared errors of \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), and \( \hat{\beta}_{A3,2} \) are at least twice as large as those of \( \hat{\beta}_{A1,2} \) and \( \hat{\beta}_{A2,2} \).

It is somewhat surprising that \( \hat{\beta}_{M,2} \) has a smaller Monte Carlo mean squared error than \( \hat{\beta}_{M,2} \). The modification associated with \( \alpha = 4 \) resulted in considerably smaller Monte Carlo bias and mean squared error for \( \hat{\beta}_{M,2} \) than those given in Wolter and Fuller (1982b) for the maximum likelihood estimator without the \( \alpha \)-modification.

We recall that the estimators \( \hat{\beta}_{M} \) and \( \hat{\beta}_{A3} \) use a similar type of bias adjustment. The small sample behavior of \( \hat{\beta}_{M} \) and \( \hat{\beta}_{A3} \) are also similar.

The distribution of the five iterative estimators of \( \beta_2^0 \) are all positively skewed. The estimators \( \hat{\beta}_{A1,2} \) and \( \hat{\beta}_{A2,2} \) have a more symmetric distribution than the estimators \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), and \( \hat{\beta}_{A3,2} \).

The variance of the asymptotic distribution is 0.1948 for \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{A2,2} \), and \( \hat{\beta}_{A3,2} \). The Monte Carlo variances of all five iterative estimators are considerably larger than the asymptotic variance.

Statistics analogous to Student's t were also computed for \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{M,2} \), \( \hat{\beta}_{A1,2} \), \( \hat{\beta}_{A2,2} \), and \( \hat{\beta}_{A3,2} \). The covariance matrix of each
iterative estimator was estimated by the inverse of

\[
\sum_{t=1}^{33} \frac{\hat{\sigma}_{vt}^2}{\hat{\sigma}_{vt}^2} (1, x_t, x_t^2)'(1, x_t, x_t^2), \tag{4.60}
\]

where \(\hat{\sigma}_{vt}^2\) and \(x_t\) are the estimators of \(\sigma_{vt}^2\) and \(x_t\) at the fourth iteration in the computation of each estimator. The t-statistics for all five iterative estimators of \(\beta_2^0\) had negatively skewed distributions with heavy tails in comparison to the \(N(0,1)\) distribution. The sample percentiles of the statistics deviated considerably from percentiles of the \(N(0,1)\) distribution. Wolter and Fuller (1982b) introduced a modified estimator of the covariance matrix. Their Monte Carlo study showed that the modified covariance estimator produced t-statistics whose distributions agree reasonably well with the \(N(0,1)\) distribution. We conjecture that a modification similar to that of Wolter and Fuller would improve the small sample behavior of the t-statistics for \(\hat{\beta}_{A1}\), \(\hat{\beta}_{A2}\), and \(\hat{\beta}_{A3}\).

Three conclusions are possible from the results of our Monte Carlo experiment. These are:

1) The ordinary least squares estimator \(\hat{\beta}_L\) is not practical because of its large bias.
2) The adjusted estimators \(\hat{\beta}_{A1}\) and \(\hat{\beta}_{A2}\) with the \(\alpha\)-modification are recommended for practical use.
3) The inverse of the matrix (4.60) is not recommended as an estimator of the covariance matrix of \(\hat{\beta}_M\), \(\tilde{\beta}_M\), \(\hat{\beta}_{A1}\), \(\hat{\beta}_{A1}\), \(\hat{\beta}_{A2}\), or \(\hat{\beta}_{A3}\).
In this chapter, we consider the estimation of the errors-in-variables model with a nonlinear functional relationship, when there are available observations on variables outside the relationship of interest. Such an additional information enables us to estimate the unknown parameters in the relationship without knowledge of the error covariance matrix. However, as in Chapter IV, the assumption of decreasing errors in the explanatory variables seems to be necessary to obtain consistent estimators for the general nonlinear model.

A. Introduction

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 \( n = 1, 2, \ldots, \infty \), \( a_n = o(n) \), and \( b_n = o(n) \). We assume the existence of a sequence of experiments indexed by \( n \).

The functional relation of interest is

\[
y^0_t = f(x^0_t; \xi^0) + e^0_t, \quad t = 1, 2, \ldots, b_n, \quad (5.1)
\]

where \( x^0_t \) are \( 1 \times q \) vectors of fixed constants belonging to a parameter space \( \Xi \); a convex subset of \( q \)-dimensional Euclidean space, \( \xi^0 \) is a \( k \times 1 \) vector belonging to a parameter space \( \Omega \); a convex
compact subset of k-dimensional Euclidean space, \( f(z; \theta) \) is continuous on \( \Gamma \subset \Omega \), and \( e^0_t \) are errors in the equation. Our observations are \( (Y_{nt}, X_{nt}) \), where

\[
Y_{nt} = y^0_t + e^*_n, \\
X_{nt} = x^0_t + u_{nt}, \quad t = 1, 2, \ldots, b_n, \tag{5.2}
\]

and \( e^*_n \) and \( u_{nt} \) are the measurement errors. In addition, it is assumed that there are available observations

\[
\omega_{nt} = \omega^0_t + \xi_{nt}, \quad t = 1, 2, \ldots, b_n, \tag{5.3}
\]

where \( \omega^0_t \) are 1 x p vectors of constants, \( p > k \) and \( \xi_{nt} \) are random errors. The model is said to be just identified by the counting rule if \( p = k \), and to be over identified by the counting rule if \( p > k \).

The nonlinear errors-in-variables model defined above differs from the nonlinear simultaneous equation system in econometrics. If the errors-in-variables model is linear in two variables and is given by

\[
Y_t = x_t \beta + e_t, \\
X_t = x_t + u_t, \quad t = 1, 2, \ldots, n,
\]
then substituting the second equation into the first we have

\[ Y_t = X_t \beta + v_t , \quad t = 1,2,\ldots,n, \]

where the \( v_t = e_t - u_t \beta \) are independently and identically distributed with zero mean. Note that \( v_t \) is correlated with \( X_t \). It is assumed that the additional variable \( W_t \) is correlated with \( X_t \) but is uncorrelated with \( (e_t, u_t) \), and with this information \( \beta \) can be estimated consistently. The linear errors-in-variables model is equivalent to the simultaneous equation model with \( Y_t \) and \( X_t \) as endogenous variables and \( W_t \) as an exogenous variable outside the equation of interest. For another type of correspondence between the linear errors-in-variables model and the linear simultaneous equation model, obtained by treating the estimated reduced form coefficients as observations in the errors-in-variables model, see Fuller (1977) and Anderson (1976). When the model is nonlinear in the variables, the equivalence of the two models no longer holds. An equation in the nonlinear simultaneous equation system has the form

\[ Y_t = g(X_t; \beta) + v_t , \quad t = 1,2,\ldots,n, \tag{5.4} \]

where \( X_t \) is an endogenous variable, and we assume for simplicity that there is no exogenous variable in the equation. The nonlinear errors-in-variables model with \( b_n = n \) and \( a_n = 1 \) is of the form
\[ Y_t = f(x_t; \beta) + \varepsilon_t \]
\[ X_t = x_t + u_t \quad t = 1, 2, \ldots, n \quad (5.5) \]

In the model (5.4), for the two observed variables \( Y_t \) and \( X_t \), the variable \( Y_t - g(X_t; \beta) \) is independently and identically distributed with zero mean. But, in (5.5), for the observations \( Y_t \) and \( X_t \),

\[ Y_t - f(X_t; \beta) = \varepsilon_t - \frac{\partial f}{\partial x} (X_t^*; \beta) u_t \quad (5.6) \]

where \( X_t^* \) is a random variable lying on the segment joining \( x_t \) and \( X_t \), and we have assumed \( f \) is continuously differentiable. Since \( X_t^* \) and, thus, \( \frac{\partial f}{\partial x} (X_t^*; \beta) \) depend on \( u_t \) the right-hand side of (5.6) does not have zero mean unless \( \frac{\partial f}{\partial x} (\cdot; \beta) \) is constant. But if \( \frac{\partial f}{\partial x} (\cdot; \beta) \) is constant, then the model (5.5) is linear in \( X_t \).

Therefore, the nonlinear errors-in-variables model differs from the nonlinear simultaneous equation system in the sense that the relation to be estimated is the one between the unobserved true values in the former and is the one between observed values in the latter. To avoid the trivial case, we assume, from now on, that the model (5.1) is nonlinear in \( x_t^0 \).

In the model specified by (5.1), (5.2), and (5.3), we assume that \( (\varepsilon_t^0, \varepsilon_{nt}^*, u_{nt}, \xi_{nt}) \) are independently and identically distributed with mean zero and covariance matrix
Interpretations of the Assumption (5.7) are the same as those given for the model in Chapter IV. Note that we are interested in the relationship between \( y^0_t \) and \( x^0_t \), and that \( W_{nt} \) does not appear in the relationship. The \( W_{nt} \) may be measurements of \( x^0_t \) by a different measuring procedure from that used for \( X_{nt} \). Each \( W_{nt} \) is not necessarily the mean of \( a_n \) observations. Also, the term \( \epsilon_{nt} \) does not entirely consist of measurement error.

Letting

\[
e_{nt} = e^0_t + \epsilon_{nt},
\]

\[
\sigma_{ee} = \sigma^0_{ee} + \sigma^* \]

where

\[
\begin{bmatrix}
\sigma^0_{ee} & 0 & 0 & 0 \\
0 & \sigma^*_{een} & \epsilon_{eun} & \epsilon_{ern} \\
0 & \epsilon_{uen} & \epsilon_{uun} & \epsilon_{urn} \\
0 & \epsilon_{ren} & \epsilon_{run} & \epsilon_{rrn}
\end{bmatrix},
\]

\[
\begin{bmatrix}
\sigma^*_{een} & \epsilon_{eun} \\
\epsilon_{uen} & \epsilon_{uun}
\end{bmatrix} = 0(a_n^{-1}) .
\]
we can write our model as

\[ Y_{nt} = f(x_t^0, g_t^0) + e_{nt}, \]

\[ X_{nt} = x_t^0 + u_{nt}, \]

\[ W_{nt} = x_t^0 + \xi_{nt}, \quad t = 1, 2, \ldots, b_n, \]

\[
\begin{bmatrix}
    e_{nt} \\
    u_{nt}' \\
    \xi_{nt}'
\end{bmatrix} \sim \mathcal{N}
\begin{bmatrix}
    0 \\
    0 \\
    0
\end{bmatrix},
\begin{bmatrix}
    \Sigma_n & \Sigma_{ren} \\
    \Sigma_{ren}' & \Sigma_{rrn}
\end{bmatrix}, \quad (5.8)
\]

where

\[ \Sigma_n = \begin{bmatrix} \sigma_{een} & \xi_{eun} \\ \xi_{uen} & \xi_{uun} \end{bmatrix}, \]

\[ \Sigma_{ren} = (\xi_{ren}, \xi_{run}). \]

We assume that \( b_n \) is known, but that \( a_n \) is unobservable. That is, we can use the value \( b_n \) in our inference, but not the actual value of \( a_n \).

If the error in the relationship exists, then \( \sigma_{ee}^2 > 0 \), and thus \( \sigma_{een} = 0(1) \). If the relationship is perfect, then \( \sigma_{een} = O(a_n^{-1}) \).
This difference introduces some difficulty in deriving asymptotic results. When the errors $e_t^0, \ t = 1,2, \ldots, t_n,$ are present, $b_n^{1/2}$ is an obvious choice for the normalizing constant to obtain a limiting distribution of an estimator. However, the contribution of the $b_n$ measurement errors each with variance of $O(a_n^{-1})$ is $n^{-1/2}$. Hence, the measurement error contribution would disappear in the limiting results, if the normalizing constant $b_n^{1/2}$ was used. But, it is desirable to keep the effect of measurement errors in the limiting results. Also, we are interested in the asymptotic results with the normalizing constant $n^{1/2}$. Therefore, for the purpose of investigating the limiting behavior of estimators, the model with errors in the relationship is of little interest. Hence, from now on, we assume that there are no errors in the relationship. That is, we assume $\sigma_{e e} = O(a_n^{-1})$. This assumption does not necessarily restrict the range of applications of our asymptotic results. As we discussed in Chapter IV, the asymptotic results with normalizing constant $n^{1/2}$ are applicable when either measurement errors are small or the number $b_n$ of data points is large. Thus, we may apply our results to the case where the error in the relationship and the measurement errors are considered to be small and the number of observations is large. Hence, we assume that the term $e_{nt}$ includes the error in the relationship and has a variance of $O(a_n^{-1})$. This is an unpleasing aspect of the nonlinear model. We recall that in the linear model the instrumental variable estimation is useful regardless of the existence of errors in the equation. The instrumental variable estimator of $\beta_0$ which will be defined in the
next section can be shown to be consistent for $\hat{\beta}^0$ under appropriate assumptions, even when $\sigma_{\text{een}} = O(1)$, i.e., when an error in the relationship exists. However, the properties other than consistency do not hold if $\sigma_{\text{een}} = O(1)$, since the orders of the error in the equation and the measurement errors are different.

Some care is required in presenting the assumptions on $E_{\text{rrn}}$. As we noted earlier, the variable $W_{nt}$ does not appear in the equation of interest. Thus, multiplying $W_{nt}$ by a constant, say $b_n^{-1}$, does not change the relationship. Also, as we will see, our estimation procedure is invariant under this type of transformation of $W_{nt}$. Hence, there is some arbitrariness associated with the order of $W_{nt}$. We note that the true values $\chi_t^0$ are fixed. As in the linear case, for an instrumental variable $W_{nt}$ to produce a consistent estimator, the random part $\epsilon_{nt}$ with zero mean cannot dominate the fixed part which helps the estimation procedure. This condition may be written as

$$T_n' E_{\text{rrn}} T_n = O(1), \quad (5.9)$$

where $T_n$ satisfies

$$T_n' \left( b_n^{-1} \frac{b_n}{n} \sum_{t=1}^{n} W_t^0 W_t^0 \right) T_n = I.$$ 

Under the Assumption (5.9), the order of $b_n^{-1} \sum_{t=1}^{n} W_t' W_{nt}$ is the same as
the order of

\[ b_n \sum_{t=1}^{t_n-1} \frac{w_t^r w_t^s}{m_{ww}}. \]

To avoid the ambiguity due to the arbitrariness of the order of \( W_{nt} \) and to simplify our discussion, we assume

\[ \lim_{n \to \infty} b_n \sum_{t=1}^{t_n-1} \frac{w_t^r w_t^s}{m_{ww}} = m_{ww}, \quad (5.10) \]

where \( m_{ww} \) is positive definite. With the Assumption (5.10), the condition (5.9) is equivalent to assuming that the order of \( \Sigma_{rrn} \) is at most one. If \( W_{nt} \) can be considered as a mean of \( a_n \) observations, or if \( Z_{nt} \) are measurement errors, then \( \Sigma_{rrn} = O(a_n^{-1}) \). Also, if \( W_{nt} \) have no errors, then \( \Sigma_{rrn} = 0 \), and the condition \( \Sigma_{rrn} = O(1) \) still holds. Thus, the result derived using the order \( \Sigma_{rrn} = O(1) \) is also valid for the case where either \( \Sigma_{rrn} = O(a_n^{-1}) \) or \( \Sigma_{rrn} = O(1) \). Note also that if \( \Sigma_n = O(a_n^{-1}) \) and \( \Sigma_{rrn} = O(1) \), then \( \Sigma_{rrn} = O(a_n^{-1/2}) \).

In the analysis of the linear instrumental variable model, it has often been assumed, without loss of generality, that

\[ b_n \sum_{t=1}^{t_n-1} \frac{W_t^r W_t^s}{W_{nt} W_{nt}} = I. \quad (5.11) \]

If \( W_{nt} \) are independent of \( (e_n, u_n) \), we can always make a transformation to satisfy (5.11) without altering the model assumption. But, the assumption of zero correlation between \( (e_n, u_n) \) and \( \Sigma_{nt} \).
without independence may be destroyed by a transformation used to obtain (5.11). It is still possible to use the form (5.11) to simplify the computations. We do not assume that the condition (5.11) holds.

We will use the same notations as those in Chapter IV to denote the partial derivatives of the function \( f(x; \theta) \). Thus, for example, \( \frac{\partial^2 f}{\partial \theta \partial x} (z; \theta) \) is the matrix of second partial derivatives with respect to the elements of \( \theta \) evaluated at \((z; \theta)\), and \( \frac{\partial f}{\partial x} \) is the partial derivative with respect to the elements of \( x \) evaluated at \((x^0_t; \theta^0)\).

In this chapter, the order, or the order in probability statements are taken as \( n \) tends to infinity, unless specified otherwise.

B. The Instrumental Variable Estimator

1. The estimator

We introduce an estimator of \( \theta^0 \) in the model (5.8) which does not require the knowledge of the error covariance matrix. Our estimator can be considered as a version of the two-stage least squares estimator. In the context of the nonlinear simultaneous equation model, T. Amemiya (1974) introduced the nonlinear two-stage least squares estimator. His model is

\[
Y_t = f(z_t; \theta^0) + \varepsilon_t \quad t = 1, 2, \ldots, n,
\]

where \( z_t \) consists partly of endogenous variables and partly of exogenous variables. Given the vector \( x_t \) of exogenous variables in
the system T. Amemiya defined the nonlinear two-stage least squares estimator of $\bar{\theta}^0$ to be the value of $\bar{\theta}$ that minimizes

$$
\phi(\bar{\theta}) = \left\{ \sum_{t=1}^{n} [Y_t - f(\xi_t; \bar{\theta})] \sigma_t \right\} \left\{ \sum_{t=1}^{n} \sigma_t \sigma_t' \right\}^{-1} \left\{ \sum_{t=1}^{n} \sigma_t' [Y_t - f(\xi_t; \bar{\theta})] \right\}.
$$

It is easy to see that for the linear simultaneous equation model the estimator minimizing $\phi(\bar{\theta})$ reduces to the usual two-stage least squares estimator.

As we discussed earlier, the nonlinear errors-in-variables model is different from the nonlinear simultaneous equation model. In our nonlinear functional relationship model, $Y_{nt} - f(X_{nt}; \bar{\theta}^0)$ is not an identically distributed random variable with zero mean. However, we minimize a function analogous to $\phi(\bar{\theta})$ to obtain an estimator of $\bar{\theta}^0$ in the model (5.8).

We define the instrumental variable estimator $\hat{\theta}_n$ of $\bar{\theta}^0$ to be the value of $\bar{\theta}$ in $\Theta$ that minimizes

$$
Q_n(\theta) = \left\{ \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \theta)] \omega_{nt} \right\} \left\{ \sum_{t=1}^{b_n} \omega_{nt} \omega_{nt}' \right\}^{-1} \left\{ \sum_{t=1}^{b_n} \omega_{nt}' [Y_{nt} - f(X_{nt}; \theta)] \right\}.
$$

(5.12)

It is to be understood that in (5.12) and in the ensuing development $X_{nt}$ is replaced by its projection onto the space $\Gamma$ whenever $X_{nt}$ is outside of $\Gamma$. 
2. Consistency

In this section, we derive the consistency of the estimator \( \hat{\theta}_n \) of \( \theta^0 \) in the model (5.8). We introduce the following assumptions.

**Assumption 5.1.** The model (5.8) holds with \( \Xi_n = O(a_n^{-1}) \).

**Assumption 5.2.** \( a_n = o(n) \) and \( b_n = o(n) \).

**Assumption 5.3.** \( \lim_{n \to \infty} \Xi_{nn} = \Xi \).

**Assumption 5.4.** \( \lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} \xi_t' \xi_t = \Xi_{\omega \omega} \), where \( \Xi_{\omega \omega} \) is positive definite.

