The use of a weighting function in measurement error regression

Nancy Ann Eyink Hasabelnaby
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/8651

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

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

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

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" × 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The use of a weighting function in measurement error regression

Hasabelnaby, Nancy Ann Eyink, Ph.D.
Iowa State University, 1987
PLEASE NOTE:

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

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

_____________________________________________________

_____________________________________________________

_____________________________________________________

_____________________________________________________

_____________________________________________________

_____________________________________________________

UMI
The use of a weighting function in measurement error regression

by

Nancy Ann Eyink Hasabelnaby

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

1987
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>I. INTRODUCTION AND LITERATURE REVIEW</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>B. Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>1. No measurement error in the independent variable</td>
<td>2</td>
</tr>
<tr>
<td>a. Homogeneous error variances</td>
<td>3</td>
</tr>
<tr>
<td>b. Non-homogeneous error variances</td>
<td>3</td>
</tr>
<tr>
<td>2. Measurement error present in the independent variable</td>
<td>4</td>
</tr>
<tr>
<td>a. Homogeneous error variances</td>
<td>4</td>
</tr>
<tr>
<td>b. Non-homogeneous error variances</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. USE OF WEIGHTING FUNCTIONS IN THE NON-HOMOGENEOUS MEASUREMENT ERROR MODEL</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Error in the Regression Equation</td>
<td>11</td>
</tr>
<tr>
<td>B. No Error in the Regression Equation</td>
<td>33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. EXTENSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Error in the Equation and Measurement Error Variances Decreasing</td>
</tr>
<tr>
<td>B. Variance Functions and Estimated Variances as Weights</td>
</tr>
<tr>
<td>1. Simple linear regression to estimate variance functions</td>
</tr>
<tr>
<td>2. Non-linear regression to estimate variance functions</td>
</tr>
<tr>
<td>3. A note on variance functions</td>
</tr>
</tbody>
</table>
### IV. EXAMPLE: NON-HOMOGENEOUS ERROR VARIANCES
- A. Functionally Related Model 99
- B. Description of Hog Data 100
- C. Analysis of Data 101

### V. MONTE CARLO SIMULATIONS 110

### VI. COMPUTER PROGRAM FOR WEIGHTING BY ESTIMATED VARIANCES 121
- A. Model With an Error in the Equation 123
- B. Model With No Error in the Equation 126
- C. Estimated True Values and Standardized Residuals 127
- D. Second Round Estimation - Weighting by Estimated Variances 129

### VII. REFERENCES 133

### VIII. ACKNOWLEDGEMENTS 136

### IX. APPENDIX A: CONVERGENCE IN PROBABILITY AND THE WEAK LAW OF LARGE NUMBERS 137

### X. APPENDIX B: CENTRAL LIMIT THEOREMS 141
I. INTRODUCTION AND LITERATURE REVIEW

A. Introduction

The problem of estimating the parameters of a regression equation in which both the dependent and independent variables have been measured with error has been studied since the latter part of the nineteenth century. Most of the past work has been done on the univariate linear model with constant error variances. More recently, work has been done on multivariate, non-linear, and non-constant error variance models. As an extension of ordinary weighted regression, we consider the problem of using weighting functions in the univariate errors-in-variables model when there are non-constant error variances. We also consider under what conditions estimated variances may be used as weights. Theoretical results require the use of large sample theory. We also describe a computer program that constructs estimates by weighting the observations by estimated variances.

B. Literature Review

We present here the general univariate linear errors-in-variables model. Assume

\[ y_t = x_t \beta + q_t, \quad t = 1, 2, \ldots, n, \]  

(1.1)

where \( x_t \) is an \((1 \times k)\) vector, \( \beta \) is a \((k \times 1)\) vector of unknown parameters, and \( q_t \) is the error in the regression equation. The \( x_t \)
may be either fixed or random.

We assume that we are unable to observe \( z_t = (y_t, x_t) \) directly but instead we observe \( Z_t = (Y_t, X_t) \) where

\[
Y_t = y_t + w_t
\]

\[ t = 1, 2, \ldots, n \]  \hspace{1cm} (1.2)

\[
X_t = x_t + u_t
\]

with \( a_t = (w_t, u_t) \) the vector of measurement errors. We assume that the \( q_t \) are independent \((0, \sigma_q)\) random variables, \( E(a_t) = 0 \), \( E(a_t'a_t) = \Sigma_{aatt} \), with \( \text{vech} \Sigma_{aatt} = [\Sigma_{wtt}, \Sigma_{utt}, (\text{vech} \Sigma_{uutt})']' \), and the \( q_t \) are independent of \((a_j, x_j)\) for all \( t \) and \( j \).

Frequently, it is convenient to combine the two kinds of error in \( Y_t \). We denote this combined error by \( e_t = q_t + w_t \), where

\( E(e_t^2) = \sigma_{eett} \). In this notation, \( E_t = (e_t, u_t) \) and

\( E(e_t'e_t) = \Sigma_{eett} \).

We now present a literature review of some special cases of the above model.

1. No measurement error in the independent variable

First, let us review models in which \( x_t \) has been measured without error so that \( \Sigma_{uutt} = 0 \), \( t = 1, 2, \ldots, n \). There are two main cases.
### a. Homogeneous error variances

In this case, we have $E(e_t^2) = \sigma_{ee}$, $t = 1, 2, \ldots, n$, and it is also assumed that the errors are uncorrelated so that $E(e_t e_j) = 0$ for $t \neq j$. If we assume that $x_1, \ldots, x_n$ are fixed real-valued vectors, this model is called the fixed linear model. There are numerous references for the above model. See, for example, Searle (1971).

On the other hand, if $x_t = (1, x_{2t})$, $t = 1, 2, \ldots, n$, where $x_{21}, \ldots, x_{2n}$ is a sample from a distribution with finite second moments, the model is called the random linear regression model. A good reference is Graybill (1976).

### b. Non-homogeneous error variances

Assume $E(e'e) = \Sigma_{ee}$ where $e = (e_1, e_2, \ldots, e_n)$ and $x_t$, $t = 1, 2, \ldots, n$, are fixed real-valued vectors. This model is called the general linear model. Two general cases have been studied. In the first case $\Sigma_{ee}$ is known, positive definite (or known up to a scalar multiple). In this case, the best linear unbiased estimator is

$$\hat{\beta} = (X'\Sigma_{ee}^{-1}X)^{-1}X'\Sigma_{ee}^{-1}Y$$

where $X = (x_1', x_2', \ldots, x_n')'$ is of full rank, $Y = (y_1, y_2, \ldots, y_n)'$ and $\Sigma_{ee}$ is a generalized inverse of the matrix $\Sigma$.

In this case,

$$\text{Var}(\hat{\beta}) = (X'\Sigma_{ee}^{-1}X)^{-1}.$$  

For more information, see Graybill (1976).
In the second case, \( \Sigma_{ee} \) is unknown and it is assumed that an estimator of \( \Sigma_{ee} \), namely \( \hat{\Sigma}_{ee} \), is available. Then, under certain conditions an unbiased estimator is

\[
\hat{\beta} = (X' \hat{\Sigma}_{ee}^{-1} X)^{-1} X' \hat{\Sigma}_{ee}^{-1} Y,
\]

which is called the weighted regression estimator. Often \( \hat{\Sigma}_{ee} \) is given as a function of a small number of parameters and these are estimated from residuals obtained in an ordinary least squares regression. The variance of the weighted regression estimator has been studied by Williams (1967) and Bement and Williams (1969).

Finally, information on multivariate linear models can be found in Johnson and Wichern (1982), and good discussions of non-linear models can be obtained in Bard (1974) and Gallant (1986).

2. **Measurement error present in the independent variable**

We now consider models in which the independent variables have been measured with error.

a. **Homogeneous error variances**

Assume \( \Sigma_{aat} = \Sigma_{aa} \), \( t = 1, 2, \ldots, n \). If the \( x_t \), \( t = 1, 2, \ldots, n \), are fixed real-valued vectors, the model (1.1)-(1.2) is called the functional model while if \( x_t = (1, x_{2t}^\top) \), \( t = 1, 2, \ldots, n \), where the \( x_{2t} \) have finite second moments, the model is called the structural model. In either the functional or structural cases, in order to estimate \( \beta \) consistently, additional assumptions need to be
made. Generally, assumptions are made about the variances and covariances. Some of the most common assumptions are:

i) $\Sigma_{uu}$, $\Sigma_{ue}$ known

ii) $\sigma_{qq} > 0$ and an unbiased estimator of $\Sigma_{aa}$, namely $S_{aa}$, is available

iii) $q_t \equiv 0$, $\Sigma_{aa} = \sigma^2 T_{aa}$ and $T_{aa}$ known

iv) $q_t \equiv 0$, $\Sigma_{aa} = T^{-1} \Omega_{aa}$, $\Omega_{aa}$ a fixed positive semi-definite matrix, $S_{aa}$ an unbiased estimator of $\Omega_{aa}$.

For cases i) and iii) see Moran (1971) and Fuller (1987, Sections 2.2 and 2.3). For case ii), Fuller (1987, Section 2.2) considers the estimator

$$\hat{\gamma} = (M_{XX} - S_{uu})^{-1}(M_{XY} - S_{uw}),$$

where $M_{ZZ} = \frac{1}{n} \sum_{t=1}^{n} Z_t Z_t'$ and $M_{XX}$, $M_{XY}$, $S_{uu}$, and $S_{uw}$ are the appropriate submatrices of $M_{ZZ}$ and $S_{aa}$. Under certain conditions

$$n^{1/2}(\hat{\gamma} - \gamma) \xrightarrow{L} N(0, \Gamma_{\beta\beta})$$

where $\Gamma_{\beta\beta}$ is the asymptotic covariance matrix. For the multivariate result, see Dahm and Fuller (1986).
For case iv), Fuller (1987, Section 2.3) has shown that with

\[ \hat{\beta} = (\mathbf{M}_{XX} - \bar{\lambda}T^{-1}\mathbf{S}_{uu})^{-1}(\mathbf{M}_{XY} - \bar{\lambda}T^{-1}\mathbf{S}_{uw}), \]  

(1.4)

where \( \bar{\lambda} \) is the smallest root of \( |\mathbf{M}_{ZZ} - \lambda T^{-1}\mathbf{S}_{aa}| = 0 \), as \( v = nT + \infty \)

\[ \Gamma_v^{-1/2}(\hat{\beta} - \beta) \xrightarrow{L} N(0, I) \]

and \( \Gamma_v \) is the asymptotic covariance matrix.

For multivariate and non-linear results in the case of homogeneous error variances, see Amemiya (1982), Amemiya and Fuller (1984), Anderson (1984), and Fuller (1987, Sections 3.2-3.3 and Chapter 4). For the special case called the factor analysis model in which \( \text{E}(x_t u_t) = 0 \), \( t = 1, 2, \ldots, n \), and \( \mathbf{S}_{uu} \) is diagonal, see Anderson and Rubin (1956) and Lawley and Maxwell (1971).

b. Non-homogeneous error variances

In all of the above cases, for which there is measurement error in the observed independent variable \( X_t \), it is assumed that the error variances are constant. Under the assumptions that the errors are normal and that \( x_t \) is independent of \( (q_j, a_j, q_j a_j) \) for all \( t \) and \( j \), Miller (1986, Section V.C) has shown that a \( t \)-test or \( F \)-test for homogeneity of variance for

\[ v_t = e_t - u_t \hat{\beta}, \quad t = 1, 2, \ldots, n, \]
is obtained by the regression of \((\hat{v}_t - \bar{v})^2\) on \(\hat{x}_t\), where \(\hat{v}_t\) and \(\hat{x}_t\) are estimates, obtained under homogeneous variance assumptions, of \(v_t\) and \(x_t\) respectively.

For the remainder of this chapter, we concentrate on work that has been done under the assumption that error variances are non-homogeneous. We will use the notation \(\bar{g} = (1, -\bar{g}')'\).

Sprent (1966) presented a general estimator for the functional relationship with no error in the equation, i.e., \(q_t = 0\). His estimator requires that the matrices \(x_{aatt}\) be known for each observation. Sprent used the estimator \(\hat{g}\) which minimizes (w.r.t. \(g\))

\[
f(g) = \sum_{t=1}^{n} (g'x_{aatt}g)^{-1}(g'Z_tZ_tg) .
\]

He showed this estimator to be the generalized least squares estimator for the given problem. He did not discuss the distributional properties of his estimator. Dolby (1972) showed that Sprent's estimator is identical with the maximum likelihood procedure under the assumption of normal \(a_t\).

Booth (1973) developed an estimator using the quantity

\[
\bar{x}_{aa} = n^{-1}\sum_{t=1}^{n} x_{aatt}
\]

under the assumption \(q_t = 0\). His estimator minimizes (w.r.t. \(g\))

\[
g(g) = (g'\bar{x}_{aa}g)^{-1}(g'M_{zz}g) .
\]

The estimator is
\[ \hat{\beta} = (H_{XX} - \hat{\lambda}_u \hat{\Sigma}_{uu})^{-1}(H_{XY} - \hat{\lambda}_u \hat{\Sigma}_{uw}) \]

where \( \hat{\lambda} \) is the smallest root of \( |M_{ZZ} - \hat{\lambda} \hat{\Sigma}_{aa}| = 0 \) and \( \hat{\Sigma}_{uu} \) and \( \hat{\Sigma}_{uw} \) are the appropriate submatrices of \( \hat{\Sigma}_{aa} \). In addition to other assumptions, Booth assumed that the \( a_t \) have bounded third and fourth moments and that the elements of \( x_t \) are bounded by a finite constant for \( t = 1, 2, \ldots \). Under these assumptions, he showed that \( \hat{\beta} \) is consistent, and he derived the mean and variance of the limiting distribution with and without the assumption of normality.

Then, Booth (1973) also derived a two-step estimator \( \hat{\beta}^* \) which requires an initial consistent estimator such as his \( \hat{\beta} \) and also requires that each \( \hat{\Sigma}_{at} \) be known. Under the assumption of multivariate normality for the errors, he showed that \( \hat{\beta}^* \) is asymptotically normal. His estimator is

\[
\hat{\beta}^* = [n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-1}_{vtt} X'_t X_t - \lambda^* n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-2}_{vtt} \hat{\sigma}' Z'_t a \hat{\Sigma}_{utt}]^{-1} \\
\times [n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-1}_{vtt} X'_t Y_t - \lambda^* n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-2}_{vtt} \hat{\sigma}' Z'_t a \hat{\Sigma}_{utt}]
\]

where \( \hat{\sigma}_{vtt} = \hat{\sigma}' \hat{\Sigma}_{att} \hat{\sigma} \), \( \lambda^* = (1, -\hat{\beta}^{'})' \), and \( \lambda^* \) is the smallest root of

\[
|n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-1}_{vtt} Z'_t Z_t - \lambda n^{-1} \sum_{t=1}^{n} \hat{\sigma}^{-2}_{vtt} \hat{\sigma}' Z'_t a \hat{\Sigma}_{utt}| = 0.
\]
Mak (1983) investigated the distributional properties of Sprent's generalized least squares estimator. He showed that Sprent's estimator is consistent under certain conditions, and he derived the asymptotic variance of the estimator assuming that the third and fourth moments of $a_t$ exist.

Chan and Mak (1984) estimated the structural parameters (i.e., those parameters that enter the distribution of the observations for infinitely many $t$) in a multivariate linear functional relationship ($y_t$ is $r$-dimensional) with heterogeneous error variances. They assume $q_t \equiv 0$ and the $\Sigma_{aatt}$ to be known or to be given functions of the same unknown parameters $\theta = (\theta_0', \ldots, \theta_s')'$. They showed that the maximum likelihood estimators for $\theta = (\beta_0', \beta_1')$ and $\theta$ are the roots of a rather complex system of equations. In general, the system must be solved numerically, and they describe an algorithm for this purpose. They presented conditions under which $\hat{\chi} = \text{vec}(\hat{\theta}')$ is consistent and showed that when it is consistent, $n^{1/2} (\hat{\chi} - \chi) \xrightarrow{L} \mathcal{N}(0, \Gamma)$ where $\Gamma$ is the asymptotic covariance matrix. A consistent estimator of $\Gamma$ is given.

Fuller (1984) considers the case for which we have available estimator matrices of $\Sigma_{aatt}$, denoted by $\hat{\Sigma}_{aatt}$. The $\hat{\Sigma}_{aatt}$ are such that $E(\hat{\Sigma}_{aatt}) = \Sigma_{aatt}$ or $E(\hat{\Sigma}_{t=1}^n a_{aatt}) = \Sigma_{t=1}^n \Sigma_{aatt}$. He assumed that the random part of $x_t$ is a random sample from a distribution with finite $4 + \delta$ ($\delta > 0$) moments and that the $\hat{\Sigma}_{aatt}$ have bounded $2 + \delta$ moments. Fuller (1984) showed that the estimator
\[ \hat{\beta} = \hat{M}_{xx}^{-1} M_{xy} \]

where \( \hat{M}_{zz} = n^{-1} \sum_{t=1}^{n} (Z_t'Z_t - \hat{\Sigma}_{zz}) \) and \( \hat{M}_{xx} \) and \( \hat{M}_{xy} \) are the appropriate submatrices of \( \hat{M}_{zz} \) is such that

\[ n^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} N(0, \hat{M}^{-1}_{xx}GM_{xx}^{-1}) , \]

where

\[ \hat{M}_{xx} = \text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} x_t'x_t , \]

\[ G = \text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} d_t'd_t , \]

\[ d_t' = X_t'\nu_t - [\hat{\Sigma}_{uttt} - \hat{\Sigma}_{utt\hat{\beta}}] , \]

\[ \nu_t = Y_t - X_t\hat{\beta} = q_t + w_t - u_t\hat{\beta} . \]

He also presented an estimator of the asymptotic variance of \( \hat{\beta} \).
II. USE OF WEIGHTING FUNCTIONS IN THE NON-HOMOGENEOUS MEASUREMENT ERROR MODEL

Fuller's estimator (1.5) can be improved when the error variances are non-constant by the use of weights to construct a generalized least squares estimator. We now consider the use of estimated weights in a generalized form of (1.5). The $x_t$ may be either fixed or random, and we assume, as in Fuller (1984), that we have estimators of $\Sigma_{\alpha t}$. The cases of presence and absence of an error in the equation must be treated separately. Results for the two cases appear below as Theorems 2.1 and 2.2, respectively.

A. Error in the Regression Equation

We begin by considering the case for which we have an error in the equation. Discussion of the theorem and its implications follow the proof.

**Theorem 2.1** Let $y_t = x_t \beta + q_t$, $Z_t = z_t + a_t$

\[(q_t, a_t) \sim \text{Ind}(0, \text{block diag}\{\sigma_{qq}^2, \Sigma_{\alpha t}\}) \quad (2.1)\]

where $Z_t = (Y_t, X_t)$, $a_t = (w_t, u_t)$ is the vector of measurement errors, $q_t$ is the error in the equation, and $x_t$ is the $k$-dimensional row vector of true values. Assume that estimators of $\Sigma_{\alpha t}$, denoted by $\hat{\Sigma}_{\alpha t}$, $t = 1, 2, ..., n$, are symmetric positive semi-definite.
matrices. Let \( \hat{c}_t = \text{vech} \Sigma_{at} \) and \( \hat{c}_t = \text{vech} \Sigma_{aat} \). Assume \( \sigma_{qq} > 0 \) is unknown and that there is a set of estimated weights \( \hat{\pi}_t \), \( t = 1, 2, \ldots, n \), available. Let

\[
\hat{M}_{z\pi z} = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t (Z_t Z_t' - \hat{\Sigma}_{aat})
\]

and

\[
\hat{\beta} = \hat{M}_{x\pi x}^{-1} \hat{M}_{x\pi y}
\]

where \( \hat{M}_{x\pi x} \) and \( \hat{M}_{x\pi y} \) are the appropriate submatrices of \( \hat{M}_{z\pi z} \). Let \([q_t, a_t, (c_{at} - c_{at})', (x_t - \mu_{xt})]\), \( t = 1, 2, \ldots, n \), be independent with bounded \( 4 + \delta \) moments \( (\delta > 0) \), \( E\{(q_t, a_t, q_t a_t)|x_t\} = 0 \), \( E\{x_t - \mu_{xt}\} = 0 \), and

\[
\lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \hat{\pi}_t E\{\hat{c}_{at} - c_{at}\} = 0 . \tag{2.2}
\]

Let \( \{\mu_{xt}, c_{at}', \pi_t\} \) be a fixed sequence indexed by \( t \), where \( \{\pi_t\} \) is bounded above and below by fixed positive numbers. Let

\[
\text{plim } n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \gamma_{1t} \gamma_{1t}' = \text{plim } n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \chi_t \chi_t', \quad j = 1, 2 \tag{2.3}
\]

where

\[
\chi_t = [Z_t, (\text{vech} Z_t Z_t)'), (\text{vech} z_t z_t'), \hat{c}_{at}']'.
\]
Let
\[ \lim_{n \to +\infty} n^{-1} \sum_{t=1}^{n} \pi_t x_t x_t = \lim_{n \to +\infty} M_{x \times x} = \bar{M}_{x \times x}, \] (2.4)

\[ \lim_{n \to +\infty} G = \bar{G} \] (2.5)

be positive definite, where
\[ G = n^{-1} \sum_{t=1}^{n} \pi_t^2 d_t' d_t, \]

\[ d_t' = X_t' v_t - \bar{X}_{uvtt}, \]

\[ v_t = q_t + w_t - u_t, \]

\[ \bar{X}_{uvtt} = (\bar{X}_{uvtt} - \bar{X}_{uuut}), \]

and \( \bar{X}_{uvtt} \) and \( \bar{X}_{uuut} \) are submatrices of \( \bar{\Sigma}_{aatt} \). Assume
\[ \lim_{n \to +\infty} n^{-1/2} \sum_{t=1}^{n} (\pi_t' \pi_t) d_t = 0 \] (2.6)

and
\[ \lim_{n \to +\infty} n^{-1} \sum_{t=1}^{n} \pi_t u_t u_t' = \bar{M}_{\mu \times \mu}. \] (2.7)

Then, as \( n \to +\infty \)
\[ \hat{v}_{\beta \beta}^{-1/2} (\hat{\beta} - \beta) \xrightarrow{L} N(0, 1), \]

where
\[ \hat{v}_{\beta \beta} = n^{-1} \hat{M}_{\mu \times \mu} \hat{G} \hat{M}_{\mu \times \mu}, \]

\[ \hat{G} = (n - k)^{-1} \sum_{t=1}^{n} \pi_t^2 d_t' d_t, \]
\[ \hat{d}_t^l = X_t^l v_t - \hat{\Sigma}_{uvtt}, \]

\[ \hat{\Sigma}_{uvtt} = \hat{\Sigma}_{aatt}(1, - \hat{\beta}'), \]

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

**Proof.** We begin by writing

\[ n^{1/2} (\hat{\beta} - \beta) = n^{1/2} (\hat{M}^{-1} \hat{M}_{xy} - \beta) \]

\[ = n^{1/2} \hat{M}^{-1} M_{xy} \beta \]

\[ = n^{1/2} \hat{M}^{-1} [n^{-1} \sum_{t=1}^{n} \hat{\pi}(X_t'v_t - \hat{\Sigma}_{uvtt}) \]

\[ - n^{-1} \sum_{t=1}^{n} \hat{\pi}(X_t'X_t - \hat{\Sigma}_{uvtt}) \beta] \]

\[ = n^{1/2} \hat{M}^{-1} [n^{-1} \sum_{t=1}^{n} \hat{\pi}(X_t'v_t - \hat{\Sigma}_{uvtt})] \]

where \( v_t = Y_t - X_t \beta \) and \( \hat{\Sigma}_{uvtt} = \hat{\Sigma}_{autt} - \hat{\Sigma}_{uutt} \beta \). An even more compact expression is

\[ n^{1/2} (\hat{\beta} - \beta) = \hat{M}^{-1} (n^{-1} \sum_{t=1}^{n} \hat{\pi} \hat{d}_t^l), \]  

(2.8)

where \( \hat{d}_t^l = X_t^l v_t - \hat{\Sigma}_{uvtt} \). Now,
\[ \hat{M}_{xt} = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t (x_t' x_t - \hat{\Sigma}_{utt}) \]

\[ = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t (x_t' x_t + x_t' u_t + u_t' x_t + u_t' u_t - \hat{\Sigma}_{utt}) \]

\[ = n^{-1} \sum_{t=1}^{n} \pi_t x_t' x_t + n^{-1} \sum_{t=1}^{n} \pi_t (x_t' u_t + u_t' x_t + u_t' u_t - \hat{\Sigma}_{utt}) \]

\[ + n^{-1} \sum_{t=1}^{n} (\pi_t - \hat{\pi}_t) (x_t' x_t - \hat{\Sigma}_{utt}). \]

The matrix \( M_{xt} \) is

\[ n^{-1} \sum_{t=1}^{n} \pi_t x_t' x_t = \widetilde{M}_{xt} + o_p(1) \]

by Assumption (2.4). Also

\[ n^{-1} \sum_{t=1}^{n} \pi_t \left[ x_t' u_t + u_t' x_t + u_t' u_t - \hat{\Sigma}_{utt} - (\hat{\Sigma}_{utt} - \hat{\Sigma}_{utt}) \right] = o_p(1) \]

since \( \{\pi_t\} \) are bounded, \( x_t' u_t, u_t' x_t - \hat{\Sigma}_{utt} \) and \( \hat{\Sigma}_{utt} - \hat{\Sigma}_{utt} \) have \( 2 + \frac{1}{2} \delta \) moments, and \( n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\Sigma}_{utt} - \hat{\Sigma}_{utt}) = o_p(1) \) by (2.2). In addition,

\[ n^{-1} \sum_{t=1}^{n} (\pi_t - \hat{\pi}_t) (x_t' x_t - \hat{\Sigma}_{utt}) = o_p(1) \]

by Assumption (2.3). Therefore,

\[ \hat{M}_{xt} = \widetilde{M}_{xt} + o_p(1). \] (2.9)

Also,
\[ n^{-1/2} \sum_{t=1}^{n} \hat{\pi}_t \mathbf{d}'_t = n^{-1/2} \sum_{t=1}^{n} \pi_t \mathbf{d}'_t + n^{-1/2} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) \mathbf{d}'_t \]

and

\[ n^{-1/2} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) \mathbf{d}'_t = o_p(1) \text{ by Assumption (2.6)} \]

so that

\[ n^{-1/2} \sum_{t=1}^{n} \hat{\pi}_t \mathbf{d}'_t = n^{-1/2} \sum_{t=1}^{n} \pi_t \mathbf{d}'_t + o_p(1) . \]  \hspace{1cm} (2.10)

Note that \( n^{-1/2} \sum_{t=1}^{n} \mathbf{d}'_t = o_p(1) \) because \( \mathbb{E}(n^{-1/2} \sum_{t=1}^{n} \mathbf{d}'_t) = o(1) \).

