Measurement error models for time series

John Lamont Eltinge
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/8639

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" x 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" x 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.
Measurement error models for time series

Eltinge, John Lamont, Ph.D.
Iowa State University, 1987

Copyright ©1987 by Eltinge, John Lamont. All rights reserved.
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
Measurement error models for time series

by

John Lamont Eltinge

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

Copyright © John Lamont Eltinge, 1987. All rights reserved.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE AND NOTATION FOR MEASUREMENT ERROR MODELS AND TIME SERIES MODELS</td>
<td>10</td>
</tr>
<tr>
<td>2.1. Notation and Some Previous Results for Measurement Error Models</td>
<td>11</td>
</tr>
<tr>
<td>2.2. Notation and Models for Time Series</td>
<td>22</td>
</tr>
<tr>
<td>2.3. Identification Issues for Measurement Error Models with Serially Correlated Errors or True Values</td>
<td>28</td>
</tr>
<tr>
<td>2.4. Previous Work in Time Domain Estimation for Measurement Error Models</td>
<td>36</td>
</tr>
<tr>
<td>2.4.1. Previous work in time domain estimation for structural models and factor analysis models</td>
<td>37</td>
</tr>
<tr>
<td>2.4.2. Previous work in time domain estimation for general &quot;signal plus noise&quot; models</td>
<td>44</td>
</tr>
<tr>
<td>2.5. Previous Work in Frequency Domain Estimation for Measurement Error Models</td>
<td>53</td>
</tr>
<tr>
<td>3. ASYMPTOTIC PROPERTIES OF SUMS OF LINEAR AND BILINEAR FUNCTIONS OF CORRELATED OBSERVATIONS</td>
<td>63</td>
</tr>
<tr>
<td>3.1. Definitions and Notation for Vector Martingales</td>
<td>64</td>
</tr>
<tr>
<td>3.2. Asymptotic Properties of the Sum of a Sample Mean and a Sample Variance</td>
<td>73</td>
</tr>
<tr>
<td>3.3. Asymptotic Properties of the Sum of a Linear Function and an Unweighted Bilinear Function</td>
<td>104</td>
</tr>
<tr>
<td>3.4. Asymptotic Properties of the Sum of a Linear Function and a Weighted Bilinear Function</td>
<td>156</td>
</tr>
<tr>
<td>4. ESTIMATION BY THE METHOD OF MOMENTS</td>
<td>243</td>
</tr>
<tr>
<td>4.1. Unweighted Estimators</td>
<td>244</td>
</tr>
<tr>
<td>4.2. Weighted Estimators</td>
<td>264</td>
</tr>
</tbody>
</table>
### 11. APPENDIX C. MAXIMUM LIKELIHOOD ESTIMATION FOR MEASUREMENT ERROR MODELS WITH COMBINED TIME SERIES AND CROSS SECTIONAL DATA

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1. N Independent Cross-Sectional Units</td>
<td>457</td>
</tr>
<tr>
<td>11.2. Nested Inputs and Errors</td>
<td>460</td>
</tr>
<tr>
<td>11.3. Multi-Stage Nested Inputs and Errors</td>
<td>472</td>
</tr>
<tr>
<td>11.4. Crossed Inputs and Errors</td>
<td>481</td>
</tr>
</tbody>
</table>
1

1. INTRODUCTION

A common multiple regression model may be written in the form,

\[ y_t = x_t \beta + q_t, \quad t=1, 2, \ldots, T \]  \hspace{1cm} (1.1)

where \( y_t \) is a \( 1 \times r \) vector of "dependent variables," \( x_t \) is a \( 1 \times k \) vector of "independent variables," \( \beta \) is a fixed \( k \times r \) matrix of regression coefficients, and \( q_t \) is a \( 1 \times r \) "error in equation" term associated with the deviation of the vector \( y_t \) from the linear function \( x_t \beta \). There are many practical situations in which neither \( y_t \) nor \( x_t \) may be observed directly; instead, one may record observations \( Z_t = (Y_t, X_t) \), \( t=1, 2, \ldots, T \), where

\[ Y_t = y_t + w_t, \]  \hspace{1cm} (1.2.a)

\[ X_t = x_t + u_t, \]  \hspace{1cm} (1.2.b)

and \( a_t = (w_t, u_t) \) represents a vector of "measurement errors." Model (1.1)-(1.2) is often called a "measurement error model" or an "errors-in-variables model."

Given the observations \( \{(Y_t, X_t), t=1, 2, \ldots, T\} \) and the measurement error model (1.1)-(1.2), one may wish to estimate the regression coefficient \( \beta \) as well as parameters associated with the sequences \( \{x_t\} \), \( \{q_t\} \) and \( \{a_t\} \). In addition, one may wish to
obtain a predictor of \((y_t, x_t)\) that has smaller mean squared prediction error than the observation \((\hat{y}_t, \hat{x}_t)\). The fundamental difficulty with such estimation and prediction problems is that the application of standard regression methods to model (1.1)-(1.2) leads to unsatisfactory results. In particular, ordinary least squares regression of \(Y_t\) on \(X_t\) leads to a biased estimator of \(\beta\). Section 2.1 will review the errors-in-variables literature in greater detail, but for the moment it suffices to note that much previous work with measurement error models has developed estimation and prediction procedures under the assumption that the observations \(Z_t = (\hat{Y}_t, \hat{X}_t)\) are mutually uncorrelated. For the discussion below, denote such models as "uncorrelated errors-in-variables models."

In some practical cases, the observations \(Z_t\) may be taken in sequence, so that there may be serial correlations in either the "true value" sequence \(z_t = (y_t, x_t)\) or the measurement error sequence \(e_t = (w_t, u_t)\); denote the resulting form of model (1.1)-(1.2) as a "correlated errors-in-variables model." The statistical literature has noted the presence of serially correlated measurement errors in a number of widely studied social and economic data series. Morgenstern (1949) discussed several sources of such correlated errors, and devoted particular attention to correlated sampling errors that are present in rotation survey data. Pierce (1981) described five major sources of error in economic data: "conceptual error," due to the lack of precise definitions of the "true" phenomena of interest; "transitory error," a special type of conceptual error that arises because certain phenomena
exist in one or more time periods but are deemed to be non-representative of the "true" system of interest; "sampling error," which will in general be serially correlated for overlapping surveys; "seasonal adjustment error," attributable to imperfect seasonal adjustment, which leads inherently to serial correlations in the resulting errors; and "reporting error," any discrepancy between an individual respondent's "true" or "correct" response and the final recorded response. In this final category, Bailar (1986) mentioned serially correlated components due to "recall error" and "time in sample error."

For the present work, one may consider each of Pierce's five error types to be forms of "measurement error," because each potentially constitutes a component of the difference between an observed variable and the corresponding "true" or "latent" variable. Further, it is apparent from the above descriptions that errors in any of these five categories may exhibit serial correlation. For a given data series, a practical point of interest is to identify the predominant sources of measurement error and to establish whether the dominant errors exhibit pronounced patterns of serial correlation. To date, the bulk of the literature on such errors has addressed sampling errors that arise from overlapping surveys. For example, Hausman and Watson (1985) considered measurement errors in unemployment data gathered through the Current Population Survey. They attributed this error to sampling error and, due to the rotation design of the Current Population Survey, they modeled the error as a moving average process. They concluded, "the
white-noise measurement error process may be a very poor approximation to the process generating measurement error in many economic time series." Similar reasoning led Miazaki (1985) to use a moving average model for sampling errors in data from the National Crime Survey. Of course, nonsampling errors associated with "recall error," "time in sample error," interviewer effects, and data processing effects need not follow a pure moving average process. If such nonsampling errors contribute a substantial component to the total measurement error in a socioeconomic series, this total measurement error also may not follow a pure moving average process. Consequently, the analysis of socioeconomic data series may require fairly general models for serially correlated measurement error terms.

The discussion above indicates that some economic and social data may be described by forms of the measurement error model (1.1)-(1.2) in which the measurement errors are serially correlated. Industrial process control provides a second area of application for correlated measurement error models. Several authors have noted the tendency of some measuring instruments to "drift" or "wander" over time in a serially correlated, nondeterministic pattern. For example, Bhuyan (1985) mentioned serial correlation as one of several aspects of measurement error problems in quality control work. Also, Taguchi (1986) discussed a control charting problem in which both "true values" and measurement errors were serially correlated. If such measurement errors arise in a single-variable process control system, one may describe the resulting observations with a univariate "signal plus noise" model discussed in
Sections 2.4 and 5.2 below. In some cases, however, consideration of several product or process characteristics may lead to more complicated modeling and control methods. Examples of such methods include inferential process control and Kalman filtering.

Stephanopoulos (1984, pp. 438-447) described "inferential process control" as an attempt to control a vector $\mathbf{x}_t$ of characteristics which are difficult or impossible to measure precisely during production, e.g., chemical composition of a product. Control actions are therefore determined by a vector $\mathbf{y}_t$ of observations on "secondary characteristics" such as temperature. Given a locally linear relationship between $\mathbf{y}_t$ and $\mathbf{x}_t$ described by model (1.1)-(1.2.a), linear least squares prediction of $\mathbf{x}_t$ is a relatively simple problem, provided one knows $\beta$ and the parameters of the $\{x_t\}$, $\{q_t\}$ and $\{w_t\}$ processes. Stephanopoulos (1984, p. 440) notes that in practice, such parameters may be unknown, so that parameter estimation is a necessary first step in inferential process control. Given a set of imperfect observations $\mathbf{X}_t$ on the variables $\mathbf{x}_t$, parameter estimation for inferential process control is a special case of estimation for model (1.1)-(1.2).

Kalman filter methods in process control are also closely related to model (1.1)-(1.2). Chapter 5 will give a detailed discussion of Kalman filters, but for the present it suffices to state that Kalman filtering is a method of linear least squares prediction based on a "state-space model" of the observation process; and that estimation of $\beta$ in model (1.1)-(1.2) constitutes a special case of the estimation of the coefficient matrix of the measurement equation of a state-space
model. To date, many applications of Kalman filter methods to process control [c.f. Phadke (1981), Maybeck (1982, pp. 70-71) and Crowder (1986)] have assumed that the coefficient matrix of the measurement equation is known a priori. Extension of these Kalman filter methods to state-space models with unknown measurement equation coefficients requires estimation of these coefficients from either experimental data [c.f. Goodwin and Payne (1977)] or observational data. In the latter case, the estimation problem may then be equivalent to estimation for model (1.1)-(1.2).

The preceding discussion indicates that errors-in-variables models for serially correlated observations have a number of applications in the social sciences and engineering. Consequently, it is desirable to develop efficient estimation procedures for such models. The remaining chapters of this dissertation contribute to this goal according to the following outline.

First, Chapter 2 provides a review of some notation and some previous literature associated with measurement error models and time series models, including some elementary definitions and results for uncorrelated errors-in-variables models; some elementary definitions for time series models; a brief discussion of identification issues for correlated measurement error models; and a review of some previous work in time-domain and frequency-domain estimation for correlated measurement error models.

Second, the asymptotic properties of method-of-moments estimators for model (1.1)-(1.2) are investigated. The large-sample behavior of
errors-in-variables estimators is generally dependent on the large-sample behavior of the sum of a linear function and a bilinear function of some random components of the associated model. Therefore, Chapter 3 develops some results on almost sure convergence and asymptotic normality of the sum of a linear function and a bilinear function of a sequence of serially correlated random vectors. The results for unweighted functions are slight extensions of similar results in Fuller, Hasza and Goebel (1981) and Hannan (1970, Chapters IV and VII), while the results for weighted functions require some additional work.

Chapter 4 uses the results of Chapter 3 to derive the asymptotic properties of some standard unweighted method-of-moments estimators of the regression coefficient \( \hat{\beta} \) in model (1.1)-(1.2). The results of Chapter 4 indicate that in the presence of serially correlated errors which satisfy mild regularity conditions, standard errors-in-variables estimators are consistent and asymptotically normal. However, the covariance matrices of the resulting asymptotic distributions reflect the serial covariance structure of the errors, so reported standard errors must be adjusted accordingly. Given a set of weight matrices satisfying certain asymptotic conditions, large-sample properties of weighted estimators of \( \hat{\beta} \) are derived in Chapter 4.

Third, Chapters 5 and 6 address maximum likelihood approaches to the structural (random \( x_t \)) and functional (fixed \( x_t \)) forms, respectively, of model (1.1)-(1.2). In each case, general formulas for the matrices of first and second derivatives of the normal log-likelihood function are presented. The cumbersome and high-dimensional
nature of these matrices suggests the pursuit of alternative modeling approaches. In the structural case, two modeling approaches are considered. First, autoregressive moving average models for multivariate "signal plus noise" processes are studied through a multivariate extension of a previously known univariate result. This approach is theoretically appealing, but it has a number of practical limitations. Second, a state-space representation of the structural form of model (1.1)-(1.2) is developed. This representation leads to iterative low-dimensional computational formulas for the matrices of the first and second derivatives of the normal structural likelihood function. Use of these formulas in a Newton-Raphson procedure for maximum likelihood estimation is considered. In Chapter 6, similar state-space arguments lead to iterative formulas for derivative matrices, and a Newton-Raphson approach to maximum likelihood estimation is suggested.

The results presented in this dissertation are by no means definitive. Practical applications of the methods proposed here require additional research in the following areas. First, one may study additional forms of method-of-moments estimation. Asymptotic results similar to Theorem 4.1 could be studied for instrumental-variable estimators. An errors-in-variables extension of the Durbin method [c.f. Fuller (1976, p. 352)] may lead to method-of-moments estimators of autoregressive and moving-average parameters of the $x_t$ and $q_t$ processes. Small-sample properties of errors-in-variables estimators for serially correlated observations may be considered. Optimal
weighting procedures for the weighted estimators in Chapter 4 also
deserve further study.

Second, the maximum likelihood work of Chapters 5 and 6 cannot be
reduced to practice until there is a considerable amount of additional
work based on specific forms of identifying information and specific
autoregressive moving average models for the \( x_t \), \( q_t \), and \( a_t \)
processes. Questions remain regarding the topology of the likelihood
surfaces for the normal structural and functional forms of model (1.1)–
(1.2). Local and global maxima, boundary cases and concavity properties
are of particular interest. Also, the numerical maximum-likelihood
methods proposed in Chapters 5 and 6 may be modified to improve conver­
gence properties and computational efficiency. In addition, one may
investigate the asymptotic properties of maximum likelihood estimators
for the normal structural and functional forms of model (1.1)–(1.2).

Finally, simulations and data analyses may offer some guidance
regarding practical application of method-of-moments or maximum-
likelihood estimation to serially correlated forms of model (1.1)–(1.2).

In summary, the presence of serially correlated measurement errors
in some social, economic and engineering data leads one to consider
estimation of model (1.1)–(1.2) for correlated observations. The
present work contributes to an understanding and solution of this
problem, but many practical questions await further study.
Chapter 1 noted that in both the social sciences and engineering, if data are collected over time, then the resulting observations may contain serially correlated measurement errors. The remainder of this dissertation addresses parameter estimation in the presence of such errors.

To this end, the present chapter discusses some notation and literature associated with univariate and multivariate "signal plus noise" models. Section 2.1 reviews some previous work with measurement error models for uncorrelated observations; Chapters 4, 5 and 6 will extend some of these results to models with serially correlated observations. Section 2.2 introduces some notation and results required for a discussion of correlated observations. Section 2.3 discusses the identification status of some measurement error models under various assumptions about the availability of auxiliary information or about the covariance structure of the "true value" and "error" series. Finally, there has been a considerable amount of previous work devoted to the estimation of the parameters of various "signal plus noise" models for serially correlated observations; Section 2.4 discusses such results in time domain estimation, while Section 2.5 reviews some similar work in the frequency domain.
2.1. Notation and Some Previous Results for Measurement Error Models

This section reviews some definitions, notation and results developed previously in the literature of measurement error models for uncorrelated observations. This review will be limited to points most directly related to the estimators and results considered in Chapters 4, 5 and 6 below. For a more complete discussion of the literature for measurement error models, the reader may consult Amemiya (1982), Anderson (1984), Fuller (1987) and additional references given in these works. Unless noted otherwise, the definitions and notation presented in this section are borrowed from Fuller (1987).

A commonly studied measurement error model may be written

\[ Z_t = z_t + a_t, \quad t=1, 2, \ldots, T \]  

where \( Z_t = (y_t, x_t) \) is a \( 1 \times p \) vector of observed values, \( z_t = (y_t, x_t) \) is a \( 1 \times p \) vector of "true values," \( a_t = (w_t, u_t) \) is a \( 1 \times p \) vector of "measurement errors," \( y_t, x_t \) and \( w_t \) are \( 1 \times r \) subvectors, \( x_t \) and \( u_t \) are \( 1 \times k \) subvectors, and \( p = r + k \). The terms "measurement error model" and "errors-in-variables model" are sometimes reserved for models (2.1) in which the "true values" \( z_t = (y_t, x_t) \) satisfy the model

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

(2.2)
where $\beta$ is a $k \times r$ matrix of regression coefficients and $\{q_t\}$ is a sequence of $1 \times r$ "error in equation" vectors. Much of the errors-in-variables literature to date devotes principal attention to the estimation of the regression coefficient matrix $\beta$, and gives secondary attention to the estimation of parameters of the $\{x_t\}$ or $\{q_t\}$ sequences and to the estimation of the "true" vectors $\{z_t\}$.

Various subclasses of the errors-in-variables model arise through assumptions about the component vectors $\{a_t\}$, $\{x_t\}$ and $\{q_t\}$. As a general rule, the measurement errors $a_t$ are assumed to be a sequence of uncorrelated random vectors with mean $0_{1 \times p}$ and common $p \times p$ covariance matrix, say. In some cases, sampling or other considerations may imply that while $E(a_t) = 0_{1 \times p}$ for all $t$, $\Sigma_{a_t} = \text{Var}(a_t)$ is not constant with respect to $t$; one may call the result a "heteroscedastic errors-in-variables model."

Assumptions about the sequence $\{x_t\}$ allow one to define the "structural," "functional" and "ultrastructural" models. If one considers $\{x_t\}$ to be a sequence of random vectors with a common mean, then model (2.1)-(2.2) is called a "structural" errors-in-variables model; common assumptions for the structural model are that the random vectors $\{x_t\}$ are uncorrelated, have a common $1 \times k$ mean $\mu_x$, and have a common $k \times k$ covariance matrix $\Sigma_{xx}$. On the other hand, in some cases it may be reasonable to consider $\{x_t\}$ to be a fixed sequence of $1 \times k$ vectors; under this assumption, model (2.1)-(2.2) is called a "functional" errors-in-variables model. Dolby (1976) noted that one may extend the structural and functional models to an
"ultrastructural" model in which the $x_t$ are random vectors with possibly different means $\mu_{xt}$ and common covariance matrix $\Sigma_{xx}$. Note that if $\mu_{xt} = \mu_x$ for all $t$, then the ultrastructural model reduces to the structural model, while if $\Sigma_{xx} = 0_{k \times k}$, then $x_t = \mu_{xt}$ for all $t = 1, 2, \ldots, T$, and the ultrastructural model reduces to the functional model.

Finally, one may consider assumptions about the "error in equation" sequence $\{q_t\}$. Recall that $q_t$ is equal to the difference between the "true value" $y_t$ and the linear combination $x_t \beta$. Previous literature has generally assumed that $\{q_t\}$ is a sequence of uncorrelated random vectors which are independent of $\{x_t\}$ and $\{a_t\}$ and which have mean $0_{1 \times r}$ and common $r \times r$ covariance matrix $\Sigma_{qq}$. If $q_t = 0_{1 \times r}$ with probability one for all $t = 1, 2, \ldots, T$, then the resulting model is called the "model with no error in the equation."

Equivalently, the model with no error in the equation arises if one chooses not to distinguish between $w_t$, the measurement error in the observation $y_t$, and $q_t$, the error in the equation. In either case, one may define a "total error" vector at time $t$,

$$
\xi_t \equiv (e_t, u_t)
$$

$$
= a_t + (q_t, 0_{1 \times k})
$$

$$
= Z_t - x_t(\beta, I_r), \ t = 1, 2, \ldots, T,
$$
where \( e_t = w_t + q_t \). Then model (2.1)-(2.2) may be rewritten,

\[
Z_t = z_t + \varepsilon_t ,
\]

(2.3)

\[
y_t = x_t \beta ;
\]

(2.4)

and \( \varepsilon_t \) is generally assumed to be a sequence of uncorrelated \( p \)-dimensional random vectors with mean \( 0_{1 \times p} \) and common covariance matrix \( \Sigma_{\varepsilon \varepsilon} \). In the remainder of this work, the phrase, "model with an error in the equation" will apply to model (2.1)-(2.2) only, while the phrase, "model with no error in the equation" will apply to model (2.3)-(2.4) only. A practical distinction between these two models arises when multiple independent observations of the same \( z_t \) vector allow one to estimate \( \Sigma_{\varepsilon \varepsilon} \). Such an estimate yields direct estimates of 

\[
\Sigma_{uu} = \text{Var}(u_t), \quad \Sigma_{ww} = \text{Var}(w_t) \quad \text{and} \quad \Sigma_{uw} = \text{Cov}(u_t, w_t) = \text{Cov}(u_t, e_t)
\]

but not of \( \Sigma_{qq} \) or \( \Sigma_{ee} = \text{Var}(e_t) = \Sigma_{qq} + \Sigma_{ww} \). This fact may lead to slightly different estimation procedures for models (2.1)-(2.2) and (2.3)-(2.4), respectively.

It is sometimes useful to rewrite models (2.1)-(2.2) and (2.3)-(2.4) in general matrix form; the required definitions and notation are summarized below and are reviewed in greater detail in Appendix 4.A of Fuller (1987). Let \( \beta \) have \((i,j)\)-th element equal to \( \beta_{ij} \), \(i\)-th row equal to \( \beta_i \), and \(j\)-th column equal to \( \beta_j \), so that 

\[
\text{vec}(\beta) = (\beta_{1}', \beta_{2}', \ldots, \beta_{p}')'.
\]

Recall that
is the p-dimensional observation taken at time \( t \), and define

\[
Y_{i,t} = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})',
\]

\[
X_{j,t} = (X_{1j}, X_{2j}, \ldots, X_{Tj})',
\]

and

\[
Z = (Z_1', Z_2', \ldots, Z_T')',
\]

\[
= (Y, X)
\]

\[
= (Y_{,1}', Y_{,2}', \ldots, Y_{,r}', X_{,1}', X_{,2}', \ldots, X_{,k}').
\]

Note that \( Y_{,i} \) is a \( T \times 1 \) vector consisting of the \( i \)-th element of each \( Y_t \) vector, \( t=1, 2, \ldots, T \) and that \( X_{,j} \) is a \( T \times 1 \) vector consisting of the \( j \)-th element of the \( X_t \) vector, \( t=1, 2, \ldots, T \). Thus, \( Z \) is a \( T \times p \) matrix with \( t \)-th row equal to \( Z_t \); \( i \)-th column equal to \( Y_{,i} \), \( i=1, 2, \ldots, r \); and \( j \)-th column equal to \( X_{,j} \), \( j=r+1, \ldots, p \). Define analogous matrices for other variables similarly, e.g.,

\[
q_{,i} = (q_{1i}, q_{2i}, \ldots, q_{Ti})'.
\]
Also, to maintain clear terminology, the vectors $x_t$, $a_t$ and $q_t$ will be called the "components" of the observation $Z_t$ under model (2.1)-(2.2), while the variables $Z_{t1}$, $Z_{t2}$, ..., $Z_{tp}$ will be called the "elements" of the $Z_t$ vector. A similar distinction between "component" and "element" will apply to model (2.3)-(2.4).

Given the notation above, model (2.1)-(2.2) may be rewritten

$$Z = z + a \quad \text{(2.5)}$$

$$y = x^\xi + q \quad \text{(2.6)}$$

Expression (2.6) may also be written in the forms,

$$\text{vec}(y) = (\xi' \ a \ I_p)\text{vec}(x) + \text{vec}(q)$$

$$= (I_p \ a \ x)\text{vec}(\xi) + \text{vec}(q) \quad \text{(2.7)}$$

Similarly, model (2.3)-(2.4) implies that

$$Z = x(\xi, I_k) + \xi \quad \text{(2.8)}$$

or equivalently,
Each of these re-expressions of models (2.1)-(2.2) and (2.3)-(2.4) will be useful at various points in the work below.

As noted in Chapter 1, the fundamental difficulty associated with models (2.1)-(2.2) and (2.3)-(2.4) is that standard regression methods lead to unsatisfactory estimators of $\beta$ and of parameters of the component processes. This difficulty has led to the development of a large body of literature associated with errors-in-variables estimation. This literature is discussed thoroughly in the references mentioned at the beginning of this section. However, as an introduction to some ideas employed in Chapters 4, 5 and 6, it is useful to review briefly estimation for three forms of the uncorrelated errors-in-variables model with a single dependent variable: the homoscedastic model with an error in the equation; the homoscedastic model with no error in the equation; and the heteroscedastic model with known error variances. The discussion of these models is drawn from Sections 2.2, 2.3 and 3.1, respectively, of Fuller (1987).

First, for the homoscedastic model (2.1)-(2.2) with an error in the equation, and with a single "dependent variable" $Y$, replicated observations may permit the construction of an estimator

$$S_{aa} = \begin{pmatrix} S_{ww} & S_{wu} \\ S_{uw} & S_{uu} \end{pmatrix}$$

(2.10)
which is unbiased for $\text{Var}(a_t') = \Sigma_{a'a}$ . For this case Fuller (1987, p. 107) suggests that one estimate the regression coefficient $\beta$ with

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

(2.11)

where

$$M_{XY} = T^{-1} \sum_{t=1}^{T} X_t' Y_t$$

(2.12.a)

and

$$M_{XX} = T^{-1} \sum_{t=1}^{T} X_t' X_t$$

(2.12.b)

Under the ultrastructural form of model (2.1)-(2.2) with independent observations, Theorem 2.2.1 of Fuller (1987) gives conditions under which estimator (2.11) is consistent and asymptotically normal.

Corollary 4.1.1 below will generalize Theorem 2.2.1 of Fuller (1987) to a form of model (2.1)-(2.2) for serially correlated components.

Second, for the homoscedastic uncorrelated errors-in-variables model with no error in the equation, knowledge of sampling procedures or previous experience may permit one to know the covariance matrix $\text{Var}(\epsilon_t') = \Sigma_{\epsilon \epsilon}$ up to a scalar multiple. More formally, one may write

$$\Sigma_{\epsilon \epsilon} = \begin{pmatrix} \sigma_{ee} & \Sigma_{eu} \\ \Sigma_{ue} & \Sigma_{uu} \end{pmatrix} = \sigma^2 T_{\epsilon \epsilon} = \sigma^2 \begin{pmatrix} T_{ee} & T_{eu} \\ T_{ue} & T_{uu} \end{pmatrix}$$
and assume that $T_{\varepsilon \varepsilon}$ is known, while $\sigma^2$ is unknown. Theorem 2.3.1 of Fuller (1987) indicates that for the functional model (2.3)-(2.4) with $r=1$ and the errors $\varepsilon_t$ distributed as normal independent $(0, \Sigma_{\varepsilon \varepsilon})$ random vectors with $T_{\varepsilon \varepsilon}$ known, the maximum likelihood estimator of $\beta$ is

$$\hat{\beta} = (M_{XX} - \hat{\lambda}_T \Sigma_{uu})^{-1}(M_{XY} - \hat{\lambda}_T \Sigma_{ue})$$

(2.13)

where $\hat{\lambda}$ is the smallest root of the determinantal equation

$$|M_{ZZ} - \hat{\lambda}_T \Sigma_{\varepsilon \varepsilon}| = 0$$

and

$$M_{ZZ} = T^{-1} \sum_{t=1}^{T} z_t' z_t$$

Section 2.3.2 of Fuller (1987) indicates that expression (2.13) is also the maximum likelihood estimator of $\beta$ for the homoscedastic, uncorrelated structural model (2.3)-(2.4) with normal $x_t$ and $\varepsilon_t$. Also, sufficient conditions for the consistency and asymptotic normality of a generalization of estimator (2.13) are given in Theorem 2.3.2 of Fuller (1987). Theorem 4.2 below will discuss the asymptotic behavior of a form of estimator (2.13) under model (2.3)-(2.4) for serially correlated observations.

The estimators reviewed to this point have been constructed with the assumption that $\text{Var}(\varepsilon_t')$ is constant with respect to $t$. In some practical cases, the errors $\varepsilon_t$ may be heteroscedastic; for example,
the observations $Z_t$ may represent sample means computed from unequal numbers $n_t$ of observations. Hence, heteroscedastic versions of models (2.1)-(2.2) and (2.3)-(2.4) may be of interest. Booth (1973), Fuller and Booth (1983), Fuller (1984, 1987) and Hasabelnaby (1987) have addressed estimation for these heteroscedastic errors-in-variables models. In particular, Section 3.1.6 of Fuller (1987) noted that for the functional form of model (2.3)-(2.4) with $r=1$, if the errors $\varepsilon_t$ are normally and independently distributed with mean zero and covariance matrices $\Sigma_{\varepsilon t t}$, and if the matrices $\Sigma_{\varepsilon t t}$ are known, then maximum likelihood estimation of $\beta$ may be reduced to the problem of solving for the unknown $\beta$ in the equation

$$
\sum_{t=1}^{T} \left( \sigma_{v t t}^{-1} \left[ X_t (Y_t - X_t \beta) - \gamma_t (\Sigma_{u t t} - \Sigma_{u u t t} \beta) \right] \right) = 0 \quad (2.14)
$$

where

$$
\Sigma_{\varepsilon t t} = \begin{pmatrix} \sigma_{\varepsilon t t} & \Sigma_{u t t} \\ \Sigma_{u t t} & \Sigma_{u u t t} \end{pmatrix},
$$

$$
\sigma_{v t t} = (1, -\beta') \Sigma_{\varepsilon t t} (1, -\beta')',
$$

and

$$
\gamma_t = \sigma_{v t t}^{-1} [(1, -\beta') Z_t] \sigma_{v t t}^{-1}.
$$

Due to the definitions of $\sigma_{v t t}$ and $\gamma_t$, expression (2.14) is not a

Chapter 6 below will develop estimators for the homoscedastic normal functional model (2.3)-(2.4) with serially correlated errors. Derivative computations will lead to an expression similar to (2.14) and to an iterative maximum likelihood procedure similar to the heteroscedastic maximum likelihood procedure.

For an uncorrelated heteroscedastic form of model (2.1)-(2.2), Section 3.1.1 of Fuller (1987) suggested the estimator

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

where

$$\hat{M}_{x\pi z} = \begin{pmatrix} M_{x\pi y} & M_{y\pi x} \\
M_{y\pi y} & M_{x\pi x}
\end{pmatrix}$$

$$= T^{-1} \sum_{t=1}^{T} \pi_t (Z'_t Z_t - \hat{E}_{\pi t}) ,$$

$\hat{E}_{\pi t}$ is an estimator of $\text{Var}(a_t) = E_{\pi t}$, and $\{\pi_t, t=1, 2, \ldots, T\}$ is a set of estimated weights. Theorem 3.1.1 of Fuller (1987) discussed the asymptotic properties of the estimator $\hat{\beta}$ under a fairly general ultrastructural form of the uncorrelated errors-in-variables model (2.1)-(2.2). Section 4.2 below will discuss similar weighted estimators for the homoscedastic correlated form of model (2.3)-(2.4).
2.2. Notation and Models for Time Series

As noted in Chapter 1, a set of observations $Z_t$ may be taken over time, and this may result in serial correlation of some of the components of $Z_t$. Description of the moment structure of $\text{vec}(Z)$ then requires definitions and notation like those used in standard time series texts. Examples of such texts are Hannan (1970), Anderson (1971), Fuller (1976), and Box and Jenkins (1976).

The sequence of random vectors $\{Z_t\}$ is said to be second-order stationary if the following matrices are identical for all $t$:

$$
\mathbb{E}_Z \equiv \mathbb{E}(Z_t) \quad (2.15)
$$

and

$$
\Gamma_{ZZ}(h) = \text{Cov}(Z_t, Z_{t+h}) \quad (2.16)
$$

i.e., the mean and the lagged covariance structure of $\{Z_t\}$ are constant over time. Similarly, $\{Z_t\}$ is said to be fourth order stationary if for all $t$ the matrices (2.15), (2.16),

$$
\mathbb{M}_{ZZZ}(0, h, \ell) \equiv \mathbb{E}(Z_t \ a\ Z_{t+h} \ a\ Z_{t+\ell}) \quad (2.17)
$$

and

$$
\mathbb{K}_{ZZZ}(0, g, h, \ell) \equiv \mathbb{E}(Z_t \ a\ Z_{t+g} \ a\ Z_{t+h} \ a\ Z_{t+\ell}) \quad (2.18)
$$
are identical for all \( t \), i.e., the mean and lagged second, third and fourth moments are constant over time.

In the work below, the observations \( Z_t \) will be permitted to have nonconstant mean, but unless stated otherwise, it will be assumed that the sequence \( \{Z_t - E(Z_t)\} \) is second-order stationary. Assumptions of fourth-order stationarity will be stated explicitly as needed.

Let \( \Gamma_{ZZ} = \text{Var}[\text{vec}(Z)] \) be a \( T_p \times T_p \) matrix with \((i,j)\)-th \( T \times T \) block equal to \( \Gamma_{Z_iZ_j} = \text{Cov}(Z_i, Z_j) \), which in turn has \((s,t)\)-th element equal to \( \Gamma_{Z_iZ_j}(t-s) \). It is also useful to define \( \Gamma_{ZZ}(\ell) \), a \( p \times p \) matrix with \((i,j)\)-th element equal to \( \Gamma_{Z_iZ_j}(\ell) \). Define covariance and cross-covariance matrices for the random components \( \varepsilon_t \), \( a_t \), \( q_t \) and \( (x_t - \varepsilon_t) \) similarly. Note that the dimensions of \( \Gamma_{ZZ} \) and \( \Gamma_{Z_iZ_j} \) are functions of \( T \); this dependency will generally be suppressed in the notation to avoid excessive subscripting.

A fairly general model for a second-order stationary vector time series is a \( k \)-dimensional autoregressive moving average of order \((p, q)\), or \( \text{ARMA}_k(p, q) \), defined by

\[
\sum_{j=0}^{p} \phi_j u_{t-j} = \sum_{i=0}^{q} \theta_i c_{t-i},
\]

where \( u_t \) is the \( 1 \times k \) vector of interest observed at time \( t \); \( \phi_j \) and \( \theta_i \) are \( k \times k \) real matrices for all \( j=0, 1, \ldots, p \) and all \( i=0, 1, \ldots, q \); \( \phi_0 = I_k \); \( \theta_0 = I_k \); and \( \{c_t\} \) is a sequence of uncorrelated \( 1 \times k \) \((0, \Sigma_{cc})\) random vectors. Such a sequence \( \{c_t\} \) is sometimes called a \( k \)-dimensional white noise process. Hannan (1969)
defines the \( k \times k \) matrix generating functions for the autoregressive and moving average parameter matrices to be

\[
\Phi(\lambda) = \sum_{j=0}^{P} \phi_j \lambda^j \quad \text{and} \quad \Theta(\lambda) = \sum_{i=0}^{Q} \theta_i \lambda^i,
\]

respectively. Hannan (1969) also notes that if the process represented by (2.19) is second-order stationary with no zeros of the equation \(|\Phi(\lambda)| = 0\) on or inside the unit circle and with no zeros of the equation \(|\Theta(\lambda)| = 0\) inside the unit circle, then the matrix spectral density of \( \{u_t\} \) is

\[
f(w) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Sigma_{uu}(h)\exp(-iwh)
\]

\[
= (2\pi)^{-1} \{\phi(\exp(iw))\}^{-1} \Theta(\exp(iw)) \Sigma_{cc} \Theta^* \exp(iw) \{\Theta^* \exp(iw)\}^{-1} ,
\]

where \( A^* \) denotes the complex conjugate of the transpose of \( A \).

Two special cases of model (2.19) are of particular interest. The autoregressive model of order \( p \) is represented by

\[
\sum_{j=0}^{p} \phi_j u_{t-j} = c_t' ,
\]

i.e., the moving average parameters \( \theta_i \), \( i=1, 2, \ldots, q \), are assumed to be null. The moving average model of order \( q \) is represented by
i.e., the autoregressive parameters \( \phi_j \), \( j = 1, 2, \ldots, p \), are assumed to be null. As noted in Miazaki (1985), a moving average model may be appropriate for sampling errors associated with the use of rotation samples.

Box and Jenkins (1976) introduced a generalization of model (2.19) which they called an autoregressive integrated moving average (or ARIMA) model. Define the backshift operator \( B \) by the relation \( Bu_t = u_{t-1} \), and define the polynomials \( \Phi(B) \) and \( \Theta(B) \) of orders \( p \) and \( q \), respectively, by the relations \( \Phi(B) = \sum_{j=0}^{p} \phi_j B^j \) and \( \Theta(B) = \sum_{i=1}^{q} \theta_i B^i \). Then model (2.19) may be rewritten,

\[
\Phi(B)u_t' = \Theta(B)c_t' \quad (2.23)
\]

Following Gohberg et al. (1982), define the polynomial \( \Phi(B) \) to be comonic if \( \Phi(0) = I_k \). Unless noted otherwise, the work below will assume that \( \Phi(0) = I_k \) and \( \Theta(0) = I_k \), so that both \( \Phi(B) \) and \( \Theta(B) \) are comonic polynomials in the backshift operator \( B \).

Note that the backshift operator \( B \) commutes with any \( k \times k \) real matrix \( \psi = (\psi_{ij}) \), say, because \( \psi u_t' \) has \( i \)-th element equal to \( \sum_{j=1}^{k} \psi_{ij} u_{t-j} \) and thus \( \psi Bu_t' \) and \( B\psi u_t' \) both have \( i \)-th element equal to \( \sum_{j=1}^{k} \psi_{ij} u_{t-1-j} \). Define the difference operator \( \nabla \) by the relation \( \nabla = (1 - B) \), so that \( \nabla u_t = u_t - u_{t-1} \). This operator also commutes with any \( k \times k \) real matrix. If one believes that the \( d \)-th order
differences of \( u_t \), \( \nabla^d u_t = \sum_{\ell=0}^{d-1} \ell \nabla u_{t-\ell} \), follow an autoregressive moving average model (2.19), then one may represent the resulting relationship in the form

\[
\nabla(B) \nabla^d u_t = \nabla(B) c_t.
\]

Box and Jenkins (1976) defined \( u_t \) satisfying (2.24) to be an ARIMA\((p, d, q)\) model.

Many authors have proposed methods of estimating the parameters \( (p, d, q, \sum_{\infty} c \{\phi_j\}_{j=1}^p, \{ \theta_j\}_{1=1}^{q} ) \). The estimation of the parameters \( p, d \) and \( q \) is beyond the scope of the current work. Useful references include Akaike (1976) and Box and Jenkins (1976). One may use either time-domain or frequency-domain methods to estimate the parameters \( \sum_{\infty} c \{\phi_j\}_{j=1}^p \) and \( \{ \theta_j\}_{1=1}^{q} \). Akaike (1974), Box and Jenkins (1976), Brillinger (1976), Fuller (1976), and Hannan (1970) present general discussions of the estimation of these parameters, while Chapters 4, 5 and 6 below address estimation for certain "signal plus noise" models.

In general, a vector autoregressive moving average process need not have a unique representation (2.19) for a given pair \( (p, q) \). This clearly may cause difficulties in parameter estimation, so it is useful to know the circumstances under which representation (2.19) is unique.

Akaike (1974) noted that for the case \( k = 1 \), representation (2.19) is unique if the autoregressive and moving average orders \( p \) and \( q \), respectively, are minimal \( \text{i.e., if the process has no other} \)
representation (2.19) with smaller autoregressive or moving average orders]. For processes of general dimension $k$, Hannan (1969, p. 224) defined the $k \times k$ matrix $d$ to be a common left divisor of the $k \times k$ matrices $\phi$ and $\theta$ if there exist $k \times k$ matrices $\tilde{\phi}$ and $\tilde{\theta}$ such that $\phi = d \tilde{\phi}$ and $\theta = d \tilde{\theta}$; and defined $d$ to be a greatest common left divisor for $\phi$ and $\theta$ "if it is a common left divisor and any other common left divisor has $d$ as a right multiple." Hannan (1969, p. 224) then proved that the representation (2.19) of an autoregressive moving average is unique if and only if $I_k$ is a greatest common left divisor of the matrix polynomials $\phi(\lambda)$ and $\theta(\lambda)$, and $C^1(\theta) \cap C^1(\phi) = \phi$, where $C^1(\theta)$ is the $k$-dimensional null space of the columns of $\theta$ and $C^1(\phi)$ is the $k$-dimensional null space of the columns of $\phi$. Hannan (1969, p. 224) also noted that "requiring that $I$ be a G.C.L.D. is, in a sense, a trivial requirement for it merely eliminates the possibility that the equation system has been written in a redundant form."

Akaike (1974, p. 369) noted that a sufficient condition for the uniqueness of representation of (2.19) is that $\text{Var}(\tilde{\mathbf{w}}_t)$ be nonsingular, where $\tilde{\mathbf{w}}_t$ is the state vector of the state-space representation of the autoregressive moving average process. Akaike referred to this condition as "block identifiability" of a process. Brockwell and Davis (1987, p. 418) and Dunsmuir and Hannan (1976) contain additional discussion of identification issues for multivariate autoregressive moving average processes.
For the remainder of this work, it will be assumed that any vector autoregressive moving average process under discussion satisfies the necessary and sufficient conditions of Hannan (1969) and thus has unique representation (2.19).

2.3. Identification Issues for Measurement Error Models with Serially Correlated Errors or True Values

Issues of model identification have played a prominent role in the literature associated with errors-in-variables models (2.1)-(2.2) and (2.3)-(2.4). Fuller (1987, pp. 9-10) gave the following definition of model identification for a random vector \( Z \) with distribution function \( F_z(z; \theta) \) and a parameter vector \( \theta \) belonging to some parameter space \( \Theta \).

**Definition 2.1.** [Definition 1.1.1 of Fuller (1987)]. Let \( Z \) be the vector of observable random variables and let \( F_z(z; \theta) \) be the distribution function of \( Z \) for parameter \( \theta \) evaluated at \( Z = z \). The parameter \( \theta \) is identified if, for any \( \theta_1 \in \Theta \) and \( \theta_2 \in \Theta \), \( \theta_1 \neq \theta_2 \) implies that \( F_z(z; \theta_1) \neq F_z(z; \theta_2) \) for some \( z \). If the vector \( \theta \) is identified, we also say that the model is identified.

Recall that if the random vector \( Z \) has a multivariate normal distribution, then the distribution function \( F_z(z; \theta) \) of \( Z \) is determined by the first two moments of \( Z \), \( E(Z) \) and \( \text{Var}(Z) \). Thus, identification issues for the normal models (2.1)-(2.2) and (2.3)-(2.4) reduce to the question of whether there exists a one to one onto
transformation from the population moments $[E(Z), \text{Var}(Z)]$ to the unknown elements of the parameter vector $\theta$. Some authors [e.g., Anderson and Deistler (1985, p. 13)] have mentioned the possibility of using third- and higher-order moments for identification of non-normal versions of models (2.1)-(2.2) and (2.3)-(2.4), but this does not appear to have been pursued in detail. Also, as noted by Gleser (1983, p. 57), relatively few results are available regarding identification of functional errors-in-variables models. Consequently, the present discussion of identification issues will be restricted to normal structural versions of models (2.1)-(2.2) and (2.3)-(2.4).

Many authors [e.g., Fuller (1987, pp. 9-11)] have noted that given only the independent observations $Z_t$, $t=1, 2, \ldots, T$, the structural models (2.1)-(2.2) and (2.3)-(2.4) are not identified. To see this, note that if $\{Z_t, t=1, 2, \ldots, T\}$ is a set of independent and identically distributed normal random vectors that satisfy model (2.3)-(2.4), then for all $t$,

$$
E(Z_t) = \mu_{\theta} = \mu_{Z_t}(\theta, I_k)
$$

(2.25)

and

$$
\text{Var}(Z_t) = \text{Var}(Z_t)
$$
Given the general unidentified status of models (2.1)-(2.2) and (2.3)-(2.4), one may consider three general approaches to "estimation" for these models. First, even without an identified model, one may be able to establish certain bounds for regression parameters and model component variances. For examples of this approach, see Kalman (1982, 1983), Klepper and Leamer (1984) and Fuller (1987, pp. 10-11, 100-103).

Second, one may observe a sequence of "instrumental variables" $W_t$, such that $\text{Cov}(W_t, x_t) \neq 0$ and $\text{Cov}(W_t, e_t) = 0$. Subject to conditions on the $W_t$ and $x_t$ sequences, models (2.1)-(2.2) and (2.3)-(2.4) may then be identified; see Fuller (1987, Sections 1.4 and 2.4) for details of the resulting estimation procedures. Third, if some auxiliary information regarding certain model parameters is available,
then the remaining parameters of models (2.1)-(2.2) or (2.3)-(2.4) may be identified. For example, Section 2.1 above noted that if
\[ \Gamma_{ee}(0) = \sigma^2 \Gamma_{ee} \] and \( \Gamma_{ee} \) is assumed to be known, then \( \sigma^2, \beta \) and \( \Gamma_{xx}(0) \) are identified under the normal structural model (2.3)-(2.4).

Another example of model identification through the use of auxiliary information is as follows. Consider the normal structural model (2.1)-(2.2) with an error in the equation, and assume that

\[
\Gamma_{aa}(0) = \begin{pmatrix} \Gamma_{ww}(0) & \Gamma_{wu}(0) \\ \Gamma_{uw}(0) & \Gamma_{uu}(0) \end{pmatrix}
\]

is known. Assume also that \( \Gamma_{xx}(0) \) is positive definite. (This final condition may be assumed without loss of generality, because the condition \( |\Gamma_{xx}(0)| = 0 \) would indicate redundancies in the vector.) Then as noted in Fuller (1987, pp. 105-107), the equations

\[
 \mu_Z = (\mu_Y, \mu_X) = \mu_X(\beta, I_k)
\]

(2.27)

\[
 \Gamma_{YY}(0) = \beta' \Gamma_{xx}(0) \beta + \Gamma_{ww}(0) + \Gamma_{qq}(0)
\]

\[
 \Gamma_{XY}(0) = \Gamma_{xx}(0) \beta + \Gamma_{uw}(0)
\]

\[
 \Gamma_{XX}(0) = \Gamma_{xx}(0) + \Gamma_{uu}(0)
\]
have no more than one solution \( \{ \mu_x, \beta, \Gamma_{xx}(0), \Gamma_{qq}(0) \} \) in terms of the other parameters \( \{ \mu_z, \Gamma_{yy}(0), \Gamma_{xy}(0), \Gamma_{zx}(0), \Gamma_{za}(0) \} \), namely,

\[
\mu_x = \mu_z \tag{2.28}
\]

\[
\Gamma_{xx}(0) = \Gamma_{xx}(0) - \Gamma_{uu}(0)
\]

\[
\beta = [\Gamma_{xx}(0) - \Gamma_{uu}(0)]^{-1}[\Gamma_{xy}(0) - \Gamma_{uw}(0)]
\]

\[
\Gamma_{qq}(0) = \Gamma_{yy}(0) - \Gamma_{uw}(0)
\]

\[
- [\Gamma_{xy}(0) - \Gamma_{uw}(0)]'[\Gamma_{xx}(0) - \Gamma_{uu}(0)]^{-1}[\Gamma_{xy}(0) - \Gamma_{uw}(0)].
\]

Hence, given \( \Gamma_{za}(0) \), the parameters \( \mu_x, \beta, \Gamma_{xx}(0) \), and \( \Gamma_{qq}(0) \) are identified under the normal structural model (2.1)-(2.2) with an error in the equation. A similar argument indicates that if \( \Gamma_{uu}(0) \) and \( \Gamma_{ue}(0) \) are known, then under model (2.3)-(2.4) with no error in the equation, the parameters \( \mu_x, \beta, \Gamma_{xx}(0) \) and \( \Gamma_{ee}(0) \) are identified.

Note that the identification arguments using expressions (2.27) and (2.28) required only the assumption of homogeneous lag-zero covariances \( \Gamma_{xx}(0), \Gamma_{qq}(0) \) and \( \Gamma_{za}(0) \); a common regression parameter matrix \( \beta \); and a common mean vector \( \mu_x \). These identification arguments required no assumptions regarding the lagged covariance structure of the \( x \) or \( z \) components. Thus, the arguments regarding identification
of \( \{ \mu_x, \beta, \Gamma_{xx}(0), \Gamma_{qq}(0) \} \) given \( \Gamma_{aa}(0) \) apply to serially correlated forms of the normal structural model (2.1)-(2.2); and similarly for identification of \( \{ \mu_x, \beta, \Gamma_{xx}(0), \Gamma_{ee}(0) \} \) given \( \{ \Gamma_{uu}(0), \Gamma_{ue}(0) \} \) under the normal structural model (2.3)-(2.4). However, two remaining identification issues for serially correlated versions of the normal structural models (2.1)-(2.2) and (2.3)-(2.4) are as follows.

(i) Identification of the parameters of the lagged covariance structure of the \( x_t \) and \( \xi_t \) components through the availability of auxiliary estimates of some parameters.

(ii) Determination of the identification status of models (2.1)-(2.2) and (2.3)-(2.4) without knowledge of any elements of \( \Gamma_{aa} \) or \( \Gamma_{ee} \), but with restrictions on the covariance structure of the \( x \) and \( \xi \) processes.

Regarding issue (i), let \( \gamma_x, \gamma_q \) and \( \gamma_a \) represent parameter vectors that determine the distributions of \( x_c, q \) and \( a \), respectively; and let \( \gamma_x, \gamma_q \) and \( \gamma_a \) belong to parameter spaces \( A_x, A_q \) and \( A_a \), respectively. Recall from Section 2.2 that a vector autoregressive moving average model need not have a unique representation. Hence, if \( x_t \) and \( q_t \) followed vector autoregressive moving average models, knowledge of the observations \( \{ Z_t, t=1, 2, \ldots, T \} \) and of the entire covariance matrix \( \Gamma_{aa} \) would not necessarily assure identification of the autocovariance parameters of the \( x_t \) and \( q_t \) models. Nonetheless, one may note the relations
Expressions (2.28) and (2.29) imply that if the parameters $\alpha_x$ and $\alpha_q$ have a one-to-one onto relationship with the lagged covariance matrices 
\[
\{\Gamma_{xx}(\ell), 0 < \ell < L_x\} \quad \text{and} \quad \{\Gamma_{qq}(\ell), 0 < \ell < L_q\},
\]
respectively; if $T > \max(L_x, L_q)$; and if the matrices 
\[
\{\Gamma_{aa}(\ell), 0 < \ell < \max(L_x, L_q)\}
\]
are known, then the parameters $\alpha_x$ and $\alpha_q$, as well as $\nu_x$ and $\beta$, are identifiable from the observations \{\(z_t, t=1, 2, \ldots, T\)\}.

In practice, it is unlikely that one will know the parameters $\Gamma_{aa}(0)$ a priori, but re-interview or other replicated-observation methods may lead to estimates $\hat{\Gamma}_{aa}(\ell)$ of $\Gamma_{aa}(\ell)$. Then by Definition 2.1, identification issues center on the distribution of the random vector
\[
Z^* = \{\text{vec}(Z)'', \text{vech}[\Gamma_{aa}(0)]', \text{vec}[\hat{\Gamma}_{aa}(1), \hat{\Gamma}_{aa}(2), \ldots, \hat{\Gamma}_{aa}(L^*)]'\}
\]
for some positive integer $L^*$. If the parameterization of the models for $x$, $q$ and $a$ are such that each possible distribution of $Z^*$ corresponds to a single element of $A_x \times A_q \times A_a$, then the model for $Z^*$ is identified.

Regarding issue (ii), several authors have considered the identification status of correlated errors-in-variables models under restrictive parameterizations of the $x_t$, $q_t$ and $a_t$ processes. Under some such
restrictive parameterizations, models (2.1)-(2.2) and (2.3)-(2.4) may be identifiable from the observations \( \{Z_t, t=1, 2, \ldots, T\} \), without any auxiliary information regarding the values of specific nonnull elements of \( \Gamma_{\omega \omega} \). References on this matter include Maravall and Aigner (1977), Maravall (1979), Söderström (1980), Anderson and Deistler (1984) and Nowak (1985). Since the present work will not give any direct extensions or applications of the results developed in these references, it suffices to note that the authors mentioned above have demonstrated that in some cases, restrictions on component model parameterizations may lead to the identification of models (2.1)-(2.2) and (2.3)-(2.4) without any a priori estimates of specific nonnull parameters. For example, Nowak (1985) discussed versions of model (2.3)-(2.4) in which \( r = 1 \) and each of \( x_t, q_t, \omega_t \) and \( u_t \) follow independent autoregressive moving average models. For these models, he considered some conditions on the autoregressive and moving average orders of the component processes and showed that these conditions were sufficient for the identification of \( \beta \) as well as some parameters of the \( x_t, q_t, \omega_t \) and \( u_t \) processes.

Identification issues will not be central to the estimation work in Chapters 4, 5 and 6 for the following reasons. Chapter 4 will address only estimation of the regression parameter \( \beta \), and identifying information in the form of known or estimated \( \Gamma_{uu}(0) \) and \( \Gamma_{ue}(0) \) will be assumed. Chapters 5 and 6 will assume the availability of identifying information as described in either (i) or (ii) above, and will discuss properties of the normal structural- and functional-model
likelihood functions at a sufficiently high level of generality that the
details of the identifying information will not intrude. For specific
models of the $x_t$, $q_t$ and $a_t$ component processes, practical
application of the methods outlined in Chapters 5 and 6 would require
use of specific identifying information or restrictions, but details of
these applications will be deferred to future work.

2.4. Previous Work in Time Domain Estimation for
Measurement Error Models

The literature of time domain estimation for measurement error
models falls into two broad categories. First, there is a relatively
small amount of work on the estimation of the regression parameters and
component process parameters of the structural model (2.3)-(2.4) and on
estimation of closely related factor analysis models. For the most
part, the methods considered are slight adaptations of estimation
procedures originally developed for measurement error models or factor
analysis models with independent and identically distributed components.
Subsection 2.4.1 reviews these adaptations.

To facilitate discussion of more general measurement error models,
define a "signal plus noise" model to be any model that represents an
observation as the sum of a "signal" or "true value" component of
primary interest, and a "noise" or "error" component of secondary
interest. Then one may consider multivariate "signal plus noise" models
to be forms of model (2.1) without the linear restrictions (2.2) on the
elements of the "signal" component. Moreover, many results presented in
subsequent chapters are extensions of results developed previously for univariate "signal plus noise" models. Therefore, Subsection 2.4.2 reviews some previous work in time domain estimation for such models.

2.4.1. Previous work in time domain estimation for structural models and factor analysis models

To date, estimators for the structural model (2.3)-(2.4) with autocorrelated observations have been quite similar to estimators for the uncorrelated structural model. Deaton (1985) considered observations obtained with the same instrument from several independent, non-overlapping samples taken on T separate occasions. He asserted that in some cases, the analysis of the resulting cohort means over time may provide a satisfactory substitute for panel data commonly obtained from overlapping surveys that are subject to major attrition problems.

To consider Deaton's methods, restrict attention to a single cohort and let \( Z_t = (Y_t, X_t) \) be the sample mean of observations taken at time \( t \), where \( r = 1 \) is the dimensionality of \( Y_t \). Deaton (1985) noted that given the independence of the samples over time and the use of the same survey instrument on each sampling occasion, one may reasonably assume that \( \{a_t = (w_t, u_t), t=1, 2, ..., T\} \) is a sequence of independent and identically distributed \((0, \Sigma_{aa})\) random vectors. These assumptions allow one to estimate \( \Sigma_{aa} \) from the cross-sectional data from the \( T \) surveys. Deaton (1985) thus concluded that in either the functional or the structural case, one may use the methods of Fuller (1975) to estimate \( \beta \); the resulting estimators are equivalent to those
given in Section 2.1 for the case \( r = 1 \). Under the additional assumption of normality of \( \varepsilon_t \) and the implicit assumption that the errors \( q_t \) were mutually uncorrelated over time, he derived an expression for the variance of the asymptotic distribution of his proposed estimators.

Deaton also discussed briefly the instrumental variable estimation of \( \beta \), and noted that if the \( \{e_t\} \) are not serially correlated and the \( \{x_t\} \) are serially correlated, then lagged values of \( X_t \) may serve as instrumental variables for the estimation of \( \beta \). One may find similar remarks in other econometric and time series work, e.g., Kmenta (1971, Section 9.1) and Brillinger (1981, p. 323). Since much of the asymptotic work associated with this instrumental variable approach has appeared in the frequency domain literature, further discussion of this subject will be deferred to the following section.

Finally, Deaton (1985) addressed estimation for models of change. Define \( x^*_t = x_t - x_{t-1} \); define \( y^*_t, u^*_t, e^*_t \) similarly, and consider the model

\[
\begin{align*}
y^*_t &= x^*_t \beta + q^*_t \\
Y^*_t &= y^*_t + w^*_t \\
X^*_t &= x^*_t + u^*_t.
\end{align*}
\]

Note that if \( a_t = (w_t, u_t) \) is a sequence of uncorrelated \( (0, \Sigma_{aa}) \) random vectors, then \( a^*_t = (w^*_t, u^*_t) = a_t - a_{t-1} \) follows a first order
moving average model such that

\[ \text{Var}(a_t) = 2\Sigma \]

and

\[ \text{Cov}(a_t, a_{t-1}) = \Sigma_{aa}. \]

Deaton (1985) thus proposed the estimator

\[
\tilde{\beta}_\Delta = (\mathbf{M}_{X^*X^*}(0) - 2\Sigma_{uu})^{-1}(\mathbf{M}_{X^*Y^*}(0) - 2\Sigma_{ue})
\]

where

\[ \mathbf{M}_{X^*Y^*}(0) = T^{-1} \sum_{t=1}^{T} X_t Y_t \]

and \( \mathbf{M}_{X^*X^*}(0) \) is defined similarly. He then concluded that the covariance matrix of the asymptotic distribution of \( T^{1/2}(\tilde{\beta}_\Delta - \beta_\Delta) \) is

\[
(\Sigma_{X^*X^*} - 2\Sigma_{uu})^{-1} C(\Sigma_{X^*X^*} - 2\Sigma_{uu})^{-1},
\]

where

\[
C = [\tilde{\mathbf{M}}_{X^*X^*}(0) + 2\Sigma_{uu}] [\sigma_q q + 2(1, -\beta') \Sigma_{aa}(1, -\beta')']
\]

\[
- (1, -\beta') \Sigma_{aa}(1, -\beta')' [\tilde{\mathbf{M}}_{X^*X^*}(1) + \tilde{\mathbf{M}}_{X^*X^*}(-1) - \Sigma_{uu}]
\]

\[
+ 14 \Sigma_{uv\Sigma_{uu}'}
\]
\[ \tilde{\mathbf{X}}_{\mathbf{X^*X^*}}(0) = \lim_{T \to \infty} \mathbf{X}_{\mathbf{X^*X^*}}(0), \]

and

\[ \tilde{\mathbf{X}}_{\mathbf{X^*X^*}}(-1) = \left[ \tilde{\mathbf{X}}_{\mathbf{X^*X^*}}(1) \right]' = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbf{X}^*_t \mathbf{X}^*_{t-1}. \]

This again is a special case of the results of Section 4.1.

Hence, Deaton (1985) reached the conclusion that even if the measurement errors in a set of original observations are not serially correlated, interest in models for change in level may require parameter estimation in the presence of serially correlated errors.

An alternative approach to estimation for model (2.3)-(2.4) in the case \( r = k = 1 \) appears in de Leeuw and McKelvey (1983). These authors proposed the use of Wald's method of grouping to estimate \( \beta \). They did not, however, present any small-sample or asymptotic properties of the resulting estimator. This is of particular concern because, as noted in Fuller (1987, pp. 73-74), the method of grouping may lead to inconsistent estimators if the grouping method is dependent on the errors \( e_t \) and \( u_t \). De Leeuw and McKelvey (1983) do not give enough mathematical detail to permit a definitive assessment of their estimation procedure, but it appears that their grouping method is in general dependent on the errors \( e_t \) and \( u_t \). Grouping methods independent of \( e_t \) and \( u_t \) lead to procedures that are special cases of the instrumental variable estimation as described by Fuller (1987, p. 74).
As noted in Fuller (1987), structural measurement error models are closely related to factor analysis models. Therefore, it is desirable to review some previous work with "dynamic factor models," i.e., factor analysis models in which one or more factors may be serially correlated.

Stone (1947) applied principal component factor analysis to a multivariate time series of income and expenditure data. His analysis, however, was based only on the matrix of sums of lag zero squares and cross products, and he took no explicit account of the serial correlation of his observations. Cattell (1952, pp. 104-105, 363) proposed a "P-technique" to analyze repeated multivariate measures on a single unit.

Given a sequence of observations \( \{Z_t = (Z_{t1}, Z_{t2}, ..., Z_{tp})\} \), lags \( \ell_2, \ell_3, ..., \ell_p \) are chosen and a factor analysis is performed on the resulting lagged observations \( Z_t^* = (Z_{t1}, Z_{t-\ell_2}, Z_{t-\ell_3}, ..., Z_{t-\ell_p}) \). The resulting analysis is based only on the matrix of sums of lag zero squares and cross products of the \( Z_t^* \) vectors. Cattell (1952, p. 105) suggested that the lags \( \ell_2, \ell_3, ..., \ell_p \) be chosen to maximize the resulting factor loadings. Cattell (1963, pp. 187-190) suggested an iterative procedure to obtain a different set of lags for each factor.

Molenaar (1985) criticized the entire Cattell P-technique approach because it permitted a given factor \( f_i \) to have a nonzero loading for a given variable \( Z_j \) at only one time lag, say \( \ell_2 \). As an alternative, Molenaar (1985) suggested the factor model,

\[
Z_t' = \sum_{s=0}^{q} a_s x_{t-s} + e_t', \quad t = \ldots, -1, 0, 1, \ldots \tag{2.30}
\]
where \( \{Z_t'\} \) is a \( p \times 1 \) series of observations, \( \{x_t'\} \) is a \( k \times 1 \) "factor series" assumed to be distributed i.i.d. \( N_k(0, I_k) \), \( \{a_s\} \) is a fixed sequence of \( p \times k \) factor loadings, and \( \{e_t'\} \) is serially correlated normal error series such that

\[
\Gamma_{e e}(s) = \text{Cov}(e_t', e_{t+s}) = \text{diag}[\gamma_{ee11}(s), \gamma_{ee22}(s), \ldots, \gamma_{eepp}(s)].
\]

Hence, the common part of \( Z_t' \) is a general vector moving average of order \( q \), and the unique part of \( Z_t' \) is an error term correlated across time but uncorrelated across vector components.

Given some integer \( m > q \), one may rewrite model (2.30) in the form

\[
\tilde{Z}_t' = \tilde{a} \tilde{x}_t' + \tilde{\xi}_t' \tag{2.31}
\]

where

\[
\tilde{Z}_t = (Z_t', Z_{t-1}', \ldots, Z_{t-m}'),
\]

\[
\tilde{x}_t = (x_t', x_{t-1}', \ldots, x_{t-m-s}'),
\]

\[
\tilde{\xi}_t = (\xi_t', \xi_{t-1}', \ldots, \xi_{t-m}'), \text{ and}
\]
Note that the matrix \( \tilde{a} \) has dimension \( p(m+1) \times k(m+s+1) \) and is partitioned into blocks of dimension \( p \times k \), so that there are \( m+1 \) rows of blocks and \( m+s+1 \) columns of blocks.

Model (2.31) is a special type of state-space model; Chapter 5 addresses more general types of state-space models and their application to structural measurement error models. Molenaar (1985) suggested the use of LISREL software by Jöreskog and Sörbom (1978) to obtain maximum likelihood estimates of the parameters of model (2.31), but he did not present the computational details of this proposal.

Box and Tiao (1977) and Peña and Box (1987) have also discussed factor analysis and principal-component analysis of multivariate time series. The particular techniques that they used are beyond the scope of the present work, but it is worthwhile to note that for one data set, Peña and Box found that a dynamic factor model gave much more satisfactory results than a general multivariate ARIMA model.

For the most part, the methods reviewed above constitute fairly direct extensions of procedures developed originally for structural or
factor models with independent observations. The following subsection addresses a related area in which authors have devoted explicit attention to the autocovariance structure of observations.

2.4.2. Previous work in time domain estimation for general "signal plus noise" models

In addition to the work with structural and factor analysis models discussed above, there is a considerable body of literature devoted to the following related problem. Let

$$X_t = X_t + u_t,$$  \hspace{1cm} (2.32)

where \{X_t\} is a sequence of \(1 \times k\) observations, \{x_t\} is the associated sequence of "true values" of interest, and \{u_t\} is a sequence of measurement errors associated with sampling error, instrument error, or some other source. Given replicated observations, one may be able to estimate the parameters of the \(u_t\) process. Principal interest may then focus on estimation of true \(x_t\) values, if \{x_t\} is a fixed sequence; or on estimation of the parameters of the \{x_t\} process and prediction of current or future \(x_t\) values, if \{x_t\} is a sequence of random vectors. As indicated in the following chapters, many of the ideas associated with estimation for model (2.32) will be useful in the development of estimators for models (2.1)-(2.2) and (2.3)-(2.4). Therefore, it is desirable to review some of the literature associated with model (2.32).
First, several authors have addressed modeling and estimation problems for observed autoregressive or autoregressive moving average signals corrupted by white noise.

For the case $k = 1$, Walker (1960) noted that if $x_t$ follows an AR($p$) process and $u_t$ follows an independent white noise process, then $X_t$ follows an ARMA($p$, $p$) process. Walker (1960) used this result and properties of the sample autocovariance function to obtain method-of-moments estimators of the parameters $\gamma_{XX}(0)$, $\gamma_{uu}(0)$, and $\phi_{xj}$, $j=1, 2, ..., p$. He then assessed the asymptotic properties of such estimators, with principal attention to the case $p = 1$.

Anderson et al. (1969) considered the case of general dimension $k$ under the assumption that $x_t$ follows an AR$_k(1)$ process and $u_t$ follows an independent $k$-dimensional white noise process. The authors presented method-of-moments estimators of the parameters of the $x_t$ and $u_t$ processes, and gave conditions under which these estimators are strongly consistent. In addition, they assessed the effect of parameter estimation on the performance of linear least squares predictors of $X_t$ and established conditions under which the contribution of parameter estimation error to prediction error becomes negligible with increasing sample size.

A special case of the results in Kashyap (1970) indicated that if $x_t$ follows an ARMA$_k(p, p-1)$ process and $u_t$ follows an independent $k$-dimensional white noise process, then $X_t$ follows an ARMA$_k(p, p)$ process. In fact, Kashyap (1970) stated his results and proofs for the case in which parameters of the $x_t$ and $u_t$ processes are time-
inhomogeneous. The present work will not pursue this more general case, but such results for time-inhomogeneous processes may be of value for survey sampling with nonconstant sample characteristics, as well as for the engineering problems within Kashyap's original domain of application.

Pagano (1974) considered the same problem as did Walker (1960). Under the assumption of normality, he used a nonlinear least squares procedure to obtain strongly consistent and asymptotically efficient estimators of the parameters of the $x_t$ and $u_t$ processes. He further sketched an extension of his results to the case in which $x_t$ follows an ARMA$(p, q)$ process, $p < q$, and $u_t$ follows an independent white noise process. Box and Jenkins (1976, Appendix A.4.4) generalized the modeling results of Walker (1960) and Pagano (1974). The details of this generalization are of considerable interest to the present work, so discussion will be deferred to Chapter 5.

Sakai and Arase (1979) again considered the case in which $x_t$ is an autoregressive process and $u_t$ is a white noise process. For the estimation of the parameters of the $x_t$ and $u_t$ processes, they considered two methods, the first based on a solution to the Yule-Walker equations and the second based on a modified least squares procedure. Simulation results from fifty runs of $T = 2000$ observations each led Sakai and Arase (1979) to conclude that their modified least squares estimator was superior to the Yule-Walker-type estimator in terms of mean squared error.
The publications discussed above restricted the measurement errors \( \{u_t\} \) to be a sequence of uncorrelated random variables. As noted in Chapter 1, however, many measurement processes in the social sciences and engineering lead to measurement errors that are serially correlated. This fact has received considerable attention in the survey sampling literature.

Initial work on this problem treated a "true" value \( x_t \) as a fixed real number, generally identified with a finite population mean at time \( t \). Jessen (1942) considered the special case of sampling on two occasions with unequal numbers of observations, and studied the optimal allocation of units to overlapping and nonoverlapping sample groups. Patterson (1950) considered sampling on \( T \) occasions under several schemes of partial replacement of units. The simplest such sampling plan required the replacement of a proportion \( u \) of sampling units on each successive sampling occasion. Also, Patterson (1950) assumed that for a given \( i \) the differences \( x_{ti} - x_t \), \( t=1, 2, \ldots \), followed a first-order autoregressive process, where \( x_{ti} \) was the value of the \( i \)-th population unit at time \( t \), and \( x_t \) was the corresponding finite population mean. Under the resulting error model, he developed optimal estimators of the fixed \( x_t \) values and of the differences \( x_t - x_{t-1} \). He further considered the optimal estimation of \( x_t \) under generalizations of the partial replacement plan to the cases of nonconstant sample sizes, nonconstant sampling error variances, and nonconstant replacement proportions. Patterson (1950) concluded with brief discussions of optimal sample size selection and of nonautoregressive errors.
Subsequent to Patterson (1950), several other authors [e.g., Gurney and Daly (1965)] have also pursued a fixed-x approach, but Blight and Scott (1973) noted that in some survey sampling problems, it may be reasonable to consider the \( \{x_t\} \) sequence to be a realization of a stochastic process. Thus, Blight and Scott (1973) retained the Patterson (1950) assumptions regarding sampling errors and rotation patterns, but replaced the assumption of fixed \( x_t \) values with the assumption that \( x_t \) followed a first-order autoregressive process independent of the sampling error process. Given the parameters of the resulting model, they derived minimum mean squared error predictors of \( x_t \) and \( x_t - x_{t-1} \), and developed a formula for the optimal number of units to replace on each sampling occasion.

Scott and Smith (1974) extended the results of Blight and Scott (1973) in two directions. First, for nonoverlapping surveys (and hence, uncorrelated \( u_t \)) and a general nonstationary \( x_t \) process with known parameters, they presented a Kalman filter approach to the prediction of true \( x_t \) values. Second, for general covariance stationary \( x_t \) and \( u_t \) processes, they applied the methods of Whittle (1963) to obtain minimum mean squared error linear predictors of \( x_t \) and of linear combinations, \( \sum_j d_j x_{t+j} \), say.

Scott, Smith and Jones (1977) noted that the methods of Patterson (1950) and of Blight and Scott (1973) required one to distinguish among survey respondents according to their membership in previous samples. Confidentiality restrictions or simple recordkeeping problems may make such distinctions impossible. In such cases, it may be preferable to
use the more general time series methods of Scott and Smith (1974),
which required only overall sample means for each sampling occasion, but
which also relied more heavily on modeling assumptions for the $x_t$
and $u_t$ processes. Scott, Smith and Jones (1977) also extended some of
the results of Scott and Smith (1974) to multi-stage surveys in which
units may overlap at one or more stages of sampling.

Smith (1978) provided a brief, accessible review of the fixed-$x$ and
random-$x$ approaches to model (2.32). In addition, he expanded the
Scott, Smith and Jones (1977) critique of the fixed-$x$ approach, and
presented theoretical and practical arguments in favor of time series
modeling of $x_t$.

Under the assumption of known process parameters, Jones (1979)
compared the relative variances of the Patterson (1950), Blight and
Scott (1973) and Scott and Smith (1974) estimators of level $x_t$ and
change $x_t - x_{t-1}$ under the Blight and Scott (1973) models for $x_t$
and $u_t$. For the estimation of $x_t$, he found that the Blight and
Scott (1973) and Scott and Smith (1974) methods were generally superior
to those of Patterson (1950), especially if the variance of $x_t$ was
small compared to the variance of $u_t$. Further, the Blight and Scott
(1973) estimator of $x_t$ was somewhat superior to the Scott and Smith
(1974) estimator. Jones (1979) noted similar results for estimators of
change, but observed that if the deviations $x_{ti} - x_t$ of the i-th
unit's value $x_{ti}$ from the finite-population mean $x_t$ were strongly
autocorrelated, then the Scott and Smith (1974) method could perform
quite poorly compared to the Patterson (1950) and Blight and Scott
Jones (1979) closed by noting that if only a relatively short series is available, the time series methods of Blight and Scott (1973) and Scott and Smith (1974) may lose some of their theoretical efficiency through preliminary detrending, deseasonalizing, and estimation of the parameters of the $x_t$ and $u_t$ processes.

Jones (1980) used general least squares theory to obtain best linear unbiased estimators of $x_t$ under model (2.32) with fixed-$x$ and random-$x$ assumptions. He then discussed the estimators of Patterson (1950), Gurney and Daly (1965), Blight and Scott (1973), and Scott and Smith (1974) as special cases of his generalized least squares estimators. Also, he sketched a general mixed linear model approach to the estimation of $x_t$ with nonstationary mean, but indicated a general preference to analyze the stationary differences of such nonstationary series.

Smith and Brunsdon (1986) applied the results of Scott, Smith and Jones (1977) to the analysis of categorical data under an additive-logistic transformation. They suggested the application of this method to small area estimation and also discussed briefly the use of combined time series and cross-sectional data in small-area estimation.

The survey sampling papers reviewed above devoted primary interest to the estimation of true $x_t$ values, given the parameters of the $x_t$ and $u_t$ processes. In practice, the parameters of the $u_t$ process are generally estimable from replicated survey data, but estimation of the parameters of the $x_t$ process requires additional work. Miazaki (1985) extended the nonlinear least squares estimation procedure of Pagano...
in the following manner. Let (2.32) hold with $k = 1$, assume
that $x_t$ follows an AR($p_x$) model,

$$
\sum_{j=0}^{p_x} \phi_j x_{t-j} = g_t
$$

and that $u_t$ follows an MA($q_u$) model,

$$
u_t = \sum_{i=0}^{q_u} \theta_i c_{t-i},$$

where \{g_t\} and \{c_t\} are mutually uncorrelated sequences of mutually
uncorrelated $(0, \sigma_{gg})$ and $(0, \sigma_{cc})$ random variables, respectively.
Then

$$
\sum_{j=0}^{p_x} \phi_j x_{t-j} = g_t + \sum_{j=0}^{p_x} \phi_j \sum_{i=0}^{q_u} \theta_i c_{t-j-i}
$$

has covariance function equal to zero at lags greater than $p_x + q_u$, so $x_t$
follows an ARMA($p_x$, $p_x + q_u$) model. Chapter 5 will address
this argument in considerable detail, but for the moment it suffices to
note that Miazaki (1985) used results associated with expression (2.33)
to develop a nonlinear least squares procedure to estimate the
parameters

$$
\{\sigma_{gg}, \phi_0, \phi_1, \phi_2, \ldots, \phi_{p_x}, \sigma_{cc}, \theta_0, \theta_1, \theta_2, \ldots, \theta_{q_u}\}
$$

from the observations $x_t$ and preliminary survey design-based esti-
mators \( \{ \hat{\sigma}_{cc}, \hat{\theta}_{u1}, \hat{\theta}_{u2}, \ldots, \hat{\theta}_{uq_u} \} \) of the noise process parameters.

She then found conditions under which the resulting final parameter estimators were consistent and asymptotically normal. In addition, Miazaki (1985) presented conditions under which least squares prediction of \( x_t \) with estimated process parameters results in an additional \( O_p(T^{-1/2}) \) prediction error term.

Binder and Hidiroglou (1987, pp. 29-40) considered model (2.32) under the assumptions that \( k = 1 \), that \( x_t \) follows an ARMA\((p_x', q_x')\) model, and that \( u_t \) follows an ARMA\((p_u', q_u')\) model. They suggested that the model for the resulting observations \( X_t \) is most easily represented in state-space form, rather than through a direct extension of the Miazaki (1985) arguments associated with (2.33) above; in Chapter 5 below, a comparison of multivariate ARMA and state-space models for "signal plus noise" observations leads to the same conclusion. Binder and Hidiroglou (1987) then applied the methods of Harvey and Philips (1979) and Harvey (1984) to outline an iterative numerical procedure for the computation of maximum likelihood estimates of the parameters of the ARMA\((p_x', q_x')\) model for \( x_t \).

The preceding review of time domain prediction and estimation procedures for measurement error models leads to several conclusions. First, the literature on the "signal plus noise" model (2.32) indicates considerable interest in the analysis of such models, with principal interest directed toward the estimation of true \( x_t \) values or the estimation of the parameters of an autoregressive moving average model for \( x_t \). Moreover, the data analyses presented in this literature...
reinforce the conclusion of Chapter 1 that many data series obtained through survey methods do indeed contain serially correlated measurement errors.

The work of Deaton (1985) and de Leeuw and McKelvey (1983) provide some indication of possible extensions of common errors-in-variables methods to models with serially correlated errors, but examination of these papers also indicates that considerably more work remains to be done for the development and assessment of efficient estimation procedures for model (2.1)-(2.2). Finally, the dynamic factor analysis results of Molenaar (1985) provide some suggestion of the importance of state-space methods for models with serially correlated components.

Chapters 4, 5 and 6 below use the methods reviewed in this section to develop method-of-moments and maximum likelihood estimators for the parameters of model (2.1)-(2.2).

2.5. Previous Work in Frequency Domain Estimation for Measurement Error Models

As indicated by Hannan (1970, p. 325), the statistical analysis of time series is often divided into two categories: "time-domain analysis" and "frequency-domain analysis." Time domain analysis addresses the behavior of a sequence of observations \( \{Z_t, t=1, 2, \ldots, T\} \) with particular attention to the first two population moments, \( E(Z_t) \) and \( \tau_{ZZ} \); sample forms of these moments; and the normal likelihood function of \( Z \). On the other hand, for a second-order stationary time series \( \{Z_t\} \), frequency domain analysis is based on the Fourier transformation
of the function \( \{\Gamma_{zz}(h), h \in \mathbb{Z}\} \). This transform is often called the "spectral density function" of \( Z \); and may be written,

\[
f(w) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma_{zz}(h) \exp(-iwh)
\]

where \( i = (-1)^{1/2} \) and \( w \in \mathbb{R} \). Hannan (1970), Anderson (1971, Chapters 8, 9 and 10), Fuller (1976, Chapters 3, 4 and 7) and Brillinger (1981) provide detailed discussions of the theory and methods of frequency domain analysis.

Empirical work in the frequency domain is often based on an estimator of the spectral density function. Brillinger (1981, pp. 238-248) suggests that the \((j, \ell)\)-th element of \( f(w) \) be estimated by

\[
\hat{f}_{j, \ell}^{(T)}(w) = 2\pi T^{-1} \sum_{t=1}^{T-1} W_{j, \ell}^{(T)}(w - 2\pi t^{-1}) I_{j, \ell}^{(T)}(2\pi t^{-1})
\]

where

\[
I_{j, \ell}^{(T)}(w) = (2\pi T)^{-1} \sum_{t=0}^{T-1} Z_{t+1, j} \exp(-iwt); \quad \sum_{t=0}^{T-1} Z_{t+1, \ell} \exp(-iwt)
\]

the bar "\(^{-}\)" denotes complex conjugation; \( 1 \leq j, \ell \leq p \); \( w \in \mathbb{R} \); and \( \{W^{(T)}(\cdot)\} \) is any one of a class of weight function discussed in detail by Brillinger (1981, Chapter 7).

Several authors have considered frequency-domain approaches to errors-in-variables models and related problems. The fundamental idea for such work employs two steps. First, one establishes certain
Several authors have considered frequency-domain approaches to errors-in-variables models and related problems. The fundamental idea for such work employs two steps. First, one establishes certain relationships among the spectral densities of the observed components \([e.g., \mathbf{Z} = (\mathbf{Y}, \mathbf{X})]\) and unobserved components \([e.g., \mathbf{y}, \mathbf{x}, \mathbf{e} \text{ and } \mathbf{u}]\). Second, one uses these relationships and the estimated spectral densities of the observed variables to obtain consistent estimators of \(\mathbf{\beta}\) and other parameters of interest. The forms of the resulting frequency-domain procedures are generally very similar to the forms of the corresponding time domain procedures.

Hannan (1963) discussed an ultrastructural form of model (2.3)-(2.4) with no error in the equation under the restrictions that \(r = 1\) and that \(\mathbf{x}, \mathbf{u}\) and \(\mathbf{e}\) are mutually uncorrelated. He considered two approaches to estimation: the method of instrumental variables and a method that employed auxiliary information regarding \(f_{uu}(0)\). For the instrumental variable case, Hannan (1963) assumed the availability of a sequence of \((l\times k)\)-dimensional "instrumental variables" \(\mathbf{W}_t\) such that the \((k\times k)\)-dimensional cross-spectral densities \(f_{\mathbf{Wx}}(\lambda)\) and \(f_{\mathbf{WW}}(\lambda)\) were nonsingular; and the cross-spectral densities \(f_{\mathbf{Wu}}(\lambda)\) and \(f_{\mathbf{We}}(\lambda)\) were identically zero. He then based his estimation method on the observation that

\[
f_{\mathbf{WX}}(\lambda) = f_{\mathbf{WX}}(\lambda)\mathbf{\beta}
\]

for all real numbers \(\lambda\). Given the initial consistent estimator,
based on time-domain instrumental-variable procedures [Fuller (1987, Section 2.4)] and given estimated spectral densities \( \hat{f}_{ZZ}(\lambda) \) and \( \hat{f}_{WZ}(\lambda) \) defined by (2.34), Hannan (1963) defined

\[
\hat{f}_{VY}(\lambda) = \hat{f}_{YY}(\lambda) - \hat{\beta}(1)\hat{f}_{XY}(\lambda) - \hat{\beta}(1)\hat{f}_{XX}(\lambda)\hat{\beta}(1) + \hat{\beta}(1)\hat{f}_{XX}(\lambda)\hat{\beta}(1),
\]

an estimate of the spectral density of \( v = Y - XB = e - uB \). He used these initial estimators to define a second-stage estimator

\[
\hat{\beta}(2) = \frac{1}{(2T)^{-1}} \sum_{t=-T+1}^{T} \hat{f}_{XW}(\lambda_t)[\hat{f}_{WW}(\lambda_t)]^{-1}\hat{f}_{WX}(\lambda_t)[\hat{f}_{VV}(\lambda_t)]^{-1}
\]

where \( \lambda_t = \pi t T^{-1} \). Note that this estimator employs the weights \( [\hat{f}_{VV}(\lambda_t)]^{-1} \). Section 4.2 will discuss a time domain estimator that uses a very similar weighting scheme. Hannan (1963) also discussed some asymptotic properties of the estimator (2.36).

In addition to instrumental variable estimation, Hannan (1963) discussed estimation of \( \beta \) in the presence of auxiliary information regarding the covariance structure of \( u \). In particular, under the restrictions that \( r = k = 1 \), that \( \{u_t\} \) formed a sequence of uncorrelated \( [0, \Gamma uu(0)] \) random variables, and that \( \text{Cov}(u_t, e_s) = 0 \) for all integers \( t \) and \( s \), he noted the relationships
Given spectral density estimates \( \hat{f}_{zz}(\lambda) \) and an initial consistent estimator \( \hat{\beta}^{(3)} \) of \( \beta \), Hannan (1963) defined

\[
s_{uu,t} = 2\pi \left[ \hat{f}_{xx}(\lambda_t) - \beta^{(3)} R[\hat{f}_{xy}(\lambda_t)] \right], \quad t = -T+1, -T+2, \ldots, T;
\]

\[
s_{uu} = (2\pi)^{-1} \sum_{t=-T+1}^{T} s_{uu,t};
\]

and

\[
\hat{f}_{xx}(\lambda) = \hat{f}_{xx}(\lambda) - (2\pi)^{-1}s_{uu},
\]

where \( R(a) \) denotes the real part of a complex number \( a \). He then used expressions (2.37) and (2.38) to propose several estimators of \( \beta \), including

\[
\hat{\beta}^{(4)} = \left\{ \sum_{t=-T+1}^{T} \hat{f}_{xx}(\lambda_t) \hat{f}_{xx}(\lambda_t) \hat{f}_{vv}(\lambda_t) + (4\pi^2)^{-1} \beta^{(3)} s_{uu} \right\}^{-1}
\]

\[
\times \left\{ \sum_{t=-T+1}^{T} \hat{f}_{xx}(\lambda_t) \hat{f}_{xx}(\lambda_t) \hat{f}_{vv}(\lambda_t) + (4\pi^2)^{-1} \beta^{(3)} s_{uu} \right\}^{-1}
\]

\[
(2.39)
\]
and

\[ \hat{\beta}(5) = \left\{ \sum_{t=-T+1}^{T} \hat{F}_{XX}(\lambda_t) \left[ \hat{F}_{YV}(\lambda_t) \right]^{-1} \right\}^{-1} \left\{ \sum_{t=-T+1}^{T} \hat{F}_{XY}(\lambda_t) \left[ \hat{F}_{VY}(\lambda_t) \right]^{-1} \right\}, \quad (2.40) \]

where \( \hat{F}_{VV}(\lambda) \) was defined by a form of expression (2.35) in which \( \hat{\beta}(1) \) was replaced by \( \hat{\beta}(3) \). Hannan (1963) closed his discussion by presenting some asymptotic properties of the estimators (2.39) and (2.40).

Parzen (1967) considered some extensions of Hannan (1963) through the use of "bispectral" methods. Given a sequence \( \{Z_t\} \) of mean-zero, third-order stationary random vectors, Parzen (1967) defined the "bispectral density"

\[ q_z(\lambda, \omega) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \exp[-i(j\lambda + \ell\omega)]E[Z(t)Z(t+j)Z(t+j)], \]

For model (2.3)-(2.4) and extensions thereof, he noted that if the "signal" \( x \) is nonnormal and has nonzero third cumulant function, and if the "noise" \( \xi \) is normal (and thus has third cumulant function identically equal to zero), then one could use an estimated bispectral density function to obtain estimates of \( \beta \) and related parameters. Parzen (1967) discussed several estimation procedures of this type, but did not evaluate the asymptotic or small-sample properties of the resulting estimators.

In addition to the work of Hannan (1963) and Parzen (1967) with model (2.3)-(2.4), Hinich (1983) presented a frequency-domain approach
to the following related continuous-time problem. For all real numbers \( t \), let \( y(t) \) and \( x(t) \) denote the "output" and "input," respectively, of a system at time \( t \), and assume that

\[
y(t) = \int_{s=-\infty}^{\infty} \beta(s)x(t-s)ds.
\]

(2.41)

The function \( \beta(s) \) is called the "impulse response function" associated with expression (2.41). The "transfer function" \( B(w) \) is defined to be the Fourier transform of \( \beta(s) \):

\[
B(w) = \int_{s=-\infty}^{\infty} \beta(s)\exp(-2\pi is)ds
\]

(2.42)

where \( i = (-1)^{1/2} \). Also, the "gain" associated with expression (2.41) is \( |B(w)| \), the absolute value of expression (2.42). Hinich (1983) considered the "noisy" observations

\[
Y(n\tau) = y(n\tau) + e(n\tau)
\]

(2.43.a)

and

\[
X(n\tau) = x(n\tau) + u(n\tau)
\]

(2.43.b)

\( n=1, 2, \ldots, N \), where \( \tau > 0 \) is known and \( \text{Cov}[e(n\tau), u(m\tau)] \) is assumed to equal zero for all \( m,n = 1, 2, \ldots, N \). He noted that this final assumption implied that the cross-spectral density \( f_{XY}(w) \) of \( X \) and \( Y \) is equal to the cross-spectral density \( f_{XY}(w) \) of \( x \) and
This result let Hinich (1983) to derive an estimator of \( |B(w)| \) based only on the real and imaginary parts of the estimated cross-spectral density \( \hat{f}_{XY}(w) \) computed from the observations \( \{[X(nt), Y(nt)], n=1, 2, \ldots, T\} \). He also presented some asymptotic properties of this estimator of \( |B(w)| \) and discussed extensions of his method for data of the form (2.43) observed in \( P \) disjoint segments of length \( N \) each.

There has also been some frequency-domain analysis of two classes of problems closely related to the errors-in-variables model (2.3)-(2.4): factor analysis models and general "signal plus noise" models. First, as indicated in Peña and Box (1987), a number of authors have considered a frequency-domain approach to factor analysis models. In particular, Brillinger (1981, p. 354) considered the factor model

\[
Z_t' = \mu' + \sum_{h=-\infty}^{\infty} A_{t-h} x_h' + e_t', \quad t \in \mathbb{Z},
\]

where \( \{Z_t'\} \) is a \( p \times 1 \) series of observations, \( \{x_t'\} \) is a \( k \times 1 \) "factor series" with zero mean, \( \{A_t\} \) is a sequence of \( p \times k \) "factor loading" matrices, \( \{e_t'\} \) is a \( p \times 1 \) serially correlated error series, and \( \mu' \) is a \( p \times 1 \) mean vector. He suggested that the factor loadings \( \{A_t\} \) be estimated through a principal-component analysis of the estimated spectral density matrix.

Second, there is a considerable body of literature associated with the frequency domain analysis of the general "signal plus noise" model.
introduced in Section 2.4. Restrictive assumptions [e.g., a priori knowledge or control of the $x_t$ values] make much of this literature tangential to the problems addressed in this dissertation. However, Förstner (1985) did give a frequency-domain approach to the Walker (1960) problem reviewed in Section 2.4. In particular, Förstner (1985) considered model (2.44) for univariate observations $X_t$ such that $x_t$ followed a $p$-th order autoregressive model

$$X_t = x_t + u_t$$

(2.44)

$$\sum_{j=0}^{p} \phi_j x_{t-j} = g_t$$

and $\{u_t\}$ was a sequence of uncorrelated $(0, \sigma_{uu})$ random variables which were independent of the $x_t$. Under the assumption that the autoregressive coefficients $\{\phi_j, j=0, 1, \ldots, p\}$ were known a priori, Förstner used an estimated power spectrum to develop an iterative procedure for the estimation of $\sigma_{uu}$ and $\sigma_{gg} = \text{Var}(g_t)$.

The time- and frequency-domain approaches to time series analysis are not entirely distinct. For example, theoretical discussion of time-domain estimators may be facilitated by frequency-domain arguments. [See, for example, Theorem IV.8 of Hannan (1970, p. 216).] Similarly, a single data analysis may rely on both time- and frequency-domain estimators. Nonetheless, this dissertation will restrict itself to the time domain analysis for the following reasons. First, comments by Hannan (1970, p. 325), Anderson (1971, p. 549), Tukey (1980), Nerlove
and Pinto (1984) and others indicate that time-domain estimation methods may in general have greater small-sample efficiency than frequency-domain methods. Second, a number of the time-domain methods suggested below are fairly direct extensions of methods proposed previously for estimation of uncorrelated measurement error models. Additional research is required to develop a careful comparison of the frequency-domain methods reviewed in this section and the time-domain methods discussed in Chapters 4, 5 and 6 below.
3. ASYMPTOTIC PROPERTIES OF SUMS OF LINEAR AND BILINEAR FUNCTIONS OF CORRELATED OBSERVATIONS

The study of multiple time series often requires one to consider the asymptotic properties of regression estimators and of sample covariances; see, e.g. Hannan (1970, pp. 215-229, 415-448), Hannan and Heyde (1972), and Hannan (1976). Results for regression estimators are special cases of the behavior of a linear function of a sequence of serially correlated random vectors. Similarly, results for sample covariances are special cases of the properties of a bilinear function of a sequence of serially correlated random vectors. Examination of the estimators of Chapter 4 requires one to consider the asymptotic behavior of the sum of a regression estimator and a sample covariance, or, more generally, the sum of a linear function and a bilinear function of a sequence of serially correlated random vectors. Section 3.1 reviews some necessary probabilistic background. Section 3.2 addresses the asymptotic behavior of the sum of a sample mean and a sample variance of a univariate time series. Section 3.3 extends the results of the previous section to the sum of a linear function and an unweighted bilinear function of a vector process. The results of Section 3.3 will be applied in Section 4.1 to determine the asymptotic behavior of some standard errors-in-variables estimators in the presence of serially correlated observations. Finally, Section 3.4 addresses weighted linear and bilinear functions; Section 4.2 will apply the results of this
section to establish the asymptotic properties of weighted errors-in-
variables estimators of regression parameters.