**Assumption 5.5.** The \( \beta \)-parameter space \( \Omega \) is a compact convex subset of \( k \)-dimensional Euclidean space.

**Assumption 5.6.** The \( x \)-parameter space \( \Gamma \) is a convex subset of \( q \)-dimensional Euclidean space.

**Assumption 5.7.** The partial derivatives of order one of \( f(x; \theta) \) with respect to \( x \) and \( \theta \) exist and are continuous on \( \Gamma \times \Omega \).

**Assumption 5.8.** There exists a constant \( K_1 \) such that for all \( x \) in \( \Gamma \) and all \( \theta \) in \( \Omega \) the absolute value of each element of \( \xi_x(x; \theta) \) is bounded by \( K_1 \).

**Assumption 5.9.** Uniformly for all \( \theta \) in \( \Omega \),

\[
\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(x_t^0; \theta) [f_{\beta}(x_t^0; \theta)]' = \Xi_{\omega \beta} \theta.
\]

**Assumption 5.10.** Uniformly for all \( \theta \) in \( \Omega \),

\[
\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} \xi_{\theta_t} \xi_{\theta_t} = \Xi_{\theta \theta} \theta.
\]
where the rank of the \( k \times p \) matrix \( b(\theta) \) is \( k \) for all \( \theta \) in \( \Theta \).

Note that by Lemma 2 of Jennrich (1969), under Assumptions 5.5 and 5.7, a measurable function \( \hat{\theta}_n \) of \( Y_{nt} \), \( X_{nt} \) and \( \omega_{nt} \) which minimizes (5.12) always exists. The following theorem shows the consistency of \( \hat{\theta}_n \).

**Theorem 5.1.** Let Assumptions 5.1 through 5.10 hold. Then, as \( n \to \infty \),

\[
\hat{\theta}_n \xrightarrow{P} \theta^0 .
\]

**Proof.** By Assumption 5.7 and Lemma 3 of Jennrich (1969), for any \( \theta \) in \( \Theta \),

\[
Y_{nt} - f(X_{nt}; \theta) = e_{nt} + f(\theta^0; \theta) - f(X_{nt}; \theta)
\]

\[
= e_{nt} + f(\theta^0; \theta) - f(\theta^0; \theta) - f(X_{nt}; \theta)
\]

\[
= e_{nt} - f'(\theta^0; \theta^*) (\theta^* - \theta^0) - f_x(\theta^0; \theta)\omega_{nt}
\]

where \( X^*_{nt} \) is on the segment joining \( X^0_t \) and \( X_{nt} \), and \( \theta^* \) is on the segment joining \( \theta^0 \) and \( \theta \). Note that by Assumptions 5.5 and 5.6, \( \mathcal{F} \) and \( \Theta \) are convex, and thus \( X^*_{nt} \) is in \( \mathcal{F} \) and \( \theta^* \) is in \( \Theta \). We have

\[
b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \theta)]\omega_{nt} = b_n^{-1} \sum_{t=1}^{b_n} e_{nt}(\theta^0 + X_{nt})
\]
By Assumptions 5.1 and 5.4, and by the Chebyshev's inequality, as $n \to \infty$,

$$
\frac{b_n}{n} \sum_{t=1}^{b_n} \xi_t (x^*_t; \theta) u^*_t (y^0_t + r_{nt}) 
\to \sum_{t=1}^{b_n} \xi_t (x^*_t; \theta) (y^0_t + r_{nt}) .
$$

(5.13)

Also, by Assumptions 5.1 and 5.3, and by the Cauchy-Schwarz inequality,

$$
\frac{b_n}{n} \sum_{t=1}^{b_n} e_{nt} \frac{y^0_t}{\sqrt{b_n}} \xrightarrow{p} 0 .
$$

(5.14)

Also, by Assumptions 5.1 and 5.3, and by the Cauchy-Schwarz inequality,

$$
E\left\{ \left| \frac{b_n}{n} \sum_{t=1}^{b_n} e_{nt} r_{nt} \right| \right\} \leq \frac{b_n}{n} \sum_{t=1}^{b_n} \left[ E\left\{ e_{nt}^2 \right\} E\left\{ r_{nt}^2 \right\} \right]^{1/2}
\leq O(a_n^{-1/2}) ,
$$

where the norm is the Euclidean norm. Thus, by Assumption 5.2, as $n \to \infty$,
\[ b_n^{-1} \sum_{t=1}^{n} e_{nt} X_{nt} \xrightarrow{p} 0 \]  

Let \( f_{x_{nt}; \beta} \), \( u_{nti} \), \( w_{ti}^0 \) and \( r_{nti} \) be the \( i \)-th elements of \( f_{x_{nt}; \beta} \), \( u_{nt} \), \( w_0^t \) and \( r_{nt} \), respectively. Also, let 
\[ \sigma_{u_{nti}} = \mathbb{V}[u_{nti}], \quad \sigma_{r_{nti}} = \mathbb{V}[r_{nti}] \]  

Then, by Assumptions 5.1, 5.3, 5.4, and 5.8, and the Cauchy-Schwarz inequality, for \( i = 1, 2, \ldots, p \), 

\[ \mathbb{E}\{ b_n^{-1} \sum_{t=1}^{n} f_{x_{nt}; \beta} u_{nti} (w_{ti}^0 + r_{nti}) \} \]

\[ < b_n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{q} \mathbb{E}\{ |f_{x_{nt}; \beta} u_{nti} w_{ti}^0| + |f_{x_{nt}; \beta} u_{nti} r_{nti}| \} \]

\[ < b_n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{q} \frac{1}{2} \left( \mathbb{E}[u_{nti}^2]^{1/2} (|w_{ti}^0| + |w_{ti}^0|) + \mathbb{E}[r_{nti}^2]^{1/2} \right) \]

\[ = \frac{1}{2} \left( \sum_{i=1}^{q} \sigma_{u_{nti}} + b_n^{-1} \sum_{t=1}^{n} |w_{ti}^0| \right) \]

\[ < \frac{1}{2} \left( \sum_{i=1}^{q} \sigma_{u_{nti}} + (b_n^{-1} \sum_{t=1}^{n} w_{ti}^0)^{1/2} \right) \]
\[ = M_1 a_n^{-1/2} \], \hspace{1cm} (5.16)\]

where \( M_1 \) is some constant. Therefore, by Assumption 5.2, for every \( \varepsilon > 0 \), there exists an \( N_1(\varepsilon) \) such that if \( n > N_1(\varepsilon) \) for all \( \xi \) in \( \xi \),

\[ \Pr\left \{ \left| b_n^{-1} \sum_{t=1}^{b_n} f(x_{nt};\xi) u_{nt} W_{nt} \right| > \varepsilon \right \} < \varepsilon \], \hspace{1cm} (5.17)\]

where the norm is the Euclidean norm. By Assumption 5.10, uniformly for all \( \xi \) in \( \xi \),

\[ b_n^{-1} \sum_{t=1}^{b_n} f_\xi(x_t^0;\xi) x_t^0 \rightarrow h(\xi^*) \]. \hspace{1cm} (5.18)\]

Let \( m_{\beta ij}(\xi) \) be the \((i,j)\)-th element of \( m_{\beta \xi}(\xi) \), and let \( f_{\beta i}(x_t^0;\xi^*) \) be the \(i\)-th element of \( f_{\beta}(x_t^0;\xi^*) \). By Assumptions 5.7 and 5.9, \( m_{\beta ij}(\xi) \) are continuous on \( \xi \), since \( m_{\beta ij}(\xi) \) are the uniform limits of continuous functions. Thus, by Assumption 5.5 there exists an \( M_2 \) such that for \( i,j = 1,2,\ldots,k \), and for all \( \xi \) in \( \xi \),

\[ m_{\beta ij}(\xi) < M_2 < \infty \]. \hspace{1cm} (5.19)\]
It follows from (5.19) and Assumption 5.3 that for $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, p$ for all $\xi$ in $\mathcal{G}$, and for some constant $M_3 < \infty$,

$$E[b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(x^0_t, \xi^*)_t\xi_{ntj}] = 0,$$

(5.20)

Thus, there exists an $N_2(\varepsilon)$ such that for all $\xi$ in $\mathcal{G}$,

$$\mathbb{P}\left[ b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(x^0_t, \xi^*)_t\xi_{ntj} > \varepsilon \right] < \varepsilon.$$

(5.21)

Also, by Assumption 5.5,

$$|\xi - \xi^0| < M_4$$

(5.22)

for all $\xi$ in $\mathcal{G}$ and some constant $M_4 < \infty$. It follows from Assumptions 5.2, 5.3, and 5.4 that
where $\overline{m}_{WW} = \overline{m}_{WW} + \Sigma_{rr}$ is positive definite.

Therefore, by (5.13), (5.14), (5.15), (5.17), (5.18), (5.21), (5.22) and (5.23), for every $\varepsilon > 0$ there exists an $N_3(\varepsilon)$ such that if $n > N_3(\varepsilon)$, then for all $\hat{g}$ in $\mathcal{G}$

$$P\left[|Q_n(\hat{g}) - (\hat{g} - \hat{g}^0)'h(\hat{g}^*)\overline{m}_{WW}^{-1}[h(\hat{g}^*)]'(\hat{g} - \hat{g}^0)| > \varepsilon \right] < \varepsilon , \quad (5.24)$$

where $\hat{g}^*$ is on the segment joining $\hat{g}^0$ and $\hat{g}$. Since $\hat{g}_n$ is in $\mathcal{G}$ for all realizations, it follows from (5.24) that if $n > N_3(\varepsilon)$ then

$$P\left[|Q_n(\hat{g}_n) - (\hat{g}_n - \hat{g}^0)'h(\hat{g}_n^*)\overline{m}_{WW}^{-1}[h(\hat{g}_n^*)]'(\hat{g}_n - \hat{g}^0)| > \varepsilon \right] < \varepsilon , \quad (5.25)$$

where $\hat{g}_n^*$ is on the segment joining $\hat{g}^0$ and $\hat{g}_n$. By Assumptions 5.4, 5.5, 5.7, and 5.10, and by the argument used to obtain (5.19), there exists an $M_5 > 0$ such that for all $\hat{g}$ in $\mathcal{G}$ the smallest eigenvalue $\lambda_m(\hat{g})$ of $h(\hat{g})\overline{m}_{WW}^{-1}[h(\hat{g})]'$ satisfies

$$\lambda_m(\hat{g}) > M_5 . \quad (5.26)$$

By (5.24), as $n \to \infty$,
Since \( \hat{\beta}_n \) minimizes \( Q_n(\beta) \) over \( \beta \),

\[
Q_n(\hat{\beta}_n) < Q_n(\beta_0) \tag{5.28}
\]

It follows from (5.27) and (5.28) that there is an \( N_4(\varepsilon) \) such that if \( n > N_4(\varepsilon) \), then

\[
P\{ |Q_n(\hat{\beta}_n)| > \varepsilon \} < P\{ |Q_n(\beta_0)| > \varepsilon \} < \varepsilon \tag{5.29}
\]

By (5.25), (5.26), and (5.29), if \( n > \max\{N_3(\varepsilon), N_4(\varepsilon)\} \), then

\[
P\{ |\hat{\beta}_n - \beta_0|^2 > 2M_5^{-1}\varepsilon \} < P\{ (\hat{\beta}_n - \beta_0) \bar{h}(\beta_0) \hat{m}_{WW}^{-1} [\hat{h}(\beta_0)]'(\hat{\beta}_n - \beta) > 2\varepsilon \}
\]

\[
< P\{ |(\hat{\beta}_n - \beta_0) \bar{h}(\beta_0) \hat{m}_{WW}^{-1} [\hat{h}(\beta_0)]'(\hat{\beta}_n - \beta_0) - Q_n(\hat{\beta}_n) | > \varepsilon \}
\]

\[
+ P\{ |Q_n(\hat{\beta}_n)| > \varepsilon \}
\]

\[
< 2\varepsilon .
\]

Thus, the result holds. \(\square\)

Note that in Theorem 5.1 we did not use any relation between \((e_{nt}, \xi_{nt})\) and \(\xi_{nt}\) such as zero correlation or independence. We
recall that in the linear model instrumental variables have to be uncorrelated with errors in the variables contained in the equation to produce a consistent estimator. However, in our model, the errors \( (e_{nt}, u_{nt}) \) are decreasing, and thus the correlation between \( (e_{nt}, u_{nt}) \) and \( \xi_{nt} \) also tends to zero. This is why we were able to obtain the consistency of \( \hat{\xi}_n \) without an explicit assumption of the zero correlation.

Also, Theorem 5.1 is applicable to both the case with \( \Sigma_{rrn} = O(1) \) and the case with \( \Sigma_{rrn} = O(a^{-1}) \), as long as Assumption 5.3 holds.

As we mentioned earlier, the consistency of \( \hat{\xi}_n \) can be proved for the case with \( \sigma_{een} = O(1) \) under an additional assumption of zero correlation between \( e_{nt} \) and \( \xi_{nt} \). The zero correlation is used only to obtain (5.15). The rest of the proof is exactly the same as the proof of Theorem 5.1.

Under the assumption that the variances of \( (e_{nt}, u_{nt}) \) decrease, the ordinary least squares estimator \( \hat{\xi}_L \) obtained by minimizing

\[
\frac{b_n}{\sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \beta)]^2}
\]

is also consistent for \( \xi^0 \). Wolter and Fuller (1982b) showed that under certain assumptions including that \( \Sigma_n = O(a^{-1}) \),

\[
\hat{\xi}_L - \xi^0 = o_p(\max [n^{-1/2}, a^{-1}])
\]
3. Limiting distribution

We introduce further assumptions.

Assumption 5.11. \( \theta^0 \) is an interior point of \( \Theta \).

Assumption 5.12. \( f_{\theta^0}(z; \theta) \) and \( f_{\theta^0}'(z; \theta) \) exist and are continuous on \( \Theta \).

Assumption 5.13. There exist constants \( K_2 \) and \( K_3 \) such that uniformly for all \( z \) in \( \mathcal{Z} \) and all \( \theta \) in \( \Theta \) the absolute value of each element of \( f_{\theta^0}(z; \theta) \) is bounded by \( K_2 \) and the absolute value of each element of \( f_{\theta^0}'(z; \theta) \) is bounded by \( K_3 \).

Later, we will use different sets of assumptions to derive the limiting distribution, the order, and the bias expression of \( \hat{\theta}_n \). A term playing an important role in the derivations is

\[
-\frac{1}{b^n} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \theta^0)] W_{nt}.
\]

The order of this term varies under different assumptions. Hence, it is convenient to establish an expansion of \( \hat{\theta}_n \) in terms of this quantity. The following Lemma 5.1 provides such an expansion, and will be used repeatedly in later discussions.

Lemma 5.1. Let Assumptions 5.1 through 5.13 hold. Assume that

\[
-\frac{1}{b^n} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \theta^0)] W_{nt} = O_p(c_n^{-1}), \quad (5.30)
\]

where \( c_n^{-1} = o(1) \). Then,
\[ \hat{\beta}_n - \beta^0 = \{h^0 E_{WW}^{-1} h^0^\prime\}^{-1} h^0 E_{WW}^{-1} \left[ b_n \sum_{t=1}^{b_n} W_{nt}^\prime \{Y_{nt} - f(X_{nt}; \hat{\beta}_n^0)\} \right] + o_p(c_n^{-1}), \]

where

\[ h^0 = h(\beta^0), \]

\[ E_{WW} = E_{WW} + \xi_{rr}. \]

and \( h(\beta) \) is defined in Assumption 5.10.

**Proof.** Let \( f_{\beta_i}(z; \beta) \) be the \( i \)-th element of \( f_{\beta}(z; \beta) \) and let

\( f_{\beta_i}(z; \beta) \) be the \( i \)-th row of \( f_{\beta_{\beta_i}}(z; \beta) \). Also let

\[ E_{WW} = b_n^{-1} \sum_{t=1}^{b_n} W_{nt}^\prime W_{nt}. \]

By Assumptions 5.5, 5.6, and 5.12, and by Lemma 3 of Jennrich (1969), for \( i = 1, 2, \ldots, k, \)

\[ \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \hat{\beta}_n)] W_{nt} \}_{E_{WW}}^{-1} \{b_n^{-1} \sum_{s=1}^{b_n} W_{ns} f_{\beta_{\beta_i}}(X_{ns}; \hat{\beta}_n) \} \]
where $\xi_n^*$ is on the segment joining $\xi^0$ and $\hat{\xi}_n$. The estimator $\hat{\xi}_n$ minimizes $Q_n(\xi)$ in (5.12) over $\Theta$, and $Q_n(\xi)$ is differentiable on $\Theta$ by Assumption 5.8. Thus, if $\hat{\xi}_n$ is an interior point of $\Theta$, then

$$
\frac{\partial Q_n(\hat{\xi}_n)}{\partial \xi} = -2 \left[ b_n^{-1} \sum_{t=1}^{b_n} f(\xi_{nt};\hat{\xi}_n) \tilde{W}_{nt} \right]^{-1} \left[ b_n^{-1} \sum_{s=1}^{b_n} \tilde{W}_{ns} f(\xi_{ns};\hat{\xi}_n) \right] .
$$

By Assumption 5.11, there is a $\delta > 0$ such that if $|\xi - \xi^0| < \delta$, then $\xi$ is an interior point of $\Theta$. Since $\hat{\xi}_n \overset{P}{\longrightarrow} \xi^0$ by Theorem 5.1, for every $\varepsilon > 0$ given, there is an $N_1$ such that if $n > N_1$, then
By (5.32) and (5.33), for every $\epsilon > 0$ and every $\alpha > 0$, if $n > N_1$, then

$$p\left(\frac{\hat{a}_n - \hat{b}}{\hat{a}} < \epsilon \right) \geq 1 - \epsilon \ .$$

(5.33)

Hence, for every $\alpha > 0$,

$$\left\{ b^{-1} \sum_{t=1}^{b_n} f(x_{nt}; \hat{g}_n) \right\}^{m_{WW}} \left\{ b^{-1} \sum_{t=1}^{b_n} [Y_{ns} - f(x_{ns}; \hat{g}_n)] \right\} \rightarrow o_p(n^{-\alpha}) \ .$$

(5.34)

It follows from (5.24) in the proof of Theorem 5.1 that

$$Q_n(g^*_n) = (g^*_n - g^0)' \hat{h}(g^{**}) \hat{h}(g^0) = o_p(1) \ .$$

(5.35)

where $g^{**}$ is on the segment joining $g^0$ and $g^*$. Since by Theorem 5.1 $g^*_n = g^0 + o_p(1)$, we obtain from (5.26) and (5.35) that

$$Q_n(g^*_n) = o_p(1) \ .$$
Hence, by (5.23) and the definition (5.12) of \( Q_n(\xi) \),

\[
\frac{b_n}{n} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \xi^*_n)]W_{nt} = o_p(1). \tag{5.36}
\]

Letting \( W_{nm} \) be the \( m \)-th element of \( W_{ns} \) and \( f_{\beta \beta ij}(X_{ns}; \xi^*_n) \) be the \((i,j)\)-th element of \( f_{\beta \beta}(X_{ns}; \xi^*_n) \), we have by Assumptions 5.4 and 5.13,

\[
\left| \frac{b_n}{n} \sum_{s=1}^{b_n} W_{ns} f_{\beta \beta ij}(X_{ns}; \xi^*_n) \right| < \frac{b_n}{n} \sum_{s=1}^{b_n} |W_{ns}| K_2
\]

\[
< K_2 \left( \frac{b_n}{n} \sum_{s=1}^{b_n} W_{nm}^2 \right)^{1/2}
\]

\[
= K_2.
\]

Thus,

\[
\frac{b_n}{n} \sum_{s=1}^{b_n} W_{ns} f_{\beta \beta i}(X_{ns}; \xi^0) = o_p(1). \tag{5.37}
\]