See the discussion on the random variables \( \pi_t \mathbf{d}'_t \) that follows. Thus, by (2.9) and (2.10)

\[ n^{1/2} (\hat{\beta} - \beta) = n^{-1/2} \sum_{t=1}^{n} \pi_t \mathbf{d}'_t + o_p(1) . \]  \hspace{1cm} (2.11)

Consider the random variables \( \pi_t \mathbf{d}'_t \),

\[ \pi_t \mathbf{d}'_t = \pi_t (\mathbf{x}'_t \mathbf{v}_t - \mathbf{\xi}_{uvt}) \]

\[ = \pi_t [\mathbf{x}'_t q_t + \mathbf{x}'_t \mathbf{v}_t + \mathbf{u}'_t q_t + \mathbf{u}'_t \mathbf{v}_t - \hat{\mathbf{\xi}}_{uvt} - (\mathbf{x}'_t \mathbf{u}_t + \mathbf{u}'_t \mathbf{u}_t - \hat{\mathbf{\xi}}_{uvt}) \beta] \]

\[ = \pi_t [(\mathbf{x}_t - \mu_{xt})' q_t + (\mathbf{x}_t - \mu_{xt})' \mathbf{v}_t + \mu_{xt}' (q_t + \mathbf{w}_t) \]

\[ + \mathbf{u}'_t q_t + \mathbf{u}'_t \mathbf{w}_t - (\hat{\mathbf{\xi}}_{uvt} - \mathbf{\xi}_{uvt}) - \mathbf{\xi}_{uvt}] \]

\[ - \pi_t [(\mathbf{x}_t - \mu_{xt})' \mathbf{u}_t + \mu_{xt}' \mathbf{u}_t + \mathbf{u}'_t \mathbf{u}_t - (\hat{\mathbf{\xi}}_{uvt} - \mathbf{\xi}_{uvt}) - \mathbf{\xi}_{uvt}] \beta \]
which, with \( \{u_t, \pi_t, \pi_t^0\} \) fixed, is a combination of elements of 
\([q_t, a_t, (c_{at} - c_{at})'], (x_t - u_{xt})] \), \( t = 1, 2, \ldots, n \). Since the 
\([q_t, a_t, (c_{at} - c_{at})'], (x_t - u_{xt})] \) are independent, the \( \pi_t d_t \), \( t = 1, 2, \ldots, n \) are independent. Also,

\[
E(\pi_t d'_t) = E\{E[\pi_t (x_t q_t + x_t w_t + u_t q_t + u_t w_t - \hat{\Sigma}_{uwt})] \\
- \pi_t (x_t u_t + u_t w_t - \hat{\Sigma}_{uwt}) \beta | x_t] \}
\]

\[
= E\{E[\pi_t x_t (q_t + w_t - u_t \beta) + \pi_t u_t q_t | x_t] \}
\]

\[
+ E\{E[\pi_t (u_t w_t - \hat{\Sigma}_{uwt}) | x_t] \}
\]

\[
- E\{E[\pi_t (u_t u_t - \hat{\Sigma}_{uwt}) \beta | x_t] \} .
\]

The first term of the expression for \( E(\pi_t d'_t) \) is zero because 
\( E\{(q_t, a_t, q_t a_t) | x_t\} = 0 \) and \( \{\pi_t\} \) is a fixed sequence. Note that 
\( E(\pi_t d'_t) \) is not necessarily zero since \( E(a'_t a_t - \hat{\Sigma}_{aatt}) = 0 \) may not be true. However,

\[
\lim_{n \to \infty} E(n^{-1/2} \sum_{t=1}^{n} \pi_t d'_t) = 0
\]

since
\[
\lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \pi_t E(a_t' a_t - \hat{a}_{att} a_{att})
\]

\[
= \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \pi_t E(a_t' a_t - \hat{a}_{att} a_{att} + \hat{a}_{att} - \hat{a}_{att})
\]

\[
= \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \pi_t E(\hat{a}_{att} - \hat{a}_{att}) = 0
\]

by Assumption 2.2. Now,

\[
|\pi_t d_t'|^{2+1/2} = |\pi_t ((x_t - \mu_t)' q_t + (x_t - \mu_t)' w_t + \beta_t w_t + u_t q_t
\]

\[
+ (u_t w_t - \hat{u}_{utt}) - (\hat{w}_{utt} - \hat{u}_{utt}) - \pi_t \{(x_t - \mu_t) u_t
\]

\[
+ \mu_t u_t + (u_t u_t - \hat{u}_{utt}) - (\hat{u}_{utt} - \hat{u}_{utt})\} \| \|^{2+1/2}
\]

where, for a vector \( d_t \), \(|d_t|\) denotes the vector of absolute values of the elements of \( d_t \). Since \( \{\pi_t\} \) is bounded above and below by fixed positive numbers, there exist real numbers \( \pi_L \) and \( \pi_U \) such that \( 0 < \pi_L < \pi_t < \pi_U \) for all \( t \). Therefore, using the upper bound
\[
| \pi_t a_t^t |^{2+1/2} \delta < \eta_U^{2+1/2} \delta \left( |(x_t - \mu_{xt})' q_t |^{2+1/2} \delta + |(x_t - \mu_{xt})' w_t |^{2+1/2} \delta \\
+ |a_t' q_t |^{2+1/2} \delta + |\mu_{xt} q_t |^{2+1/2} \delta + |\mu_{xt} w_t |^{2+1/2} \delta \\
+ |a_t' w_t - \Sigma_{uwt} t |^{2+1/2} \delta + |\hat{\Sigma}_{uwt} t - \Sigma_{uwt} |^{2+1/2} \delta \\
+ \left[ |(x_t - \mu_{xt}) u_t |^{2+1/2} \delta + |\mu_{xt} u_t |^{2+1/2} \delta \\
+ |a_t' u_t - \Sigma_{uut} t |^{2+1/2} \delta + |\hat{\Sigma}_{uut} t - \Sigma_{uut} |^{2+1/2} \delta \right] \| \beta \|^{2+1/2} \delta \right) \\
\]

and thus

\[
E( | \pi_t a_t^t |^{2+1/2} \delta )
\]

\[
< \eta_U^{2+1/2} \delta \left( E[ |(x_t - \mu_{xt})' q_t |^{2+1/2} \delta ] + E[ |(x_t - \mu_{xt})' w_t |^{2+1/2} \delta ] \\
+ E( |a_t' q_t |^{2+1/2} \delta ) + E( |\hat{\Sigma}_{uwt} t - \Sigma_{uwt} |^{2+1/2} \delta ) \\
+ E[ |(x_t - \mu_{xt}) u_t |^{2+1/2} \delta ] \| \beta \|^{2+1/2} \delta \\
+ E( |\hat{\Sigma}_{uut} t - \Sigma_{uut} |^{2+1/2} \delta ) \| \beta \|^{2+1/2} \delta + E( |\mu_{xt} q_t |^{2+1/2} \delta ) \\
+ E( |\mu_{xt} w_t |^{2+1/2} \delta ) + E( |\mu_{xt} u_t |^{2+1/2} \delta ) \| \beta \|^{2+1/2} \delta \\
+ E( |a_t' w_t - \Sigma_{uwt} t |^{2+1/2} \delta ) + E( |a_t' u_t - \Sigma_{uut} t |^{2+1/2} \delta ) \| \beta \|^{2+1/2} \delta \right). \\
(2.12)
\]
The elements of the first 6 terms in (2.12) are all finite since 
\[ q_t, a_t, (c_{at} - c_{at}), (x_t - \mu_{xt}) \], \ t = 1, 2, \ldots, n \, , \ are \ from \ a 
distribution with bounded \( 4 + \delta \) \ moments and since \( |\beta|^{2+1/2} \delta \) is 
finite. The next three terms in (2.12) which involve \( \mu_{xt} \) are finite 
for the above reasons and since \( \mu_{xt} \) is a fixed sequence. Finally, 
consider \( E(\mid u_t'w_t - E_{uwt} \mid ^{2+1/2} \delta) \). Now,

\[ u_t'w_t = E_{uwt} + \eta_t \]

where \( \eta_t \) is a zero mean random variable with bounded \( 2 + 1/2 \delta \) 
moments since \( a_t = (w_t, u_t) \) has bounded \( 4 + \delta \) \ moments. Thus,

\[ E(\mid u_t'w_t - E_{uwt} \mid ^{2+1/2} \delta) = E(\mid \eta_t \mid ^{2+1/2} \delta) < \infty. \]

In a similar way, we can show \( E(\mid u_t'u_t - E_{uwt} \mid ^{2+1/2} \delta) < \infty. \) Therefore,
\( E(\mid \pi_t d_t \mid ^{2+1/2} \delta) \) is finite, \( t = 1, 2, \ldots, n \). However, the moments 
are not bounded because the fixed part \( \mu_{xt} \) is not bounded. Let
\( X_t = \pi_t \lambda d_t \) where \( \lambda \) is a fixed \( 1 \times k \) row vector, \( \lambda \neq 0 \). Note 
that

\[ d_t = u_t'(q_t + w_t - u_t \beta) + (x_t - \mu_{xt})(q_t + w_t - u_t \beta) \]

\[ + u_t'q_t + (u_t'w_t - E_{uwt}) - (u_t'u_t - E_{uwt}) \beta \]

\[ - (E_{uwt} - E_{uwt}) + (E_{uwt} - E_{uwt}) \beta \]
\[ = (\mu^t, I, I, -I)(\nu, \nu(x_t - \mu^t), v_t u_t - \Sigma v, \Sigma vut - \Sigma vut) \]

where \( I \) is the \( k \times k \) identity matrix. Therefore,

\[
d_t' (\text{def.}) = b_t e^t,\]

where \( b_t = (\mu^t, I, I, -I) \) and \( e^t \) is the vector of random parts of \( d_t \) listed above. Thus,

\[
X_t = \pi_t \lambda b_t e^t = c_t e^t,\]

where \( c_t = \pi_t \lambda b_t \). The \( e^t \) are independent with bounded \( 2 + \frac{1}{2} \delta \) moments because the \( d_t \) are independent and we have removed the fixed, unbounded \( \mu^t \). Now

\[
n^{-1} v_n = n^{-1} \sum_{t=1}^{n} c_t E(e_t' e_t) c_t' = n^{-1} \sum_{t=1}^{n} \pi^2 E(d_t' d_t) \lambda_t' = \lambda \lambda' \xrightarrow{P} \lambda \lambda' \]

so that \( n^{-1} v_n \) is bounded above and below for large \( n \). Let

\[
M_{cc} = n^{-1} \sum_{t=1}^{n} c_t' c_t.\]

The elements of \( M_{cc} \) are multiples of \( n^{-1} \sum_{t=1}^{n} \pi^2 \mu^t \mu^t, n^{-1} \sum_{t=1}^{n} \pi^2 \mu^t \mu^t \).
and $n^{-1} \Sigma_{t=1}^{n} \pi_{t}^2$. Note that the elements of

$$n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$$

are bounded above and below for large $n$ since the $\{\pi_{t}\}$ are bounded and $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t} \rightarrow \bar{N}_{\mu \Sigma \mu}$ by (2.7) and $n^{-1} \sum_{t=1}^{n} \pi_{t}^2$ is bounded since the $\{\pi_{t}\}$ are bounded. We now show that the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded (in absolute value).

Since the $\{\pi_{t}\}$ are bounded, the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded if the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded. Since the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded by (2.13), the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded. Now, suppose $n^{-1} \sum_{t=1}^{n} a_{t}^2$ is bounded for some $a_{t}$. Then

$$n^{-1} \sum_{t=1}^{n} |a_{t}| = n^{-1} \sum_{t=1}^{n} |a_{t}| + n^{-1} \sum_{t=1}^{n} |a_{t}|$$

$$< n^{-1} \sum_{t=1}^{n} 1 + n^{-1} \sum_{t=1}^{n} a_{t}^2$$

$$< 1 + n^{-1} \sum_{t=1}^{n} a_{t}^2$$

is bounded above. Let $a_{t}^2$ be the diagonal element of $\mu'_{t} \mu_{t}$, then the elements of $n^{-1} \sum_{t=1}^{n} \pi_{t}^2 \mu'_{t} \mu_{t}$ are bounded in absolute value. Thus, by Theorem 10.1 and its multivariate extension (see Chapter X, Appendix B),
as \( n \to \infty \)

\[
\sum_{t=1}^{n} \pi_t^2 (d_t' d_t) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_t d_t' \xrightarrow{\text{L}} N(0, I).
\]  \( (2.14) \)

We next show that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \pi_t^2 [d_t' d_t - E(d_t' d_t)] = 0.
\]

Since \( \{\pi_t\} \) is bounded above and below, it is enough to show that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} [d_t' d_t - E(d_t' d_t)] = 0.
\]

Now, the \( i \)-th element of the vector \( d_t \) is

\[
d_{ti} = \mu_{x_{ti}} v_{ti} + (x_{ti} - \mu_{x_{ti}}) v_{ti} + (u_{ti} v_{ti} - \sigma_{uv_{ti}}) - (\sigma_{uv_{ti}} - \sigma_{uv_{ti}}).
\]

Thus, \( d_t' d_t - E(d_t' d_t) \) contains terms such as

\[
\mu_{x_{ti}}^2 (v_{ti}^2 - \sigma_{uv_{ti}}^2) = \mu_{x_{ti}}^2 v_{ti}^2,
\]  \( (2.15) \)

\[
\mu_{x_{ti}} [v_{ti}^2 u_{ti} - E(v_{ti}^2 u_{ti})] = \mu_{x_{ti}} u_{ti}^3,
\]  \( (2.16) \)

and terms with all random parts such as

\[
(x_{ti} - \mu_{x_{ti}})^2 v_{ti}^2 - E((x_{ti} - \mu_{x_{ti}})^2 v_{ti}^2) = v_{ti}^4
\]  \( (2.17) \)

where we have dropped the subscript \( i \). Consider the first term listed
and let $\gamma = \frac{1}{4} \delta$. Then $E\{(r_t^2)^{1+\gamma}\}$ is bounded by the assumption of $4+\delta$ moments on the errors and estimated covariances. Let $0 < \delta < 1$. We use Theorem 5.2.3 of Chung (1974, p. 111).

Let $x_{ct} = \mu_t^2 (\nu_t^2 - \sigma_{vtt}) = \mu_t^2 r_t^2$ where $x_{ct}$ is the $x$ of Chung's theorem. Then

$$\sum_{t=1}^{n} \int_{\{x_{ct} \mid |x_{ct}| > n\}} dF(x_t) < \sum_{t=1}^{n} \mu_t^{2n-1} \int_{\{r_t \mid r_t^2 > n\}} r_t^2 dF(r_t).$$

Let $A_t = \sup_{1 \leq t \leq n} \mu_t^2$. The set of $x$ such that $|x_{ct}| = \mu_t^2 r_t^2 > n$ is a smaller set than the set $|\sup_{1 \leq t \leq n} \mu_t^2 r_t^2| > n$. Also,

$$\{x: |\sup_{t} \mu_t^2 r_t^2| > n\} = \{x: r_t^2 > (\sup_{t} \mu_t^2)^{-1} n\}
= \{x: r_t^2 > [\sup_{t} \mu_t^2]^{-1} = A_t^{-1}\}.$$

Thus,

$$\sum_{t=1}^{n} \int_{\{x_{ct} \mid |x_{ct}| > n\}} dF(x_t) < \sum_{t=1}^{n} \mu_t^{2n-1} \int_{\{r_t \mid r_t^2 > A_t^{-1}\}} r_t^2 dF(r_t)
< n^{-1} A_t^{-\gamma} \sum_{t=1}^{n} \mu_t^{2} \int_{\{r_t \mid r_t^2 > A_t^{-1}\}} (r_t^2)^{1+\gamma} dF(r_t).$$

Now, we know that $E\{(r_t^2)^{1+\gamma}\}$ and $n^{-1} \sum_{t=1}^{n} \mu_t^{2}$ are bounded. Thus, if we can show that $A_t + 0$, then
\[ \sum_{t=1}^{n} \int_{\mathcal{X}_{c}} dF_{x_{t}}(x_{c}) + 0 . \]

We have that

\[ n^{-1} \sum_{t=1}^{n} \mathbb{E} x_{t}^{2} \xrightarrow{P} \mathbb{E}_{x_{wx}} > 0 . \]

Let \( S_{n} = n^{-1} \sum_{t=1}^{n} \pi_{t} x_{t}^{2} \). Then,

\[ S_{n} = n^{-1}(n-1)S_{n-1} + n^{-1} \pi_{n} x_{n}^{2} \]

and

\[ S_{n-1} \xrightarrow{P} \mathbb{E}_{x_{wx}} . \]

Thus \( n^{-1} \pi_{n} x_{n}^{2} \xrightarrow{P} 0 \), or \( n^{-1} x_{n}^{2} \xrightarrow{P} 0 \) with \( \{\pi_{t}\} \) bounded above and below. We now need the following lemma.

**Lemma 2.1.** Let \( \{g_{t}\} \) be a sequence of r.v.'s such that

\[ n^{-1} \sum_{t=1}^{n} g_{t} \xrightarrow{P} A < \infty . \]

Then \( \sup_{1 \leq t \leq n} |g_{t}| = o_{P}(n) \).

**Proof.** Let \( S_{n} = n^{-1} \sum_{t=1}^{n} g_{t} \). Then as we have seen above,

\[ S_{n} = n^{-1}(n-1)S_{n-1} + n^{-1} g_{n} \quad \text{and} \quad n^{-1} g_{n} \xrightarrow{P} 0 \] or instead,

\[ n^{-1}|g_{n}| \xrightarrow{P} 0 . \] Now, let \( \varepsilon > 0 \) be given. By definition, there exists an integer \( N(\varepsilon) \) such that \( P\{n^{-1}|g_{n}| > \varepsilon\} < \varepsilon \) if \( n > N(\varepsilon) \).

The set for which \( m^{-1}|g_{m}| > \varepsilon \) \((m > n)\) is smaller than the set for which \( n^{-1}|g_{n}| > \varepsilon \). Therefore, we have
\[ P\{m^{-1}\left|g_n\right| > \varepsilon\} < \varepsilon \text{ for all } m > n > N(\varepsilon) . \]

Choose \( M > N(\varepsilon) \) such that

\[ P\left\{ M^{-1} \sup_{1 \leq t \leq M} \left|g_t\right| > \varepsilon \right\} < \varepsilon . \]

Then,

\[ P\left\{ n^{-1} \sup_{1 \leq t \leq n} \left|g_t\right| > \varepsilon \right\} < \varepsilon \]

for every \( n > M \). Since \( \varepsilon \) is arbitrary, the result follows.

Applying the above lemma to our situation, we have

\[ \sup_{1 \leq t \leq n} x_t^2 = o_p(n) \]

that is,

\[ n^{-1} \sup_{1 \leq t \leq n} x_t^2 \xrightarrow{P} 0 . \]

Since we also have \( \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \mu'_t \mu_t = \bar{M} \mu^T \mu \), we can use the above lemma to show

\[ n^{-1} \sup_{1 \leq t \leq n} \mu_t^2 \xrightarrow{P} 0 . \]  \hspace{1cm} (2.18)

That is, \( A_t \to 0 \). Thus, the first condition of Chung's theorem is satisfied. To show the second condition,
by the same arguments. Therefore, conditions (i) and (ii) of Chung's theorem are established and

\[
\text{plim} \ n^{-1} \sum_{t=1}^{n} \mu_{t}^{2}(v_{t}^{2} - \sigma_{vtt}^{2}) = 0.
\]

For the other two types of terms (2.16) and (2.17), we can show that the conditions are satisfied by similar arguments.

For terms of the form \( \mu_{t} s_{t}^{3} \), we note that \( E(|s_{t}^{3}|^{1+\gamma}) < \infty \) and that \( B_{t} = \sup_{1 \leq t \leq n} (n^{-1}|\mu_{t}|) + 0 \) by the following. If \( |\mu_{t}| < 1 \) for every \( t \), then \( \sup_{1 \leq t \leq n} (n^{-1}|\mu_{t}|) < n^{-1} + 0 \). If \( |\mu_{t}| > 1 \) for some \( t \), then \( \sup_{t} (n^{-1}|\mu_{t}|) < \sup_{t} (n^{-1}\mu_{t}^{2}) = A_{t} + 0 \) by (2.18).

For those terms containing only random parts as in (2.17), we note that \( E(|w_{t}^{\gamma}|^{1+1/4}) < K < \infty \) and thus

\[
\sum_{t=1}^{n} \int_{|w_{t}^{\gamma}| > n} dF_{w} < \sum_{t=1}^{n} \int_{|w_{t}^{\gamma}| > n} n^{-(1+\gamma)}|w_{t}^{\gamma}|^{1+\gamma}dF_{w},
\]

\[
< n^{-(1+\gamma)} \sum_{t=1}^{n} \int_{|w_{t}^{\gamma}| > n} |w_{t}^{\gamma}|^{1+\gamma}dF_{w}.
\]
\[ \gamma_n = \gamma_{n_k} = 0 , \]

\[ n^{-2} \sum_{t=1}^{n} \int |w_t^4|^2 \text{d}F_w = n^{-2} \sum_{t=1}^{n} \int |w_t^{1+\gamma}|w_t^{1-\gamma} \text{d}F_w \]

\[ n^{-2} \sum_{t=1}^{n} \int |w_t^{1-\gamma}|w_t^{1+\gamma} \text{d}F_w \]

\[ n^{-2} \sum_{t=1}^{n} \int |w_t^{1+\gamma}|w_t^{1-\gamma} \text{d}F_w \]

\[ \text{and the conditions of Chung's theorem are satisfied for all terms.} \]

Therefore,

\[ \text{plim } \frac{1}{n} \sum_{t=1}^{n} \pi_t^2 d_t d_t' = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \pi_t^2 \mathbb{E}(d_t d_t') . \quad (2.19) \]

Now,

\[ d_t' = X_t' \nu_t - \hat{E}_{uvtt} \]

\[ = X_t' (\gamma_t - X_t \beta) - (\hat{E}_{uvtt} - \hat{E}_{uutt} \beta) \]

\[ = X_t' \nu_t - \hat{E}_{uvtt} - (X_t' X_t - \hat{E}_{uutt})(\hat{\beta} - \beta) \]

\[ d_t' = (X_t' X_t - \hat{E}_{uutt})(\hat{\beta} - \beta) . \quad (2.20) \]
We wish to show that

\[(n - k)^{-1} \sum_{t=1}^{n} \hat{\pi}_{t}^2 \hat{d}'_{t} \hat{d}_{t} = (n - k)^{-1} \sum_{t=1}^{n} \pi_{t}^2 d'_{t} d_{t} + o_{p}(1). \quad (2.21)\]

By equality (2.20),

\[(n - k)^{-1} \sum_{t=1}^{n} \pi_{t}^2 d'_{t} d_{t} = (n - k)^{-1} \sum_{t=1}^{n} \pi_{t}^2 (\hat{X}_{t}' \hat{X}_{t} - \hat{B}_{t} d_{t})^{-1} d_{t} \]

\[= (n - k)^{-1} \sum_{t=1}^{n} \pi_{t}^2 (\hat{X}_{t}' \hat{X}_{t} - \hat{B}_{t} d_{t})^{-1} d_{t} \]

\[= (n - k)^{-1} \sum_{t=1}^{n} \pi_{t}^2 (\hat{X}_{t}' \hat{X}_{t} - \hat{B}_{t} d_{t})^{-1} d_{t} \]

Thus, it remains to show that the terms other than \( n^{-1} \sum_{t=1}^{n} \pi_{t}^2 d'_{t} d_{t} \) are \( o_{p}(1) \).

Recall (2.11) that

\[n^{1/2} (\hat{\beta} - \beta) = n^{1/2} \Xi_{n}^{-1} \sum_{t=1}^{n} \pi_{t} d'_{t} + o_{p}(1). \]

By Assumption (2.5), \( \Xi_{n}^n \pi_{t}^2 E(d'_{t} d_{t}) = o_{p}(n) \). Therefore, by (2.14) we have \( n^{-1} \sum_{t=1}^{n} \pi_{t} d'_{t} = o_{p}(n^{-1/2}) \) and thus

\[(\hat{\beta} - \beta) = o_{p}(n^{-1/2}). \quad (2.22)\]
Therefore, to complete the proof of (2.21), we need to show that

\[(n-k)^{-1} \sum_{t=1}^{n} \hat{\delta}_t^2 (x_t' x_t - \hat{\Sigma}_{uut}) \quad \text{and} \quad (n-k)^{-1} \sum_{t=1}^{n} \hat{\delta}_t^2 (x_t' x_t - \hat{\Sigma}_{uut})' (x_t' x_t - \hat{\Sigma}_{uut}) \quad \text{are } o_p(n^{1/2}) \quad \text{. Now,} \]

\[\begin{align*}
d_t &= \mathbf{x}_t' v_t - \hat{\Sigma}_{uvtt} \\
    &= \mathbf{x}_t' (y_t - \bar{x}_t) - (\hat{\Sigma}_{uvtt} - \hat{\Sigma}_{uutt}) \\
    &= (0, I) \mathbf{z}_t' \mathbf{z}_t (1, -\beta)' - (0, I) \mathbf{z}_t' \mathbf{z}_t (1, -\beta)' \\
    &= (0, I) \mathbf{z}_t' \mathbf{z}_t - \hat{\Sigma}_{aatt} (1, -\beta)' .
\end{align*}\]

Thus,

\[(n-k)^{-1} \sum_{t=1}^{n} \hat{\delta}_t^2 (x_t' x_t - \hat{\Sigma}_{uut}) \]

\[= (n-k)^{-1} \sum_{t=1}^{n} \hat{\delta}_t^2 (0, I) \mathbf{z}_t' \mathbf{z}_t - \hat{\Sigma}_{aatt} (1, -\beta)' \]

\[\times (0, I) \mathbf{z}_t' \mathbf{z}_t - \hat{\Sigma}_{aatt} (0, I)' \]

\[= (n-k)^{-1} \sum_{t=1}^{n} \hat{\delta}_t^2 (0, I) \mathbf{z}_t' \mathbf{z}_t - \hat{\Sigma}_{aatt} (1, -\beta)' \]

\[\times (0, I) \mathbf{z}_t' \mathbf{z}_t - \hat{\Sigma}_{aatt} (0, I)' + o_p(1) \]

\[= o_p(1) \]
and (2.21) follows. Also,

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} (\hat{\pi}^2_{t,t} \hat{d}'_{t,t} - \pi^2_{t,t} d'_t d_t) = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} (\hat{\pi}^2_{t,t} - \pi^2_{t,t}) (0, I)[Z_t'Z_t - \hat{\pi}^2_{eatt}](1, -\beta')' \times (1, -\beta')[Z_t'Z_t - \hat{\pi}^2_{eatt}](0, I)' = 0
\]

(2.23)

by (2.3) and because we have a linear function of the elements of \(\hat{\gamma}'_{eatt}^\prime\). Combining (2.21) and (2.23), we have

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} (\hat{\pi}^2_{t,t} \hat{d}'_{t,t} - \pi^2_{t,t} d'_t d_t) = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} (\hat{\pi}^2_{t,t} - \pi^2_{t,t}) (0, I)[Z_t'Z_t - \hat{\pi}^2_{eatt}](1, -\beta')' \times (1, -\beta')[Z_t'Z_t - \hat{\pi}^2_{eatt}](0, I)' = 0
\]

(2.24)

Recall that \(\hat{\gamma}_{eatt} = n^{-\frac{1}{2}} \hat{\gamma}_{eatt} \). Then, combining Assumption (2.5), (2.9), (2.11), (2.14), and (2.24), we have that

\[
\hat{\gamma}_{eatt} \sim N(0, I) \quad \square
\]
Note that the specification of the model in Theorem 2.1 contains both the functional and structural models as special cases. In the functional or fixed model, \( x_t = \mu_{xt} \) are fixed vectors. In the structural or random model, \( \mu_{xt} = (1, \mu') \) where \( \mu' \) is fixed over \( t \) and the lower right \((k-1) \times (k-1)\) portion of the variance-covariance matrix of \( x_t \), \( \Sigma_{\mu xt} \), is positive definite. Our specification admits the general situation where the mean of \( x_t \) may depend on the index \( t \). Clearly, to obtain a limiting normal distribution, some conditions need to be imposed on the sequence of the means of the true values. One such condition that works is

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t \mu'_{xt} \mu_{xt} = M_{\mu \mu}.
\]

Clearly, we also need to have the probability limit of \( M_{\mu \mu} \) be positive definite to obtain a limiting distribution for \( \hat{\beta} \).

Note also that our definition of the estimator \( \hat{\beta} \) permits the matrices \( \hat{\Sigma}_{\mu att} \) to be singular. For example, it is well-known that if the model contains an intercept then one of the elements (usually designated to be the first element) of the vector \( x_t \), is always one and the corresponding row and column of \( \hat{\Sigma}_{\mu att} \) are zero vectors. Clearly, we need to place some assumptions on the estimated error variances \( \hat{\Sigma}_{\mu att} \) and the weights to obtain a limiting distribution.

We also need to have our estimated quantities \( n^{-1/2} \sum_{t=1}^{n} \pi_t \hat{d}_t \) and \( G = (n-k)^{-1} \sum_{t=1}^{n} \pi^2 \hat{d} \hat{d}' \) behave in the limit as the quantities they are estimating, namely \( n^{-1/2} \sum_{t=1}^{n} \pi_t d_t \) and \( G = n^{-1} \sum_{t=1}^{n} \pi^2 d' d_t \). To obtain
this result, we have assumed (2.3) and (2.6). Clearly, we also need to have $\lim_{n \to \infty} G$ exist and be positive definite.

As we can see, an important aspect of Theorem 2.1 is that it provides an estimator of the parameters that can be used without having to completely specify the distribution of $(x_t, u_t)$. Thus, it may be used in a wide range of applications.

Note that the weights will usually be related to the error variances in the model. Often, as in ordinary generalized least squares, the estimation procedure will be an iterative process. The first iteration would use $\hat{\pi}_t = \pi_t = 1$, and then an estimator of the "best" weight would be used as the $\hat{\pi}_t$ for succeeding iterations. More discussion of weights may be found in Chapter 3.

B. No Error in the Regression Equation

We now consider the case when we have no error in the regression equation, that is, $q_t = 0$. In Theorem 2.2, we show that for this case the estimator of $\beta$ is asymptotically normal either as the sample size becomes large or (and) as the error variances become small compared to the variation in $x_t$. To take care of both of these limit types, the sequence of estimators is indexed by $\nu$, where $\nu = n_T \nu$, $n_\nu$ is the sample size, and for a particular $\nu$, the error covariance matrix $\Sigma_{\nu} = \Sigma_\nu' \Sigma_\nu^{-1}$ is a fixed matrix multiplied by $T_\nu^{-1}$. The sequences $\{n_\nu\}$ and $\{T_\nu\}$ are assumed to be non-decreasing in $\nu$. For simplicity, we will omit the subscript $\nu$ on $n$ and $T$ from now on.
Because we assume that there is no error in the equation, we now have an estimate for the entire variance structure. This assumption leads us to the use of the root of a determinantal equation in defining our estimator. As a result of using the root equation, we obtain an additional term in the expression of the random variable \( d_t \) as compared to Theorem 2.1. Therefore, additional assumptions are needed to get convergence to a normal distribution.

