Asymptotics for pooled marginal slicing estimator based on SIR\(_\alpha\) approach

Jérôme Saracco

Institut de Mathématiques de Bourgogne, UMR CNRS 5584, Université de Bourgogne, BP 47870, 20 rue Alain Savary, 21 078 Dijon Cedex.

Received 4 September 2003
Available online 2 December 2004

Abstract

Pooled marginal slicing (PMS) is a semiparametric method, based on sliced inverse regression (SIR) approach, for achieving dimension reduction in regression problems when the outcome variable \(y\) and the regressor \(x\) are both assumed to be multidimensional. In this paper, we consider the SIR\(_\alpha\) version (combining the SIR-I and SIR-II approaches) of the PMS estimator and we establish the asymptotic distribution of the estimated matrix of interest. Then the asymptotic normality of the eigenprojector on the estimated effective dimension reduction (e.d.r.) space is derived as well as the asymptotic distributions of each estimated e.d.r. direction and its corresponding eigenvalue.

© 2004 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: primary 62F12; secondary 62H10; 62E20

Keywords: Asymptotics; Dimension reduction subspaces; Eigen-elements; Pooled marginal slicing; Semiparametric regression model; Sliced inverse regression

1. Introduction

The existing multivariate sliced inverse regression (SIR) methods (see [1,17]) are some adaptations and extensions of the univariate usual SIR approach (introduced by Li [16] and called SIR-I hereafter) when the response variable \(y\) is assumed to be \(q\)-dimensional \((q > 1)\). These methods are named: complete slicing method, marginal slicing method,
pooled marginal slicing (PMS) method and alternating SIR method. Barreda et al. [3] focus on some extensions of the existing multivariate SIR approaches by using the SIR$_x$ method which is not blind for symmetric dependencies.

All these multivariate methods examine the relationship between the response variable $y$ and a $p$-dimensional regressor variable $x$ via a semiparametric regression approach. The corresponding model assumes that the dependency between the regressor and the response variable is described by the $K$ linear combinations $\beta'_1x, \ldots, \beta'_Kx$. In other words, the distribution of $y$ given $x$ depends only on these $K$ linear combinations. The associated dimension reduction model can be written this way:

$$y = \begin{cases} 
  g_1(\beta'_1x, \ldots, \beta'_Kx, \varepsilon_1) \\
  \vdots \\
  g_q(\beta'_1x, \ldots, \beta'_Kx, \varepsilon_q).
\end{cases} \quad (1)$$

The error terms $\varepsilon_j$ are assumed independent of $x$ and the $g_j$'s are unknown real-valued functions. Clearly, the $\beta_k$ are not identifiable. Given a sample $\{(y_i, x_i), i, \ldots, n\}$ of independent observations of $(y, x)$, the objective of the multivariate SIR approaches is to find the effective dimension reduction (e.d.r.) space, namely the linear subspace $\mathcal{B}$ spanned by the unknown $\beta$-vectors, without assuming the functional form of the $g_j$'s and without estimating the link functions $g_j$. When $K$ is small ($K \ll p$), the data can be effectively reduced by projecting $x$ along the e.d.r. directions (a basis of the e.d.r. space) for further study of their relationship with $y$. Nonparametric smoothing methods (kernel, smoothing splines, ...) can be used in order to estimate the $g_j$'s.

This paper focuses on the asymptotic normality of the PMS estimator based on the SIR$_x$ approach. In Section 2, we give an overview of the univariate SIR$_x$ method and we describe the corresponding PMS estimator. In Section 3, we state the main results. The asymptotic distribution of the estimated matrix of interest is obtained in Theorem 1. Then the asymptotic normality of the eigenprojector on the estimated e.d.r. space is derived in Theorem 2, as well as the asymptotic distributions of each estimated e.d.r. direction and its corresponding eigenvalue. The proofs are in the Appendix A.

2. Pooled marginal slicing estimator

First, we give a panorama of the univariate SIR$_x$ approach, then we describe the corresponding PMS estimator.