3.1. Definitions and Notation for Vector Martingales

The results presented below develop the asymptotic properties of
quadratic and linear forms of vector martingales. Following Billingsley
(1979, p. 407), let \{Z_t, t > 1\} be a sequence of \(1 \times p\) random
vectors defined on the probability space \((\Omega, F, P)\), let \{\(F_t, t > 1\)\}
be a sequence of sub-\(\sigma\)-fields of \(F\), and define the sequence
\(\{(Z_t, F_t), t > 1\}\) to be a martingale if for all \(t=1, 2, \ldots\),

1. \(F_t \subseteq F_{t+1}\); \hspace{1cm} (3.1.a)

2. \(Z_t\) is measurable \(F_t\); \hspace{1cm} (3.1.b)

3. \(E(|Z_{ti}|) < \infty, i=1, 2, \ldots, p\); \hspace{1cm} (3.1.c)

4. \(E(Z_{t+1}|F_t) = Z_t\) almost surely. \hspace{1cm} (3.1.d)

For \(p = 1\), define \(\{(Z_t, F_t), t > 1\}\) to be a submartingale if
conditions (3.1.a)-(3.1.c) hold and

\[E[Z_{t+1}|F_t] > Z_t\] \hspace{1cm} (3.1.e)

almost surely for \(t=1, 2, \ldots\). Note that (3.1.d) implies (3.1.e), so
that any martingale is a submartingale. A slight generalization of Theorem 35.1 of Billingsley (1979) may be stated as follows.

**Lemma 3.1.** Let \( \{Z_t^k\} \) be a sequence of \( 1 \times p \) random vectors defined on \((\Omega, F, P)\); let \( D \) be a convex subset of \( \mathbb{R}^p \) such that \( P[Z_t^k \in D] = 1 \) for all \( t = 1, 2, ... \); and let \( \phi: D \to \mathbb{R}^1 \) be a convex function such that \( \phi(Z_t^k) \) is integrable for all \( t \).

1. If \( \{(Z_t^k, F_t^k)\} \) is a martingale, then \( \{\phi(Z_t^k), F_t^k\} \) is a submartingale.
2. If \( p = 1 \), \( \{(Z_t, F_t)\} \) is a submartingale, and \( \phi(\cdot) \) is nondecreasing on \( D \), then \( \{\phi(Z_t), F_t\} \) is a submartingale.

Lemma 3.1.a is a vector generalization of Theorem 35.1(i) of Billingsley, but has essentially the same proof by Jensen's inequality. Lemma 3.1.b is identical to Theorem 35.1(ii) of Billingsley (1979).

Theorem 35.4 of Billingsley (1979) provides a useful tool for proving the convergence of functions of martingale differences, and is stated below for reference.

**Lemma 3.2.** [Theorem 35.4 of Billingsley (1979).] Let \( \{(Z_t, F_t)\} \) be a submartingale. If \( K = \sup_t E[|Z_t|] < \infty \), then \( Z_t + Z \) almost surely, where \( Z \) is a random variable satisfying \( E[|Z|] < K \).

A multivariate generalization of Lemma 3.2 for martingales is given in Lemma 3.3.
Lemma 3.3. Let \((Z_t, F_t)\) be a \(p\)-dimensional martingale such that

\[
K = \max \sup \{E[|Z_t|] \mid 1 \leq i \leq p\} < \infty.
\]

Then \(Z_t + Z\) with probability one, where \(Z = (Z_1, Z_2, \ldots, Z_p)\) is a \(1 \times p\) random vector such that \(E[|Z_i|] < K\) for all \(i = 1, 2, \ldots, p\).

Proof. If \((Z_t, F_t)\) is a martingale, then \((Z_t^i, F_t)\) is a submartingale for each \(i = 1, 2, \ldots, p\), so Lemma 3.2 implies that

\[
\lim_{t \to \infty} Z_t^i(w) = Z_i(w) \quad \text{for all } w \in A_i \subseteq \Omega, \quad \text{where } P(A_i) = 1 \text{ and } E[|Z_i|] < K.
\]

Let \(A = \bigcap_{i=1}^p A_i\). Then \(P(A) = 1\), and \(w \in A\) implies that \(\lim_{t \to \infty} Z_t^i(w) = Z_i(w)\), \(i = 1, 2, \ldots, p\), so the result follows.

For a martingale \((Z_t, F_t)\), one may define the martingale difference sequence \(\{d_t = Z_t - Z_{t-1}, t = 2, 3, \ldots\}\). One may note from Billingsley (1979, p. 408) that for all \(t\),

\[
\sigma(d_1, d_2, \ldots, d_t) = \sigma(Z_1, Z_2, \ldots, Z_t),
\]

so \(d_t\) is measurable \(F_t\); that for all \(i = 1, 2, \ldots, p\) and all \(t = 1, 2, \ldots\), \(E(|d_{ti}|) < \infty\); and that for all \(t = 2, 3, \ldots\),

\[
E(d_t \mid F_{t-1}) = E(Z_t \mid F_{t-1}) - Z_{t-1}
\]

\[
= Z_t - Z_{t-1}
\]

\[
= 0_{1 \times p}.
\]
Conversely, it follows from the remarks in Billingsley (1979, p. 408) that for any sequence \( \{d_t\} \) which satisfies the conditions of the preceding sentence, if

\[
Z_t \equiv \sum_{s=1}^{t} d_s, \quad t=1, 2, \ldots,
\]

then the sequence \( \{Z_t, F_t\} \) forms a martingale.

The sequences of random vectors studied below are functions of a special type of martingale difference sequence. Let \( \{c_t\} \) be a set of \( 1 \times p \) random vectors defined on a common probability space \( (\Omega, F, P) \). Let \( F_t = \sigma(c_s, s < t) \), the sub-\( \sigma \)-algebra of \( F \) generated by the random vectors \( \{c_s, s < t\} \) and assume that the following conditional expectations exist and are equal to finite constant matrices, as indicated, for all integers \( t \):

\[
\begin{align*}
E[c_t | F_{t-1}] &= 0_{1 \times p} \quad (3.2.a) \\
E[c_t' c_t | F_{t-1}] &= \Sigma_{cc} \quad (3.2.b) \\
E[c_t' c_t | F_{t-1}] &= M_3 \quad (3.2.c) \\
E[c_t' c_t | F_{t-1}] &= K. \quad (3.2.d)
\end{align*}
\]

One may infer several moment properties of the sequence \( \{c_t\} \) from conditions (3.2) above. If condition (3.2.a) holds, then
If conditions (3.2a)-(3.2b) hold, then

\[ E(c'_t c_t) = E[E( c_t | F_{t-1} )] \]

\[ = \Sigma_{cc} \] \hspace{1cm} (3.4)

and for \( s < t \),

\[ E(c'_s c_t) = E[E(c'_s c_t | F_s)] \]

\[ = E(c'_s E(c_t | F_s)) \]

\[ = E(c'_s E(E(c'_t | F_{t-1}) | F_s)) \]

\[ = E(c'_s 0) \]

\[ = 0 \]

by Theorem 34.4 of Billingsley (1979). Thus, the sequence \( \{ c_t \} \) satisfying conditions (3.2.a)-(3.2.b) is a sequence of uncorrelated \( (0, \Sigma_{cc}) \) vectors. Under conditions (3.2.a)-(3.2.d), similar conditioning arguments establish that
\[ E[c'_r \circ c'_s \circ c'_t] = \begin{cases} 0 & \text{if } r = s = t \\ M_3 & \text{otherwise} \end{cases} \quad (3.5.a) \]

and

\[ E[c'_s \circ c'_s \circ c'_t \circ c'_t] = \begin{cases} K & \text{if } s = t \\ \Sigma_{cc} & \Sigma_{cc} \text{ otherwise} \end{cases} \quad (3.5.b) \]

This final result implies that \( \{c'_s \circ c'_s - \Sigma_{cc}\} \) is a sequence of uncorrelated random matrices with mean zero and common finite covariance matrix.

Lemma 3.4 shows that the sample mean of a sequence \( \{c^t\} \) satisfying conditions (3.2.a)-(3.2.b) converges to zero with probability one.

**Lemma 3.4.** Let \( \{c^t\} \) be a sequence of \( 1 \times p \) random vectors that satisfy conditions (3.2.a)-(3.2.b). Then

\[ \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} c^t = 0_{1 \times p} \]

with probability one.

**Proof.** As noted above, conditions (3.2.a)-(3.2.b) imply that \( \{c^t\} \) is a sequence of uncorrelated \( (0, \Sigma_{cc}) \) random vectors. The result then follows by Theorem 5.1.2 of Chung (1974).
Lemma 3.5 establishes that the sample mean of the sequence \{c_t'c_t\} converges to \(\Sigma_{cc}\) with probability one. Its proof follows closely that of Lemma 3.4 and, therefore, is not given.

**Lemma 3.5.** Let \(\{c_t\}\) be a sequence of \(1 \times p\) random vectors that satisfy conditions (3.2.a)-(3.2.d). Then

\[
\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} c_t'c_t = \Sigma_{cc}
\]

with probability one.

**Proof.** Omitted. 

Some correlated time series may be written as infinite moving averages of random vectors \(\{c_t\}\) satisfying (3.2). The following two definitions allow one to discuss such moving averages more precisely. These definitions are adapted from Fuller (1976, p. 72) and Hannan (1970, p. 209).

**Definition 3.1.** [c.f. Definition 2.8.1 of Fuller (1976)].

Let \(\{A_h, h \in \mathbb{Z}\}\) be an infinite sequence of \(p \times p\) real matrices such that \(A_h\) has \((i,j)\)-th element equal to \(A_{hij}\), \(1 \leq i, j \leq p\), \(h \in \mathbb{Z}\). Define the sequence \(\{A_h\}\) to be absolutely summable if

\[
\sum_{h=-\infty}^{\infty} |A_{hij}| < \infty
\]

for all \(1 \leq i, j \leq p\).
Definition 3.2. [c.f. expression (IV.3.4) of Hannan (1970, p. 209)]. A sequence of $1 \times p$ random vectors $\{e_t\}$ is a linear process if it can be written in the form

$$e_t' = \sum_{h=0}^{\infty} A_h c_{t-h}'$$

for some sequence of $1 \times p$ random vectors $\{c_t\}$ satisfying conditions (3.2) and some absolutely summable sequence of $p \times p$ real matrices $\{A_h, h=0, 1, 2, \ldots\}$.

Let $B$ be a $p \times p$ real matrix and let $e_t' = \sum_{h=0}^{\infty} A_h c_{t-h}'$ be a $p$-dimensional stochastic process that satisfies Definition 3.2. Each element of $B$ is finite, so $\{(B A_h), h=0, 1, 2, \ldots\}$ is an absolutely summable sequence, and thus

$$B e_t' = \sum_{h=0}^{\infty} (B A_h) c_{t-h}'$$

is also a $p$-dimensional linear process. Thus, a $p \times p$ linear transformation of a linear process is itself a linear process. As noted in Hannan (1970, p. 211), if $e_t'$ is a linear process, then it is fourth-order stationary such that

$$\Gamma_{e'e'}(\ell) = \text{Cov}(e_t', e_{t+\ell}') = \sum_{j=-\infty}^{\infty} A_j A_{j+\ell} A'_{j+\ell}$$

for $\ell > 0$; $\Gamma_{e'e'}(\ell) = [\Gamma_{e'e'}(-\ell)]'$ for $\ell < 0$; and $\{\Gamma_{e'e'}(\ell)\}$ is absolutely summable; the sequence $\{e_t\}$ has an absolutely continuous
spectrum with continuous spectral density

\[ f_{\xi}(\lambda) \equiv (2\pi)^{-1} \sum_{\xi=\infty}^{\infty} |G_\xi(\lambda)| \exp(-i\lambda \xi); \]

and the covariance function and the third-order and fourth-order cumulant functions of \( \{\xi_n\} \) are absolutely summable. Thus for all \( 1 \leq i, j, k, m \leq p \),

\[ \sum_{t=-\infty}^{\infty} |G_{\xi i j}(t)| < \infty, \quad (3.6.a) \]

\[ \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} |M_{\xi i j k}(0, s, t)| < \infty, \quad (3.6.b) \]

and

\[ \sum_{q=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} |K_{\xi i j k m}(0, q, s, t)| < \infty, \quad (3.6.c) \]

where the third-cumulant function \( \{M_{\xi i j}(0, s, t)\} \) is a doubly-infinite matrix sequence of dimension \( p^2 \times p \) defined by

\[ M_{\xi i j k}(0, s, t) = E(\xi_i \xi_j \xi_k), \quad (3.7.a) \]

\( 1 \leq i, j, k \leq p \); \( s, t \in \mathbb{Z} \); and the fourth-cumulant function \( \{K_{\xi i j k}(0, q, s, t)\} \) is a three-fold infinite matrix sequence of dimension \( p^2 \times p^2 \) defined by
Given the definitions and results presented above, the following three sections develop some properties of linear and bilinear functions of vector martingale difference sequences.

3.2. Asymptotic Properties of the Sum of a Sample Mean and a Sample Variance

The following three sections address the asymptotic properties of the sum of a linear function and a bilinear function of a sequence of serially correlated random vectors. This section considers the simple example of such a sum defined by

\[
\tilde{\epsilon} + \tilde{\Gamma}(0),
\]

where

\[
\tilde{\epsilon} = T^{-1} \sum_{t=1}^{T} \epsilon_t,
\]

\[
\tilde{\Gamma}(0) = T^{-1} \sum_{t=1}^{T} \epsilon_t^2.
\]
and \( \{ \varepsilon_t \} \) is a sequence of serially correlated random variables. Although the behavior of \( \varepsilon + \tilde{\Gamma}(0) \) is of relatively little intrinsic interest, the study of this case provides an introduction to several ideas and arguments that will be required to establish asymptotic results for more complicated functions of correlated random vectors.

Assume that \( \{ \varepsilon_t \} \) is a univariate linear process that satisfies Definition 3.2 and let \( E(\varepsilon \varepsilon') = \Gamma \) be a \( T \times T \) matrix with \((s,t)\)-th element equal to \( \Gamma(t-s) \). Then \( E(\varepsilon^2_t) = 0 \) and \( E(\varepsilon_t^2) = \Gamma(0) \), so

\[
E[\varepsilon + \tilde{\Gamma}(0)] = \Gamma^{-1} \left[ \sum_{t=1}^{T} E(\varepsilon_t) + \sum_{t=1}^{T} E(\varepsilon_t^2) \right]
\]

\[
= \Gamma(0) .
\]

By the fourth-order stationarity of \( \{ \varepsilon_t \} \), one may define the third-cumulant function,

\[
M(0, 0, t) = E(\varepsilon_s \varepsilon_s \varepsilon_{s+t}) ,
\]

and fourth-cumulant function,

\[
k(0, 0, t, t) = E(\varepsilon_s \varepsilon_s \varepsilon_{s+t} \varepsilon_{s+t}) - [\Gamma(0)]^2 - 2[\Gamma(t)]^2 .
\]

These functions are special univariate cases of the third and fourth cumulant functions defined in expressions (3.7). Then
\[
\text{Var}(\bar{e}) = T^{-2} \prod_{t=-T}^{T} \Gamma(t)
\]

\[
= T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \Gamma(t - s)
\]

\[
= T^{-2} \sum_{t=-T+1}^{T-1} (T - |t|) \Gamma(h)
\]

\[
\text{Cov}[\bar{r}(0), \bar{e}] = T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \text{Cov}(e_s^2, e_t^2)
\]

\[
= T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} E(e_s^2 e_t^2)
\]

\[
= T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} M(0, 0, t-s)
\]

\[
= T^{-1} \sum_{t=-T+1}^{T-1} (T - |t|) M(0, 0, t)
\]

and

\[
\text{Var}[\bar{r}(0)] = T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \text{Cov}(e_s^2, e_t^2)
\]

\[
= T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \{2 \Gamma(t - s)^2 + k(0, 0, t-s, t-s)\}
\]

\[
= T^{-2} \sum_{t=-T+1}^{T-1} (T - |t|) \{2 \Gamma(t)^2 + k(0, 0, t, t)\}.
\]

Thus,

\[
\text{Var}[\bar{e} + \bar{r}(0)] = \text{Var}(\bar{e}) + 2 \text{Cov}(\bar{e}, \bar{r}(0)) + \text{Var}[\bar{r}(0)]
\]
\[
Z(t) = \frac{1}{T} \sum_{t=-T+1}^{T} \left( T - |t| \right) \left( \Gamma(t) + 2 \Pi(0, 0, t) + 2|\Gamma(t)|^2 + k(0, 0, t, t) \right).
\]

Kronecker's Lemma [Fuller (1976), p. 109] and the absolute summability of \( \Gamma(t) \), \( M(0, 0, t) \) and \( k(0, 0, t, t) \) imply that

\[
\lim_{T \to \infty} \frac{T}{T} \left[ \varepsilon + \tilde{\Gamma}(0) \right] = \lim_{T \to \infty} \sum_{t=-T+1}^{T} \left( 1 - \frac{1}{T} |t| \right) \left( \Gamma(t) + 2 \Pi(0, 0, t) + 2|\Gamma(t)|^2 + k(0, 0, t, t) \right)
\]

\[
= \sum_{t=-\infty}^{\infty} \left( \Gamma(t) + 2 \Pi(0, 0, t) + 2|\Gamma(t)|^2 + k(0, 0, t, t) \right). \quad (3.9)
\]

Now consider the asymptotic distribution of

\[
T^{-1/2} \left[ \varepsilon + \tilde{\Gamma}(0) - \Gamma(0) \right].
\]

In addition to the moment and martingale conditions presented above, asymptotic normality results in the time series literature generally require an additional assumption. Assumptions that have been made include independence and an identical distribution for the \( c_t \) variables [c.f. Hannan (1970, pp. 209, 228)]; ergodicity of the \( \epsilon_t \) process [c.f. Hannan (1970), p. 201] and Hannan (1976)]; or the existence of bounded \( 4 + 2v \) moments of the \( c_t \) variables [c.f. Fuller, Hasza and Goebel (1981)]. The work below will employ the latter.
moment assumption. Given the condition of the finite $4 + 2\nu$ moments, the following lemma will be useful in establishing Lindeberg-type conditions for random sequences.

**Lemma 3.6.** Let $X$ be a random variable defined on a probability space $(\Omega, F, P)$ such that $E[|X|^{2+\nu}] < \infty$ for some $\nu > 0$. Let $c$ be some positive real number.

a. Let $F_1$ be a sub-$\sigma$-algebra of $F$. Then

$$E[X^2 I(|X| > c) | F_1] < c^{-\nu} E[|X|^{2+\nu} | F_1]$$

with probability one.

b. In addition,

$$E[X^2 I(|X| > c)] < c^{-\nu} E[|X|^{2+\nu}] .$$

**Proof.** If $|X| > c > 0$, then $c^\nu < |X|^\nu$ and $X^2 < c^{-\nu} |X|^{2+\nu}$. Thus,

$$X^2 I(|X| > c) < c^{-\nu} |X|^{2+\nu} I(|X| > c) < c^{-\nu} |X|^{2+\nu}$$

with probability one. Then by Theorem 34.2 (iii) of Billingsley (1979),

$$E[X^2 I(|X| > c) | F_1] < c^{-\nu} E[|X|^{2+\nu} | F_1]$$

with probability one, so part (a) is established. Part (b) then follows from Theorem 34.4 of Billingsley (1979). □
In the study of the asymptotic behavior of a sequence of random variables \( \{T_e; T \in \mathbb{Z}^+\} \), say, it is often useful to direct initial attention to a doubly infinite array of random variables \( \{M, T \in \mathbb{Z}^+\} \). If \( S_{MT} \xrightarrow{P} S_M \) as \( T \to \infty \); if \( S_M \xrightarrow{P} S_\ldots \) as \( M \to \infty \); and if \( D_{MT} \equiv (T_e - S_{MT}) \xrightarrow{P} 0 \) uniformly in \( T \) as \( M \to \infty \); then one may argue that \( T_e \xrightarrow{P} S_\ldots \) as \( T \to \infty \). This partial extension of Slutsky's Theorem [Property A.14.9 of Bickel and Doksum (1977, p. 461)] is stated and proved more formally by Fuller (1976, p. 248) and is repeated below for convenient reference.

**Lemma 3.7.** [Lemma 6.3.1 of Fuller (1976, p. 248)].

Let the random variables \( T_e \) be defined by

\[
T_e = S_{MT} + D_{MT}
\]

for \( M \in \mathbb{Z}^+ \) and \( T \in \mathbb{Z}^+ \). Let

\[
\text{plim}_{M \to \infty} D_{MT} = 0 \quad (3.10.a)
\]

uniformly in \( T \),

\[
S_{MT} \xrightarrow{P} S_M \quad \text{as} \quad T \to \infty \quad (3.10.b)
\]

and

\[
S_M \xrightarrow{P} S_\ldots \quad \text{as} \quad M \to \infty \quad (3.10.c)
\]
Then

\[ \xi_T \xrightarrow{f} \xi, \quad \text{as } T \to \infty. \quad (3.11) \]

**Proof.** See Fuller (1976, pp. 248-249).

The following lemma uses Lemmas 3.6 and 3.7 to establish the asymptotic normality of

\[ T^{1/2} \left[ \tilde{\varepsilon} + \tilde{\Gamma}(0) - \Gamma(0) \right]. \]

**Lemma 3.8.** Let \( \varepsilon_t \) be generated by a univariate linear process,

\[ \varepsilon_t = \sum_{h=0}^{\infty} a_h \varepsilon_{t-h}, \quad (3.12) \]

where the real-valued weights \( a_h \) and the sequence of random variables \( \{ \varepsilon_t \} \) satisfy the conditions of Definition 3.2. Assume that \( \sigma_{\varepsilon \varepsilon} > 0 \) and that for some \( \nu > 0 \),

\[ \sup_t \{ \mathbb{E}[|\varepsilon_t|^{4+2\nu}] \} \leq K < \infty. \quad (3.13) \]

Then

\[ T^{1/2} \left[ \varepsilon + \tilde{\Gamma}(0) - \Gamma(0) \right] \xrightarrow{f} \mathcal{N}(0, \sigma^2) \quad (3.14) \]

as \( T \to \infty \), where
The proof of Lemma 3.8 has the same structure as the proofs of more complicated central limit results presented in Sections 3.3 and 3.4, so it is useful to outline this proof before presenting it in detail.

First, result (3.14) is established for sequences \( \{ \epsilon_t \} \) which follow a finite moving average model,

\[
\epsilon_t = \sum_{h=0}^{M} a_h c_{t-h},
\]

say. Under the finite moving average model, denote the sample and population moments with the subscript \( M \), e.g., \( \Gamma_M(0) = \mathbb{E}(\epsilon_t^2) \). Some preliminary arguments show that under the finite moving-average model,

\[
T[\bar{\epsilon}_M + \Gamma_M(0) - \Gamma_M(0)] = \sum_{t=1}^{T} [Z_t c_t - \Gamma_M(0)] + o_p(T^{1/2}),
\]

where \( Z_t = e_M + f_t + g_t \), \( e_M \) is a constant, \( f_t \) is a \( \sigma(c_t) \)-measurable random variable, and \( g_t \) is a \( \sigma(c_s, s < t-1) \)-measurable random variable. The notational complexity introduced by the terms \( e_M, f_t, \) and \( g_t \) is not necessary at this point, but it will be required for the proofs of the central limit results of Sections 3.3 and 3.4. Examination of the properties of \( (e_M c_t), (f_t c_t), \) and \( (g_t c_t) \) indicate that the sequence \( \{ Z_t c_t - \Gamma_M(0), t=1, 2, \ldots \} \) satisfies the two conditions of Theorem 1 given in Scott (1973). Theorem 1 of Scott (1973) then establishes the asymptotic normality of
\[ T^{-\frac{1}{2}} \sum_{t=1}^{T} [Z_t c_t - \Gamma_M(0)] \] and hence, of \[ T^{\frac{1}{2}} [\hat{c} + \hat{\Gamma}_M(0) - \Gamma_M(0)] \] when \( \{c_t\} \) follows a finite moving average model.

Next, one may note that given the same \( \{c_t\} \) process and an extended set of absolutely summable weights \( \{a_h\} \), then uniformly in \( T \), the difference between \[ T^{\frac{1}{2}} [\hat{c} + \hat{\Gamma}_M(0) - \Gamma(0)] \] under the infinite moving average model (3.12) and \[ T^{\frac{1}{2}} [\hat{c}_M + \hat{\Gamma}_M(0) - \Gamma_M(0)] \] under the finite moving average model (3.15) becomes asymptotically negligible as \( M \) increases without bound. The asymptotic normality of \[ T^{\frac{1}{2}} [\hat{c} + \hat{\Gamma}_M(0) - \Gamma(0)] \] under model (3.12) then follows from the finite moving average result and Lemma 3.7.

**Proof of Lemma 3.8.** Consider first the case in which \( c_t \) has a finite moving average representation,

\[ c_t = \sum_{h=0}^{M} a_h c_{t-h}, \quad (3.16) \]

where \( a_0 \neq 0 \) and \( a_M \neq 0 \). Fix \( M \) temporarily. Under model (3.16), denote the moments and other parameters of the \( \{c_t\} \) process with the subscript \( M \), e.g.,

\[ \Gamma_M(\tau) = \text{Cov} \left( \sum_{h=0}^{M} a_h c_{t-h}, \sum_{h=0}^{M} a_h c_{t+\tau-h} \right), \]

\[ = \sigma \sum_{h=0}^{M} a_h^2, \quad \tau = 0, 1, \ldots, M. \] Similarly, let
\[ \bar{e}_M = T^{-1} \sum_{t=1}^{T} \sum_{h=0}^{M} ( \sum_{c} a_{h} c_{t-h} ) \]

and

\[ \bar{\tilde{\tau}}_M(0) = T^{-1} \sum_{t=1}^{T} \sum_{h=0}^{M} ( \sum_{c} a_{h} c_{t-h} )^2 \]

i.e. the first and second sample moments, respectively, of \( e_t \) under model (3.16). Let \( b = (M + 1)^{-1} \). Then under model (3.16),

\[ T[\bar{e}_M + \bar{\tilde{\tau}}_M(0)] = \sum_{t=1}^{T} (1 + e_t) e_t \]

\[ = \sum_{t=1}^{T} \sum_{g=0}^{M} \sum_{h=0}^{M} (b + a_{c_t-g} a_{c_{t-h}} + \sum_{g=1}^{M-1} \sum_{h=0}^{M} (b + a_{c_t-g} a_{c_{t-h}}) + \sum_{g=0}^{M-1} \sum_{h=g+1}^{M} (b + a_{c_t-g} a_{c_{t-h}}) ) \]

\[ = \sum_{t=1}^{T} \sum_{h=0}^{M} (b + a_{c_t-h} a_{c_{t-h}} + \sum_{g=1}^{M-1} \sum_{h=0}^{M} (b + a_{c_t-g} a_{c_{t-h}}) + \sum_{g=0}^{M-1} \sum_{h=g+1}^{M} (b + a_{c_t-g} a_{c_{t-h}}) ) \]

because

\[ \sum_{g=1}^{M-1} \sum_{h=0}^{M} (b + a_{c_t-g} a_{c_{t-h}}) = \sum_{g=1}^{M-1} \sum_{h=0}^{M} (b + a_{c_t-g} a_{c_{t-h}}) \]
and

\[
\sum_{g=0}^{M-1} \sum_{h=g+1}^{M-1} (b + ac_{t-g})a_{h}c_{t-h}
\]

\[
= \sum_{g=0}^{M-1} \sum_{h=g+1}^{M-1} ba_{h}c_{t-h} + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} a_{h}c_{t-h}a_{g}c_{t-g}
\]

\[
= \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} hba_{h}c_{t-h} + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} a_{h}c_{t-h}a_{g}c_{t-h}
\]

\[
= \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} [hb + a_{h}c_{t-g}]a_{h}c_{t-h} + Mba_{M}c_{t-M}.
\]

Then

\[
T[\bar{\varepsilon}_{M} + \bar{P}_{M}(0) - \Gamma_{M}(0)] = T^{-1} M \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} \{ (b + ac_{t})a_{h} + \sum_{g=0}^{h-1} \sum_{t=1}^{T} \sum_{h=0}^{T} (b + ac_{t+g})a_{h}
\]

\[
+ hba_{h} + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M-1} a_{h}c_{t+g}a_{h} + Mba_{M}c_{t} + (T - M) \Gamma_{M}(0)
\]

\[
= T^{T-t} + \sum_{t=T-M+1}^{T-t-1} \sum_{h=0}^{T} (b + ac_{t})a_{h} + \sum_{h=0}^{T} \sum_{g=h+1}^{M-1} (b + ac_{t+g})a_{h}
\]

\[
+ (T-t)ba_{h} + \sum_{h=0}^{T} \sum_{g=h+1}^{M-1} a_{h}c_{t+g}a_{h} + (T - t)ba_{T-t}c_{h} - M_{T}(0).
\]

Since \( M \) is finite and is not a function of \( T \), expression (3.18) is \( o_{p}(T^{1/2}) \), so
\[
\sum_{t=1}^{T} [e_t + e_t^2 - \Gamma_M(0)] = \sum_{t=1}^{T} [Z_t c_t - \Gamma_M(0)] + o_p(T^{1/2})
\] (3.19)

where

\[ Z_t = e_M + f_t + g_t, \]
\[ e_M = b(M + 1) \sum_{h=0}^{M} a_h = \sum_{h=0}^{M} a_h, \]
\[ f_t = c_t \sum_{h=0}^{M} a_h^2, \]
\[ g_t = 2 \sum_{h=0}^{M} \sum_{h=0}^{M-h} c_{t-h} a_h^2. \]

Now \( e_M \) is fixed, \( f_t \) is a mean-zero random variable that is measurable \( \sigma(c_t) \), and \( g_t \) is a mean-zero random variable that is measurable \( \mathbb{F}_{t-1} \). Martingale conditions (3.2) imply that

\[ \mathbb{E}[e_M c_t | \mathbb{F}_{t-1}] = 0; \]
\[ \mathbb{E}[f_t c_t | \mathbb{F}_{t-1}] - \Gamma_M(0) = \sum_{h=0}^{M} a_h^2 \sigma_{c_h} - \Gamma_M(0). \]
because

\[ \Gamma_{M}(0) = \text{Var}(e_t) \]

\[ = \text{Cov}(\Sigma_{g=0}^{M} a_{g} c_{t-g}, \Sigma_{h=0}^{M} a_{h} c_{t-h}) \]

\[ = \Sigma_{h=0}^{M} a_{h}^{2} \sigma_{c} ; \]

and

\[ E[g_{t} c_{t} | F_{t-1}] = g_{t} E[c_{t} | F_{t-1}] \]

\[ = 0 . \]

Also, conditions (3.2) imply that the following moments are constant over \( t \):

\[ V_{M_{f} c_{c}} = \text{Cov}(c_{t}, f_{t} c_{t}) \]

\[ = E(c_{t}^{2} f_{t}) \]

\[ = E(c_{t}^{3} \Sigma_{h=0}^{M} a_{h}^{2}) \]
\[ V_{Mcf} = \text{Var}(f_{c_t} c_t) \]

\[ = (\sum_{h=0}^{M} a_h^2)^2 \text{Var}(c_t^2) \]

\[ = (\sum_{h=0}^{M} a_h^2)^2[K - \sigma_{cc}^2] ; \]

and

\[ V_{Mgg} = E[g_t^2] \]

\[ = 4\sigma_{cc}^2 \sum_{h=1}^{M-1} (\sum_{h=0}^{M-1} a_{h+z} a_h)^2 . \]

Note that for \( z = M \),

\[ \sum_{h=0}^{M-1} a_{h+z} a_h = a_M a_0 \neq 0 \]

because \( a_M \neq 0 \) and \( a_0 \neq 0 \) by assumption. Also, \( \sigma_{cc} > 0 \), so

\[ V_{Mgg} > 0 . \]

In addition, one may apply martingale conditions (3.2) to find that

\[ d^2_{Mtt} = E[(Z_{c_t} c_t - \Gamma_M(0))^2|F_{t-1}] \]

\[ = \text{Var}[Z_{c_t} c_t|F_{t-1}] \]
\[= \text{Var}[\{e_t + f_t + g_t\} | F_{t-1}]\]

\[= \sigma_{cc}(e_t + g_t)^2 + 2(e_t + g_t)\text{Cov}(c_t, f_t c_t) + \text{Var}(f_t c_t)\]

\[= \sigma_{cc}(e_t + g_t)^2 + 2(e_t + g_t)V_{Mcf} + V_{Mcf},\]

so

\[E\{[Z_{tc} - \Gamma_M(0)]^2\} = E(E\{[Z_{tc} - \Gamma_M(0)]^2 | F_{t-1}\})\]

\[= \sigma_{cc}(e_t^2 + V_{Mgg}) + 2e_t V_{Mcf} + V_{Mcf}.\]

The equality \(E[Z_{tc} - \Gamma_M(0) | F_{t-1}] = 0\) implies that

\[E\left(\sum_{s=1}^{t} [Z_{tc} - \Gamma_M(0)] | F_{t-1}\right) = \sum_{s=1}^{t} [Z_{tc} - \Gamma_M(0)]\]

and

\[\text{Var}\left(\sum_{s=1}^{t} [Z_{tc} - \Gamma_M(0)] | F_{t-1}\right) = \text{Var}[Z_{tc} - \Gamma_M(0) | F_{t-1}],\]

so

\[s^2_{Mtt} \equiv \text{Var}\left(\sum_{s=1}^{t} [Z_{tc} - \Gamma_M(0)]\right)\]

\[= E\{\text{Var}[Z_{tc} - \Gamma_M(0) | F_{t-1}]\} + \text{Var}\left(\sum_{s=1}^{t-1} [Z_{tc} - \Gamma_M(0)]\right)\]
\[ t = \sum_{s=1}^{t} \mathbb{E}\{\text{Var}[Z_s c_s - \Gamma_M(0)|F_{s-1}]\} \]

\[ = \sum_{s=1}^{t} \left[ \sigma_{cc}^2 (e_M^2 + V_{Mgg}) + 2e_M V_{Mcf} + V_{Mcf}\right] \]

\[ = t\left[ \sigma_{cc}^2 (e_M^2 + V_{Mgg}) + 2e_M V_{Mcf} + V_{Mcf}\right] . \quad (3.20) \]

As noted above, \( V_{Mgg} > 0 \); and \( \sigma_{cc} > 0 \) by assumption. Also,

\[ \varnothing < \text{Var}[(e_M + f_t)Z_t] \]

\[ = \text{Var}(e_M Z_t + f_t Z_t) \]

\[ = \text{Var}(e_M F_t) + 2e_M \text{Cov}(f_t, Z_t) + \text{Var}(f_t Z_t) \]

\[ = e_M^2 \sigma_{cc}^2 + 2e_M V_{Mcf} + V_{Mcf} . \]

Thus,

\[ s_{M}^2 = \left[ \sigma_{cc}^2 (e_M^2 + V_{Mgg}) + 2e_M V_{Mcf} + V_{Mcf}\right] > 0 . \quad (3.21) \]

Now the sequence \( \{Z_t c_t - \Gamma_M(0)\} \) will satisfy the assumptions of Theorem 1 of Scott (1973) if the following conditions hold:

\[ s_{M}^{-2} \mathbb{E}_{MTT} \sum_{t=1}^{T} \mathbb{E}\{[Z_t c_t - \Gamma_M(0)]^2|F_{t-1}\} \rightarrow 1 \text{ as } T \rightarrow \infty, \quad (3.22) \]
and

\[ s_{\text{MTT}}^{-2} \sum_{t=1}^{T} \mathbb{E}[(Z_{t}c_{t} - \Gamma_{M}(0))^{2} | F_{t-1}] > \epsilon s_{\text{MTT}} | F_{t-1} \] \[ \rightarrow 0 \]

as \( T \to \infty \). \hspace{1cm} (3.23)

Result (3.19) implies that

\[ s_{\text{MTT}}^{-2} \sum_{t=1}^{T} \mathbb{E}[(Z_{t}c_{t} - \Gamma_{M}(0))^{2} | F_{t-1}] - 1 \]

\[ = s_{M}^{-2} T^{-1} \sum_{t=1}^{T} \left( \mathbb{E}[(Z_{t}c_{t} - \Gamma_{M}(0))^{2} | F_{t-1}] - s_{M}^{2} \right) \]

\[ = s_{M}^{-2} T^{-1} \sum_{t=1}^{T} \left[ \sigma_{cc}(e_{M} + g_{t})^{2} + 2(e_{M} + g_{t})V_{Mcf} + V_{Mcf} \right. \]

\[ - \sigma_{cc}(e_{M}^{2} + V_{Mg}) - 2e_{M}V_{Mcf} - V_{Mcf} \]

\[ = s_{M}^{-2} T^{-1} \sum_{t=1}^{T} \left[ \sigma_{cc}(2e_{M}g_{t} + g_{t}^{2} - V_{Mg}) + 2g_{t}V_{Mcf} \right] \]

Thus, condition (3.21) is satisfied if

\[ T^{-1} \sum_{t=1}^{T} (g_{t}^{2} - V_{Mg}) \rightarrow 0 \] \hspace{1cm} (3.24)

and

\[ T^{-1} \sum_{t=1}^{T} g_{t} \rightarrow 0 \] \hspace{1cm} (3.25)
Now $g_t^2$ is a sequence of random variables with mean $\mu_{gg}$ and common finite variance. Moreover, since the sequence $\{c_t\}$ is uncorrelated, the definition of $g_t$ implies that for $|t - s| > M$, $g_t^2$ and $g_s^2$ are uncorrelated. If $T = Mq + w$, $0 < w < M$, one may write (3.24) as the sum of $M$ separate terms, with the $i$-th term equal to

$$T^{-1} \sum_{j=0}^{q-1} (s_{jM+i}^2 - \mu_{gg}) \text{ if } 0 < i < w,$$

and

$$T^{-1} \sum_{j=0}^{q-1} (s_{jM+i}^2 - \mu_{gg}) \text{ if } w < i < M.$$

By Lemma 3.4, each of the $M$ sums converges almost surely to zero as $T \to \infty$. Since $M$ is constant with respect to $T$, it follows that convergence in (3.24) holds with probability one. Almost sure convergence in (3.25) is established similarly. Now consider the convergence of

$$s_{MTT}^{-2} \sum_{t=1}^{T} E\{[Zc_t - \Gamma_M(0)]^2 I[|Zc_t - \Gamma_M(0)| > \epsilon s_{MTT}|F_{t-1}]\}$$

$$= T^{-1}s_{MTT}^{-2} \sum_{t=1}^{T} E\{[Zc_t - \Gamma_M(0)]^2 I[|Zc_t - \Gamma_M(0)| > \epsilon T s_{MTT}^{-2}|F_{t-1}]\}.$$ 

(3.26)

By Lemma 3.6.b,
The expectation of (3.26) is bounded above by

\[ E\left( [Z_t c_t - \Gamma_M(0)]^2 \mathbb{I}\left[ |Z_t c_t - \Gamma_M(0)| > \epsilon \cdot T \cdot s_M^2 \right] F_{t-1} \right) \]

\[ < \left( \epsilon \cdot T \cdot s_M^2 \right)^{-\nu} E\left\{ \left| Z_t c_t - \Gamma_M(0) \right|^{2+\nu} \left| F_{t-1} \right\} \],

so the expectation of (3.26) is bounded above by

\[ T^{-1} s_M^{-2} \left( \epsilon \cdot T \cdot s_M^2 \right)^{-\nu} \sum_{t=1}^{T} E\left\{ \left| Z_t c_t - \Gamma_M(0) \right|^{2+\nu} \right\} . \]

Assumption (3.13) and the definitions of \( Z_t \) and \( c_t \) imply that

\[ \text{sup}_t E\{ |Z_t c_t - \Gamma_M(0)|^{2+\nu} \} \equiv B_M < \infty. \]

Then by the Markov inequality,

\[ P\left( s_{\text{M}T}^{-2} \sum_{t=1}^{T} E\left\{ [Z_t c_t - \Gamma_M(0)]^2 \mathbb{I}\left[ |Z_t c_t - \Gamma_M(0)| > \epsilon s_{\text{MT}} \left| F_{t-1} \right| \right] > \eta \right) \]

\[ < \eta^{-1} T^{-1} s_M^{-2} \left( \epsilon \cdot T \cdot s_M^2 \right)^{-\nu} \sum_{t=1}^{T} E\left\{ \left| Z_t c_t - \Gamma_M(0) \right|^{2+\nu} \right\} \]

\[ < \eta^{-1} s_M^{-2-2\nu} \epsilon^{-\nu} v^{-\nu} B_M, \]

which converges to zero as \( T \to \infty \). Thus, conditions (3.22) and (3.23) are satisfied, so by Theorem 1 of Scott (1973),

\[ s_{\text{M}T}^{-1} \sum_{t=1}^{T} [Z_t c_t - \Gamma_M(0)] \xrightarrow{f} \mathcal{N}(0, 1) \]
as $T \to \infty$. It follows from condition (3.19) that under model (3.16),

$$T^{1/2} [\bar{\varepsilon}_M + \bar{\Gamma}_M(0) - \Gamma_M(0)] \xrightarrow{\mathcal{L}} N(0, \Sigma_M)$$

as $T \to \infty$.

Now consider the more general case

$$\varepsilon_t = \sum_{h=0}^{\infty} \alpha_h \varepsilon_{t-h}.$$

Unless indicated otherwise, retain the notation given above. Note that

$$T^{1/2} [\bar{\varepsilon} + \bar{\Gamma}(0) - \Gamma(0)]$$

$$= T^{1/2} [\bar{\varepsilon}_M + \bar{\Gamma}_M(0) - \Gamma_M(0)]$$

$$+ T^{1/2} \left\{ [\bar{\varepsilon} - \bar{\varepsilon}_M] + [\bar{\Gamma}(0) - \bar{\Gamma}_M(0)] - [\Gamma(0) - \Gamma_M(0)] \right\}$$

$$= T^{1/2} [\bar{\varepsilon}_M + \bar{\Gamma}_M(0) - \Gamma_M(0)] + T^{-1/2} \sum_{t=1}^{T} \sum_{h=M+1}^{\infty} \alpha_h \varepsilon_{t-h}$$

$$+ T^{-1/2} \sum_{t=1}^{T} \sum_{h=0}^{h=M} \left( \sum_{h=0}^{M} \alpha_h \varepsilon_{t-h} \right)^2 - \sum_{h=0}^{M} \alpha_h \varepsilon_{t-h}^2 - T^{1/2} [\Gamma(0) - \Gamma_M(0)]$$

$$= T^{1/2} [\bar{\varepsilon}_M + \bar{\Gamma}_M(0) - \Gamma_M(0)] + T^{-1/2} \left\{ W_{TM1} + W_{TM2} + W_{TM3} \right\},$$

where
Then by Lemma 3.7, the asymptotic distribution of $T^{1/2} [\epsilon + \Gamma(0) - \Gamma(0)]$ under the infinite moving average model will be established if

$T^{1/2} [W_{TM1} + W_{TM2} + W_{TM3}]$ satisfies condition (3.10.a) uniformly in $T$ and if the $N(0, \sigma^2_h)$ distribution converges completely to the $N(0, G_1)$ distribution as $M \to \infty$.

First, since the $c_t$ are uncorrelated,

$T^{-1/2} W_{TM1} = T^{-1/2} \sum_{t=1}^{T-M-1} \sum_{h=M+1}^{T} a_{t-h} c_t$

has variance equal to

$= T^{-1} \sum_{s=-\infty}^{T-M-1} \sum_{t=\max(1,s+M+1)}^{T} [\sum_{t-s}^{T} a_t]^2 \sigma^2_{cc}$

$= T^{-1} \sum_{s=-\infty}^{T-M-1} \sum_{t=\max(1,s+M+1)}^{T} [\sum_{t-s}^{T} a_t]^2 \sigma^2_{cc}$

$+ T^{-1} \sum_{s=-M+1}^{T-M-1} \sum_{t=M+1}^{T-s} [\sum_{t-s}^{T} a_t]^2 \sigma^2_{cc}$

(3.27)
By the absolute summability of the sequence \( \{a_n\} \), one may write

\[
|T^{-1} \sum_{s=-\infty}^{-M} \sum_{t=1}^{T} a_{t-s}^2| = |T^{-1} \sum_{s=-\infty}^{-M} \sum_{t=1}^{T} a_{t-s} a_{t-s'}| \\
= (\sum_{\ell=0}^{\infty} |a_{\ell}|) T^{-1} \sum_{t=1}^{T} \sum_{s=-\infty}^{-M} |a_{t-s'}| \\
< (\sum_{\ell=1}^{\infty} |a_{\ell}|) \sum_{h=M+1}^{\infty} |a_h| 
\]

(3.29)

because the inequalities \( t > 1 \) and \(-\infty < s < -M\) imply that \( M + 1 < t - s < \infty \). Thus, expression (3.27) may be made arbitrarily small by choice of sufficiently large \( M \). Also, \( t > M + 1 \) implies that \( T - t + M - 1 < T - 2 \), so

\[
|T^{-1} \sum_{s=-M+1}^{T-M-1} \sum_{t=M+1}^{T-2} a_t^2| = |T^{M+1} \sum_{t=M+1}^{T+M-1} a_t^2 T^{-1} \min(T-t, T-\ell) + M-1| \\
< (\sum_{t=M+1}^{T+M-1} |a_t| |a_{\ell}| T^{-1} (T - 2)) \\
< (\sum_{t=M+1}^{\infty} |a_t|)^2 \cdot \sum_{t=M+1}^{\infty} |a_{\ell}| T^{-1} \min(T-t, T-\ell) + M-1 
\]

(3.30)

Thus, expression (3.28) also may be made arbitrarily small by choice of sufficiently large \( M \). Moreover, expressions (3.29) and (3.30) are not functions of \( T \), so uniformly in \( T \), \( E[T^{-1}w_{TM1}^2] \) converges to zero as \( M \to \infty \). Then by the Markov inequality, \( T^{-1/2}w_{TM1} \) converges in probability to zero uniformly in \( T \) as \( M \to \infty \).
Next, note that $T^{-1/2}w_{TM} = T^{-1/2} \sum_{t=1}^{T} \sum_{g=0}^{M-1} a_{g} a_{t} c_{g-h} c_{t-h}$ has mean zero and variance less than or equal to

$$\sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{g=0}^{M-1} a_{g} a_{t} a_{s} a_{h} |E(c_{t-g} c_{t-h} c_{s-g} c_{s-h})|$$

For a given quadruple $(g, h, s, m)$, the indices $t-g$, $t-h$, $s-g$, $s-m$ are all equal for at most one pair $(t, s)$, in which case the expectation above equals $K = E(c_{t-g} c_{t-h} c_{s-g} c_{s-m})$; and the indices $t-g$, $t-h$, $s-g$, $s-m$ are equal in pairs only for at most $T$ pairs $(t, s)$, in which case the expectation above equals $\sigma_{cc}^2$. Let $b_3 = \max(K, \sigma_{cc}^2)$. Then expression (3.31) is bounded above by

$$b_3 \sum_{g=0}^{M-1} \sum_{h=0}^{M-1} |a_{g}|^2 |a_{h}|^2.$$  

The absolute summability of the sequence $\{a_h\}$ implies that $[\sum_{g=0}^{\infty} |a_{g}|^2]$ is finite and that $[\sum_{h=0}^{\infty} |a_{h}|^2]$ may be made arbitrarily small by choice of sufficiently large $M$. Thus, expression (3.32) can be made arbitrarily small by choice of sufficiently large $M$, so uniformly in $T$, $E[T^{-1}w_{TM}^2]$ converges to zero as $M \to \infty$. Then by the Markov inequality, $T^{-1/2}w_{TM}$ converges in probability to zero uniformly in $T$ as $M \to \infty$. 
Now note that

\[
E(W_{TM3}) = T \left\{ \sum_{g=M+1}^{\infty} \sum_{h=M+1}^{\infty} a_g a_h E(c_{t-g} c_{t-h}) - [\Gamma(0) - \Gamma_M(0)] \right\}
\]

\[
= T \sum_{h=M+1}^{\infty} \sum_{h=0}^{M} a_h^2 - \sum_{h=0}^{\infty} \sum_{h=0}^{M} a_h^2
\]

\[
= 0.
\]

Also,

\[
T^{-1/2} W_{TM3} = T^{-1/2} \sum_{t=1}^{T} \sum_{g=M+1}^{\infty} \sum_{h=M+1}^{\infty} a_g c_{t-g} a_h c_{t-h}
\]

and has variance less than or equal to

\[
T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=M+1}^{\infty} \sum_{\ell=M+1}^{\infty} \sum_{m=M+1}^{\infty} \left| a_s a_h a_m \right| \left| E(c_{t-g} c_{t-h} c_{s-\ell} c_{s-m}) \right|
\]

\[
< T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{m=M+1}^{\infty} \left| a_s a_h a_m \right| \left| E(c_{t-g} c_{t-h} c_{s-\ell} c_{s-m}) \right|
\]

Since this final expression is identical to expression (3.31), the arguments given following (3.31) indicate that uniformly in \( T \),

\( E[T^{-1/2} W_{TM3}] \) converges to zero as \( M \to \infty \). Then by the Markov inequality, \( T^{-1/2} W_{TM3} \) converges in probability to zero uniformly in \( T \) as \( M \to \infty \). Thus, \( T^{1/2} [W_{TM1} + W_{TM2} + W_{TM3}] \) satisfies condition (3.10.a) uniformly in \( T \), so it remains only to show that the \( N(0, s_h^2) \) distribution converges completely to the \( N(0, G_1) \) distribution as \( M \to \infty \).
Let $Y_M$ be distributed as a normal $(0, \tilde{s}_M^2)$ random variable and let $Y_*$ be distributed as a normal $(0, G_1)$ random variable. Recall that $\tilde{s}_M^2 > 0$ for all $M$ and that $G_1 > 0$. Then for any real number $c$,

$$P[Y_M < c] = P[Z < \tilde{s}_M c]$$

and

$$P[Y_* < c] = P[Z < G_1^{1/2} c] ,$$

where $Z$ is distributed as a normal $(0, 1)$ random variable. Thus, the complete convergence of the $N(0, \tilde{s}_M^2)$ distribution to the $N(0, G_1)$ distribution as $M \to \infty$ will be established if

$$\lim_{M \to \infty} \tilde{s}_M^2 = G_1 . \quad (3.33)$$

For a given $M$, expressions (3.16) and (3.21) imply that

$$\tilde{s}_M^2 = T^{-1} \text{Var} \left\{ \sum_{t=1}^T \left[ Z_{ct} c_t - r_M(0) \right] \right\}$$

$$= \sigma_{cc} (e^2_M + V_{Mgg}) + 2 e_M V_{Mcf} + V_{Mcf}$$

$$\leq \sum_{h=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2 4 \sigma_{cc}^2 \left( \sum_{h=0}^{M-1} a_h \right)^2$$

$$+ 2\left( \sum_{h=0}^{M-1} a_h \right)^2 \sum_{l=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2$$

$$+ 2\left( \sum_{h=0}^{M-1} a_h \right)^2 \sum_{l=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2 ,$$

$$= \sigma_{cc} (e^2_M + V_{Mgg}) + 2 e_M V_{Mcf} + V_{Mcf}$$

$$\leq \sum_{h=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2 4 \sigma_{cc}^2 \left( \sum_{h=0}^{M-1} a_h \right)^2$$

$$+ 2\left( \sum_{h=0}^{M-1} a_h \right)^2 \sum_{l=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2$$

$$+ 2\left( \sum_{h=0}^{M-1} a_h \right)^2 \sum_{l=0}^{M-1} \left( \sum_{h=0}^{M-1} a_h \right)^2 .$$
which, as \( M \to \infty \), has limit

\[
\sigma_{cc}^2 \left( \sum_{h=0}^{\infty} a_h \right)^2 + 4\sigma^2 \sum_{h=0}^{\infty} \left( \sum_{\ell=1}^{h} a_{h+\ell} \right)^2 \\
+ 2M_3 \left( \sum_{h=0}^{\infty} a_h \right) \left( \sum_{\ell=0}^{\infty} a_\ell^2 \right) + \left( \sum_{h=0}^{\infty} a_h^2 \right)^2[k - \sigma_{cc}^2].
\] (3.34)

The absolute summability of the sequence \( \{s_h\} \) and the existence of common finite fourth moments of the \( c_t \) allow the exchange of summation and expectation operations, so for \( t > 0 \),

\[
\Gamma(t) = \text{Cov}(s_t, s_{t+s})
\]

\[
= \text{Cov}(\sum_{h=0}^{\infty} a_h c_{s-h}, \sum_{\ell=0}^{\infty} a_\ell c_{s+t-\ell})
\]

\[
= \sigma_{cc} \sum_{h=0}^{\infty} a_h^2 h + t;
\]

and

\[
\sum_{t=-\infty}^{\infty} \Gamma(t) = \sigma_{cc} \left\{ \sum_{t=0}^{\infty} \Gamma(t) \right\} - \Gamma(0)
\]

\[
= \sigma_{cc} \left\{ \sum_{t=0}^{\infty} \sum_{h=0}^{\infty} a_h^2 h + t \right\} - \sum_{h=0}^{\infty} a_h^2
\]

\[
= \sigma_{cc}^2 \sum_{h=0}^{\infty} a_h^2.
\]

This final equality follows from an algebraic identity for absolutely summable sequences \( \{a_h\} \) and \( \{b_h\} \),
\[
(\sum_{h=0}^{\infty} b_h)(\sum_{a, \xi} a_{h, \xi}) = \sum_{h=0}^{\infty} b_h[\sum_{a, \xi} a_{h, \xi}] = \sum_{h=0}^{\infty} \sum_{a, \xi} b_h a_{h, \xi} + \sum_{h=0}^{\infty} b_h a_{h, h + t} = \sum_{h=0}^{\infty} b_h a_{h, h + t} = 2 \sum_{h=0}^{\infty} b_h a_{h, h + t} - \sum_{h=0}^{\infty} b_h a_{h}.
\]

(3.35)

In addition,

\[
\sum_{t=1}^{\infty} [\Gamma(t)]^2 = 2 \sum_{t=1}^{\infty} [\Gamma(t)]^2 + [\Gamma(0)]^2 = 2 \sigma_{cc}^2 \sum_{t=1}^{\infty} [\sum_{h=0}^{\infty} a_{h, h + t}]^2 + \sigma_{cc}^2 \sum_{h=0}^{\infty} (\sum_{a} a_{h, h}^2).
\]

Similarly,

\[
M(0, 0, t) = E(e_s e_s e_s)
\]

\[
= E[(\sum_{a, \xi} a_{h, \xi}^s)(\sum_{a, \xi} a_{h, \xi}^s)(\sum_{s, \xi} a_{i, s + t - i}^s)] = \sum_{h=0}^{\infty} (a_{h, h}^s)^2.
\]

So

\[
\sum_{t=1}^{\infty} M(0, 0, t) = \{2[\sum_{t=0}^{\infty} M(0, 0, t)] - M(0, 0, 0)\}
\]
\[ M_3 \left( \sum_{h=0}^{\infty} a^2_{h} a^3_{h+t} \right) - \sum_{h=0}^{\infty} a^3_{h} \]

\[ = M_3 \left( \sum_{h=0}^{\infty} a^2_{h} \right) \left( \sum_{h=0}^{\infty} a^3_{h} \right) . \]

Also,

\[ E(\epsilon_s \epsilon_s \epsilon_{s+t} \epsilon_{s+t}) \]

\[ = \sum_{h=0}^{\infty} \sum_{l=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a^s_{h} a^s_{l} a^j_{i} a^j_{i} E(c_s-h c_s-l c_{s+t-i} c_{s+t-j}) \]

For a given pair \((s, t)\), the indices \(s-h, s-\ell, s+t-i, \) and \(s+t-j\)
are all equal if and only if \(i = j = h+t = \ell+t\), in which case
\[ E(c_s-h c_s-l c_{s+t-i} c_{s+t-j}) = \kappa ; \]
and the indices \(s-h, s-\ell, s+t-i, \) and \(s+t-j\)
are equal in pairs only, if \(h = \ell \neq i-t = j-t, \)
\(h = i-t \neq \ell = j-t, \) or \(h = j-t \neq \ell = i-t, \) in which cases
\[ E(c_s-h c_s-\ell c_{s+t-i} c_{s+t-j}) = \sigma^2 \; \text{cc}. \]
In all other cases,
\[ E(c_s-h c_s-\ell c_{s+t-i} c_{s+t-j}) = 0. \] Thus, for \(t > 0, \)

\[ E(\epsilon_s \epsilon_s \epsilon_{s+t} \epsilon_{s+t}) \]

\[ = \kappa \sum_{h=0}^{\infty} a^2_{h} a^2_{h+t} + \sigma^2 \sum_{i=0}^{\infty} a^2_{i} a^2_{i} + 2 \sum_{h=0}^{\infty} \sum_{i=0}^{\infty} a^3_{h} a^3_{h+t} \]

\[ - \sum_{i \# h}^{\infty} a^3_{i} a^3_{i+t} . \]
Then for \( t > 0 \),

\[
k(0, 0, t, t) = \sum_{s} \varepsilon_s \varepsilon_{s+t} - [\Gamma(t)]^2 - 2[\Gamma(0)]^2
\]

\[
= K \sum_{h} a_{h}^2 + \sum_{h} \sum_{h+t} a_{h}^2 - \sum_{h} \sum_{h+t} a_{h}^2
\]

\[
= K \sum_{h} a_{h}^2 + \sum_{h} \sum_{h+t} a_{h}^2
\]

\[
= (K - 3a^2) \left( \sum_{h} a_{h}^2 \right)
\]

so

\[
\sum_{t=0}^{\infty} k(0, 0, t, t) = (K - 3a^2) \left( \sum_{h} a_{h}^2 \right)^2
\]

where the final equality again follows from expression (3.35). Finally,

\[
\sum_{t=\infty}^{\infty} \{ \Gamma(t) + 2M(0, 0, t) + 2[\Gamma(t)]^2 + k(0, 0, t, t) \}
\]
\[ \begin{align*}
&= \sigma_{cc}^2 \left( \sum_{h=0}^{\infty} a_h \right)^2 + 2M_3 \left( \sum_{h=0}^{\infty} a_h^2 \right) \left( \sum_{\ell=0}^{\infty} a_\ell \right) + 4\sigma_{cc}^2 \sum_{t=1}^{\infty} \sum_{h=0}^{\infty} a_h a_{h+t}^2 \\
&\quad + 2\sigma_{cc}^2 \left( \sum_{h=0}^{\infty} a_h^2 \right)^2 + (K - 3\sigma_{cc}^2) \left( \sum_{h=0}^{\infty} a_h^2 \right)^2 \\
&= \sigma_{cc}^2 \left( \sum_{h=0}^{\infty} a_h \right)^2 + 2M_3 \left( \sum_{h=0}^{\infty} a_h^2 \right) \left( \sum_{\ell=0}^{\infty} a_\ell \right) \\
&\quad + 4\sigma_{cc}^2 \sum_{t=1}^{\infty} \sum_{h=0}^{\infty} a_h a_{h+t}^2 + (K - \sigma_{cc}^2) \left( \sum_{h=0}^{\infty} a_h^2 \right)^2,
\end{align*} \]

which is equal to expression (3.34). Thus, condition (3.33) is established, so by Lemma 3.7,

\[ T^{1/2} [\tilde{e} + \tilde{r}(0) - \Gamma(0)] \xrightarrow{p} N(0, \Gamma_1) \]

as \( T \to \infty \). \( \square \)

Lemma 3.8 established weak consistency and asymptotic normality properties of the sample moment function \( \tilde{e} + \tilde{r}(0) \). Also, under the conditions of Lemma 3.8, the functions \( \tilde{e} \) and \( \tilde{r}(0) - \Gamma(0) \) each converge with probability one to zero, so \( \tilde{e} + \tilde{r}(0) - \Gamma(0) \) also converges with probability one to zero. These almost sure convergence results follow immediately from Theorems IV.5 and IV.6 of Hannan (1970), so detailed consideration of convergence with probability one will be deferred to the following section.

Finally, in practice one often uses the estimator
\[ \hat{\gamma}(0) = T^{-1} \sum_{t=1}^{T} (\varepsilon_t - \bar{\varepsilon})^2 \]

in place of

\[ \bar{\gamma}(0) = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 . \]

Now

\[ \hat{\gamma}(0) - \bar{\gamma}(0) = T^{-1} \sum_{t=1}^{T} \left[ (\varepsilon_t - \bar{\varepsilon})^2 - \varepsilon_t^2 \right] \]

\[ = T^{-1} \sum_{t=1}^{T} \left[ -2 \bar{\varepsilon} \varepsilon_t + \bar{\varepsilon}^2 \right] \]

\[ = - T \bar{\varepsilon}^2 , \]

\[ E(\bar{\varepsilon}) = 0 , \]

and

\[ \text{Var}(\bar{\varepsilon}) = T^{-2} \sum_{h=-T+1}^{T-1} |T - h| \Gamma(h) . \]

Thus, under the conditions of Lemma 3.8

\[ \lim_{T \to \infty} T E(\bar{\varepsilon}^2) = \sum_{h=-\infty}^{\infty} \Gamma(h) < \infty , \]

so

\[ T^{1/2} \bar{\varepsilon}^2 = o_p(T^{-1/2}) \]

and
Thus, under the conditions of Lemma 3.8, \( T^{1/2} [\bar{\varepsilon} + \hat{\Gamma}(0) - \Gamma(0)] \) and \( T^{1/2} [\bar{\varepsilon} + \hat{\Gamma}(0) - \Gamma(0)] \) have the same limiting distribution. Moreover, the almost sure convergence of \( \bar{\varepsilon} \) to zero and of \( \bar{\varepsilon} + \hat{\Gamma}(0) - \Gamma(0) \) to zero imply that under the conditions of Lemma 3.8, \( \bar{\varepsilon} + \hat{\Gamma}(0) - \Gamma(0) \) converges to zero with probability one.

3.3. Asymptotic Properties of the Sum of a Linear Function and an Unweighted Bilinear Function

The preceding section presented some asymptotic properties of

\[
\bar{\varepsilon} + \hat{\Gamma}(0) = T^{-1} l^t \xi + T^{-1} l^t \xi',
\]

where \( \{\varepsilon_t\} \) was a univariate linear process. This section develops similar properties for a more general sum of a linear function and an unweighted bilinear function of a multivariate linear process.

An outline of this section is as follows. First, some preliminary remarks lead to the evaluation of the mean and variance of the sum of a linear function and a bilinear function of a multivariate linear process. Second, condition (3.39.a) imposes a Grenander-type regularity condition on the weights of the linear function of interest. Lemma 3.9 gives a technical result associated with condition (3.39.a). Third, expression (3.42) gives the limiting value of the variance of the sum of the linear function and the bilinear function under study. Fourth,
Lemmas 3.10 and 3.11 review some previously known results regarding almost sure convergence properties of a linear function and a bilinear function of serially correlated random vectors. Finally, Theorem 3.1 gives conditions for the asymptotic normality of the sum of a linear function and a bilinear function of a realization of a multivariate linear process. The results of this section will be used in Section 4.1 to establish the strong consistency and asymptotic normality of some standard errors-in-variables estimators in the presence of serially correlated observations.

Let \( \{ \mu_t', \, t > 1 \} \) be a sequence of fixed \( k \times 1 \) vectors and let

\[
\varepsilon_t = (u_t', v_t')' = \sum_{j=0}^{\infty} A_j c_{t-j} \]

\[
= [(\sum_{j=0}^{\infty} B_j c_{t-j})', (\sum_{j=0}^{\infty} D_j c_{t-j})']' \tag{3.36}
\]

be a \( p \times 1 \) linear process, where \( u_t' \) is \( k \times 1 \), \( v_t' \) is \( r \times 1 \), \( p = r + k \), \( B_j \) is \( k \times p \), \( D_j \) is \( r \times p \), and \( \{ A_j = (B_j', D_j')' \} \) is an absolutely summable sequence. Let \( X_t = u_t + u_t \) and let \( X, X_j \), and similar matrices be as defined in Chapter 2. Then \( (I_r \otimes X') \text{vec}(v) \) is an \( rk \times 1 \) vector with \((i,j)\)-th double-subscripted entry equal to

\[
X_{ij} v_i = u_{ij} v_i + u_{ij} v_i = \sum_{t=1}^{T} u_{tj} v_{ti} + \sum_{t=1}^{T} u_{tj} v_{ti}.
\]

Since the vectors \( u_t \) are fixed,
\[ E[\mathbf{X}' \mathbf{v} . i] = E[\mu_j \mathbf{v} . i] + E[\mathbf{u}' \mathbf{v} . i] = T \Gamma_{ij}v_i(0) , \]

so \[ E[(I_T \otimes \mathbf{X}')\mathbf{v} . i] = T \Gamma_{ij}v_i(0) . \]

Also,

\[ \text{Var}[(I_T \otimes \mathbf{X}')\mathbf{v} . i] = \text{Var}[(I_T \otimes \mu')\mathbf{v} . i] + \text{Cov}[(I_T \otimes \mu')\mathbf{v} . i, (I_T \otimes \mathbf{u}')\mathbf{v} . i] + \text{Cov}[(I_T \otimes \mathbf{u}')\mathbf{v} . i, (I_T \otimes \mathbf{u}')\mathbf{v} . i] + \text{Var}[(I_T \otimes \mathbf{u}')\mathbf{v} . i] . \]

Consider each of the terms of \[ \text{Var}[(I_T \otimes \mathbf{X}')\mathbf{v} . i] \] individually.

First, \[ \text{Var}[(I_T \otimes \mu')\mathbf{v} . i] \] has \([(i,j), (l,m)]\)-th element equal to

\[ \mu_{j,l} = \sum_{s=1}^{m_2} \sum_{t=1}^{T} \Gamma_{vii}(s-t) u_{sm} , \]

where \( m_{1t} = \max(1, 1+t) \) and \( m_{2t} = \min(T, T+t) \). Thus,

\[ \text{Var}[(I_T \otimes \mu')\mathbf{v} . i] = \sum_{t=-T+1}^{T-1} \Gamma_{vii}(t) \sum_{s=m_{1t}}^{m_{2t}} \mu_{s-t,l} u_{sm} . \]

Next, \[ \text{Var}[(I_T \otimes \mathbf{u}')\mathbf{v} . i] \] has \([(i,j), (l,m)]\)-th element equal to
where the third equality follows from equation (1.5.1) of Hannan (1970) and $k_{uvuvjimz}$ is the fourth cross-cumulant function for the indicated elements of the $e_t$ process. Let $k_{uvuv}(0,0,t,t)$ be an $rk \times rk$ matrix with $[(i,j), (z,m)]$-th element equal to $k_{uvuvjimz}(0,0,t,t)$. An $rk \times rk$ matrix with $[(i,j), (z,m)]$-th entry equal to $\Gamma_{uvj}(s-t)\Gamma_{vim}(s-t)$ has $(i,z)$-th $k \times k$ block equal to

$$\Gamma_{uvj}(s-t)\Gamma_{vim}(s-t) = \Gamma_{uvj}(s-t)\Gamma_{vim}(t-s)'$$

by the skew-symmetry of $\Gamma_{e_t}(s-t)$, so the entire $rk \times rk$ matrix equals
Thus,

\[ \text{Var}[(I_r \ast u') \vec{v}] \]

\[ = \sum_{t=1}^{T} \sum_{s=1}^{T} \{ \Gamma_{vv}^u(s-t) \ast \Gamma_{uu}^u(s-t) + \vec{\Gamma}_{vuv}^u(s-t) \vec{\Gamma}_{uvu}^u(t-s) \}' \]

\[ + \mathbf{k}_{uvuv}(0, 0, s-t, s-t) \]

\[ = \sum_{t=-T+1}^{T-1} (T - |t|) \{ \Gamma_{vv}^u(t) \ast \Gamma_{uu}^u(t) + \vec{\Gamma}_{vuv}^u(t) \vec{\Gamma}_{uvu}^u(-t) \}' \]

\[ + \mathbf{k}_{uvuv}(0, 0, t, t) . \]

This is a vector version of a special case of formula (IV.3.3) of Hannan (1970).

Finally,

\[ \text{Cov}[(I_r \ast u') \vec{v}, (I_r \ast u') \vec{v}] \]

has \([(i,j), (l,m)]\)-th element equal to

\[ \text{Cov}[u_j^v \vec{v}_l, u_m^v \vec{v}_l] = \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Cov}(u_{tj}^v v_t, v_{sl}^v s_m^v) \]

\[ = \sum_{t=1}^{T} \sum_{s=1}^{T} E(u_{tj}^v v_t v_{sl}^v s_m^v) . \]
Recalling the third-order stationarity of $\xi_t$, let

$$M_{vuv}(0, 0, s-t) = E[v_t' a u_t' a v_s']$$

be an $rk \times r$ matrix with $[(i,j), \xi]$-th element equal to $E[v_{ti}' u_t v_{sj}']$. Then

$$\text{Cov}[(I^a u')vec(v), (I^a u')vec(v)] = \sum_{t=1}^{T} \sum_{s=1}^{T} M_{vuv}(0, 0, s-t) a_{s-t} a_{s}$$

$$= \sum_{t=-T+1}^{T-1} \sum_{s=m_{2t}}^{m_{2t}} M_{vuv}(0, 0, t) a_{s-t} a_{s}$$

Thus,

$$\text{Var}[(I^a X')vec(v)]$$

$$= \sum_{t=-T+1}^{T-1} \sum_{s=m_{2t}}^{m_{2t}} \{ \Gamma_{vv}(t) a_{s-t} a_{s} \}$$

$$+ \Gamma_{vv}(t) a_{s} \Gamma_{uu}(t) + \text{vec}[\Gamma_{vv}(t)]\text{vec}[\Gamma_{uu}(-t)]'$$

$$+ M_{vuv}(0, 0, t) a_{s} + [M_{vuv}(0, 0, t) a_{s}]' + k(0, 0, t, t)$$

A limiting value of $T^{-1} \text{Var}[(I^a X')vec(v)]$ need not exist unless the fixed sequence $\{a_t\}$ satisfies some regularity conditions. A simplified form of Grenander's conditions for the fixed vectors $a_t = (a_{t1}, a_{t2}, \ldots, a_{tk})$ is imposed by assuming that
exists and is finite for all \( j, m = 1, 2, \ldots, k \) and \( \lambda = 0, 1, 2, \ldots \).

Lemma 3.9 demonstrates that condition (3.39.a) gives rise to a second limiting condition.

**Lemma 3.9.** Let the sequence of \( 1 \times k \) vectors \( \{\mu_t\} \) satisfy condition (3.39.a). Then

\[
\lim_{T \to \infty} T^{-1} \left\{ \sum_{t=1}^{T} \mu_{tj}^2 \right\} = 0 , \quad j = 1, 2, \ldots, k . \tag{3.39.b}
\]

**Proof.** Pick \( \eta > 0 \). Let \( \delta = 3^{-1} \eta \). By condition (3.39.a), there exists some \( T_\delta \) such that \( T > T_\delta \) implies that

\[
\left| T^{-1} \sum_{t=1}^{T} \mu_{tj}^2 - \tilde{\mu}_{jj}(0) \right| < \delta
\]

for \( j = 1, 2, \ldots, k \). Let

\[
T_\eta = \max \{ T_\delta + 1, \delta^{-1}[\tilde{\mu}_{jj}(0) + \delta] \} .
\]

For \( T > T_\eta \), repeated application of the triangle inequality implies that for all \( j = 1, 2, \ldots, k \),

\[
\delta > \left| T^{-1} \sum_{t=1}^{T} \mu_{tj}^2 - \tilde{\mu}_{jj}(0) \right|
\]
\[ T^{-1} \mu_{ij}^2 - \sum_{t=1}^{T-1} \mu_{ij}^2 - M_{jj}(0) \leq T^{-1} \sum_{t=1}^{T-1} \mu_{ij}^2 \]
\[ T^{-1} \mu_{ij}^2 - \delta - T^{-1} [\tilde{M}_{jj}(0) + \delta] \]
\[ T^{-1} \mu_{ij}^2 - 2\delta, \]
so
\[ T^{-1} \mu_{ij}^2 < 3\delta = \eta. \]

Thus, for all \( T > T_\eta \),
\[ T^{-1} \{ \max_{T_\eta \leq T} (\mu_{ij}^2) \} < \eta. \]

Since each \( \mu_{ij} \) is finite,
\[ \lim_{T \to \infty} T^{-1} \{ \max_{1 \leq t \leq T_\eta} (\mu_{ij}^2) \} = 0 \]
so the result follows. \( \square \)

Let \( \tilde{M}_{\mu j} (\ell) \) be the \( k \times k \) matrix with \((j,m)\)-th element equal to \( \tilde{M}_{\mu jm} (\ell) \). The inequality \( 2|\mu_{ij}^{t+\ell,\mu}| < \mu_{ij}^2 + \mu_{ij}^{t+\ell,\mu} \) implies that \( 2|\tilde{M}_{\mu jm}(\ell)| < \tilde{M}_{\mu jm}(0) + \tilde{M}_{\mu mm}(0) \) for all \( \ell \in \mathbb{Z} \). By (3.39.a), \( \tilde{M}_{\mu jm}(0) \) is bounded by some real number \( B \) for all \( j=1, 2, \ldots, k \), so
\[ |\tilde{M}_{\mu jm}(\ell)| < B \] \( (3.39.c) \)
for all $j, m = 1, 2, \ldots, k$ and $\ell = \ldots -1, 0, 1, \ldots$, i.e. the elements of $\{\bar{M}_{\mu \mu}(\ell)\}$ are uniformly bounded.

Moreover, by definition of $\bar{M}_{\mu \mu j}(0) = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mu_{tj}^2$, there exists some $T_j$ such that $T > T_j$ implies that

$$|T^{-1} \sum_{t=1}^{T} \mu_{tj}^2| < \bar{M}_{\mu \mu j}(0) + |T^{-1} \sum_{t=1}^{T} \mu_{tj}^2 - \bar{M}_{\mu \mu j}(0)| < \bar{M}_{\mu \mu j}(0) + 1.$$

Thus, for all positive integers $t$ and all $1 < j < k$,

$$t^{-1} \sum_{s=1}^{t} \mu_{s j}^2 < B_1$$

where

$$B_1 = \max \left\{ \max_{1 < j < k} \left( \max_{1 < t < T_j} \left( t^{-1} \sum_{s=1}^{t} \mu_{s j}^2 \right) \right), \bar{M}_{\mu \mu j}(0) + 1 \right\}.$$

Finally, note that condition (3.39.a) and the inequality $|\mu_{tj}| < (\mu_{tj}^2 + 1)$ imply that

$$\bar{\mu} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mu_t$$

exists and is finite.

Now apply conditions (3.39) to evaluate
First, the absolute summability of $\bar{\Gamma}_V(t)$ and the uniform boundedness of $\bar{M}_\mu(t)$ imply that the matrices $\bar{\Gamma}_V(t) \circ \bar{M}_\mu(t)$ are absolutely summable. Let $T_0$ be an integer such that $1 < T_0 < T$. Let $m_{1t} = \max(1, t+1)$ and $m_{2t} = \min(T, T+t)$. Then

$$\lim_{T \to \infty} T^{-1} \text{Var}[(I_T \circ X')\vec{v}(w)].$$

$$\begin{align*}
\text{has } [(i,j), (l,m)]-\text{th element equal to} \quad & \\
& \\
& T_0^{-1} T^{-1} \sum_{t=-T_0+1}^{T-1} \Gamma_{V\bar{V}}(t) \left[ \sum_{s=m_{1t}}^{m_{2t}} \mu_{s-t,j} \mu_{s-m} \right] - \lim_{t \to -\infty} \Gamma_{V\bar{V}}(t) \circ \bar{M}_\mu(t) \\
& + \sum_{t=T_0}^{T-1} \Gamma_{V\bar{V}}(t) \left[ \sum_{s=m_{1t}}^{m_{2t}} \mu_{s-t,j} \mu_{s-m} \right] \\
& + \sum_{t=-T_0}^{T_0} \sum_{s=m_{1t}}^{m_{2t}} \Gamma_{V\bar{V}}(t) \bar{M}_{\mu j}(t) - \sum_{t=-T_0}^{T_0} \Gamma_{V\bar{V}}(t) \bar{M}_{\mu j m}(t). \\
& \text{Pick } \varepsilon > 0 \text{ and consider each of these five addends individually. The} \\
& \text{absolute summability of } \bar{\Gamma}_V(t) \circ \bar{M}_\mu(t) \text{ implies that there exists some} \\
& T_1 \varepsilon \text{ such that if } T_0 > T_1 \varepsilon, \text{ then} \\
\end{align*}
\[ \left| \sum_{t=T_0}^{\infty} \Gamma_{\nu \nu \nu}(t) \mu(t) \nu(t) \right| < 5^{-1} \epsilon \]

and

\[ \left| \sum_{t=-\infty}^{T} \Gamma_{\nu \nu \nu}(t) \mu(t) \nu(t) \right| < 5^{-1} \epsilon \]

for all $1 \leq i, \ell \leq r$ and all $1 \leq j, m \leq k$. Next, recall that

\[ m_{1t} = \max(1, t+1), \quad m_{2t} = \min(T, T+t), \quad \text{and} \quad \mu_{hj} \mu_{h+t,m} < \mu_{hj}^2 + \mu_{h+t,m}^2, \]

so

\[ m_{2t} \]

\[ \left| \sum_{s=m_{1t}}^{T} \mu_{s-t,j} \mu_{s+m} \right| < \sum_{h=1}^{T} (\mu_{hj}^2 + \mu_{hm}^2). \]  

(3.40)

Thus,

\[ \left| \sum_{t=T_0}^{T-1} \Gamma_{\nu \nu \nu}(t) \left[ \mu_{s-t,j} \mu_{s+m} \right] \right| \]

\[ < \sum_{t=T_0}^{T-1} \left| \Gamma_{\nu \nu \nu}(t) \right| \cdot \sum_{h=1}^{T-t} \left( \mu_{hj}^2 + \mu_{h+t,m}^2 \right) \]

\[ < \sum_{t=T_0}^{T-1} \left| \Gamma_{\nu \nu \nu}(t) \right| \cdot \sum_{h=1}^{T} \left( \mu_{hj}^2 + \mu_{hm}^2 \right) \]

\[ < 2B_1 \sum_{t=T_0}^{T-1} \left| \Gamma_{\nu \nu \nu}(t) \right|, \]
where $B_1$ is the uniform bound on $T_{t=1}^{T} u_{hj}^2$ and $T_{t=1}^{T} u_{hm}^2$.

Similarly,

$$
\left| \sum_{t=-T+1}^{-T_0} \sum_{s=m_1t}^{m_2t} \Gamma_{vvi}(t)[T_{t=1}^{T} \sum_{h=1}^{T} \mu_{hj} + T_{t=1}^{T} \mu_{hm}] \right|
\leq \sum_{t=-T+1}^{-T_0} \sum_{s=m_1t}^{m_2t} |\Gamma_{vvi}(t)| \cdot |T_{t=1}^{T} \sum_{h=1}^{T} \mu_{hj} + T_{t=1}^{T} \mu_{hm}|
\leq 2B_1 \sum_{t=-T+1}^{-T_0} |\Gamma_{vvi}(t)| .
$$

By the absolute summability of $\Gamma_{vvi}(t)$, there exists some $T_{2\epsilon}$ such that if $T_0 > T_{2\epsilon}$, then

$$
2B_1 \sum_{t=T_0}^{\infty} |\Gamma_{vvi}(t)| < 5^{-1}\epsilon
$$

and

$$
2B_1 \sum_{t=-\infty}^{-T_0} |\Gamma_{vvi}(t)| < 5^{-1}\epsilon
$$

for all $1 < i, \ell < r$.

Now let $T_{0\epsilon} = \max(T_{1\epsilon}, T_{2\epsilon})$. By the absolute summability of $\Gamma_{vvi}(t)$, define

$$
\bar{V} \equiv \max \{ \sum_{t=-\infty}^{\infty} |\Gamma_{vvi}(t)| \} + 1 .
$$
By condition (3.39.a), for each $t = -T_0\varepsilon + 1, -T_0\varepsilon + 2, \ldots, T_0\varepsilon - 1$, there exists some $T_{1t\varepsilon}$ such that $T > T_{1t\varepsilon}$ implies that

$$|T^{-1} \sum_{h=1}^{T} \mu_{h+t,m} - \overline{M}_{\mu jm}(t)| < (10)^{-1} \varepsilon.$$ 

Let

$$T_{3\varepsilon} = \max \{ T_{1t\varepsilon} \}.$$ 

Since $T_0\varepsilon$ is fixed, $\lim_{T \to \infty} T^{-1}(T - T_0\varepsilon) = 1$, so there exists some $T_{4\varepsilon}$ such that $[1 - T^{-1}(T - T_0\varepsilon)]B_1 < (10)^{-1} \varepsilon$ for $T > T_{4\varepsilon}$. Let $T_\varepsilon = \max(T_{3\varepsilon} + T_0\varepsilon, T_{4\varepsilon})$. Then for $T > T_\varepsilon$ and $t > 1$,

$$|T^{-1} \sum_{s=m+1}^{m+T} \mu_{s-t,j} \mu_{sm} - \overline{M}_{\mu jm}(t)|$$

$$= |T^{-1} \sum_{s=m+1}^{T} \mu_{s-t,j} \mu_{sm} - \overline{M}_{\mu jm}(t)|$$

$$= |T^{-1} \sum_{h=1}^{T-t} \mu_{h+t,m} - \overline{M}_{\mu jm}(t)|$$

$$< |(T - t)^{-1} T-t \sum_{h=1}^{T-t} \mu_{h+t,m} - \overline{M}_{\mu jm}(t)|$$

$$+ [1 + T^{-1}(T - T_0\varepsilon)](T - t)^{-1} \sum_{h=1}^{T-t} \mu_{h+t,m}$$

$$< |(T - t)^{-1} T-t \sum_{h=1}^{T-t} \mu_{h+t,m} - \overline{M}_{\mu jm}(t)| + [1 - T^{-1}(T - T_0\varepsilon)]B_1$$
Similar arguments establish that

\[ |T^{-1} \sum_{s=m}^{m_2t} \mu_{s-t,j} \mu_{sm} - \bar{\mu}_{\mu jm}(t)| < 5^{-1} \epsilon \]

for \( T > T_\epsilon \) and \( t < 1 \) also. Thus, for \( T > T_\epsilon \),

\[ 0 \leq T_0 \leq \sum_{t=-T_0+1}^{m_2t} \Gamma_{\nu\nu}(t)[T^{-1} \sum_{s=m}^{m_1t} \mu_{s-t,j} \mu_{sm} - \bar{\mu}_{\mu jm}(t)] < 5^{-1} \epsilon . \]

Hence, for \( T > T_\epsilon \),

\[ T^{-1} \sum_{t=-T+1}^{m_2t} \Gamma_{\nu\nu}(t) \in \sum_{s=m}^{m_1t} \mu_{s-t,j} \mu_{sm} - \bar{\mu}_{\mu jm}(t) \]

has each element bounded in modulus by \( \epsilon \), so

\[ \lim_{T \to \infty} T^{-1} \sum_{t=-T+1}^{m_2t} \Gamma_{\nu\nu}(t) \in \sum_{s=m}^{m_1t} \mu_{s-t,j} \mu_{sm} - \bar{\mu}_{\mu jm}(t). \quad (3.41) \]

Similarly,

\[ \lim_{T \to \infty} \left( T^{-1} \sum_{s=m}^{m_2t} M_{\nu\nu}(0,0,t) \in \sum_{s=m}^{m_1t} \mu_{s} \right) = \left[ \sum_{t=\infty}^{\infty} M_{\nu\nu}(0,0,t) \right] \in \bar{\mu} \]
by the absolute summability of $M_{uv}(0, 0, t)$ and the fact that for any
fixed $t$,
\[
\lim_{T \to \infty} T^{-1} \sum_{m=1}^{m_{2T}} \mu_s = \bar{\mu}.
\]
Thus,
\[
G \equiv \lim_{T \to \infty} \{ T^{-1} \text{Var}[(I_{r} \otimes \mathbf{x}') \text{vec}(v)] \}
\]
\[
= \sum_{t=-\infty}^{\infty} \{ \tau_{uv}(t) \sigma [\bar{M}_{u}u(t) + \tau_{uu}(t)] + \text{vec}[\tau_{uv}(t)][\text{vec}[\tau_{vv}(t)]'
\]
\[
+ [M_{uv}(0, 0, t)]^{\bar{\mu}} + \bar{\mu}'[M_{uv}(0, 0, t)]' + k_{uvuv}(0, 0, t, t) \}.
\]
\[
(3.42)
\]
Given the results above, one may develop some convergence proper­
ties of $(I_{r} \otimes \mathbf{x}') \text{vec}(v)$. First, Lemma 3.10 and 3.11 give conditions
under which this function converges almost surely to $\text{vec}[\tau_{vv}(0)]$. 
Lemma 3.10(1) is a slight variant of Theorems IV.5 and IV.9 of Hannan
(1970), and the proof of Lemma 3.10(2) follows closely the proof of
Hannan's Theorem IV.5. Similarly, Lemma 3.10(2) is part of Theorem
IV.6 of Hannan (1970), and the proof of Lemma 3.10(2) is a more
detailed version of the corresponding part of the proof of Hannan's
Theorem IV.6. A detailed presentation of these results at this point is
desirable because of the application of these results in Chapter 4 and
because Section 3.4 requires a generalization of Lemma 3.11 that is not generally available in the current literature.

**Lemma 3.10.** Let \( \{u_t\} \) be a sequence of fixed \( 1 \times k \) vectors that satisfy condition (3.39.a) and let \( \xi_t = (u_t, v_t) \) be a \( 1 \times p \) fourth-order stationary process with mean zero, where \( u_t \) is \( 1 \times k \), \( v_t \) is \( 1 \times r \), and \( p = r + k \). Assume that for some \( \alpha > 0 \) and for all \( 1 \leq i, j, l, m \leq p \), the covariance function \( \Gamma_{ij}(t) \) and the fourth-order cumulant function \( k_{ijkl}(0, 0, s, t) \) satisfy the conditions

\[
\lim_{T \to \infty} \frac{1}{T^{1+\alpha}} \sum_{t=-T+1}^{T-1} |\Gamma_{ij}(t)| < \infty \tag{3.43.a}
\]

and

\[
\lim_{T \to \infty} \frac{1}{T^{1+\alpha}} \sum_{t=-T+1}^{T-1} |k_{ijkl}(0, 0, s, t)| < \infty \tag{3.43.b}
\]

Let \( X = \xi + u \). Then

\( i \) \( T^{-1}[I_r \otimes y']vec(v)] + O_{rk \times 1} \) with probability one;

\( ii \) \( T^{-1}[I_r \otimes u']vec(v)] - vec[\Gamma_{uv}(0)] + O_{rk \times 1} \) with probability one;

and thus
(iii) \( T^{-1}[(I_r \circ X') \text{vec}(v)] - \text{vec}(\Pi_{uv}(0)) + \Theta_{rk \times 1} \) with probability one.

Proof. Since almost sure convergence of sequences of random vectors is an element-by-element property, it suffices to establish results (i), (ii), and (iii) for the case \( r = k = 1 \).

To develop some necessary notation, choose some \( \beta > 1 \) such that \( \beta \alpha > 1 \). For each nonnegative integer \( M \), let \( T(M) \) be the smallest integer greater than or equal to \( M^\beta \). Note that

\[
M^\beta < T(M) < M^\beta + 1,
\]

so that

\[
0 < T(M + 1) - T(M) < (M + 1)^\beta + 1 - M^\beta.
\]

Also, since \( \beta > 0 \), \( M^\beta \) and \( T(M) \) increase without bound as \( M \to \infty \). Moreover, for all positive integers \( M \),

\[
T(M)^{-1}[T(M + 1) - T(M)] < M^{-\beta}[T(M + 1) - T(M)]
\]

\[
< M^{-\beta}[(M + 1)^\beta + 1 - M^\beta]
\]

\[
= (1 + M^{-1})^\beta + M^{-\beta} - 1
\]
for some positive number $K_2$, where the final inequality follows from the fact that $\beta > 1$.

To establish result (i), define

$$W_T = T^{-1} \sum_{t=1}^T u_t v_t$$

and note that for $T(M) < T < T(M + 1)$, $T^{-1}T(M) < 1$ and

$$|W_T| = |W_T - T^{-1}T(M)W_{T(M)} + T^{-1}T(M)W_{T(M)}|$$

$$< |W_T - T^{-1}T(M)W_{T(M)}| + |W_{T(M)}|.$$ 

Thus, to prove that $W_T \to 0$ almost surely as $T \to \infty$, it suffices to show that

$$\lim_{M \to \infty} |W_{T(M)}| = 0$$ 

and that
122

\[ \lim \{ \max_{T(M) < T < T(M+1)} |W_T - T^{-1}T(M)W_T(M)| \} = 0 \text{ almost surely.} \]

Note that

\[ E(W_T) = 0 \]

and that by (3.37) and the definition of \( B_1 \),

\[ \text{Var}(W_T) = T^{-2} \text{Var}(\mu'\nu) \]

\[ = T^{-1} \sum_{t=-T+1}^{T-1} \Gamma_{\nu\nu}(t) [T^{-1} \sum_{s=m_1,1}^{m_2,1} \mu_{s-c} \mu_s] \]

\[ < T^{-1} \sum_{t=-T+1}^{T-1} |\Gamma_{\nu\nu}(t)| [T^{-1} \sum_{h=1}^{T} \mu_h^2] \]

\[ < B_1 T^{-1} \sum_{t=-T+1}^{T-1} |\Gamma_{\nu\nu}(t)| . \]

Under condition (3.43.a), there exists some \( \alpha > 0 \) and some \( K_1 > 0 \) such that

\[ T^{-1} \sum_{t=-T+1}^{T-1} |\Gamma_{\nu\nu}(t)| < K_1 T^{-\alpha} \]

and hence

\[ \text{Var}(W_T) < B_1 K_1 T^{-\alpha} \]
for all positive integers $T$. Pick $\epsilon > 0$ and note that by the Chebyshev inequality,

$$P[|W_{T(M)}| > \epsilon] < \epsilon^{-2} \frac{\text{Var}[W_{T(M)}]}{\epsilon} < \epsilon^{-2} B_1 K_1 [T(M)]^{-\alpha}$$

$$< \epsilon^{-2} K_1 M^{-\alpha \beta}.$$ 

Thus,

$$\sum_{M=1}^{\infty} P[|W_{T(M)}| > \epsilon] < \epsilon^{-2} K_1 \sum_{M=1}^{\infty} M^{-\alpha \beta} < \infty$$

because $\alpha \beta > 1$. Then by the first Borel-Cantelli lemma [Theorem 4.2.1 of Chung (1974)],

$$P[|W_{T(M)}| > \epsilon \text{ for infinitely many positive integers } M] = 0.$$ 

Since $\epsilon > 0$ was arbitrary, it follows that $|W_{T(M)}|$ converges to zero almost surely as $M \to \infty$.