**Theorem 2.2.** Let the model (2.1) hold with \( q_t = 0 \) and

\[
\Sigma_{\text{aaatt}} = T^{-1} \Omega_{\text{aaatt}} \quad \text{for} \quad t = 1, 2, \ldots, n.
\]

Let \( \hat{\Theta} = \hat{M}^{-1} \hat{M} \), where

\[
\hat{M}_{Z \pi Z} = \hat{M}_{Z \pi Z} - \hat{\lambda}_{\text{e a a m}}.
\]

\[
(M_{Z \pi Z}, \hat{\lambda}_{\text{e a a m}}) = n^{-1} \sum_{t=1}^{n} t(Z_t^t, \Sigma_{\text{aaatt}})
\]

and \( \hat{\lambda} \) is the smallest root of \( |\hat{M}_{Z \pi Z} - \hat{\lambda}_{\text{e a a m}}| = 0 \). Let

\[
[T^{1/2} a_t, T(C_{at} - c_{at})', (x_t - u_{xt})], \quad t = 1, 2, \ldots, n,
\]

be independent random variables with bounded \( 4 + \delta \) moments \( (\delta > 0) \),

\[
E(x_t - u_{xt}) = 0, \quad E(a_t | x_t) = 0,
\]
and

$$\lim_{\nu \to \infty} T_n^{1/2} \sum_{t=1}^{n} \pi_t \mathbb{E}(\hat{c}_{at} - c_{at}) = 0,$$  \hspace{1cm} (2.25)$$

where \( \nu = T_n \). Assume that

$$n^{-1} \sum_{t=1}^{n} (\pi_t^{(j)} - \pi_t^{(j)}) (z_t^{(j)} y_t, T_a^{(j)} a_t^{(j)}, T_{a_{a_{a_{a}}}}^{(j)}) = o_p(n^{-1/2}), \ j = 1, 2 \hspace{1cm} (2.26)$$

$$\lim_{\nu \to \infty} n^{-1/2} \sum_{t=1}^{n} (\pi_t^{(j)} - \pi_t^{(j)}) \hat{\xi}_t^{(j)} = 0, \ j = 1, 2 \hspace{1cm} (2.27)$$

$$\lim_{\nu \to \infty} n^{-1/2} \sum_{t=1}^{n} (\pi_t^{(j)} - \pi_t^{(j)}) \hat{\xi}_t^{(j)} \hat{\xi}_t^{(j)} = 0, \ j = 1, 2 \hspace{1cm} (2.28)$$

where \( \hat{\xi}_t^{(j)} = [T^{1/2}(\text{vec } z_t^{(j)} a_t^{(j)}), T(c_{at} - \text{vech } a_t^{(j)}), T(c_{at} - c_{at}^{(j)})] \), and

$$\lim_{\nu \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^{(j)} \hat{\phi}_t^{(j)} = \lim_{\nu \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^{(j)} \hat{\phi}_t^{(j)} \hspace{1cm} (2.29)$$

where \( \hat{\phi}_t^{(j)} = [\text{vech}(z_t^{(j)} a_t^{(j)}), T^{1/2} \text{vec}(z_t^{(j)} a_t^{(j)}), T \text{vech}(a_t^{(j)} a_t^{(j)}), T c_{at}^{(j)}] \). Let \( \{\mu_t, T_{a_{a_{a_{a}}}}^{(j)} \} \) be a fixed sequence, where \( \{\pi_t^{(j)}\} \) is bounded above and below by fixed positive numbers. Let

$$\lim_{\nu \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^{(j)} \hat{x}_t^{(j)} = \bar{M}^{|x|}_{\mu \pi \mu}, \hspace{1cm} (2.30)$$

$$\lim_{\nu \to \infty} T_n^{-1} \sum_{t=1}^{n} \pi_t \hat{x}_t = \lim_{\nu \to \infty} T_n^{-1} \sum_{t=1}^{n} \pi_t \hat{x}_t \hspace{1cm} (2.31)$$

$$\lim_{\nu \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t \hat{x}_t \hat{x}_t = \lim_{\nu \to \infty} M_{x \pi x} = \bar{M}^{|x|}_{x \pi x}, \hspace{1cm} (2.32)$$
\[ \lim_{\nu \to \infty} G = \overline{G}, \quad (2.33) \]

with \( \overline{M}_{\times \times} \) and \( \overline{G} \) positive definite, and

\[ G = n^{-1} \sum_{t=1}^{n} \pi \Sigma \{ d_t^t d_t \} \]

\[ \overline{M}_{\times \times} = n^{-1} \sum_{t=1}^{n} \pi \overline{x}_t \overline{x}_t \]

\[ d'_t = \overline{x}'_t - \frac{1}{\Sigma \overline{u}_t \overline{u}_t} \left( \nu^2 - \sigma_{\overline{u}_t \overline{u}_t} \right) \sigma_{\overline{v}_t \overline{v}_t}^{-1} \]

then,

\[ \lim_{\nu \to \infty} \left\{ \frac{1}{2} (\overline{G} - \beta) - n^{-1} \sum_{t=1}^{n} \pi d_t' \right\} = 0 \quad (2.34) \]

where \( \Gamma_{\nu} = n^{-1} \sum_{t=1}^{n} \pi \sigma_{\overline{x}_t \overline{x}_t}^{-1} \). Furthermore, if \( n \to \infty \) as \( \nu \to \infty \), then

\[ \overline{G} = n^{-1} \sum_{t=1}^{n} \pi \Sigma \{ d_t^t d_t \} \]

\[ \overline{G} = (n-k)^{-1} \sum_{t=1}^{n} \pi \Sigma \{ d_t^t d_t \} \]

where

\[ \overline{G} = n^{-1} \sum_{t=1}^{n} \pi \Sigma \{ d_t^t d_t \} \]
\[
\tilde{\alpha}_t = x'_t v_t - \sum_{uv} \tilde{\sigma}_{uv} - \tilde{\alpha}_t^{-1} (\tilde{G}_t - \tilde{\alpha}_t) \tilde{\sigma}_{uv}.
\]

\[
(\tilde{\sigma}_{uv}, \tilde{\sigma}_{uv}^{-1}) = n^{-1} \sum_{t=1}^{n} \tilde{\sigma}_{uv} (\tilde{\sigma}_{uv}^{-1} \tilde{\sigma}_{uv}^{-1})
\]

\[
(\tilde{\sigma}_{uv}, \tilde{\sigma}_{uv}^{-1}) = (\tilde{\alpha}' \hat{Z}_{a \alpha} \tilde{\alpha}, \tilde{\alpha}' \hat{Z}_{a \alpha} \tilde{\alpha})
\]

\[
\tilde{v}_t = y_t - x'_t \tilde{\alpha}
\]

and

\[
\tilde{\alpha}' = (1, -\tilde{\alpha}')
\]

**Proof.** Consider \( \hat{M}_{Z \pi Z} \) and \( \hat{E}_{a \alpha a} \). By definition,

\[
\hat{M}_{Z \pi Z} - \hat{M}_{Z \pi Z} - \hat{E}_{a \alpha a} = 
\]

\[
= n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t z'_t z_t - \hat{\pi}_t z'_t z_t - \hat{\pi}_t \hat{E}_{a \alpha a})
\]

\[
= n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t - \hat{\pi}_t) z'_t z_t
\]

\[
+ n^{-1} \sum_{t=1}^{n} \hat{\pi}_t (z'_t a_t + a'_t z_t - a'_t a_t)
\]

\[
+ n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t - \hat{\pi}_t) (z'_t a_t + a'_t z_t + a'_t a_t)
\]

\[
= o_p(n^{-1/2}) + o_p(n^{-1/2} T^{-1/2}) + o_p(n^{-1/2} T^{-1/2}) = o_p(n^{-1/2}) \quad (2.36)
\]
by Assumptions (2.26) and (2.27), \{\pi_t\} bounded, and the distributional assumptions. Also,

\[
\hat{\Sigma}_{a\pi a..} - \tilde{\Sigma}_{a\pi a..} = n^{-1} \sum_{t=1}^{n} \pi_t \hat{\Sigma}_{att} - n^{-1} \sum_{t=1}^{n} \pi_t \tilde{\Sigma}_{att}
\]

\[
= n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\Sigma}_{att} - \tilde{\Sigma}_{att}) + n^{-1} \sum_{t=1}^{n} (\pi_t - \pi) \tilde{\Sigma}_{att}
\]

\[
= O_p(n^{-1/2}T^{-1})
\]  

(2.37)

by (2.25), (2.26), \{\pi_t\} bounded, and the distributional assumptions.

Consider \( |\hat{M}_{Z\pi Z} - \lambda \tilde{\Sigma}_{a\pi a..}| = 0 \). The smallest root \( \lambda \) such that

\[
|\hat{M}_{Z\pi Z} + \Sigma_{a\pi a..} - \lambda \tilde{\Sigma}_{a\pi a..}| = 0
\]

is \( \lambda = 1 \) because \( z_t = (x_t, x_t^2) \) so that \( |\hat{M}_{Z\pi Z}| = 0 \). Thus, \( \lambda - 1 \) is a continuous function of the elements of \( \hat{\theta} = (\hat{M}_{Z\pi Z} - \hat{M}_{Z\pi Z} - \Sigma_{a\pi a..}, \hat{\Sigma}_{a\pi a..} - \Sigma_{a\pi a..}) \). \( \hat{\theta} \) has continuous first derivatives in a region about \( 0 \). Consider \( T \) fixed and \( n \to \infty \), then \( \hat{\theta} \xrightarrow{P} 0 \). By (2.36) and (2.37), \( \hat{\theta} = O_p(n^{-1/2}) \) and by considering a Taylor expansion,

\[
\lambda - 1 = O_p(n^{-1/2})
\]  

(2.38)

Now,

\[
(\hat{\theta} - \theta) = \hat{M}_{M\pi M}^{-1} \hat{M}_{M\pi Y} - \theta
\]

\[
= \hat{M}_{M\pi M}^{-1} (\hat{M}_{M\pi Y} - \hat{\Sigma}_{a\pi a..} - \hat{M}_{M\pi X^\theta} + \hat{\Sigma}_{a\pi a..} \theta)
\]
\begin{equation}
\begin{split}
\mathbb{M}_{\text{Xmx}}^{-1} &= \mathbb{M}_{\text{Xmx}}^{-1} [n^{-1} \sum_{t=1}^{n} \pi_t (\mathbf{X}' \mathbf{Y}_t - \mathbf{X}' \mathbf{\hat{y}}) - \hat{\lambda} (\hat{\Sigma}_{\text{uutt}} - \hat{\Sigma}_{\text{uutt}})] \\
&= \mathbb{M}_{\text{Xmx}}^{-1} [n^{-1} \sum_{t=1}^{n} \pi_t (\mathbf{X}' \mathbf{v}_t - \hat{\Sigma}_{\text{uutt}}^\prime)] \\
\end{split}
\end{equation}

where \( \mathbf{v}_t = \mathbf{y}_t - \mathbf{X}_t \hat{\beta} \) and \( \hat{\Sigma}_{\text{uutt}} = \hat{\Sigma}_{\text{uutt}} - \hat{\Sigma}_{\text{uutt}} \). Therefore,

\begin{equation}
(\hat{\beta} - \beta) = \mathbb{M}_{\text{Xmx}}^{-1} [n^{-1} \sum_{t=1}^{n} \pi_t (\mathbf{X}' \mathbf{v}_t - \hat{\Sigma}_{\text{uutt}}^\prime - (\hat{\lambda} - 1) \hat{\Sigma}_{\text{uutt}}^\prime)] \\
- (\hat{\lambda} - 1)(\hat{\Sigma}_{\text{uutt}} - \hat{\Sigma}_{\text{uutt}}^\prime)) \quad (2.39)
\end{equation}

Now, note that

\begin{equation}
\begin{split}
n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\Sigma}_{\text{uutt}} - \hat{\Sigma}_{\text{uutt}}^\prime) \\
&= n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\Sigma}_{\text{uatt}} - \hat{\Sigma}_{\text{uatt}}) \mathbf{z} + n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\Sigma}_{\text{uatt}} - \hat{\Sigma}_{\text{uatt}})^\prime \mathbf{z} \\
&= o_p(n^{-1/2} T^{-1}) + o_p(n^{-1/2} T^{-1}) \\
&= o_p(n^{-1/2} T^{-1}) \quad (2.40)
\end{split}
\end{equation}

by (2.27), \{\pi_t\} bounded and the distributional assumptions. Also,

\begin{equation}
\begin{split}
\mathbb{M}_{\text{Xmx}}^{-1} &= \mathbb{M}_{\text{Xmx}}^{-1} - \hat{\lambda} \hat{\Sigma}_{\text{uutt}} \\
&= n^{-1} \sum_{t=1}^{n} \pi_t \mathbf{X}' \mathbf{X}_t - \hat{\lambda} n^{-1} \sum_{t=1}^{n} \pi_t \mathbf{X}' \mathbf{X}_t \\
\end{split}
\end{equation}
\begin{align*}
\frac{1}{n} \sum_{t=1}^{n} \frac{1}{\gamma_t} x_t' x_t + \frac{1}{n} \sum_{t=1}^{n} \pi_t [x_t' u_t + u_t' x_t + u_t' u_t - E_{uut}]
&= \lambda (\hat{E}_{uut} - E_{uut}) - (\lambda - 1) \hat{E}_{uut} \\
&+ \frac{1}{n} \sum_{t=1}^{n} \pi_t (\hat{y}_t - \hat{y}_t) (x_t' y_t - y_t' x_t + y_t' y_t + y_t' u_t - \hat{y}_t' u_t - \hat{y}_t) \\
&\overset{p}{=} 0(n^{-1/2})
\end{align*}

by (2.26), (2.27), (2.32), and the distributional assumptions. Thus, by (2.38)-(2.41), a simpler expression is

\begin{align*}
(\hat{\beta} - \beta) &= \frac{1}{n} \sum_{t=1}^{n} \pi_t (x_t' y_t - E_{yyt} - (\lambda - 1) E_{uut}) + o_p(n^{-1/2}) \\
&\overset{p}{=} 0(n^{-1/2})
\end{align*}

Also,

\begin{align*}
\frac{1}{n} \sum_{t=1}^{n} \pi_t (x_t' y_t - E_{uut})
&= \frac{1}{n} \sum_{t=1}^{n} \pi_t [x_t' y_t - E_{uut} - (\lambda - 1) E_{uut}]
\\
&+ \frac{1}{n} \sum_{t=1}^{n} \pi_t (\hat{y}_t - \hat{y}_t) (x_t' y_t - E_{yyt}) - \frac{1}{n} \sum_{t=1}^{n} \pi_t E_{yyt} (\lambda - 1) \\
&\overset{p}{=} 0(n^{-1/2})
\end{align*}

by (2.27), (2.31), (2.38), and the distributional assumptions.
Therefore,

$$(\tilde{g} - g) = 0_p(\nu^{-1/2}) .$$  \hfill (2.43)

Now, \( \tilde{g} = \tilde{M}_{x, y}^{-1} \tilde{M}_{x, y} \) implies \( \tilde{M}_{x, y} \tilde{g} = \tilde{M}_{x, y} \) or instead

$$\tilde{M}_{x, y}, \tilde{M}_{x, y} \tilde{g} = 0$$

where \( \tilde{g}' = (1, -\tilde{g}) \). But \( (\tilde{M}_{x, y}, \tilde{M}_{x, y}) \) is a linear combination of the rows of \( (M_{x, y}, M_{x, y}) \) so we also have

$$(\tilde{M}_{y, x}, \tilde{M}_{y, x}) \tilde{g} = 0 .$$

Thus, we have

$$M_{x, y} \tilde{g} = (n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} Z_{t}^{'Z_{t}} - \lambda n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} a_{t} \hat{a}_{t} \tilde{g}) = 0 .$$

Multiplying by \( g' \) on the left, we have

$$g'(n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} Z_{t}^{'Z_{t}} - \lambda n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} a_{t} \hat{a}_{t} \tilde{g})$$

$$= n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} [(Z_{t} a)'Z_{t} \tilde{g} - \lambda g' \hat{a}_{t} \hat{a}_{t} \tilde{g}]$$

$$= n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} [v_{t} Z_{t} \tilde{g} + v_{t} Z_{t} (\tilde{g} - \tilde{g}) - \lambda g' \hat{a}_{t} \hat{a}_{t} \tilde{g} - \lambda g' \hat{a}_{t} \hat{a}_{t} (\tilde{g} - \tilde{g})]$$

$$= n^{-1} \sum_{t=1}^{n} \hat{\pi}_{t} [v_{t}^{2} - \lambda g' \hat{a}_{t} \hat{a}_{t} \tilde{g} + (v_{t} Z_{t} - \lambda g' \hat{a}_{t} \hat{a}_{t}) (\tilde{g} - \tilde{g})]$$

$$= 0 .$$
Rearranging, we obtain

\[ \lambda [n^{-1} \sum_{t=1}^{n} \hat{\pi}_t (g^{'} \hat{a}_{att} g - g^{'} \hat{\pi}_{aatt}) (\tilde{g} - \tilde{g})] \]

\[ = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \left[ v_t^2 - v_t x_t (\tilde{g} - \tilde{g}) - v_t u_t (\tilde{g} - \tilde{g}) \right] \]

and thus

\[ \lambda - 1 = \left( n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \left[ g^{'} \hat{a}_{att} g - g^{'} \hat{\pi}_{aatt} (\tilde{g} - \tilde{g}) \right] \right)^{-1} \]

\[ = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \left[ v_t^2 - a^{'} \hat{a}_{aatt} g - v_t x_t (\tilde{g} - \tilde{g}) \right] \]

\[ - g^{'} (a^{'} u_t - \hat{\pi}_{aatt}) (\tilde{g} - \tilde{g}) \]  \quad (2.44)

Now, note that

\[ n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \left[ v_t^2 - g^{'} \hat{a}_{aatt} g - v_t x_t (\tilde{g} - \tilde{g}) + g^{'} (a^{'} u_t - \hat{\pi}_{aatt}) (\tilde{g} - \tilde{g}) \right] \]

\[ = n^{-1} \sum_{t=1}^{n} \hat{\pi}_t \left[ v_t^2 - g^{'} \hat{a}_{aatt} g - v_t x_t (\tilde{g} - \tilde{g}) \right] \]

\[ - n^{-1} \sum_{t=1}^{n} \hat{\pi}_t g^{'} (a^{'} u_t - \hat{\pi}_{aatt}) (\tilde{g} - \tilde{g}) \]

\[ + n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) \left[ g^{'} (a^{'} a_t - \hat{\pi}_{aatt}) g - g^{'} (\hat{\pi}_{aatt} - \pi_{aatt}) g \right. \]

\[ - v_t x_t (\tilde{g} - \tilde{g}) - g^{'} (a^{'} u_t - \hat{\pi}_{aatt}) (\tilde{g} - \tilde{g}) \]
by (2.27), (2.43), and the distributional assumptions. Thus,

\[ n^{-1} \sum_{t=1}^{n} \pi_t [v_t^2 - \sigma_{\varepsilon_{\text{att}}} - v_t \pi_t (\hat{a} - \bar{a}) + \sigma'(a_t - \hat{a}_{\text{att}}) (\hat{b} - \bar{b})] \]

\[ = \bar{M}_{V_{\text{att}}} - \pi V_{\text{att}} - \bar{M}_{V_{\text{att}}} (\hat{a} - \bar{a}) + o_p(\nu^{1/2}). \]  

(2.45)

Also,

\[ n^{-1} \sum_{t=1}^{n} \pi_t [\hat{a}' \hat{\Sigma}_{\text{att}} - \hat{a}' \hat{\Sigma}_{\text{att}} (\hat{\theta} - \bar{\theta})] \]

\[ = n^{-1} \sum_{t=1}^{n} \pi_t \hat{a}' \hat{\Sigma}_{\text{att}} \pi_t \hat{a}' - n^{-1} \sum_{t=1}^{n} \pi_t \hat{a}' \hat{\Sigma}_{\text{att}} \pi_t \hat{a}' \hat{\Sigma}_{\text{att}} - n^{-1} \sum_{t=1}^{n} \pi_t \hat{a}' \hat{\Sigma}_{\text{att}} (\hat{\theta} - \bar{\theta}) - n^{-1} \sum_{t=1}^{n} (\pi_t - \pi_t) \hat{a}' \hat{\Sigma}_{\text{att}} (\hat{\theta} - \bar{\theta}) \]

\[ = n^{-1} \sum_{t=1}^{n} \pi_t \hat{a}' \hat{\Sigma}_{\text{att}} \pi_t \hat{a}' + o_p(\nu^{1/2}). \]  

(2.46)

by Assumption: (2.26) and (2.31), (2.43), and the distributional assumptions. Now,

\[ (\hat{\theta} - \bar{\theta}) = \frac{1}{\bar{M}_{X_{\text{att}}}} [\bar{M}_{X_{\text{att}}} - \bar{M}_{X_{\text{att}} \bar{\theta}}]. \]  

(2.47)

However, another expression for \( \bar{M}_{X_{\text{att}}} \) is
\[
\hat{\eta}_{xmx} = \hat{\eta}_{xmx} - \lambda \hat{\xi}_{uvtc} \\
= n^{-1} \sum_{t=1}^{n} \pi_t (x'_{t} x_t + x'_{t} u_t + u'_{t} x_t + u'_{t} u_t - \hat{\lambda} \hat{\xi}_{uvtc}) \\
= n^{-1} \sum_{t=1}^{n} \pi_t x'_{t} x_t + n^{-1} \sum_{t=1}^{n} \pi_t (x'_{t} u_t + u'_{t} x_t + (u'_{t} u_t - \hat{\xi}_{uvtc}) \\
- (\hat{\xi}_{uvtc} - \hat{\xi}_{uvtc}) - (\lambda - 1) (\hat{\xi}_{uvtc} - \hat{\xi}_{uvtc}) - (\lambda - 1) \hat{\xi}_{uvtc} } \\
+ n^{-1} \sum_{t=1}^{n} (\pi_t - \pi_t) (x'_{t} x_t + x'_{t} u_t + u'_{t} x_t + u'_{t} u_t - \hat{\lambda} \hat{\xi}_{uvtc}) \\
= \mathbf{M}_{xmx} + o_P(n^{-1/2}) \quad (2.48)
\]

by (2.26), (2.27), (2.38) and the distributional assumptions. Also,

\[
\hat{\eta}_{xmy} - \hat{\eta}_{xmx} \beta = n^{-1} \sum_{t=1}^{n} \pi_t (x'_{t} v_t - \hat{\lambda} \hat{\xi}_{uvtc}) \\
= n^{-1} \sum_{t=1}^{n} \pi_t (x'_{t} v_t + u'_{t} v_t - \hat{\lambda} \hat{\xi}_{uvtc}) \\
= n^{-1} \sum_{t=1}^{n} \pi_t x'_{t} v_t + n^{-1} \sum_{t=1}^{n} \pi_t (u'_{t} v_t - \hat{\lambda} (\hat{\xi}_{uvtc} - \hat{\xi}_{uvtc}) \beta - \hat{\lambda} \hat{\xi}_{uvtc}) \\
+ n^{-1} \sum_{t=1}^{n} (\pi_t - \pi_t) [x'_{t} v_t + (u'_{t} v_t - \hat{\xi}_{uvtc}) \\
- (\hat{\xi}_{uvtc} - \hat{\xi}_{uvtc}) \beta - (\lambda - 1) \hat{\xi}_{uvtc} ] \\
= \mathbf{M}_{xmv} + o_P(n^{-1/2}) \quad (2.49)
\]
by (2.26), (2.27), (2.38) and the distributional assumptions. Combining (2.47)-(2.49), we have

\[(\bar{\beta} - \beta) = \mathbf{M}_x^{-1} \mathbf{M}_x + o_p(\nu^{-1/2}). \quad (2.50)\]

Then, by (2.44)-(2.46) and (2.50), we have

\[
\hat{\lambda} - 1 = [n^{-1} \sum_{t=1}^n \hat{\pi}_t \hat{g}' \hat{a}^{\text{att}} \hat{a}]^{-1} [\mathbf{M}_v - \sigma_{v \pi \nu}.. - \mathbf{M}_v \mathbf{M}^{-1}_x \mathbf{M}_x] \]

\[
+ o_p(\nu^{-1/2}). \quad (2.51)
\]

However,

\[
\hat{\sigma}_{v \pi \nu} = n^{-1} \sum_{t=1}^n \hat{\pi}_t \hat{g}' \hat{a}^{\text{att}} \hat{a} = n^{-1} \sum_{t=1}^n \hat{\pi}_t \hat{g}' \hat{a}^{\text{att}} \hat{a} + o(1)
\]

\[
= \sigma_{v \pi \nu} + o(1) \quad (2.52)
\]

by Assumption (2.31). Thus, by (2.51) and (2.52)

\[
\hat{\lambda} - 1 = \sigma_{v \pi \nu}^{-1} [\mathbf{M}_v - \sigma_{v \pi \nu}.. - \mathbf{M}_v \mathbf{M}^{-1}_x \mathbf{M}_x] + o_p(\nu^{-1/2}). \quad (2.53)
\]

Now, from (2.42)

\[
(\bar{\beta} - \beta) = n^{-1} \mathbf{M}_x^{-1} \sum_{t=1}^n \hat{\pi}_t [\mathbf{X}' \nu_t - \sum_{uvtt} - (\hat{\lambda} - 1) \mathbf{Z}_{uvtt}] + o_p(\nu^{-1/2})
\]

\[
= n^{-1} \mathbf{M}_x^{-1} \sum_{t=1}^n \hat{\pi}_t [\mathbf{X}' \nu_t - \sum_{uvtt} - (\hat{\lambda} - 1) \mathbf{Z}_{uvtt}]
\]
\[ + n^{-1} \sum_{t=1}^{n} (\pi_t - \hat{\pi}_t)(x_t'v_t + (u_t'v_t - \Sigma_{uvt}) - (\Sigma_{uatt} - \Sigma_{uatt})g \]

\[ - (\lambda - 1)(\Sigma_{uvt} - u_t'g) + (\lambda - 1)u_t'g] + o_p(\nu^{-1/2}) \]

\[ = n^{-1} \sum_{t=1}^{n} \pi_t [x_t'v_t - \Sigma_{uvt} - (\lambda - 1)\Sigma_{uvt}] + o_p(\nu^{-1/2}) \]

(2.54)

by (2.27), (2.38) and the distributional assumptions. Therefore, combining (2.53) and (2.54), we obtain

\[ (\widehat{\beta} - \beta) = n^{-1} \sum_{t=1}^{n} \pi_t (x_t'v_t - \Sigma_{uvt}) \]

\[ - \sigma_{v,v}^{-1} \sum_{t=1}^{n} \pi_t \sigma_{v,v}^{-1} - \Sigma_{v,v}^{-1} (\nu^2 - \sigma_{v,v}^{-1}) \Sigma_{u,v} \]

\[ + o_p(\nu^{-1/2}) \]

\[ = n^{-1} \sum_{t=1}^{n} \pi_t (x_t'v_t - \Sigma_{uvt} - \sigma_{v,v}^{-1} (\nu^2 - \sigma_{v,v}^{-1}) \Sigma_{u,v}) \]

\[ + n^{-1} \sigma_{v,v}^{-1} \Sigma_{v,v}^{-1} \Sigma_{u,v} + o_p(\nu^{1/2}) \].