2.1. Overview of the univariate SIR$_x$ method

We give an overview of the univariate SIR$_x$ approaches (that is when $q = 1$). While there are several possible variations, the basic principle of SIR methods (SIR-I, SIR-II or SIR$_x$) is to reverse the role of $y$ and $x$. Instead of regressing the univariate $y$ on the multivariate $x$, the covariable $x$ is regressed on the response variable $y$. The SIR-I estimates based on the first moment $\mathbb{E}(x|y)$ have been studied extensively (see for instance [2,6,8,13,14,16,18,24]). But this approach is “blind” for symmetric dependencies (see [7,15]). Then, SIR-II estimates
based on the inverse conditional second moment $\sqrt{\mathbb{V}(x|y)}$ have been suggested (see for instance [7,15,16]). Hence these two approaches concentrate on the use of the inverse conditional moments $\mathbb{E}(x|y)$ or $\sqrt{\mathbb{V}(x|y)}$ to find the e.d.r. space. For increasing the chance of discovering all the e.d.r. directions, the idea of the SIR$_x$ method is to conjugate these informations: if an e.d.r. direction can only be marginally detected by SIR-I or SIR-II, a suitable combination of these two methods may sharpen the result.

Let us now recall the geometric properties of the model. Let $T$ denote a monotonic transformation of $y$. In order to conjugate information from the SIR-I and SIR-II approaches, Li [16] considered, for $\varpi \in [0, 1]$, the eigen-decomposition of the matrix

$$\Sigma^{-1} M_{\varpi},$$

where $\mu = \mathbb{E}(x)$, $\Sigma = \sqrt{\mathbb{V}(x)}$, and $M_{\varpi} = (1 - \varpi)M_I \Sigma^{-1} M_I + \varpi M_{\Pi}$. The matrices $M_I$ and $M_{\Pi}$ are respectively the matrices used in the usual SIR-I and SIR-II approaches. They are defined as follows: $M_I = \sqrt{\mathbb{V}(\mathbb{E}(x|T(y)))}$ and $M_{\Pi} = \mathbb{E}\left\{ (\sqrt{\mathbb{V}(x|T(y))} - \mathbb{E}(\sqrt{\mathbb{V}(x|T(y))})) \Sigma^{-1} (\sqrt{\mathbb{V}(x|T(y))} - \mathbb{E}(\sqrt{\mathbb{V}(x|T(y))}))' \right\}$. It can be shown that, under the linearity condition (2) defined in Remark 3, the eigenvectors associated with the largest $K$ eigenvalues of $\Sigma^{-1} M_{\varpi}$ are some e.d.r. directions. Let us remark that, when $\varpi = 0$ (resp. $\varpi = 1$), SIR$_x$ is equivalent to SIR-I (resp. SIR-II).

Li [16] proposed a transformation $T$, called a slicing, which categorizes the response $y$ into a new response with $H > K$ levels. The support of $y$ is partitioned into $H$ non-overlapping slices $s_1, \ldots, s_h, \ldots, s_H$. With such transformation $T$, the matrices of interest are now written as

$$M_I = \sum_{h=1}^H p_h (m_h - \mu)(m_h - \mu)' \quad \text{and} \quad M_{\Pi} = \sum_{h=1}^H p_h (V_h - \overline{V}) \Sigma^{-1} (V_h - \overline{V}),$$

where $p_h = P(y \in s_h)$, $m_h = \mathbb{E}(x|y \in s_h)$, $V_h = \sqrt{\mathbb{V}(x|y \in s_h)}$ and $\overline{V} = \sum_{h=1}^H p_h V_h$.

So, it is straightforward to estimate these matrices by substituting empirical versions of the moments for their theoretical counterparts, and therefore to obtain the estimation of the e.d.r. directions. Each estimated e.d.r. direction converges to an e.d.r. direction at rate $\sqrt{n}$, see for instance [16,18]. Asymptotic normality of the SIR$_x$ estimates has been studied by Gannoun and Saracco [10].