Next, note that for all $T$ such that $T(M) < T < T(M + 1)$,

$$W_T - T(M) T^{-1} W_{T(M)} = T^{-1} \sum_{t = T(M) + 1}^{T(M+1)} W_t v_t.$$
Next, note that

\[ \ast_{\mathcal{H}} I_{+} = I_{+} \]

\[ \frac{3}{2} \left[ (W)I - (I + W)I \right] (0) \frac{1}{2} (W)I > 0 \]

\[ \ast_{\mathcal{H}} I_{+} = \frac{3}{2} \left[ (W)I - (I + W)I \right] (0) \frac{1}{2} (W)I > 0 \]

\[ \frac{3}{2} \left[ (W)I - (I + W)I \right] (0) \frac{1}{2} (W)I > 0 \]

\[ \{ z \left| (W)I_{+} - I_{+} \right| \} \]

so, \( \ast_{\mathcal{H}} \left[ \frac{3}{2} \left[ (W)I - (I + W)I \right] \right] > 0 \)

so, \( (0)^{\ast_{\mathcal{H}}} > 0 \)

and \( \left[ \frac{3}{2} \left[ (W)I - (I + W)I \right] \right] > 0 \)

Now

\[ \ast_{\mathcal{H}} I_{+} = \frac{3}{2} \left[ (W)I - (I + W)I \right] (0) \frac{1}{2} (W)I > 0 \]

\[ \frac{3}{2} \left[ (W)I - (I + W)I \right] (0) \frac{1}{2} (W)I > 0 \]

\[ \{ z \left| (W)I_{+} - I_{+} \right| \} \]

\[ \{ z \left| (W)I_{+} - I_{+} \right| \} \]
\[ T(M)^{-1} \left[ \sum_{t=1}^{T(M+1)} \mu_t^2 - \sum_{t=1}^{T(M)} \mu_t^2 \right] + T(M)^{-1} \left[ T(M + 1) - T(M) \right] (T(M + 1) - T(M))^{-1} \sum_{t=1}^{T(M+1)} \mu_t^2 \]

\[ + \left| T(M + 1)^{-1} \sum_{t=1}^{T(M+1)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

\[ + \left| T(M)^{-1} \sum_{t=1}^{T(M)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

\[ < M^{-1} K_2 B_1 + \left| T(M + 1)^{-1} \sum_{t=1}^{T(M+1)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

\[ + \left| T(M)^{-1} \sum_{t=1}^{T(M)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

where the final inequality follows from the definitions of \( K_2 \) and \( B_1 \). By the definition of \( \bar{\mu}_{\mu\mu}(0) \) and the fact that \( T(M) \) is increasing without bound in \( M \),

\[ \left| T(M + 1)^{-1} \sum_{t=1}^{T(M+1)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

and

\[ \left| T(M)^{-1} \sum_{t=1}^{T(M)} \mu_t^2 - \bar{\mu}_{\mu\mu}(0) \right| \]

may be made arbitrarily small by choice of sufficiently large \( M \).

Thus, there exists some positive number \( K_3 \) such that
for all positive integers \( M \). Thus,

\[
E\{ \max_{T(M) < T < T(M+1)} \left| W_T - T(M)I_T^{-1}W_{T(M)} \right|^2 \} < \Gamma(0)K_2K_3M^{-2}
\]

for all positive integers \( M \), so by the Markov inequality,

\[
P\{ \max_{T(M) < T < T(M+1)} \left| W_T - T(M)I_T^{-1}W_{T(M)} \right| > \varepsilon \} < \varepsilon^{-2} \Gamma(0)K_2K_3M^{-2}
\]

and hence

\[
\sum_{M=1}^{\infty} P\{ \max_{T(M) < T < T(M+1)} \left| W_T - T(M)I_T^{-1}W_{T(M)} \right| > \varepsilon \} < \varepsilon^{-2} \Gamma(0)K_2K_3 \sum_{M=1}^{\infty} M^{-2} < \infty.
\]

Then by the first Borel-Cantelli lemma,

\[
P\{ \max_{T(M) < T < T(M+1)} \left| W_T - T(M)I_T^{-1}W_{T(M)} \right| > \varepsilon \text{ for infinitely many positive integers } M \} = 0.
\]

Since \( \varepsilon > 0 \) was arbitrary, it follows that
\[
\max_{T(M) < T < T(M+1)} \left| W_T - T(M)T^{-1}w_{T(M)} \right|
\]

converges to zero with probability one as \( M \to \infty \). Thus, \( |W_T| \)
converges to zero with probability one as \( T \to \infty \), so result (i) is
established.

Now define

\[
Y_T = T^{-1} \sum_{t=1}^{T} u_t v_t - \Gamma_{uv}(0)
\]

= \( T^{-1} W'v - \Gamma_{uv}(0) \)

and note that for \( T(M) < T < T(M + 1) \),

\[
|Y_T| < \left| Y_T - T^{-1}T(M)Y_{T(M)} \right| + \left| Y_{T(M)} \right|.
\]

Thus, to prove that \( Y_T \to 0 \) almost surely as \( T \to \infty \), it suffices to
show that

\[
\lim_{M \to \infty} |Y_{T(M)}| = 0 \quad \text{almost surely}
\]

and that

\[
\lim_{M \to \infty} \max_{T(M) < T < T(M+1)} \left| Y_T - T^{-1}T(M)Y_{T(M)} \right| = 0 \quad \text{almost surely}.
\]
Note that $E(Y_\pi) = 0$ and that by expression (3.38),

$$\text{Var}(Y_\pi) = T^{-2} \text{Var}(u'v)$$

$$= T^{-2} \sum_{t=-T+1}^{T-1} (T - |t|) \left[ \Gamma_{vv}(t) \Gamma_{uu}(t) + [\Gamma_{uv}(t)]^2 \right]$$

$$+ k_{uvuv}(0, 0, t, t)$$

$$< T^{-1} \sum_{t=-T+1}^{T-1} \left[ |\Gamma_{vv}(t)| |\Gamma_{uu}(t)| + [\Gamma_{uv}(t)]^2 \right]$$

$$+ |k_{uvuv}(0, 0, t, t)|.$$
Thus,

\[ \sum_{M=1}^{\infty} P[|Y_{T(M)}| > \varepsilon] < e^{-2\varepsilon K}, \sum_{M=1}^{\infty} M^{-a\beta} < \infty \]

because \( a\beta > 1 \). Then by the first Borel-Cantelli lemma,

\[ P[|Y_{T(M)}| > \varepsilon \text{ for infinitely many positive integers } M] = 0. \]

Since \( \varepsilon > 0 \) was arbitrary, it follows that \( |Y_{T(M)}| \) converges to zero almost surely as \( M \to \infty \).

Next, note that for all \( T \) such that \( T(M) < T < T(M + 1) \),

\[ Y_T - T(M)T^{-1}Y_{T(M)} = T^{-1} \sum_{t=T(M)+1}^{T} d_t, \]

where \( d_t = u_t v_t - \Gamma_{uv}(0) \). Moreover, \( |d_s d_t| < 2^{-1}(d_s^2 + d_t^2) \), so

\[ \max_{T(M) < T < T(M+1)} \left[ \left| Y_T - T(M)T^{-1}Y_{T(M)} \right|^2 \right] \]

\[ \leq T(M)^{-2} \left[ \sum_{t=T(M)+1}^{T(M+1)} |d_t|^2 \right] \]

\[ \leq T(M)^{-2} \sum_{s=T(M)+1}^{T(M+1)} \sum_{t=T(M)+1}^{T(M+1)} |d_s d_t| \]

\[ \leq T(M)^{-2} \left[ (T(M + 1) - T(M)) \sum_{t=T(M)+1}^{T(M+1)} d_t^2 \right]. \]

Now \( E(d_t) = 0 \), so by (3.38),
Thus, for all positive integers \( M \),

\[
E\left( \max_{T(M) < T < T(M+1)} \left| Y_T - T(M)T^{-1}Y_{T(M)} \right|^2 \right) < T(M)^{-2}[T(M-1) - T(M)]^2\{\gamma_{vv}(0)\gamma_{uu}(0) + [\gamma_{uv}(0)]^2 + k_{uvuv}(0, 0, 0)\} \\
< M^{-2}K_5
\]

where the final inequality follows from expression (3.44) and

\[
K_5 \equiv K_2^2\{\gamma_{vv}(0)\gamma_{uu}(0) + [\gamma_{uv}(0)]^2 + k_{uvuv}(0, 0, 0)\}.
\]

Then by the Markov inequality,

\[
P\left\{ \max_{T(M) < T < T(M+1)} \left| Y_T - T(M)T^{-1}Y_{T(M)} \right| > \epsilon \right\} < \epsilon^{-2}E\left( \max_{T(M) < T < T(M+1)} \left| Y_T - T(M)T^{-1}Y_{T(M)} \right|^2 \right) \\
< \epsilon^{-2}K_5M^{-2},
\]

and hence
\begin{equation}
\sum_{M=1}^{\infty} P\{ \max_{T(M) < T < T(M+1)} [ |Y_T - T(M)T^{-1}Y_{T(M)}| ] > \varepsilon \} < \varepsilon^{-2} K_5 \sum_{M=1}^{\infty} M^{-2} < \infty.
\end{equation}

Then by the first Borel–Cantelli lemma,

\begin{equation}
P\{ \max_{T(M) < T < T(M+1)} [ |Y_T - T(M)T^{-1}Y_{T(M)}| ] > \varepsilon \\
\text{for infinitely many positive integers } M \} = 0.
\end{equation}

Since \( \varepsilon > 0 \) was arbitrary, it follows that

\begin{equation}
\max_{T(M) < T < T(M+1)} [ |W_T - T(M)T^{-1}W_{T(M)}| ]
\end{equation}

converges to zero with probability one as \( M \to \infty \). Thus, \( |Y_T| \)
converges to zero with probability one as \( T \to \infty \), so result (ii) is
established.

Result (iii) follows immediately from results (i) and (ii).

One may now apply Lemma 3.10 to the special case of a linear
process which satisfies Definition 3.2.

Lemma 3.11. Let \( \{ u_t \} \) be a sequence of fixed \( 1 \times k \) vectors that
satisfies condition (3.39.a). Let \( \zeta' = (u_t, v_t)' \) be a \( p \times 1 \) linear
process defined by (3.36), and define \( X = y + u \). Then as \( T \to \infty \),
\[ T^{-1}(I_r \odot X') \text{vec}(v) - \text{vec}[\Gamma_{uv}(0)] \to 0_{rk \times l} \]

with probability one.

**Proof.** Recall that a linear process that satisfies Definition 3.2 has a covariance function and a fourth-order cumulant function that are absolutely summable. Thus, for \( \alpha = 1 \) and for all \( 1 < i, j, k, m < p \),

\[
\lim_{T \to \infty} T^{-1+\alpha} \sum_{t=-T+1}^{T-1} |\Gamma_{\epsilon\epsilon i j}(t)| = \lim_{T \to \infty} T^{-1} \sum_{t=-T+1}^{T-1} |\Gamma_{\epsilon\epsilon i j}(t)| < \infty
\]

and

\[
\lim_{T \to \infty} T^{-1+\alpha} \sum_{t=-T+1}^{T-1} k_{\epsilon\epsilon\epsilon i j m}(0, 0, t, t) = \lim_{T \to \infty} T^{-1} \sum_{t=-T+1}^{T-1} k_{\epsilon\epsilon\epsilon i j m}(0, 0, t, t) < \infty.
\]

Hence, conditions (3.43) hold for \( \alpha = 1 \). The result then follows from Lemma 3.10. \( \square \)

In addition to the almost sure convergence of \( T^{-1}(I_r \odot X') \text{vec}(v) - \text{vec}[\Gamma_{uv}(0)] \to 0_{rk \times l} \), one may study the asymptotic distribution of this function. In particular, one may wish to establish an asymptotic normality result analogous to Lemma 3.8 in the preceding section. Since \( (I_r \odot X') \text{vec}(v) \) is an \( rk \)-dimensional random vector, the associated asymptotic normality arguments are simplified
The following lemma due to Varadarajan (1958) justifies this reduction from a problem of multivariate asymptotic normality to a problem of univariate asymptotic normality. The version stated below is repeated from Fuller (1976, p. 200).

Lemma 3.12. [Theorem 5.3.3 of Fuller (1976)]. Let \( \{Z_t; t=1, 2, \ldots\} \) be a sequence of \( m \)-dimensional random vectors with distribution functions \( \{F_{Z_t}(z)\} \). Let \( F_{Y_t}(y) \) be the distribution function of \( Y_t = \delta'Z_t \), where \( \delta \) is a fixed \( m \)-dimensional real vector. A necessary and sufficient condition for \( F_{Z_t}(z) \) converge to a \( m \)-variate distribution function \( F_{Z}(z) \) is that \( F_{Y_t}(y) \) converge to a limit for each \( \delta \in \mathbb{R}^m \).


Theorem 3.1 uses Lemmas 3.6, 3.7 and 3.12 to establish the asymptotic normality of

\[
T^{-1/2} \left( (I_r \otimes \mathbf{X}) \text{vec}(\mathbf{v}) - T \text{vec}[\Gamma_{\text{inv}}(0)] \right)
\]

under fairly general conditions.

Theorem 3.1. Let \( \mathbf{y}_t \) be a sequence of fixed \( 1 \times k \) vectors satisfying conditions (3.39) and let \( \mathbf{z}_t = (\mathbf{u}_t, \mathbf{v}_t) \) be a linear process of dimension \( 1 \times p \), where \( \mathbf{u}_t \) is \( 1 \times k \), \( \mathbf{v}_t \) is \( 1 \times r \), and
\[ p = k + r \]. Assume also that for some \( \nu > 0 \),

\[
\max_{1 \leq i \leq p} \sup_{t} \{ |c_{ti}|^{4+2\nu} \} = K_{\nu} < \infty
\]  

(3.46)

where \( \{c_{ti}\} \) is the underlying process that generates \( \{e_{ti}\} \). Define

\[ X_{t} = u_{t} + u_{t} \]. Then as \( T \to \infty \),

\[
T^{-1/2} \{ (I_{r} * X') \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)] \} \xrightarrow{d} N(0, G),
\]  

(3.47)

where \( G \) is defined by formula (3.42) above.

An outline of the proof of Theorem 3.1 is as follows. The Varadarajan (1958) result quoted in Lemma 3.12 allows one to restrict attention to the asymptotic properties of \( \delta'(I_{r} * X') \text{vec}(v) \). For this univariate random variable, the asymptotic normality proof is similar to the proof of Lemma 3.8. First, the central limit result is established for sequences \( \{(u_{t}, v_{t})\} \) which follow a finite moving-average model,

\[
(u_{t}, v_{t})' = \sum_{h=0}^{M} A_{h}c_{t-h}'.
\]  

(3.48)

Under this model define \( \Gamma_{uv}(0) = E[u_{t}'v_{t}] \). Then some preliminary arguments show that

\[
\delta'(I_{r} * X') \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)]
\]

\[ = \sum_{t=1}^{T} (Z_{t}c_{t}' - \text{vec}[\Gamma_{uv}(0)]) + o_{p}(T^{1/2}), \]
where \( Z_t = \mathbf{e}_t + \mathbf{f}_t + g_t \), \( \mathbf{e}_t \) is a fixed vector, \( \mathbf{f}_t \) is a \( \sigma(c_t) \)-measurable random vector, and \( g_t \) is a \( \sigma(c_s, s < t-1) \)-measurable random vector. Examination of the properties of \( e_t c'_t \), \( f_t c'_t \) and \( g_t c'_t \) indicate that the sequence \( \{Z_t c'_t - \text{vec}[\Gamma_{uv}(0)], t=1, 2, \ldots\} \) satisfies two conditions given in Scott (1973). Theorem 1 of Scott (1973) then establishes the asymptotic normality of 
\[
T^{-1/2} \sum_{t=1}^{T} \{Z_t c'_t - \text{vec}[\Gamma_{uv}(0)]\}
\]
and hence of
\[
T^{-1/2} \hat{\xi} \{(I_r \otimes \mathbf{X'})\text{vec}(w) - T \text{vec}[\Gamma_{uv}(0)]\}
\]
when \( \{(u_t, v_t)\} \) follows a finite moving average model.

Next, one may consider the same \( \{c_t\} \) process, an extended, absolutely summable set of weight matrices \( \{A_h\} \), and a linear process
\[
(u_t, v_t)' = \sum_{h=0}^{\infty} A_h c'_{t-h}.
\] (3.49)

Uniformly in \( T \), the difference between
\[
T^{-1/2} \hat{\xi} \{(I_r \otimes \mathbf{X'})\text{vec}(w) - \text{vec}[\Gamma_{uv}(0)]\}
\]
under the infinite moving average model (3.49) and the finite moving average model (3.48) becomes asymptotically negligible as \( M \) increases without bound. The asymptotic normality of \( T^{-1/2} \hat{\xi} \{(I_r \otimes \mathbf{X'})\text{vec}(w) - T \text{vec}[\Gamma_{uv}(0)]\} \) under model (3.49) then follows from the finite moving average result and Lemma 3.7. The multivariate asymptotic normality of
\[
T^{-1/2} \{(I_r \otimes \mathbf{X'})\text{vec}(w) - \text{vec}[\Gamma_{uv}(0)]\}
\]
under model (3.49) then follows from Lemma 3.12.
Proof of Theorem 3.1. Let $\delta$ be an arbitrary $rk \times 1$ real vector with double-subscripted $(i,j)$-th element $\delta_{ij}$ and note that

$$\delta'\{(X_{r} \otimes X')\text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)]\}$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} [X'_{ij} v_{i} - T \Gamma_{uvj}(0)]$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{t=1}^{T} \delta_{ij} [X'_{ij} v_{ti} - \Gamma_{uvj}(0)]. \quad (3.50)$$

By Lemma 3.12, result (3.47) will be established if for all $\delta \in \mathbb{R}^{rk}$, $T^{-1/2} \delta'\{(X_{r} \otimes X')\text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)]\}$ converges in law to a normal $(0, \delta'G\delta)$ distribution as $T \to \infty$. For the time being, fix $i$, $j$, and some positive integer $M$, and consider the case in which $(u_{t}, v_{t})'$ has a finite moving average representation,

$$(u_{t}, v_{t})' = \sum_{h=0}^{M} A_{h} c'_{t-h}$$

$$= [(\sum_{h=0}^{M} B_{h} c'_{t-h})', (\sum_{h=0}^{M} D_{h} c'_{t-h})']', \quad (3.51)$$

where $A_{h} = (B_{h}'', D_{h}')'$ is $p \times p$, $B_{h}$ is $k \times p$ with $h$-th row equal to $B_{h}^{*}$, and $D_{h}$ is $r \times p$ with $h$-th row equal to $D_{h}^{*}$. Fix $M$ temporarily. Under model (3.51), denote the moments and other parameters of the $(u_{t}, v_{t})$ process with the subscript $M$, e.g.,

$$\Gamma_{uv}(z) = \text{Cov}(\sum_{h=0}^{M} B_{h} c'_{t-h}, \sum_{h=0}^{M} D_{h} c'_{t+h})$$
$$h = -M, -M+1, \ldots, M. \text{ Let } y_{tj} = (M+1)^{-1} u_{tj}. \text{ Then}$$

$$T \sum_{t=1}^{T} x_{tj} y_{ti} = \sum_{t=1}^{T} (\mu_{tj} + u_{tj}) y_{ti}$$

$$= \sum_{t=1}^{T} \sum_{g=0}^{M} B_{g} c_{t-g}^{'} \sum_{h=0}^{M} D_{h} c_{t-h}^{'}$$

$$= \sum_{t=1}^{T} \sum_{g=0}^{M} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$= \sum_{t=1}^{T} \sum_{g=0}^{M-1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$+ \sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$= \sum_{t=1}^{T} \sum_{g=0}^{M-1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$+ \sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$+ \sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$= \sum_{t=1}^{T} \sum_{g=0}^{M-1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$+ \sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$+ \sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

because

$$\sum_{g=0}^{M-1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'}) = \sum_{h=0}^{g=1} \sum_{g=0}^{M-1} \sum_{t=1}^{T} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$\sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'}) = \sum_{h=0}^{g=1} \sum_{g=0}^{M-1} \sum_{t=1}^{T} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$

$$\sum_{t=1}^{T} \sum_{g=0}^{g=1} \sum_{h=0}^{M} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'}) = \sum_{g=0}^{g=1} \sum_{h=0}^{M} \sum_{t=1}^{T} (y_{tj} + B_{g} c_{t-g}^{'} D_{h} c_{t-h}^{'})$$
and

\[ \sum_{g=0}^{M-1} \sum_{h=g+1}^{M-1} (y_{tj} + B_{gj} c'_{t-g}) D_{h} c'_{t-h} \]

\[ = \sum_{h=1}^{M} \sum_{g=0}^{h-1} y_{tj} D_{h} c'_{t-h} + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M} B_{gj} c'_{t-h} D_{h} c'_{t-g} \]

\[ = \sum_{h=1}^{M} h y_{tj} D_{h} c'_{t-h} + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M} D_{gj} c'_{t-g} B_{h} c'_{t-h} \]

\[ = (\sum_{g=1}^{M-1} g D_{g} c'_{t-g}) B_{0j} c'_{t} + \sum_{h=1}^{M} (h y_{tj} D_{h} + \sum_{g=h+1}^{M} D_{gj} c'_{t-g} B_{h}) c'_{t-h} \]

\[ + M y_{tj} D_{M} c'_{t-M} . \]

Then

\[ \sum_{t=1}^{T} X_{t} v_{t} - T \Gamma_{Muvj_{1}}(0) = \sum_{t=0}^{T-M} \sum_{h=0}^{h} (y_{t+h,j} + B_{hj} c'_{t}) D_{h} \]

\[ + \sum_{h=0}^{M-1} \sum_{g=h+1}^{M} (y_{t+h,j} + B_{gj} c'_{t+g}) D_{h} + (\sum_{g=1}^{M-1} g D_{g} c'_{t-g}) B_{0j} \]

\[ + \sum_{h=1}^{M} (h y_{t+j} D_{h} + \sum_{g=h+1}^{M} D_{gj} c'_{t-g} B_{h}) \]

\[ + M y_{t+M,j} D_{M} c'_{t} - (T - M) \Gamma_{Muvj_{1}}(0) \quad (3.52) \]
\[
T_{t} = T_{t-t} + \sum_{t=T-M+1}^{T-t} \left\{ \sum_{h=0}^{T-t} \left( y_{t+h,j} + \delta_{ij} c' \right) D_{hi} \right\}
\]

\[
T_{t-t-1} = T_{t-t} + \sum_{h=0}^{T-t} \sum_{g=h+1}^{T-t} \left( y_{t+h,j} + \delta_{ij} c' \right) D_{hi} + \left( \sum_{g=1}^{T-t} \delta_{ij} c' \right) B_{ij}
\]

\[
T_{t-t-1} + \sum_{h=1}^{T-t} h y_{t+h,j} D_{hi} + \sum_{g=h+1}^{T-t} \left( \sum_{g=1}^{T-t} \delta_{ij} c' \right) B_{ij}
\]

\[
+ (T-t)y_{T,t} D_{T-t,i} c' - M \Gamma_{Muvji}(0).
\]

(3.53)

Since \( M \) is finite and is not a function of \( T \), condition (3.39.a) implies that expression (3.53) is \( o_p(T^{1/2}) \), so

\[
T+M \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{T} \delta_{ij} [X_{t} v_{t} - \Gamma_{Muvji}(0)] = T \sum_{t=1}^{T} (Z_{t} c' - \gamma_{M6}) + o(T^{1/2})
\]

(3.54)

where

\[
Z_{t} = e_{t} + f_{t} + g_{t},
\]

\[
e_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} \left\{ \sum_{h=0}^{M-1} y_{t+h,j} D_{hi} + \sum_{h=0}^{M-1} (M-h) y_{t+h,j} D_{hi} + \sum_{h=1}^{M} h y_{t+h,j} D_{hi} \right\}
\]

\[
e_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} y_{t+h,j} (M+1) D_{hi}
\]

\[
e_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} y_{t+h,j} D_{hi}
\]

\[
e_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} y_{t+h,j} D_{hi}
\]

\[
f_{t} = c'_{t} \left( \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} B'_{ij} D_{hi} \right)
\]
\[ g_t = \sum_{r} \sum_{k} \delta_{ij} \left( \sum_{h=0}^{M-1} \delta_{ij} (B_{ij} + B_{ij}^r) \right) \]

\[ = \sum_{r} \sum_{k} \delta_{ij} \left( \sum_{h=0}^{M-1} \delta_{ij} (B_{ij} + B_{ij}^r) \right) \]

\[ = \sum_{r} \sum_{k} \delta_{ij} \left( \sum_{h=0}^{M-1} \delta_{ij} (B_{ij} + B_{ij}^r) \right) \]

\[ = \sum_{r} \sum_{k} \delta_{ij} \left( \sum_{h=0}^{M-1} \delta_{ij} (B_{ij} + B_{ij}^r) \right) \]

and

\[ \gamma_{M_0} = \sum_{i=1}^{r} \sum_{j=1}^{k} \gamma_{M_{ij}}(0) = \delta^r \text{vec} [\Gamma_{M_{ij}}(0)] . \]

Now \( e_t \) is a fixed vector, \( f_t \) is a random vector that is measurable \( \sigma(e_t) \), and \( g_t \) is a random vector that is measurable \( F_{t-1} \).

Martingale conditions (3.2) imply that \( E[e_t e_t^t | F_{t-1}] = 0 \); \n
\[ E[f_t e_t^t | F_{t-1}] - \gamma_{M_0} = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} \left[ \sum_{h=0}^{M} h_{ij} \gamma_{M_{ij}}(0) \right] = 0 \]

because

\[ \gamma_{M_{ij}}(0) = \text{Cov}(u_{tj}, v_{ti}) \]

\[ = \text{Cov}(\sum_{g=0}^{M} \delta_{ij} (B_{ij}^g), \sum_{h=0}^{M} \delta_{ij} (B_{ij}^h)) \]

\[ = \sum_{h=0}^{M} \delta_{ij} \text{Cov}(B_{ij}^h, B_{ij}^h) \]

\[ = \sum_{h=0}^{M} \delta_{ij} \text{Cov}(B_{ij}^h, B_{ij}^h) \]
and

\[ E[g_t c'_t | F_{t-1}] = g_t E(c'_t | F_{t-1}) = 0. \]

Also, the following moment matrices are constant over \( t \):

\[ C_{Mcfc} = \text{Cov}(c'_t, f_t c'_t); \]

\[ V_{Mcf} = \text{Var}(f_t c'_t) = \text{Var}[c_t (\Sigma \Sigma \delta_{ij} \Sigma h_j D_i c'_t)]; \]

and

\[ V_{Mcf} = E(g_t \Sigma g'_t). \]

In addition,

\[ d^2_{Mtt} = E[(Z_t c'_t - \gamma_{M0})^2 | F_{t-1}] \]

\[ = \text{Var}[Z_t c'_t | F_{t-1}] \]

\[ = \text{Var}[(e_t + f_t + g_t)c'_t | F_{t-1}] \]

\[ = (e_t + g_t) \Sigma c_t (e_t + g_t)' + 2(e_t + g_t)\text{Cov}(c'_t, f_t c'_t) + \text{Var}(f_t c'_t) \]

\[ = (e_t + g_t) \Sigma c_t (e_t + g_t)' + 2(e_t + g_t)C_{Mfc} + V_{Mcf}; \quad (3.55) \]
so

\[ E[(Z_t c'_t - \gamma_{M0})^2] = E[E[(Z_t c'_t - \gamma_{M0})^2|F_{t-1}]] \]

\[ = e_t \sum_{c_t} e'_t + V_{Mcg} + 2e_t c_{Mcf} + V_{Mcf}. \]

Then \( E[Z_t c'_t - \gamma_{M0}|F_{t-1}] = 0 \) implies that

\[ E[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{M0})|F_{t-1}] = \sum_{s=1}^{t-1} (Z_s c'_s - \gamma_{M0}) \]

and

\[ \text{Var}[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{M0})|F_{t-1}] = \text{Var}[(Z_t c'_t - \gamma_{M0})|F_{t-1}], \]

so

\[ s^2_{Mtt} = \text{Var}[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{M0})] \]

\[ = E[\text{Var}[(Z_t c'_t - \gamma_{M0})|F_{t-1}]] + \sum_{s=1}^{t-1} \text{Var}[(Z_s c'_s - \gamma_{M0})|F_{t-1}] \]

\[ = \sum_{s=1}^{t} E[\text{Var}[(Z_s c'_s - \gamma_{M0})|F_{s-1}]] \]

\[ = \sum_{s=1}^{t} [e_s \sum_{c_s} e'_s + V_{Mcg} + 2e_s c_{Mcf} + V_{Mcf}] \].

Conditions (3.2), (3.39.a), and (3.39.d) imply that
exists and is finite. Assume first that $s_M^2 = 0$. Since $E(Z_s c_s' - \gamma_{M\delta}) = 0$, it follows from the definition of $s_{Mtt}^2$ that

$$\lim_{t \to \infty} t^{-1} \frac{\sum_{s=1}^{t} [ \sum_{s} (Z_s c_s' - \gamma_{M\delta})]^2}{s} = 0,$$

so by the Chebyshev inequality,

$$\lim_{t \to \infty} t^{-1/2} \frac{\sum_{s=1}^{t} (Z_s c_s' - \gamma_{M\delta})}{s} = 0.$$

Thus, in a trivial sense,

$$T^{-1/2} \sum_{s=1}^{T} (Z_s c_s' - \gamma_{M\delta}) - \mathcal{F} \Rightarrow N(0, \bar{s}_M^2)$$

when $\bar{s}_M^2 = 0$. Now assume that $\bar{s}_M^2 > 0$. Note that the inequalities

$$2ab < a^2 + b^2 < 2 \max(a^2, b^2)$$

imply that

$$|a + b|^{2+\nu} = |a^2 + 2ab + b^2|^{1+2^{-1}\nu} < 2^{1+2^{-1}\nu}|a^2 + b^2|$$
Given the above results, the sequence \( \{Z_t c'_t - \gamma_{M_g}\} \) will satisfy the assumptions of Theorem 1 of Scott (1973) if the following conditions hold:

\[
\sum_{t=1}^{T} \mathbb{E}((Z_t c'_t - \gamma_{M_g})^2 | F_{t-1}) \overset{P}{\to} 1 \quad \text{as} \quad T \to \infty
\]

and

\[
\sum_{t=1}^{T} \mathbb{E}((Z_t c'_t - \gamma_{M_g})^2 \mathbb{I}(|Z_t c'_t - \gamma_{M_g}| > s_{IT}^T | F_{t-1}|) \overset{P}{\to} 0 \quad \text{as} \quad T \to \infty
\]

Results (3.55) and (3.57) imply that (3.60) will be established if

\[
T^{-1} \sum_{t=1}^{T} g_t c_t - V_{M_g} \overset{P}{\to} 0
\]

and

\[
T^{-1} \sum_{t=1}^{T} e_t c_t \overset{P}{\to} 0
\]

and

\[
T^{-1} \sum_{t=1}^{T} g_t \overset{P}{\to} 0
\]

Now \( g_t c_t c'_t \) is a sequence of random variables with mean \( V_{M_g} \) and common finite variance. Moreover, since the sequence \( \{c'_t c_t\} \) is
uncorrelated, the definition of \( g_t^* \) implies that for \( |t - s| > M \), \( g_t \Sigma g_t^* + \Sigma g_s g_s^* \) are uncorrelated. If \( T = Nq + w, 0 < w < M \), then one may write (3.62) as the sum of \( M \) separate terms, with the \( i \)-th term equal to

\[
-T^{-1} \sum_{j=0}^{q-1} (g_{jM+i} + \Sigma g_{jM+i}^* - V_{Mg}) \text{, if } 0 < i < w;
\]

or

\[
-T^{-1} \sum_{j=0}^{q-1} (g_{jM+i} + \Sigma g_{jM+i}^* - V_{Mg}) \text{, if } w < i < M.
\]

By Lemma 3.4, each of these \( M \) sums converges almost surely to zero as \( T \rightarrow \infty \). Since \( M \) is constant with respect to \( T \), it follows that the convergence in (3.61) holds with probability one. Almost sure convergence in (3.64) is established similarly. Also, for \( i = 1, 2, \ldots, M \), \( i < w \) implies that

\[
\text{Var}[T^{-1} \sum_{j=0}^{q} e_{jM+i} + \Sigma e_{jM+i}^*] = T^{-2} \sum_{j=0}^{q} e_{jM+i} + V_{Mg} \Sigma e_{jM+i}^*\]

and similarly for \( i > w \). Condition (3.39.b) implies that expression (3.65) converges to zero, so result (3.63) follows by Chebyshev's inequality.

Lemma 3.6 implies that

\[
\varepsilon_{\text{MTT}}^{-2} \sum_{t=1}^{T} E[(Z_t c_t^* - \gamma_{M\delta})^2 \delta_t^2 | Z_t c_t^* - \gamma_{M\delta} | > \varepsilon_{\text{MTT}}^2]
\]
By condition (3.46), the variables \(|(f_t + g_t)c_t' - \gamma_{M6}|\) have a common finite \(2 + \nu\) moment, so

\[
\lim_{T \to \infty} T^{-1-2^{-1-1} \nu} T \sum_{t=1}^{T} E[|(f_t + g_t)c_t' - \gamma_{M6}|^{2+\nu}] = 0 .
\]

Also, (3.39.b) implies that

\[
\lim_{T \to \infty} T^{-2-1 \nu} \max_{1 \leq t \leq T} (|e_{t, l}|^{\nu}) = 0 , \quad l = 1, 2, ..., r ,
\]

and (3.39.a) implies that \(\lim_{T \to \infty} T^{-2-1 \nu} \sum_{l=1}^{r} |e_{t, l}|^{2+\nu}\) is finite for all \(l = 1, 2, ..., r\). Thus,

\[
\lim_{T \to \infty} T^{-1-2^{-1} \nu} T \sum_{t=1}^{T} |e_{t, l}|^{2+\nu} = 0 ,
\]

for all \(l = 1, 2, ..., r\), so the existence of common finite \(2 + \nu\) moments of \(c_t\) implies that
\[
\lim_{T \to \infty} \left\{ T^{-1-2^{-1}} \sum_{t=1}^{T} \mathbb{E} \left[ \left| e_t c_t' \right|^{2+\nu} \right] \right\} = 0.
\]

Then (3.58) implies that (3.66) converges to zero as \(T \to \infty\), so condition (3.61) follows from the Markov inequality. Then by Theorem 1 of Scott (1973),

\[
s_{MTT}^{-1} \sum_{t=1}^{T} \frac{(Z_t c_t' - \gamma_{M0})}{2} \xrightarrow{d} N(0, 1)
\]

under model (3.51). Then by condition (3.57) and the assumption that \(s_M^2 > 0\),

\[
T^{-1/2} \sum_{t=1}^{T} (Z_t c_t' - \gamma_{M0}) \xrightarrow{d} N(0, s_M^2)
\]

as \(T \to \infty\). This result, and conclusion (3.58) for the case \(s_M^2 = 0\), imply that conclusion (3.57) holds regardless of whether \(s_M^2\) is positive or zero.

Now consider the more general case

\[
(u_t, v_t)' = \sum_{h=0}^{\infty} A_h c_t' = \left[ \sum_{h=0}^{\infty} B_h c_t', \sum_{h=0}^{\infty} D_h c_t' \right]'.
\]

Unless indicated otherwise, retain the notation given above. Note that the difference between \((I_r \otimes X')\text{vec}(v) - T\text{vec}(I_r \otimes 0)\) under model (3.68) and \((I_r \otimes X')\text{vec}(v) - T\text{vec}(I_r \otimes 0)\) under model (3.51) has
(i,j)-th element equal to

\[
\begin{align*}
T \sum_{t=1}^{\infty} \left[ \sum_{h=0}^{n} \{ u_{tj} + \sum_{h=0}^{n} B_{h} c_{t-h}' \} \sum_{h=0}^{n} D_{h} c_{t-h}' \} - \Gamma_{uvj}(0) \right] \\
- \sum_{t=1}^{\infty} \left[ \sum_{h=0}^{n} \{ u_{tj} + \sum_{h=0}^{n} B_{h} c_{t-h}' \} \sum_{h=0}^{n} D_{h} c_{t-h}' \} - \Gamma_{uvj}(0) \right] \\
= \sum_{t=1}^{\infty} \left( \sum_{h=0}^{n} \left( D_{h} c_{t-h}' \right) + \sum_{h=0}^{n} B_{h} c_{t-h}' \right) D_{h} c_{t-h}' \\
+ \sum_{h=0}^{n} \left( \sum_{g=0}^{n} B_{g} c_{t-g}' \right) D_{h} c_{t-h}' + \sum_{h=0}^{n} \left( \sum_{g=0}^{n} B_{g} c_{t-g}' \right) D_{h} c_{t-h}' \\
- \left( \sum_{h=0}^{n} \left( \sum_{g=0}^{n} B_{g} c_{t-h}' \right) - \left( \sum_{h=0}^{n} \left( \sum_{g=0}^{n} B_{g} c_{t-h}' \right) \right) \right) \\
= W_{TM1ij} + W_{TM2ij} + W_{TM3ij} + W_{TM4ij},
\end{align*}
\]

say, where

\[
W_{TM1ij} = \sum_{t=1}^{\infty} u_{tj} \left( \sum_{h=0}^{n} D_{h} c_{t-h}' \right),
\]

\[
W_{TM2ij} = \sum_{t=1}^{\infty} \sum_{h=0}^{n} B_{h} c_{t-h}' \left( \sum_{g=0}^{n} D_{h} c_{t-h}' \right),
\]

\[
W_{TM3ij} = \sum_{t=1}^{\infty} \sum_{g=0}^{n} B_{g} c_{t-g}' \left( \sum_{h=0}^{n} D_{h} c_{t-h}' \right),
\]

and

\[
W_{TM4ij} = \sum_{t=1}^{\infty} \sum_{g=0}^{n} B_{g} c_{t-g}' \left( \sum_{h=0}^{n} D_{h} c_{t-h}' \right) - \sum_{h=0}^{n} B_{h} c_{t-h}' \left( \sum_{g=0}^{n} D_{h} c_{t-h}' \right) \left( \sum_{h=0}^{n} \left( \sum_{g=0}^{n} B_{g} c_{t-h}' \right) \right). \]
Then by Lemma 3.7, the asymptotic distribution of
\( \Delta' \{ (I_r \otimes X') \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)] \} \) under model (3.68) will be established if

\[
T^{-1/2} \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} (W_{TM1ij} + W_{TM2ij} + W_{TM3ij} + W_{TM4ij})
\]
satisfies condition (3.10.a) uniformly in \( T \) and if the \( N(0, \bar{s}_{M}^2) \) distribution converges completely to the \( N(0, \Delta'G\Delta) \) distribution as \( M \to \infty \).

First, for any \( 1 < i < r, \ 1 < j < k \),

\[
T^{-1/2} W_{TM1ij} = T^{-1/2} \sum_{t=1}^{T} \sum_{h=M+1}^{\infty} \nu_{tj} \sum_{s=-\infty}^{t-1} \mu_{tj} D_{t-s, i} c'_{s}.
\]

Because the \( c_{s} \) vectors have mean zero and are uncorrelated, \( T^{-1/2} W_{TM1ij} \) has mean zero and variance equal to

\[
\begin{align*}
T^{-1} \sum_{s=-\infty}^{t=\max(1,s+M+1)} & \sum_{t=1}^{T-M-1} \sum_{\xi=\max(1,s+M+1)}^{T} \{ D_{t-s, i} \sum_{\xi}^{D'_{i}} \sum_{s=-\infty}^{t-1} \mu_{tj} \mu_{kJ} \} \\
& = T^{-1} \sum_{s=-\infty}^{t=\max(1,s+M+1)} \sum_{t=1}^{T-M-1} \sum_{\xi=\max(1,s+M+1)}^{T} \{ D_{t-s, i} \sum_{\xi}^{D'_{i}} \sum_{s=-\infty}^{t-1} \mu_{tj} \mu_{kJ} \} \\
& + T^{-1} \sum_{s=-\infty}^{t=M+1} \sum_{t=M+1}^{T-M-1} \sum_{\xi=\max(1,s+M+1)}^{T} \{ D_{t-s, i} \sum_{\xi}^{D'_{i}} \sum_{s=-\infty}^{t-1} \mu_{tj} \mu_{kJ} \} .
\end{align*}
\]

(3.69) (3.70)

Note that the inequalities \( t > 1 \) and \( -\infty < s < -M \) imply that
\[ \begin{align*}
M+1 < t-s < \infty. \ Let \ D_{hi}^m \ denote \ the \ m-th \ element \ of \ the \ row \ vector \\
\text{Denote the m-th element of the row vector } D_{hi}, \ m=1, 2, \ldots, p. \ Then \ expression \ (3.69) \ is \ bounded \ in \ modulus \ by \\
\left| T^{-1} \sum_{s=\infty}^{t-s} D_{t-s,im} \sum_{n=1}^{\infty} D_{t-s,im} n^{m} c c m n D_{t-s,im} n^{m} c c m n \right| \\
< p \sum_{m=1}^{\infty} n=1 h=M+1 D_{hi}^m \left| \sum_{t=1}^{T} D_{t-s,im} n^{m} c c m n D_{t-s,im} n^{m} c c m n \right| .
\end{align*} \]

Now for all \( T \in \mathbb{Z}^+ \),
\[\begin{align*}
T^{-1} \sum_{t=1}^{T} D_{t-s,im} n^{m} c c m n D_{t-s,im} n^{m} c c m n < T \sum_{d=-T+1}^{m} D_{d-h,in} \left| \sum_{t=1}^{T} D_{t-s,im} n^{m} c c m n D_{t-s,im} n^{m} c c m n \right| \\
< B_1 \sum_{d=-T+1}^{m} D_{d-h,in} \left| D_{d-h,in} \right| \\
< B_1 \sum_{d=-T+1}^{m} D_{d-h,in} \left| D_{d-h,in} \right| \\
= B_2,
\end{align*}\]

say. \ This \ final \ expression \ is \ finite \ by \ the \ absolute \ summability \ of \\
\{D_h\} \ and \ the \ definition \ of \ B_1. \ Hence, \ expression \ (3.71) \ is \ bounded 
uniformly \ in \ \( T \) \ by \ the \ expression,
\[\begin{align*}
P \sum_{m=1}^{\infty} n=1 h=M+1 D_{hi}^m \left| \sum_{t=1}^{T} D_{t-s,im} n^{m} c c m n D_{t-s,im} n^{m} c c m n \right| .
\end{align*}\]
The absolute summability of \( \{D_n\} \) then implies that uniformly in \( T \), expression (3.71) may be made arbitrarily small by choice of sufficiently large \( M \). Also, expression (3.70) is bounded in modulus by

\[
\frac{T+M-1}{T+M} \frac{T+M-1}{T+M-1} \min(T-t, T-\ell) \sum_{t=M+1}^{T+M-1} \sum_{\ell=M+1}^{T+M-1} \left| D_{t,\ell} \sum_{i=1}^{\infty} \sum_{s=-M}^{M} \sum_{j} \nu_{t+s,j} \nu_{\ell+s,j} \right|
\]

Also, expression (3.70) is bounded in modulus by

\[
\sum_{t=M+1}^{T+M-1} \sum_{\ell=M+1}^{T+M-1} \min(T-t, T-\ell) \sum_{i=1}^{\infty} \sum_{s=-M}^{M} \sum_{j} \nu_{t+s,j} \nu_{\ell+s,j} \]

Hence, uniformly in \( T \), expression (3.70) converges to zero as \( M \) increases without bound. Then uniformly in \( T \), \( E[T^{-1/2}W_{TMij}] \) converges to zero as \( M \to \infty \), so by the Markov inequality, \( T^{-1/2}W_{TMij} \) converges in probability to zero uniformly in \( T \) as \( M \to \infty \).

Next, note that

\[
T^{-1/2}W_{TM2ij} = T^{-1/2} \sum_{t=1}^{T} \sum_{g=M+1}^{\infty} \sum_{h=0}^{M} B_{ij} c_t' D_{hi} c_{t-h}'
\]

has mean zero and variance less than or equal to
For a given quadruple \((g, h, \lambda, m)\), the indices \(t-g\), \(t-h\), \(s-\lambda\), \(s-m\) are all equal for at most one pair \((t, s)\), in which case the expectation above is bounded above by some \(b_1 > 0\); and the indices \(t-g\), \(t-h\), \(s-\lambda\), and \(s-m\) are equal in pairs for at most \(T\) pairs \((t, s)\), in which case the expectation above is bounded above by some \(b_2 > 0\); otherwise, the expectation above equals zero. Let \(b_3 = \max(b_1, b_2)\). Then (3.73) is bounded above by

\[
b_3 \sum_{h=0}^{\infty} \sum_{n=1}^{\infty} |D_{\text{hin}}| \sum_{m=0}^{\infty} \sum_{b=1}^{\infty} |D_{\text{mib}}| \sum_{g=M+1}^{\infty} \sum_{q=1}^{\infty} |B_{gjq}| \sum_{\ell=M+1}^{\infty} \sum_{a=1}^{\infty} |B_{\lambda ja}|,
\]

which is not a function of \(T\). The absolute summability of \(A_j\) implies that the second and third factors above are finite and that the fourth and fifth factors can be made arbitrarily small by choice of sufficiently large \(M\). Hence, uniformly in \(T\), \(E[T^{-1/2}W_{TM2ij}]\) converges to zero as \(M + \infty\), so by the Chebyshev inequality, \(T^{-1/2}W_{TM2ij}\) converges in probability to zero uniformly in \(T\) as \(M + \infty\). Similar arguments establish that \(T^{-1/2}W_{TM3ij}\) and
$T^{-1/2} W_{TM4ij}$ each converge in probability to zero uniformly in $T$ as $M \to \infty$. Since $r$ and $k$ are fixed finite integers, it follows that

$$T^{-1/2} \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij}(W_{TM1ij} + W_{TM2ij} + W_{TM3ij} + W_{TM4ij})$$

satisfies condition (3.10.a) uniformly in $T$. Thus, it remains only to establish that the $N(0, \tilde{s}^2)$ distribution converges completely to the $N(0, \delta'^G\delta)$ distribution as $M \to \infty$. An argument similar to the one given preceding expression (3.33) indicates that it suffices to show that

$$\lim_{M \to \infty} \tilde{s}^2 = \delta'^G\delta.$$

For a given $M$, (3.56) implies that

$$\tilde{s}^2 = \lim_{T \to \infty} T^{-1} \text{Var}[\sum_{t=1}^{T} (Z_t e'_t - \gamma_s)]$$

$$= \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \left[ e_t \xi_t e'_t + \nu_{MCG} + 2e_t C_{MCF} + V_{MCF} \right]$$

$$= \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \left[ \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{m=1}^{M} \sum_{n=1}^{M} (\xi_{t,\mu} e'_t + \nu_{MCG} + 2e_t C_{MCF} + V_{MCF}) \right]$$

$$+ \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \left[ \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=0}^{D \delta'} M_{CF} + V_{MCG} + V_{MCF} \right]$$. 


Also,\n\[
V_{MCG} = \mathbb{E}[g_t^c g'_t^c]\n\]
\[
= \sum_i \sum_j \sum_k \sum_m \sum_h \sum_{n=0}^{M-h} \delta_{ij} \delta_{km} \delta_{m}^n (n-h) D_{hi} \sum_{c} D_{c n}^{'}
\]
\[
+ \sum_i \sum_j \sum_m \sum_h \sum_{n=0}^{M-h} \delta_{ij} \delta_{km} \delta_{m}^{n} [2 \mathbb{E}_{MCG} + \mathbb{V}_{MCG} + \mathbb{V}_{MCF}]
\]
\[
\text{which, as } M \to \infty, \text{ has limit}
\]
\[
= \delta'(\mathbb{E}_{\nu V}(s) \ast \mathbb{E}_{\nu U}(s)) \delta_{ij} \delta_{km} \delta_{m}^{n} (n-h) D_{hi} \sum_{c} D_{c n}^{'}
\]
Similar arguments establish that
\[ \lim_{M \to \infty} \Sigma_{i=1}^{r} \Sigma_{j=1}^{k} \Sigma_{l=1}^{m} \Sigma_{h=0}^{n} \delta_{ij} \delta_{lm} \mu_{ijlm} (n - h) D_{h} \Sigma_{i} D_{l}' \lim_{s \to -\infty} \delta' \Sigma_{i} \Gamma_{uv}(s) \mu_{ijlm}(s) \delta, \]

\[ = \delta' \Sigma_{i} \Gamma_{uv}(s) \mu_{ijlm}(s) \delta, \]

\[ = \delta' \Sigma_{i} \mu_{ijlm}(s) \delta, \]

\[ \lim_{M \to \infty} \Sigma_{i=1}^{r} \Sigma_{j=1}^{k} \Sigma_{l=1}^{m} \Sigma_{h=0}^{n} \delta_{ij} \delta_{lm} D_{h} \Sigma_{i} D_{l}' \]

\[ = \delta' \Sigma_{i} \mu_{ijlm}(s) \delta, \]

\[ \text{and} \]

\[ \lim_{M \to \infty} V_{M \text{cf}} = \lim_{M \to \infty} \text{Var}[c_{i} \Sigma_{i=1}^{r} \Sigma_{j=1}^{k} \Sigma_{l=1}^{m} \Sigma_{h=1}^{n} \delta_{ij} \delta_{lm} D_{h} \Sigma_{i} D_{l}' c_{i}'], \]

\[ = \delta' \Sigma_{i=1}^{r} \Sigma_{j=1}^{k} \Sigma_{l=1}^{m} \Sigma_{h=1}^{n} \delta_{ij} \delta_{lm} \Sigma_{i} \mu_{ijlm}(s) \delta, \]

\[ \text{Thus, } \lim_{M \to \infty} \Sigma_{i=1}^{r} \Sigma_{j=1}^{k} \Sigma_{l=1}^{m} \Sigma_{h=1}^{n} \delta_{ij} \delta_{lm} \mu_{ijlm}(s) \delta, \]

\[ = \delta' \Sigma_{i} \mu_{ijlm}(s) \delta, \]

\[ \text{for all } \delta \in \mathbb{R}^{r \times k}. \]

Then by Lemma 3.7,

\[ T^{-1/2} \delta' (I_{r} \otimes X') \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)] \xrightarrow{\mathcal{D}} N(0, \delta' \delta) \text{ as } T \to \infty. \]

Since the vector \( \delta \) was an arbitrary element of \( \mathbb{R}^{r \times k} \), it follows from Lemma 3.12 that

\[ T^{-1/2} (I_{r} \otimes X') \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)] \xrightarrow{\mathcal{D}} N_{rk}(0, \delta). \]
3.4. Asymptotic Properties of the Sum of a Linear Function and a Weighted Bilinear Function

The preceding section addressed the asymptotic behavior of

\[ \text{Vec}(X'v) = (I^B X') \text{vec}(v), \]

an unweighted bilinear function of \( X \) and \( v \). As indicated in Section 4.2 below, the behavior of a weighted function of \( X \) and \( v \) may also be of interest in errors-in-variables estimation for serially correlated observations. Consequently, this section considers the \( r_k \times 1 \) vector \((I^B X') \Pi(T) \text{vec}(v)\) where \( \Pi(T) \) is an \( r_T \times r_T \) matrix of weights. In order to use relatively simple asymptotic arguments while permitting the matrix \( \Pi(T) \) to have a fairly general form, some parts of this section will also address two related functions. These functions are \((I^B X') \Pi \text{vec}(v)\) and \((I^B X') \tilde{\Pi} \text{vec}(v)\), where \( \Pi \) is an \( r_T \times r_T \) matrix with a given element equal to the limiting value of the corresponding element of \( \Pi(T) \) as \( T \to \infty \), and \( \tilde{\Pi} \) is an \( r_T \times r_T \) block-Toeplitz matrix that is closely related to the matrix \( \Pi \).

An outline of the results of this section is as follows. First, some preliminary remarks lead to the evaluation of the mean and variance of \((I^B X') \Pi(T) \text{vec}(v)\) in expressions (3.78) and (3.82), respectively. Second, Lemma 3.13 addresses the convergence properties of

\[ (I^B X')[\Pi(T) - \Pi] \text{vec}(v) - E((I^B X')[\Pi(T) - \Pi] \text{vec}(v)) \]

and of
Third, Lemma 3.14 gives conditions for the almost sure convergence of

$$T^{-1}E((I_{r} \otimes X')[(\pi(T) - \pi)vec(v)])$$

to zero. Lemma 3.15 applies Lemma 3.14 to the case in which the random vectors $\xi_t$ are fourth-order stationary and have absolutely summable covariance function, third-cumulant function, and fourth-cumulant function. Fourth, Lemma 3.16 gives conditions under which

$$T^{-1}var((I_{r} \otimes X')\pi vec(v))$$

converges to a limiting matrix, and gives an explicit expression for this limiting covariance matrix. This limiting covariance matrix depends on the entries of the weight matrix $\pi$.

Fifth, Lemma 3.17 addresses the convergence properties of

$$(I_{r} \otimes X')[\pi - \pi]vec(v) - E((I_{r} \otimes X')[\pi - \pi]vec(v)),$$

$$(I_{r} \otimes X')[(\pi(T) - \pi)vec(v) - E((I_{r} \otimes X')[(\pi(T) - \pi)vec(v)),$$

$$T^{-1}E((I_{r} \otimes X')[\pi - \pi]vec(v)),$$

and

$$T^{-1}E((I_{r} \otimes X')[(\pi(T) - \pi)vec(v)).$$

Sixth, Theorem 3.2 gives conditions under which
\[ T^{-1/2} \{ (I_r \otimes X') \Sigma(T) \text{vec}(v) - E[(I_r \otimes X') \Sigma(T) \text{vec}(v)] \} \]

is asymptotically normal. Finally, Corollary 3.2.1 applies the results of Lemma 3.17 and Theorem 3.2 to establish conditions under which

\[ T^{-1/2} \{ (I_r \otimes X') \Sigma(T) \text{vec}(v) - E[(I_r \otimes X') \Sigma(T) \text{vec}(v)] \} \]

and

\[ T^{-1/2} \{ (I_r \otimes X') \Sigma(T) \text{vec}(v) - E[(I_r \otimes X') \Sigma(T) \text{vec}(v)] \} \]

are asymptotically normal.

Let \( \pi_{(T)} \) contain \( r^2 \times r^2 \) blocks with \((i,j)\)-th such block equal to \( \pi_{(T)}^{ij} \), which in turn has \((s,t)\)-th element equal to \( \pi_{(T)}^{ij}(s,t) \).

Define \( \pi_{(T)}^{(s,t)} \) to be the \( r \times r \) matrix with \((i,j)\)-th element equal to \( \pi_{(T)}^{ij}(s,t) \). An important example of \( \pi_{(T)} \) that is discussed in Chapter 4 is \( \pi_{(T)} = \Sigma^{-1} \). Since \( \Sigma^{-1} \) is in general not of block-Toeplitz form [see Theorem 8.15.4 of Graybill (1983) for a counterexample with \( r = 1 \)], \( \pi_{(T)} \) will not be assumed to be of block-Toeplitz form. Instead, assume that for any \( 1 < i, j < r \),

\[ \sum_{s=1}^{T} |\pi_{(T)}^{ij}(s,t)| \text{ is uniformly bounded in } t, T \in \mathbb{Z}^+ ; \quad (3.74.a) \]

\[ \sum_{t=1}^{T} |\pi_{(T)}^{ij}(s,t)| \text{ is uniformly bounded in } s, T \in \mathbb{Z}^+ ; \quad (3.74.b) \]

and that for any \( \varepsilon > 0 \) and any positive integer \( t \), there exists some
positive integer $T > t \epsilon + t$, such that for all $T > t \epsilon + t$,

\[
\sum_{s=T \epsilon + t}^{T} \left| \pi_{ij}(T)(s, t) \right| < \varepsilon \quad \text{(3.75.a)}
\]

and

\[
\sum_{s=T \epsilon + t}^{T} \left| \pi_{ij}(T, s) \right| < \varepsilon \quad \text{(3.75.b)}
\]

In addition, assume that for all $1 < i, j < r$ and all fixed $s, t \in \mathbb{Z}^+$,

\[
\pi_{ij}(s, t) = \lim_{T \to \infty} \pi_{ij}(T)(s, t) \quad \text{(3.76)}
\]

exists and is finite.

Appendix B gives some definitions and properties associated with sequences of matrices that satisfy conditions (3.74), (3.75), and (3.76).

Consider the case in which $Z(t)$ has block-Toeplitz form with $t$-th diagonal, subdiagonal or superdiagonal filled with $r \times r$ blocks equal to $\tilde{\lambda}(t)$, where $\tilde{\lambda}(t)$ is constant over $T$. If $\tilde{\lambda}(t)$ is absolutely summable over $t$, then for any $s \in \mathbb{Z}$ and any $1 < i, j < r$,

\[
\sum_{t=1}^{T} \left| \pi_{ij}(t-s) \right| < \sum_{\ell=-\infty}^{\infty} \left| \pi_{ij}(\ell) \right| < \infty \quad \text{(3.77)}
\]
so conditions (3.74) and (3.75) follow immediately. Moreover, condition (3.76) holds trivially.

Now consider some random functions of $\pi(T)$. Let $\pi_i(T)$ be a $T \times T$ matrix equal to $[\pi_{i1}(T), \pi_{i2}(T), \ldots, \pi_{ir}(T)]$, $i=1, 2, \ldots, r$. Similarly, define

$$\pi_{ij}(T) = [\pi_{i1j}(T), \pi_{i2j}(T), \ldots, \pi_{irj}(T)]'$$

$j=1, 2, \ldots, r$. Then the $r \times 1$ vector $(I_r \otimes X')\pi(T)\text{vec}(v)$ has $(i,j)$-th element equal to

$$X' \cdot \pi_{ij}(T)\text{vec}(v) = \sum_{h=1}^{r} X' \cdot \pi_{ih}(T)\text{vec}(v)$$

$$= \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} X_{tj} \pi_{ih}(T)(t, s) v_{sh}.$$ 

Assume throughout this section that $\xi_t = (u_t, v_t)$ is a $1 \times p$ second-order stationary process with mean zero and absolutely summable covariance function. Then

$$E[(I_r \otimes X')\pi(T)\text{vec}(v)] = (I_r \otimes u')\sum \pi(T)E[\text{vec}(v)] + E(I_r \otimes u')\pi(T)\text{vec}(v)$$

$$= E[(I_r \otimes u')\pi(T)\text{vec}(v)]$$

$$= T \text{vec}[\Gamma_{u\pi v}(T)(0)],$$
say, where $\Gamma_{uvv}(T)(0)$ is a $k \times r$ matrix with $(j,i)$-th element equal to

$$
T \Gamma_{uvvji}(T)(0) = \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} E[u, \tau_{ih}(T)(t, s)v_{sh}]
$$

$$
= \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{uvjh}(s - t) \tau_{ih}(T)(t, s)
$$

$$
= \sum_{h=1}^{r} \text{tr}(\Gamma_{uvjh} \tau_{ih}(T)) .
$$

Consider separately the addends in

$$
\text{Var}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)]
$$

$$
= \text{Var}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)]
$$

$$
+ \text{Cov}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v), (I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)]
$$

$$
+ \text{Cov}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v), (I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)]
$$

$$
+ \text{Var}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)]
$$

First,

$$
\text{Var}[(I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)] = (I_r \otimes T) \tau_{T}(\cdot) \Gamma_{uvv} \tau_{T}(\cdot) (I_r \otimes T) \tau_{T}(\cdot) \text{vec}(v)
$$
has [(i,j), (l,m)]-th element equal to

\[
\hat{\mu}_{j}^{\pi_{l}} \in_{\pi_{l}} \nu_{(T)} \pi_{l}(T) \hat{\mu}_{l,m}
\]

\[
= \sum_{r} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \hat{\mu}_{r}^{T} \pi_{l}(T)(t, s) \Gamma_{v_{gh}(q - s)} \pi_{h \xi(T)}(q, n) \hat{\mu}_{l,m}.
\]

(3.79)

Second, \(\text{Var}[(I_{r} \pm u')_{N_{l}(T)} \text{vec}(v)]\) has [(i,j), (l,m)]-th element equal to

\[
\text{Cov}[u'_{j} \pi_{l}(T) \text{vec}(v), u'_{m} \pi_{l}(T) \text{vec}(v)]
\]

\[
= \sum_{r} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \text{Cov}(u'_{j} \pi_{l}(T) \nu_{s}, u'_{m} \pi_{l}(T) \nu_{h})
\]

\[
= \sum_{r} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \pi_{l}(T)(t, s) \pi_{h \xi(T)}(q, n) \text{Cov}(u_{r} \nu_{s}, u_{m} \nu_{h})
\]

\[
= \sum_{r} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \sum_{T} \pi_{l}(T)(t, s) \pi_{h \xi(T)}(q, n)
\]

\[
x \{ \Gamma_{\nu_{gh}(n-t)} \nu_{gh}(q-s) + \Gamma_{\nu_{jh}(q-t)} \nu_{gm}(n-s)
\]

\[
+ k_{uvuvjgsh}(0, s-t, n-t, q-t) \}.
\]

(3.80)

Finally,

\[
\text{Cov}[(I_{r} \pm u')_{N_{l}(T)} \text{vec}(v), (I_{r} \pm u')_{N_{l}(T)} \text{vec}(v)]
\]
has \([(i,j), (r,m)]\)-th element equal to

$$\text{Cov}[u_{ij}^T \pi_{T}(T) \text{vec}(v), u_{jm}^T \pi_{T}(T) \text{vec}(v)]$$

$$= \sum_{g=1}^{r} \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi_{ig}(T)(t, s) \pi_{hj}(T)(q, n)$$

$$\times M_{vuvgh}(0, t-s, q-s) u_{nm}$$

(3.81)

where the third-cumulant matrix \(M_{uv}\) is as defined in Section 3.1.

Thus, \(\text{Var}[(I_{r} \otimes X') \pi_{T}(T) \text{vec}(v)]\) has \([(i,j), (r,m)]\)-th element equal to

$$= \sum_{g=1}^{r} \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi_{ig}(T)(t, s) \pi_{hj}(T)(q, n)$$

$$\times \{ \Gamma_{vuvgh}(q-s)[u_{nm} u_{tj} + \Gamma_{uu} m(n-t)] + \Gamma_{uv} (q-t) \Gamma_{uv} (s-n)$$

$$+ \Gamma_{vuvgh}(0, t-s, q-s) u_{qm} + \mu_{tj} \Gamma_{vuvgh}(0, n-s, q-s)$$

$$+ k_{uvuvgh}(0, s-t, n-t, q-t) \}$$

(3.82)

Now consider the asymptotic behavior of \((I_{r} \otimes X') \pi_{T}(T) \text{vec}(v)\).

Under condition (3.76), let \(\pi\) be a \(Tr \times Tr\) matrix with double-subscripted \([(i, s), (j, t)]\)-th element equal to

$$\pi_{ij}(s, t) = \lim_{T \to \infty} \pi_{ij}(T)(s, t)$$
Define $\Gamma_{\mu \nu}(0)$ to be a $k \times r$ matrix with $(j,i)$-th element defined by expression (3.78) with the entries $\pi_{ih}(t, s)$ replaced by the entries $\pi_{ih}(t, s)$. Note that for fixed $s$ and $t$, $\pi_{ij}(s, t)$ is constant with respect to $T$, so asymptotic arguments involving $\pi$ are generally simpler than corresponding arguments with $\pi(T)$. Now

$$T^{-1}\{(I_r \otimes X')\pi(T)\text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(T)(0)]\}$$

$$= T^{-1}\{(I_r \otimes X')\pi \text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(0)]\}$$

$$+ T^{-1}\{(I_r \otimes X')[-\pi \text{vec}(w) - T \text{vec}[\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)]\}.$$

Thus,

$$T^{-1}\{(I_r \otimes X')\pi(T)\text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(T)(0)]\}$$

converges to $0_{rk \times 1}$ with probability one as $T \to \infty$ if

$$T^{-1}\{(I_r \otimes X')\pi \text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(0)]\}$$

and

$$T^{-1}\{(I_r \otimes X')[-\pi \text{vec}(w) - T \text{vec}[\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)]\} \quad (3.83)$$

each converge to $0_{rk \times 1}$ with probability one as $T \to \infty$. Similarly,
\[ T^{-1/2} \{(I_r \otimes X') \pi(T) \text{vec}(v) - T \text{vec}[\Gamma_{\Pi T V}(T)(0)]\} \]

and

\[ T^{-1/2} \{(I_r \otimes X') \pi \text{vec}(v) - T \text{vec}[\Gamma_{\Pi T V}(0)]\} \]

have the same limiting distribution provided

\[ T^{-1/2} \{(I_r \otimes X') [\pi(T) - \pi] \text{vec}(v) - T \text{vec}[\Gamma_{\Pi T V}(T)(0) - \Gamma_{\Pi T V}(0)]\} \] \hspace{1cm} (3.84)

converges in probability to \( O_{r k \times 1} \) as \( T \to \infty \).

The following lemma gives conditions for convergence of expressions (3.83) and (3.84).

**Lemma 3.13.** Let \( \{y_t\} \) be a sequence of fixed \( 1 \times k \) vectors that satisfy condition (3.39.a). Let \( \{\xi_t\} \) be a \( 1 \times p \) fourth-order stationary process with mean zero and covariance function and fourth-cumulant function that are absolutely summable. Let \( \{\pi(T)\} \) and \( \{\pi\} \) be sequences of \( Tr \times Tr \) matrices that satisfy conditions (3.74)-(3.76).

a. Assume that there exists some \( \beta > 0 \) and \( M > 0 \) such that for all \( 1 < i, j < r \) and all \( T \in \mathbb{Z}^+ \),

\[ T \sum_{s=1}^{T} |\pi_{ij}(T)(s, t) - \pi_{ij}(s, t)| < MT^{-\beta} \text{ for all } 1 < t < T; \] \hspace{1cm} (3.85.a)

and
Then as \( T \to \infty \),

\[
T^{-1} \{(I_r \otimes X')[\pi(T) - \pi]\text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)]\}
\]

converges to \( O_{rk \times 1} \) with probability one, and

\[
\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)
\]

converges to \( O_{rk \times 1} \).

b. Assume that as \( T \to \infty \), for all \( 1 \leq i, j \leq r \),

\[
\sum_{s=1}^{T} |\pi_{ij}(T)(s, t) - \pi_{ij}(s, t)|
\]

(3.86.a)

converges to zero uniformly in \( t \), and

\[
\sum_{t=1}^{T} |\pi_{ij}(T)(s, t) - \pi_{ij}(s, t)|
\]

(3.86.b)

converges to zero uniformly in \( s \). Then as \( T \to \infty \),

\[
T^{-1/2} \{(I_r \otimes X')[\pi(T) - \pi]\text{vec}(v) - T \text{vec}[\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)]\}
\]

converges to \( O_{rk \times 1} \) in probability, and

\[
\Gamma_{\mu \nu}(T)(0) - \Gamma_{\mu \nu}(0)
\]

converges to \( O_{rk} \).
Proof. As for the proof of Lemma 3.10, it suffices to prove convergence with probability one, or in probability, for the case \( r = k = 1 \).

a. First, recall that \( E(\mu'^T \pi(T) \nu) = 0_{rk \times l} = E(\mu' \nu) \). By expression (3.78),

\[
E(T^{-1} \mu'[\pi(T) - \pi] \nu) = T^{-1} \sum_{t=1}^{T} \mathbb{E} (\sum_{s=1}^{T} \gamma_{uv}(s-t)[\pi(T)(t, s) - \pi(t, s)]) .
\]

Conditions (3.85) imply that there exists some \( M_1 > 0 \) such that

\[
\left| \sum_{s=1}^{T} \gamma_{uv}(s-t)[\pi(T)(t, s) - \pi(t, s)] \right| < M_1 T^{-\beta}
\]

for all \( T \in \mathbb{Z}^+ \) and all \( 1 < t < T \), so expression (3.87) is bounded in modulus by \( M_1 T^{-\beta} \). Thus, \( \gamma_{uv}(T)(0) - \gamma_{uv}(0) \) converges to zero as \( T \to \infty \).

Next, expression (3.79) implies that

\[
\operatorname{Var}(T^{-1} \mu'[\pi(T) - \pi] \nu) = T^{-1} \sum_{t=1}^{T} \mathbb{E} (\sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \sum_{l=1}^{T} \gamma_{uv}(q-s)[\pi(T)(q, n) - \pi(q, n)][T^{-1} \mu_{n'l}(T)] .}
\]

(3.88)
By conditions (3.85) and Result 10.3.b, there exists some $M_2 > 0$ such that for all $T \in \mathbb{Z}^+$,

\[ \sum_{T} \sum_{T} \sum_{T} \sum_{n=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \left| \pi(T)(t, s) - \pi(t, s) \right| \Gamma_{vv}(q-s) \left| \pi(T)(q, n) - \pi(q, n) \right| < M_2 T^{-2\beta} \text{ for all } 1 < n < T; \text{ and} \]

\[ \sum_{T} \sum_{T} \sum_{T} \sum_{n=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \left| \pi(T)(t, s) - \pi(t, s) \right| \Gamma_{vv}(q-s) \left| \pi(T)(q, n) - \pi(q, n) \right| < M_2 T^{-2\beta} \text{ for all } 1 < t < T. \]

Also, $\left| u_t u_n \right| < 2^{-1}(u_t^2 + u_n^2)$, and by the arguments following (3.39.c), $T^{-1} \sum_{t=1}^{T} u_t^2$ is uniformly bounded, so a second application of Result 10.3.b implies that there exists some $M_3 > 0$ such that expression (3.88) is bounded above by $M_3 T^{-2\beta - 1}$. Then by the Chebyshev inequality,

\[ \Pr \left[ T^{-1} \left| u'_T \right| \left| \pi(T) - \pi \right| > \varepsilon \right] < \varepsilon^{-2} \text{Var} \left[ T^{-1} u'_T \left| \pi(T) - \pi \right| \right] \]

\[ < \varepsilon^{-2} M_3 T^{-2\beta - 1} \]

for all $T \in \mathbb{Z}^+$, so

\[ \sum_{T=1}^{\infty} \Pr \left[ T^{-1} \left| u'_T \right| \left| \pi(T) - \pi \right| > \varepsilon \right] < \sum_{T=1}^{\infty} \varepsilon^{-2} M_3 T^{-2\beta - 1} < \infty. \]
because $\beta > 0$. Then by the first Borel-Cantelli lemma,

$$\Pr\{T^{-1}|y'[\pi(T) - y]v| > \epsilon \text{ for infinitely many positive integers } T\} = 0.$$  

Since $\epsilon > 0$ was arbitrary, it follows that $T^{-1}y'[\pi(T) - y]v$ converges almost surely to zero as $T \to \infty$.

It remains to prove that

$$T^{-1} \{y'[\pi(T) - y]v - T[\Gamma_{uvv}(T)(0) - \Gamma_{uvv}(0)]\}$$

converges almost surely to zero. Expression (3.80) implies that

$$\text{Var}\{T^{-1}y'[\pi(T) - y]v\}$$

$$= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} [\pi(T)(t, s) - \pi(t, s)][\pi(T)(q, n) - \pi(q, n)]$$

$$\times \Gamma_{uu}(n-t)\Gamma_{vv}(q-s), \quad (3.89.a)$$

$$= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} [\pi(T)(t, s) - \pi(t, s)][\pi(T)(q, n) - \pi(q, n)]$$

$$\times \Gamma_{uv}(q-t)\Gamma_{vu}(n-s) \quad (3.89.b)$$

$$+ T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} [\pi(T)(t, s) - \pi(t, s)]$$

$$\times [\pi(T)(q, n) - \pi(q, n)] k_{uvuv}(0, s-t, n-t, q-t). \quad (3.89.c)$$
By two applications of Result 10.3, there exist \( M_4 > 0 \), \( M_5 > 0 \) and \( M_6 > 0 \) such that for all \( T \in \mathbb{Z}^+ \),

\[
\sum_{q=1}^{T} \sum_{n=1}^{T} \left| \sum_{s=1}^{T} \left[ \pi_{(T)}(q, n) - \pi(q, n) \right] \Gamma_{uu}(n-t) \right| < M_4 T^{-B} \quad \text{for all } 1 < t < T;
\]

\[
\sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \sum_{q=1}^{T} \left[ \pi_{(T)}(q, n) - \pi(q, n) \right] \pi_{(T)}(t, s) - \pi(t, s) \right| \Gamma_{uu}(n-t) \left| \Gamma_{vv}(q-s) \right| < M_5 T^{-2B}.
\]

Thus, expression (3.89.a) is bounded in modulus by \( M_5 T^{-2B-1} \).

A similar argument establishes that there exists some \( M_7 > 0 \) such that expression (3.89.b) is bounded in modulus by \( M_7 T^{-2B-1} \).

Conditions (3.85) imply that

\[
\left| \pi_{(T)}(t, s) - \pi(t, s) \right| < M T^{-B}
\]

for all \( T \in \mathbb{Z}^+ \) and all \( 1 < s, t < T \). Since

\[
\sum_{n=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \left| k_{uvuv}(0, s, n, q) \right| < \infty,
\]
there exists some $M_9 > 0$ such that for all $T \in \mathcal{Z}$,

$$
\sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} |k_{uvuv}(0, s-t, n-t, q-t)| < M_9
$$

for all $1 < t < T$. Thus, for all $T \in \mathcal{Z}$,

$$
\sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \pi(T)(t, s) - \pi(t, s) \right| \left| \pi(T)(q, n) - \pi(q, n) \right| \times |k_{uvuv}(0, s-t, n-t, q-t)| < M_T^2 - 2M_9
$$

so expression (3.89.c) is bounded in modulus by $M_9^2 T^{-2\beta-1}$. Hence, there exists some $M_9 > 0$ such that

$$
\text{Var}\{T^{-1}u'[\pi(T) - \bar{\pi}]v\} < M_9 T^{-2\beta-1}.
$$

By the Chebyshev inequality,

$$
\Pr\{T^{-1}\left| u'[\pi(T) - \bar{\pi}]v - T[\Gamma_{uvv}(T)(0) - \Gamma_{uvv}(0)] \right| > \epsilon\} < \epsilon^{-2} \text{Var}\{T^{-1}u'[\pi(T) - \bar{\pi}]v\}
$$

$$
< \epsilon^{-2} M_9 T^{-2\beta-1}
$$
for all \( T \in \mathbb{Z}^+ \), so

\[
\sum_{T=1}^{\infty} \Pr \left( \left| T^{-1} \left( \pi(T) - \pi(T) - T \left( \Gamma_{uv}(T) - \Gamma_{uv}(0) \right) \right) \right| > \epsilon \right) < \sum_{T=1}^{\infty} \epsilon^{-2} 9^{-2\beta-1} < \infty
\]

because \( \beta > 0 \). Then by the first Borel-Cantelli lemma,

\[
\Pr \left( \left| T^{-1} \left( \pi(T) - \pi(T) - T \left( \Gamma_{uv}(T) - \Gamma_{uv}(0) \right) \right) \right| > \epsilon \text{ for infinitely many positive integers } T \} = 0 .
\]

Since \( \epsilon > 0 \) was arbitrary, it follows that

\[
T^{-1} \left[ \pi(T) - \pi(T) - T \left( \Gamma_{uv}(T) - \Gamma_{uv}(0) \right) \right]
\]

converges almost surely to zero as \( T \to \infty \).

b. The proof of part (b) is similar to the proof of part (a) and thus requires only a brief outline. Conditions (3.86) and Result 10.3.a imply that as \( T \to \infty \), \( \sum_{s=1}^{T} T_{uv}(s-t)[\pi(T)(t, s) - \pi(t, s)] \) converges to zero uniformly in \( t \), so that expression (3.87) converges to zero as \( T \to \infty \). Similarly, conditions (3.86) and Result 10.3.a imply that as \( T \to \infty \),
The uniform boundedness of \( T^{-1} \sum_{t=1}^{T} u_t^2 \) then implies that expression (3.88) converges to zero. Then by Chebyshev inequality, \( T^{-1} \mu'[\pi(T) - \pi]v \) converges to zero in probability.

Also, conditions (3.86) and Result 10.3.b imply that as \( T \to \infty \),

\[
\sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \pi(T)(t,s) - \pi(t,s) \right| \Gamma_{\pi(T)}(q,n) - \pi(q,n) \Gamma_{\pi(T)}(q,s)\]

converges to zero uniformly in \( 1 < t < T \), so that expression (3.89.a) converges to zero. A similar argument indicates that expression (3.89.b) converges to zero.

Finally, the inequality

\[
\sum_{s=0}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left| k_{uvv}(0,s,n,q) \right| < \infty
\]

and conditions (3.86) imply that expression (3.89.c) converges to zero as \( T \to \infty \). Thus, \( \text{Var}[T^{-1}u'[\pi(T) - \pi]v] \) converges to zero as \( T \to \infty \), so by the Chebyshev inequality,
\[ T^{-1}w'[\pi(T) - \pi]v - [\Gamma_{uvw}(T)(0) - \Gamma_{uvw}(0)] \]

converges to zero in probability as \( T \to \infty \).

Inspection of the proof of Lemma 3.13.a indicates that the conclusion of part (a) holds if one replaces conditions (3.85) with the condition that there exist some \( M > 0 \) such that for all \( T \in \mathbb{Z}^+ \) and all \( 1 < s, t < T \),

\[
|\pi_{ij}(T)(t, s) - \pi_{ij}(t, s)| < MT^{-\beta}
\]

for some \( \beta > 2^{-1} \).

Moreover, note that conditions (3.85) imply conditions (3.86). Hence, the conclusions of Lemma 3.13.b hold under the hypotheses of Lemma 3.13.a.

Given the results of Lemma 3.13, one may now study the asymptotic behavior of \( T^{-1}[(I_r \otimes X')\tilde{\pi}(v)] - \text{vec}[\Gamma_{uvw}(0)] \). As a first step in this work, Lemmas 3.14 and 3.15 address almost sure convergence of this function to zero.

**Lemma 3.14.** Let \( \{\mu_\varepsilon\} \) be a sequence of fixed \( 1 \times k \) vectors with uniformly bounded elements that satisfy condition (3.39.a). Let \( \xi = (u, v) \) be a \( 1 \times p \) fourth-order stationary process with mean zero, where \( u \) is \( 1 \times k \), \( v \) is \( 1 \times r \), and \( p = r + k \). Assume also that for some \( \alpha < 1 \), \( K_1 > 0 \) and for all \( 1 < i, j, l, m < p \), and all positive integers \( T \), the covariance function \( \Gamma_{\xi}(t) \) and the
fourth-order cumulant function $k_{\epsilon \epsilon \epsilon \epsilon}(0, q, s, t)$ satisfy the conditions,

$$
\sum_{t=-T+1}^{T-1} |\Gamma_{\epsilon \epsilon \epsilon \epsilon}(t)| < K_1 T^\alpha
$$

and

$$
\sum_{q=-T+1}^{T-1} \sum_{s=-T+1}^{T-1} \sum_{t=-T+1}^{T-1} |k_{\epsilon \epsilon \epsilon \epsilon}(0, q, s, t)| < K_1 T^\alpha .
$$

Moreover, assume that for some $\eta < 2^{-1}\min\{1, 1-\alpha\}$, $K_2 > 0$, the real numbers $\pi_{ij}(s, t)$ satisfy the conditions,

$$
\sum_{s=1}^{T} |\pi_{ij}(s, t)| < K_2 T^{\eta} \text{ for all } T \in \mathbb{Z}^+ \text{ and all } 1 < t < T ; \quad (3.90.a)
$$

and

$$
\sum_{t=1}^{T} |\pi_{ij}(s, t)| < K_2 T^{\eta} \text{ for all } T \in \mathbb{Z}^+ \text{ and all } 1 < s < T . \quad (3.90.b)
$$

Let $X = \mu + u$. Then as $T \to \infty$,

(i) $T^{-1}[((I_r \circ \mu')_n \text{ vec}(v)) - \circ_{r \times 1}$ with probability one;

(ii) $T^{-1}[(I_r \circ u')_n \text{ vec}(v)] - \text{ vec}[(I_n \circ \mu_n(0))] + \circ_{r \times 1}$ with probability one;

and thus
(iii) \( T^{-1}[(I_n \otimes X')_I \vec{w} - \vec{X}_{\Delta \mathcal{W}}(0)] + O_{r \times 1} \) with probability one.

Proof. As for Lemma 3.10, it suffices to establish results (i), (ii) and (iii) for the case \( r = k = 1 \).

To develop some necessary notation, choose some \( \beta > 1 \) such that \( \beta(1 - \alpha - 2\gamma) > 1 \) and \( 2\beta(\eta - 1) < -1 \). For each nonnegative integer \( M \), let \( T(M) \) be the smallest integer greater than or equal to \( M^\beta \). As in the proof of Lemma 3.10, one may obtain the inequalities

\[
M^\beta < T(M) < M^\beta + 1
\]

and

\[
T(M)^{-1}[T(M + 1) - T(M)] < M^{-\beta}K_3
\]

for some \( K_3 > 0 \) and for all \( M \in \mathbb{Z}^+ \). Moreover, the inequality \( \beta > 1 \) implies that for \( M \in \mathbb{Z}^+ \),

\[
T(M + 1) < (M + 1)^\beta + 1 < (M + 2)^\beta,
\]

so for any \( \delta > 0 \), there exists some \( K_4 > 0 \) such that for all \( M \in \mathbb{Z} \),

\[
T(M)^{-1}[T(M + 1) - T(M)][T(M + 1)]^\delta \leq K_3M^{-\beta}[M + 2]^{\delta \beta}
\]
< K_4 M^{-\beta} M \beta \\
= K_4 M^{\beta (\delta - 1)} . \quad (3.91)

To establish result (i), define

\[ W_T = T^{-1} \sum_{\nu} \pi_{\nu} T \]

\[ = T^{-1} \sum_{t=1}^{T(M)} \sum_{s=1}^{T(M)} \mu_{\nu}(t, s)v_s \]

and note that for \( T(M) < T < T(M + 1) \), \( T^{-1} T(M) < 1 \) and

\[ |W_T| < |W_T - T^{-1} T(M) W_{T(M)}| + |W_{T(M)}| . \]

Thus, to prove that \( W_T \to 0 \) almost surely as \( T \to \infty \), it suffices to show that

\[ \lim_{M \to \infty} |W_{T(M)}| = 0 \text{ almost surely} \]

and that

\[ \lim_{M \to \infty} \max_{T(M) < T < T(M + 1)} \{|W_T - T^{-1} T(M) W_{T(M)}|\} = 0 \text{ almost surely.} \]

Note that \( E(W_T) = 0 \) and that for \( B_2 \) equal to a uniform bound on \( \mu_{\nu} \), expression (3.79) implies that
\text{Var}(W_T) = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mu_t \pi(t, s) \Gamma_{vv}(q-s) \pi(q, n) \mu_n

< \frac{T^{-2} B^2_2}{Z^2} \sum_{q=1}^{T} \sum_{s=1}^{T} |\Gamma_{vv}(q-s)| \pi(t, s) \sum_{n=1}^{T} |\pi(q, n)|

< \frac{T^{-2} B^2_2 k^2 T^2 n}{2} \sum_{q=1}^{T} \sum_{s=1}^{T} |\Gamma_{vv}(q-s)|

< \frac{T^{-2} B^2_2 k^2 T^{-1}}{2} \sum_{q=1}^{T} |\Gamma_{vv}(q-s)|

< \frac{B_2^2 k^2 T^{(\alpha + 2) n - 1}}{2}.

(3.92)

Pick \( \epsilon > 0 \) and note that by the Chebyshev inequality,

\[ \mathbb{P}\left[ |W_{T(M)}| > \epsilon \right] < \frac{\epsilon^{-2} \text{Var}(W_{T(M)})}{2} \]

\[ < \frac{\epsilon^{-2} B^2_2 k^2 k_1 (T(M))^{(\alpha + 2) n - 1}}{2} \]

\[ < \frac{\epsilon^{-2} B^2_2 k^2 k_1 M^{-\beta(1-\alpha-2n)}}{2}. \]

Thus,

\[ \sum_{M=1}^{\infty} \mathbb{P}\left[ |W_{T(M)}| > \epsilon \right] < \frac{\epsilon^{-2} B^2_2 k^2 k_1 \sum_{M=1}^{\infty} M^{-\beta(1-\alpha-2n)}}{2} < \infty \]

because \( \beta(1-\alpha-2n) > 1 \). Then by the first Borel-Cantelli lemma,

\[ \mathbb{P}\left[ |W_{T(M)}| > \epsilon \text{ for infinitely many positive integers } M \right] = 0. \]
Since $\epsilon > 0$ was arbitrary, it follows that $|W_{T(M)}|$ converges to zero almost surely as $M \to \infty$.

Next, note that for all $T$ such that $T(M) < T < T(M + 1)$,

$$W_T - T^{-1}T(M)W_{T(M)} = T^{-1}[\sum_{t=T(M)+1}^{T} \sum_{s=1}^{T(M)} \mu_t \pi(t, s)v_s]$$

$$+ T^{-1}[\sum_{t=1}^{T(M)} \sum_{s=T(M)+1}^{T} \mu_t \pi(t, s)v_s]$$

$$+ T^{-1}[\sum_{s=T(M)+1}^{T(M)+1} \sum_{t=T(M)+1}^{T} \mu_t \pi(t, s)v_s].$$

Now

$$\max_{T(M) < T < T(M+1)} \{ |T^{-1}[\sum_{t=T(M)+1}^{T} \sum_{s=1}^{T(M)} \mu_t \pi(t, s)v_s]|^2 \}$$

$$< T(M)^{-2} \sum_{t=T(M)+1}^{T(M+1)} \sum_{s=1}^{T(M)} |\mu_t \pi(t, s)v_s|^2$$

$$< T(M)^{-2} B^2 \sum_{t=T(M)+1}^{T(M+1)} \sum_{s=1}^{T(M)} |\pi(t, s)v_s|^2$$

$$< T(M)^{-2} B^2 \sum_{s=1}^{T(M+1)} \sum_{t=T(M)+1}^{T(M+1)} \sum_{q=1}^{T(M+1)} \sum_{n=T(M)+1}^{T(M+1)} |\pi(t, s)\pi(q, n)v_s v_q|^2.$$

This final expression has expectation bounded above by

$$T(M)^{-2} B^2 r_{vv}(0)[\sum_{s=1}^{T(M)} \sum_{t=T(M)+1}^{T(M+1)} |\pi(t, s)|^2]$$

$$< T(M)^{-2} B^2 r_{vv}(0)[T(M+1) - T(M)]^2 \kappa^2_2[T(M+1)]^2 n. \quad (3.93)$$
By a similar argument,

\[
\max_{T(M) < T < T(M+1)} \left\{ \left| T^{-1} \sum_{T(M)+1}^{T(M+1)} \sum_{t=T(M)+1}^{T(M+1)} \mu_t \pi(t, s)v_s \right|^2 \right\}
\]

also has expectation bounded above by expression (3.93). In addition,

\[
\max_{T(M) < T < T(M+1)} \left\{ \left| T^{-1} \sum_{T(M)+1}^{T(M+1)} \sum_{t=T(M)+1}^{T(M+1)} \mu_t \pi(t, s)v_s \right|^2 \right\}
\]

\[< T(M)^{-2}B_2^2 \sum_{s=T(M)+1}^{T(M+1)} \sum_{t=1}^{T(M)} \pi(t, s) \left| v_s \right|^2 \]

\[< T(M)^{-2}B_2^2K^2T^2n \sum_{s=T(M)+1}^{T(M+1)} \left| v_s \right|^2 .
\]

This final expression also has expectation bounded above by expression (3.93). Thus,

\[
\max_{T(M) < T < T(M+1)} \left\{ \left| W_T - T^{-1}T(M)W_{T(M)} \right|^2 \right\}
\]

has expectation bounded above by

\[
K_5 T(M)^{-2}[T(M+1) - T(M)]^2[T(M+1)]^2n , \quad (3.94)
\]

where \( K_5 \) is some positive real number. By inequality (3.91), expression (3.73) is bounded above by

\[
K_6^2B(\eta-1)
\]
for all \( T \), where \( K_6 \) is another positive real number. Then by the Chebyshev inequality,

\[
P\left( \max_{T(M) < T < T(M+1)} \left| W_T - T^{-1}T(M)W_{T(M)} \right| > \varepsilon \right) < \varepsilon^{-2}K_6 \, M^{2\beta(\eta-1)}.
\]

Thus,

\[
\sum_{M=1}^{\infty} P\left( \max_{T(M) < T < T(M+1)} \left| W_T - T^{-1}T(M)W_{T(M)} \right| > \varepsilon \right) < \varepsilon^{-2}K_6 \sum_{M=1}^{\infty} M^{2\beta(\eta-1)} < \infty
\]

because \( 2\beta(\eta-1) < -1 \). Then by the first Borel-Cantelli lemma,

\[
P\left( \max_{T(M) < T < T(M+1)} \left| W_T - T^{-1}T(M)W_{T(M)} \right| > \varepsilon \text{ for infinitely many positive integers } M \right) = 0.
\]

Since \( \varepsilon > 0 \) was arbitrary, it follows that

\[
\max_{T(M) < T < T(M+1)} \left\{ \left| W_T - T^{-1}T(M)W_{T(M)} \right| \right\}
\]

converges to zero almost surely as \( M \to \infty \). Thus, \( |W_T| \) converges to zero almost surely as \( T \to \infty \), so result (i) is established.

Now define
and note that for $T(M) < T < T(M+1)$, $T^{-1}T(M) < 1$ and

$$|Y_T| < |Y_T - T^{-1}T(M)Y_{T(M)}| + |Y_{T(M)}|.$$ 

Thus, to prove that $|Y_T| \to 0$ almost surely as $T \to \infty$, it suffices to show that

$$\lim_{M \to \infty} |Y_{T(M)}| = 0 \text{ almost surely}$$

and that

$$\lim_{M \to \infty} \left\{ \max_{T(M) < T < T(M+1)} [ |Y_T - T^{-1}T(M)Y_{T(M)}|] \right\} = 0 \text{ almost surely.}$$

Note that $E(Y_T) = 0$ and that by expression (3.80),

$$\text{Var}(Y_T) = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) \{ \Gamma_{vv}(q-s) \Gamma_{uu}(n-t) \}$$

$$+ \Gamma_{vu}(q-t) \Gamma_{uv}(n-s) + k_{uvuv}(0, s-t, n-t, q-t).$$

Now
$T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \pi(t, s) \pi(q, n) \Gamma_{uv}(q-s) \Gamma_{vu}(n-t) \right|$

$< T^{-1} \Gamma_{uu}(0) \sum_{q=1}^{T} \sum_{s=1}^{T} \sum_{t=1}^{T-1} \left| \Gamma_{vv}(q-s) \right| \left| \pi(q, n) \right|$

$< T^{-1} \Gamma_{uu}(0) K_{2} T^{n} \left\{ \sum_{q=1}^{T} \sum_{s=-T+1}^{T-1} \left| \Gamma_{vv}(q-s) \right| \left| \pi(q, n) \right| \right\}$

$< T^{-1} \Gamma_{uu}(0) K_{1} K_{2}^{2} n^{n+1}$.

(3.95)

A similar argument indicates that for some $K_{7} > 0$, 

$T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \pi(t, s) \pi(q, n) \Gamma_{uv}(q-t) \Gamma_{vu}(n-s) \right| < K_{7}^{2} n^{n+1}$

for all $T \in \mathbb{Z}^{+}$. In addition, conditions (3.90) imply that

$\left| \pi(s, t) \right| < K_{2} T^{n}$ for all $1 < s, t < T$ and all $T \in \mathbb{Z}^{+}$, so

$T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \left| \pi(t, s) \pi(q, n) k_{uvuv}(0, s-t, n-t, q-t) \right|$

$< K_{2}^{2} n^{2} + 2 n \sum_{s=-T+1}^{T-1} \sum_{q=-T+1}^{T-1} \sum_{n=-T+1}^{T-1} \left| k_{uvuv}(0, s-t, n-t, q-t) \right|$

$< K_{1} K_{2}^{2} n^{n+1}$. 

Thus, there exists some $K_{8} > 0$ such that for all $T \in \mathbb{Z}^{+}$,

$\text{Var}(Y_{T}) < K_{8} T^{\alpha+2 n^{n-1}}$.
Pick \( \epsilon > 0 \) and note that by the Chebyshev inequality,

\[
P[|Y_{T(M)}| > \epsilon] < \epsilon^{-2} \text{Var}[Y_{T(M)}] < K_\delta \delta^{\alpha+2\eta-1},
\]

so

\[
\sum_{M=1}^{\infty} P[|Y_{T(M)}| > \epsilon] < \epsilon^{-2} K_\delta \sum_{M=1}^{\infty} \delta^{\alpha+2\eta-1} < \infty
\]

because \( \delta^{\alpha+2\eta-1} < -1 \). Then by the first Borel-Cantelli lemma,

\[
P[|Y_{T(M)}| > \epsilon \text{ for infinitely many positive integers } M] = 0.
\]

Since \( \epsilon > 0 \) was arbitrary, it follows that \( |Y_{T(M)}| \) converges to zero almost surely as \( M \to \infty \).

