Nonlinear measurement error analysis for system monitoring

Jean Elizabeth Pelkey
Iowa State University

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

Recommended Citation
Pelkey, Jean Elizabeth, "Nonlinear measurement error analysis for system monitoring " (1997). Retrospective Theses and Dissertations. 11737.
https://lib.dr.iastate.edu/rtd/11737
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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 copyright 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 original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 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.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
Nonlinear measurement error analysis for system monitoring

by

Jean Elizabeth Pelkey

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics
Major Professors: Yasuo Amemiya and Derrick Rollins

Iowa State University
Ames, Iowa
1997

Copyright © Jean Elizabeth Pelkey, 1997. All rights reserved.
This is to certify that the doctoral dissertation of

Jean Elizabeth Pelkey

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College
## TABLE OF CONTENTS

1 GENERAL INTRODUCTION ................................................. 1
   Discussion of Models for Processes in Steady State .......... 3
   Discussion of Models for Processes in Non-Steady State ...... 5
   Literature Review .................................................. 9
      Linearly constrained models steady state processes ...... 10
      Linearly constrained models for non-steady state processes 13
      Nonlinearly constrained models for steady state processes 14
      Nonlinearly constrained models for non-steady state processes 15
   Dissertation Organization ........................................ 16
   Bibliography ...................................................... 16

2 ERRORS-IN-VARIABLES ANALYSIS OF NONLINEAR SYSTEMS 20
   Abstract .......................................................... 20
   Introduction ...................................................... 20
   Estimation Problem .............................................. 24
   Testing Problem ................................................. 31
   Simulation Study ................................................ 38
   Bibliography .................................................... 41
   Appendix ......................................................... 43

3 NONLINEAR ERRORS-IN-VARIABLES ANALYSIS WITH MEASUREMENT ERROR BIASES 48
Abstract .................................................. 48
Introduction ............................................ 49
Estimability ............................................. 51
Estimation Problem .................................... 54
Testing For System Anomaly ......................... 63
Simulation Study ....................................... 64
Bibliography ........................................... 67
Appendix .................................................. 69

4 GENERAL CONCLUSIONS ............................... 107
1 GENERAL INTRODUCTION

In the chemical process industry, the performance of a process (i.e. the quality of the end product) depends on the process control engineer's ability to accurately measure and adjust the process. For example, if a process engineer has no control over the concentration of the ingredients going into a reactor, or the temperature at which the reaction within the reactor takes place, then controlling the reaction and, hence, the quality of the end product is difficult. In order to control the process, measurements are taken at points within the process and are used to estimate the true state of the process. Examples of measurements taken within the process include the temperature of a stream, the flow rate of a stream, the pressure of a stream or the concentration of a chemical specie in a stream.

Chemical engineers classify measurement errors into two types: random and systematic measurement errors. A systematic error represents either an instrument bias or some system anomaly. An instrument bias occurs when an instrument consistently reports measurements that are either higher or lower than the true value. This type of error differs from random measurement errors in that it is a consistent measurement bias that occurs, e.g., when the instrument is out of calibration. Once the instrument is out of calibration, it will remain out of calibration until it is recalibrated. A system anomaly occurs, e.g., if a material leak exists somewhere in the process. The loss of process materials due to a leak is costly in terms of unrecoverable revenues, environmental pollution and poor quality of the product. A random measurement error is the measurement variability due to various factors.
The measurements collected from the chemical process are associated with a fixed but unknown component corresponding to the true but unknown value of the process variable. As an example, suppose we take measurements of the flow rate of a stream. The true flow rate is an unknown value or unknown parameter for which the engineer would like to obtain a reliable estimate even when systematic errors such as instrument biases are present. Once the engineer has a reliable estimate of the true flow rate of the stream, the flow rate can be adjusted so that it is operating at the desired level. The true value of the process variable may change over time, as with the non-steady state process, or it may remain constant over time, as with a process assumed to be in steady state. If the true value of the process variables are oscillating around some fixed point, the process is said to be in pseudo-steady state. Steady state and pseudo-steady state are closely related and methods to address steady state problems can often be applied to pseudo-steady state problems with minor adjustments. Thus, this thesis will loosely use the term “steady state” to mean either steady state or pseudo-steady state. Practical problems in system monitoring include detection and correction of systematic errors, and estimation of the true values of the process variables (i.e. the process parameters).

In the chemical engineering literature, methods of detecting and identifying process leaks and instrument biases are called gross error detection (GED). In addition to detecting systematic biases and process leaks, GED methods seek to accurately estimate the unknown parameters when process leaks and instrument biases are present. Methods of estimating the true value of the process variables when only random measurement errors are assumed to exist are referred to as data reconciliation (DR).

The purpose of the work described in this thesis is to formulate the engineering system monitoring problem as a statistical problem, and to develop statistical procedures appropriate for problem solving. In particular, processes having nonlinear physical constraints are considered. The physical constraints are energy and material balances derived from engineering knowledge of the system. The physical constraints include, for example, the
chemical reactions. Procedures will be developed for systems in steady state as well as those operating in the non-steady state.

**Discussion of Models for Processes in Steady State**

Process measurements collected are subject to random and systematic errors. For the steady state process, the model describing the measurements can be expressed as

\[ Y = \mu + \delta + \epsilon, \]  

where \( Y \), a \( p \times 1 \) vector with components \( Y_i, \) \( i = 1, 2, \ldots, p \), contains measurements corresponding to a \( p \times 1 \) unobservable true process value \( \mu \), \( \delta \) is a \( p \times 1 \) vector of systematic biases, and \( \epsilon \) represents a random error term. Typically, \( Y_i \) is an average of \( n \) measurements taken over a short span of time or some representative value of the data after a time trend has been removed. Thus, a condition used later, that \( \text{Var}(\epsilon) = \Psi = O\left(\frac{1}{n}\right) \) for some \( n \), often makes sense in practice. Also, it is often reasonable to assume that an accurate estimate of \( \Psi \) is available, or that \( \Psi \) is known.

The physical constraints in the system correspond to a set of known restrictions that \( \mu \) must satisfy. A set of linear homogeneous constraints which can represent total mass balances, for example, can be expressed as

\[ A\mu = 0, \]  

where \( A \) is a full row rank \( r \times p \) matrix of known constants representing \( r \) linear restrictions. A more general nonlinear restriction for representing any type of mass or energy balance can be written as

\[ f(\mu) = 0, \]  

where \( f(\mu) \) is an \( r \)-valued known function with no redundancy or inconsistency. Each of the \( r(< p) \) elements of \( f \) can be a nonlinear function of the \( p \) true values in \( \mu \).
very special case of the nonlinear restriction is a bilinear restriction, where $f(\mu)$ includes linear combinations of cross products of the elements in $\mu$. Bilinear restrictions often appear in the context of component mass balances.

Model (1.1) with known restrictions (1.2) or (1.3), and with known $\Psi$ is not generally identified in the sense that not all elements of $\mu$ and $\delta$ are estimable. Note that all $p$ elements of $\mu$ are unknown and to be estimated. On the other hand, some of the elements of $\delta$ are known to be zero (i.e. insignificant), if the corresponding instrument is known to have no bias. For the linear model (1.1) - (1.2), it follows readily from linear model theory that a linear function of $\delta$ is estimable (and testable) if and only if it is a linear function of the elements of $A\delta$, and that all elements of $\mu$ are estimable if and only if the measurement biases are such that $L\delta = 0$, where the rows of $L$ span the orthogonal complement of the row space of $A$. For the general restriction (1.3), a general estimability condition cannot be written down without specifying the form of $f(\mu)$. However, the fact that not all elements of $\delta$ are estimable and that the estimability of $\mu$ depends on the structure of $\delta$ still holds for the nonlinear model.

The system leak problem corresponds to a violation of some of the $r$ constraints, because a material or energy balance physical equality must hold without some system anomaly. A restriction suspected for a leak is not used in estimation. Development of checking or testing procedures for a restriction or for the existence of a leak is desired.

In some applications, restrictions contain some unknown parameters. For example, a set of linear restrictions may be of the form (Rollins and Davis [20])

$$Y = \mu + \delta + \epsilon,$$

\[ A_1\mu + A_2\theta = 0, \]  

where $A_1$ is a full row rank $r \times p$ known matrix, and $A_2$ is a full column rank $r \times k$ known matrix, and $\theta$ is a $k \times 1$ vector of unknown parameters. Then, there is an $r \times r$
matrix $Q$ such that

$$QA_2 = \begin{pmatrix} A_{21}^* \\ O \end{pmatrix},$$

(1.5)

where $A_{21}^*$ is $k \times k$ nonsingular. Let

$$A_1^* = \begin{pmatrix} A_{11}^* \\ A_{12}^* \end{pmatrix} = QA_1,$$

where $A_{11}^*$ is $k \times p$ and $A_{12}^*$ is $(r - k) \times p$. Then, the restriction (1.4) is equivalent to

$$A_{12}^* \mu = 0,$$

$$A_{11}^* \mu = \theta^*$$

where $\theta^* = A_{21}^* \theta$ is unknown. Hence, this restriction is equivalent to having only $r - k$ known restrictions. Thus, $\theta$ is non-estimable if $A_{11}^* \mu$ is not estimable.

The identification and leak problems discussed above force us to consider only a certain practical special case of model (1.1) with restriction (1.2) or (1.3). In the first paper of this dissertation, it is assumed that $\delta = 0$, i.e., the instruments are properly calibrated. Under this assumption, the estimation of $\mu$ and the leak detection problem are discussed for the nonlinear model. In practice, the developed procedures can be useful for the cases with nonzero $\delta$, depending on the configuration of known zero biases and the restrictions.

**Discussion of Models for Processes in Non-Steady State**

In a non-steady process, the true values are changing over time, and the corresponding measurements cannot be combined easily. Let $\mu_t$ and $Y_t$ be the true value and the corresponding measurement of the process at time $t$. The measurement model becomes

$$Y_t = \mu_t + \delta + \epsilon_t, \ t = 1, 2, \ldots, T,$$

(1.6)
where \( \delta \) is the systematic instrument bias free of \( t \), and \( \epsilon_t \) is the random error with zero mean. The linear physical balance restriction is

\[
A \mu_t = 0, \quad t = 1, 2, \ldots, T,
\]

(1.7)

where \( A \) is a full row rank \( r \times p \) known matrix. Then, by the linear model theory, even with \( T \) observations, only a linear function of \( A \delta \) is an estimable function of \( \delta \), and the entire \( \delta \) is not, in general, estimable. Also, all elements of \( \mu_t \) cannot be estimable unless \( L \delta = 0 \) with \( L \) being a \((p - r) \times p\) orthogonal complement of \( A \). Thus, for the linear restrictions, the identification problem remains, even with \( T \) observations and a non-steady state process.

With \( T \) observations and true values, estimation of unknown parameters in the restrictions makes more sense than the steady state case. Suppose model (1.6) holds with

\[
A_1 \mu_t + A_2 \theta = 0,
\]

(1.8)

where \( A_1 \) is \( r \times p \) of rank \( r \), \( A_2 \) is \( r \times k \) of rank \( k \), \( A_1 \) and \( A_2 \) are known and \( \theta \) is a \( k \times 1 \) vector of unknown parameters. To address the estimability issues for this model, let \( R \) be an \( p \times p \) nonsingular matrix such that

\[
R = \begin{pmatrix}
A_1 \\
L_1
\end{pmatrix},
\]

and consider a transformed model under (1.7)

\[
\begin{pmatrix}
Y_{1t}^* \\
Y_{2t}^*
\end{pmatrix} = RY_t.
\]

Then,

\[
Y_{1t}^* = A_1 \mu_t + A_1 \delta + \epsilon_{1t}^*,
\]

(1.9)

\[
Y_{2t}^* = L_1 \mu_t + \delta_t^* + \epsilon_{2t}^*,
\]

(1.10)
where $\delta^*_1 = A_1 \delta$ and $\delta^*_2 = L_1 \delta$. Thus, the identification of $\theta$ and $\delta^*_1$ needs to be made only through $Y^*_1$. For all $t = 1, 2, \ldots, T$, we write

$$Y^*_1 = A^* \begin{pmatrix} \theta \\ \delta \end{pmatrix} + \epsilon^*,$$

(1.11)

where $Y^*_1 = (Y^*_{11}, Y^*_{12}, \ldots, Y^*_{1T})'$, and

$$A^* = \begin{bmatrix} -A_2 & A_1 \\ -A_2 & A_1 \\ \vdots & \vdots \\ -A_2 & A_1 \end{bmatrix}.$$  

(1.12)

Hence, the only estimable functions of $\delta$ is a linear combination of the elements of $L_2 A_1 \delta$, where the rows of the $(r-k) \times r$ $L_2$ span the orthogonal complement of $A_2$. In general, $\mu_t$, $\theta$ and $\delta$ are not estimable.

From (1.12), we see that the non-identifiability of $\theta$ and $\delta$ are closely related to the fact that $E(Y^*_1)$ is free of $t$. In particular, the rank of $Tr \times (p+k)$ $A^*$ in (1.12) is $r$ which is less than $p$. Suppose that the physical restriction depends on $t$ and is given by $A_{1t} \mu_t + A_{2t} \theta = 0$. Then, (1.12) becomes

$$A^* = \begin{bmatrix} -A_{21} & A_{11} \\ -A_{22} & A_{12} \\ \vdots & \vdots \\ -A_{2T} & A_{1T} \end{bmatrix},$$

(1.13)

so that $(\theta', \delta')'$ would be estimable if $A^*$ has full column rank, i.e., if $A_{1t}$ and $A_{2t}$ vary sufficiently over $t$. This fact can be used to describe intuitively why a model with sufficiently nonlinear restrictions yield the estimability of all parameters.

The model for the nonlinear process in non-steady state, assuming no leaks exist, is given by

$$Y_t = \mu_t + \delta + \epsilon_t,$$

(1.14)
where \( f(\mu_t, \theta) \) is an \( r \times 1 \) vector of nonlinear constraints, and \( \theta \) is a \( k \times 1 \) vector of unknown parameters. A precise set of conditions required for the estimability of all parameters is given in the second paper of this dissertation. Heuristically, we consider a linearization of the restrictions in (1.14). Let

\[
F_{\mu t} = \left. \frac{\partial f(\mu, \theta)}{\partial \mu} \right|_{(\mu, \theta)},
\]

\[
F_{\theta t} = \left. \frac{\partial f(\mu, \theta)}{\partial \theta} \right|_{(\mu, \theta)}.
\]

Then, a linearized restriction is approximately,

\[
F_{\mu t} (\mu_t - \mu_t^0) + F_{\theta t} (\theta - \theta^0) = 0,
\]

where \( \mu_t^0 \) and \( \theta^0 \) are true values. This leads to the \( A^* \) matrix of (1.13) with

\[
A_{1t} = F_{\mu t},
\]

\[
A_{2t} = F_{\theta t}.
\]

For nonlinear \( f(\mu_t, \theta) \), \( F_{\mu t} \) varies over \( t \), and we expect all elements of \( \theta \) and \( \delta \) to be identified. Then, from an equation corresponding to (1.10), all \( \mu_t \) are estimable. That is, with sufficiently nonlinear restrictions, all parameters, including the systematic biases for all \( p \) instruments and the restriction parameter \( \theta \), can be estimated. The full column rank condition for \( A^* \) with (1.15) as elements can often be verified in practice.

The condition that \( A^* \) be full column rank means that the columns of \( A^* \) are linearly independent. Two columns of \( A^* \) are linearly dependent, for example, if two or more columns are multiples of each other. This would occur if two elements of \( \mu_t \), \( \mu_{it} \) and \( \mu_{jt} \) enter the same subset of restrictions linearly. Or, this would occur if, for some \( k \), \( \mu_{it} \) and \( \theta_k \) enter the same set of restrictions linearly. A related issue concerning the nonlinear case is that of how to express the restrictions. Since \( f(\mu_t, \theta) \) has a general
form, one element of $\theta$ may even be the leak for a particular restriction. That is, if the exact $i^{th}$ restriction is $g(\mu_t, \theta_0) = 0$, then we could write the $i^{th}$ restriction to be fitted as $f_i(\mu_t, \theta) = g_i(\mu_t, \theta_0) - \nu_i = 0$, and include $\theta_0$ and $\nu_i$ in $\theta$. Hence, the leak could be included in the general formulation of the nonlinear model. However, $\nu_i$ appears linearly in the restrictions.

**Literature Review**

This section mostly reviews the chemical engineering literature which has limited technical relevance to the statistical problem addressed in the two papers constituting this thesis.

Much of the previous work in the area of gross error detection (GED) and data reconciliation (DR) in the chemical engineering literature has concentrated on finding solutions to the steady state model described by (1.1) - (1.2) (e.g., [5], [6], [8], [9], [13], [14], [15], [19], [21], [20], [25], [26], [27], [29]), since it is the simplest case and imposes the fewest computational problems. Applications of the models in the chemical process industry are limited because most chemical reactions are described by nonlinear equations. As computers become faster and more efficient, more complex problems are being solved. Researchers have begun to look at GED and DR for non-steady state systems with linear constraints (e.g., Almasy [1], Darouach and Zasadzinski [7], Narasimhan and Mah [16], Rollins and Devanathan [22]), as well as to consider the problem of bilinear constraints for processes in steady state (e.g., Crowe [4], Kuiper et al. [23], Rollins and Roelfs [24], Tamhane and Mah [29]). Limited work has been done to solve the problem of nonlinear processes in both steady state (e.g., Kim et al. [10], Pai and Fisher [17], Tjoa and Biegler [30] ) and non-steady state (e.g., Albuquerque and Biegler [2], Kim et al. [11], Liebman et al. [12], Ramamurthi and Bequette [18]).

Very little work has been done to consider the statistical properties of the estimates
of the unknown parameters in the nonlinear model. In fact, most research conducted in GED and DR has considered some variation of the model (1.1) - (1.2) as will be seen in the review given below. We begin by discussing previous research for both the steady and non-steady state models with linear restrictions and then present work done for the same two models with nonlinear restrictions.

**Linearly constrained models steady state processes**

Mah and Tamhane [14] developed the Measurement Test (MT) to test for gross errors in a chemical process operating in steady state. They consider gross errors “caused by non-random events such as instrument biases, malfunctioning measuring devices, [and] incomplete or inaccurate process models.” The work, however, considers only instrument biases and malfunctioning measuring devices. The test targets statistical outliers in the process data by comparing the maximum standardized residual to a percentile of the normal distribution chosen to control the Type I error rate. The disadvantages of this test are: (1) it makes multiple comparisons so the exact Type I error rate is not easily calculated; and (2) the test has low power [8]. Rollins [19] has shown that the MT does not maintain an overall level of significance, \( \alpha \), when multiple biases and leaks are present in the system. Iordache et al. [9] studied the performance of the MT when one instrument bias is present. Their study indicates the power of the test increases as the ratio of the size of the gross error to the standard deviation of the measurement error increases.

Narasimhan and Mah [15] proposed the Generalized Likelihood Ratio (GLR) method which also looks at the residual vectors for the restricted model (1.1) - (1.2). This method looks to improve GED over the MT by conducting a single size \( \alpha \) test for gross errors using a \( \chi^2 \) test. A second improvement is the GLR attempts to identify whether the source of the gross error is due to a process leak, a bias in the measurement device, or some other cause. The authors accomplish this by incorporating into the model the
effect of a measurement bias and process leak. The (1.1) - (1.2) is transformed from a
restricted model to an unrestricted model using the transformation \( r = Ay \). If a leak
is present in the process, the restrictions imposed on the model will not be met. The
restrictions when a leak is present are \( A\mu = \gamma \) where \( \gamma_i = 0 \) if there is no leak in the
\( i^{th} \) restriction and \( \gamma_i \neq 0 \) if the leak is in the \( i^{th} \) restriction. If there are no leaks in
the process, then \( A\mu = 0 \). Assuming \( \epsilon \) is normally distributed with mean zero and
covariance matrix \( \Psi \), the transformed vector \( r \) is normally distributed with mean vector
\( A\mu + A\delta \) and covariance matrix \( A\Psi A' \). The method then uses a \( \chi^2 \) test for \( H_0 : \mu_r = 0 \).
The authors show that the GLR test is equivalent to the MT when there are no process
leaks. Thus, as shown in [19], the GLR method will have high Type I error rate.

If the hypothesis is rejected Narasimhan and Mah [15] provide a method for es­
timating the process leak and measurement biases. This method is called the serial
compensation strategy (SCS). The method involves a sequence of GLR tests to identify
all biases and leaks. If the SCS method concludes that a measurement is biased or that
a leak exists in the process, it estimates the bias or leak and adjusts the corresponding
process variable so the physical constraints are met. Rollins and Davis [20] pointed out
the drawbacks of the SCS method. First, the identification process is subject to large
Type I errors when multiple biases or leaks exist. Second, the SCS-adjusted estimates of
the process variables can have larger errors than the original estimates because of errors
induced by previously estimated variables. Finally, the distribution of the estimates of
the process variables has not been determined.

Rollins and Davis [20] developed an unbiased estimation technique (UBET) to look
for biased measurements and process leaks when the covariance matrix is known. The
method uses the transformation above. The method proposes a series of tests to deter­
mine whether there are biases or leaks in the system and the location of the problem.
The first component of the method is a global test (GT). The GT is performed using
the residual vector of the transformed model to form a \( \chi^2 \) test statistic.
The second step of the UBET is to locate the errors by testing the hypothesis $H_0 : l^T \mu_r = 0$ vs $H_a : l^T \mu_r \neq 0$, where $l$ is a vector of ones and zeros representing the balance equation at a node. Such a hypothesis test is conducted at each node in the process. Rejecting $H_0$ at a particular node is considered evidence of a possible instrument bias or process leak associated with that portion of the system represented by the test. Once the location of the process leaks and biases are identified, maximum likelihood estimation techniques are used to estimate the leak or bias and these estimates are used to adjust the values of the process variables so that the constraint equations are met.

A second article by Rollins and Davis [21] extended the work for the UBET method discussed above with the additional condition of the covariance matrix, $\Psi$, is unknown.

Serth and Heenan [26] proposed the modified iterative measurement test (MIMT) to identify a single potential gross error. The method consists of the following steps: (1) calculate the least squares estimates for model (1.1) - (1.2); (2) use the criteria proposed in the MT for identifying a single suspect measurement; (3) remove the suspect measurement from the data set and re-estimate the least squares estimates for the model. This procedure is continued until the final iteration indicates no suspicious measurements. The estimates of $\mu$ are the estimates computed during the final iteration.

Crowe et al. [6] extended the work done by Mah and Tamhane [14] to include the estimation and testing of unmeasured process parameters. Unmeasured process parameters are parameters in the restrictions imposed on the model for which measurements are not taken. Crowe provided a method for reducing the dimensionality of the problem.

Rosenberg et al. [25] used the work proposed by Crowe et al [6] and proposed a simple additional rule to improve GED abilities. If the true values of the process parameters are known to fall within certain reliable bounds, then one can conclude a gross error is present if the parameter estimate obtained by methods described in [6] falls outside the boundary values.

Tamhane et al. [27] proposed a method for detecting gross errors in process. This
method uses a Bayesian approach to solving the problem. The method is restricted to finding measurement biases. The method combines a one time Bayesian GED test with additional sequential tests to find the source of the gross errors. One problem with actually applying this technique to practical applications is the specification of the prior distribution for all the unknown parameters. A second problem is the method does not adequately model instruments that fail more often as time increases. Finally, there is a question of how to update the parameters of the prior distribution when an instrument fails. If the assumption is made that the instrument failure was found as soon as it occurred, then all previous data can be used to update the parameters. If the instrument failed a long time before it was detected, then updating the parameters of the prior distribution is difficult. Results of a simulation study [28] show the procedure does better than the MT when gross errors are infrequent but the results rely heavily on accurate estimates of the parameters in the prior distribution.

**Linearly constrained models for non-steady state processes**

Limited work has been done in the area of GED and DR for chemical processes in non-steady state with linear process constraints. Narasimhan and Mah [16] extended the GLR method from the steady state condition to the non-steady state condition. One disadvantage of this method, however, it is restricted to processes that are operating around a steady state point (i.e., processes in pseudo-steady state). The method does not apply to processes that are changing from one steady state to another steady state.

A second approach to estimating process parameters for a dynamic process is proposed by Almasy [1] and utilizes the Kalman filter approach. This approach is somewhat restrictive in that it requires specification of the error correlations between time $t$ and $t+1$. Results are dependent upon correctly identifying these correlations structures.

A third approach to linear dynamic systems is presented by Darouach and Zasadzinski [7]. The method provides a recursive algorithm for online estimation of a generalized
least squares problem when the process is in transition from one steady state to another but is computationally intensive. This method is novel because it can be implemented online by the chemical processing industry. The authors do not address the topic of GED.

Rollins and Devanathan [22] developed a method that is computationally faster than the Darouch and Zasadzinski method but provided estimators that had larger variances. The two methods provide comparable variances of the estimates for relatively small measurement variances. Rollins and Devanathan [22] developed maximum likelihood estimators for the unknown parameters in (1.6) - (1.8) without δ assuming normally distributed errors. They also showed the MLE is an unbiased estimator of $\mu_i$ as well as provided a formula for estimating the covariance matrix of $\hat{\mu}_i$.

**Nonlinearly constrained models for steady state processes**

The research problem of nonlinear constrained models has been divided into two portions in the engineering literature, that is, (1) processes that are bilinear; and (2) processes that are nonlinear but not bilinear. Most of the research published for nonlinear processes in steady state addresses bilinearly constrained models. For example, Crowe et al. [4] optimized the weighted least squares Lagrangian function. The method utilizes first-order derivatives in the recursive estimation techniques to arrive at the final estimates. Rollins and Roelfs [24] propose two methods for solving bilinear problems. The first, the two stage approach (TSA), extends the work of Rollins and Davis [20]. The second method uses a first order Taylor series expansion about the true mean $\mu$ and is called the linearization approach (LA). Results of a simulation study in [23] show that both methods do well in estimating and detecting multiple systematic biases when the true value of the process parameters are similar in size.

To consider the general nonlinear case of a process in steady state, Tamhane and Mah [29] extended the MT and global test from the case of a model with linear restrictions to
the case of a model with bilinear restrictions. Pai and Fisher [17] extended the bilinear
results obtained by Crowe et al. [4] to nonlinear constrained problems by using a first-
order Taylor series expansion around the true mean to obtain bilinear constraints. The
authors then applied Crowe's methods to obtain estimates of the process variables.

Tjoa and Biegler [30] and Kim et al. [10] suggested computational algorithms to
obtain a point estimate for the true value. Their methods differ in the computational
methods used to converge to the parameter estimates. Tjoa and Biegler use a reduced
Hessian approach to successive quadratic programming. Kim et al. used a nonlinear
programming algorithm for optimization.

Nonlinearly constrained models for non-steady state processes

Estimation of measured and unmeasured parameters for chemical processes assumed
to be in non-steady state and modeled by nonlinear constraint equations is likely the
toughest problem for researchers working in the area of GED and DR. Some researchers
have attacked the problem by linearizing the nonlinear constraints and using the tech­
niques developed for linearly constrained equations. Almasy [1] proposed a method called
'dynamic balancing', which uses only the linear constraints and ignores the nonlinear
constraints.

Research published beginning in late 1980's discuss computational issues associated
with the model fitting methods proposed by Britt and Luecke [3]. Liebman et al. [12]
and Kim et al. [11] use Nonlinear Programming (NLP) techniques to carry out the op­
timization. The authors demonstrated that the NLP method provided better estimates
than the linearization methods, especially in regions where the response surface is highly
nonlinear.
Dissertation Organization

As can be seen in the above review, the engineering literature on the model with nonlinear restrictions is limited to ad-hoc methods or computational methods for obtaining an estimate. No systematic statistical approach has been explored for either steady state or non-steady state problems. This dissertation attempts to take such an approach.

This thesis is divided into four chapters. Chapters 2 and 3 each contain a paper intended for submission to a refereed publication. Both papers deal with nonlinear restrictions. The purpose of the first paper is to provide statistical methods for DR and GED for processes in steady state or pseudo-steady state conditions. The second paper develops statistical procedures for the process operating in non-steady state. Chapter 4 gives general conclusions.

Bibliography


2 ERRORS-IN-VARIABLES ANALYSIS OF NONLINEAR SYSTEMS

A paper to be submitted to Biometrika

Jean E. Pelkey

Abstract

Observations from a large engineering system such as a chemical process are measured with error. A problem of interest is the estimation of the underlying process using the theoretical mass balances and other constraints inherent in the system. Another problem is that of detecting measurement biases and process leaks. This paper develops a method for a system with nonlinear constraints using the nonlinear errors-in-variables techniques. The nonlinearity presents a difficulty in that the usual maximum likelihood estimator has a large bias. A bias-adjusted estimator is developed and its properties are discussed. Statistical tests for checking measurement biases and process leaks are also discussed. Theoretical results using the so-called small-error asymptotics are presented. The usefulness of the methods is illustrated using a simulation study.

Introduction

In many engineering processes, some liquid, chemical, or product is processed through a complex system. To assure the quality of the product as well as the quality of the sys-
tem, the engineers must monitor and adjust the process. For this purpose, a processing system is equipped with a number of on-line data collection stations. The underlying process variables satisfy physical balances or constraints which are implied by the nature of the engineering system. Any deviation from the balance indicates the existence of some abnormality or system flaw such as a leak in the process. One difficulty in the engineering process monitoring is that the actual observations or readings from the monitoring stations are subject to measurement error due to instrumental, calibration, and human errors, and that some measurement instruments may possess systematic biases in addition to measurement variability. Another difficulty is associated with the ??? of nonlinear implicit equations, in which the physical balances or constraints are often expressed. Thus, one objective in the system monitoring can be formulated as the detection of possible process leaks and/or measurement biases based on observations subject to error and based on known constraints for error-free variables (i.e., the true values). A second objective of practical interest is to estimate the true values of the process variables subject to the physical balance constraints under the assumption of no leaks or measurement biases in the process. The first objective, sometimes called gross error detection in the chemical engineering community, can be treated as a statistical testing problem. The second objective, referred to as data reconciliation, is that of statistical estimation.