By Assumption 5.12,
where $\xi_{nt}$ is on the segment joining $\xi_t^0$ and $\xi_{nt}$, and $f_{\beta x i}$ is the $i$-th row of $f_{\beta x}$. By Assumptions 5.4 and 5.13, and the argument used in (5.16),

$$
b_n^{-1} \sum_{t=1}^{b_n} f_{\beta x i}(\xi_{nt}^0; \xi_{nt}) \xi_{nt} u_{nt} W_n = 0 (a_n^{-1/2}) . \quad (5.39)$$

By (5.20) in the proof of Theorem 5.1,

$$
b_n^{-1} \sum_{t=1}^{b_n} f_{\beta i}(\xi_t^0; \xi_{nt}) \xi_{nt} = 0 (a_n^{-1/2}) . \quad (5.40)$$

By Assumption 5.10, for every $\varepsilon > 0$, there exists an $N_1(\varepsilon)$ such that if $n > N_1(\varepsilon)$, then for all $\xi$ in $\Omega$,

$$
|\text{vec}[b_n^{-1} \sum_{t=1}^{b_n} f_{\beta}(\xi_t^0; \xi_t) \xi_t^0 - b(\xi)]| < 2^{-1} \varepsilon ,
$$
and, thus,

\[ |\text{vec}\left( b_n^{-1} \sum_{t=1}^{b_n} f(\mathbf{x}_t^0; g^*_t) \mathbf{w}_t^0 - h(g^*) \right)| < 2^{-1} \varepsilon, \quad (5.41) \]

where we have used the convexity of \( g \). Since \( h(g) \) is the uniform limit of continuous functions, \( h(g) \) is continuous. Since

\[ g^*_n \xrightarrow{p} g^0 \]

by Theorem 5.1, there exists an \( N_2(\varepsilon) \) such that if

\[ n > N_2(\varepsilon), \]

then

\[ P\left( |h(g^*) - h(g^0)| > 2^{-1} \varepsilon \right) < \varepsilon. \quad (5.42) \]

By (5.41) and (5.42), if \( n > \max[N_1(\varepsilon), N_2(\varepsilon)] \), then

\[ P\left( \left| \text{vec}\left( b_n^{-1} \sum_{t=1}^{b_n} f(\mathbf{x}_t^0; g^*_t) \mathbf{w}_t^0 - h(g^0) \right) \right| > \varepsilon \right) \]

\[ < P\left( \left| b_n^{-1} \sum_{t=1}^{b_n} f(\mathbf{x}_t^0; g^*_t) \mathbf{w}_t^0 - h(g^0) \right| > 2^{-1} \varepsilon \right) + P\left( |h(g^*) - h(g^0)| > 2^{-1} \varepsilon \right) \]

\[ < \varepsilon. \]

Therefore,
By (5.38), (5.39), (5.40), and (5.43),

\[ b_n^{-1} \sum_{t=1}^{b_n} f_\theta (\xi_t; \xi_t^0) \xi_t^0 = h(\xi^0) + o_p(1). \]  

(5.44)

Similarly,

\[ b_n^{-1} \sum_{t=1}^{b_n} f_\theta (\xi_n^0; \xi_n^0) \xi_n^0 = h^0 + o_p(1). \]  

(5.45)

Note that by Assumptions 5.4 and 5.10, \( h^0 \) is nonsingular.

Hence, by (5.23), (5.31), (5.34), (5.36), (5.37), (5.44), (5.45), and Theorem 5.1, the result holds.

Note that if \( \xi_{rrn} = 0(a_n^{-1}) \) in Lemma 5.1 then \( \xi_{rr} = 0 \) and thus \( \bar{m}_{WW} = \bar{m}_{WW} \).

In the linear instrumental variable estimation, it is assumed that the instrumental variables are either independent of, or uncorrelated with, errors in the variables contained in the relationship. In the
derivation of the limiting distribution of \( \hat{\beta}_n \), the argument is somewhat simpler if \( \xi_{nt} \) are independent of \((e_{nt}, u_{nt})\). In practice, instrumental variables are often measured independently of other variables, and thus the assumption of independent errors holds. However, there are situations where we are not willing to assume the independence of \( \xi_{nt} \) and \((e_{nt}, u_{nt})\). Without the independence, some assumptions on the higher moments of \( \xi_{nt} \) and \((e_{nt}, u_{nt})\) are required to obtain the limiting distribution of \( \hat{\beta}_n \).

We introduce assumptions which will be used to derive the limiting distribution of \( \hat{\beta}_n \).

**Assumption 5.14.** \( f_{XX}(z; \theta) \) exists and is continuous on \( \Gamma \times \Omega \).

**Assumption 5.15.** There exists a constant \( K_4 \) such that the absolute value of each element of \( f_{XX}(z; \theta) \) is bounded by \( K_4 \) for all \( z \) in \( \Gamma \) and all \( \theta \) in \( \Omega \).

**Assumption 5.16.** Let \( w_0^{t_1} \) and \( f_{x_1}(x_0^{t_1}; \theta_0) \) be the \( i \)-th elements of \( w_0^{t_1} \) and \( f_{x_1}(x_0^{t_1}; \theta_0) \), respectively. Then, for all \( i_1, i_2 = 1, 2, \ldots, p \), and for all \( j_1, j_2 = 1, 2, \ldots, q \),

\[
\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} w_0^{t_1} w_0^{t_2} f_{x_1}(x_0^{t_1}; \theta_0)^{-1} f_{x_2}(x_0^{t_2}; \theta_0)^{-1} = M_4(i_1, i_2, j_1, j_2).
\]

Also,

\[
\lim_{n \to \infty} a_n \sum_{n-\infty} = 0.
\]

**Assumption 5.17.** \( E\left[ \frac{1}{n} (e_{nt}, u_{nt})^4 \right] = O(1) \).
Assumption 5.18. \( E\{\left| r_{nt} \right|^4 \} = O(1) \). Also, for \( i = 1, 2, \ldots, p, \)

\[
\lim_{n \to \infty} a_n V\{r_{nt}, (e_{nt}, u_{nt})'\} = \Psi_i.
\]

The next theorem gives the limiting distribution of \( \hat{\theta}_n \) with the normalizing constant \( \frac{1}{\sqrt{n}} \).

**Theorem 5.2.** Let Assumptions 5.1 through 5.16 hold. Assume that \( \sigma = o(n^{1/2}) \). (5.46)

Also, assume either

(a) \((e_{nt}, u_{nt})\) and \( \xi_{nt} \) are independent, or

(b) \((e_{nt}, u_{nt})\) and \( \xi_{nt} \) are uncorrelated, and Assumptions 5.17 and 5.18 hold.

Then,

\[
\frac{1}{\sqrt{n}} (\hat{\theta}_n - \theta^0) \xrightarrow{L} N(\Psi, \Sigma),
\]

where

\[
\Psi = \left\{ h^0 \frac{\Sigma_{WW}}{\Sigma_{WW}} h^0 \right\}^{-1} h^0 \Sigma_{WW} \frac{\Sigma_{WW}}{\Sigma_{WW}} h^0, \quad \left\{ h^0 \frac{\Sigma_{WW}}{\Sigma_{WW}} h^0 \right\}^{-1},
\]

and for case (a),
\[ V_0 = \lim_{n \to \infty} b_n^{1/2} \sum_{t=1}^n a_{nt} v_{nt} W_{nt}^{2}, \]

\[ \sigma^2_{\text{vnt}} = [1, -x_{xt}^0] E_n [1, -x_{xt}^0]' , \]

and for case (b),

\[ V_0 = \lim_{n \to \infty} b_n^{1/2} \sum_{t=1}^n V\{ \tilde{W}_{nt}^{2} [1, -f_{x}(x_{t}^0; \theta^0)] a_{nt}^{1/2} (e_{nt}', u_{nt}') \} . \]

**Proof.** By Assumption 5.14,

\[ Y_{nt} = f(x_{nt}; \theta^0) = e_{nt} - f_{x}(x_{t}^0; \theta^0) u_{nt}' - \frac{1}{2} u_{nt}^2 f_{xx}(x_{nt}; \theta^0) u_{nt}' , \]

(5.47)

where \( x_{nt}^* \) is on the segment joining \( x_{t}^0 \) and \( x_{nt} \). Let \( f_{xxij} \) be the \((i,j)\)-th element of \( f_{xx} \), let \( u_{nti} \) be the \(i\)-th element of \( u_{nt} \), and let \( (\tilde{w}_{t}^0 + r_{ntm}) \) be the \(m\)-th element of \( \tilde{W}_{nt} = w_{t}^0 + r_{nt} \). Then, for either the case (a) or (b), and for \( i,j = 1,2,\ldots,q, \) and \( m = 1,2,\ldots,p, \)

\[ E \{ b_n^{-1} \sum_{t=1}^n f_{xxij}(x_{nt}^*; \theta^0) u_{nti} u_{ntj} w_{t}^0 \} . \]
\[ K_4 \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \]

\[ \leq K_4 \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \]

Thus,

\[ \leq K_4 \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \]

For case (a), by the independence of \( u_{nt} \) and \( \xi_{nt} \) and Assumption 5.15,

\[ E\left[ \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \right] \leq K_4 \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \]

\[ \leq K_4 \left( \sum_{t=1}^{b_n} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \left( \sum_{n=1}^{t} \frac{1}{n} \right)^{1/2} \]

\[ = 0_{p}(a_n^{-1}) \]
For case (b), by Assumption 5.17,

$$E\left[ b_n^{-1} \sum_{t=1}^{n} \mathcal{E}_{nt}^j (x_t^*: \xi^0) u_{nt} u_{ntj} y_{ntm} \right] \leq K_4 b_n^{-1} \sum_{t=1}^{n} E\left[ |u_{ntj} y_{ntm}| \right]$$

$$= K_4 \left[ E[u_{ntj}^2 y_{ntm}^2] E[r_{ntm}^2] \right]^{1/2} .$$

$$= O(a_n^{-1}) .$$  \hspace{1cm} (5.50)

By (5.49) and (5.50), for both cases (a) and (b),

$$b_n^{-1} \sum_{t=1}^{n} u_{ntj}^2 (x_t^*: \xi^0) y_{ntj} y_{ntm} = O(a_n^{-1}) .$$  \hspace{1cm} (5.51)

For case (a), by the independence of \((e_{nt} , u_{nt} \) and \( \xi_{nt} \), and by Assumption 5.8,

$$E\left\{ b_n^{-1} \sum_{t=1}^{n} \left[ e_{nt} - f_x (x_t^*, \xi_t^0) u_{nt} \right] \xi_{nt} \right\}^2$$

$$= b_n^{-2} \sum_{t=1}^{n} E\left\{ (e_{nt} - f_x (x_t^*, \xi_t^0) u_{nt} )^2 \right\} E[r_{ntm}^2] = O(n^{-1}) .$$  \hspace{1cm} (5.52)
For case (b), by the zero correlation, and Assumptions 5.8, 5.17, and 5.18,

\[
\begin{align*}
E\left[ \left( b_n^{-1} \sum_{t=1}^{b_n} \left( e_{nt} - \frac{f(x_{zt}; \xi^0)u_{nt}'}{\mathcal{r}_{ntm}} \right)^2 \right) \right] \\
= b_n^{-2} \sum_{t=1}^{b_n} E\left[ \left( e_{nt} - \frac{f(x_{zt}; \xi^0)u_{nt}'}{\mathcal{r}_{ntm}} \right)^2 \right] \\
< b_n^{-2} \sum_{t=1}^{b_n} \left[ E\left[ \left( e_{nt} - \frac{f(x_{zt}; \xi^0)u_{nt}'}{\mathcal{r}_{ntm}} \right)^4 \right] \right]^{1/2} \mathcal{E}\left[ \mathcal{r}_{ntm} \right]^{1/2} \\
= o(n^{-1}) \quad (5.53)
\end{align*}
\]

Since \( a_n^{-1} = o(n^{-1/2}) \), by (5.47), (5.48), (5.51), (5.52), (5.53), and Assumption 5.4, for both cases (a) and (b),

\[
\begin{align*}
b_n^{-1} \sum_{t=1}^{b_n} \left[ Y_{nt} - f(x_{nt}; \xi^0) \right] \\
= b_n^{-1} \sum_{t=1}^{b_n} \left( x_{nt}^0 + x_{nt} \right)' [1, - f(x_{zt}; \xi^0)] (e_{nt}, u_{nt}') + o_p(n^{-1/2}) \\
= o_p(n^{-1/2}) \quad (5.54)
\end{align*}
\]
We observe that each element of

\[(x_0^0 + \xi_{nt})', [1, - f_x(x_0^0, \xi^0)](e_{nt}, \xi_{nt})',\]

is a linear combination of elements of a random vector

\[x^*_nt = (e_{nt}, \xi_{nt}, e_{nt} \xi_{nt}, \xi_{nt} \xi_{nt}).\]

If we let \(A_t\) be the coefficient matrix of the linear combinations, then by Assumptions 5.4 and 5.16, the limit

\[\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} \text{vec} A_t \text{vec} A_t' = \mathbb{H}\]

exists. Also, \(\frac{1}{n^2} \xi^*_nt\) are independently and identically distributed.

For case (a), by the independence and Assumption 5.16, the mean of \(\frac{1}{n^2} \xi^*_nt\) is zero, and the covariance matrix of \(\frac{1}{n} \xi^*_nt\) converges to some matrix. For case (b), by the zero correlation and Assumption 5.18, the mean of \(\frac{1}{n^2} \xi^*_nt\) is zero, and the covariance matrix of \(\frac{1}{n} \xi^*_nt\) converges to some matrix. Hence, by Theorem 2.12, the limit

\[\xi_0 = \lim_{n \to \infty} b_n^{-1} V[(x_0^0 + \xi_{nt})', [1, - f_x(x_0^0, \xi^0)]\frac{1}{n^2}(e_{nt}, \xi_{nt})']\]

exists and
\[ n^{\frac{1}{2}} b_n^{-1} \sum_{t=1}^{b_n} (x_t^0 + x_{nt})'[1, - \frac{1}{2} \chi(x_t^0; \xi^0)](e_{nt}', u_{nt})' \xrightarrow{L} N(0, \Sigma_v) \]  

(5.55)

Note that for case (a), by the independence and Assumption 5.3,

\[ V_0 = \lim_{n \to \infty} \left[ b_n^{-1} \sum_{t=1}^{b_n} \sigma_t^2 x_t^0, x_t^0 + b_n^{-1} \sum_{t=1}^{b_n} \sigma_t^2 x_{rr} \right]. \]

Hence, the result follows from (5.54), (5.55), and Lemma 5.1.

If \( \Sigma_{rrn} = o(1) \), then the result of Theorem 5.2 holds with \( \Sigma_{rr} = 0 \). It can be shown that if \( \Sigma_{rrn} = O(a_n^{-1}) \) then the limiting result for \( n^{\frac{1}{2}} (\hat{e}_n - \xi^0) \) is valid under either the assumption of independence between \( u_{nt} \) and \( x_{nt} \) or the assumption that

\[ E\left[ \left| \frac{1}{n} a_n^{-1} u_{nt} \right|^4 \right] = O(1). \]

That is, if the variance of the random part \( x_{nt} \) is decreasing, then with the higher moment assumption, \( \hat{w}_{nt} \) need not be uncorrelated with \( (e_{nt}, u_{nt}) \) to produce an asymptotically normal estimator. In practice, we recommend that the zero correlation or the independence of \( \hat{w}_{nt} \) and \( (e_{nt}, u_{nt}) \) be examined before applying the result of Theorem 5.2.

The following theorem provides a consistent estimator of the covariance matrix of the limiting distribution of \( \hat{\xi}_n \). The estimator
is identical for the case with independence of $\xi_{nt}$ and $(e_{nt}, u_{nt})$
and the case only with zero correlation between $\xi_{nt}$ and $(e_{nt}, u_{nt})$.
The consistency of the covariance matrix estimator requires stronger
assumptions than those used to obtain the limiting distribution.

**Assumption 5.19.** There exists a constant $K_5$ such that the absolute
value of each element of $\xi^0_t (\xi^0, \xi^0)$ is bounded by $K_5$ for all $t$.

**Assumption 5.20.** For $i = 1, 2, \ldots, p$,

$$b_n^{-1} \sum_{t=1}^{bn} (w_i^0)^4 = O(1).$$

The following theorem repeats the results in Theorem 5.2 under
stronger assumptions, and presents a consistent estimator of the
covariance matrix of the limiting distribution of $\hat{\Sigma}_n$.

**Theorem 5.3.** Let Assumptions 5.1 through 5.20 hold. Assume that

$$a_n^{-1} = o(n^{-1/2}).$$

Also, assume either

(a) $(e_{nt}, u_{nt})$ and $\xi_{nt}$ are independent, or

(b) $(e_{nt}, u_{nt})$ and $\xi_{nt}$ are uncorrelated, and for $i = 1, 2, \ldots, p$,

$$E\left[\left\{ a_n^{1/2} (e_{nt}, u_{nt}) r_{nti} \right\}^4 \right] = O(1).$$

(5.56)
Then,

$$\frac{1}{\sqrt{n}} (\hat{\theta}_n - \theta^0) \xrightarrow{L} N(0, \Psi) ,$$

and

$$n \hat{\Psi} (\hat{\theta}_n) \xrightarrow{P} \Psi ,$$

where

$$\hat{\Psi}(\hat{\theta}_n) = \{ \hat{\Phi}_m^{-1} \hat{h} \}^{-1} \hat{\Phi}_m^{-1} \hat{\Psi}_m^{-1} \hat{h} \{ \hat{h} \Phi_{WW} \hat{h} \}^{-1} ,$$

$$\hat{h} = b_n^{-1} \sum_{t=1}^{b_n} \xi_{t} (x_{nt}, \hat{\theta}_n) W_{nt} ,$$

$$\Phi_{WW} = b_n^{-1} \sum_{t=1}^{b_n} W_{nt} W_{nt} ,$$

$$\hat{\Psi}_o = b_n^{-2} \sum_{t=1}^{b_n} W_{nt} W_{nt} [Y_{nt} - f(x_{nt}; \hat{\theta}_n)]^2 .$$

**Proof.** By the argument used to obtain (5.44),

$$\hat{h} = h(\theta^0) + o_p(1) .$$

(5.57)
By Assumption 5.4 and the law of large numbers,

\[ m_{WW} = \bar{m}_{WW} + o_p(1) \]  

(5.58)

We observe that

\[ Y_{nt} - f(X_{nt}; \hat{\theta}_n) = e_{nt} - \xi_X(x_t^0; \theta^0)u_{nt} - \xi_{\theta}^{'}(x_t^0; \theta^0) (\hat{\theta}_n - \theta^0) \]

\[ - \frac{1}{2} u_{nt} \xi_{xx} (X_{nt}; \theta^*)u_{nt}' - \frac{1}{2} (\hat{\theta}_n - \theta^0)' \xi_{\theta \theta} (X_{nt}; \theta^*) (\hat{\theta}_n - \theta^0) \]

\[- (\hat{\theta}_n - \theta^0)' \xi_{\theta x} (X_{nt}; \theta^*)u_{nt}' \],

(5.59)

where \( X^*_t \) is on the segment joining \( x_t^0 \) and \( X_{nt} \), and \( \theta^*_n \) is on the segment joining \( \theta^0 \) and \( \hat{\theta}_n \). By Theorem 5.2,

\[ \hat{\theta}_n - \theta^0 = O_p(n^{-1/2}) \].