Next, note that for all \( T \) such that \( T(M) < T < T(M+1) \),

\[
Y_T - T^{-1}T(M)Y_{T(M)} = [T^{-1} \sum_{t=T(M)+1}^{T} \sum_{s=1}^{T(M)} u_t \pi(t, s) v_s] + T^{-1} \sum_{t=1}^{T(M)} \sum_{s=T(M)+1}^{T} u_t \pi(t, s) v_s + T^{-1} \sum_{t=T(M)+1}^{T} \sum_{s=T(M)+1}^{T} u_t \pi(t, s) v_s.
\]

Now
The fourth-order stationarity of $\xi_t$ implies that $E[|u_{t,n,s,q}|]$ is bounded above by some $K_9 > 0$ for all $t, n, s$ and $q$, so expression (3.96) has expectation bounded above by

$$K_9 T(M)^{-2} \sum_{s=1}^{T(M)} \sum_{q=1}^{T(M+1)} \sum_{t=T(M)+1}^{T(M+1)} \sum_{n=T(M)+1}^{T(M+1)} |\pi(t, s) \pi(q, n)|.$$ (3.97)

By a similar argument,

$$\max_{T(M) < T < T(M+1)} \{ |T^{-1} \sum_{t=T(M)+1}^{T(M+1)} \sum_{s=T(M)+1}^{T(M+1)} u_{t,n,s,q} |^2 \}$$

also has expectation bounded above by expression (3.97). In addition,

$$\max_{T(M) < T < T(M+1)} \{ |T^{-1} \sum_{t=T(M)+1}^{T(M+1)} \sum_{s=T(M)+1}^{T(M+1)} u_{t,n,s,q} |^2 \}$$
This final expression also has expectation bounded above by expression (3.97). Thus, there is some $K_{10} > 0$ such that

$$\max_{T(M) < T < T(M+1)} \left\{ \left| Y_T - T^{-1} T(M) Y_{T(M)} \right|^2 \right\}$$

has expectation bounded above by

$$K_{10} T(M)^{-2} [T(M + 1) - T(M)]^2 \leq [T(M + 1)]^2 < K_{11} M^2 2^B(\eta-1)$$

for all $T$, where $K_{11}$ is some other positive real number. Then by the Chebyshev inequality,

$$\sum_{M=1}^{\infty} P\left[ \max_{T(M) < T < T(M+1)} \left| Y_T - T^{-1} T(M) Y_{T(M)} \right| > \epsilon \right] < \epsilon^{-2} \sum_{M=1}^{\infty} K_{11} M^2 2^B(\eta-1)$$

because $2^B(\eta-1) < 1$. Then by the first Borel-Cantelli lemma,

$$P\left[ \max_{T(M) < T < T(M+1)} \left| Y_T - T^{-1} T(M) Y_{T(M)} \right| > \epsilon \text{ for infinitely many positive integers } M \right] = 0.$$  

Since $\epsilon > 0$ was arbitrary, it follows that

$$\max_{T(M) < T < T(M+1)} \left\{ \left| Y_T - T^{-1} T(M) Y_{T(M)} \right| \right\}$$
converges to zero almost surely as $M \to \infty$. Thus, $|Y_T|$ converges to zero almost surely as $T \to \infty$, so result (ii) is established.

Result (iii) follows immediately from results (i) and (ii). \[\square\]

For fourth-order stationary process with absolutely summable covariance function and fourth-order cumulant function, one may state a special case of Lemma 3.14 as follows.

**Lemma 3.15.** Let $\{u_t\}$ be a sequence of fixed $1 \times k$ vectors with uniformly bounded elements that satisfy condition (3.39.a), and let $\varepsilon_t = (u_t, v_t)$ be a $1 \times p$ fourth-order stationary process with mean zero, where $u_t$ is $1 \times k$, $v_t$ is $1 \times r$, and $p = r + k$. Assume also that for some finite $K_1$ and for all $1 < i, j, k, m < p$, the covariance function $\Gamma_{ij}(t)$ and the fourth-cumulant function $k_{(0,q,s,t)}(t)$ satisfy the conditions

$$
\sum_{t=-\infty}^{\infty} \left| \Gamma_{ij}(t) \right| < \infty
$$

and

$$
\sum_{q=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \left| k_{(0,q,s,t)}(0, q, s, t) \right| < \infty.
$$

Moreover, assume that for some $\eta < 2^{-1}$ and some $K_2 > 0$, the real numbers $\pi_{ij}(s, t)$ satisfy the conditions

$$
T \sum_{s=1}^{T} \left| \pi_{ij}(s, t) \right| < K_2 \eta^n \text{ for all } T \in \mathbb{Z}^+ \text{ and all } 1 < t < T;
$$
and

\[ T \sum_{t=1}^{T} | \pi_{ij}^{(s,t)}(s, t) | < K_2 T^n \] for all \( T \in \mathbb{Z}^+ \) and all \( 1 < s < T \).

Let \( X = \mu + u \). Then as \( T \to \infty \),

(i) \( T^{-1}[(I_2 \ast u')\Sigma \text{vec}(v)] + O_{rk \times 1} \) with probability one,

(ii) \( T^{-1}[(I_2 \ast u')\Sigma \text{vec}(v)] - \Gamma_{uv}(0) + O_{rk \times 1} \) with probability one;

and thus,

(iii) \( T^{-1}[(I_2 \ast \chi')\Sigma \text{vec}(v)] - \Gamma_{uv}(0) + O_{rk \times 1} \) with probability one.

**Proof.** The hypotheses given above satisfy the conditions of Lemma 3.14 with \( \alpha = 0 \), so the results follow from the conclusions of Lemma 3.14. □

Now consider the asymptotic behavior of

\[ T^{-1/2} \left[ (I_2 \ast \chi')\Sigma \text{vec}(v) - T \text{vec}[\Gamma_{uv}(0)] \right]. \]

Expression (3.82) indicates that
\[ T^{-1} \text{Var}[(I_r \otimes X') \otimes \text{vec}(\varphi)] \quad (3.98) \]

has \(((i, j), (\xi, m))\)-th element equal to

\[ T^{-1} \sum_{g=1}^{r} \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi_{ig}(t, s) \pi_{hj}(q, n) \]

\[ \times \{ \Gamma_{vvh}(q-s)(\mu_{nm} \nu_{tj} + \Gamma_{uu}(n-t)) \} + \Gamma_{uvhn}(q-t) \Gamma_{uvn}(s-n) \]

\[ + M_{vvgh}(0, t-s, q-s) \mu_{nm} + \nu_{tj} M_{vvgh}(0, n-s, q-s) \]

\[ + k_{uvvgh}(0, s-t, n-t, q-t) \} . \quad (3.99) \]

The following lemma gives conditions under which expression (3.99) converges. The lemma also permits one to evaluate the resulting limit.

**Lemma 3.16.** Assume that \( \{\mu_{ct}\} \) and \( \{\nu_{ct}\} \) satisfy the conditions of Lemma 3.15. Assume moreover that the real numbers \( \pi_{ij}(s, t) \), \( 1 \leq i, j \leq r \), \( s, t \in \mathbb{Z}^+ \), satisfy the following conditions.

1. For all \( d \in \mathbb{Z} \),

\[ \pi_{ij}(d) \equiv \lim_{s \to \infty} \pi_{ij}(s, s+d) \quad (3.100.a) \]

exists and is finite.
(ii) There exists an absolutely summable sequence of positive real numbers \( \{M_{md}, d \in \mathbb{Z}\} \) such that

\[
\left| \tilde{\pi}_{ij}(s, s+d) - \pi_{ij}(d) \right| < \frac{1}{100} M_{md} \tag{3.100.b}
\]

for all \( s \in \mathbb{Z}^+ \) and all \( 1 \leq i, j \leq r \).

(iii) For all \( 1 \leq i, j \leq r \),

\[
\sum_{d=-\infty}^{\infty} \left| \pi_{ij}(d) \right| < \infty. \tag{3.100.c}
\]

Then as \( T \to \infty \), expression (3.99) converges to the limit

\[
\sum_{r=1}^{r} \sum_{h=1}^{r} \sum_{d=-\infty}^{\infty} \sum_{g=1}^{\infty} \pi_{ig}(s) \mu_{jg}(d-q) \left\{ \Gamma_{vvgm}(q-s) \left[ \tilde{M}_{ujjm}(d) + \tilde{M}_{uujm}(d) \right] + \Gamma_{uvjh}(q) \Gamma_{uvjm}(s-d) + M_{vuvj}h(0, -s, q-s) \mu_{ujjm} \right. \\
\left. + \tilde{M}_{jvvgmh}(0, d-s, q-s) + k_{uvuvjgmh}(0, s, d, q) \right\}. \tag{3.101}
\]

Proof. Since the lemma asserts only element-by-element convergence of the \( r^2 k^2 \) sequences represented by (3.99), it suffices to prove the result for the case \( r = k = 1 \).

In the proof below, part (a) establishes that
part (b) establishes that

\[ \lim_{T \to \infty} T^{-1} \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \pi(q, n) \Gamma_{vv}(q-s) \mu_n \mu_t \]

\[ = \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{\pi}(s) \Gamma_{vv}(q-s) \tilde{\pi}(d-q) \mu_{uv}(d) ; \quad (3.102.a) \]

part (c) establishes that

\[ \lim_{T \to \infty} T^{-1} \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \pi(q, n) \Gamma_{uu}(q-t) \Gamma_{uv}(s-n) \]

\[ = \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{\pi}(s) \Gamma_{vv}(q-s) \tilde{\pi}(d-q) \Gamma_{uv}(q) ; \quad (3.102.b) \]

part (d) establishes that

\[ \lim_{T \to \infty} T^{-1} \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \pi(q, n) M_{uvv}(0, t-s, q-s) \mu_n \mu_t \]

\[ = \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{\pi}(s) M_{uvv}(0, -s, q-s) \tilde{\pi}(d-q) \mu ; \quad (3.102.d) \]

and part (e) establishes that
\[
\lim_{T \to \infty} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) k_{uvuv}(0, s-t, n-t, q-t) \\
= \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \pi(s) \pi(d-q) k_{uvuv}(0, s, d, q). \\
\] (3.102.e)

a. Consider first the convergence of

\[
T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) \Gamma_{vv}(q-s) \mu_n \mu_t. \\
\] (3.103)

For each \( t, n \in \mathbb{Z}^+ \), define

\[
\alpha(t, n) = \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \pi(t, s) \Gamma_{vv}(q-s) \pi(q, n). \\
\]

Since \( \Gamma_{vv}(q-s) \) satisfies conditions (10.3) trivially, two applications of Result 10.4 imply that each element of \( \{\alpha(t, n); t, n \in \mathbb{Z}^+\} \) exists; that this set satisfies conditions (10.4) as well; that for all \( d \in \mathbb{Z} \),

\[
\overline{\alpha}(d) \equiv \lim_{s \to -\infty} \alpha(s, s+d) = \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \pi(s) \Gamma_{vv}(q-s) \pi(d-q); \\
\]

that \( \overline{\alpha}(d) \) is absolutely summable; and that there exists an absolutely summable sequence \( \{M_{\alpha d}, d \in \mathbb{Z}\} \) such that

\[
|\alpha(s, s+d) - \overline{\alpha}(d)| < M_{\alpha d} \\
\]

for all \( s \in \mathbb{Z}^+ \) and all \( d \in \mathbb{Z} \).
One may establish the convergence of expression (3.103) by showing that each of the three addends in (3.104) converges to zero as \( T \to \infty \). Two applications of Result 10.4.c imply that

\[
\lim_{T \to \infty} \sum_{t=1}^{T} \sum_{n=1}^{T} \left| \alpha(t, n) - \sum_{s=1}^{T} \sum_{q=1}^{T} \pi(t, s) \Gamma_{\nu \nu}(q-s) \pi(q, n) \right| = 0 .
\]

so

\[
\lim_{T \to \infty} \left| \sum_{t=1}^{T} \sum_{n=1}^{T} \left[ \alpha(t, n) - \sum_{s=1}^{T} \sum_{q=1}^{T} \pi(t, s) \Gamma_{\nu \nu}(q-s) \pi(q, n) \right] \mu_{t} \mu_{n} \right| < B_{2} \lim_{T \to \infty} \sum_{t=1}^{T} \sum_{n=1}^{T} \left| \alpha(t, n) - \sum_{s=1}^{T} \sum_{q=1}^{T} \pi(t, s) \Gamma_{\nu \nu}(q-s) \pi(q, t) \right| = 0 ,
\]

where \( B_{2} \) is a uniform bound on \( |\mu_{t}| \), \( t \in \mathbb{Z}^{+} \). Thus, the first addend of (3.104) converges to zero as \( T \to \infty \).

Now pick \( \varepsilon > 0 \). The absolute summability of \( M_{\alpha d} \) implies that there exists some \( T_{\varepsilon} > 0 \) such that

\[
\sum_{d=1}^{T_{\varepsilon}} \left[ M_{\alpha d} + M_{\alpha,-d} \right] < 2^{-1} B_{2}^{-2} \varepsilon ,
\]

and that

\[
\bar{M} = \sum_{d=-\infty}^{\infty} M_{\alpha d} + 1
\]
is finite. Moreover, the uniformity in \( \epsilon \) of the convergence

\[
\lim_{s \to \infty} a(s, s+\epsilon) = \overline{a}(\epsilon)
\]

implies that there exists some \( S_{1\epsilon} \) such that if \( s > S_{1\epsilon} \), then

\[
|a(s, s+d) - \overline{a}(d)| < s^{-1}B^{-2}T^{-1} \epsilon.
\]

Let

\[
A_1 = \max_{1 \leq s \leq S_{1\epsilon}} \max_{-T_{1\epsilon} \leq d \leq T_{1\epsilon}} |a(s, s+d) - \overline{a}(d)|,
\]

define \( T_{2\epsilon} = \frac{8B^2}{2T_{1\epsilon} S_{1\epsilon} A_1^{-1}} \), and note that if \( T > T_{2\epsilon} \), then

\[
\frac{2B^2}{T_{1\epsilon} S_{1\epsilon} A_1^{-1}} < 4^{-1} \epsilon.
\]

Then \( T > T_{2\epsilon} \) implies that

\[
\left| \sum_{d=-T+1}^{T-1} \sum_{s=T+1}^{S_{1\epsilon}} (a(s, s+d) - \overline{a}(d)) \right| B^2
\]

\[
< \sum_{d=-T+1}^{T-1} \sum_{s=T+1}^{S_{1\epsilon}} \left| a(s, s+d) - \overline{a}(d) \right| |B^2|
\]

\[
< \sum_{d=T_{1\epsilon}}^{\infty} \sum_{s=S_{1\epsilon}+1}^{T_{1\epsilon}} (M{ad} + M\alpha_d) + \sum_{d=-T_{1\epsilon}}^{T_{1\epsilon}} \sum_{s=1}^{S_{1\epsilon}} \left| a(s, s+d) - \overline{a}(d) \right|
\]

\[
+ \left| a(s, s+d) - \overline{a}(d) \right|
\]
Thus, the second addend of expression (3.104) converges to zero as $T \to \infty$.

Note that since $|\mu| < B_2$, $|\bar{M}_\mu(d)| < B_2^2$ for all $d \in \mathbb{Z}$. By the absolute summability of $\bar{\alpha}(z)$, there exists some $T_{3\varepsilon} > 0$ such that

$$
\sum_{d=T_{3\varepsilon}}^{\infty} \left[ |\bar{\alpha}(z)| + |\bar{\alpha}(-z)| \right] < B_2^{-2}\varepsilon,
$$

and the number

$$
A_2 \equiv \left[ \sum_{d=-\infty}^{\infty} |\bar{\alpha}(d)| \right] + 1
$$

is finite. Recall that for any given $d \in \mathbb{Z}$,

$$
\lim_{T \to \infty} T^{-1} \sum_{s=m_2d}^{m_2d} \mu_s \mu_{s+d} = \lim_{T \to \infty} T^{-1} \sum_{s=m_1d}^{m_1d} \mu_s \mu_{s+d} = \bar{M}_\mu(d),
$$

so there exists some $T_{4\varepsilon}$ such that if $T > T_{4\varepsilon}$, then
for all $-T_3 \epsilon < d < T_3 \epsilon$. Let $T_5 \epsilon = \max(T_3 \epsilon, T_4 \epsilon)$ and note that if $T > T_5 \epsilon$, then

$$| \sum_{d=-T+1}^{m_2d} \alpha(d)T^{-1} \sum_{s=m_1d}^{m_2d} \mu_s \mu_{s+d} - \sum_{d=-\infty}^{\infty} \alpha(d)\overline{\nu}(d)|$$

$$\leq 4B^2_2 \sum_{d=T_3 \epsilon}^{\infty} \left[ |\alpha(d)| + |\overline{\alpha}(d)| \right]$$

$$+ \sum_{d=-T_3 \epsilon+1}^{T_3 \epsilon-1} |\alpha(d)| \cdot \left[ \sum_{s=m_1d}^{m_2d} \mu_s \mu_{s+d} - \overline{\nu}(d) \right]$$

$$\leq 4B^2_2 T_3 \epsilon^{-2} + A_2 A_2^{-1} T_3 \epsilon^{-1}$$

$$= \epsilon.$$
so condition (3.102.a) is established.

b. Now consider the limit of

\[ T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) \Gamma_{vv}(q-s) \Gamma_{uu}(q-n) \]

as \( T \to \infty \).

By an argument similar to that for expression (3.104),

\[
|T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \Gamma_{vv}(q-s) \pi(q, n) \Gamma_{uu}(t-n) - \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{\pi}(s) \Gamma_{vv}(q-s) \tilde{\pi}(d-q) \Gamma_{uu}(d)|
\]

\[
< T^{-1} \sum_{t=1}^{T} \sum_{n=1}^{T} \left[ \alpha(t, n) - \sum_{s=1}^{T} \sum_{q=1}^{T} \pi(t, s) \Gamma_{vv}(q-s) \pi(q, n) \Gamma_{uu}(t-n) \right]
\]

\[
+ \left| \sum_{d=-T+1}^{d=m} \sum_{s=m}^{d} \alpha(s, s+d) - \alpha(d) \right| \Gamma_{uu}(d)
\]

\[
+ \left| \sum_{d=-T+1}^{d=T-1} \left[ T - |d| \right] \alpha(d) \right| \Gamma_{uu}(d) - \sum_{d=-\infty}^{\infty} \alpha(d) \Gamma_{uu}(d) \right|
\]

(3.106)

Recall that the arguments for the convergence of the first two addends of (3.104) to zero involved the \( u_s \) terms only through the remark that \( |u_s u_{s+d}| \) was uniformly bounded. Since \( |\Gamma_{uu}(d)|, d \in \mathbb{Z} \) is also uniformly bounded, these same arguments indicate that the first two addends of (3.106) converge to zero as \( T \to \infty \). Since
\[
\lim_{T \to \infty} \sum_{d=-T+1}^{T-1} [T - |d|] \tilde{\alpha}(d) \Gamma_{uu}(d) = \sum_{d=-\infty}^{\infty} \tilde{\alpha}(d) \Gamma_{uu}(d),
\]

the final addend of (3.106) also converges to zero as \( T \to \infty \), so

\[
\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \Gamma_{vv}(q-s) \pi(q, n) \Gamma_{uu}(t-n) = \sum_{d=\infty}^{\infty} \sum_{s=\infty}^{\infty} \sum_{q=\infty}^{\infty} \tilde{\pi}(s) \Gamma_{vv}(q-s) \tilde{\pi}(d-q) \Gamma_{uu}(d),
\]

and condition (3.102.b) is established.

c. Now consider the limit of

\[
T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) \Gamma_{uv}(q-t) \Gamma_{uv}(s-n)
\]

as \( T \to \infty \). For each \( t, q \in \mathbb{Z}^+ \), define

\[
\beta(t, q) = \sum_{s=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \Gamma_{uv}(s-n) \pi(q, n).
\]

Since \( \Gamma_{uv}(s-n) \) satisfies conditions (10.3) trivially, two applications of Result 10.4 imply that each element of \( \{\beta(t, q) : s, n \in \mathbb{Z}^+\} \) exists; that this set satisfies conditions (10.4) as well; that for all \( q \in \mathbb{Z} \),

\[
\bar{\beta}(q) \equiv \lim_{t \to \infty} \beta(t, t+q) = \sum_{s=-\infty}^{\infty} \sum_{d=-\infty}^{\infty} \tilde{\pi}(s) \Gamma_{uv}(s-d) \tilde{\pi}(d-q)
\]

exists and is finite; that \( \bar{\beta}(q) \) is absolutely summable; and that there
exists an absolutely summable sequence \( \{ \beta_q, q \in \mathbb{Z} \} \) such that

\[
| \beta(t, t+q) - \beta(q) | < \beta_q
\]

for all \( t \in \mathbb{Z}^+ \) and all \( q \in \mathbb{Z} \).

Now

\[
\left| \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \Gamma_{uv}(s-n) \pi(q, n) \Gamma_{uv}(q-t) \right|
\]

\[
- \sum_{d=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \pi(s) \Gamma_{uv}(s-d) \pi(d-q) \Gamma_{uv}(q)
\]

\[
= \left| \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \Gamma_{uv}(s-n) \pi(q, n) \Gamma_{uv}(q-t) \right|
\]

\[
- \sum_{q=-\infty}^{\infty} \beta(q) \Gamma_{uv}(q)
\]

\[
< \left| \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \Gamma_{uv}(s-n) \pi(q, n) \Gamma_{uv}(q-t) \right|
\]

\[
- \sum_{t=1}^{\infty} \sum_{q=1}^{\infty} \beta(t, q) \Gamma_{uv}(q-t)
\]

\[
+ \left| \sum_{t=1}^{\infty} \sum_{q=1}^{\infty} \left[ \beta(t, q) - \beta(q-t) \right] \Gamma_{uv}(q-t) \right|
\]

\[
+ \left| \sum_{t=1}^{\infty} \sum_{q=1}^{\infty} \beta(q-t) \Gamma_{uv}(q-t) - \sum_{q=-\infty}^{\infty} \beta(q) \Gamma_{uv}(q) \right| . \quad (3.108)
\]

One may establish the convergence of expression (3.108) by showing that each of the three addends in (3.108) converge to zero as \( T \to \infty \). Two
applications of Result 10.4.c imply that the first addend of expression (3.108) converges to zero. An argument similar to the argument for the second addend of expression (3.104) implies that the second addend of expression (3.108) also converges to zero as $T \to \infty$. Finally, the third addend of expression (3.108) is

$$
|T^{-1} \sum_{t=1}^{T} \sum_{q=1}^{T} \tilde{B}(q-t) \Gamma_{uv}(q-t) - \sum_{q=1}^{\infty} \tilde{B}(q) \Gamma_{uv}(q)|
$$

$$
= \left| \sum_{q=-T+1}^{T-1} |q| \tilde{B}(q) \Gamma_{uv}(q) - \sum_{q=1}^{\infty} \tilde{B}(q) \Gamma_{uv}(q) \right|
$$

which converges to zero as $T \to \infty$, by the absolute summability of $\tilde{B}(q)$ and $\Gamma_{uv}(q)$. Hence, condition (3.102.c) is established.

d. Now consider the limit of

$$
T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{q=1}^{T} \sum_{n=1}^{T} \pi(t, s) \pi(q, n) M_{vuv}(0, t-s, q-s) u_n
$$

as $T \to \infty$. For each $s, n \in \mathbb{Z}^+$, define

$$
\delta(s, n) = \sum_{t=1}^{\infty} \sum_{q=1}^{\infty} \pi(t, s) \pi(q, n) M_{vuv}(0, t-s, q-s) u_n
$$

Since for each $s$, $M_{vuv}(0, t-s, q-s)$ satisfies conditions (10.3), arguments similar to those of Result 10.4 imply that each element of $\{\delta(s, n): s, n \in \mathbb{Z}^+\}$ exists; that this set satisfies conditions (10.4) as well; that for all $n \in \mathbb{Z}$. 


\[ \delta(n) \equiv \lim_{s \to \infty} \delta(s, s+n) \]

\[ = \sum_{t=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \pi(t) \mu_{vuv}(0, -t, q-t) \tilde{\pi}(n-q) \tilde{\mu} \]

exists and is finite, that \( \delta(n) \) is absolutely summable; that there exists an absolutely summable sequence \( \{M_{\delta n}, n \in \mathbb{Z}\} \) such that

\[ |\delta(s, s+n) - \delta(n)| < M_{\delta n} \]

for all \( s \in \mathbb{Z}^+ \) and all \( n \in \mathbb{Z} \). Now

\[ |T^{-1} \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \pi(q, n) \mu_{vuv}(0, t-s, q-s) \mu_n \]

\[ - \sum_{n=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \pi(t) \mu_{vuv}(0, -t, q-t) \tilde{\pi}(n-q) \tilde{\mu} | \]

\[ = |T^{-1} \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} \sum_{q=1}^{\infty} \sum_{n=1}^{\infty} \pi(t, s) \pi(q, n) \mu_{vuv}(0, t-s, q-s) \mu_n \]

\[ - \sum_{n=-\infty}^{\infty} \delta(n) \tilde{\mu} | \]

(3.109)
One may establish the convergence of expression (3.109) to zero by showing that (3.110.a), (3.110.b) and (3.110.c) each converge to zero as $T \to \infty$. Arguments similar to those for Result 10.4.c imply that expression (3.110.a) converges to zero as $T \to \infty$. An argument similar to the argument for the second addend of expression (3.104) implies that (3.110.b) also converges to zero as $T \to \infty$. Also, expression (3.110.c) equals

$$
\left| \sum_{s=1}^{T-1} \sum_{n=1}^{T-1} \delta(n-s) \mu_n - \sum_{n=-\infty}^{\infty} \delta(n) \bar{\mu} \right|
$$

$$
= \left| \sum_{n=-T+1}^{T-1} \delta(n) \left( \sum_{\ell=1}^{m_2} \mu_\ell \right) - \sum_{n=-\infty}^{\infty} \delta(n) \bar{\mu} \right| \quad (3.111)
$$

$$
\leq \sum_{n=T}^{\infty} \left( |\delta(n)| + |\delta(-n)| \right) |\bar{\mu}|
$$

$$
+ \left| \sum_{n=-T+1}^{T-1} \delta(n) \left( \sum_{\ell=1}^{m_2} \mu_\ell \right) - \bar{\mu} \right| \quad (3.112.a)
$$

$$
+ \left| \sum_{n=-T+1}^{T-1} \delta(n) \left( \sum_{\ell=1}^{m_2} \mu_\ell \right) - \bar{\mu} \right| \quad (3.112.b)
$$
where \( m_{1n} = \max(1, l+n) \) and \( m_{2n} = \min(T, T+n) \). Pick \( \varepsilon > 0 \).

Recall that \( |\mu_t| < B_2 \) for all \( t \), so that \( |\overline{\mu}| < B_2 \) as well. Also, recall that \( |m_{2n} - m_{1n}| < T \) for all \( n = -T+1, -T+2, \ldots, T-1 \). By the absolute summability of \( \delta(n) \), there exists some \( T_{1\varepsilon} \in \mathbb{Z}^+ \) such that

\[
\sum_{n=T_{1\varepsilon}}^{\infty} \max(1, n) |\delta(n)| + |\delta(-n)| < (B_2 + 1)^{-\varepsilon}.
\]

Hence, expression (3.112.a) converges to zero as \( T \to \infty \). Moreover, if \( T > T_{1\varepsilon} \),

\[
\left| \sum_{n=-T+1}^{T-1} \delta(n) \left( T^{-1} \sum_{\ell=m_{1n}}^{m_{2n}} \mu_{\ell} \right) - \overline{\mu} \right| \\
< \left| \sum_{n=-T_{1\varepsilon}+1}^{T_{1\varepsilon}-1} \delta(n) \left( T^{-1} \sum_{\ell=m_{1n}}^{m_{2n}} \mu_{\ell} \right) - \overline{\mu} \right| \\
+ 2 B_2 \left( \sum_{n=T_{1\varepsilon}}^{T} \max(1, n) |\delta(n)| + |\delta(-n)| \right) \\
< \left| \sum_{n=-T_{1\varepsilon}+1}^{T_{1\varepsilon}-1} \delta(n) \left( T^{-1} \sum_{\ell=m_{1n}}^{m_{2n}} \mu_{\ell} \right) - \overline{\mu} \right| + \varepsilon. \tag{3.113}
\]

Now let \( 0 < n < T_{1\varepsilon} \). Then \( m_{1n} = l+n \) and \( m_{2n} = T \), so

\[
\left| \left( T^{-1} \sum_{\ell=m_{1n}}^{m_{2n}} \mu_{\ell} \right) - \overline{\mu} \right| = \left| \left( T^{-1} \sum_{t=l+n}^{T} \mu_{t} \right) - \overline{\mu} \right|
\]
Since \( n \) is fixed, expression (3.114.a) converges to zero as \( T \to \infty \).

By the definition of \( \bar{u} \), expression (3.114.b) converges to zero as \( T \to \infty \). Hence, for each \( n=1, 2, ..., T_{1\varepsilon}^{-1} \),

\[
\lim_{T \to \infty} \left| \frac{1}{T} \sum_{k=1}^{T} \left| \mu_k \right| - \bar{u} \right| = 0 . 
\] (3.115)

A similar proof establishes condition (3.115) for each \( n = -T_{1\varepsilon}^{-1} + 1, -T_{1\varepsilon}^{-1} + 2, ..., 0 \). Since \( T_{1\varepsilon} \) is fixed with respect to \( T \), it follows that

\[
\lim_{T \to \infty} \left| \sum_{n=-T_{1\varepsilon}^{-1} + 1}^{T_{1\varepsilon}^{-1}} \delta(n) \left( \frac{1}{T} \sum_{k=m_{1n}}^{m_{2n}} \left| \mu_k \right| - \bar{u} \right) \right| = 0 . 
\]

This result and expression (3.113) imply that expression (3.112.b) converges to zero as \( T \to \infty \). Hence, expression (3.110.c) converges to zero as \( T \to \infty \), so expression (3.109) also converges to zero as \( T \to \infty \) and condition (3.102.d) is established.

e. Condition (3.102.e) is established by arguments similar to those for conditions (3.102.c) and (3.102.d).
As indicated by expression (3.102.a)-(3.102.e), parts (a) through (e) of the proof above establish the convergence of expression (3.99) to expression (3.101) as \( T \to \infty \), so the lemma is proved. \( \square \)

Lemma 3.16 established some convergence properties of \( \text{Var}[T^{-1/2}(I_r \otimes X')_{\pi} \text{vec}(v)] \). It is also of interest to study the convergence of \( T^{-1/2}(I_r \otimes X')_{\pi} \text{vec}(v) \) and its expectation. Recall that

\[
E[T^{-1}(I_r \otimes X')_{\pi} \text{vec}(v)] = \text{vec} \{ \Gamma_{\pi T V}(0) \}
\]

has \((i,j)\)-th double-subscripted element equal to

\[
\Gamma_{uvvij}(0) = T^{-1} \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{uvjh}(s-t)_{\pi h}(t, s).
\]

Now conditions (3.100) and Result 10.4 imply that

\[
\lim_{T \to \infty} \Gamma_{uvvij}(0) = \lim_{T \to \infty} T^{-1} \sum_{h=1}^{r} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{uvjh}(s-t)_{\pi h}(t, s)
\]

\[
= \sum_{h=1}^{r} \sum_{d=-\infty}^{\infty} \Gamma_{uvjh}(d)_{\pi h}(d).
\]

Let \( \pi \) be the \( Tr \times Tr \) matrix with \([(i, s), (j, s+d)]\)-th element equal to \( \pi_{ij}(d), 1 < s < T, -T + 1 < d < T - 1 \); let \( \pi_{dij} \) be a \( T \times T \) matrix with \((s, s+d)\)-th element equal to \( \pi_{ij}(d), 1 < s < T, -T + 1 < d < T - 1 \); and let \( \pi_{ij}(d) \) be an \( r \times r \) matrix with \((i,j)\)-th
element equal to \( \bar{\pi}_{ij}(d) \), \( d \in \mathbb{Z} \), where \( \bar{\pi}_{ij}(d) \) is defined in (3.100.a). Also, define \( \Gamma_{uvw}^{(0)} \) to be an \( r \times k \) matrix with \((i,j)\)-th element equal to

\[
\Gamma_{uvwji}^{(0)} = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{h=1}^{s-t} \Gamma_{vhj}^{(s-t)} \bar{\pi}_{ih}^{(s-t)}
\]

\[= T^{-1} E[X_{ij} \vec{v}] \] \hspace{1cm} (3.116)

for all \( 1 < j < k \) and \( 1 < i < r \). The convergence properties of

\[
T^{-1/2} \{(I_r \otimes X')(\bar{\pi} - \bar{\bar{\pi}})\vec{v} - T \vec{v} \} \rightarrow 0_{rk \times 1}
\]

\[\rightarrow O_{rk \times 1} \] \hspace{1cm} (3.119.a)
\[
\Gamma_{\sim u \pi v}(0) - \Gamma_{\sim u \pi v}(0) + O_{rk \times 1};
\]
and
\[
T^{-1}\{(I - X')(\bar{π} - \bar{π})\text{vec}(v) - T \text{vec}[\Gamma_{\sim u \pi v}(0) - \Gamma_{\sim u \pi v}(0)]\} + O_{rk \times 1}
\]
with probability one.

b. Assume moreover that \{\pi_i j(T)(s, t)\} and \{\pi_i j(s, t)\} satisfy conditions (3.86). Then as \( T \rightarrow \infty \),
\[
T^{-1/2} [(I - X')(\pi(\hat{T}) - \bar{π})\text{vec}(v) - T \text{vec}[\Gamma_{\sim u \pi v}(T)(0) - \Gamma_{\sim u \pi v}(0)]])
\]
converges to \( O_{rk \times 1} \) in probability, and
\[
\Gamma_{\sim u \pi v}(T)(0) - \Gamma_{\sim u \pi v}(0) \text{ converges to } O_{rk \times 1}.
\]

c. Assume moreover that \{\pi_i j(T)(s, t)\} and \{\pi_i j(s, t)\} satisfy conditions (3.85). Then the conclusions of part (b) hold, and in addition
\[
T^{-1}\{(I - X')(\pi(\hat{T}) - \bar{π})\text{vec}(v) - T \text{vec}[\Gamma_{\sim u \pi v}(T)(0) - \Gamma_{\sim u \pi v}(0)]\} + O_{rk \times 1}
\]
with probability one.

Proof.

a. Expressions (3.78) and (3.116) imply that expression (3.117) has expectation equal to zero. Lemma 3.14 implies that

\[ T^{-1} \text{Var}(I(T \cdot \mathbf{X'})(\mathbf{y} - \bar{y})\text{vec}(\mathbf{v})) \] converges to zero as \( T \to \infty \), so expression (3.119.a) follows by the Chebyshev inequality. Also, expression (3.119.b) follows from expressions (3.100) and (3.116), Result 10.4, and the fact that

\[
\lim_{T \to \infty} \frac{r}{T} \sum_{h=1}^{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E} \mathbb{E} \mathbb{E} g_{u(vj)}(s-t) p_{lh}(s-t) = \frac{r}{T} \lim_{T \to \infty} \sum_{h=1}^{T} \sum_{d=-T+1}^{T-1} |d| g_{uvj}(d) p_{lh}(d) \]

where the final equality follows from the fact that \( \{g_{u(vj)(d)}\} \) and \( \{p_{lh}(d)\} \) are absolutely summable.

The almost sure convergence in (3.119.c) follows from an application of Lemma 3.14 with \( \alpha = \eta = 0 \).

b. To establish the convergence of expression (3.119.d) and (3.119.e), note that

\[
T^{-1/2} [(I(T \cdot \mathbf{X'})(\bar{y}(T) - \bar{y})\text{vec}(\mathbf{v}) - T \text{vec}[\Gamma_{uvv}(0) - \Gamma_{uv}(0)] (3.120.a)
\]
\[ T^{-1/2} \left( (I_T \otimes X') \left[ \delta(T) - \frac{1}{T} \right] \text{vec}(v) - T \text{vec}[\Gamma_{u\nu T}(0) - \Gamma_{u\nu}](0) \right) \]  
\[ + T^{-1/2} \left( (I_T \otimes X') \left[ \phi - \frac{1}{T} \right] \text{vec}(v) - T \text{vec}[\Gamma_{u\nu T}(0) - \Gamma_{\nu}](0) \right) \]  
(3.120.b)
and

\[ \Gamma_{u\nu T}(0) - \Gamma_{\nu}(0) \]  
(3.121.a)

\[ = \Gamma_{u\nu T}(0) - \Gamma_{\nu}(0) \]  
(3.121.b)

\[ + \Gamma_{u\nu}(0) - \Gamma_{\nu}(0). \]  
(3.121.c)

Conditions (3.86) and Lemma 3.13.b imply that as \( T \to \infty \), expression (3.120.b) converges in probability to \( 0_{r_k \times 1} \) and expression (3.121.b) converges to \( 0_{r_k \times 1} \). Also, Lemma 3.17.a implies that as \( T \to \infty \), expression (3.120.c) converges in probability to \( 0_{r_k \times 1} \) and expression (3.121.c) converges to \( 0_{r_k \times 1} \). Hence, as \( T \to \infty \), expression (3.120.a) converges in probability to \( 0_{r_k \times 1} \) and expression (3.121.a) converges to \( 0_{r_k \times 1} \), so part (b) is established.

c. Recall that conditions (3.85) imply conditions (3.86). Hence, under the assumptions of part (c), the conclusions of part (b) still hold. To establish the almost sure convergence in expression (3.119.f), note that
Conditions (3.85) and Lemma 3.13.a imply that as $T \to \infty$, expression (3.122.b) converges to $0_{rk\times 1}$ with probability one. Also, Lemma 3.17.a implies that as $T \to \infty$, expression (3.122.c) converges to $0_{rk\times 1}$ with probability one. Hence, as $T \to \infty$, expression (3.122.a) converges to $0_{rk\times 1}$ with probability one. \hfill \Box

The preceding sequence of lemmas addresses the convergence with probability one and the convergence in probability of various bilinear functions of $X$ and $\nu$. It remains to establish the asymptotic normality of such functions. Theorem 3.2 addresses bilinear functions with weight matrix $\tilde{\pi}$, while Corollary 3.2.1 addresses bilinear functions with weight matrices $\tilde{\pi}$ and $\tilde{\pi}(T)$.

**Theorem 3.2.** Let $\{\pi_{i, j}(d)\}$ and $\{\pi_{i, j}^{(d)}(d)\}$ satisfy the conditions of Lemma 3.16. Assume that the sequence of random vectors $\{\zeta_{\ell}\}$ satisfies the definition of linear process given by Definition 3.2. Define $\tilde{\pi}$ to be a $\text{Tr} \times \text{Tr}$ matrix with $[(i, s), (j, s+d)]$-th element equal to $\pi_{i, j}(d)$, $1 \leq s < T$, $-T + 1 < d < T - 1$, as defined in (3.100.a); and
define $\tilde{\pi}_\infty(0)$ as in (3.116). Assume also that for some $v > 0$,

$$\max_{1 \leq i \leq p} \sup_{t} \left| c_{t,i} \right|^{4+2v} \leq K_v < \infty,$$

where $\{c_{t,i}\}$ is the underlying process that generates $\{\tilde{\epsilon}_t\}$. Let $X_t = y_t + \tilde{\epsilon}_t$. Then as $T \to \infty$,

$$T^{-1/2} \left( (I_r \otimes \Sigma) \tilde{\pi} \text{vec}(\vec{v}) - T \text{vec}[\tilde{\pi}_\infty(0)] \right) \xrightarrow{D} N_{rk}(0, G) , \quad (3.123)$$

where $G$ is an $rk \times rk$ matrix with $(i, j), (k, m)$-th element given by expression (3.101).

The proof of Theorem 3.2 is similar to the proof of Theorem 3.1, except that introduction of the weights $\tilde{\pi}_{ij}(d)$ requires some additional notational complexity and leads to two separate applications of Lemma 3.7. Part (A) of the proof develops some notation, and notes that the Varadarajan (1958) result quoted in Lemma 3.12 allows one to restrict attention to the asymptotic properties of

$$T^{-1/2} \tilde{\pi} \left( (I_r \otimes \Sigma) \tilde{\pi} \text{vec}(\vec{v}) - T \text{vec}[\tilde{\pi}_\infty(0)] \right) \quad (3.124)$$

for an arbitrary $rk \times 1$ real vector.

Part (B) establishes a convergence result for expression (3.124) under the assumption that

$$\tilde{\pi}(d) = 0_{rk} \text{ for } |d| > L . \quad (3.125)$$
Let \( \bar{\bar{T}} \) denote the matrix \( \bar{T} \) under condition (3.125). Under condition (3.125) and a linear-process model for \( \pi \), one may define \( \bar{\bar{\pi}}(0) \) by expression (3.116). The asymptotic normality of

\[
T^{-1/2} \left\{ (T \times X') \bar{\bar{T}} vec(\nu) - T vec[\bar{\bar{\pi}}(0)] \right\}
\]

then follows from two arguments. First, part (B.1) considers the case in which \( \{ \pi = (u_t, v_t) \} \) follows a finite moving-average model,

\[
(u_t, v_t)' = \sum_{h=0}^{M} A_h c_t'.
\]

Under model (3.127) and condition (3.125), one may define \( \bar{\bar{\pi}}(0) \) by expression (3.116). Then some preliminary arguments show that

\[
T^{-1/2} \left\{ (T \times X') \bar{\bar{T}} vec(\nu) - T vec[\bar{\bar{\pi}}(0)] \right\} = T^{-1/2} \sum_{t=1}^{T} \{ Z_t c_t' - E(Z_t c_t') \} + o_p(1),
\]

where \( Z_t = e_t + f_t + g_t \), \( e_t \) is a fixed vector, \( f_t \) is a \( \sigma(c_t) \) measurable random vector, and \( g_t \) is a \( \sigma(c_s, s < t-1) \) measurable random vector. Examination of the properties of \( e_t c_t', f_t c_t' \) and \( g_t c_t' \) indicate that the sequence
Theorem 1 of Scott (1973) then indicates that as $T \to \infty$, expression (3.129) converges in law to some normal $(0, \tilde{s}^2_{LM})$ distribution. Hence, expression (3.128) also has this limiting normal distribution.

Part (B.2) establishes that given a particular sequence \( \{a_h, h \in Z\} \), and a particular \( \{c_t\} \) process, the difference between expressions (3.126) and (3.128) becomes negligible uniformly in $T$ as $M \to \infty$. It follows from Lemma 3.7 that as $T \to \infty$, expression (3.126) converges to a normal $(0, \bar{s}^2_L)$ distribution, where $\bar{s}^2_L = \lim_{M \to \infty} \bar{s}^2_{LM}$.

Part (C) establishes that given a particular sequence \( \{\pi(d)\} \), the difference between expressions (3.124) and (3.126) becomes negligible uniformly in $T$ as $L \to \infty$. It follows from Lemma 3.7 that as $T \to \infty$, expression (3.124) has a limiting normal $(0, \tilde{s}^2)$ distribution, where $\tilde{s}^2 = \lim_{L \to \infty} \tilde{s}^2_L$.

Additional arguments indicate that $\tilde{s}^2 = \delta' \Gamma_{\pi} \delta$, so result (3.123) then follows from Lemma 3.12.

Given the above outline, one may now consider the proof of Theorem 3.2 in detail.

**Proof of Theorem 3.2.**

(A). In general, the model for the $p \times 1$ linear process \( \{s_t'\} \) may be written in the form,
\[ e_t' = (u_t', v_t')' \]
\[
= \sum_{h=0}^{\infty} A_h c_{t-h}'
\]
\[
= \left[ \left( \sum_{h=0}^{\infty} R_h c_{t-h}' \right)', \left( \sum_{h=0}^{\infty} D_h c_{t-h}' \right)' \right]',
\]

where \( R_h \) is \( k \times p \), \( D_h \) is \( r \times p \), \( \{A_h = (R_h', D_h')'\} \) is an absolutely summable sequence, and the sequence of random vectors \( \{c_t\} \) satisfies conditions (3.2). Let \( R_{h,z} \) denote the \( z \)-th row of \( R_h \), let \( R_{h,z,m} \) denote the \((z, m)\)-th element of \( R_{h,z} \), and define \( D_{h,z} \) and \( D_{h,z,m} \) similarly. Let \( \delta \) be an arbitrary \( r k \times 1 \) real vector with double-subscripted \((i, j)\)-th element \( \delta_{i,j} \). By the Varadarajan (1958) result quoted in Lemma 3.12, it suffices to show that for any \( \delta \in \mathbb{R}^{rk} \), as \( T \to \infty \),

\[
T^{-1/2} \delta' \left\{ \left( \mathbf{I}_r \otimes \mathbf{K} \right) \sum_{\mathbf{w}} \mathbf{w}' \mathbf{w} \right\} \mathbf{1} \to N \left( \mathbf{0}, \delta' \left[ C_{\gamma} \right] \delta \right).
\]

Result (3.131) will be established by several steps in parts (B) and (C) below.

(B) For the time being, let \( M \) be some fixed positive integer, and consider the case in which \( e_t = (u_t, v_t) \) has a finite moving average representation,
Under model (3.132), denote the moments and other parameters of the 
$$(u_t, v_t)$$ process with the subscript $M$, e.g.

$$
(\mathbf{u}_t, \mathbf{v}_t)' = \sum_{h=0}^{M} \mathbf{A}_h \mathbf{c}'_{t-h}
$$

$$
= \left[ (\sum_{h=0}^{M} \mathbf{B}_h \mathbf{c}'_{t-h})', (\sum_{h=0}^{M} \mathbf{D}_h \mathbf{c}'_{t-h})' \right]' .
$$

(B.1). Under condition (3.134), note that for $T > L$,

$$
\gamma(\ell) = \mathbf{0}_{r \times r} \quad \text{for all } |\ell| > L ,
$$

where $L$ is some positive integer. Let $\mathbf{\Sigma}_L$ denote the $r \times r$ matrix $\mathbf{\Sigma}$ under condition (3.134), and define $\Gamma_{L \times L \times L \times L}$ from expressions (3.116), (3.133) and (3.134). Development of the central limit result under condition (3.134) will now proceed in two steps.

First, part (B.1) establishes asymptotic normality under conditions (3.132) and (3.134). Second, part (B.2) uses Lemma 3.7 to extend the result of part (B.1) to model (3.130) under condition (3.134).
\[ T^{-1/2} \delta'(I_r \otimes X') \mathbf{\Xi} \mathbf{\Psi}_L \mathbf{\vec{v}} = E[(I_r \otimes X') \mathbf{\Xi} \mathbf{\Psi}_L \mathbf{\vec{v}}] \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} T^{-1/2} \sum_{h=1}^{r} \left[ (X' \mathbf{\Xi} \mathbf{\Psi}_L \mathbf{\vec{v}})_{ih} - \mathbf{E}(u' \mathbf{\Xi} \mathbf{\Psi}_L \mathbf{\vec{v}}) \right] \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} T^{-1/2} \sum_{h=1}^{T} \sum_{t=1}^{T} \left[ X_{sj} \mathbf{\Xi} \mathbf{\Psi}_L (t-s) \mathbf{\vec{v}}_{ts} - \mathbf{E}(u_{sj} \mathbf{\Xi} \mathbf{\Psi}_L (t-s) \mathbf{\vec{v}}_{th}) \right] \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} T^{-1/2} \sum_{h=1}^{T} \sum_{s=1}^{T} \mathbf{\Xi} \mathbf{\Psi}_L (t-s) [X_{sj} \mathbf{\vec{v}}_{th} - \mathbf{\Gamma}_{uvjh}(t-s)] \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} T^{-1/2} \sum_{h=1}^{T} \sum_{d=-T+1}^{T-1} \mathbf{\Xi} \mathbf{\Psi}_L (d) [X_{sj} \mathbf{\vec{v}}_{s+d,h} - \mathbf{\Gamma}_{uvjh}(d)] \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \delta_{ij} T^{-1/2} \sum_{h=1}^{T} \sum_{d=-L}^{L} \mathbf{\Xi} \mathbf{\Psi}_L (d) [X_{sj} \mathbf{\vec{v}}_{s+d,h} - \mathbf{\Gamma}_{uvjh}(d)] \]

where \( m_{1d} = \max(1, 1+d) \), \( m_{2d} = \min(T, T+d) \) and the final equality of (3.135) follows from condition (3.134). If \( 0 < d < L \), then \( m_{1d} = 1+d \) and \( m_{2d} = T \), so

\[ \sum_{s=1}^{m_{1d}} X_{sj} \mathbf{\vec{v}}_{s+d,h} - \sum_{s=1}^{T} X_{sd} \mathbf{\vec{v}}_{s+d,h} = \sum_{s=1}^{d} X_{sj} \mathbf{\vec{v}}_{s+d,h} . \]  

(3.136)

Since \( L \) is fixed, the bounds \( 0 < d < L \) and \( |u_{sj}| < B \); and the absolute summability of \( \{A_n\} \) imply that expression (3.136) is \( O_p(1) \), and that the square of expression (3.136) has expectation that is \( O(1) \). A similar argument indicates the same conclusions hold for \(-L < d < 0\). Hence, for \(-L < d < L\),
and the square of the $O(T^{-1/2})$ term in expression (3.137) has expectation that is $O(T^{-1})$. Under model (3.132),

$$X_{s_j} v_{s+d,h} = (u_{s_j} + u_{s_j}) v_{s+d,h}$$

$$= [u_{s_j} + \sum_{\xi=0}^{M} B_{x_j} c'_{\xi}] \sum_{m=0}^{M} D_{m} c'_{s+d-m}$$

so

$$T^{-1/2} \sum_{s=-L}^{T} \sum_{s=1}^{T} \tilde{\pi}_h(d)X_{s_j} v_{s+d,h}$$

$$= T^{-1/2} \sum_{d=-L}^{T} \sum_{s=1}^{T} \tilde{\pi}_h(d)[u_{s_j} + \sum_{\xi=0}^{M} B_{x_j} c'_{\xi}] \sum_{m=0}^{M} D_{m} c'_{s+d-m} \cdot (3.138)$$

Now for $T > M + L$,

$$T^{-1/2} \sum_{d=-L}^{T} \sum_{s=1}^{T} \tilde{\pi}_h(d)u_{s_j} \sum_{m=0}^{M} D_{m} c'_{s+d-m}$$

$$= T^{-1/2} \sum_{n=1-L-M}^{T+L} \min(L,n-1+M) \min(T+T-d-n) \tilde{\pi}_h(d)u_{n-d+m} \sum_{m=0}^{M} D_{m} c'_{n} \cdot (3.139)$$
An argument similar to the argument for expression (3.137) indicates that for fixed \( L \) and \( M \), expression (3.139) is equal to

\[
T^{-1/2} \sum_{n=1}^{T} \sum_{d=-L}^{L} \sum_{m=0}^{M} h(d) \mu_{n-d+m,j} D_{n,j} c_n \bigg) + o(T^{-1/2}) ,
\]

and that the square of the \( O(T^{-1/2}) \) term in expression (3.140) has expectation that is \( O(T^{-1}) \). One may similarly evaluate the four-fold sum of expression (3.138). Assume for the moment that \( s > L + M \) and consider the following three cases.

**Case 1.** Let \( s - \ell = s + d - m \), or equivalently,

\[
m = d + \ell .
\]

Under condition (3.141), three equivalent sets of bounds on the indices \( d \), \( \ell \) and \( m \) are:

\[
-L < d < L , \ 0 < m < M , \text{ and } 0 < \ell < M ;
\]

\[
-L < d < L , \ 0 < d+\ell < M , \text{ and } 0 < \ell < M ;
\]

and

\[
-B_{LM1} < d < B_{LM1} \text{ and } \max(0, -d) < \ell < \min(M, M-d) ,
\]

where \( B_{LM1} = \min(L, M) \). Define \( m_3d = \max(0, -d) \) and \( m_4d = \min(M, M-d) \).
Case 2. Let \( s - \ell > s + d - m \), or equivalently,

\[
m > d + \ell .
\]

Under condition (3.142), three equivalent sets of bounds on the indices \( d \), \( \ell \) and \( m \) are:

\[-L < d < L , \quad 0 < \ell < M , \quad \text{and} \quad 0 < m < M ;\]

\[-L < d < L , \quad 0 < \ell < M , \quad \text{and} \quad \max(0, d+\ell+1) < m < M ;\]

and

\[-L < d < B_{LM2} , \quad 0 < \ell < \min(M, M-d-1) , \quad \text{and} \quad \max(0, d+\ell+1) < m < M ,\]

where \( B_{LM2} = \min(L, M-1) \). Define \( m_d = \min(M, M-d-1) \) and \( m_{\ell M} = \max(0, d+\ell+1) \).

Case 3. Let \( s - \ell < s + d - m \), or equivalently,

\[
m - d < \ell .
\]

Under condition (3.143), three equivalent sets of bounds on the indices \( d \), \( \ell \) and \( m \) are:

\[-L < d < L , \quad 0 < m < M , \quad \text{and} \quad 0 < \ell < M ;\]

\[-L < d < L , \quad 0 < m < M , \quad \text{and} \quad \max(0, m-d+1) < \ell < M ;\]
and

\[-B_{L2} < d < L, \quad 0 < m < \min(M, M+d-1), \quad \text{and} \quad \max(0, m-d+1) < \ell < M.\]

Define \(m^\_d = \min(M, M+d-1)\) and \(m^\_8d = \max(0, m-d+1)\). Then an argument similar to the argument for expression (3.137) indicates that for fixed \(L\) and \(M\),

\[
T^{-1/2} \sum_{d=-L}^{L} \sum_{s=1}^{M} \sum_{l=0}^{M} B_{hj} c_{s-l} D_{hj} c_{s+d-m} .
\]

\[
= T^{-1/2} \sum_{d=-B_{L2}}^{B_{L2}} \sum_{l=1}^{m_4d} \sum_{m=0}^{m_5d} B_{L1} \sum_{n=1}^{m_4d} \sum_{d=-B_{L1}}^{B_{L1}} \sum_{m=0}^{m_5d} B_{hj} c_{n+d+m} c_{n}.
\]

\[
+ \sum_{d=-B_{L2}}^{B_{L2}} \sum_{m=0}^{m_7d} \sum_{\ell=0}^{\ell_8d} \sum_{m=0}^{m_8d} B_{hj} c_{n+d-m} D_{hj} c_{n}.
\]

\[
+ O_p(T^{-1/2}) .
\]

(3.144)

Moreover, the square of the \(O_p(T^{-1/2})\) term of expression (3.144) has expectation that is \(O(T^{-1})\). Expressions (3.138), (3.139) and (3.144) imply that
and that the square of the $O_p(T^{-1/2})$ term in (3.145) has expectation that is $O(T^{-1})$. Since expression (3.135) has expectation equal to zero, it follows that

$$T^{-1/2} \delta' \{(I_r \otimes X')X \vec{v} - E[(I_r \otimes X')\vec{v}]\}$$

$$= T^{-1/2} \sum_{n=1}^{L} \sum_{i=1}^{r} \sum_{j=1}^{r} \sum_{h=1}^{M} \{ \sum_{d=-L}^{L} \sum_{m=0}^{M} \pi_i h(d) \mu_{n-d+m,j} D_{mh} \}$$

$$+ [E \sum_{d=-L}^{L} \sum_{m=0}^{M} \pi_i h(d) B_{lij} c_n' D_{d+n} D_{mh}]$$

$$+ O_p(T^{-1/2}), \quad (3.145)$$
\[ L^{m7d} L_{d=-B_{LM2}}^{m0} m_{8dm} \]

\[ + \left( \sum_{d=-B_{LM2}}^{m0} \sum_{\theta=m}^{8dm} \right) \]

\[ \pi_{1h}(d) B_{d+\theta}^{\gamma_{LM6}} \]

\[ \sum_{\theta=m}^{8dm} \sum_{d=-B_{LM2}}^{m0} \]

\[ E_{\gamma_{LM6}}^{T-1/2} \]

\[ = T^{-1/2} \sum_{t=1}^{T} (Z_{t} c_{t}' - \gamma_{LM6}) + O_{p}(T^{-1/2}) \]  \hspace{1cm} (3.146)

where

\[ Z_{t} = e_{t} + f_{t} + g_{t} \]

\[ e_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \delta_{ij} \left( \sum_{d=-L}^{m0} \sum_{\theta=m}^{3d} \pi_{1h}(d) B_{d+\theta}^{\gamma_{LM1}} \right) \]

\[ f_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \delta_{ij} \left( \sum_{d=-B_{LM1}}^{m=3d} \sum_{\theta=m}^{6d} \pi_{1h}(d) B_{d+\theta}^{\gamma_{LM2}} \right) \]

\[ g_{t} = \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \delta_{ij} \left( \sum_{d=-L}^{m=0} \sum_{\theta=m}^{6d} \pi_{1h}(d) B_{d+\theta}^{\gamma_{LM2}} \right) \]

\[ \gamma_{LM6} = E[f_{t} c_{t}'] \]
and the square of the expectation of the \( O(T^{-1/2}) \) term in (3.146) has expectation that is \( O(T^{-1}) \).

Now \( e_t \) is a fixed vector, \( f_t \) is a random vector that is measurable \( o(c_t) \), and \( g_t \) is a random vector that is measurable \( F_{t-1} \). Martingale conditions (3.2) and the definition of \( \gamma_{LM6} \) imply that \( E[e_t c'_t | F_{t-1}] = 0 \); \( E[f_t c'_t | F_{t-1}] - \gamma_{LM6} = E[f_t c'_t] - \gamma_{LM6} = 0 \); and \( E[g_t c'_t | F_{t-1}] = g_t E[c'_t | F_{t-1}] = 0 \). Hence, \( E[Z_t c'_t - \gamma_{LM6} | F_{t-1}] = 0 \).

Also, the following moment matrices are all finite and are constant over \( t \):

\[
C_{LMcfc} = \text{Cov}(c'_t, f_t c'_t); \\
V_{LMcf} = \text{Var}(f_t c'_t)
\]
\[ \text{and} \]

\[ V_{LMCG} = E(g_t \Sigma g'_t). \]

Also,

\[ d^2_{LMtt} \equiv E[(Z_t e'_t - \gamma_{LM})^2 | F_{t-1}] \]

\[ = \text{Var}[Z_t e'_t | F_{t-1}] \]

\[ = \text{Var}[(e_t + f_t + g_t) e'_t | F_{t-1}] \]

\[ = (e_t + g_t) \Sigma_{cc} (e_t + g_t)' + 2(e_t + g_t) \text{Cov}(e'_t, f'_t e'_t) + \text{Var}(f'_t e'_t) \]

\[ = (e_t + g_t) \Sigma_{cc} (e_t + g_t)' + 2(e_t + g_t) C_{LMtc} + V_{LMcf}; \]

so

\[ E[(Z_t e'_t - \gamma_{LM})^2] = E[E[(Z_t e'_t - \gamma_{LM})^2 | F_{t-1}]] \]

\[ = e_t \Sigma_{cc} e'_t + V_{LMcg} + 2e_t C_{LMtc} + V_{LMcf}. \]
Then $E[Z_t c'_t - \gamma_{LM\delta}|F_{t-1}] = 0$ implies that

$$E[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{LM\delta})|F_{t-1}] = \sum_{s=1}^{t-1} (Z_s c'_s - \gamma_{LM\delta})$$

and

$$\text{Var}[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{LM\delta})|F_{t-1}] = \text{Var}[(Z_t c'_t - \gamma_{LM\delta})|F_{t-1}],$$

so

$$s^2_{LMt} = \text{Var}[\sum_{s=1}^{t} (Z_s c'_s - \gamma_{LM\delta})] = \text{Var}[(Z_t c'_t - \gamma_{LM\delta})|F_{t-1}] = \sum_{s=1}^{t-1} \text{Var}[(Z_s c'_s - \gamma_{LM\delta})|F_{s-1}]$$

$$= \sum_{s=1}^{t} \left[ e_s c'_s + V_{LMcg} + 2e_s c'_{LMcf} + V_{LMcf} \right].$$

Conditions (3.2), (3.39.a), and (3.39.d) imply that

$$s^2_{LM} = \lim_{t \to \infty} \left[ t^{-1} s^2_{LMt} \right] > 0$$

exists and is finite. Assume first that $s^2_{LM} = 0$. Since $E(Z_s c'_s - \gamma_{LM\delta}) = 0$, it follows from the definition of $s^2_{LMt}$ that
\[
\lim_{t \to \infty} t^{-1} \mathbb{E} \left[ \left( \sum_{s=1}^{t} (Z_{s} c'_{s} - \gamma_{LM}) \right)^2 \right] = 0 ,
\]
so by the Chebyshev inequality,
\[
\text{plim}_{t \to \infty} t^{-1/2} \sum_{s=1}^{t} (Z_{s} c'_{s} - \gamma_{LM}) = 0 .
\]

Thus, in a trivial sense,
\[
T^{-1/2} \sum_{s=1}^{T} (Z_{s} c'_{s} - \gamma_{LM}) \xrightarrow{d} N(0, \bar{s}^2_{LM})
\]
when \( \bar{s}^2_{LM} = 0 . \)

Now assume that \( \bar{s}^2_{LM} > 0 . \) Given the above results, the sequence \( \{Z_{t} c'_{t} - \gamma_{LM}\} \) will satisfy the assumptions of Theorem 1 of Scott (1973) if the following conditions hold:
\[
s^{-2}_{LM} T \sum_{t=1}^{T} \mathbb{E}[(Z_{t} c'_{t} - \gamma_{LM})^2 | F_{t-1}] \xrightarrow{P} 1 \text{ as } T \to \infty \quad (3.153)
\]
and
\[
s^{-2}_{LM} T \sum_{t=1}^{T} \mathbb{E}[(Z_{t} c'_{t} - \gamma_{LM})^2 I(|Z_{t} c'_{t} - \gamma_{LM}| > \varepsilon s_{LM})) | F_{t-1}] \xrightarrow{P} 0 \text{ as } T \to \infty \quad (3.154)
\]

Results (3.149) and (3.151) imply that (3.153) will be established if
\[
T^{-1} \sum_{t=1}^{T} \left[ g_{t} \Delta c_{t} c'_{t} - V_{LMc} \right] \xrightarrow{P} 0 ; \quad (3.155)
\]
Now \( g_t \cdot \Sigma \Sigma g'_t \) is a sequence of random variables with mean \( \nu_{LMt} \) and common finite variance. Moreover, since the sequence \( \{c'_t, c_t\} \) is uncorrelated, the definition of \( g_t \) implies that for 
\[ |t-s| > 2L + M, \quad g_t \cdot \Sigma \Sigma g'_t \quad \text{and} \quad g_s \cdot \Sigma \Sigma g'_s \quad \text{are uncorrelated.} \]
Then an argument similar to the argument for (3.62) in the proof of Theorem 3.1 indicates that (3.155) may be written as the sum of no more than \( 2L + M \) separate terms, where each of these \( 2L + M \) sums has mean zero and is equal to a sum of uncorrelated random vectors. It follows that expression (3.155) converges to zero with probability one. Almost sure convergence in (3.156) and (3.157) is established similarly.

Lemma 3.6 implies that

\[
s_{LMt}^{-2} \sum_{t=1}^{T} E[(Z_t c'_t - \gamma_{LMS}^t)^2 I(|Z_t c'_t - \gamma_{LMS}^t| > \epsilon_{LMTT}^t)]
\]

\[
< s_{LMt}^{-2} \sum_{t=1}^{T} (\epsilon s_{LMt}^{-1})^{-\nu_{LMS}} E(|Z_t c'_t - \gamma_{LMS}^t|^{2+\nu})
\]

\[
= (s_{LMt}^{-2-\nu})(\epsilon^{-\nu}) \sum_{t=1}^{T} E(|Z_t c'_t - \gamma_{LMS}^t|^{2+\nu})
\]
By condition (3.46), the variables \(|(f_t + g_t)c_t' - \gamma_{LM5}|\) have a common finite \(2 + \nu\) moment, so

\[
\lim_{T \to \infty} T^{-1-2^{-1}\nu} \sum_{t=1}^{T} \mathbb{E}[(f_t + g_t)c_t' - \gamma_{LM5}]^{2+\nu} = 0.
\]

Also, the uniform finite bound on all \(|\mu_{tj}|\) and the absolute summability of \(\{A_n\}\) imply that \(e_{tj}\) is uniformly bounded, so for all \(s=1, 2, ..., r\), the existence of common finite \(2 + \nu\) moments of \(c_t\) implies that

\[
\lim_{T \to \infty} T^{-1-2^{-1}\nu} \sum_{t=1}^{T} \mathbb{E}[e_t c_t']^{2+\nu} = 0.
\]

Then the assumption that

\[
s_{LM}^2 \equiv \lim_{T \to \infty} s_{LM}^2 > 0 \quad (3.159)
\]

implies that (3.158) converges to zero as \(T \to \infty\), so condition (3.154) follows from the Markov inequality. Then by Theorem 1 of Scott (1973),

\[
s_{LM}^{-1} \sum_{t=1}^{T} (Z_t c_t' - \gamma_{LM5}) \xrightarrow{f} N(0, 1)
\]
under model (3.132). Then by condition (3.159),

$$T^{-1/2} \sum_{t=1}^{T} (Z_{t} e_{t} - \gamma_{LM} \delta) \xrightarrow{d} N(0, \overline{s}_{LM}^2)$$  \hspace{1cm} (3.160)

as \( T \to \infty \) under model (3.132). This result, and conclusion (3.152) for the case \( \overline{s}_{LM}^2 = 0 \), imply that conclusion (3.160) holds regardless of whether \( \overline{s}_{LM}^2 \) is positive or zero.

(B.2). Under constraint (3.134), one may now extend the asymptotic normality result of part (B.1) from the finite moving-average model (3.132) to the infinite moving average model (3.130). Note that for \( M > L \), expressions (3.135)-(3.140) imply that the difference between

$$(I_{L} \otimes X') \overline{\pi}_{L} \text{vec}(\nu) - T \text{vec}[\Gamma_{LM} \overline{u}^2 (0)]$$

under model (3.130) and

$$(I_{L} \otimes X') \overline{\pi}_{L} \text{vec}(\nu) - T \text{vec}[\Gamma_{LMu} \overline{u}^2 (0)]$$

under model (3.132) has \((i,j)\)-th element equal to
\[
\begin{align*}
&\sum_{d=-L}^{L} \sum_{h=1}^{r} \sum_{s=1}^{T} \tilde{p}_{th}(d) \left[ \mu_{sj} + \sum_{l=0}^{\infty} B_{kj}^{s} c'_{s-l} \right] \left[ \sum_{m=0}^{\infty} D_{m} c'_{s+d-m} \right] \\
&\quad - \left[ \mu_{sj} + \sum_{l=0}^{\infty} B_{kj}^{s} c'_{s-l} \right] \left[ \sum_{m=0}^{\infty} D_{m} c'_{s+d-m} \right] \\
&\quad - \left[ \sum_{l=\max(0,-d)}^{\min(M,M-d)} B_{kj}^{s} c'_{s-l} D'_{l+d,h} \right] - \left[ \sum_{l=\max(0,-d)}^{\min(M,M-d)} B_{kj}^{s} c'_{s-l} D'_{l+d,h} \right] \\
&\quad + O(1) \\
&= \sum_{d=-L}^{L} \sum_{h=1}^{r} \sum_{s=1}^{T} \tilde{p}_{th}(d) \left[ \mu_{sj} \sum_{m=M+1}^{\infty} D_{m} c'_{s+d-m} \right] \\
&\quad + \left( \sum_{l=M+1}^{\infty} \sum_{m=0}^{\infty} B_{kj}^{s} c'_{s-l} D_{m} c'_{s+d-m} \right) \\
&\quad + \left( \sum_{l=0}^{\infty} \sum_{m=M+1}^{\infty} B_{kj}^{s} c'_{s-l} D_{m} c'_{s+d-m} \right) \\
&\quad + \left( \sum_{l=M+1}^{\infty} \sum_{m=M+1}^{\infty} B_{kj}^{s} c'_{s-l} D_{m} c'_{s+d-m} \right) \\
&\quad - \left( \sum_{l=\min(M,M-d)+1}^{\infty} B_{kj}^{s} c'_{s-l} D'_{l+d,h} \right) + O(1) \\
&= \sum_{d=-L}^{L} \sum_{h=1}^{r} \sum_{s=1}^{T} \tilde{p}_{th}(d) \left[ W_{TLM1ijhd} + W_{TLM2ijhd} \right] \\
&\quad + W_{TLM3ijhd} + W_{TLM4ijhd} + O(1) \\
&= \sum_{s=1}^{T} \mu_{sj} \left( \sum_{m=M+1}^{\infty} D_{m} c'_{s+d-m} \right) \\
\end{align*}
\]

where

\[ W_{TLM1ijhd} = \sum_{s=1}^{T} \mu_{sj} \left( \sum_{m=M+1}^{\infty} D_{m} c'_{s+d-m} \right) \]}
and the square of the $O_p(1)$ term in (3.161) has expectation that is $O(1)$. Then by Lemma (3.7), the asymptotic distribution of

$$T^{-1/2} [((I_r \otimes X')_{1T} \vec{w}(v) - T \vec{w}(\Sigma_{L \otimes T}(0))]$$

under model (3.130) and condition (3.134) will be established if for all $i, j, h$ and $d$, and for all fixed $L$,

$$\lim_{T \to \infty} T^{-1/2} [W_{TLM1ijhd} + W_{TLM2ijhd} + W_{TLM3ijhd} + W_{TLM4ijhd}] = 0 \quad (3.162)$$

uniformly in $M$ and if the $N(0, \overline{s}_{LM}^2)$ distribution converges completely to the $N(0, \overline{s}_{LM}^2)$ distribution as $M \to \infty$, where

$$\overline{s}_{LM}^2 \equiv \lim_{M \to \infty} s_{LM}^2 \quad \text{(3.163)}$$

First, for any $1 \leq h, i \leq r$; $1 \leq j \leq k$; and $-L \leq d \leq L$. 
Because the $c_t$ vectors have mean zero and are uncorrelated, expression (3.164) has mean zero and has variance equal to

$$T^{-1} \sum_{s=1}^{T-M-1+d} T \sum_{t=-\infty}^{s=\max(1,M+1+t-d)} \sum_{t=d-M}^{s+1} \left[ \sum_{u=1}^{s+d-t} h_u c_{s+d-t} \right] c_t'$$

(3.165)

Note that the inequalities $s > 1$ and $-\infty < t < d-M$ imply that $M+1 < s+d-t < \infty$. Thus, expression (3.165) is bounded in modulus by

$$|T^{-1} \sum_{t=-\infty}^{s=\max(1,M+1+t-d)} \sum_{t=d-M+1}^{s+1} \left[ \sum_{u=1}^{s+d-t} h_u c_{s+d-t} \right] c_t'|$$

(3.167)
where the second inequality follows from the uniform bound $|\mu_{b,j}| < B_2$. By the absolute summability of $\{D_{l}\}$, expression (3.167) is finite and may be made arbitrarily small by choice of sufficiently large $M$. Also, expression (3.166) is bounded in modulus by

$$|T^{-1}\sum_{t=d-M+1}^{T-M-1+d} \sum_{s=M+1}^{T-s} \sum_{s=M+1}^{T-s} \sum_{a=1}^{p} \sum_{b=1}^{p} D_{sh} \sum_{c=1}^{c} \sum_{h_b}^{h_b} \mu_{s-d+t, j} \mu_{s-d+t, j} |$$

$$< B_2^2 \sum_{a=1}^{p} \sum_{b=1}^{p} \sum_{s=M+1}^{T-M-1+d} \sum_{a=1}^{p} \sum_{b=1}^{p} \sum_{s=M+1}^{T-M-1+d} |D_{sh} \sum_{c=1}^{c} \sum_{h_b}^{h_b} |.$$

(3.168)

Note that expression (3.168) is not a function of $T$. By the absolute summability of $\{D_{l}\}$, expression (3.168) may be made arbitrarily small by choice of sufficiently large $M$. Hence, uniformly in $T$, expression (3.166) converges to zero as $M \to \infty$. Thus, uniformly in $T$, $E[T^{-1}W_{TLMijhd}^2]$ converges to zero as $M \to \infty$. Then by the Markov inequality, uniformly in $T$, $T^{-1/2}W_{TLMijhd}$ converges in probability to zero as $M \to \infty$.

Next, note that $E[W_{TLM2ijhd} + W_{TLM3ijhd} + W_{TLM4ijhd}] = 0$. Now for $d > 0$,
Var\[T^{-1/2}W_{TLM2jhd} \]

\[= \text{Var}[T^{-1/2} \sum_{s=1}^{M} \sum_{t=1}^{t} \sum_{g=0}^{g} \sum_{m=0}^{m} B_{s-g} c_{s} D_{m} c_{s+d-m}] \]

\[< T^{-1} \sum_{s=1}^{M} \sum_{t=1}^{t} \sum_{g=0}^{g} \sum_{m=0}^{m} \sum_{n=0}^{n} \sum_{q=1}^{q} \sum_{w=1}^{w} \sum_{a=1}^{a} \sum_{b=1}^{b} \]

\[|B_{gjq} D_{mh} A_{njh} \cdot E(c_{s-g}, c_{s+d-m}, w, c_{s-d-n}, b)| \]

\[< T^{-1} \sum_{t=1}^{t} \sum_{s=1}^{M} \sum_{g=0}^{g} \sum_{m=0}^{m} \sum_{n=0}^{n} \sum_{q=1}^{q} \sum_{w=1}^{w} \sum_{a=1}^{a} \sum_{b=1}^{b} \]

\[|B_{gjq} D_{mh} A_{njh} \cdot E(c_{s-g}, c_{s+d-m}, w, c_{s-d-n}, b)| . \ (3.169) \]

An argument similar to the argument given for expression (3.73) indicates that there exists some finite positive real number \(b_4\) such that expression (3.169) is bounded above by

\[b_4 [\sum_{m=0}^{\infty} \sum_{w=1}^{r} |D_{mh} w|][\sum_{n=0}^{\infty} \sum_{b=1}^{r} |D_{nh} b|][\sum_{g=0}^{M+1-L} \sum_{q=1}^{r} B_{gjq}][\sum_{r=1}^{M+1-L} \sum_{a=1}^{r} B_{gja}] , \]

which is not a function of \(T\). The absolute summability of \(\{A_{g}\}\) implies that the second and third factors of (3.170) are finite and that the fourth and fifth factors of (3.170) can be made arbitrarily small by choice of sufficiently large \(M\). Thus, for \(d > 0\), uniformly in \(T\), \(\text{Var}[T^{-1/2}W_{TLM2jhd}]\) may be made arbitrarily small by choice of sufficiently large \(M\). A similar argument for \(d > 0\) indicates that
uniformly in $T$, $\text{Var}[T^{-1/2}W_{TLM2ijhd}]$ is bounded above by expression (3.170). Thus, for all $-L < d < L$, uniformly in $T$, $\text{Var}[T^{-1/2}W_{TLM2ijhd}]$ may be made arbitrarily small by choice of sufficiently large $M$. Similar arguments indicate that $\text{Var}[T^{-1/2}W_{TLM3ijhd}]$ and $\text{Var}[T^{-1/2}W_{TLM4ijhd}]$ are each bounded above by expression (3.170). Hence,

$$E(T^{-1}[W_{TLM2ijhd} + W_{TLM3ijhd} + W_{TLM4ijhd}]^2)$$

is bounded above by expression (3.170) multiplied by nine, and hence may be made arbitrarily small, uniformly in $T$, by choice of sufficiently large $M$. Then by the Markov inequality,

$$T^{-1/2}[W_{TLM2ijhd} + W_{TLM3ijhd} + W_{TLM4ijhd}]$$

converges in probability to zero, uniformly in $T$, as $M \to \infty$.

Thus, condition (3.162) is satisfied and it remains only to demonstrate that the $N(0, \bar{s}^2)$ distribution converges completely to the $N(0, \bar{s}^2)$ distribution as $M \to \infty$. An argument similar to the one given preceding expression (3.33) indicates that the complete convergence result follows immediately from the existence of

$$\bar{s}^2_L \equiv \lim_{M \to \infty} s^2_{LM}$$

$$= \lim_{M \to \infty} \lim_{T \to \infty} T^{-1}s^2_{LMTT}$$

$$= \lim_{M \to \infty} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left[ e_{tcc} e'_t + V_{LMcg} + 2e_{t} C_{LMcf} + V_{LMcf} \right]. \quad (3.171)$$
Part (C) below will discuss the limit (3.171) in greater detail.

(C). Finally, note that under model (3.130), if \( T > L + 1 \), then expression (3.135) indicates that

\[
T^{-1/2} \left\{ \left( I - \mathbf{x}' \right) \mathbf{v} - T \mathbf{v} \mu_{uv} \right\} \mathbf{v} = T^{-1/2} \left\{ \left( I - \mathbf{x}' \right) \mathbf{v} - \mathbf{v} \mu_{uv} \right\} \mathbf{v} - \mathbf{E} \left\{ \left( I - \mathbf{x}' \right) \mathbf{v} \mu_{uv} \right\} \mathbf{v}
\]

has \((i,j)\)-th element equal to

\[
T^{-1/2} \sum_{h=1}^{T-1} \sum_{d=L+1}^{T-1} \sum_{s=m_{1d}}^{m_{2d}} \left[ \mathbf{X}_{s_j} \mathbf{v}_{s+d,h} - \Gamma_{uvj} (d) \right] + \sum_{d=-T+1}^{L-1} \sum_{s=m_{1d}}^{m_{2d}} \left[ \mathbf{X}_{s_j} \mathbf{v}_{s+d,h} - \Gamma_{uvj} (d) \right].
\]

(3.172)

By definition of \( \Gamma_{uvj} (d) \), \( \mathbf{E} \left[ \mathbf{X}_{s_j} \mathbf{v}_{s+d,h} - \Gamma_{uvj} (d) \right] = 0 \), so expression (3.172) has expectation equal to zero. The uniform bound on \( |u_{t,j}| < B_2 \) and the linear-process assumptions on \( \{ \xi_t \} \) imply that there exists some \( K_1 > 0 \) such that for any \( d \in \mathcal{Z} \) and all \( T \in \mathcal{Z}^+ \),

\[
\text{Var} \left\{ T^{-1/2} \sum_{s=m_{1d}}^{m_{2d}} \left[ \mathbf{X}_{s_j} \mathbf{v}_{s+d,h} - \Gamma_{uvj} (d) \right] \right\} < K_1.
\]
Thus, the expectation of the square of expression (3.172) is bounded in modulus by

\[ r_{k_1} \left\{ \sum_{d=L+1}^{T-1} \left[ \left| \sum_{i=1}^{d} h_i (d) \right| + \left| \sum_{i=1}^{d} h_i (-d) \right| \right] \right\} < r_{k_1} \left\{ \sum_{d=L+1}^{\infty} \left[ \left| \sum_{i=1}^{d} h_i (d) \right| + \left| \sum_{i=1}^{d} h_i (-d) \right| \right] \right\} . \]

(3.173)

Expression (3.173) may be made arbitrarily small by choice of sufficiently large \( L \), so the expectation of the square of expression (3.172) converges to zero uniformly in \( T \) as \( L \to \infty \). Then by the Markov inequality, expression (3.172) converges in probability to zero uniformly in \( T \) as \( L \to \infty \).

Thus, by Lemma 3.12, the convergence result (3.131) will be established if the normal \((0, \bar{s}_L^2)\) distribution converges completely to the normal \((0, \bar{s}' \bar{G} \bar{d})\) distribution as \( L \to \infty \). As argument similar to the one given preceding expression (3.33) indicates that this complete convergence result will hold if

\[ \lim_{L \to \infty} \bar{s}_L^2 = \bar{s}' \bar{G} \bar{d} \quad (3.174) \]

In the arguments that follow, the uniform bound \( |u_j^*| < B_2 \), the absolute summability of \( \{A_n\} \), and Fubini's Theorem imply that \( \lim_{L \to \infty} \bar{s}_L^2 \) exists and is finite, and that limits with respect to \( T \), \( L \) and \( M \) may be exchanged. Now by expression (3.171),

\[ \lim_{L \to \infty} \bar{s}_L^2 = \lim_{L \to \infty} \left[ \lim_{M \to \infty} \left( \lim_{T \to \infty} \sum_{t=1}^{T} \left( e_t e_t^* + 2 e_t^* c_t L M c + V_{LMc} \right) \right) \right] . \]

(3.171)
Note that by the definition of $e_t$,

\[
\lim \left( \lim \left( \lim \left( T^{-1} \sum_{t=1}^T e_t e_t' \right) \right) \right)
\]

\[
= \lim \left( \lim \left( \lim \left( T^{-1} \sum_{t=1}^T \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N \sum_{r=1}^L \sum_{s=1}^M \sum_{t=1}^T \sum_{d=1}^L \sum_{m=1}^M \sum_{a=1}^A \sum_{b=1}^B \sum_{h=1}^H \sum_{w=1}^W \delta_{ij} \delta_{ab} \right) \right) \right)
\]

\[
= \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N \sum_{r=1}^L \sum_{s=1}^M \sum_{t=1}^T \sum_{d=1}^L \sum_{m=1}^M \sum_{a=1}^A \sum_{b=1}^B \sum_{h=1}^H \sum_{w=1}^W \delta_{ij} \delta_{ab}
\]

\[
\times \left( \lim \left( \lim \left( \lim \left( \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N \sum_{r=1}^L \sum_{s=1}^M \sum_{t=1}^T \sum_{d=1}^L \sum_{m=1}^M \sum_{a=1}^A \sum_{b=1}^B \sum_{h=1}^H \sum_{w=1}^W \delta_{ij} \delta_{ab} \right) \right) \right) \right)
\]

\[
= \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N \sum_{r=1}^L \sum_{s=1}^M \sum_{t=1}^T \sum_{d=1}^L \sum_{m=1}^M \sum_{a=1}^A \sum_{b=1}^B \sum_{h=1}^H \sum_{w=1}^W \delta_{ij} \delta_{ab}
\]

\[
\times \left( \lim \left( \lim \left( \lim \left( \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N \sum_{r=1}^L \sum_{s=1}^M \sum_{t=1}^T \sum_{d=1}^L \sum_{m=1}^M \sum_{a=1}^A \sum_{b=1}^B \sum_{h=1}^H \sum_{w=1}^W \delta_{ij} \delta_{ab} \right) \right) \right) \right)
\]

Similar arguments indicate that

\[
(3.175)
\]
\[
\lim_{L \to \infty} \lim_{M \to \infty} \lim_{T \to \infty} \sum_{\epsilon \in C_L M_c f_c} \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \sum_{a=1}^{r} \sum_{b=1}^{k} \sum_{w=1}^{r} \sum_{d=1}^{m} \sum_{m=-\infty}^{\infty} [\pi_{i h}(d) \pi_{a w}(-m) \Gamma_{v v h w}(m-d) \Gamma_{u v j b}(\ell)] \quad (3.176)
\]

\[
\lim_{L \to \infty} \lim_{M \to \infty} \lim_{T \to \infty} \sum_{\epsilon \in C_L M_c f_c} \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \sum_{a=1}^{r} \sum_{b=1}^{k} \sum_{w=1}^{r} \sum_{d=1}^{m} \sum_{m=-\infty}^{\infty} \sum_{t=1}^{T} \sum_{\epsilon \in C_L M_c f_c} \pi_{i h}(d) \pi_{a w}(-m) [\Gamma_{v v h w}(0, -d, m-d) \bar{\mu}] \]

\[
+ \bar{\mu}_j M_{v v h b w}(0, -d, m-d) \quad (3.177)
\]

and

\[
\lim_{L \to \infty} \lim_{M \to \infty} \lim_{T \to \infty} \sum_{\epsilon \in C_L M_c f_c} \sum_{i=1}^{r} \sum_{j=1}^{k} \sum_{h=1}^{r} \sum_{a=1}^{r} \sum_{b=1}^{k} \sum_{w=1}^{r} \sum_{d=1}^{m} \sum_{m=-\infty}^{\infty} [\pi_{i h}(d) \pi_{a w}(-m) [\Gamma_{u v j w}(m) \Gamma_{u v b h}(d-\ell)] + k_{u v u v j b h w}(0, d, \ell, m)] . \quad (3.178)
\]

Inspection of expression (3.101) indicates that expressions (3.175), (3.176), (3.177) and (3.178) sum to $\delta[G_\pi] \hat{\delta}$, so by Lemma 3.12, result (3.131) is established.

Theorem 3.2 established the asymptotic normality of
under the conditions of Lemma 3.16 and under a linear-process assumption for the sequence \( \{e_t\} \). The following corollary extends this result to bilinear functions with the less restrictive weight matrices \( \pi \) and \( \pi(T) \).

**Corollary 3.2.1.** Let \( \{u_t\} \), \( \{e_t\} \), \( \{\pi_{ij}(s, t)\} \) and \( \{\pi_{ij}(d)\} \) satisfy the conditions of Lemma 3.16, and Theorem 3.2, and define \( \pi \) and \( \Gamma_{uv}(0) \) as in Theorem 3.2.

a. Then as \( T \to \infty \),

\[
T^{-1/2} \left\{ (I_T \otimes X')\pi \vec{v} - T \vec{\Gamma}_{uv}(0) \right\} \xrightarrow{d} N_{rk}(0, \Sigma_{uv}).
\]

b. Assume moreover that \( \{\pi_{ij}(t, t)\} \) and \( \{\pi_{ij}(s, t)\} \) satisfy conditions (3.86). Then as \( T \to \infty \),

\[
T^{-1/2} \left\{ (I_T \otimes X')\pi(T) \vec{v} - T \vec{\Gamma}_{uv(T)}(0) \right\} \xrightarrow{d} N_{rk}(0, \Sigma_{uv}).
\]

**Proof.** Part (a) follows immediately from Theorem 3.2 and Lemma 3.17.a. Part (b) follows immediately from Theorem 3.2 and Lemma 3.17.b.

This chapter has developed some asymptotic properties of the sum of a linear function and a bilinear function of a realization of a multivariate linear process. The following chapter will use these properties
to establish asymptotic results for unweighted and weighted errors-in-variables estimators in the presence of serially correlated observations.
4. ESTIMATION BY THE METHOD OF MOMENTS

This chapter discusses the asymptotic behavior of a number of estimators of the regression coefficient \( \beta \) in the correlated errors-in-variables model (2.3)-(2.4). For each of the estimators under consideration, some standard Taylor expansion arguments indicate that the large-sample properties of interest depend only on the large-sample properties of the first and second sample moments of the components of model (2.3)-(2.4). It is thus found that if the random components \( (x_t - \mu_{xt}, \varepsilon_t) \) are assumed to follow a linear process model that satisfies Definition 3.2, and the fixed vectors \( \mu_{xt} = E(x_t) \) are assumed to satisfy the Grenander-type condition (3.39.a), then the asymptotic properties of the estimators follow immediately from the results of Chapter 3.

Under these assumptions, Section 4.1 gives the large-sample properties of three estimators which are functions of the unweighted first and second sample moments of the observations \( \{Z_t, t=1, 2, ..., T \} \). These estimators were originally derived in the literature under the assumption that the observations \( Z_t \) were mutually uncorrelated, but the results of Section 4.1 indicate that the same estimators retain some desirable asymptotic properties under a correlated form of model (2.3)-(2.4).

Section 4.2 addresses a class of weighted estimators of \( \beta \). These weighted estimators may be motivated by considering similar weighted estimators for the heteroscedastic model reviewed in Section 2.5.
above. Theorem 4.3 gives conditions for the consistency and asymptotic normality of a weighted estimator of $\hat{\beta}$ under the correlated errors-in-variables model (2.3)-(2.4).

4.1. Unweighted Estimators

This section addresses the asymptotic behavior of some unweighted errors-in-variables estimators. A correlated ultrastructural form of the model (2.3)-(2.4) is assumed. Theorem 4.1 gives conditions under which a multivariate version of the estimator (2.11) of $\hat{\beta}$ for known $\Gamma_{uu}(0)$ and known $\Gamma_{ue}(0)$ is strongly consistent and asymptotically normal, and gives an explicit expression for the covariance matrix of the associated asymptotic distribution. Corollary 4.1.1 extends the results of Theorem 4.1 to the case in which the lag-zero covariance matrices $\Gamma_{uu}(0)$ and $\Gamma_{ue}(0)$ are estimated. Next, maximum likelihood estimation of $\hat{\beta}$ for the uncorrelated normal structural model (2.3)-(2.4) is reviewed. Theorem 4.2 establishes the strong consistency and asymptotic normality of this estimator under the conditions of Theorem 4.1. In sum, Theorem 4.1, Corollary 4.1.1, and Theorem 4.2 indicate that a number of unweighted estimators derived originally for uncorrelated errors-in-variables models retain some desirable properties in the presence of serial correlation. However, the covariance matrices of the asymptotic distributions of the unweighted estimators reflect the autocovariance structure of the errors $\varepsilon$, so reported standard errors and related statistics must be adjusted accordingly.
The results presented below rely on a common set of conditions on the fixed and random terms of the ultrastructural errors-in-variables model (2.3)-(2.4) with no error in the equation, i.e.,

\[ Y_t = X_t \beta + e_t, \tag{4.1.a} \]
\[ X_t = X_t + u_t. \tag{4.1.b} \]

where all definitions and notation remain the same as in Section 2.1 unless noted otherwise. Although the first theorem addresses an estimator generally associated with model (2.1)-(2.2) with an error in the equation, no separate analysis of the "error in equation" terms \( q_t \) will be given. Hence, it will suffice to state error conditions only for the "total error" term \( \varepsilon_t = a_t + (q_t, 0, \ldots, 0) \), and results will be stated only for the model with no error in the equation.

Let \( u_{xt} = E(x_t) \) and let \( x_{ct} = x_t - u_{xt} \). It will be assumed that \( \{(x_{ct}, \varepsilon_t), t=1, 2, \ldots, T\} \) is a sequence of \([1 \times (k+p)]\) - dimensional random vectors which follow the linear process model,

\[ (x_{ct}', \varepsilon_t) = \sum_{h=0}^{\infty} A_h c_{t-h} \tag{4.2} \]

where the sequence of \((k+p) \times (k+p)\) real matrices \( \{A_h\} \) is absolutely summable and the sequence of \(1 \times (k+p)\) random vectors \( \{c_t\} \) satisfies the following conditions.
(i) The random vectors \( \mathbf{c}_t \) are defined on a common probability space \((\Omega, \mathcal{F}, P)\).

(ii) The following conditional expectations exist and are equal to finite constant matrices, as indicated, for all integers \( t \):

\[
\mathbb{E}[c_t | \mathcal{F}_{t-1}] = \mathbf{0}_{1 \times (k+p)} \quad (4.3.a)
\]

\[
\mathbb{E}[c_t' c_t | \mathcal{F}_{t-1}] = \Sigma_{cc} \quad (4.3.b)
\]

\[
\mathbb{E}[c_t' \mathbf{x}_t a_t c_t | \mathcal{F}_{t-1}] = \mathbf{M}_3 \quad (4.3.c)
\]

\[
\mathbb{E}[c_t' \mathbf{x}_t a_t c_t' \mathbf{c}_t | \mathcal{F}_{t-1}] = \mathbf{K} \quad (4.3.d)
\]

where \( \mathcal{F}_t = \sigma(\mathbf{c}_s, s < t) \), the sub-\(\sigma\)-algebra of \( \mathcal{F} \) generated by the random vectors \( \{\mathbf{c}_s, s < t\} \).