In this paper, the statistical data reconciliation and gross error detection for the so-called "steady state" or "pseudo steady state" processes with nonlinear physical constraints are discussed. For any data collection station, the true values of a steady state process are believed to be either constant over time or to follow a simple trend (at least within a short time span). For such a process, a number of observations taken at a station in a short time interval can be summarized using one value by averaging the observations, possibly after removing the trend. Often, such an averaging is automatically performed by a measurement instrument. For most processes, physical laws or material
energy balances imply that the true process values corresponding to the observed summary values satisfy known nonlinear constraints. Consider such a process in steady state that is monitored using \( p \) measurement stations. Let \( \mu_i \) denote the true process value (possibly adjusted for a trend) at the \( i^{\text{th}} \) station, and let \( Y_i \) be a summary measurement of the process at the \( i^{\text{th}} \) station. Writing \( Y = (Y_1, Y_2, \ldots, Y_p)' \) and \( \mu = (\mu_1, \mu_2, \ldots, \mu_p)' \), the model for the summary value \( Y \) monitoring a steady state process is

\[
Y = \mu + \epsilon, \tag{1}
\]

\[
f(\mu) = 0,
\]

where \( \epsilon \) is the measurement error vector, and the \( r \)-valued function \( f(\mu) \) represents the physical constraints. In general, each component of \( f \) is a nonlinear function of \( \mu \) and is expressed in an implicit functional form.

If the measurement instrument at a monitoring station has a systematic bias, then the corresponding element of \( \epsilon \) has nonzero mean. If some anomaly or leak exists in the process, some components of \( f(\mu) \) are nonzero, i.e., some constraints do not hold. It is generally impossible to distinguish between the bias and the constraint violation based only on the observation \( Y \), because the bias detection is possible only through the knowledge of the correct constraints. On the other hand, the instrument bias can often be found and corrected from other sources of information, e.g., calibration samples and instrument testing. Hence, we assume that correct instrument calibration has been performed, and that

\[
E(\epsilon) = 0. \tag{2}
\]

Imposing this assumption does not mean that our procedures are not useful unless proper instrument calibration is performed. If a particular one of the \( r \) constraints was suspected for violation, and if some systematic measurement bias is also suspected, then we would still be able to conclude that either this physical balance does not hold or the measurement instruments for values appearing in this constraint have a systematic bias.
Without additional outside information, distinguishing these two would not be possible. Thus, we develop our procedures under assumption (2), and focus on the leak detection. Since each element of $Y$ is an average or some summary value based on a number of measurements, an estimate of the variance-covariance matrix of $e$ is usually available based on the measurements. In addition, the instrument variability can be estimated or documented using a calibration experiment. Thus, we assume initially that $\Psi = \text{Var}(e)$ is known and positive definite, and later extend our results to include the variability due to estimating $\Psi$.

In the model described by (1)-(2), the leak detection problem is to test whether suspected components of $f(\mu)$ are in fact zero. The data reconciliation problem is the estimation of $\mu$ in model (1)-(2) where $f(\mu) = 0$ is the constraint known to hold. Note that the only observation is one $p \times 1$ vector of summary $Y$. However, the fact that $Y$ is some kind of average implies that the elements of $\Psi$ can be made small by increasing the number of measurements used to obtain $Y$. Also, most instruments used in practice have small measurement error variability. Hence, in deriving approximate statistical results, we will assume that $\Psi = O(1/n)$ for some index $n$ and consider asymptotics as $n \to \infty$. Such results will be applicable in practical situations with small but non-negligible measurement errors or large number of measurements at each station. See e.g., Amemiya and Fuller, [3], Stefanski and Carroll [15] and Carroll et al., [7].

The errors-in-variables problems have been discussed widely. For the traditional and linear errors-in-variables, see, e.g., Fuller [8]. The multivariate linear errors-in-variables analysis with known or estimated error covariance matrix has some similarity to our problem, and is reviewed in, e.g., Gleser [9], Anderson [4], and Amemiya and Fuller [2]. For the general nonlinear errors-in-variables problem, Carroll et al. [7] gives an overview. For the nonlinear problem with known or estimated error covariance matrix, see Wolter and Fuller [17] and Amemiya and Fuller [3]. We follow the latter work closely. In particular, Amemiya and Fuller [3] considered the non-steady state model
with one nonlinear implicit constraint involving an unknown parameter. They discussed the statistical properties of the maximum likelihood estimator (MLE) and proposed a bias adjusted estimator with smaller bias than the MLE, but did not consider the testing problem. The engineering literature on the nonlinear process monitoring has concentrated on developing and discussing different optimization routines to obtain point estimates for models (1)-(2) or the one corresponding to the non-steady state model, see, e.g., Albuquerque Biegler [1], Kim et al. [10], Kim et al. [11], Liebman et al. [12], Tjoa and Biegler [16]. Most algorithmic routines are similar to the one proposed by Britt and Luecke [6].

The work presented in this paper will consider model (1)-(2) for arbitrary \( r(< p) \), and both testing (leak detection) and estimation (data reconciliation) problems. Since the test procedures utilize estimators, the estimation problem is discussed first in the next section. The following section addresses the testing problem. Derivations of all results are discussed in the appendix.

### Estimation Problem

Consider model (1) where the nonlinear constraints \( f(\mu) = 0 \) and assumption (2) are known to hold, and a positive definite \( \Psi = \text{Var}(\epsilon) \) is available. A natural estimator of \( \mu \) in such a situation is obtained by minimizing

\[
Q(\mu) = (y - \mu)' \Psi^{-1}(y - \mu),
\]

over the parameters space (the range of possible values) \( \Gamma \) for \( \mu \), subject to

\[
f(\mu) = 0.
\]

We denote this estimator by \( \hat{\mu} \). If we assume the measurement error, \( \epsilon \), is normally distributed, then \( \hat{\mu} \) is the MLE of \( \mu \). In practice, the normality of \( \epsilon \) is often reasonable, because \( \epsilon \) is some average over a number of measurement errors. However, we consider
the use of \( \hat{\mu} \) without the assumption of normal \( \epsilon \), and assess the properties under general conditions.

It turns out that this \( \hat{\mu} \) has considerable bias in estimating \( \mu \) due to the nonlinearity of \( f(\mu) \) if the nonlinearity of \( f(\mu) \) is non-negligible. This is an important property of \( \hat{\mu} \) that is not discussed in the engineering literature. To characterize and assess the bias of \( \hat{\mu} \), we consider an asymptotic expansion of \( \hat{\mu} \) around the true value \( \mu \) using the setup \( \Psi = O(\frac{1}{n}) \). The expansion is expressed as the sum of two terms, the first is of order \( O_p(\frac{1}{\sqrt{n}}) \) and the second is of order \( O_p(\frac{1}{n}) \). The first term has no bias, but the second term represents the effect of the nonlinearity of \( f \).

**Theorem 1** Let the model (1)-(2) hold, and assume

(i) \( \text{Var}(\epsilon) = \Psi = \frac{1}{n} \Sigma \), where \( \Sigma \) is positive definite and the fourth moments of \( n^{\frac{1}{2}} \epsilon \) exist.

(ii) The partial derivatives of order three or less of \( f(\mu) \) exist and are continuous on the parameter space for \( \mu, \Gamma \), a subset of \( p \)-dimensional Euclidean space and the true value of \( \mu \) is in the interior of \( \Gamma \).

(iii) \( \Phi = F \Psi F' \) is positive definite, where \( F = \frac{\partial f(\mu)}{\partial \mu} \).

Then,

\[
\hat{\mu} - \mu = d_1 + d_2 + O_p \left( \frac{1}{n \sqrt{n}} \right),
\]

\[
d_1 = (I - \Psi F' \Phi^{-1} F) \epsilon = O_p \left( \frac{1}{\sqrt{n}} \right),
\]

\[
d_2 = -\Psi F' \Phi^{-1} c - V_{d_1} G' \Phi^{-1} F \epsilon = O_p \left( \frac{1}{n} \right),
\]

where

\[
c = \frac{1}{2} [d_1' A_1 d_1, \ldots, d_1' A_r d_1]',
\]
\[ A_i = \frac{\partial^2 f_i(\mu)}{\partial \mu \partial \mu'}, i = 1, 2, \ldots, r, \]

\[ f(\mu) = (f_1(\mu), f_2(\mu), \ldots, f_r(\mu))', \]

\[ V_{d_i} = (\Psi - \Psi F' \Phi^{-1} F \Psi) = \text{Var}(d_i), \]

\[ G = \begin{bmatrix} d_1' A_1 \\ \vdots \\ d_r' A_r \end{bmatrix}. \]

Furthermore,

\[ E(d_1) = 0, \]

\[ E(d_2) = -\Psi F' \Phi^{-1} B(\mu), \]

\[ B(\mu) = \frac{1}{2} \begin{bmatrix} \text{tr}(A_1 V_{d_1}) \\ \text{tr}(A_2 V_{d_1}) \\ \vdots \\ \text{tr}(A_r V_{d_1}) \end{bmatrix}. \]

This result implies that an approximate bias of \( \hat{\mu} \) is \( E(d_2) \), which is a function \( B(\mu) \).

Note that \( B(\mu) \) is a function of the second derivatives of \( f \), \( A_i, i = 1, 2, \ldots, r \), and \( V_{d_i} \), the variance-covariance matrix of \( d_i \). Hence, the bias vanishes only if either the second derivatives of \( f_i(\mu) \) are all zero (i.e., \( f(\mu) \) is linear in \( \mu \)) or the error variances corresponding to the components of \( \mu \) appearing nonlinearly in \( f \) are all zero. In fact, if \( f(\mu) \) is linear in \( \mu \), then \( d_2 = 0 \). Thus, the approximate bias given in Theorem 1 is considered a difficulty associated with the nonlinearity of \( f \).

Following the idea given by Amemiya and Fuller [3], it is possible to develop a bias adjusted estimator which has smaller bias than \( \hat{\mu} \). Given \( \hat{\mu} \), let \( \hat{\mu} \) be the value of \( \mu \) minimizing \( Q(\mu) \) in (3) over \( \Gamma \) subject to

\[ f(\mu) - B(\hat{\mu}) = 0, \]
where \( B(\hat{\mu}) \) is \( B(\mu) \) defined in (8) evaluated at \( \hat{\mu} \). Note that \( \hat{\mu} \) satisfies (9) instead of (4). In practice, \( B(\mu) \) is small so that \( \hat{\mu} \) nearly satisfies (4). However, we will show that \( \hat{\mu} \) has better statistical properties than \( \hat{\mu} \) in terms of estimating \( \mu \). By considering the asymptotic expansion of \( \hat{\mu} \) around \( \mu \) up to the same order of approximation as done in Theorem 1, we can compare the approximate bias of \( \hat{\mu} \) to that of \( \hat{\mu} \).

**Theorem 2** Under the conditions of Theorem 1,

\[
\hat{\mu} - \mu = d_1 + d_3 + O_p \left( \frac{1}{n\sqrt{n}} \right),
\]

where \( d_1 \) is given in Theorem 1, and

\[
d_3 = -\Psi F'\Phi^{-1}(c - B(\mu)) - V_{d_1} G'\Phi^{-1}F\epsilon
\]

\[
E(d_3) = 0.
\]

The first term in the expansion of \( \hat{\mu} \) is the same as the first term in (5). Since \( E(d_3) = 0 \) whereas \( E(d_2) \neq 0 \), \( \hat{\mu} \) has smaller bias than \( \hat{\mu} \) up to this order of approximation. If \( f(\mu) \) is linear in \( \mu \), then (9) reduces to (4), and \( \hat{\mu} \) and \( \hat{\mu} \) are identical.

To make statistical inferences for the elements of \( \mu \) such as testing and constructing confidence intervals, we assume that

(iv) \( \sqrt{n}\epsilon \xrightarrow{d} N(0, \Sigma) \), as \( n \to \infty \).

This assumption is reasonable in practice, since \( \epsilon \) is some kind of average. Under this assumption \( \sqrt{n}d_1 \) in Theorems 1 and 2 converges to a normal random vector, and \( \hat{\mu} \) and \( \hat{\mu} \) are approximately normally distributed if \( \Psi \) is small. Hence, approximate confidence regions for \( \mu \) can be obtained based on either \( \hat{\mu} \) or \( \hat{\mu} \) and based on normality. To carry out such inference procedures, some variability estimates are required. A naive estimate of the variability can be obtained by evaluating \( V_{d_1} = Var(d_1) \) at \( \hat{\mu} \) or \( \hat{\mu} \).
But, the expansions in Theorems 1 and 2 suggest a better estimate by incorporating the variability in the higher-order second terms. For this, we note that $d_3 = d_2 - E(d_2)$, and that $d_1$ and $d_2$ are uncorrelated in the limit under (iv). Thus, approximate mean squared errors (MSE) of $\hat{\mu}$ and $\tilde{\mu}$ are

$$\text{MSE}(\hat{\mu}) = \text{Var}(d_1) + \text{Var}(d_2) + E(d_2)E(d_2)',$$

$$\text{MSE}(\tilde{\mu}) = \text{Var}(d_1) + \text{Var}(d_2),$$

where the expressions for $\text{Var}(d_1) = V_{d_1}$ and $E(d_2)$ are given in (6) and (7), and

$$\text{Var}(d_2) = \Psi F'\tilde{\Phi}^{-1}V_c\tilde{\Phi}^{-1}F'\Psi + V_{d_1}(\sum_{j=1}^r s_j' \otimes A_j)(\Psi \otimes V_{d_1})(\sum_{j=1}^r s_j' \otimes A_j)'V_{d_1}. \quad (13)$$

$$V_c = \frac{1}{2} \begin{bmatrix}
tr(A_1 V_{d_1} A_1 V_{d_1}) & \cdots & tr(A_1 V_{d_1} A_r V_{d_1}) \\
\vdots & \ddots & \vdots \\
tr(A_r V_{d_1} A_1 V_{d_1}) & \cdots & tr(A_r V_{d_1} A_r V_{d_1})
\end{bmatrix},$$

$$\tilde{\Phi}^{-1} = \left( s_1, s_2, \ldots, s_r \right).$$

Note that the difference $\text{MSE}(\hat{\mu}) - \text{MSE}(\tilde{\mu})$ is non-negative definite, and is nonzero unless $B(\mu) = 0$. If $f(\mu)$ is linear in $\mu$, then the approximations in (11) and (12) reduce to $\text{Var}(d_1)$. For nonlinear $f$, (11) and (12) provide better approximations to the true MSE of $\hat{\mu}$ and $\tilde{\mu}$ than $V_{d_1}$, incorporating higher order terms. We suggest using the estimates of the MSE obtained by evaluating (11) and (12) at $\mu = \hat{\mu}$ and $\mu = \tilde{\mu}$, respectively, in making inferences for $\mu$, e.g., confidence regions. Even with the biased $\tilde{\mu}$, the use of this MSE estimate is expected to produce a good approximate coverage probability, especially compared to that using only $V_{d_1}$. Because of the larger bias and "larger" MSE for $\tilde{\mu}$, we expect the confidence regions using $\tilde{\mu}$ and MSE($\tilde{\mu}$) to provide a smaller confidence region than that constructed using $\hat{\mu}$ and MSE($\hat{\mu}$), keeping the same approximate level of coverage.
The results presented in Theorems 1 and 2 were derived assuming \( \Psi \) is known. This assumption is valid in a number of situations. In many cases the manufacturer's specifications or plant data provide information about the variability in the measurement devices. If this information is based on a large study, then it could be used to develop \( \Psi \) and the assumption of known \( \Psi \) would not be unreasonable. Or, if this information is unavailable, we might conduct a designed experiment for the sole purpose of determining \( \Psi \). If the range over which the experiment is designed is representative of the operating range of the process, and if the sample size is large enough, then we might consider an estimate of \( \Psi \) derived from this experiment to be known, and hence, the results reported in Theorems 1 and 2 would apply.

If an estimate of \( \Psi \) is available along with some measure of variability associated with the estimate, then the results of Theorems 1 and 2 can be modified to include the variability due to the estimation of \( \Psi \). Consider the model described by (1) - (2) where \( \Psi \) is unknown. An estimate of \( \mu \) is obtained by minimizing the function

\[
Q(\mu) = (y - \mu)' \hat{\Psi}^{-1} (y - \mu),
\]

subject to (4), where \( \hat{\Psi} \) is an estimate of \( \Psi \). We use \( \hat{\mu}_\phi \) to denote the estimate that minimizes (14) subject to (4). Also, we define the bias adjusted estimator to be the value of \( \mu \) that minimizes (14) subject to

\[
f(\mu) - B(\hat{\mu}_\phi) = 0,
\]

where \( B(\hat{\mu}_\phi) \) is \( B(\mu) \) in (8) evaluated at \( \mu = \hat{\mu}_\phi \) and \( \Psi = \bar{\Psi} \).

**Theorem 3** Assume that (i) - (iii) in Theorem 1 hold. In addition, assume

\((v)\)  \( \hat{\Psi} \) is a \( p \times p \) positive definite matrix independent of \( \varepsilon \) satisfying \( n(\bar{\Psi} - \Psi) = O_p(\frac{1}{\sqrt{n}}) \) as \( n \to \infty \), where \( d \) is some index tending to infinity.
Then,

\[ \hat{\mu}_{\hat{\Psi}} - \mu = d_1 + d_2 + d_4 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{d}} \right] \right), \]

(16)

\[ \hat{\mu}_{\hat{\Psi}} - \mu = d_1 + d_3 + d_4 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{d}} \right] \right), \]

(17)

where \( d_1, d_2 \) and \( d_3 \) are given in Theorems 1 and 2, and

\[ d_4 = (I_p - \Psi F \Phi^{-1} F)(\hat{\Psi} - \Psi)F \Phi^{-1} F \epsilon = O_p \left( \frac{1}{\sqrt{nd}} \right), \]

(18)

\[ E(d_4) = 0. \]

(19)

Thus, the additional uncertainty attributed to estimation of the error variance covariance matrix \( \Psi \) translates into an additional term in the expansions given in (16) and (17). To obtain an explicit form of the additional variability, assume that

(vi) \( nd\hat{\Psi} \sim W(n\Psi, d) \),

where \( W(\cdot) \) denotes the Wishart distribution. Under assumptions (i) - (vi),

\[ MSE(\hat{\mu}_{\hat{\Psi}}) = MSE(\hat{\mu}) + Var(d_4), \]

(20)

and

\[ MSE(\hat{\mu}_{\hat{\Psi}}) = MSE(\hat{\mu}) + Var(d_4), \]

(21)

where \( MSE(\hat{\mu}) \) and \( MSE(\hat{\Psi}) \) are given in (11) and (12), and

\[ Var(d_4) = \frac{2}{d} \left[ I_p - \Psi F \Phi^{-1} F \right] \sum_{i=1}^{r} C_{ii} \left[ I_p - F' \Phi^{-1} F \Psi \right], \]

\[ C = \left( \Phi^{-\frac{1}{2}} F \otimes I_p \right) P_{K_p} [\Psi \otimes \Psi] P_{K_p} \left( F' \Phi^{-\frac{1}{2}} \otimes I_p \right) \]

\[ = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rr} \end{bmatrix}, \]
\[ P_{K_p} = K_p \left( K_p' K_p \right)^{-1} K_p', \]

and \( K_p \) is a \( p^2 \times \frac{p(p+1)}{2} \) matrix consisting of zeros and ones satisfying

\[ \text{vec}(\hat{\Sigma}) = K_p \text{vech}(\hat{\Sigma}). \]

Testing Problem

A second objective of interest to the process engineers, the detection of instrument biases and process leaks, can be formulated as a problem of statistical hypothesis testing. For this formulation, we assume that the \( r \) restrictions in \( f(\mu) \) can be partitioned into two subsets: \( f_1(\mu) \) containing \( r - s \) restrictions and \( f_2(\mu) \) containing the remaining \( s \) restrictions, and write \( f(\mu) = \begin{bmatrix} f_1(\mu) \\ f_2(\mu) \end{bmatrix} \). We assume that the restrictions in \( f_1(\mu) \) are known to hold, and that \( f_2(\mu) \) contains the restrictions suspect for violation. For example, a portion of the process corresponding to some restrictions may have recently been checked for biases in the measurement instruments and for process "leaks," and proper corrections have been made. Then, \( f_2(\mu) \) corresponds to the portion of the system for which the calibration problems or "leaks" have not been checked. Thus, the problem of interest is to test the hypotheses

\[ H_0 : f_2(\mu) = 0, \quad (22) \]

\[ H_a : f_2(\mu) \neq 0, \quad (23) \]

given the knowledge of \( f_1(\mu) = 0 \). To develop procedures for testing (22) - (23), it is convenient to consider full and reduced models. The full model assumes that (22) is not known to hold, and has only \( f_1(\mu) = 0 \) as restrictions on the model. The reduced model assumes \( H_0 \) is true, and has both \( f_1(\mu) = 0 \) and \( f_2(\mu) = 0 \) as restrictions, i.e., \( f(\mu) = 0 \). In this section, we assume that \( \Psi \) is known. The two estimation
procedures discussed previously can be applied for the full and reduced models. We use \( \hat{\mu}_R \) to denote the estimate of \( \mu \) obtained by minimizing (3) subject to (4), and \( \hat{\mu}_R \) the estimate minimizing (3) subject to (9) with \( \hat{\mu} = \hat{\mu}_R \). For the full model, \( \hat{\mu}_F \) is the estimate of \( \mu \) obtained by minimizing (3) subject to \( f_1(\mu) = 0 \), and \( \hat{\mu}_F \) represents the estimate minimizing (3) subject to \( f_1(\mu) - B_1(\hat{\mu}_F) = 0 \), where \( B_1(\mu) \) is \( B(\mu) \) using only \( f_1(\mu) = 0 \) as restrictions. Three types of general techniques for developing a test procedure are the likelihood ratio test (LRT), the Wald test (W), and the Lagrange multiplier test (LM). In the context of our problem, the LRT and W can be derived easily, but the application of the LM idea requires some derivation.

Under the normality of \( \epsilon \), the LRT statistic for testing (22) versus (23) is

\[
\text{LRT} = \begin{pmatrix} y - \hat{\mu}_R \end{pmatrix}' \Psi^{-1} \begin{pmatrix} y - \hat{\mu}_R \end{pmatrix} - \begin{pmatrix} y - \hat{\mu}_F \end{pmatrix}' \Psi^{-1} \begin{pmatrix} y - \hat{\mu}_F \end{pmatrix}.
\] (24)

Using the bias adjusted estimator \( \hat{\mu} \), we can consider a modified LRT statistic

\[
\tilde{\text{LRT}} = \begin{pmatrix} y - \hat{\mu}_R \end{pmatrix}' \Psi^{-1} \begin{pmatrix} y - \hat{\mu}_R \end{pmatrix} - \begin{pmatrix} y - \hat{\mu}_F \end{pmatrix}' \Psi^{-1} \begin{pmatrix} y - \hat{\mu}_F \end{pmatrix}.
\] (25)

As a practical consideration, the LRT requires fitting both the full and reduced models. Consequently, for its implementation in an on-line fashion to monitor a chemical process, the LRT may not be very practical. From this point of view, the LM and W tests are more appealing, since each test requires only one model fitting.

The W test statistic can be obtained using only the estimates under the full model. To obtain a good variability estimate used in the W test, we can use the MSE estimates developed in (20) and (21). Thus, the W test statistic using \( \hat{\mu}_F \) is

\[
\tilde{W} = f_2(\hat{\mu}_F) \hat{H}_{22}^{-1} f_2(\hat{\mu}_F),
\] (26)

where \( \hat{H}_{22} = \hat{F}_2 MSE \hat{F}_2' \), \( \hat{F}_2 \) is obtained by evaluating the \( s \times p \) \( F_2 \) in \( F = (F_1', F_2')' \) at \( \hat{\mu}_F \), and \( \hat{MSE} \) in (20) for fitting only \( f_1(\mu) = 0 \) evaluated at \( \hat{\mu}_F \). Similarly, using \( \hat{\mu}_F \) and (21), we can define

\[
\tilde{W} = f_2(\hat{\mu}_F) \hat{H}_{22}^{-1} f_2(\hat{\mu}_F).
\] (27)
The LM test can be obtained by fitting only the reduced model. The idea behind the LM test is described in Silvey [14]. Since our problem involves two sets of implicit restrictions, we need to modify the general formula to obtain test statistics for our problem. First, we consider using $\hat{\mu}_R$ obtained by minimizing (3) subject to (4) containing both $f_1(\mu)$ and $f_2(\mu)$. Consider this minimization as solving the derivative equations for the Lagrangian function

$$\frac{1}{2} (y - \mu)' \Psi^{-1} (y - \mu) + \lambda_1' f_1(\mu) + \lambda_2' f_2(\mu).$$

Then, the solutions $\hat{\mu}_R$ and $\hat{\lambda} = (\hat{\lambda}_1', \hat{\lambda}_2')'$ satisfy

$$\hat{F}' \hat{\lambda} = \Psi^{-1} (y - \hat{\mu}_R),$$

$$f(\hat{\mu}_R) = 0,$$

where $\hat{F}$ is $F$ evaluated at $\hat{\mu}_R$. It follows that

$$\hat{\lambda} = (\hat{F}' \Psi \hat{F}')^{-1} \hat{F}' (y - \hat{\mu}_R).$$

If assumption (iv) holds, and if the reduced model holds, i.e., $f_2(\mu) = 0$, then by Theorem 1, $\hat{\lambda}$ is approximately normally distributed with a covariance matrix estimated by $(\hat{F}' \Psi \hat{F}')^{-1}$. Since we are testing only the $f_2$ part of the restrictions, the LM test statistic can be obtained using only the $\hat{\lambda}_2$ part of $\hat{\lambda}$. Hence, a natural extension of the standard LM test statistic to our problem is

$$\lambda_2' \hat{E}_0^{-1} \hat{\lambda}_2,$$

where $\hat{E}_0$ is the $s \times s$ lower right corner of $(\hat{F}' \Psi \hat{F}')^{-1}$. Note that if we write

$$\hat{F}' \Psi \hat{F}' = \begin{pmatrix} \hat{K}_{11} & \hat{K}_{12} \\ \hat{K}_{21} & \hat{K}_{22} \end{pmatrix},$$

then $\hat{E}_0^{-1} = \hat{K}_{22,1} = \hat{K}_{22} - \hat{K}_{21} \hat{K}_{11}^{-1} \hat{K}_{12}$. However, a better estimate of the covariance matrix can be obtained using the expansion in Theorem 1. For this, let

$$\hat{LM} = \lambda_2' \hat{E}^{-1} \lambda_2,$$
where \( \hat{E} \) is the \( s \times s \) lower right corner of

\[
(\hat{F} \Psi \hat{F}^\prime)^{-1} + (\hat{F} \Psi \hat{F}^\prime)^{-1} \hat{F} \hat{\text{Var}}_{d_{2}} \hat{F}^\prime (\hat{F} \Psi \hat{F}^\prime)^{-1},
\]

(30)

and \( \hat{\text{Var}}_{d_{2}} \) is \( \text{Var}(d_{2}) \) evaluated at \( \hat{\mu}_{R} \). Using the bias adjusted estimator, \( \hat{\mu}_{R} \), we can similarly form the LM test statistic

\[
\hat{L}M = \hat{\lambda}_{2}^\prime \hat{E}^{-1} \hat{\lambda}_{2},
\]

where \( \hat{E} \) is (30) evaluated at \( \hat{\mu}_{R} \), and \( \hat{\lambda} = (\hat{F} \Psi \hat{F}^\prime)^{-1} \hat{F}(y - \hat{\mu}_{R}) \).

To derive and compare the properties of the six test statistics under \( H_{0} \) and \( H_{a} \), we consider asymptotic expansions under assumption (i) and a contiguous alternative. For this purpose, we assume that the true \( \mu \) satisfies

\[
f_{2}(\mu) = \frac{1}{\sqrt{n}} \delta,
\]

(31)

where the \( s \times 1 \) vector \( \delta \) is zero under \( H_{0} \), and \( n \) is given in (i).