(2.55)

Now,

\[ \tilde{M}_{x,m} = M_{x,m} + o_p(n^{-1/2}) = \tilde{M}_{x,m} + o_p(1) \text{ by (2.32)}, \]
\[ \sigma_{\text{v.vw}..} = n^{-1} \sum_{t=1}^{n} \pi_t a'_t a_t = o_p(1) \]

because \( \{\pi_t\} \) bounded and the bounded moments assumption,

\[ M_{\text{v.vw}..} = n^{-1} \sum_{t=1}^{n} \pi_t a'_t a_t x_t = o_p(1) \]

by the distributional assumptions, and

\[ \tilde{\Sigma}_{\text{uvw}..} = n^{-1} \sum_{t=1}^{n} \pi_t \tilde{\Sigma}_{\text{uatt}..} = o_p(1) \]

by the same reasoning as for \( \sigma_{\text{v.vw}..} \). Thus, by (2.55) and the last four order results,

\[ (\bar{g} - \bar{g}) = \bar{M}_{x\text{vwx}..}^{-1} [n^{-1} \sum_{t=1}^{n} \pi_t a'_t] + o_p(\nu^{-1/2}) \] (2.56)

Consider the random variables \( T^{1/2} \pi_t d'_t \). By definition,

\[ T^{1/2} \pi_t d'_t = T^{1/2} \pi_t [x'_t - \tilde{\Sigma}_{\text{uvt}..}^{-1} (\nu^2 - \tilde{\Sigma}_{\text{uvt}..})] \]

\[ = \pi_t (T^{1/2} (x'_t - \mu'_x)') a_t \sigma_t + T^{1/2} \mu'_x a_t \sigma_t + T^{1/2} \tilde{T}_{a'a'} \tilde{T}^{-1/2} a_t \sigma_t \]

\[ - T^{1/2} T (\tilde{\Sigma}_{\text{uatt}..} - \tilde{\Sigma}_{\text{uatt}..}) g - T^{1/2} T \tilde{\Sigma}_{\text{uatt}..} g \]

\[ - \sigma_{\text{v.vw}..} T^{1/2} (g' T a_t \sigma_t - a'_t (\tilde{\Sigma}_{\text{uatt}..} - \tilde{\Sigma}_{\text{uatt}..}) g) \]
Thus, $T^{1/2} \pi_t d_t'$ is a linear combination of elements of $[T^{1/2} a_t, T(c_{at} - c_{at})', (x_t - \mu_{xt})]$ so the $T^{1/2} \pi_t d_t'$, $t = 1, 2, \ldots, n$, are independent random variables. As in Theorem 2.1, we can show that the $T^{1/2} \pi_t d_t'$ have $2 + \frac{1}{2} \delta$ moments. Now,

$$d_t' = \lambda_{xt} v_t' + (x_t - \mu_{xt})' v_t + (w_t' v_t - \tau_{uvtt}) - (\Sigma_{uuvtt} - \Sigma_{uvtt})$$

$$\quad - \sigma_{vuv..}(v_t^2 - \sigma_{uvtt}) \Sigma_{uuv..} + \sigma_{vuv..}(\sigma_{uvtt} - \sigma_{uvtt}) \Sigma_{uv..}$$

$$= (\lambda_{xt}^t, I, I, -I, -I, 1)[v_t, v_t(x_t - \mu_{xt}), v_t u_t - \Sigma_{uvtt},$$

$$\Sigma_{uvtt} - \Sigma_{uvtt}, \sigma_{vuv..}(v_t^2 - \sigma_{uvtt}) \Sigma_{v..}^v.,$$

$$\sigma_{vuv..}(\sigma_{uvtt} - \sigma_{uvtt}) \Sigma_{v..}^v.],$$

where $I$ is the $k \times k$ identity matrix. Let

$$c_t = \pi_t^\lambda (\mu_{xt}^t, I, I, -I, -I, I)$$

where $\lambda$ is a fixed row vector and let $e_{xt}^t$ be the vector of random parts of $T^{1/2} d_t'$ above. Thus,

$$c_t e_{xt}^t = T^{1/2} \pi_t \lambda d_t'$$

Now, $e_{xt}^t$ has bounded $2 + \frac{1}{2} \delta$ moments,

$$n^{-1} \Sigma_{t=1}^n c_t^E(e_{xt}^t e_{xt}^t) c_t^t = n^{-1} \Sigma_{t=1}^n \pi_t^2 E(d_t' d_t') \lambda'$$

$$= T^{\lambda} G^{\lambda'} \frac{p}{n+\delta} T^{\lambda} G^{\lambda'} \sim \lambda$$

by Assumption (2.33).
Thus, $n^{-1} \sum_{t=1}^{n} c_t E(e_{x_t} e_{x_t'}) c_t^t$ is bounded above and below for large $n$. Also, $n^{-1} \sum_{t=1}^{n} c_t c_t^t$ is bounded by the same reasoning as in Theorem 2.1 using (2.30). Thus, by Theorem 10.1,

$$\left( T^\lambda (G\lambda')^{-1/2} T^{1/2} \sum_{t=1}^{n} \pi_t d_t \right)$$

$$= (G\lambda')^{-1/2} \sum_{t=1}^{n} \pi_t d_t \overset{L}{\rightarrow} N(0, 1). \quad (2.57)$$

By the multivariate extension theorem and Theorem 10.2, we have if $n \rightarrow \infty$ as $\nrightarrow \infty$,

$$\left( n^{-1} M_{x \times \pi x}^{-1} \tilde{\Sigma}_{x \times \pi x}^{-1} \right)^{-1/2} \left( \tilde{\Sigma} - \beta \right) \overset{L}{\rightarrow} N(0, I)$$

or in the notation of the theorem,

$$\frac{1}{\Gamma_{\nu}^{1/2}} \left( \tilde{\Sigma} - \beta \right) \overset{L}{\rightarrow} N(0, I). \quad (2.58)$$

We now consider the estimation of the covariance matrix. Let

$$\tilde{\Sigma}_{\beta} = n^{-1} M_{x \times \pi x}^{-1} \tilde{\Sigma}_{x \times \pi x}^{-1},$$

where $\tilde{\Sigma}_{\beta}$ is defined in (2.35). Then, if the probability limits exist,

$$\text{plim} \, n T \tilde{\Sigma}_{\beta} = \text{plim} \, \frac{1}{\Gamma_{\nu}^{1/2}} \left( \tilde{\Sigma} - \beta \right) \overset{L}{\rightarrow} N(0, I).$$

(2.59)
because \( \lim_{\gamma \to \infty} \tilde{M} \chi_{\text{tr}} = \tilde{M} \chi_{\text{tr}} \) by Assumption (2.32). We now consider

\[
T\tilde{G} = (n-k)^{-1} \sum_{t=1}^{n} \tilde{\sigma}_{tt} \tilde{\tilde{X}}_{tt}^t,
\]

where

\[
\tilde{\tilde{X}}_{tt}^t = X_t' \tilde{v}_t - \tilde{\Sigma}_{ttt} - \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

\[
= X_t' \tilde{v}_t - \tilde{\Sigma}_{ttt} - \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt} + X_t' (\tilde{v}_t' - v_t') + (\tilde{\Sigma}_{ttt} - \tilde{\Sigma}_{ttt}) - \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

\[
+ \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

\[
= d_t' + X_t' (\tilde{v}_t' - v_t') + (\tilde{\Sigma}_{ttt} - \tilde{\Sigma}_{ttt}) - \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

\[
+ \tilde{\sigma}_{ttt} (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

\[
- (\tilde{v}_t^2 - \tilde{\sigma}_{ttt}) (\tilde{\sigma}_{ttt} - \tilde{\sigma}_{ttt}) \tilde{\Sigma}_{ttt}.
\]

(2.60)

(def.)

\[
= d_t' + b_t' .
\]

Now,

\[
\tilde{\sigma}_{ttt} = n^{-1} \tilde{\Sigma}_{ttt} \tilde{X}_{ttt}^t \tilde{X}_{ttt}^t.
\]
\[
\begin{align*}
&= n^{-1} \sum_{t=1}^{n} \pi_t \hat{\xi}_{\text{aatt}} + n^{-1} \sum_{t=1}^{n} \pi_t (\hat{\xi}_t - \pi_t) \hat{\xi}_{\text{aatt}} \\
&= n^{-1} \sum_{t=1}^{n} \pi_t \hat{\xi}_{\text{aatt}} + O_p(v^{-1/2}) \quad \text{by Assumption (2) and } \hat{\xi}_t = O_p(1) \\
&= n^{-1} \sum_{t=1}^{n} \pi_t [a_t \hat{\xi}_{\text{aatt}} + a_t \hat{\xi}_{\text{aatt}} (\hat{\xi} - \xi) + (\hat{\xi} - \xi) \hat{\xi}_{\text{aatt}}] \\
&= n^{-1} \sum_{t=1}^{n} \pi_t a_t \hat{\xi}_{\text{aatt}} + O_p(v^{-1/2}) \\
\end{align*}
\]

since \( \hat{\xi} - \xi = O_p(v^{-1/2}) \). Also

\[
\begin{align*}
\hat{\sigma}_{\nu_{\mu}..} &= n^{-1} \sum_{t=1}^{n} \pi_t a_t \hat{\xi}_{\text{aatt}} + n^{-1} \sum_{t=1}^{n} \pi_t a_t (\hat{\xi}_{\text{aatt}} - \xi_{\text{aatt}}) \xi + O_p(v^{-1/2}) \\
&= \sigma_{\nu_{\mu}..} + O_p(v^{-1/2}) \\
\end{align*}
\]

by Assumptions (2.25) and the moment assumptions for \( \hat{\xi}_{\text{aatt}} \). Similarly,

\[
\hat{\xi}_{\mu_{\nu}..} = \xi_{\mu_{\nu}..} + O_p(v^{-1/2}) \quad \text{so that}
\]

\[
\begin{align*}
\hat{\sigma}_{\nu_{\mu}..} \hat{\xi}_{\mu_{\nu}..} - \sigma_{\nu_{\mu}..} \xi_{\mu_{\nu}..} &= O_p(v^{-1/2}) . \\
\end{align*}
\]

Consider the following expansions of terms in \( \hat{d}_t \). We have

\[
\begin{align*}
d_t &= X_t' \nu - \sum_{t=1}^{n} (\nu_t^2 - \sigma_{\nu_{\mu}..}) \xi_{\nu_{\mu}..} \\
\end{align*}
\]
\begin{equation}
\begin{aligned}
\bar{\Sigma}_{uvtt} - \Sigma_{uvtt} &= \hat{\Sigma}_{uatt}(\bar{g} - g), \\
\sigma^{-1}_{\nu \mu ..}(\bar{v}^2 - v^2)\Sigma_{\nu \mu ..} &= [\bar{g}'Z_t Z_t \bar{g} - g'Z_t Z_t g] \sigma^{-1}_{\nu \mu ..}\Sigma_{\nu \mu ..}
\end{aligned}
\end{equation}

\begin{align}
X'_t(\bar{v}_t - v_t) &= X'_t(z_t \bar{g} + a_t \bar{g} - a_t g) \\
&= (x_t + u'_t)'[z_t(\bar{g} - g) + a_t(\bar{g} - g)] \text{ since } z_t \bar{g} = 0 \\
&= - (x_t + u'_t)'[x_t(\bar{g} - g) + u_t(\bar{g} - g)] \\
&= -[\mu'_t + (x_t - \mu_{xt}) + u'_t]\mu_{xt} + (x_t - \mu_{xt})'u_t \\
&= -[\mu'_t \mu_{xt} + (x_t - \mu_{xt})' + u'_t \mu_{xt} + (x_t - \mu_{xt})'u_t] \\
&+ u'_t(x_t - \mu_{xt}) + u'_t u_t(\bar{g} - g), \\
&\quad (2.63)
\end{align}

\begin{align}
&\quad (x_t - \mu_{xt})'(x_t - \mu_{xt}) + (x_t - \mu_{xt})'u_t + u'_t \mu_{xt} \\
&\quad + u'_t(x_t - \mu_{xt}) + u'_t u_t(\bar{g} - g),
\end{align}

\begin{align}
(-\hat{\Sigma}_{aatt} - \Sigma_{aatt})g - g'(a_t a_t - \Sigma_{aatt})g^{-1} \sigma^{-1}_{\nu \mu ..}\Sigma_{\nu \mu ..} \\
&\quad - g'(-\hat{\Sigma}_{aatt} - \Sigma_{aatt})g^{-1} \sigma^{-1}_{\nu \mu ..}\Sigma_{\nu \mu ..},
\end{align}

\begin{align}
X'_t(\bar{v}_t - v_t) &= X'_t(z_t \bar{g} + a_t \bar{g} - a_t g) \\
&= (x_t + u'_t)'[z_t(\bar{g} - g) + a_t(\bar{g} - g)] \text{ since } z_t \bar{g} = 0 \\
&= - (x_t + u'_t)'[x_t(\bar{g} - g) + u_t(\bar{g} - g)] \\
&= -[\mu'_t + (x_t - \mu_{xt}) + u'_t]\mu_{xt} + (x_t - \mu_{xt})'u_t \\
&= -[\mu'_t \mu_{xt} + (x_t - \mu_{xt})' + u'_t \mu_{xt} + (x_t - \mu_{xt})'u_t] \\
&+ u'_t(x_t - \mu_{xt}) + u'_t u_t(\bar{g} - g), \\
&\quad (2.63)
\end{align}
\[
\hat{\Sigma}_t^2 - \hat{\Sigma}_{vtt}^2 = \hat{\Sigma}_t^2 (Z_t'Z_t - \hat{\Sigma}_{aatt}) \hat{\Sigma}_t
\]

(2.67)

We will have

\[
\text{plim } (n-k)^{-1} \sum_{t=1}^{n} \hat{\Sigma}_t^2 d_t' d_t = \text{plim } (n-k)^{-1} \sum_{t=1}^{n} \hat{\Sigma}_t^2 d_t' d_t
\]

(2.68)

if the right hand probability limit exists and if

\[
\text{plim } (n-k)^{-1} \sum_{t=1}^{n} \hat{\Sigma}_t^2 (d_t'b_t + b_t'd_t + b_t'b_t) = 0 .
\]

(2.69)

By assumptions (2.33) and (2.28) and the moment properties of the random variables,

\[
\text{plim } (n-k)^{-1} \sum_{t=1}^{n} \hat{\Sigma}_t^2 d_t' d_t = \lim G = \tilde{G} .
\]
Details are given in the discussion associated with (2.70). Because

\[ T^{\frac{1}{2}} (\tilde{g} - g) = o_p(n^{-\frac{1}{2}}), \]

\[ T^{\frac{1}{2}} (\tilde{\Sigma}_{uv..} - \Sigma_{uv..}) = o_p(n^{-\frac{1}{2}}), \]

we only need to show that the multipliers for these quantities are \( o_p(n^{\frac{1}{2}}) \) to show (2.69). Products of (2.62) with any of the expressions (2.63)-(2.67) and products of any two of the expressions (2.63)-(2.67) are \( O_p(1) \) by Assumption (2.29). Hence, result (2.69) follows.

To show the existence of the right hand limit of (2.68), we need to show

\[ \lim_{n \to \infty} n^{-1} T \sum_{t=1}^{n} (\pi_t^2 - \tilde{\pi}_t^2) d_t' d_t = 0. \quad (2.70) \]

Because

\[ d_t' = x_t' a_{\tilde{g}} + (u_t' a_{\tilde{g}} - \Sigma_{uatt}) g - g'(a_t' a_t - \hat{\Sigma}_{aatt}) \tilde{g}^{-1} \sum_{uv..}, \]

\[ = x_t' a_{\tilde{g}} + (u_t' a_{\tilde{g}} - \Sigma_{uatt}) g - \hat{\Sigma}_{uatt} g' a_{\tilde{g}} \]

\[ - g'(a_t' a_t - \hat{\Sigma}_{aatt}) \tilde{g}^{-1} \sum_{uv..}, \]

\[ + g'(\hat{\Sigma}_{aatt} - \tilde{\Sigma}_{aatt}) g\Sigma_{uv..}^{-1} \sum_{uv..}, \]
(2.70) follows from Assumption (2.28) and the fact that \( \sigma \), \( \sigma_{\nu^T \nu} \), and \( \Sigma \) are all \( O(1) \). Finally, it can be shown that

\[
\lim_{n \to \infty} n^{-1} T \sum_{t=1}^{n} \pi_t^2 d_t d_t' = \lim_{n \to \infty} n^{-1} T \sum_{t=1}^{n} \pi_t^2 E(d_t d_t'),
\]

(2.71)

by the reasoning as used to show (2.19) in the proof of Theorem 2.1 because \( \sigma_{\nu^T \nu} \), \( \Sigma \) are bounded and with \( r_t^2 = \nu_t^2 - \sigma_{\nu t t} \), \( E(|T_{t t}^2|^{1+\gamma}) \) is bounded. Note that the right hand limit exists by Assumption (2.33). Thus, by (2.68), (2.70), and (2.71) if \( n \to \infty \) as \( \nu \to \infty \),

\[
\lim_{n \to \infty} (n-k)^{-1} T \sum_{t=1}^{n} \pi_t^2 d_t d_t' = \lim_{n \to \infty} n^{-1} T \sum_{t=1}^{n} \pi_t^2 E(d_t d_t'),
\]

or in the notation of the theorem statement

\[
\lim_{\nu \to \infty} \widetilde{T}_G = \lim_{\nu \to \infty} T_G.
\]

Therefore, \( \lim_{\nu \to \infty} n \tilde{\nu}_{\beta \beta} = \lim_{\nu \to \infty} n \tilde{\gamma}_{\nu} \). Thus, if \( n \to \infty \) as \( \nu \to \infty \),

\[
\tilde{\nu}_{\beta \beta}^{1/2} (\tilde{\nu} - \beta) \xrightarrow{L} N(0, I).
\]

Note that in the proof of Theorem 2.2, in addition to \( 4 + \delta \)
bounded moments, \( \{ \nu_{x_t} \}, \ t = 1, 2, \ldots, n \), bounded is sufficient to show expression (2.69). For example, consider expression (2.63) and the product of (2.62) with (2.63). We have
\[(n-k)^{-1} \sum_{t=1}^{n} d_t^t X_t (\bar{\nu}_t - \nu_t)\]

\[= (n-k)^{-1} \sum_{t=1}^{n} d_t^t X_t [X_t (\bar{\gamma} - \beta)] .\]

The matrix of multipliers for \((\bar{\gamma} - \beta)\) in this sum has \(ij\)-th element equal to

\[(n-k)^{-1} \sum_{t=1}^{n} d_t^t X_t^j X_t^k .\]  \hspace{1cm} (2.72)

This matrix multiplier will be \(O_p(1)\) under a number of conditions. First, the assumption that the random components have \(4 + \delta\) moments means that any product of three or four random components is \(O_p(1)\), by the weak law of large numbers. If the elements of the fixed component \(u_{xt}\) are bounded, then

\[u_{xt}^i u_{xtj}^j u_{xtk}^k\]

is bounded and the variance of a term such as

\[(n-k)^{-1} \sum_{t=1}^{n} u_{xti} u_{xtj} u_{xtk} v_t \]

is \(O(n^{-1})\). Therefore, the term is \(O_p(n^{-1/2})\) since \(E(v_t) = 0\). A term such as

\[(n-k)^{-1} \sum_{t=1}^{n} u_{xti} u_{xtj} v_t^i v_t^j\]

is \(O_p(1)\) because the variance is \(O(n^{-1})\).
III. EXTENSIONS

In this chapter, we consider some extensions of Theorems 2.1 and 2.2. In the model of Theorem 2.1 we have an error in the equation. However, Theorem 2.1 contains no assumption about the magnitude of the error variances as a function of \( t \). In this section we consider the situation in which the error variances are getting smaller as \( t \) increases. Then we consider the problem of using estimated variances as weights.

A. Error in the Equation and Measurement

Error Variances Decreasing

When there is an error in the equation and the error variances are getting smaller as \( t \) increases, we consider two possibilities. The first is that \( \sigma^2 \) does not decrease as \( t \) increases while the second is that \( \sigma^2 \) decreases at the same rate as \( \tau \) as \( t \) increases. The two cases differ in the assumptions required to obtain the limiting properties of the random variables \( \pi_t d_t \) and \( T^{1/2} \pi_t d_t \). The random variables \( d_t \) are defined in Theorem 2.1. In this section, as in Theorem 2.2, the sequences of estimators is indexed by \( v \), where \( v = n_T \). When \( \sigma^2 \) does not decrease, we use steps nearly the same as those of Theorem 2.1 to obtain the limiting properties of the estimators. When \( \sigma^2 \) does not decrease, terms containing \( q_t \) dominate those that do not contain \( q_t \) as \( v \to \infty \). Therefore, although the measurement error variances decrease, \( \sigma^2 \) dominates the variance of \( \hat{\beta} \) and we still obtain \( (\hat{\beta} - \beta) = O_p(n^{-1/2}) \). The above result is
described in Theorem 3.1 below.

**Theorem 3.1.** Let model (2.1) hold where

\[
\Sigma_{a_t} = T^{-1} \Omega_{a_t} , \quad t = 1, 2, \ldots, n .
\]  

Assume \( q_{q_t} > 0 \) is unknown and that a set of estimated weights \( \hat{\pi}_t \), \( t = 1, 2, \ldots, n \), is available. Let \([q_t, T^{1/2} a_t, T(\hat{c}_t - c_{at})', (x_t - \hat{\mu}_{xt})] \), \( t = 1, 2, \ldots, n \), be independent with bounded 4 + \( \delta \) moments \( \delta > 0 \), \( E\{(q_t, a_t, q_t a_t) | x_t\} = 0 \), \( E\{x_t - \hat{\mu}_{xt}\} = 0 \), and

\[
\lim_{n \to \infty} n^{-1/2} T \sum_{t=1}^{n} \pi_t E\{\hat{c}_t - c_{at}\} = 0 .
\]

Let \( \{(\mu_{xt}, c_{at}, \pi_t)\} \) be a fixed sequence indexed by \( t \), where \( \{\pi_t\} \) is bounded above and below by fixed positive numbers. Let

\[
\text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t \hat{\gamma}_t = \text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t \hat{\gamma}_t' , \quad j = 1, 2 ,
\]  

where

\[
\hat{\gamma}_t = [Z_t, (\text{vech} Z_t)', (\text{vech} z_t z_t)', T_{at}]' .
\]

Let

\[
\text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t x_t x_t^t = \text{plim}_{n \to \infty} M_{x x} = \bar{M}_{x x}
\]

and
\[ \text{plim } G = \bar{G} \quad (3.3) \]

be positive definite, where \( G \) is defined in Theorem 2.1. Assume

\[ \text{plim } n^{-\frac{1}{2}} \sum_{t=1}^{n} (\pi_t - \pi)^d_t = 0 \quad (3.4) \]

and

\[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t \mu_t \mu_t = \bar{\mu} \mu \mu \cdot \]

Then, if \( n \to \infty \) as \( v \to \infty \)

\[ \hat{\beta} \beta \beta \left( \hat{\beta} - \beta \right) \xrightarrow{L} N(0, 1) , \]

where \( \hat{\beta} \) and \( \hat{\beta} \beta \beta \) are defined in Theorem 2.1.

**Proof.** As in the proof of Theorem 2.1,

\[ n^{1/2} (\hat{\beta} - \beta) = n^{-1/2} \bar{M}_{\pi x} \sum_{t=1}^{n} \pi_t (X_t'v - Z_{ut} t) \]

\[ = n^{-1/2} \bar{M}_{\pi x} \sum_{t=1}^{n} \pi_t d_t' . \]

Also,

\[ \bar{M}_{x \pi x} = \bar{M}_{x \pi x} + o_p(1) \]

and
\[
\begin{align*}
\sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} &= \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} + o(1).
\end{align*}
\]

Thus,
\[
\begin{align*}
n^{1/2} \left( \mathbf{\hat{\beta}} - \mathbf{\beta} \right) &= \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} + o(1) .
\end{align*}
\]

Likewise,
\[
\begin{align*}
n^{1/2} \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} &= \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} + o(1) .
\end{align*}
\]

is bounded above and below. Also \( n^{-1} \sum_{t=1}^{n} \mathbf{c}'_{t} \mathbf{c}_{t} \), where
\[
\mathbf{c}_{t} = \pi_{t} \left( \mathbf{y}_{t}' - \mathbf{I}, \mathbf{I}, -\mathbf{I} \right),
\]

is bounded in absolute value. Thus, by the central limit theorem of Theorem 10.1 and its multivariate extension (see Chapter X, Appendix B),
\[
\begin{align*}
[ \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \mathbf{d}'_{t} \mathbf{d}_{t} ]^{-1/2} \sum_{t=1}^{n} \mathbf{d}'_{t} \xrightarrow{L} N(0, \mathbf{I}) .
\end{align*}
\]

See the discussion of (2.12)-(2.14).

We can also show that
\[
\begin{align*}
\text{plim } n^{-1} \sum_{t=1}^{n} \left( \pi_{t}^{2} \mathbf{d}'_{t} \mathbf{d}_{t} - \mathbf{E}(\mathbf{d}'_{t} \mathbf{d}_{t}) \right) &= 0 \\
\text{plim } n^{-1} \sum_{t=1}^{n} \left( \pi_{t}^{2} \mathbf{d}'_{t} \mathbf{d}_{t} - \mathbf{E}(\mathbf{d}'_{t} \mathbf{d}_{t}) \right) &= 0
\end{align*}
\]

and
\[
\begin{align*}
\text{plim } (n-k)^{-1} \sum_{t=1}^{n} \pi_{t}^{2} \mathbf{d}'_{t} \mathbf{d}_{t} - \text{plim } (n-k)^{-1} \sum_{t=1}^{n} \pi_{t}^{2} \mathbf{d}'_{t} \mathbf{d}_{t} = 0
\end{align*}
\]
by the steps used in the proof of Theorem 2.1. Combining (3.5) through (3.9), we have the desired result. In showing (3.7) through (3.9), \( T \) enters only when it is necessary to obtain the moments of \( a_t \) or to use the assumptions on the \( a_t \). In such cases, we multiply the appropriate random variable by \( T^{-\gamma} T^\gamma \) where \( \gamma \) is some exponent and \( T^\gamma \) appears in the assumptions about the \( a_t \). Therefore, some expressions will include \( T^{-\gamma} \) and the order in probability for those terms will be smaller than those without \( T^{-\gamma} \). The overall order results for the estimators are the same as in Theorem 2.1.

Also note that by (3.5)

\[
n^{1/2} (\hat{\beta} - \beta) = n^{1/2} MX^{-1} n^{-1} \sum_{t=1}^{n} \pi_t (X'_{tt} \beta - \Sigma_{uvtt}) + o_P(1),
\]

and because

\[
n^{-1} \sum_{t=1}^{n} \pi_t (X'_{tt} \beta - \Sigma_{uvtt})
\]

\[
= n^{-1} \sum_{t=1}^{n} \pi_t (X'_{tt} q_t + u' q_t + X'_{tt} (w_t - u_t \beta) - \Sigma_{uvtt})
\]

is dominated by the \( X'_{tt} q_t \) term, then

\[
n^{-1} \sum_{t=1}^{n} \pi_t (X'_{tt} \beta - \Sigma_{uvtt}) = o_P(n^{-1/2})
\]

and

\[
(\hat{\beta} - \beta) = o_P(n^{-1/2}) \quad (3.10)
\]
as in Theorem 2.1.

We now consider the case in which both the variance of the error in the equation and the measurement error variances decrease at the same rate as \( t \) increases. In this case, the terms containing \( q_t \) do not dominate the remaining terms, and we obtain \( (\hat{\theta} - \theta) = O_p(\sqrt{t}) \) as in Theorem 2.2. In Theorem 3.2, Assumption (3.14) takes the place of Assumption (2.3) of Theorem 2.1.

**Theorem 3.2.** Let the model (2.1) hold. Let

\[
\sigma^2_{qq} = T^{-1} \omega_{qq}, \tag{3.11}
\]

\[
\Sigma_{aatt} = T^{-1} \omega_{aatt}, \quad t = 1, 2, \ldots, n. \tag{3.12}
\]

Let

\[
[T^{1/2} q_t, T^{1/2} a_t, T(\hat{c}_t - c_t), (x_t - \mu_t)], \quad t = 1, 2, \ldots, n,
\]

be a random sample from a distribution with bounded \( 4 + \delta \) (\( \delta > 0 \)) moments, where

\[
\lim_{n \to \infty} T^{-1/2} \sum_{t=1}^{n} T \mathbb{E} \left[ (\hat{c}_t - c_t) \right] = 0 \tag{3.13}
\]

and \( \nu = T^{1/2} \). Let

\[
\Sigma_{aatt} = T^{-1} \omega_{aatt}, \quad t = 1, 2, \ldots, n.
\]
\[ \hat{\rho}_t = [(\text{vech } Z_t' Z_t)'', \text{vec}(T^{1/2} Z_t' q_t)'', \text{vec}(T^{1/2} Z_t' a_t)'', Tc'_{a_t}]' \]

and assume for \( j = 1, 2 \) that

\[
\text{plim } n^{-1} \sum_{t=1}^{n} \pi_t \hat{\rho}_t \hat{\rho}_t' = \text{plim } n^{-1} \sum_{t=1}^{n} \pi_t \hat{\rho}_t \hat{\rho}_t'. \tag{3.14}
\]

Assume

\[
\text{plim } n^{-1/2} T^{1/2} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) d_t = 0, \tag{3.15}
\]

where \( d_t = X_t' v_t - \sum_{t=1}^{m} v_t t_t \). Finally, assume (2.4), (2.5), and (2.7).

Then, if \( n \to \infty \) as \( v \to \infty \)

\[
\hat{v}_{\beta \beta}^{1/2} (\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, I)
\]

where \( \hat{v}_{\beta \beta} \) and \( \hat{\beta} \) are defined in Theorem 2.1.