(iii) There exists some \( \nu > 0 \) such that

\[
\sup_{t} \max_{1 \leq i \leq k+p} \{ \mathbb{E}[|c_{ti}|^{4+\nu}] \} \equiv B_\nu < \infty \quad (4.4)
\]

Note that conditions (4.2)-(4.3) imply that the \((\mathbf{x}_c, \mathbf{c})\) process is fourth-order stationary. Hence, one may define the covariance function and the third- and fourth-order cumulant functions of the \((\mathbf{x}_c, \mathbf{c})\) process as indicated in Section 3.1.
In addition to the sequence \( \{(x_t, e_t)\} \), the asymptotic results presented below require the definition of a sequence of \( r \)-dimensional random vectors

\[
v_t = Y_t - X_t \beta = e_t - u_t \beta,
\]

t = 1, 2, \ldots, T. Define the covariance and cumulant functions of the \( v_t \) sequence in accordance with the notation developed in Section 3.1, e.g.,

\[
\Gamma_{uv}(t) = \text{Cov}(u_t, v_{s+t}) = \Gamma_{ue}(t) - \Gamma_{uu}(t) \beta \quad \text{and}
\]

\[
\Gamma_{vv}(t) = \text{Cov}(v_t, v_{s+t}) = \Gamma_{ee}(t) - \beta' \Gamma_{ue}(t) - \Gamma_{eu}(t) \beta + \beta' \Gamma_{uu}(t) \beta.
\]

Finally, the ultrastructural results presented below require conditions on the sequence of mean vectors \( \{\mu_t\} \). A simplified form of Grenander's conditions may be imposed by assuming that

\[
\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mu_{x,t+t_i, x,t+t_j} = \mu_{\mu ij}(\ell) \quad (4.5)
\]

exists and is finite for all \( i, j = 1, 2, \ldots, k \) and all integers \( \ell \). Define \( \tilde{M}_{\mu \mu}(\ell) \) to be the \( k \times k \) matrix with \((i,j)\)-th element equal to \( \tilde{M}_{\mu \mu ij}(\ell) \). Recall from expression (3.39.d) that condition (4.5) implies that
exists and is finite.

Given the conditions stated above, one may obtain two useful results regarding the second sample moments of the observations \( \{Z_t\} \) and the random components \( \{\varepsilon_t\} \). Define

\[
\mathbf{M}_{zz} = \left( \begin{array}{cc} \mathbf{M}_{YY} & \mathbf{M}_{YX} \\ \mathbf{M}_{XY} & \mathbf{M}_{XX} \end{array} \right) = T^{-1} \sum_{t=1}^{T} Z_t' Z_t,
\]

\[
\mathbf{M}_{z\varepsilon} = T^{-1} \sum_{t=1}^{T} Z_t' \varepsilon_t,
\]

\[
\mathbf{M}_{\varepsilon\varepsilon} = T^{-1} \sum_{t=1}^{T} \varepsilon_t' \varepsilon_t,
\]

and

\[
\bar{\mathbf{M}}_{zz}(0) = (\mathbf{\beta}, \mathbf{I}_k)' [\Gamma_{xx}(0) + \bar{\mathbf{M}}_{\mu\mu}(0)](\mathbf{\beta}, \mathbf{I}_k).
\]

Result 4.1 establishes two almost sure convergence results and two order in probability results that will be useful in subsequent proofs.

**Result 4.1.** Let model (4.1) hold and assume that conditions (4.2)-(4.3) and (4.5) are satisfied.

a. Then with probability one,

\[
\lim_{{T \to \infty}} \mathbf{M}_{zz} = \bar{\mathbf{M}}_{zz}(0) + \Gamma_{\varepsilon\varepsilon}(0),
\]

and
b. Moreover,

\[ \lim_{T \to \infty} M_{e e}^{Z e} = \Gamma_e (0) . \]

Proof. Result 4.1.a follows immediately from Lemma 3.11. Result 4.1.b follows immediately from expression (3.42), the observation that

\[ E(M_{e e}) = \Gamma_e (0) = E(M_{Z e}) , \]

and the Chebyshev inequality. □

Recall that expression (2.11) provided an estimator of \( \mu \) for the model (2.1)-(2.2) with an error in the equation and with a single "dependent variable" \( Y_t \) at time \( t \). A direct extension of the estimator (2.11) to the present case of multivariate \( Y_t \) and serially correlated errors may be written,

\[ \text{vec}(\hat{\mu}) = \text{vec}(M_{X X} - \hat{\Gamma}_{u u}(0))^{-1}[M_{X Y} - \hat{\Gamma}_{u e}(0)] \]  \hspace{1cm} (4.6)

where \( \hat{\Gamma}_{u u}(0) \) and \( \hat{\Gamma}_{u e}(0) \) are estimators of \( \Gamma_{u u}(0) \) and \( \Gamma_{u e}(0) \), respectively. In order to simplify the discussion, the estimator (4.6) will be discussed in two steps. First, Theorem 4.1 will establish the simpler asymptotic properties of the estimator.
Second, Corollary 4.1.1 will give conditions under which the estimator 
(4.6) has asymptotic properties similar to the asymptotic properties of 
the estimator (4.7).

Theorem 4.1. Let model (4.1) hold. Let the random vectors $(x_n, e_n)$ 
satisfy conditions (4.2)-(4.4) and let the fixed vectors $Y_{xt}$ satisfy 
conditions (4.5). Assume that $M_{xx}(0) = \Gamma_{xx}(0) + M_{mu}(0)$ is positive 
definite; and that the cross-covariance function between $x_c$ and $e$ 
and the third- and fourth-order cross-cumulant functions between $x_c$ 
and $e$ are identically zero. Define $\beta$ by expression (4.7).

a. Then as $T \to \infty$, $\hat{\beta}$ converges to $\beta$ with probability one.

b. Moreover, as $T \to \infty$,

$$T^{1/2} \frac{1}{\sqrt{n}} \text{vec}[\hat{\beta} - \beta] \xrightarrow{D} N_{rk}\left\{ \Gamma_{xx}(0)^{-1} G \Gamma_{xx}(0)^{-1} \right\}$$

where

$$G = \sum_{t=-\infty}^{\infty} \{ \Gamma_{vv}(t) \ast [M_{mu}(t) + \Gamma_{xx}(t) + \Gamma_{uu}(t)] + \text{vec}[\Gamma_{uv}(t)] \text{vec}[\Gamma_{uv}(-t)]' + M_{vv}(0, 0, t) \ast [M_{vv}(0, 0, t)]' + k_{vvuv}(0, 0, t, t) \}.$$
Proof.

a. By Result 4.1.a, \( M_{XX} \sim \Gamma_{uu}(0) \) converges with probability one to \( \bar{M}_{xx}(0) \) as \( T \to \infty \). Since \( \bar{M}_{xx}(0) \) is positive definite, one need only consider the case in which \( M_{XX} \sim \Gamma_{uu}(0) \) is positive definite. Now

\[
\hat{\beta} - \beta = [M_{XX} - \Gamma_{uu}(0)]^{-1}[M_{XY} - \Gamma_{ue}(0)] - \beta
\]

\[
= [M_{XX} - \Gamma_{uu}(0)]^{-1}[M_{XY} - M_{XX}\hat{\beta} - \Gamma_{ue}(0) + \Gamma_{uu}(0)]
\]

\[
= [M_{XX} - \Gamma_{uu}(0)]^{-1}[M_{Xv} - \Gamma_{uv}(0)].
\]

Result (a) then follows from the almost sure convergence of

\( M_{XX} \sim \Gamma_{uu}(0) \) to \( \bar{M}_{xx}(0) \), the almost sure convergence of \( M_{Xv} \sim \Gamma_{uv}(0) \) to zero, and the continuity of \( \hat{\beta} - \beta \) in the elements of \( M_{XX} \sim \Gamma_{uu}(0) \) and of \( M_{Xv} \sim \Gamma_{uv}(0) \).

b. By Result 4.1.a,

\[
M_{XX} \sim \Gamma_{uu}(0) = \bar{M}_{xx}(0) + \bar{M}_{\mu\mu}(0) + o_p(1) = \bar{M}_{xx}(0) + o_p(1). \quad (4.9)
\]

Since \( \bar{M}_{xx}(0) \) is positive definite, it follows that the probability that \( M_{XX} \sim \Gamma_{uu}(0) \) is positive definite is increasing to one as \( T \to \infty \). Hence, one need only consider the case in which \( M_{XX} \sim \Gamma_{uu}(0) \) is positive definite. A first-order expansion of a matrix inverse indicates that
Moreover, Result 4.1.b and the relations

\[ M_{Xv} = T^{-1} \Sigma \prod_{t=1}^{T} X_t' v_t \]

\[ = M_{Xe}(I, -\beta')' , \text{ and} \]

\[ \Gamma_{uv}(0) = \Gamma_{uv}(0)(I, -\beta')' \]

imply that

\[ M_{Xv} - \Gamma_{uv}(0) = O_p(T^{-1/2}) . \]  (4.11)

Then by (4.10) and the arguments of part (a),

\[ \text{vec}(\hat{\beta} - \beta) = \text{vec}([M_{xx}(0)]^{-1} + o_p(1))[M_{Xv} - \Gamma_{uv}(0)] \]

\[ = (I_r \otimes [M_{xx}(0)]^{-1}) \text{vec}[M_{Xv} - \Gamma_{uv}(0)] + o_p(T^{-1/2}) . \]  (4.12)

By Theorem 3.1 and the assumption that the cross-covariance function of \( x_c \) and \( z \) and the third- and fourth-order cross-cumulant functions between \( x_c \) and \( z \) are identically zero,

\[ T^{1/2} \text{vec}[M_{Xv} - \Gamma_{uv}(0)] \overset{d}{\rightarrow} N_{rk}(\Theta, G) \text{ as } T \rightarrow \infty . \]
Thus, as \( T \to \infty \),

\[
T^{1/2} \{ I_r \oplus \begin{bmatrix} \mathbf{M}_{xx}(0) \end{bmatrix}^{-1} \} \text{vec}[\mathbf{M}_{xy} - \Gamma_{uv}(0)] \xrightarrow{\text{d}} N_{rk}(0, \{ I_r \oplus \begin{bmatrix} \mathbf{M}_{xx}(0) \end{bmatrix}^{-1} \} \mathbf{G} \{ I_r \oplus \begin{bmatrix} \mathbf{M}_{xx}(0) \end{bmatrix}^{-1} \})
\]

and the result follows from expression (4.12). \( \square \)

The covariance matrices \( \Gamma_{uu}(0) \) and \( \Gamma_{ue}(0) \) are generally not known a priori, but they may be estimated from replicated observations or other forms of auxiliary data. Corollary 4.1.1 assesses the asymptotic distribution of estimators of the form (4.6) which employ estimators of the indicated lag-zero covariances.

**Corollary 4.1.** Let the assumptions of Theorem 4.1 hold. Let \( \{ \hat{\Gamma}_{uu}(T)(0) \} \) be a sequence of \( k \times r \) random matrices and let \( \{ \hat{\Gamma}_{ue}(T)(0) \} \) be a sequence of \( k \times k \) random matrices such that as \( T \to \infty \),

\[
T^{1/2} \{ \text{vec}[\hat{\Gamma}_{ue}(T)(0) - \Gamma_{ue}(0)]', \text{vec}[\hat{\Gamma}_{uu}(T)(0) - \Gamma_{uu}(0)]' \}' \xrightarrow{\text{d}} N_{kr+k^2}(0, \varphi) . \tag{4.13}
\]

Assume that for all \( T \), \( \hat{\Gamma}_{uu}(T)(0) \) and \( \hat{\Gamma}_{ue}(T)(0) \) are independent of \( Z \). Define
\[
\hat{\beta}_E = [M_{XX} - \hat{\Gamma}_{uu}(T)(0)]^{-1}[M_{XY} - \hat{\Gamma}_{ue}(T)(0)].
\]

a. Then as \( T \to \infty \),
\[
T^{1/2}\text{vec}(\hat{\beta}_E - \beta) \xrightarrow{p} N_{r_k}(0, \{I_r \oplus \bar{M}_{xx}(0)\}^{-1}\bar{G}_E\{I_r \oplus \bar{M}_{xx}(0)\}^{-1})
\]

where
\[
G_E = G + [(I_r, -\beta') \oplus I_k] \psi[(I_r, -\beta')' \oplus I_k]
\]

and \( G \) is defined by expression (4.8).

b. Moreover, if with probability one,
\[
\lim_{T \to \infty} \{[\hat{\Gamma}_{ue}(T)(0) - \Gamma_{ue}(0)], [\hat{\Gamma}_{uu}(0) - \Gamma_{uu}(0)]\} = 0 \tag{4.14}
\]

then with probability one,
\[
\lim_{T \to \infty} (\hat{\beta}_E - \beta) = 0.
\]

**Proof.**

a. By the hypotheses of this corollary,
\[
\hat{\Gamma}_{uu}(T)(0) - \Gamma_{uu}(0) = O_p(T^{-1/2}), \text{ and }
\]
Then by expression (4.9),

\[ \bar{M}_{\text{xx}} - \hat{\bar{\rho}}_{\text{uu}}(T)(0) = [\bar{M}_{\text{xx}} - \bar{\rho}_{\text{uu}}(0)] - [\hat{\bar{\rho}}_{\text{uu}}(T)(0) - \bar{\rho}_{\text{uu}}(0)] \]

\[ = \bar{M}_{\text{xx}}(0) + o_p(1). \]

A repetition of the arguments preceding expression (4.10) indicates that one need only consider the case in which \( \bar{M}_{\text{xx}} - \hat{\bar{\rho}}_{\text{uu}}(T)(0) \) is positive definite, and that

\[ [\bar{M}_{\text{xx}} - \hat{\bar{\rho}}_{\text{uu}}(T)(0)]^{-1} = [\bar{M}_{\text{xx}}(0)]^{-1} + o_p(1). \]

Then a repetition of the arguments in expression (4.12) indicates that

\[ \text{vec}(\beta_e - \beta) = \text{vec}\{[\bar{M}_{\text{xx}} - \hat{\bar{\rho}}_{\text{uu}}(T)(0)]^{-1}[\bar{M}_{\text{xy}} - \hat{\bar{\rho}}_{\text{ue}}(T)(0)] - \beta\} \]

\[ = \text{vec}\{[\bar{M}_{\text{xx}} - \hat{\bar{\rho}}_{\text{uu}}(T)(0)]^{-1}[[\bar{M}_{\text{xy}} - \bar{M}_{\text{xx}}\beta] \]

\[ - [\hat{\bar{\rho}}_{\text{ue}}(T)(0) - \hat{\bar{\rho}}_{\text{uu}}(T)(0)\beta]\}\}

\[ = \text{vec}\{[\bar{M}_{\text{xx}}(0)]^{-1} + o_p(1)}[[\bar{M}_{\text{xy}} - \bar{\rho}_{\text{uv}}(0)] \]

\[ - [\hat{\bar{\rho}}_{\text{ue}}(T)(0) - \hat{\bar{\rho}}_{\text{ue}}(0)] - [\hat{\bar{\rho}}_{\text{uu}}(T)(0) - \bar{\rho}_{\text{uu}}(0)\beta]\} \]
\[ = \{I_r \circ \left[ \begin{bmatrix} \mathbf{M}_{xx}(0) \end{bmatrix} \right]^{-1}\} \left\{ \text{vec} \left[ \mathbf{M}_{xv} - \mathbf{\Gamma}_{uv}(0) \right] - \text{vec} \left\{ \mathbf{\Gamma}_{ue}(T)(0) - \mathbf{\Gamma}_{ue}(0) - \left[ \begin{bmatrix} \mathbf{\Gamma}_{uu}(T)(0) - \mathbf{\Gamma}_{uu}(0) \end{bmatrix} \right] \right\} + \mathbf{0}_p(T^{-1/2}) \right. \]

Conclusion (a) then follows from Theorem 4.1, condition (4.13) and the independence of \( \left\{ \mathbf{\Gamma}_{uu}(T)(0), \mathbf{\Gamma}_{ue}(T)(0) \right\} \) from \( \{Z_t\} \).

b. Conclusion (b) follows immediately from the proof of Theorem 4.1.a and assumption (4.14).

If the entire covariance matrix \( \mathbf{\Gamma}_{\varepsilon\varepsilon}(0) \) is known or estimated consistently, then one may derive estimators of \( \mathbf{\beta} \) that may have greater asymptotic efficiency than the estimators (4.6) or (4.7) under the uncorrelated model (4.1). In particular, for the normal structural model (4.1), Fuller (1987, pp. 300-305) presented the following development of the maximum likelihood estimator of \( \mathbf{\beta} \).

Assume that \( \mathbf{\Gamma}_{uu}(0) \) and \( \mathbf{\Gamma}_{xx}(0) \) are nonsingular. Then it follows from the discussion in Fuller (1987, p. 300) that under the structural model (4.1), the matrix \( \mathbf{\Gamma}_{ZZ}(0) \) is also nonsingular. Let \( \mathbf{S}_{\varepsilon\varepsilon} \) be an estimator of \( \mathbf{\Gamma}_{\varepsilon\varepsilon}(0) \) such that \( d_\varepsilon \mathbf{S}_{\varepsilon\varepsilon} \) has Wishart distribution with \( d_\varepsilon \) degrees of freedom and parameter matrix \( \mathbf{\Gamma}_{\varepsilon\varepsilon}(0) \). Assume also that \( \mathbf{S}_{\varepsilon\varepsilon} \) is independent of \( \mathbf{Z} \). By Result 4.1, the probability that \( \mathbf{M}_{ZZ} \) is positive definite converges to one as \( T \rightarrow \infty \). Thus, without loss of generality, one may define \( \hat{\lambda}_1 < \hat{\lambda}_2 < \ldots < \hat{\lambda}_p \) to be the eigenvalues...
of the matrix $S_{ee}$ in the metric $M_{zz}$, i.e., the roots of the determinantal equation $|S_{ee} - \lambda^{-1}M_{zz}| = 0$; and one may define $R = (R_1, R_2, \ldots, R_p)$ to be the $p \times p$ matrix of eigenvectors of $S_{ee}$ in the metric $M_{zz}$. [For some background on eigenvalues and eigenvectors of one matrix in the metric determined by a second, positive definite, matrix, see Appendix 4.A of Fuller (1987).] Define the $p \times k$ matrix

$$\hat{A} = \begin{bmatrix} \hat{A}'_{rk} & \hat{A}'_{kk} \end{bmatrix}$$

$$= M_{zz}[(1 - \hat{\lambda}_{1}^{-1})^{1/2}R_1, (1 - \hat{\lambda}_{2}^{-1})^{1/2}R_2, \ldots, (1 - \hat{\lambda}_{k}^{-1})^{1/2}R_k]$$

where $\hat{A}_{rk}$ is $r \times k$ and $\hat{A}_{kk}$ is $k \times k$. Then expression (4.1.29) of Fuller (1987, p. 299) indicates that the maximum likelihood estimator of $\beta$ is

$$\hat{\beta} = (\hat{A}_{rk} \hat{A}_{kk})^{-1} \hat{A}_{rk}$$

(4.15)

For observations which follow a correlated form of model (4.1) with no error in the equation, estimators of the matrix $\Gamma_{ee}(0)$ may be available from one of two sources. First, a replicated-observation study involving a large number of cross-sectional units in a single period may lead to an estimator with the Wishart distribution attributed to the matrix $d_f S_{ee}$ discussed above. On the other hand, if an estimator $\hat{\Gamma}_{ee}(0)$ is computed from observations on overlapping sets of
units at several time periods, it may not be reasonable to assume that \( \widehat{\Sigma}_{\text{ee}}(0) \) is distributed as a multiple of a Wishart matrix, but it may be reasonable to assume for large samples, \( T^{1/2} \text{vec}[\widehat{\Sigma}_{\text{ee}}(0) - \Sigma_{\text{ee}}(0)] \) is approximately distributed as a normal random vector. Therefore, one may consider the following generalization of the estimator (4.15). Let \( \{\widehat{\Sigma}_{\text{ee}(T)}(0)\} \) represent a sequence of \( p \times p \) symmetric random matrices such that

\[
T^{1/2} \text{vec}[\widehat{\Sigma}_{\text{ee}(T)}(0) - \Sigma_{\text{ee}}(0)] \xrightarrow{\mathbb{P}} N_p(0, 0) \quad (4.16)
\]

as \( T \to \infty \). Let \( \lambda_1(T) < \lambda_2(T) < \ldots < \lambda_p(T) \) be the eigenvalues of \( \widehat{\Sigma}_{\text{ee}(T)}(0) \) in the metric \( M_{zz} \) and let

\[
R(T) = (R_{.1}(T), R_{.2}(T), \ldots, R_{.p}(T)) \quad (4.17)
\]

be the \( p \times p \) matrix of eigenvalues of \( \widehat{\Sigma}_{\text{ee}(T)}(0) \) in the metric \( M_{zz} \). Define the \( p \times k \) matrix

\[
\hat{A}_T = [\hat{A}_{rk}(T), \hat{A}_{kk}(T)]'
\]

\[
= M_{zz} [(1 - \lambda_1(T)^{-1})^{1/2} \mathbf{R}_{.1}(T), (1 - \lambda_2(T)^{-1})^{1/2} \mathbf{R}_{.2}(T), \ldots, (1 - \lambda_k(T)^{-1})^{1/2} \mathbf{R}_{.k}(T)]
\]

(4.18)

where \( \hat{A}_{rk}(T) \) is \( r \times k \) and \( \hat{A}_{kk}(T) \) is \( k \times k \). Then a
generalization of expression (4.15) may be written,

\[ \hat{\theta}_{(T)} = (\hat{A}_{rk(T)} \hat{A}^{-1}_{kk(T)})' \cdot \]  

(4.19)

Theorem 4.2 presented below gives conditions under which the estimator (4.19) is strongly consistent and asymptotically normal. Almost all of its proof is identical to the proof of the analogous uncorrelated-model results presented in Theorems 4.1.4 and 4.1.5 of Fuller (1987). Therefore, details of the proof of the present theorem will be given only at points of divergence from the proofs of the uncorrelated-model theorems.

Some notation required for the statement and proof of Theorem 4.2 is as follows. Recall that \( \Gamma_{vv}(0) \) is assumed to be nonsingular, and define

\[ \varrho_t = u_t - v_t \Gamma_{v v}(0)^{-1} \Gamma_{v u}(0) \]

\[ = u_t - (e_t - u_t \beta) \Gamma_{v v}(0)^{-1} \Gamma_{v u}(0) \]

\[ = \varrho_{-t} A \rho, \]

(4.20)

where

\[ A \rho = \{ -[\Gamma_{v v}(0)]^{-1} \Gamma_{v u}(0) \}' , \{ \Gamma_k + \beta [\Gamma_{v v}(0)]^{-1} \Gamma_{v u}(0) \}' \} \cdot \]  

(4.21)
Under the conditions of Theorem 4.1, \((x_{ct}, \rho_c, v_t)\) follows a linear process model and is fourth-order stationary. Thus, one may define the lagged covariance function of the \((x_{ct}, \rho_c, v_t)\) process, e.g.,

\[ \Gamma_{\rho^p}(t) = \text{Cov}(\rho_s', \rho_{s+t}') \]

\[ = \Lambda' \Gamma_{\rho^\epsilon}(t) \Lambda \rho \]

and

\[ \Gamma_{v^p}(t) = \text{Cov}(v_s', v_{s+t}') \]

\[ = \Lambda' \Gamma_{v^\epsilon}(t)(I_{(\Gamma', -\beta')'}) \]

The lagged third- and fourth-order cumulant functions of \((x_{ct}, \rho_c, v_t)\) may be defined similarly, in accordance with the definitions in expressions (3.7). Also, note that \(E(\rho_c) = 0\). Finally, under the conditions of Theorem 4.1, the cross-covariance function between \(x_{ct}\) and \((\rho_c, v_t)\) and the third- and fourth-order cross-cumulant functions between \(x_{ct}\) and \((\rho_c, v_t)\) are all identically zero, e.g.,

\[ \Gamma_{xp}(s) = 0, \quad \Gamma_{vxp}(0, s, t) = 0 \quad \text{and} \quad k_{xv\rho v}(0, q, s, t) = 0 \]

for all integers \(q\), \(s\) and \(t\). With these remarks in mind, one may now present some asymptotic properties of the estimator \(\hat{\beta}_{(T)}\).

**Theorem 4.2.** Let the conditions of Theorem 4.1 hold. Assume that \(\Gamma_{vv}(0)\) is nonsingular. Assume that \(\{\hat{\Gamma}_{vv}(T)(0)\}\) is a sequence of
p \times p \text{ random matrices which satisfy condition (4.16) and which are independent of } Z. \text{ Assume also that }

\[
\lim_{T \to \infty} \mathbb{E}_{\mathbb{E}}(T)(0) = \mathbb{E}_{\mathbb{E}}(0) \tag{4.22}
\]

with probability one. Define \( \hat{\beta}_T \) by expressions (4.17)-(4.19). Then

\[
\lim_{T \to \infty} \hat{\beta}_T = \beta \tag{4.23}
\]

with probability one. Also, as \( T \to \infty \),

\[
T^{1/2} \text{vec}(\hat{\beta}_T - \beta) \xrightarrow{L^2} \mathcal{N}_{rk}[0, \{I_r \otimes \bar{M}_{xx}(0)\}^{-1} G_2 \{I_r \otimes \bar{M}_{xx}(0)\}^{-1}] \tag{4.24}
\]

where

\[
G_2 = G + [(I_r, -\beta') \otimes A \rho] \Theta [(I_r, -\beta') \otimes A \rho],
\]

\[
G = \sum_{t=-\infty}^{\infty} \left[ \Gamma_{\mu\mu}(t) \otimes \bar{M}_{\mu\mu}(t) + \Gamma_{\rho\rho}(t) + \Gamma_{\rho\rho}(-t) \right] + \text{vec}[\Gamma_{\rho\rho}(t)] \text{vec}[\Gamma_{\rho\rho}(-t)]' + \bar{M}_{\rho\rho}(0, 0, t)' + \bar{M}_{\rho\rho}(0, 0, t)' \text{ vec}[\bar{M}_{\rho\rho}(0, 0, t)'],
\]

and \( A \rho \) is defined by expression (4.21).
Proof. The strong consistency of $\hat{\beta}(T)$ follows immediately from Result 4.1, condition (4.22), and an extension of the proof of Theorem 4.1.4 of Fuller (1987, pp. 303-305) to the case of singular $\Gamma_{e}(0)$. Also, a repetition of the arguments associated with expressions (4.1.50)-(4.1.58) in the proof of Theorem 4.1.5 of Fuller (1987, pp. 306-307) indicates that

$$\hat{\beta}(T) - \beta = [\tilde{M}_{xx}(0)]^{-1}[M_{xv} + M_{pv} - \hat{\Gamma}_{pv}(T)(0)] + o_p(T^{-1/2})$$

$$= [\tilde{M}_{xx}(0)]^{-1}[M_{xv} + M_{pv} - \Gamma_{\sim pv}(0) - \hat{\Gamma}_{pv}(T)(0) - \Gamma_{\sim pv}(0)] + o_p(T^{-1/2})$$

where

$$M_{xv} = T^{-1} \sum_{t=1}^{T} x'_t v_t,$$

$$M_{pv} = T^{-1} \sum_{t=1}^{T} p'_t v_t,$$

$$\Gamma_{pv}(0) = A'_p \Gamma_{\epsilon e}(0)(I_r, -\beta')',$$

and

$$\hat{\Gamma}_{pv}(T)(0) = A'_p [\hat{\Gamma}_{\sim pv}(T)(0)](I_r, -\beta')'.$$

Thus,
\[ T^{1/2} \text{vec}(\hat{\beta}(T) - \beta) = [I_x \otimes H_{xx}(0)]^{-1}\{T^{1/2} \text{vec}[M_{xv} + M_{pv} - \Gamma_{pv}(0)] + T^{1/2} \text{vec}[\hat{\Gamma}_{pv}(T)(0) - \Gamma_{pv}(0)]\} + o_p(1). \] (4.25)

Now by Theorem 3.1 and the conditions of Theorem 4.1,

\[ T^{1/2} \text{vec}[M_{xv} + M_{pv} - \Gamma_{pv}(0)] \xrightarrow{d} N(0, \Sigma_{\rho}). \] (4.26)

as \( T \to \infty \). Also,

\[ T^{1/2} \text{vec}[\hat{\Gamma}_{pv}(T)(0) - \Gamma_{pv}(0)]\]

\[ = T^{1/2} \text{vec}\left\{ A' \left[ \hat{\Gamma}_{\rho}(\hat{\xi}(T)) - \Gamma_{\rho}(0) \right] (I_x, -\beta')' \right\} \]

\[ = T^{1/2} \left[ (I_x, -\beta') \otimes A'_\rho \right] \text{vec}[\hat{\Gamma}_{\rho}(\hat{\xi}(T))(0) - \Gamma_{\rho}(0)]. \] (4.27)

By condition (4.16), expression (4.27) converges in law to a normal \( (0, [(I_x, -\beta') \otimes A'_\rho \otimes [(I_x, -\beta')' \otimes A'_\rho]) \) random vector as \( T \to \infty \). The conclusion (4.24) then follows from expressions (4.25), (4.26) and the independence of \( \hat{\Gamma}_{\rho}(\hat{\xi}(T))(0) \) from \( Z \).

Theorem 4.1, Corollary 4.1.1, and Theorem 4.2 indicate that a number of estimators developed originally for the uncorrelated errors-in-variables models (2.1)-(2.2) or (2.3)-(2.4) retain some desirable properties when the random components follow a correlated linear process.
model. Note, however, that for each estimator in question, the covariance matrix of the resulting asymptotic distribution is dependent on the serial covariance structure of the errors.

In addition to the standard estimation procedures developed previously for the uncorrelated errors-in-variables model, one may consider two other approaches to estimation for correlated forms of models (2.1)-(2.2) and (2.3)-(2.4): weighted estimation and full-information maximum likelihood estimation. Section 4.2 below discusses some asymptotic properties of a weighted estimator of \( \beta \), while Chapters 5 and 6 address the likelihood approach.

4.2. Weighted Estimators

As noted in Section 2.1, a number of authors have proposed weighted method-of-moments estimators of the regression coefficient \( \beta \) for heteroscedastic forms of the errors-in-variables models (2.1)-(2.2) and (2.3)-(2.4). One may consider similar weighted estimators for measurement error models with serially correlated errors.

In particular, let \( \pi(T) \), \( T = 1, 2, \ldots \) be a sequence of \( Tr \times Tr \) weight matrices with \( [(i, s), (j, t)] \)-th element equal to \( \pi_{ij}(T)(s, t) \), \( i, j = 1, 2, \ldots, r \). Assume that the array \( \{\pi_{ij}(T)(s, t)\} \) satisfies the following conditions.

1. For all \( 1 < i, j < r \) and all positive integers \( s, t \),

\[
\pi_{ij}(s, t) = \lim_{T \to \infty} \pi_{ij}(T)(s, t) \quad (4.28.a)
\]
exists and is finite.

(ii) There exist some $\beta > 0$ and $M > 0$ such that for all $1 < i, j < r$ and all positive integers $T$,

\[ \sum_{s=1}^{T} |\pi_{ij}(T)(s, t) - \pi_{ij}(s, t)| < MT^{-\beta} \quad (4.28.b) \]

and

\[ \sum_{s=1}^{T} |\pi_{ij}(T)(t, s) - \pi_{ij}(t, s)| < MT^{-\beta} \quad (4.28.c) \]

for all $1 < t < T$.

(iii) For any integer $d$,

\[ \pi_{ij}(d) = \lim_{s \to \infty} \pi_{ij}(s, s+d) \quad (4.28.d) \]

exists and is finite, and \{\pi_{ij}(d)\} is absolutely summable over $d$.

(iv) There exists an absolutely summable sequence of positive real numbers \{M_{pd}, d \in \mathbb{Z}\} such that for all $1 < i, j < r$ and all positive integers $s$,

\[ |\pi_{ij}(s, s+d) - \pi_{ij}(d)| < M_{pd} \quad (4.28.e) \]

Given a set of matrices $\pi(T)$, a weighted version of the regression coefficient estimator (4.7) may be written,
Theorem 4.3 gives conditions under which the weighted estimator \( \tilde{\beta}_\pi \) is strongly consistent and asymptotically normal.

**Theorem 4.3.** Let the assumptions of Theorem 4.1 hold and let the elements of the \( \{\mu_t\} \) sequence be uniformly bounded. Assume that the weight matrices \( \pi(T) \) satisfy conditions (4.28). Let \( \tilde{M}_-(0) \) be an \( r_k \times r_k \) -dimensional matrix with \((i,j), (m,n)\)-th element equal to

\[
\sum_{d=-\infty}^{\infty} \pi^{-1}(d)[\tilde{M}_{ij}(d) + \tilde{C}_{ij}(d)] ,
\]

and assume that \( \tilde{M}_-(0) \) is positive definite. Define \( \tilde{\beta}_\pi \) by expression (4.29).

a. Then \( \lim_{T \to \infty} \tilde{\beta}_\pi = \beta \) with probability one.

b. Also, as \( T \to \infty \),

\[
T^{1/2} \text{vec}(\tilde{\beta}_\pi - \beta) \xrightarrow{\mathcal{D}} N_{r_k X \times \pi \times \pi} \left[ \begin{array}{c} \tilde{M}_-(0) \end{array} \right]^{-1} \cdot \left[ \begin{array}{c} \tilde{M}_-(0) \end{array} \right]^{-1}
\]

where \( G_\pi \) is an \( r_k \times r_k \) -dimensional matrix with \((i,j), (m,n)\)-th
Proof. Note first that conditions (4.28.b) and (4.28.c) are identical to conditions (3.85), and recall that conditions (3.85) imply conditions (3.86). Also, note that condition (4.28.d) is equivalent to conditions (3.100.a) and (3.100.c), and that condition (4.27.e) is identical to condition (3.100.b). Next, note that
Lemma 3.14 and Lemma 3.17.c imply that expressions (4.31.a), (4.31.c), and (4.31.d) each converge to zero with probability one as $T \to \infty$; and that expression (4.31.b) converges to zero as $T \to \infty$. Then as $T \to \infty$

$$T^{-1}((I_r \odot X)' \pi(T)(I_r \odot X) - E[(I_r \odot u)' \pi(T)(I_r \odot u)] - \bar{M}_-(0)$$

(4.30)

$$= T^{-1}((I_r \odot x)' \pi(T)(I_r \odot x) - E[(I_r \odot x)' \pi(T)(I_r \odot x)])$$

(4.31.a)

$$+ T^{-1}E[(I_r \odot x)' \pi(T)(I_r \odot x)] - \bar{M}_-(0)$$

(4.31.b)

$$+ T^{-1}[(I_r \odot x)' \pi(T)(I_r \odot x) + (I_r \odot u)' \pi(T)(I_r \odot x)]$$

(4.31.c)

$$+ T^{-1}[(I_r \odot u)' \pi(T)(I_r \odot u) - E[(I_r \odot u)' \pi(T)(I_r \odot u)]}$$

(4.31.d)

converges to $\bar{M}_-(0)$ with probability one. Now $\bar{M}_-(0)$ is assumed to be positive definite, so the probability that expression (4.32) is positive definite increases to one as $T \to \infty$. Hence, one need only consider the case in which expression (4.32) is positive definite. A first-order expansion of a matrix inverse then implies that the inverse of expression (4.32) equals

$$[\bar{M}_-(0)]^{-1} + o(1) .$$
Note that

\[
\text{vec}(\tilde{\beta}_T - \beta) = ((I_T \odot X)'\pi(T)(I_T \odot X) - E[(I_T \odot u)'\pi(T)(I_T \odot u)])^{-1} \\
\times ((I_T \odot X)'\pi(T)\text{vec}(Y) - E[(I_T \odot u)'\pi(T)\text{vec}(e)]) - \text{vec}(\beta)
\]

\[
= ((I_T \odot X)'\pi(T)(I_T \odot X) - E[(I_T \odot u)'\pi(T)(I_T \odot u)])^{-1} \\
\times ((I_T \odot X)'\pi(T)\text{vec}(v) - E[(I_T \odot u)'\pi(T)\text{vec}(v)])
\]

(4.33)

\[
= ([\bar{M} - (0)]^{-1} + o_p(1)) \\
\times T^{-1}((I_T \odot X)'\pi(T)\text{vec}(v) - E[(I_T \odot u)'\pi(T)\text{vec}(v)])
\]

(4.34)

Now Lemma 3.14 and Lemma 3.17.c imply that

\[
T^{-1}((I_T \odot X)'\pi(T)\text{vec}(v) - E[(I_T \odot u)'\pi(T)\text{vec}(v)])
\]

converges to zero with probability one as \( T \to \infty \). This result, and the almost sure convergence of expression (4.32) to the positive definite matrix \( \bar{M} - (0) \) imply that expression (4.33) converges to zero with probability one as \( T \to \infty \), so conclusion (a) is established. Moreover, Lemma 3.16 and Lemma 3.17.b imply that

\[
T^{-1/2}((I_T \odot X)'\pi(T)\text{vec}(v) - E[(I_T \odot u)'\pi(T)\text{vec}(v)]) = o_p(1)
\]

(4.36)
Expressions (4.34) and (4.36) imply that

\[
T^{1/2} \text{vec}(\bar{\beta}_\pi - \beta)
\]

\[
= \left[ N_{\pi \pi}(0) \right]^{-1} T^{-1/2} \left\{ (I - X)' \bar{\pi}(T) \text{vec}(\nu) - E\left[ (I - U)' \bar{\pi}(T) \text{vec}(\nu) \right] \right\} + o_p(1).
\]

Result (b) then follows immediately from expression (4.37) and Corollary 3.2.1.

Theorem 4.3 gives some indication of the asymptotic behavior of a weighted estimator of \( \beta \) for model (4.1). Additional research is required before one may make practical recommendations regarding such weighted estimation procedures. In particular, Theorem 4.3 may be extended to the case in which the expectation matrices in expression (4.29) are replaced by consistent estimators. Also, optimality properties of various weight matrices \( \bar{\pi}(T) \) may be studied; extensions of some heteroscedastic-model arguments of Fuller (1987, p. 223) suggest the use of \( \bar{\pi}(T) = \frac{1}{\text{var}_v} \).
5. MODELS AND ESTIMATORS FOR THE STRUCTURAL CASE

This chapter addresses estimation for the structural measurement error model, i.e., the model in which \( x_t \) follows a second-order stationary process. Section 5.1 presents the matrices of first and second derivatives required for a direct Newton-Raphson approach to maximum likelihood estimation of identified parameters. Although this approach has the advantage of generality, much applied time series work relies on the assumption that components of interest follow autoregressive moving average models. Under such assumptions for \( x_t \) and \( e_t \), one may characterize the resulting observations \( z_t \) with either higher-order autoregressive moving average models or with state-space models. Sections 5.2, 5.3 and 5.4 address these two approaches as follows.

First, Section 5.2 considers simple additive "signal plus noise" models in which an observation \( x_t \) is the sum of a "true value" \( x_t \) and a "measurement error" \( u_t \). A univariate result in Box and Jenkins (1976) is extended to show that if \( x_t \) and \( u_t \) are two mutually uncorrelated vector autoregressive moving average processes, then \( x_t \) is also a vector autoregressive moving average process. However, the applicability of this result is limited by the dependence of the autoregressive and moving average orders of \( x_t \) on both the orders and the coefficient matrices of the \( x_t \) and \( u_t \) processes. Therefore, Section 5.3 uses state-space models to develop an alternative characterization of the same \( x_t \) process and, more generally, of any linear transformation of a vector autoregressive moving average process. This characterization
suggests the use of standard time series software to estimate the parameters of such "unobserved component" time series models, but this approach requires a considerable amount of auxiliary information. Consequently, Section 5.4 outlines a more direct approach to maximum likelihood estimation of identified parameters of the structural models (2.1)-(2.2) and (2.3)-(2.4). These two models are re-expressed in state-space form, and iterative expressions for low-dimensional derivative computations are given. These results lead to a possible Newton-Raphson procedure for maximum likelihood estimation in the normal structural case. Practical application of the results of this chapter are dependent on a number of issues, including identifiability of the process parameters, topology of the normal structural likelihood surface, and convergence properties of modified Newton-Raphson procedures. These issues are dependent on the details of the specific autoregressive and moving average models used for the $x_t$ and $e_t$ processes, so a thorough consideration of these issues will be deferred to future work with specific models and specific data sets.

5.1. Maximum Likelihood Estimation for General Second-Order Stationary Covariance Structure

In measurement error models for independent and identically distributed observations, the method of maximum likelihood and modifications thereof are frequently proposed for the estimation of $\theta$ and some nuisance parameters, e.g., $\text{Var}(x_t)$. Similarly, time series analysis often uses approximate maximum likelihood estimators. Therefore, it is
worthwhile to consider maximum likelihood estimation in the present case as well.

Recall from Chapter 4 the measurement error model,

$$\text{vec}(Z) = [((\beta, I_K)' \otimes I_T) \text{vec}(x) + \text{vec}(\varepsilon)] . \tag{5.1}$$

The definitions and notation of Chapter 4 are retained. If $[\text{vec}(x)', \text{vec}(\varepsilon)']'$ follows a multivariate normal distribution with mean $0$ and covariance matrix block diag$(\Gamma_{XX}, \Gamma_{EE})$, then it follows that $\text{vec}(Z)$ is normally distributed with mean zero and covariance matrix $\Gamma_{ZZ}$, where

$$\Gamma_{ZZ} = [((\beta, I_K)' \otimes I_T) \Gamma_{XX} [(\beta, I_K) \otimes I_T] + \Gamma_{EE}] . \tag{5.2}$$

Let $\alpha_{\xi}$ be a $1 \times L_{\xi}$ vector with elements equal to a minimal set of parameters for the covariance matrix $\Gamma_{XX}$, let $\alpha_{\varepsilon}$ be a $1 \times L_{\varepsilon}$ vector with elements equal to a minimal set of parameters for the covariance matrix $\Gamma_{EE}$, let $L = rk + L_{\xi} + L_{\varepsilon}$, and let $\gamma = [\text{vec}(\beta)', \alpha_{\xi}, \alpha_{\varepsilon}] = (\alpha_1, \alpha_2, \ldots, \alpha_L)$, say. The likelihood function of $\gamma$ is then

$$L(\gamma; Z) = (2\pi)^{-L/2} |\Gamma_{ZZ}|^{-1/2} \exp[-(2^{-1})\text{vec}(Z)' \Gamma_{ZZ}^{-1} \text{vec}(Z)] . \tag{5.3}$$

Inspection of (5.3) suggests one difficulty in developing maximum likelihood estimators for model (5.1). A common first step in maximum
likelihood estimation is to reduce the original observation vector to a sufficient statistic of much lower dimensionality. In this case, however, inspection of (5.2) indicates that \( \Gamma_{ZZ} \) does not have block-diagonal form. Consequently, the usual sufficiency and data-reduction arguments based on the factorization theorem [Lehmann (1983), p. 39] cannot be applied in this case. Due to the structure of (5.2), each element of \( \Gamma_{ZZ}^{-1} \) is a function of \( \beta \), so that reduction is not possible even for the case in which only \( \beta \) is unknown.

Despite this problem, one may study the likelihood function in some detail. Let

\[
g(\zeta; Z) = -2\ln[L(\zeta; Z)] - (T_p)\ln(2\pi) = \ln|\Gamma_{Z}\Gamma_{Z}^{-1} \Gamma_{Z}^{-1}\vec{\sigma}^{(Z)}| + \vec{\sigma}^{(Z)}'\Gamma_{Z}^{-1}\vec{\sigma}^{(Z)}.\]

Throughout this chapter assume that the elements of \( \Gamma_{X} \) are twice continuously differentiable functions of \( \zeta \) and that the elements of \( \Gamma_{Z} \) are twice continuously differentiable functions of \( \zeta \). (These conditions will be satisfied if, for example, \( \zeta \) and \( \epsilon \) follow autoregressive moving average processes; \( \zeta \) represents the distinct elements of the residual variance matrix and of the autoregressive and moving average coefficient matrices of the \( \zeta \) process; and \( \zeta \) represents the distinct elements of the residual variance matrix and of the autoregressive and moving average coefficient matrices of the \( \epsilon \) process.) Then inspection of expression (5.2) indicates that the elements of \( \Gamma_{ZZ} \) are twice continuously differentiable with respect to \( \zeta \). Hence, one may use \( g(\zeta; Z) \) to obtain a matrix of second
derivatives associated with (5.3) and to develop possible numerical methods for the estimation of unknown components of $g$. This work requires several results on matrix differentiation; some elementary general results are reviewed in Appendix A, while more specific results are developed below. To simplify the development below, assume that the elements of $g$, $\alpha_x$ and $\alpha_e$ are functionally unrelated.

Consider first the differentiation of $\Gamma_{zz}$ with respect to $\alpha_z$, $z=1,2,\ldots,L$. Let $\frac{\partial (1)}{\partial \alpha_z} = \frac{\partial g}{\partial \alpha_z}$. Then by Result 9.4,

$$\frac{\partial}{\partial \alpha_z} \{ \Gamma_{zz} \} = \frac{\partial}{\partial \alpha_z} \{ ( ( g, I_k' )', \alpha I_T ) \Gamma_{xx} \} \left[ ( g, I_k' ) \alpha I_T \right] + \Gamma_{ee}$$

$$= \{ ( ( g,(1,z) \alpha I_T )' \alpha I_T \Gamma_{xx} \} \left[ ( g, I_k' ) \alpha I_T \right] + \Gamma_{ee}$$

$$= \{ ( ( g,(1,z) \alpha I_T )' \alpha I_T \Gamma_{xx} \} \left[ ( g, I_k' ) \alpha I_T \right] + \frac{\partial}{\partial \alpha_z} \left[ \Gamma_{ee} \right].$$

(5.4)

Note that $\frac{\partial}{\partial \alpha_z} \{ g \} = \tilde{A}_{ij}$, where $\tilde{A}_{ij}$ is an $r \times k$ matrix with a one in the $(i,j)$-th entry and zeros elsewhere. The assumption that $g$, $\alpha_x$ and $\alpha_e$ are functionally unrelated implies that

$$\frac{\partial}{\partial \alpha_z} \{ \Gamma_{zz} \} = \{ ( \tilde{A}_{ij}, 0 )', \alpha I_T \} \Gamma_{xx} \left[ ( g, I_k' ) \alpha I_T \right] + \frac{\partial}{\partial \alpha_z} \left[ \Gamma_{ee} \right].$$
\[
\frac{\partial}{\partial \alpha_{x_1}} (\Gamma_{ZZ}) = \left[ (\tilde{A}_{d_{ij}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{jk}}, 0_k) \right] \cdot (\Gamma_T),
\]

and

\[
\frac{\partial}{\partial \alpha_{x_1}} (\Gamma_{Zz}) = \frac{\partial}{\partial \alpha_{x_1}} (\Gamma_{zz}).
\]

Similarly, one may obtain the following second derivatives:

\[
\frac{\partial^2}{\partial \alpha_{x_1} \partial \alpha_{x_2}} (\Gamma_{ZZ}) = \left[ (\tilde{A}_{d_{ij}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{jk}}, 0_k) \right] \cdot (\Gamma_T),
\]

\[
+ \left[ (\tilde{A}_{d_{lm}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{kl}}, 0_k) \right] \cdot (\Gamma_T),
\] (5.5)

\[
\frac{\partial^2}{\partial \alpha_{x_1} \partial \alpha_{x_j}} (\Gamma_{Zz}) = \frac{\partial^2}{\partial \alpha_{x_1} \partial \alpha_{x_j}} (\Gamma_{zz}),
\]

\[
\frac{\partial^2}{\partial \alpha_{x_1} \partial \alpha_{x_j}} (\Gamma_{Zz}) = \left[ (\tilde{A}_{d_{ij}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{jk}}, 0_k) \right] \cdot (\Gamma_T),
\]

and

\[
\frac{\partial^2}{\partial \alpha_{x_1} \partial \alpha_{x_j}} (\Gamma_{Zz}) = \left[ (\tilde{A}_{d_{ij}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{jk}}, 0_k) \right] \cdot (\Gamma_T)
\]

\[
+ \left[ (\tilde{A}_{d_{ij}}, 0) \right] \cdot (\Gamma_T)_{xx} \left[ (\tilde{A}_{d_{jk}}, 0_k) \right] \cdot (\Gamma_T).
\]

All other second order partial derivatives equal zero.

The first and second partial derivatives of \( \text{vec}(Z) \cdot r^{-1} \text{vec}(Z) \) follow from repeated application of Results 9.3, 9.5 and 9.6:
\[
\frac{\partial}{\partial \theta_{i,j}} \left[ \text{vec}(Z)^{\top} \Gamma_{ZZ}^{-1} \text{vec}(Z) \right] = - \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \theta_{i,j}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
= - \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ}^{-1} \left( (\Delta_{i,j}, 0)^{\top} \otimes I_{T} \right) \Gamma_{xx} \left( (\theta, I_{k}) \otimes I_{T} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
+ \left( (\theta, I_{k}) \otimes I_{T} \right) \Gamma_{xx} \left( (\Delta_{i,j}, 0)^{\top} \otimes I_{T} \right) \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
= - 2 \text{vec}(Z)^{\top} \Gamma_{ZZ}^{-1} \left( (\Delta_{i,j}, 0)^{\top} \otimes I_{T} \right) \Gamma_{xx} \left( (\theta, I_{k}) \otimes I_{T} \right) \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
\frac{\partial}{\partial x_{1}} \left[ \text{vec}(Z)^{\top} \Gamma_{ZZ}^{-1} \text{vec}(Z) \right] = - \text{vec}(Z)^{\top} \left[ \frac{\partial}{\partial x_{1}} \left( \Gamma_{xx} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
\frac{\partial}{\partial \theta_{e_{1}}} \left[ \text{vec}(Z)^{\top} \Gamma_{ZZ}^{-1} \text{vec}(Z) \right] = - \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \theta_{e_{1}}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
\frac{\partial^{2}}{\partial \theta_{i,j} \partial \theta_{l,m}} \left[ \text{vec}(Z)^{\top} \Gamma_{ZZ}^{-1} \text{vec}(Z) \right] = \frac{\partial}{\partial \theta_{i,j}} \left\{ - \text{vec}(Z)^{\top} \left( \Gamma_{ZZ} \right) \frac{\partial}{\partial \theta_{l,m}} \left( \Gamma_{ZZ} \right) \right\} \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
= - \text{vec}(Z)^{\top} \left[ \left( \Gamma_{ZZ} \right) \frac{\partial}{\partial \theta_{i,j}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
+ \left( \Gamma_{ZZ} \right) \frac{\partial}{\partial \theta_{l,m}} \left( \Gamma_{ZZ} \right) \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
= - \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ} \frac{\partial}{\partial \theta_{i,j}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
- \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ} \frac{\partial}{\partial \theta_{l,m}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]

\[
+ \text{vec}(Z)^{\top} \left[ \Gamma_{ZZ} \frac{\partial^{2}}{\partial \theta_{i,j} \partial \theta_{l,m}} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \text{vec}(Z)
\]
\[
= \text{vec}(z)' \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \beta_{lm}} (\Gamma_{ZZ}) \\
+ \left[ \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) - \frac{\partial^2}{\partial \beta_{ij} \partial \beta_{jm}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \text{vec}(z),
\]

\[
\frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} \left[ \text{vec}(z)' \Gamma_{ZZ}^{-1} \text{vec}(z) \right]
= \text{vec}(z)' \Gamma_{ZZ}^{-1} \left[ 2 \frac{\partial}{\partial \alpha_{ei}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{ej}} (\Gamma_{ee}) \\
+ \left[ \frac{\partial}{\partial \alpha_{ei}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{ej}} (\Gamma_{ee}) - \frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \text{vec}(z)
\]

\[
= \text{vec}(z)' \Gamma_{ZZ}^{-1} \left[ 2 \frac{\partial}{\partial \alpha_{ei}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{ej}} (\Gamma_{ee}) \\
- \frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \text{vec}(z)
\]

\[
\frac{\partial^2}{\partial \alpha_{x_i} \partial \alpha_{x_j}} \left[ \text{vec}(z)' \Gamma_{ZZ}^{-1} \text{vec}(z) \right]
= \text{vec}(z)' \Gamma_{ZZ}^{-1} \left[ 2 \frac{\partial}{\partial \alpha_{x_i}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{x_j}} (\Gamma_{ZZ}) - \frac{\partial^2}{\partial \alpha_{x_i} \partial \alpha_{x_j}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \text{vec}(z),
\]
\[
\frac{\partial^2}{\partial \alpha_{x} \partial \alpha_{e}} \left[ \text{vec}(Z)' \Gamma_{ZZ}^{-1} \text{vec}(Z) \right]
\]

\[= \text{vec}(Z)' \Gamma_{ZZ}^{-1} \left[ \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{ZZ} \right) \Gamma_{ZZ}^{-1} \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{ZZ} \right) \right] \]

\[+ \left[ \frac{\partial}{\partial \rho_{ij}} \Gamma_{zz} \right] \Gamma_{zz}^{-1} \left[ \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right] - \frac{\partial^2}{\partial \alpha_{x} \partial \alpha_{e}} \left( \Gamma_{zz} \right) \Gamma_{zz}^{-1} \text{vec}(Z) \]

\[= \text{vec}(Z)' \Gamma_{ZZ}^{-1} \left[ \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \Gamma_{zz}^{-1} \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \right] \]

\[= \text{vec}(Z)' \Gamma_{ZZ}^{-1} \left[ \left( \frac{\partial}{\partial \alpha_{e}} \Gamma_{ee} \right) \Gamma_{ee}^{-1} \left( \frac{\partial}{\partial \alpha_{e}} \Gamma_{ee} \right) \right] \]

\[+ \left[ \frac{\partial}{\partial \rho_{ij}} \Gamma_{zz} \right] \Gamma_{zz}^{-1} \left[ \frac{\partial}{\partial \alpha_{e}} \Gamma_{zz} \right] - \frac{\partial^2}{\partial \alpha_{e} \partial \alpha_{e}} \left( \Gamma_{ee} \right) \Gamma_{ee}^{-1} \text{vec}(Z) \]

\[= 2 \text{vec}(Z)' \Gamma_{ZZ}^{-1} \left[ \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \Gamma_{zz}^{-1} \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \right] \Gamma_{zz}^{-1} \text{vec}(Z) \]

and

\[
\frac{\partial^2}{\partial \alpha_{x} \partial \alpha_{e}} \left[ \text{vec}(Z)' \Gamma_{ZZ}^{-1} \text{vec}(Z) \right]
\]

\[= \text{vec}(Z)' \Gamma_{ZZ}^{-1} \left[ \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \Gamma_{zz}^{-1} \left( \frac{\partial}{\partial \alpha_{x}} \Gamma_{zz} \right) \right] \]

\[+ \left[ \frac{\partial}{\partial \alpha_{e}} \Gamma_{ee} \right] \Gamma_{ee}^{-1} \left[ \frac{\partial}{\partial \alpha_{e}} \Gamma_{ee} \right] - \frac{\partial^2}{\partial \alpha_{x} \partial \alpha_{e}} \left( \Gamma_{ee} \right) \Gamma_{ee}^{-1} \text{vec}(Z) \]
Repeated application of Result 9.8 allows one to compute first and second partial derivatives of \( \ln |r_{\omega Z}| \), namely

\[
\frac{\partial}{\partial \theta_{ij}} [\ln |r_{\omega Z}|] = \text{tr} \left[ \left[ \frac{\partial}{\partial \theta_{ij}} (r_{\omega Z}) \right] r_{\omega Z}^{-1} \right] \\
= \text{tr} \left\{ [(\tilde{\alpha}_{ij}, 0)' \ast I_T] r_{\omega Z}^{-1} \left[ (\alpha, I_k) \ast I_T \right] r_{\omega Z}^{-1} \right\} \\
+ \left[ (\alpha, I_k) \ast I_T \right] r_{\omega Z}^{-1} \left[ [(\tilde{\alpha}_{ij}, 0)' \ast I_T] r_{\omega Z}^{-1} \right]\right. \\
= 2 \text{tr} \left\{ [(\tilde{\alpha}_{ij}, 0)' \ast I_T] r_{\omega Z}^{-1} \left[ (\alpha, I_k) \ast I_T \right] r_{\omega Z}^{-1} \right\} ,
\]

\[
\frac{\partial}{\partial \alpha_{xi}} [\ln |r_{\omega Z}|] = \text{tr} \left[ \left[ \frac{\partial}{\partial \alpha_{xi}} (r_{\omega Z}) \right] r_{\omega Z}^{-1} \right] \\
= \text{tr} \left\{ [(\alpha, I_k)' \ast I_T] r_{\omega Z}^{-1} \left[ \frac{\partial}{\partial \alpha_{xi}} (r_{\omega Z}) \right] [(\alpha, I_k) \ast I_T] r_{\omega Z}^{-1} \right\} ,
\]

\[
\frac{\partial}{\partial \epsilon_{xi}} [\ln |r_{\omega Z}|] = \text{tr} \left[ \left[ \frac{\partial}{\partial \epsilon_{xi}} (r_{\omega Z}) \right] r_{\omega Z}^{-1} \right] \\
= \text{tr} \left\{ [\frac{\partial}{\partial \epsilon_{xi}} (r_{\omega Z})] r_{\omega Z}^{-1} \right\} \\
= \text{tr} \left\{ \left[ \frac{\partial}{\partial \epsilon_{xi}} (r_{\omega Z}) \right] r_{\omega Z}^{-1} \right\} ,
\]
\[
\frac{\partial^2}{\partial \beta_{ij} \partial \beta_{lm}} [\ln |\Gamma_{ZZ}|] = \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \beta_{ij} \partial \beta_{lm}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \beta_{lm}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1},
\]
\[
\frac{\partial^2}{\partial \alpha_{xi} \partial \alpha_{xj}} [\ln |\Gamma_{ZZ}|] = \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \alpha_{xi} \partial \alpha_{xj}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \alpha_{xi}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{xj}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1},
\]
\[
\frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} [\ln |\Gamma_{ZZ}|] = \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \alpha_{ei}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{ej}} (\Gamma_{ee}) \right] \Gamma_{ZZ}^{-1},
\]
\[
\frac{\partial^2}{\partial \beta_{ij} \partial \alpha_{x\ell}} [\ln |\Gamma_{ZZ}|] = \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \beta_{ij} \partial \alpha_{x\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{x\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1},
\]
\[
\frac{\partial^2}{\partial \beta_{ij} \partial \alpha_{e\ell}} [\ln |\Gamma_{ZZ}|] = - \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \beta_{ij} \partial \alpha_{e\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{e\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1},
\]
and
\[
\frac{\partial^2}{\partial \alpha_{xi} \partial \alpha_{e\ell}} [\ln |\Gamma_{ZZ}|] = - \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \alpha_{xi} \partial \alpha_{e\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \right\} \\
- \left[ \frac{\partial}{\partial \alpha_{xi}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{e\ell}} (\Gamma_{ZZ}) \right] \Gamma_{ZZ}^{-1}.
\]

The derivative results developed above applied to the structural measurement error model with no intercept and with the mean of each \( x_t \) equal to \( \mathbf{0}_{1 \times k} \). Now consider the more general model.
\[ Z_t = (\beta_0, 0_1 \times k) + \varepsilon_t (\beta, \mathbf{I}_k) + \varepsilon_t \quad (5.6) \]

\[ = [(\beta_0 + \mu_\infty \beta), \mu_\infty] + (\varepsilon_t - \mu_\infty)(\beta, \mathbf{I}_k) + \varepsilon_t \]

\[ = \mu_Z + (\varepsilon_t - \mu_\infty)(\beta, \mathbf{I}_k) + \varepsilon_t, \quad t=1, 2, \ldots, T, \]

where \( \mu_Z = [(\beta_0 + \mu_\infty \beta), \mu_\infty] \), \( \beta_0 \) is an unconstrained \( 1 \times r \) vector, \( \mu_\infty \) is an unconstrained \( 1 \times k \) vector, \( \beta \) is an unconstrained \( k \times r \) matrix, and \( (\varepsilon_t - \mu_\infty) \) and \( \varepsilon_t \) satisfy the descriptions of \( \varepsilon_t \) and \( \varepsilon_t \), respectively, following (5.1). Note that the transformation from \( \mathbb{R}^p \) to \( \mathbb{R}^p \) represented by

\[ \mu_Z' = [(\beta_0 + \mu_\infty \beta), \mu_\infty]' = \begin{bmatrix} I_r & \beta' \\ 0 & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \beta_0' \\ \mu_\infty' \end{bmatrix} \quad (5.7) \]

is nonsingular and has Jacobian equal to 1. The functional invariance of maximum likelihood estimation implies that the maximum likelihood estimators \( \hat{\mu}_Z, \hat{\beta}_1 \) and \( \hat{\mu}_x, \hat{\beta}_0, \hat{\alpha}_1 \) are equivalent in the sense that \( \hat{\mu}_Z, \hat{\beta}_0, \hat{\beta}_0, \hat{\alpha}_1 \) and \( \hat{\mu}_x \) satisfy relation (5.7), where \( \hat{\alpha}_1 \) is the maximum likelihood estimator of the vector of unknown parameters \( \alpha \) in \( \alpha = [\text{vec}(\beta), \alpha_x, \alpha_\varepsilon] \). Therefore, the derivative and information matrices for the estimation of the unknown parameters of model (5.6) may be obtained by augmenting the corresponding matrices for the zero-mean model (5.1).
The likelihood function for model (5.6) is

\[ L_{\mu}[g, \mu_Z]; Z] = (2\pi)^{-1} p^T \left| \Gamma_{ZZ} \right|^{-1} \exp\left\{ -\frac{1}{2} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right]' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right] \right\} \]

and interest focuses on the first and second partial derivatives of

\[ \ln \left| \Gamma_{ZZ} \right| + \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right]' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right]. \quad (5.8) \]

By Result 9.9,

\[ \frac{\partial}{\partial \mu_Z} \left\{ \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right]' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right] \right\} \]

\[ = -2 \left( \mathbf{1}_p \& \mathbf{1}_T \right)' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right] \]

and

\[ \frac{\partial^2}{\partial \mu_Z^2} \left\{ \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right]' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left[ \text{vec}(Z) - \left( \mu'_Z \& \mathbf{1}_T \right) \right] \right\} \]

\[ = -2 \left( \mathbf{1}_p \& \mathbf{1}_T \right)' \left( \frac{1}{\Gamma_{ZZ}} \right)^{-1} \left( \mathbf{1}_p \& \mathbf{1}_T \right). \]

Also,
\[ \frac{\partial^2}{\partial \alpha_i \partial \mu_Z} \left\{ \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} = -2(I_p \otimes \mathbf{1}_T)' \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \frac{\partial}{\partial \alpha_i} \left( \Gamma_{ZZ} \right) \end{array} \right] \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right]\]

Therefore,

\[ \frac{\partial^2}{\partial \theta_{ij} \partial \mu_Z} \left\{ \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} = -2(I_p \otimes \mathbf{1}_T)' \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \frac{\partial}{\partial \theta_{ij}} \left( \Gamma_{ZZ} \right) \end{array} \right] \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right]\]

and

\[ \frac{\partial^2}{\partial x_i \partial \mu_Z} \left\{ \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\} = -2(I_p \otimes \mathbf{1}_T)' \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \frac{\partial}{\partial x_i} \left( \Gamma_{ZZ} \right) \end{array} \right] \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right]\]

Last, the results given previously for the first and second derivatives of \[ \vec{Z} \] with respect to \[ \mathbf{a} \] remain the same for

\[ \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right\] \Gamma_{ZZ}^{-1} \left[ \begin{array}{c} \vec{Z} \\ \end{array} \right] - (\mu_Z' \otimes \mathbf{1}_T) \right] ,
except that \( \text{vec}(z) \) is replaced by \( \text{vec}(z) - (\mu' \circ \mathbf{1}_I) \) in the derivative expressions.

Therefore, the vector of first partial derivatives of 
\[-2 \Delta \log L \mu ((a, \mu, Z), Z) \] with respect to \((a, \mu, Z)\) may be written
\[-2 \Delta \log L \mu ((a, \mu, Z), Z) \]
\[
\frac{\partial}{\partial \mu} \begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix} ,
\]

where:
\[
a = \frac{-2 \Delta \log L \mu ((a, \mu, Z), Z)}{\partial \text{vec} \mu} \]

has, in double-subscripted vector notation, \((i,j)\)-th element equal to
\[
\frac{\partial}{\partial \mu_{ij}} \left[ \Delta \log L \mu \right] + \frac{\partial}{\partial \mu_{ij}} \left\{ \left[ \text{vec}(z) - (\mu' \circ \mathbf{1}_I) \right] \Gamma_{zz}^{-1} \left[ \text{vec}(z) - (\mu' \circ \mathbf{1}_I) \right] \right\}
\]
\[
= \text{tr} \left\{ \left[ \frac{\partial}{\partial \mu_{ij}} \left( \Gamma_{zz} \right) \right] \Gamma_{zz}^{-1} \right\} - \left[ \text{vec}(z) - (\mu' \circ \mathbf{1}_I) \right] \Gamma_{zz}^{-1} \left[ \frac{\partial}{\partial \mu_{ij}} \left( \Gamma_{zz} \right) \right] \Gamma_{zz}^{-1} \left[ \text{vec}(z) - (\mu' \circ \mathbf{1}_I) \right] ;
\]
\[
b = \frac{-2 \Delta \log L \mu ((a, \mu, Z), Z)}{\partial a'_{ij}} \]

has \(i\)-th element equal to
\[ \frac{\partial}{\partial \alpha_i} \left( \frac{\partial}{\partial \alpha_i} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \right) + \frac{\partial}{\partial \alpha_i} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \]

\[ = \text{tr} \left\{ \left[ \frac{\partial}{\partial \alpha_i} \left( \Gamma_{ZZ} \right) \right] \Gamma_{ZZ}^{-1} \right\} \]

\[ + \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_i} \left( \Gamma_{ZZ} \right) \Gamma_{ZZ}^{-1} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) ; \]

\[ c = \frac{-2 \Im \left( L_{\mu \left( \left[ (a, \mu_Z) ; Z \right] \right)} \right)}{\partial \alpha_c} \]

has i-th element equal to

\[ \frac{\partial}{\partial \alpha_i} \left( \frac{\partial}{\partial \alpha_i} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \right) + \frac{\partial}{\partial \alpha_i} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \]

\[ = \text{tr} \left\{ \left[ \frac{\partial}{\partial \alpha_i} \left( \Gamma_{ee} \right) \right] \Gamma_{ZZ}^{-1} \right\} \]

\[ + \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_i} \left( \Gamma_{ee} \right) \Gamma_{ZZ}^{-1} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) ; \]

and

\[ d = \frac{-2 \Im \left( L_{\mu \left( \left[ (a, \mu_Z) ; Z \right] \right)} \right)}{\partial \mu'_Z} \]

\[ = -2 \left( I_p \otimes I_T \right) \Gamma_{ZZ}^{-1} \left( \text{vec}(Z) - \left( \mu'_Z \otimes I_T \right) \right) \cdot \]

Similarly, the matrix of second partial derivatives,

\[ \begin{align*}
- \frac{\partial^2 \Im \left( L_{\mu \left( \left[ (a, \mu_Z) ; Z \right] \right)} \right)}{\partial (a, \mu_Z)'^{\top} \partial (a, \mu_Z)} \cdot
\end{align*} \]
may be written

\[
\begin{array}{cccc}
A & B & C & D \\
B' & E & F & G \\
C' & F' & H & K \\
D' & G' & K' & L \\
\end{array}
\]

where

\[
A = \frac{(-2) \delta^2 \ln \{L_u[(a, \mu_Z); Z]\}}{\delta \text{vec}(\beta) \delta \text{vec}(\beta)}
\]

has, in double-subscripted notation, \([i,j], (l,m)]-th element equal to

\[
\begin{align*}
&\text{tr}[[\frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ})]^{-1} \frac{\partial}{\partial \beta_{lm}} (\Gamma_{ZZ})]^{-1} - 2\text{tr}[[\frac{\partial^2}{\partial \beta_{ij} \partial \beta_{lm}} (\Gamma_{ZZ})]^{-1} \\
&\quad + [\text{vec}(Z) - (\mu'_Z \circ \mathbf{1}_T)]^T \Gamma_{ZZ}^{-1} [\frac{\partial}{\partial \beta_{ij} \partial \beta_{lm}} (\Gamma_{ZZ})] \\
&\quad - 2[\frac{\partial}{\partial \beta_{ij}} (\Gamma_{ZZ})]^{-1} \frac{\partial}{\partial \beta_{lm}} (\Gamma_{ZZ})]^{-1} \text{vec}(Z) - (\mu'_Z \circ \mathbf{1}_T) \]
\]

\[
B = \frac{(-2) \delta^2 \ln \{L_u[(a, \mu_Z); Z]\}}{\delta \mathcal{X} \delta \text{vec}(\beta)}
\]

has \([i, (i,j)]-th element equal to
\[
\begin{align*}
\text{tr}\{ & \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \} \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \Gamma_{ZZ}^{-1} - \text{tr}\{ \frac{\partial^2}{\partial \alpha_{i,j} \partial \beta_{i,j}} (\Gamma_{ZZ}) \} \Gamma_{ZZ}^{-1} \\
& + \{ \text{vec}(Z) - (\mu_Z' \otimes I_T) \} \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \\
& - \frac{\partial^2}{\partial \alpha_{i,j} \partial \beta_{i,j}} (\Gamma_{ZZ}) \Gamma_{ZZ}^{-1} \{ (I_T \otimes I_T) \} ; \\
C &= \frac{(-2) \partial^2 \ln\{ L_{\mu} [(g, \mu_Z); Z] \}}{\partial \beta_{i,j} \partial \text{vec}(\vec{Z})}.
\end{align*}
\]

has \((i, (i,j))\)-th element equal to

\[
\begin{align*}
\text{tr}\{ & \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \} \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{EE}) \Gamma_{ZZ}^{-1} \\
& + 2 \{ \text{vec}(Z) - (\mu_Z' \otimes I_T) \} \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{EE}) \Gamma_{ZZ}^{-1} \\
& \times \frac{\partial}{\partial \beta_{i,j}} (\Gamma_{ZZ}) \Gamma_{ZZ}^{-1} \{ \text{vec}(Z) - (\mu_Z' \otimes I_T) \} ; \\
D &= \frac{(-2) \partial^2 \ln\{ L_{\mu} [(g, \mu_Z); Z] \}}{\partial \beta_{i,j} \partial \text{vec}(\vec{Z})}.
\end{align*}
\]

has \((i,j)\)-th row equal to

\[
\begin{align*}
2 \{ \text{vec}(Z) - (\mu_Z' \otimes I_T) \} \Gamma_{ZZ}^{-1} \frac{\partial}{\partial \alpha_{i,j}} (\Gamma_{ZZ}) \Gamma_{ZZ}^{-1} (I_T \otimes I_T) ; \\
E &= \frac{(-2) \partial^2 \ln\{ L_{\mu} [(g, \mu_Z); Z] \}}{\partial \alpha_{i,j} \partial \alpha_{i,j}}.
\end{align*}
\]

has \((i,j)\)-th element equal to
\[ \text{tr}\left[\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1}\right] - \text{tr}\left[\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1}\right] \\
+ \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right]' \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right] \\
- 2\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right] \\
\times \Gamma_{ZZ}^{-1} \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right] ; \\
F = \frac{(-2)3^2\ln[L_{\mu}\left((\alpha, \mu_{Z}); Z\right)]}{3_{\epsilon}3_{\epsilon}} \]

has \((i,j)\)-th element equal to

\[ \text{tr}\left[\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1}\right] - \text{tr}\left[\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1}\right] \\
- 2\left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_j}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right] \\
\times \Gamma_{ZZ}^{-1} \left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right] ; \\
G = \frac{(-2)3^2\ln[L_{\mu}\left((\alpha, \mu_{Z}); Z\right)]}{3_{\epsilon}3_{\epsilon}} \]

has \(i\)-th row equal to

\[ 2\left[\text{vec}(Z) - (\mu^t_{Z} \triangleq 1^T)\right]' \Gamma_{ZZ}^{-1} \left[\frac{3}{3_{x_i}}(I_{ZZ})\right] \Gamma_{ZZ}^{-1} \left(\Gamma_{p} \triangleq 1^T\right) ; \\
H = \frac{(-2)3^2\ln[L_{\mu}\left((\alpha, \mu_{Z}); Z\right)]}{3_{\epsilon}3_{\epsilon}} \]

has \((i,j)\)-th element equal to
\[\begin{align*}
\text{tr} & \left[ \frac{\partial}{\partial \alpha_{ij}} (\Gamma_{\varepsilon \varepsilon}) \right] \Gamma_{\varepsilon Z}^{-1} \left[ \frac{\partial}{\partial \alpha_{ij}} (\Gamma_{\varepsilon \varepsilon}) \right] \Gamma_{\varepsilon Z}^{-1} - \text{tr} \left[ \frac{\partial^2}{\partial \alpha_{ij} \partial \alpha_{ij}} (\Gamma_{\varepsilon \varepsilon}) \right] \Gamma_{\varepsilon Z}^{-1} \\
+ [\text{vec}(Z) - (\mu'_{Z} \otimes I_{T})]' & \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{ij}} (\Gamma_{\varepsilon \varepsilon}) \right] \Gamma_{ZZ}^{-1} [\text{vec}(Z) - (\mu'_{Z} \otimes I_{T})] ;
\end{align*}\]

has \(i\)-th row equal to

\[\begin{align*}
[\text{vec}(Z) - (\mu'_{Z} \otimes I_{T})]' & \Gamma_{ZZ}^{-1} \left[ \frac{\partial}{\partial \alpha_{ij}} (\Gamma_{\varepsilon \varepsilon}) \right] \Gamma_{ZZ}^{-1} (I_{P} \otimes I_{T}) ;
\end{align*}\]

and

\[\begin{align*}
L & = \frac{(-2) \alpha^2 \ln \{L_{u}[(\alpha, \mu_{Z}); Z]\}}{\partial \mu_{Z}^i \partial \mu_{Z}} \\
& = 2(I_{P} \otimes I_{T})' \Gamma_{ZZ}^{-1} (I_{P} \otimes I_{T}) .
\end{align*}\]

Given the derivative vector \((a', b', c', d')'\) and the Fisher information matrix \(I_{F}\) outlined above, one may develop a general Newton-Raphson procedure for the maximum likelihood estimation of identified elements of the parameter vector \((\alpha, \mu_{Z})\). Beyond the usual identification and differentiability requirements, no specific parametric structure for \(\Gamma_{XX}\) and \(\Gamma_{\varepsilon \varepsilon}\) was assumed, so the resulting maximum likelihood procedure could, in principle, apply to a very broad class of
normal $x_t$ and $e_t$ processes, including nonstationary normal processes. However, such a maximum likelihood estimation procedure is of little value unless the likelihood function $L(\theta, \mu_Z; Z)$ satisfies certain regularity conditions. Additional research will be required to develop the relation between the surface determined by $L(\theta, \mu_Z; Z)$ and the functional dependence of $\Gamma_{xx} = \Gamma_{xx}(\theta, \mu_Z), \Gamma_{ee} = \Gamma_{ee}(\theta, \mu_Z)$,

$$\Gamma_{vv} = [(I_T, -\beta')a_L][\Gamma_{ee}(\theta, \mu_Z)][(I_T, -\beta')a_L],$$

and $v = (y - I_T \mu_Z) + (x - I_T \mu_Z) \beta$ on $(\theta, \mu_Z)$. An understanding of the likelihood surface may then offer insight into the concavity of $L(\theta, \mu_Z; Z)$ with respect to $(\theta, \mu_Z)$; the existence of global or local maxima of $L(\theta, \mu_Z; Z)$ with respect to $(\theta, \mu_Z)$; boundary-case maxima; convergence of Newton-Raphson or other recursive procedures; and asymptotic behavior of the resulting maximum likelihood estimators.

The comments given above indicate some of the difficulties encountered with maximum likelihood estimation for model (2.3)-(2.4) at a high level of parametric generality. In addition, time-domain analysis of correlated observations often restricts "signal" and "noise" processes to have parametric structures associated with autoregressive integrated moving average (ARIMA) models or with state-space models. The following three sections address estimation procedures for structural measurement error models with such restricted $x_t$ and $e_t$ component process parametrizations.

Box and Jenkins (1976) maintained that many phenomena in engineering and the social sciences may be approximated well by the ARIMA models presented in Section 2.2. As discussed in Chapter 1, some such phenomena may also be modeled as the sum of a "true value" term and an "error" term. In such cases, it may be very useful to characterize the relationship between these two modeling approaches. Given such a relationship, one could, in principle, use standard time series software to estimate the parameters of the observed "noisy" process and then combine these "first stage" estimates with knowledge of the "noise process" parameters to obtain estimates of the "true process" parameters. This is roughly the approach of Miazaki (1985, p. 42 ff.).

This section assesses the applicability of this approach to structural measurement error models and other multivariate "signal plus noise" models. Subsection 5.2.1 reviews and provides some slight extensions of previous work in ARIMA modeling of univariate "signal plus noise" processes. Subsection 5.2.2 extends these results to multivariate processes and notes some parameterization issues which limit the practical applicability of the multivariate extensions. This limitation suggests that one consider an alternative modeling approach to multivariate "signal plus noise" models, as discussed in Section 5.3.
5.2.1. Univariate case

For some univariate "signal plus noise" models, it may be reasonable to model the "signal" series as an autoregressive moving average process and to model the "noise" series as an independent autoregressive moving average process. It is then of interest to determine the model of the resulting "signal plus noise" observations.

Several authors, including Pagano (1974), Box and Jenkins (1976) and Miazaki (1985) have addressed this issue. In particular, Box and Jenkins (1976, Appendix A.4.4) present the following argument for a univariate time series model.

Let \( x_t \) follow an ARIMA\((p_x, d, q_x)\) model,

\[
\phi_x(B) \nabla^d x_t = \theta_x(B) g_t
\]

(5.9)

where \( g_t \) is a univariate \((0, \sigma_g)\) white noise process and \( \phi_x(B) \) and \( \theta_x(B) \) are polynomials in \( B \) of order \( p_x \) and \( q_x \), respectively. Assume that the observed values \( X_t \) are the sum of the "true" \( x_t \) and a serially correlated "noise" term \( u_t \),

\[
X_t = x_t + u_t
\]

(5.10)

where \( u_t \) follows an ARMA\((p_u, q_u)\) process,

\[
\phi_u(B) u_t = \theta_u(B) c_t
\]

(5.11)
$\phi_{u}(B)$ and $\theta_{u}(B)$ are polynomials in $B$ of orders $p_{u}$ and $q_{u}$, respectively, and the $(0, \sigma_{cc})$ white noise process $c_{t}$ is independent of $g_{t}$. Then

$$\phi_{u}(B)\phi_{x}(B)v^{d}x_{t} = \phi_{u}(B)\phi_{x}(x_{t} + u_{t})$$

$$= \phi_{u}(B)\phi_{x}(B)v^{d}x_{t} + \phi_{u}(B)\phi_{x}(B)v^{d}u_{t}$$

$$= \phi_{u}(B)\theta_{x}(B)g_{t} + \phi_{x}(B)\theta_{u}(B)v^{d}c_{t}$$

(5.12)

where the last equality follows from the commutativity of the univariate polynomials $\phi_{x}(B)$, $v^{d} = (1 - B)^{d}$, and $\phi_{u}(B)$, and from the difference equations (5.9) and (5.11). Note that $\phi_{x}(B) = \phi_{u}(B)\phi_{x}(B)$ is a polynomial in $B$ of order $P_{x} = p_{u} + p_{x}$, $\phi_{u}(B)\theta_{x}(B)$ is a polynomial in $B$ of order $p_{u} + q_{x}$, and $\phi_{x}(B)\theta_{u}(B)v^{d}$ is a polynomial in $B$ of order $p_{x} + q_{u} + d$. Now consider the process

$$a_{t} = a_{t1} + a_{t2}$$

where

$$a_{t1} = \phi_{u}(B)\theta_{x}(B)g_{t}$$

and

$$a_{t2} = \phi_{x}(B)\theta_{u}(B)v^{d}c_{t}$$

Exclude the trivial cases in which $\sigma_{gg} = 0$ or $\sigma_{cc} = 0$. Since the
autocovariance function $\gamma_1(h)$ of $a_{t1}$ equals zero at lags greater than $p_u + q_x$ and the autocovariance function $\gamma_2(h)$ of $a_{t2}$ equals zero at lags greater than $d + p_x + q_u$, the autocovariance function $\gamma_a(h)$ of $a_t$ equals zero at lags greater than

$$Q_X \equiv \max(p_u + q_x, d + p_x + q_u).$$

Then by Theorem II.10' of Hannan (1970, p. 66), it follows that $a_t$ follows a moving average process of order less than or equal to $Q_X$,

$$a_t = \sum_{j=0}^{Q_X} \theta_j X_{t-j},$$

say, where $\{X_t\}$ is some sequence of mutually uncorrelated $(0, \sigma_{Xt}^2)$ random variables, $\theta_j X_j \in \mathbb{R}$ for all $j$, $\theta_{X0} = 1$, and the roots of the polynomial $\theta_X(\lambda) = \sum_{j=0}^{Q_X} \theta_j X_j \lambda^j$ all fall on or outside the unit circle. Moreover, since

$$f_a(w) = f_{a1}(w) + f_{a2}(w),$$

for all $w \in [-\pi, \pi]$, where $f_a(w)$, $f_{a1}(w)$ and $f_{a2}(w)$ are the spectral densities of the $a_t$, $a_{t1}$ and $a_{t2}$ processes, respectively, $\theta_X(\lambda)$ will have all roots outside the unit circle if either $\phi_u(\lambda) \theta_X(\lambda)$ has no roots on the unit circle; or if $d = 0$ and $\phi_X(\lambda) \theta_u(\lambda)$ has no roots on the unit circle. Similarly, $\phi_X(\lambda)$ will have all its roots outside the unit circle provided that each of $\phi_X(\lambda)$
and \( \phi_u(\lambda) \) have all their roots outside the unit circle. Box and Jenkins thus conclude that the "noisy" observations \( X_t \) follow an ARIMA\((P_X, d, Q_X)\) model,

\[
\phi_X(B) \nabla^d X_t = \theta_X(B) f_t.
\]

Note that the orders \( P_X \) and \( Q_X \) are not necessarily minimal. For example, suppose that \( \phi_u(B) \phi_x(B) \), \( \phi_u(B) \theta_x(B) \), and \( \phi_x(B) \theta_u(B) \) have a common invertible polynomial factor \( \alpha(B) \) of order \( A > 1 \), i.e.

\[
\phi_u(B) \phi_x(B) = \alpha(B) \psi_1(B),
\]

\[
\phi_u(B) \theta_x(B) = \alpha(B) \psi_2(B), \quad \text{and}
\]

\[
\phi_x(B) \theta_u(B) = \alpha(B) \psi_3(B).
\]

Then one may rewrite expression (5.12) as

\[
\alpha(B) \psi_1(B) \nabla^d X_t = \alpha(B) \psi_2(B) g_t + \alpha(B) \psi_3(B) \nabla^d c_t,
\]

which is equivalent to

\[
\psi_1(B) \nabla^d X_t = \psi_2(B) g_t + \psi_3(B) \nabla^d c_t,
\]

an ARIMA\((P_X - A, d, Q_X - A)\) model.
Finally, the arguments given above extend to the sum of three or more ARIMA processes. Let \( x_{t1}, i = 1, 2, \ldots, s \), follow mutually uncorrelated univariate ARIMA\((p^*_i, d^*_i, q^*_i)\) processes,

\[
\phi_i(B)^{(d^*_i)} x_{t1} = \theta_i(B) g_{t1}.
\]

where \( \{g_{ti}, t \in \mathbb{Z}, i = 1, 2, \ldots, s\} \) is a set of \( s \) mutually uncorrelated sequences of mutually uncorrelated \( (0, \sigma_{g_{ti}}) \) random variables. Define

\[
X_t = \sum_{i=1}^{s} x_{t1},
\]

\[
d = \max_{1 \leq i \leq s} \{d^*_i\},
\]

and \( c^*_i = d - d^*_i \). Let \( \phi_X(B) \) be the comonic least common multiple of the polynomials \( \phi_i(B), i = 1, 2, \ldots, s \), let \( P_X \) be the degree of \( \phi_X(B) \), and define the polynomial \( \psi_i(B) \) of degree \( r^*_i \) by the relations

\[
\phi_X(B) = \psi_i(B) \phi_i(B),
\]

\( i = 1, 2, \ldots, s \). Then

\[
\phi_X(B)^{d}_X X_t = \sum_{i=1}^{s} \psi_i(B)^{(c^*_i)} = \sum_{i=1}^{s} \psi_i(B)^{(c^*_i)} \theta_i(B) g_{t1}.
\]
Define

\[ a_t = \sum_{i=1}^{s} a_{ti} \]

where

\[ a_{ti} = \psi_i(B) \theta_i(B) v(c_i) g_{ti}, \quad i=1, 2, \ldots, s. \]

Now \( \psi_i(B) \theta_i(B) v(c_i) \) is a polynomial in \( B \) of order \( r_i + q_i + c_i \), so the autocovariance function \( \gamma_j(h) \) of \( a_{tj} \) equals zero at lags greater than \( r_i + q_i + c_i \), and the autocovariance function \( \gamma_a(h) \) of \( a_t \) equals zero at lags greater than

\[ Q_X \equiv \max_{1 \leq i \leq s} \{ r_i + q_i + c_i \}. \]

Then again by Theorem II.10' of Hannan (1970, p. 66), \( a_t \) follows a moving average process of order less than or equal to \( Q_X \).

\[ a_t = \sum_{j=0}^{Q_X} \theta_j f_{t-j}, \]

say, where \( \{f_t\} \) is some sequence of mutually uncorrelated \( (0, \sigma_f) \) random variables, \( \theta_j \in \mathbb{R} \) for all \( j \), \( \theta_{X0} = 1 \), and the roots of the polynomial \( \theta_X(\lambda) \) all fall on or outside the unit circle.

Moreover, all the roots of \( \theta_X(\lambda) \) will fall outside the unit circle provided that at least one of the processes \( a_{tj} \) has strictly positive
spectral density $f_{a_j}(\omega)$ for all $\omega \in [-\pi, \pi]$. Since $\min_{1 \leq i \leq s} \{c_i\} = 0$, this final condition will hold if for all $i = 1, 2, \ldots, s$, $\phi_{x_i}(\lambda)$ and $\theta_{x_i}(\lambda)$ have all their roots outside the unit circle; in fact, it is sufficient simply to require that $\phi_{x_{i_0}}(\lambda)$ and $\theta_{x_{i_0}}(\lambda)$ have all their roots outside unit circle for some $i_0$ such that $c_{i_0} = 0$.

Thus, $X_t$ follows an ARIMA($p_i$, $d$, $q_i$) process,

$$\phi_X(B)V^d X_t = \theta_X(B)f_t,$$

say. If the polynomials $\phi_i(B)$ and $\theta_i(B)$, $i = 1, 2, \ldots, s$, have no roots on the unit circle, then the order $d$ is minimal. Moreover, the orders $P_X$ and $Q_X$ are minimal if a greatest common divisor of $\phi_X(B)$ and $\{\phi_i(B) \cdot \theta_i(B), i = 1, 2, \ldots, s\}$ is constant. Since the $\psi_i(B)$ have greatest common divisor equal to unity, this final condition will be satisfied if for all $i$, $\phi_i(B)$ and $\theta_i(B)$ have greatest common divisor equal to unity, i.e. if the orders $(p_i, q_i)$ are minimal for all $i = 1, 2, \ldots, s$.

To establish bounds for $P_X$ and $Q_X$ that are global, i.e. not a function of the coefficient values of $\phi_i(\lambda)$ and $\theta_i(\lambda)$, define the ordered $p_i$ values $p(1) < p(2) < \ldots < p(s)$ and define $d(i)$ and $q(i)$, $i = 1, 2, \ldots, s$ similarly. Note that

$$p(s) < P_X < \sum_{i=1}^{s} p_i;$$

$P_X$ achieves its upper bound if the $\phi_i(\lambda)$ have greatest common divisor
equal to unity; and \( P_X \) achieves its lower bound if for some \( j=1, 2, \ldots, s \), \( \phi_j(\lambda) \) is a least common multiple of \( \{\phi_i(\lambda), i=1, 2, \ldots, s\} \). Also, \( r_1 = P_X - p_1 \), so for all \( i \),

\[
P(s) - P(1) < P(s) - p_i < r_i < \sum_{i=1}^{s} p_j < \sum_{j=2}^{s} p(j) .
\]

Similarly,

\[
0 < c_i = d(s) - d_i < d(s) - d(1)
\]

for all \( i \). Thus,

\[
P(s) - P(1) + q(s) < P(s) - P(1) + d(s) + \max_{1 \leq i \leq s} \{q_i - d_i\} < Q_X
\]

\[
= \max_{1 \leq i \leq s} \{r_1 + q_1 + c_1\}
\]

\[
< \max_{1 \leq i \leq s} \left\{ \left( \sum_{j=1}^{s} p_j \right) + q_i - d_i \right\} + d(s)
\]

\[
< \left\{ \sum_{j=2}^{s} p(j) \right\} + q(s) - d(1) + d(s) .
\]

The inner set of bounds for \( Q_X \) are in general more restrictive than the outer bounds, but the outer bounds are somewhat simpler to evaluate.
In closing, note that the orders $P_X$ and $Q_X$ are in general functions of the coefficients of the component polynomials $\phi_i(B)$ and $\theta_i(B)$, as well as the component orders $(p_i, d_i, q_i)$, $i = 1, 2, \ldots, s$. This interdependence of order and polynomial coefficients becomes even more pronounced in the multivariate case.

5.2.2. **Multivariate case**

The preceding subsection established that the sum of two or more mutually uncorrelated univariate ARIMA processes is itself a univariate ARIMA process, and established minimal orders $(P_X, d, Q_X)$ of the resulting process.