**Theorem 4** Assume (i), (ii), and (iii) in Theorem 1 hold, and assume that the true \( \mu \) satisfies (31). Then

\[
\hat{L}RT = n [x + c_{0}]^\prime \Phi^{-1} [x + c_{0}] - n [x + c_{F}]^\prime \left( \begin{array}{cc} \Phi_{11}^{-1} & O \\ O & O \end{array} \right) [x + c_{F}] + O_{p} \left( \frac{1}{n} \right),
\]

\[
\hat{L}RT = n [x + c_{0} - B(\mu)]^\prime \Phi^{-1} [x + c_{0} - B(\mu)]
\]

\[
- n [x + c_{F} - B(\mu)]^\prime \left( \begin{array}{cc} \Phi_{11}^{-1} & O \\ O & O \end{array} \right) [x + c_{F} - B(\mu)] + O_{p} \left( \frac{1}{n} \right),
\]

\[
\hat{W} = n \left[ \left( \begin{array}{c} 0 \\ \frac{1}{\sqrt{n}} \delta \end{array} \right) + Fd_{1F} + Fd_{2F} + c_{F} \right]^\prime \left( O, I_{s} \right)^\prime \Omega \left( O, I_{s} \right)
\]

\[
\left[ \left( \begin{array}{c} 0 \\ \frac{1}{\sqrt{n}} \delta \end{array} \right) + Fd_{1F} + Fd_{2F} + c_{F} \right] + O_{p} \left( \frac{1}{n} \right),
\]
\[
\tilde{W} = n \left[ \left( \frac{0}{\sqrt{n} \delta} \right) + Fd_{1_F} + Fd_{3_F} + c_F \right] \right]^{\prime} \left( O, I_s \right)^{\prime} \Omega \left( O, I_s \right) \\
\left[ \left( \frac{0}{\sqrt{n} \delta} \right) + Fd_{1_F} + Fd_{3_F} + c_F \right] + O_p \left( \frac{1}{n} \right),
\]

\[
\tilde{\Lambda} = n \left[ x + c_0 \right]^{\prime} \Omega \left[ x + c_0 \right] + O_p \left( \frac{1}{n} \right),
\]

\[
\tilde{\Lambda} = n \left[ x + c_0 - B(\mu) \right]^{\prime} \Omega \left[ x + c_0 - B(\mu) \right] + O_p \left( \frac{1}{n} \right),
\]

where

\[
x = F\varepsilon + \left( \frac{0}{\sqrt{n} \delta} \right),
\]

\[
c_0 = \frac{1}{2} \left[ w'A_w w \quad w'A_{2w} \quad \ldots \quad w'A_r w \right]^{\prime},
\]

\[
w = d_1 - \Sigma F^{\prime} \Phi^{-1} \left( \frac{0}{\sqrt{n} \delta} \right),
\]

\[
c_F = \frac{1}{2} [d_{1_F} A_1 d_{1_F}, \ldots, d_{1_F} A_r d_{1_F}],
\]

\[
\Omega = \Phi^{-1} - \left( \Phi^{-1}_{11} 0 \\
0 0 \right),
\]

\[
\Phi = \left( \Phi_{11} \Phi_{12} \\
\Phi_{21} \Phi_{22} \right),
\]

and \(d_{1_F}, d_{2_F} \) and \(d_{3_F}\) are \(d_1, d_2 \) and \(d_3\), respectively, computed for the full model (i.e., using only \(f_1(\mu) = 0\)).
The expansions in Theorem 4 lead to a number of results comparing the six test statistics. First, we observe that the expansions in the theorem include the terms of order $O_p(\frac{1}{\sqrt{n}})$. If we ignore such terms, all six statistics are equal to $x'\Omega x$. Thus, under assumption $(iv)$, all six statistics converge in distribution to $\chi^2_2(\delta'\Phi_{22,1}^{-1}\delta)$ where $\Phi_{22,1} = \Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12}$. Hence, the percentiles of $\chi^2$ distribution are used as approximate reference points for all six test statistics.

To obtain a meaningful comparison, we consider the full expansions in Theorem 4 including the higher order terms. Note that the first order approximation to all six test statistics is a noncentral chi-square distribution. Hence, it seems insightful to obtain a higher order approximate expression for a quantity corresponding to the noncentrality parameter for each statistic. Noting that every term in the expansion in Theorem 4 is a quadratic form in some random vector, an approximate noncentrality parameter can be obtained by replacing the random vector by its expectation in every quadratic form. Denoting such an expression by $\phi$, we obtain

$$\phi(LRT) = n \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) + M \right]^{'} \Phi^{-1} \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) + M \right]$$

$$- n B(\mu)' \begin{pmatrix} \Phi_{21}^{-1} & 0 \\ 0 & 0 \end{pmatrix} B(\mu),$$

$$\phi(W) = n \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) \right]^{'} \Omega \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) \right],$$

$$\phi(W) = n \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) \right]^{'} \begin{pmatrix} O & O \\ O & \Phi_{22,1}\Phi_{11}^{-1} \end{pmatrix} \left[ \left( \frac{0}{\sqrt{n}\delta} \right) + B(\mu) \right].$$
\[
\phi(\bar{L}M) = n \left[ \left( \frac{0}{\sqrt{n} \delta} \right) + B(\mu) + M \right] \Omega \left[ \left( \frac{0}{\sqrt{n} \delta} \right) + B(\mu) + M \right],
\]

where

\[
\phi(\bar{L}M) = n \left[ \left( \frac{0}{\sqrt{n} \delta} \right) + M \right] \Omega \left[ \left( \frac{0}{\sqrt{n} \delta} \right) + M \right],
\]

In assessing these expressions, we first consider the null distribution by setting \( \delta = 0 \). With \( \delta = 0 \), \( \phi(\bar{L}RT) \) and \( \phi(\bar{L}M) \) reduce to zero, while the four others are positive due to the nonlinearity of \( f \). Hence, we expect the \( \chi^2 \) approximation of the null distribution to be better for \( \bar{L}RT \) and \( \bar{L}M \) using the bias adjusted estimators than for the four others. Note that, for the \( W \) test, the use of the bias adjusted estimator does not completely remove the deviation from the central \( \chi^2 \) due to nonlinearity of \( f \). Under \( H_a : \delta \neq 0 \), it can be seen that \( \phi(\bar{L}RT) > \phi(\bar{L}M) \). Thus, \( \bar{L}RT \) appears to be more powerful than \( \bar{L}M \).

But, a close examination of \( \phi(\bar{L}RT) \) and \( \phi(\bar{L}M) \) shows that the difference is only in a certain weight matrix and is generally small. Recall that the LM test requires estimation of only one model, the reduced model, while the LRT requires estimation of two sets of parameter estimates. Thus, for the purposes of on-line hypothesis testing, the use of the LM test using the bias adjusted estimator might be useful. Otherwise, we suggest the use of the modified LRT with the bias adjusted estimator.
Simulation Study

A numerical study was conducted using simulated measurements for the continuous stir tank reactor process given in Kim et al. [10]. The process is a first order irreversible reaction \( A \rightarrow B \) in which species A is converted to species B and the rate of reaction is given by \( k \), which is a nonlinear function of temperature. The five true values of the measurements are \( \mu_A^0 \): inlet concentration of species A, \( \mu_A \): outlet concentration of species A, \( \mu_B \): outlet concentration of species B, \( \mu_T^0 \): inlet temperature, and \( \mu_T \): outlet temperature. Three constraints describing the steady state model are:

\[
\begin{align*}
    f_1(\mu) &= \theta_1(\mu_A^0 - \mu_A) - k\mu_A = 0, \quad (33) \\
    f_2(\mu) &= -\theta_1\mu_B + k\mu_A = 0, \quad (34) \\
    f_3(\mu) &= \theta_1(\mu_T - \mu_T^0) - \theta_2k\mu_A = 0, \quad (35) \\
    k &= \theta_3e^{-\theta_4\left(\frac{\mu_T}{T_T} - 1\right)}. 
\end{align*}
\]

where \( \theta_1, \theta_2, \theta_3, \theta_4 \) and \( \theta_5 \) are all known constants determined by the chemical engineer using the engineering knowledge of the process and \( \mu = (\mu_A, \mu_A^0, \mu_B, \mu_T^0, \mu_T)' \). The restrictions (33) and (34) describe the changes in \( \mu_A \) and in \( \mu_B \) over time, respectively, assuming the process is in steady state. These restrictions hold because there is no accumulation or loss of mass or energy, depletion over time in the steady state process. The third restriction (35) is the energy balance on the tank contents.

In our simulation, the true values of \( \mu \) was set to be the first set of conditions given in Table 7 in Kim et al. [10], i.e.,

\[
\mu = (0.882 \text{ mol/L}, 1.0 \text{ mol/L}, 0.118 \text{ mol/L}, 0.54710 \text{ K}/1000, 0.66509 \text{ K}/1000)',
\]

where, in our study, temperatures \( \mu_T \) and \( \mu_T^0 \) were scaled by 1000 K to be of the same order as the other parameters. We generated observed measurements by adding normal errors to \( \mu \). Following Kim et al.[10], the five errors were assumed to be independent.
The error variances were also taken to be in a range suggested by Kim et al. [10], and were set equal to $(0.05)^2 \text{mol}^2/L^2$ for concentrations and $(0.005)^2 K^2/1000^2$ for temperature. From this model, 1000 samples were generated.

For estimating the true value $\mu$, we note that $Y$ is an unbiased estimator of $\mu$, and that an exact confidence interval for each $\mu_i$ can be constructed based on $Y_i$, the corresponding observation, and the known error variance. The length of such an interval is a normal percentile times the error standard deviation. We consider using the bias adjusted estimator $\tilde{\mu}$ along with higher order estimated MSE based on (12) for constructing an approximate confidence interval for $\mu_i$. Table 2.1 gives the error standard deviation, the Monte Carlo MSE for $\tilde{\mu}$, and the average of the estimated MSE. As can be seen in Table 2.1, the estimated MSE with the higher order terms provides good approximation to the true MSE on average, despite the highly nonlinear restrictions. Also, the estimated MSE of $\tilde{\mu}$ is much smaller than the error standard deviation. Thus,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Error STD</th>
<th>$\sqrt{MC\text{MSE}}$</th>
<th>$\sqrt{AVEM\text{SE}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_A$</td>
<td>0.05</td>
<td>0.00646</td>
<td>0.00645</td>
</tr>
<tr>
<td>$\mu_T$</td>
<td>0.005</td>
<td>0.00070</td>
<td>0.00065</td>
</tr>
<tr>
<td>$\mu_B$</td>
<td>0.05</td>
<td>0.00139</td>
<td>0.00137</td>
</tr>
<tr>
<td>$\mu_{A0}$</td>
<td>0.05</td>
<td>0.00632</td>
<td>0.00633</td>
</tr>
<tr>
<td>$\mu_{T0}$</td>
<td>0.005</td>
<td>0.00087</td>
<td>0.00086</td>
</tr>
</tbody>
</table>

the length of a confidence interval based on $\tilde{\mu}_i$ is much narrower than that based on $Y_i$. Hence, the data reconciliation using the restrictions provides much more information about the true value $\mu$ than the use of the observation $Y$ as an estimate. Table 2.2 presents the percentage of times the true value was contained in the approximate 95% confidence interval based on $\tilde{\mu}_i$ and the estimated higher order MSE. The use of the higher order MSE seems to produce reasonably accurate coverage probability by a much narrower interval than that using $Y_i$. 
Table 2.2 Simulated coverage probability for $\hat{\mu}$

<table>
<thead>
<tr>
<th></th>
<th>$\mu_A$</th>
<th>$\mu_T$</th>
<th>$\mu_B$</th>
<th>$\mu_{A_0}$</th>
<th>$\mu_{T_0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}$</td>
<td>95.1</td>
<td>92.1</td>
<td>93.9</td>
<td>94.8</td>
<td>94.4</td>
</tr>
</tbody>
</table>

For the testing problem, we considered testing the validity of (35), i.e.,

$$H_0 : f_3(\mu) = 0,$$  \hspace{1cm} (36)

$$H_a : f_3(\mu) \neq 0.$$  \hspace{1cm} (37)

The three statistics using the bias adjusted estimate $\hat{\mu}$ described in Section 3 were computed and compared to the upper 0.05 point of the chi-squared distribution with 1 degree of freedom. The number of times $H_0$ was rejected is summarized in Table 2.3.

This corresponds to the null case and the chi-squared approximation is good for the three tests. In order to compare the power of the six tests, 1000 samples were generated from the same structure as before, except $\mu_{T_0}$ was decreased by $2\sigma$ to $0.53935K/1000$ so that $f_3(\mu) = -0.07$. Table 2.4 presents the power of the three test for this particular alternative. We see that all the tests have high power.
Bibliography


Appendix

The proofs of Theorems 1 - 3 are given in this appendix. The proof of Theorem 3 is only sketched, as it mostly duplicates those for Theorems 1 and 2.

The following lemma will be used in the proof of Theorem 1.

Lemma 1 Under the assumptions of Theorem 1, \( \hat{\mu} \) is a consistent estimator of \( \mu \), as \( n \to \infty \).

Proof. Note that the minimization of \( Q(\mu) \) in (3) is equivalent to the minimization of \( Q^*(\mu) = n^{-1}Q(\mu) = (y - \mu)' \Sigma^{-1}(y - \mu) \). For the true value \( \mu_0 \), \( y \xrightarrow{P} \mu_0 \) and \( Q^*(\mu_0) \xrightarrow{P} 0 \). Since \( \hat{\mu} \) minimizes \( Q^*(\mu) \) over all \( \mu \) satisfying \( f(\mu) = 0 \), and since \( \mu_0 \) satisfies \( f(\mu) = 0 \),

\[
0 \leq Q^*(\hat{\mu}) \leq Q^*(\mu_0),
\]

and \( Q^*(\hat{\mu}) \xrightarrow{P} 0 \). Since \( y \xrightarrow{P} \mu_0 \), \( \hat{\mu} \xrightarrow{P} \mu_0 \).

Proof of Theorem 1

Consider the lagrangian function associated with the minimization of \( Q^*(\mu) \),

\[
Q(\mu, \alpha) = (y - \mu)' \Sigma^{-1}(y - \mu) + \alpha' f(\mu),
\]

where \( \alpha \) is an \( r \times 1 \) vector of lagrange multipliers. By Lemma 1, we can assume \( \hat{\mu} \) is in the interior of the parameter space with probability approaching one as \( n \to \infty \). Thus, we can act as if \( \hat{\mu} \) satisfies the derivative equations for (A.1),

\[
-\Sigma^{-1}(y - \hat{\mu}) + \hat{F}' \alpha = 0,
\]

\[
f(\hat{\mu}) = 0,
\]

where \( \hat{F} = \frac{\partial f(\mu)}{\partial \mu} |_{\mu=\hat{\mu}} \). By the mean value theorem, for the true value of \( \mu \), there exists a point \( \bar{\mu} \) on the line segment joining \( \hat{\mu} \) and \( \mu \) such that

\[
f(\hat{\mu}) = f(\mu) + \hat{F}(\hat{\mu} - \mu),
\]

(A.4)
where \( \tilde{F} = \left. \frac{\partial \tilde{f}(\mu)}{\partial \mu} \right|_{\mu = \hat{\mu}} \). Since both \( \hat{\mu} \) and the true value \( \mu \) satisfy (4), (A.4) simplifies to

\[
\tilde{F}'(\hat{\mu} - \mu) = 0.
\]

Hence,

\[
\tilde{F}(y - \hat{\mu}) = \tilde{F}\epsilon - \tilde{F}(\hat{\mu} - \mu) = \tilde{F}\epsilon.
\]  
(A.5)

Premultiplying (A.2) by \( \tilde{F}\Sigma \), we obtain

\[
-\tilde{F}(y - \hat{\mu}) + \tilde{\Pi}\alpha = 0,
\]

where \( \tilde{\Pi} = \tilde{F}\Sigma\tilde{F}' \). By assumption (iii), \( \tilde{\Pi} \) is positive definite for large \( n \), so that

\[
\alpha = \tilde{\Pi}^{-1}\tilde{F}(y - \hat{\mu}).
\]  
(A.6)

Combining (A.5) and (A.6), we see that

\[
\alpha = \tilde{\Pi}^{-1}\tilde{F}\epsilon.
\]  
(A.7)

Substituting (A.7) into (A.2) and solving for \( y - \hat{\mu} \), we obtain

\[
y - \hat{\mu} = \Sigma\tilde{F}'\tilde{\Pi}^{-1}\tilde{F}\epsilon.
\]  
(A.8)

It follows from assumptions (i) and (iii) and Lemma 1, that

\[
\hat{\mu} - \mu = \left( I_p - \Sigma\tilde{F}'\tilde{\Pi}^{-1}\tilde{F} \right)\epsilon = O_p \left( \frac{1}{\sqrt{n}} \right).
\]  
(A.9)

In turn, it then follows that

\[
\hat{\mu} - \mu = d_1 + O_p \left( \frac{1}{n} \right).
\]  
(A.10)

We next use the first order expansion of \( \hat{\mu} - \mu \) obtained in (A.10) to derive a higher order approximation of \( \hat{\mu} - \mu \). To do this, we consider an approximation of \( \hat{\mu} - \mu \) which
includes all terms of order $O_p\left(\frac{1}{n}\right)$. Expanding $f(\hat{\mu})$ around $\mu$ to the third order terms, we can show that

$$f(\hat{\mu}) = f(\mu) + F(\hat{\mu} - \mu) + \frac{1}{2} \left[ (\hat{\mu} - \mu)'A_1(\hat{\mu} - \mu) \right] + \frac{1}{3} \left[ (\hat{\mu} - \mu)'A_2(\hat{\mu} - \mu) \right] + \cdots + O_p\left(\frac{1}{n^{3/2}}\right),$$

(A.11)

where $A_i$ is defined in Theorem 1. Noting that $f(\hat{\mu}) = 0$ and $f(\mu) = 0$, and using (A.10), we have

$$0 = F(\hat{\mu} - \mu) + c + O_p\left(\frac{1}{n^{3/2}}\right),$$

(A.12)

where $c$, defined in Theorem 1, is $O_p\left(\frac{1}{n}\right)$. Premultiplying (A.2) by $F\Sigma$, and using (A.12), we can show

$$\alpha = \hat{\Pi}^{-1}F(y - \hat{\mu})$$

(A.13)

$$= \hat{\Pi}^{-1}[v + c] + O_p\left(\frac{1}{n^{3/2}}\right),$$

(A.14)

where $\hat{\Pi} = F\Sigma \hat{F}'$ and $v = Fe$. Substituting (A.13) in (A.2),

$$\hat{\mu} - \mu = e - \Sigma \hat{F}' \hat{\Pi}^{-1}(v + c) + O_p\left(\frac{1}{n^{3/2}}\right).$$

(A.15)

The expansion (5) for $\hat{\mu} - \mu$ follows because

$$\hat{F}' = F' + G + O_p\left(\frac{1}{n}\right),$$

(A.16)

$$\hat{\Pi} = F\Sigma (F + G)' + O_p\left(\frac{1}{n}\right),$$

$$\hat{\Pi}^{-1} = (F\Sigma F)^{-1} - (F\Sigma F)^{-1} F\Sigma G' (F\Sigma F)^{-1} + O_p\left(\frac{1}{n}\right).$$

The expectation results can be derived by direct evaluation.

Proof of Theorem 2

For the consistency of $\hat{\mu}$, let $\mu_*$ be the value closest (in Euclidean norm) to the true value $\mu$ among all $\mu^*$ satisfying $f(\mu^*) = B(\hat{\mu})$. Then, by the argument used in Lemma
1, \( \hat{\mu} - \mu_n \xrightarrow{P} 0 \). Since \( B(\hat{\mu}) = O_p \left( \frac{1}{n} \right) \), \( \mu_n \xrightarrow{P} \mu \), and \( \hat{\mu} \) is consistent. To obtain the expansion, consider the lagrangian function

\[
Q(\mu, \alpha) = (y - \mu)'\Sigma^{-1}(y - \mu) + \alpha'[f(\mu) - B(\hat{\mu})],
\]

(A.17)

where

\[
B(\hat{\mu}) = \frac{1}{2} \left[ \text{tr}(\hat{A}_1 \hat{V}_{d_1}), \text{tr}(\hat{A}_2 \hat{V}_{d_2}), \ldots, \text{tr}(\hat{A}_r \hat{V}_{d_r}) \right]',
\]

\[
\hat{V}_{d_i} = \frac{1}{n} \left[ \Sigma - \Sigma \hat{F}' (\hat{F}' \Sigma \hat{F}')^{-1} \hat{F} \Sigma \right],
\]

and \( \hat{A}_i \) is evaluated at \( \hat{\mu} \). For \( \hat{\mu} \) in the interior of the parameter space, the derivative equations are

\[
-S^{-1}(y - \hat{\mu}) + \hat{F}' \alpha = 0,
\]

(A.18)

\[
f(\hat{\mu}) - B(\hat{\mu}) = 0.
\]

(A.19)

Since \( B(\hat{\mu}) = O_p \left( \frac{1}{n} \right) \), the argument used for obtaining (A.10) leads to

\[
\hat{\mu} - \mu = d_1 + O_p \left( \frac{1}{n} \right).
\]

(A.20)

Using (A.20) and an expansion of \( f(\hat{\mu}) \) as in (A.11), we can show

\[
B(\hat{\mu}) = \hat{F}(\hat{\mu} - \mu) + c + O_p \left( \frac{1}{n^{1/2}} \right).
\]

(A.21)

Writing \( \hat{F}(\hat{\mu} - \mu) = \hat{F} \epsilon - \hat{F}(y - \hat{\mu}) \), and solving for \( \hat{F}(y - \hat{\mu}) \), we see that

\[
\hat{F}(y - \hat{\mu}) = v + c - B(\hat{\mu}) + O_p \left( \frac{1}{n^{1/2}} \right).
\]

(A.22)

Premultiplying (A.18) by \( \hat{F} \Sigma \), and using (A.21), we obtain

\[
\alpha = \hat{\Pi}^{-1} \hat{F}(y - \hat{\mu})
\]

(A.23)

\[
= \hat{\Pi}^{-1} [v + c - B(\mu)] + O_p \left( \frac{1}{n^{1/2}} \right),
\]

(A.24)

where \( \hat{\Pi} = \hat{F} \Sigma \hat{F}' \). Using (A.23) in (A.18), and using the argument at the end of the previous proof, we obtain the expansion result.
Proof of Theorem 3

The consistency of \( \hat{\mu}_\psi \) follows from the argument in Lemma 1 and the fact that \( n \hat{\Psi} \xrightarrow{p} \Sigma \). Applying the steps to obtain (A.10) to the derivative equations

\[
\begin{align*}
-\hat{\Sigma}^{-1}(y - \hat{\mu}_\psi) + \hat{F}'\alpha &= 0, \\
f(\hat{\mu}_\psi) &= 0,
\end{align*}
\tag{A.25, A.26}
\]

with \( \hat{\Sigma} = n \hat{\Psi} \), we obtain

\[
\hat{\mu}_\psi - \mu = d_1 + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nd}} \right] \right).
\]

Following the steps to obtain (A.13), the result for this case can be shown to be

\[
\alpha = \hat{\Pi}_{\psi}^{-1}(v + c) + O_p \left( \frac{1}{n\sqrt{n}} \right),
\tag{A.27}
\]

where \( \hat{\Pi}_{\psi} = F\hat{\Sigma}\hat{F}' \). Using (A.27) in (A.25), we have

\[
\hat{\mu}_\psi - \mu = \epsilon - \hat{\Sigma}\hat{F}\hat{\Pi}_{\psi}^{-1}(v + c) + O_p \left( \frac{1}{n\sqrt{n}} \right).
\tag{A.28}
\]

We note (A.16) and

\[
\hat{\Pi} = F\hat{\Sigma}\hat{F}'
\]

\[
= F\Sigma F' + F\Sigma G' + F(\hat{\Sigma} - \Sigma) F' + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nd}} \right] \right)
\]

\[
\hat{\Pi}^{-1} = (F\Sigma F')^{-1} - (F\Sigma F')^{-1} \left[ F\Sigma G' + F(\hat{\Sigma} - \Sigma) F' \right] (F\Sigma F')^{-1}
\]

\[
+ O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nd}} \right] \right).
\]

Thus, the expansions for \( \hat{\mu}_\psi - \mu \) follows from (A.28) and the argument at the end of Theorem 1 proof. The expectation of \( d_1 \) is zero because of the independence of \( \hat{\Psi} \) and \( \epsilon \). The derivation of the expansion for \( \hat{\mu}_\psi \) follows the derivation for \( \hat{\mu}_\tilde{\phi} \) and Theorem 2.
3 NONLINEAR ERRORS-IN-VARIABLES ANALYSIS WITH MEASUREMENT ERROR BIASES

A paper to be submitted to Journal of the American Statistical Association

Jean E. Pelkey

Abstract

Statistical analysis of measurements taken for the purpose of monitoring a large engineering process such as a chemical processing system is discussed. The process variables are known to satisfy certain nonlinear physical restrictions involving unknown parameters provided the variables are measured without error. In practice, all variables are measured with error and some system anomaly, e.g., one or more leaks in the process, leads to the violation of some of the restrictions. In engineering problems, it is often possible to obtain an accurate estimate of the variance-covariance matrix of the measurement errors through replication or past experience. However, systematic measurement error bias can be present for some of the variables due to calibration or human errors. Thus, a statistical problem in monitoring such processes consists of unknown parameter estimation and detecting possible measurement biases and/or violation of restrictions. Nonlinear errors-in-variables analysis can be used to approach this problem. It is shown that, in a model with sufficient nonlinearity, systematic bias of all instruments and all restriction parameters can be estimated. Properties of the maximum likelihood estimators and its bias-adjusted modification are discussed. Inference procedures are proposed for the systematic measurement bias and the restriction violation.
Introduction

In a chemical process, accurate statistical parameter estimation and timely detection of systematic biases and process "leaks" are important tools used by process control engineers in system monitoring. Inaccurate estimation can lead to unnecessary adjustments to the process that can adversely affect the quality of the product. The measurements taken at points within the process are subject to measurement errors. In addition to random variability, a systematic measurement bias can exist due to instrument calibration error. In an engineering process, the true error-free values satisfy some known physical restrictions. These constraints are called material and energy balance equations, and may contain unknown parameters to be estimated. One problem of interest to the engineers is the estimation of the true process values in the process based on the measurements and the material and energy balance equations. A second problem is that of estimating and checking the instrument biases. Another problem of interest is to detect the violation of a restriction due to system anomaly such as process leaks. In the engineering literature, the first problem is called data reconciliation and the second and third are jointly referred to as gross error detection. This engineering classification arose, because only linear restrictions have been considered. For the linear model, it is known that the whole model is not identified and most parameters are not estimable. In particular, the estimation of all instrument biases is impossible. Thus, the gross error detection is concerned with detection of some error without specifying whether it is due to measurement bias or system anomaly. In this paper, a model with nonlinear restrictions is considered. It is shown that all parameters including the true process values, all measurement biases, and restriction parameters can be estimated with sufficiently nonlinear restrictions.

Consider a process that is monitored over \( T \) time points using \( p \) measurement stations. Let \( \mu_{it} \) denote the true value of the process at the \( i^{th} \) collection station at time \( t \).
Let $Y_t = (Y_{t1}, Y_{t2}, \ldots, Y_{tp})'$ and $\mu_t = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{pt})'$. Assume that the $p$ values in $\mu_t$ satisfy $r$ physical nonlinear restrictions containing $k$ parameters, and that $\mu_t$ is a fixed vector in a parameter space $\Upsilon$. The process monitoring model is

$$Y_t = \mu_t + \gamma + \epsilon_t,$$  \hspace{1cm} (1)

$$f(\mu_t, \theta) = 0,$$

where $\epsilon_t$ is the $p \times 1$ measurement error vector at time $t$ with $E(\epsilon_t) = 0$ and $V(\epsilon_t) = \Psi$, $\gamma$ is a $p \times 1$ vector of instrument biases belonging to a parameter space $\Gamma$, the $r$-valued function $f(\mu_t, \theta)$ is a function defined on $\Upsilon \times \Theta$ representing the physical constraints at time $t$, and $\theta$ is a $k \times 1$ vector of unknown parameters belonging to a parameter space $\Theta$. It is assumed that $\Upsilon$ and $\Gamma$ are subsets of $p$-dimensional Euclidean space, and that $\Theta$ is a subset of $k$-dimensional Euclidean space. We consider the case for which $f(\mu_t, \theta)$ is generally nonlinear in both $\mu_t$ and $\theta$. It is assumed that $r < p$, and that the constraints $f(\mu_t, \theta) = 0$ is written in a consistent and non-redundant manner. In engineering applications, the measurement error covariance matrix $V(\epsilon_t) = \Psi$ is often known or estimated with acceptable accuracy based on replication or specification. In this paper, we assume that $\Psi$ is a known positive definite matrix. Model (1) contains a large number of unknown parameters to be estimated. The $p \times 1$ vector $\mu_t$ represents the true but unknown value of those parameters for which measurements are taken at time $t$, the $p \times 1$ vector $\gamma$ corresponds to a vector of systematic biases (e.g. instrument biases), and the $k \times 1$ vector $\theta$ represents those unknown values determining the restrictions.

In a simple engineering system such as a pipeline system, the material in the system does not change its form drastically, and the physical constraints represent simple mass balances. For such a case, the constraint $f(\mu_t, \theta)$ does not contain an unknown parameter $\theta$, is linear in $\mu_t$, and is of the form

$$A\mu_t = 0,$$  \hspace{1cm} (2)
where $A$ is a known $r \times p$ matrix of rank $r$. For example, in a pipeline system with $r$ processing units, each of the $r$ constraints corresponds to the equality between the input and output. The model with such a linear restriction (2) has been discussed widely in the chemical engineering literature. See, e.g., Almasy [2], Darouach and Zasadzinski [7], Narasimhan and Mah [11], Rollins and Devanathan [13]. However, in a more complex engineering system, the restrictions represent material and energy balance equations that can be quite nonlinear and can involve parameters specific to a system. For the model with nonlinear restrictions, recent work in the chemical engineering literature assumes no measurement bias, i.e. $\gamma = 0$, and has concentrated on finding efficient optimization routines based on the methods proposed by Britt and Luecke ([4]) (e.g., Albuquerque and Biegler [1], Kim et al. [8], Kim et al. [9], Liebman et al. [10]). They do not discuss statistical inference procedures or statistical properties of the estimators. Statistical analysis of model (1) is that of nonlinear errors-in-variables. See, e.g., Fuller [7] and Carroll et al. [5]. Model (1) is similar to that discussed by Amemiya and Fuller [3] and we follow their approach for some issues. However, no work in nonlinear measurement error analysis directly addresses the estimation of measurement bias, $\gamma$.

This paper considers the general model (1) with the instrument bias $\gamma$ and the unknown parameter $\theta$, assuming that $f(\mu, \theta)$ is nonlinear in $\mu_i$. The issue of estimability of $\theta$ and $\gamma$ is an important point of discussion, and is addressed in the next section. It turns out that $\gamma$ is generally non-estimable unless $f(\mu, \theta)$ is sufficiently nonlinear in $\mu_i$. Estimation of $\mu_i$, $\gamma$ and $\theta$ is discussed fully in the following section. Sections on testing and simulation follow. The appendix contains all derivations.