**Proof.** Consider

\[
\nu^{1/2} (\hat{\beta} - \beta) = \nu^{1/2} \hat{M}_{\times \nu} (\hat{M}_{\times \pi y} - \hat{M}_{\times \nu \pi x})
\]

\[
= \nu^{1/2} \hat{M}_{\times \nu} n^{-1} \sum_{t=1}^{n} \pi_t d_t'
\]

\[
= \nu^{1/2} \hat{M}_{\times \nu} n^{-1} \sum_{t=1}^{n} \pi_t d_t' + \nu^{1/2} \hat{M}_{\times \nu} n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) d_t'.
\]
Now, as in the discussion of (2.9),

\[ \hat{M}_{\mathbf{x} \mathbf{m}} = \bar{M}_{\mathbf{x} \mathbf{m}} + o_p(1) \]

and by Assumption (3.15)

\[ n^{-1} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) d_t' = o_p(1). \]

Therefore,

\[ \nu^{1/2} (\hat{\beta} - \beta) = \nu^{1/2} \frac{1}{n} \sum_{t=1}^{n} \pi_t d_t' + o_p(1). \quad (3.16) \]

The random variables \( T^{1/2} \pi_t d_t' \) are independent with finite \( 2 + \frac{1}{2} \delta \) moments and such that \( \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \pi_t E(d_t') = 0 \). We can write

\[
\begin{align*}
d_t' &= (\nu_t, I, I, -I)[v_t, v_t(x_t - \mu_{xt}), v_t u_t - \gamma_{vut}, \gamma_{vut} - \gamma_{vut}'] \\
&= b_t e_t'.
\end{align*}
\]

Let \( X_t = \pi_t \lambda d_t' = c_t e_t' \) where \( c_t = \pi_t \lambda b_t \). Then, as in the discussions of (2.14) and (2.57), \( n^{-1} \sum_{t=1}^{n} c_t e_t' \) is bounded above and below for large \( n \), and \( n^{-1} \sum_{t=1}^{n} e_t' e_t \) is bounded in absolute value. Therefore, as \( n \to \infty \)
We next show that

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^2 [ \mathbf{d}_t' \mathbf{d}_t - \mathbf{E}(\mathbf{d}_t' \mathbf{d}_t)] = 0.
\]  

With $T$ fixed, this situation is clearly the same as that of Theorem 2.1. If $T$ increases, the term $T(\mathbf{u}_t^2 - T) = T(\mathbf{y}_t^2 - \sigma_{\mathbf{y}_t^2})$ is the same as that used to show (2.71) in Theorem 2.2 except for the term involving $T(q_t^2 - \sigma_{q_t^2})$. However, by the moment assumptions on $T_{1/2} q_t$, it follows that $\mathbf{E}\{|T(q_t^2 - \sigma_{q_t^2})|^{1+\gamma}\}$ is bounded and therefore, $T(q_t^2 - \sigma_{q_t^2})$ satisfies the conditions of Chung's theorem as do the other terms of $\mathbf{d}_t' \mathbf{d}_t$ in (2.71), and (3.18) follows by the reasoning used to show (2.19) of Theorem 2.1.

Now, we have by Assumption (3.14) that

\[
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^2 \mathbf{d}_t' \mathbf{d}_t = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_t^2 \mathbf{d}_t' \mathbf{d}_t.
\]  

because

\[
\mathbf{d}_t' = \mathbf{x}_t' \mathbf{v}_t - \mathbf{z}_{\mathbf{v}_t} - \mathbf{z}_{\mathbf{u}_t}
\]

\[
= \mathbf{x}_t' \mathbf{q}_t + \mathbf{x}_t' \mathbf{a}_t (1, -\mathbf{g}')' - (0, 1) \mathbf{z}_{\mathbf{a}_t} (1, -\mathbf{g}')'.
\]  

\[
(3.20)
\]
Finally, to show that the estimated variance can be used to normalize 
\((\hat{\theta} - \theta)\), we need to show that

\[
\lim_{V \to \infty} n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 \hat{d}_t' \hat{d}_t = \lim_{V \to \infty} n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 \hat{d}_t' \hat{d}_t . \tag{3.21}
\]

By (3.16), \(T^{1/2} (\hat{\theta} - \theta) = O_p(n^{-1/2})\). It remains to show that

\[n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 (X'_t X'_t - \hat{\Sigma}_{uutt})^2\]

and

\[n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 \frac{1}{2} d'_t (X'_t X'_t - \hat{\Sigma}_{uutt})\]

are \(O_p(n^{1/2})\). Now, \(n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 (X'_t X'_t - \hat{\Sigma}_{uutt})^2 = O_p(1)\) by (3.14). Using (3.20) we have

\[
n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 (X'_t X'_t - \hat{\Sigma}_{uutt})
\]

\[= n^{-1} \sum_{t=1}^{n} \hat{\pi}_t^2 \left( T^{1/2} X'_t q_t + T^{1/2} X'_t a_t (1, -\theta')' - T^{1/2} (0, 1) \hat{\Sigma}_{uatt} (1, -\theta')' [X'_t X'_t - \hat{\Sigma}_{uutt}] \right) = O_p(1)
\]

by Assumption (3.14). Therefore, (3.21) follows. Combining (3.18), (3.19) and (3.21),

\[
\lim_{V \to \infty} \hat{T} \hat{\theta} = \lim_{V \to \infty} T \theta .
\]

Therefore, in the notation of the theorem,
Theorem 3.2 is for the situation in which no estimator of $\sigma_q$ is available. This lack of an estimator of $\sigma_q$ separate from $\sigma_{wtt}$ is the most common situation in practice. Finally, we note that Theorem 2.2 contains the case in which the error variances do not decrease. This special case is obtained by fixing $T$.

B. Variance Functions and Estimated Variances as Weights

We now consider the problem of constructing appropriate estimated weights $\hat{\pi}_t$. Note that we can write $Y_t = X_t \hat{\beta} + v_t$ where $v_t = e_t - u_t \hat{\beta}$ and $e_t = q_t + w_t$. The variance of $v_t$ is comparable to the variance of the error in the equation for the $t$-th observation in the general linear model. Thus, it is reasonable to construct an estimated generalized least squares estimator by weighting the observations by an estimator of $\sigma_{vtt}$. Thus, in the model of Theorem 2.1 we could use as weights the
\[ \hat{\sigma}_{v\nu t} = \hat{\sigma}_{q\nu} + (1, -\hat{\beta}') \sum_{a\nu t\nu t}(1, -\hat{\beta}')', \ t = 1, 2, \ldots, n, \]

where

\[ \hat{\sigma}_{q\nu} = n^{-1} \sum_{t=1}^{n} \left[ v_{t}^{2} - (1, -\hat{\beta}') \sum_{a\nu t\nu t}(1, -\hat{\beta}')' \right], \]

\[ v_{t} = Y_{t} - X_{t} \hat{\beta}, \] and \( \hat{\beta} \) is an initial estimate obtained using \( \tau_{t} = 1 \) in Theorem 3.1. Under the assumptions of Theorem 3.1, \( \hat{\sigma}_{q\nu} \) is a consistent estimator of \( \sigma_{q\nu} \). These weights do not minimize the variance of the limiting distribution. The weights that do minimize the variance of the limiting distribution depend on the true but unknown \( \nu_{t} \). Therefore, without additional assumptions, we are unable to construct a best weight, and we consider the use of \( \hat{\sigma}_{v\nu t} \).

In order to continue our discussion, we need order results for sums of multiples of random variables, each of which has the same order in probability. We consider cases in which the multipliers are either constants or random variables. Results for these two cases are given in Lemmas 3.1 and 3.3 respectively. Finally, Lemma 3.4 combines elements of Lemmas 3.1 and 3.3.

**Lemma 3.1.** Let \( \{a_{t}\} \) be a sequence of constants such that \( \sum_{t=1}^{\infty} |a_{t}| \) converges. Also, let \( \{H_{tn} : t=1, 2, \ldots, n; n=1, 2, \ldots\} \) be a triangular array of random variables such that \( E(H_{tn}^2) = O(b_n^2) \), \( t=1, 2, \ldots, n \) where \( b_n \to 0 \). Then,
Proof. It is enough to show that $E[(\sum_{t=1}^{n} a_t H_{tn})^2] = O(b_n^2)$ (see Chapter 9, Appendix A). Now,

$$E[(\sum_{t=1}^{n} a_t H_{tn})^2] = \sum_{t=1}^{n} a_t^2E(H_{tn}^2) + 2 \sum_{t=1}^{n} \sum_{j=t+1}^{n} a_t a_j E(H_{tn}H_{jn}). \quad (3.22)$$

By definition, there exists a positive real number $M_1$ such that

$$E(H_{tn}^2) < M_1b_n^2 \text{ for all } n. \quad (3.23)$$

Now, by the Cauchy-Schwarz inequality and inequality (3.23),

$$E(H_{tn}H_{jn}) < \sqrt{E(H_{tn}^2)E(H_{jn}^2)} \quad \text{for all } t,j$$

$$< M_1b_n^2 \text{ for all } n. \quad (3.24)$$

Therefore, combining (3.22)-(3.24), we obtain

$$E[(\sum_{t=1}^{n} a_t H_{tn})^2] < M_1b_n^2(\sum_{t=1}^{n} a_t^2 + 2 \sum_{t=1}^{n} \sum_{j=t+1}^{n} a_t a_j)$$

$$= M_1b_n^2(\sum_{t=1}^{n} a_t^2). \quad (3.25)$$

However, since $\sum_{t=1}^{\infty} |a_t|$ converges, we have that

$$|\sum_{t=1}^{n} a_t| < K \text{ for all } n \text{ for some real number } K.$$
Therefore,

\[ E[(\sum_{t=1}^{n} a_t H_{tn})^2] < K^2 n^2 b_n^2 \]

or in other notation, \( E[(\sum_{t=1}^{n} a_t X_{tn})^2] = O(b_n^2) \).

Before generalizing Lemma 3.1 to the situation in which the multipliers are random variables, we prove another order result.

**Lemma 3.2.** Let \( \{H_t\} \) be a sequence of random variables such that \( E(|H_t|^s) = O(b_n^s) \) for some \( s > 0 \) and let \( b_n \) be a sequence of positive real numbers. Then, \( H_t = O_p(b_n) \).

**Proof.** By definition, \( E(|H_t|^s) = O(b_n^s) \) implies that \( E(|H_t|^s) < M_1 b_n^s \) for some positive real number \( M_1 \). By Chebyshev's inequality, for any \( M_2 > 0 \),

\[
P(|H_t| > M_2 b_n) < \frac{M_2^{-s} b_n^{-s} E(|H_t|^s)}{M_2^{-s} M_1^s}.
\]

Therefore, given \( \epsilon > 0 \) choose \( M_2 \) such that \( M_2^s > \epsilon^{-1} M_1^s \). Then,

\[
P(|H_t| > M_2 b_n) < \epsilon \text{ and } H_t = O_p(b_n) \text{ by definition.}
\]

**Lemma 3.3.** Let \( \{U_t\} \) be a sequence of random variables with bounded second moments. Also, let \( \{H_{tn}: t=1, 2, \ldots, n; n=1, 2, \ldots\} \) be a triangular array of random variables such that \( E(H_{tn}^2) = O(b_n^2) \), \( t=1, \ldots, n \).
2, ..., n, where \{b_n\} is a sequence of positive real numbers. Assume 
(U_t, H_{tn}^2), t=1, 2, ..., n, are independent. Then,

\[ n^{-1} \sum_{t=1}^{n} U_t H_{tn} = O_p(b_n). \]

**Proof.** By the Cauchy-Schwarz inequality,

\[ E(\left| U_t H_{tn} \right|) < [E(U_t^2)E(H_{tn}^2)]^{1/2}. \quad (3.26) \]

By assumption, there exists a \( K_1 \) such that \( E(U_t^2) < K_1^2 \) for all \( t \).

Also, by assumption, there exists an \( M_1 > 0 \) such that \( E(H_{tn}^2) < M_1 b_n^2 \). Therefore,

\[ E(\left| U_t H_{tn} \right|) < K_1 M_1 b_n \]

or instead

\[ E(\left| U_t H_{tn} \right|) = O(b_n). \quad (3.27) \]

Therefore,

\[ E(\left| n^{-1} \sum_{t=1}^{n} U_t H_{tn} \right|) < n^{-1} \sum_{t=1}^{n} E(\left| U_t H_{tn} \right|) < K_1 M_1 b_n \]

by (3.27). Thus, by Lemma 3.2,

\[ n^{-1} \sum_{t=1}^{n} U_t H_{tn} = O_p(b_n). \]

\( \square \)
Lemma 3.4. Let \( \{a_t\} \) be a sequence of constants such that 
\[ n^{-1} \sum_{t=1}^{n} |a_t| \] 
is bounded. Also, let \( \{U_t\} \) be a sequence of random variables with bounded second moments and let \( \{H_{tn}; t=1, 2, \ldots; n=1, 2, \ldots\} \) be a triangular array of random variables such that 
\[ E(H_{tn}^2) = O(b_n^2), \quad t=1, 2, \ldots, n, \] 
where \( \{b_n\} \) is a sequence of positive real numbers. Assume \( \{U_t, H_{tn}; t=1, 2, \ldots, n\} \) are independent. Then,

\[ n^{-1} \sum_{t=1}^{n} a_t U_t H_{tn} = O_p(b_n). \]

Proof. From the proof of Lemma 3.3 we have \( E(|U_t H_{tn}|) = O(b_n) \). To prove Lemma 3.4, it is enough to show that 
\[ E(|n^{-1} \sum_{t=1}^{n} a_t U_t H_{tn}|) = O(b_n). \]

By definition, there exists an \( M_1 > 0 \) such that

\[ E(|U_t H_{tn}|) > M_1 b_n. \] (3.28)

Thus,

\[
E(|n^{-1} \sum_{t=1}^{n} a_t U_t H_{tn}|) < n^{-1} \sum_{t=1}^{n} |a_t| E(|U_t H_{tn}|)
\]

\[
< M_1 b_n (n^{-1} \sum_{t=1}^{n} |a_t|)
\]

\[
< M_1 K b_n \text{ for some } K.
\]

\[ \square \]
We now use the above lemmas to show that for a given form of \( \hat{\pi}_t \), then the assumptions involving \( \hat{\pi}_t - \pi_t \) in Theorems 2.1 through 3.2 will be satisfied. In some situations, this will then enable us to find a better weight than \( \hat{\sigma} \) defined previously.

**Theorem 3.3.** Let \( \{a_t\} \) be a sequence of constants such that 
\[
n^{-1} \sum_{t=1}^{n} |a_t| \text{ is bounded. Also, let } \{U_t\} \text{ be a sequence of random variables with bounded second moments and let the model (1.1)-(1.2) hold with}
\]

\[
\mathbb{E} \left( \Delta_{\text{att}} - \Delta_{\text{att}} \right) = T^{-1} \mathbb{X}_{\text{att}} , \quad t=1, 2, \ldots, n.
\]

Assume that 

\[
T^{-1} = o(n^{-1}) . \tag{3.29}
\]

Also, assume that \( \hat{\pi}_t \), the estimator of \( \pi_t \), \( t=1, 2, \ldots, n \), is such that \( \hat{\pi}_t - \pi_t \) can be written in the form 

\[
(\hat{\pi}_t - \pi_t) = \sum_{j=1}^{\ell} a_{tj} U_{tj} H_{tjn} \tag{3.30}
\]

where \( \ell \) is some integer, the \( \{a_{tj}\} \) are sequences of constants, the \( \{U_{tj}\} \) are sequences of random variables, and the \( \{H_{tjn}: t=1, 2, \ldots, n; n=1, 2, \ldots\} \) are triangular arrays of random variables such that 

\[
\mathbb{E}(H_{tjn}^2) = O(T^{-1}) , \quad t=1, 2, \ldots, n .
\]

Also, assume that one of two
conditions is met for each \( j \):

i) The sequence \( \{U_{tj}\} \) is a sequence of constant random variables with the same value for all \( t \) and the sequence \( \{a_{tj}\} \) is bounded, or

ii) The sequences \( \{a_{tj}\} \) and \( \{U_{tj}\} \) are such that

\[
\sum_{t=1}^{n} a_{tj} U_{tj} = O_p(1)
\]

and the triangular array \( \{H_{tjn}; t=1, 2, \ldots, n; n=1, 2, \ldots\} \)

\( \equiv \{H_{jn}; t=1, 2, \ldots, n; n=1, 2, \ldots\} \), i.e., the \( H_{tjn} \) array does not depend on \( t \).

Then,

\[
\text{plim } n^{-1/2} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) a_t U_t = 0.
\]

**Proof.** We write

\[
n^{-1/2} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) a_t U_t = n^{1/2} \sum_{t=1}^{\ell} \left[ \sum_{j=1}^{n-1} (\hat{\pi}_j U_j H_{tjn} - \pi_j a_t U_t) \right]
\]

\[
= n^{1/2} \sum_{j=1}^{\ell} \left[ \sum_{t=1}^{n-1} a_t - a_{tj} U_t U_j H_{tjn} \right].
\]

For a particular \( j \), \( j=1, 2, \ldots, \ell \), consider the expression
Consider case i) when \( \{U_{t,j}\} \) is a constant sequence with constant value, say \( c_j \). First consider the sequences of constants, \( \{a_{t,j}\} \).

By assumption, there exists a real number \( M_j < \infty \) such that
\[
|a_{t,j}| < M_j \text{ for all } t.
\]
Therefore,
\[
n^{-1} \sum_{t=1}^{n} |a_{t,j}a_t| < M_j n^{-1} \sum_{t=1}^{n} |a_t| < L < \infty
\]
by the assumption that \( n^{-1} \sum_{t=1}^{n} |a_t| \) is bounded and \( L \) is a real number. Thus, \( n^{-1} \sum_{t=1}^{n} |a_{t,j}a_t| \) is bounded for every \( j \). Then, since
\[
n^{-1} \sum_{t=1}^{n} a_{t,j}a_t U_{t,j} U_t H_{t,j} = c_j (n^{-1} \sum_{t=1}^{n} a_{t,j}a_t U_t H_{t,j}),
\]
Lemma 3.4 applies and for such \( U_{t,j} \),
\[
n^{-1} \sum_{t=1}^{n} a_{t,j}a_t U_{t,j} U_t H_{t,j} = o_p(T^{-1/2}) = o_p(n^{-1/2})
\]
by (3.29).

Next, consider case ii) where the triangular array of \( H_{t,j} \) does not depend on \( t \). Then,
\[
n^{-1} \sum_{t=1}^{n} a_{t,j}a_t U_{t,j} U_t H_{t,j} = H_{j,n} n^{-1} \sum_{t=1}^{n} a_{t,j}a_t U_{t,j} U_t
\]
\[
= o_p(T^{-1/2})
\]
by the assumption that \( n^{-1} \sum_{t=1}^{n} a_{t} a_{t} U_{t} U_{t} = o_{p}(1) \). Therefore, by (3.29) we have in both cases i) and ii) that

\[
n^{-1} \sum_{t=1}^{n} a_{t} a_{t} U_{t} U_{t} = o_{p}(n^{-1/2})
\]
or instead, that

\[
n^{-1/2} \sum_{t=1}^{n} (\hat{\pi}_{t} - \pi_{t}) a_{t} U_{t} = o_{p}(1).
\]

1. Simple linear regression to estimate variance functions

We now look at some specific cases of Theorem 3.3. Consider a linear variance function and the use of simple linear regression to estimate the function. If we let \( \pi_{t} \) be the inverse of the variance function for the \( t \)-th observation, we need to show that the estimator \( \hat{\pi}_{t} \) satisfies Assumptions (3.2) and (3.4) above. In this particular case, we will not use Theorem 3.3 but will derive the results directly.

Consider the simple univariate case with fixed \( x_{t} \). Let

\[
Y_{t} = X_{t} \beta + \nu_{t}
\]

where \( X_{t} = (1, X_{t}) \), \( \beta = (\beta_{0}, \beta_{1})' \), and \( \nu_{t} = q_{t} + w_{t} - u_{t} \beta_{1} \).

Assume that

\[
E(\nu_{t}^2) = \sigma_{\nu t t} = c_{0} + c_{1} x_{t} \quad (c_{1} \neq 0)
\]

and

\[
E(\delta_{a t t}^{-1}) = T^{-1} \Omega_{a t t}^{-1} \quad , \quad t=1, 2, \ldots, n.
\]
Note that (3.31) and (3.32) imply that \((c^0, c^1) = O(T^{-1})\). While this implies that the parameters become smaller as \(T \to \infty\), we assume that the parameters are constant in the estimation procedure for a particular sample. Assume that \([q_t, T^{1/2} a_t, T(c_{at} - c_{at}'), (x_t - y_{xt})], t = 1, 2, \ldots, n\), are independent with bounded eighth moments. Assume also that we have initial estimates \((\hat{\beta}_0, \hat{\beta}_1)\) obtained by assuming equal variances and unit weights in Theorem 3.1. The initial estimates are given by

\[
\hat{\beta} = (\hat{M}_{XX} - \hat{7}_{uu})^{-1}(\hat{M}_{XY} - \hat{7}_{uw}),
\]

where \(\hat{M}_{ZZ} = n^{-1} \sum_{t=1}^{n} Z_t Z_t'\) and \(\hat{7}_{aa} = n^{-1} \sum_{t=1}^{n} \hat{e}_{att}\). Also let

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

\[
\hat{\sigma}_{qq} = n^{-1} \sum_{t=1}^{n} [\hat{v}_t^2 - \hat{g}' \hat{7}_{aatt} \hat{g}],
\]

\[
\hat{g} = (1, - \hat{\beta}')
\]

\[
\hat{7}_{eett} = \text{diag}(\hat{\sigma}_{qq}, 0) + \hat{7}_{aatt},
\]

\[
\hat{\sigma}_{vvtt} = \hat{g}' \hat{7}_{eett} \hat{g},
\]

\[
\hat{\sigma}_{uutt} = \hat{\sigma}_{aett} - \hat{\sigma}_{uutt} \hat{\beta}_1.
\]
In the arguments of Result 3.1 we require the second moments of \( \hat{x}_t \) to exist. To guarantee the existence we modify the estimator by bounding the adjustment applied to \( X_t \) to create \( \hat{x}_t \) by a multiple of the variance of \( u_t \). This bound could be imposed in practice because, under normality for \( u_t \), the estimator of \( u_t \) should seldom exceed \( 3\sigma_{u_t}^{1/2} \). Let

\[
\hat{x}_t = \begin{cases} 
X_t - \frac{\hat{\delta}_t \hat{v}_t}{p} \sigma_{u_t}^{1/2} & \text{if } |\hat{\delta}_t \hat{v}_t| < p \sigma_{u_t}^{1/2} \\
X_t - \text{sgn}(\hat{\delta}_t \hat{v}_t)(p \sigma_{u_t}^{1/2}) & \text{otherwise ,}
\end{cases}
\]  

(3.33)

where \( \hat{\delta}_t = \sigma_{v_t}^{-1} \), \( p \) is some reasonable constant and \( \text{sgn}(\cdot) \) denotes the sign function. Then, our new estimator of \( \sigma_{v_t} \) is \( \hat{c}_0 + \hat{c}_1 \hat{x}_t \) where \( \hat{c}_0 \) and \( \hat{c}_1 \) are the ordinary least squares estimates obtained by regressing \( (\hat{v}_t - \hat{v})^2 \) on \( (1) \) and \( \hat{x}_t \), where \( (1) \) denotes a column of ones.

**Result 3.1.** Assume the model of Theorem 3.1 holds. Assume (3.31)-(3.33) and let

\[
T^{-1} = o(n^{-1}) .
\]

(3.34)

Assume \([q_t, T^{1/2} a_t, T(c_{at} - c_{at})', (x_t - u_{xt})], t=1, 2, \ldots, n, \) are independent with bounded eighth moments. Assume

\[
\pi_t^{-1} = \sigma_{v_t} = \hat{c}_0 + \hat{c}_1 \hat{x}_t
\]

is bounded away from zero. Also assume

\[
n^{-1} \sum_{t=1}^{n} |u_{xt}|^{2+j} = o(1) ,
\]

(3.35)
\[ n^{-1} \sum_{t=1}^{n} \left| \mu_{xt} - n^{-1} \sum_{t=1}^{n} \mu_{xt} \right|^{2+j} = O(1) , \quad (3.36) \]

\[ n^{-1} \sum_{t=1}^{n} \pi_{t}^{(1+j)} x_{t} \hat{y}_{t} = O(p) \quad (1) , \quad (3.37) \]

and \( \text{plim}_{n \to \infty} n^{-1} \sum_{t=1}^{n} \pi_{t} x_{t} \hat{y}_{t} \) exists, \( j = 1, 2 \). Also assume

\[ n^{-1/2} \sum_{t=1}^{n} \pi_{t}^{2} x_{t} \delta_{t} = O(p) \quad (1) . \quad (3.38) \]

Let

\[ \pi_{t}^{-1} = \sigma_{\text{vtt}}^{-1} = \hat{c}_{0} + \hat{c}_{1} \hat{x}_{t} \quad (3.39) \]

where

\[ \hat{c}_{1} = [n^{-1} \sum_{t=1}^{n} (x_{t} - \bar{x})^2]^{-1} n^{-1} \sum_{t=1}^{n} (x_{t} - \bar{x})(\hat{v}_{t} - \bar{v})^2 , \]

\[ \hat{c}_{0} = n^{-1} \sum_{t=1}^{n} (\hat{v}_{t} - \bar{v})^2 - \bar{x} \hat{c}_{1} . \]

Then, (3.2) and (3.4) hold for \( \pi_{t}^{-1} \) defined in (3.39).

\textbf{Proof.} We begin by writing

\[ \pi_{t}^{-1} = \sigma_{\text{vtt}}^{-1} = (\hat{c}_{0} + \hat{c}_{1} \hat{x}_{t})^{-1} \]
\[ = (c_0 + c_1 x_t)^{-1} - (c_0 + c_1 x_t)^{-2} (c_1 - c_0) \]
\[ - x_t (c_0 + c_1 x_t)^{-2} (c_1 - c_1) - c_1 (c_0 + c_1 x_t)^{-2} (x_t - x_t) \]
\[ + \text{remainder} . \]

Ignoring this remainder, we have
\[ \hat{\pi}_t - \pi_t = -\pi_t (c_0 - c_0) - x_t \pi_t (c_1 - c_1) - c_1 \pi_t (x_t - x_t) . \] (3.40)

By (3.34),
\[ E\{|x_t - x_t|^2\} \leq E\{(u_t + p \sigma_{uutt}^{1/2})^2\} \]
\[ = \sigma_{uutt} + 2p E(u_t \sigma_{uutt}^{1/2}) + p^2 E(\sigma_{uutt}) \]
\[ = O(T^{-1}) . \]

Thus,
\[ \hat{x}_t - x_t = O_p(T^{-1/2}) . \] (3.41)

To determine whether Assumptions (2.3) and (2.6) are satisfied, we also need to obtain the orders of \( \hat{c}_0 - c_0 \) and \( \hat{c}_1 - c_1 \).
Now, let

\[
\tilde{u}_t = \begin{cases} 
\hat{\delta}_t \sqrt{\nu} & \text{if } |\hat{\delta}_t \sqrt{\nu}| < p \sqrt{\sigma_{W\nu}} \\
\text{sgn}(\hat{\delta}_t \sqrt{\nu}) p \sqrt{\sigma_{W\nu}} & \text{otherwise}.
\end{cases}
\]

Then,

\[
\hat{x}_t = X_t - \tilde{u}_t
\]

\[
= x_t + u_t - \tilde{u}_t.
\]

Thus, we have

\[
\hat{x}_t - \bar{x} = (x_t - \bar{x}) + (u_t - \bar{u}) - (\tilde{u}_t - \bar{u}). \tag{3.42}
\]

Also, we have \(\hat{v}_t = v_t - X_t (\hat{\beta} - \beta)\) and therefore,

\[
(\hat{v}_t - \bar{v})^2 = (v_t - \bar{v})^2 + (u_t - \bar{u})^2 (\hat{\beta}_1 - \beta_1)^2 + (x_t - \bar{x})^2 (\hat{\beta}_1 - \beta_1)^2
\]

\[
- 2(v_t - \bar{v})(u_t - \bar{u})(\hat{\beta}_1 - \beta_1)
\]

\[
- 2(v_t - \bar{v})(x_t - \bar{x})(\hat{\beta}_1 - \beta_1)
\]

\[
+ 2(u_t - \bar{u})(x_t - \bar{x})(\hat{\beta}_1 - \beta_1). \tag{3.43}
\]
By assumption, \( u_t - \bar{u} = O_p(T^{-1/2}) \). Since

\[
E[|\hat{u}_t - \hat{\bar{u}}|^2] \leq E\left(p \sigma_{\hat{u}_{tt}}^{1/2} + p n^{-1} \sum_{t=1}^{n} \sigma_{\hat{u}_{tt}}^{1/2}\right) = O(T^{-1})
\]

we also have \( \hat{u}_t - \hat{\bar{u}} = O_p(T^{-1/2}) \).