(5.60)

Using (5.59) and (5.60), we obtain for both the cases (a) and (b),

\[ b_n Y_n = b_n^{-1} \sum_{t=1}^{b_n} \bar{W}_{nt} W_{nt} [Y_{nt} - f(X_{nt}; \hat{\theta}_n)]^2 \]

\[ = b_n^{-1} \sum_{t=1}^{b_n} \bar{W}_{nt} W_{nt} [e_{nt} - \xi_X(x_t^0; \theta^0)u_{nt}']^2 + O_p(n^{-1/2}a_n^{-1/2}) \],

(5.61)
where we have used, for example, if (a) holds, then

\[ E\left[ b_n^{-1} \sum_{t=1}^{b_n} \left| \xi_{nt} \xi_{nt} u_{nt} f_{xx}(x_{nt}; \hat{g}, \hat{g}'; \xi_t^0; \xi^0)(\xi_{nt} - \xi^0) \right| \right] \\
= O\left( \max\{n^{-1/2} a_n, a_n^2 \} \right) , \]

and if (b) holds, then

\[ E\left[ b_n^{-1} \sum_{t=1}^{b_n} \left| \xi_{nt} \xi_{nt} u_{nt} f_{xx}(x_{nt}; \hat{g}, \hat{g}'; \xi_t^0; \xi^0)(\xi_{nt} - \xi^0) \right| \right] \\
< b_n^{-1} \sum_{t=1}^{b_n} \xi_{nt}^2 \left| \xi_{nt} \xi_{nt} u_{nt} u_{nt} \right|^2 E\left[ \left| u_{nt} u_{nt} \right|^2 \right]^{1/2} \\
= o\left( a_n^{3/2} \right) \\
= o\left( n^{-1/2} a_n^{-1/2} \right) . \]

We note that by Theorem 5.2 \( \Psi \) can always be written in the form of

\[ \{ h_0, h_0 \}^{-1} h_0 \begin{pmatrix} \mathbb{I} & -1 \\ \mathbb{I} & -1 \end{pmatrix} \begin{pmatrix} h_0 \mathbb{I} & -1 \\ \mathbb{I} \end{pmatrix} \}^{-1} \begin{pmatrix} h_0 \mathbb{I} & -1 \\ \mathbb{I} \end{pmatrix} \].
where

\[ v_o = \lim_{n \to \infty} a_n b_n^{-1} \sum_{t=1}^{b_n} E\{ \tilde{W}'_t \tilde{W}_t [e_{nt} - f_x(x_t^0; \hat{\beta}_n^0)u_t']^2 \} \quad (5.62) \]

For case (a), by the independence and Assumptions 5.17, 5.18, and 5.20,

\[ v(a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_t \tilde{W}_t [e_{nt} - f_x(x_t^0; \hat{\beta}_n^0)u_t']^2) = O(b_n^{-1}) \quad (5.63) \]

For case (b), by (5.56) and Assumptions 5.17, 5.18, and 5.20, the expression (5.63) holds. By (5.62) and (5.63),

\[ a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_t \tilde{W}_t [e_{nt} - f_x(x_t^0; \hat{\beta}_n^0)]^2 = v_o + O(p_n^{-1/2}) \quad (5.64) \]

By (5.61) and (5.64),

\[ n v_o = a_n b_n^{-1} \sum_{t=1}^{b_n} \tilde{W}'_t \tilde{W}_t [Y_{nt} - f(x_{nt}; \hat{\beta}_n)]^2 \]
\[ \hat{\theta}_n = \nu_0 + o_p\left(b_n^{1/2}\right) \quad (5.65) \]

The result follows from \((5.57), (5.58), (5.65),\) and Theorem 5.2. \(\square\)

We have shown that the estimator of the variance of the approximate distribution of \(\hat{\theta}_n\) is the same for both the cases (a) and (b) in Theorem 5.3. Thus, the inference procedure based on the asymptotic theory is the same for both the cases. As we have seen, the proofs for the case without independence of \((e_{nt}, u_{nt})\) and \(\xi_{nt}\) are somewhat more tedious and troublesome than the cases with the independence. However, all the results holding for the independent case still hold for the case without the independence under the assumptions of zero correlation and the existence of higher moments for the errors.

In later sections, we will consider estimation of the error covariance matrix \(\xi\) and the true values \(\xi^0\) based on \(\hat{\theta}_n\), and will derive an improved estimator of \(\xi^0\) using \(\hat{\theta}_n\) as an initial estimator. Therefore, it is convenient to derive the order of \((\hat{\theta}_n - \xi^0)\) without the assumptions which are needed to obtain the limiting distribution of \(\hat{\theta}_n\).

**Theorem 5.4.** Let Assumptions 5.1 through 5.15 hold. Assume either

(a) \((e_{nt}, u_{nt})\) and \(\xi_{nt}\) are independent, or

(b) \(E\left| a_n u_{nt}\right|^4 = O(1)\).

Then,
\[ \hat{\xi}_n - \xi^0 = O_p(\max[n^{-1/2}, a_n^{-1}]) . \]

Proof. For case (a), the result follows from (5.47), (5.48), (5.51), (5.52), and Lemma 5.1. For case (b), we observe that by Assumption 5.8, for \( m = 1, 2, \ldots, p \),

\[
E[|b_n^{-1} \sum_{t=1}^{b_n} (e_{nt} - f_x(x_t^0; \xi^0)u'_nt)u_{ntm}|] \\
< b_n^{-1} \sum_{t=1}^{b_n} E[|e_{nt}r_{ntm}|] + b_n^{-1} \sum_{t=1}^{b_n} \sum_{i=1}^q |f_{x_i}(x_t^0; \xi^0)|E[|u_{nti}r_{ntm}|] \\
< [E[e_{nt}^2]E[r_{ntm}^2]^{1/2} + \sum_{i=1}^q \sum_{t=1}^{b_n} \sum_{i=1}^q E[u_{nti}^2]E[r_{ntm}^2]^{1/2}]^{1/2} \\
= O(a_n^{-1}) ,
\]

where \( f_{x_i} \) is the \( i \)-th element of \( f_x \). Thus,

\[
b_n^{-1} \sum_{t=1}^{b_n} (e_{nt} - f_x(x_t^0; \xi^0)u'_nt)u_{nt} = O_p(a_n^{-1}) .
\]
By (5.47), (5.48), (5.50), (5.66), and Lemma 5.1, the result also holds for the case (b).

4. Bias

In Chapter IV, we discussed the bias of the maximum likelihood type estimator of the nonlinear functional relationship model with known error covariance matrix. The bias was of \( O(a_n^{-1}) \), and was due to the nonlinearity of the functional form. The instrumental variable estimator \( \hat{\theta}_n \) in this chapter does not adjust for the bias due to the curvature. Thus, we expect the estimator \( \hat{\theta}_n \) to have a bias which has a larger order than \( n^{-1} \). In order to simplify our derivation of the asymptotic bias of \( \hat{\theta}_n \), we introduce the following assumptions.

Assumption 5.21. Let \( \xi_{xi}(z; \theta) \) be the i-th row of \( \xi_{xx}(z; \theta) \), \( i = 1, 2, \ldots, q \). Then,

\[
\xi_{xxi}(z; \theta) = \frac{\partial}{\partial z} \xi_{xi}(z; \theta)
\]

exists and is continuous on \( \Gamma \times \Theta \).

Assumption 5.22. There exists a constant \( K_6 \) such that for every \( i = 1, 2, \ldots, q \), all \( z \) in \( \Gamma \), and all \( \theta \) in \( \Theta \), the absolute value of each element of \( \xi_{xxi}(z; \theta) \) is bounded by \( K_6 \).

In the next theorem, we derive the asymptotic bias of the instrumental variable estimator \( \hat{\theta} \).

Theorem 5.5. Let Assumptions 5.1 through 5.15, 5.21, and 5.22 hold.
Assume either

(a) \((e_{nt}, u_{nt})\) and \(\xi_{nt}\) are independent, and

\[
E\left[ |a_n^{1/2} u_{nt}|^3 \right] = O(1) ,
\]  

(5.67)

or

(b) \((e_{nt}, u_{nt})\) and \(\xi_{nt}\) are uncorrelated, and

\[
E\left[ |a_n^{1/2} (e_{nt}, u_{nt})|^6 \right] = O(1) .
\]  

(5.68)

Then, by taking the expectation of the expansion of \((\hat{\xi}_n - \xi^0)\) through the terms of \(O_p(a_n^{-1})\), the bias of \(\hat{\xi}_n\) is, for case (a),

\[
- \frac{1}{2} \left\{ h^0 \ E_{WW}^{-1} h^0 \right\}^{-1} h^0 \ E_{WW}^{-1} \left\{ b_n \sum_{t=1}^{b_n} w_t^0 \ tr [f_{xxt} \xi_{uun}] \right\} ,
\]

and for case (b),

\[
- \frac{1}{2} \left\{ h^0 \ E_{WW}^{-1} h^0 \right\}^{-1} h^0 \ E_{WW}^{-1} \left\{ b_n \sum_{t=1}^{b_n} w_t^0 \ tr [f_{xxt} \xi_{uun}] + b_n \sum_{t=1}^{b_n} \xi_{nt} \right\} ,
\]

where \(\xi_{nt}\) is a \(p \times 1\) vector with the \(i\)-th element being
Proof. We observe that (5.68) implies (5.67). By Assumption 5.21,

\[ \text{tr}\left[ g^0 \mathbf{E}[r_{nti} u'_{nt} u_{nt}] \right]. \]

where \( X^*_{nt} \) is on the segment joining \( x_t^0 \) and \( x_{nt}^0 \). We note that

\[ b_n^{-1} \sum_{t=1}^{b_n} \left[ e_{nt} - \xi_x(x_t^0; g^0)u'_{nt}\right] x_t^0 = o_p(n^{-1/2}) , \]

and that

\[ b_n^{-1} \sum_{t=1}^{b_n} u_{nt}f(x_t^0; g^0)u'_{nt}u_{nt} = o_p(a_{-1}) . \]

By (5.67) and Assumption 5.22, for \( i = 1,2,\ldots,q \),
For case (a), by (5.52),

\[
\sum_{t=1}^{b_n} u_{nt} f(x^t_{nt}; \xi^0_{nt}) u_{nt} \xi_{nt} = 0(a^{-3/2}_n) .
\]  

(5.72)

For case (b), by (5.53), the zero correlation, and the argument used to obtain (5.53), the expression (5.73) holds. For case (a), by (5.67), the independence, and Assumption 5.22,

\[
E\{b_n^{-1} \sum_{t=1}^{b_n} [u_{nt} f(x^t_{nt}; \xi^0_{nt}) u_{nt} \xi_{nt}]\} = 0(a^{-1}_n) ,
\]  

(5.74)

and for \( i = 1, 2, \ldots, q, \)

\[
E\{b_n^{-1} \sum_{t=1}^{b_n} [u_{nt} f(x^t_{nt}; \xi^0_{nt}) u_{nt} \xi_{nt}]\} = 0(a^{-3/2}_n) .
\]  

(5.75)

For case (b), by (5.68), Assumption 5.22, and the Cauchy-Schwartz inequality, the expressions (5.74) and (5.75) hold. Therefore, by (5.69), (5.70), (5.71), (5.72), (5.73), (5.74), (5.75), and Lemma 5.1,

\[
\hat{\xi}_n - \xi^0 = \{h^0 \tilde{M}_w^{-1} h^0\}^{-1} h^0 \tilde{M}_w^{-1} \Delta \hat{\xi} + o_p(a^{-1}_n) ,
\]  

(5.76)
where
\[
\Delta_1 \hat{\beta} = b_n^{-1} \sum_{t=1}^{b_n} (y_t^0 + \xi_{nt})' [e_{nt} - \hat{f}_x(x_t; \beta^0)u_{nt}'], \quad \text{if } a_n^{-1} = o(n^{-1/2}),
\]
\[
= 0, \quad \text{if } n^{-1/2} = o(a_n^{-1}).
\]

\[
\Delta_2 \hat{\beta} = -\frac{1}{2} b_n^{-1} \sum_{t=1}^{b_n} (y_t^0 + \xi_{nt})' u_{nt} f_0(x_t; \beta^0) u_{nt}'.
\]

Since under either (a) or (b),
\[
E[\Delta_1 \hat{\beta}] = 0,
\]
the results follow by taking the expectation of the leading term of (5.76).

We point out that if \( \Sigma_{rrn} = 0(a_n^{-1}) \) then under either (a) or (b) the bias of \( \hat{\beta}_n \) is given by
\[
-\frac{1}{2} \left[ b_{-1}^0 \Sigma_{ww}^{-1} b_{-1}^0 \right]^{-1} \left[ b_{-1}^0 \Sigma_{ww}^{-1} b_{-1}^0 \right] \sum_{t=1}^{b_n} y_t^0' \text{tr} \left[ f_0(x_t; \beta^0) u_{nn} \right].
\]

The asymptotic bias derived in Theorem 5.5 is a function of the second derivative \( f_{xx}(x_t^0; \beta^0) \). Thus, as in Chapter IV, the bias of \( O(a_n^{-1}) \) in \( \hat{\beta}_n \) can be considered as the bias due to the curvature of the function \( f(x; \beta^0) \). For the model linear in all elements of
$\hat{\beta}_n^0$, the bias of $\hat{\beta}_n$ is $o(a_n^{-1})$. We note that in the bias of $\hat{\beta}_n$
the covariance matrix multiplying $f_0^0$ is $\Sigma_{uu}$ while the multiplier
is $A_{nt}^{-1}$ in the bias of the maximum likelihood type estimator considered
in Chapter IV.

C. An Estimator of the Error Covariance Matrix

In the previous sections, we discussed the instrumental variable
estimator $\hat{\beta}_n$ of $\beta^0$ in the model (5.8). We assumed that the
covariance matrix $\Sigma_n$ of $(e_{nt}, u_{nt})$ is unknown. In this section, we
introduce an estimator of $\Sigma_n$, and discuss properties of the
estimator. The estimator will be used in a later section to obtain an
improved estimator of $\beta^0$.

In the linear errors-in-variables model with instrumental
variables, a general unknown error covariance matrix cannot be estimated
without an additional assumption such as zero correlation between errors
in $Y_{nt}$ and $X_{nt}$. For the estimation of the error covariance matrix,
the residual

$$\hat{\nu}_{nt} = Y_{nt} - X_{nt} \hat{\beta}_{IV},$$

where $\hat{\beta}_{IV}$ is the instrumental variable estimator, plays an important
role in the linear model. The sample covariances of $\hat{\nu}_{nt}$ with
$(Y_{nt}, X_{nt}, \hat{\nu}_{nt})$ provide estimates of linear combinations of the
elements in the error covariance matrix. In the nonlinear model, the
residual
\[ \hat{v}_{nt} = Y_{nt} - f(X_{nt}; \hat{\beta}_n) \]

contains the information on the measurement errors. As we observed earlier,

\[ v^0_{nt} = Y_{nt} - f(X_{nt}; \beta^0_0), \quad t = 1, 2, \ldots, b_n. \]

are not identically distributed for the models nonlinear in \( \mathbf{x}_t^0 \). This fact enables us to estimate the error covariance matrix in the nonlinear model without additional knowledge on \( \mathbf{E}_n \).

We observe that under the assumptions of Theorem 5.4,

\[ \hat{\beta}_n - \beta^0 = O_p(\max[n^{-1/2}, a_n^{-1}]), \]

\[ \hat{v}_{nt} = Y_{nt} - f(X_{nt}; \hat{\beta}_n) \]

\[ = e_{nt} - \frac{1}{\mathbf{x}_t} u'^{\prime} + O_p(\max[n^{-1/2}, a_n^{-1}]), \]

and \( \frac{1}{\mathbf{x}_t} \) varies over \( t = 1, 2, \ldots, b_n \), for the models nonlinear in \( \mathbf{x}_t^0 \). Hence, the expected values of \( \hat{v}_{nt}^2 \) are approximately given by

\[ \sigma_{vnt}^2 = [1, -\frac{1}{\mathbf{x}_t}] \Sigma_0 [1, -\frac{1}{\mathbf{x}_t}]' \]

\[ = [1, -\frac{1}{\mathbf{x}_t}] \Sigma_0 [1, -\frac{1}{\mathbf{x}_t}] \vec{\Sigma}_n \]
where

\[ \eta_t = ((E_t^0 \circ E_t^0) \triangleleft_q + 1 , \]

and \( \triangleleft_q + 1 \) was defined in Definition 2.3. The vector \( \eta_t \) can be estimated by

\[ \hat{\eta}_{nt} = (\hat{E}_{nt} \circ \hat{E}_{nt} \triangleleft_q + 1 , \]

where

\[ \hat{E}_{nt} = [1, - f^0(\hat{X}_n)] . \]

The expressions (5.77) and (5.78) suggest that the estimated coefficients in the regression of \( \hat{\eta}^2_{nt} \) on \( \hat{\eta}_{nt} \) estimate the elements of \( \hat{\Sigma}_n \). In order to investigate the properties of such an estimator of \( \hat{\Sigma}_n \), we introduce an assumption.

**Assumption 5.23.** The limit

\[ H_0 = \lim_{n \to \infty} b_n \left\{ \sum_{t=1}^{b_n} (E_t^0 \circ E_t^0) \circ (E_t^0 \circ E_t^0) \right\} . \]
exists, and the matrix

\[ H = \mathcal{Q}_{q+1} \cdot H \cdot \mathcal{Q}_{q+1} \]

is nonsingular.

Note that for a 1 x (q+1) vector \( \mathbf{A} \) each element of \((\mathbf{A} \otimes \mathbf{A}) \cdot \mathcal{Q}_{q+1}\) has the form of either \( a_i^2 \) or \( 2a_i a_j \), \( i \neq j \), where \( a_i \) is the \( i \)-th element of \( \mathbf{A} \). Thus, the multiplication of \( \mathcal{Q}_{q+1} \) combines repeated terms in \((\mathbf{A} \otimes \mathbf{A})\).

If the model is linear in one of the elements of \( \mathcal{E}_t^0 \), then the corresponding element of \( \mathcal{E}_t^0 \) is a constant. Hence, the model linear in at least one element of \( \mathcal{E}_t^0 \) does not satisfy the nonsingularity of \( H \) in Assumption 5.23. An intuitive interpretation of Assumption 5.23 is that the sequence \( \mathcal{E}_t^0 \) is chosen in a way such that for large \( n \) the vectors \((\mathcal{E}_t^0 \otimes \mathcal{E}_t^0) \cdot \mathcal{Q}_{q+1}\), \( t = 1,2,\ldots, b_n \), do not concentrate on any subspaces of less than full dimension in the \( \frac{1}{2}(q+1)(q+2) \)-dimensional Euclidean space. Hence, intuitively speaking, Assumption 5.23 is satisfied if the elements of \( \mathcal{E}_t^0 \) vary sufficiently, but without increasing in absolute magnitude.

We define the estimator \( \text{vech} \mathcal{E}_n \) of \( \text{vech} \mathcal{E}_n \) to be the estimated regression vector given by the regression of \( \mathcal{E}_{nt}^2 \) on \( \mathcal{E}_{nt} \), \( t = 1,2,\ldots, b_n \). That is,

\[ \text{vech} \mathcal{E}_n = \left( \sum_{t=1}^{b_n} \mathcal{E}_{nt} \mathcal{E}_{nt} \right)^{-1} \sum_{t=1}^{b_n} \mathcal{E}_{nt} \mathcal{E}_{nt}^2. \]
We recall that $\Sigma_n$ is of $O(a_n^{-1})$. Thus, it is desirable that the estimator $\hat{\Sigma}_n$ satisfies

$$a_n (\hat{\Sigma}_n - \Sigma) \xrightarrow{P} 0 \quad (5.79)$$

The following theorem shows that our estimator $\hat{\Sigma}_n$ satisfies the consistency condition (5.79).