**Remark 1.** The practical choice of the slicing function $T$ is discussed in [15,16,19]. Note that the user has to fix the slicing strategy and the number $H$ of slices. The SIR theory makes no assumption about the slicing strategy. In practice, there are naturally two possibilities: to fix the width of the slices or to fix the number of observations per slice. In their investigation of SIR-I, various researchers have preferred the second approach. Then, from the sample point of view, the slices are such that the number of observations in each slice is as close to each other as possible. In order to avoid artificial reduction of dimension, $H$ must be greater than $K$. Also, in order to have at least two cases in each slice, $H$ must be less than $\lceil n/2 \rceil$ where $[a]$ denotes the integer part of $a$. Li [16] noticed that the choice of the slicing is less crucial than the choice of a bandwidth as in kernel-based methods. Simulation studies show that the influence of “slicing parameter” is small when the sample size is greater than 100. Note that, in order to avoid the choice of a slicing, kernel-based estimate of SIR-I has
been investigated, see [2,23]. However, these methods are hard to implement with regard to
basic Slicing one and are computationally slow. Moreover, Bura [4] and Bura and Cook [5]
proposed a parametric version of SIR-I. Note that determining the number $K$ (of indices) is
considered by Li [16], Schott [20] and Ferré [9], for the SIR-I method.

**Remark 2.** The practical choice of $\alpha$ can be based on the test approach proposed by Saracco
[19], this approach does not require the estimation of the link function. Two cross-validation
criteria have been also developed by Gannoun and Saracco [11] to select the parameter $\alpha$, note that these criteria require the kernel smoothing estimation of the link function.

**Remark 3.** Note that one crucial condition for the success of SIR methods is
\[ \mathbb{E}(b'x|\beta_1x, \ldots, \beta_Kx) \text{ is linear for any } b. \] (2)

Note that (2) is satisfied when $x$ has an elliptically symmetric distribution. It does not seem
possible to verify (2), this involves the unknown directions of main interest as a start. As
Li [16] pointed out, this linearity condition is not a severe restriction. Using a Bayesian
argument of Hall and Li [12], we can infer that (2) holds approximatively for many high
dimensional data sets.

2.2. Pooled marginal slicing estimator based on SIR$_{\alpha}$

In the multidimensional framework of model (1), only the basic SIR-I theory has been
used. Li et al. [17] and Aragon [1] considered several estimation methods in this multivariate
context. These methods are named: complete slicing method, marginal slicing method, PMS
method and alternating SIR method. Barreda et al. [3] introduced their corresponding SIR$_{\alpha}$
version. Hereafter, we give a short description of the PMS method based on the SIR$_{\alpha}$
approach.

In the following, let $y^{(j)}$ denote the $j$th component of the $q$-dimensional vector $y$, let $y_i$
be the $i$th observed $q$-dimensional vector $y$, let $y_i^{(j)}$ denote the $j$th component of $y_i$, and let
$x_i$ be the $i$th observed $p$-dimensional vector $x$.

2.2.1. Population version

The idea of this method is to consider the $q$ univariate SIR$_{\alpha}$ methods of each component
$y^{(j)}$ of $y$ on $x$ (based on a specific slicing $T_j$) and to combine the corresponding $M_{\alpha}$ matrices
(denoted by $M_{\alpha}^{(j)}$) in the following pooling:

\[ M_{\alpha,P} = \sum_{j=1}^{q} w_j M_{\alpha}^{(j)} \] (3)