To extend these results to multivariate time series, consider first the case in which $u_t$ follows a $k$-dimensional pure moving average process

$$
u_t' = \theta(B)c_t' = \sum_{i=0}^{q_u} \theta_i c_{t-1}'.$$

Assume that $x_t$ follows an autoregressive moving average process

$$\phi_k(B)x_t' = \theta_k(B)g_t'$$

where

$$\phi_k(B) = \sum_{j=0}^{P_X} \phi_k^j B^j,$$
and

\[ \theta_X(B) = \sum_{i=0}^{q_X} \theta_{X_i} B^i; \]

and that \{c_t\} and \{g_t\} are mutually uncorrelated sequences of mutually uncorrelated \( k \)-dimensional \((0, \Sigma_{c_c})\) and \((0, \Sigma_{g_g})\) random vectors, respectively. Then

\[ a_t' = \phi_X(B) \varepsilon_t' \]

\[ = \theta_X(B) g_t' + \phi_X(B) \varepsilon_t d \theta_u(B) c_t'. \quad (5.14) \]

A repetition of the arguments following (5.13) indicates that the right-hand side of (5.14) has autocovariance function equal to zero at lags greater than \( Q_X = \max(q_X, p_X + d + q_u) \), so by Theorem II.10' of Hannan (1970, p. 66) one may represent \( a_t \) as a moving average process of order \( Q_X \),

\[ a_t' = \sum_{i=0}^{Q_X} \theta_{X_i} f_{t-i}', \]

where \( \{f_t\} \) is a sequence of uncorrelated \((0, \Sigma_{f_f})\) \( 1 \times k \) random vectors, \( |\Sigma_{f_f}| > 0 \), \( \theta_{X_0} = I_k \), and all the roots of the polynomial

\[ |\sum_{i=0}^{Q_X} \theta_{X_i} X_i| \]

fall on or outside the unit circle. Thus, \( x_t \) follows an ARIMA\((p_X, d, Q_X)\) process. Inspection of (5.14) indicates that \( x_t \) and \( x_t' \) have the same autoregressive parameters. Now consider the
right-hand side of (5.14),

\[ a'_t = \theta^{(u)}_x (B) g'_t + \phi^{(d)}_x (B) \varphi^{(d)}_u (B) c'_t \]

\[ = \theta^{(u)}_x (B) f'_t. \quad \text{(5.15)} \]

If \( 0 < \xi < Q_x \), the second part of (5.15) implies that

\[ \text{Cov}(a'_t, a'_{t+\xi}) = \sum_{j=2}^{Q_x} \theta^{(u)}_x, j-\xi \sum_{k} \phi^{(d)}_u (B) c'_t, \phi^{(d)}_u (B) c'_{t+\xi}. \]

while the first part of (5.15) implies that

\[ \text{Cov}(a'_t, a'_{t+\xi}) = \text{Cov}[\theta^{(u)}_x (B) g'_t, \theta^{(u)}_x (B) g'_{t+\xi}] \]

\[ + \text{Cov}[\phi^{(d)}_x (B) \varphi^{(d)}_u (B) c'_t, \phi^{(d)}_x (B) \varphi^{(d)}_u (B) c'_{t+\xi}] \]

\[ = \sum_{j=2}^{Q_x} \theta^{(u)}_x, j-\xi \sum_{k} \phi^{(d)}_u (B) c'_t, \phi^{(d)}_u (B) c'_{t+\xi}. \]

where

\[ M_\xi = \text{Cov}[\phi^{(d)}_x (B) \varphi^{(d)}_u (B) c'_t, \phi^{(d)}_x (B) \varphi^{(d)}_u (B) c'_{t+\xi}] \]

\[ = \{ \sum_{h=0}^{p_x} \sum_{i=0}^{d} \sum_{j=0}^{q_u} \theta_{xh}^{(d)} \theta_{ui}^{(d)} \varphi_{cj}^{(d)} \varphi_{lj}^{(d)} \theta_{kr}^{(d)} \theta_{mr}^{(d)} \} \sum_{s} \text{I}[0 < \xi < p_x + d + q_u] \]
and

\[ S_{hijz} = \{(r, s, w): 0 < r < p_x, 0 < s < d, 0 < w < q_u, r + s + w = h + i + j + z\} . \]

Note that \( M_L \) is not a function of \( S_X \) or \( S_X' \). Since

\[ \text{Cov}(a_t, a_{t+L}) = 0 \text{ for } |t| > Q_X, \]

it follows from the skew-symmetry of the multivariate autocovariance function that the parameters of the \( \{X_t\}, \{x_t\} \) and \( \{u_t\} \) processes satisfy the relations

\[ \phi_{Xj} = \phi_{Xj}, \quad j = 0, 1, \ldots, p_X ; \]

\[ Q_X \]

\[ \sum_{j=0}^{p_X} \theta_X, j - 2 \sum_{j=0}^{p_X} \theta_X' j = \left[ \sum_{j=0}^{p_X} \theta_X, j - 2 \sum_{j=0}^{p_X} \theta_X' j \right] I[\| < q_X] + M_L, \quad l = 0, 1, \ldots, Q_X . \]

(5.16)

The derivation above establishes the existence of a solution \( \{\theta_{Xj}, j = 1, 2, \ldots, Q_X; \Sigma_{Xj} \} \) to relations (5.16). This does not, however, establish the uniqueness of the solution.

If \( \{\theta_{Xj}, j = 1, 2, \ldots, Q_X; \Sigma_{Xj} \} \) are solutions to (5.16), the relations

\[ \phi_{Xj} = \phi_{Xj}, \quad j = 1, 2, \ldots, p_X ; \]

\[ 0 \quad Q_X \quad q_X \quad X_j \]

\[ \sum_{j=0}^{p_X} \theta_X, j - 2 \sum_{j=0}^{p_X} \theta_X, j - 2 \sum_{j=0}^{p_X} \theta_X' j = \left[ \sum_{j=0}^{p_X} \theta_X, j - 2 \sum_{j=0}^{p_X} \theta_X' j \right] I[\| < q_X] + M_L, \quad l = 0, 1, \ldots, Q_X . \]
allow one to express the parameters of the \( \{ x_t \} \) process as functions of the parameters of the \( \{ X_t \} \) and \( \{ u_t \} \) processes. Moreover, given estimates of the parameters of the \( \{ X_t \} \) and \( \{ u_t \} \) processes, an iterative application of (5.17) allows one to obtain associated estimates of the parameters of the \( \{ x_t \} \) process. This is a multivariate generalization of the estimation procedure in Miazaki (1985). Modifications of the iterative application of (5.17) may be required to ensure convergence of the associated sequence of estimates. Also, as noted for (5.16), a solution of (5.17) is not necessarily unique, nor do the roots of the characteristic polynomial of such a solution necessarily fall within the unit circle.

Kashyap and Rao (1976, pp. 29-31) provide a simplified version of the above results for the case \( d = 0, \quad q = 0, \quad p_x = q_x + 1 \).

Arguments similar to those given in Section 5.2.1 allow one to extend the results above to the sum of one ARIMA\( _k \) process and several IMA\( _k \) processes. Let

\[
X_t = \sum_{i=1}^{s} x_{t1},
\]

where \( x_{t1} \) follows an ARIMA\( _k(p_1, d_1, q_1) \) process.
$x_{tj}$ follows an $IMA(d_j, q_j)$ process,

$$(d_j) \quad x'_{tj} = \theta_j(B)g'_{tj} , \quad j=2, 3, \ldots, s ;$$

and the $\{g'_{tj}\}$ are mutually uncorrelated sequences of mutually uncorrelated $(0, \Sigma_j)$ $1 \times k$ random vectors, $j=1, 2, \ldots, s$. Let $d = \max \{d_i\}$ and let $c_i = d - d_i$, $i=1, 2, \ldots, s$. Now consider

$$a'_t = \phi_1(B)V^d x'_t = \sum_{i=1}^{s} a'_{ti}$$

where

$$a'_{t1} = \theta_1(B)g'_{t1}$$

and

$$a'_{ti} = \phi_1(B)\theta_i(B)V^{c_i} g'_{ti} , \quad i=2, 3, \ldots, s.$$ 

Because $\phi_1(B)\theta_i(B)V^{c_i}$ is a polynomial in $B$ of degree $p_1 + c_i + q_i$, the autocovariance function of $a_t$ equals zero at lags greater than

$$Q_X = \max\{q_1 + c_1, p_1 + \max_{2\leq i \leq s} \{c_i + q_i\}\}.$$
Then by Theorem II.10' of Hannan (1970, p. 66), one may represent \( a_t \)
as a \( k \)-dimensional moving average process of order \( Q_x \),

\[
a_t' = \sum_{i=1}^{Q_x} \theta_{xi} f_{t-i},
\]

where \( \{f_t\} \) is a sequence of uncorrelated \( (0, \Sigma_{xf}) 1 \times k \) random vectors, \( \Sigma_{xf} > 0, \theta_{x0} = I_k \), and all the roots of the polynomial \( \Sigma_{x1} \theta_{xi} \) fall on or outside the unit circle. Thus, \( X_t \) follows an ARIMA\(_k\)(\( p_x \), \( d \), \( Q_x \)) model with the same autoregressive parameters as \( x_{1t} \) and with moving average parameters that satisfy the equations similar to (5.16).

Note that if \( Z_s > 0 \) and all the roots of \( \Sigma_{x}^1 \theta_{xi} \) fall outside the unit circle, then the spectral density of \( a_{t1} \) is nonsingular for all \( w \in [-\pi, \pi] \). Similarly, if all the roots of \( |\phi_1(\lambda)| \) fall outside the unit circle; and if for some \( i_0 \in \{2, 3, \ldots, s\}, |\Sigma_{x0}^1| > 0, c_{i_0} = 0 \), and all the roots of

\[
|\Sigma_{x0}^1 \theta_{i0} \lambda^h |
\]

fall outside the unit circle; then the spectral density of \( a_{t0}^1 \) is nonsingular for all \( w \in [-\pi, \pi] \). In either case, the spectral density of \( a_t \) is nonsingular for all \( w \in [-\pi, \pi] \), and thus all the roots of \( \Sigma_{x1} \theta_{xi} \) fall outside the unit circle.

Now consider the more general case in which \( \{u_t\} \) may follow an autoregressive moving average process,
\[ \phi_u(B)u_t = \theta_u(B)c_t', \]

where

\[ \phi_u(B) = \sum_{j=0}^{P_u} \phi_{u_j} B^j. \]

In addition to the definitions and assumptions made previously in this subsection, assume that the roots of the polynomials \( |\phi_\nu(\lambda)| \) and \( |\phi_\nu(\lambda)| \) all fall outside the unit circle, so that \( \phi_\nu(B) \) and \( \phi_u(B) \) are invertible. Also, assume that \( d = 0 \); the results below extend readily to the case of nonzero \( d \) in a manner analogous to the results above for vector moving average \( u_t \), but this requires additional notational complexity without offering much additional insight into the problem.

In the univariate case, derivation of the autoregressive order of \( X_t \) relied heavily on the commutativity of the polynomials \( \phi_\nu(B) \), \( \psi^d = (1 - B)^d \), and \( \phi_u(B) \). Such commutativity does not hold for the general multivariate case, because

\begin{align*}
\phi_u(B)\phi_\nu(B) &= \left( \sum_{i=0}^{P_u} \phi_{u_i} B^i \right) \left( \sum_{j=0}^{P_\nu} \phi_{\nu_j} B^j \right) \\
&= \sum_{i=0}^{P_u} \sum_{j=0}^{P_\nu} \phi_{u_i} \phi_{\nu_j} B^{i+j} \\
&= \sum_{i=0}^{P_u} \sum_{j=0}^{P_\nu} \phi_{u_i} \phi_{\nu_j} B^{i+j} \\
&= \sum_{i=0}^{P_u} \sum_{j=0}^{P_\nu} \phi_{u_i} \phi_{\nu_j} B^{i+j}.
\end{align*}

and similarly

\[ \phi_\nu(B)\phi_u(B) = \sum_{i=0}^{P_\nu} \sum_{j=0}^{P_u} \phi_{\nu_i} \phi_{u_j} B^{i+j}. \]
If \( \phi_{ui} \) and \( \phi_{xj} \) commute for all \( i \) and \( j \) (e.g., if the \( \phi_{ui} \) and \( \phi_{xj} \) are all diagonal), then \( \phi_{ui}(B) \), \( \phi_{x}(B) \) and \( \nu \) all commute and the Box-Jenkins result holds for multivariate time series provided that the white noise series \( \varepsilon_t \) and \( c_t \) have the same dimensionality as the observations \( X_t \). In general, however, \( \phi_{ui} \) and \( \phi_{xj} \) need not commute, so that \( \phi_{ui}(B)\phi_{x}(B) \) need not equal \( \phi_{x}(B)\phi_{ui}(B) \) and the Box-Jenkins proof does not extend directly to multivariate time series.

Instead, one may note that the commutativity of univariate \( \phi_{x}(\lambda) \) and \( \phi_{u}(\lambda) \) was not of interest in itself, but only because it allowed one to find a polynomial \( \phi_{x}(\lambda) \) of minimal order such that \( \phi_{x}(\lambda) \) and \( \phi_{u}(\lambda) \) were each right factors of \( \phi_{x}(\lambda) \). Now consider the multivariate case. Recall from Section 2.2 that a \( k \times k \) matrix polynomial

\[
\phi(\lambda) = \sum_{j=0}^{p} \phi_j \lambda^j
\]

is called a comonic polynomial if \( \phi_0 = I_k \). Thus, \( \phi_{x}(\lambda) \) and \( \phi_{u}(\lambda) \) are comonic polynomials because \( \phi_{x}0 = I_k \) and \( \phi_{u}0 = I_k \). Theorems 9.8 and 9.10 of Gohberg et al. (1982) establish the existence of a comonic polynomial \( \phi_{x}(\lambda) \) of finite minimal degree \( P_x \), say, which is a least common right multiple of the comonic polynomials \( \phi_{x}(\lambda) \) and \( \phi_{u}(\lambda) \).

Given such a \( \phi_{x}(\lambda) \), there exist comonic polynomials

\[
\psi_{x}(\lambda) = \sum_{j=0}^{R} \psi_{xj} \lambda^j \quad \text{and} \quad \psi_{u}(\lambda) = \sum_{j=0}^{R} \psi_{uj} \lambda^j
\]

such that

\[
\phi_{x}(\lambda) = \psi_{x}(\lambda)\phi_{x}(\lambda) = \psi_{u}(\lambda)\phi_{u}(\lambda) ,
\]

(5.18)
where \( R_x = P_x - p_x \) and \( R_u = P_x - p_u \).

The product

\[
\psi_x(B)\phi_x(B) = (\sum_{i=0}^{R_x} \psi_i B^i)(\sum_{j=0}^{P_x} \phi_j B^j)
\]

\[
= \sum_{h=0}^{P_x} \left[ \min(p_x, h) \right] \psi_{x,h-j} \phi_{x,h-j} b^h,
\]

(5.19)

and similarly for \( \psi_u(B)\phi_u(B) \). Thus, \( \psi_x(B) \) and \( \psi_u(B) \) satisfy (5.18) if and only if

\[
\min(p_x, h) \sum_{j = \max(0, h - R_x)}^{\min(p_x, h)} \psi_{x,j} \phi_{x,j} = \sum_{j = \max(0, h - R_u)}^{\min(p_u, h)} \psi_{u,j} \phi_{u,j}
\]

(5.20)

for all \( h = 0, 1, \ldots, P_x \), in which case \( \phi_{x,h} \) is equal to the \( h \)-th entry of (5.20), \( h = 0, 1, \ldots, P_x \). Section 5.3 below establishes that the minimal degree \( P_x \) of \( \phi_x(\lambda) \) satisfies

\[ P_x < k \cdot \max\{p_x, p_u, q_x + 1, q_u + 1\} \].

In general, however, \( P_x \) is a function of the coefficients of \( \phi_x(\lambda) \) and \( \phi_u(\lambda) \) as indicated by Theorems 9.8 and 9.10 of Gohberg et al. (1982) or by the following argument.

Let \( \psi_{\alpha} = [\psi_{\alpha 0}, \psi_{\alpha 1}, \ldots, \psi_{\alpha R_x}]' \) and let \( \phi_{\alpha} \) be a \([(P_x + 1)k] \times [(R_x + 1)k] \) matrix with \( h \)-th \( k \times (R_x + 1)k \) block equal to:
Define \( \tilde{\phi}_x \) and \( \tilde{\phi}_u \) similarly. Then one may rewrite (5.20) as

\[
\tilde{\phi}_x \tilde{\psi}_x' = \tilde{\phi}_x' \tilde{\psi}_x'' \quad \text{(5.22)}
\]

Let \( 1\tilde{\phi}_x \) denote the lower-left \( P_k \times k \) block of \( \tilde{\phi}_x \), let \( 2\tilde{\phi}_x \) denote the lower right \( P_k \times R_k \) block of \( \tilde{\phi}_x \) and define \( 1\tilde{\phi}_u \) and \( 2\tilde{\phi}_u \) similarly. Let \( 2\tilde{\psi}_x = (\tilde{\psi}_x 1, \tilde{\psi}_x 2, ..., \tilde{\psi}_x R_k) \) and define \( 2\tilde{\psi}_u \) similarly. Then under the constraint that \( \phi_x(B) \), \( \tilde{\psi}_x(B) \), \( \phi_u(B) \) and \( \tilde{\psi}_u(B) \) are comonic polynomials, equation (5.22) is equivalent to

\[
\begin{bmatrix}
2\tilde{\phi}_x' \\
-2\tilde{\psi}_x'
\end{bmatrix} = \begin{bmatrix}
1\tilde{\phi}_x' & -1\tilde{\phi}_u'
\end{bmatrix},
\]

i.e.
\[
\begin{bmatrix}
I_k & 0 & \ldots & 0 & | & I_k & 0 & \ldots & 0 \\
\tilde{\phi}'_{x1} & I_k & \ldots & 0 & | & \tilde{\phi}'_{u1} & I_k & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & I_k & | & \vdots & \vdots & \ddots & I_k \\
0 & \tilde{\phi}'_{x_{p_x}} & \ldots & \tilde{\phi}'_{x_{p_x}} & | & 0 & \tilde{\phi}'_{u_{p_x}} & \ldots & \tilde{\phi}'_{u_{p_x}} \\
0 & 0 & \ldots & \tilde{\phi}'_{x_{p_x}} & | & 0 & 0 & \ldots & \tilde{\phi}'_{u_{p_x}} \\
\end{bmatrix}
\]

= \begin{bmatrix}
\psi'_{x1} \\
\psi'_{x2} \\
\vdots \\
\psi'_{x_{p_x}} \\
\psi'_{u_{p_x}} \\
\end{bmatrix}

where \( p = \max(p_x, p_u) \), \( \tilde{\phi}_x = 0 \) for \( j > p_x \) and \( \tilde{\phi}_u = 0 \) for \( j > p_u \). By standard linear model theory, equation (5.23) has a solution \( [\tilde{\phi}_x, \tilde{\phi}_u]' \) if and only if

\[
C[1, \tilde{\phi}_x] \subset C[2, \tilde{\phi}_u],
\]

(5.24)
where \( C[W] \) denotes the column space of the matrix \( W \). Note that \([2\phi_x', -2\phi_u']\) is \( P_k \times (2P_X - P_X - P_u) \), \([2\psi_x', -2\psi_u']\) is \((2P_X - P_X - P_u)k \times k\), and \([1\phi_x', -1\phi_u']\) is \( P_Xk \times k\). The existence of a finite minimal degree \( P_X \) of a least common right multiple of \( \phi_x(\lambda) \) and \( \phi_u(\lambda) \) implies that the containment relation (5.24) is satisfied for sufficiently large \( P_X \); indeed, the development above establishes that one may equivalently define \( P_X \) to be the smallest integer such that (5.24) is satisfied.

Now consider the moving average part of the \( X_t \) process. Given a comonic least common multiple \( \hat{\phi}_x(B) \) of \( \phi_x(B) \) and \( \phi_u(B) \) and associated comonic polynomials \( \hat{\psi}_x(B) \) and \( \hat{\psi}_u(B) \), note that

\[
\mathbf{a}_t' = \hat{\phi}_x(B) \mathbf{x}_t' \\
= \hat{\psi}_x(B) \hat{\phi}_x(B) \mathbf{x}_t' + \hat{\psi}_u(B) \hat{\phi}_u(B) \mathbf{u}_t' \\
= \hat{\psi}_x(B) \theta_x(B) \mathbf{g}_t' + \hat{\psi}_u(B) \theta_u(B) \mathbf{e}_t' .
\]

The right-hand side of this expression has autocovariance function equal to zero at lags greater than \( Q_x = \max\{R_x + q_x, R_u + q_u\} \), so by Theorem II.10' of Hannan (1970, p. 66) one may write

\[
\mathbf{a}_t' = \theta_x(B) \mathbf{f}_t' \\
\]

where \( \theta_x(B) = \sum_{l=0}^{Q_x} \theta_{xl} B^l \), \( \theta_{x0} = I_k \), \( \{f_t'\} \) is a sequence of
uncorrelated \((0, \Sigma_{\xi^f})\) random vectors, \(|\Sigma_{\xi^f}| > 0\), and the polynomial \(\theta_X(\lambda)\) has all its roots on or outside the unit circle. Arguments similar to those used for the case of vector moving average \(u_t\) establish that the covariance structure of \(a_t\) is characterized by the matrices

\[
\text{Cov}(a_t', a_{t+\ell}) = L_{\ell}
\]

\[
= M_{\ell} + N_{\ell}, \quad 0 < \ell < Q,
\]

where

\[
L_{\ell} = \sum_{j=0}^{Q_X} \theta_{X,j} \times \xi^f \theta_{X,j}',
\]

\[
M_{\ell} = \text{Cov}\left[\psi_X(B) \theta_X(B) g_{t} ', \psi_X(B) \theta_X(B) g_{t+\ell}'\right]
\]

\[
= \left[ \sum_{i=0}^{r_X} \sum_{j=0}^{q_X-i} \psi_{X,i} \theta_{X,j} \xi^g \theta_{X,n-i} \psi_{X,n} \right] I[\ell < q_X + r_X],
\]

\[
S_{ij \ell} = \{(m, n) : 0 < m < r_X, 0 < n < q_X, m + n = i + j + \ell\},
\]

\[
N_{\ell} = \text{Cov}\left[\psi_u(B) \theta_u(B) c_{t} ', \psi_u(B) \theta_u(B) c_{t+\ell}'\right]
\]

\[
= \left[ \sum_{i=0}^{r_u} \sum_{j=0}^{q_u-i} \psi_{u,i} \theta_{u,j} \xi^c \theta_{u,n-i} \psi_{u,n} \right] I[\ell < q_u + r_u],
\]

\[
T_{ij \ell} = \{(m, n) : 0 < m < r_u, 0 < n < q_u, m + n = i + j + \ell\}.
\]
If one writes \( L_\mathbf{X} \) in greater detail, (5.25) becomes

\[
\sum_{\ell \in \mathcal{F}} \sum_{X_0} \theta_{X_0} = M_\mathbf{X} + N_\mathbf{X} 
\]

The derivation above establishes that a solution \( \{ \theta_{X_j}, j = 1, 2, \ldots, Q_\mathbf{X} \} \), \( \sum_{\ell \in \mathcal{F}} \) of (5.25) exists. Equations (5.26) indicate that the system is not linear in these parameters, but the same equations suggest a possible iterative procedure to obtain such a solution. Of course, this solution need not be unique, and modifications in the iterative procedure may be required to ensure convergence of the associated sequence of estimates.

Finally, consider the spectral density of \( \mathbf{a}_t \), which has the form

\[
f_a(\omega) = \sum_{h=0}^{Q_\mathbf{X}} \omega_{X_h} \exp(i\omega h) \sum_{\ell \in \mathcal{F}} \sum_{X_0} \theta_{X_0} \exp(-i\omega h)
\]

by Theorem 4.4.1 of Fuller (1976, p. 165). Let

\[
\mathbf{a}_t' = \mathbf{a}_t - \sum_{X_0} \mathbf{X}_h \theta_{X_h} \mathbf{g}_t
\]
which have spectral densities

\[
\begin{align*}
R_x & \quad q_x \\
F_1(w) & = \left[ \sum_{j} \psi_j \exp(iwj) \right] \left[ \sum_{h} \varphi_h \exp(iwh) \right] \sum_{gg} \\
& \times \left[ \sum_{h} \varphi'_{-h} \exp(-iwh) \right] \left[ \sum_{j} \psi'_{-j} \exp(-iwj) \right] \\
R_u & \quad q_u
\end{align*}
\]

and

\[
\begin{align*}
R_x & \quad q_x \\
F_2(w) & = \left[ \sum_{j} \psi_j \exp(iwj) \right] \left[ \sum_{h} \varphi_h \exp(iwh) \right] \sum_{cc} \\
& \times \left[ \sum_{h} \varphi'_{-h} \exp(-iwh) \right] \left[ \sum_{j} \psi'_{-j} \exp(-iwj) \right],
\end{align*}
\]

respectively. The inequalities \(|F_1(w)| > 0\) and \(|F_2(w)| > 0\) for all \(w \in [-\pi, \pi]\) imply that \(|F_a(w)| > 0\) for all \(w \in [-\pi, \pi]\) under either of the following sets of conditions:

(a) \(|\Sigma_{gg}| > 0\), and all the roots of \(|\Sigma_{j=0}^{R_x} \psi_j \lambda_j^j| = 0\) and \(|\Sigma_{h=0}^{q_x} \varphi_h \lambda_h^h| = 0\) fall outside the unit circle; or

(b) \(|\Sigma_{cc}| > 0\) and all the roots of \(|\Sigma_{j=0}^{R_u} \psi_j \lambda_j^j| = 0\) and \(|\Sigma_{h=0}^{q_u} \varphi_h \lambda_h^h| = 0\) fall outside the unit circle.

In either case it then follows that all roots of the polynomial \(|\Sigma_{h=0}^{R} \varphi_h \lambda_h^h| \) fall outside the unit circle.
Example 5.1. To illustrate some of the ideas presented above, consider
the special case \( p = p = q = q = 1 \) for general \( k \).

First let \( R = R = 1 \). Then \( P = 2 \) and the nontrivial parts of
(5.20) become

\[
\psi'_x + \psi'_x = \psi'_u + \psi'_u
\]

and

\[
\psi'_x \psi'_x = \psi'_u \psi'_u.
\]

This system is equivalent to

\[
\psi'_x = \psi'_u + \psi'_u - \psi'_x
\]

and

\[
(\psi'_x - \psi'_u) \psi'_u = \psi'_x (\psi'_x - \psi'_u), \quad (5.27)
\]

which has a solution \((\psi'_x, \psi'_x)\) if and only if

\[
C[\psi'_x (\psi'_x - \psi'_u)] \subseteq C[\psi'_x - \psi'_u]. \quad (5.28)
\]

Under condition (5.28),

\[
\hat{\psi}'_u = (\psi'_x - \psi'_u) \psi'_x (\psi'_x - \psi'_u).
\]
and

\[ \hat{\phi}_x = (\phi_x - \phi_u) \phi_x (\phi_x - \phi_u) \]

provide a solution to (5.27), where \( A^\dagger \) denotes the Moore-Penrose inverse of the matrix \( A \). In this case, \( x_t = x_t + u_t \) follows an autoregressive moving average process of order \( (2, 2) \),

\[ \phi_x(B)x_t = \theta_x(B)f_t, \]

with parameters that satisfy the equations

\[ \hat{\phi}_x = \hat{\phi}_u + \hat{\phi}_u \]

\[ = (\phi_x - \phi_u) \phi_x (\phi_x - \phi_u) + \phi_u \]

\[ \hat{\phi}_x = \hat{\phi}_u \hat{\phi}_u \]

\[ = \hat{\phi}_u (\phi_x - \phi_u) \phi_x (\phi_x - \phi_u); \]

\[ \Sigma_{X} \theta'_{X} = M_2 + N_2 \]

\[ = \Sigma_{gg} \theta'_{X} \hat{\phi}_x + \Sigma_{cc} \theta'_{u} \hat{\phi}_u; \]

\[ \Sigma_{X} \theta'_{X} = - \Sigma_{X} \Sigma \theta'_{X} + M_1 + N_1 \]
If condition (5.28) does not hold, consider the case $R_x = R_u = 2$. Then $P_x = 3$ and the nontrivial parts of expression (5.20) become,

$$\psi'_{x1} + \hat{\psi}'_{x1} = \psi'_{u1} + \hat{\psi}'_{u1},$$

$$\psi'_{x2} + \hat{\psi}'_{x1} \psi'_{x1} = \psi'_{u2} + \hat{\psi}'_{u1} \hat{\psi}'_{u1},$$

and

$$\hat{\psi}'_{x1} \psi'_{x2} = \hat{\psi}'_{u1} \psi'_{u2};$$

or equivalently,

$$\psi'_{x1} = \psi'_{u1} + \hat{\psi}'_{u1} - \hat{\psi}'_{x1},$$

$$\psi'_{x2} = \psi'_{u2} + \hat{\psi}'_{u1} \hat{\psi}'_{u1} - \hat{\psi}'_{x1} (\psi'_{u1} + \hat{\psi}'_{u1} - \hat{\psi}'_{x1}).$$
This linear system has a solution if and only if

\[ C[(\phi'_{x1} - \phi'_{u1})^2(\phi'_{x1} - \phi'_{u1})] \subset C[\phi'_{x1} - \phi'_{u1}, \phi'_{x1} - \phi'_{u1}] \tag{5.30} \]

Note that \( \phi'_{x1}(\phi'_{x1} - \phi'_{u1}) = (\phi'_{x1} - \phi'_{u1})A \) implies that

\[ (\phi'_{x1})^2(\phi'_{x1} - \phi'_{u1}) = \phi'_{x1}(\phi'_{x1} - \phi'_{u1})A \], so condition (5.28) is a special case of condition (5.30).

Under condition (5.30),

\[ [\hat{\psi}_{u1}, \hat{\psi}_{u2}]' = [\phi'_{x1}(\phi'_{x1} - \phi'_{u1}), \phi'_{x1} - \phi'_{u1}]^\top(\phi'_{x1} - \phi'_{u1}) \]

provides a solution to (5.29). In this case, \( X_c \) follows an ARMA(3, 3) model with parameters that satisfy the equations

\[ \phi'_{x1} = \hat{\psi}_{u1} + \phi'_{u1} ; \]

\[ \phi'_{x2} = \hat{\psi}_{u2} + \phi'_{u1}\hat{\psi}_{u1} ; \]

\[ \sum_{\delta \xi \kappa} \theta'_{x3} = M_3 + N_3 \]

\[ = \sum_{\delta \gamma \xi \kappa} \theta'_{x1}\hat{\psi}_{x2} + \sum_{\delta \phi \xi \kappa} \theta'_{u1}\hat{\psi}_{u2} ; \]
\[ \Sigma_{\xi f} \xi^f \psi^t = - \theta^1 \Sigma_{\xi f} \xi^f \xi^t \psi^t + M_2 + N_2 \]

\[ = - \theta^1 \Sigma_{\xi 1} \xi^f \xi^t \xi^3 + \Sigma_{\xi 2} \psi^t \xi^1 + \theta^1 \Sigma_{\xi 1} \psi^t \xi^1 + \Sigma_{\xi 2} \theta^1 \Sigma_{\xi 1} \psi^2 \]

\[ + \Sigma_{\xi c c} (\psi^t \xi^1 + \theta^1 \psi^1) + (\psi^1 \xi^1 + \theta^1) \Sigma_{\xi c} \theta^1 \psi^1 \]

\[ \Sigma_{\xi f} \xi^f \xi^1 = - \theta^1 \Sigma_{\xi 1} \xi^f \xi^2 - \theta^1 \Sigma_{\xi 2} \xi^f \xi^3 + M_1 + N_1 \]

\[ = - \theta^1 \Sigma_{\xi 1} \xi^f \xi^2 - \theta^1 \Sigma_{\xi 2} \xi^f \xi^3 \]

\[ + \Sigma_{\xi 2} (\psi^2 \xi^1 + \theta^1 \xi^1) + (\psi^1 \xi^1 + \theta^1) \Sigma_{\xi 2} (\psi^2 \xi^1 + \theta^1 \xi^1) \]

\[ + (\psi^2 \xi^1 + \psi^1 \xi^1) \Sigma_{\xi c c} \theta^1 \psi^1 + \Sigma_{\xi c c} (\psi^1 \xi^1 + \theta^1) \]

\[ + (\psi^1 \xi^1 + \theta^1) \Sigma_{\xi c c} (\psi^2 \xi^1 + \theta^1 \xi^1) + (\psi^2 \xi^1 + \psi^1 \xi^1) \Sigma_{\xi c c} \theta^1 \psi^1 ; \]

and

\[ \Sigma_{\xi f} = - \theta^1 \Sigma_{\xi 1} \xi^f \xi^1 - \theta^1 \Sigma_{\xi 2} \xi^f \xi^2 - \theta^1 \Sigma_{\xi 3} \xi^f \xi^3 + M_0 + N_0 \]
This subsection and example indicate that one may extend the approach of Box and Jenkins (1976, Appendix A.4.4) to multivariate "signal plus noise" models. However, the autoregressive and moving average orders of the resulting observations are dependent on both the orders and coefficients of the underlying component processes. The following section establishes that the multivariate autocorrelated measurement error model (2.3)-(2.4) can also be characterized as an ARMA model; however, this result is obtained more easily through some state-space arguments of Akaike (1974) rather than through the direct extension of the Box-Jenkins derivation suggested above.


The preceding section noted some practical limitations on autoregressive integrated moving average models for multivariate "signal
plus noise" processes. The remainder of this chapter suggests that state-space models provide a flexible alternative approach to such time series. Subsection 5.3.1 presents some general definitions and notation associated with state-space models, and notes the relationship between state-space models and autoregressive moving average models. This subsection also notes that under the restrictive assumption of known transition equation parameters, one may use standard multiple ARMA time series software to estimate the parameters of a state-space measurement matrix. Subsection 5.3.2 outlines a procedure to use PROC STATESPACE of SAS to estimate the regression coefficients of a standard measurement error model under the restrictive assumption of known \( \mathbf{x}_t \) and \( \varepsilon_t \) process parameters. The restrictive nature of the estimation procedures outlined in Section 5.3.2 indicates that current standard multiple time series software has limited practical value in estimation of the structural model (2.3)-(2.4). Under considerably less restrictive conditions, Section 5.4 outlines a state-space approach to the maximum likelihood estimation of regression coefficients and autocovariance parameters for the structural model.

5.3.1. General state-space models

Before reviewing the results of Akaike (1974), it is necessary to introduce some notation associated with state-space models and Kalman filters. For a more thorough review of state-space models, see Kalman (1960), Kalman and Bucy (1961), Harrison and Stevens (1971, 1976), Sallas and Harville (1981), Meinhold and Singpurwalla (1983), Harvey
A simple state-space model arises from two equations, a "measurement equation" and a "transition equation". The "measurement equation" may be written

\[ Z_t' = B_t W_t' , \quad t=1, 2, \ldots, T , \quad (5.31) \]

where \( Z_t' \) is a \( p \times 1 \) vector of observations, \( B_t \) is a fixed \( p \times K \) matrix, and \( W_t' \) is a \( K \times 1 \) state vector governed by the "transition equation" below. Hence, each element of \( Z_t' \) is a linear combination of several "unobserved components" contained in the state vector \( W_t' \). In the measurement error case, some of the unobserved components of \( W_t' \) are associated with a "true value" or "signal" of interest, while other components are associated with "measurement error" or "noise" terms.

Some state-space models add a separate sequence of uncorrelated errors to the right-hand side of expression (5.31). It would be possible to parameterize measurement error models in this way. However, given serially correlated measurement errors, notational convenience suggests that one absorb the errors into the state vector \( W_t' \). The "transition equation" may be written

\[ W_{t+1}' = A_t W_t' + C_t g_t' , \quad t=1, 2, \ldots, T , \quad (5.32) \]

where \( A_t \) and \( C_t \) are fixed \( K \times K \) and \( K \times m \) matrices,
respectively, and \( g_t \) is an \( m \)-dimensional sequence of uncorrelated 
\((0, S_{ggt})\) random vectors. It is assumed that an initial \( K \)-dimensional 
state vector \( W_0 \) is distributed with mean zero and covariance matrix 
\( \Sigma_{W_0} \) and is uncorrelated with the \( g_t \) sequence.

Assume for the time being that the matrices \( \Sigma_{W_0}, \Sigma_{ggt}, A_t, B_t, C_t, t=1, 2, \ldots, T \), are known. Let \( \hat{W}_t \) denote the minimum 
mean squared error linear predictor of \( W_t \) based on observations 
\( Z_1, Z_2, \ldots, Z_t \), where the term "linear" here means "linear in the 
observations \( Z_i \)." Also, let \( P_t = \text{Var}(\hat{W}_t - W_t) \). Then the minimum 
mean squared error linear predictor of \( W_{t+1} \) based on observations 
\( \{Z_i, i=1, 2, \ldots, t\} \) is

\[
\hat{W}_{t+1|t} = A_t \hat{W}_t.
\]  
(5.33)

Let

\[
P_{t+1|t} = \text{Var}(\hat{W}_{t+1|t} - W_{t+1})
\]

\[
= A_t P_t A'_t + C_t \Sigma_{ggt} C'_t.
\]  
(5.34)

Equations (5.33) and (5.34) are called the "prediction equations" of the 
Kalman filter. The Kalman "updating equations" allow one to "update" 
the prediction equations with an additional observation \( Z_{t+1} \):

\[
\hat{W}_{t+1} = \hat{W}_{t+1|t} + P_{t+1|t} B_{t+1|t}^{-1} (Z_{t+1} - B_{t+1} \hat{W}_{t+1|t})
\]  
(5.35)
and

\[ P_{t+1} = P_{t+1}^t - P_{t+1}^t B_{t+1}^t A_{t+1}^{-1} B_{t+1}^t P_{t+1}^t \]

where

\[ A_{t+1} = B_{t+1} P_{t+1}^t B_{t+1}^t \]

is the covariance matrix of the "innovation", \( d_{t+1} \), at time \( t + 1 \), and

\[ d_{t+1} = Z_{t+1}^t - Z_{t+1}^t \]

\[ = Z_{t+1}^t - B_{t+1}^t \hat{W}_{t+1}^t \]

\[ = B_{t+1}^t \hat{W}_{t+1}^t - B_{t+1}^t \hat{W}_{t+1}^t \]

\[ = B_{t+1}^t (\hat{W}_{t+1} - \hat{W}_{t+1}^t) . \]

Akaike (1974) has noted that any stationary multivariate autoregressive moving average model may be written in state-space form. Let \( x_t^k \) follow an \( \text{ARMA}_k(p, q) \) model,

\[ \phi(B)x_t^k = \psi(B)g_t^k \]

(5.36)

where \( g_t^k \) is a \( k \)-dimensional \( (0, \Sigma_g) \) white noise sequence, \( \phi(B) \)

and \( \psi(B) \) are \( k \)-dimensional polynomials of order \( p \) and \( q \),
respectively, and the roots of the characteristic polynomials

\[ | \sum_{j=0}^{p} \phi_j x^j | = 0 \quad \text{and} \quad | \sum_{i=0}^{q} \theta_i x^i | = 0 \]

all fall outside the unit circle. By the Wold decomposition, \( x_t \) may be written as infinite moving average,

\[ x_t^* = \sum_{i=0}^{\infty} \psi_i g_{t+i} \]

where \( \psi_0 = I_k \), \( \psi_i = 0 \) for \( i < 0 \), and, for nonsingular \( \Sigma_{gg} \), \( \psi_i \) is defined by the recursive relation,

\[ \psi_i = \theta_i - \sum_{j=1}^{p} \phi_j \psi_{i-j} \]  \hspace{1cm} (5.37)

For singular \( \Sigma_{gg} \), the \( \psi_i \) are not defined uniquely, but the matrices defined by (5.37) satisfy the Wold decomposition. Let \( x_{t+n}^*|t \) represent the minimum mean squared error \( n \)-step ahead linear predictor of \( x_{t+n}^* \), and define \( g_{t+n}|t \) similarly, where in both cases, "linear" refers to linearity in the observations \( x_t^* \). Note that \( g_{t+n}|t = 0 \) for \( n > 0 \) and \( x_{t+n}^*|t = x_{t+n}^* \) for \( n < 0 \). Then (5.36) implies that

\[ \sum_{j=0}^{p} \phi_j x_{t+n-j}^*|t = \sum_{i=0}^{q} \theta_i g_{t+n-i}|t \]
which in turn implies that

\[ x^*_{t+n} | t = - \sum_{j=0}^{p} \phi_j x^*_{t+n-j} | t \]

for \( n > M \equiv \max(p, q+1) \). Also, the Wold decomposition implies that, given an infinite "past history",

\[ x^*_{t+n} | t = \sum_{j=n}^{\infty} \phi_j x^*_{t+n-j} \quad \text{and} \]

\[ x^*_{t+n} | t+1 = \sum_{j=n-1}^{\infty} \phi_j g^*_{t+(n-1)+1-j} \]

\[ = \psi_{n-1} g^*_{t+1} + \sum_{j=n}^{\infty} \phi_j g^*_{t+n-j} \]

\[ = \psi_{n-1} g^*_{t+1} + x^*_{t+n} | t \].

Therefore, if one defines

\[ W_t = (x^*_{t} | t, x^*_{t+1} | t, \ldots, x^*_{t+M-1} | t) \]

and \( \phi_j = 0 \) for \( j > p \), one may write

\[ x^*_{t} = EW'_{t} \quad (5.38) \]

and

\[ W'_{t+1} = AW'_{t} + CG'_{t} \]
as the measurement and transition equations, respectively, of a state-space model, where $B$, $A$ and $C$ are, respectively, matrices of dimensions $k \times M_k$, $M_k \times M_k$, and $M_k \times k$ such that

$$B = [I_k, 0_{k \times (M-1)k}],$$

$$A = \begin{pmatrix} 0 & I_k & 0 & 0 & \ldots & 0 \\ 0 & 0 & I_k & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & I_k \\ -\phi_M & -\phi_{M-1} & -\phi_{M-2} & -\phi_{M-3} & \ldots & -\phi_1 \end{pmatrix},$$

and

$$C = [\psi_0', \psi_1', \ldots, \psi_{M-2}', \psi_{M-1}'].$$

This is a special case of the "linearization" of a comonic polynomial, as discussed in Gohberg et al. (1982, p. 187). Hence, Akaike (1974) concluded that any $k$-dimensional stationary autoregressive moving average may be written as a $kM$-dimensional state space model.

Conversely, Akaike (1974) also demonstrated that an $L$-dimensional time-homogeneous state-space model with a $p$-dimensional observation vector may be written as an ARMA$_p(L, L-1)$ model. Write the measurement and transition equations as
\[ Z'_t = BW'_t \]

and

\[ W'_{t+1} = AW'_t + Cg'_t, \quad (5.40) \]

respectively, where \( g'_t \) is a \( k \)-dimensional \((0, \Sigma_g)\) white noise sequence. In some cases, \( g_t = Z_{t+1} - Z_{t+1|t} \), the "innovation sequence." Let

\[ p(\lambda) = |\lambda I_L - A| = \sum_{j=0}^{L} a_j \lambda^{L-j}, \quad a_0 = 1, \]

be the characteristic polynomial of the matrix \( A \). The Hamilton-Cayley equation [Halmos (1974), p. 115] implies that

\[ p(A) = \sum_{j=0}^{L} a_j A^{L-j} = 0 \]

Iteration of (5.40) implies that

\[ W'_{t+j} = AW'_{t+j-1} + Cg'_{t+j} \]

\[ = A(AW'_{t+j-2} + Cg'_{t+j-1}) + Cg'_{t+j} \]

\[ = A^2 W'_{t} + A^{j-1} Cg'_{t+1} + \ldots + ACg'_{t+j-1} + Cg'_{t+j} \]

\[ = A^j W'_{t} + \sum_{i=1}^{j} A^{j-i} Cg'_{t+i} . \]
Let

$$Z_j = B(A^j + a_1 A^{j-1} + \ldots + a_{j-1} A + a_j L)C = B(\sum_{i=0}^{j} a_i A^{j-i})C,$$

\(j=1, 2, \ldots, L\),

and note that

$$Z_{t+j} = BW_{t+j} = B(A^j W_t + \sum_{i=1}^{j} A^{j-i} \theta_{t+i})$$

for \(j=1, 2, \ldots, L\). Then

$$\sum_{j=0}^{L} a_{L-j} Z'_{t+j} = a_L Z'_t + \sum_{j=1}^{L} a_{L-j} Z'_{t+j}$$

$$= a_L BW'_t + \sum_{j=1}^{L} a_{L-j} B(A^j W'_t + \sum_{i=1}^{j} A^{j-i} \theta_{t+i})$$

$$= a_L BW'_t + B(\sum_{j=1}^{L} a_{L-j} A^j)W'_t + \sum_{i=1}^{L} a_{L-j} B(A^{j-i} \theta_{t+i})$$

$$= B(a_L L + \sum_{j=1}^{L} a_{L-j} A^j)W'_t + \sum_{i=1}^{L} a_{L-j} \theta_{Z, L-i} \theta_{t+i}$$

$$= \sum_{i=1}^{L} \theta_{Z, L-i} \theta_{t+i},$$

(5.41)

where the last equality follows from the Hamilton-Cayley equation.

Thus, the observed process \(Z_t\) follows a \(p\)-dimensional autoregressive moving average model of order \((L, L-1)\).
Cain (1987) noted that the orders \((L, L-1)\) are not necessarily minimal. Define the minimal polynomial \(m(\lambda) = \sum_{i=0}^{L'} m_i \lambda^{M-i}\) of \(A\) to be a polynomial such that

\[
m(A) = \sum_{i=0}^{L'} m_i A^{M-i} = 0,
\]

\(m_i \in \mathbb{R}\) for all \(i\), \(m_0 = 1\), and no polynomial \(n(\lambda)\) with real scalar coefficients and with degree less than \(L'\) satisfies the equation \(n(A) = 0\). By the Hamilton-Cayley equation, \(L'\) is bounded above by \(L\), while trivially \(L'\) is bounded below by one. Hence, in the argument above, one may then replace the coefficients

\(\{a_i, i=0, 1, ..., L\}\) with \(\{m_i, i=0, 1, ..., L'\}\) and reduce the ARMA order of the \(Z_t\) process accordingly. However, the minimal degree \(L'\) is a function of the entries of \(A\), and it appears that the ARMA orders \((L, L-1)\) are the smallest "global" values attainable, i.e. there exist \(L\)-dimensional time-homogeneous state-space models for which the ARMA orders \((L, L-1)\) are minimal.

Although Akaike (1974) did not state so explicitly, it is clear from the two results above that any linear transformation of a multivariate autoregressive moving average process is itself a multivariate autoregressive moving average process. To see this, let \(x_t^*\) follow an ARMA \((p_x, q_x)\) process and let \(Z_t = B_1 x_t^*\) for some \(p \times k\) real matrix \(B_1\). In (5.40) let \(B = [B_1, 0_{p \times (M-1)k}]\), and let all other symbols be as defined above (5.40). Then \(Z_t\) follows an ARMA \((L, L-1)\) process, where \(L = M \times k\) and \(M = \max(p_x, q_x + 1)\).
Although the closure of multivariate ARMA processes under linear transformations may be somewhat satisfying mathematically, it is of limited value for moderate to large values of \( p \), \( q \), and \( k \), due to the resulting high order \( L \). Nonetheless, this closure property does suggest one way in which one could use standard multivariate time series software to estimate the "measurement matrix" \( B \) if the parameters of the \( \mathbf{x}_t \) process were known, i.e.,

1. Given the parameters of the ARMA\(_r(p, q)\) process \( \mathbf{x}_t \), compute the coefficients \( a_1, a_2, \ldots, a_L \) of the characteristic polynomial \( p(\lambda) \) given below (5.40), and compute \( C \).

2. Fit the observations \( Z_t \) to an ARMA\(_p(L, L-1)\) model, with the \( j \)-th autoregressive parameter constrained to equal \( a_{L-j} I_p \).

3. Given the resulting moving average parameter estimates \( \hat{\theta}_{2j}, j=1, 2, \ldots, L \), use a nonlinear least squares fitting routine to fit the free parameters, \( \hat{\theta}_z \), say, of \( B_1(\hat{\varphi}) \) to the equations

\[
\hat{\theta}_z = B_1(\hat{\varphi})(A_j^{j} + a_1 A_j^{j-1} + \ldots + a_L I_L)C.
\]

This procedure is somewhat similar to the nonlinear fitting procedure suggested by Miazaki (1985, p. 42 ff.) for a simpler univariate problem.

Based on the results of this subsection, one may use either a state-space model or a high-order autoregressive moving average model to describe a linear transformation of a vector autoregressive moving
average process. In particular, this conclusion applies to "signal plus noise" observations for which the mutually uncorrelated "signal" and "noise" components each follow vector autoregressive moving average models. The following subsection develops further state-space representations of the structural measurement error model (2.3)-(2.4).

5.3.2. Application to measurement error models

For model (2.3)-(2.4), assume that $x_t$ follows an ARMA$(p_x, q_x)$ model independent of $\varepsilon_t$, which follows an ARMA$(p_e, q_e)$ model. Then $x_t^* = (x_t, \varepsilon_t)$ follows an ARMA$(p+q, q+q)$ model, where $p_1 = \max(p_x, p_e)$ and $q_1 = \max(q_x, q_e)$. Hence, the observations $Z_t$ may be written

$$Z_t' = (Y_t, X_t)' = [(x_t B + e_t), (x_t + u_t)]'$$

$$= (\beta, I_p)'x_t^* + \varepsilon_t'$$

$$= [(\beta, I_p)', I_p]x_t^*.$$

Then by the comments at the end of the preceding section, $Z_t$ follows an ARMA$(P, K)$ process, where $K = (p + k)\max(p_1, q_1 + 1)$. Moreover, the algorithm presented at the end of Section 5.3.a could be used to estimate the parameters of $\beta$, given the parameters of the $x_t$ and $\varepsilon_t$ processes.
Similarly, the arguments of the preceding section indicate that the ARMA$_p$(K, K) observations $Z_t$ may also be represented in state-space form (5.31)-(5.32) with

$$B_Z = [I_p, O_{p \times Kp}],$$

$$A_Z = \begin{bmatrix}
0 & I_p & 0 & 0 & \cdots & 0 \\
0 & 0 & I_p & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & I_p \\
-\phi_{Z,K} & -\phi_{Z,K-1} & -\phi_{Z,K-2} & -\phi_{Z,K-3} & \cdots & -\phi_{Z,1}
\end{bmatrix},$$

and

$$C_Z = [I_p, \psi_{Z1}', \cdots, \psi_{Z,K-1}', \psi_{ZK}'],$$

where $\{\phi_{Z1}, i=1, 2, \ldots, K\}$ and $\{\theta_{Z1}, i=1, 2, \ldots, K\}$ are the $p \times p$ autoregressive and moving average parameter matrices, respectively, of the $Z_t$ process, and $\{\psi_{Z1}, i=1, 2, \ldots, K\}$ are defined from $\phi_{Z1}$ and $\theta_{Z1}$ by (5.37). The form of $B$ permits one to use PROC STATESPACE to estimate the parameters $\phi_{Z1}$ and $\psi_{Z1}$, and hence $\theta_{Z1} = \sum_{j=0}^{K} \psi_{Z1-j}$ as well. This suggests an estimation procedure for $B$, i.e.,
1. Given the parameters of the ARMA\_p+k(p, q) process x^*, compute the associated matrices A^* and C^* and compute the coefficients a^*_1, a^*_2, ..., a^*_K of the characteristic polynomial p^*(\lambda) of A^*, where K = (p + k) \* \text{max}(p, q + 1).

2. Use PROC STATESPACE to fit the observations Z_t to an ARMA(K, K) model by means of the A_Z and C_Z matrices presented above, with the j-th autoregressive parameter constrained to equal a^*_k-j I_p. Note that the entire matrix A_Z is constrained a priori; only the matrix C_Z, or equivalently, the moving average parameters \hat{\theta}_{zj}, are fit in this step.

3. Given the resulting moving average parameter estimates \hat{\theta}_{zj}, j=1, 2, ..., K, use a least squares fitting routine to fit the free parameters \theta to the equations

\[ \hat{\theta}_{zj} = \left\{ ((\theta, I_k)')', I_p \right\} \left[ (A^*)^j + a_1(A^*)^j-1 + ... + a_j I_k \right] C^*, \]

\[ j=1, 2, ..., K. \] (5.42)

Because A^* and C^* are assumed to be known, the least squares fit of (5.42) simplifies considerably. Partition \hat{\theta}_{zj} = (\hat{\theta}_{z1j}', \hat{\theta}_{z2j}')', where \hat{\theta}_{z1j} is p \times (p + k) and \hat{\theta}_{z2j} is k \times (p + k). Also, define

\[ M_j = (M_{1j}', M_{2j}', M_{3j}', M_{4j}')' \]

\[ = [(A^*)^j + a_1(A^*)^j-1 + ... + a_j I_k] C^*, \]
where $M_{1j}$ is $k \times (p + k)$, $M_{2j}$ is $r \times (p + k)$, $M_{3j}$ is $k \times (p + k)$, $M_{4j}$ is $(K - p - k) \times (p + k)$ and $r = p - k$. Then one may rewrite model (5.42) as

$$\hat{\theta}_{Zj} = [((\beta, I_1)', I_p)(M_{1j}', M_{2j}', M_{3j}')']$$

so the part of model (5.42) that is relevant to estimation of $\beta$ may be rewritten,

$$\hat{\theta}_{Zj} - M_{2j} = \beta'M_{1j} \quad j = 1, 2, \ldots, K.$$

Thus, given "observed" $\hat{\theta}_{Zj}$ and known $M_{1j}$, $M_{2j}$, the estimation of $\beta$ reduces to an exercise in multivariate regression with "dependent variable" $\hat{\theta}_{Zj} - M_{2j}$ and "independent variable" $M_{1j}$, $j = 1, 2, \ldots, K$.

Therefore, one may in principle use PROC STATESPACE and PROC GLM to estimate the parameter $\beta$ of model (2.3)-(2.4), provided that the $x_t$ and $e_t$ follow autoregressive moving average processes with known orders and parameters. In practice, this result will be useful only when $p$, $k$, $p_x$, $q_x$, $p_e$ and $q_e$ are such that the resulting state-space dimension $K$ is small enough to be handled efficiently by PROC STATESPACE. This concern would be particularly acute if the number of observation times, $T$, were relatively small. An additional limitation of the approach outlined above is the requirement that the autoregressive and moving average parameters of the $x_t$ and $e_t$
processes be known a priori; in some practical cases this requirement may be unrealistic. The following section outlines an alternative estimation procedure that requires only enough auxiliary information to identify the parameters of a "signal plus noise" model.

5.4. Maximum Likelihood Estimation for the Structural Model

Through a State-Space Representation

If \( x_t \) and \( \varepsilon_t \) follow autoregressive moving average models, one may write the structural errors-in-variables model (2.3)-(2.4) in a state-space form that is of lower dimension than the state-space forms of Section 5.3. Retain the notation and assumptions of Section 5.3.

Define

\[
\begin{align*}
W_t &= (x_t^*|t, x_{t+1}^*|t, \ldots, x_{t+J-1}^*|t), \\
G_{t-1} &= x_t^*|t - x_t^*|t-1, \\
x_t^* &= (x_t, \varepsilon_t), \\
B_W &= [(g, I_k)', I_p', O_p \times (J-1)(p+k)].
\end{align*}
\]
and \( C_w = [I_{p+k}, \gamma_{1}, \ldots, \gamma_{J-2}, \gamma_{J-1}]' \), where \( J = \max(p, q_1 + 1) \), \( \{\gamma_j, j=1, 2, \ldots, J\} \) and \( \{\theta_j, j=1, 2, \ldots, J-1\} \) are the \((p+k)\times(p+k)\) autoregressive and moving average parameter matrices, respectively, of the \( z^* \) process, \( \gamma_j = 0 \) for \( j > p \), \( \theta_j = 0 \) for \( j > q_1 \), and \( \{\gamma_j, j=0, 1, \ldots, J-1\} \) are defined from \( \gamma_j \) and \( \theta_j \) by (5.37). Then

\[
Z_t' = B_w w_t^* \tag{5.43}
\]

and

\[
w_{t+1}^* = A_w w_t^* + C_w g_t^* \tag{5.44}
\]

constitute the measurement and transition equations, respectively, of a state-space representation of a structural model (2.3)-(2.4). Hence, the estimation of \( \beta \) is a special case of the estimation of the unknown parameters of the measurement equation of a state-space model. Similarly, the estimation of unknown parameters of \( \gamma = [\text{vec}(\beta)', \gamma_x, \gamma_c] \)
constitutes a special case of the estimation of the unknown parameters of a Kalman filter.

Note that the state vector $\hat{w}_t^k$ has dimension $(p + k)J = (p + k)\max(p_1, q_1 + 1)$. By contrast, the state-space representation of a structural model (2.3)-(2.4) in Section 5.3 has a state vector of dimension $p(p + k)\max(p_1, q_1 + 1)$. Since $p$ is always at least two, model (5.43)-(5.44) may be of considerably lower dimension than the corresponding state-space model of Section 5.3. Depending on the availability of computational power and the magnitude of $p$ and $J$, this may imply that model (5.43)-(5.44) is preferable for the sake of computational efficiency. In addition, the use of PROC STATESPACE requires that the first $p$ entries of the state vector $\hat{w}_t^k$ constitute the observation at time $t$; the use of a general measurement coefficient matrix $B_w$ as in (5.43) is not permitted. Therefore, PROC STATESPACE cannot be used directly to fit model (5.43)-(5.44). Instead, one may pursue an alternative maximum-likelihood fitting procedure.

Given the state-space model (5.43)-(5.44), one may write down the associated Kalman filter prediction and updating equations. Because model (5.43)-(5.44) restricts the coefficient matrices of the measurement and transition matrices to be constant over time, the prediction and updating equations for model (5.43)-(5.44) differ slightly from (5.33)-(5.35) above. The prediction equations are now

$$\hat{w}_{t+1|t} = A_w \hat{w}_t$$

(5.45)
and

\[ P_{t+1|t} = \text{Var}(\hat{W}_{t+1|t} - W_{t+1}) = A_t P_{t} A_t' + \Sigma \Sigma_{gg} W_{t+1} \]

where \( \Sigma_{gg} = \text{Var}(g_{t} g_{t}') \); and the updating equations are

\[ \hat{W}_{t+1} = \hat{W}_{t+1|t} + P_{t+1|t} B_t' A_t^{-1} (Z_{t+1} - B_t \hat{W}_{t+1}) \]

and

\[ F_{t+1} = F_{t+1|t} - F_{t+1|t} B_t' A_t^{-1} B_t F_{t+1} \]

where

\[ A_{t+1} = B_t F_{t+1|t} B_t' \]

The innovations

\[ d_{t} = Z_{t+1} - B_t \hat{W}_{t+1} \]

form a sequence of uncorrelated \((0, A_{t})\) random vectors. If vec(x, \( \varepsilon \)) follows a multivariate normal distribution, then the \( d_{t} \) are independent and have joint likelihood function \( L(g; d) \) such that

\[ -2 \ln L(g; d) = Tr \ln(2\pi) = \sum_{t=1}^{T} L_c(g; d_t) \]

where
and, as in Section 5.1, α is a minimal set of parameters that
determines β, β*, θ*, and ξ. Following an idea from Wincek
and Reinsel (1986, Section 5), one may use the iterative structure of
equations (5.46)-(5.49) to obtain iterative expressions for the
derivatives of (5.51) with respect to the unknown elements of α.

By Result 9.4,

\[
\frac{\partial \alpha}{\partial \alpha_h} = \left[ \frac{\partial W}{\partial \alpha_h} \right]_t |t-1 B_t' W + W_t' P_t |t-1 \frac{\partial W}{\partial \alpha_h} |t-1 B_t' W + W_t' \left[ \frac{\partial F}{\partial \alpha_h} |t-1 \right] B_t'.
\]

Also, the relation

\[
\hat{W}_t |t-1 = A_W [\hat{W}_t |t-2 + P_t |t-2 B_t' A_t^{-1} d_t |t-1]
\]

implies that

\[
\frac{\partial \hat{W}_t |t-1}{\partial \alpha_h} = \left( \frac{\partial A_W}{\partial \alpha_h} \right) [\hat{W}_t |t-2 + P_t |t-2 B_t' A_t^{-1} d_t |t-1] + A_W M_{2ht},
\]

where

\[
M_{2ht} = \frac{\partial W_t |t-2}{\partial \alpha_h} + \left( \frac{\partial F_t |t-2}{\partial \alpha_h} \right) B_t' A_t^{-1} d_t |t-1 + P_t |t-2 \left( \frac{\partial W_t |t-1}{\partial \alpha_h} \right) A_t^{-1} d_t |t-1
\]

\[
- P_t |t-2 B_t' A_t^{-1} \left( \frac{\partial \theta}{\partial \alpha_h} \right) A_t^{-1} d_t |t-1 + P_t |t-2 B_t' A_t^{-1} \left( \frac{\partial \xi}{\partial \alpha_h} \right) A_t^{-1} d_t |t-1.
\]
Similarly,

\[
\frac{3p_t|t-1}{3\alpha_h} = \left(\frac{3A_w}{3\alpha_h}\right)p_{t-1}A'_w + A_w(p_{t-1}\left(\frac{3A_w}{3\alpha_h}\right))' + A_w\left(\frac{3p_{t-1}}{3\alpha_h}\right)A'_w \\
+ \left(\frac{3C_w}{3\alpha_h}\right)\Sigma G_w C'_w + C_w(\Sigma G_w)\left(\frac{3C_w}{3\alpha_h}\right)' + C_w(\Sigma G_w)\left(\frac{3G_w}{3\alpha_h}\right)C'_w;
\]

\[
\frac{3p_t}{3\alpha_h} = \left(\frac{3p_t}{3\alpha_h}\right)_{t-1} - \frac{3p_{t-1}}{3\alpha_h}B_w^tB_w^{\dagger-1}B_w^t|t-1 - p_{t-1}B_w^{\dagger-1}B_w^t\left(\frac{3p_{t-1}}{3\alpha_h}\right) \\
- p_{t-1}\left[\frac{3}{3\alpha_h}(B_w^{t-1}B_w^t)p_{t-1}';
\]

\[
\frac{\partial}{\partial \alpha_h}(B_w^{t-1}B_w^t) = \left(\frac{\partial}{\partial \alpha_h}B_w^t\right)^{-1}B_w^t + B_w^{t-1}\left(\frac{\partial}{\partial \alpha_h}B_w^t\right) - B_w^{t-1}\left(\frac{3A_w}{3\alpha_h}\right)A_w^{t-1}B_w^t;
\]

and

\[
\frac{3d_t^t}{3\alpha_h} = -\left(\frac{3B_w}{3\alpha_h}\right)^{\dagger}t - B_w\left(\frac{3B_w|t-1}{3\alpha_h}\right)'.
\]

Second derivative results also follow an iterative pattern. Since

\[
\frac{3^2B_w}{3\alpha_h3\alpha_l} = 0 \quad \text{and} \quad \frac{3^2A_w}{3\alpha_h3\alpha_l} = 0
\]

for any \( h \) and \( i \),
$$\frac{\partial^2 \Delta_c}{\partial \alpha_h \partial \alpha_i} = \left( \frac{\partial B_h}{\partial \alpha_h} \right) \left[ \left( \frac{\partial P_t}{\partial \alpha_i} \right) B_w + \frac{\partial B_w'}{\partial \alpha_i} \right]$$

$$+ \left[ \frac{\partial B_w}{\partial \alpha_i} \right] P_t + \frac{\partial P_t}{\partial \alpha_i} \left( \frac{\partial B_w}{\partial \alpha_h} \right) + \frac{\partial B_w'}{\partial \alpha_h} \left( \frac{\partial P_t}{\partial \alpha_i} \right) B_w'$$

$$+ B_w \frac{\partial P_t}{\partial \alpha_i} \left( \frac{\partial B_w'}{\partial \alpha_h} \right) + B_w \left( \frac{\partial^2 P_t}{\partial \alpha_i \partial \alpha_i} \right) B_w' ;$$

$$\frac{\partial^2 P_t}{\partial \alpha_h \partial \alpha_i} = \left( \frac{\partial A_w}{\partial \alpha_i} \right) \left[ \left( \frac{\partial P_t}{\partial \alpha_h} \right) A_w' + \frac{\partial A_w'}{\partial \alpha_i} \right] + \left[ \left( \frac{\partial A_w}{\partial \alpha_i} \right) P_t - 1 + A_w \left( \frac{\partial P_t}{\partial \alpha_i} \right) \right] \left( \frac{\partial A_w}{\partial \alpha_i} \right) ;$$

$$+ \left( \frac{\partial A_w}{\partial \alpha_i} \right) \left( \frac{\partial P_t}{\partial \alpha_i} \right) A_w' + A_w \left( \frac{\partial P_t}{\partial \alpha_i} \right) A_w' + \frac{\partial^2 P_t}{\partial \alpha_i \partial \alpha_i} A_w'$$

$$+ \left( \frac{\partial^2 C_w}{\partial \alpha_h \partial \alpha_i} \right) \Sigma_{gg} C_w' + C_w \Sigma_{gg} \left( \frac{\partial^2 C_w}{\partial \alpha_h \partial \alpha_i} \right) ;$$

$$+ \left( \frac{\partial C_w}{\partial \alpha_i} \right) \left( \frac{\partial \Sigma_{gg}}{\partial \alpha_h} \right) C_w' + C_w \left( \frac{\partial \Sigma_{gg}}{\partial \alpha_h} \right) C_w' + \left( \frac{\partial C_w}{\partial \alpha_i} \right) \left( \frac{\partial \Sigma_{gg}}{\partial \alpha_i} \right) C_w'$$

$$+ \left( \frac{\partial C_w}{\partial \alpha_i} \right) \Sigma_{gg} C_w' + C_w \left( \frac{\partial \Sigma_{gg}}{\partial \alpha_i} \right) C_w' + \left( \frac{\partial^2 C_w}{\partial \alpha_i \partial \alpha_i} \right) C_w' ;$$
\[ M_{I - V}(\frac{u_{be}}{\frac{V}{e}}) - (\frac{u_{be}}{\frac{V}{e}}) \frac{T_{be}}{e} + (\frac{T_{be}}{e}) \frac{u_{be}}{\frac{V}{e}} \frac{V_{ae}}{e} = \frac{T_{be} u_{be}}{I_{-3} d_{2e}} \]

\[ M_{I - V}(\frac{u_{be}}{\frac{V}{e}}) - (\frac{u_{be}}{\frac{V}{e}}) \frac{T_{be}}{e} + (\frac{T_{be}}{e}) \frac{u_{be}}{\frac{V}{e}} \frac{V_{ae}}{e} \]

\[ [(\frac{u_{be}}{\frac{V}{e}}) \frac{T_{be}}{e} + \frac{M_{I - V}(\frac{u_{be}}{\frac{V}{e}})}{\frac{V}{e}} - (\frac{T_{be}}{e})] = \frac{T_{be} u_{be}}{M_{ae} \frac{V_{ae}}{e} (I_{-3} d_{2e})} \]

\[ I_{-3} \frac{a(M_{I - V}(\frac{u_{be}}{\frac{V}{e}}))}{\frac{V}{e}} - I_{-3} \frac{a}{\frac{V}{e}} \]

\[ (\frac{T_{be}}{e}) \frac{u_{be}}{\frac{V}{e}} \]

\[ [(\frac{u_{be}}{\frac{V}{e}}) \frac{T_{be}}{e} + \frac{M_{I - V}(\frac{u_{be}}{\frac{V}{e}})}{\frac{V}{e}} - (\frac{T_{be}}{e})] = \frac{T_{be} u_{be}}{M_{ae} \frac{V_{ae}}{e} (I_{-3} d_{2e})} \]

\[ I_{-3} \frac{a(M_{I - V}(\frac{u_{be}}{\frac{V}{e}}))}{\frac{V}{e}} - I_{-3} \frac{a}{\frac{V}{e}} \]

\[ \frac{T_{be}}{e} \frac{u_{be}}{\frac{V}{e}} \]

\[ [(\frac{u_{be}}{\frac{V}{e}}) \frac{T_{be}}{e} + \frac{M_{I - V}(\frac{u_{be}}{\frac{V}{e}})}{\frac{V}{e}} - (\frac{T_{be}}{e})] = \frac{T_{be} u_{be}}{M_{ae} \frac{V_{ae}}{e} (I_{-3} d_{2e})} \]
Finally, by Results 9.7 and 9.8,

\[
\frac{\partial^2 \hat{W}_t(t-1)}{\partial a_h \partial \alpha_{\Phi}} = \text{tr}\left[\frac{\partial^2 \hat{\Lambda}}{\partial \alpha_{\Phi}}\right] + 2\left(\frac{\partial \hat{w}_t}{\partial a_h}\right)\hat{d}_t' - d_t \hat{a}_{\Phi}(t) \left(\frac{\partial \alpha_{\Phi}}{\partial a_h}\right)^{-1} d_t'.
\]
and

\[
\frac{\partial^2 \mathcal{L}(\theta; \mathcal{d}_t)}{\partial \theta_h \partial \theta_l} = \text{tr}\left[\left(\frac{\partial^2 \Delta_c}{\partial \theta_h \partial \theta_l}\right)\Delta_c^{-1}\right] - \text{tr}\left[\left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\left(\frac{\partial \Delta_c}{\partial \theta_l}\right)\Delta_c^{-1}\right]
\]

\[+ 2\left[\left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\Delta_c\frac{\partial \theta_t}{\partial \theta_h} - \left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\Delta_c\right]
\]

\[- \left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\Delta_c\frac{\partial \theta_t}{\partial \theta_h} + \left(\frac{\partial \Delta_c}{\partial \theta_l}\right)\Delta_c^{-1}\Delta_c\frac{\partial \theta_t}{\partial \theta_l}\right]
\]

\[\frac{\partial^2 \mathcal{L}(\theta; \mathcal{d}_t)}{\partial \theta_h \partial \theta_l} = d_t\Delta_c^{-1}\left[-\left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\left(\frac{\partial \Delta_c}{\partial \theta_l}\right) - \left(\frac{\partial \Delta_c}{\partial \theta_h}\right)\Delta_c^{-1}\left(\frac{\partial \Delta_c}{\partial \theta_l}\right)\right] + \frac{\partial^2 \Delta_c}{\partial \theta_h \partial \theta_l}\Delta_c^{-1}\Delta_c\frac{\partial \theta_t}{\partial \theta_l}.
\]

Expressions (5.51) and (5.52) and the derivative results developed above may lead to an iterative procedure for the minimization of expression (5.51) with respect to the unknown but identified elements of \( \theta = \{\text{vec}(\beta)\}', \alpha, \gamma \} \). As in Section 5.1, however, this statement is contingent upon the likelihood function \( L(\theta; Z) \) satisfying certain regularity conditions; additional research will be required to develop the details of such conditions. With this contingency in mind, one may consider a specific Newton-Raphson procedure for maximum likelihood estimation of unknown but identified elements of \( \theta \).

Let \( \theta = \{\text{vec}(\beta)' , \alpha , \gamma \} \) be a \( 1 \times L \) vector, where \( L = L_\beta + L_\alpha + L_\gamma \), \( L_\beta = \text{rk} \), \( \alpha \) is a \( 1 \times L_\alpha \) vector of unknown but identified elements of \( \alpha \), and \( \gamma \) is a \( 1 \times L_\gamma \) vector of unknown but identified elements of \( \gamma \). Let \( f, f_1, f_2 \) be \( 1 \times L \) vectors such that
\[ f = f_1 + f_2 ; \]

\( f_1 \) has \( i \)-th element equal to

\[
f_{1i} = \frac{3}{\partial \alpha_i} \left[ \ln |\Gamma_{zz}^i| \right]
= \frac{3}{\partial \alpha_i} \left[ \sum_{t=1}^{T} \ln |\Lambda_t^i| \right]
= \sum_{t=1}^{T} \text{tr} \left[ \Lambda_t^{-1}(\frac{\partial \Lambda_t^i}{\partial \alpha_i}) \right],
\]

for \( i = 1, 2, \ldots, L \). Similarly, define \( F, F_1 \) and \( F_2 \) to be \( L \times L \) matrices such that

\[ F = F_1 + F_2 ; \]
$F_1$ has $(i, j)$-th element equal to

$$F_{1ij} = \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \left[ \ln |\Gamma_{ZZ}| \right]$$

$$= \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \left[ \sum_{t=1}^{T} \ln |\Lambda_t| \right]$$

$$= \sum_{t=1}^{T} \left[ \text{tr} \left( \frac{\partial^2 \Lambda_t}{\partial \alpha_i \partial \alpha_j} \right) \Lambda_t^{-1} \right] - \text{tr} \left[ \left( \frac{\partial \Lambda_t}{\partial \alpha_i} \right) \Lambda_t^{-1} \left( \frac{\partial \Lambda_t}{\partial \alpha_j} \right) \Lambda_t^{-1} \right]$$

for $1 < i, j < L$; and $F_2$ has $(i, j)$-th element equal to

$$F_{2ij} = \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \left[ \text{vec}(Z)' \Lambda_t^{-1} \text{vec}(Z) \right]$$

$$= \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \left[ \sum_{t=1}^{T} d_t \Lambda_t^{-1} d_t' \right]$$

$$= \sum_{t=1}^{T} \left[ \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} (d_t \Lambda_t^{-1} d_t') \right]$$

for $1 < i, j < L$. Define $\mathbf{f}_1(n)$ to be the matrix $f_1$ evaluated at the point $\hat{\alpha}_1 = \alpha_1(n)$, and define $\mathbf{f}_2(n)$, $\mathbf{f}(n)$, $F_1(n)$, $F_2(n)$, and $F(n)$ similarly. Then the formula

$$\hat{\alpha}_1(n+1) = \hat{\alpha}_1(n) + [\mathbf{F}(n)]^{-1} \mathbf{f}(n) \quad (5.53)$$

provides an iterative Newton-Raphson algorithm to compute maximum likelihood estimates of $\alpha_1$. Additional research is required to develop
modifications of expression (5.53) which will assure convergence of the sequence \(\{\hat{a}_1(n)\}\) to an \(a^*_t\) which minimizes expression (5.51).

In the procedure outlined above, the notational complexity of the required derivatives is approximately the same as in Section 5.1. However, the matrices here have dimension \((p+k)J \times (p+k)J\) or less and the only inversion required is for matrices of dimension \(p \times p\). The resulting reduction in numerical complexity offers the potential for considerably more efficient computational results.

Strictly speaking, for \(q_1 > 1\), the procedure outlined above requires an "infinite past" in order to lead to an exact maximum likelihood procedure. In practice, the procedure is initialized with selected quantities \(\hat{W}_0\) and \(P_0\). In keeping with the initial values of Wincek and Reinsel (1986), one may use \(\hat{W}_0 = 0\) and \(P_0 = 0\).

The work above addressed the structural model (2.3)-(2.4) with no error in the equation. For model (2.1)-(2.2) with an error in the equation, if \(x_t\), \(a_t\) and \(q_t\) follow three mutually uncorrelated ARMA\(_p\)(\(p\), \(q\)) processes, respectively, then one may again represent the \(Z_t\) process in state-space form, but with \(x^*_t = (x_t, e_t)\) replaced by \(\tilde{x}_t = (x_t, a_t, q_t)\) so that

\[
Z'_t = [(\bar{\beta}, I_k)', I_p, (\bar{X}_t, 0_{r \times k}', \tilde{a}_t)', (\tilde{x}_t)']' \cdot
\]

The resulting state-space model replaces (5.43)-(5.44) with

\[
Z'_t = \tilde{\bar{W}}_t \tilde{W}'_t
\]  
(5.54)
\[ \tilde{\tilde{W}}_{t+1} = \tilde{A}_w \tilde{W}_t + \tilde{A}_s \tilde{s}_t, \quad (5.55) \]

where \( \tilde{W}_t \) and \( \tilde{s}_t \) are defined analogously to \( W_t^* \) and \( s_t^* \);

\[
\tilde{B}_w = [(\beta, I_{2} \cdot (r, O_{2 \times 2}), O_{p \times 2(K-1)p}],
\]

\[
\tilde{A}_w = \begin{bmatrix}
0 & I_{2p} & 0 & 0 & \cdots & 0 \\
0 & 0 & I_{2p} & 0 & \cdots & 0 \\
& & & & \ddots & \ddots \\
& & & \ddots & \ddots & \ddots \\
& & -\tilde{\beta}_K & -\tilde{\beta}_{K-1} & -\tilde{\beta}_{K-2} & -\tilde{\beta}_{K-3} & \cdots & -\tilde{\beta}_1 \\
\end{bmatrix}, \quad (5.56)
\]

\[
\tilde{C}_w = [I_{2p}, \tilde{C}_1, \ldots, \tilde{C}_K]',
\]

\( \{\tilde{\beta}_j, j=1, 2, \ldots, K\} \) and \( \{\tilde{\beta}_j, j=0, 1, \ldots, K-1\} \) are the \( 2p \times 2p \) autoregressive and moving average parameter matrices, respectively, of the \( \tilde{X}_t \) process, \( \tilde{\beta}_j = 0 \) for \( j > p_2 \), \( \tilde{\beta}_j = 0 \) for \( j > p_1 \);

\( \{\tilde{\beta}_j, j=0, 1, \ldots, K\} \) are defined from \( \tilde{\beta}_j \) and \( \tilde{\beta}_j \) by (5.37); and

\( K = \max(p_2, q_2 + 1), \quad p_2 = \max(p_x, p_a, p_q), \) and \( q_2 = \max(q_x, q_a, q_q) \).

The innovation sequence, likelihood function and derivatives then follow in a manner analogous to the results for the model with no error in the equation.

Thus, for some identified structural models with or without an error in the equation, state-space representations and innovation sequence arguments may lead to Newton-Raphson procedures for maximum likelihood
estimation of regression coefficients and autocovariance parameters. Compared to the methods of Section 5.3, the procedures presented in this section require considerably less prior knowledge of autocovariance parameters, and permit the use of lower-dimensional models for the "signal plus noise" observations.
6. MODELS AND ESTIMATORS FOR THE FUNCTIONAL CASE

The preceding chapter addressed parameter estimation for the structural measurement error model, i.e., for the case in which the "true x" values are a realization of a k-dimensional stochastic process. In some cases, it may be more reasonable to consider the \{x_t}\ to be a fixed sequence of 1 \times k vectors; the resulting "functional" measurement error model requires estimation procedures that are somewhat different from those for the structural model.

Section 6.1 presents the matrices of first and second derivatives required for direct Newton-Raphson computation of maximum likelihood estimators of unknown but identified parameters. The relationship between the resulting estimators and the weighted estimators of Section 4.2 is discussed. As in Chapter 5, considerations of parsimony as well as common time series practice suggest that one examine the case in which \( x_0 \) follows an autoregressive moving average model. This permits one to use a modified state-space approach to obtain a relatively simple maximum likelihood estimation procedure. Section 6.2 outlines the computations required for this method.

6.1. Maximum Likelihood Estimation for General Second-Order Stationary Covariance Structure

As in Chapter 5, consider first the measurement-error model

\[
\text{vec}(Z) = [(\beta, I_k)' \cdot I_T]\text{vec}(x) + \text{vec}(\varepsilon),
\]

(6.1)
but assume now that \( x \) is a \( T \times k \) matrix of fixed real numbers; that \( \text{vec}(e) \) has a \( Tp \)-dimensional normal \( (0, \Gamma_{
abla e}) \) distribution; that \( \alpha \) is a minimal set of parameters for the nonsingular covariance matrix \( \Gamma_{
abla e} \); and that \( \beta, \alpha \), and \( x \) are functionally unrelated. Then \( \text{vec}(Z) \) has a \( Tp \)-dimensional normal distribution with mean \( [(\beta, I_k)' \ast I_T]\text{vec}(x) \) and positive definite covariance matrix \( \Gamma_{
abla e} \), so that the likelihood function of \( (\beta, \alpha, x) \), \( L(\beta, \alpha, x; Z) \), satisfies the equation

\[
-2 \ln L(\beta, \alpha, x; Z) = Tp \ln(2\pi) - \ln |\Gamma_{\nabla e}(\alpha)| + \{\text{vec}(Z) - [(\beta, I_k)' \ast I_T]\text{vec}(x)\}'[\Gamma_{\nabla e}(\alpha)]^{-1}\ \times \\{\text{vec}(Z) - [(\beta, I_k)' \ast I_T]\text{vec}(x)\}.
\]

Therefore, maximum likelihood estimation requires one to minimize

\[
g_1(\alpha) + g_2(\beta, \alpha, x; Z) = \ln |\Gamma_{\nabla e}(\alpha)|
\]

with respect to the unknown elements of \( \beta, \alpha \) and \( x \), where

\[
g_1(\alpha) = \ln |\Gamma_{\nabla e}(\alpha)|
\]

and

\[
g_2(\beta, \alpha, x; Z)
\]
355

\( g_2(\beta, \alpha_e, \mathbf{x}; \mathbf{Z}) \)

\[
= \{\text{vec}(\mathbf{Z}) - [(\beta, I_k)' \circ I_T]\text{vec}(\mathbf{x})\}'[I_{\alpha_e}^{-1} \circ \alpha_e]' \times \{\text{vec}(\mathbf{Z}) - [(\beta, I_k)' \circ I_T]\text{vec}(\mathbf{x})\}.
\]

Consider first the estimation of \( \mathbf{x} \). To simplify notation, \( I_{\alpha_e}(\alpha_e) \) will be written as \( I_{\alpha_e} \), and the dependence of \( I_{\alpha_e} \) on the elements of \( \alpha_e \) will be implicitly understood. For given \( \beta, \alpha_e \), and \( \mathbf{Z} \), formula (6.4) implies that minimization of (6.2) with respect to \( \mathbf{x} \) is achieved by a least-squares fit of \( I_{\alpha_e}^{-1/2} \text{vec}(\mathbf{Z}) \) to

\[
I_{\alpha_e}^{-1/2}[(\beta, I_k)' \circ I_T]\text{vec}(\mathbf{x}),
\]

which has solution

\[
\text{vec}(\mathbf{x}) \equiv [(\beta, I_k) \circ I_T]I_{\alpha_e}^{-1}[(\beta, I_k)' \circ I_T]^{-1}[(\beta, I_k) \circ I_T]I_{\alpha_e}^{-1}\text{vec}(\mathbf{Z})
\]

\( (6.5) \)

where, as throughout this chapter, it is assumed that

\[
[(\beta, I_k) \circ I_T]I_{\alpha_e}^{-1}[(\beta, I_k)' \circ I_T]
\]

is nonsingular. Then for given \( \beta, \alpha_e \) and \( \mathbf{Z} \),

\[
g_3(\beta, \alpha_e; \mathbf{Z}) \equiv \min_{\mathbf{x}} g_2(\beta, \alpha_e, \mathbf{x}; \mathbf{Z})
\]

\[
= \{\text{vec}(\mathbf{Z}) - [(\beta, I_k)' \circ I_T]\text{vec}(\mathbf{x})\}'I_{\alpha_e}^{-1}\{\text{vec}(\mathbf{Z}) - [(\beta, I_k)' \circ I_T]\text{vec}(\mathbf{x})\}.
\]
where the equality preceding formula (6.6) follows from Result 9.12.

Recall that formula (6.5) is equivalent to a least squares solution of the fit of \( \frac{1}{2} \) vec(Z) to \( \frac{1}{2} \) vec([β, I_k]' a I_T) and that formula (6.7) gives an associated residual sum of squares. Consider the special case in which the \( \xi_e \) are independent and identically distributed p-dimensional normal \((0, \Sigma_{\xi_e})\) random vectors, and assume that \( \Sigma_{\xi_e} \) is nonsingular. Then \( \Sigma_{\xi_e} = \Sigma_{\xi_e} a I_T \), \( \Sigma_{\xi_e} = \Sigma_{\xi_e} a I_T \), and \( \Sigma_{\xi_e} = \Sigma_{\xi_e} a I_T \). Hence, one may rewrite expression (6.5) to read
\[ \text{vec}(\hat{x}) = \{(a, I_k) \Sigma^{-1} (\beta, I_k) \}' a I_T \}^{-1} \{(a, I_k) \Sigma^{-1} (\beta, I_k) \}' a I_T \} \text{vec}(Z) \]

\[ = \{(a, I_k) \Sigma^{-1} (\beta, I_k) \}'^{-1} \{(a, I_k) \Sigma^{-1} (\beta, I_k) \}' a I_T \} \text{vec}(Z) \]

so that

\[ \hat{x} = Z \Sigma^{-1} (\beta, I_k) \}' [(a, I_k) \Sigma^{-1} (\beta, I_k) ]^{-1} \]

and thus

\[ \hat{x}_t = Z_t \Sigma_t^{-1} (\beta, I_k) \]' ([a, I_k] \Sigma_t^{-1} (\beta, I_k)]^{-1} \]

This is the standard result for prediction of true \( x_t \) values in the functional measurement error model with independent and identically distributed errors \( \varepsilon_t \) [c.f. expression (4.1.5) of Fuller (1987, p. 294)]. Similarly, \( \Gamma_\varepsilon = \Sigma_\varepsilon a I_T \) implies that one may rewrite expression (6.7) as

\[ \text{vec}(Z)' [(I_r, -\beta') a I_T] \{(I_r, -\beta') a I_T \} \Sigma_\varepsilon a I_T [(I_r, -\beta') a I_T]^{-1} \]

\[ \times [(I_r, -\beta') a I_T] \text{vec}(Z) \]

\[ = \text{vec}(Z)' [(I_r, -\beta') [(I_r, -\beta') \Sigma_\varepsilon (I_r, -\beta')]'^{-1} \]

\[ \times (I_r, -\beta') a I_T \text{vec}(Z) \]
This is the standard formula for the residual sum of squares in the prediction of true $z$ in the functional model with independent and identically distributed errors $\varepsilon_t$ [c.f. expression (4.1.6) of Fuller (1987, p. 294)].