**Theorem 5.6.** Let Assumptions 5.1 through 5.15, 5.17, and 5.23 hold. Then,

$$\hat{\Sigma}_n - \Sigma = O_p(\max\{n^{3/2}, a_n^{1/2}\}) \quad (5.80)$$

**Proof.** By Assumption 5.17 and Theorem 5.4,

$$\hat{\beta}_n - \beta^0 = O_p(\max\{n^{1/2}, a_n^{-1}\}) \quad (5.81)$$

We observe that

$$\hat{\eta}_{nt} = e_{nt} - \beta_0 x_{nt} - \xi_{nt} (\hat{\beta}_n - \beta^0) - \frac{1}{2} u_{nt}\xi_{nt} (\hat{\xi}_n^*; \xi_{nt}^*) u_{nt}$$

$$- \frac{1}{2} (\hat{\xi}_n - \xi_0)\xi_{nt} (\hat{\xi}_n^*; \xi_{nt}^*) (\hat{\xi}_n^* - \xi_0) - (\hat{\xi}_n - \xi_0)\xi_{nt} (\hat{\xi}_n^*; \xi_{nt}^*) u_{nt},$$

where $x_{nt}^*$ is on the segment joining $x_0^*$ and $x_{nt}$, and $\xi_{nt}^*$ is on the segment joining $\xi_0$ and $\xi_{nt}$. Hence, it follows from (5.80),
(5.81), and Assumptions 5.8, 5.9, 5.13, 5.15, and 5.17 that

$$
\mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} \mathbf{v}_{t}^{2} = \mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} \mathbf{t}_{t} \operatorname{vech} \mathbf{\Sigma}_{t} + o_{p}(\max[\mathbf{a}_{n}^{-3/2}, \mathbf{a}_{n}^{-1/2}])
$$

$$
= o_{p}(\mathbf{a}_{n}^{-1})
$$

where we have used, for example,

$$
\mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} ((\mathbf{\tilde{\xi}}_{n} - \mathbf{\xi}_{0})' \mathbf{u}^{0} \mathbf{u}^{0} \mathbf{u}^{0}) = o_{p}(\max[\mathbf{a}_{n}^{-1}, \mathbf{a}_{n}^{-2}])
$$

$$
\mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{f}_{t}^{0} \mathbf{u}_{t}^{0} \mathbf{u}_{t}' = o_{p}(\mathbf{a}_{n}^{-3/2})
$$

$$
\mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{f}_{t}^{0} \mathbf{u}_{t}^{0} \mathbf{u}_{t}^{0} \mathbf{u}_{t}^{0} = o_{p}(\max[\mathbf{a}_{n}^{-1/2}, \mathbf{a}_{n}^{-3/2}])
$$

$$
\mathbf{E}[\mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} (\mathbf{e}_{nt} - \mathbf{\xi}_{nt} \mathbf{u}_{nt})^{2}] = \mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} \mathbf{n}_{t} \sigma_{vnt}^{2}
$$

$$
= \mathbf{b}^{-1} \sum_{t=1}^{n} \mathbf{n}_{t}^{'} \mathbf{t}_{t} \operatorname{vech} \mathbf{\Sigma}_{n}
$$
\[ V\{b_n^{-1} \sum_{t=1}^{b_n} \hat{n}_t^t e_{nt} - \xi_0 e_{nt} - \xi_{nt}\}^2 = 0(n^{-1} a_n^{-1}) , \]

and in \( \hat{n}_n \) for some \( X^{**} \) and \( \beta^{**} \),

\[ f(x; \hat{\beta}_n) = f(x; \beta^0) + f(x; X^{**}; \beta^{**}) u_n^t + (\hat{\beta}_n - \beta^0)' f(x; X^{**}; \beta^{**}) . \] (5.83)

Also, by (5.83) and Assumptions 5.17 and 5.23,

\[ b_n^{-1} \sum_{t=1}^{b_n} \hat{n}_t^t \hat{n}_n = b_n^{-1} \sum_{t=1}^{b_n} \hat{n}_t^t \hat{n}_n + O_p(\max[n^{-1/2}, a_n^{-1}]) \]

\[ = H + O_p(1) , \] (5.84)

where \( H \) is nonsingular. Therefore, the result follows from (5.82) and (5.84).

We recall that the assumption \( a_n^{-1} = o(n^{-1/2}) \) was required to obtain the limiting distribution of \( n^{1/2}(\hat{\beta}_n - \beta^0) \). The limiting distribution of \( \hat{\xi}_n \) with normalizing constant \( n^{1/2} \) can be derived under a weaker assumption on the rate of increase of \( a_n \).

**Theorem 5.7.** Let Assumptions 5.1 through 5.15, 5.17, and 5.23 hold.

Also, let
\[ G_n = a_n^2 \operatorname{vech}(e_{nt} u_{nt}')(e_{nt} u_{nt}) - \Sigma_n ). \]

Assume

\[ a_n^{-1} = o(n^{-1/2}), \quad (5.85) \]

\[ \lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} (n_t n_t') (n_t n_t') = A_1, \quad (5.86) \]

\[ \lim_{n \to \infty} G_n = G. \]

Then,

\[ n^{1/2} \operatorname{vech}(\hat{\Sigma}_n - \Sigma_n) \xrightarrow{L} N(0, \Psi_\sigma), \]

where

\[ \Psi_\sigma = \Sigma^{-1} A_2 \Sigma^{-1}, \]

\[ A_2 = \lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} n_t n_t' n_t n_t' = \Sigma_n n_t n_t'. \]

Proof. By (5.82), (5.83), and (5.85),
\[ b_n^{-1} \sum_{t=1}^{b_n} \left( \hat{v}_{nt}^2 - \hat{\sigma}_{nt} \text{vech } \hat{\Sigma}_n \right) = n^{-1} \sum_{t=1}^{b_n} \eta_{nt}' \eta_{nt} \text{vech} \left( (e_{nt}' u_{nt})' (e_{nt}' u_{nt}) - \Sigma_n \right) + o_p(n^{-1/2}). \]  

(5.87)

It follows from (5.86), (5.87), and Theorem 2.12,

\[
\frac{1}{n^{1/2}} b_n^{-1} \sum_{t=1}^{b_n} \left( \hat{v}_{nt}^2 - \hat{\sigma}_{nt} \text{vech } \hat{\Sigma}_n \right) \xrightarrow{L} N(\mathbf{Q}, A_2). \]

(5.88)

The result follows from (5.84) and (5.88).

Without knowledge of a particular form of \( G_n \), we do not have a consistent estimator of the covariance matrix \( \Sigma_d \) of the limiting distribution of \( \text{vech } \Sigma_n \). Thus, to make inferences on \( \Sigma_n \), we need more information on distributional properties of \( (e_{nt}', u_{nt}) \).

It should be pointed out that with a finite number of observations the estimator \( \hat{\Sigma}_n \) may not be a nonnegative definite matrix. To use \( \hat{\Sigma}_n \) in the derivation of estimators of \( \Sigma_t^0 \) and \( \Sigma_t^0 \), we replace \( \hat{\Sigma}_n \) which is not nonnegative definite by its closest positive semidefinite matrix, i.e., the matrix with the same eigenvectors as \( \Sigma_n \) and zeros.
replacing the negative eigenvalues of $\Sigma_n$. In order to simplify our theoretical development, we assume, from now on, that $\Sigma_n$ is nonnegative definite. It is understood that the inverse of a matrix is replaced by the Moore-Penrose generalized inverse whenever the matrix is singular in the ensuing sections.

D. An Estimator of the True Value

Having obtained an estimator $\hat{\Sigma}_n$ of the error covariance matrix in the previous section, we can now obtain an estimator of the unknown true value $x_0^0$ which utilizes both $Y_{nt}$ and $X_{nt}$. Let $\hat{x}_t$ be the value of $x_t$ contained in $\mathcal{X}$ that minimizes

$$D_n(x_t) = [Y_{nt} - f(x_t; \beta_n), x_{nt} - x_t] \Sigma_n^{-1} [Y_{nt} - f(x_t; \beta_n), x_{nt} - x_t]' .$$

We introduce additional assumptions.

**Assumption 5.24.** For all $n$, $\Sigma_n$ is nonsingular.

**Assumption 5.25.** For all $n$, the $x_t^0$, $t = 1, 2, \ldots, b_n$, are interior points of $\mathcal{X}$.

The property of the estimator $\hat{x}_t$ is given in the following theorem.

**Theorem 5.8.** Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\hat{x}_t - x_t^0 = \delta_n + o_p(\max\{n^{-1/2}, a_n^{-1}\}) ,$$
where

\[ \delta_{nt} = (e_{nt}, u_{nt}) \Sigma_n^{-1} [f_0^t, I] \Lambda_n^{-1} \]

\[ = u_{nt} - v^* \sigma_{vnt}^{-2} \Sigma \nabla \Sigma^{-1} v^* \]

\[ \Lambda_{nt} = [f_0^t, I] \Sigma_n^{-1} [f_0^t, I]' \]

\[ v^*_{nt} = e_{nt} - f_0^t u_{nt} \]

\[ \Sigma_{vnt} = [1, -f_0^t] \Sigma_n [Q, I]' \]

and \( \sigma_{vnt}^2 \) is given in (5.77).

**Proof.** By Assumption 5.24 and Theorem 5.6, \( \Sigma_n \) is nonsingular for large \( n \). Since \( \hat{X}_t \) minimizes \( D_n(x_t) \) over \( \Gamma \),

\[ D_n(\hat{X}_t) < D_n(x_0^t) \]  \hspace{1cm} (5.89)

Observe that

\[ Y_{nt} - f(x_0^t; \hat{\beta}_n) = e_{nt} - f_0^t (\hat{\beta}_n - \beta_0) - \frac{1}{2} (\hat{\beta}_n - \beta_0)' f_0^t (\hat{\beta}_n - \beta_0) \]

\[ = \frac{1}{2} (\hat{\beta}_n - \beta_0)' f_0^t (\hat{\beta}_n - \beta_0) \]  \hspace{1cm} (5.90)

where \( \hat{\beta}_n \) is on the segment joining \( \beta_0 \) and \( \hat{\beta}_n \). Since \( \hat{\beta}_n \) is in \( \Omega \), and since \( \Omega \) is convex, \( \hat{\beta}_n \) is also in \( \Omega \). Thus, by Assumptions
By Theorem 5.6,

\[ a_n \hat{\xi}_n = O_p(1) . \]  

(5.92)

It follows from (5.89), (5.91), and (5.92) that

\[ D_n(\hat{\xi}_t) = D_n(\hat{\xi}^0) = O_p(1) . \]  

(5.93)

We observe that

\[ D_n(\hat{\xi}_t) = (\hat{\xi}_nt - \hat{\xi}_t)\hat{\Sigma}_{\text{ee.un}}^{-1}(\hat{\xi}_nt - \hat{\xi}_t)' + \sigma_{\text{ee.un}}^{-2} \left[ Y_{nt} - f(\hat{\xi}_t; \hat{\theta}_n) - \gamma_n(\hat{\xi}_nt - \hat{\xi}_t) \right]^2 , \]  

(5.94)

where

\[ \hat{\Sigma}_{\text{ne.un}}^{-1} = \begin{pmatrix} \hat{\sigma}_{\text{een}} & \hat{\xi}_{\text{eeun}} \\ \hat{\xi}_{\text{een}} & \hat{\xi}_{\text{uen}} \end{pmatrix} , \]

\[ \hat{\sigma}_{\text{ee.un}} = \hat{\sigma}_{\text{een}} - \hat{\xi}_{\text{een}} \hat{\xi}_{\text{uen}}^{-1} \hat{\xi}_{\text{uen}} , \]
\[ \hat{X}_n = \hat{X}_{eun} \hat{X}_{uun}^{-1}. \]

By (5.93) and (5.94),

\[ (\hat{X}_{nt} - \hat{X}_t) \hat{X}_{uun}^{-1} (\hat{X}_{nt} - \hat{X}_t)' = \mathcal{O}_p(1). \]  

(5.95)

Since by Theorem 5.6,

\[ a \hat{X}_{nt} - a \hat{X}_t = \mathcal{O}_p(1), \]

it follows from Assumption 5.24 and (5.95) that

\[ \hat{X}_t - \hat{X}_{nt} = \mathcal{O}_p(a_n^{-1/2}). \]

Since

\[ \hat{X}_{nt} - \hat{X}_t = \mathcal{O}_p(a_n^{-1/2}), \]

we have

\[ \hat{X}_t - \hat{X}_t^0 = \mathcal{O}_p(a_n^{-1/2}). \]  

(5.96)
Let $f_{xi}$ be the $i$-th element of $f_x$, let $f_{xxi}$ be the $i$-th row of $f_{xx}$, and let $I_i$ be the $i$-th row of the $q \times q$ identity matrix. Then, for $i = 1,2,\ldots,q$,

\[
[Y_{nt} - f(x_t; \hat{\beta}_n), x_{nt} - x_t]_{x_n^{-1}}^i[f_{xi}(x_t; \hat{\beta}_n), I_i]'
\]

\[
= [Y_{nt} - f(x_t^0; \hat{\beta}_n), u_{nt}]_{x_n^{-1}}^i[f_{xi}(x_t^0; \hat{\beta}_n), I_i]'
\]

\[
+ (x_t - x_t^0)[f_{xxi}(x_t^0; \hat{\beta}_n), \mathcal{Q}]_{x_n^{-1}}^i[Y_{nt} - f(x_t^0; \hat{\beta}_n), x_{nt} - x_t^0]'
\]

\[
- [f'_x(x_t^0; \hat{\beta}_n), \mathcal{Q}]_{x_n^{-1}}^i[f_{xi}(x_t^0; \hat{\beta}_n), I_i]'
\]

(5.97)

where $x_t^0$ is on the segment joining $x_t^0$ and $x_t$. The left hand side of (5.97) is a multiple of the derivative of $D_n(x_t)$ evaluated at $x_t$. Therefore, by Assumption 5.25 and the argument used in (5.34) for $i = 1,2,\ldots,q$, the left hand

\[
[Y_{nt} - f(x_t^0; \hat{\beta}_n), x_{nt} - x_t]_{x_n^{-1}}^i[f_{xi}(x_t^0; \hat{\beta}_n), I_i]' = o_p(a_n^{-2})
\]

(5.98)

It follows from (5.80), (5.90), (5.96), (5.98), and Theorem 5.6 that

\[
\hat{x}_t - x_t^0 = (e_{nt}', u_{nt})_{x_n^{-1}}[f_{x,t}^0, I]_{A_{nt}^{-1}} + o_p(\max\{n^{-1/2}, a_n^{-1}\})
\]

\[
= u_{nt} + [v_{nt}', \mathcal{Q}]_{x_n^{-1}}[f_{x,t}^0, I]_{A_{nt}^{-1}} + o_p(\max\{n^{-1/2}, a_n^{-1}\})
\]
Using Theorem 2.1, we obtain the result.

If we ignore the terms of $O_p(\max[n^{-1/2}, a_n^{-1}])$, then the error in $\hat{x}_t$ as an estimator of $x_t^0$ is $\delta_{nt}$ with mean zero and covariance matrix $A_{nt}^{-1}$. Since by Theorem 5.6

$$a_n \hat{\xi}_n - a_n \xi_n = O_p(\max[a_n^{-1/2}, b_n^{-1/2}]),$$

the errors of $O_p(a_n^{-1})$ in $(\hat{x}_t - x_t^0)$ include the terms due to the error in $\hat{\xi}_n$ along with the terms due to the curvature of $f$. The unknown true value $y_t^0$ can be estimated by

$$\hat{y}_t = f(\hat{x}_t; \hat{\xi}_n).$$

Corollary 5.8.1. Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\hat{y}_t - y_t^0 = \delta^*_{nt} + O_p(\max[n^{-1/2}, a_n^{-1}]),$$

where

$$\delta^*_{nt} = e_{nt} - v_{nt} \frac{\sigma^2}{nt} v_{nt}.$$
\[ \sigma_{\text{vent}} = [1, -\frac{t^0}{\sqrt{2}x_t}] \Sigma_n (1, Q)' \]

**Proof.** We observe that by Theorem 5.8,

\[
\hat{y}_t - y_t^0 = f^0 \hat{\delta} + O_p(\max[n^{-1/2}, a_n^{-1}])
\]

\[
= e_{nt} + \{\frac{t^0}{\sqrt{2}x_t} [\Sigma_{nt}^{-1} (f^0, Q) - I] \nu_{nt} + O_p(\max[n^{-1/2}, a_n^{-1}]) \}.
\]

Thus, the result follows from Theorem 2.1.

In practice, the minimization of \( D_n(x_t) \) for each \( t \) may not be feasible. For computational purposes, we present the following local approximation to \( \hat{x}_t \). Let

\[
\tilde{x}_t = \hat{x}_{nt} - u_{nt},
\]

where

\[
\hat{u}_{nt} = v_{nt} \Sigma_{nt}^{-2} \nu_{nt},
\]

\[
\hat{v}_{nt} = v_{nt} - f(x_{nt} : \hat{\theta}_n),
\]

\[
\hat{\sigma}_{vnt}^2 = [1, -\frac{t}{\sqrt{2}x_t} (x_{nt} : \hat{\theta}_n)] \Sigma_n [I, -\frac{t}{\sqrt{2}x_t} (x_{nt} : \hat{\theta}_n)]',
\]

\[
\tilde{\Sigma}_{vunt} = [1, -\frac{t}{\sqrt{2}x_t} (x_{nt} : \hat{\theta}_n)] \Sigma_n [Q, I]'.
\]
The next theorem shows that the estimator $\tilde{x}_t$ is equivalent to $\hat{x}_t$ up to terms of $O_p\left(a_n^{-1/2}\right)$.

**Theorem 5.9.** Let Assumption 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$\tilde{x}_t - x_t = \delta_{nt} + O_p\left(\max\{n^{-1/2}, a_n^{-1}\}\right),$$

where $\delta_{nt}$ is defined in Theorem 5.8.

**Proof.** By (5.80) and (5.81),

$$\hat{v}_{nt} = v_n + O_p\left(\max\{n^{-1/2}, a_n^{-1}\}\right),$$

$$f_x(x_{nt}; \hat{\theta}_n) = f^0_{xt} + O_p\left(a_n^{-1/2}\right).$$

Hence, the result follows from Theorem 5.6.

**Corollary 5.9.1.** Let Assumptions 5.1 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Also, let

$$\tilde{y}_t = f(\tilde{x}_t; \hat{\theta}_n).$$

Then,

$$\tilde{y}_t - y_t = \delta^*_{nt} + O_p\left(\max\{n^{-1/2}, a_n^{-1}\}\right),$$

where $\delta^*_{nt}$ is defined in Corollary 5.8.1.
Proof. The result follows from Theorem 5.9 and the argument used in the proof of Corollary 5.8.1.

E. The Modified Instrumental Variable Estimator

1. Introduction

In the previous sections, we discussed the properties of the instrumental variable estimator \( \hat{\beta}_n \) and the estimators of the error covariance matrix and of the true values based on \( \hat{\beta}_n \). The estimator \( \hat{\beta}_n \) was introduced as an analogue of the two-stage least squares estimator in the simultaneous equation system. As we discussed earlier, the quantities \( [Y_{nt} - f(X_{nt};g^0)] \), \( t = 1,2,\ldots,b_n \), are not identically distributed. The estimator \( \hat{\beta}_n \) was obtained by minimizing \( Q_n(\beta) \) which does not take into account the differences among the variances of \( [Y_{nt} - f(X_{nt};g^0)] \). Having obtained an estimator of \( \Sigma_n \), we now seek to improve the estimator of \( g^0 \) by modifying \( Q_n(\beta) \) with estimated variances of \( [Y_{nt} - f(X_{nt};g^0)] \). That is, we expect to obtain an improvement analogous to that of the estimated generalized least squares estimator over the ordinary least squares estimator.