for positive weights $w_j$ and parameters $\alpha_j$. Note that, in the $M_{\alpha,P}$ matrix, the $\alpha$ index
stands for the vector $(\alpha_1, \ldots, \alpha_q)$ and the $P$ index stands for “pooled”. Each transformation
$T_j$ categorizes each response $y^{(j)}$ into a new response with $H_j > K$ levels. We assume
that the support of each $y^{(j)}$ is partitioned into $H_j$ fixed slices $s_1^{(j)}, \ldots, s_{h_j}^{(j)}, \ldots, s_{H_j}^{(j)}$. For
\[ j = 1, \ldots, q, \] the matrices \( M_{x_j}^{(j)} \) are defined as follows:
\[
M_{x_j}^{(j)} = (1 - z_j) M_1^{(j)} \Sigma^{-1} M_1^{(j)} + z_j M_{II}^{(j)}
\]
with
\[
M_1^{(j)} = \mathbb{V}(\mathbb{E}(x|T_j(y^{(j)}))) = \sum_{h=1}^{H_j} p_h^{(j)} (m_h^{(j)} - \mu)(m_h^{(j)} - \mu)',
\]
\[
M_{II}^{(j)} = \mathbb{E} \left\{ \left( \mathbb{V}(x|T_j(y^{(j)})) - \mathbb{E}(\mathbb{V}(x|T_j(y^{(j)}))) \right) \Sigma^{-1} \times \left( \mathbb{V}(x|T_j(y^{(j)})) - \mathbb{E}(\mathbb{V}(x|T_j(y^{(j)}))) \right)' \right\}
\]
\[
= \sum_{h=1}^{H_j} p_h^{(j)} \left( V_h^{(j)} - \overline{V}^{(j)} \right) \Sigma^{-1} \left( V_h^{(j)} - \overline{V}^{(j)} \right)
\]
\[
= \sum_{h=1}^{H_j} K_h^{(j)} \Sigma^{-1} K_h^{(j)},
\]
where \( p_h^{(j)} = P(y^{(j)} \in s_h^{(j)}) \), \( m_h^{(j)} = \mathbb{E}(x|y^{(j)} \in s_h^{(j)}) \), \( V_h^{(j)} = \mathbb{V}(s_h^{(j)}) \) and \( \overline{V}^{(j)} = \sum_{h=1}^{H_j} p_h^{(j)} V_h^{(j)} \) and \( K_h^{(j)} = \sqrt{p_h^{(j)}} (V_h^{(j)} - \overline{V}^{(j)}) \).

Under the linearity condition (2), the eigenvectors associated with the largest \( K \) eigenvalues of \( \Sigma^{-1} M_{x,p} \) are e.d.r. directions. In the following, we assume that these \( K \) e.d.r. directions, denoted by \( b_1, \ldots, b_K \), span the e.d.r. space.

2.2.2. Sample version

Let \( \mathbb{1}[] \) be the indicator function and let \( \mathbb{1}_h^{(j)} = \mathbb{1}[y^{(j)} \in s_h^{(j)}] \). Then \( p_h^{(j)} = \mathbb{E}(\mathbb{1}_h^{(j)}) \), \( m_h^{(j)} = \mathbb{E}(x \mathbb{1}_h^{(j)}) / p_h^{(j)} \) and \( V_h^{(j)} = \mathbb{E}(xx' \mathbb{1}_h^{(j)}) / p_h^{(j)} - \left( \mathbb{E}(x \mathbb{1}_h^{(j)}) / p_h^{(j)} \right)^2 \). So, it is straightforward to estimate these matrices and therefore the e.d.r. directions.

Let \( \{y_i, x_i\}, i = 1, \ldots, n \) be a sample of observations from model (1) and let \( \mathbb{1}_{hi}^{(j)} = \mathbb{1}[y_i^{(j)} \in s_h^{(j)}] \). The empirical mean and covariance matrix of the \( x_i \)'s are given by \( \overline{x} = n^{-1} \sum_{i=1}^{n} x_i \) and \( \overline{xx'} = n^{-1} \sum_{i=1}^{n} x_i x_i' \). Moreover, let us write

\[
\mathbb{1}_{hi}^{(j)} = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{hi}^{(j)} x_i x_i', \overline{xx'} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{xx'} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{\mathbb{1}_{hi}^{(j)}} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{\mathbb{1}_{hi}^{(j)}} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{\mathbb{1}_{hi}^{(j)}} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{\mathbb{1}_{hi}^{(j)}} = n^{-1} \sum_{i=1}^{n} x_i x_i', \overline{\mathbb{1}_{hi}^{(j)}} = n^{-1} \sum_{i=1}^{n} x_i x_i'.
\]