Throughout this chapter assume that the elements of $\Gamma_{ee}$ are twice continuously differentiable functions of $\varphi_e$. (This condition will be satisfied if, for example, $\varepsilon_t$ follows an autoregressive moving average process and $\varphi_e$ represents the distinct elements of the residual variance matrix and the autoregressive and moving average coefficient matrices of the $\varepsilon_t$ process.) Then inspection of expressions (6.2)-(6.4) indicates that these expressions are twice continuously differentiable with respect to $[\text{vec}(\beta)'$, $\varphi_e]$. Hence, one may use matrix differentiation arguments to develop numerical methods for the estimation of unknown but identified elements of $[\text{vec}(\beta)'$, $\varphi_e]$. Now consider the estimation of $\beta$. Since $g_1(\varphi_e)$ of expression (6.3) is not a function of $\beta$, maximum likelihood estimation of $\beta$ for a given $\varphi_e$ requires one to minimize $g_3(\beta, \varphi_e; Z)$ with respect to $\beta$. One may take a number of approaches to this problem. First, let

$$\nu = (\nu_1, \nu_2, ..., \nu_T)'$$

$$= Z(I_T, -\beta)'$$
\[ Y - X_\beta = (X_\beta + \epsilon) - (X + u)_\beta = \epsilon - u_\beta, \]

where each \( v_t, t=1, 2, ..., T \), is a \( 1 \times r \) vector. Note that

\[ \text{vec}(v) = [(I_r, -\beta') \otimes I_T] \text{vec}(v) \]

and that \( \Sigma_{vv} = \text{Var}[\text{vec}(v)] \) equals

\[ [(I_r, -\beta') \otimes I_T] \Sigma_{\epsilon \epsilon} [(I_r, -\beta')' \otimes I_T]. \]

Formula (6.6) then becomes

\[ \text{vec}(v)' \Sigma_{vv}^{-1} \text{vec}(v), \]

or, if the dependency on \( \beta \) and \( \epsilon_{-\epsilon} \) is made notationally explicit,

\[ \text{vec}[v(\beta)]' \Sigma_{vv}^{-1}(\beta, \epsilon_{-\epsilon})^{-1} \text{vec}[v(\beta)]. \quad (6.8) \]

One may then evaluate matrix derivatives and develop a direct

Newton-Raphson approach to the estimation of \( \beta \) and \( \epsilon_{-\epsilon} \) by minimizing

(6.8). Let \( \epsilon_{-\epsilon_k} \) be the \( k \)-th element of \( \epsilon_{-\epsilon} \). Recall from Appendix A...
\[
\frac{\partial}{\partial \alpha_{el}} [\ln |\Gamma_{ee}|] = \text{tr} \left\{ \left[ \frac{\partial}{\partial \alpha_{el}} (\Gamma_{ee}) \right]^{-1} \right\};
\]

\[
\frac{\partial^2}{\partial \alpha_{el} \partial \alpha_{ej}} [\ln |\Gamma_{ee}|] = \text{tr} \left\{ \left[ \frac{\partial^2}{\partial \alpha_{el} \partial \alpha_{ej}} (\Gamma_{ee}) \right]^{-1} \right\} - \text{tr} \left\{ \left[ \frac{\partial}{\partial \alpha_{el}} (\Gamma_{ee}) \right]^{-1} \frac{\partial}{\partial \alpha_{ej}} (\Gamma_{ee}) \right\};
\]

and

\[
\frac{\partial}{\partial \beta'_{ij}} [(I, -\beta') \circ I_T] = (0_{r \times r}, -\tilde{\alpha}'_{di}) \circ I_T;
\]

where, as in Chapter 5, \( \tilde{\alpha}'_{di} \) is a \( k \times r \) matrix with a one in the \((i,j)\)-th position and zeros elsewhere. By Result 9.4,

\[
\frac{\partial}{\partial \beta'_{ij}} [\Gamma_{vv}] = \frac{\partial}{\partial \beta'_{ij}} \left\{ \left( (I, -\beta') \circ I_T \right) \Gamma_{ee} \left( (I, -\beta')' \circ I_T \right) \right\}
\]

\[
= \left[ (0, -\tilde{\alpha}'_{di}) \circ I_T \right] \Gamma_{ee} \left[ (I, -\beta')' \circ I_T \right]
\]

\[
+ \left[ (I, -\beta') \circ I_T \right] \Gamma_{ee} \left[ (0, -\tilde{\alpha}'_{di})' \circ I_T \right]
\]

\[
= -[\tilde{\alpha}'_{di} \circ I_T] [\Gamma_{ue} - \Gamma_{uu} (\beta \circ I_T)] - [\Gamma_{eu} - (\beta' \circ I_T) \Gamma_{uu}] [\tilde{\alpha}'_{di} \circ I_T];
\]
\[
\frac{\partial^2}{\partial \beta_i \partial \beta_j} \{ \Gamma_{vv} \} = \left\{ \begin{array}{l}
(0, \tilde{\Delta}'_{ij}) \alpha I_T \Gamma_{cc}[(0, \tilde{\Delta}_{km})' \alpha I_T] \\
+ [(0, \tilde{\Delta}'_{km}) \alpha I_T \Gamma_{cc}[(0, \tilde{\Delta}'_{ij})' \alpha I_T] \\
= [\tilde{\Delta}'_{ij} \alpha I_T \Gamma_{uu}[(\tilde{\Delta}'_{km}) \alpha I_T] + [\tilde{\Delta}'_{km} \alpha I_T \Gamma_{uu}[(\tilde{\Delta}'_{ij}) \times I_T] ; \\
(6.9)
\end{array} \right.
\]

\[
\frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} \{ \Gamma_{vv} \} = \left\{ \begin{array}{l}
[(I_r, -\beta') \alpha I_T]\frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} (\Gamma_{cc})][(I_r, -\beta')' \alpha I_T] ; \\
(6.9)
\end{array} \right.
\]

and

\[
\frac{\partial^2}{\partial \alpha_{ei} \partial \beta_{ij}} \{ \Gamma_{vv} \} = \left\{ \begin{array}{l}
(0, -\tilde{\Delta}'_{ij}) \alpha I_T]\frac{\partial^2}{\partial \alpha_{ei} \partial \beta_{ij}} (\Gamma_{cc})][(I_r, -\beta')' \alpha I_T] \\
+ [(I_r, -\beta') \alpha I_T]\frac{\partial^2}{\partial \alpha_{ei} \partial \beta_{ij}} (\Gamma_{cc})][(0, -\tilde{\Delta}'_{ij})' \alpha I_T] . \\
\end{array} \right.
\]

Result 9.5 implies that

\[
\frac{\partial}{\partial \beta_{ij}} \mathrm{vec}(\nu) = \frac{\partial}{\partial \beta_{ij}} \{ [(I_r, -\beta') \alpha I_T] \mathrm{vec}(Z) \}
\]

\[
= [(0, -\tilde{\Delta}'_{ij}) \alpha I_T] \mathrm{vec}(Z)
\]

\[
= - [\tilde{\Delta}'_{ij} \alpha I_T] \mathrm{vec}(X) ,
\]

\[a \ Tr \times 1 \ \text{vector with the } j-\text{th } T \times 1 \ \text{block equal to } X_{ij} \ \text{and with}\]
zeros elsewhere. Then by Result 9.7,

\[
\frac{\partial}{\partial \beta_{ij}} \left[ \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) \right] = \left[ \frac{\partial}{\partial \beta_{ij}} \text{vec}(v)' \right] \Gamma_{vv}^{-1} \text{vec}(v) + \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} \text{vec}(v) \right] \\
- \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v) \\
= -2 \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} \right] \text{vec}(X) \\
- 2 \text{vec}(v)' \Gamma_{vv}^{-1} \left[ (I - \gamma') \times I \right] \text{vec}(X)
\]

\[
\frac{\partial^2}{\partial \beta_{ij} \partial \beta_{km}} \left[ \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) \right] = \frac{\partial}{\partial \beta_{ij}} \left\{ 2 \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{km}} \text{vec}(v) \right] - \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{km}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v) \right\} \\
= 2 \left\{ \frac{\partial}{\partial \beta_{ij}} \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{km}} \text{vec}(v) \right] \\
- \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v) \right\} \\
- 2 \left[ \frac{\partial}{\partial \beta_{ij}} \text{vec}(v)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{km}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v) \right] \\
- \text{vec}(v)' \left[ -2 \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{ij}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \beta_{km}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \right] \\
+ \Gamma_{vv}^{-1} \left[ \frac{\partial^2}{\partial \beta_{ij} \partial \beta_{km}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v) \right\};
\]
\[ \frac{3}{\partial \alpha \varepsilon \ell} \left[ \text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \right] = -\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \ell} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \]

\[ = -\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \left( I_{\varepsilon} - \beta' \right) \cdot I \Gamma_{\varepsilon \varepsilon} \left[ \frac{3}{\partial \alpha \varepsilon \ell} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \left( I_{\varepsilon} - \beta' \right) \cdot I \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) ; \]

\[ \frac{3^2}{\partial \alpha \varepsilon \ell \partial \alpha \varepsilon \varepsilon} \left[ \text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \right] \]

\[ = -\frac{3}{\partial \alpha \varepsilon \ell} \left[ -\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \right] \]

\[ = 2\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \ell} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \]

\[ -\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3^2}{\partial \alpha \varepsilon \ell \partial \alpha \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) ; \]

and

\[ \frac{3^2}{\partial \alpha \varepsilon \ell \partial \beta \varepsilon \varepsilon} \left[ \text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \right] \]

\[ = -\frac{3}{\partial \alpha \varepsilon \ell} \left[ 2\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \beta \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \text{vec}(v) \right] - \text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \beta \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \]

\[ = -2\text{vec}(v)' \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \ell} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \beta \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \]

\[ - \text{vec}(v)' \left[ -2 \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \alpha \varepsilon \ell} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3}{\partial \beta \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \right. \]

\[ + \Gamma_{\varepsilon \varepsilon}^{-1} \left[ \frac{3^2}{\partial \alpha \varepsilon \ell \partial \beta \varepsilon \varepsilon} \left( \Gamma_{\varepsilon \varepsilon} \right) \right] \Gamma_{\varepsilon \varepsilon}^{-1} \text{vec}(v) \]
Model (6.1) did not include an explicit intercept term. To handle this case, one may extend model (6.1) to read

\[ \text{vec}(Z) = (\beta_0, 0_{1\times k})' \circ 1_T + [(\beta, I_k) \circ 1_T] \text{vec}(x) + \text{vec}(e) \]  

(6.10)

where \( \beta_0 \) is an unconstrained \( 1 \times r \) vector. Then the results derived above for minimization of (6.4) with respect to \( \beta \) and for matrix derivatives continue to hold with \( Z \) replaced by \( Z - y_0 \), where

\[ y_0 = 1_T(\beta_0, 0_{1\times k}) \]. By Results 9.1 and 9.9,

\[ \frac{\partial}{\partial \beta_0} [\text{vec}(\nu)' \Gamma^{-1}_{VV} \text{vec}(\nu)] = -2(1_T \circ 1_T)' \Gamma^{-1}_{VV} \text{vec}(\nu) \]

where now

\[ \nu = Y - (1_T \circ 1_T) - X_2 \]

\[ = [Z - y_0](I_T, -\beta_0)' \]

and
\[ \frac{\partial}{\partial \theta_0} \text{vec}(v) = I_T \ast I_T. \]

Similarly,

\[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) = -2(\mathbf{I} \ast \mathbf{1}_T)' \Gamma_{vv}^{-1}(\mathbf{I} \ast \mathbf{1}_T), \]

\[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) = \frac{\partial}{\partial \theta_i} \left\{ -2(\mathbf{I} \ast \mathbf{1}_T)' \Gamma_{vv}^{-1} \text{vec}(v) \right\} \]

\[ = -2(\mathbf{I} \ast \mathbf{1}_T)' \left\{ \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \theta_i} (\Gamma_{vv}) \right] \Gamma_{vv}^{-1} \text{vec}(v) + \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \theta_j} \text{vec}(v) \right] \right\}, \]

\[ = 2(\mathbf{I} \ast \mathbf{1}_T)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \theta_i} (\Gamma_{vv}) \right] \Gamma_{vv}^{-1} \text{vec}(v) + \left[ \frac{\partial}{\partial \theta_j} \text{vec}(v) \right], \]

and

\[ \frac{\partial^2}{\partial \alpha \partial \theta_0} \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) = \frac{\partial}{\partial \alpha} \left\{ -2(\mathbf{I} \ast \mathbf{1}_T)' \Gamma_{vv}^{-1} \text{vec}(v) \right\} \]

\[ = 2(\mathbf{I} \ast \mathbf{1}_T)' \Gamma_{vv}^{-1} \left[ \frac{\partial}{\partial \alpha} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} \text{vec}(v). \]

Let \( g = \{ \text{vec}(\beta) \}, \alpha_0, \beta_0 \), let \( L(g, x; Z) \) be the likelihood function for model (6.10), and let

\[ L_m(g, Z) = \max_x L(g, x; Z). \]

Then the vector of first partial derivatives of \(-2\ln[L_m(g, Z)]\) with respect to \( g \) is \( (a', b', c')' \), where

\[ a = \frac{(-2) \partial \ln[L_m(g; Z)]}{\partial \text{vec}(\beta)} \quad (6.11) \]
has double-subscripted \((i,j)\)-th element equal to

\[
2\text{vec}(v)' \Gamma^{-1}_{\prec \prec \prec} \frac{\partial}{\partial \beta_{ij}} \text{vec}(v) - \text{vec}(v)' \Gamma^{-1}_{\prec \prec \prec} \frac{\partial}{\partial \beta_{ij}} (\Gamma_{\prec \prec \prec}) \Gamma^{-1}_{\prec \prec \prec} \text{vec}(v);
\]

\[
b = \frac{(-2) \delta\{\ln[I_m(g; Z)]\}}{\delta \alpha'^i} \]

has \(i\)-th element equal to

\[
\text{tr}\left[\left[\frac{\partial}{\partial \alpha_e} (\Gamma_{\prec \prec \prec})\right]^{-1}\right] - \text{vec}(v)' \Gamma^{-1}_{\prec \prec \prec} \frac{\partial}{\partial \alpha_e} (\Gamma_{\prec \prec \prec}) \Gamma^{-1}_{\prec \prec \prec} \text{vec}(v);
\]

and

\[
c = \frac{(-2) \delta\{\ln[I_m(g; Z)]\}}{\delta \beta_j} \]

\[
= -2(\Gamma_r \otimes \Gamma_r)' \Gamma^{-1}_{\prec \prec \prec} \text{vec}(v).
\]

Similarly, the matrix of second partial derivatives

\[
-\frac{\partial^2 \ln[I_m(g; Z)]}{\partial \alpha' \partial \alpha}
\]

may be written

\[
2^{-1}\begin{pmatrix}
A & B & C \\
B' & D & E \\
C' & E' & F
\end{pmatrix},
\]

where
\[
A = \frac{(-2) \alpha^2 \ln(L_m [g; Z])}{\alpha[\text{vec}(g)]^2 \alpha[\text{vec}(g)]'}
\]

has, in double-subscripted notation, \([(i,j), (\ell,m)]\)-th element equal to

\[
2 \left[ \frac{\partial}{\partial_{ij}} \text{vec}(v) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{km}} \text{vec}(v) \right] - 2 \text{vec}(v)' \left[ \frac{\partial}{\partial_{ij}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \text{vec}(v) \right]
+ \left[ \frac{\partial}{\partial_{km}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \text{vec}(v) \right]
+ \text{vec}(v)' \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{km}} \left( \Gamma_{\nu v} \right) \right] - \left[ \frac{\partial^2}{\partial_{ij} \partial_{km}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \text{vec}(v)
\]

\[
B = \frac{(-2) \alpha^2 \ln(L_m [g; Z])}{\alpha[\text{vec}(g)]^2 \alpha[\ell]}
\]

has \([(i,j), \ell]\)-th element equal to

\[
- 2 \text{vec}(v)' \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{\ell l}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \text{vec}(v) \right]
+ \text{vec}(v)' \Gamma^{-1}_{\nu v} \left[ 2 \frac{\partial}{\partial_{\ell l}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \left( \Gamma_{\nu v} \right) \right] - \left[ \frac{\partial^2}{\partial_{\ell l} \partial_{ij}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \text{vec}(v)
\]

\[
C = \frac{(-2) \alpha^2 \ln(L_m [g; Z])}{\alpha[\text{vec}(g)]^2 \alpha[0]}
\]

has \((i,j)\)-th row equal to

\[
2 \left( \text{vec}(I_r = I_T) \right)' \Gamma^{-1}_{\nu v} \left[ \frac{\partial}{\partial_{ij}} \left( \Gamma_{\nu v} \right) \right] \Gamma^{-1}_{\nu v} \text{vec}(v) + \left[ \frac{\partial}{\partial_{ij}} \text{vec}(I_T) \right] \text{vec}(\mathbf{1})
\]
has $(i,j)$-th element equal to

\[
vec(v)' \Gamma_{vv}^{-1} \left[ 2 \left( \frac{\partial}{\partial \alpha_{ei}} \left( \Gamma_{vv} \right) \right) \Gamma_{vv}^{-1} \left( \frac{\partial}{\partial \alpha_{ej}} \left( \Gamma_{vv} \right) \right) - \frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} \left( \Gamma_{vv} \right) \right] \Gamma_{vv}^{-1} vec(v)
\]

\[
- \text{tr} \left[ \left( \frac{\partial}{\partial \alpha_{ei}} \left( \Gamma_{ee} \right) \right) \Gamma_{ee}^{-1} \left( \frac{\partial}{\partial \alpha_{ej}} \left( \Gamma_{ee} \right) \right) \Gamma_{ee}^{-1} \right]
\]

\[
+ \text{tr} \left[ \left( \frac{\partial^2}{\partial \alpha_{ei} \partial \alpha_{ej}} \left( \Gamma_{ee} \right) \right) \Gamma_{ee}^{-1} \right];
\]

\[
E = \frac{(-2)^2 \partial \ln \{L_m | g; Z\}}{\partial \alpha^i \partial \alpha^j}
\]

has $i$-th row equal to

\[
2 (I_r \otimes I_t)' \Gamma_{vv}^{-1} \left( \frac{\partial}{\partial \alpha_{ei}} \left( \Gamma_{vv} \right) \right) \Gamma_{vv}^{-1} vec(v);
\]

and

\[
F = \frac{(-2)^2 \partial \ln \{L_m | g; Z\}}{\partial \theta^i \partial \theta^j}
\]

\[
= - 2 (I_r \otimes I_t)' \Gamma_{vv}^{-1} (I_r \otimes I_t).
\]

The matrix derivatives $(a', b', c')'$ and $I_F$ developed above may lead to a general Newton-Raphson algorithm for maximum likelihood estimation of unknown but identified elements of
\( \mathbf{a} = \{ \text{vec}(\beta) \}' \), \( \alpha \), \( \beta_0 \). Of course, such a recursive procedure is of little value unless the reduced likelihood function \( L_m(\mathbf{a}; \mathbf{Z}) \) satisfies certain regularity conditions. Additional research will be required to develop the relation between the surface determined by \( L_m(\mathbf{a}; \mathbf{Z}) \) and the functional dependence of \( \Gamma_{\mathbf{a}, \alpha} = \Gamma_{\mathbf{a}, \alpha} (\alpha) \),

\[
\Gamma_{\mathbf{v}, \mathbf{v}} = [(\mathbf{I}_r, -\beta') \mathbf{a} \mathbf{L}_r] \Gamma_{\mathbf{a}, \alpha} (\alpha) [(\mathbf{I}_r, -\beta')' \mathbf{a} \mathbf{L}_r],
\]

and \( \mathbf{v} = \mathbf{Y} - \mathbf{L}_r \beta_0 - \mathbf{X} \beta \) on \( \mathbf{a} \). An understanding of the reduced likelihood surface may then offer insight into the concavity of \( L_m(\mathbf{a}; \mathbf{Z}) \) with respect to \( \mathbf{a} \); the existence of global or local maxima of \( L_m(\mathbf{a}; \mathbf{Z}) \) with respect to \( \mathbf{a} \); boundary-case maxima; convergence of Newton-Raphson or other recursive procedures; and asymptotic behavior of the resulting maximum likelihood estimators.

For the reasons noted above, the present work will not pursue further the estimation of unknown but identified elements of \( \mathbf{a} \) at the current level of generality. However, given a known covariance matrix \( \Gamma_{\mathbf{a}, \alpha} \), one may study maximum likelihood estimation of regression coefficients in additional detail.

First assume that \( r = k = 1 \) and that \( \Gamma_{\mathbf{a}, \alpha} = 0_{T \times T} \). By formula (6.7),

\[
g_3(\beta, \alpha; \mathbf{Z}) = \mathbf{v}' \Gamma_{\mathbf{v}, \mathbf{v}}^{-1} \mathbf{v}
\]
By Theorem 12.2.13 of Graybill (1983), there exists a $T \times T$ matrix $Q$ such that $Q^* = \Lambda_{ee} = \text{diag}(\lambda_{ee1}, \lambda_{ee2}, \ldots, \lambda_{eett})$ and $Q^T Q' = I_T$. Let $Y^* = (Y^*_1, Y^*_2, \ldots, Y^*_T)' = QY$ and $X^* = (X^*_1, X^*_2, \ldots, X^*_T)' = QX$. One may then rewrite $g_2(\beta, \gamma; Z)$ as

$$[Y - \beta X]' \{Q^{-1} Q^T + \beta^2 Q^{-1} Q Q^{-1} \}^{-1} [Y - \beta X]$$

$$= [Y^* - \beta X^*]' \{\Lambda_{ee} + \beta^2 I_T \}^{-1} [Y^* - \beta X^*]$$

$$= \sum_{t=1}^{T} (Y^*_t - \beta X^*_t)^2 (\lambda_{eett} + \beta^2)^{-2}. \quad (6.12)$$

Thus, the transformation $Z^* = QZ$ has reduced the error structure from one of serial correlation to one of heteroscedasticity. Therefore, the methods of Section 3.1.6 of Fuller (1987) are applicable here, provided $\Lambda_{ee}$ and $\Lambda_{uu}$ are known. Formula (6.12) implies that given $\Lambda_{ee}$, the maximum likelihood estimation of $\beta$ reduces to a problem of iteratively reweighted least squares.

For computational purposes, one may construct $\Lambda_{ee}$ and $Q$ as follows. Assume that $\Lambda_{uu}$ is positive definite and let $\Lambda_{uu} = \text{diag}(\lambda_{uu1}, \lambda_{uu2}, \ldots, \lambda_{uut})$, where $\lambda_{uui}$ is the $i$-th largest eigenvalue of $\Lambda_{uu}$, and let $Q_1 = (q_{11}, q_{12}, \ldots, q_{1T})'$, where $q_{11}$
is an eigenvector of $\Gamma_{uu}$ corresponding to $\lambda_1$, and $Q_1$ is orthogonal. Let $M = A_{uu}^{-1/2} Q_1 \Gamma_{ee} Q_1 A_{uu}^{-1/2}$ and define

$\Lambda_{ee} = \text{diag}(\lambda_{ee11}, \lambda_{ee22}, \ldots, \lambda_{eeTT})$ and $Q_2 = (q_{21}, q_{22}, \ldots, q_{2T})'$

to be matrices of ordered eigenvalues and associated orthonormal eigenvectors, respectively, of $M$. Then

$$Q = Q_2 A_{uu}^{-1/2} Q_1$$

satisfies the following relationships:

$$Q \Gamma_{uu} Q' = Q_2 A_{uu}^{-1/2} Q_1 \Gamma_{ee} Q_1 A_{uu}^{-1/2} Q_2$$

$$= Q_2 A_{uu}^{-1/2} A_{uu}^{-1/2} Q_2$$

$$= Q_2 Q_2' = I_T;$$

and

$$Q \Gamma_{ee} Q' = Q_2 A_{uu}^{-1/2} Q_2 \Gamma_{ee} Q_1 A_{uu}^{-1/2} Q_2$$

$$= \Lambda_{ee}.$$

It follows that $\Lambda_{ee}$ and $Q$ are matrices of eigenvalues and associated orthonormal eigenvectors of $\Gamma_{ee}$ in the metric determined by $\Gamma_{uu}$.

For more background on metrics determined by positive definite matrices, see Appendix 4.A of Fuller (1987).
In the special case presented above, model (6.1) led to minimization of (6.12). For model (6.10) including an intercept, \( r = k = 1 \), and \( \Gamma_{ue} = 0_{T \times T} \),

\[
v' \Gamma_{\omega \omega}^{-1} v = \sum_{t=1}^{T} (Y_t^* - \beta_0 X_0^* - \beta X_t^*)^2 (\lambda_{ett} + \beta^2)^{-2} \tag{6.13}
\]

where \( X_0^* = (X_{01}^*, X_{02}^*, ..., X_{0T}^*)' = Q_{1,T} \). Thus, the estimation of \( \beta_0 \) and \( \beta \) in this case again reduces to a problem in iteratively reweighted least squares.

For the single-relation model with nonzero \( \Gamma_{ue} \), a slightly more general iterative procedure is required for the estimation of \( \beta \) when \( \Gamma_{\omega \omega} \) is known. Recall that \( \varphi \) defined in expression (6.11) is the vector of first partial derivatives of \( L_m(\varphi; Z) \) with respect to the elements of \( \text{vec}(\varphi) \). Note that for general \( r \) and \( k \), \( \varphi = 0_{rk \times 1} \) if and only if

\[
0 = \text{vec}(v)' \Gamma_{\omega \omega}^{-1} \left[ 2 \left[ \frac{3}{\beta_{ij}} \text{vec}(v) \right] - \left[ \frac{3}{\beta_{ij}} (\Gamma_{\omega \omega}) \right] \Gamma_{\omega \omega}^{-1} \text{vec}(v) \right] \\
= -2 \text{vec}(v)' \Gamma_{\omega \omega}^{-1} \left[ \begin{bmatrix} \tilde{A}'_{1j} \otimes I_T \end{bmatrix} \text{vec}(X) \right. \\
\left. + \begin{bmatrix} (0, -\tilde{A}'_{1j}) \otimes I_T \end{bmatrix} \Gamma_{\omega \omega} \begin{bmatrix} (I_r, -\beta') \otimes I_T \end{bmatrix} \Gamma_{\omega \omega}^{-1} \text{vec}(v) \right] \\
= -2 \text{vec}(v)' \Gamma_{\omega \omega}^{-1} \left[ \begin{bmatrix} \tilde{A}'_{1j} \otimes I_T \end{bmatrix} \text{vec}(X) - \left[ \Gamma_{\omega \omega} - \Gamma_{uu} (\varphi \otimes I_T) \right] \Gamma_{\omega \omega}^{-1} \text{vec}(v) \right] \\
\tag{6.14}
\]
for all \( l < i < k, \ l < j < r \). For the case \( r = k = 1 \), this condition reduces to

\[
0 = v' \Gamma_{\nu \nu}^{-1} X
\]

\[
= v' \Gamma_{\nu \nu}^{-1} \{ X - \{ \Gamma_{\nu e} - \Gamma_{\nu u} \beta \} \Gamma_{\nu \nu}^{-1} v \}
\]

\[
= X' \Gamma_{\nu \nu}^{-1} Y - X' \Gamma_{\nu \nu}^{-1} X \beta + v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu u} \Gamma_{\nu \nu}^{-1} v \beta - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu \nu}^{-1} v
\]

\[
= [X' \Gamma_{\nu \nu}^{-1} Y - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu u} \Gamma_{\nu \nu}^{-1} v] - [X' \Gamma_{\nu \nu}^{-1} X - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu \nu}^{-1} v] \beta .
\]

If \( X' \Gamma_{\nu \nu}^{-1} X - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu \nu}^{-1} v \) is nonzero, this final condition is equivalent to

\[
\beta = [X' \Gamma_{\nu \nu}^{-1} X - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu u} \Gamma_{\nu \nu}^{-1} v]^{-1} [X' \Gamma_{\nu \nu}^{-1} Y - v' \Gamma_{\nu \nu}^{-1} \Gamma_{\nu \nu}^{-1} v] .
\]  \( (6.15) \)

Because \( v \) and \( \Gamma_{\nu \nu} \) are functions of \( \beta \), \( (6.15) \) does not provide an explicit expression for the maximum likelihood estimator of \( \beta \), but it does suggest an iterative procedure to estimate \( \beta \). For such an iterative procedure let \( \hat{\beta}_{(i)} \) denote the estimate of \( \beta \) obtained in the \( i \)-th step. In particular, following the modification for the heteroscedastic case suggested by Fuller (1987, Section 3.1.6), let

\[
\hat{\beta}_{(i)} = [H_{(i-1)}^{-1}] R_{(i-1)}^{-1},
\]  \( (6.16) \)
where

\[ H(i) = X'[\hat{\Gamma}_{vv(i)}]^{-1}X - \lambda(i) \hat{v}(i)' [\hat{\Gamma}_{vv(i)}]^{-1}\hat{\Gamma}_{uu}[\hat{\Gamma}_{vv(i)}]^{-1}\hat{v}(i), \]

\[ R(i) = X'[\hat{\Gamma}_{vv(i)}]^{-1}Y - \lambda(i) \hat{v}(i)' [\hat{\Gamma}_{vv(i)}]^{-1}\hat{\Gamma}_{ue}[\hat{\Gamma}_{vv(i)}]^{-1}\hat{v}(i), \]

\[ \hat{v}(i) = Y - \hat{X}_e(i), \]

\[ \hat{\Gamma}_{vv(i)} = \Gamma_{ee} - \beta(i)\Gamma_{ue} - \Gamma_{eu}\hat{\beta}(i) + [\hat{\beta}(i)]^2\Gamma_{uu}, \]

\[ \lambda(i) = \max(0, c(i) - \alpha^{-1}), \]

\( \alpha \) is a preselected positive constant, and \( c(i) \) is the smallest root of the determinantal equation

\[ |(Y, X)'[\hat{\Gamma}_{vv(i)}]^{-1}(Y, X) - c[(\hat{v}(i)'[\hat{\Gamma}_{vv(i)}]^{-1} \otimes I_2|\hat{\Gamma}_{ee}[\hat{\Gamma}_{vv(i)}]^{-1}\hat{v}(i) \otimes I_2]| = 0. \]

The modification with \( \lambda(i) \) assures that \( H(i) \) is strictly positive with probability one.

Note that for \( r = k = 1 \),

\[ A = \frac{(-2)\beta^2\ln[L_n(g; Z)]}{\beta^2}. \]
\[= 2 \left[ \frac{\partial}{\partial \beta} (v) \right] \Gamma^{-1}_{\nu \nu} \left[ \frac{\partial}{\partial \beta} (v) \right] \]
\[- 4w' \Gamma^{-1}_{\nu \nu} \left[ \frac{\partial}{\partial \beta} (\Gamma_{\nu \nu}) \right] \Gamma^{-1}_{\nu \nu} \left[ \frac{\partial}{\partial \beta} (v) \right] \]
\[+ v' \Gamma^{-1}_{\nu \nu} \left[ 2 \left( \frac{\partial}{\partial \beta} (\Gamma_{\nu \nu}) \right) \right] \Gamma^{-1}_{\nu \nu} \left[ \frac{\partial}{\partial \beta} (\Gamma_{\nu \nu}) \right] - \frac{\partial^2}{\partial \beta^2} (\Gamma_{\nu \nu}) \Gamma^{-1}_{\nu \nu} \]
\[= 2 \xi' \Gamma^{-1}_{\nu \nu} \xi - 4w' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} - 2 \beta \Gamma_{\nu uu} \right] \Gamma^{-1}_{\nu \nu} \xi \]
\[+ 2w' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} - 2 \beta \Gamma_{\nu uu} \right] \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} - 2 \beta \Gamma_{\nu uu} \right] \Gamma^{-1}_{\nu \nu} v \]
\[- 2w' \Gamma^{-1}_{\nu \nu} \Gamma^{-1}_{\nu \nu} \Gamma^{-1}_{\nu \nu} v \]  \hspace{1cm} (6.17)

Now

\[E \left\{ w' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} - 2 \beta \Gamma_{\nu uu} \right] \Gamma^{-1}_{\nu \nu} \right\} \]
\[= E \left\{ w' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} \right] \Gamma^{-1}_{\nu \nu} (\xi + u) \right\} \]
\[= E \left\{ w' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e} + \Gamma_{\nu u} \right] \Gamma^{-1}_{\nu \nu} u \right\} \]
\[= 2 \text{tr} \{ E(uw') \} \Gamma^{-1}_{\nu \nu} \Gamma^{-1}_{\nu \nu} \]
\[= 2 \text{tr} \{ \Gamma_{\nu \nu} \Gamma^{-1}_{\nu \nu} \Gamma^{-1}_{\nu \nu} \}, \text{ and} \]
Thus, (6.17) has the same expectation as $2[X'\Gamma_{VV}^{-1}X - V'\Gamma_{VV}^{-1}\Gamma_{UV}^{-1}v]$.

Thus, under the conditions of Theorem 4.1, $T^{-1}H(i)$ is a consistent estimator of $T^{-2}E(A)$, where $A = (-2)\bar{\theta}^2\log[I_m(a; Z)]/\bar{\theta}^2$. Also, inspection of the equations preceding (6.15) indicates that $\hat{\beta}(i-1) - H(i-1)\hat{\beta}(i)$ is an estimator of the variable $\beta$ defined in expression (6.11). It follows that (6.16) defines an approximate Newton-Raphson algorithm for maximum likelihood estimation of $\beta$.

The iterative estimation procedure is slightly more complicated if one includes a nonzero intercept in the model. Note that

$$0 = c = 1'\Gamma_{VV}^{-1}v = 1'\Gamma_{VV}^{-1}(y - X\theta - 1\theta_0)$$

if and only if

$$\theta_0 = (1'\Gamma_{VV}^{-1}1_T)^{-1}1'\Gamma_{VV}^{-1}(y - X\theta)$$  \hspace{1cm} (6.18)

provided $1'\Gamma_{VV}^{-1}1_T$ is nonzero. Hence, simultaneous iterative solution of $a = 0$ and $c = 0$ for $\theta_0$ and $\beta$ leads one to the equation
\[ 0 = \mathbf{v}' \Gamma_{\omega \omega}^{-1} \{ \mathbf{X} - [\Gamma_{\omega \epsilon} - \Gamma_{\omega \omega}] \Gamma_{\omega \omega}^{-1} \mathbf{v} \} \]

\[ = \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{X} - \mathbf{I}_T \mathbf{v}_0 - \mathbf{X} \mathbf{v}_0] - \mathbf{v}' \Gamma_{\omega \omega}^{-1} [\Gamma_{\omega \epsilon} - \Gamma_{\omega \omega}] \Gamma_{\omega \omega}^{-1} \mathbf{v} \]

\[ = \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] (\mathbf{X} - \mathbf{X} \mathbf{v}_0) - \mathbf{v}' \Gamma_{\omega \omega}^{-1} [\Gamma_{\omega \epsilon} - \Gamma_{\omega \omega}] \Gamma_{\omega \omega}^{-1} \mathbf{v} \]

\[ = \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] \mathbf{X} - \mathbf{v}' \Gamma_{\omega \omega}^{-1} [\Gamma_{\omega \epsilon} - \Gamma_{\omega \omega}] \Gamma_{\omega \omega}^{-1} \mathbf{v} \]

\[ - \left\{ \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] \mathbf{X} - \mathbf{v}' \Gamma_{\omega \omega}^{-1} [\Gamma_{\omega \epsilon} - \Gamma_{\omega \omega}] \Gamma_{\omega \omega}^{-1} \mathbf{v} \right\} \mathbf{v}_0. \]

If \( \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] \mathbf{X} - \mathbf{v}' \Gamma_{\omega \omega}^{-1} \mathbf{v} \) is nonzero, then this last equality is equivalent to

\[ \mathbf{v}_0 = \left\{ \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] \mathbf{X} - \mathbf{v}' \Gamma_{\omega \omega}^{-1} \mathbf{v} \right\}^{-1} \]

\[ \times \left\{ \mathbf{X}' \Gamma_{\omega \omega}^{-1} [\mathbf{I}_T - \mathbf{I}_T (\mathbf{I}_T \Gamma_{\omega \omega}^{-1} \mathbf{I}_T)^{-1} \mathbf{I}_T \Gamma_{\omega \omega}^{-1}] \mathbf{X} - \mathbf{v}' \Gamma_{\omega \omega}^{-1} \mathbf{v} \right\} \] (6.19)

As for (6.15) above, \( \mathbf{v} \) and \( \Gamma_{\omega \omega} \) are functions of \( \mathbf{v}_0 \), so (6.19) suggests an iterative procedure for the estimation of \( \mathbf{v}_0 \). Let

\[ \mathbf{v}_0^{(i)} = \left( \mathbf{H}_0^{(i-1)} \right)^{(-1)} \mathbf{E}_0^{(i-1)}, \]

where
\[ \tilde{h}(i) = x' \tilde{\Gamma}_{vv(i)}^{-1} [I_T - I_T (I_T' \tilde{\Gamma}_{vv(i)}^{-1} I_T) ^{-1} I_T' \tilde{\Gamma}_{vv(i)}^{-1} ] x \]

\[ - \tilde{\gamma}(i) \tilde{\nu}(i) \tilde{\Gamma}_{vv(i)}^{-1} \tilde{\Gamma}_{uu} \tilde{\nu}(i) \]

\[ \tilde{k}(i) = x' \tilde{\Gamma}_{vv(i)}^{-1} [I_T - I_T (I_T' \tilde{\Gamma}_{vv(i)}^{-1} I_T) ^{-1} I_T' \tilde{\Gamma}_{vv(i)}^{-1} ] y \]

\[ - \tilde{\gamma}(i) \tilde{\nu}(i) \tilde{\Gamma}_{uu} \tilde{\nu}(i) \]

\[ \tilde{\nu}(i) = y - x \tilde{\beta}(i) - I_T \tilde{\beta}_0(i) \]

\[ = [I_T - I_T (I_T' \tilde{\Gamma}_{vv(i)}^{-1} I_T) ^{-1} I_T' \tilde{\Gamma}_{vv(i)}^{-1} ] (y - x \tilde{\beta}(i)) \]

\[ \tilde{\Gamma}_{vv(i)} = \Gamma_{ee} - \tilde{b}(i)(\Gamma_{ue} + \Gamma_{eu}) + \tilde{b}_0^2 \Gamma_{uu} \]

\[ \tilde{\gamma}(i) = \max(0, c(i) - \alpha^{-1}) \]

\( \alpha \) is a preselected positive constant, and \( c(i) \) is the smallest root of the determinantal equation,

\[ |(y, x)' \tilde{\Gamma}_{vv(i)}^{-1} [I_T - I_T (I_T' \tilde{\Gamma}_{vv(i)}^{-1} I_T) ^{-1} I_T' \tilde{\Gamma}_{vv(i)}^{-1} ] (y, x) \]

\[ - c[\tilde{\nu}(i) \tilde{\Gamma}_{vv(i)}^{-1} \tilde{\nu}(i) \tilde{\Gamma}_{uu} \tilde{\nu}(i)\tilde{\Gamma}_{uu}] = 0. \]

Thus, for the single-relation functional model either with or without an intercept, if \( \Gamma_{ee} \) is known, one may obtain an explicit iterative
procedure for the maximum likelihood estimation of regression coefficients.

For \( r = 1 \) and general \( k \), condition (6.14) is equivalent to

\[
0 = \nu' \Gamma^{-1}_{\nu \nu} X_i - \left[ \Gamma_{\nu e i l} - \Gamma_{\nu u i} (\beta \otimes I_T) \right] \Gamma^{-1}_{\nu \nu} v
\]

\[
= (Y - X\beta)' \Gamma^{-1}_{\nu \nu} X_i - \nu' \Gamma^{-1}_{\nu \nu} \left[ \Gamma_{\nu e i l} - \Gamma_{\nu u i} (\beta \otimes I_T) \right] \Gamma^{-1}_{\nu \nu} v
\]

(6.20)

for \( i = 1, 2, \ldots, k \), where

\[
\Gamma_{\nu e i l} = \text{Cov}(u_{i}, e_{i}) \quad \text{is a} \quad T \times T \quad \text{matrix, and}
\]

\[
\Gamma_{\nu u i} = \text{Cov}(u_{i}, \text{vec}(u))
\]

\[
= \begin{bmatrix}
\Gamma_{\nu u 11} & \Gamma_{\nu u 12} & \cdots & \Gamma_{\nu u 1k} \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{\nu u k1} & \Gamma_{\nu u k2} & \cdots & \Gamma_{\nu u kk}
\end{bmatrix}
\quad \text{is a} \quad T \times Tk \quad \text{matrix.}
\]

Since \( \beta \) is now a \( k \times 1 \) vector, one requires additional work to isolate \( \beta \) in expression (6.20). Define

\[
v^* = (v^*_1, v^*_2, \ldots, v^*_k)' = \Gamma^{-1}_{\nu \nu} v \quad \text{and note that}
\]

\[
\Gamma_{\nu u i} (\beta \otimes I_T) = \sum_{k=1}^{k} \Gamma_{\nu u i k} \beta_k.
\]

Then
Therefore, condition (6.20) is equivalent to

\[
0 = x'x^{-1} - v'v^{-1}[(s \in V) - w_{uii}]x^{-1}v
\]

\[
= x'x^{-1} - x'x^{-1}y - v'v^{-1}xw_{uii}x^{-1}v + v'v^{-1}w_{uii}(y \in V)
\]

\[
= [x'x^{-1} - v'v^{-1}xw_{uii}x^{-1}v] - [x'x^{-1} - v'v^{-1}w_{uii}(y \in V)]
\]

(6.21)

for all \( i = 1, 2, \ldots, k \). Note that
and similarly for \( (I_k \oplus v^*) \approx_{uu} \), so condition (6.21) is equivalent to

\[
0_{k \times 1} = [X' \approx_{vv}^{-1} Y - (I_k \oplus \approx_{vv}^{-1} v)' \approx_{uu}^{-1} \approx_{uu} (I_k \oplus \approx_{vv}^{-1} v)] b' .
\]

If \( X' \approx_{vv}^{-1} X - (I_k \oplus \approx_{vv}^{-1} v)' \approx_{uu} (I_k \oplus \approx_{vv}^{-1} v) \) is nonsingular, this final condition is equivalent to

\[
b' = [X' \approx_{vv}^{-1} X - (I_k \oplus \approx_{vv}^{-1} v)' \approx_{uu} (I_k \oplus \approx_{vv}^{-1} v)]^{-1}
\times [X' \approx_{vv}^{-1} Y - (I_k \oplus \approx_{vv}^{-1} v)' \approx_{uu}^{-1} \approx_{uu}^{-1} v] .
\] (6.22)

As for the single-relation case, \( v \) and \( \Gamma_{vv} \) are functions of \( \beta \), so (6.22) gives only an iterative procedure for the maximum likelihood estimation of \( \beta \).

For the case of general \( r \) and \( k \), isolation of \( \beta \) in expression (6.14) requires additional algebra. First, recall that \( \tilde{A}_{ij} \) is an \( r \times k \) matrix with a one in the \((i,j)\)-th position and zeros elsewhere. Let \( \tilde{e}_{ri} \) be an \( r \times 1 \) vector with a one in the \( i \)-th position and zeros elsewhere, and define \( \tilde{e}_{kj} \) similarly. Then \( \tilde{A}_{ij} = \tilde{e}_{ri} \tilde{e}_{kj} \). Define \( \tilde{T}_{vv} = \approx_{vv}^{-1} \), a \( Tr \times Tr \) matrix with \((i,j)\)-th \( T \times T \) block equal to
Let $T_{vvij}$, which in turn has $(s,t)$-th element equal to $T_{vvij}(s,t)$. Let $\Gamma_{vv.i} = T_{vv}((e_i \otimes I_T)$, a $Tr \times T$ matrix equal to the $i$-th $Tr \times T$ block of $T_{vv}$, $i=1, 2, ..., r$. Similarly, let $\Gamma_{uu.j} = (e_k \otimes I_T)\Gamma_{uu}$ and $\Gamma_{ue.j} = (e_k \otimes I_T)\Gamma_{ue}$. Also, define $v^* = (v_{*1}, v_{*2}, ..., v_{*r})$ to be a $T \times r$ matrix such that

$$\text{vec}(v^*) = (v_{*1}^t, v_{*2}^t, ..., v_{*r}^t)^t = \Gamma_{vv}^{-1}\text{vec}(v);$$

and let

$$w^* = (v^*_{*1}, v^*_{*2}, ..., v^*_{*r})^t,$$

and

$$w^*_j = (v^*_{*j1}, v^*_{*j2}, ..., v^*_{*jT})^t,$$

$$= (e_{rj} \otimes I_T)\Gamma_{vv}^{-1}v$$

$$= (e_{rj} \otimes I_T)\Gamma_{vv}^{-1}\text{vec}(v) = T_{vv.j} \text{vec}(w).$$

Condition (6.14) is then equivalent to

$$0 = \text{vec}(v)^t \Gamma_{vv.j} X_i - v^*_j [\Gamma_{ue.i} - \Gamma_{uu.i} (\beta \otimes I_T)]\text{vec}(v^*).$$

(6.23)

for all $i = 1, 2, ..., k$ and $j = 1, 2, ..., r$. Now
where $\beta_{\ell}$ is the $\ell$-th row of $\beta$. Hence,

$$
\lambda^{\ast \ast} \Gamma_{\ast \ast \ast \ast \ast} (\beta \ast I_T) \text{vec}(v^\ast)
= \sum_{\ell=1}^{k} \lambda^{\ast \ast \ast \ast \ast} (\beta_{\ell} \ast I_T) \text{vec}(v^\ast)
= \sum_{\ell=1}^{k} \sum_{r=1}^{m} \lambda^{\ast \ast \ast \ast \ast} (\beta_{\ell} \ast I_T) \text{vec}(v^\ast)
= \sum_{r=1}^{m} \sum_{m=1}^{r} \lambda^{\ast \ast \ast \ast \ast} (I_k \ast v^\ast) \beta_{m}
= \sum_{r=1}^{m} \lambda^{\ast \ast \ast \ast \ast} (I_k \ast v^\ast) \text{vec}(\beta)
= \lambda^{\ast \ast \ast \ast \ast} \text{vec}(\beta),
$$

where

$$
\tilde{v}^\ast = [I_k \ast v^\ast_1, I_k \ast v^\ast_2, \ldots, I_k \ast v^\ast_r]
= [I_k \ast [T_{v_1v_1} \text{vec}(v)], I_k \ast [T_{v_2v_2} \text{vec}(v)],
\ldots, I_k \ast [T_{vrv_r} \text{vec}(v)]}
= [I_k \ast T_{v_1v_1}, I_k \ast T_{v_2v_2}, \ldots, I_k \ast T_{vrv_r}] [I_{r_k} \ast \text{vec}(v)]
$$
\[ \vec{v} = T_{\nu v} [ I_{rk} \otimes \text{vec}(v) ] , \]

and

\[ T_{\nu v} = [ I_k \otimes T_{\nu v1} , I_k \otimes T_{\nu v2} , \ldots , I_k \otimes T_{\nu vr} ] . \]

Also,

\[ \text{vec}(v)' T_{\nu v,j} X_{,i} = X_{,i}' T_{\nu v,j} \text{vec}(Y - XB) \]

\[ = X_{,i}' T_{\nu v,j} [ \text{vec}(Y) - (I_r \otimes X) \text{vec}(\delta) ] . \]

Thus, condition (6.23) is equivalent to

\[ 0 = X_{,i}' T_{\nu v,j} [ \text{vec}(Y) - (I_r \otimes X) \text{vec}(\delta) ] \]

\[ - v^* T_{\nu v,j} \text{vec}(v^*) + v^* T_{\nu v,j} \text{vec}(\delta) \]

\[ = [ X_{,i}' T_{\nu v,j} \text{vec}(Y) - v^* T_{\nu v,j} \text{vec}(v^*) ] \]

\[ - [ X_{,i}' T_{\nu v,j} (I_r \otimes X) - v^* T_{\nu v,j} \text{vec}(\delta) ] vec(\delta) \quad (6.25) \]

for all \( i = 1, 2, \ldots, k \) and \( j = 1, 2, \ldots, r \). Now

\[ X_{,i}' T_{\nu v,j} [ \text{vec}(Y) - (I_r \otimes X) \text{vec}(\delta) ] \]
is the \((j,i)\)-th element of

\[
(I_\Gamma \bowtie X)' \gamma_{\nu \nu} [\vec{v}(\gamma) - (I_\Gamma \bowtie X) \vec{v}(\gamma)];
\]

\[\tilde{\nu}_* \Gamma . j^\alpha e_i . \vec{v}(\nu^*)\]

is the \((j,i)\)-th element of

\[\tilde{\nu}_* \Gamma . \vec{v}(\nu^*)\]

and

\[\tilde{\nu}_* \Gamma . j^\alpha u_i . \vec{v}(\nu^*)\]

is the \((j,i)\)-th element of

\[\tilde{\nu}_* \Gamma \tilde{\nu}_* \vec{v}(\nu^*)\]

Hence, condition (6.25) is equivalent to

\[
(I_\Gamma \bowtie X)' \gamma_{\nu \nu}^{-1} [\vec{v}(\gamma) - \tilde{\nu}_* \Gamma \vec{v}(\nu^*)] = 0_{r^k \times 1}.
\]  

(6.26)
If
\[
M_1 = (I_r \otimes X)' \tilde{\Gamma}_{\nu \nu}^{-1} (I_r \otimes X) - \tilde{\nu}^* \tilde{\Gamma}_{uu} \tilde{\nu}^*
\]
is nonsingular, then condition (6.26) is equivalent to
\[
vec(\hat{\beta}) = [(I_r \otimes X)' \tilde{\Gamma}_{\nu \nu}^{-1} (I_r \otimes X) - \tilde{\nu}^* \tilde{\Gamma}_{uu} \tilde{\nu}^*]^{-1}
\times [(I_r \otimes X)' \tilde{\Gamma}_{\nu \nu}^{-1} vec(y) - \tilde{\nu}^* \tilde{\Gamma}_{ue} vec(\nu^*)]
\]
\[
= [(I_r \otimes X)' \tilde{\Gamma}_{\nu \nu}^{-1} (I_r \otimes X) - [I_{rk} \otimes vec(\nu)]' \tilde{\Gamma}_{\nu \nu}^{-1} \Gamma_{\nu \nu} \tilde{\Gamma}_{\nu \nu}^{-1} [I_{rk} \otimes vec(\nu)]]^{-1}
\times [(I_r \otimes X)' \tilde{\Gamma}_{\nu \nu}^{-1} vec(y) - [I_{rk} \otimes vec(\nu)]' \tilde{\Gamma}_{\nu \nu} \tilde{\Gamma}_{\nu \nu}^{-1} \Gamma_{\nu \nu} vec(\nu)]
\]
\[
(6.27)
\]
As for the single-relation model, the dependence of \( \nu \) and \( \Gamma_{\nu \nu} \)
on \( \beta \) implies that the estimation of \( \beta \) through (6.27) may be pursued
through an iterative procedure. Let
\[
vec(\hat{\beta}_i) = [\hat{R}_{(i-1)}^{-1}]' \hat{R}_{(i-1)} = \hat{R}_{(i-1)}^{-1} \hat{R}_{(i-1)},
\]
(6.28)
where
\[
\hat{R}_{(i)} = (I_r \otimes X)' [\tilde{\Gamma}_{\nu \nu(i)}]^{-1} (I_r \otimes X) - \lambda^{(i)} \tilde{\nu}^* (i) \tilde{\Gamma}_{uu} \tilde{\nu}^* (i)
\]
\[
\hat{H}_{(i)} = (I_r \otimes X)' [\tilde{\Gamma}_{\nu \nu(i)}]^{-1} vec(\tilde{\nu}^*(i)) - \lambda^{(i)} \tilde{\nu}^* (i) \tilde{\Gamma}_{ue} vec(\tilde{\nu}^*(i)),
\]
\[ \text{vec}[\hat{\psi}(i)] = [\hat{\Gamma}_{\text{vv}}(i)]^{-1}\text{vec}[Y - \hat{X}_g(i)], \]

\[ \hat{\Gamma}_{\text{vv}}(i) = \Gamma_{ee} - [\hat{g}(i) \otimes I_T]'\Gamma_{uu} - \Gamma_{eu}[\hat{g}(i) \otimes I_T] \]

\[ + [\hat{g}(i) \otimes I_T]'\Gamma_{ee}[\hat{g}(i) \otimes I_T], \]

\( \hat{\psi}(i) \) is defined from \( \text{vec}[\hat{\psi}(i)] \) by equation (6.24),

\[ \hat{\lambda}(i) = \max(0, c(i) - \alpha^{-1}), \]

\( \alpha \) is a preselected positive constant, and \( c(i) \) is the smallest root of the determinantal equation,

\[ |[\text{vec}(Y), (I_r \otimes X)]'[\hat{\Gamma}_{\text{vv}}(i)]^{-1}[\text{vec}(Y), (I_r \otimes X)] \]

\[ - c[\text{vec}[\hat{\psi}], \hat{\psi} _{<}(i)]'\Gamma_{ee}[\text{vec}[\hat{\psi}], \hat{\psi} _{>}(i)]| = 0. \]

To assess the behavior of estimator (6.28), note that the \([(j,i), (m,j)]\)-th element of \((I_r \otimes X)'\hat{\Gamma}_{\text{vv}}^{-1}(I_r \otimes X)\) equals

\[ X'_i \Gamma_{\text{vvjm}} X_{.j} \], where \( \Gamma_{\text{vvjm}} \) is a \( T \times T \) matrix defined analogously to \( \Gamma_{\text{vv.m}} \). Also, the \([(j,i), (m,j)]\)-th element of \( \hat{\psi}^* \Gamma_{uu} \hat{\psi}^* \) equals

\[ \hat{\psi}^* _{.j} \Gamma_{uu} \hat{\psi}^* _{.m} \], so the \([(j,i), (m,j)]\)-th element of \( M_1 \) equals

\[ X'_i \Gamma_{\text{vvjm}} X_{.j} - \psi^* _{.j} \Gamma_{uu} \psi^* _{.m} \]. On the other hand, the \([(j,i), (m,j)]\)-th element of \( 2^{-1}A \) equals
\[
\left[\frac{\partial}{\partial \beta_{ij}} (\nu)\right] \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{km}} (\nu)\right] \\
- v' \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{ij}} (\Gamma_{vv})\right] \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{km}} (\nu)\right] + \left[\frac{\partial}{\partial \beta_{km}} (\Gamma_{vv})\right] \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{ij}} (\nu)\right] \\
+ v' \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{ij}} (\Gamma_{vv})\right] \Gamma_{vv}^{-1} \left[\frac{\partial}{\partial \beta_{km}} (\nu)\right] - 2^{-1} \left[\frac{\partial^2}{\partial \beta_{ij} \partial \beta_{km}} (\Gamma_{vv})\right] \Gamma_{vv}^{-1} v \\
= x' \left(\tilde{A}_{ij} \otimes I_T\right) \Gamma_{vv} \left(\tilde{A}_{km} \otimes I_T\right) x \\
- v^* \left(\left[\tilde{A}_{ij} \otimes I_T\right] \Gamma_{ue} - \Gamma_{uu} (\beta \otimes I_T)\right) \\
+ \left[\Gamma_{ue} - (\beta' \otimes I_T) \Gamma_{uu}\right] \left[\tilde{A}_{ij} \otimes I_T\right] \Gamma_{vv} \left[\tilde{A}_{km} \otimes I_T\right] x \\
+ \left[\tilde{A}_{km} \otimes I_T\right] \left[\Gamma_{ue} - \Gamma_{uu} (\beta \otimes I_T)\right] + \left[\Gamma_{ue} - (\beta' \otimes I_T) \Gamma_{uu}\right] x \left[\tilde{A}_{km} \otimes I_T\right] \Gamma_{vv} \left[\tilde{A}_{ij} \otimes I_T\right] x \\
+ v^* \left(\left[\tilde{A}_{ij} \otimes I_T\right] \Gamma_{ue} - \Gamma_{uu} (\beta \otimes I_T)\right) \\
+ \left[\Gamma_{ue} - (\beta' \otimes I_T) \Gamma_{uu}\right] \left[\tilde{A}_{ij} \otimes I_T\right] \\
+ \Gamma_{vv} \left[\tilde{A}_{km} \otimes I_T\right] \left[\Gamma_{ue} - \Gamma_{uu} (\beta \otimes I_T)\right] + \left[\Gamma_{ue} - (\beta' \otimes I_T) \Gamma_{uu}\right] x \left[\tilde{A}_{km} \otimes I_T\right] \\
- 2^{-1} \left[\tilde{A}_{ij} \otimes I_T\right] \Gamma_{uu} \left[\tilde{A}_{km} \otimes I_T\right] + \left[\tilde{A}_{km} \otimes I_T\right] \Gamma_{uu} \left[\tilde{A}_{ij} \otimes I_T\right]\right) v^*
Now

\[ E\{v^*_j \Gamma_{uei} - \Gamma_{uui} (g \circ I_T) \}_{\tilde{\varphi}} \cdot X \}

= E\{w^*_{ij} \tilde{\varphi}(g_{ij} \circ I_T) \tilde{\varphi}^{-1}(g_{ik} \circ I_T)(x + u) \}

= \text{tr} (E(uv') \tilde{\varphi}^{-1}(g_{ij} \circ I_T) \tilde{\varphi}^{-1}(g_{ik} \circ I_T))
Similar relations hold for the other six similar terms of (6.29). It follows that (6.29) has the same expectation as

\[ x_i'X_{i\cdot}^{-1} - v^*_j\Gamma_{v^*_j\cdot}^{-1} \]

so that under the conditions of Theorem 4.1, \( T^{-1}H_{(i)} \) is a consistent estimator of \( T^{-1}2^{-1}E(A) \). Also, inspection of the equations preceding (6.26) indicates that \( \hat{H}_{(i-1)} - \hat{H}_{(i-1)}\text{vec}(\hat{\beta}_{(i)}) \) is an estimator of \( a \). It follows that (6.28) defines an approximate Newton-Raphson algorithm for maximum likelihood estimation of \( \beta \).
6.2. Maximum Likelihood Estimation Through a Modified State-Space Approach

The preceding section outlined a general approximate Newton-Raphson approach to maximum likelihood estimation for the functional model. After using the transformation \( \mathbf{v}_t = \mathbf{Y}_t - \mathbf{X}_t \mathbf{\beta} = \mathbf{e}_t - \mathbf{u}_t \mathbf{\beta} \) to "remove" the fixed \( \mathbf{x}_t \) vector from the observations, the reduced likelihood function \( L_m(\mathbf{\alpha}; \mathbf{Z}) \) for the model with no intercept satisfies the equations

\[
-2 \ln L_m(\mathbf{\alpha}; \mathbf{Z}) = \text{Tr} [\mathbf{Y}' \mathbf{Y} - (\mathbf{I}_T \otimes \mathbf{X}) \mathbf{V}]
\]

Direct computation of the resulting matrices of first and second derivatives proceeded without conceptual difficulty, but required inversion of \( T_p \times T_p \) matrices and other computations that may present numerical problems for moderate to large \( T \). Therefore, one may wish to modify the state-space approach of Section 5.4 to obtain an alternative computational procedure for maximum likelihood estimation. Recall that in Section 5.4, the state-space approach to the likelihood function allowed one to write the determinant \( |\mathbf{I}_{22}| \) as the product
of $T$ determinants $|\Lambda_{ct}|$, where each $\Lambda_{ct}$ was $p \times p$; and to write the quadratic form $\vec(Z)' \Sigma^{-1}_{ZZ} \vec(Z)$ as the sum of $T$ quadratic forms $d_t \Lambda_{ct}^{-1} d_t'$, where $\{\Lambda_{ct}\}$ is the same set of $p \times p$ matrices mentioned above, and $\{d_t\}$ is a set of $1 \times p$ random vectors. Since the covariance matrices in the log-determinant and quadratic-form parts of expression (6.30) do not match as they did in Section 5.4, a decomposition of expression (6.30) will require two separate steps. The first step will lead to the "block diagonalization" of $\Sigma_{ee}$, so that

$$\ln |\Sigma_{ee}| = \sum_{t=1}^{T} \ln |\Lambda_{ct}|,$$

where each $\Lambda_{ct}$ is a $p \times p$ matrix. The second step will lead to the transformation

$$\vec(v)' \Sigma^{-1}_{vv} \vec(v) = \sum_{t=1}^{T} d_t \Lambda_{vt}^{-1} d_t',$$

where $\{d_t\}$ is an "innovation sequence" associated with $\vec(v)$, and $\Lambda_{vt} = \text{var}(d_t)$, $t=1, 2, \ldots, T$.

Assume that $z_t$ follows an ARMA$_p(p_{z}, q_{z})$ process,

$$\sum_{j=0}^{p_{z}} \alpha_{zj} z_{t-j} = \sum_{i=0}^{q_{z}} \theta_{zj} c_{t-i}, \quad (6.31)$$

say, and consider first the decomposition of $\ln |\Sigma_{ee}|$. If $z_t$ were observed, the resulting state-space model would be a special case of model (5.38)-(5.39), i.e.,
\( \varepsilon_t' = B \varepsilon_t' \), \( (6.32) \)

\[
W_t' = A \varepsilon_t' + C \varepsilon_t', \quad t = 1, 2, \ldots, T,
\]

where

\[
W_t = (\varepsilon_1|t, \varepsilon_{t+1}|t, \ldots, \varepsilon_{t+M-1}|t);
\]

\[B_{\varepsilon} = [I_p, 0_p \times (M-1)_k];\]

\[
A_{\varepsilon} = \begin{bmatrix}
0 & I_p & 0 & 0 & \cdots & 0 \\
0 & 0 & I_p & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & I_p \\
-\hat{\phi}_{eM} & -\hat{\phi}_{e,M-1} & -\hat{\phi}_{e,M-2} & -\hat{\phi}_{e,M-3} & \cdots & -\hat{\phi}_{e1}
\end{bmatrix};
\]

\[C_{\varepsilon} = [\psi_{\varepsilon0}', \psi_{\varepsilon1}', \ldots, \psi_{\varepsilon,M-1}'];\]

\[\psi_{\varepsilon0} = I_p;\]

\[\psi_{\varepsilon i} = \theta_{\varepsilon i} - \sum_{j=1}^{p_{\varepsilon}} \phi_{\varepsilon j} \psi_{\varepsilon, i-j}, \quad i = 1, 2, \ldots, M-1; \quad (6.34)\]

\[\psi_{\varepsilon i} = 0 \text{ for } i < 0;\]
$\theta_{\epsilon i} = 0$ for $i > q_\epsilon$;

$\phi_{\epsilon j} = 0$ for $j > p_\epsilon$;

$e_{t+1|t}$ is the $i$ step ahead linear predictor of $e_{t+1}$, and

$M = \max(p_\epsilon, q_\epsilon + 1)$. The results of Section 5.3 then indicate that

$$\ln|\Gamma_{\sim \epsilon \epsilon}| = \sum_{t=1}^{T} \ln|\Lambda_{\sim t}|,$$

(6.35)

where

$$\Lambda_{\sim t} = B\varepsilon_t |t-1| B',$$

$$P_{t|t-1} = A\varepsilon_t |t-1| A' + C\varepsilon_{cc} C',$$

(6.36)

$$\Sigma_{cc} = \text{Var}(e'_t)$$

and

$$P_t = P_t |t-1| - P_t |t-1| B'\Lambda_{\sim t}^{-1} B P_t |t-1|.$$  

(6.37)

Hence, by Result 9.8,

$$\frac{\partial}{\partial \alpha_{\epsilon l}} [\ln|\Gamma_{\sim \epsilon \epsilon}|] = \sum_{t=1}^{T} \frac{\partial}{\partial \alpha_{\epsilon l}} [\ln|\Lambda_{\sim t}|] = \sum_{t=1}^{T} \text{tr}\left\{\Lambda_{\sim t}^{-1} \frac{\partial}{\partial \alpha_{\epsilon l}} (\Lambda_{\sim t})\right\}$$

and
As in Section 5.4, one may now obtain iterative expressions for the first and second derivatives of $A_{e\epsilon t}$ with respect to the elements of $e_{e\epsilon}$. First, note that $B C = I$, so

$$A_{e\epsilon t} = B A_{e\epsilon t-1} A'B' + \Sigma_{e\epsilon\epsilon}.$$  \hspace{1cm} (6.38)

Moreover, $B A_{e\epsilon} = [O_{p \times p}$ $I, O_{p \times (M - 2)p}]$, so

$$\frac{\partial A_{e\epsilon t}}{\partial e_{e\epsilon i}} = B e_{e\epsilon t-1} A'B' + \frac{\partial \Sigma_{e\epsilon\epsilon}}{\partial e_{e\epsilon i}}$$

and

$$\frac{\partial^2 A_{e\epsilon t}}{\partial e_{e\epsilon i} \partial e_{e\epsilon j}} = B e_{e\epsilon t-1} A'B' + \frac{\partial^2 \Sigma_{e\epsilon\epsilon}}{\partial e_{e\epsilon i} \partial e_{e\epsilon j}}.$$  

By Result 9.4,

$$\frac{\partial A_{e\epsilon t-1}}{\partial e_{e\epsilon i}} = \frac{\partial A_{e\epsilon}}{\partial e_{e\epsilon i}} t-1 A'e + A e_{e\epsilon t-1} A' + A e_{e\epsilon t-1} (\frac{\partial A_{e\epsilon}}{\partial e_{e\epsilon i}})' + \frac{\partial \Sigma_{e\epsilon\epsilon}}{\partial e_{e\epsilon i}} C'$$

$$+ C e_{e\epsilon} (\frac{\partial C_{e\epsilon\epsilon}}{\partial e_{e\epsilon i}})' + C e_{e\epsilon\epsilon} (\frac{\partial C_{e\epsilon\epsilon}}{\partial e_{e\epsilon i}})'$$  \hspace{1cm} (6.39)

and
\[
\frac{\partial^2 P_t}{\partial a_{ei} \partial a_{ej}} = \left(\frac{\partial^2 A}{\partial a_{ei} \partial a_{ej}}\right) + \left(\frac{\partial^2 P_t}{\partial a_{ei} \partial a_{ej}}\right)A' + \left(\frac{\partial^2 P_t}{\partial a_{ei} \partial a_{ej}}\right)A' + \left(\frac{\partial^2 P_t}{\partial a_{ei} \partial a_{ej}}\right) + \left(\frac{\partial^2 C}{\partial a_{ei} \partial a_{ej}}\right)C' + \left(\frac{\partial^2 C}{\partial a_{ei} \partial a_{ej}}\right)C' + \left(\frac{\partial^2 C}{\partial a_{ei} \partial a_{ej}}\right)C' + \left(\frac{\partial^2 C}{\partial a_{ei} \partial a_{ej}}\right)
\]

(6.40)

because \( \frac{\partial^2 A}{\partial a_{ei} \partial a_{ej}} \) and \( \frac{\partial^2 C}{\partial a_{ei} \partial a_{ej}} \) are both null matrices. Similarly,

\[
\frac{\partial P_t}{\partial a_{ei}} = \left(\frac{\partial P_t}{\partial a_{ei}}\right) - \left(\frac{\partial P_t}{\partial a_{ei}}\right)B' \cdot A^{-1}B \cdot P_t | t-1 + \left(\frac{\partial P_t}{\partial a_{ei}}\right)B' \cdot A^{-1}B \cdot P_t | t-1 - \left(\frac{\partial P_t}{\partial a_{ei}}\right)B' \cdot A^{-1}B \cdot \left(\frac{\partial P_t}{\partial a_{ei}}\right)
\]

(6.41)
\[
\frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} = \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right)_t - \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right) B_t \Lambda_{et}^{-1} B_t e P_t | t-1 \\
- \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right) B_t \left[ \Lambda_{et}^{-1} \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} B_t e P_t | t-1 + \Lambda_{et}^{-1} B_t e \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right) \right] \\
+ \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right) B_t \Lambda_{et}^{-1} \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} B_t e P_t | t-1 \\
+ P_t | t-1 \Lambda_{et} B_t e \Lambda_{et}^{-1} \left[ \Lambda_{et} \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} B_t e P_t | t-1 \\
+ \left( \frac{\partial^2 \Lambda_{et}}{\partial \alpha_i \partial \alpha_j} \right) \Lambda_{et}^{-1} B_t e P_t | t-1 - \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} \left( \frac{\partial \Lambda_{et}}{\partial \alpha_j} \right) \Lambda_{et}^{-1} B_t e P_t | t-1 \\
+ \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} B_t e \left( \frac{\partial p_t}{\partial \alpha_j} \right) - \left( \frac{\partial p_t}{\partial \alpha_j} \right) B_t \Lambda_{et}^{-1} B_t e \left( \frac{\partial p_t}{\partial \alpha_j} \right) \\
- P_t | t-1 \Lambda_{et} B_t e \left( \frac{\partial \Lambda_{et}}{\partial \alpha_i} \right) \Lambda_{et}^{-1} B_t e \left( \frac{\partial p_t}{\partial \alpha_i} \right) + B_t e \left( \frac{\partial^2 p_t}{\partial \alpha_i \partial \alpha_j} \right) \right]. \\
(6.42)
\]

Now consider the decomposition of \( \text{vec}(v)' \Gamma_{vv}^{-1} \text{vec}(v) \). Since

\[ v_t' = e_t' - \bar{\theta}' u_t' = (I_r, -\bar{\theta}') e_t', \]

a state-space model for the \( v_t \) is the same as for \( e_t \), except that the measurement equation (6.32) is replaced by

\[ v_t' = B_t W_t, \]

(6.43)
where

\[ B_v = [(I_r, -g'), \, 0_{r \times (M-1)p}] . \]

A second application of the results of Section 5.3 then indicates that

\[ \text{vec}(v)' \Gamma^{-1} \text{vec}(v) \overset{\sim}{=} \sum_{t=1}^{T} d \Lambda^{-1} d' \]  \hspace{1cm} (6.44)

where

\[ d'_t = v'_t - v'_t | t \]

\[ \hat{w}'_{t+1} | t = A \hat{w}'_t , \hspace{1cm} (6.46) \]

\[ \hat{w}'_t = \hat{w}'_{t-1} + R \hat{w}'_{t-1} B' \Lambda^{-1} d' \]

\[ A_{vt} = B_v R | t-1 B_v' \]

\[ R | t-1 = A_v R | t-1 A_v' + C_v C_v C_v' \]  \hspace{1cm} (6.49)

and

\[ R_t = R_t | t-1 - R_t | t-1 B' \Lambda^{-1} B R_t | t-1 . \]  \hspace{1cm} (6.50)
Hence, by Result 9.7,

\[
\frac{\partial}{\partial \alpha_i} [\text{vec}(v)' \Gamma^{-1} \text{vec}(v)]
\]

\[
= \sum_{t=1}^{T} \frac{\partial}{\partial \alpha_i} [d_t \Lambda_t^{-1} d_t']
\]

\[
= \sum_{t=1}^{T} \left[ \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} d_t' + d_t \Lambda_t^{-1} \left( \frac{3d_t}{\partial \alpha_i} \right)' - d_t \Lambda_t^{-1} \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} d_t' \right]
\]

and

\[
\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} [\text{vec}(v)' \Gamma^{-1} \text{vec}(v)]
\]

\[
= \sum_{t=1}^{T} \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \left( d_t \Lambda_t^{-1} d_t' \right)
\]

\[
= \sum_{t=1}^{T} \left[ \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} d_t' + \left( \frac{3d_t}{\partial \alpha_j} \right) \Lambda_t^{-1} d_t' - \left( \frac{3d_t}{\partial \alpha_i} \right) \left( \frac{3d_t}{\partial \alpha_j} \right)' + \left( \frac{3d_t}{\partial \alpha_j} \right) \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} d_t' \right]
\]

\[
+ \left[ \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} \left( \frac{3d_t}{\partial \alpha_j} \right) \Lambda_t^{-1} d_t' - \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} \left( \frac{3d_t}{\partial \alpha_j} \right)' \right]
\]

\[
= \sum_{t=1}^{T} \left[ \left( \frac{3d_t}{\partial \alpha_i} \right) \Lambda_t^{-1} \left( \frac{3d_t}{\partial \alpha_j} \right) \Lambda_t^{-1} d_t' \right]
\]

As for \( \Lambda_{ct} \), one may obtain iterative expressions for the first and second derivatives of \( \Lambda_{vt} \) with respect to the elements of \( \alpha_c \) and \( \beta_c \), namely,
\[
\begin{align*}
\frac{\partial^2 \mathcal{A}_{vt}}{\partial \alpha_i \partial \alpha_j} &= B_v \left( \frac{3R_v}{\partial \alpha_i} \right) B'_v + R_v \left( \frac{3R_v}{\partial \alpha_i} \right) B'_v + R_v \left( \frac{3R_v}{\partial \alpha_i} \right) \left( \frac{3R_v}{\partial \alpha_j} \right) \left( \frac{3R_v}{\partial \alpha_j} \right) ; \\
\frac{\partial^2 \mathcal{A}_{vt}}{\partial \beta_{ij} \partial \beta_{km}} &= B_v \left( \frac{3R_v}{\partial \beta_{ij}} \right) B'_v + R_v \left( \frac{3R_v}{\partial \beta_{ij}} \right) B'_v + R_v \left( \frac{3R_v}{\partial \beta_{ij}} \right) \left( \frac{3R_v}{\partial \beta_{km}} \right) ; \end{align*}
\]

Since the algebraic structure of formulas (6.49) and (6.50) is the same as in formulas (6.36) and (6.37), respectively, and since \( B_v \) is not a function of \( \alpha \), it follows that

\[
\frac{3R_v}{\partial \alpha_i} , \frac{3R_v}{\partial \alpha_i} , \frac{3R_v}{\partial \alpha_i} , \text{ and } \frac{3R_v}{\partial \alpha_i \partial \alpha_j}
\]

are given by formulas (6.39), (6.40), (6.41) and (6.42), respectively, but with the replacement of \( P_v \) by \( R_v \); \( P_v \) by \( R_v \); \( \Lambda_{et} \) by \( \Lambda_{vt} \); and \( B_v \) by \( B_v \). Also,

\[
\frac{3R_v}{\partial \beta_{ij}} = \Lambda_v \left( \frac{3R_v}{\partial \beta_{ij}} \right) A'_v ;
\]
\[
\frac{\partial^2 R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} = \Delta e \frac{\partial^2 R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \cdot \Delta e ;
\]
\[
\frac{\partial^2 R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} = \frac{\partial^2 R_t | t-1}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial^2 (B_v R_t | t-1)}{\partial \beta_{ij}} = \frac{\partial^2 R_t | t-1}{\partial \beta_{ij}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij}} \Delta e ;
\]
\[
\frac{\partial^2 (B_v R_t | t-1)}{\partial \beta_{ij} \partial \alpha_{x \ell}} = \frac{\partial^2 R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial R_t | t-1}{\partial \beta_{ij}} = \frac{\partial (B_v R_t | t-1)}{\partial \beta_{ij}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij}} \Delta e ;
\]
\[
\frac{\partial^2 R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} = \frac{\partial^2 R_t | t-1}{\partial \alpha_{x \ell} \partial \beta_{ij}} \frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \beta_{ij} \partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell}} = \frac{\partial (B_v) R_t | t-1}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial^2 (B_v R_t | t-1)}{\partial \alpha_{x \ell} \partial \beta_{ij}} = \frac{\partial^2 (B_v R_t | t-1)}{\partial \alpha_{x \ell} \partial \beta_{ij}} \Delta e + \frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell} \partial \beta_{ij}} \Delta e ;
\]
\[
\frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell}} = \frac{\partial (B_v) R_t | t-1}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell}} = \frac{\partial (B_v) R_t | t-1}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \alpha_{x \ell}} \Delta e ;
\]
\[
\frac{\partial (B_v R_t | t-1)}{\partial \alpha_{x \ell}} = \frac{\partial (B_v) R_t | t-1}{\partial \alpha_{x \ell}} \Delta e + \frac{\partial R_t | t-1}{\partial \alpha_{x \ell}} \Delta e ;
\]
where, as indicated above, \( a_z \) may be any element of
\[ a = \text{vec}(G)' , \ a_z \].

Finally, consider the derivatives associated with the innovation
vector \( d_{t+1}' \) defined by (6.45) and the predicted state vectors \( \hat{\mathbf{W}}_{t+1} | t \)
and \( \hat{\mathbf{W}}_t \) defined by (6.46) and (6.47), respectively. Now

\[ B_v A_e = [O_{r \times p} , \ (I_r , -\beta) , \ 0_{r \times (n-2)p}]\]

\[ = D_v , \]

say, so by (6.45), (6.46) and the identity \( v_t' = y_t' - \beta' x_t' \),

\[ d_{t+1}' = v_{t+1}' - B_v A_e \hat{w}_t \]

\[ = y_{t+1}' - \beta' x_{t+1}' - D_v \hat{w}_t . \]  \hspace{1cm} (6.51)

Thus,

\[ \frac{\partial d_{t+1}'}{\partial \beta_{ij}} = -\left(\frac{\partial \beta}{\partial \beta_{ij}}\right)' x_{t+1}' - \left(\frac{\partial D_v}{\partial \beta_{ij}}\right)' \hat{w}_t - D_v \left(\frac{\partial \hat{w}_t}{\partial \beta_{ij}}\right)' ; \]

\[ \frac{\partial d_{t+1}'}{\partial a_{\varepsilon z}} = - D_v \left(\frac{\partial \hat{w}_t}{\partial a_{\varepsilon z}}\right)' ; \]
\[
\frac{\partial^2 d'_{t+1}}{\partial \beta_{ij} \partial \beta_{km}} = - \left( \frac{\partial d_{V}}{\partial \beta_{ij}} \right) \left( \frac{\partial \hat{W}_{t}'}{\partial \beta_{km}} \right)' - \left( \frac{\partial d_{V}}{\partial \beta_{km}} \right) \left( \frac{\partial \hat{W}_{t}'}{\partial \beta_{ij}} \right)' - D_{V} \left( \frac{\partial^2 \hat{W}_{t}'}{\partial \beta_{ij} \partial \beta_{km}} \right) ;
\]

\[
\frac{\partial^2 d'_{t+1}}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} = - D_{V} \left( \frac{\partial^2 \hat{W}_{t}'}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} \right) ;
\]

and

\[
\frac{\partial^2 d'_{t+1}}{\partial \beta_{ij} \partial \alpha_{\epsilon L}} = - \left( \frac{\partial d_{V}}{\partial \beta_{ij}} \right) \left( \frac{\partial \hat{W}_{t}'}{\partial \alpha_{\epsilon L}} \right)' - D_{V} \left( \frac{\partial \hat{W}_{t}'}{\partial \beta_{ij} \partial \alpha_{\epsilon L}} \right)' .
\]

Routine application of Results 9.5 and 9.7 also implies that

\[
\frac{\partial \hat{W}_{t+1}'}{\partial \alpha_{\epsilon L}} = \left( \frac{3 \alpha_{\epsilon L}}{\partial \alpha_{\epsilon L}} \right) \hat{W}_{t} + A \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon L}} \right) ;
\]

\[
\frac{\partial \hat{W}_{t+1}'}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} = \left( \frac{3 \alpha_{\epsilon L}}{\partial \alpha_{\epsilon L}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M}} \right) + \left( \frac{3 \alpha_{\epsilon M}}{\partial \alpha_{\epsilon M}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon L}} \right) + A \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M} \partial \alpha_{\epsilon L}} \right) ;
\]

\[
\frac{\partial \hat{W}_{t+1}'}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} = \left( \frac{3 \alpha_{\epsilon L}}{\partial \alpha_{\epsilon L}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M}} \right) + \left( \frac{3 \alpha_{\epsilon M}}{\partial \alpha_{\epsilon M}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon L}} \right) + A \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M} \partial \alpha_{\epsilon L}} \right) ;
\]

\[
\frac{\partial \hat{W}_{t+1}'}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} = \left( \frac{3 \alpha_{\epsilon L}}{\partial \alpha_{\epsilon L}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M}} \right) + \left( \frac{3 \alpha_{\epsilon M}}{\partial \alpha_{\epsilon M}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon L}} \right) + A \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M} \partial \alpha_{\epsilon L}} \right) ;
\]

and

\[
\frac{\partial \hat{W}_{t+1}'}{\partial \alpha_{\epsilon L} \partial \alpha_{\epsilon M}} = \left( \frac{3 \alpha_{\epsilon L}}{\partial \alpha_{\epsilon L}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M}} \right) + \left( \frac{3 \alpha_{\epsilon M}}{\partial \alpha_{\epsilon M}} \right) \hat{W}_{t} \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon L}} \right) + A \left( \frac{\partial \hat{W}_{t}}{\partial \alpha_{\epsilon M} \partial \alpha_{\epsilon L}} \right) ;
\]
Hence, expressions (6.35) and (6.44) imply that expression (6.30) equals

\[
\frac{\partial^2 (B' \Delta^{-1} d')}{\partial \alpha \partial \alpha_m} = \left( \frac{\partial B}{\partial \alpha_m} \right)' \Delta_{tt}^{-1} \left[ -\left( \frac{\partial \Delta_{tt}}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' + \left( \frac{\partial d'}{\partial \alpha_m} \right) \right] \\
+ \left( \frac{\partial B}{\partial \alpha_m} \right)' \Delta_{tt}^{-1} \left[ -\left( \frac{\partial \Delta_{tt}}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' + \left( \frac{\partial d'}{\partial \alpha_m} \right) \right] \\
+ B' \Delta^{-1} \left[ \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' \right] \\
- \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' + \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' - \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \\
+ \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' - \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \\
+ \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} d' - \left( \frac{\partial \Delta}{\partial \alpha_m} \right) \Delta_{tt}^{-1} \left( \frac{\partial \Delta}{\partial \alpha_m} \right) .
\]

or, to make explicit the dependence on the parameter vector
\[
\alpha = [\text{vec}(\beta)', \alpha_0] ,
\]

\[
\sum_{t=1}^{T} \left[ \ln \Lambda_{et}(\beta) + d_t' \Delta_{tt}^{-1} d_t' \right] ,
\]

This result and the derivative results developed above may lead to an
iterative procedure for the minimization of expression (6.52) with
respect to the unknown but identified elements of \( \alpha \). As in Section
6.1, this statement is contingent upon the reduced likelihood function
\( L_m(\alpha; \mathbf{Z}) \) satisfying certain regularity conditions; additional research
will be required to develop the details of such conditions. With this
contingency in mind, one may consider a specific Newton-Raphson procedure for maximum likelihood estimation of unknown but identified elements of \( \alpha \).

In notation similar to that of Chapter 5, let \( \alpha_1 = \text{vec}(\beta)' \), \( \alpha_{e1} \) be a \( 1 \times L \) vector, where \( L = L_\beta + L_\alpha \), \( L_\beta = \text{rk} \), and \( \alpha_{e1} \) is a \( 1 \times L \) vector of unknown, but identified, elements of \( \alpha_e \). Let \( f \), \( f_1 \) and \( f_2 \) be \( 1 \times L \) vectors such that \( f = f_1 + f_2 \); \( f_1 \) has \( i \)-th element equal to

\[
f_{1i} = \sum_{t=1}^{T} \text{tr}[A^{-1}_{et}[\frac{\partial}{\partial \alpha_{i}} (A_{et})]]
\]

for \( i = L_\beta + 1, L_\beta + 2, \ldots, L \) and zero otherwise; and \( f_2 \) has \( i \)-th element equal to

\[
f_{2i} = \sum_{t=1}^{T} \frac{\partial (d_t A_{t\gamma}^{-1} d'_t)}{\partial \alpha_i}
\]

for \( i = 1, 2, \ldots, L \). Similarly, define \( F \), \( F_1 \), and \( F_2 \) to be \( L \times L \) matrices such that

\[
F = F_1 + F_2
\]

\( F_1 \) has \((i,j)\)-th element equal to

\[
F_{1ij} = \sum_{t=1}^{T} \left\{ \text{tr}[\left(\frac{\partial^2 A_{et}}{\partial \alpha_i \partial \alpha_j} A_{et}^{-1}\right)] - \text{tr}[\left(\frac{\partial A_{et}}{\partial \alpha_i} A_{et}^{-1}\right) \left(\frac{\partial A_{et}}{\partial \alpha_j} A_{et}^{-1}\right)] \right\}
\]
for \( L_{\beta} + 1 < i, j < L \) and zero otherwise; and \( F_2 \) has \((i,j)\)-th element equal to

\[
F_{2ij} = \sum_{t=1}^{T} \frac{2(d_{t}A_{\beta}^{-1}d')}{\partial a_{i} \partial a_{j}}
\]

for \( 1 < i, j < L \). Define \( f_1(n) \) to be the matrix \( f_1 \) evaluated at the point \( a = a(n) \), and define \( f_2(n) \), \( f(n) \), \( F_1(n) \), \( F_2(n) \) and \( F(n) \) similarly. Then the iterative formula

\[
\hat{a}(n+1) = \hat{a}(n) + [F(n)]^{-1}f(n)
\]

provides an iterative Newton-Raphson algorithm to compute maximum likelihood estimates of \( a_{1} \). Additional research is required to develop modifications of expression (6.53) which will assure convergence of the \( \{\hat{a}_1(n)\} \) sequence to an \( \hat{a}_{1}^{*} \) which minimizes expression (6.52).

The work above addressed the functional model with no error in the equation. Now consider model (2.1)-(2.2) with an error in the equation, and assume that \( a_{t} \) and \( q_{t} \) follow mutually uncorrelated ARMA \((p, q)\) and ARMA \((p, q)\) processes,

\[
\sum_{j=0}^{p_a} \theta_{aj} a_{t-j} = \sum_{i=0}^{q_a} \theta_{ai} c_{t-i}
\]

and

\[
\sum_{j=0}^{p_q} \phi_{qj} q_{t-j} = \sum_{i=0}^{q_q} \phi_{qi} c_{t-i}
\]
respectively, where \{c_{1t}\} and \{c_{2t}\} are two mutually uncorrelated sequences of mutually uncorrelated \(p\)-dimensional \((0, \Sigma_{11})\) and \(q\)-dimensional \((0, \Sigma_{22})\) random vectors, respectively. Then \(b_t = (a_t, q_t)\) follows an ARMA\(_{p+r}(p_h, q_h)\) process

\[
\Sigma_{hj} b'_t = \Sigma \phi_{hj} c'_t - \Sigma \theta_{hi} c'_{t-i} ,
\]

where

\[
\phi_{hj} = \text{block diag}(\phi_{aj}, \phi_{qj}) , \quad j=0, 1, \ldots, p_h ;
\]

\[
\theta_{ai} = \text{block diag}(\theta_{ai}, \theta_{qi}) , \quad i=0, 1, \ldots, q_h ;
\]

\[
\Sigma_{cc} = \text{block diag}(\Sigma_{11}, \Sigma_{22}) ;
\]
\[ p_h = \max(p_a, p_q) ; \]

and

\[ q_h = \max(q_a, q_q) . \]

One may again pursue a modified state-space approach to the minimization of

\[ \ln \left| I_n + \text{vec}(Y - XB)' \Sigma^{-1} \text{vec}(Y - XB) \right| , \]

with respect to unknown parameters. Replace state-space models (6.32) and (6.43) with

\[ \varepsilon_t' = \tilde{B}_\varepsilon \tilde{\varepsilon}_t , \]

\[ \tilde{W}_{t+1} = A_h \tilde{W}_t + C_h c_t' ; \quad (6.54) \]

and

\[ v_t' = \tilde{B}_v \tilde{v}_t , \quad (6.55) \]

respectively, where

\[ \tilde{W}_t = \langle h_t | t', h_{t+1} | t', \ldots, h_{t+K-1} | t' \rangle ; \]

\[ \tilde{B}_\varepsilon = [ \mathbf{1}_p, (\mathbf{1}_r, 0_{r \times K})', 0_{p \times (K-1)(r+p)} ] ; \]
\[ A_h = \begin{bmatrix} 0 & I_{p+r} & 0 & 0 & \cdots & 0 \\ 0 & 0 & I_{p+r} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I_{p+r} \\ -\delta_{hK} & -\delta_{h,K-1} & -\delta_{h,K-2} & -\delta_{h,K-3} & \cdots & -\delta_{h1} \end{bmatrix} ; \]

\[ C_h = [\psi'_h, \psi_h^1, \ldots, \psi_h^{K-1}]' ; \]

\[ \psi'_h = I_{p+r} ; \]

\[ \psi_h^i = \theta_{hi} - \sum_{j=1}^{p_h} \psi_{hi} \psi_{h,i-j}, \quad i=1, 2, \ldots, K-1 ; \]

\[ \psi_{hi} = 0 \text{ for } i < 0 ; \]

\[ \theta_{hi} = 0 \text{ for } i > q_h ; \]

\[ \phi_{hj} = 0 \text{ for } j > p_h ; \]

\[ \tilde{W}_v = [(I_r, -\delta'_r, I_r), \theta_r \times (K-1)(p+r)] ; \]

\[ K = \max(p_h, q_h + 1) . \]
One may then find that

$$\mathcal{A}_{e t} = \mathcal{B}_{e t} \mathcal{E}_{t-1} \mathcal{B}'_t \mathcal{E}'_t,$$

and

$$\text{vec}(\mathbf{Y} - \mathbf{x}_t) (\mathbf{R}^{-1} \text{vec}(\mathbf{Y} - \mathbf{x}_t) = \sum_{t=1}^{T} \mathbf{d}_{t} \mathbf{\Lambda}_{vt}^{-1} \mathbf{d}'_t,$$

where

$$\mathbf{\tilde{P}}_{t+1|t} = \mathbf{A}_{h} \mathbf{\tilde{P}}_{t+1} + \mathbf{C}_{h} \mathbf{\tilde{C}}_{h};$$

$$\mathbf{\tilde{P}}_{t} = \mathbf{\tilde{P}}_{t|t-1} - \mathbf{\tilde{P}}_{t|t-1} \mathbf{B}_{e}^{-1} \mathbf{\tilde{P}}_{t|t-1};$$

$$\mathbf{\tilde{d}}_{t+1} = \mathbf{v}'_{t+1} - \mathbf{\tilde{B}}_{v} \mathbf{\tilde{W}}_{t+1|t};$$

$$\mathbf{\tilde{W}}_{t+1|t} = \mathbf{A}_{h} \mathbf{\tilde{W}}_{t|t};$$

$$\mathbf{\tilde{W}}_{t} = \mathbf{\tilde{W}}_{t|t-1} + \mathbf{\tilde{R}}_{t|t-1} \mathbf{\tilde{V}}_{v}^{-1} \mathbf{\tilde{d}}_{t};$$

$$\mathbf{\tilde{A}}_{v t} = \mathbf{\tilde{B}}_{v} \mathbf{\tilde{R}}_{t|t-1} \mathbf{\tilde{B}}'_{v};$$

$$\mathbf{\tilde{R}}_{t|t-1} = \mathbf{\tilde{A}}_{h} \mathbf{\tilde{R}}_{t-1} + \mathbf{C}_{h} \mathbf{\tilde{C}}_{h};$$

\[ \text{vec}(\mathbf{Y} - \mathbf{x}_t) (\mathbf{R}^{-1} \text{vec}(\mathbf{Y} - \mathbf{x}_t) = \sum_{t=1}^{T} \mathbf{d}_{t} \mathbf{\Lambda}_{vt}^{-1} \mathbf{d}'_t, \]
Derivatives of the likelihood function and the associated Newton-Raphson maximum likelihood procedure then follow in a manner analogous to the corresponding results for the model with no error in the equation.

The methods of this section extend to other models for which the components of $\tilde{\xi}_t$ permit a time-invariant state-space representation or a slight variation thereof [e.g., series with nonconsecutive observations, as discussed in Wincek and Reinsel (1986)]. This category includes many models used in practice; however, the more general approach of Section 6.1 remains valuable for two reasons. First, the linearization arguments of Section 6.1 indicate the similarity between the iterative estimator (6.27) and the weighted estimator of Section 4.2 in the case $\tilde{\pi}(T) = \tilde{\Gamma}_{vV}^{-1}$. Second, and more generally, there may be cases in which a sampling design is sufficiently complex that the resulting composite error $\tilde{\xi}_t$ does not admit readily a state-space representation (6.32). In this case, if $\tilde{\Gamma}_{vV}$ is nonetheless a known function of an identified low-dimensional parameter $\tilde{\alpha}_t$, the more general Newton-Raphson approach of Section 6.1 may be preferable.


I wish to express my sincere thanks to Professor Wayne A. Fuller for his advice and encouragement during the production of this dissertation. His expert advice and insight have been a constant source of inspiration throughout my years at Iowa State University.

I also thank the other members of my graduate committee: Professors David A. Harville, Glen D. Meeden, Richard J. Tondra, and Stephen B. Vardeman. In addition, I thank Professors Yasuo Amemiya and Herbert T. David for many helpful comments and I thank Professor Bryan Cain for a helpful discussion of matrix polynomials.

I am grateful to the Survey Section of the Statistical Laboratory for the use of its word processor. I thank Ms. Christine Olson for her extremely efficient and cheerful typing of this dissertation. I also thank Ms. Helen Nelson for her assistance.

With deep gratitude I acknowledge the love and support of my wife and parents.

This research was supported in part by a National Science Foundation Graduate Fellowship and by the United States Bureau of the Census Joint Statistical Agreement 87-1.
9. APPENDIX A.

SOME USEFUL MATRIX RESULTS

9.1. Matrix Differentiation

The maximum likelihood work in Chapters 5 and 6 required several results in matrix differentiation. These results are not original to this dissertation, but are presented here for convenient reference. The first such result is a multivariate chain rule given by Williamson and Trotter (1979, Theorem 9.2.1, p. 250).

**Result 9.1.** Let \( f: \mathbb{R}^n \rightarrow \mathbb{R}^m \) be continuously differentiable at \( x \in \mathbb{R}^n \) and let \( g: \mathbb{R}^m \rightarrow \mathbb{R}^p \) be continuously differentiable at \( f(x) \in \mathbb{R}^m \). If the composition \( g \circ f \) is defined on an open set containing \( x \), then \( g \circ f \) is continuously differentiable at \( x \), and

\[
(g \circ f)'(x) = g'[f(x)]f'(x),
\]

where \( (g \circ f)'(x) \), \( g'[f(x)] \) and \( f'(x) \) have \((i,j)\)-th elements

\[
\frac{\partial (g \circ f)(x)}{\partial x_j}, \quad \frac{\partial g_i[f(x)]}{\partial y_j}, \quad \text{and} \quad \frac{\partial f_i(x)}{\partial x_j},
\]

respectively, and \( (g \circ f)(x), \ g[f(x)] \), and \( y = f(x) \) have \( i \)-th elements \( (g \circ f)_i(x), \ g_i[f(x)] \), and \( y_i = f_i(x) \), respectively.

**Proof.** See Williamson and Trotter (1979, p. 250). □

The remaining results of this section apply Lemma 9.1 to parts of the multivariate normal log-likelihood function. Let

\[
g = (\alpha_1, \alpha_2, ..., \alpha_p)'
\]

be a vector of variables, let \( a = a(g) \) be a
scalar function of \( \mathbf{a} \), and let \( \mathbf{B} = \mathbf{B}(\mathbf{a}) \) be an \( r \times k \) matrix function of \( \mathbf{a} \) with \((i,j)\)-th element equal to \( B_{ij}(\mathbf{a}) \). Following Graybill (1983, Definitions 10.8.2 and 10.8.3), define

\[
\frac{\partial \mathbf{a}(\mathbf{a})}{\partial \mathbf{a}} = \left[ \frac{\partial \mathbf{a}(\mathbf{a})}{\partial a_1}, \frac{\partial \mathbf{a}(\mathbf{a})}{\partial a_2}, \ldots, \frac{\partial \mathbf{a}(\mathbf{a})}{\partial a_p} \right]^T ;
\]  

(9.1)

and

\[
\frac{\partial \mathbf{B}(\mathbf{a})}{\partial a_h} = \left[ \frac{\partial B_{ij}(\mathbf{a})}{\partial a_h} \right], \quad h=1, 2, \ldots, p ,
\]  

(9.2)

i.e., an \( r \times k \) matrix with \((i,j)\)-th element equal to \( \frac{\partial B_{ij}(\mathbf{a})}{\partial a_h} \).

Assume throughout this appendix that all matrices of interest are functions of \( \mathbf{a} \) that are defined on some common open set \( A \) contained in \( \mathbb{R}^p \) and are continuously differentiable at some common point \( \mathbf{x}_0 \in A \). The results below are with reference to differentials evaluated at this point \( \mathbf{x}_0 \).

Result 9.2 addresses differentiation of quadratic forms.

Result 9.2. Let \( \mathbf{a}(\mathbf{a}) = [a_1(\mathbf{a}), a_2(\mathbf{a}), \ldots, a_k(\mathbf{a})]^T \) be \( k \times 1 \) and let \( \mathbf{B}(\mathbf{a}) = [B_{ij}(\mathbf{a})] \) be a \( k \times k \) matrix. Then

\[
\frac{\partial}{\partial a_h} [\mathbf{a}(\mathbf{a})^T \mathbf{B}(\mathbf{a}) \mathbf{a}(\mathbf{a})]
\]

\[
= 2\left[ \frac{\partial \mathbf{a}(\mathbf{a})}{\partial a_h} \right]^T \mathbf{B}(\mathbf{a}) \mathbf{a}(\mathbf{a}) + \mathbf{a}(\mathbf{a})^T \left[ \frac{\partial \mathbf{B}(\mathbf{a})}{\partial a_h} \right] \mathbf{a}(\mathbf{a}) .
\]
Proof. Assume without loss of generality that \( B(\alpha) \) is symmetric. By Result 9.1,

\[
\frac{\partial}{\partial \alpha_h} [a(\alpha)'B(\alpha)c(\alpha)] = \sum_{i=1}^{k} \frac{\partial a_i(\alpha)}{\partial \alpha_h} \frac{\partial \alpha_i(\alpha)}{\partial \alpha_h} + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\partial a_i(\alpha)}{\partial \alpha_h} \frac{\partial \alpha_j(\alpha)}{\partial \alpha_h} - \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\partial a_i(\alpha)}{\partial \alpha_h} \frac{\partial \alpha_j(\alpha)}{\partial \alpha_h} + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\partial \alpha_i(\alpha)}{\partial \alpha_h} \frac{\partial \alpha_j(\alpha)}{\partial \alpha_h}
\]

where \( B_i \) is the i-th row of \( B \) and the second equality follows from Theorems 10.8.2 and 10.8.4 of Graybill (1983).

Result 9.3 is a slight extension of Result 9.2 that is useful in computing second derivatives; Result 9.4 is a further generalization to higher dimensions.

Result 9.3. Let \( a(\alpha) \) and \( c(\alpha) \) be \( k \times 1 \) vectors and let \( B(\alpha) \) be a \( k \times k \) matrix. Then

\[
\frac{\partial}{\partial \alpha_h} [a(\alpha)'B(\alpha)c(\alpha)] = \frac{\partial a(\alpha)}{\partial \alpha_h}'B(\alpha)c(\alpha) + a(\alpha)'B(\alpha) \frac{\partial c(\alpha)}{\partial \alpha_h} + a(\alpha)'B(\alpha) \frac{\partial c(\alpha)}{\partial \alpha_h}
\]

Proof. Note that
\[ a' Bc = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i B_{ij} c_j \]

so

\[ \frac{\partial (a' Bc)}{\partial B_{ij}} = a_i c_j \]

Result 9.1 then implies that

\[ \frac{\partial}{\partial \alpha_h} [a(\alpha)' B(\alpha) c(\alpha)] = \sum_{i=1}^{k} \left[ \frac{\partial a_i(\alpha)}{\partial \alpha_h} \right] \frac{\partial B_{ii}(\alpha)}{\partial \alpha_h} + \sum_{i=1}^{k} \left[ \frac{\partial a_i(\alpha)}{\partial \alpha_h} \right] \frac{\partial c_i(\alpha)}{\partial \alpha_h} + \sum_{i=1}^{k} \sum_{j=1}^{k} \left[ \frac{\partial a_i(\alpha)}{\partial \alpha_h} \right] \frac{\partial B_{ij}(\alpha)}{\partial \alpha_h} + a(\alpha)' B(\alpha) \frac{\partial c(\alpha)}{\partial \alpha_h} + a(\alpha)' B(\alpha) \frac{\partial c(\alpha)}{\partial \alpha_h} + a(\alpha)' B(\alpha) \frac{\partial c(\alpha)}{\partial \alpha_h} \]

A similar proof establishes Result 9.3 for symmetric \( B \).