**Estimability**

To discuss the estimability of parameters in model (1), we first consider simple linear constraints. Suppose that the $r$ constraints in $f(\mu, \theta)$ are linear in both $\mu_i$ and $\theta$, and
of the form
\[ f(\mu_t, \theta) = A_1 \mu_t + A_2 \theta = 0, \] (3)

where \( A_1 \) is \( r \times p \) known matrix of rank \( r \), and \( A_2 \) is an \( r \times k \) known matrix of rank \( k \).

Then, let \( R \) be a \( p \times p \) nonsingular matrix such that
\[ R = \begin{pmatrix} A_1 \\ L_1 \end{pmatrix}, \]
\[ L_1' A_1 = O, \]

and consider a transformed observation
\[ \begin{pmatrix} Y_{1t}^* \\ Y_{2t}^* \end{pmatrix} = R Y_t. \]

Then, under model (1) and restriction (3),
\[ Y_{1t}^* = A_1 \mu_t + A_1 \gamma + \epsilon_{1t}^* = -A_2 \theta + \gamma_1^* + \epsilon_{1t}^*, \]
\[ Y_{2t}^* = L_1 \mu_t + \gamma_2^* + \epsilon_{2t}^*, \]

where \( \gamma_1^* = A_1 \gamma \) and \( \gamma_2^* = L_1 \gamma \). Thus, \( L_1 \mu_t, t = 1, 2, \ldots, T, \) and \( \gamma_2^* \) are non-estimable, and the identification of \( \theta \) and \( \gamma_1^* \) needs to be made only through \( Y_{1t}^* \). For all \( t = 1, 2, \ldots, T \), we write
\[ Y_t^* = S^* \begin{pmatrix} \theta \\ \gamma \end{pmatrix} + \epsilon^*, \]

where \( Y_t^* = (Y_{1t}^*, Y_{12}^*, \ldots, Y_{1T}^*)' \), and
\[ S^* = \begin{bmatrix} -A_2 & A_1 \\ -A_2 & A_1 \\ \vdots & \vdots \\ -A_2 & A_1 \end{bmatrix}. \] (4)
The rank of $Tr \times (p + k) S^*$ is $r$ which is less than $p$. Hence, only linear combinations of the $r$ elements of $A_1 \gamma - A_2 \theta$ are estimable. For a function only of the instrument bias $\gamma$, an estimable function has to be a linear combination of the elements of $L_2 A_1 \gamma$, where the rows of the $(r-k) \times r L_2$ span the orthogonal complement of $A_2$. Thus, the whole $\gamma$ is non-estimable even if $\theta$ is not present. For the model with linear restrictions, $\mu_t, \theta$ and $\gamma$ are not generally estimable.

From (4), we see that the non-identifiability of $\theta$ and $\gamma$ are closely related to the fact that $E(Y_{ij}^*)$ is free of $t$. Suppose that the physical restriction depends on $t$ and is given by $A_{1t} \mu_t + A_{2t} \theta = 0$. Then, (4) becomes

$$S^* = \begin{bmatrix} -A_{21} & A_{11} \\ -A_{22} & A_{12} \\ \vdots & \vdots \\ -A_{2T} & A_{1T} \end{bmatrix},$$

which is of full column rank if $A_{1t}$ and $A_{2t}$ vary sufficiently over $t$. This fact can be used to describe intuitively why a model with sufficiently nonlinear restrictions yield the estimability of all parameters.

For the nonlinear model, assumption (v) of Theorem 1 in the next section gives a precise condition that allows estimation of all parameters. An intuitive interpretation of the condition is that $S$, the matrix $Tr \times (p + k)$ defined by $S = [S_\mu, S_\theta]$ has full column rank, where $S_\mu = [F_{\mu1}' F_{\mu2}' \ldots F_{\muT}']'$, is $Tr \times p$, $S_\theta = [F_{\theta1}' F_{\theta2}' \ldots F_{\thetaT}']'$ is $Tr \times k$, $F_{\theta t} = \frac{\partial f(\mu_t, \theta)}{\partial \theta}$ and $F_{\mu t} = \frac{\partial f(\mu_t, \theta)}{\partial \mu_t}$. Thus, $S$ is the matrix of first partial derivatives of $f(\mu_t, \theta)$ with respect to $\mu_t$ and $\theta$ for all $t = 1, 2, \ldots, T$. For the linear model (3), $F_{\mu t} = A_1$ and $F_{\theta t} = A_2$ do not depend on $t$, $rank(S) = r < p + k$, and the condition does not hold. But, if $f(\mu_t, \theta)$ is nonlinear in $\mu_t$, then $F_{\mu t}$ varies over $t$, and $S_\mu$ is expected to be of full column rank. Also, if the parameterization $\theta$ is given in a non-redundant fashion, we expect $S_\theta$ to be of full column rank. Hence, with the nonlinearity of $f(\mu_t, \theta)$ in $\mu_t$, we might expect the estimability of all elements of $\theta$ and
It is shown in the next section that, under the nonlinearity condition, all \( \mu_t, \theta, \) and \( \gamma \) can be estimated.

We point out that the above full column rank condition concerns mainly the nonlinearity of \( f(\mu, \theta) \) in \( \mu \), and not in \( \theta \). Consider linear restrictions of the form

\[
f(\mu, \theta) = a(\theta) + A(\theta) \mu,
\]

where \( a(\theta) \) and \( A(\theta) \) are nonlinear functions of \( \theta \). Then, \( S \) for this case has rank at most \( r + k \), and \( S \) does not have full column rank. Thus, the nonlinearity in \( \mu \) plays the key condition in achieving the estimability of all parameters and in particular of \( \gamma \).

Suppose that elements \( \mu_{it} \) and \( \mu_{jt} \) appear in an identical subset of the \( r \) restrictions, and that both \( \mu_{it} \) and \( \mu_{jt} \) linearly enter each of the restrictions in the set. Then, the two columns of \( S_\mu \) corresponding to \( \mu_{it} \) and \( \mu_{jt} \) are a multiple of each other, and the full column rank condition for \( S_\mu \) is violated. In general, the full rank condition for \( S \) may appear somewhat too technical, but is, in fact, easily verified for a particular applied problem.

**Estimation Problem**

For the model (1) with known \( \Psi \), a natural estimation procedure is to minimize the function

\[
Q(\mu, \gamma, \theta) = \sum_{i=1}^{T} (Y_i - \mu_i - \gamma)\Psi^{-1}(Y_i - \mu_i - \gamma),
\]

with respect to \( (\mu, \gamma, \theta) \) in \( \Gamma \times \Gamma \times \Theta \), subject to

\[
f(\mu, \theta) = 0, \ t = 1, 2, \ldots, T.
\]

The estimators obtained in this way are denoted by \( \hat{\mu}, \hat{\gamma} \) and \( \hat{\theta} \). If \( \varepsilon_t \) are normally distributed, \( \hat{\mu}, \hat{\gamma} \), and \( \hat{\theta} \) are the maximum likelihood estimators. We derive properties of the estimators without assuming the normality of \( \varepsilon_t \). In this type of nonlinear measurement error analysis, the so-called small error asymptotics is often used to obtain
insightful properties of the estimators. See, e.g., Amemiya and Fuller [3] and Stefanski and Carroll [14]. In such a setup, the limit is taken over a sequence of decreasing $\Psi$ and increasing $T$. Asymptotic results obtained in this manner are applicable in practice when the elements of $\Psi$ are small or when $Y_t$ is some summary value based on many readings. In engineering applications, $Y_t$ is often an average or some summary of $n$ measurements taken during a short time period, where the process is believed to be in a steady state period. Then, the assumption that $T \to \infty$, and $\Psi = O \left( \frac{1}{n} \right)$, $n \to \infty$, may be reasonable. Under this assumption, the first theorem gives an asymptotic expansion of $\hat{\mu}_t$, $\hat{\gamma}$ and $\hat{\theta}$ up to higher order terms.

**Theorem 1** Let model (1) hold, and assume

(i) The measurement errors $\epsilon_t$, $t = 1, 2, \ldots, T$, are independently distributed for all $t$, the fourth moments of $n^t \epsilon_t$ exist, and $\text{Var}(\epsilon_t) = \Psi = \frac{1}{n} \Sigma$ where $\Sigma$ is positive definite. Also, $n = o(T)$, i.e., $\frac{n}{T} \to 0$.

(ii) The partial derivatives of $f(\mu_t, \theta)$ with respect to $\mu_t$ of order three or less exist and are continuous on $\Upsilon \times \Theta$.

(iii) The partial derivatives of $f(\mu_t, \theta)$ with respect to $\theta$ of order two or less exist and are continuous on $\Upsilon \times \Theta$.

(iv) For $t = 1, 2, \ldots, T$, $\Phi_{\mu_t} = F_{\mu_t} \Sigma F'_{\mu_t}$ is positive definite where $F_{\mu_t} = \frac{\partial f(\mu_t, \theta)}{\partial \mu_t}$ evaluated at the true values of $\mu_t$ and $\theta$.

(v) As $T \to \infty$, $m \to m_0$ for a positive definite $m_0$, where the $(p+k) \times (p+k)$ matrix

$$m = \begin{pmatrix} m_{\mu \mu} & m_{\mu \theta} \\ m_{\theta \mu} & m_{\theta \theta} \end{pmatrix}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} F'_{\mu_t} \\ -F'_{\theta_t} \end{pmatrix} \Phi^{-1}_{\mu_t} \begin{pmatrix} F_{\mu_t} & -F_{\theta_t} \end{pmatrix},$$
\[
F_{\theta t} = \frac{\partial f(\mu_t, \theta)}{\partial \theta},
\]

and all quantities are evaluated at the true values of \( \mu_t \) and \( \theta \).

In addition, assume that a technical condition (a) given in the appendix holds. Then, as \( n \to \infty \) and \( T \to \infty \),

\[
\hat{\mu}_t - \mu_t = d_{1t} + d_{2t} + d_{3t} + O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right),
\]

\[
\hat{\theta} - \theta = \tau_1 + \tau_2 + O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right),
\]

\[
\hat{\gamma} - \gamma = \Lambda_1 + \Lambda_2 + O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right),
\]

where

\[
d_{1t} = \left( I_p - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \epsilon_t + O_p \left( \frac{1}{\sqrt{n}} \right),
\]

\[
d_{2t} = - \left( I_p - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_1 - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} F_{\theta t} \tau_1 = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

\[
d_{3t} = - \left( I_p - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_2 - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} [F_{\theta t} \tau_2 + c_t]
\]

\[-nV_{d_{1t}} G'_{\mu t} \Phi^{-1}_{\mu t} v_t = O_p \left( \frac{1}{n} \right),
\]

\[
\tau_1 = q_{\theta t, \mu}^{-1} [m_{\theta t} - m_{\theta \mu} m_{\mu t}^{-1} m_{\mu \mu}] = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

\[
\tau_2 = q_{\theta t, \mu}^{-1} [m_{\theta c} - m_{\theta \mu} m_{\mu t}^{-1} m_{\mu c}] = O_p \left( \frac{1}{n} \right),
\]

\[
\Lambda_1 = m_{\mu \mu}^{-1} [m_{\mu \mu} - m_{\mu \mu} \tau_1] = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

\[
\Lambda_2 = m_{\mu \mu}^{-1} [m_{\mu c} - m_{\mu \mu} \tau_2] = O_p \left( \frac{1}{n} \right),
\]
where

\[ v_t = F_{\mu t} \epsilon_t, \]

\[ c_t = \frac{1}{2} [d'_{1t} A_{1t} d_{1t}, d'_{2t} A_{2t} d_{2t}, \ldots, d'_{rt} A_{rt} d_{rt}], \]

\[ A_{it} = \frac{\partial^2 f_i(\mu_{it}, \theta)}{\partial \mu_i \partial \mu_i}, \ i = 1, 2, \ldots, r, \]

\[ V_{d_{1t}} = \text{Var}(d_{1t}) = \frac{1}{n} \left( \Sigma - \Sigma F'_{\mu t} \Phi_{\mu t}^{-1} F_{\mu t} \Sigma \right), \quad (10) \]

\[ G_t = [A_{1t} d_{1t}, A_{2t} d_{1t}, \ldots, A_{rt} d_{1t}], \]

\[
\begin{pmatrix}
  m_{\mu u} & m_{\mu c} \\
  m_{\theta u} & m_{\theta c}
\end{pmatrix}
= \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix}
  F'_{\mu t} \\
  -F'_{\theta t}
\end{pmatrix}
\Phi_{\mu t}^{-1} \left( v_t, c_t \right),
\]

\[ q_{\theta \theta, \mu} = m_{\theta \theta} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu \theta}. \]

Furthermore,

\[ E(d_{1t}) = 0, \quad (11) \]

\[ E(d_{2t}) = 0, \quad (12) \]

\[ E(d_{3t}) = - \left( I_p - \Sigma F'_{\mu t} \Phi_{\mu t}^{-1} F_{\mu t} \right) E(A_2) - \Sigma F'_{\mu t} \Phi_{\mu t}^{-1} [F_{\theta t} E(\tau_2) + B(\mu_t, \theta)], \quad (13) \]

\[ E(\tau_1) = 0, \quad (14) \]

\[ E(\tau_2) = q_{\theta \theta, \mu}^{-1} \left[ m_{\theta B} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu B} \right], \quad (15) \]

\[ E(A_1) = 0, \quad (16) \]
\[ E(A_2) = m_{\mu_2}^{-1} \left( m_{\mu_2} q_{\theta_0, \mu}^{-1} m_{\theta B} - m_{\mu B} \right), \]  

where

\[
B(\mu_i, \theta) = \frac{1}{2} \left[ \text{tr}(A_{1i} V_{d_{1i}}), \text{tr}(A_{2i} V_{d_{1i}}), \ldots, \text{tr}(A_{ri} V_{d_{1i}}) \right]',
\]

\[
\begin{pmatrix}
  m_{\mu B} \\
  m_{\theta B}
\end{pmatrix}
= \frac{1}{T} \sum_{i=1}^{T} \begin{pmatrix}
  F'_{\mu i} \\
  -F'_{\theta i}
\end{pmatrix}
\Phi_{\mu i}^{-1} B(\mu_i, \theta).
\]

In the expansions in Theorem 1, \( d_{1i}, \tau_1 \) and \( A_1 \) are the leading terms with zero expectation. The term \( d_{2i} \) represents the first order effect of estimation of \( \theta \) and \( \gamma \) in estimation of \( \mu_i \). The terms \( d_{3i}, \tau_2 \) and \( A_2 \) are quadratic functions of \( e_i \)'s, depend on the second derivatives of \( f(\mu_i, \theta) \) with respect to \( \mu_i \), and have nonzero expected values. Thus, \( E(d_{2i}), E(\tau_2) \) and \( E(A_2) \) represent biases due to the nonlinearity of \( f(\mu_i, \theta) \) in \( \mu_i \). Under the general identification condition (v), these biases do not vanish. Note that the instrument bias \( \gamma \) enters the model linearly. But, \( \hat{\gamma} \) still possesses the nonlinearity bias, because the estimation of \( \gamma \) is possible only through the nonlinearity of \( f(\mu_i, \theta) \).

For a similar model Amemiya and Fuller [3] propose a second estimator that adjusts for the nonlinearity bias. In their model, \( \gamma = 0 \) and their adjusted (BA) estimator minimizes (5) subject to the restriction

\[ f(\mu_i, \theta) - B(\mu_i, \hat{\theta}) = 0. \]  

Note that \( B(\mu_i, \hat{\theta}) \) is known since \( V_{d_{1i}} = \Psi - \Psi F'_{\mu i} \Phi_{\mu i}^{-1} F_{\mu i} \Psi \) with \( \Phi_{\mu i} = F_{\mu i} \Psi F'_{\mu i} \), and \( F_{\mu i} \) and \( A_{it} \) can be evaluated at \( \mu_i \) and \( \hat{\theta} \). Amemiya and Fuller [3] show that their bias adjusted estimator, \( (\hat{\mu}_i, \hat{\theta}) \), has smaller bias than \( (\mu_i, \hat{\theta}) \) for their model. For model (1), we take a slightly different approach to bias adjustment. Given \( \mu_i \) and \( \hat{\theta} \), choose \( \Delta_i \) so that \( F_{\mu i} \Delta_i = -B(\mu_i, \hat{\theta}) \), where \( F_{\mu i} \) is \( F_{\mu i} \) evaluated at \( \mu_i \). For example, \( \Delta_i = -\Psi F'_{\mu i} \Phi_{\mu i}^{-1} B(\mu_i, \hat{\theta}) \) with \( \Phi_{\mu i} = F_{\mu i} \Psi F'_{\mu i} \) satisfies \( F_{\mu i} \Delta_i = -B(\mu_i, \hat{\theta}) \). Then, we adjust the observations to \( Y_i' = Y_i + \Delta_i \) and define \( (\mu_i, \gamma, \hat{\theta}) \) as the estimator that
minimizes the function
\[
Q(\mu_t, \gamma, \theta, \alpha_t) = \sum_{i=1}^{T} (Y_i - \mu_t - \gamma)' \Psi^{-1} (Y_i - \mu_t - \gamma)
\]  
(19)
subject to \( f(\mu_t, \theta) = 0 \). It is shown in Theorem 2 below that \( (\hat{\mu}_t, \hat{\gamma}, \hat{\theta}) \) has smaller bias than \( (\hat{\mu}_t, \hat{\gamma}, \hat{\theta}) \).

**Theorem 2** Under the conditions of Theorem 1,
\[
\begin{align*}
\hat{\mu}_t - \mu_t &= d_{4t} + d_{2t} + d_{4t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \\
\hat{\theta} - \theta &= \tau_1 + \tau_3 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \\
\hat{\gamma} - \gamma &= \Lambda_1 + \Lambda_3 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
\end{align*}
\]  
(20)
(21)
(22)

where
\[
d_{4t} = - \left( I_p - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_3 - \Sigma F'_{\mu t} \Phi^{-1}_{\mu t} [F_{\theta t} \tau_3 + c_t] \\
- nV_{d_{4t}} G'_{\mu t} \Phi^{-1}_{\mu t} v_t = O_p \left( \frac{1}{n} \right),
\]  
(23)
\[
\tau_3 = \frac{1}{\sigma_{\theta \mu}} \left[ (m_{\theta e} - m_{\theta B}) - m_{\theta \mu} m_{\mu}^{-1} (m_{\mu c} - m_{\mu B}) \right] = O_p \left( \frac{1}{n} \right),
\]  
(24)
\[
\Lambda_3 = m_{\mu}^{-1} [m_{\mu c} - m_{\mu B} - m_{\mu \theta} \tau_3] = O_p \left( \frac{1}{n} \right).
\]  
(25)

In addition,
\[
E(d_{4t}) = - \Sigma \Phi_{\mu t}^{-1} B(\mu_t, \theta),
\]
\[
E(\tau_3) = 0,
\]
\[
E(\Lambda_3) = 0.
\]

Thus, comparing the approximate bias of \( (\hat{\mu}_t, \hat{\gamma}, \hat{\theta}) \) and \( (\hat{\mu}_t, \hat{\gamma}, \hat{\theta}) \) given in Theorems 1 and 2, we see that, up to the order of approximation given, the bias adjustment removes the nonlinearity bias in \( \hat{\gamma} \) and \( \hat{\theta} \), and a part of the bias in \( \hat{\mu}_t \).
The expansions in Theorems 1 and 2 can also be used to obtain higher order approximate mean squared error (MSE) of each estimator. To obtain explicit formulas, assume the normality of \( \epsilon_t \), which may be reasonable if \( \epsilon_t \) is some average value based on many errors. By Theorem 1,

\[
MSE(\hat{\mu}_t) = Var(d_{1t}) + Var(d_{2t}) + Var(d_{3t}) + E(d_{2t})E(d_{2t})' + E(d_{3t})E(d_{3t})',
\]

\[
MSE(\hat{\theta}) = Var(\tau_1) + Var(\tau_2) + E(\tau_2)E(\tau_2)',
\]

\[
MSE(\hat{\gamma}) = Var(\Lambda_1) + Var(\Lambda_2) + E(\Lambda_2)E(\Lambda_2)',
\]

where \( Var(d_{1t}) = V_{d_{1t}}, E(d_{2t}), E(d_{3t}), E(\Lambda_1), E(\Lambda_2), E(\tau_1) \) and \( E(\tau_2) \) are given in Theorem 1. Also,

\[
Var(\tau_1) = \frac{1}{nT}q_{\theta,\mu}^{-1},
\]

\[
Var(\Lambda_1) = \frac{1}{nT} \left[ m_{\mu}^{-1} + m_{\mu}^{-1}m_{\mu}q_{\theta,\mu}^{-1}m_{\mu}m_{\mu}^{-1} \right],
\]

\[
Var(d_{2t}) = \frac{1}{nT} \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right) \left[ m_{\mu}^{-1} + m_{\mu}^{-1}m_{\mu}q_{\theta,\mu}^{-1}m_{\mu}m_{\mu}^{-1} \right] \times \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right)' + \frac{1}{nT} \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\theta}q_{\theta,\mu}^{-1}F_{\theta} \Phi_{\mu}^{-1} F_{\mu} \Sigma
\]

\[
- \frac{2}{nT} \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right) \left[ \frac{1}{nT} \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right) \left[ m_{\mu}^{-1} + m_{\mu}^{-1}m_{\mu}q_{\theta,\mu}^{-1}m_{\mu}m_{\mu}^{-1} \right] \times \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right)' \right] m_{\mu}^{-1}m_{\mu}q_{\theta,\mu}^{-1}F_{\theta} \Phi_{\mu}^{-1} F_{\mu} \Sigma,
\]

\[
Var(d_{3t}) = \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right) Var(\Lambda_2) \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right)' + \Sigma F'_{\mu} \Phi_{\mu}^{-1} Var(c_t + F_{\theta_t} \tau_2) \Phi_{\mu}^{-1} F_{\mu} \Sigma
\]

\[
+ nV_{d_{1t}} \left( \sum_{i=1}^r s_{it} \otimes A_{it} \right) \left( \Phi_{\mu} \otimes V_{d_{1t}} \right) \left( \sum_{i=1}^r s_{it} \otimes A_{it} \right)' V_{d_{1t}}
\]

\[
+ 2 \left( I_p - \Sigma F'_{\mu} \Phi_{\mu}^{-1} F_{\mu} \right) Cov(\Lambda_2, c_t + F_{\theta_t} \tau_2) \Phi_{\mu}^{-1} F_{\mu} \Sigma,
\]

\[
Var(\tau_2) = q_{\theta,\mu}^{-1} \left[ V_{\theta c} + m_{\theta}m_{\mu}^{-1}V_{\mu c}m_{\mu}^{-1}m_{\theta} - 2V_{\theta c}m_{\mu}^{-1}m_{\theta} \right] q_{\theta,\mu}^{-1},
\]
where $\Phi_{\mu t}^{-1} = [s_{1t}, s_{2t}, \ldots, s_{rt}]$, $V_{\theta t c} = \frac{1}{T_2} \sum_{t=1}^{T} F_{\theta t}^r \Phi_{\mu t}^{-1} V_{c_t} \Phi_{\mu t}^{-1} F_{\theta t}$,

$V_{\mu c} = \frac{1}{T_2} \sum_{t=1}^{T} F_{\mu t}^r \Phi_{\mu t}^{-1} V_{c_t} \Phi_{\mu t}^{-1} F_{\mu t}$, $V_{\theta c} = -\frac{1}{T_2} \sum_{t=1}^{T} F_{\theta t}^r \Phi_{\mu t}^{-1} V_{c_t} \Phi_{\mu t}^{-1} F_{\mu t}$,

$Var(A_2) = m_{\mu t}^{-1} [V_{\mu c} + m_{\mu t} Var(\tau_2) m_{\mu t} - 2 (V_{\mu c} - V_{\mu c} m_{\mu t}^{-1} m_{\mu t} \theta) q_{\theta \mu}^{-1} \theta_{\mu}] m_{\mu t}^{-1}$,

$Cov(A_2, c_t + F_{\theta t} \tau_2) = \frac{1}{T} m_{\mu t}^{-1} \left[ F_{\mu t}^r \Phi_{\mu t}^{-1} V_{c_t} + m_{\mu t} q_{\theta \mu}^{-1} \left( F_{\theta t} + F_{\mu t} m_{\mu t}^{-1} m_{\mu t} \theta \right) \Phi_{\mu t}^{-1} V_{c_t} \right]$

$$+ m_{\mu t}^{-1} \left( (V_{\mu c} - V_{\mu c} m_{\mu t}^{-1} m_{\mu t} \theta) q_{\theta \mu}^{-1} \theta_{\mu} - m_{\mu t} Var(\tau_2) F_{\theta t} \right),$$

$Var(c_t + F_{\theta t} \tau_2) = V_{c_t} + F_{\theta t} Var(\tau_2) F_{\theta t}^{-1}$

$$- \frac{2}{T} V_{c_t} \Phi_{\mu t}^{-1} \left[ F_{\theta t} - F_{\mu t} m_{\mu t}^{-1} m_{\mu t} \theta \right] q_{\theta \mu}^{-1} \theta_{\mu} F_{\theta t}.$$
Theorem 3 Assume the conditions of Theorem 1 hold. Then,

\[ \sqrt{nT} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\gamma} - \gamma \end{pmatrix} \overset{d}{\to} N_{p+k}(0, m_0^{-1}), \]

\[ \sqrt{nT} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\gamma} - \gamma \end{pmatrix} \overset{d}{\to} N_{p+k}(0, m_0^{-1}), \]

where \( m_0 \) is given in assumption (v).

Thus, if the error variances are not large or if \( T \) is large, then the limiting normal distribution can be used to make inferences about \( \theta \) and \( \gamma \). Note that if we set

\[ \Pi_{\theta, \gamma} = \begin{pmatrix} \text{Var}(\tau_1) & \text{Cov}(\tau_1, \Lambda_1) \\ \text{Cov}(\Lambda_1, \tau_1) & \text{Var}(\Lambda_1) \end{pmatrix}, \]

\[ \text{Cov}(\tau_1, \Lambda_1) = -\frac{1}{nT}m_{\mu_\mu}m_{\mu_\delta}q_{\delta, \mu}, \]

then

\[ nT\Pi_{\theta, \gamma} \overset{p}{\to} m_0^{-1}. \]

To improve the normal approximation, we recommend the use of the higher order MSE rather than the first order variance as the variability estimate. For this, MSE's in (25), (26), (28) and (29) evaluated at appropriate estimates can be used for, e.g., constructing confidence regions. Recall that \( \gamma \) represents the systematic biases of the \( p \) instruments.

Thus, the problem of checking the calibration of all \( p \) instruments is to test

\[ H_0 : \gamma = 0, \]

\[ H_a : \gamma \neq 0. \]

An appropriate procedure for this problem using the bias adjusted estimator is to reject \( H_0 \) if

\[ \hat{\gamma}'\overline{MSE}(\hat{\gamma})^{-1}\hat{\gamma} > \chi^2_{\alpha, p}, \]

where \( \overline{MSE}(\hat{\gamma}) \) is (29) evaluated at \((\hat{\mu}, \hat{\gamma}, \hat{\theta})\).
Testing For System Anomaly

Another problem of interest to the engineers is to detect any anomaly in the system such as process leaks. The problem of detecting process leaks corresponds to testing whether some subset of the $r$ restrictions in (6) hold. We assume that no system anomaly was suspected for $T$ time points, but that at a particular time $t_0$ some restrictions are suspect. Typically, $t_0$ is $T + 1$ or a new time point when a system is altered after time $T$. We partition $f(\mu_t, \theta)$ into two subsets, $f(\mu_t, \theta) = \begin{pmatrix} f_1(\mu_t, \theta) \\ f_2(\mu_t, \theta) \end{pmatrix}$. Given that $f(\mu_t, \theta) = 0, t = 1, 2, \ldots, T$ and that $f_1(\mu_{t_0}, \theta) = 0$, we wish to test

$$H_0 : f_2(\mu_{t_0}, \theta) = 0,$$
$$H_a : f_2(\mu_{t_0}, \theta) \neq 0.$$
n \to \infty,

\begin{align*}
f_2(\hat{\mu}_{t_0}, \hat{\Theta}_T) &= f_2(\mu_{t_0}, \theta) + F_{2\mu_{t_0}}(d_{2t_0} + d_{3t_0} + d_{5t_0}) + F_{2\delta_{t_0}}(\tau_1 + \tau_3) \\
&\quad + c_{2t_0} + o_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
\end{align*}

where $d_{1t_0}$ and $d_{2t_0}$ are $d_1$ and $d_2$ in Theorem 1 evaluated at $t = t_0$,

\begin{align*}
d_{5t_0} &= -(I_p - \sum F'_{1\mu_{t_0}} \Phi^{-1}_{11\mu_{t_0}} F_{1\mu_{t_0}}) \Lambda_3 \\
&\quad - \sum F'_{1\mu_{t_0}} \Phi^{-1}_{11\mu_{t_0}} [F_{\delta_{t_0} \tau_3} + c_{1t_0} - B_1(\mu_{t_0}, \theta_T)] \\
&\quad - nV_{1,d_{1t_0}} G'_{1t_0} \Phi^{-1}_{11\mu_{t_0}} F_{1\mu_{t_0}} e_{t_0},
\end{align*}

\begin{align*}
c_{2t_0} &= \frac{1}{2} \left[ d_{1_{r+s+t_0}}' A_{r-s+t_0} d_{1t_0}, d_{1_{r+s+2t_0}}' A_{r-s+2t_0} d_{2t_0}, \ldots, d_{1_{r+t_0}}' A_{r+t_0} d_{1t_0} \right]',
\end{align*}

$\tau_1, \tau_3, \Lambda_1$ and $\Lambda_3$, are given in Theorems 1 and 2, and the quantities with subscript 1 are those in Theorems 1 and 2 using only $f_1$. Furthermore,

\begin{equation*}
E(d_{5t_0}) = 0.
\end{equation*}

Theorem 4 has shown that $f_2(\hat{\mu}_{t_0}, \hat{\Theta}_T)$ is a good estimate of $f_2(\mu_{t_0}, \theta)$, although a small bias term $E(c_{2t_0})$ remains. For a variability estimate, we suggest $\tilde{H}$ obtained by computing the variance of the expansion terms in Theorem 4 and evaluating it at the appropriate estimates. Then, a Wald test rejects the hypothesis of $f_2(\mu_{t_0}, \theta) = 0$ if

\begin{equation*}
[f_2(\hat{\mu}_{t_0}, \hat{\Theta}_T)]' \tilde{H}^{-1} f_2(\hat{\mu}_{t_0}, \hat{\Theta}_T) > \chi^2_{2,s},
\end{equation*}

with $s$ being the dimension of $f_2$.