By substituting empirical versions of these moments for their theoretical counterparts, \( M_1^{(j)} \) and \( M_{II}^{(j)} \) are then estimated by

\[
\hat{M}_1^{(j)} = \sum_{h=1}^{H_j} \hat{m}_h^{(j)} \left( \left( \mathbb{1}_{hi}^{(j)} / \mathbb{1}_h^{(j)} \right) - \overline{x} \right) \left( \left( \mathbb{1}_{hi}^{(j)} / \mathbb{1}_h^{(j)} \right) - \overline{x} \right)'
\]
and

\[
\hat{M}_{II}^{(j)} = \sum_{h=1}^{H_j} \hat{K}_h^{(j)} \overline{xx'} \hat{K}_h^{(j)},
\]
where

\[
\hat{K}_h^{(j)} = \left( \mathbb{1}_h^{(j)} \right)^{-1/2} \left( \mathbb{1}_{hi}^{(j)} - \overline{\mathbb{1}_{hi}^{(j)}} \right).
\]
In the following, we denote the estimated e.d.r. directions by \( \hat{b}_1, \ldots, \hat{b}_K \) which are the eigenvectors associated with the \( K \) largest eigenvalues of \( \hat{\Sigma}^{-1} \hat{M}_{\alpha, \rho} \). The e.d.r. space \( E \) is estimated by \( \hat{E} \), the linear subspace generated by the \( \hat{b}_k \)'s.

**Remark 4.** From a practical point of view, as in [1], two kinds of weights \( w_j \) can be used: equal weights or weights proportional to the major eigenvalues found by a preliminary univariate SIR\( _\alpha \) analysis of each component of \( y \).

**Remark 5.** In (3), the parameters \( \alpha_j \) are individually chosen for each univariate SIR\( _\alpha \) method. For the choice of the \( \alpha_j \)'s, the method based on the test approach of Saracco [19] which does not require the estimation of the link function can be used.

### 3. Asymptotic results

In the sequel, the notation \( X_n \rightarrow_d X \) means that \( X_n \) converges in distribution to \( X \) as \( n \rightarrow \infty \). Let \( D_1 \otimes D_2 \) denote the Kronecker product of the matrices \( D_1 \) and \( D_2 \) (see [22] for some useful properties of the Kronecker product). From now on, for each \( s \times s \) matrix \( D = (d^{(jk)}) \), let \( \text{vec}(D) = (d^{(11)}, \ldots, d^{(s1)}, d^{(21)}, \ldots, d^{(ss)})' \) be the \( s^2 \)-dimensional column vector of all elements of \( D \).

The assumptions which are necessary to state our results are gathered together below for easy reference.

(A1). \( \{(y_i, x_i), i = 1, \ldots, n\} \) is a sample of independent observations from model (1).

(A2). The supports of each component \( y^{(j)} \) of \( y \) are partitioned into \( H_j \) fixed slices \( s_1^{(j)}, \ldots, s_h^{(j)}, \ldots, s_{H_j}^{(j)} \) such that \( p^{(j)} \neq 0 \).

(A3). The covariance matrix \( \Sigma \) is positive definite.

(A4). We assume that \( K \) is known and the \( K \) largest eigenvalues of \( \Sigma^{-1} M_{\alpha, \rho} \) are non-null and satisfy: \( \lambda_1 > \ldots > \lambda_K > \lambda_{K+1} \), where \( K + 1 \leq p \).