We have discovered another unpleasant property of the estimator \( \hat{\beta}_n \). The derivation of the limiting distribution of \( n^{1/2}(\hat{\beta}_n - g^0) \) required a rather strong assumption for the rate of increase in \( a_n \). The error variances were required to decrease at a faster rate than \( n^{-1/2} \). This assumption about \( a_n \) may be weakened by considering the second derivative adjustment used in Chapter IV. We also expect that such an adjustment will reduce the bias in \( \hat{\beta}_n \) due to the
nonlinearity of the relationship.

In the derivation of the properties of \( \hat{\beta}_n \), we considered two cases depending on whether \((e_{nt}, u_{nt})\) and \(z_{nt}\) are independent or uncorrelated. Such separate considerations made our discussion rather tedious. We had to introduce different assumptions and proofs for different cases. But, every result we derived was equally valid for the two cases with appropriate adjustment in the assumptions. As shown in Theorem 5.3, the inferences on \( \hat{\beta}^0 \) based on the asymptotic results for \( \hat{\beta}_n \) are identical for the independent case and the uncorrelated case. Thus, in order to simplify our discussion on the modified estimator of \( \hat{\beta}^0 \), we consider only the case where \( W_{nt} \) are fixed constants.

The assumption of fixed \( W_{nt} \) does not necessarily mean that we observe fixed constants \( W_{nt} \) without measurement errors. The \( W_{nt} \) may have random components which are independent of \((e_{nt}, u_{nt})\). In such a case, we can make inferences on the equation of interest, conditional on the observed \( W_{nt} \), without altering the nature of the equation. Also, as we have seen, the estimated covariance matrix of the estimator for unconditional inference has the same form as that for conditional inference. Thus, our asymptotic results under the assumption of fixed \( W_{nt} \) can be applied to the case with the assumption that \( W_{nt} \) are independent of \((e_{nt}, u_{nt})\). The results in this section would hold for the cases with only the zero correlation of \( W_{nt} \) and \((e_{nt}, u_{nt})\) and with higher moment assumptions. Since our purpose in this section is the presentation of a modified estimator, we concentrate on the case with fixed \( W_{nt} \), and avoid the complication due to the differences in
the assumptions on $W_{nt}$.

We also transform $W_{nt}$ so that for all $n$

$$b_n^{-1} \sum_{t=1}^{b_n} W_{nt} W_{nt} = I.$$

Since $W_{nt}$ are fixed and, thus, independent of $(Y_{nt}, X_{nt})$, such a transformation does not change our model specification.

The assumption of fixed $W_{nt}$ may be considered as a special case of the general model discussed in previous sections, where $r_{nt} = 0$ for all $t$ and $n$. But, we choose to introduce some assumptions to clarify the model considered in this section.

Assumption 5.1a. The observations satisfy

$$Y_{nt} = f(x^0_t; \theta^0) + e_{nt},$$

$$X_{nt} = x^0_t + u_{nt}, \quad t = 1, 2, \ldots, b_n,$$

$$(e_{nt}, u_{nt})' \sim \mathcal{N}(0, \Sigma_n),$$

$$\Sigma_n = O(a_n^{-1}).$$

Assumption 5.4a. The $W_{nt}$ form a triangular array of fixed constants satisfying, for all $n$,
Assumption 5.10a. Uniformly for all \( \xi \) in \( \mathcal{G} \),

\[
\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} \mathcal{W}^t \mathcal{W}_{nt} = I.
\]

Assumption 5.16a. Let \( W_{nt} \) and \( f^0_{xti} \) be the \( i \)-th elements of \( \mathcal{W}_{nt} \) and \( f^0_{xt} \), respectively. Then, for all \( i_1, i_2 = 1,2,\ldots,p \), and for all \( j_1, j_2 = 1,2,\ldots,q \),

\[
\lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{b_n} W_{nti} W_{nti_1} f^0_{xtj_1} f^0_{xtj_2} = m_4(i_1, i_2, j_1, j_2).
\]

Also,

\[
\lim_{n \to \infty} a_n \sum_{n} \xi = \xi.
\]
Assumption 5.20a. For \( i = 1, 2, \ldots, p, \)

\[
\begin{align*}
\frac{b_n}{n} \sum_{t=1}^{b_n} (W_{nt})^4 \to 0(1).
\end{align*}
\]

Under Assumption 5.4a, the instrumental variable estimator \( \hat{\beta}_n \)
minimizes

\[
\begin{align*}
Q_n(\beta) &= \left\{ \sum_{t=1}^{b_n} (Y_{nt} - f(X_{nt}; \beta))^2 \right\} \left\{ \sum_{t=1}^{b_n} W_{nt} (Y_{nt} - f(X_{nt}; \beta))^2 \right\}.
\end{align*}
\]

We summarize the properties of \( \hat{\beta}_n \) for the case with fixed \( W_{nt} \). By

Theorem 5.1, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, and

5.10a, \( \hat{\beta}_n \) is consistent for \( \beta^0 \). By Theorem 5.2, if Assumptions

5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, and 5.16a

hold, and if \( \sigma_n^{-1} = o(n^{-1/2}) \), then

\[
\begin{align*}
\frac{1}{n^{1/2}} (\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, \Psi),
\end{align*}
\]

where

\[
\Psi = (\eta^0, \eta^0) -1 \eta^0 \sigma^{-1} \eta^0, \eta^0, (\eta^0, \eta^0) -1
\]
\[ m = \lim_{n \to \infty} b^{-1}_n \sum_{t=1}^{b_n} a_n v_{nt} W' W_{nt}, \]

\[ b^0 = h(\theta^0), \]

\[ \sigma^2_{v_{nt}} = [1, -x^0_{xt}] E_n [1, -x^0_{xt}]'. \]

By Theorem 5.4, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, and 5.11 through 5.15,

\[ \hat{\theta}_n - \theta^0 = O_p \left( \max \left\{ n^{-1/2}, a_n^{-1} \right\} \right). \] (5.100)

Also, by Theorem 5.5, if Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.21, and 5.22 hold, and if

\[ E \left[ a_n^{1/2} u_{nt}^3 \right] = O(1), \]

then the bias of \( \hat{\theta}_n \) through terms of \( O_p(a_n^{-1}) \) is

\[-\frac{1}{2} \left\{ b^0 \right\}^{-1} b^0 \left\{ \frac{b_n}{n} \sum_{t=1}^{b_n} W' \text{tr} \left[ x_{xt} x_{xt} u_{unt} \right] \right\}. \] (5.101)
By Theorem 5.7, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, and 5.23,

\[ \hat{\Sigma}_n - \Sigma_n = o_p(\max[a_n^{-3/2}, n^{-1/2}a_n^{-1/2}]) \]  \hspace{1cm} (5.102)

Finally, by Theorem 5.8, under Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25,

\[ \hat{\Sigma}_t - \Sigma_t = \delta_{nt} + o_p(\max[n^{-1/2}, a_n^{-1}]) \]  \hspace{1cm} (5.103)

2. The estimator

The covariance matrix of the limiting distribution of \( \hat{\beta}_n \) given in (5.99) has the form of the covariance matrix of the ordinary least squares estimator with heteroscedastic residuals. The variance of \( [Y_{nt} - f(X_{nt}; \beta^0)] \) is approximately \( \sigma^2_{vnt} \), which can be estimated using \( \hat{\Sigma}_n \) and \( \hat{\Sigma}_t \). Also, we notice from (5.101) that the source of the bias of \( \hat{\beta}_n \) is the same as that for the maximum likelihood type estimator discussed in Chapter IV with \( \Sigma_{uun} \) replacing \( A_{nt}^{-1} \). Thus, the bias of \( 0(a_n^{-1}) \) may be removed by an adjustment similar to that considered in Chapter IV. Even though there are more than one possible type of adjustment, we consider only one of such adjustments in this chapter. We choose the adjustment based on the local quadratic approximation because of its relative simplicity.

Let
\[ Y^*_{nt} = \hat{Y}_{nt} + d\hat{Y}_{nt}, \]

where

\[ d\hat{Y}_{nt} = \frac{1}{2} \text{tr}[\hat{g}_{xx}(\hat{x}_t; \hat{g}_n)\hat{\Sigma}_{uun}], \]

\[ \hat{\Sigma}_{uun} = \begin{bmatrix} \hat{\Sigma}_{een} & \hat{\Sigma}_{eun} \\ \hat{\Sigma}_{uen} & \hat{\Sigma}_{uun} \end{bmatrix}. \]

The modified instrumental variable estimator \( \hat{g}_n \) is defined to be the value of \( g \) contained in \( \hat{g} \) that minimizes

\[ Q_n(\hat{g}) = \left\{ b_n^{-1} \sum_{t=1}^{b_n} [Y^*_{nt} - f(\hat{x}_t; \hat{g}_n)]\hat{\Sigma}_{nt}\right\}^{-1} \left\{ b_n^{-1} \sum_{t=1}^{b_n} \hat{\Sigma}_{vnt} W' W_t\right\}^{-1} \]

\[ \left\{ b_n^{-1} \sum_{t=1}^{b_n} W_t [Y^*_{nt} - f(\hat{x}_t; \hat{g}_n)]\right\}, \quad (5.104) \]

where

\[ \hat{\Sigma}_{vnt} = [1, - F_{x}(\hat{x}_t; \hat{g}_n)]\hat{\Sigma}_{nt}[1, - F_{x}(\hat{x}_t; \hat{g}_n)]', \]

The weight matrix.
\[ b_n^{-1} \sum_{t=1}^{b_n} \sigma^2_{vnt} \tilde{W}_v^{nt} \tilde{W}^{nt} \]

estimates the variance of

\[ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(x_{nt}; \beta_0)] \tilde{W}^{nt} \cdot \]

3. Preliminary lemmas

We present three lemmas which will be used to study the asymptotic properties of \( \hat{\beta}_n \).

**Lemma 5.2.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24 and 5.25 hold. Then,

\[ b_n^{-1} \sum_{t=1}^{b_n} (\hat{x}_t - \tilde{x}_t')(\hat{x}_t - \tilde{x}_t) = o_p(a_n^{-1}) \cdot \]

**Proof.** By (5.93) and (5.94),

\[ \text{tr}\{\hat{\Sigma}_{vnt}^{-1} b_n^{-1} \sum_{t=1}^{b_n} (x_{nt} - \hat{x}_t)(x_{nt} - \hat{x}_t)\} < b_n^{-1} \sum_{t=1}^{b_n} D_n(\tilde{x}_t^0) \cdot \]

(5.105)
By (5.90), (5.100), (5.102), and Assumption 5.13,

\[ b_n \sum_{t=1}^{b_n} D_n(x_t^0) = \text{tr}\{ \hat{\Sigma}_n^{-1} [Y_{nt} - f(x_t^0; \hat{\beta}_n), u_{nt}]' [Y_{nt} - f(x_t^0; \hat{\beta}_n), u_{nt}] \} \]

\[ = \text{tr}\{ [a_n \hat{\Sigma_n}]^{-1} b_n \sum_{t=1}^{b_n} a_n (e_{nt}, u_{nt})' (e_{nt}, u_{nt}) \} \]

\[ + O_p(\max\{n^{-1/2}, a_n^{-1/2}\}) \]

\[ = O_p(1). \quad (5.106) \]

Since by (5.102) each element of \( \hat{\Sigma}_{uun} \) is \( O_p(a_n^{-1}) \), there is a matrix \( \hat{\Sigma}_n \) such that

\[ (a_n \hat{\Sigma}_n)^{-1} = \hat{\Sigma}_n^{-1} \quad (a_n \hat{\Sigma}_n)^{-1} \]

\[ \hat{\Sigma}_n = O_p(1). \quad (5.107) \]

Hence, by (5.105) and (5.106),

\[ \text{tr}\{ b_n^{-1} \sum_{t=1}^{b_n} \hat{\Sigma}_n (x_{nt} - \hat{x}_t)' (x_{nt} - \hat{x}_t) \hat{\Sigma}_n \} = O_p(a_n^{-1}). \]
Thus, the average of $b_n$ values of squares of the elements of 
$(\hat{X}_{nt} - \hat{X}_t)^T_n$ is $0_{a_n}$. Therefore, by the Cauchy-Schwarz inequality,

$$b_n^{-1} \sum_{t=1}^{b_n} (\hat{X}_{nt} - \hat{X}_t)'(\hat{X}_{nt} - \hat{X}_t)^T_n = 0_{a_n}.$$  
(5.108)

It follows from (5.107) and (5.108) that

$$b_n^{-1} \sum_{t=1}^{b_n} (\hat{X}_{nt} - \hat{X}_t)'(\hat{X}_{nt} - \hat{X}_t) = 0_{a_n}.$$  
(5.109)

We observe that

$$b_n^{-1} \sum_{t=1}^{b_n} (\hat{X}_t - \hat{X}_t)^t(\hat{X}_t - \hat{X}_t) - b_n^{-1} \sum_{t=1}^{b_n} (\hat{X}_{nt} - \hat{X}_t)'(\hat{X}_{nt} - \hat{X}_t) - b_n^{-1} \sum_{t=1}^{b_n} (\hat{X}_{nt} - \hat{X}_t)'u_{nt}$$

$$- b_n^{-1} \sum_{t=1}^{b_n} u_{nt}'(\hat{X}_{nt} - \hat{X}_t) + b_n^{-1} \sum_{t=1}^{b_n} u_{nt}u_{nt}.$$  
(5.110)

By Assumption 5.1a,
By (5.109), (5.111) and the Cauchy-Schwarz inequality,

\[
\sum_{t=1}^{b} a_n^{-1} u_n' u_n = O_p(a_n^{-1}).
\]

The result follows from (5.109) through (5.112). \(\square\)

**Lemma 5.3.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

\[
\sum_{t=1}^{b} a_n^{-1} \left( \hat{x}_n - \hat{x}_t \right)' u_n = O_p(a_n^{-1}).
\]

**Proof.** We observe that

\[
\tilde{f}_x^{(\hat{x}_t; \hat{\theta}_n)} - \tilde{f}_x^{(\hat{x}_t; \hat{\phi}_n)} = \tilde{f}_x^{(\hat{x}_t; \hat{\phi}_n)}(\hat{x}_t - \hat{x}_t)' + \tilde{f}_x^{(\hat{x}_t; \hat{\theta}_n)}(\hat{\sigma}_n - \tilde{\phi}_n),
\]

(5.113)

where \(\hat{x}_t^*\) is on the segment joining \(\hat{x}_t^0\) and \(\hat{x}_t\), and \(\tilde{\phi}_n^*\) is on the
segment joining $\xi^0$ and $\hat{\xi}_n$ . Hence, the result follows from (5.100), (5.113), Assumptions 5.13 and 5.15, the Cauchy–Schwarz inequality, and Lemma 5.2.

**Lemma 5.4.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.17, 5.19, 5.23, 5.24, and 5.25 hold. Then,

$$b_n^{-1} \sum_{t=1}^{b_n} [a_n (\hat{\sigma}^2_{vnt} - \sigma^2_{vnt})]^2 = o_p (\max [b_n^{-1}, a_n^{-1}]) .$$

**Proof.** We observe that

$$\hat{\sigma}^2_{vnt} - \sigma^2_{vnt} = [1, - \hat{f}_x (\hat{x}_t; \hat{\beta}_n)] (\hat{\xi}_n - \xi_n) [1, - \hat{f}_x (\hat{x}_t; \hat{\beta}_n)]'$$

$$- [1, - \hat{f}_x (\hat{x}_t; \hat{\beta}_n)] \xi_n [0, \hat{f}_x (\hat{x}_t; \hat{\beta}_n) - f^0_{xt}]'$$

$$- [0, \hat{f}_x (\hat{x}_t; \hat{\beta}_n) - f^0_{xt}] \xi_n [1, - f^0_{xt}]' .$$

Hence, by Assumptions 5.1a and 5.8, for some constants $K_1^\theta$ and $K_2^\theta$ ,

$$|a_n (\hat{\sigma}^2_{vnt} - \sigma^2_{vnt})| < K_1^\theta \| \text{vech}[a_n (\hat{\xi}_n - \xi_n)] \| + K_2^\theta \| \hat{f}_x (\hat{x}_t; \hat{\beta}_n) - f^0_{xt} \| .$$

Therefore, the result follows from (5.102), Lemma 5.3, and the Cauchy–Schwarz inequality. \qed
4. Consistency

To show the consistency of the modified instrumental variable estimator \( \tilde{\theta}_n \), we introduce an additional assumption.

**Assumption 5.26.** The limit

\[
\lim_{n \to \infty} \frac{m_n}{m} = \bar{m}
\]

is positive definite, where

\[
m_n = b_n^{-1} \sum_{t=1}^{n} a_n v_{nt}^2 W_t W_t'
\]

Note that the existence of the limit in Assumption 5.26 follows from Assumptions 5.1a and 5.16a. But, the nonsingularity of the limit \( \bar{m} \) will be needed for the consistency of \( \tilde{\theta}_n \) defined in (5.104). In the following theorem, we show the consistency of \( \tilde{\theta}_n \). Note that Assumptions 5.20a and 5.25 are used, but Assumption 5.16a is not used to obtain the consistency.

**Theorem 5.10.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.23 through 5.26 hold. Then,

\[
\tilde{\theta}_n \xrightarrow{p} \theta^0
\]

**Proof.** The estimator \( \tilde{\theta}_n \) minimizes over \( \Omega \).
\[ Q^*_n (g) = a_n^{-1} Q_n^* (g) \]

\[ = \left\{ b_n^{-1} \sum_{t=1}^{b_n} \left[ Y_{nt} - f(\hat{X}_{nt}; \hat{g}) + d_{nt} \hat{Y}_{nt} \right] \right\}^{-1} \]

\[ \left\{ b_n^{-1} \sum_{t=1}^{b_n} \left[ Y_{nt} - f(\hat{X}_{nt}; \hat{g}) + d_{nt} \hat{Y}_{nt} \right] \right\}^{-1} \]

\[ (5.114) \]

where

\[ \hat{m}_n = b_n^{-1} \sum_{t=1}^{b_n} a_n \sigma^2 \hat{w}_t \hat{w}_t^\top \]

Let \( W_{nti} \) be the \( i \)-th element of \( W_{nt} \), and let \( \hat{m}_{nij} \) and \( m_{nij} \) be the \((i,j)\)-th elements of \( \hat{m}_n \) and \( m_n \), respectively. Then, by Assumption 5.20a, Lemma 5.4, and the Cauchy-Schwarz inequality, for \( i, j = 1, 2, \ldots, p, \)

\[ |\hat{m}_{nij} - m_{nij}| = b_n^{-1} \sum_{t=1}^{b_n} a_n (\hat{\sigma}^2_{vnt} - \sigma^2_{vnt}) W_{nti} W_{ntj} \]

\[ < \left\{ b_n^{-1} \sum_{t=1}^{b_n} [a_n (\hat{\sigma}^2_{vnt} - \sigma^2_{vnt})] \right\}^{1/2} \left\{ b_n^{-1} \sum_{t=1}^{b_n} (w_{nti})^4 \right\}^{1/4} \left\{ b_n^{-1} \sum_{t=1}^{b_n} (w_{ntj})^4 \right\}^{1/4} \]
Thus,

\[ \hat{m}_n - m = o_p(\max\{b_n^{-1/2}, a_n^{-1/2}\}) \quad (5.115) \]

By Assumption 5.26,

\[ \lim_{n \to \infty} m_n = \hat{m} \quad (5.116) \]

It follows from (5.115) and (5.116) that

\[ \hat{m}_n = \hat{m} + o_p(1) \quad (5.117) \]

By (5.102) and Assumptions 5.4a and 5.15,

\[ b_n^{-1} \sum_{t=1}^{b_n} |(d^\hat{Y}_{nt})W_{nt}| < \frac{1}{2} K_4 \operatorname{tr}\{\hat{W}^{\hat{W}_un} b_n^{-1} \sum_{t=1}^{b_n} |W_{nt}|}\]

\[ = o_p(a_n^{-1}) \quad (5.118) \]

By (5.114), (5.117), (5.118), and the argument used in the proof of Theorem 5.1, for every \( \varepsilon > 0 \) there is an \( N(\varepsilon) \) such that if \( n > N(\varepsilon) \) then for all \( \xi \) in \( \xi \)
\[
P\left[ Q_n^{**}(\theta) - (\theta - \theta_0)' h(\theta^*) \bar{m}^{-1}[h(\theta^*)]'(\theta - \theta_0) > \varepsilon \right] < \varepsilon, \quad (5.119)
\]

where \( \theta^* \) is on the segment joining \( \theta_0 \) and \( \theta \). Note that by Assumption 5.26 \( \bar{m} \) is positive definite. Hence, by the argument used in (5.26), the smallest eigenvalue of \( h(\theta) \bar{m}^{-1}[h(\theta)]' \) is bounded below for all \( \theta \) in \( \Theta \) by a positive constant. Thus, the consistency of \( \hat{\theta}_n \) follows from the argument used in the proof of Theorem 5.1.