**Simulation Study**

Finite sample properties of the estimators were evaluated by a simulation study. Consider model (1), where $p = 2$, $r = 1$, $k = 1$, $\mu_t = (\mu_{t_1}, \mu_{t_2})$, and the restriction is

$$\mu_{t_1} = \theta \mu_{t_2}.$$
Given $2 \times 1$ observations $Y_t$, $t = 1, 2, \ldots, T$, the relationship parameter $\theta$, the $2 \times 1$ instrument bias $\gamma = (\gamma_1, \gamma_2)'$, and $2T$ true values are estimated. The true bias was taken to be $\gamma = (-2.5, 1.7)$, and two choices of the true value of $\theta$ were 1 and 2. For the true values $\mu_t$, we first chose 25 values that were equally spaced within the interval $(-1, 1)$. Then, each of the 25 values were replicated either 3 or 8 times to generate $T = 75$ or $T = 200$ values of $\mu_{2t}$. The true values of $\mu_{1t}$ are given by $\mu_{1t} = \theta \mu_{2t}^2$. The replicate structure of the true values were unknown in the estimation, i.e., $T$ different $\mu_t$'s are to be estimated. The true $\epsilon_t$ was generated as $N(0, \sigma^2 I_2)$. For each of the two sample sizes, $T = 75$ and $T = 200$, two choices of $\sigma^2$ were chosen via comparison to the variability among the true $\mu_{2t}$. That is, for each $T$, $\sigma^2$ was chosen to be either $M = \frac{1}{3}$ or $M = \frac{1}{9}$ of the variance among $\mu_{2t}$'s. Hence, there were 8 cases depending on two values each of $\theta$, $T$, and $M$. For each of the 8 cases, 1000 samples were generated. From each sample, the maximum likelihood estimates (MLE) (the natural estimator minimizing (5) and the bias adjusted (BA) estimator (minimizing (19)) were obtained.

The results of the simulation study are presented in Table 3.1. The organization of the table is as follows. Columns 1, 2, and 3 give the true value of $\theta$, the total sample size, $T$, and the ratio $M$ of the error variance to the variance of $\mu_{2t}$, respectively. The fourth column shows the theoretical approximate bias of the MLE of $(\gamma_1, \gamma_2, \theta)$ which was obtained using the results of Theorem 1. Columns 5 and 6 indicate the Monte Carlo bias and MSE for the MLE of $(\gamma_1, \gamma_2, \theta)$. Columns 7 and 8 report the same quantities for the BA estimator. The entries reported in Columns 4 through 8 are multiplied by 100 to make comparisons easier.

It can be seen in Table 3.1 that the actual bias of the MLE is even larger than the asymptotic bias based on the expansion in Theorem 1. The bias increases, as the error variance ratio increases, or as the curvature or degree of nonlinearity, $\theta$ increases. Increasing the sample size $T$ does not directly decrease the bias which is more closely associated with the error variability and the nonlinearity. The bias adjustment is suc-
Table 3.1 Bias (×100) and MSE (×100) for \( \gamma_1, \gamma_2 \) and \( \theta \).

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( T )</th>
<th>( M )</th>
<th>Theoretical Bias of MLE</th>
<th>Bias of MLE</th>
<th>MSE of MLE</th>
<th>Bias of BA</th>
<th>MSE of BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>( \frac{1}{3} )</td>
<td>-10.5</td>
<td>-31.9</td>
<td>14.3</td>
<td>-11.9</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>-0.19</td>
<td>0.40</td>
<td>-0.25</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.0</td>
<td>71.7</td>
<td>97.4</td>
<td>31.2</td>
<td>44.3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>( \frac{1}{9} )</td>
<td>-3.5</td>
<td>-5.3</td>
<td>0.5</td>
<td>-1.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>-0.190</td>
<td>0.130</td>
<td>-0.190</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.7</td>
<td>8.7</td>
<td>3.4</td>
<td>3.4</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>( \frac{1}{3} )</td>
<td>-10.5</td>
<td>-30.3</td>
<td>10.5</td>
<td>-9.5</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0.003</td>
<td>0.14</td>
<td>0.011</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.0</td>
<td>66.4</td>
<td>59.0</td>
<td>24.3</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>( \frac{1}{9} )</td>
<td>-3.5</td>
<td>-5.1</td>
<td>0.4</td>
<td>-1.0</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0.044</td>
<td>0.051</td>
<td>0.042</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.7</td>
<td>8.1</td>
<td>1.7</td>
<td>2.9</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>( \frac{1}{3} )</td>
<td>-17.8</td>
<td>-40.2</td>
<td>19.1</td>
<td>-20.5</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>-0.113</td>
<td>0.215</td>
<td>-0.125</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.8</td>
<td>86.8</td>
<td>119.6</td>
<td>46.1</td>
<td>61.2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>( \frac{1}{9} )</td>
<td>-5.9</td>
<td>-9.8</td>
<td>1.3</td>
<td>-3.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>-0.125</td>
<td>0.090</td>
<td>-0.126</td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.3</td>
<td>20.6</td>
<td>10.5</td>
<td>8.9</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>( \frac{1}{3} )</td>
<td>-17.8</td>
<td>-39.2</td>
<td>16.4</td>
<td>-18.9</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0.129</td>
<td>0.081</td>
<td>0.130</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.8</td>
<td>82.9</td>
<td>84.9</td>
<td>40.5</td>
<td>29.7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>( \frac{1}{9} )</td>
<td>-5.9</td>
<td>-9.7</td>
<td>1.1</td>
<td>-3.3</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>-0.034</td>
<td>0.034</td>
<td>-0.032</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.3</td>
<td>20.7</td>
<td>6.8</td>
<td>8.9</td>
<td>2.9</td>
<td></td>
</tr>
</tbody>
</table>
cessful in reducing the bias as well as the MSE, although some bias remains. For the model used, the theoretical bias of the MLE of $\gamma_2$ is zero. Even for such a parameter, the bias adjustment does not increase the variability of the estimator.

**Bibliography**


Appendix

Proof of Theorem 1

To prove Theorem 1, we consider a reparameterization of the model (1). We define $\xi_t = \mu_t + \gamma$ and write (1) as

$$Y_t = \xi_t + \epsilon_t$$

$$f(\xi_t - \gamma, \theta) = 0. \quad (A.1)$$

Examining (A.1), we note that the vector $\xi_t$ denotes unknown values for which measurements are made and $\gamma$ and $\theta$ both represent unknown values with no associated measurements. Thus, we could combine $\gamma$ and $\theta$ into one vector, $\pi = \begin{pmatrix} \gamma \\ \theta \end{pmatrix}$, where $\pi$ is a $(p + k) \times 1$ vector, with (A.1) as

$$f(\xi_t, \pi) = 0. \quad (A.2)$$

The minimization problem described in (5) - (6) can be carried out with respect to $\xi_t$ and $\pi$. Let $\mu_t^0$ and $\pi_0 = (\gamma_0', \pi_0')'$ be the true values of $\mu_t$ and $\pi_0$, and let $\xi_t^0 = \mu_t^0 + \gamma_0$. Also, let $\Xi$ be the parameter space of $\pi$. Throughout this proof, the conditions of Theorem 1 hold. The technical assumption mentioned in this theorem is

(a) For all $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ and a $T_\epsilon > 0$ such that if $T > T_\epsilon$,

$$R_T(\pi) = \frac{1}{T} \sum_{t=1}^{T} \inf_{\xi_t \in \Xi} (\xi_t - \xi_t^0)' \Sigma^{-1} (\xi_t - \xi_t^0) > \delta_\epsilon,$$

for all $\pi \in \Xi$ satisfying $|\pi_0 - \pi| > \epsilon$, where $\Xi_\pi = \{\xi : f(\xi, \pi) = 0\}$.

Theorem 1 is proved using a series of lemmas. The first lemma gives the consistency of $\hat{\pi}$. 

Lemma 1 As \( n \to \infty \) and \( T \to \infty \), \( \hat{\pi} \xrightarrow{P} \pi_0 \).

Proof: Let
\[
P(\pi) = \frac{1}{T} \sum_{t=1}^{T} \inf_{\xi_t, \epsilon_t} (Y_t - \xi_t)' \Sigma^{-1} (Y_t - \xi_t).\]

Then \( \hat{\pi} \) minimizes \( P(\pi) \), implying that
\[
P(\hat{\pi}) \leq P(\pi_0) \leq \frac{1}{T} \sum_{t=1}^{T} \epsilon_t' \Sigma^{-1} \epsilon_t = O_p \left( \frac{1}{n} \right).\]

Next, using Minkowski's Inequality with \( p = 2 \),
\[
(\xi_t - \xi_{t_0})' \Sigma^{-1} (\xi_t - \xi_{t_0}) \leq 4 \left[ (Y_t - \xi_t)' \Sigma^{-1} (Y_t - \xi_t) + \epsilon_t' \Sigma^{-1} \epsilon_t \right],
\]
so that
\[
R_T(\hat{\pi}) = \frac{1}{T} \sum_{t=1}^{T} \left[ \inf_{\xi_t, \epsilon_t} (\xi_t - \xi_{t_0})' \Sigma^{-1} (\xi_t - \xi_{t_0}) \right] \leq \frac{4}{T} \sum_{t=1}^{T} \left[ \inf_{\xi_t, \epsilon_t} \left( (Y_t - \xi_t)' \Sigma^{-1} (Y_t - \xi_t) + \epsilon_t' \Sigma^{-1} \epsilon_t \right) \right] = 8 \frac{T}{T} \sum_{t=1}^{T} \epsilon_t' \Sigma^{-1} \epsilon_t = O_p \left( \frac{1}{n} \right). \tag{A.3}
\]

Thus, as \( n \to \infty \) \( R_T(\hat{\pi}) \) can be made arbitrarily small and assumption (a) guarantees \( \hat{\pi} \xrightarrow{P} \pi_0 \).

The next lemma gives the consistency of \( \hat{\xi}_t \).

Lemma 2 For any \( \epsilon > 0 \), as \( n \to \infty \) and \( T \to \infty \),
\[
P(|\xi_t - \xi_0^t| < \epsilon \text{ for all } T) \to 0.
\]

Proof. By (A.3),
\[
TQ(\hat{\theta}, \hat{\gamma}) = O_p \left( \frac{T}{n} \right),
\]
and the result follows.

Given the consistency of \( \hat{\pi} \) and \( \hat{\xi}_t \), we can consider the lagrangian function

\[
Q(\xi_t, \pi, \alpha_t) = \sum_{i=1}^{T} \left[ (y_t - \xi_t)'\Sigma^{-1}(y_t - \xi_t) + \alpha_t'f(\xi_t, \pi) \right],
\]

where \( \alpha_t, t = 1, 2, \ldots, T \) are \( r \times 1 \) vectors of lagrangian multipliers used to incorporate the \( r \) restrictions, \( f(\xi_t, \pi) = 0 \), into the minimization problem. Differentiating (A.4) with respect to \( \xi_t, \pi \) and \( \alpha_t \), we obtain the derivative equations

\[
\begin{align*}
-\Sigma^{-1}(y_t - \hat{\xi}_t) + \hat{F}'_{\xi_t}\alpha_t &= 0, \\
f(\hat{\xi}_t, \hat{\pi}) &= 0, \\
\sum_{i=1}^{T} \hat{F}'_{\pi_t}\alpha_t &= 0,
\end{align*}
\]

where \((\hat{\xi}_t, \hat{\pi})\) minimizes (A.4), \( \hat{F}_{\xi_t} = \frac{\partial f(\xi_t, \pi)}{\partial \xi_t} |(\xi_t, \pi) \) and \( \hat{F}_{\pi_t} = \frac{\partial f(\xi_t, \pi)}{\partial \pi_t} |(\xi_t, \pi) \). By the mean value theorem, there exists a point \((\hat{\xi}_t, \hat{\pi})\) on the line segment joining \((\xi_t, \pi)\) and \((\hat{\xi}_t, \hat{\pi})\) such that

\[
f(\hat{\xi}_t, \hat{\pi}) = f(\xi_t, \pi) + \hat{F}_{\xi_t}(\hat{\xi}_t - \xi_t) + \hat{F}_{\pi_t}(\hat{\pi} - \pi),
\]

where \( \hat{F}_{\xi_t} = \frac{\partial f(\xi_t, \pi)}{\partial \xi_t} |(\hat{\xi}_t, \hat{\pi}) \) and \( \hat{F}_{\pi_t} = \frac{\partial f(\xi_t, \pi)}{\partial \pi_t} |(\hat{\xi}_t, \hat{\pi}) \).

Using (A.2) and (A.6), (A.8) simplifies to

\[
\hat{F}_{\xi_t}(\hat{\xi}_t - \xi_t) + \hat{F}_{\pi_t}(\hat{\pi} - \pi) = 0.
\]

Premultiplying (A.5) by \( \hat{F}_{\xi_t}\Sigma \), we obtain

\[
\hat{\Phi}_{\xi_t}\alpha_t = \hat{F}_{\xi_t}(y_t - \hat{\xi}_t),
\]

where \( \hat{\Phi}_{\xi_t} = \hat{F}_{\xi_t}\Sigma\hat{F}'_{\xi_t} \). Writing \( y_t - \hat{\xi}_t \) as

\[
y_t - \hat{\xi}_t = (y_t - \xi_t) - (\hat{\xi}_t - \xi_t) = \epsilon_t - (\hat{\xi}_t - \xi_t),
\]

\[
(A.11)
\]
we can express (A.10) as

\[ \Phi_{\xi t} \alpha_t = \tilde{F}_{\xi t} \epsilon_t - F_{\xi t} (\xi_t - \xi_t), \]

and, using (A.9) to obtain an expression for \( \tilde{F}_{\xi t} (\xi_t - \xi_t) \), it follows that

\[ \Phi_{\xi t} \alpha_t = \tilde{F}_{\xi t} \epsilon_t + F_{\pi t} (\pi_t - \pi). \]

Assuming \( F_{\xi t} \) is of rank \( r \) in a neighborhood of \( (\xi_t, \pi_t) \), then \( \Phi_{\xi t}^{-1} \) exists, and,

\[ \alpha_t = \Phi_{\xi t}^{-1} \left[ \tilde{F}_{\xi t} \epsilon_t + F_{\pi t} (\pi_t - \pi) \right]. \]  

(A.12)

We next investigate the order of \( \dot{\xi}_t - \xi_t \) and \( \pi_t - \pi_t \). Substituting (A.12) into (A.7) we obtain

\[ \sum_{t=1}^{T} \tilde{F}_{\pi t}^\prime \Phi_{\xi t}^{-1} \left[ \tilde{F}_{\xi t} \epsilon_t + F_{\pi t} (\pi_t - \pi) \right] = 0, \]

or, multiplying by \( \frac{1}{T} \),

\[ \bar{m}_{\pi v} (\pi_t - \pi) = \bar{m}_{\pi v}, \]

where \( \bar{m}_{\pi \pi} = \frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi t}^\prime \Phi_{\xi t}^{-1} F_{\pi t} \), and \( \bar{m}_{\pi v} = -\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi t}^\prime \Phi_{\xi t}^{-1} \tilde{F}_{\xi t} \epsilon_t \). Assuming condition (vi) of Theorem 1 holds and using Lemma 2, it follows that \( \bar{m}_{\pi \pi}^{-1} \) exists so that

\[ (\pi_t - \pi) = \bar{m}_{\pi \pi}^{-1} \bar{m}_{\pi v}. \]  

(A.13)

Expanding \( \bar{m}_{\pi \pi}^{-1} \) and \( \bar{m}_{\pi v} \) around \( (\xi_t, \pi_t) \) and using assumptions (ii) - (vi) of Theorem 1, we can show that \( \bar{m}_{\pi \pi}^{-1} = O_\phi (1) \) and \( \bar{m}_{\pi v} = O_\phi \left( \frac{1}{\sqrt{n}} \right) \). Thus, using (A.13), it follows that \( \pi_t - \pi = O_\phi \left( \frac{1}{\sqrt{n}} \right) \). To find the order of \( \dot{\xi}_t - \xi_t \), we use assumptions (ii) - (vi) of Theorem 1, as well as the order of \( \pi_t - \pi \) and (A.9), to show that \( \dot{\xi}_t - \xi_t = O_\phi \left( \frac{1}{\sqrt{n}} \right) \).

Now that we have crude estimates of the orders of \( \pi_t - \pi \) and \( \dot{\xi}_t - \xi_t \), we find can approximations for \( \pi_t - \pi \), and \( \dot{\xi}_t - \xi_t \). For this, we use a general lemma:
Lemma 3 Assume

(i) $g(\xi_t, \pi)$ is a function for which the first order derivatives with respect to $\xi_t$ and $\pi$ exist and are continuous in a neighborhood of $(\xi_t, \pi)$.

(ii) $\xi^*_t = \xi_t + O_p\left(a^{-1}_n\right)$, where $a_n \to \infty$ as $n \to \infty$.

(iii) $\pi^* = \pi + O_p\left(b^{-1}_n\right)$ where $b_n \to \infty$ as $n \to \infty$.

Then $g(\xi^*_t, \pi^*) = g(\xi_t, \pi) + O_p\left(\max[a^{-1}_n, b^{-1}_n]\right)$, or, equivalently, $g(\xi^*_t, \pi^*) - g(\xi_t, \pi) = O_p\left(\max[a^{-1}_n, b^{-1}_n]\right)$.

Proof: Expanding $g(\xi^*_t, \pi^*)$ around $(\xi_t, \pi)$ we can show that

$$g(\xi^*_t, \pi^*) = g(\xi_t, \pi) + \frac{\partial g(\xi_t, \pi)}{\partial \xi_t^*}(\xi^*_t - \xi_t) + \frac{\partial g(\xi_t, \pi)}{\partial \pi^*}(\pi^* - \pi) + O_p\left(\max[a^2_n, b^2_n, a_n, b_n]\right).$$

Using assumptions (ii) and (iii) of Theorem 1, $\frac{\partial g(\xi_t, \pi)}{\partial \xi^*_t}$ and $\frac{\partial g(\xi_t, \pi)}{\partial \pi^*}$ are bounded. By assumption (ii) and (iii) above, $\frac{\partial g(\xi_t, \pi)}{\partial \pi^*}(\pi^* - \pi) = O_p\left(b^{-1}_n\right)$ and $\frac{\partial g(\xi_t, \pi)}{\partial \xi^*_t}(\xi^*_t - \xi_t) = O_p\left(a^{-1}_n\right)$. Combining all terms of order $O_p\left(\max[a^{-1}_n, b^{-1}_n]\right)$ or higher into the remainder term, it follows that

$$g(\xi^*_t, \pi^*) = g(\xi_t, \pi) + O_p\left(\max[a^{-1}_n, b^{-1}_n]\right).$$

Lemma 4 Assume conditions (ii) and (iii) of Theorem 1 hold. Then,

$$\bar{F}_{xt} = F_{xt} + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (A.14)$$

$$\bar{F}_{xt} = F_{xt} + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (A.15)$$

$$\bar{F}_{xt} = F_{xt} + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (A.16)$$

$$\bar{F}_{xt} = F_{xt} + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (A.17)$$

$$\bar{F}_{xt} = F_{xt} + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (A.18)$$
**Proof:** Using the notation of Lemma 3 to prove (A.14), we let \( \xi_t^* = \hat{\xi}_t, \pi^* = \hat{\pi}, \) and \( g(\xi_t, \pi) = F_{\xi t}. \) By assumptions \((ii)\) and \((iii)\) of Theorem 1, condition \((i)\) of Lemma 3 is met. In (A.13) and discussions immediately following, we have shown that \( \hat{\pi} = \pi + O_p \left( \frac{1}{\sqrt{n}} \right) \) and \( \hat{\xi}_t = \xi_t + O_p \left( \frac{1}{\sqrt{n}} \right). \) From Lemmas 1, we know that, as \( n \to \infty, \) \( \hat{\pi} \to \pi \) and \( \hat{\xi}_t \to \xi_t \) for all \( t = 1, 2, \ldots, T. \) Since the point \((\hat{\xi}_t, \hat{\pi})\) lies on the line segment connecting \((\xi_t, \pi)\) and \((\hat{\xi}_t, \hat{\pi})\), it follows that \( \pi = \pi + O_p \left( \frac{1}{\sqrt{n}} \right) \) and \( \xi_t = \xi_t + O_p \left( \frac{1}{\sqrt{n}} \right) \) for all \( t = 1, 2, \ldots, T. \) Thus, by Lemma 3, (A.14) holds.

The proof of (A.15) mirrors the proof of (A.14). To prove (A.15), let \( \xi_t^* = \hat{\xi}_t \) and \( \pi^* = \hat{\pi}. \) Then, by the argument given above the result follows. The proofs of (A.16) - (A.17) are also similar to that of (A.14). For (A.16), we let \( \xi_t^* = \hat{\xi}_t, \pi^* = \hat{\pi} \) and \( g(\xi_t, \pi) = F_{\xi t}. \) For (A.17), define \( \xi_t^* = \hat{\xi}_t, \pi^* = \hat{\pi} \) and \( g(\xi_t, \pi) = F_{\xi t}. \) The result follows.

To show (A.18), we define

\[
\hat{R}_{\xi t} = \hat{F}_{\xi t} - F_{\xi t} = O_p \left( \frac{1}{\sqrt{n}} \right),
\]

\[
\hat{R}_{\xi t} = \hat{F}_{\xi t} - F_{\xi t} = O_p \left( \frac{1}{\sqrt{n}} \right).
\]

Then,

\[
\Phi_{\xi t} = \left( F_{\xi t} + \hat{R}_{\xi t} \right) \Sigma \left( F_{\xi t} + \hat{R}_{\xi t} \right)
\]

\[
= F_{\xi t} \Sigma F_{\xi t} + \hat{R}_{\xi t} \Sigma F_{\xi t} + F_{\xi t} \Sigma \hat{R}_{\xi t} + \hat{R}_{\xi t} \Sigma \hat{R}_{\xi t}.
\]

\[
= \Phi_{\xi t} + O_p \left( \frac{1}{\sqrt{n}} \right).
\]

Lemma 5 Let \( \Phi_{\xi t} = \Phi_{\xi t} + O_p \left( \frac{1}{\sqrt{n}} \right) \) where \( \Phi_{\xi t} \) is defined above and \( \Phi_{\xi t} = F_{\xi t} \Sigma F_{\xi t}. \)

Assume

(iv) \( F_{\xi t} \) has rank \( r \) in a neighborhood of \((\xi_t, \pi).\)
Then, $\Phi^{-1}_{\xi t}$ and $\Phi^{-1}_{\xi t}$ exist and
\[
\Phi^{-1}_{\xi t} = \Phi^{-1}_{\xi t} + O_p \left( \frac{1}{\sqrt{n}} \right).
\]  

**Proof:** Let $g(\xi_t, \pi) = \Phi^{-1}_{\xi t}$. By assumption (ii) of Theorem 1, the first derivatives of $g$ exist and are continuous in a neighborhood of $(\xi_t, \pi)$. From the proof of Lemma 4, we know that conditions (ii) and (iii) of Lemma 3 are met for $\xi_t^* = \xi_t$ and $\pi^* = \pi$. Thus, by Lemma 3, (A.19) holds.

**Lemma 6** Assume condition (vi) of Theorem 1 and Lemma 2 hold. Then, $m^{-1}_{\pi}$ exists and
\[
m^{-1}_{\pi} = m^{-1}_{\pi} + O_p \left( \frac{1}{\sqrt{n}} \right),
\]

where $m_{\pi} = \frac{1}{T} \sum_{t=1}^{T} F'_{\pi t} \Phi^{-1}_{\xi t} F_{\pi t}$.

**Proof:**

To show (A.20), we define $g(\xi_t, \pi) = m^{-1}_{\pi}$, $\tilde{R}_{\Phi_t} = \Phi^{-1}_{\xi t} - \Phi^{-1}_{\xi t} = O_p \left( \frac{1}{\sqrt{n}} \right)$, $\tilde{R}_{\pi} = \tilde{F}_{\pi t} - F_{\pi t} = O_p \left( \frac{1}{\sqrt{n}} \right)$, and $\tilde{R}_{\pi} = \tilde{F}_{\pi t} - F_{\pi t} = O_p \left( \frac{1}{\sqrt{n}} \right)$. Again by assumption (ii) of Theorem 1, the first derivatives of $g$ exist and are continuous in a neighborhood of $(\xi_t, \pi)$. Then, using the results of Lemma 2, it follows that
\[
m_{\pi} = \frac{1}{T} \sum_{t=1}^{T} \left[ F_{\pi t} + \tilde{R}_{\pi} \right]' \left[ \Phi^{-1}_{\xi t} + \tilde{R}_{\Phi t} \right] \left[ F_{\pi t} + \tilde{R}_{\pi} \right]
\]
\[
\quad = m_{\pi} + O_p \left( \frac{1}{\sqrt{n}} \right).
\]
By assumption (vi) of Theorem 1, $m_{\pi}$ is of full rank and $m^{-1}_{\pi}$ exists. It follows from Lemma 3 that (A.20) holds.

Next, we consider an approximation to $\tilde{m}_{\pi t}$. Using (A.15), (A.17) and (A.19), we see that
\[
\tilde{m}_{\pi t} = -\frac{1}{T} \sum_{t=1}^{T} \left[ F_{\pi t} + \tilde{R}_{\pi} \right]' \left[ \Phi^{-1}_{\xi t} + \tilde{R}_{\Phi t} \right] \left[ F_{\pi t} + \tilde{R}_{\pi} \right] \epsilon_{\pi t}
\]
\[ \hat{\pi} - \pi = m_\pi^{-1} m_{\pi\nu} + O_p \left( \frac{1}{n} \right) \]

Thus, using (A.20) and (A.21) in (A.13), we can show that

\[ \hat{\pi} - \pi = \beta_1 + O_p \left( \frac{1}{n} \right), \]

where \( \beta_1 = m_\pi^{-1} m_{\pi\nu} = O_p \left( \frac{1}{\sqrt{n}T} \right) \).

To find an approximation for \( \hat{\xi}_t - \xi_t \), we substitute (A.12) into (A.5) to obtain

\[ -\Sigma^{-1}(y_t - \hat{\xi}_t) + F'_{\xi_t} \Phi_{\xi_t}^{-1} \left[ F_{\xi_t} \epsilon_t + \bar{F}_{\pi t}(\hat{\pi} - \pi) \right] = 0. \]  

Using (A.11), substituting (A.22) into (A.23), and solving for \( \hat{\xi}_t - \xi_t \), we see that

\[ \hat{\xi}_t - \xi_t = \epsilon_t - \Sigma F'_{\xi_t} \Phi_{\xi_t}^{-1} [F_{\xi_t} \epsilon_t + \bar{F}_{\pi t} \beta_1] + O_p \left( \frac{1}{n} \right). \]

Substituting (A.14), (A.16), (A.17) and (A.19) into (A.24) and combining all terms of order \( O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right) \) or higher into the remainder term, it can be shown that

\[ \hat{\xi}_t - \xi_t = \epsilon_t - \Sigma F'_{\xi_t} \Phi_{\xi_t}^{-1} F_{\xi_t} \epsilon_t + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right) \]

\[ = d_{\xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right), \]

where \( d_{\xi t} = (I_p - \Sigma F'_{\xi_t} \Phi_{\xi_t}^{-1} F_{\xi_t}) \epsilon_t = O_p \left( \frac{1}{\sqrt{n}} \right) \).