Let us denote by \( \Omega = \{\hat{\lambda}_1, \ldots, \hat{\lambda}_K\} \) the set of the \( K \) eigenvalues associated with the e.d.r. space, that is, the \( K \) largest eigenvalues of \( \Sigma^{-1} M_{\alpha, \rho} \). Let \( P = \sum_{\hat{\lambda}_k \in \Omega} P_{\hat{\lambda}_k} \) be the \( \Sigma \)-orthogonal eigenprojector on the e.d.r. space, where \( P_{\hat{\lambda}_k} = \hat{b}_k \hat{b}_k' \Sigma \). Let \( \hat{\Omega} = \{\hat{\lambda}_1, \ldots, \hat{\lambda}_K\} \) be the set of the \( K \) largest eigenvalues of \( \hat{\Sigma}^{-1} \hat{M}_{\alpha, \rho} \). The \( \hat{\Sigma} \)-orthogonal eigenprojector onto the estimated e.d.r. space \( \hat{E} \) is \( \hat{P} = \sum_{\hat{\lambda}_k \in \hat{\Omega}} \hat{P}_{\hat{\lambda}_k} \) where \( \hat{P}_{\hat{\lambda}_k} = \hat{b}_k \hat{b}_k' \hat{\Sigma} \).

Theorem 1 gives the asymptotic distribution of \( \hat{\Sigma}^{-1} \hat{M}_{\alpha, \rho} \). Starting from this limit distribution, the asymptotic distributions of the eigenelements describing the estimated e.d.r. space are derived in Theorem 2. These eigenelements are the eigenprojector onto the estimated e.d.r. space, the estimated e.d.r. directions and their corresponding eigenvalues.

**Theorem 1.** Under assumptions (A1), (A2) and (A3),

\[
\sqrt{n} \left( \hat{\Sigma}^{-1} \hat{M}_{\alpha, \rho} - \Sigma^{-1} M_{\alpha, \rho} \right) \rightarrow_d \Phi,
\]
where $\Phi$ is such that $\text{vec}(\Phi)$ is normally distributed with mean zero and covariance matrix $C$ defined at (7).

**Theorem 2.** Under the assumptions (A1), (A2), (A3) and (A4), we have

(i) $\sqrt{n} \left( \hat{P} - P \right) \rightarrow_d \Phi_P$, where $\Phi_P$ is such that $\text{vec}(\Phi_P)$ is normally distributed with mean zero and covariance matrix $C_P$ defined at (10).

(ii) $\sqrt{n}(\hat{b}_k - b_k) \rightarrow_d \Phi_{b_k}$, where $\Phi_{b_k}$ has the normal distribution with mean zero and covariance matrix $C_{b_k}$ defined at (14).

(iii) $\sqrt{n}(\hat{\lambda}_k - \lambda_k) \rightarrow_d \Phi_{\lambda_k}$, where $\Phi_{\lambda_k}$ has a normal distribution with mean zero and variance $C_{\lambda_k} = [b_k' \otimes b_k' \Sigma][b_k \otimes \Sigma b_k]$.

These theorems generalize the results obtained by Gannoun and Saracco [10] in which the semiparametric model contains only one dependent variable $y$ (that is for $q = 1$).

**Acknowledgments**

The author is grateful to the associate editor and anonymous referee for contributing to the improvement of this paper through their useful remarks and detailed comments.

**Appendix A.**

Throughout the proof, $I_s$ denotes the $s \times s$ identity matrix, and $0_{s_1,s_2}$ stands for the $s_1 \times s_2$ null matrix.

**A.1. Proof of Theorem 1**

The proof relies on the Central Limit Theorem and the Delta method (see for example [21, Theorem A, p. 122]). It is divided into five steps.

**Step 1:** Application of the Central Limit Theorem.

For $i = 1, \ldots, n$ and $j = 1, \ldots, q$, define the $(dq + p + p^2)$-dimensional column vector

$$U_i^{(j)} = \left( \theta_{1i}^{(j)}, \ldots, \theta_{Hji}^{(j)}, x_i', \theta_{1i}^{(j)}, \ldots, x_i' \theta_{Hji}, \text{vec}(x_i'x_i' \theta_{1i}^{(j)}), \ldots, \text{vec}(x_i'x_i' \theta_{Hji}) \right)' .$$

Let $d_q = \sum_{j=1}^{q} H_j + p H_j + p^2 H_j$.