Now, looking at the denominator of \( c_1 \),

\[
n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 = n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 + 2n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(u_t - \bar{u})
\]

\[
- 2n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(\hat{u}_t - \hat{\bar{u}})
\]

\[
- 2n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})(\hat{u}_t - \hat{\bar{u}})
\]

\[
+ n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})^2 + n^{-1} \sum_{t=1}^{n} (\hat{u}_t - \hat{\bar{u}})^2
\]

(3.44)

Note that

\[
x_t - \bar{x} = \mu_{xt} + (x_t - \mu_{xt}) + n^{-1} \sum_{t=1}^{n} \mu_{xt} + n^{-1} \sum_{t=1}^{n} (x_t - \mu_{xt})
\]

Since \( n^{-1} \sum_{t=1}^{n} \mu_{xt} \) is bounded by the arguments for Theorem 2.1 and \( x_t - \mu_{xt} \) has bounded \( 1 + \frac{1}{4} \delta \) moments (assume \( \delta < 4 \)), we apply Lemmas 3.1 and 3.3 to expression (3.44) and obtain
\[ n^{-1} \sum_{t=1}^{n} (\hat{x}_t - \bar{x})^2 = n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 + O_p(T^{-1/2}). \quad (3.45) \]

Using (3.42) and (3.43), we can write the numerator of \( \hat{c}_1 \) as

\[ \begin{align*}
&n^{-1} \sum_{t=1}^{n} (\hat{x}_t - \bar{x})(\hat{v}_t - \bar{v})^2 = n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})^2 \\
&\quad + n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(u_t - \bar{u})^2(\beta_1 - \beta_1)^2 \\
&\quad + n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^3(\beta_1 - \beta_1)^2 \\
&\quad - 2n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})(u_t - \bar{u})(\beta_1 - \beta_1) \\
&\quad - 2n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2(v_t - \bar{v})(\beta_1 - \beta_1) \\
&\quad + 2n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})(x_t - \bar{x})^2(\beta_1 - \beta_1) \\
&\quad + n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})(v_t - \bar{v})^2 \\
&\quad + n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})^3(\beta_1 - \beta_1)^2 \\
&\quad + n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})(x_t - \bar{x})^2(\beta_1 - \beta_1)^2 \\
&\quad - 2n^{-1} \sum_{t=1}^{n} (v_t - \bar{v})(u_t - \bar{u})^2(\beta_1 - \beta_1) \\
&\quad - 2n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(u_t - \bar{u})(v_t - \bar{v})(\beta_1 - \beta_1)
\end{align*} \]
\[\begin{align*}
+ 2n^{-1} & \sum_{t=1}^{n} (u_t - \bar{u})^2(x_t - \bar{x})(\beta_1 - \beta_1) \\
- n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(v_t - \bar{v})^2 \\
- n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(u_t - \bar{u})^2(\beta_1 - \beta_1)^2 \\
- n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(x_t - \bar{x})^2(\beta_1 - \beta_1)^2 \\
+ 2n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(v_t - \bar{v})(u_t - \bar{u})(\beta_1 - \beta_1) \\
+ 2n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(v_t - \bar{v})(x_t - \bar{x})(\beta_1 - \beta_1) \\
- 2n^{-1} & \sum_{t=1}^{n} (\tilde{u}_t - \bar{u})(u_t - \bar{u})(x_t - \bar{x})(\beta_1 - \beta_1) .
\end{align*}\]

Since \((\beta_1 - \beta_1) = O_p(n^{-1/2})\), \(u_t - \bar{u} = O_p(T^{-1/2})\), \(\tilde{u}_t - \bar{u} = O_p(T^{-1/2})\), and \(n^{-1} \sum_{t=1}^{n} u_{xt}\) and \(n^{-1} \sum_{t=1}^{n} x_{xt}^2\) are bounded, then terms containing at least one of \(u_t - \bar{u}\) and \(\tilde{u}_t - \bar{u}\) are \(O_p(\max\{n^{-1}, T^{-1/2}\})\) by Lemmas 3.1 and 3.3. Now,

\[n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^3 = n^{-1} \sum_{t=1}^{n} \left[ (x_t - \mu_{xt})^3 + (x_t - \mu_{xt}) \sum_{t=1}^{n} (x_t - \mu_{xt}) \right] = O_p(1) (3.47)\]
by Assumption (3.36) and the assumption of bounded eighth moments of
\((x_t - \mu_{xt})\). Also,

\[
\begin{align*}
& n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 (v_t - \bar{v}) = n^{-1} \sum_{t=1}^{n} [(\mu_{xt} - \mu_{xt}) + (x_t - \mu_{xt})] \\
& \quad - n^{-1} \sum_{t=1}^{n} (x_t - \mu_{xt})^2 (v_t - \bar{v}) \\
& = O_p(n^{-1/2}) \tag{3.48}
\end{align*}
\]

because \(E(q_t, a_t \mid x_t) = 0\) implies that \(E(v_t \mid x_t) = 0\),

\([x_t - \mu_{xt} - n^{-1} \sum_{t=1}^{n} (x_t - \mu_{xt})]^2\) has bounded \(2 + \frac{1}{2} \delta\) moments and

\(n^{-1} \sum_{t=1}^{n} [\mu_{xt} - n^{-1} \sum_{t=1}^{n} \mu_{xt}]^4 = O(1)\) by (3.36). Thus

\[
\text{Var}(n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 (v_t - \bar{v})) = O(n^{-1}).
\]

Therefore, (3.46) becomes

\[
\begin{align*}
& n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})^2 = n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})^2 \\
& \quad + O_p(\max(n^{-1}, T^{-1/2})) \tag{3.49}
\end{align*}
\]

Combining (3.45) and (3.47), we have

\[
\hat{c}_1 = [n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2]^{-1} n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})^2
\]
\[
= \left[ n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})^2 \right]^{-1} p n^{-1} \sum_{t=1}^{n} (x_t - \bar{x})(v_t - \bar{v})^2 \\
+ O_p \left( \max \{ n^{-1}, T^{-1/2} \} \right) \\
= c_1 + o_p(1) \quad (3.50)
\]

because we assumed in (3.34) that \( T^{-1} = o(n^{-1}) \).

To obtain an appropriate order expression for \( \hat{c}_0 - c_0 \), we write

\[
\hat{c}_0 = n^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2 - \bar{x} c_1 \\
= n^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2 - \bar{x} c_1 + [n^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2 - n^{-1} \sum_{t=1}^{n} (v_t - \bar{v})^2] \\
- (\bar{x} - \bar{v})c_1 - \bar{x}(\hat{c}_1 - c_1) \\
= c_0 + [n^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2 - n^{-1} \sum_{t=1}^{n} (v_t - \bar{v})^2] \\
- (\bar{x} - \bar{v})c_1 - \bar{x}(\hat{c}_1 - c_1) \quad (3.51)
\]

From (3.43) and arguments used to show (3.49), we obtain

\[
n^{-1} \sum_{t=1}^{n} (\hat{v}_t - \bar{v})^2 = n^{-1} \sum_{t=1}^{n} (v_t - \bar{v})^2 + O_p \left( \max \{ n^{-1}, T^{-1/2} n^{-1/2} \} \right). \quad (3.52)
\]

Also, note that
\[
\tilde{x} - x = n^{-1} \sum_{t=1}^{n} (u_t - \tilde{u}_t) = O_p(T^{-1/2}) \quad (3.53)
\]

and

\[
\hat{x} = n^{-1} \sum_{t=1}^{n} x_t = n^{-1} \sum_{t=1}^{n} \mu_{xt} + n^{-1} \sum_{t=1}^{n} (x_t - \mu_{xt}) = O_p(1) \quad (3.54)
\]

Combining (3.50)-(3.54) and the assumption that \( \text{plim} \ n^{-1} n^{\tau} \xi_{t=1} \xi_{t=1} \xi_{t=1} \xi_{t=1} \xi_{t=1} \xi_{t=1} \xi_{t=1} \xi_{t=1} \) exists, we have

\[
\hat{c}_0 - c_0 = O_p(\max(n^{-1}, T^{-1/2})) = o_p(1) \quad (3.55)
\]

Now, using expression (3.40), for \( j=1 \) Assumption (2.3) becomes

\[
\text{plim} \ n^{-1} \sum_{t=1}^{n} \pi_{xt} \chi_{t} \chi_{t} \ = \ - \left[ \text{plim} \ (\hat{c}_0 - c_0) n^{-1} \sum_{t=1}^{n} \pi_{xt} \chi_{t} \chi_{t} \right] + \text{plim} \ (\hat{c}_1 - c_1) n^{-1} \sum_{t=1}^{n} \pi_{xt} \chi_{t} \chi_{t} \ + \ c_1 \text{plim} n^{-1} \sum_{t=1}^{n} \pi_{xt} \chi_{t} \chi_{t} = 0 \quad (3.56)
\]

By expressions (3.50) and (3.55) and Assumption (3.37), the probability limits of the first two terms of (3.56) are zero. Now, the random parts of \( \pi_{xt} \chi_{t} \chi_{t} \) have bounded second moments. Then, by Assumptions (3.35) and (2.7), expression (3.41), and Lemmas 3.1, 3.3, and 3.4, we obtain
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) \hat{\chi}_t \hat{\chi}'_t = 0.
\]

Therefore,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} (\hat{\pi}_t - \pi_t) \hat{\chi}_t \hat{\chi}'_t = 0 \tag{3.57}
\]

and Assumption (2.3) is satisfied for \( j = 1 \).

When \( j = 2 \) Assumption (2.3) becomes

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} (\hat{\pi}_t^2 - \pi_t^2) \hat{\chi}_t \hat{\chi}'_t = 0 \tag{3.58}
\]

Note that

\[
\hat{\pi}_t^2 = (c_0 + c_1 \hat{x}_t)^{-2} = \pi_t^2 - 2(c_0 + c_1 x_t)^{-3}(c_0 - c_0)
\]

\[
- 2(c_0 + c_1 x_t)^{-3} x_t(c_1 - c_1) - 2(c_0 + c_1 x_t)^{-3} c_1(\hat{x}_t - x_t)
\]

\[
+ \text{remainder}
\]

\[
= \pi_t^2 - 2\pi_t^3(c_0 - c_0) - 2\pi_t^3 x_t(c_1 - c_1) - 2\pi_t^3 c_1(\hat{x}_t - x_t)
\]

ignoring the remainder. Therefore, (3.58) follows by Assumption (3.37) and the arguments and assumptions used to show (3.57).

Finally, we need to show that
\[
\lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} (\pi_t - \pi_t^*) d_t^* = 0. \tag{3.59}
\]

Using expression (3.40), we write (3.39) as

\[
\lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} (\pi_t^* - \pi_t) d_t^* \\
= -\left[ \lim_{n \to \infty} (c_0^* - c_0)n^{-1/2} \sum_{t=1}^{n} \pi_t^* d_t^* + \lim_{n \to \infty} (c_1^* - c_1)n^{-1/2} \sum_{t=1}^{n} \pi_t^* x_t d_t^* \\
- c_1 \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} \pi_t^* (x_t^* - x_t)d_t^* \right] \tag{3.60}
\]

where \( d_t^* = x_t^* v_t - \sum_{u \neq t} x_u d_t \) has finite fourth moments. Recall that from (3.50) and (3.55) we have

\[
\hat{c}_0 - c_0 = o_p(\max(n^{-1}, T^{-1/2})) = o_p(n^{-1/2})
\]

and

\[
\hat{c}_1 - c_1 = o_p(\max(n^{-1}, T^{-1/2})) = o_p(n^{-1/2})
\]

by Assumption (3.34). Using Assumptions (2.7), (3.38), and the fact that the random parts of \( d_t^* \) have bounded fourth moments, we apply Lemmas 3.3 and 3.4 to expression (3.60) and obtain (3.59).

For Theorems 2.2 and 3.2 when we either have no error in the equation or \( \sigma_{qq} = T^{-1/2} \), we note that \( T^{1/2} d_t \) has finite 4 + \( \delta \) moments. Thus, for \( \pi_t^* \) defined by (3.39),
\[
\lim_{n \to \infty} n^{-1/2} \prod_{t=1}^{n} \left( \frac{\hat{\pi}_t - \pi_t}{d_t} \right) = \lim_{n \to \infty} n^{-1/2} \prod_{t=1}^{n} \left( \frac{\hat{\pi}_t - \pi_t}{d_t \sqrt{T}} \right) = 0
\]

by the same arguments used to show Result 3.1. Note also that in these two cases instead of the definition of \( \hat{x}_t \) in (3.33), we may simply define

\[
\hat{x}_t = \hat{X}_t - \hat{v}_t \delta_t , \quad (3.61)
\]

for then, (3.41) may be shown as follows,

\[
\hat{x}_t - x_t = u_t - \hat{v}_t \delta_t \\
= u_t - \hat{v}_t \delta_t - \hat{v}_t (\delta_t - \delta_t) - (\hat{v}_t - v_t) \delta_t \\
= 0_p (T^{-1/2}).
\]

We need to retain Assumptions (3.35) and (3.36) on the \( u_{xt} \).

2. **Non-linear regression to estimate variance functions**

In Section 3.B.1 we considered the variance function to be linear. We may, however, have non-linear variance functions such as

\[
\sigma_{vtt} = (\theta_0 + \theta_1 x_t)^2
\]

or in terms of the weights
We first outline the Gauss-Newton estimation procedure in the usual non-linear model case. Then we present an important property of the Gauss-Newton estimator.

We assume the model

\[ W_t = f(x_t; \theta_0) + \eta_t, \quad t=1, 2, \ldots, \ (3.62) \]

where the \( \eta_t \) are independent \( (0, \sigma^2) \) random variables, \( \theta_0 \) is a \( r \)-dimensional parameter contained in the parameter space \( \Theta \), \( x_t \) is an observable \( k \)-dimensional vector, and \( f(x_t; \theta) \) has continuous third derivatives with respect to \( \theta \) for all \( \theta \in \Theta \) and all \( x_t \). Assume that the vectors \( \{x_t\} \) form a fixed sequence, and assume that we have available an initial estimator of \( \theta_0 \), say \( \hat{\theta} \), that satisfies \( (\hat{\theta} - \theta_0) = o_p(b_n) \) where \( \lim_{n \to \infty} b_n = 0 \). Let \( f_j(x_t; \hat{\theta}) \) denote the first partial derivative of \( f(x_t; \theta) \) with respect to the \( j \)-th element of \( \theta \) evaluated at \( \theta = \hat{\theta}, \ x = x_t \). Also, let \( F(\hat{\theta}) \) be the \( n \times r \) matrix with \( t_j \)-th element given by \( f_j(x_t; \hat{\theta}) \).

Using the principle of least squares, we would choose as our estimator of \( \theta_0 \) the \( \hat{\theta} \) in \( \Theta \) that minimizes

\[ Q(\hat{\theta}) = n^{-1} \sum_{t=1}^{n} [W_t - f(x_t; \hat{\theta})]^2. \]

However, if we expand \( f(x_t; \theta_0) \) about \( \hat{\theta} \), we get
\[ f(x_t; \theta_0) = f(x_t; \hat{\theta}) + \sum_{j=1}^{r} f_j(x_t; \hat{\theta})(\theta_{0j} - \hat{\theta}_j) + \text{remainder} \quad (3.63) \]

where \( \theta_{0j} \) and \( \hat{\theta}_j \) are the \( j \)-th elements of \( \theta_0 \) and \( \hat{\theta} \), respectively. Thus, the sum of squares \( Q(\theta) \) is approximated by

\[ Q^*(\theta) = n^{-1} \sum_{t=1}^{n} \left[ W_t - f(x_t; \hat{\theta}) - \sum_{j=1}^{r} f_j(x_t; \hat{\theta})(\theta_{0j} - \hat{\theta}_j) \right]^2, \quad (3.64) \]

where we have ignored the remainder in (3.63).

Minimizing \( Q^*(\theta) \) with respect to \( \hat{\theta} = (\hat{\theta} - \hat{\theta}) \), we obtain the one-step Gauss-Newton estimator \( \tilde{\theta} \), where

\[ \tilde{\theta} = \hat{\theta} + \hat{\delta}, \quad (3.65) \]

\[ \hat{\delta} = \left[ F'(\hat{\theta})F(\hat{\theta}) \right]^{-1} F'(\hat{\theta})b, \]

and \( b \) is the \( n \times 1 \) vector with the \( t \)-th element given by

\[ b_t = W_t - f(x_t; \hat{\theta}). \]

Now, we state without proof a theorem found in detail in Fuller (1976, Section 5.5).

**Theorem 3.4.** Assume that the non-linear model (3.62) holds along with certain regularity conditions on the first through third derivatives of \( f(x_t; \theta) \) and that \( (\hat{\theta} - \theta_0) = O_p(b_n) \). Then, if \( \tilde{\theta} \) is the one-step
Gauss-Newton estimator of (3.65),

\[
(\hat{\theta} - \theta_0) = \left[ \mathbf{F}'(\theta_0)\mathbf{F}(\theta_0) \right]^{-1} \mathbf{F}'(\theta_0) \eta + O_p(\max\{b_n^2, b_n n^{-1/2}\}),
\]

where \( \eta \) is the \( n \times 1 \) vector with \( t \)-th element \( \eta_t \).

This theorem implies that (in most circumstances) the error in the one-step Gauss-Newton estimator will be no larger than that in the initial estimator.

For our particular situation, this means that if \( \eta_t = f(x_t; \theta_0) \) where \( f \) is a non-linear function and we have some initial estimate \( \hat{\theta} \) of \( \theta_0 \), estimated using the \( \hat{x}_t \) in place of the true \( x_t \), such that

\[
(\hat{\theta} - \theta_0) = O_p(n^{-1/2})
\]

then the one-step Gauss-Newton estimator \( \tilde{\theta} \) is such that

\[
(\tilde{\theta} - \theta_0) = O_p(n^{-1/2}). \quad (3.66)
\]

Under the appropriate definition of \( \hat{x}_t \) assuming fixed \( x_t \), and assuming either (3.33) or (3.61), we still obtain

\[
(\hat{x}_t - x_t) = O_p(T^{-1/2}) = O_p(n^{-1/2}) \quad (3.67)
\]

if \( T^{-1} = o(n^{-1}) \).
Now, write

\[ \hat{\pi}_t = f(\hat{x}_t; \hat{\theta}) \]

\[ = f(x_t; \theta_0) + \sum_{i=1}^{k} f^{(i)}(x_t; \theta_0)(x_{ti} - \hat{x}_{ti}) \]

\[ + \sum_{j=1}^{r} f^{(j)}(x_t; \theta_0)(\hat{\theta}_j - \theta_{0j}) \]  \hspace{1cm} (3.68)

+ remainder

where \( f^{(i)}(x_t; \theta_0) \) denotes the first derivative of \( f(x; \theta) \) with respect to the \( i \)-th component of \( \hat{x}_t \) evaluated at \( x_t \) and \( \theta_0 \). Note that with \( \hat{\pi}_t = f(x_t; \theta_0) \),

\[ \hat{\pi}_t - \pi_t = \sum_{i=1}^{k} f^{(i)}(x_t; \theta_0)(x_{ti} - \hat{x}_{ti}) + \sum_{j=1}^{r} f^{(j)}(x_t; \theta_0)(\hat{\theta}_j - \theta_{0j}) \]  \hspace{1cm} (3.69)

ignoring the remainder. Thus, (3.69) is in the form of (3.30) of Theorem 3.3 if we make the appropriate assumptions on the first derivatives. To apply Theorem 3.3, note that for the first \( k \) terms of (3.69) which involve \( (x_{ti} - \hat{x}_{ti}) \), the \( f^{(i)}(x_t; \theta_0) \) must satisfy the conditions of the \( a_{ij} \) needed to prove case i) of Theorem 3.3. For the next \( r \) terms which involve \( (\hat{\theta}_j - \theta_{0j}) \), the \( f^{(j)}(x_t; \theta_0) \) must satisfy the conditions of the \( a_{ij} \) for case ii) of Theorem 3.3.
Therefore, given $a_t$ and $U_t$ which meet the conditions of Theorem 3.3, we need to assume that

i) the sequences of constants $\{f^{(i)}(x_t; \theta_0)\}$ are bounded, $i=1, 2, ..., k$; and

ii) for each $j$, $j=1, 2, ..., r$, the sequences of constants $\{f^{(j)}(x_t; \theta_0)\}$ are such that

$$n^{-1/2} \sum_{t=1}^{n} f^{(j)}(x_t; \theta_0) a_t U_t = O_p(1).$$

(3.70)

Note that we may iterate the Gauss-Newton procedure and at each step obtain estimators which are at least as good as those of the previous step. Thus, under the same conditions (3.70), Theorem 3.3 also applies when we have iterative Gauss-Newton estimation with a fixed finite number of steps.

Consider as an example the following variance function

$$\sigma_{\varepsilon \varepsilon} = [\theta(c_0 + c_1 x_t)]^2, \quad \theta > 1.$$

Now, let $(\hat{\theta}, \hat{c}_0, \hat{c}_1)$ be the vector of iterated Gauss-Newton estimators of $(\theta, c_0, c_1)$. Then,

$$\hat{\pi}_t = [\hat{\theta}(\hat{c}_0 + \hat{c}_1 x_t)]^{-2}$$
\[ \left(\theta_0 + c_1x_t\right)^{-2} - 2\theta_0^{-3}\left(\theta_0 + c_1x_t\right)^{-2}(\theta - \theta) \]

\[ - 2\theta^{-2}\left(\theta_0 + c_1x_t\right)^{-3}(\hat{\theta} - \theta_0) - 2\theta^{-2}x_t\left(\theta_0 + c_1x_t\right)^{-3}(\hat{\theta} - \theta_0) \]

\[ - 2\theta^{-2}c_1\left(\theta_0 + c_1x_t\right)^{-3}(\hat{x}_t - x_t) + \text{remainder}, \]

\[ = \pi_t - 2\theta^{-1}\pi_t(\theta - \theta) - 2\pi_t\left(\theta_0 + c_1x_t\right)^{-1}(\hat{x}_t - \theta_0) \]

\[ - 2\pi_t x_t\left(\theta_0 + c_1x_t\right)^{-1}(\hat{c}_t - c_1) - 2\pi_tC_1\left(\theta_0 + c_1x_t\right)^{-1}(\hat{x}_t - x_t) \]

(3.71)

ignoring the remainder. These arguments lead to the following result.

Although the model in (3.73) is not identified, we can assume that we have additional information. See the example in Chapter IV.

Result 3.2. Assume the model of Theorem 3.1 holds. Assume

\[ [q_t, T^{1/2}a_t, T(c_{at} - c_{at})', (x_t - \mu_t)], t=1, 2, \ldots, n, \] are independent with bounded eighth moments. Assume

\[ \Sigma_{aatt} = T^{-1}T_{aatt}, \quad t=1, 2, \ldots, n, \quad (3.72) \]

where \( T^{-1} = \sigma(n^{-1}) \). Assume that we have fixed \( x_t, t=1, 2, \ldots, n, \) and for

\[ \pi_t = \sigma_{vtt}^{-1}[\theta(c_0 + c_1x_t)]^{-2}, \quad \theta > 1 \quad (3.73) \]
that the sequence \( \{\pi_t\} \) is bounded above and below by fixed positive numbers. Assume that we have additional information so that the model \( (3.73) \) is identified. Let the estimator of \( x_t \) defined in \( (3.33) \) be used to obtain the iterative Gauss-Newton estimator of \( \sigma_{\text{vtt}} \), say \( \hat{\sigma}_{\text{vtt}} = [\hat{\theta}(c_0 + c_1x_t)]^2 \) where we assume that

\[
[(\hat{\theta} - \theta), (c_0 - c_0), (c_1 - c_1)] = o_p(1) .
\]

Also assume

\[
n^{-1} \sum_{t=1}^{n} (c_0 + c_1x_t)^{-1} \hat{\chi}_t \hat{\chi}'_t = o_p(1)
\]

\[
n^{-1} \sum_{t=1}^{n} (c_0 + c_1x_t)^{-1} d_t = o_p(1)
\]

\[
n^{-1} \sum_{t=1}^{n} x_t (c_0 + c_1x_t)^{-1} \hat{\chi}_t \hat{\chi}'_t = o_p(1)
\]

\[
n^{-1} \sum_{t=1}^{n} x_t (c_0 + c_1x_t)^{-1} d_t = o_p(1)
\]

and

\[
n^{-1} \sum_{t=1}^{n} \left|(c_0 + c_1x_t)^{-1}\right| \tag{3.74}
\]

is bounded. Then, the conditions of Theorem 3.3 are satisfied for \( a_t^U_t = \hat{\chi}_t \hat{\chi}'_t \) and for \( a_t U_t = d_t \). Therefore, \( (3.2) \) and \( (3.4) \) of Theorem 3.1 hold.

3. A note on variance functions

We note that in either expression \( (3.33) \) or \( (3.61) \) for \( \hat{x}_t \), we have used
\[ \hat{\sigma}_t = \sigma_{uvtt} \hat{\sigma}_{vvtt} ^{-1} \]

for the simple univariate case. This becomes \( \hat{\sigma}_t = \sigma_{uvtt} \hat{\sigma}_{vvtt} ^{-1} \) in the general univariate case. Now, if the variance model assumes that

\[ \Sigma_{uvtt} = f(x_t) \]

and

\[ \sigma_{vvtt} = f(x_t) \]

where \( f(x_t) \) is some function, linear or non-linear, of \( x_t \), then we may replace \( \hat{\sigma}_t \) by

\[ \hat{\sigma} = \Sigma_{uv} \hat{\sigma}_{vv} ^{-1} \]

where \( \Sigma_{uv} = n^{-1} \sum_{t=1}^{n} \Sigma_{uvtt} \) and \( \hat{\sigma}_{vv} = n^{-1} \sum_{t=1}^{n} \hat{\sigma}_{vvtt} \).
IV. EXAMPLE: NON-HOMOGENEOUS ERROR VARIANCES

In this chapter, we consider a situation in which the error variances are non-constant. A specific type of estimators $\hat{\Sigma}_{aatt}^t$, $t=1, 2, \ldots, n$, is considered, and the variance function assumed is non-linear.

A. Functionally Related Model

Fuller (1984) considered models in which the covariance matrices of the measurement error are known functions of observable variables or are known functions of the expectation of observable random variables. He considered estimator matrices of the form

$$\hat{\Sigma}_{aatt} = \sum_{i=1}^{m} \psi_i' \psi_i$$

where $m < k + 1$, and the $\psi_i$ are observable vectors. The $\hat{\Sigma}_{aatt}$ are such that either

$$E(\hat{\Sigma}_{aatt}) = \Sigma_{aatt}$$

or

$$E(\sum_{t=1}^{n} \hat{\Sigma}_{aatt}) = \sum_{t=1}^{n} \Sigma_{aatt}.$$  

Note that if the $\Sigma_{aatt}$ are known, then they can be written in the form (4.1), as can any variance-covariance matrix. Moreover, the $\psi_i$ may
contain fixed and random parts and may be functions of $Z_t$. The model of (1.1)-(1.2) with estimator matrices of the form (4.1) is called the functionally related model.

B. Description of Hog Data

We consider data from an interview-reinterview survey of Iowa farmers conducted by the Statistical Laboratory at Iowa State University in 1970. The survey is described in Battese, Fuller, and Hickman (1972, 1976).

The survey was a personal-interview farm survey for estimating land use, crop acreage, livestock numbers and farm labor. Most of the questionnaire items were taken from the USDA's 1970 June acreage, livestock and labor enumeration survey questionnaire. The 1970 interview-reinterview survey involved drawing an area sample of farm operators in each of three geographic areas in Iowa. Interviews with eligible farm operators were obtained once in the first week of September 1970 and again one month later. Items on the trial-2 questionnaires were constructed so that the exact question was either (i) repeated with reference to the date of the first interview or (ii) tied to the date of the second interview with other items included to obtain any changes in the inventory of the variable in question.

In trial 1, 92.0% of the eligible farm operators were actually interviewed, and in trial 2, 91.8% of those assigned for the trial were interviewed. The total number of farm operators interviewed twice was 262. Of the 21 variates on which data were collected, we consider only
two. These variables are

\[ x_t = \text{number of breeding hogs the } t\text{-th farm operator had on hand September 1, 1970.} \]

and

\[ y_t = \text{number of sows giving birth to baby pigs between June 1 and August 31, 1970.} \]

C. Analysis of Data

Since we have two observations for each individual for each variable, we consider the model for the means of the two responses. Let \( \mathbf{x}_t = (1, x_t) \) and \( \beta' = (\beta_0, \beta_1) \), with \( x_t \) fixed. Then consider the model

\[ y_t = x_t \beta + q_t , \]

\[ \bar{y}_t = y_t + \bar{q}_t , \quad (4.3) \]

\[ \bar{x}_t = x_t + \bar{u}_t , \]

where

\[ \bar{y}_t = \frac{1}{2} (y_{t1} + y_{t2}) , \]

\[ \bar{x}_t = \frac{1}{2} (x_{t1} + x_{t2}) , \]
\[ \bar{w}_t = \frac{1}{2} (w_{t1} + w_{t2}) , \]
\[ \bar{u}_t = \frac{1}{2} (u_{t1} + u_{t2}) , \]
\[ Y_{tj} = y_t + w_{tj} , \]
\[ X_{tj} = x_t + u_{tj} , \]

and j=1, 2 denote the first and second observation on the t-th respondent respectively. We assume that the two vectors
\[ T^{1/2}(w_{t1}, u_{t1}) \] and \[ T^{1/2}(w_{t2}, u_{t2}) \] are independent with zero mean vector and bounded \( 8 + \delta \) moments, \( \delta > 0 \). We also assume that the \( q_t \) are independent with zero means and bounded \( 4 + \delta \) moments and that \( q_t \) is independent of \( (w_{j1}, u_{j1}, w_{j2}, u_{j2}) \) for all \( t \) and \( j \). Let \( E(\bar{w}_t, \bar{u}_t)'(\bar{w}_t, \bar{u}_t) = \Sigma = T^{-1} \Omega T^{-1} \), \( t=1, 2, \ldots, n \), where \( T^{-1} = o(n^{-1}) \) and assume \( \sigma^2 > 0 \) is unknown.