We next consider higher order approximations for \( \hat{\xi}_t - \xi_t \) and \( \hat{\pi} - \pi \). Expanding \( f(\hat{\xi}_t, \hat{\pi}) \) around \((\xi_t, \pi)\) up to the order \( O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right) \), we get

\[ f(\hat{\xi}_t, \hat{\pi}) = f(\xi_t, \pi) + F_{\xi t} (\hat{\xi}_t - \xi_t) + F_{\pi t} (\hat{\pi} - \pi) \]

\[ + \frac{1}{2} \begin{bmatrix} (\hat{\xi}_t - \xi_t)' A_{1t}(\hat{\xi}_t - \xi_t) \\ (\hat{\xi}_t - \xi_t)' A_{2t}(\hat{\xi}_t - \xi_t) \\ \vdots \\ (\hat{\xi}_t - \xi_t)' A_{rt}(\hat{\xi}_t - \xi_t) \end{bmatrix} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \]
where $A_{it} = \frac{\partial^2 f_i(\xi, \pi)}{\partial \xi_i \partial \xi_i}, i = 1, 2, \ldots, r$ is a $p \times p$ matrix of second partial derivatives of $f_i(\xi, \pi)$ with respect to $\xi_i$. Substituting (A.2), (A.6) and (A.25) into (A.26), we obtain

$$0 = F_{it}(\hat{\xi}_t - \xi_t) + F_{\pi t}(\hat{\pi} - \pi) + c_t + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \quad \text{(A.27)}$$

where $c_t = \frac{1}{2} \left[ d'_{it} A_{1t} d_{1t}, d'_{it} A_{2t} d_{2t}, \ldots, d'_{it} A_{rt} d_{rt} \right]' = O_p \left( \max \left[ \frac{1}{n}, \frac{1}{nT} \right] \right)$. Using (A.11) and (A.27), we can show that

$$F_{it}(y_t - \hat{\xi}_t) = v_t - F_{it} \left( \hat{\xi}_t - \xi_t \right) = v_{\xi t} + F_{\pi t}(\hat{\pi} - \pi) + c_t + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \quad \text{(A.28)}$$

where $v_{\xi t} = F_{it} v_t = O_p \left( \frac{1}{\sqrt{n}} \right)$. Premultiplying (A.5) by $F_{it} \Sigma$, we obtain

$$\hat{\Phi}_{it} \alpha_t = F_{it}(y_t - \hat{\xi}_t),$$

where $\hat{\Phi}_{it} = F_{it} \Sigma \hat{F}'_{it}$. Using assumption (iv) of Lemma 5, it follows that $\hat{\Phi}_{it}$ is positive definite, $\hat{\Phi}_{it}^{-1}$ exists in a neighborhood of $(\xi_t, \pi)$, and

$$\alpha_t = \hat{\Phi}_{it}^{-1} F_{it}(y_t - \hat{\xi}_t). \quad \text{(A.29)}$$

Substituting (A.28) into (A.29), it can be shown that

$$\alpha_t = \hat{\Phi}_{it}^{-1} \left[ v_{\xi t} + F_{\pi t}(\hat{\pi} - \pi) + c_t \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \quad \text{(A.30)}$$

Putting (A.30) into (A.7), and multiplying by $\frac{1}{T}$, we see that

$$0 = \frac{1}{T} \sum_{i=1}^{T} \hat{F}'_{\pi t} \hat{\Phi}_{it}^{-1} \left[ v_{\xi t} + F_{\pi t}(\hat{\pi} - \pi) + c_t \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),$$

or,

$$\hat{m}_{\pi t}(\hat{\pi} - \pi) = \hat{m}_{\pi v} + \hat{m}_{\pi c} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),$$

where $\hat{m}_{\pi v} = \frac{1}{T} \sum_{i=1}^{T} \hat{F}'_{\pi t} \hat{\Phi}_{it}^{-1} F_{\pi t}$, $\hat{m}_{\pi c} = -\frac{1}{T} \sum_{i=1}^{T} \hat{F}'_{\pi t} \hat{\Phi}_{it}^{-1} v_{\xi t}$, and

$$\hat{m}_{\pi c} = -\frac{1}{T} \sum_{i=1}^{T} \hat{F}'_{\pi t} \hat{\Phi}_{it}^{-1} c_t. \quad \text{Using condition (vi) of Theorem 1, } \hat{m}_{\pi v}^{-1} \text{ exists and,}$$

$$\hat{\pi} - \pi = \hat{m}_{\pi v}^{-1} \left[ \hat{m}_{\pi v} + \hat{m}_{\pi c} \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \quad \text{(A.31)}$$
To find an expansion for (A.31), we expand each element around \((\xi_t, \pi)\). Using (A.16) and assumption \((iv)\) of Lemma 5, we can show that

\[
\Phi_{\xi t}^{-1} = \Phi_{\xi t}^{-1} + O_p \left( \frac{1}{\sqrt{n}} \right).
\]

(A.32)

Then, using (A.15), (A.16), (A.32) and condition \((vi)\) of Theorem 1, it can be shown that

\[
\dot{m}_{\pi \pi} = m_{\pi \pi} + O_p \left( \frac{1}{\sqrt{n}} \right),
\]

(A.33)

\[
\dot{m}_{\pi \pi}^{-1} = m_{\pi \pi}^{-1} + O_p \left( \frac{1}{\sqrt{n}} \right),
\]

(A.34)

\[
\dot{m}_{\pi \pi} = m_{\pi \pi} + O_p \left( \frac{1}{n \sqrt{n}} \right),
\]

(A.35)

where \(m_{\pi \pi} = -\frac{1}{T} \sum_{t=1}^{T} F'_{\pi t} \Phi_{\xi t}^{-1} c_t\).

To find an approximation for \(\dot{m}_{\pi \pi}\), we consider higher order approximations for \(\dot{F}_{\pi t}\), \(\dot{F}_{\xi t}\) and \(\Phi_{\xi t}^{-1}\). Expanding \(\dot{F}_{\xi t}\) around \((\xi_t, \pi)\), we obtain

\[
\dot{F}_{\xi t} = F_{\xi t} + \left[ (\dot{\xi}_t - \xi_t)' A_{1t} \right] + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right).
\]

(A.36)

Substituting (A.25) into (A.36), we can show that

\[
\dot{F}_{\xi t} = F_{\xi t} + G_{\xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right),
\]

(A.37)

where

\[
G_{\xi t} = \begin{bmatrix}
d'_{r t} A_{1t} \\
d'_{r t} A_{2t} \\
\vdots \\
d'_{r t} A_{rt}
\end{bmatrix} = O_p \left( \frac{1}{\sqrt{n}} \right).
\]
Premultiplying the transpose of (A.37) by \( F^T \Sigma \), we find that
\[
\hat{\Phi}_{\xi t} = \Phi_{\xi t} + F^T \Sigma G'_t \Phi_{\xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right).
\]
Assuming (iv) of Lemma 5 holds, \( \hat{\Phi}_{\xi t} \) is positive definite and \( \hat{\Phi}_{\xi t}^{-1} \) exists in a neighborhood of \((\xi_t, \pi)\) so that,
\[
\hat{\Phi}_{\xi t}^{-1} = \Phi_{\xi t}^{-1} - \Phi_{\xi t}^{-1} F^T \Sigma G'_t \Phi_{\xi t}^{-1} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right).
\] (A.38)

Next, expanding \( \hat{F}_{\pi t} \) around \((\xi_t, \pi)\), we see that
\[
\hat{F}_{\pi t} = F_{\pi t} + \left[ \begin{array}{c} \frac{\partial^2 f_1(\xi_t, \pi)}{\partial \xi_t \partial \xi_t^T} \\
\frac{\partial^2 f_2(\xi_t, \pi)}{\partial \xi_t \partial \xi_t^T} \\
\vdots \\
\frac{\partial^2 f_k(\xi_t, \pi)}{\partial \xi_t \partial \xi_t^T} \end{array} \right] \left( \xi_t - \xi_t \right) + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right) + \frac{\partial^2 f_1(\xi_t, \pi)}{\partial \xi_t^2 \partial \xi_t} \left( \xi_t - \xi_t \right) + \frac{\partial^2 f_2(\xi_t, \pi)}{\partial \xi_t^2 \partial \xi_t} \left( \xi_t - \xi_t \right). \] (A.39)

Substituting (A.25) into (A.39) and combining all terms of order \( O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right) \) or higher into the remainder term, we can show that
\[
\hat{F}_{\pi t} = F_{\pi t} + U_{\pi \xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right), \] (A.40)

where
\[
U_{\pi \xi t} = \left[ \begin{array}{c} d'_{\xi t} \frac{\partial^2 f_1(\xi_t, \pi)}{\partial \xi_t \partial \xi_t^T} \\
\vdots \\
d'_{\pi t} \frac{\partial^2 f_k(\xi_t, \pi)}{\partial \xi_t \partial \xi_t^T} \end{array} \right] = O_p \left( \frac{1}{\sqrt{n}} \right) + \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{n}} \right) + \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{nT}} \right) \] (A.41)

Using the expansion (A.38) and (A.40) in \( \bar{m}_{\pi \nu} \) and combining all terms of order \( O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right) \) or higher into the remainder term, it can be shown that
\[
\bar{m}_{\pi \nu} = -\frac{1}{T} \sum_{t=1}^{T} \left[ F_{\pi t} + U_{\pi \xi t} \right] \left[ \Phi_{\xi t}^{-1} - \Phi_{\xi t}^{-1} F^T \Sigma G'_t \Phi_{\xi t}^{-1} \right] v_{\xi t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right)
= -\frac{1}{T} \sum_{t=1}^{T} F_{\pi t} \hat{\Phi}_{\xi t}^{-1} v_{\xi t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right)
= m_{\pi \nu} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \] (A.41)
Thus, combining the results of (A.34), (A.35) and (A.41) into (A.31), we obtain

\[
\tilde{\pi} - \pi = m_{\pi\pi}^{-1} \left[ m_{\pi v} + m_{\pi c} \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right)
\]

\[
= m_{\pi\pi}^{-1} m_{\pi v} + m_{\pi\pi}^{-1} m_{\pi c}
\]

\[
= \beta_1 + \beta_2 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
\]

(A.42)

where \( \beta_2 = m_{\pi\pi}^{-1} m_{\pi c} = O_p \left( \frac{1}{n} \right) \) and \( m_{\pi v} = -\frac{1}{T} \sum_{t=1}^{T} F'_{\pi t} \Phi_{\xi t}^{-1} c_t = O_p \left( \frac{1}{n} \right) \).

To derive a higher order approximation for \( \xi_t - \xi_i \), we substitute (A.42) into (A.30) to obtain

\[
\alpha_t = \Phi_{\xi t}^{-1} \left[ v_{\xi t} + F_{\pi t} (\beta_1 + \beta_2) + c_t \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right).
\]

(A.43)

Substituting (A.37), (A.38) and (A.43) into (A.5), using (A.11) and solving for \( \xi_t - \xi_i \), we can show that

\[
\dot{\xi}_t - \xi_t = \epsilon_t - \Sigma [F_{\xi t} + G_{\xi t}] \left[ \Phi_{\xi t}^{-1} - \Phi_{\xi t}^{-1} F_{\pi t} \Sigma G'_{\xi t} \Phi_{\xi t}^{-1} \right] \times \left[ v_{\xi t} + c_t + F_{\pi t} (\beta_1 + \beta_2) \right] + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right)
\]

\[
= d_{1t} + l_{2t} + l_{3t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
\]

(A.44)

where \( l_{2t} = -\Sigma F'_{\xi t} \Phi_{\xi t}^{-1} F_{\pi t} \beta_1 = O_p \left( \frac{1}{\sqrt{nT}} \right) \) and

\[
l_{3t} = -\Sigma F'_{\xi t} \Phi_{\xi t}^{-1} [c_t + F_{\pi t} \beta_2] - n \Sigma V_{dt} G'_{\xi t} \Phi_{\xi t}^{-1} v_{\xi t} = O_p \left( \frac{1}{n} \right).
\]

Now that we have obtained expansions for \( \dot{\xi}_t - \xi_t \) and \( \dot{\pi} - \pi \), we show that the expansions for the original unknown parameters are those given in Theorem 1. To do this, recall that the reparameterization was \( \pi = \begin{pmatrix} \gamma \\ \theta \end{pmatrix} \) and \( \xi_t = \mu_t + \gamma \). Hence, if we obtain an expansion of \( \dot{\gamma} - \gamma \) from \( \dot{\pi} - \pi \), then we have an expansion for \( \dot{\mu}_t - \mu_t = (\dot{\xi}_t - \xi_t) - (\dot{\gamma} - \gamma) \).
By making the appropriate substitutions and taking derivatives using the chain rule, we can show that

\[
\begin{align*}
F_{\xi t} &= F_{\mu t}, \quad (A.45) \\
F_{\gamma t} &= -F_{\mu t}, \\
F_{\pi t} &= [F_{\gamma t}, F_{\theta t}] \\
&= [-F_{\mu t}, F_{\theta t}]. \quad (A.46)
\end{align*}
\]

Using (A.45) - (A.46), the approximation given in (A.42) for \( \hat{\pi} - \pi \) can be expressed in terms of \( \hat{\mu}_t - \mu_t, \hat{\gamma} - \gamma \) and \( \hat{\theta} - \theta \) as

\[
\begin{pmatrix}
\hat{\gamma} - \gamma \\
\hat{\theta} - \theta
\end{pmatrix} = \begin{pmatrix}
m_{\mu\mu} & m_{\mu\theta} \\
m_{\theta\mu} & m_{\theta\theta}
\end{pmatrix}^{-1} \begin{pmatrix}
m_{\mu \nu} + m_{\mu \epsilon} + R_{\mu \nu} \\
m_{\theta \nu} + m_{\theta \epsilon} + R_{\theta \nu}
\end{pmatrix}, \quad (A.47)
\]

where \( R_{\mu \nu} = O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right), R_{\theta \nu} = O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right), v_t = F_{\mu t} \epsilon_t, m_{\mu\mu} = \frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi_{\mu t}^{-1} F_{\mu t}, m_{\mu\theta} = -\frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi_{\mu t}^{-1} F_{\theta t}, m_{\theta\mu} = m'_{\mu\theta}, m_{\theta\theta} = \frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} F_{\theta t}, m_{\mu\nu} = \frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi_{\mu t}^{-1} v_t, m_{\mu\epsilon} = \frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi_{\mu t}^{-1} c_t, m_{\theta\nu} = -\frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} v_t, \text{ and } m_{\theta\epsilon} = -\frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} c_t.

Solving for \( \hat{\theta} - \theta \) and \( \hat{\gamma} - \gamma \) in (A.47), it can be shown that

\[
\hat{\theta} - \theta = q^{-1}_{\theta, \mu} \left[ m_{\theta \nu} + m_{\theta \epsilon} - m_{\theta \mu} m_{\mu \nu}^{-1} (m_{\mu \nu} + m_{\mu \epsilon}) \right] + O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right), \quad (A.48)
\]

where \( q_{\theta, \mu} = m_{\theta \mu} - m_{\theta \epsilon} m_{\mu \nu}^{-1} m_{\mu \epsilon} \), and

\[
\hat{\gamma} - \gamma = m_{\mu \nu}^{-1} \left[ m_{\mu \nu} + m_{\mu \epsilon} - m_{\mu \theta} (\hat{\theta} - \theta) \right] + O_p \left( \max \left[ \frac{1}{n \sqrt{n}}, \frac{1}{n \sqrt{T}} \right] \right). \quad (A.49)
\]

Defining

\[
\tau_1 = q^{-1}_{\theta, \mu} \left[ m_{\theta \nu} - m_{\theta \mu} m_{\mu \nu}^{-1} m_{\mu \nu} \right] = O_p \left( \frac{1}{\sqrt{nT}} \right)
\]

and

\[
\tau_2 = q^{-1}_{\theta, \mu} \left[ m_{\theta \epsilon} - m_{\theta \mu} m_{\mu \nu}^{-1} m_{\mu \epsilon} \right] = O_p \left( \frac{1}{n} \right),
\]
(A.48) can be written as
\[ \dot{\theta} - \theta = \tau_1 + \tau_2 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \] (A.50)

Letting
\[ \Lambda_1 = m_{\mu\mu}^{-1} \left[ m_{\mu\nu} - m_{\mu\theta} \tau_1 \right] = O_p \left( \frac{1}{\sqrt{n}} \right) \]
and
\[ \Lambda_2 = m_{\mu\mu}^{-1} \left[ m_{\mu\nu} - m_{\mu\theta} \tau_2 \right] = O_p \left( \frac{1}{n} \right), \]
we can express (A.49) as
\[ \dot{\gamma} - \gamma = \Lambda_1 + \Lambda_2 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \] (A.51)

Now that we have expressions (A.50) and (A.51) we look for an expression for \( \dot{\mu}_t - \mu_t \).

First, considering \( l_{2t} \) and \( l_{3t} \),
\[
l_{2t} = -\Sigma F'_{\xi_t} \Phi^{-1}_{\xi_t} F_{\pi_t} \beta_1
= -\Sigma F'_{\mu\mu} \Phi^{-1}_{\mu\mu} \left[ -F_{\mu\mu} \Phi_{\mu\mu} \right]
= -\Sigma F'_{\mu\mu} \Phi^{-1}_{\mu\mu} \left[ -F_{\mu\mu} \Phi_{\mu\mu} \right].
\] (A.52)

Next,
\[
l_{3t} = -\Sigma F'_{\xi_t} \Phi^{-1}_{\xi_t} [c_t + F_{\pi_t} \beta_2] - n V_{d_{1t}} G'_{\xi_t} \Phi^{-1}_{\xi_t} v_t
= -\Sigma F'_{\mu\mu} \Phi^{-1}_{\mu\mu} \left[ c_t + [-F_{\mu\mu}, F_{\theta t}] \left( m_{\mu\mu} \ m_{\mu\theta} \right)^{-1} \left( m_{\mu\nu} \ m_{\theta\theta} \right) \right]
= -\Sigma F'_{\mu\mu} \Phi^{-1}_{\mu\mu} \left[ c_t - F_{\mu\mu} \Lambda_2 + F_{\theta t} \tau_2 \right] - n V_{d_{1t}} G'_{\mu\mu} \Phi^{-1}_{\mu\mu} v_t. \] (A.53)

Thus, using (A.52), (A.53) in (A.44), and noting that \( \dot{\mu}_t - \mu_t = \dot{\xi}_t - \xi - (\dot{\gamma} - \gamma) \), we can show that
\[ \dot{\mu}_t - \mu_t = d_{1t} + d_{2t} + d_{3t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right), \] (A.54)
where
\[ d_{2t} = - \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_1 - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\theta t} \tau_1 \]
and
\[ d_{3t} = - \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_2 - \sum F'_{\mu t} \Phi^{-1}_{\mu t} \left[ c_t + F_{\theta t} \tau_2 \right] - n V_{d_{1t}} G'_{t} \Phi^{-1}_{\mu t} v_t. \]

To complete the proof of Theorem 1, we find the expectation of \( d_{1t} - d_{3t}, \tau_1, \tau_2, \Lambda_1 \) and \( \Lambda_2. \) First,
\[
E(d_{1t}) = E \left[ \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \varepsilon_t \right] = \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E[\varepsilon_t] = 0. \tag{A.55}
\]

Next, in finding \( E(\tau_1), \) we note that \( q_{\theta \theta, \mu}, m_{\theta \mu} \) and \( m_{\mu \mu}^{-1} \) are constant. Thus,
\[
E(\tau_1) = E \left[ q_{\theta \theta, \mu}^{-1} \left( m_{\theta \mu} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu \mu} \right) \right] = q_{\theta \theta, \mu}^{-1} E(\theta_{\theta \mu}) - m_{\theta \mu} m_{\mu \mu}^{-1} E(\mu_{\mu \mu}) = 0, \tag{A.56}
\]
since
\[
E(\theta_{\theta \mu}) = E \left[ \frac{1}{T} \sum_{t=1}^{T} F'_{\theta t} \Phi^{-1}_{\mu t} v_t \right] = \frac{1}{T} \sum_{t=1}^{T} F'_{\theta t} \Phi^{-1}_{\mu t} E[v_t] = 0,
\]
and
\[
E(\mu_{\mu \mu}) = E \left[ \frac{1}{T} \sum_{t=1}^{T} F'_{\mu t} \Phi^{-1}_{\mu t} v_t \right] = \frac{1}{T} \sum_{t=1}^{T} F'_{\mu t} \Phi^{-1}_{\mu t} E[v_t] = 0. \tag{A.57}
\]
Also,

\[
E(\tau_2) = \mathbb{E}\left[ q_{\hat{\theta}, \mu}^{-1} \left( m_{\hat{\theta}c} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu c} \right) \right]
\]

\[
= q_{\hat{\theta}, \mu}^{-1} \mathbb{E}\left[ m_{\hat{\theta}c} \right] - m_{\theta \mu} m_{\mu \mu}^{-1} \mathbb{E}\left[ m_{\mu c} \right]
\]

\[
= q_{\hat{\theta}, \mu}^{-1} \left( -\frac{1}{T} \sum_{t=1}^{T} F'_{\theta} \Phi^{-1}_{\mu} \mathbb{E}[c_t] - m_{\theta \mu} m_{\mu \mu}^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} F'_{\mu} \Phi^{-1}_{\mu} \mathbb{E}[c_t] \right) \right) \quad \text{(A.58)}
\]

To find \( E[c_t] \), we note that the \( i \)th element in the vector \( c_t \) is quadratic in \( d_{1t} \), i.e.,

\[ c_i = \frac{1}{2} d_{1i}' A_{ii} d_{1t}. \]

Taking the expected value of \( c_{it} \), it can be shown that

\[
E(c_{it}) = \mathbb{E}\left( \frac{1}{2} d_{1i}' A_{ii} d_{1t} \right)
\]

\[
= \frac{1}{2} \text{tr} \left[ A_{ii} \text{Var}(d_{1t}) \right] + \frac{1}{2} E[d_{1t}]' A_{ii} E[d_{1t}]
\]

\[
= \frac{1}{2} \text{tr} \left[ A_{ii} \text{Var}(d_{1t}) \right],
\]

since \( E[d_{1t}] = 0 \). Thus,

\[
E[c_{it}] = \frac{1}{2} \begin{bmatrix}
\text{tr}(A_{i1} \text{Var}(d_{1t})) \\
\text{tr}(A_{i2} \text{Var}(d_{1t})) \\
\vdots \\
\text{tr}(A_{iT} \text{Var}(d_{1t}))
\end{bmatrix},
\]

where \( \text{V}_d_{1t} = \text{Var}(d_{1t}) \), and

\[
\text{Var}(d_{1t}) = \text{Var} \left[ \left( I_p - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \right) \varepsilon_t \right]
\]

\[
= \left( I_p - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \right) \text{Var}(\varepsilon_t) \left( I_p - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \right)'
\]

\[
= \frac{1}{n} \left( I_p - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \right) \Sigma \left( I_p - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \right)'
\]

\[
= \frac{1}{n} \left[ \Sigma - \Sigma F'_{\mu} \Phi^{-1}_{\mu} F_{\mu} \Sigma \right]. \quad \text{(A.59)}
\]

Defining \( B(\mu, \theta) = E[c_t] \), we see that (A.58) simplifies to

\[
E(\tau_2) = q_{\hat{\theta}, \mu}^{-1} \left[ m_{\hat{\theta}B} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu B} \right],
\]
where $m_{\theta B} = - \frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi^{-1}_{\mu t} B(\mu_t, \theta)$ and $m_{\mu B} = \frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi^{-1}_{\mu t} B(\mu_t, \theta)$. Next, we consider $E(A_1)$ and $E(A_2)$. First,

$$E(A_1) = E \left[ m^{-1}_{\mu t} \left( m_{\mu \nu} - m_{\mu \theta} \tau_1 \right) \right] = m^{-1}_{\mu t} \left[ E(m_{\mu \nu}) - m_{\mu \theta} E(\tau_1) \right] = 0,$$

by (A.56) and (A.57). Also,

$$E(A_2) = E \left[ m^{-1}_{\mu t} \left( m_{\mu c} - m_{\mu \theta} \tau_2 \right) \right] = m^{-1}_{\mu t} \left( E(m_{\mu c}) - m_{\mu \theta} E(\tau_2) \right)$$

$$= m^{-1}_{\mu t} \left[ m_{\mu B} - m_{\mu \theta} q_{\theta B}^{-1} \left( m_{\theta B} - m_{\theta \mu} m_{\mu B} \right) \right]. \quad (A.60)$$

To find an expression for $E(d_{2t})$, we note that

$$E(d_{2t}) = E \left[ - \left( I_p - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_1 - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\theta t} \tau_1 \right]$$

$$= - \left( I_p - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E(A_1) - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\theta t} E(\tau_1)$$

$$= 0. \quad (A.61)$$

Next,

$$E(d_{3t}) = E \left[ - \left( I_p - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Lambda_2 - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} \left[ c_t + F_{\theta t} \tau_2 \right] - nV_{d_{3t}} \frac{G'_{\mu t} \Phi^{-1}_{\mu t} v_t}{v_t} \right]$$

$$= - \left( I_p - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E(A_2) - \Sigma F_{\mu t} \Phi^{-1}_{\mu t} \left[ E(c_t) + F_{\theta t} E(\tau_2) \right]$$

$$- nV_{d_{3t}} E \left[ G'_{\mu t} \Phi^{-1}_{\mu t} v_t \right], \quad (A.62)$$

where an expression for $E(\tau_2)$ and $E(A_2)$ are given in (A.58) and (A.60), respectively. Looking at $E \left[ G'_{\mu t} \Phi^{-1}_{\mu t} v_t \right]$, we write $\Phi^{-1}_{\mu t} = (s_{1t}', s_{2t}', \ldots, s_{rt}')$, where for all $i = 1, 2, \ldots, r$, $s_{it}$ is a $1 \times r$ vector corresponding to the $i^{th}$ row of $\Phi^{-1}_{\mu t}$. Using this notation it can be shown that

$$G'_{\mu t} \Phi^{-1}_{\mu t} v_t = \sum_{i=1}^{r} A_{it} d_{3t} s_{it} v_t. \quad (A.63)$$
Thus,

\[
E \left[ G'_t \Phi^{-1}_{\mu t} v_t \right] = E \left[ \sum_{i=1}^r A_i d_{it} s_{it} v_t \right] \\
= \sum_{i=1}^r A_i E \left[ d_{it} s_{it} v_t \right] \\
= \sum_{i=1}^r A_i E \left[ d_{it} v_t' \right] s'_t \\
= \sum_{i=1}^r A_i \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E \left[ \varepsilon_t \varepsilon_t' \right] F'_{\mu t} s'_t \\
= \frac{1}{n} \sum_{i=1}^r A_i \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Sigma F'_{\mu t} s'_t \\
= 0, \quad (A.64)
\]

since

\[
\left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \Sigma F'_{\mu t} = 0. \quad (A.65)
\]

Using (A.64) in (A.62) above, it follows that

\[
E(d_{yt}) = - \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E(\Lambda_2) - \sum F'_{\mu t} \Phi^{-1}_{\mu t} [B(\mu_t, \theta) + F'_{\theta t} E(\tau_2)].
\]

Proof of Theorem 2

To prove Theorem 2, we reparameterize the model as

\[
Y'_t = \xi_t + \varepsilon_t \\
f(\xi_t, \pi) = 0, \quad (A.66)
\]

where \(Y'_t = Y_t + \Delta_t\) and \(\pi\) is defined in the proof of Theorem 1. Since \(\Delta_t = O_p \left( \frac{1}{n} \right)\), and since this minimization has the same form as that in Theorem 1, the consistency results corresponding to Lemmas 1 and 2 follow. The problem of minimizing (19) subject to \(f(\mu_t, \theta) = 0\) is equivalent to minimizing over \(\xi_t, \pi\) and \(\alpha_t\), the lagrangian function

\[
Q(\xi_t, \pi, \alpha_t) = \sum_{t=1}^T \left( (y_t^* - \xi_t)' \Sigma^{-1}(y_t^* - \xi_t) + \alpha_t f(\xi_t, \pi) \right). \quad (A.67)
\]
Differentiating (A.67) with respect to $\xi_t$, $\pi$ and $\alpha_t$, we obtain the derivative equations

$$ -\Sigma^{-1}(y_t - \tilde{\xi}_t) + \tilde{F}'_\xi \alpha_t = 0, \quad (A.68) $$

$$ f(\tilde{\xi}_t, \tilde{\pi}) = 0, \quad (A.69) $$

$$ \sum_{t=1}^{T} \tilde{F}'_{\pi t} \alpha_t = 0, \quad (A.70) $$

where $(\tilde{\xi}_t, \tilde{\pi})$ is the value of $(\xi_t, \pi)$ that minimizes (A.67). There exists a point $(\tilde{\xi}_t, \tilde{\pi})$ on the plane connecting $(\xi_t, \pi)$ and $(\tilde{\xi}_t, \tilde{\pi})$ such that

$$ f(\tilde{\xi}_t, \tilde{\pi}) = f(\xi_t, \pi) + \tilde{F}'_{\xi t}(\tilde{\xi}_t - \xi_t) + \tilde{F}'_{\pi t}(\tilde{\pi} - \pi), $$

where $\tilde{F}'_{\xi t} = \frac{\partial f(\xi_t, \pi)}{\partial \xi_t}(\tilde{\xi}_t, \tilde{\pi})$ and $\tilde{F}'_{\pi t} = \frac{\partial f(\xi_t, \pi)}{\partial \pi}(\tilde{\xi}_t, \tilde{\pi})$. Following the steps outlined in (A.9) - (A.13) of the proof of Theorem 1, we can show that

$$ \tilde{\pi} - \pi = \tilde{m}_{\pi \pi}^{-1} \tilde{m}_{\pi \nu}, \quad (A.71) $$

where $\tilde{m}_{\pi \pi} = \frac{1}{T} \sum_{t=1}^{T} \tilde{F}'_{\pi t} \tilde{F}_{\xi t}$ and $\tilde{m}_{\pi \nu} = -\frac{1}{T} \sum_{t=1}^{T} \tilde{F}'_{\pi t} \tilde{F}_{\xi t} \xi_t$. Using (A.71) and conditions (ii) - (vi) of Theorem 1, it follows that $\tilde{\pi} - \pi = O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\tilde{\xi}_t - \xi_t = O_p\left(\frac{1}{\sqrt{n}}\right)$. Continuing to follow the steps of the proof of Theorem 1, we obtain expansions for $\tilde{\pi} - \pi$ and $\tilde{\xi}_t - \xi_t$ given by

$$ \tilde{\xi}_t - \xi_t = d_{1t} + O_p\left(\frac{1}{n}\right), \quad (A.72) $$

$$ \tilde{\pi} - \pi = \beta_1 + O_p\left(\frac{1}{n}\right). $$

Next, we find higher order approximations for both $\tilde{\pi} - \pi$ and $\tilde{\mu}_t - \mu_t$. Expanding $f(\tilde{\xi}_t, \tilde{\pi})$ around $(\xi_t, \pi)$ up to the order $O_p\left(\max\left[\frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}}\right]\right)$, we obtain

$$ f(\tilde{\xi}_t, \tilde{\pi}) = f(\xi_t, \pi) + F_{\xi t}(\tilde{\xi}_t - \xi_t) + F_{\pi t}(\tilde{\pi} - \pi) + \frac{1}{2} \begin{bmatrix} (\tilde{\xi}_t - \xi_t)' A_{11}(\tilde{\xi}_t - \xi_t) \\ (\tilde{\xi}_t - \xi_t)' A_{21}(\tilde{\xi}_t - \xi_t) \\ \vdots \\ (\tilde{\xi}_t - \xi_t)' A_{n1}(\tilde{\xi}_t - \xi_t) \end{bmatrix} + O_p\left(\max\left[\frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}}\right]\right). \quad (A.73) $$
where $A_{it}$ is as defined in (A.26). Substituting (A.66) and (A.69) into (A.73), and using (A.72) to approximate $\hat{\xi}_t - \xi_t$, it follows that

$$0 = F_{\epsilon t}(\hat{\xi}_t - \xi_t) + F_{\pi t}(\hat{\pi} - \pi) + c_t + O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right), \quad (A.74)$$

where $c_t$ is defined in (A.28). Using the fact that

$$y^*_t - \hat{\epsilon}_t = (y_t - \epsilon_t) - (\hat{\epsilon}_t - \epsilon_t) + (y^*_t - y_t)$$

$$= \epsilon_t - (\hat{\epsilon}_t - \epsilon_t) + \Delta_t,$$

we can show that

$$F_{\epsilon t}(y^*_t - \hat{\epsilon}_t) = v_{\epsilon t} - F_{\epsilon t}(\hat{\epsilon}_t - \epsilon_t) - B(\hat{\epsilon}_t, \hat{\pi}). \quad (A.75)$$