**Result 9.4.** Let \( A(\alpha) \) be a \( k \times r \) matrix and let \( B(\alpha) \) be a \( k \times k \) matrix. Then

\[ \frac{\partial}{\partial \alpha_h} [A(\alpha)' B(\alpha) A(\alpha)] = \left[ \frac{\partial A(\alpha)}{\partial \alpha_h} \right]' B(\alpha) A(\alpha) + A(\alpha)' B(\alpha) \left[ \frac{\partial A(\alpha)}{\partial \alpha_h} \right] + A(\alpha)' B(\alpha) \left[ \frac{\partial A(\alpha)}{\partial \alpha_h} \right] A(\alpha) \quad (9.3) \]
Proof. This result follows immediately from Results 9.2 and 9.3, and the observation that expression (9.3) has \((i,j)\)-th entry equal to

\[
\frac{\partial A_{i}(\alpha)}{\partial \alpha_{h}} \mathbf{B}(\alpha) A_{j}(\alpha) + A_{i}(\alpha)' \mathbf{B}(\alpha) \frac{\partial A_{j}(\alpha)}{\partial \alpha_{h}} + A_{i}(\alpha)' \frac{\partial \mathbf{B}(\alpha)}{\partial \alpha_{h}} A_{j}(\alpha)
\]

where \(A_{i}(\alpha)\) is the \(i\)-th column of \(A(\alpha)\).

Result 9.5 applies the chain rule to matrix products and their traces.

Result 9.5. Let \(A(\alpha)\) be an \(r \times k\) matrix and let \(B(\alpha)\) be a \(k \times q\) matrix.

(i) Then

\[
\frac{\partial}{\partial \alpha_{h}} \left[ A(\alpha) B(\alpha) \right] = \left[ \frac{\partial A(\alpha)}{\partial \alpha_{h}} \right] B(\alpha) + A(\alpha) \left[ \frac{\partial B(\alpha)}{\partial \alpha_{h}} \right].
\]

(ii) If \(r = q\), then

\[
\frac{\partial}{\partial \alpha_{h}} \text{tr}[A(\alpha) B(\alpha)] = \text{tr} \left\{ \left[ \frac{\partial A(\alpha)}{\partial \alpha_{h}} \right] B(\alpha) + A(\alpha) \left[ \frac{\partial B(\alpha)}{\partial \alpha_{h}} \right] \right\}.
\]

Proof. Part (i) follows immediately from Theorem 10.8.1 of Graybill (1983) and the observation that the \((i,j)\)-th element of \(A(\alpha) B(\alpha)\) equals \(A_{i}(\alpha) B_{j}(\alpha)\), where \(A_{i}(\alpha)\) is the \(i\)-th row of \(A(\alpha)\) and \(B_{j}(\alpha)\) is the \(j\)-th column of \(B(\alpha)\). To establish part (ii), note that
\[ \frac{\partial \text{tr}[A(\omega)B(\omega)]}{\partial \alpha_h} = \sum_{i=1}^{r} \sum_{j=1}^{k} A_{ij}(\omega)B_{ij}(\omega) \]

\[ = \sum_{i=1}^{r} \sum_{j=1}^{k} \left( \frac{\partial A_{ij}(\omega)}{\alpha_h}B_{ij}(\omega) + A_{ij}(\omega) \frac{\partial B_{ij}(\omega)}{\alpha_h} \right) \]

\[ = \text{tr} \left\{ \left[ \frac{\partial A(\omega)}{\alpha_h} \right] B(\omega) + A(\omega) \left[ \frac{\partial B(\omega)}{\alpha_h} \right] \right\}. \]

Results 9.6 and 9.7 apply the chain rule to matrix inverses and associated quadratic and bilinear forms.

**Result 9.6.** [Result 4.A.9 of Fuller (1987).] Let \( B(\omega) \) be a \( k \times k \) nonsingular matrix. Then

\[ \frac{\partial}{\partial \alpha_h} [B(\omega)^{-1}] = -B(\omega)^{-1} \frac{\partial B(\omega)}{\alpha_h} B(\omega)^{-1}. \]

**Proof.** See Fuller (1987, p. 390).

**Result 9.7.** Let \( A(\omega) \) be a \( k \times r \) matrix and let \( B(\omega) \) be a \( k \times k \) nonsingular matrix. Then

\[ \frac{\partial}{\partial \alpha_h} [A(\omega)'B(\omega)^{-1}A(\omega)] = \left[ -\frac{\partial A(\omega)}{\alpha_h} \right]'B(\omega)^{-1}A(\omega) + A(\omega)'B(\omega)^{-1} \frac{\partial A(\omega)}{\alpha_h} \]

\[ - A(\omega)'B(\omega)^{-1} \frac{\partial B(\omega)}{\alpha_h} B(\omega)^{-1}A(\omega). \]

**Proof.** This follows immediately from Results 9.4 and 9.6.
Result 9.8 applies the chain rule to the logarithm of the determinant of a nonsingular matrix.

**Result 9.8.** Let \( B(\alpha) \) be a nonsingular \( k \times k \) matrix that is twice continuously differentiable with respect to \( \alpha \). Then

\[
\frac{\partial \ln |B(\alpha)|}{\partial \alpha_h} = \text{tr} \left\{ B(\alpha)^{-1} \left[ \frac{\partial B(\alpha)}{\partial \alpha_h} \right] \right\}
\]

and

\[
\frac{\partial^2 \ln |B(\alpha)|}{\partial \alpha_h \partial \alpha_m} = \text{tr} \left\{ \left[ \frac{\partial B(\alpha)}{\partial \alpha_h \partial \alpha_m} \right] B^{-1}(\alpha) \right\} - \text{tr} \left\{ \left[ \frac{\partial B(\alpha)}{\partial \alpha_h} \right] B(\alpha)^{-1} \left[ \frac{\partial B(\alpha)}{\partial \alpha_m} \right] B(\alpha)^{-1} \right\}.
\]

**Proof.** Part (i) is stated and proved as Result 4.4.8 in Fuller (1987). To establish part (ii), note that by Result 9.5 (ii),

\[
\frac{\partial}{\partial \alpha_h} \text{tr} \left\{ B(\alpha)^{-1} \left[ \frac{\partial^2 B(\alpha)}{\partial \alpha_h \partial \alpha_m} \right] \right\}
\]

\[
= \text{tr} \left\{ \frac{\partial}{\partial \alpha_h} \left[ B(\alpha)^{-1} \left[ \frac{\partial^2 B(\alpha)}{\partial \alpha_h \partial \alpha_m} \right] \right] \right\}
\]

\[
= \text{tr} \left\{ B(\alpha)^{-1} \left[ \frac{\partial^2 B(\alpha)}{\partial \alpha_h \partial \alpha_m} \right] \right\} - \text{tr} \left\{ \left[ \frac{\partial B(\alpha)}{\partial \alpha_h} \right] B(\alpha)^{-1} \left[ \frac{\partial B(\alpha)}{\partial \alpha_m} \right] B(\alpha)^{-1} \right\},
\]

so the result follows from part (i).

Finally, Result 9.9 applies the chain rule to two simple Kronecker products.
**Result 9.9.** Let $A$ be a $T_p \times T_p$ fixed matrix and let $B$ be a $k \times T_p$ fixed matrix. Then

(i) \[ \frac{\partial}{\partial \alpha} \left[ (\alpha \cdot 1_T)'A(\alpha \cdot 1_T) \right] = 2(I_p \cdot 1_T)'A(\alpha \cdot 1_T) \]

and

(ii) \[ \frac{\partial}{\partial \alpha} \left[ B(\alpha \cdot 1_T) \right] = B(I_p \cdot 1_T) . \]

**Proof.** Let $A(i,j)$ be the $(i,j)$-th $T \times T$ block of $A$ and let $\alpha$ be a $p \times p$ matrix with $(i,j)$-th element equal to $1_T A(i,j) 1_T$. Then

\[ (\alpha \cdot 1_T)'A(\alpha \cdot 1_T) = \alpha' \alpha , \]

so by Theorem 10.8.2 of Graybill (1983),

\[ \frac{\partial}{\partial \alpha} \left[ (\alpha \cdot 1_T)'A(\alpha \cdot 1_T) \right] = 2\alpha' \alpha = 2(I_p \cdot 1_T)'A(\alpha \cdot 1_T) . \]

Similarly, let $B(j)$ be the $j$-th $k \times T$ block of $B$ and let $b$ be a $k \times p$ matrix with $j$-th column equal to $B(j) 1_T$. Then

\[ \frac{\partial}{\partial \alpha} \left[ B(\alpha \cdot 1_T) \right] = \frac{\partial}{\partial \alpha} \left[ b \alpha \right] = b \]
9.2. Positive Definite Matrix Differences

 Chapters 4 and 5 present some errors-in-variables parameter estimators that have been modified to ensure that certain associated matrix differences are positive definite with probability one. The following two results provide the algebraic background for such modifications.

**Result 9.10.** Let $M$ be a real nonnull symmetric $n \times n$ matrix with eigenvalues $d_1 > d_2 > \ldots > d_n$.

(i) If $d_n > 0$, then $I_n - cM$ is positive definite if and only if $c < d_n^{-1}$.

(ii) If $d_1 < 0$, then $I_n - cM$ is positive definite if and only if $c > d_1^{-1}$.

(iii) If $d_1 > 0 > d_n$, then $I_n - cM$ is positive definite if and only if $d_n^{-1} < c < d_1^{-1}$.

**Proof.** Since $M$ is a real symmetric matrix, we may write its spectral decomposition as $MQ' = D = \text{diag}(d_1, d_2, \ldots, d_n)$, where $Q$ is an orthogonal $n \times n$ matrix and $d_i$ is real for all $i = 1, 2, \ldots, n$. Since $x'(I - cM)x = x'Q(I - cD)Q'x = y'(I - cD)y$ for $y = Qx$, it
follows that $I - cM$ is positive definite if and only if
$I - cD = \text{diag}(1 - cd_1, 1 - cd_2, \ldots, 1 - cd_n)$ is positive definite.

This last condition is true if and only if $1 - cd_i > 0$ for all $i = 1, 2, \ldots, n$. For positive $d_i$, $1 - cd_i > 0$ if and only if $c < d_i^{-1}$, so we must have $c < \left[\min_{d_i > 0} (d_i^{-1})\right]$, where the minimum over an empty set is taken to be $+\infty$. For negative $d_i$, $1 - cd_i > 0$ if and only if $c > d_i^{-1}$, so we must have $c > \left[\max_{d_i < 0} (d_i^{-1})\right]$, where the maximum over an empty set is taken to be $-\infty$. Thus $I_n - cM$ is positive definite if and only if

$$\left[\max_{d_i < 0} (d_i^{-1})\right] < c < \left[\min_{d_i > 0} (d_i^{-1})\right],$$

and results (i)-(iii) follow.

**Result 9.11.** Let $A$ be an $n \times n$ symmetric positive definite matrix and let $B$ be an $n \times n$ symmetric matrix. Let $d_1 > d_2 > \ldots > d_n$ be the roots of the matrix $B$ in the metric $A$. Then $A - cB$ is positive definite if and only if

$$\left[\max_{d_i < 0} (d_i^{-1})\right] < c < \left[\min_{d_i > 0} (d_i^{-1})\right]$$

where the maximum over an empty set is $-\infty$ and the minimum over an empty set is $+\infty$.

**Proof.** Note that $|B - dA| = 0$ if and only if

$$|A^{-1/2} BA^{-1/2} - dI| = 0.$$ Thus, $d_i$, $i = 1, 2, \ldots, n$ are the
eigenvalues of $A^{-1/2}BA^{-1/2}$. Also, the nonsingularity of $A$ implies that $A - cB$ is positive definite if and only if $I - cA^{-1/2}BA^{-1/2}$ is positive definite. The result then follows from Result 9.10.

9.3. Alternative Representation of Projection Matrices

Chapter 6 requires the maximization of the normal functional model density function $f(Z; \beta, \sigma^2, x)$ with respect to the fixed parameter matrix $x$. The following result allows one to express

$$\inf_{\text{vec}(x) \in \mathbb{R}^k} \{-2 \ln f(Z; \beta, \sigma^2, x) - Tp \ln(2\pi)\}$$

in useful form.

Result 9.12. Let $\beta$ be a $k \times r$ matrix, let $p = r + k$, and let $\Gamma_{ee}$ be a $Tp \times Tp$ symmetric positive definite matrix. Let $A^-$ be any generalized inverse of a matrix $A$. Then

$$\begin{align*}
I_T &= \left[\left(\beta, I_k\right)' \cdot I_T\right]\left[\left(\beta, I_k\right)' \cdot I_T\right]^{-1}\left[\left(\beta, I_k\right)' \cdot I_T\right]\left[\left(\beta, I_k\right)' \cdot I_T\right]^{-1} \\
&= \Gamma_{ee}\left[\left(\beta, I_k\right)' \cdot I_T\right]\left[\left(\beta, I_k\right)' \cdot I_T\right]^{-1} \\
&\times \left[\left(\beta, I_k\right)' \cdot I_T\right]^{-1}
\end{align*}$$

Proof. The claim is equivalent to the statement,
\[
I_T - \Gamma^{-1/2}_{e\in E} [(\beta, I_k)' \circ I_T] \{(\beta, I_k) \circ I_T\}^{-1} [(\beta, I_k)' \circ I_T]^{-1} \\
\times [(\beta, I_k) \circ I_T \Gamma^{-1/2}_{e\in E} \\
= \Gamma^{1/2}_{e\in E} [(I_T', -\beta')' \circ I_T] \{(I_T', -\beta')' \circ I_T\}^{-1} [(I_T', -\beta')' \circ I_T]^{-1} \\
\times [(I_T', -\beta')' \circ I_T \Gamma^{1/2}_{e\in E}.
\]

By inspection, the left-hand side of the last equation equals the matrix of the orthogonal projection onto \(C^{1} \{(I_T', -\beta')' \circ I_T\}\), while the right-hand side equals the matrix of the orthogonal projection onto \(C^{1} \{(I_T', -\beta')' \circ I_T\}\). Note that

\[
\{\Gamma^{1/2}_{e\in E} [(I_T', -\beta')' \circ I_T]\}' \{\Gamma^{1/2}_{e\in E} [(I_T', -\beta')' \circ I_T]\}
\]

\[
= [(\beta, I_k) (I_T', -\beta')]' \circ I_T = 0_{T_p \times T_p}.
\]

Thus, the elements of \(C^{1} \{(I_T', -\beta')' \circ I_T\}\) are orthogonal to the elements of \(C^{1} \{(I_T', -\beta')' \circ I_T\}\), so \(C^{1} \{(I_T', -\beta')' \circ I_T\}\) is isomorphic to \(C^{1} \{(I_T', -\beta')' \circ I_T\}\) and the result follows. \(\square\)
10. APPENDIX B.

SEQUENCES OF ARRAYS OF REAL NUMBERS

The arguments in Section 3.4 use several properties of sequences of $T \times T$ arrays of real numbers. These properties are summarized here for convenient reference.

Throughout this appendix, let \( \{a_{Tst}; 1 \leq s, t \leq T\} \), \( \{b_{Tst}; 1 \leq s, t \leq T\} \), and \( \{c_{Tst}; 1 \leq s, t \leq T\} \) be three sequences of $T \times T$ arrays of real numbers, $T=1, 2, \ldots$. Also, let \( \{a_{st}; s, t \in \mathbb{Z}^+\} \), \( \{b_{st}; s, t \in \mathbb{Z}^+\} \), and \( \{c_{st}; s, t \in \mathbb{Z}^+\} \) be three doubly semi-infinite arrays of real numbers, where $\mathbb{Z}^+$ is the set of all positive integers.

Define \( \{a_{st}\} \) to be row-absolutely summable if

\[
\sum_{s=1}^{\infty} |a_{st}| \text{ is finite and uniformly bounded in } t \in \mathbb{Z}^+; \tag{10.1.a}
\]

and define \( \{a_{st}\} \) to be column-absolutely summable if

\[
\sum_{t=1}^{\infty} |a_{st}| \text{ is finite and uniformly bounded in } s \in \mathbb{Z}^+. \tag{10.1.b}
\]

For such sequences, the following result is a slight variant on the fact that the convolution of two absolutely summable sequences is absolutely summable.

**Result 10.1.** Assume that \( \{a_{qs}; q, s \in \mathbb{Z}^+\} \) and \( \{b_{st}; s, t \in \mathbb{Z}^+\} \) are doubly semi-infinite arrays that satisfy conditions (10.1). Then for
all \( q, t \in \mathbb{Z}^+ \),

\[
c_{qt} \equiv \sum_{s=1}^{\infty} a_{qs} b_{st} = \lim_{T \to \infty} \sum_{s=1}^{T} a_{qs} b_{st}
\]

exists and is finite; and the array \( \{c_{qt}; q, t \in \mathbb{Z}^+\} \) satisfies conditions (10.1).

Proof. By conditions (10.1), there exists some finite real number \( A \) such that

\[
\sum_{q=1}^{\infty} |a_{qs}| < A \text{ for all } s \in \mathbb{Z}^+ ;
\]

\[
\sum_{s=1}^{\infty} |a_{qs}| < A \text{ for all } q \in \mathbb{Z}^+ ;
\]

\[
\sum_{s=1}^{\infty} |b_{st}| < A \text{ for all } t \in \mathbb{Z}^+ ;
\]

and

\[
\sum_{t=1}^{\infty} |b_{st}| < A \text{ for all } s \in \mathbb{Z}^+ .
\]

Then for all \( q, t \in \mathbb{Z}^+ \),

\[
\sum_{s=1}^{T} |a_{qs} b_{st}| < A \sum_{s=1}^{T} |b_{st}| < A^2 .
\]
Hence,

\[ \sum_{s=1}^{\infty} |a_{qs} b_{st}| = \lim_{T \to \infty} \sum_{s=1}^{T} |a_{qs} b_{st}| < A^2, \]

so for all \( q, t \in \mathbb{Z}^+ \), \( \sum_{s=1}^{\infty} a_{qs} b_{st} \) exists and is finite. Next, condition (10.1) implies that for all \( T, S \in \mathbb{Z}^+ \),

\[ \sum_{q=1}^{T} |a_{qs}| < A \quad \text{for all } s \in \mathbb{Z}^+; \]

and

\[ \sum_{s=1}^{S} |b_{st}| < A \quad \text{for all } t \in \mathbb{Z}^+. \]

Thus, for all \( T, S \in \mathbb{Z}^+ \)

\[ \sum_{q=1}^{T} \sum_{s=1}^{S} |a_{qs} b_{st}| = \sum_{s=1}^{S} \left( \sum_{q=1}^{T} |a_{qs}| b_{st} \right) \]

\[ < \sum_{s=1}^{S} A b_{st} \]

\[ < A^2, \]

and the row-absolute summability of \( \{c_{qm}\} \) follows. The column-absolute summability of \( \{c_{qt}\} \) is established similarly. \( \square \)
For the arrays \( \{a_{Tst}\} \) and \( \{b_{Tst}\} \), results analogous to Result 10.1 generally require conditions similar to row-absolute summability and column-absolute summability, e.g., conditions that

\[
\sum_{s=1}^{T} |a_{Tst}| \text{ is uniformly bounded in } t, T \in \mathbb{Z}^+; \quad (10.2.a)
\]

and

\[
\sum_{t=1}^{T} |a_{Tst}| \text{ is uniformly bounded in } s, T \in \mathbb{Z}^+. \quad (10.2.b)
\]

Result 10.2. Assume that \( \{a_{Tst}\} \) and \( \{b_{Tst}\} \) satisfy conditions (10.2), and assume that for all \( s, t \in \mathbb{Z}^+ \), \( a_{st} \equiv \lim_{T \to \infty} a_{Tst} \) and \( b_{st} \equiv \lim_{T \to \infty} b_{Tst} \) exist and are finite.

a. Then \( \{a_{st}\} \) and \( \{b_{st}\} \) satisfy conditions (10.1).

b. Let \( c_{Tqt} \equiv \sum_{s=1}^{T} a_{Tqs} b_{Tst} \), \( 1 < q, t < T, T \in \mathbb{Z}^+ \). Then the arrays \( \{c_{Tqt}, 1 < q, t < T\} \), \( T \in \mathbb{Z} \) satisfy conditions (10.2).

c. Assume that for any positive integer \( q \) and any \( \epsilon > 0 \), there exists some positive integer \( T_{q\epsilon} \) such that

\[
\sum_{s=1}^{T} |a_{Tqs}| < \epsilon \text{ for all } T > T_{q\epsilon} + q.
\]

Then for any fixed positive integers \( q \) and \( t \), \( c_{qt} = \lim_{T \to \infty} c_{Tqt} \) exists, is finite, and is equal to \( \sum_{s=1}^{\infty} a_{qs} b_{st} \).
d. Under the hypothesis of part (c), the array \( \{c_{qt}; q, t \in \mathbb{Z}^+\} \) satisfies conditions (10.1).

**Proof.**

a. Under conditions (10.2), there exists some \( A > 0 \) such that
\[
\sum_{s=1}^{T} |a_{Tst}| < A \quad \text{for all } T \text{ and all } 1 < t < T; \quad \text{and}
\]
\[
\sum_{t=1}^{T} |a_{Tst}| < A \quad \text{for all } T \text{ and all } 1 < s < T. \quad \text{Pick } \epsilon > 0 \text{ and } N \in \mathbb{Z}^+. \]
The existence of limits \( a_{st} = \lim_{T \to \infty} a_{Tst} \) implies that for each pair of positive integers \((s, t)\), there exists some positive integer \( T_{st} \in \mathbb{Z}^+ \) such that \( |a_{st} - a_{Tst}| < \epsilon N^{-1} \) for all \( T > T_{st} \).

Define \( T_N = \max \{T_{st}: 1 \leq s, t \leq N\} \). Then for all \( T > T_N \) and \( 1 < s, t < N \), \( |a_{st} - a_{Tst}| < \epsilon N^{-1} \) and thus
\[
\sum_{s=1}^{N} |a_{st}| < \sum_{s=1}^{N} |a_{st} - a_{Tst}| + \sum_{s=1}^{N} |a_{Tst}| < \epsilon + A. \]

Since \( \epsilon > 0 \) and \( N \in \mathbb{Z}^+ \) were arbitrary, it follows that \( \sum_{s=1}^{N} |a_{st}| < A \) for all \( N \) and \( s \), so \( \sum_{s=1}^{\infty} |a_{st}| < A \) for all \( t \in \mathbb{Z}^+ \). Conditions (10.1.b) for \( \{a_{st}\} \) and (10.1) for \( \{b_{st}\} \) follow similarly.

b. Let \( A \) be a finite uniform bound on \( \sum_{s=1}^{T} |a_{Tst}| \), \( \sum_{t=1}^{T} |a_{Tst}| \), \( \sum_{s=1}^{T} |b_{Tst}| \) and \( \sum_{t=1}^{T} |b_{Tst}| \). Then for all \( 1 < t < T, T \in \mathbb{Z}^+ \),
\[
\sum_{q=1}^{T} |c_{Tqt}| = \sum_{q=1}^{T} \sum_{s=1}^{T} a_{Tqs} b_{Tst}.\]
c. Pick $\varepsilon > 0$. It follows from Results 10.1 and 10.2.a that there exists some positive integer $T_{1\varepsilon}$ such that

$$
\sum_{s=T_{1\varepsilon}+1}^{\infty} |a_{qs}| < (A + 1)^{-1}\varepsilon,
$$

$$
\sum_{s=T_{1\varepsilon}+1}^{\infty} |b_{st}| < (A + 1)^{-1}\varepsilon, \text{ and}
$$

$$
\sum_{s=T_{1\varepsilon}+1}^{\infty} |a_{qs}b_{st}| < \varepsilon
$$

for all $q, t \in \mathbb{Z}^+$. Moreover, under the hypotheses of part (c), there exists some $T_{q\varepsilon}$ such that

$$
\sum_{s=T_{q\varepsilon}+q}^{T} |a_{Tqs}| < A^{-1}\varepsilon
$$

for all $T > T_{q\varepsilon} + q$. Let $T_{2\varepsilon} = \max\{T_{1\varepsilon}, T_{q\varepsilon} + q\}$. The existence of limits $a_{qs} = \lim_{T\to\infty} a_{Tqs}$ and $b_{st} = \lim_{T\to\infty} b_{Tst}$ implies that there exists some $T_{3\varepsilon}$ such that if $1 < q, s, t < T_{2\varepsilon}$
and $T > T_{3\epsilon}$, then $|a_{qs} - a_{Tqs}| < (A + 1)^{-1}\epsilon$ and $|b_{st} - b_{Tst}| < (A + 1)^{-1}\epsilon$. Let $T_{4\epsilon} = \max(T_{2\epsilon}, T_{3\epsilon})$. Then for $T > T_{4\epsilon}$,

$$
\left| \sum_{s=1}^{\infty} (a_{qs} b_{st} - c_{Tqs}) \right| < \left| \sum_{s=1}^{T} (a_{qs} b_{st} - a_{Tqs} b_{Tst}) \right| + \sum_{s=T+1}^{\infty} |a_{qs} b_{st}|
$$

$$
< \left| \sum_{s=1}^{T} (a_{qs} b_{st} - a_{Tqs} b_{Tst}) \right| + \epsilon
$$

$$
= \left| \sum_{s=1}^{T} [a_{qs} b_{st} - (a_{Tqs} - a_{qs} + a_{qs}) (b_{Tst} - b_{st} + b_{st})] \right| + \epsilon
$$

$$
= \left| \sum_{s=1}^{T} [a_{qs} b_{st} - (a_{Tqs} - a_{qs}) b_{st} - a_{qs} (b_{Tst} - b_{st}) - (a_{Tqs} - a_{qs}) (b_{Tst} - b_{st}) - a_{qs} b_{st}] \right| + \epsilon
$$

$$
< \sum_{s=1}^{T_{2\epsilon}} |a_{Tqs} - a_{qs}| \cdot |b_{st}| + \sum_{s=1}^{T_{2\epsilon}} |a_{qs}| \cdot |b_{Tst} - b_{st}|
$$

$$
+ \sum_{s=1}^{T_{2\epsilon}} |a_{Tqs} - a_{qs}| \cdot |b_{Tst} - b_{st}| + \sum_{s=T_{2\epsilon}+1}^{T} |a_{Tqs} - a_{qs}| \cdot |b_{Tst} - b_{st}|
$$

$$
+ \epsilon
$$
\[ \begin{align*}
&< \sum_{s=1}^{T_{2\epsilon}} [(A + 1)^{-1} \varepsilon] |b_{st}| + \sum_{s=1}^{T_{2\epsilon}} |a_{qs}| [((A + 1)^{-1} \varepsilon] \\
&+ \sum_{s=1}^{T_{2\epsilon}} [(A + 1)^{-1} \varepsilon] |b_{Tst} - b_{st}| + \sum_{s=T_{2\epsilon}+1}^{T} |b_{st}| 2a \\
&+ \sum_{s=T_{2\epsilon}+1}^{T} |a_{qs}| 2A + \sum_{s=T_{2\epsilon}+1}^{T} |a_{Tqs} - a_{qs}| 2A + \varepsilon \\
&< (A + 1)^{-1} \varepsilon A + A(A + 1)^{-1} \varepsilon + (A + 1)^{-1} \varepsilon 2A + 2A(A + 1)^{-1} \varepsilon \\
&+ (A + 1)^{-1} \varepsilon 2A + [(A + 1)^{-1} \varepsilon + (A + 1)^{-1} \varepsilon] 2A + \varepsilon \\
&< 13 \varepsilon .
\end{align*} \]

Since \( \varepsilon > 0 \) was arbitrary, the result follows.

d. The final conclusion follows immediately from Results 10.1, 10.2.a, and 10.2.c.

Some intuitive discussion of Results 10.1 and 10.2 is in order.

Let \( A_T \) be a \( T \times T \) matrix with \((q,s)\)-th element equal to \( a_{Tqs} \) and define \( B_T \) similarly from the array \( \{b_{Tst}\} \). Then \( C_T = A_T B_T \) has \((q,t)\)-th element equal to \( c_{Tqt} = \sum_{s=1}^{T} a_{Tqs} b_{Tst} \).

Conditions (10.2) place uniform bounds on sums of absolute values of elements of a given row or column of \( A_T \) or \( B_T \). Result 10.2.b establishes the resulting uniform bounds on absolute row and column sums of \( C_T \). The additional hypothesis given in Result 10.2.c requires that...
elements far to the right of the diagonal of any given q-th row of $A_T$ contribute negligibly to $\sum_{s=1}^{T}|a_{Tqs}|$.

In some parts of Section 3.4, the limiting behavior of functions of $A_T$, $B_T$ and $C_T = A_T B_T$ is closely related to the behavior of functions of the associated limits $a_{qs} = \lim_{T \to \infty} a_{Tqs}$, $b_{st} = \lim_{T \to \infty} b_{Tst}$, and $c_{qt} = \sum_{s=1}^{\infty} a_{qs} b_{st}$. Results 10.1, 10.2.a and 10.2.d outline the effect of conditions (10.2) on such limiting arrays.

Section 3.4 also employs arrays with absolute row and column sums that converge to zero at various rates. The following result addresses three forms of such convergence.

**Result 10.3.** Let $\{a_{Tst}; 1 < s, t < T\}$ and $\{b_{Tst}; 1 < s, t < T\}$ be two sequences of $T \times T$ arrays of real numbers, $T=1, 2, ...$. Let $c_{Tqt} = \sum_{s=1}^{T} a_{Tqs} b_{Tst}$.

a. Assume that as $T \to \infty$, $\sum_{q=1}^{T}|a_{Tqs}|$ converges to zero uniformly in $s$; $\sum_{s=1}^{T}|a_{Tqs}|$ converges to zero uniformly in $q$; and assume that $\{b_{Tst}\}$ satisfies condition (10.2). Then $\sum_{q=1}^{T}|c_{Tqt}|$ converges to zero uniformly in $t$ and $\sum_{t=1}^{T}|c_{Tqt}|$ converges to zero uniformly in $q$.

b. Assume that there exist real numbers $K_1 > 0$, $K_2 > 0$, $\alpha$ and $\beta$ such that

$$\sum_{s=1}^{T}|a_{Tst}| < K_1 T^\alpha$$

and

$$\sum_{s=1}^{T}|b_{Tst}| < K_2 T^\beta$$
for all \( 1 < t < T \); and

\[
\sum_{t=1}^{T} |a_{Tst}| < K_1 T^\alpha \quad \text{and} \quad \sum_{t=1}^{T} |b_{Tst}| < K_2 T^\beta
\]

for all \( 1 < s < T \). Then for all \( T \),

\[
\sum_{q=1}^{T} |c_{Tqt}| < K_1 K_2 T^{\alpha+\beta} \quad \text{for all} \quad 1 < t < T
\]

and

\[
\sum_{t=1}^{T} |c_{Tqt}| < K_1 K_2 T^{\alpha+\beta} \quad \text{for all} \quad 1 < s < T.
\]

c. Assume that there are real numbers \( K_1 > 0, K_2 > 0, 0 < \lambda < 1 \), \( \alpha \) and \( \beta \) such that for all \( T \),

\[
\sum_{s=1}^{T} |a_{Tst}| < K_1 \lambda^{\alpha T} \quad \text{and} \quad \sum_{s=1}^{T} |b_{Tst}| < K_2 \lambda^{\beta T}
\]

for all \( 1 < t < T \); and

\[
\sum_{t=1}^{T} |a_{Tst}| < K_1 \lambda^{\alpha T} \quad \text{and} \quad \sum_{t=1}^{T} |b_{Tst}| < K_2 \lambda^{\beta T}
\]

for all \( 1 < s < T \). Then for all \( T \),

\[
\sum_{q=1}^{T} |c_{Tqt}| < K_1 K_2 \lambda^{(\alpha+\beta)T} \quad \text{for all} \quad 1 < t < T;
\]
and

\[ \sum_{t=1}^{T} |c_{Tqt}| < K_1 K_2^{(\alpha+\beta)T} \] for all \( 1 < t < T \).

**Proof.**

**a.** Under condition (10.2), let \( B > 0 \) be a uniform bound on \( \sum_{s=1}^{T} |b_{Tst}| \). By the uniform convergence of \( \sum_{q=1}^{T} |a_{Tqs}| \) to zero, there exists some \( T_0 \) such that for \( T > T_0 \), \( \sum_{q=1}^{T} |a_{Tqs}| < B^{-1} \varepsilon \) for all \( 1 < s < T \). Then for \( T > T_0 \),

\[
\sum_{q=1}^{T} \left| \sum_{s=1}^{T} b_{Tst} \right| |a_{Tqs}| b_{Tst} < \varepsilon
\]

for all \( 1 < t < T \). Uniform convergence of \( \sum_{t=1}^{T} |c_{Tqt}| \) to zero is established similarly.

**b.** Note that

\[
\sum_{q=1}^{T} \left| \sum_{s=1}^{T} a_{Tqs} \right| b_{Tst} < \varepsilon
\]
for all $1 < t < T$. The result for $\sum_{t=1}^{T} |c_{Tqt}|$ is established similarly.

c. Note that

$$\sum_{q=1}^{T} |c_{Tqt}| < \sum_{s=1}^{T} \left| b_{Tst} \right| \left( \sum_{q=1}^{T} |a_{Tqs}| \right)$$

$$< \sum_{s=1}^{T} \left| b_{Tst} \right| K_{1}^{T}$$

$$< K_{1}K_{2}T^{\alpha+\beta}$$

for all $1 < t < T$. The result for $\sum_{t=1}^{T} |c_{Tqt}|$ follows similarly.

Finally, some arrays $\{a_{st}; s, t \in \mathbb{Z}^{+}\}$ may, in the limit, have structure similar to Toeplitz matrices. In particular, consider the following conditions.
(i) For all \(d \in \mathbb{Z}\),

\[
\tilde{a}(d) = \lim_{s \to \infty} a_{s, s+d}
\]
exists and is finite. \hspace{1cm} (10.3.a)

(ii) There exists an absolute summable sequence of positive real numbers \(\{M_{ad}, d \in \mathbb{Z}\}\) such that

\[
|a_{s, s+d} - \tilde{a}(d)| < M_{ad} \text{ for all } s \in \mathbb{Z}^+ .
\hspace{1cm} (10.3.b)

(iii) Also,

\[
\sum_{d=-\infty}^{\infty} |\tilde{a}(d)| < \infty .
\hspace{1cm} (10.3.c)

For notational convenience, define \(a_{st} = 0\) if \(s < 0\) or \(t < 0\).

The following result gives some useful properties of arrays that satisfy conditions (10.3).

**Result 10.4.** Let \(\{a_{qs}; q, s \in \mathbb{Z}^+\}\) and \(\{b_{st}; s, t \in \mathbb{Z}^+\}\) be two doubly semi-infinite arrays of real numbers that satisfy conditions (10.3). Then the following conditions are satisfied.

a. The convergence of \(a_{s, s+d}\) to \(\tilde{a}(d)\) as \(s \to \infty\) is uniform in \(d\).

b. The arrays \(\{a_{qs}\}\) and \(\{b_{st}\}\) satisfy conditions (10.1).

c. Define \(c_{qt} = \sum_{s=1}^{\infty} a_{qs} b_{st}\). The convergence of \(\sum_{s=1}^{T} a_{qs} b_{st}\) to \(c_{qt}\) is uniform in \(q, t \in \mathbb{Z}^+\); and the differences
The array \( \{c_{qt}; q, t \in \mathbb{Z}^+\} \) satisfies conditions (10.3); and the limits \( c(d) = \lim_{s \to \infty} c_{s, s+d} \), \( d \in \mathbb{Z}^+ \), are equal to
\[
\sum_{k=-\infty}^{\infty} \bar{a}(k) \delta(d - k).
\]

**Proof.**

a. By condition (10.3.b), there exists some \( D_\varepsilon \) such that \( |d| > D_\varepsilon \) implies that
\[
|a_{s, s+d} - \bar{a}(d)| < M_d < \varepsilon \quad \text{for all } s \in \mathbb{Z}^+.
\]

By condition (10.3.a), for each \( d \) such that \( |d| < D_\varepsilon \), there exists some \( S_{d, \varepsilon} \) such that if \( s > S_{d, \varepsilon} \), then
\[
|a_{s, s+d} - \bar{a}(d)| < \varepsilon. \quad \text{Let}
\]
\[
S_\varepsilon = \max_{-D_\varepsilon < d < D_\varepsilon} \{S_{d, \varepsilon}\}.
\]

Then for all \( d \in \mathbb{Z}^+ \), \( s > S_\varepsilon \) implies that \( |a_{s, s+d} - \bar{a}(d)| < \varepsilon \) and the result follows.
b. Note first that

\[ |a_{s,s+d}| < |\tilde{a}(d)| + |a_{s,s+d} - \tilde{a}(d)| \]

\[ < |\tilde{a}(d)| + M_{ad} , \]

so

\[ \sum_{t=1}^{\infty} |a_{st}| = \sum_{d=-s+1}^{\infty} |a_{s,s+d}| \]

\[ \leq \sum_{d=-\infty}^{\infty} (|\tilde{a}(d)| + M_{ad}) . \]

The column-absolute summability of \( \{a_{3t}\} \) then follows from conditions (10.3.b) and (10.3.c). A similar proof establishes the row-absolute summability of \( \{a_{st}\} \).

c. The existence of \( c_{qt} \) follows from part (b) and Result 10.1. Also, part (b) implies that there exists some \( M_b > 1 \) such that

\[ |b_{st}| < M_b \text{ for all } s, t \in \mathbb{Z} . \]

Pick \( \varepsilon > 0 \). By conditions (10.3.b) and (10.3.c), there exists some \( T_0 \varepsilon \) such that

\[ \sum_{s=T_0 \varepsilon}^{\infty} (|\tilde{a}(s)| + M_{as}) < M_b^{-1} \varepsilon . \]

From the proof of part (b),

\[ |a_{q,q+d}| < |\tilde{a}(d)| + M_{ad} , \text{ so for } T > T_0 \varepsilon \]
so \( \sum_{s=q}^{T} a_{qs} b_{st} \) converges to \( c_{qt} \) uniformly in \( q, t \in \mathbb{Z}^+ \). Let

\[
A = \sum_{\ell=-\infty}^{\infty} \left[ |\tilde{a}(\ell)| + |\tilde{b}(\ell)| + M_{a\ell} + M_{b\ell} \right] + 1,
\]

and note that \( A \) constitutes a uniform bound on the absolute row and column sums of \( \{a_{qs}\} \) and \( \{b_{st}\} \) as described in conditions (10.1). Thus, for any \( s \in \mathbb{Z}^+ \),

\[
\sum_{q=1}^{T} |a_{qs}| < A \quad (10.4a)
\]

and

\[
\sum_{t=1}^{T} |b_{st}| < A \quad (10.4b)
\]

By conditions (10.3.b) and (10.3.c), there exists some \( T_{1\varepsilon} \) such that

\[
\sum_{\ell=T_{1\varepsilon}+1}^{\infty} \left[ |\tilde{a}(\ell)| + M_{a\ell} \right] < A^{-l_{2}-1\varepsilon}.
\]

Now \( q < T \) and \( s > T + T_{1\varepsilon} + 1 \) imply that \( s-q > T_{1\varepsilon} \), so
Let $T_{2\epsilon} = \max(T_{1\epsilon}, T_{1\epsilon}A^22\epsilon^{-1})$. Then for $T > T_{2\epsilon}$,

$$T^{-1}T_{1\epsilon}A^2 < [T_{1\epsilon}A^22\epsilon^{-1}]^{-1}T_{1\epsilon}A^2 = 2^{-1}\epsilon,$$

so for $T > T_{2\epsilon}$,
Thus,

\[
\lim_{T \to \infty} T^{-1} \sum_{q=1}^{T} \sum_{t=1}^{T} \left| c_{qt} - \sum_{s=1}^{T} a_{qs} b_{st} \right| = 0 .
\]

By conditions (10.3) and part (a), there exists some \( M_1 > 1 \) such that

\[
\sum_{d=-\infty}^{\infty} |a(d)| < M_1 ; \quad \sum_{d=-\infty}^{\infty} |b(d)| < M_1 ; \quad \sum_{d=-\infty}^{\infty} |\tilde{a}(d)| < M_1 ; \quad \sum_{d=-\infty}^{\infty} |\tilde{b}(d)| < M_1 ;
\]

and \( |a_{s,s+d} - \tilde{a}(d)| < M_1 \) and \( |b_{s,s+d} - \tilde{b}(d)| < M_1 \) for all \( s \in \mathbb{Z}^+ \) and all \( d \in \mathbb{Z} \). Then there exists some \( T_1 \epsilon \) such that

\[
\sum_{d=T_1 \epsilon}^{\infty} \left( |\tilde{a}(\lambda)| + |\tilde{a}(-\lambda)| \right) < M_1^{-1} \epsilon \]
so that for any \( d \in \mathbb{Z} \),

\[
\sum_{\ell=1}^{\infty} \left[ |\tilde{a}(\ell)\tilde{b}(d-\ell)| + |\tilde{a}(-\ell)\tilde{b}(\ell-d)| \right] < \varepsilon.
\]

Now

\[
\begin{align*}
\sum_{\ell=1}^{\infty} a_{s,s+\ell} b_{s+\ell,s+d} &= \sum_{\ell=1}^{\infty} \tilde{a}(\ell)\tilde{b}(d-\ell) \\
&= \sum_{\ell=1}^{\infty} \tilde{a}(\ell)[\tilde{b}(d-\ell) + b_{s+\ell,s+d} - \tilde{b}(d-\ell)] \\
&\quad - \sum_{\ell=1}^{\infty} \tilde{a}(\ell)\tilde{b}(d-\ell) \\
&= \sum_{\ell=1}^{\infty} \tilde{a}(\ell)[b_{s+\ell,s+d} - \tilde{b}(d-\ell)] \quad (10.5.a) \\
&\quad + \sum_{\ell=1}^{\infty} [a_{s,s+\ell} - \tilde{a}(\ell)]\tilde{b}(d-\ell) \quad (10.5.b) \\
&\quad + \sum_{\ell=1}^{s} [a_{s,s+\ell} - \tilde{a}(\ell)][b_{s+\ell,s+d} - \tilde{b}(d-\ell)] \quad (10.5.c) \\
&\quad - \sum_{\ell=1}^{s} \tilde{a}(\ell)\tilde{b}(d-\ell). \quad (10.5.d)
\end{align*}
\]

Fix \( d \in \mathbb{Z} \). By part (a), there exists some \( S_{d,\varepsilon} \) such that if

\( s > S_{d,\varepsilon} \), then \(|a_{s,s+\ell} - \tilde{a}(\ell)| < M_{1}\varepsilon \) and

\(|b_{s+d,s+\ell} - \tilde{b}(d-\ell)| < M_{1}\varepsilon \) for all \( \ell \in \mathbb{Z} \), which in turn implies
that expressions (10.5.a), (10.5.b) and (10.5.c) are each bounded in modulus by $\epsilon$. Moreover, if $s > T_1\epsilon$, then

$$
\mathcal{L}^s_{-\infty} \left| \tilde{a}(\ell) \tilde{b}(d - \ell) \right| < M \mathcal{L}^s_{-\infty} |\tilde{a}(\ell)|
$$

$$
< \epsilon.
$$

Thus, if $s > \max\{T_1\epsilon, T_2\epsilon\}$, then $|c_{s,s+d} - \sum_{l=-\infty}^{\infty} \tilde{a}(\ell) \tilde{b}(d - \ell)| < 4\epsilon$, so for each $d \in \mathcal{D}$, $\tilde{c}(d) \equiv \lim_{s \to \infty} c_{s,s+d}$ exists, is finite, and is equal to $\sum_{l=-\infty}^{\infty} \tilde{a}(\ell) \tilde{b}(d - \ell)$.

To show that $\{c_{st}\}$ satisfies conditions (10.3.b), note that

$$
|c_{s,s+d} - \tilde{c}(d)| < \sum_{l=1-s}^{\infty} \left| \tilde{a}(\ell) \tilde{b}(d - \ell) \right| + \left| [a_{s+\ell} - \tilde{a}(\ell)] \tilde{b}(d - \ell) \right|
$$

$$
+ \left| \tilde{a}(\ell) [b_{s+\ell,s+d} - \tilde{b}(d - \ell)] \right|
$$

$$
+ \left| [a_{s+\ell} - \tilde{a}(\ell)] [b_{s+\ell,s+d} - \tilde{b}(d - \ell)] \right|
$$

$$
+ \sum_{l=-\infty}^{\infty} \left| \tilde{a}(\ell) \tilde{b}(d - \ell) \right|
$$

$$
< \sum_{l=-\infty}^{\infty} \left[ 2 \left| \tilde{a}(\ell) \tilde{b}(d - \ell) \right| + M_{a,\ell} \left| \tilde{b}(d - \ell) \right| \right]
$$

$$
+ \left| \tilde{a}(\ell) \right| M_{b,d-\ell} + M_{a,\ell} M_{b,d-\ell}
$$
The result \( \sum_{d=-\infty}^{\infty} M_{cd} < \infty \) then follows from the fact that the convolution of two absolutely summable sequences is itself absolutely summable [c.f. Fuller (1976, p. 28)]. This same fact also implies that \( \{c(d), d \in \mathbb{Z}\} \) satisfies condition (10.3.c). \( \square \)
APPENDIX C.

MAXIMUM LIKELIHOOD ESTIMATION FOR MEASUREMENT ERROR MODELS
WITH COMBINED TIME SERIES AND CROSS SECTIONAL DATA

The estimation procedures considered to this point have been based on observations from a single realization of a multivariate time series $Z_t$, $t=1, 2, \ldots, T$. In practice, this would correspond to the observation of a single "unit" over time, where the "unit" may be a single sampling unit or the aggregate of several sampling units. Some studies in engineering and the social sciences may allow one to observe simultaneously $N$ "cross sectional" units over $T$ time periods. It is therefore desirable to consider measurement error models for combined time series and cross sectional data.

For the non-measurement error case, there exists a considerable body of literature devoted to the estimation of regression coefficients and residual variance components; see, e.g., Balestra and Nerlove (1966), Swamy and Arora (1972), Fuller and Battese (1974), and Anderson and Hsiao (1981, 1982). In the above-mentioned papers, the Helmert transformation and similar analysis of variance ideas play central roles in the development of estimators. In addition, Henderson (1971) has noted that for simple error structures, the problem of regression estimation with combined time series and cross-sectional data is a special case of the general mixed linear model.

In the analysis of measurement error models for combined time series and cross-sectional data, one again may note many similarities to
general linear model theory and the analysis of variance. In particular, one may use Helmert transformations to decompose the cross-sectional covariance structure of the $x$ and $\zeta$ components and thus obtain independent innovation sequences of low dimension. As in Chapter 5, given a sufficiently regular likelihood surface, these innovation sequences may lead to a Newton-Raphson procedure for maximum likelihood estimation of $\beta$ and identified parameters of the $x$ and $\epsilon$ processes.

The details of such a procedure are dependent on the covariance structure among the cross-sectional units. The sections below address some simple cases of such cross-sectional covariance structures and their associated innovation sequences.

11.1. $N$ Independent Cross-Sectional Units

Let $z_{ti}$, $t=1, 2, \ldots, T$, $i=1, 2, \ldots, N$, be the $NT$ $p$-dimensional observations taken on $N$ cross-sectional units at $T$ times. Generalize model (2.3)-(2.4) to the following model:

$$z_{ti} = x_{ti}(\beta, I_k) + \epsilon_{ti}, \quad (11.1)$$

where $\beta$ is a $k \times r$ matrix of unknown regression parameters, and for each $i$, \{ $x_{ti}$, $i=1, 2, \ldots, T$ \} is an independent realization of a stationary normal ARMA$_k(p_x, q_x)$ process, \{ $\epsilon_{ti}$, $i=1, 2, \ldots, T$ \} is an independent realization of a stationary normal ARMA$_p(p_\epsilon, q_\epsilon)$ process, and $x_{ti}$ is independent of $\epsilon_{sj}$ for all $i, j, s, \text{ and } t$. Let
and define \( x_t \), \( x \), \( \varepsilon_t \), and \( \varepsilon \) similarly. Denote the vectors of parameters of the \( x \) and \( \varepsilon \) processes as \( \alpha_x \) and \( \alpha_e \), respectively, and assume that \( \alpha_x \), \( \alpha_e \) and \( \beta \) are functionally unrelated.

Assume that \( x \) and \( \varepsilon \) have a joint normal distribution, let \( \ell(\alpha_x, \alpha_e, \beta; z_{.1}) \) be the log-likelihood function for \( z_{.1} \), \( i=1, 2, \ldots, N \), and let \( \ell(\alpha_x, \alpha_e, \beta; z_{.}) \) be the log-likelihood function for \( z_{.} \).

The independence assumptions stated above imply that

\[
\ell(\alpha_x, \alpha_e, \beta; z_{.}) = \sum_{i=1}^{N} \ell(\alpha_x, \alpha_e, \beta; z_{.1}).
\]

Therefore, if we let \( \theta = [\text{vec}(\beta)', \alpha_x, \alpha_e] \),

\[
\frac{\partial}{\partial \theta} [\ell(\theta; z_{.})] = \sum_{i=1}^{N} \frac{\partial}{\partial \theta} [\ell(\theta; z_{.1})].
\]

For each \( i \), \( \ell(\theta; z_{.1}) \) has the same form as the likelihood (5.3) for a single realization in Section 5.1. Therefore, requirements of computational tractability again indicate that it is preferable to use the log-likelihood of an innovation sequence \( \{d_{t1}, t=1, 2, \ldots, T\} \),

\[
\begin{align*}
\hat{z}_{.1} &= (z_{11}, z_{21}, \ldots, z_{1N}), \\
\hat{z}_t &= (z_{t1}, z_{t2}, \ldots, z_{tN}), \\
\hat{z}_t &= (z_{.1}, z_{.2}, \ldots, z_{.N}),
\end{align*}
\]
rather than using \( z(\delta; z_{i1}) \) directly. To develop the innovation sequences, note that for each \( i=1, 2, \ldots, N \), the model for \( \{z_{ti}, t=1, 2, \ldots, T\} \) may be written in state-space form

\[
Z'_{ti} = B_{W}w_{ti}'
\]

\[
w_{ti}' = A_{W}w_{ti}' + C_{W}g_{ti}' , \quad t=1, 2, \ldots, T ,
\]

where \( A_{W} , B_{W} \), and \( C_{W} \) are defined preceding (5.43)-(5.44), and \( w_{ti}' \) and \( g_{ti}' \) are defined by the expressions for \( w_{ti} \) and \( g_{ti} \) preceding (5.43)-(5.44), but with \( f_{c}^{*} \) replaced by \( x_{i}^{*} = (x_{ti}, s_{ti}) \).

For each \( i=1, 2, \ldots, N \) and each \( t=1, 2, \ldots, T \), one then obtains the prediction equation

\[
\hat{w}_{t+1|i} = A_{W}w_{t+1|i}'
\]

and the updating equation

\[
\hat{w}_{t+1,i} = \hat{w}_{t+1|i} + P_{t+1|i} B_{W}^{-1}(Z'_{t+1,i} - B_{W}w_{t+1|i}|t,i)
\]

where \( P_{t+1|i} \), \( P_{t+1} \), and \( A_{t+1} \) are given by expressions (5.46), (5.48) and (5.49), respectively. The innovations

\[
d_{ti}' = z_{ti}' - B_{W}w_{t|t-1,i}'
\]
\[ i=1, 2, \ldots, N, \quad t=1, 2, \ldots, T, \text{ form a sequence of normal and independent } (0, \Lambda_t) \text{ random vectors. Hence, one may rewrite the log-likelihood function of } Z_{..} \text{ in the form,} \]

\[ f(g; Z_{..}) = \sum_{i=1}^{N} f(g; Z_{...,i}) \]

\[ = -2^{-1}NT \ln(2\pi) - 2^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \ln |\Lambda_t| + d_{t,i} \Lambda_t^{-1} d_{t,i}' \right\}. \]

Consequently, derivative computations and the resulting Newton-Raphson maximum likelihood estimation procedure follow the same pattern as in Section 5.4, but with the time index \( t \) replaced by the time and cross sectional unit indices \( t \) and \( i \), and with summation over \( t \) replaced by summation over \( t \) and \( i \).

11.2. Nested Inputs and Errors

The preceding section extended the innovation-sequence approach of Section 5.4 from the case of a single realization of a process to the case of \( N \) independent realizations of the given stochastic process. In practice, observations may be correlated across cross-sectional units as well as across time. The present section addresses a nested-term model for such cross-sectional correlation.

Retain the notation of Section 11.1 and model (11.1), but now assume that

\[ x_{ti} = x_{0t} + x_{1ti} \quad (11.2) \]
and

\[ \varepsilon_{ti} = \varepsilon_{0t} + \varepsilon_{lti}, \quad t=1, 2, \ldots, T, \ i=1, 2, \ldots, N, \]

where \( \{x_{0t}, t=1, 2, \ldots, T\} \) is a realization of a stationary ARMA_{k}(p_{x0}, q_{x0}) process; \( \{\varepsilon_{0t}, t=1, 2, \ldots, T\} \) is a realization of a stationary ARMA_{p}(p_{e0}, q_{e0}) process; for each \( i=1, 2, \ldots, N \), \( \{x_{lti}, t=1, 2, \ldots, T\} \) is an independent realization of a stationary ARMA_{k}(p_{x1}, q_{x1}) process and \( \{\varepsilon_{lti}, t=1, 2, \ldots, T\} \) is an independent realization of an ARMA_{p}(p_{e1}, q_{e1}) process; and \( x_{0t}, \varepsilon_{0s}, x_{lti}, \) and \( \varepsilon_{lmj} \) are independent for all \( t, s, l, i, m \) and \( j \).

One now requires an additional transformation to obtain \( N \) independent "innovation sequences" from which the log-likelihood of \( Z \) may be derived. Let

\[ \mathbf{x}_{0t} = (x_{0t1}, x_{0t2}, \ldots, x_{0tk}), \]

\[ \mathbf{x}_{lti} = (x_{lti1}, x_{lti2}, \ldots, x_{ltik}), \]

\[ \mathbf{\varepsilon}_{0t} = (\varepsilon_{0t1}, \varepsilon_{0t2}, \ldots, \varepsilon_{0tp}), \]

\[ \mathbf{\varepsilon}_{lti} = (\varepsilon_{lti1}, \varepsilon_{lti2}, \ldots, \varepsilon_{ltip}), \]

\[ \mathbf{x}_{lt} = (x_{lt1}, x_{lt2}, \ldots, x_{ltN}). \]
\[ x_{1,i} = (x_{11i}, x_{12i}, \ldots, x_{1Ti}) , \]

and define \( \varepsilon_{1t} \) and \( \varepsilon_{1,i} \) similarly. Let

\[ \Gamma_{\text{xx00}}(h) = \text{Cov}(x'_0, x'_0, t+h) \]

\[ \Gamma_{\text{xx11}}(h) = \text{Cov}(x'_{1t}, x'_{1,t+h}, i) \]

and define \( \Gamma_{\varepsilon \varepsilon 00}(h) \) and \( \Gamma_{\varepsilon \varepsilon 11}(h) \) similarly. Since \( \{x_{1,i}, i=1, 2, \ldots, N\} \) are \( N \) independent realizations of the same stationary process, \( \Gamma_{\text{xx11}}(h) \) is the same for all \( i \); \( \text{Cov}(x'_{1t}, x'_{1s}) = 0 \) for all \( i \neq j \) and all \( t \) and \( s \); and similarly for \( \Gamma_{\varepsilon \varepsilon 11}(h) \) and \( \text{Cov}(\varepsilon_{1t}, \varepsilon_{1s}) \). Thus,

\[ \text{Cov}(x'_{1t}, x'_{1,t+h}, i) = I_N \otimes \Gamma_{\text{xx11}}(h) . \]

Note that \( x'_t = l_N \otimes x'_0 + x'_{1t}, \) so

\[ \text{Cov}(x'_t, x'_{t+h}) = \text{Cov}(l_N \otimes x'_0, l_N \otimes x'_{0,t+h}) + \text{Cov}(x'_{1t}, x'_{1,t+h}, i) \]

\[ = (l_N \otimes I_N) \otimes \Gamma_{\text{xx00}}(h) + I_N \otimes \Gamma_{\text{xx11}}(h) . \] (11.3)

Therefore, the first step in obtaining \( N \) independent "innovation sequences" for maximum likelihood estimation is to find a transformation that will reduce the covariance structure (11.3) to block-diagonal
form. Once may obtain such a transformation from some elementary matrix results discussed in Morrison (1976, pp. 289–290).

Lemma 11.1. Let $a$ and $b$ be real numbers and let $N$ be a positive integer. Then $a(I_N I_N') + b I_N$ has one eigenvalue equal to $Na + b$, with corresponding normalized eigenvector $m_1 = N^{-1/2} I_N$ and has $(N - 1)$ eigenvalues equal to $b$, with corresponding eigenvectors equal to any $N - 1$ vectors that span $\mathbb{C}^I(I_N)$.

Proof. Note that $I_N I_N' = N$, so

$$[a(I_N I_N') + b I_N](N^{-1/2} I_N)$$

$$= Na(N^{-1/2} I_N) + b(N^{-1/2} I_N) = (Na + b)(N^{-1/2} I_N) ,$$

and the first part of the lemma is established. For any $m \in \mathbb{C}^I(I_N)$, $I_N m = 0$, so $[a(I_N I_N') + b I_N] m = 0 + bm$, so the second part of the lemma follows.

Note that one normalized spanning set for $\mathbb{C}^I(I_N)$ is 

$\{m_2, m_3, \ldots, m_N\}$, where

$$m_2 = 2^{-1/2} (1, -1, 0, \ldots, 0)' ,$$

$$m_3 = 2^{-1/2} (0, 1, -1, 0, \ldots, 0)' , \ldots ,$$
each $m_i$, $i=2, 3, \ldots, N$ is $N \times 1$. Let $Q_N$ be an $N \times N$ matrix with $i$-th row equal to $m_i$. By Lemma 11.1 and the usual spectral decomposition of a real symmetric matrix,

$$Q_N [a(I_{N \times N} I'_1) + bI_N]Q'_N = \text{diag}[(aN + b), b, b, \ldots, b].$$

Therefore, by expression (11.3),

$$\text{Cov}[(Q_N \boxtimes_k x'_i, (Q_N \boxtimes_k x'_i) + h, \ldots, h) = (Q_N \boxtimes_k x'_i)(I_{N \times N} I'_{N \times N} + I_N \boxtimes_{1 \times 11}(h))(Q_N \boxtimes_k x'_i)$$

$$= [Q_N (I_{N \times N} I'_N) Q'_N] \boxtimes_{1 \times 00}(h) + I_N \boxtimes_{1 \times 11}(h)$$

$$= (e_{N1} e'_{N1} + [N \boxtimes_{1 \times 00}(h)] + I_N \boxtimes_{1 \times 11}(h),$$

where $e_{N1}$ is an $N \times 1$ vector with unity in its first row and zeros elsewhere. Similarly,

$$\text{Cov}[(Q_N \boxtimes_p e'_i, (Q_N \boxtimes_p e'_i) + h, \ldots, h) = (e_{N1} e'_{N1} + [N \boxtimes_{e \in 00}(h)] + I_N \boxtimes_{e \in 11}(h).$$
It follows from model (11.1)-(11.2) that

\[ Z_t = x_t [I_N \oplus (g, I_k)] + \varepsilon_t. \]

The work above and the independence of \( x_t \) from \( \varepsilon_t \) then imply that

\[
\text{Cov}[(Q_N \oplus I_p)z_t', (Q_N \oplus I_p)z_{t+h}']
\]

\[
= (Q_N \oplus I_p)[I_N \oplus (g, I_k)'] \text{Cov}(x_t', x_{t+h}')[I_N \oplus (g, I_k)](Q_N \oplus I_p)
\]

\[
+ \text{Cov}[(Q_N \oplus I_p)\varepsilon_t, (Q_N \oplus I_p)\varepsilon_{t+h}']
\]

\[
= [Q_N \oplus (g, I_k)'] (Q_N' \oplus I_k) \text{Cov}[(Q_N \oplus I_k)x_t', (Q_N \oplus I_k)x_{t+h}']
\]

\[
\times (Q_N \oplus I_k)[Q_N' \oplus (g, I_k)]
\]

\[
+ \text{Cov}[(Q_N \oplus I_p)\varepsilon_t, (Q_N \oplus I_p)\varepsilon_{t+h}']
\]

\[
= [I_N \oplus (g, I_k)'] \{(e_1 e_1') \oplus [N \Gamma_{xx00}(h)] + I_N \oplus \Gamma_{xx11}(h) \}
\]

\[
\times [I_N \oplus (g, I_k)]
\]

\[
+ \{(e_1 e_1') \oplus [N \Gamma_{ee00}(h)] + I_N \oplus \Gamma_{ee11}(h) \}
\]
The matrix following the final equality is block diagonal with blocks of dimension \( p \times p \). Define \( \tilde{Z}'_t. = (\tilde{Z}'_{t1}, \tilde{Z}'_{t2}, \ldots, \tilde{Z}'_{tN})' = (Q_N \otimes I_p)Z'_t. \), and define \( \tilde{Z}'_t. \) and \( \tilde{Z}'_i. \) accordingly. Similarly, let \( \tilde{X}'_t. = (Q_N \otimes I_k)x'_t. \) and \( \tilde{\xi}'_t. = (Q_N \otimes I_p)\xi'_t. \), and define \( \tilde{X}'_{t1}, \tilde{X}'_{ti}, \tilde{X}'_i, \tilde{\xi}'_{t1}, \tilde{\xi}'_{ti}, \tilde{\xi}'_i, \) and \( \tilde{\xi}'_i. \) accordingly. It follows from the block-diagonal structure of (11.4) that the Helmert transformation from \( Z'_t. \) to \( \tilde{Z}'_t. \) has yielded \( N \) uncorrelated \( pT \times 1 \) vectors \( \tilde{Z}'_i. \). The covariance structure of \( \tilde{Z}'_i. \) is described by

\[
\text{Cov}(\tilde{Z}'_{t1}, \tilde{Z}'_{t+h,1}) = (\beta, I_k)'[N \Gamma_{xx00}(h) + \Gamma_{xx11}(h)](\beta, I_k)
\]

\[+ [N \Gamma_{e00}(h) + \Gamma_{e11}(h)] ,
\]

and the covariance structure of \( \tilde{Z}'_i, i=2, 3, \ldots, N \) is described by

\[
\text{Cov}(\tilde{Z}'_{ti}, \tilde{Z}'_{t+h,i}) = (\beta, I_k)'\Gamma_{xx11}(h)(\beta, I_k) + \Gamma_{e11}(h) .
\]

To develop the innovation sequence associated with the transformed observation vector \( \tilde{Z}'_1. \), define

\[\tilde{X}'_{0t} = (X'_{0t}, \zeta'_{0t}) .\]
\[ \tilde{x}_{11} = \left( \tilde{x}_{111}, \tilde{x}_{112} \right), \]

\[ W_{11} = \left( \tilde{x}_{011} | t, \tilde{x}_{111} | t, \tilde{x}_{011+1} | t, \tilde{x}_{111+1} | t, \ldots, \right), \]

\[ \tilde{x}_{011+J-1} | t, \tilde{x}_{111+J-1} | t \right), \]

\[ g_{-1,-1} = \left( \tilde{x}_{01t} | t, \tilde{x}_{11t} | t \right) - \left( \tilde{x}_{01t-1}, \tilde{x}_{11t-1} \right), \]

\[ B_1 = [N^{1/2} (\mathbb{E}, I_k)'', N^{1/2} I_p', (\mathbb{E}, I_k)'', I_p', 0_{p \times 2(J-1)(p+k)}], \]

\[ A_1 = \begin{bmatrix} 0 & I_{2(p+k)} & 0 & 0 & \cdots & 0 \\ 0 & 0 & I_{2(p+k)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I_{2(p+k)} \\ -g_{-1,j} & -g_{-1,j-1} & -g_{-1,j-2} & -g_{-1,j-3} & \cdots & -g_{-1,J-1} \end{bmatrix}, \]

\[ C_1 = \left[ I_{2(p+k)}, \tilde{x}_{111}', \ldots, \tilde{x}_{11J-2}', \tilde{x}_{11J-1}' \right], \]

\[ J = \max(p_{0x}, q_{0x+1}, p_{0e}, q_{0e+1}, p_{1x}, q_{1x+1}, p_{1e}, q_{1e+1}), \]

\( g_{1j}, j=1, 2, \ldots, J \) and \( g_{11}, j=1, 2, \ldots, J-1 \) are the \( 2(p+k) \times 2(p+k) \) dimensional autoregressive and moving average parameter matrices, respectively, of the \( (\tilde{x}_{01}, \tilde{x}_{11}) \) process,
\[ \Phi_{ij}^* = 0 \text{ for } j > \max(p_{x0}, p_{e0}, p_{x1}, p_{e1}), \]

\[ \Phi_{ij} = 0 \text{ for } j > \max(q_{x0}, q_{e0}, q_{x1}, q_{e1}) + 1, \]

\( \{ \Psi_{ij}, j=0, 1, \ldots, J-1 \} \) are defined from \( \Phi_{ij}^* \) and \( \Phi_{ij} \) by expression (5.37), and \( p_{x0}, p_{e0}, p_{x1}, p_{e1}, q_{x0}, q_{e0}, q_{x1} \) and \( q_{e1} \) are the autoregressive and moving average orders of the designated processes. Then

\[ \tilde{Z}_{t1} = B_{1} \hat{W}_{t1}^* \]

and

\[ \hat{W}_{t+1,1} = A_{1} \hat{W}_{t1}^* + C_{1} \tilde{g}_{t1}^* \]

constitute the measurement and transition equations, respectively, of a state-space representation of the transformed observation vector \( \tilde{Z}_{t1} \). Hence, the associated Kalman filter prediction equations are

\[ \hat{W}_{t+1,1|t} = A_{1} \hat{W}_{t1}^* \]

and

\[ P_{t+1,1|t} = \text{Var}(\hat{W}_{t+1,1|t} - \hat{W}_{t+1,1}) \]

\[ = A_{1} P_{t1} A_{1}^* + C_{1} \Sigma_{gg} C_{1}^* \]
where \( \Sigma_{g1} = \text{Var}(g_i^1) \); and the updating equations are

\[
\hat{w}_{t+1,1} = \hat{w}_{t+1,1} \| t + P_{t+1,1} \| t B_1' A_{t+1,1}^{-1} (\tilde{z}_{t+1,1} - B_1 \hat{w}_{t+1,1} \| t)
\]

and

\[
P_{t+1,1} = P_{t+1,1} \| t - P_{t+1,1} \| t B_1' A_{t+1,1}^{-1} B_1 P_{t+1,1} \| t
\]

where

\[
A_{t+1,1} = B_1 P_{t+1,1} \| t B_1'.
\]

Consequently, the innovations

\[
d_{t+1} = \tilde{z}_{t+1} - B_1 \hat{w}_{t+1,1} \| t-l, \quad t=1, 2, ..., T,
\]

form a sequence of independent normal \((0, A_{t+1,1})\) random vectors.

For each \( i=2, 3, ..., N \), one may develop the innovation sequence associated with the transformed observation vector \( \tilde{z}_{i,t} \) by defining

\[
\tilde{x}_{t+1} = (\tilde{x}_{t+1,1}, \tilde{x}_{t+1,1}^1, ..., \tilde{x}_{t+1,i}^1)
\]

\[
\tilde{w}_{t+1,1} = (\tilde{w}_{t+1,1}^1, ..., \tilde{w}_{t+1,i}^1)
\]

\[
\tilde{e}_{t+1,1} = \tilde{x}_{t+1,1} - \tilde{w}_{t+1,1} \| t-l
\]
\[ B_2 = [(\theta, \mathbf{I}_k)', \mathbf{I}_p', \mathbf{0}_{p \times (k-1)(p+k)}] , \]

\[
A_2 = \begin{bmatrix}
0 & \mathbf{I}_{p+k} & 0 & 0 & \ldots & 0 \\
0 & 0 & \mathbf{I}_{p+k} & 0 & \ldots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \ldots & \mathbf{I}_{p+k} \\
-\tilde{\omega}_k & -\tilde{\omega}_{k-1} & -\tilde{\omega}_{k-2} & -\tilde{\omega}_{k-3} & \ldots & -\tilde{\omega}_1 \\
\end{bmatrix}
\]

\[ C_2 = [\mathbf{I}_{p+k}', \tilde{\omega}_1', \ldots, \tilde{\omega}_{k-2}', \tilde{\omega}_{k-1}']' \]

\[ K = \max(p_{1x}, q_{1x}+1, p_{1e}, q_{1e}+1) , \]

\{\tilde{\omega}_j, j=1, 2, \ldots, K\} and \{\theta^*_j, j=1, 2, \ldots, K-1\} are the \((p+k) \times (p+k)\) dimensional autoregressive and moving average parameter matrices, respectively, of the \(\hat{Z}_{t_1}^*\) process, \(\tilde{\omega}_j = 0\) for \(j > \max(p_{1x}, p_{1e})\), \(\theta^*_j = 0\) for \(j > \max(q_{1x}, q_{1e}) + 1\), \(\{\tilde{\omega}_j, j=0, 1, \ldots, K-1\}\) are defined from \(\tilde{\omega}_j^*\) and \(\theta^*_j\) by expression (5.37), and \(p_{1x}, p_{1e}, q_{1x}\) and \(q_{1e}\) are the autoregressive and moving average orders of the designated processes. Then

\[ \hat{Z}_{t_1}' = B_2 W_{t_1}^* \]

and
\[ W_{t+1,1}^t = A_2 W_{t2}^t + C_2 g_{t2}^* \]

constitute the measurement and transition equations, respectively, of a state-space representation of the transformed observation vector \( \tilde{Z}_{t1} \), \( i=2, 3, \ldots, N \). Hence for \( i=2, 3, \ldots, N \), the associated Kalman filter prediction equations are

\[ \hat{W}_{t+1,1|t} = A_2 \hat{W}_{t2}^t \]

and

\[ P_{t+1,2|t} = \text{Var}(\hat{W}_{t+1,1|t} - \hat{W}_{t+1,1}^t) \]

\[ = A_2 P_{t2} A_2' + C_2 \Sigma_{gg22} C_2' \]

where \( \Sigma_{gg22} = \text{Var}(g_{t2}^*) \); and the updating equations are

\[ \hat{W}_{t+1,1} = \hat{W}_{t+1,1|t} + P_{t+1,2|t} B_{t2} A_{t2}^{-1} (\tilde{Z}_{t+1,1} - B_2 \hat{W}_{t+1,1|t}) \]

and

\[ P_{t+1,2} = P_{t+1,2|t} - P_{t+1,2|t} B_{t2} A_{t2}^{-1} B_{t2} P_{t+1,2|t} \]

where

\[ A_{t+1,2} = B_2 P_{t+1,2|t} B_{t2}^{-1} \]
Consequently, the innovations

$$d_i' = \hat{Z}_n^i - \hat{B}_2'Z_{i1}'|_{t-1}, \quad t=1, 2, \ldots, N, \quad i=2, 3, \ldots, N,$$

form \((N-1)\) independent sequences of independent normal \((0, A_{c2})\) random vectors.

Let \(L(g; Z)\) denote the multivariate normal likelihood function of \(\text{vec}(Z)\). Then by the innovation arguments above,

$$-2 \ln[L(g; Z)] - NT\ln(2\pi)$$

$$= \sum_{t=1}^{T} \left[ \ln|A_{c1}^{-1}| + d_i A_{c1}^{-1}d_i' \right] + \sum_{i=2}^{N} \sum_{t=1}^{T} \left[ \ln|A_{c2}^{-1}| + d_i A_{c2}^{-1}d_i' \right].$$

Derivative computations and the resulting Newton-Raphson maximum likelihood estimation procedure then follow from this log-likelihood expression in the same patterns as in Section 5.4.

11.3. Multi-Stage Nested Inputs and Errors

Retain the notation and assumptions of the preceding sections and model (11.1), but now assume that a total of \(M\) units are observed simultaneously over \(T\) time periods such that each sampling unit falls into one of \(N\) groups, \(M_i\) units are contained in the \(i\)-th group, \(i=1, 2, \ldots, N\); and \(M = \sum_{i=1}^{N} M_i\). Let \(x_{ij}\) be the \(k\)-dimensional input vector for unit \(j\) in group \(i\) at time \(t\), define the \(p\)-dimensional error vector \(z_{ij}\) similarly, replace model (11.1) with the
model,

$$Z_{tij} = x_{tij} (\theta, I_k) + \xi_{tij},$$  \hspace{1cm} (11.5)

and replace model (11.2) with the model,

$$x_{tij} = x_{0t} + x_{lti} + x_{2tij},$$  \hspace{1cm} (11.6)

$$\xi_{tij} = \xi_{0t} + \xi_{lti} + \xi_{2tij}, \quad t=1, 2, \ldots, T; \quad j=1, 2, \ldots, M_j; \quad i=1, 2, \ldots, N;$$

where $x_{0t}$, $x_{lti}$, $\xi_{0t}$, and $\xi_{lti}$ follow the models indicated below (11.2); for each $i=1, 2, \ldots, N$ and $j=1, 2, \ldots, M_j$,

\{x_{2tij}, \ t=1, 2, \ldots, T\} is an independent realization of a stationary ARMA_{p|\phi, q_x|\theta} process and \{\xi_{2tij}, \ t=1, 2, \ldots, T\} is an independent realization of a stationary ARMA_{p|\phi, q_\xi|\theta} process; and $x_{0t}$,

$x_{lti1}$, $x_{lti2}$, $\xi_{lt1}$, $\xi_{lt2}$, $\xi_{lt1}'$, $\xi_{lt2}'$, $\xi_{lt1}''$, $\xi_{lt2}''$, $\xi_{lt1}'''$, $\xi_{lt2}'''$, $\xi_{lt1}''''$, $\xi_{lt2}''''$, $\xi_{lt1}'''', \xi_{lt2}'''', \xi_{lt1}''''', \xi_{lt2}'''''$ are independent for all $t$, $t'$, $t''$, $i$, $i'$, $j$, $s$, $s'$, $s''$, $z$, $z'$ and $m$.

In order to obtain independent innovation sequences, one requires a generalization of the Helmert transformation given in the preceding section. Define

$$x_{2ti.} = (x_{2ti1}, x_{2ti2}, \ldots, x_{2tii})', \quad x_{2ti.} = (x_{2ti1}, x_{2ti2}, \ldots, x_{2tiN})'.$$
\[ \mathbf{x}_{2i} = (x'_{21i}, x'_{22i}, \ldots, x'_{2Ti})', \]

and define \( \mathbf{e}_{2i} \), \( \mathbf{e}_{2t} \), and \( \mathbf{e}_{2i} \) similarly. Let

\[ \Gamma_{xx22}(h) = \text{Cov}(\mathbf{x}_{2tij}, \mathbf{x}_{2t+h,ij}) \]

and define \( \Gamma_{ee22}(h) \) similarly.

Since \( \{\mathbf{x}_{2ij}, i=1, 2, \ldots, N, j=1, 2, \ldots, M\} \) are \( M \) independent realizations of the same stationary process, \( \Gamma_{xx22}(h) \) is the same for all \( i \) and \( j \); \( \text{Cov}(\mathbf{x}_{2tij}, \mathbf{x}_{2sij}) = 0 \) for all \( t \) and \( s \) unless \( i = \ell \) and \( j = m \); and similarly for \( \Gamma_{ee22}(h) \) and \( \text{Cov}(\mathbf{e}_{2tij}, \mathbf{e}_{2sij}) \).

Now \( \mathbf{x}_{t} = (I_M \otimes I_k)x_{0t} + (\mathbf{M}_1 \otimes I_k)x_{1t} + x_{2t} \),

where

\[ \mathbf{M}_1 = \begin{pmatrix} I_{M_1} & 0 & \cdots & 0 \\ 0 & I_{M_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{M_N} \end{pmatrix}, \]

an \( M \times N \) matrix. Thus,
Thus, in order to obtain $M$ uncorrelated $k$-dimensional time series of length $T$ each, we must find an $M \times M$ matrix $Q$ that simultaneously diagonalizes the matrices $\{I_M, I'_M\}$, $(R'_1 R'_1)$ and $I_M$. By Corollary 12.2.12.1 of Graybill (1983, p. 408), such a matrix $Q$ exists if and only if the latter three matrices commute. Note that

$$R'_1 R'_1 = \text{block diag}(I_{M_1}, I'_{M_1}, I_{M_2}, I'_{M_2}, \ldots, I_{M_N}, I'_{M_N}),$$

so

$$\begin{pmatrix}
\begin{array}{c}
M_1(I_{M_1}, I'_{M_1}) \\
M_2(I_{M_2}, I'_{M_2}) \\
\vdots \\
M_N(I_{M_N}, I'_{M_N})
\end{array}
\end{pmatrix}.$$
The symmetry of $R_i R'_i$ and $l_M l'_M$ implies that these two matrices commute if and only if their product is symmetric; by inspection, this latter condition holds if and only if the group sizes $M_i$, $i=1, 2, \ldots, N$ are equal. This result closely parallels the well-known analysis of variance result that the usual partition of sums of squares into within-group and between-group sums of squares is "orthogonal" if and only if all group sizes are equal. Therefore, it does not appear that the methods of the preceding section extend directly to multi-stage nested designs with unequal group sizes.

For equal group sizes, $M_i = M$, $i=1, 2, \ldots, N$, note that

$$l_M l'_M = I_N \otimes I_M \quad \text{and} \quad R_i R'_i = I_N \otimes I_M.$$ 

In this case, (11.7) becomes

$$(1, 1') \otimes (1, 1') \otimes I_N \otimes I_M \sim \gamma_{xx00}(h) + I_N \otimes I_M \otimes \gamma_{xx11}(h) + I_N \otimes I_M \otimes \gamma_{xx22}(h).$$

(11.8)

Premultiplication by $Q_N \otimes Q_M \otimes I_k$ and postmultiplication by $Q_N' \otimes Q_M' \otimes I_k$ transforms (11.8) into

$$(N M) (e_{N1} e'_{N1}) \otimes (e_{M1} e'_{M1}) \otimes \gamma_{xx00}(h) + (N) I_N \otimes (e_{M1} e'_{M1}) \otimes \gamma_{xx11}(h) + I_N \otimes I_M \otimes \gamma_{xx22}(h).$$

(11.9)
Results analogous to (11.7)-(11.9) hold for the covariance structure of \( \varepsilon_{t..} \). Define

\[
\tilde{z}_{t..} = (\tilde{z}_{t1..}, \tilde{z}_{t2..}, \ldots, \tilde{z}_{tN..})' = (Q_N \otimes Q_M \otimes I_p)z_{t..};
\]

and define \( \tilde{z}_{..i..} \), \( \tilde{z}_{..t..} \), and \( \tilde{z}_{..tij} \) accordingly. Similarly, let

\[
\tilde{z}_{t..} = (Q_N \otimes Q_M \otimes I_k)x_{t..} \quad \text{and} \quad \tilde{e}_{t..} = (Q_N \otimes Q_M \otimes I_p)e_{t..} ;
\]

and define \( \tilde{x}_{..i..}, \tilde{x}_{..t..}, \tilde{e}_{..i..}, \tilde{e}_{..t..} \), and \( \tilde{e}_{..tij} \) accordingly. It follows from (11.9) that

\[
\text{Cov}(\tilde{z}_{t..}, \tilde{z}_{t+h..})
\]

\[
= (Q_N \otimes Q_M \otimes I_p) \text{Cov}(z_{t..}, z_{t+h..})(Q_N \otimes Q_M \otimes I_p)
\]

\[
= (Q_N \otimes Q_M \otimes I_p)[(I_N \otimes I_M \otimes (\beta, I_k))' \text{Cov}(x_{t..}, x_{t+h..})]
\]

\[
\times [I_N \otimes I_M \otimes (\beta, I_k)] + \text{Cov}(e_{t..}, e_{t+h..})(Q_N \otimes Q_M \otimes I_p)
\]

\[
= [Q_N \otimes Q_M \otimes (\beta, I_k)'] [(I_{11} N) @ (I_{11} M) @ I_{00} (h)]
\]

\[
+ I_N @ (I_{11} M) @ I_{00} (h) + I_N @ I_M @ I_{22} (h) [Q_N @ Q_M @ (\beta, I_k)]
\]

\[
+ [Q_N @ Q_M @ I_p] [(I_{11} N) @ (I_{11} M) @ I_{00} (h)]
\]

\[
+ I_N @ (I_{11} M) @ I_{00} (h) + I_N @ I_M @ I_{22} (h) [Q_N @ Q_M @ I_p]
\]
Thus, the transformation from \( Z \) to \( \tilde{Z} \) has yielded \( N \times M \) uncorrelated \( p \times 1 \) vectors \( \tilde{Z}_{ij} \), \( i=1, 2, \ldots, N \), \( j=1, 2, \ldots, M \). The covariance structure of \( \tilde{Z}_{11} \) is described by

\[
\text{Cov}(\tilde{Z}_{t11}, \tilde{Z}_{t+h,11}) = N M \{(g, I_k)' \Gamma_{xx00}(h)(g, I_k) + \Gamma_{ee00}(h)\}
\]

\[
+ M\{(g, I_k)' \Gamma_{xx11}(h)(g, I_k) + \Gamma_{ee11}(h)\}
\]

\[
+ \{(g, I_k)' \Gamma_{xx22}(h)(g, I_k) + \Gamma_{ee22}(h)\}.
\]

For \( i \neq 1 \), the covariance structure of \( \tilde{Z}_{i1} \) is described by
The remaining \((N - 1)M\) vectors \(\tilde{z}_{ij}, i \neq 1\), have covariance structure described by

\[
\text{Cov}(\tilde{z}_{tij}, \tilde{z}_{t+h,ij}) = (\mathbf{g}_{k}, \mathbf{I}_{k})' \Gamma^{-xx_{22}}(h)(\mathbf{g}_{k}, \mathbf{I}_{k}) + \Gamma^{-\epsilon_{22}}(h).
\]

The derivation of innovation sequences and the associated Newton-Raphson maximum likelihood estimation procedure then follows by arguments similar to those given for the simpler nested-term model of Section 11.2.

In closing this section, note that the methods developed above extend directly to cross sectional data with three or more levels of sampling, provided that within each level, all group sizes are equal. For example, given a third level of sampling for which the group size was \(L\), the observation vector \(Z_{t\ldots}\) would be transformed to

\[
\tilde{Z}_{t\ldots} = (Q_N \otimes Q_M \otimes Q_L \otimes \mathbf{I}_p)Z_{t\ldots}
\]

Given assumptions on \(x_{3tij}\) and \(\xi_{3tij}\) analogous to those given following (11.6),
\[
\text{Cov}(\tilde{Z}_t, \ldots, \tilde{Z}_{t+h}, \ldots) = (\text{NML})\{(e_{11}e_{11}'), (e_{11}e_{11}'), (e_{11}e_{11}') \}
\]

\[
+ \text{ML}(e_{11}e_{11}') \sim \{ (e_{11}e_{11}')' \sim (\Gamma_{xx00}(h)) (e_{11}e_{11}') + \Gamma_{\varepsilon \varepsilon 00}(h) \}
\]

\[
+ \text{L}(e_{11}e_{11}') \sim \{ (e_{11}e_{11}')' \sim (\Gamma_{xx11}(h)) (e_{11}e_{11}') + \Gamma_{\varepsilon \varepsilon 11}(h) \}
\]

\[
+ \text{L}(e_{11}e_{11}') \sim \{ (e_{11}e_{11}')' \sim (\Gamma_{xx22}(h)) (e_{11}e_{11}') + \Gamma_{\varepsilon \varepsilon 22}(h) \}
\]

\[
+ \text{L}(e_{11}e_{11}') \sim \{ (e_{11}e_{11}')' \sim (\Gamma_{xx33}(h)) (e_{11}e_{11}') + \Gamma_{\varepsilon \varepsilon 33}(h) \}.
\]

The innovation sequences of the NML uncorrelated $p \times 1$ vectors $\tilde{Z}_{i1}$ may then be obtained by methods similar to those used for the two-stage vectors $\tilde{Z}_{i1}$ above.

Finally, note that the autoregressive moving average assumptions made on the input and error components were used only in the derivation of the innovation sequences of the $\tilde{Z}$ vectors. The covariance structures (11.4), (11.10), and (11.11) of the $\tilde{Z}$ vectors relied only on the assumptions of covariance stationarity and uncorrelatedness of the cross-sectional components given above. Therefore, the Helmert transformations proposed above can be used to decompose the covariance structure of combined time series and cross-sectional data for a wide class of balanced hierarchical time series component models.
11.4. Crossed Inputs and Errors

The three sections above addressed hierarchical time series models for the input \( x \) and error \( e \) of a given unit, and discussed the use of the Helmert transformation \( Q_N \) to decompose the resulting observations into uncorrelated vectors of length \( pT \). Use of the Helmert transformation gives similar results for models with crossed inputs and errors. For example, let model (11.5) hold, but replace model (11.6) with the model,

\[
x_{tij} = x_{0t} + x_{1ti} + x_{2j},
\]

\[
e_{tij} = e_{0t} + e_{1ti} + e_{2j}
\]

\( t=1, 2, \ldots, T; i=1, 2, \ldots, N; j=1, 2, \ldots, M; \)

where \( x_{0t}, x_{1ti}, e_{0t}, e_{1ti} \) follow the models indicated below (11.2); \( x_{2j} \) are independent \((0, \Sigma_{xx22})\) random vectors; \( e_{2j} \) are independent \((0, \Sigma_{ee22})\) random vectors; and \( x_{0t}, x_{1ti}, x_{2j}, e_{0s}, e_{1s}, e_{2m} \) are independent for all \( t, t', s, s', i, j, \ell \) and \( m \). Retain the remaining notation as given in the preceding sections, but define \( x_{2.} = (x_{21}', x_{22}', \ldots, x_{2M}') \) and define \( e_{2.} \) similarly.

Then

\[
x_{t.} = (1_N \ a \ I_M \ a \ I_k)x_{0t}' + (1_N \ a \ I_M \ a \ I_k)x_{1t}' + (1_N \ a \ I_M \ a \ I_k)x_{2.}'
\]
and

\[ \text{Cov}(\mathbf{x}'_{t..}, \mathbf{x}'_{t+h..}) \]

\[ = (I_N \ast I_M \ast I_k, \Gamma_{x00}^{-1}(h))(I_N \ast I_M \ast I_k) \]

\[ + (I_N \ast I_M \ast I_k)[I_N \ast \Gamma_{xx11}^{-1}(h)](I_N \ast I_M \ast I_k) \]

\[ + (I_N \ast I_M \ast I_k)[I_M \ast \Sigma_{xx22}^{-1}](I_N \ast I_M \ast I_k) \]

\[ = (I_N \ast I_N)[I_M \ast I_M] \ast \Gamma_{x00}^{-1}(h) \]

\[ + I_N \ast (I_M \ast I_M) \ast \Gamma_{xx11}^{-1}(h) + (I_N \ast I_N) \ast I_M \ast \Sigma_{xx22}^{-1} \cdot \]

A similar result holds for Cov(\mathbf{z}_{t..}, \mathbf{z}_{t+h..}) Define

\[ \tilde{z}_{t..}' = (Q_N \ast Q_M \ast I_p)z_{t..}' \]

Then

\[ \text{Cov}(\tilde{z}_{t..}', \tilde{z}_{t+h..}') \]

\[ = (Q_N \ast Q_M \ast I_p)[(I_N \ast I_M \ast (\mathbf{\beta}, I_k)' \text{Cov}(\mathbf{x}_{t..}', \mathbf{x}'_{t+h..})] \]

\[ \times [I_N \ast I_M \ast (\mathbf{\beta}, I_k)] + \text{Cov}(\mathbf{z}_{t..}', \mathbf{z}'_{t+h..})(Q_N \ast Q_M \ast I_p) \]
= [Q_N \otimes Q_M \otimes (\beta, I_k)'] \{(I_{1N} I_{1M}^T) \otimes (I_{1N} I_{1M}^T) \otimes \Gamma_{xx00}(h) \\
+ I_N \otimes (I_{1M} I_{1M}^T) \otimes \Gamma_{xx11}(h) + (I_{1N} I_{1M}^T) \otimes I_M \otimes \Sigma_{xx22} \}
\times [Q_N \otimes Q_M \otimes (\beta, I_k)] \\
+ [Q_N \otimes Q_M \otimes I_p] \{(I_{1N} I_{1N}^T) \otimes (I_{1N} I_{1N}^T) \otimes \Gamma_{xx00}(h) \\
+ I_N \otimes (I_{1N} I_{1N}^T) \otimes \Gamma_{xx11}(h) + (I_{1N} I_{1N}^T) \otimes I_M \otimes \Sigma_{xx22} \}[Q_N \otimes Q_M \otimes I_p] \\
= (NM)\{(e^T_{N1} e^T_{M1}) \otimes (e^T_{M1} e^T_{M1}) \otimes [(\beta, I_k)' \Gamma_{xx00}(h)(\beta, I_k) + \Gamma_{\epsilon \epsilon 00}(h)]\} \\
+ (M)\{I_{N} \otimes (e^T_{M1} e^T_{M1}) \otimes [(\beta, I_k)' \Gamma_{xx11}(h)(\beta, I_k) + \Gamma_{\epsilon \epsilon 11}(h)]\} \\
+ (N)\{(e^T_{N1} e^T_{N1}) \otimes I_M \otimes [(\beta, I_k)' \Sigma_{xx22}(\beta, I_k) + \Sigma_{\epsilon \epsilon 22}]\}. \\

Therefore, \( \tilde{Z} \) contains \( N + M - 1 \) uncorrelated \( pT \times 1 \) vectors \( \tilde{Z}_{i,11} \). The covariance structure of \( \tilde{Z}_{i,11} \) is described by \\
\[ \text{Cov}(\tilde{Z}_{t11}, \tilde{Z}_{t+h,11}) = (NM)\{(\beta, I_k)' \Gamma_{xx00}(h)(\beta, I_k) + \Gamma_{\epsilon \epsilon 00}(h)\} \\
+ (M)\{(\beta, I_k)' \Gamma_{xx11}(h)(\beta, I_k) + \Gamma_{\epsilon \epsilon 11}(h)\} \\
+ (N)\{(\beta, I_k)' \Sigma_{xx22}(\beta, I_k) + \Sigma_{\epsilon \epsilon 22}\}. \]
The covariance structure of the $M - 1$ vectors $\tilde{z}_{ij_1}$, $j \neq 1$, is described by

$$\text{Cov}(\tilde{z}_{tij_1}, \tilde{z}_{t+h,ij_1}) = N[\{\beta, I_k\}'\Sigma_{xx22}(\beta, I_k) + \Sigma_{ee22}].$$

The covariance structure of the $N - 1$ vectors $\tilde{z}_{ii_1}$, $i \neq 1$, is described by

$$\text{Cov}(\tilde{z}_{ti_1}, \tilde{z}_{t+h,i_1}) = (N)[\{\beta, I_k\}'\Gamma_{xxi1}(h)(\beta, I_k) + \Gamma_{eei1}(h)].$$

Again through the use of the Helmert matrix one obtains results that parallel closely results in the analysis of variance, in which, for example, a two-way classificatory model has one degree of freedom for a grand mean, $M - 1$ degrees of freedom for row-wise contrasts, $N - 1$ degrees of freedom for column-wise contrasts, and $(M - 1)(N - 1)$ degrees of freedom for "error". Note that model (11.12) has no "residual error" term for either $\mathbf{x}$ or $\boldsymbol{\xi}$; thus, the Helmert transformation applied to $\tilde{z}_{t..}$ for model (11.5)-(11.12) permits a "data reduction" from $NM$ to $N + M - 1$ vectors of length $pT$.

The derivation of innovation sequences and the associated Newton-Raphson maximum likelihood estimation procedure then follows a pattern similar to that given for nested-term models in the preceding sections.

Similar results may be obtained for other combinations of hierarchical and crossed-term models for $\mathbf{x}$ and $\boldsymbol{\xi}$. In each case, the salient point is that for balanced data structures, one may transform
the resulting observations $Z$ to a set of independent $pT \times 1$ vectors such that each vector follows a (possibly different) stationary time series, provided that the original components were stationary. Thus, given a balanced data structure, the measurement error model for combined time series and cross-sectional data may be reduced to a measurement error model for several independent time series following different models.