Let
\[ \psi_{t1} = \frac{1}{2} \{(Y_{t1} - Y_{t2}), (X_{t1} - X_{t2})\} . \]  \( (4.4) \)

Then, \( E(\psi_{t1}'\psi_{t1}) = \Sigma = T^{-1} \Omega T^{-1} \) by the independence of \( (w_{t1}, u_{t1}) \) and \( (w_{t2}, u_{t2}) \). Therefore, the covariance matrix of measurement errors is a known function of the expectation of observable random variables. We use Theorem 3.3 and the non-linear results of Chapter 3 to analyze the data.
First, we note that for estimation purposes, all observation for which the farmer indicated in both interviews that he/she had no breeding hogs as of September 1, 1970, were deleted. The subset of the data with zeros deleted is given in Fuller (1987, Appendix 3.A). Then, using the functionally related option of SUPER CARP (Hidiroglou et al., 1980), we obtain initial estimates for $\beta_0$ and $\beta_1$ where we use $\pi_t \equiv 1$ and estimate the $t$-th error variance by $\psi_t^\prime \psi_t$. The averages of these error variances obtained are the same as those estimated by an analysis of variance of the dependent and independent variables on individuals. For this data, $\sigma_{uut}^{-1} \sigma_{xxt}$ is approximately 0.08 where the approximation is obtained by looking at the average estimates. Thus, we obtain

$$\hat{y}_t = 0.132 + 0.2801 x_t$$

and

$$\text{vech} \{\hat{\beta}_0, 100 \hat{\beta}_1\} = (2.3864, -7.0196, 24.3443)' . \quad (4.5)$$

The estimator of variance given in (4.5) is defined in Theorem 2.1 assuming $\pi_t \equiv 1$ and using the error variance estimates found by the analysis of variance. However, by looking at the graph of $\hat{v}_t$ versus $\hat{x}_t$, where $\hat{v}_t$ and $\hat{x}_t$ are calculated from the expressions of Result 3.1 with $p = 3$, we see that the range in estimated error variances, $\hat{c}_{vxt}$, is large and that the variance seems to be related to $x_t$ (see Figure 4.1). Thus, we attempt to improve our estimate of $\beta$ by
Figure 4.1. Predicted $x_t$ and $v_t$

weighting by appropriate estimates of the $a_{vvt}$, $t=1, 2, \ldots, n$.

Let $r_t = w_t - \bar{u}_t \beta_1$. We assume that

$$E(v_t^2) = a_{vvt} = [\theta(c_0 + c_1x_t)]^2, \quad \theta > 1$$

and
Note that the model of (4.6) assumes that \( \sigma_{qtt} \) is a function of the \( x_t \).

Now, given our estimate \( (\hat{\beta}_0, \hat{\beta}_1) \) of \( (\beta_0, \beta_1) \), let

\[
\hat{r}_t = \frac{1}{2} [y_{t1} - y_{t2} - \hat{\beta}_1(x_{t1} - x_{t2})]
\]

and

\[
\hat{v}_t = \bar{y}_t - y - \hat{\beta}_1(\bar{x}_t - \bar{x}) \tag{4.7}
\]

where \( \bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t \) and \( \bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t \). Then,

\[
\hat{r}_t^2 = (0.25)[y_{t1} - y_{t2} - \hat{\beta}_1(x_{t1} - x_{t2})]^2
\]

\[
= (0.25)[\bar{y}_{t1} - \bar{y}_{t2} - \hat{\beta}_1(\bar{x}_{t1} - \bar{x}_{t2})]^2
\]

is an estimator of \( (1, -\beta_1) \Sigma_{eatt} (1, -\beta_1)' \) and

\[
\hat{v}_t^2 = [\bar{y}_t - y - \hat{\beta}_1(\bar{x}_t - \bar{x})]^2
\]

\[
= [(\bar{y}_t - \bar{y}) - \hat{\beta}_1(\bar{x}_t - \bar{x}) + (\bar{w}_t - \bar{w}) - \hat{\beta}_1(\bar{u}_t - \bar{u})]^2
\]

is an estimator of \( \sigma_{qtt} + (1, -\beta_1) \Sigma_{eatt} (1, -\beta_1)' \).
Let

\[ x_t = \bar{x}_t - \sigma_v \sigma_{vvv} t \]

where \( \sigma_{vv} = 94.9116 \) comes from the SUPER CARP run and
\[ \sigma_v = \sigma_{uu} - \beta_1 \sigma_{uu} = -15.1182 \]
with \( \sigma_{uu} \) and \( \sigma_{uu} \) the appropriate elements of \( \sum_{t=1}^{n} \psi_t \). Note that in our expression for \( x_t \) we have \( \sigma_{uu} \) and \( \sigma_{vv} \) instead of \( \sigma_{uvv} \) and \( \sigma_{vvv} \). We can use this expression here because our model (4.6) assumes that both \( \sigma_{uvv} \) and \( \sigma_{vvv} \) are proportional to the same function of the \( x_t \) (see Section 3.3).

Using PROC NLIN of SAS (1985), we fit the model

\[ r_t^2 = (c_0 + c_1 x_t)^2 + \eta_{t-1} \] (4.8)

\[ v_t^2 = [\theta(c_0 + c_1 x_t)]^2 + \eta_{t-2} , \quad \theta > 1 \]

by iterative non-linear estimation, weighting at successive steps by the predicted values from the previous step. The restriction \( \theta > 1 \) is equivalent to the restriction \( \sigma_{qqq} > 0 \). Our final estimates were

\[ (c_0, c_1, \theta) = (0.675, 0.133, 1.643) . \]

Thus, our estimate of \( \sigma_{vvv} \) is
\[ \tilde{\sigma}_{yvt} = \left[ \theta(\hat{c}_0 + \hat{c}_1 x_t) \right]^2 \]

\[ = (1.109 + 0.219 x_t)^2. \quad (4.9) \]

From Result 3.2 we know that the inverse of \( \tilde{\sigma}_{yvt} \), \( t=1, 2, \ldots, n \), above satisfies the conditions for the \( \pi_t \) in Theorem 3.4.

Incorporating these weights, we get from SUPER CARP

\[ \hat{y}_t = -0.006 + 0.287 x_t \]

and

\[ \text{vech} \ \hat{V}(\hat{\beta}_0, 100 \hat{\beta}_1) = (0.1679, -0.7404, 7.1496). \quad (4.10) \]

Comparing (4.5) and (4.10), we see that for this example, weighting by estimated variances has produced an estimated variance for the intercept that is less than one-fourteenth of the original and an estimated variance for \( \beta_1 \) that is less than one-third that before weighting.

In order to consider a residual plot after weighting, define

\[ \hat{x}_t = \bar{x}_t - \hat{\sigma}_{yvt} \tilde{\sigma}_{yvt}^{-1} \hat{v}_t \]

and

\[ \hat{v}_t = \bar{v}_t - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_t. \]
where

\[ \hat{\sigma}_{\text{vwtt}} = \hat{\sigma}_{\text{wtt}} - 2 \beta_1 \hat{\sigma}_{\text{wtt}} + \beta_1^2 \hat{\sigma}_{\text{uutt}} \]

\[ \hat{\sigma}_{\text{uvtt}} = \hat{\sigma}_{\text{wtt}} - \beta_1 \hat{\sigma}_{\text{uutt}} \]

\[ \left( \hat{\sigma}_{\text{wtt}}, \hat{\sigma}_{\text{wtt}}, \hat{\sigma}_{\text{uutt}} \right) = \frac{1}{4} \left( (y_{t1} - y_{t2})^2, (y_{t1} - y_{t2})(x_{t1} - x_{t2}), (x_{t1} - x_{t2})^2 \right) \]

If we now look at the graph of the weighted \( \hat{\nu}_t \), defined as \( \frac{1}{\hat{\sigma}_{\text{vwtt}}} \hat{\nu}_t \), versus the \( \hat{x}_t \) (see Figure 4.2), we see that most of the heterogeneity of variances for the \( \nu_t \) has been removed.
Figure 4.2. $\hat{x}_t$ and $\hat{v}_t$ after weighting

**Figure 4.2.** $\hat{x}_t$ and $\hat{v}_t$ after weighting
In previous chapters, we examined estimators for the situation in which we have unequal error variances. We examined the limiting behavior of these estimators so that we are able to use them when we have large samples. For practical situations, however, we need to know what may be considered to be a large sample. In this chapter we use Monte Carlo with using data based upon the example in Chapter 4 to explore this matter.

We selected a (nearly) systematic sample of size 50 from the 184 ordered observations of $X^*_t$ of the example of Chapter 4 and considered this sample to be the true (fixed) $x^*_t$. The sample of $x^*_t$ is contained in Table 5.1 below.

To create observations modeled after the actual data, we assume

$$ (\sigma_{\text{wtt}}, \sigma_{\text{uutt}}) = (0.13, 0.22) x^*_t, \ t=1, \ldots, 50, \quad (5.1) $$

where the coefficients on $x^*_t$ were estimated from the actual data by the regression of

$$ |Y_{t1} - Y_{t2}| \text{ on (1) and } X^*_t $$

and

$$ |X_{t1} - X_{t2}| \text{ on (1) and } X^*_t. $$
Table 5.1. Systematic sample of data, $x_t$

<table>
<thead>
<tr>
<th>Observation Number</th>
<th>$x_t$</th>
<th>Observation Number</th>
<th>$x_t$</th>
<th>Observation Number</th>
<th>$x_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>18</td>
<td>22.0</td>
<td>35</td>
<td>44.0</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>19</td>
<td>22.0</td>
<td>36</td>
<td>44.0</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>20</td>
<td>23.5</td>
<td>37</td>
<td>46.0</td>
</tr>
<tr>
<td>4</td>
<td>8.5</td>
<td>21</td>
<td>25.0</td>
<td>38</td>
<td>47.0</td>
</tr>
<tr>
<td>5</td>
<td>10.5</td>
<td>22</td>
<td>26.5</td>
<td>39</td>
<td>48.5</td>
</tr>
<tr>
<td>6</td>
<td>11.5</td>
<td>23</td>
<td>28.0</td>
<td>40</td>
<td>51.5</td>
</tr>
<tr>
<td>7</td>
<td>12.5</td>
<td>24</td>
<td>28.5</td>
<td>41</td>
<td>54.5</td>
</tr>
<tr>
<td>8</td>
<td>13.0</td>
<td>25</td>
<td>30.5</td>
<td>42</td>
<td>61.0</td>
</tr>
<tr>
<td>9</td>
<td>14.5</td>
<td>26</td>
<td>31.5</td>
<td>43</td>
<td>62.0</td>
</tr>
<tr>
<td>10</td>
<td>16.0</td>
<td>27</td>
<td>33.0</td>
<td>44</td>
<td>70.5</td>
</tr>
<tr>
<td>11</td>
<td>17.0</td>
<td>28</td>
<td>33.5</td>
<td>45</td>
<td>76.0</td>
</tr>
<tr>
<td>12</td>
<td>17.0</td>
<td>29</td>
<td>35.0</td>
<td>46</td>
<td>79.5</td>
</tr>
<tr>
<td>13</td>
<td>18.0</td>
<td>30</td>
<td>36.5</td>
<td>47</td>
<td>106.0</td>
</tr>
<tr>
<td>14</td>
<td>19.0</td>
<td>31</td>
<td>38.0</td>
<td>48</td>
<td>110.5</td>
</tr>
<tr>
<td>15</td>
<td>20.0</td>
<td>32</td>
<td>39.5</td>
<td>49</td>
<td>128.0</td>
</tr>
<tr>
<td>16</td>
<td>20.5</td>
<td>33</td>
<td>41.5</td>
<td>50</td>
<td>151.5</td>
</tr>
<tr>
<td>17</td>
<td>21.0</td>
<td>34</td>
<td>43.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relationship (5.1) also seemed reasonable from graphs of $\overline{Y}_t$ vs. $\overline{X}_t$ and of $(X_{t1} - X_{t2})$ vs. $\overline{X}_t$ which indicated that the variances were nearly multiples of $\overline{X}_t^2$.

To begin with, we generated 200 samples of size 50 according to the model

$$Y_{t1} = 0.3x_t + \omega_{t1} + \epsilon_t$$

$$Y_{t2} = 0.3x_t + \omega_{t2} + \epsilon_t$$

$$X_{t1} = x_t + \epsilon_{t1}$$

$$X_{t2} = x_t + \epsilon_{t2}$$

where
\[ w_{tj} = 0.13 \ a_{tj} x_t \]
\[ u_{tj} = 0.22 \ b_{tj} x_t \]  \hspace{1cm} (5.3)

and \((a_{t1}, a_{t2}, b_{t1}, b_{t2}, q_t)' \sim \text{NID}(0, \text{diag}(1, 1, 1, 1, 1))\).

For each sample, we then calculated

\[ \mathbf{z}_t = (\bar{y}_t, 1, \bar{x}_t) \]
\[ \Psi_t = \left[ 0.5 (y_{t1} - y_{t2}), 0, 0.5 (x_{t1} - x_{t2}) \right] \]

\[ \hat{\mathbf{m}}_{zz} = n^{-1} \left( \sum_{t=1}^{n} \mathbf{z}_t \mathbf{z}_t' - \frac{n}{2} \Psi_t \Psi_t' \right) \]
\[ \hat{\mathbf{g}} = (\hat{\beta}_0, \hat{\beta}_1)' = \mathbf{M}_{xx}^{-1} \Sigma_{x} \]

\[ \hat{d}_t = (1, \bar{x}_t)' \mathbf{v}_t - (\hat{\Sigma}_{uutt} - \hat{\Sigma}_{uwtt}) \hat{\beta} \]
\[ \hat{\mathbf{v}}_t = \bar{y}_t - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_t \]

\[ \hat{\xi}_{attt} = \Psi_t' \Psi_t \]

\[ \hat{\mathbf{g}} = (n - 2)^{-1} \sum_{t=1}^{n} \hat{d}_t' \hat{d}_t \]

and

\[ \hat{\mathbf{v}}(\hat{\mathbf{g}}) = n^{-1} \mathbf{M}_{xx}^{-1} \mathbf{G} \mathbf{M}_{xx}^{-1} \] \hspace{1cm} (5.4)
We also calculated the ordinary test statistics

\[ t_0 = [V(\hat{\beta}_0)]^{-1/2} \hat{\beta}_0 \]

and

\[ t_1 = [V(\hat{\beta}_1)]^{-1/2} (\hat{\beta}_1 - 0.3) . \]

To construct an estimator of \( v_t \), let

\[ \hat{v}_t = \bar{Y}_t - \bar{Y} - (\bar{X}_t - \bar{X}) \hat{\beta}_1 \]

\[ \hat{\Sigma}_{aa} = n^{-1} \sum_{t=1}^{n} \hat{v}_{aat} \hat{v}_{att} \]

\[ \text{vech} \ \hat{\Sigma}_{aa} = (\hat{\sigma}_{ww}, 0, \hat{\sigma}_{wu}, 0, 0, \hat{\sigma}_{uu})' \]

and

\[ \hat{\sigma}_{qq} = \max\{0, (n-2)^{-1} \sum_{t=1}^{n} v_t^2 - (\hat{\sigma}_{ww} + 2\hat{\beta}_1 \hat{\sigma}_{wu} + \hat{\beta}_1^2 \hat{\sigma}_{uu})\} . \]

Also, let

\[ \hat{x}_t = \bar{x} + [(\hat{\sigma}_{qq} + \hat{\sigma}_{ww})^{-1} \hat{\beta}_1^2 + \hat{\sigma}_{uu}]^{-1} [(\hat{\sigma}_{qq} + \hat{\sigma}_{ww})^{-1} (\bar{Y}_t - \bar{Y}) \hat{\beta}_1 + \hat{\sigma}_{uu} (\bar{x}_t - \bar{x})] \]

and

\[ r_t = \frac{1}{2} [Y_{t1} - Y_{t2} - \hat{\beta}_1 (X_{t1} - X_{t2})] \]

(5.5)
Note that our model assumes $\sigma_{\text{wutt}} = 0$ for every $t$. However, we can estimate $\sigma_{\text{wutt}}$ and have used the average estimate in defining $\hat{\sigma}_{qq}$ but have chosen not to do so in defining $\hat{x}_t$. Now, by assumption,

$$
\sigma_{\text{rutt}} = (c_0 + c_1x_t)^2, \quad c_0 > 0, \quad c_1 > 0.
$$

Thus, we regressed $|r_t|$ on (1) and $\hat{x}_t$ to obtain $\hat{c}_0$ and $\hat{c}_1$.

Then, we let

$$
\hat{\sigma}_{\text{vutt}} = \hat{\sigma}_{qq} + (\hat{c}_0 + \hat{c}_1x_t)^2. \quad (5.6)
$$

We then used the weights

$$
\pi_t = \frac{1}{\hat{\sigma}_{\text{vutt}}}, \quad t=1, 2, \ldots, n
$$

to get weighted estimates $\hat{\beta}^*$ and $\hat{V}(\hat{\beta}^*)$. We also calculated the test statistics

$$
t_0^* = \left[ V(\hat{\beta}^*) \right]^{-1/2} \hat{\beta}_0^*
$$

and

$$
t_1^* = \left[ V(\hat{\beta}_1^*) \right]^{-1/2} (\hat{\beta}_1^* - 0.3).
$$

For the 200 samples of size $n = 50$, the results are found in Tables 5.2 and 5.3. For the unweighted estimator, the theoretical
Table 5.2. Monte Carlo percentiles, mean, and variance for unweighted and weighted estimators for 200 samples, \((\beta_0, \beta_1) = (0, 0.3), n = 50\)

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>(\hat{\beta}_0)</th>
<th>(\hat{\beta}_0^*)</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_1^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-5.662</td>
<td>-2.154</td>
<td>0.1602</td>
<td>0.2047</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.985</td>
<td>-1.511</td>
<td>0.2012</td>
<td>0.2404</td>
</tr>
<tr>
<td>0.10</td>
<td>-2.473</td>
<td>-1.213</td>
<td>0.2457</td>
<td>0.2568</td>
</tr>
<tr>
<td>0.25</td>
<td>-1.494</td>
<td>-0.574</td>
<td>0.2707</td>
<td>0.2804</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.169</td>
<td>0.036</td>
<td>0.3008</td>
<td>0.2996</td>
</tr>
<tr>
<td>0.75</td>
<td>0.899</td>
<td>0.450</td>
<td>0.3452</td>
<td>0.3198</td>
</tr>
<tr>
<td>0.90</td>
<td>1.957</td>
<td>0.889</td>
<td>0.3787</td>
<td>0.3424</td>
</tr>
<tr>
<td>0.95</td>
<td>2.828</td>
<td>1.278</td>
<td>0.3912</td>
<td>0.3596</td>
</tr>
<tr>
<td>0.99</td>
<td>4.119</td>
<td>2.427</td>
<td>0.4619</td>
<td>0.3983</td>
</tr>
</tbody>
</table>

Monte Carlo mean

<table>
<thead>
<tr>
<th></th>
<th>(-0.262)</th>
<th>(-0.051)</th>
<th>0.3066</th>
<th>0.3006</th>
</tr>
</thead>
<tbody>
<tr>
<td>variance</td>
<td>3.178</td>
<td>0.687</td>
<td>0.003261</td>
<td>0.001215</td>
</tr>
</tbody>
</table>

Table 5.3. Monte Carlo percentiles for the Studentized statistics, 200 samples, size \(n = 50\)

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>(t_0)</th>
<th>(t_0^*)</th>
<th>(t_1)</th>
<th>(t_1^*)</th>
<th>Student's t (48 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-2.84</td>
<td>-26.73</td>
<td>-3.57</td>
<td>-3.25</td>
<td>-2.33</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.36</td>
<td>-4.52</td>
<td>-2.60</td>
<td>-2.01</td>
<td>-1.65</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.92</td>
<td>-3.09</td>
<td>-1.62</td>
<td>-1.63</td>
<td>-1.28</td>
</tr>
<tr>
<td>0.90</td>
<td>1.74</td>
<td>2.19</td>
<td>1.84</td>
<td>1.55</td>
<td>1.28</td>
</tr>
<tr>
<td>0.95</td>
<td>2.37</td>
<td>2.90</td>
<td>2.35</td>
<td>2.36</td>
<td>1.65</td>
</tr>
<tr>
<td>0.99</td>
<td>3.42</td>
<td>4.81</td>
<td>4.18</td>
<td>4.12</td>
<td>2.33</td>
</tr>
</tbody>
</table>
variance of \( \hat{\beta}_1 \) for a sample of size 50 is 0.002852. See Hasabelnaby (1985). The Monte Carlo variance of 0.003261 is close to the theoretical value but larger as expected.

Note that for both \( \beta_0 \) and \( \beta_1 \), weighting the samples of 50 observations greatly reduced the variance of the estimators. The variance for the weighted estimator of \( \beta_0 \) is less than one-fourth that of the unweighted estimator while for \( \beta_1 \) the weighted estimator's variance is less than one-half that of the unweighted estimator. However, the Studentized statistic for the weighted estimator of \( \beta_0 \) is more skewed to the left than that of the unweighted estimator. The Studentized statistic for the weighted estimator of \( \beta_1 \) is slightly superior to that for the unweighted estimator but the tails for both estimators are too wide.

The distribution for \( t^* \) may have become more skewed due to obtaining very small weights \( \hat{\sigma}_{vtt} \) for some observations. In the regression estimation of the variance of \( \sigma_{vtt} \), the usual regression statistics provide an estimator of the variance of the estimated \( \sigma_{vtt} \). A possible modification of the weighting procedure is to replace the estimated weight

\[
\frac{1}{\hat{\sigma}_{vtt}}
\]

with the weight

\[
\left( \hat{\sigma}_{vtt}^2 + \hat{V}[\hat{\sigma}_{vtt}] \right)^{-1}\hat{\sigma}_{vtt},
\]
where $\hat{\sigma}_{\text{vvtt}}$ is the regression estimator and $\hat{V}\{\hat{\sigma}_{\text{vvtt}}\}$ is the regression estimator of the variance of $\hat{\sigma}_{\text{vvtt}}$. This type of modification has been suggested in a number of contexts. For example, see Fuller (1980). As an approximation to this procedure we constructed weights with fixed modification. To improve upon our estimate $\hat{\sigma}_{\text{vvtt}}$, we first constructed an estimator of $\sigma_{\text{vvtt}}$. Let

$$r_t = c_0 + c_1 x_t$$

$$r^*_t = r_t(r_t + 1)^{-1}$$

$$(1)^* = (1)(r_t + 1)^{-1}$$

$$x^*_t = x_t(r_t + 1)^{-1}$$

(5.7)

where we have added one to the weight $r_t$ to ensure that we have no extremely small estimated weights. Then, we regressed $r^*_t$ on $(1)^*$ and $x^*_t$ to get estimates $(c^*_0, c^*_1)$ of $(c_0, c_1)$ where we restricted $c^*_0 > 0$ and $c^*_1 > 0$. Then, we let

$$\hat{\sigma}^*_{\text{vvtt}} = \hat{\sigma}_{qq} + (\hat{c}^*_0 + \hat{c}^*_1 x_t)^2$$

(5.8)

and used as weights

$$\hat{\pi}_t = (\hat{\sigma}^*_{\text{vvtt}} + 1)^{-1}$$
where once again we added one to ensure that no weight was extremely small. We did this for two different sets of samples – 1000 samples of size $n = 50$ and 1000 samples of size $n = 200$ where the 200 observations for each sample in the second set were generated from our 50 fixed $x_t$ with 4 observations for each $x_t$. The same model as before was used. Both weighted and unweighted estimates were calculated.

First consider the results for $\hat{\beta}_0$ which are given in Tables 5.4 and 5.5.

Note that after these modifications, for both sets of samples, the variance of the weighted estimator is at most about one-fifth that of the unweighted estimator. Also, the Studentized statistics no longer

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_0$</td>
<td>$\hat{\beta}_0^*$</td>
</tr>
<tr>
<td>0.01</td>
<td>-4.147</td>
<td>-1.703</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.745</td>
<td>-1.202</td>
</tr>
<tr>
<td>0.10</td>
<td>-2.228</td>
<td>-0.960</td>
</tr>
<tr>
<td>0.25</td>
<td>-1.221</td>
<td>-0.489</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.081</td>
<td>-0.053</td>
</tr>
<tr>
<td>0.75</td>
<td>1.004</td>
<td>0.435</td>
</tr>
<tr>
<td>0.90</td>
<td>1.947</td>
<td>0.918</td>
</tr>
<tr>
<td>0.95</td>
<td>2.479</td>
<td>1.259</td>
</tr>
<tr>
<td>0.99</td>
<td>3.577</td>
<td>1.909</td>
</tr>
</tbody>
</table>

Monte Carlo mean

<table>
<thead>
<tr>
<th></th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>variance</td>
<td>-0.099</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>2.656</td>
<td>0.550</td>
</tr>
</tbody>
</table>
Table 5.5. Monte Carlo percentiles for the Studentized statistics for the estimators of $\beta_0$ for two sets of 1000 samples each

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_0$</td>
<td>$t_{0}^{*}$</td>
</tr>
<tr>
<td>0.01</td>
<td>-2.93</td>
<td>-2.48</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.13</td>
<td>-1.81</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.64</td>
<td>-1.41</td>
</tr>
<tr>
<td>0.90</td>
<td>1.62</td>
<td>1.26</td>
</tr>
<tr>
<td>0.95</td>
<td>2.02</td>
<td>1.53</td>
</tr>
<tr>
<td>0.99</td>
<td>3.01</td>
<td>2.69</td>
</tr>
</tbody>
</table>

become so negatively skewed after weighting. Note also that the distributions of Studentized statistics for $\beta_0$ more closely resemble that of Student's t-distribution for the samples of size 200 than for those of size 50. Also for both of the sets of samples, the distributions of $t_0$ and $t_{0}^{*}$ are quite similar.