Solving (A.74) for $F_{\epsilon t}(\hat{\epsilon}_t - \epsilon_t)$ and substituting into (A.75), it follows that

$$F_{\pi t}(y^*_t - \hat{\epsilon}_t) = v_{\epsilon t} + F_{\pi t}(\hat{\pi} - \pi) + c_t - B(\hat{\epsilon}_t, \hat{\pi})$$

$$+ O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right). \quad (A.76)$$

Premultiplying (A.68) by $F_{\epsilon t} \Sigma$ we obtain

$$\Phi_{\epsilon t} \alpha_t = F_{\epsilon t}(y^*_t - \hat{\epsilon}_t),$$

where $\Phi_{\epsilon t} = F_{\epsilon t} \Sigma \hat{F}'_{\epsilon t}$. Using assumption (iv) of Lemma 5, it follows that $\Phi_{\epsilon t}$ is positive definite and $\Phi_{\epsilon t}^{-1}$ exists in a neighborhood of $(\xi_t, \pi)$, so that

$$\alpha_t = \Phi_{\epsilon t}^{-1} F_{\epsilon t}(y^*_t - \hat{\epsilon}_t). \quad (A.77)$$

Substituting (A.76) into (A.77), it follows that

$$\alpha_t = \Phi_{\epsilon t}^{-1} \left[ v_{\epsilon t} + F_{\pi t}(\hat{\pi} - \pi) + c_t - B(\hat{\epsilon}_t, \hat{\pi}) \right]$$

$$+ O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right). \quad (A.78)$$
To find a higher order approximations for \( \hat{\pi} - \pi \), we substitute (A.78) into (A.70) and multiply by \( \frac{1}{T} \) to obtain

\[
0 = \frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi}^t \tilde{\Phi}_{\xi t}^{-1} \left[ v_{\xi t} + F_{\pi t} (\hat{\pi} - \pi) + c_t - B(\xi_t, \hat{\pi}) \right] + O_p \left( \frac{1}{n^{1/2}} \right),
\]

or,

\[
\tilde{m}_{\pi \pi} (\hat{\pi} - \pi) = \tilde{m}_{\pi \nu} + \tilde{m}_{\pi c} - \tilde{m}_{\pi B} + \tilde{R}_{\nu c B},
\]

where

\[
\tilde{m}_{\pi \pi} = \frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi}^t \tilde{\Phi}_{\xi t}^{-1} F_{\pi t}, \quad \tilde{m}_{\pi \nu} = -\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\nu}^t \tilde{\Phi}_{\xi t}^{-1} v_{\xi t}, \quad \tilde{m}_{\pi c} = -\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi t} \tilde{\Phi}_{\xi t}^{-1} c_t,
\]

\[
\tilde{m}_{\pi B} = -\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{\pi t} \tilde{\Phi}_{\xi t}^{-1} B(\xi_t, \hat{\pi}) \text{ and } \tilde{R}_{\nu c B} = O_p \left( \max \left[ \frac{1}{n^{1/2}}, \frac{1}{n^{1/2}} \right] \right).
\]

Using condition (vi) of Theorem 1 and Lemma 2, it follows that \( \tilde{m}_{\pi \pi}^{-1} \) exists in a neighborhood of \( (\xi_t, \pi) \), so that

\[
(\hat{\pi} - \pi) = \tilde{m}_{\pi \pi}^{-1} [\tilde{m}_{\pi \nu} + \tilde{m}_{\pi c} - \tilde{m}_{\pi B}] + O_p \left( \max \left[ \frac{1}{n^{1/2}}, \frac{1}{n^{1/2}} \right] \right). \tag{A.79}
\]

Following steps outlined in (A.33) - (A.36), we can show that

\[
\tilde{m}_{\pi \pi}^{-1} = m_{\pi \pi}^{-1} + O_p \left( \frac{1}{\sqrt{n}} \right), \tag{A.80}
\]

\[
\tilde{m}_{\pi c} = m_{\pi c} + O_p \left( \max \left[ \frac{1}{n^{1/2}}, \frac{1}{n^{1/2}} \right] \right). \tag{A.81}
\]

To find an approximation for \( \tilde{m}_{\pi \nu} \), we complete steps similar to (A.36) - (A.40) to show that

\[
\tilde{F}_{\xi t} = F_{\xi t} + G_{\xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right), \tag{A.82}
\]

\[
\tilde{\Phi}_{\xi t}^{-1} = \Phi_{\xi t}^{-1} - \Phi_{\xi t}^{-1} F_{\xi t} \Sigma G_{\xi t}^{-1} \Phi_{\xi t}^{-1} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right), \tag{A.83}
\]

\[
\tilde{F}_{\pi t} = F_{\pi t} + U_{\xi t} + O_p \left( \max \left[ \frac{1}{n}, \frac{1}{\sqrt{nT}} \right] \right). \tag{A.84}
\]

Using expansions (A.82) - (A.84), we can show that

\[
\tilde{m}_{\pi \nu} = m_{\pi \nu} + O_p \left( \max \left[ \frac{1}{n^{1/2}}, \frac{1}{n^{1/2}} \right] \right). \tag{A.85}
\]
Finally, to find an approximation for $B(\hat{\xi}_t, \hat{\pi})$, we consider the $i^{th}$ element of $B(\hat{\xi}_t, \hat{\pi})$.

To do this, we define

$$g_i(\hat{\xi}_t, \hat{\pi}) = \frac{1}{2} tr \left[ A_i t \left( \Sigma - \Sigma \hat{F}_{\xi t} \Phi^{-1}_{\xi t} \hat{F}_{\xi t} \Sigma \right) \right],$$

$$g_i(\hat{\xi}_t, \pi) = \frac{1}{2} tr \left[ A_i t \left( \Sigma - \Sigma F'_{\xi t} \Phi^{-1}_{\xi t} F_{\xi t} \Sigma \right) \right].$$

By assumptions (ii) and (iii) of Theorem 1, the first derivatives of $g_i$ exist and are continuous in a neighborhood of $(\xi_t, \pi)$. It follows from Lemma 3 that

$$g_i(\hat{\xi}_t, \hat{\pi}) = g_i(\hat{\xi}_t, \pi) + O_p \left( \frac{1}{\sqrt{n}} \right),$$

and

$$\frac{1}{n} g_i(\hat{\xi}_t, \hat{\pi}) = \frac{1}{n} g_i(\hat{\xi}_t, \pi) + O_p \left( \frac{1}{n \sqrt{n}} \right),$$

so that,

$$B(\hat{\xi}_t, \hat{\pi}) = B(\hat{\xi}_t, \pi) + O_p \left( \frac{1}{n \sqrt{n}} \right). \quad (A.86)$$

Thus, using (A.86), we can show that

$$\hat{m}_{xB} = m_{xB} + O_p \left( \frac{1}{n \sqrt{n}} \right). \quad (A.87)$$

Substituting (A.80), (A.81), (A.85) and (A.87) into (A.79), we can show that

$$\hat{\pi} - \pi = m_{xu}^{-1} \left[ m_{xu} + m_{xc} - m_{xB} \right] + O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right)$$

$$= \beta_1 + \beta_3 + O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right), \quad (A.88)$$

where $\beta_3 = m_{xu}^{-1} \left[ m_{xc} - m_{xB} \right] = O_p \left( \frac{1}{n} \right)$.

To derive a higher order approximation for $\hat{\xi}_t - \xi$, we substitute (A.78) into (A.68) and solve for $\hat{\xi}_t - \xi_t$ to obtain

$$\hat{\xi}_t - \xi_t = y_t - \xi_t - \Sigma \hat{F}_{\xi t} \Phi^{-1}_{\xi t} \left[ v_t + F_{xt} (\hat{\pi} - \pi) + c_t - B(\hat{\xi}_t, \hat{\pi}) \right]$$

$$+ O_p \left( \max \left[ \frac{1}{n \sqrt{T}}, \frac{1}{n \sqrt{n}} \right] \right).$$
Substituting (A.88) and $y_t' - \xi_t = \epsilon_t - \Delta_t$, it follows that

$$
\dot{\xi}_t - \xi_t = \epsilon_t - \Delta_t - \Sigma \Phi^{-1}_{\xi t} \left[ v_{\xi t} + F_{\pi t} (\beta_1 + \beta_3) + c_t - B(\dot{\xi}_t, \dot{\pi}) \right]
$$

(A.89)

$$
+ O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right).
$$

Using the fact that $\Delta_t$ must satisfy $F_{\xi t} \Delta_t = -B(\dot{\xi}_t, \dot{\pi})$ we could choose, for example, $\Delta_t = -\Sigma F_{\xi t} \Phi^{-1}_{\xi t} B(\dot{\xi}_t, \dot{\pi})$. Substituting this as well as expansions (A.82), (A.83) and (A.86) into (A.89), we can show that

$$
\dot{\xi}_t - \xi_t = \epsilon_t - \Sigma [F_{\xi t} + G_{\xi t}]' [\Phi^{-1}_{\xi t} \Phi^{-1}_{\xi t} F_{\xi t} \Sigma G_{\xi t} \Phi^{-1}_{\xi t}]
$$

$$
\times [F_{\pi t} (\beta_1 + \beta_3) + c_t] + O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right) + O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right),
$$

where $l_{4 t} = -\Sigma F_{\xi t} \Phi^{-1}_{\xi t} [c_t + F_{\pi t} \beta_3] - n V_{d t} G_{\xi t} \Phi^{-1}_{\xi t} v_{\xi t} = O_p \left( \frac{1}{n} \right)$.

To obtain approximations for $\mu_t - \mu_t$, $\gamma - \gamma$ and $\theta - \theta$, we repeat steps (A.45) - (A.54) above. Using (A.46) and (A.88), we can write

$$
\begin{pmatrix}
\dot{\gamma} - \gamma \\
\dot{\theta} - \theta 
\end{pmatrix} = 
\begin{pmatrix}
m_{\mu \mu} & m_{\mu \theta} \\
m_{\theta \mu} & m_{\theta \theta}
\end{pmatrix}^{-1} 
\begin{pmatrix}
m_{\mu \nu} + m_{\mu \sigma} - m_{\mu B} + \dot{R}_{\mu \nu B} \\
m_{\theta \nu} + m_{\theta \sigma} - m_{\theta B} + \dot{R}_{\theta \nu B}
\end{pmatrix}, 
$$

(A.90)

where

$$
m_{\theta B} = -\frac{1}{T} \sum_{i=1}^{T} F'_{\mu i} \Phi^{-1}_{\mu i} B(\mu_t, \theta),
$$

$$
m_{\mu B} = \frac{1}{T} \sum_{i=1}^{T} F'_{\mu i} \Phi^{-1}_{\mu i} B(\mu_t, \theta),
$$

$$
\dot{R}_{\mu \nu B} = O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right)
$$

and $\dot{R}_{\theta \nu B} = O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right)$. Solving (A.90) for $\dot{\gamma} - \gamma$ and $\dot{\theta} - \theta$, we can show that

$$
\dot{\theta} - \theta = q_{\theta \theta, \mu}^{-1} \left[ m_{\theta \nu} + m_{\theta \sigma} - m_{\theta B} - m_{\theta \mu} m_{\mu \mu}^{-1} (m_{\mu \nu} + m_{\mu \sigma} - m_{\mu B}) \right]
$$

$$
+ O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right)
$$

$$
= \tau_1 + \tau_3 + O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right),
$$
where $\tau_3 = q_{\theta,\theta}^{-1} \left[ (m_{\theta c} - m_{\theta B}) - m_{\theta t} m_{\mu}^{-1} (m_{\mu c} - m_{\mu B}) \right] = O_p \left( \frac{1}{n} \right)$, and

$$
\hat{\gamma} - \gamma = m_{\mu}^{-1} \left[ m_{\mu v} \right] m_{\mu c} - m_{\mu B} - m_{\mu B} \left( \hat{\theta} - \theta \right) + O_p \left( \max \left[ \frac{1}{n\sqrt{T}}, \frac{1}{n\sqrt{T}} \right] \right)
$$

$$
= \Lambda_1 + \Lambda_3 + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
$$

where $\Lambda_3 = m_{\mu}^{-1} \left[ m_{\mu c} - m_{\mu B} - m_{\mu B} \tau_3 \right] = O_p \left( \frac{1}{n} \right)$.

To find an approximation for $\hat{\mu}_t - \mu_t$, we first note that $\hat{\xi}_t - \xi_t = (\hat{\mu}_t - \mu_t) - (\hat{\gamma} - \gamma)$.

Thus,

$$
\hat{\mu}_t - \mu_t = \hat{\xi}_t - \xi_t - (\hat{\gamma} - \gamma)
$$

$$
= d_{1t} + d_{2t} + d_{4t} - (\Lambda_1 + \Lambda_3) + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right). \quad (A.91)
$$

An expression for $l_{2t}$ is given in (A.52). To find an expression for $l_{4t}$, we note that

$$
l_{4t} = -\sum F_{\mu t}^{-1} \Phi_{\mu t}^{-1} \left[ c_t + [-F_{\mu t}, F_{\theta t}] \right] \left( m_{\mu B}^{-1} m_{\mu c} \right) \left( m_{\mu c}^{-1} m_{\mu B} \right)
$$

$$
- n V_{d_{2t}} G_{\mu t}^{-1} \Phi_{\mu t}^{-1} v_t.
$$

Thus, substituting (A.52) and (A.92) into (A.91), it follows that

$$
\hat{\mu}_t - \mu_t = d_{1t} + d_{2t} + d_{4t} + O_p \left( \max \left[ \frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{T}} \right] \right),
$$

where $d_{4t} = - \left( I - \sum F_{\mu t}^{-1} \Phi_{\mu t}^{-1} F_{\mu t} \right) \Lambda_3 - \sum F_{\mu t}^{-1} \Phi_{\mu t}^{-1} \left[ c_t + F_{\theta t} \tau_3 \right] - n V_{d_{2t}} G_{\mu t}^{-1} \Phi_{\mu t}^{-1} v_t$.

To complete the proof, we find $E(d_{4t})$. To make this derivation easy, we note that

$$
\tau_3 = q_{\theta,\theta}^{-1} \left[ (m_{\theta c} - m_{\theta B}) - m_{\theta t} m_{\mu}^{-1} (m_{\mu c} - m_{\mu B}) \right]
$$

$$
= q_{\theta,\theta}^{-1} \left[ m_{\theta c} - m_{\theta t} m_{\mu}^{-1} m_{\mu c} \right] - q_{\theta,\theta}^{-1} \left( m_{\theta B} - m_{\theta t} m_{\mu}^{-1} m_{\mu B} \right)
$$

$$
= \tau_2 - E(\tau_2), \quad (A.93)
$$

and, using (A.93),

$$
\Lambda_3 = m_{\mu}^{-1} \left[ m_{\mu c} - m_{\mu B} - m_{\mu B} \tau_3 \right]
$$
\[ m_{\mu}^{-1} [m_{\mu} - m_{\mu B} - m_{\mu \theta} (\tau_2 - E(\tau_2))] \]
\[ = m_{\mu}^{-1} [m_{\mu} - m_{\mu \theta} \tau_2] - m_{\mu}^{-1} [m_{\mu B} - m_{\mu \theta} E(\tau_2)] \]
\[ = \Lambda_2 - E(\Lambda_2). \]

Hence, it is easy to see that

\[ E(\tau_3) = 0, \quad (A.94) \]
\[ E(\Lambda_3) = 0. \quad (A.95) \]

Thus, using (A.64), (A.94) and (A.95), it follows that

\[ E(d_{4t}) = - \left( I_p - \sum F^t_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) E(\Lambda_3) - \sum F^t_{\mu t} \Phi^{-1}_{\mu t} [F_{\theta t} E(\tau_3) + E(c_t)] \]
\[ - n V_{d_{4t}} E(G^t_{\mu t} \Phi^{-1}_{\mu t} v_t) \]
\[ = - \sum F^t_{\mu t} \Phi^{-1}_{\mu t} E(c_t) \]
\[ = - \sum F^t_{\mu t} \Phi^{-1}_{\mu t} B(\mu_t, \theta), \]

since \( E(c_t) = B(\mu_t, \theta). \)

**Derivation of MSE(\( \hat{\mu}_t \)) and MSE(\( \hat{\mu}_3 \))**

We next derive expressions for the MSE of \( \hat{\mu}_t \) and \( \hat{\mu}_3 \). We begin by considering MSE(\( \hat{\mu}_t \)) and later look at MSE(\( \hat{\mu}_3 \)).

Using the approximation to \( \hat{\mu}_t - \mu_t \) given in (7), we can show that

\[ MSE(\hat{\mu}_t) = Var(d_{1t}) + Var(d_{2t}) + Var(d_{3t}) + 2Cov(d_{1t}, d_{2t}) \]
\[ + 2Cov(d_{1t}, d_{3t}) + 2Cov(d_{2t}, d_{3t}) \]
\[ + E(d_{1t}) E(d_{1t})' + E(d_{2t}) E(d_{2t})' + E(d_{3t}) E(d_{3t})'. \]

Expressions for \( Var(d_{1t}), E(d_{1t}), E(d_{2t}) \) and \( E(d_{3t}) \) are given in (A.59), (A.55), (A.61) and (A.62), respectively. To find the expressions for the remaining terms, it is useful to first show that \( v_t \) and \( d_{4t} \) are independent. To do this, we first note that \( v_t \) and \( d_{4t} \) are
functions of $\epsilon_t$. Since $\epsilon_t$ is assumed to be normal, it follows that $v_t$ and $d_{it}$ are jointly normal. Thus, $v_t$ and $d_{it}$ are independent if $\text{Cov}(d_{it}, v_t) = 0$.

\[
\text{Cov}(d_{it}, v_t) = \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) \text{Cov}(\epsilon_t, \epsilon_t) F\mu_t = \frac{1}{n} \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) \Sigma F\mu_t = O,
\]

by (A.65). Thus, $v_t$ and $d_{it}$ for all $t$ are independent and any functions of $v_t$ and $d_{it}$ are also independent. We also note here that $\text{Cov}(d_{is}, v_s) = O$ for all $t \neq s$ by assumption (v) of Theorem 1. We next consider an expression for $\text{Var}(d_{2t})$.

\[
\text{Var}(d_{2t}) = \text{Var} \left[ - \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) \Lambda_1 - \Sigma F\Phi^{-1}_\mu \Phi\tau_1 \right] = \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) \text{Var}(\Lambda_1) \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) + \Sigma F\Phi^{-1}_\mu \Phi\tau_1 \text{Var}(\tau_1) F\mu_t \Sigma + 2 \left( I_p - \Sigma F\Phi^{-1}_\mu F\mu_t \right) \text{Cov}(\Lambda_1, \tau_1) F\mu_t \Sigma.
\]

To find an expression for $\text{Var}(\Lambda_1)$, we note that

\[
\text{Var}(\Lambda_1) = \text{Var} \left( m_{\mu t}^{-1} [m_{\mu t} - m_{\theta t} \tau_1] \right) = m_{\mu t}^{-1} \left[ \text{Var}(m_{\mu t}) + m_{\theta t} \text{Var}(\tau_1) m_{\theta t} - 2 \text{Cov}(m_{\mu t}, \tau_1) m_{\theta t} \right] m_{\mu t}^{-1}.
\]

Using assumption (v) of Theorem 1, it follows that

\[
\text{Var}(m_{\mu t}) = \text{Var} \left( \frac{1}{T} \sum_{t=1}^T F\Phi^{-1}_\mu v_t \right) = \frac{1}{T^2} \sum_{t=1}^T F\Phi^{-1}_\mu \text{Var}(v_t) \Phi^{-1}_\mu F\mu_t = \frac{1}{nT^2} \sum_{t=1}^T F\Phi^{-1}_\mu \Phi\Phi^{-1}_\mu F\mu_t = \frac{1}{nT^2} \sum_{t=1}^T F\Phi^{-1}_\mu F\mu_t = \frac{1}{nT} m_{\mu t}.
\]
Next,

\[\text{Var}(\tau_1) = \text{Var} \left( q_{\theta \theta, \mu}^{-1} \left[ m_{\theta \theta} - m_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] \right) \]

\[= q_{\theta \theta, \mu}^{-1} \left[ \text{Var}(m_{\theta \theta}) + m_{\theta \mu} m_{\mu \theta}^{-1} \text{Var}(m_{\mu \theta}) m_{\mu \theta}^{-1} m_{\mu \theta} - 2 \text{Cov}(m_{\theta \theta}, m_{\mu \theta}) m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu}^{-1} .\]

Using assumption (v) of Theorem 1, we can show that

\[\text{Var}(m_{\theta \theta}) = \text{Var} \left( - \frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} v_t \right) \]

\[= \frac{1}{T^2} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} \text{Var}(v_t) \Phi_{\mu t}^{-1} F_{\theta t} \]

\[= \frac{1}{nT} m_{\theta \theta} ,\]

and,

\[\text{Cov}(m_{\theta \theta}, m_{\mu \theta}) = \text{Cov} \left( - \frac{1}{T} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} v_t, \frac{1}{T} \sum_{t=1}^{T} F_{\mu t} \Phi_{\mu t}^{-1} v_t \right) \]

\[= - \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} \text{Cov}(v_t, v_s) \Phi_{\mu s}^{-1} F_{\mu s} \]

\[= - \frac{1}{nT^2} \sum_{t=1}^{T} F_{\theta t} \Phi_{\mu t}^{-1} F_{\mu t} \]

\[= \frac{1}{nT} m_{\theta \mu} . \quad (A.100)\]

Thus, combining (A.99) - (A.100), it follows that

\[\text{Var}(\tau_1) = \frac{1}{nT} q_{\theta \theta, \mu}^{-1} \left[ m_{\theta \theta} + m_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} - 2 m_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu}^{-1} \]

\[= \frac{1}{nT} q_{\theta \theta, \mu}^{-1} . \quad (A.101)\]

The last expression needed in (A.98) is

\[\text{Cov}(m_{\mu \theta}, \tau_1) = \text{Cov} \left( m_{\mu \theta}, q_{\theta \theta, \mu}^{-1} \left[ m_{\theta \theta} - m_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] \right) \]

\[= \left[ \text{Cov}(m_{\mu \theta}, m_{\theta \theta}) - \text{Cov}(m_{\mu \theta}, m_{\mu \theta}) m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu}^{-1} \]

\[= \frac{1}{nT} \left[ m_{\mu \theta} - m_{\mu \theta} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu}^{-1} \]

\[= O. \quad (A.102)\]
Thus, substituting (A.99), (A.101) and (A.102) into (A.98), it follows that

\[ \text{Var}(A_1) = \frac{1}{nT} \left[ m_{\mu \mu}^{-1} + m_{\mu \theta}^{-1} m_{\theta \mu} q_{\theta \theta, \mu}^{-1} m_{\theta \mu} m_{\mu \mu}^{-1} \right]. \quad (A.103) \]

The final term needed for (A.97) is

\[ \text{Cov}(A_1, \tau_1) = \text{Cov} \left( m_{\mu \mu}^{-1} [m_{\mu \nu} - m_{\mu \theta} \tau_1], \tau_1 \right) = m_{\mu \mu}^{-1} \text{Cov} (m_{\mu \nu}, \tau_1) - m_{\mu \mu}^{-1} m_{\mu \theta} \text{Var} (\tau_1) = -m_{\mu \mu}^{-1} m_{\mu \theta} \text{Var} (\tau_1) = -\frac{1}{nT} m_{\mu \mu}^{-1} m_{\mu \theta} q_{\theta \theta, \mu}^{-1}. \quad (A.104) \]

Thus, combining results from (A.101), (A.103) and (A.104) in (A.97), we can show that

\[ \text{Var}(d_{2t}) = \frac{1}{nT} \left[ \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right) \left[ m_{\mu \mu}^{-1} + m_{\mu \theta}^{-1} m_{\theta \mu} q_{\theta \theta, \mu}^{-1} m_{\theta \mu} m_{\mu \mu}^{-1} \right] \right. \]
\[ \times \left. \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right)' + \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\theta \mu} q_{\theta \theta, \mu}^{-1} F_{\theta \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \Sigma \right. \]
\[ -2 \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right) m_{\mu \mu}^{-1} m_{\mu \theta} q_{\theta \theta, \mu}^{-1} F_{\theta \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \Sigma \right]. \quad (A.105) \]

Next we consider an expression for \( \text{Var}(d_{3t}) \). It can be shown that

\[ \text{Var}(d_{3t}) = \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right) \text{Var}(A_2) \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right)' \]
\[ + \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} \text{Var} \left( c_t + F_{\theta \mu} \tau_2 \right) \Phi^{-1}_{\mu \mu} F_{\mu \mu} \Sigma + n^2 \text{Var} (d_{3t}) \text{Var}(G'_{\mu \mu} v_t) V_{d_{3t}} \]
\[ + 2 \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right) \text{Cov}(A_2, c_t + F_{\theta \mu} \tau_2) \Phi^{-1}_{\mu \mu} F_{\mu \mu} \Sigma \]
\[ + 2n \left( I_p - \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} F_{\mu \mu} \right) \text{Cov}(A_2, G'_{\mu \mu} v_t) V_{d_{3t}} \]
\[ + 2n \Sigma F'_{\mu \mu} \Phi^{-1}_{\mu \mu} \text{Cov} \left( c_t + F_{\theta \mu} \tau_2, G'_{\mu \mu} v_t \right) V_{d_{3t}}. \quad (A.106) \]

To find \( \text{Var}(A_2) \), we first find an expression for \( \text{Var}(\tau_2) \). To do this, we show that

\[ \text{Var}(\tau_2) = \text{Var} \left( q_{\theta \theta, \mu}^{-1} \left[ m_{\theta \mu} - m_{\theta \mu} m_{\mu \mu}^{-1} m_{\mu \mu} \right] \right) = q_{\theta \theta, \mu}^{-1} \left[ \text{Var}(m_{\theta \mu}) + m_{\theta \mu} m_{\mu \mu}^{-1} \text{Var}(m_{\mu \mu}) m_{\mu \mu}^{-1} m_{\mu \theta} \right. \]
\[ -2 \text{Cov}(m_{\theta \mu}, m_{\mu \mu}) m_{\mu \mu}^{-1} m_{\mu \theta} \left] q_{\theta \theta, \mu}^{-1}. \right\} \quad (A.107) \]
Now, to find \( \text{Var}(\mathbf{m}_\theta c) \) and \( \text{Var}(\mathbf{m}_\mu c) \), we first find an expression for \( \text{Var}(c_t) \). The \( ij^{th} \) element of \( \text{Var}(c_t) \) is of the form

\[
\text{Cov}(\frac{1}{2} \mathbf{d}'_t A_{ii} \mathbf{d}_t, \frac{1}{2} \mathbf{d}'_t A_{jj} \mathbf{d}_t).
\]

Assuming \( \epsilon_t \sim N(0, \Psi) \), it follows that \( \mathbf{d}_t \sim N(0, \mathbf{V}_{d_t}) \). Consequently,

\[
\text{Cov}(\frac{1}{2} \mathbf{d}'_t A_{ii} \mathbf{d}_t, \frac{1}{2} \mathbf{d}'_t A_{jj} \mathbf{d}_t) = \frac{1}{4} \text{tr}(A_{ii} \mathbf{V}_{d_t} A_{jj} \mathbf{V}_{d_t}).
\]

Thus,

\[
\mathbf{V}_t = \frac{1}{4} \begin{bmatrix}
\text{tr}(A_{1t} \mathbf{V}_{d_t} A_{1t} \mathbf{V}_{d_t}) & \text{tr}(A_{1t} \mathbf{V}_{d_t} A_{2t} \mathbf{V}_{d_t}) & \cdots & \text{tr}(A_{1t} \mathbf{V}_{d_t} A_{rt} \mathbf{V}_{d_t}) \\
\text{tr}(A_{2t} \mathbf{V}_{d_t} A_{1t} \mathbf{V}_{d_t}) & \text{tr}(A_{2t} \mathbf{V}_{d_t} A_{2t} \mathbf{V}_{d_t}) & \cdots & \text{tr}(A_{2t} \mathbf{V}_{d_t} A_{rt} \mathbf{V}_{d_t}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{tr}(A_{rt} \mathbf{V}_{d_t} A_{1t} \mathbf{V}_{d_t}) & \text{tr}(A_{rt} \mathbf{V}_{d_t} A_{2t} \mathbf{V}_{d_t}) & \cdots & \text{tr}(A_{rt} \mathbf{V}_{d_t} A_{rt} \mathbf{V}_{d_t})
\end{bmatrix}.
\]