5. Limiting distribution

Since the estimator \( \hat{\theta}_n \) involves the term \( dY_{nt} \) which is a function of \( f_{xx}(\hat{\xi}_t; \hat{\theta}_n) \), we need assumptions on the third derivative of \( f \) to derive the limiting distribution of \( \hat{\theta}_n \).

Assumption 5.27. Let \( f_{xxi}(z; \theta) \) be the i-th row of \( f_{xx}(z; \theta) \), \( i = 1, 2, \ldots, q \). Then,

\[
f_{xxi} (z; \theta) = \frac{\partial}{\partial \theta} f_{xxi}(z; \theta)
\]

exists and is continuous on \( \Gamma \cap \Theta \), for \( i = 1, 2, \ldots, q \).

Assumption 5.28. There exists a constant \( K_7 \) such that for every \( i = 1, 2, \ldots, q \), all \( \theta \) in \( \Gamma \), and all \( \theta \) in \( \Theta \), the absolute value of each element of \( f_{xxi} (z; \theta) \) is bounded by \( K_7 \).

We recall that the condition \( a_{n}^{-1} = o(n^{-1/2}) \) was required to obtain the limiting distribution of \( \frac{1}{n^{1/2}} (\hat{\theta}_n - \theta_0) \). The next theorem shows that the condition \( a_{n}^{-1} = o(n^{-1/3}) \) is enough to derive the limiting distribution of \( \frac{1}{n^{1/2}} (\hat{\theta}_n - \theta_0) \).

Theorem 5.11. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a,
5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold.

Assume that

\[ a_n^{-1} = o(n^{-1/2}) \]  \hspace{1cm} (5.120) \]

Then,

\[ \frac{1}{\sqrt{n}} (\bar{E}_n - \bar{E}) \xrightarrow{L} N(Q, V^*) \]

where

\[ V^* = \{h^0 \bar{m}^{-1} h^0'\}^{-1} \]

\[ h^0 = h(\bar{g}^0) = \lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{\infty} \bar{g}_t W_t \]

\[ \bar{m} = \lim_{n \to \infty} b_n^{-1} \sum_{t=1}^{\infty} a_n \sigma^2 W_t W_t' \]

**Proof.** We observe that for \( i = 1, 2, \ldots, k \),

\[
{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \tilde{g}_n)] W_{nt}} \sim_{n \to \infty} \{b_n^{-1} \sum_{s=1}^{b_n} \tilde{W}'_n f_i(X_{ns}; \tilde{g}_n)\}
\]

\[
= \{b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \bar{g}^0)] W_{nt} \} \sim_{n \to \infty} \{b_n^{-1} \sum_{s=1}^{b_n} \tilde{W}'_n f_i(X_{ns}; \bar{g}^0)\}
\]
where \( \hat{g}_m^* \) is on the segment joining \( \hat{g}^0 \) and \( \hat{g}_m \). By the argument used to obtain (5.34), for every \( \alpha > 0 \) and \( i = 1, 2, \ldots, k \),

\[
\{ b_n^{-1} \sum_{t=1}^{b_n} [Y_{nt} - f(X_{nt}; \hat{g}_m^* )] \hat{w}_{nt} \} \overset{m_n}{\sim} \{ b_n^{-1} \sum_{s=1}^{b_n} \hat{w}_{ns}^t f_i(X_{ns}; \hat{g}_m^* ) \} = o_p(n^{-q}).
\]

Also, by Assumptions 5.21 and 5.27,

\[
Y_{nt}^* - f(X_{nt}; \hat{g}_m^* ) = v_{nt}^* - \frac{1}{2} u_{nt}^2 e_{nt}^0 - u_{nt}^t f_i(x_{nt}; \hat{g}_m^* ) u_{nt}^t u_{nti} + \frac{1}{2} \text{tr} \left[ f_{xxt}^0 + \sum_{i=1}^q f_{xxix}^i (x_{nt}; \hat{g}_m^* ) (x_{ti} - x_{0i}^0 ) \right] + \sum_{j=1}^k f_{xxtj}^j (x_{nt}; \hat{g}_m^* ) (\hat{g}_n^* - \hat{g}_m^* ) \hat{e}_{uni}.
\]
\[
\begin{align*}
\nu^*_{nt} &= -\frac{1}{2} \text{tr}\{g^0_{xx;\nu} (u'_t u_{nt} - E_{uun})\} \\
&\quad + \frac{1}{2} \text{tr}\{g^0_{xx;\nu} (\hat{E}_{uun} - E_{uun})\} \\
&\quad - \frac{1}{6} \sum_{i=1}^{q} u_{nt} g_{xx;\nu} (x^*_i; \hat{g}^*)_i u_{nt} u_{nt i} \\
&\quad + \frac{1}{2} \text{tr}\{\left[ \sum_{i=1}^{q} g_{xx;\nu} (x^*_i; \hat{g}^*)_i (x^*_i - x^0_i) \right] \} \\
&\quad + \sum_{j=1}^{q} g_{xx;\nu} (x^*_i; \hat{g}^*)_j (\hat{g}_{nj} - \hat{g}^0_j) \hat{E}_{uun},
\end{align*}
\]

where $x^*_i$ is on the segment joining $x^0_i$ and $x^*_i$, $x^*_i$ is on the segment joining $x^0_i$ and $\hat{x}_t$, $\hat{g}^*$ is on the segment joining $g^0_i$ and $\hat{g}_n$, $\hat{g}_n$, $x^0_t$ and $x^0_i$ are the $i$-th elements of $x^*_t$ and $x^0_t$, and $\hat{g}_n$ and $\hat{g}^0_j$ are the $j$-th elements of $\hat{g}_n$ and $\hat{g}^0_j$, and

\[
\nu^*_{nt} = e_{nt} - g^0_{xx;\nu} u'_t .
\]

By Assumptions 5.4a, 5.15, and 5.17,
Thus, 

\[ E\left[ b_n^{-1} \sum_{t=1}^{b_n} \text{tr}[\widehat{\xi}_{xxt} (u_t' u_t - \Sigma_{uun})] \mathbf{W}_{nt}^2 \right] = 0(b_n^{-1} a_{-n}^{-2}) . \]  

By Assumptions 5.1a and 5.17, for 1, j = 1, 2, ..., q,

\[ E\left[ |u_{nti} u_{ntj}|^{2} \right] < [E\left[ u_{nti}^{2} u_{ntj}^{2} \right] E\left[ |u_{nti}|^{2} \right] E\left[ |u_{ntj}|^{2} \right]]^{1/2} \]

\[ = 0(a_{-n}^{3/2}) . \]

Therefore, by Assumptions 5.4a and 5.22, for i = 1, 2, ..., q,
Since

\[ \hat{\Sigma}_{uun} = O_p\left( a_n^{-1} \right) \]  \hspace{1cm} (5.127)

it follows from Assumptions 5.4a and 5.28 and Lemma 5.2 that for

\[ i = 1, 2, \ldots, q, \]

\[ b_n^{-1} \sum_{t=1}^{b_n} \text{tr}\left( f_{xxix} (x^*_t; \beta^*_x) (x^* - x^*_0) \hat{\Sigma}_{uun} \psi_{nt} \right) W_t = O_p\left( a_n^{-3/2} \right) . \]  \hspace{1cm} (5.128)

By (5.100), (5.127), and Assumptions 5.4a and 5.28, for \( j = 1, 2, \ldots, k, \)

\[ b_n^{-1} \sum_{t=1}^{b_n} \text{tr}\left( f_{x} (x^*_t; \beta^*_x) (\hat{\beta}_{nj} - \beta^*_0) \hat{\Sigma}_{uun} \psi_{nt} \right) W_t = O_p\left( \max\left[ a_n^{-1/2} a_n^{-1}, a_n^{-2} \right] \right) . \]  \hspace{1cm} (5.129)

Since by Assumption 5.26,
\[
V\left(\sum_{t=1}^{b_n} v_n^* \tilde{w}_{nt} \right) = b_n \sum_{t=1}^{b_n} \sigma^2 v_n^* \tilde{w}_{nt} = 0(\sigma^2) ,
\]

It follows from (5.120), (5.123) through (5.126), (5.128), (5.129), and (5.130) that

\[
b_n^{-1} \sum_{t=1}^{b_n} v_n^* \tilde{w}_{nt} = o_p(n^{-1/2}) . \tag{5.130}
\]

We have

\[
b_n^{-1} \sum_{t=1}^{b_n} v_n^* \tilde{w}_{nt} = o_p(n^{-1/2}) . \tag{5.130}
\]

It follows from (5.120), (5.123) through (5.126), (5.128), (5.129), and (5.130) that

\[
b_n^{-1} \sum_{t=1}^{b_n} [v_n^* - f(\bar{z}_n ; \theta^0)] \tilde{w}_{nt} = b_n^{-1} \sum_{t=1}^{b_n} v_n^* \tilde{w}_{nt} + o_p(\max\{n^{-1/2} s_n^{-1/2} , a_n^{-3/2}\})
\]

\[
= b_n^{-1} \sum_{t=1}^{b_n} v_n^* \tilde{w}_{nt} + o_p(n^{-1/2})
\]

\[
= o_p(n^{-1/2}) . \tag{5.131}
\]
We observe that

$$b_n^{-1} \sum_{t=1}^{b_n} [y_{nt}^* - f(x_{nt}; \xi^*)]w_{nt} = b_n^{-1} \sum_{t=1}^{b_n} [y_{nt}^* - f(x_{nt}; \xi^0)]w_{nt}$$

$$- (\xi^* - \xi^0)'b_n^{-1} \sum_{t=1}^{b_n} \xi^0 (x_{nt}; \xi^*) w_{nt},$$

(5.132)

where $\xi^*_n$ is on the segment joining $\xi^0$ and $\xi^*$. By Assumptions 5.1a, 5.4a, 5.9, and 5.13,

$$b_n^{-1} \sum_{t=1}^{b_n} \xi^0 (x_{nt}; \xi^*_n) w_{nt} = b_n^{-1} \sum_{t=1}^{b_n} [\xi^0 (x_{nt}; \xi^*) w_{nt} + u_{nt}^\xi (x_{nt}; \xi^*) w_{nt}]$$

$$= o_p(1),$$

(5.133)

where $x_{nt}^*$ is on the segment joining $x_t^0$ and $x_{nt}$. Therefore, it follows from (5.100), (5.131), (5.132), and (5.133) that

$$b_n^{-1} \sum_{t=1}^{b_n} [y_{nt}^* - f(x_{nt}; \xi^*_n)]w_{nt} = o_p(\max\{n^{-1/2}, a_n^{-1}\}).$$

(5.134)
Using Assumptions 5.4a and 5.13, we have

\[ b_n \sum_{s=1}^{b_n} f_{\beta_i}(x_{ns}; \xi^*) = o_p(1). \]  

(5.135)

By the argument used to obtain (5.44) and (5.45) in the proof of Lemma 5.1,

\[ b_n \sum_{t=1}^{b_n} f_{\beta}(x_{nt}; \xi^*)w_{nt} = b^0 + o_p(1), \]  

(5.136)

\[ b_n \sum_{t=1}^{b_n} F_{\beta}(x_{nt}; \xi^0)w_{nt} = b^0 + o_p(1). \]  

(5.137)

By (5.117) and Assumption 5.26,

\[ \bar{w}_n^{-1} = \bar{w}^{-1} + o_p(1). \]  

(5.138)

It follows from (5.121), (5.122), (5.131), and (5.134) through (5.138) that
\[ \tilde{\beta}_n - \beta^0 = \left( h_0 - 1 h^0 \right)^{-1} h_0 - 1 \left( \sum_{t=1}^{b_n} W_{nt} v^*_{nt} \right) + o_p(n^{-1/2}). \] 

(5.139)

By Assumption 5.16a, Theorem 2.12, and the argument used to obtain (5.55),

\[ n^{1/2} b_n^{-1} \sum_{t=1}^{b_n} W_{nt} v^*_{nt} \xrightarrow{L} N(Q, \Sigma). \] 

(5.140)

Hence, the result follows from (5.139) and (5.140).

As we noted earlier, the modified instrumental variable estimator \( \tilde{\beta}_n \) takes into account the differences among \( \sigma^2_{vnt} \), which the instrumental variable estimator \( \hat{\beta}_n \) ignores. The following corollary shows that \( \tilde{\beta}_n \) is asymptotically more efficient than \( \hat{\beta}_n \) in the sense that the difference of the two limiting covariance matrices is positive semidefinite.

Corollary 5.11.1. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. If \( a_n^{-1} = o(n^{-1/3}) \), then

\[ n^{1/2} \left( \tilde{\beta} - \beta^0 \right) \xrightarrow{L} N(Q, \Sigma). \]
and if \( a_n^{-1} = o(n^{-1/2}) \), then

\[
n^{1/2}(\hat{\mathcal{G}}_n - \mathcal{G}^0) \xrightarrow{L} N(0, \Sigma),
\]

where \( \Sigma^* \) is given in Theorem 5.11, and \( \Sigma \) is given in (5.99).

Furthermore, the difference \( [\Sigma - \Sigma^*] \) is positive semidefinite.

**Proof.** The limiting distributions are given in Theorem 5.11 and (5.99). We observe that

\[
\Sigma - \Sigma^* = \mathcal{R}' \mathcal{M}^{-1} \mathcal{R},
\]

where

\[
\mathcal{R} = \mathcal{M} h_0' \{h_0 h_0'\}^{-1} - h_0' \{h_0 \mathcal{M}^{-1} h_0'\}^{-1}.
\]

Since \( \mathcal{M} \) is positive definite by Assumption 5.26, \( \mathcal{M}^{-1} \) is also positive definite. Thus, the result follows.

The next two theorems provide two consistent estimators of the covariance matrix of the limiting distribution of \( \hat{\mathcal{G}}_n \).

**Theorem 5.12.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19a, 5.20a, and 5.21 through 5.28 hold. Also let,

\[
\hat{\Sigma}(\hat{\mathcal{G}}_n) = \{\mathcal{M}^{-1} \mathcal{M}_n^{-1} \mathcal{M}_n\}'^{-1},
\]
where

$$\hat{\mathbf{y}} = b_n^{-1} \sum_{t=1}^{b_n} \mathcal{E}_\beta (\hat{\mathbf{y}}_t; \hat{\mathbf{y}}_n) \mathbf{W}_{nt} ,$$

and

$$\hat{\mathbf{y}} = b_n^{-2} \sum_{t=1}^{b_n} \sigma^2_{\mathbf{W}_{nt}} \mathbf{W}_{nt} \cdot$$

Then,

$$n \hat{\mathbf{y}} (\hat{\mathbf{g}}_n) \xrightarrow{P} \mathbf{Y}^* ,$$

where $\mathbf{Y}^*$ is given in Theorem 5.11.

Proof. We observe that

$$\hat{\mathbf{y}} = b_n^{-1} \sum_{t=1}^{b_n} \mathbf{W}_{nt} + b_n^{-1} \sum_{t=1}^{b_n} \mathcal{E}_\beta (\hat{\mathbf{y}}_t; \hat{\mathbf{g}}_n) (\hat{x}_t - \hat{x}_0) \mathbf{W}_{nt}$$

$$+ b_n^{-1} \sum_{t=1}^{b_n} \mathcal{E}_\beta (\hat{\mathbf{y}}_t; \hat{\mathbf{g}}_n) (\hat{\mathbf{y}}_n - \mathbf{g}_0) \mathbf{W}_{nt} ,$$

where $\hat{x}_t$ is on the segment joining $\hat{x}_0$ and $\hat{x}_t$, and $\hat{\mathbf{g}}_n$ is on the segment joining $\mathbf{g}_0$ and $\hat{\mathbf{g}}_n$. Thus, by (5.137), Lemma 5.2,
Assumptions 5.4a and 5.13,
\[ \hat{h} = h^0 + o_p(1) . \]

Also, by (5.117),
\[ n \hat{m} = \hat{m}_n \]
\[ = \hat{m} + o_p(1) . \]

Hence, the result follows.

\[ \square \]

**Theorem 5.13.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.16a, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Also, let
\[ \tilde{\nu}(\tilde{g}_n) = \left\{ \tilde{h} \tilde{\nu}_0^{-1} \hat{h}_t \right\}^{-1} , \]
where
\[ \tilde{\nu}_0 = n^{-2} \sum_{t=1}^b \tilde{w}_n^{-1} \tilde{w}_n [\tilde{y}_t - f(x_{nt}; \tilde{g}_n)^2 , \]
and \( \tilde{h} \) is defined in Theorem 5.12. Then,
\[ n \tilde{\nu}(\tilde{g}_n) \xrightarrow{P} \tilde{y}^* , \]
where $\mathbb{V}^*$ is given in Theorem 5.11.

**Proof.** By Theorem 5.11 and the argument used to obtain (5.65) in the proof of Theorem 5.3,

\[
n \tilde{\mathbb{V}}_o = m + O_p(b_n^{-1/2})
\]

\[
= \tilde{m} + o_p(1)
\]

Hence, the result follows. □

The quantity $\tilde{\mathbb{V}}_o$ seems to estimate the true variability of

\[
b_n^{-1} \sum_{t=1}^{n} \mathbb{W}'_{nt} \mathbb{V}^*
\]

more efficiently than the quantity $\tilde{m}$. Also, the estimator $\tilde{\mathbb{V}}(\tilde{e}_n)$ has a form analogous to the estimator of the covariance matrix of the two-stage least squares estimator. Thus, we prefer $\tilde{\mathbb{V}}(\tilde{e}_n)$ to $\hat{\mathbb{V}}(\hat{e}_n)$ as an estimator of the limiting covariance matrix of $\tilde{e}_n$.

**6. Bias**

As shown in (5.101), the instrumental variable estimator $\hat{e}_n$ has a bias of $O_p(a_n^{-1})$ which is due to the nonlinearity of the functional form. The next theorem shows that the modified instrumental variable estimator $\tilde{e}_n$ does not have such a bias. That is $\tilde{e}_n$ is unbiased up to the terms of $O_p(a_n^{-1})$. 
Theorem 5.14. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then, the expectation of the expansion of \((\hat{\gamma}_n - \gamma^0)\) through terms of \(O_p(a_n^{-1})\) is zero.