Now, consider the Monte Carlo results for $\beta_1$ which are found in Tables 5.6 and 5.7. After the modifications the variance of the weighted estimator of $\beta_1$ is even smaller than before. Now, the variance of the weighted estimator is nearly one-third that of the unweighted estimator. Also, we see that even for the smaller samples, the distribution of the Studentized statistic after weighting more closely resembles that of Student's t-distribution than it does before weighting. Thus, we see that for our Studentized statistics for this type of data to resemble Student's t, we need to have large samples. Also, we see that weighting improves our estimates even if we have smaller samples.
Table 5.6. Monte Carlo percentiles, mean, and variances for estimators of $\beta_1$ for the two sets of 1000 samples each ($\beta_1 = 0.3$)

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>$\hat{\beta}_1^*$</td>
<td>$\hat{\beta}_1$</td>
<td>$\hat{\beta}_1^*$</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1764</td>
<td>0.2225</td>
<td>0.2387</td>
<td>0.2661</td>
</tr>
<tr>
<td>0.05</td>
<td>0.2174</td>
<td>0.2504</td>
<td>0.2579</td>
<td>0.2768</td>
</tr>
<tr>
<td>0.10</td>
<td>0.2342</td>
<td>0.2636</td>
<td>0.2663</td>
<td>0.2825</td>
</tr>
<tr>
<td>0.25</td>
<td>0.2664</td>
<td>0.2805</td>
<td>0.2829</td>
<td>0.2916</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3030</td>
<td>0.3019</td>
<td>0.3013</td>
<td>0.3011</td>
</tr>
<tr>
<td>0.75</td>
<td>0.3393</td>
<td>0.3219</td>
<td>0.3189</td>
<td>0.3106</td>
</tr>
<tr>
<td>0.90</td>
<td>0.3702</td>
<td>0.3415</td>
<td>0.3353</td>
<td>0.3190</td>
</tr>
<tr>
<td>0.95</td>
<td>0.3914</td>
<td>0.3525</td>
<td>0.3448</td>
<td>0.3239</td>
</tr>
<tr>
<td>0.99</td>
<td>0.4357</td>
<td>0.3777</td>
<td>0.3625</td>
<td>0.3342</td>
</tr>
</tbody>
</table>

Monte Carlo mean

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>0.3029</td>
<td>0.3015</td>
<td>0.3012</td>
<td>0.3010</td>
</tr>
<tr>
<td></td>
<td>0.002854</td>
<td>0.0009613</td>
<td>0.0007028</td>
<td>0.0002023</td>
</tr>
</tbody>
</table>

Table 5.7. Monte Carlo percentiles for the Studentized statistics for the estimators of $\beta_1$ for two sets of 1000 samples each

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perentiles</td>
<td>$t_1$</td>
<td>$t_1^*$</td>
<td>$t_1$</td>
<td>$t_1^*$</td>
</tr>
<tr>
<td>0.01</td>
<td>-3.68</td>
<td>-2.64</td>
<td>-2.82</td>
<td>-2.43</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.32</td>
<td>-1.66</td>
<td>-1.92</td>
<td>-1.52</td>
</tr>
<tr>
<td>0.10</td>
<td>-1.75</td>
<td>-1.21</td>
<td>-1.46</td>
<td>-1.17</td>
</tr>
<tr>
<td>0.90</td>
<td>1.68</td>
<td>1.32</td>
<td>1.42</td>
<td>1.27</td>
</tr>
<tr>
<td>0.95</td>
<td>2.39</td>
<td>1.65</td>
<td>1.84</td>
<td>1.61</td>
</tr>
<tr>
<td>0.99</td>
<td>3.44</td>
<td>2.42</td>
<td>2.58</td>
<td>2.19</td>
</tr>
</tbody>
</table>
VI. COMPUTER PROGRAM FOR WEIGHTING BY ESTIMATED VARIANCES

In previous chapters, we have discussed the conditions under which estimated variances may be used as weights. In this chapter we discuss the basic algorithms for weighting that have been incorporated into the program EV CARP (1987). EV CARP is written in FORTRAN and designed for use on the IBM Personal Computer with math co-processor. The program is designed to estimate linear regression equations under various errors-in-variables assumptions. It also can be used to perform weighted least squares estimation. The procedures of EV CARP generalize those in Fuller (1987) and permit the observations to come from a complex survey.

Let the observation vector be

$$(Y_{ij\ell}, X_{ij\ell1}, X_{ij\ell2}, \ldots, X_{ij\ell k}) = (Y_{ij\ell}, X_{ij\ell})$$

where $i$ is the stratum identification, $i = 1, 2, \ldots, L$; $j$ is the cluster identification, $j = 1, 2, \ldots, n_i$; $\ell$ is the element-within cluster identification, $\ell = 1, 2, \ldots, m_{ij}$; $X_{ij\ell r}$ is the $ij\ell$-th observation for the $r$-th explanatory variable, $r = 1, 2, \ldots, k$; and $Y_{ij\ell}$ is the $ij\ell$-th observation for the dependent variable. From now on we will usually denote the subscript triple $ij\ell$ by $t$ where $t = 1, 2, \ldots, n$ and $n = \sum_{i=1}^{L} \sum_{j=1}^{n_i} m_{ij}$ is the total number of observations.

The matrices underlying the computations are $X'WX$ and $X'WY$ where the $rs$-th element of $X'WX$ is
and the $r$-th element of $X'WY$ is

$$
\sum_{t=1}^{n} X_{r}Y_{s}W_{t}^{*} \quad (6.2)
$$

where $r=1, 2, \ldots, k$; $s=1, 2, \ldots, k$; and $W_{t}^{*} = \hat{W}_{ij \&}$ is the sample weight associated with the $\ell$-th element in the $j$-th cluster in the $i$-th stratum. Ordinarily, the weight used is proportional to the reciprocal of the selection probability. Note that for a simple random sample the sample weights are identically equal to one, i.e., $\hat{W}_{ij \&} = 1$. The $X$ and $Y$ variables and the weights $W_{t}^{*}$ are read into the program by the user.

When the functionally related option is selected in EV CARP, the user gives the error covariance matrices by specifying $m$ vectors of $k + 1$ elements each where $m < k + 1$. The program calculates the average error covariance matrix

$$
\hat{\Sigma}_{\text{error}} = c^{-1} \sum_{t=1}^{n} W_{t}^{*} \sum_{i=1}^{m} \psi_{ti} \psi_{ti}^{*} = c^{-1} \sum_{t=1}^{n} W_{t}^{*} \psi_{ti} \psi_{ti}^{*}, \quad (6.3)
$$

where

$$
c = \sum_{i=1}^{L} \sum_{j=1}^{n_{i}} \sum_{\ell=1}^{n_{ij \&}} W_{ij \&} = \sum_{t=1}^{n_{t}} W_{t}^{*}
$$

and
\[ w_{ti} = (a_{11} R_{t1}, a_{21} R_{t2}, \ldots, a_{ki} R_{tk}, a_{k+1,1} R_{t1, k+1,1}) \]

is the specified vector for the \( i \)-th matrix, \( \{a_{ji} : j = 1, 2, \ldots, k+1\} \) are the coefficients for the variables in the \( i \)-th matrix, and the \( R \) variables in expression (6.3) can be any variables that have been read into the program. The user provides both the coefficient and the \( R \) variable for every entry in the \( m \) vectors. A particular variable can appear in more than one place in a vector and in more than one vector.

A. Model With an Error in the Equation

If the model with an error in the equation (called the EV1 model) is specified, the program uses the submatrices \( \hat{\Sigma}_{u u} \) and \( \hat{\Sigma}_{u w} \) from \( \hat{\Sigma}_{a a} \) of (6.3). Let \( \hat{\lambda} \) be the smallest root of

\[ |Z'WZ - \lambda \hat{\Sigma}_{a a}^*| = 0 \]  \hspace{1cm} (6.4)  

where

\[ Z'WZ = (\Psi, X)'W(\Psi, X) \]

\[ \text{vech } \hat{\Sigma}_{a a}^* = (\hat{\Sigma}_{ww}^*, \text{vec } \hat{\Sigma}_{uw}^*, \text{vech } \hat{\Sigma}_{uu}^*) \]

\[ \hat{\Sigma}_{ww}^* = \hat{\Sigma}_{ww} + \hat{\Sigma}_{uu} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{uw} \]

\[ \hat{\alpha} = \sum_{i=1}^{L} \sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} \hat{W}_{ij} = \sum_{i=1}^{L} \sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} \hat{W}_{ij}^* \]
$X'WX$ and $X'WY$ are defined in (6.1) and (6.2), and $\Sigma_{uu}^+$ is the Moore-Penrose generalized inverse of $\Sigma_{uu}$. 

The estimated parameter vector is

$$\hat{\beta}_{EV1} = M_{xx}^{-1} \hat{M}_{xy}$$

(6.5)

where for $\lambda > 1 - n^{-1}(k - 2)\alpha$

$$\hat{M}_{xx} = X'WX - n^{-1}[\lambda - \alpha(n - 1)] \hat{\Sigma}_{uu}$$

$$\hat{M}_{xy} = X'WY - n^{-1}[\lambda - \alpha(n - 1)] \hat{\Sigma}_{uu}$$

(6.6)

while for $\lambda < 1 - n^{-1}(k - 2)\alpha$

$$\hat{M}_{zz} = Z'WZ - [\lambda - \alpha(n - 1)] \hat{\Sigma}_{aa}$$

(6.7)

The parameter $\alpha$ can be zero or one and is specified by the user. Note that when the root $\lambda$ of the determinantal equation is less than one (minus some modifying constant if $\alpha = 1$), then the root is used in defining the estimator. This use of the root takes care of the situation when the estimators fall on the boundary of the parameter space, i.e., when the estimate of the error-in-equation variance $\sigma_{qq}$ is zero.

EV CARP offers the user the option of $\alpha = 0$ or $\alpha = 1$ in (6.6)-(6.7) above. For simple random samples ($W_{ij} = 1$) and $\alpha = 0$, this
estimator is that of Fuller (1984) given in Section 1.B.2 by equation (1.5). The \( \alpha = 1 \) modification improves the moment properties of the estimator to give finite means and variances. This modification is discussed in Fuller (1987, Section 2.5).

For the EV1 option, the estimated covariance matrix (obtained using a Taylor approximation) is computed as

\[
\hat{\Sigma}_{\text{EV1}} = c^{-2} \hat{\Sigma}_{xx} \hat{\Sigma}_{xx}^{-1}
\]  

(6.8)

where

\[
\hat{\Sigma}_{\text{EV1}} = \sum_{i=1}^{L} T_i \sum_{j=1}^{n_i} (\hat{d}_{ij} - \hat{d}_{i..})'(\hat{d}_{ij} - \hat{d}_{i..})
\]

\[
\hat{d}_{ij} = [x_t'v_t + \sum_{u=1}^{\beta} \sum_{u} \hat{\Sigma}_{u^2} - \sum_{u} \hat{\Sigma}_{u^2}] \hat{w}_t
\]

and \( T_i = [(n - k)(n_i - 1)]^{-1}(n - 1)n_i \). Except for sampling adjustments and the \( \alpha \) modification, equations (6.4)-(6.8) are those of Theorem 2.1 with \( \pi = \hat{w}_t \).

Note that we can estimate the variance of the error in the equation by

\[
\hat{\sigma}_{qq} = (1, \hat{z}')[c^{-1}Z'WZ - \sum_{i=1}^{\alpha} \hat{\Sigma}_{ii}](1, \hat{z}')'
\]

(6.9)

The program estimate of \( \sigma_{qq} \) is the maximum of the value given by (6.9) and zero.
B. Model With No Error in the Equation

If the model with no error in the equation (called the EV3 model) is selected, the program uses the entire covariance matrix \( \Sigma_{\text{a,m}} \) of (6.3) that has been entered into the program. The estimator of \( \beta \) is

\[
\hat{\beta}_{\text{EV3}} = \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy}
\]

(6.10)

where

\[
\hat{\Sigma}_{zz} = Z'WZ - \hat{\lambda}(1 - \omega^{-1})\hat{\Sigma}_{a,m}
\]

and \( \hat{\lambda} \) is the smallest root of the determinantal equation of (6.4) except that now the first element of \( \hat{\Sigma}_{a,m} \) is the value entered into the program by the user. In the EV3 case, the root is always used in defining the estimator. Note that (6.10) is nearly the same as (6.7) of the EV1 case used when \( \hat{\sigma}_{qq} = 0 \). Thus, if the root calculated in EV3 is much larger than one, this indicates that perhaps there is an error in the equation and EV1 should be used.

The estimated covariance matrix of \( \hat{\beta}_{\text{EV3}} \) is computed as

\[
V(\hat{\beta}_{\text{EV3}}) = \hat{\Sigma}_{xx}^{-2}\hat{\Sigma}_{xy}\hat{\Sigma}_{xx}^{-1}
\]

(6.11)

where

\[
\hat{G}_{\text{EV3}} = \sum_{i=1}^{L} \zeta_i \sum_{j=1}^{n_i} (\hat{d}_{ij} - \hat{d}_{i..})(\hat{d}_{ij} - \hat{d}_{i..})'
\]

\[
\hat{\Sigma}_{a,m} = \hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xy}\hat{\Sigma}_{xx}^{-1}
\]
\[
\hat{d}_{ij} = d_i = [x_t v_t - x_{uvtt} - \hat{v}^{-1} (v_{tt} - \hat{v}_{vtt} \hat{x}_{uvtt})] w_t
\]

\[
v_t = y_t - x_t \hat{\beta}
\]

\[
\hat{\sigma}_{vtt} = \lambda(1, -\hat{\beta}') \hat{Z}_{maaa}(1, -\hat{\beta}')
\]

\[
\hat{\sigma}_{vvtt} = \lambda(1, -\hat{\beta}') \hat{Z}_{maaa}(1, -\hat{\beta}')
\]

\[
\hat{Z}_{uvtt} = \lambda \hat{Z}_{maaa}(1, -\hat{\beta}')
\]

and

\[
\hat{Z}_{uvtt} = \lambda \hat{Z}_{maaa}(1, -\hat{\beta}')
\]

These equations are nearly the same as those of Theorem 2.2 with

\[
\pi_t = \hat{W}_t.
\]

C. Estimated True Values and Standardized Residuals

If requested, the standardized residuals from the regression and the estimated true values of the variables will be output. The estimates are based on the model in which the true \( x_t \) are fixed. Standardized residuals are used because the error variance matrix depends on the observation and, therefore, so does the variance of the residual \( v_t \).

The standardized residual and the vector of estimated true values output by the program are given by
\[ \hat{v}_t = \sigma_{v_{vtt}} \hat{v}_t \]

and

\[ \hat{z}_t = z_t - \hat{v}_t \sigma_{v_{vtt}} \hat{z}_{vatt} \]

where

\[ \hat{v}_t = z_t (1, -\hat{\beta}')' \]

\[ \hat{\Sigma}_{vatt} = (1, -\hat{\beta}') \hat{\Sigma}_{aatt} \]

\[ \hat{\sigma}_{v_{vtt}} = (1, -\hat{\beta}') \hat{\Sigma}_{e_{vtt}} (1, -\hat{\beta}')' \]

\[ \hat{\Sigma}_{e_{vtt}} = \begin{cases} \text{diag}(\sigma_{q_q}, 0, 0, \ldots, 0) + \hat{\Sigma}_{aatt} & \text{if EV1} \\ \lambda \hat{\Sigma}_{aatt} & \text{if EV3} \end{cases} \]

\[ \hat{\Sigma}_{aatt} = \sum_{t=1}^{m} \hat{\psi}_i' \hat{\psi}_i \]

\[ \hat{\sigma}_{q_q} = c^{-1} \sum_{t=1}^{n} \hat{v}_t \hat{v}_t' - (1, -\hat{\beta}') \hat{\Sigma}_{a_{ma}} (1, -\hat{\beta}')' . \]

Also output by the program are the estimated standard error of \( \hat{v}_t \), which is \( \hat{\sigma}_{v_{vtt}}^{1/2} \), and the estimated standard errors of the estimated true values which are calculated as the square roots of the diagonal elements of

\[ \hat{\Sigma}_{aatt} - \hat{\Sigma}_{vtt} \sigma_{v_{vtt}} \hat{\Sigma}_{vatt} . \]
The equations of (6.12) and (6.13) follow those of Fuller (1987, Section 2.2.2) where adjustments have been made to account for the heterogeneous measurement error variances.

D. Second Round Estimation - Weighting by Estimated Variances

When appropriate, the user may request a second round of estimation. For first round estimation, an estimate of the parameters is obtained using the sample weights that have been read into the program by the user. For second round estimation, the program uses the first round estimates to construct the weights to be used in the second round. The second round weights are the inverse of the $\hat{\sigma}_{\text{vvt}}$ of (6.12) above. To obtain second round estimates, the program replaces $\hat{W}_t$ by $\hat{\sigma}_{\text{vvt}}^{-1}$ in the appropriate equations for the EV1 or EV3 model. Note that because the second round weights are functions of the estimators of $\sum_{a\text{att}}$, two round estimation, as performed in EV CARP, is appropriate only when the $\psi_t$ are independent of the variables in the regression.

For the model with an error in the equation, the second round estimator is

$$\hat{\beta}_{\text{EV1}} = \hat{\mathbf{M}}^{-1} \hat{\mathbf{M}}_{\pi \pi} \hat{\mathbf{M}}_{\tau \tau}$$

(6.14)

where, if $\hat{\lambda} > 1 - n^{-1}(k - 2)\alpha$

$$\hat{\mathbf{M}}_{\pi \pi} = \mathbf{M}_{\pi \pi} - n^{-1} [n - \alpha(k - 1)] \hat{\mathbf{Z}}_{\pi \tau} \hat{\mathbf{Z}}_{\tau \pi} \hat{\mathbf{Z}}_{\pi \tau}$$
and otherwise

\[ \hat{M}_{Z\pi Z} = M_{Z\pi Z} - \left( \hat{\lambda} - \alpha(n - 1)^{-1} \right) \sum_{a=1}^{n} \hat{m}_a. \]

and

\[ M_{Z\pi Z} = \left( \sum_{t=1}^{n} \hat{\sigma}_{vtt}^{-1} \right)^{-1} \sum_{t=1}^{n} \hat{\sigma}_{vtt}^{-1} Z_t^t Z_t^t, \]

\[ \hat{\Sigma}_{a=1, \ldots} = \left( \sum_{t=1}^{n} \hat{\sigma}_{vtt}^{-1} \right)^{-1} \sum_{t=1}^{n} \hat{\sigma}_{vtt}^{-1} \hat{a} tt, \]

\[ \hat{\sigma}_{vtt} = \hat{\sigma}_{qq} + (1, -\hat{g}_{EV1}) \hat{a} tt (1, -\hat{g}_{EV1})', \]

(\hat{\sigma}_{qq}, \hat{g}_{EV1}) is the vector of first round estimates of (\sigma_{qq}, g') and \hat{\lambda} is the smallest root of

\[ |M_{Z\pi Z} - \hat{\lambda} \sum_{a=1}^{n} | = 0. \]

A second round estimator of \sigma_{qq} is the maximum of zero and

\[ \tilde{\sigma}_{qq} = (1, -\hat{g}_t')(M_{Z\pi Z} - \hat{\Sigma}_{a=1, \ldots})(1, -\hat{g}_t')'. \]

(6.15)

The estimated variance of the second round estimate \hat{g}_{EV1} is computed using equation (6.8) with the second round estimates so that (\hat{\mu}_t, \hat{g}_{EV1}) is replaced by (\hat{\sigma}_{vtt}^{-1}, \hat{g}_{EV1}).

For the model with no error in the equation, the second round estimator of \hat{g} is computed by
\[ \hat{\beta}^{-1}_{\text{EV3}} = \hat{X}^{-1}_{\text{EV3}} \]

where

\[ \hat{M}_{Z \pi Z} = M_{Z \pi Z} - \hat{\lambda}(1 - \alpha n^{-1}) \hat{\Sigma} \]

and \( \hat{\lambda} \) is the smallest root of the determinantal equation

\[ |M_{Z \pi Z} - \lambda \hat{\Sigma}| = 0. \]

The variance is computed as

\[ \hat{\mathcal{V}}(\hat{\beta}^{-1}_{\text{EV3}}) = \hat{M}^{-1}_{x \pi x} \hat{G}_{\text{EV3}} \hat{M}^{-1}_{x \pi x} \]

where \( \hat{G}_{\text{EV3}} \) is computed by the equations of (6.11) except that the old estimates and weights are replaced by the new ones.

For standardized residuals and estimated true values, second round estimates are computed from equations (6.12) replacing \( (\hat{\beta}', \hat{\sigma}_{qq}) \) by \( (\hat{\beta}', \hat{\sigma}_{qq}) \). If two round estimation is chosen, the program outputs the first round estimates of \( \hat{\beta} \) but does not output the covariance matrix of the first round estimates. The second round statistics for all requested options, including the covariance matrix of \( \hat{\beta} \), are output.

Better second round weights may be constructed outside the program. For example, for the hog data of Chapter IV, we were able to construct better weights using a non-linear variance function that depended on the true but unknown \( x_t \). Such second round weights may be
incorporated into EV CARP by reading the weights into the program as if they were "regular" first round weights and then requesting a single round of estimation.
VII. REFERENCES


Madansky, A. 1959. The fitting of straight lines when both variables are subject to error. JASA 54:173-205.


VIII. ACKNOWLEDGEMENTS

Alhamdulillah, thanks be to God. I also wish to thank Dr. Wayne A. Fuller for his help and guidance in completing this thesis. Without his help, I would never have finished. I thank the other members of my committee, Yasuo Amemiya, Noel A. C. Cressie, David A. Harville, and Richard J. Tondra, for their useful comments. I also wish to thank Christine Olson for her repeated typings and my two children, Mostafa and Fatima, for being cooperative enough to allow me to finish this research after their births.
IX. APPENDIX A: CONVERGENCE IN PROBABILITY
AND THE WEAK LAW OF LARGE NUMBERS

In this appendix we briefly review the definitions of convergence in probability and orders in probability. We also present a general form of the weak law of large numbers found in Chung (1974, p. 111).

**Definition 9.1.** The sequence of random variables \( \{X_n\} \) converges in probability if for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} P(\{|X_n - X| > \varepsilon\}) = 0.
\]

In the case, we write \( \text{plim } X_n = X \) or \( X_n \xrightarrow{P} X \).

**Definition 9.2.** We say \( X_n \) is of smaller order in probability than \( g_n \) if \( \text{plim } g_n^{-1} X_n = 0 \) and we write \( X_n = o_p(g_n) \).

**Definition 9.3.** We say \( X_n \) is at most of order in probability \( g_n \) if for every \( \varepsilon > 0 \), there exists a positive real number \( M_\varepsilon \) such that \( P(\{|X_n| > M_\varepsilon g_n\}) < \varepsilon \) for all \( n \), and we write \( X_n = O_p(g_n) \).

In addition to the above definitions, the following lemma and properties of \( o_p(\cdot) \) and \( O_p(\cdot) \) are quite useful. A proof of some of these properties is given in Fuller (1976, pp. 184-185).

**Lemma 9.1.** Let \( \{g_n\} \) and \( \{h_n\} \) be sequences of positive real numbers where \( h_n^{-1} g_n \to 0 \). If \( X_n = O_p(g_n) \), then \( X_n = o_p(h_n) \).
Proof. Given $\varepsilon > 0$, there exists an $M_{\varepsilon}$ such that

\[ P\{|X_n| > M_{\varepsilon}g_n\} < \varepsilon \text{ for all } n, \text{ or instead } P\{|X_n| > M_{\varepsilon}g_n^{-1}h_n\} < \varepsilon \]

for all $n$. However, since $h_n^{-1}g_n \to 0$ as $n \to \infty$, given $\delta > 0$ we can choose a positive integer $N$ such that $\delta > M_{\varepsilon}g_n^{-1}h_n$ for $n > N$. Then, $P\{|X_n| > \delta h_n\} < \varepsilon$ for every $n > N$, i.e., $X_n = o_p(h_n)$.

Properties of $o_p(\cdot)$ and $O_p(\cdot)$

Let \{${g_n}$\} and \{${h_n}$\} be sequences of positive real numbers, and let \{${X_n}$\} and \{${Y_n}$\} be sequences of random variables.

(i) If $X_n = o_p(g_n)$ and $Y_n = o_p(h_n)$, then

\[ X_n Y_n = o_p(g_n h_n) \]

\[ |X_n|^s = o_p(g_n^s) \text{ for } s > 0 \]

\[ X_n + Y_n = o_p(\max\{g_n, h_n\}) \]

(ii) If $X_n = O_p(g_n)$ and $Y_n = O_p(h_n)$, then

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

\[ |X_n|^s = O_p(g_n^s) \text{ for } s > 0 \]

\[ X_n + Y_n = O_p(\max\{g_n, h_n\}) \]
(iii) If $X_n = o_p(g_n)$ and $Y_n = o_p(h_n)$, then

$$X_n Y_n = o_p(g_n h_n).$$

**Lemma 9.2.** Let $\{X_n\}$ be a sequence of random variables and $\{g_n\}$ a sequence of positive real numbers. If $E(X_n^2) = O(b_n^2)$, then $X_n = O_p(b_n)$.

**Proof.** By definition, there exists $M > 0$ such that

$$E(X_n^2) < M b_n^2.$$ 

Now, by Chebyshev's inequality, for any $K > 0$,

$$P(|X_n| > K b_n) < \frac{K^{-2} b_n^{-2} E(X_n^2)}{K^{-2} M^2}.$$

Thus, given $\epsilon > 0$ choose $K$ such that $K^2 > \epsilon^{-1} M^2$.

Finally, we present a general form of the weak law of large numbers. The proof may be found in Chung (1974, p.111).

**Chung's Theorem**

Let $\{X_n\}$ be a sequence of independent random variables with distribution functions $\{F_n\}$ and let $S_n = \sum_{j=1}^n X_j$. Let $\{b_n\}$ be a
sequence of real numbers increasing to $+\infty$. Suppose

\[
(i) \quad \sum_{j=1}^{n} \int_{|x|>b_n} dF_j(x) = o(1)
\]

and

\[
(ii) \quad b_n^{-2} \sum_{j=1}^{n} \int_{|x|<b_n} x^2 dF_j(x) = o(1).
\]

Then,

\[
b_n^{-1} (S_n - a_n) \xrightarrow{P} 0,
\]

where

\[
a_n = \sum_{j=1}^{n} \int_{|x|<b_n} x dF_j(x).
\]

We note that if $a_n - E(S_n) \xrightarrow{P} 0$, then we may replace $a_n$ by $E(S_n)$ above.
X. APPENDIX B: CENTRAL LIMIT THEOREMS

In this appendix, we state the definitions of weak convergence in law, and we list some central limit theorems.

**Definition 10.1.** Let \( \{F_n\} \) be a sequence of distribution functions and \( F \) a distribution function. We say \( \{F_n\} \) converges weakly to \( F \) if and only if \( F_n(x) \rightarrow F(x) \) at all \( x \) for which \( F(\cdot) \) is continuous.

**Definition 10.2.** Let \( \{X_n\} \) be a sequence of random variables, \( \{F_n\} \) the corresponding distribution functions, and \( F \) a distribution function. Then \( \{X_n\} \) is said to converge in distribution to \( F \) if and only if \( \{F_n\} \) converges weakly to \( F \).

If \( X \) is a random variable that has the distribution function \( F \), then we also say that \( \{X_n\} \) converges in distribution to \( X \), and we write \( X_n \xrightarrow{L} X \). We now state various central limit theorems.

**Liapounov Central Limit Theorem**

Let \( \{X_t\} \) be a sequence of independent random variables with distribution functions \( \{F_t\} \). Let \( E(X_t) = \mu_t \), \( E((X_t - \mu_t)^2) = \sigma_t^2 \), and \( V_n = \sum_{t=1}^{n} \sigma_t^2 \). If for some \( n > 0 \)

\[
\lim_{n \rightarrow \infty} \frac{1}{n^{(1+n/2)}} \sum_{t=1}^{n} E|X_t - \mu_t|^{2+n} = 0
\]

then, as \( n \rightarrow \infty \),
Proof. See advanced probability texts.

Multivariate Extension of Central Limit Theorems

Let \{Z_t\} be a sequence of k-dimensional random variables with distribution function \{F_{Z_t}(z)\}. Let \(X_t\) be the distribution function of \(X_t = \lambda'Z_t\) where \(\lambda\) is a fixed non-zero vector. A necessary and sufficient condition for \(F_{Z_t}(z)\) to converge to the k-variate distribution function \(F_z(z)\) is that \(F_{X_t}(x)\) converge to a limit for each \(\lambda\).


In the theorem above, if each \(F_{X_t}(x)\) converges to a normal distribution function, then the random vector \(Z_n\) will converge in distribution to a multivariate normal.

We now list two theorems that are stated and proved in Fuller (1987, Appendix 1.C).

Theorem 10.1. Let \(\{\epsilon_t\}\) be a sequence of independently distributed \((0, \Sigma_{eett})\) p-dimensional random row vectors with uniformly bounded \(2 + \delta (\delta > 0)\) moments. Let \(\{c_t\}\) be a sequence of fixed p-dimensional row vectors with \(c_i c'_i \neq 0\). Let \(n^{-1} V_n\) be bounded above and below by positive real numbers for all \(n\) where \(V_n = \epsilon_{t=1}^n c_t^t \Sigma_{eett} c_t\) and let the elements of \(M_{cc} = n^{-1} \epsilon_{t=1}^n c_t^t c_t\) be bounded for all \(n\). Then,

\[
v_n^{-1/2} \sum_{t=1}^n (X_t - \mu_t) \xrightarrow{L} N(0, 1).
\]
Theorem 10.2. Let $Z_t = z_t + \varepsilon_t$ where the $\varepsilon_t$ are independent $p$-dimensional random row vectors with zero means, positive definite covariance matrices $\Sigma_{\varepsilon t}$, and bounded $4 + \delta (\delta > 0)$ moments. Let $z_t$ be a fixed sequence and $\bar{z} = n^{-1} \sum_{t=1}^{n} z_t$,

$$m_{zz} = (n - 1)^{-1} \sum_{t=1}^{n} (z_t - \bar{z})'(z_t - \bar{z})$$

with

$$\lim_{n \to \infty} \bar{z} = \mu_z, \quad \lim_{n \to \infty} m_{zz} = \bar{m}_{zz}, \quad \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \Sigma_{\varepsilon t} = \Sigma_{\varepsilon}$$

where $\Sigma_{\varepsilon}$ is positive definite. Let

$$\hat{\theta} = \left[ \bar{Z}, (\text{vech } m_{ZZ})' \right]'$$

$$\hat{\theta}_n = \left[ \bar{z}, \text{vech}(m_{zz} + \Sigma_{\varepsilon})' \right]'$$

Then, $G_n^{-1/2} (\hat{\theta} - \theta_n) \xrightarrow{L} N(0, I)$ where $G_n = V(\hat{\theta})$. 