Thus, looking at \( \text{Var}(\mathbf{m}_\theta c) \) and \( \text{Var}(\mathbf{m}_\mu c) \), using the fact that for \( t \neq s \), \( \text{Cov}(\epsilon_t, \epsilon_s) = 0 \), we can show that

\[
\text{Var}(\mathbf{m}_\theta c) = \text{Var}(\frac{1}{T} \sum_{t=1}^{T} F'_t \Phi^{-1}_t c_t) \\
= \frac{1}{T^2} \sum_{t=1}^{T} F'_t \Phi^{-1}_t \mathbf{V}_{c_t} \Phi^{-1}_t F'_t, \quad (A.108)
\]

and,

\[
\text{Var}(\mathbf{m}_\mu c) = \text{Var}(\frac{1}{T} \sum_{t=1}^{T} F'_t \Phi^{-1}_t c_t) \\
= \frac{1}{T^2} \sum_{t=1}^{T} F'_t \Phi^{-1}_t \mathbf{V}_{c_t} \Phi^{-1}_t F'_t. \quad (A.109)
\]

Before we find the covariance term in (A.107), we find \( \text{Cov}(c_t, c_s) \) for all \( t \) and \( s \). For \( t \neq s \), \( c_t \) is a function of \( \epsilon_t \) and \( c_s \) is a function of \( \epsilon_s \). Thus, by assumption (v) of Theorem 1, \( \text{Cov}(c_t, c_s) = 0 \). For \( t = s \), \( \text{Cov}(c_t, c_s) = \mathbf{V}_{c_t} \). Thus,

\[
\text{Cov}(\mathbf{m}_\theta c, \mathbf{m}_\mu c) = \text{Cov} \left( \frac{1}{T} \sum_{t=1}^{T} F'_t \Phi^{-1}_t c_t, \frac{1}{T} \sum_{s=1}^{T} F'_s \Phi^{-1}_s c_s \right) \\
= \frac{1}{T^2} \sum_{t=1}^{T} F'_t \Phi^{-1}_t \mathbf{V}_{c_t} \Phi^{-1}_t F'_t. \quad (A.110)
\]
Thus, using results (A.108), (A.109) and (A.110) in (A.107), it follows that

\[ \text{Var}(\tau_2) = q_{\theta \mu}^{-1} \left[ V_{\theta \mu} + m_{\theta \mu} m_{\mu \theta}^{-1} V_{\mu \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right. \]

\[ \left. - 2V_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \mu}^{-1}, \]

where \( V_{\theta \mu} = \text{Var}(m_{\theta \mu}), V_{\mu \mu} = \text{Var}(m_{\mu \mu}) \) and \( V_{\theta \mu} = \text{Cov}(m_{\theta \mu}, m_{\mu \mu}) \). To find \( \text{Var}(\Lambda_2) \), we write

\[ \text{Var}(\Lambda_2) = \text{Var} \left( m_{\mu \mu}^{-1} [m_{\mu \mu} - m_{\mu \theta} \tau_2] \right) \]

\[ = m_{\mu \mu}^{-1} [\text{Var}(m_{\mu \mu}) + m_{\mu \theta} \text{Var}(\tau_2) m_{\theta \mu} - 2\text{Cov}(m_{\mu \mu}, \tau_2) m_{\theta \mu}] m_{\mu \mu}^{-1}. \]

Looking at \( \text{Cov}(m_{\mu \mu}, \tau_2) \), we see that

\[ \text{Cov}(m_{\mu \mu}, \tau_2) = \left[ \text{Cov}(m_{\mu \mu}, m_{\theta \mu}) - \text{Var}(m_{\mu \mu}) m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \mu}^{-1} \]

\[ = \left[ V_{\theta \mu} q_{\theta \mu}^{-1} - V_{\mu \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \mu}^{-1}, \]

where \( V_{\mu \mu} = \text{Cov}(m_{\mu \mu}, m_{\mu \mu}) \). Thus,

\[ \text{Var}(\Lambda_2) = m_{\mu \mu}^{-1} [V_{\mu \mu} + m_{\mu \theta} \text{Var}(\tau_2) m_{\theta \mu} - 2\left( V_{\theta \mu} - V_{\mu \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right) q_{\theta \mu}^{-1}] m_{\mu \mu}^{-1}, \]

where \( \text{Var}(\tau_2) \) is given in (A.111).

Next, we look at

\[ \text{Var}(c_t + F_{\theta t} \tau_2) = V_{c_t} + F_{\theta t} \text{Var}(\tau_2) F_{\theta t} + 2\text{Cov}(c_t, \tau_2) F_{\theta t}. \]

First,

\[ \text{Cov}(c_t, \tau_2) = \text{Cov} \left( c_t, q_{\theta t}^{-1} \left[ m_{\theta \mu} - m_{\theta \mu} m_{\mu \theta}^{-1} m_{\mu \theta} \right] \right) \]

\[ = \text{Cov}(c_t, m_{\theta \mu}) q_{\theta t}^{-1} - \text{Cov}(c_t, m_{\mu \mu}) m_{\mu \theta}^{-1} m_{\mu \theta} q_{\theta t}^{-1}. \]

Then,

\[ \text{Cov}(c_t, m_{\theta \mu}) = \text{Cov} \left( c_t, -\frac{1}{T} \sum_{s=1}^{T} F_{\theta s} \Phi_{\mu \theta}^{-1} c_s \right) \]

\[ = -\frac{1}{T} \text{Cov}(c_t, c_t) \Phi_{\mu \theta}^{-1} F_{\theta t} \]

\[ = -\frac{1}{T} V_{c_t} \Phi_{\mu \theta}^{-1} F_{\theta t}. \]
and,

\[ \text{Cov} (c_t, m_{\mu c}) = \text{Cov} \left( c_t, \frac{1}{T} \sum_{s=1}^{T} F_{\mu s} \Phi_{\mu s}^{-1} c_s \right) \]

\[ = \frac{1}{T} V_{c_t} \Phi_{\mu t}^{-1} F_{\mu t}. \]  \hspace{1cm} (A.113)

Thus,

\[ \text{Cov}(c_t, \tau_2) = -\frac{1}{T} V_{c_t} \Phi_{\mu t}^{-1} \left[ F_{\theta t} - F_{\mu t} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu t}^{-1}, \]

and,

\[ \text{Var}(c_t + F_{\theta t} \tau_2) = V_{c_t} + F_{\theta t} \text{Var}(\tau_2) F_{\theta t} \]

\[ - \frac{2}{T} V_{c_t} \Phi_{\mu t}^{-1} \left[ F_{\theta t} - F_{\mu t} m_{\mu \theta}^{-1} m_{\mu \theta} \right] q_{\theta \theta, \mu t}^{-1} F_{\theta t}. \]  \hspace{1cm} (A.114)

To find \( \text{Var}(G_t^{i} \Phi_{\mu t}^{-1} v_t) \), we use (A.63) and assumption (v) of Theorem 1, to show that

\[ \text{Var}(G_t^{i} \Phi_{\mu t}^{-1} v_t) = \text{Var} \left( \sum_{i=1}^{r} A_{it} d_{it} s_{it} v_t \right) \]

\[ = \text{Var} \left( \sum_{i=1}^{r} A_{it} d_{it} v'_t s'_t \right) \]

\[ = \text{Var} \left( \text{vec}(A_{it} d_{it} v'_t s'_t) \right) \]

\[ = \text{Var} \left( \sum_{i=1}^{r} (s_{it} \otimes A_{it}) \text{vec}(d_{it} v'_t) \right) \]

\[ = \left( \sum_{i=1}^{r} s_{it} \otimes A_{it} \right) \text{Var} \left( \text{vec}(d_{it} v'_t) \right) \left( \sum_{i=1}^{r} s_{it} \otimes A_{it} \right)' . \]  \hspace{1cm} (A.115)

Conditioning on \( v_t \), we can show that

\[ \text{Var} \left[ \text{vec}(d_{it} v'_t) \right] = \text{Var} \left[ E \left( \text{vec}(d_{it} v'_t) \mid v_t \right) \right] + \text{E} \left[ \text{Var} \left( \text{vec}(d_{it} v'_t) \mid v_t \right) \right]. \]

Now,

\[ \text{Var} \left[ E \left( \text{vec}(d_{it} v'_t) \mid v_t \right) \right] = \text{Var} \left[ \text{vec} \left( E \left( d_{it} v'_t \right) \right) \right] \]

\[ = O, \]
since $E[d_{Hi}] = 0$. Also,

$$E[\text{Var} (\text{vec}(d_{Hi}v_i') | v_i)] = E[\text{Var} (\text{vec}(I_p d_{Hi}v_i') | v_i)]$$

$$= E[\text{Var} ([v_i \otimes I_p] \text{vec}(d_{Hi})) | v_i]$$

$$= E [(v_i \otimes I_p) \text{Var} (\text{vec}(d_{Hi}))' [v_i \otimes I_p]]$$

$$= E [(v_i \otimes I_p) (1 \otimes V_{d_{Hi}}) [v_i \otimes I_p]']$$

$$= E [(v_i \otimes V_{d_{Hi}}) [v_i \otimes I_p]']$$

$$= E (v_i v_i') \otimes V_{d_{Hi}}.$$ 

But,

$$E (v_i v_i') = E (F_{\mu \tau} (\epsilon_i' \epsilon'_i) F_{\mu \tau}')$$

$$= F_{\mu \tau} E (\epsilon_i' \epsilon'_i) F_{\mu \tau}'$$

$$= \frac{1}{n} \Phi_{\mu \tau}.$$ 

Thus, (A.115) simplifies to

$$\text{Var}(G_{Hi} \Phi_{\mu \tau}^{-1} v_i) = \frac{1}{n} \left( \sum_{i=1}^{r} s_{it} \otimes A_{it} \right) (\Phi_{\mu \tau} \otimes V_{d_{Hi}}) \left( \sum_{i=1}^{r} s_{it} \otimes A_{it} \right)'.$$ 

(A.116)

The first covariance term in (A.106) is

$$\text{Cov}(A_2, c_t + F_{\theta \tau} \tau_2) = m_{\mu \mu}^{-1} [\text{Cov}(m_{\mu \tau}, c_t) - m_{\mu \theta} \text{Cov}(\tau_2, c_t)]$$

$$+ \text{Cov}(m_{\mu \tau}, \tau_2) F_{\theta \tau}' - m_{\mu \theta} \text{Var}(\tau_2) F_{\theta \tau}'.$$ 

(A.117)

Combining (A.113), (A.114), and (A.117), it follows that

$$\text{Cov}(\Lambda_2, c_t + F_{\theta \tau} \tau_2) = \frac{1}{n} m_{\mu \mu}^{-1} \left[ F_{\mu \tau} \Phi_{\mu \tau}^{-1} V_{c_t} + m_{\mu \theta} q_{\theta \theta, \mu}^{-1} [F_{\theta \tau} + F_{\mu \tau} m_{\mu \mu}^{-1} m_{\mu \theta}] \Phi_{\mu \tau}^{-1} V_{c_t} \right]$$

$$+ m_{\mu \mu}^{-1} \left[ (V_{\mu \theta c} - V_{\mu \mu c} m_{\mu \mu}^{-1} m_{\mu \theta}) q_{\theta \theta, \mu}^{-1} \right.$$

$$\left. - m_{\mu \theta} \text{Var}(\tau_2) F_{\theta \tau}' \right].$$ 

(A.118)

where $\text{Var}(\tau_2)$ is given in (A.111). The second covariance term in (A.106) is

$$\text{Cov}(\Lambda_2, G_{Hi} \Phi_{\mu \tau}^{-1} v_i) = m_{\mu \mu}^{-1} \left[ \text{Cov}(m_{\mu c}, G_{Hi} \Phi_{\mu \tau}^{-1} v_i) \right]$$

$$- m_{\mu \theta} \text{Cov}(\tau_2, G_{Hi} \Phi_{\mu \tau}^{-1} v_i).$$
Using assumption (v) of Theorem 1, the independence of \(d_{it}\) and \(v_t\) and \(E(v_t) = 0\),

\[
\text{Cov}(m_{\mu c}, G'_t \Phi^{-1}_{\mu t} v_t) = \frac{1}{T} \text{Cov} \left( \sum_{s=1}^{T} F'_{\mu s} \Phi^{-1}_{\mu s} c_s, G'_t \Phi^{-1}_{\mu t} v_t \right)
\]

\[
= \frac{1}{T} F'_{\mu t} \Phi^{-1}_{\mu t} \text{Cov} \left( c_t, G'_t \Phi^{-1}_{\mu t} v_t \right)
\]

\[
= \frac{1}{T} F'_{\mu t} \Phi^{-1}_{\mu t} \left[ E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t - E(c_t) E(G'_t \Phi^{-1}_{\mu t} v_t) \right) \right]
\]

\[
= \frac{1}{T} F'_{\mu t} \Phi^{-1}_{\mu t} E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t \right). \quad (A.119)
\]

Conditioning on \(d_{it}\), we can show that

\[
E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t \right) = E \left[ E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t \mid d_{it} \right) \right]
\]

\[
= E \left[ c_t E(v_t) \Phi^{-1}_{\mu t} G_t \right]
\]

\[
= 0, \quad (A.120)
\]

since \(E(v_t) = 0\). Thus,

\[
\text{Cov}(m_{\mu c}, G'_t \Phi^{-1}_{\mu t} v_t) = O. \quad (A.121)
\]

Next, using (A.121), we can show that

\[
\text{Cov}(\tau_2, G'_t \Phi^{-1}_{\mu t} v_t) = q^{-1}_{\theta \theta, \mu} \left[ \text{Cov}(m_{\theta c}, G'_t \Phi^{-1}_{\mu t} v_t) - m_{\theta c} m_{\mu}^{-1} \text{Cov}(m_{\mu c}, G'_t \Phi^{-1}_{\mu t} v_t) \right]
\]

\[
= q^{-1}_{\theta \theta, \mu} \text{Cov}(m_{\theta c}, G'_t \Phi^{-1}_{\mu t} v_t).
\]

Using assumption (v) of Theorem 1, the independence of \(d_{it}\) and \(v_t\), \(E(v_t) = 0\) and (A.120), it can be shown that

\[
\text{Cov}(m_{\theta c}, G'_t \Phi^{-1}_{\mu t} v_t) = -\frac{1}{T} \sum_{s=1}^{T} \text{Cov} \left( F'_{\theta s} \Phi^{-1}_{\mu s} c_s, G'_t \Phi^{-1}_{\mu t} v_t \right)
\]

\[
= -\frac{1}{T} F'_{\theta t} \Phi^{-1}_{\mu t} \text{Cov} \left( c_t, G'_t \Phi^{-1}_{\mu t} v_t \right)
\]

\[
= -\frac{1}{T} F'_{\theta t} \Phi^{-1}_{\mu t} \left[ E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t - E(c_t) E(G'_t \Phi^{-1}_{\mu t} v_t) \right) \right]
\]

\[
= -\frac{1}{T} F'_{\theta t} \Phi^{-1}_{\mu t} E \left( c_t v'_t \Phi^{-1}_{\mu t} G_t \right)
\]

\[
= O. \quad (A.122)
\]
Thus,
\[
\text{Cov}(c_t, G_t' \Phi_{\mu t}^{-1} v_t) = 0,
\]
\[
\text{Cov}(\tau_2, G_t' \Phi_{\mu t}^{-1} v_t) = 0,
\]  \hspace{1cm} (A.123)
\[
\text{Cov}(A_2, G_t' \Phi_{\mu t}^{-1} v_t) = 0.
\]  \hspace{1cm} (A.124)

Using (A.122) and (A.123), it can be shown that
\[
\text{Cov}(c_t + F_{\theta t} \tau_2, G_t' \Phi_{\mu t}^{-1} v_t) = 0.
\]

Thus, combining the results in (A.112), (A.114), (A.116), (A.118), (A.123) and (A.124) in (A.106), we find that
\[
\text{Var}(d_{3t}) = \left( I_p - \Sigma F_{\mu t}' \Phi_{\mu t}^{-1} F_{\mu t} \right) \text{Var}(A_2) \left( I_p - \Sigma F_{\mu t}' \Phi_{\mu t}^{-1} F_{\mu t} \right)'
\]
\[
+ \Sigma F_{\mu t}' \Phi_{\mu t}^{-1} \text{Var}(c_t + F_{\theta t} \tau_2) \Phi_{\mu t}^{-1} F_{\mu t} \Sigma
\]
\[
+ n \sum_{i=1}^{r} s_{it} \otimes A_{it} \left( \Phi_{\mu t} \otimes V_{d_{it}} \right) \left( \sum_{i=1}^{r} s_{it} \otimes A_{it} \right)'
\]
\[
+ 2 \left( I_p - \Sigma F_{\mu t}' \Phi_{\mu t}^{-1} F_{\mu t} \right) \text{Cov}(A_2, c_t + F_{\theta t} \tau_2) \Phi_{\mu t}^{-1} F_{\mu t} \Sigma.
\]

We next consider the covariance terms in (A.96). Before we do this, however, we recall that \(d_{4t}\) and \(v_t\) are independent. Thus, functions of \(d_{4t}\) and \(v_t\) are also independent. In particular, \(\tau_1\) and \(\Lambda_1\) are functions of \(v_t\) and \(\tau_2\) and \(\Lambda_2\) are functions of \(d_{4t}\), so that
\[
\text{Cov}(d_{4t}, \tau_1) = 0,
\]  \hspace{1cm} (A.125)
\[
\text{Cov}(d_{4t}, \Lambda_1) = 0,
\]  \hspace{1cm} (A.126)
\[
\text{Cov}(\tau_1, \tau_2) = 0,
\]  \hspace{1cm} (A.127)
\[
\text{Cov}(\tau_1, \Lambda_2) = 0,
\]
\[
\text{Cov}(\Lambda_1, \Lambda_2) = 0,
\]
\[
\text{Cov}(\Lambda_1, \tau_2) = 0.
\]  \hspace{1cm} (A.128)

Thus, using (A.125) and (A.126) it follows that
\[
\text{Cov}(d_{4t}, d_{2t}) = -\text{Cov}(d_{4t}, \Lambda_1) \left( I_p - \Sigma F_{\mu t}' \Phi_{\mu t}^{-1} F_{\mu t} \right)' - \text{Cov}(d_{4t}, \tau_1) F_{\theta t}' \Phi_{\mu t}^{-1} F_{\mu t} \Sigma
\]
\[ = O. \]

Next,

\[
\text{Cov}(d_{1t}, d_{2t}) = -\text{Cov}(d_{1t}, \Lambda_2) \left(I_p - \sum_{\mu \tau} F'_{\mu \tau} \Phi_{\mu \tau}^{-1} F_{\mu \tau} \right)'
\]

\[
-\text{Cov}(d_{1t}, c_t + F_{\theta t} \tau_2) \Phi'_{\theta t} \Phi_{\mu \tau}^{-1} F_{\mu \tau} \Sigma
\]

\[
- n \text{Cov}(d_{1t}, G'_t \Phi_{\mu \tau}^{-1} v_t) V_{d_{1t}}.
\]

(A.129)

To find \( \text{Cov}(d_{1t}, \Lambda_2) \), we first find \( \text{Cov}(d_{1t}, \tau_2) \) since \( \Lambda_2 \) is a function of \( \tau_2 \). Thus,

\[
\text{Cov}(d_{1t}, \tau_2) = \text{Cov}(d_{1t}, m_{\theta \tau}) q_{\theta \tau, \mu}^{-1} - \text{Cov}(d_{1t}, m_{\mu \tau}) m_{\mu \tau}^{-1} m_{\mu \theta} q_{\theta \tau, \mu}^{-1},
\]

where, using assumption (v) of Theorem 1

\[
\text{Cov}(d_{1t}, m_{\theta \tau}) = -\frac{1}{T} \text{Cov} \left( d_{1t}, \sum_{s=1}^{T} F'_{\theta s} \Phi_{\mu s}^{-1} c_s \right)
\]

\[
= -\frac{1}{T} \text{Cov} \left( d_{1t}, c_t \right) \Phi_{\mu \tau}^{-1} F_{\theta t},
\]

and,

\[
\text{Cov}(d_{1t}, m_{\mu \tau}) = \frac{1}{T} \text{Cov} \left( d_{1t}, \sum_{s=1}^{T} F'_{\mu s} \Phi_{\mu s}^{-1} c_s \right)
\]

\[
= \frac{1}{T} \text{Cov} \left( d_{1t}, c_t \right) \Phi_{\mu \tau}^{-1} F_{\mu t}.
\]

To simplify \( \text{Cov}(d_{1t}, c_t) \), we find the covariance between \( d_{1t} \) and the \( i^{th} \) element of \( c_t \). For each \( i \),

\[
\text{Cov}(d_{1t}, c_{it}) = \text{Cov}(d_{1t}, \frac{1}{2} d'_{1t} A_{ii} d_{1t})
\]

\[
= \frac{1}{2} E \left( d_{1t} d'_{1t} A_{ii} d_{1t} \right) - \frac{1}{2} E \left( d_{1t} d_{1t} \right) E \left( d'_{1t} A_{ii} d_{1t} \right)
\]

\[
= \frac{1}{2} V_{d_{1t}} A_{ii} E(d_{1t})
\]

\[
= 0,
\]
so that,

\[ \text{Cov}(d_{1t}, c_t) = 0. \] \hfill (A.130)

Thus, it follows that

\[ \text{Cov}(d_{1t}, m_{\theta \mu}) = 0, \]
\[ \text{Cov}(d_{1t}, m_{\mu \sigma}) = 0, \] \hfill (A.131)
\[ \text{Cov}(d_{1t}, \tau_2) = 0. \] \hfill (A.132)

Then, using (A.131) and (A.132), it can be shown that

\[ \text{Cov}(d_{1t}, \Lambda_2) = \text{Cov}(d_{1t}, m_{\mu \sigma})m_{\mu \sigma}^{-1} - \text{Cov}(d_{1t}, \tau_2)m_{\theta \mu}m_{\mu \sigma}^{-1} \]
\[ = 0. \] \hfill (A.133)

Next, using (A.130) and (A.132),

\[ \text{Cov}(d_{1t}, c_t + F_{gt} \tau_2) = \text{Cov}(d_{1t}, c_t) + \text{Cov}(d_{1t}, \tau_2)F_{gt} \]
\[ = 0. \]

Finally, we consider \( \text{Cov}(d_{1t}, G_t^{-1} \Phi_{\mu t}^{-1} v_t) \).

\[ \text{Cov} \left[ d_{1t}, G_t^{-1} \Phi_{\mu t}^{-1} v_t \right] = E \left[ d_{1t} v'_t \Phi_{\mu t}^{-1} G_t \right] - E \left[ d_{1t} \right] E \left[ G_t^{-1} \Phi_{\mu t}^{-1} v_t \right]' \]
\[ = E \left[ d_{1t} v'_t \Phi_{\mu t}^{-1} G_t \right] \]
\[ = E \left[ E \left( d_{1t} v'_t \Phi_{\mu t}^{-1} G_t \mid d_{1t} \right) \right] \]
\[ = E \left[ d_{1t} E \left( v'_t \mid d_{1t} \right) \Phi_{\mu t}^{-1} G_t \right] \]
\[ = 0. \] \hfill (A.134)

Thus, using (A.133) - (A.134) in (A.129), it follows that

\[ \text{Cov}(d_{1t}, d_{3t}) = 0. \]
Finally, we look at $\text{Cov}(d_{2t}, d_{3t})$. Using (A.127) - (A.128), we can show that

$$
\text{Cov}(d_{2t}, d_{3t}) = n \left( I_p - \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\mu t} \right) \text{Cov}(\Lambda_1, G'_t \Phi^{-1}_{\mu t} v_t) V_{d_{3t}},
$$

$$
+ n \sum F'_{\mu t} \Phi^{-1}_{\mu t} F_{\theta t} \text{Cov}(\tau_1, G'_t \Phi^{-1}_{\mu t} v_t) V_{d_{3t}}.
$$

Working with $\text{Cov}(\tau_1, G'_t \Phi^{-1}_{\mu t} v_t)$, we obtain

$$
\text{Cov}(\tau_1, G'_t \Phi^{-1}_{\mu t} v_t) = q^{-1}_{\theta\mu} \left[ \text{Cov}(m_{\theta v}, G'_t \Phi^{-1}_{\mu t} v_t) - m_{\theta \mu} m^{-1}_{\mu \mu} \text{Cov}(m_{\mu v}, G'_t \Phi^{-1}_{\mu t} v_t) \right].
$$

Using assumption (v) of Theorem 1, we can show that

$$
\text{Cov}(m_{\theta v}, G'_t \Phi^{-1}_{\mu t} v_t) = - \frac{1}{T} \sum_{s=1}^{T} \text{Cov}(F'_{\theta s} \Phi^{-1}_{\mu t} v_s, G'_t \Phi^{-1}_{\mu t} v_t)
$$

$$
= - \frac{1}{T} F'_{\theta t} \Phi^{-1}_{\mu t} \text{Cov}(v_t, G'_t \Phi^{-1}_{\mu t} v_t),
$$

where, using the independence of $d_{1t}$ and $v_t$, $E(v_t)$ and $E(G_t) = O$, it follows that

$$
\text{Cov}(v_t, G'_t \Phi^{-1}_{\mu t} v_t) = E(v_t v'_t \Phi^{-1}_{\mu t} G_t) - E(v_t) E(G'_t \Phi^{-1}_{\mu t} v_t)'
$$

$$
= E(v_t v'_t) \Phi^{-1}_{\mu t} E(G_t)'
$$

$$
= O. \tag{A.135}
$$

Using (A.135), we can show that

$$
\text{Cov}(m_{\mu v}, G'_t \Phi^{-1}_{\mu t} v_t) = \frac{1}{T} F'_{\mu t} \Phi^{-1}_{\mu t} \text{Cov}(v_t, G'_t \Phi^{-1}_{\mu t} v_t)
$$

$$
= O. \tag{A.136}
$$

Thus,

$$
\text{Cov}(\tau_1, G'_t \Phi^{-1}_{\mu t} v_t) = O. \tag{A.137}
$$

Next, working with $\text{Cov}(\Lambda_1, G'_t \Phi^{-1}_{\mu t} v_t)$, we can show that

$$
\text{Cov}(\Lambda_1, G'_t \Phi^{-1}_{\mu t} v_t) = m^{-1}_{\mu \mu} \left[ \text{Cov}(m_{\mu v}, G'_t \Phi^{-1}_{\mu t} v_t) - m_{\theta \mu} \text{Cov}(\tau_1, G'_t \Phi^{-1}_{\mu t} v_t) \right]
$$

$$
= O,
$$
by (A.136) and (A.137). Thus, it follows that

\[ \text{Cov}(d_{2t}, d_{3t}) = 0, \]

and,

\[ \text{MSE}(\bar{\mu}_t) = V_{d_{1t}} + \text{Var}(d_{2t}) + \text{Var}(d_{3t}) + E(d_{3t})E(d_{3t})'. \]

The derivation for an expression for \( \text{MSE}(\bar{\mu}_t) \) uses much of the work shown above. Using the expansion given in (20),

\[ \text{MSE}(\bar{\mu}_t) = \text{Var}(d_{1t}) + \text{Var}(d_{2t}) + \text{Var}(d_{4t}) + 2\text{Cov}(d_{1t}, d_{2t}) \]

\[ + 2\text{Cov}(d_{1t}, d_{4t}) + 2\text{Cov}(d_{2t}, d_{4t}) \]

\[ + E(d_{1t})E(d_{1t})' + E(d_{2t})E(d_{2t})' + E(d_{4t})E(d_{4t})'. \]

Recall that \( E(d_{1t}) = 0, \ E(d_{2t}) = 0, \text{ and } E(d_{4t}) = -\Sigma F_{\mu t}' \Phi_{\mu t}^{-1} B(\mu_t, \theta). \) Also, \( d_{4t} = d_{3t} + K \) where \( K \) is a constant matrix. Thus, \( \text{Var}(d_{4t}) = \text{Var}(d_{3t}), \text{Cov}(d_{1t}, d_{4t}) = \text{Cov}(d_{1t}, d_{3t}) = 0 \) and \( \text{Cov}(d_{2t}, d_{4t}) = \text{Cov}(d_{2t}, d_{3t}) = 0 \) by (A.119), so that \( \text{MSE}(\bar{\mu}_t) \) simplifies to

\[ \text{MSE}(\bar{\mu}_t) = V_{d_{1t}} + \text{Var}(d_{2t}) + \text{Var}(d_{3t}) + E(d_{4t})E(d_{4t})', \]

where \( E(d_{3t})E(d_{3t})' - E(d_{4t})E(d_{4t})' \) is non-negative definite.

**Proof of Theorem 3**

The result follows from Liapounov's central limit theorem and the boundedness of the derivatives. \[ \blacksquare \]
Errors-in-variables analysis was considered for nonlinear engineering processes operating in both steady state and in non-steady state. For the both processes, it was shown that up to a certain order of approximation, the bias adjusted estimator has smaller bias and smaller mean squared error than the maximum likelihood estimator. As a result, inference procedures conducted using the bias adjusted estimator have better properties than those using the maximum likelihood estimator.

For the nonlinear non-steady state process, a condition for estimability of the restrictions parameters was given. This condition is important because it clarifies a point of confusion in the existing literature.