Proof. Since

\[
\max[n^{-1/2}a_n^{-1/2}, a_n^{-3/2}] = o_p(a_n^{-1})
\]

we have from (5.131) that

\[
b_n^{-1} \sum_{t=1}^{b_n} [y^*_{nt} - f(x_{nt}; \gamma^0)]\hat{w}_{nt} = b_n^{-1} \sum_{t=1}^{b_n} y^*_{nt} + o_p(a_n^{-1}) .
\]

Hence, the result follows from (5.121), (5.122), and (5.134) through (5.138). □

F. Modified Estimators of the Error Covariance Matrix and the True Value

We observe from the expressions (5.121), (5.131), and (5.134) that without the assumption \(a_n^{-1} = o(n^{-1/3})\) the modified instrumental variable estimator \(\hat{\gamma}_n\) satisfies

\[
\hat{\gamma}_n - \gamma^0 = O_p(\max\{n^{-1/2}, a_n^{-3/2}\}) .
\]  
(5.141)
As we discussed in the previous section, the estimator $\tilde{\xi}_n$ has better properties than the instrumental variable estimator $\hat{\xi}_n$. In Sections C and D, we introduced the estimation procedures based on $\tilde{\xi}_n$ for the error covariance matrix and the unknown true values. Hence, we may expect to obtain improved estimators of $\xi_n$ and $x^0_t$ by applying such procedures with $\tilde{\xi}_n$ replacing $\hat{\xi}_n$ and $y^*_n$ replacing $y_{nt}$. The modified estimator $\tilde{\xi'}_n$ of $\xi_n$ is defined by

$$\text{vech} \tilde{\xi'}_n = \left\{ \sum_{t=1}^{b_n} \tilde{\eta}'_{nt} \tilde{\xi}_n \right\}^{-1} \sum_{t=1}^{b_n} \tilde{\eta}'_{nt} \tilde{\xi}_n \ ,$$

where

$$\tilde{\eta}_{nt} = (\tilde{\varphi}_{nt} \otimes \tilde{\varphi}_{nt}) \tilde{\xi}_{q+1} \ ,$$

$$\tilde{\xi}_{nt} = [1, -\xi_{nt} (\tilde{x}_{nt}, \tilde{\xi}_n)] \ ,$$

$$\tilde{v}_{nt} = y^*_n - f(x_{nt}; \tilde{\xi}_n) \ .$$

Note that we have replaced $\tilde{x}_{nt}$ in $\tilde{\eta}_{nt}$ of (5.78) by $\tilde{x}_{nt}$. Since $\tilde{\eta}_{nt}$ or $\tilde{\eta'}_{nt}$ are used as an estimator of $\eta_{nt}$ in (5.77), we use $\tilde{x}_{nt}$ which improves over $\tilde{x}_{nt}$ as an estimator of $x^0_t$. We expect $\tilde{v}_{nt}$ to be a better estimator of $v^*_{nt} = e_{nt} - f^0_{nt} u'_{nt}$ than $v_{nt}$ used in Section C.
We observe that $\delta_{nt}$ and $v_{nt}^*$ are uncorrelated, and are independent if $(e_{nt}^*, u_{nt}^*)$ are normally distributed. This observation provides an additional intuitive justification for the use of least squares method in (5.142), since errors in $\tilde{\eta}_{nt}$ are approximately functions of $\delta_{nt}$.

The $\tilde{\eta}_{nt}$ and $\tilde{v}_{nt}$ satisfy the conditions $\hat{\eta}_{nt}$ and $\hat{v}_{nt}$ satisfied in Section C. Thus, the properties of $\tilde{\xi}_n$ obtained in Section C are also valid for $\tilde{\eta}_n$. Since errors in $\tilde{\eta}_{nt}$ as an estimator of $\eta_t$ are still of $O_p\left(a_{-1/2}^n\right)$, the estimator $\tilde{\xi}_n$ is equivalent to $\hat{\xi}_n$ up to the level of approximation considered here.

**Theorem 5.15.** Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then,

$$E - E = O_p\left(\max[a_n^{-3/2}, n^{-1/2} a_{-1/2}^n]\right).$$

If, in addition,

$$a_{-1}^n = O(n^{-1/3}),$$

then

$$n^{1/2} \text{vech} \left(\tilde{\xi}_n - \xi_n\right) \xrightarrow{L} N(0, V_{\xi}),$$

where $V_{\xi}$ is given in Theorem 5.7.

**Proof.** The results follow from (5.141) and the argument used in the
proofs of Theorem 5.6 and Theorem 5.7.

Even though the asymptotic properties of $\tilde{\Sigma}_n$ and $\hat{\Sigma}_n$ are the same, we expect $\tilde{\Sigma}_n$ to be a better estimator of $\Sigma_n$ than $\hat{\Sigma}_n$. This is because $\tilde{\nu}_{nt}$ and $\hat{\nu}_{nt}$ are superior to $\nu_{nt}$ and $\hat{\nu}_{nt}$ as estimators of $\nu_t$ and $\hat{\nu}_t$, respectively.

Using the estimator $\tilde{\Sigma}_n$, we define the modified estimator $\tilde{\Sigma}_t$ of $\Sigma_t$ to be the value of $\Sigma_t$ contained in $\tilde{\Sigma}_t$ that minimizes

$$D^*_n (\Sigma_t) = \left[ Y_{nt}^* f(\Sigma_t; \tilde{\Sigma}_n) + \Sigma_t - \tilde{\Sigma}_t \right] \Sigma_n^{-1} \left[ Y_{nt}^* f(\Sigma_t; \tilde{\Sigma}_n) + \Sigma_t - \tilde{\Sigma}_t \right]' .$$

(5.143)

As in Chapter IV, the bias of $\tilde{\Sigma}_n$ due to the curvature may help cancelling the bias in $\tilde{\Sigma}_t$. Also, the adjustment $dY_{nt}$ in (5.143) may provide a further reduction in the bias for $\tilde{\Sigma}_t$, as for the adjusted estimator based on the local quadratic approximation discussed in Chapter IV. However, the errors in $a_n \tilde{\Sigma}_n$ or $a_n \hat{\Sigma}_n$ are of

$$0_p \left( \max \left[ b_n^{-1/2}, a_n^{-1/2} \right] \right) .$$

Thus, the higher order expansions of $\tilde{\Sigma}_t$ and $\hat{\Sigma}_t$ involve the error terms in $\tilde{\Sigma}_n$ and $\hat{\Sigma}_n$, and are not as simple as those of the estimators in Chapter IV. We do not discuss the higher order results for the estimators of $\Sigma_n$ and $\Sigma_t$ in this chapter.

The property of $\tilde{\Sigma}_t$ given in Section D also holds for $\hat{\Sigma}_t$.

Theorem 5.16. Let Assumptions 5.1a, 5.2, 5.4a, 5.5 through 5.9, 5.10a, 5.11 through 5.15, 5.17, 5.19, 5.20a, and 5.21 through 5.28 hold. Then,

$$\tilde{\Sigma}_t - \Sigma_t^0 = \delta_{nt} + 0_p \left( \max \left[ n^{-1/2}, a_n^{-1} \right] \right) .$$
where \( \hat{\Delta}_{nt} \) is given in Theorem 5.8.

**Proof.** The adjustment term \( \hat{d}_{nt} \) is of \( O_p(a_n^{-1}) \). Hence, the result follows from (5.141), Theorem 5.15, and the argument used in the proof of Theorem 5.8. \( \square \)

The estimator of \( y_t^0 \) and the local approximation to \( \tilde{X}_t \) can be obtained in a manner similar to those of Section D. Also, results similar to Corollary 5.8.1, Theorem 5.9, and Corollary 5.9.1 can be derived by the argument used in Section D.

**G. An Iterative Estimation Procedure**

The results in Sections E and F suggest that a further improved estimator of \( \hat{\xi}_0 \) be obtained based on the modified estimators \( \hat{\xi}_n \), \( \hat{\xi}_n \), and \( \hat{\xi}_t \). The modified instrumental variable estimator \( \hat{\xi}_n \) was obtained by minimizing \( Q_n(\hat{\xi}) \) in (5.104) which involves the estimates \( \hat{\sigma}^{2}_{vnt} \) and \( \hat{d}_{nt} \) of \( \sigma^{2}_{vnt} \) and the bias adjustment respectively. We may expect to estimate \( \sigma^{2}_{vnt} \) and the bias adjustment more efficiently than \( \hat{\sigma}^{2}_{vnt} \) and \( \hat{d}_{nt} \) by

\[
\tilde{\sigma}^{2}_{vnt} = \left[ 1, - \frac{\partial}{\partial \xi} \hat{Q}_n(\hat{\xi}_n) \right] \tilde{\sigma}_{vnt} \left[ 1, - \frac{\partial}{\partial \xi} \hat{Q}_n(\hat{\xi}_n) \right]^T,
\]

\[
\tilde{d}_{nt} = \frac{1}{2} \text{tr} \left[ \frac{\partial^{2}}{\partial \xi \partial \xi^T} \hat{Q}_n(\hat{\xi}_n) \tilde{\xi}_{uun} \right],
\]

where \( \tilde{\xi}_{uun} \) is the portion of \( \tilde{\xi}_n \) that corresponds to \( u_{nt} \). Hence, by repeating the minimization of \( Q_n(\hat{\xi}) \) with \( \tilde{\sigma}^{2}_{vnt} \) and \( \tilde{d}_{nt} \), replacing \( \hat{\sigma}^{2}_{vnt} \) and \( \hat{d}_{nt} \), we expect to obtain an improved
estimator of $\theta^0$. We suggest that in practice such a procedure be iterated to obtain estimates of the parameters.

In the theoretical development, we have used sequences $\{a_n\}$ and $\{b_n\}$ to investigate various asymptotic properties. We assumed that the number $b_n$ of data points is known but the value of $a_n$ is unknown. Thus, we apply our estimation procedure to situations without knowledge of the decreasing error variances. When either the number of observations $N$ is large or the error variances can be considered to be small relative to total variation, the asymptotic results for our estimation procedure can be used as approximations.

We introduce an iterative estimation procedure for practical use. In the following procedure, we let $N$ be the number of observations, and denote the i-th stage estimates of $\theta^0$, $\varphi^0$, and the covariance matrix $\Sigma$ of errors in $(Y, X)$ by $\theta^{(i)}$, $\varphi^{(i)}$, and $\Sigma^{(i)}$, respectively. Also, let

$$
\Sigma^{(i)} = \begin{bmatrix}
\sigma^{(i)}_{ee} & \Sigma^{(i)}_{eu} \\
\Sigma^{(i)}_{ue} & \Sigma^{(i)}_{uu}
\end{bmatrix}.
$$

Let $I_{\text{max}}$ be a predetermined maximum number of iterations. Define $\mathbb{E}_{\mathbf{ZZ}_W}$ to be the residual mean square matrix of $\mathbf{Z}_t = (Y_t, X_t)$ obtained by regressing $\mathbf{Z}_t$ on $\mathbf{W}_t$. In our iterative procedure, we introduce certain modifications to the instrumental variable estimator and the modified instrumental variable estimator. Such modifications are
associated with a predetermined positive constant $\alpha$. The discussion of the $\alpha$-modification will be given after the presentation of the procedure.

**Procedure:**

1. Find the $\hat{\theta}$ that minimizes

$$Q(\hat{\theta}) = \sum_{t=1}^{N} \left( Y_t - f(X_t; \hat{\theta}) \right)^2 \left( \sum_{t=1}^{N} \hat{W}_t \hat{W}_t^\prime \right)^{-1} \sum_{t=1}^{N} \hat{W}_t^\prime \left( Y_t - f(X_t; \hat{\theta}) \right) + (N-p-\alpha)^{-1} \sum_{t=1}^{N} \left( Y_t - f(X_t; \hat{\theta}) \right)^2 .$$

The resulting $\hat{\theta}$ is $\hat{\theta}^{(o)}$. Let $\hat{\theta}^{(s)} = \hat{\theta}^{(o)} = Y_t$, and $i = 1$.

2. Obtain $\hat{\xi}^{(i)}$ by the ordinary least squares estimation

$$\text{vech} \hat{\xi}^{(i)} = \left( \sum_{t=1}^{N} \eta_t^{(i-1)} \eta_t^{(i-1)^\prime} \right)^{-1} \sum_{t=1}^{N} \eta_t^{(i-1)^\prime} \{ v_t^{(i-1)} \}^2 ,$$

where

$$\eta_t^{(i-1)} = \{ \hat{\xi}_t^{(i-1)} \eta_t^{(i-1)} \} \xi_{q+1} ,$$

$$\hat{\xi}_t^{(i-1)} = [1, - \hat{\xi}_t^{(i-1)} ; \hat{\xi}_t^{(i-1)}] .$$
Let $X_j$, $j = 1, 2, \ldots, q+1$, be the eigenvalues of $m''^2$ and let $Z$ be a matrix of the corresponding eigenvectors. For $j = 1, 2, \ldots, q+1$, let

$$\lambda_j^{(1)} = 0, \quad \text{if } \lambda_j^{(1)} < 0,$$

$$= \lambda_j^{(1)}, \quad \text{if } 0 < \lambda_j^{(1)} < 1,$$

$$= 1, \quad \text{if } \lambda_j^{(1)} > 1.$$

4. Compute

$$\Sigma^{(1)} = \varphi^{(1)} \Lambda^{(1)} \varphi^{(1)*},$$

where

$$\Lambda^{(1)} = \text{diag}\{\lambda_1^{(1)}, \lambda_2^{(1)}, \ldots, \lambda_{q+1}^{(1)}\}.$$

5. For each $t$, let

$$\xi_t^{(1)} = x_t - v_t^{(i-1)} \left[ \sigma^{(1)}_{vvt} \right]^{-1} \Sigma^{(1)}_{vut},$$

$$v_t^{(1)} = v_t + \frac{1}{2} \text{tr} \left[ f_{XX} \left( x_t ; \xi_{t}^{(1)} \right) \Sigma^{(1)}_{uu} \right].$$
where
\[
\sigma^{(1)}_{\nu\nu} = \varepsilon_t^{(i-1)} \varepsilon_t^{(i-1)'},
\]
\[
\varepsilon_t^{(i)} = \varepsilon_t^{(i-1)} [\varepsilon_{ue}^{(i)}, \varepsilon_{uu}^{(i)}]' .
\]

6. Find the $\hat{\beta}$ that minimizes
\[
Q^*(\hat{\beta}) = \left\{ \sum_{t=1}^{N} \left[ Y_t^{(i)} - f(x_t^{(i)}; \hat{\beta}) \right] \right\} \sum_{t=1}^{N} \sigma_{\nu\nu}^{(1)} W_t^{(i)} W_t^{(i)'}^{-1} \sum_{t=1}^{N} W_t^{(i)} [Y_t^{(i)} - f(x_t^{(i)}; \hat{\beta})]^2
\]
\[
+ (N-p-\alpha)^{-1} \sum_{t=1}^{N} \sigma_{\nu\nu}^{(i)}^{-1} [Y_t^{(i)} - f(x_t^{(i)}; \hat{\beta})]^2 .
\]

The resulting $\hat{\beta}$ is $\hat{\beta}^{(1)}$. If $i > 1$ and $|\hat{\beta}^{(i)} - \hat{\beta}^{(i-1)}|$ is less than a predetermined criterion for convergence, or if $i = I_{\text{max}}$, then stop. Otherwise, set $i = i+1$, and go to 2.

The modification of $\hat{\beta}^{(1)}$ introduced in the steps 3 and 4 ensures that the matrices $\varepsilon^{(1)}$ and $[m_{ZZ,W} - \varepsilon^{(1)}]$ are nonnegative definite.

Note that with the zero correlation between $w_t$ and $(e_t, u_t)$ the residual mean square matrix $m_{ZZ,W}$ estimates $\varepsilon_{ZZ,W} = \varepsilon_{ZZ,W} + \varepsilon$, where $\varepsilon_{ZZ,W}$ is a nonnegative definite matrix.

The modification associated with the value $\alpha$ in $Q(\hat{\beta})$ and $Q^*(\hat{\beta})$ are analogues of the modifications considered by Fuller (1977).
Fuller introduced certain modifications to the two-stage least squares estimator and the limited information maximum likelihood estimator in the linear simultaneous equation model. The modification used in $Q(\hat{\theta})$ and $Q^*(\hat{\theta})$ is the addition of a multiple of the residual sum of squares. Such a modification can be shown to be equivalent to Fuller's modification to the two-stage least squares estimator. Fuller showed that his modified two-stage least squares estimator possesses finite moments of order depending on the number of observations. We expect the minimizations of $Q(\hat{\theta})$ and $Q^*(\hat{\theta})$ to produce more stable solutions than those without modifications. For the linear simultaneous equation model, Fuller (1977) showed that the modification using $\alpha = 1$ produces an almost unbiased two-stage least squares estimator and limited information maximum likelihood estimator, and that the modification using $\alpha = 4$ produces a smaller mean squared error of the limited information estimator than any positive $\alpha$ smaller than 4. For the nonlinear functional relationship model with instrumental variables, no theoretical investigation nor Monte Carlo study for the modifications has been made. We suggest that a value of $\alpha$ satisfying $1 < \alpha < 4$ be used.

We point out that our iterative procedure is not guaranteed to converge. At each stage, the function $Q^*(\hat{\theta})$ to be minimized has different weights $\omega_v^{(i)}$ based on $E_v^{(i)}$. Hence, we recommend that a moderate number be used for $I_{\text{max}}$, the maximum number of iteration.

A finite number of iterations and the modification associated with the value $\alpha$ do not change the limiting distributions of the
estimates. Thus, by Theorem 5.13, the covariance matrix of the approximate distribution of the final estimate $\hat{g}$ from the iterative procedure is estimated by

$$
\hat{\Sigma}(\hat{g}) = \left( \hat{h} \hat{\Sigma}_o^{-1} \hat{h}' \right)^{-1},
$$

where

$$
\hat{h} = \sum_{t=1}^{N} \xi_B(\hat{x}_t; \hat{g}) \hat{w}_t,
$$

and

$$
\hat{\Sigma}_o = (N-k)^{-1} \sum_{t=1}^{N} \hat{w}_t \hat{w}' [\hat{y}_t - f(\hat{x}_t; \hat{g})]^2,
$$

and $\hat{x}_t$ and $\hat{y}_t$ are the values of $x^{(i)}_t$ and $y^{(i)}_t$ at the final stage. Also, the covariance matrix of the approximate distribution of $\hat{x}_t$ is estimated by

$$
\hat{\Sigma}(\hat{x}_t) = \hat{\Sigma}_{uu} - \hat{\Sigma}_{uvt} (\hat{\sigma}_{vvt})^{-1} \hat{\Sigma}_{vut},
$$

where $\hat{\Sigma}_{uu}$, $\hat{\Sigma}_{uvt}$, and $\hat{\sigma}_{vvt}$ are the final estimates of the respective quantities in the iterative procedure.
VI. BIBLIOGRAPHY


Clutton-Brock, M. 1967. Likelihood distributions for estimating functions when both variables are subject to error. Technometrics 9:261-269.


Madansky, W. 1959. The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc. 54:173-205.


Wald, A. 1940. Fitting of straight lines if both variables are subject to error. Ann. Math Statist. 11:284-300.


VI. ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor Wayne A. Fuller. His expert guidance has provided me with a stimulating and enjoyable experience throughout four years of graduate work.

I am grateful to the members of my graduate committee: Professors David Harville, Malay Ghosh, Glen Meeden, and Fred Wright.

Thanks go to Julie Honeick for her excellent typing of this dissertation. Thanks also go to Jane Stowe, who has provided skilled typing of several drafts of the dissertation.

Helen Nelson has frequently worked over time to print several versions of the dissertation. I wish to thank her for her efficient secretarial work as well as her friendship.

Professor Gary Koch and Ms. Rhonda Robinson of University of North Carolina introduced me to the topic of Chapter V. I am grateful for their suggestion.

I am indebted to the Survey Section of the Statistical Laboratory for providing its word processor for the production of the final version of this dissertation.