For $i = 1, \ldots, n$, define the $(d_q + p + p^2)$-dimensional column vector

$$U_i = \left( U_i^{(1)'}, \ldots, U_i^{(q)'}, x_i', \text{vec}(x_i'x_i')' \right)' .$$

Under (A1), the vectors $U_i, \ i = 1, \ldots, n$ are independent and identically distributed.

For $j = 1, \ldots, q$ and $h = 1, \ldots, H_j$, write $\tilde{m}_h^{(j)} = \mathbb{E} \left( \theta_{h}^{(j)} \right)$ and $\tilde{V}_h^{(j)} = \mathbb{E} \left( xx' \theta_{h}^{(j)} \right)$. 
Then the mean $\mu_U^{(j)}$ of $U_i^{(j)}$ is

$$\mu_U^{(j)} = \left( p_1^{(j)}, \ldots, p_{H_j}^{(j)}, \tilde{m}_1^{(j)}, \ldots, \tilde{m}_{H_j}^{(j)}, \text{vec}(\tilde{V}_1^{(j)}), \ldots, \text{vec}(\tilde{V}_{H_j}^{(j)}) \right)'.$$

Therefore, the mean $\mu_U$ of $U_i$ is

$$\mu_U = \left( \mu_U^{(1)'}, \ldots, \mu_U^{(q)'}, \mu', \text{vec}(\Sigma + \mu') \right)' .$$

To give the expression to the covariance matrix of $U_i$, we need additional notations:

- $M = \mathbb{E} \left( x(x' \otimes x') \right)$ and $N = \mathbb{E} \left( (xx') \otimes (xx') \right)$;
- for $j = 1, \ldots, q$ and $h = 1, \ldots, H_j$, $\tilde{M}_h^{(j)} = \mathbb{E} \left( x(x' \otimes x') \tilde{\eta}_h^{(j)} \right)$ and $\tilde{N}_h^{(j)} = \mathbb{E} \left( (xx') \otimes (xx') \tilde{\eta}_h^{(j)} \right)$;
- for $j \neq l$, $h = 1, \ldots, H_j$ and $k = 1, \ldots, H_k$, $\tilde{p}_{hk}^{(jl)} = \mathbb{E} \left( \tilde{\eta}_h^{(j)} \tilde{\eta}_k^{(l)} \right)$, $\tilde{m}_{hk}^{(jl)} = \mathbb{E} \left( x(x' \otimes x') \tilde{\eta}_h^{(j)} \tilde{\eta}_k^{(l)} \right)$, $\tilde{\eta}_{hk}^{(jl)} = \mathbb{E} \left( (xx') \otimes (xx') \tilde{\eta}_h^{(j)} \tilde{\eta}_k^{(l)} \right)$, and $\tilde{Q}_{hk}^{(jl)} = \mathbb{E} \left( \text{vec}(xx') \text{vec}(xx') \tilde{\eta}_h^{(j)} \tilde{\eta}_k^{(l)} \right)$.

The covariance matrix $\Sigma_U$ of $U_i$ is then

$$\Sigma_U = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{bmatrix},$$

where for $u, v = 1, 2$, the blocks $A_{uv}$ are the following:

- $A_{11} = [A_{jl}]$ where, for $j = 1, \ldots, q$,

$$[A_{jl}] = \begin{bmatrix} B_{11}^{(j)} & B_{12}^{(j)} & B_{13}^{(j)} \\ B_{12}' & B_{22}^{(j)} & B_{23}^{(j)} \\ B_{13}' & B_{23}' & B_{33}^{(j)} \end{bmatrix}$$

and for $j \neq l$,

$$[A_{jl}] = \begin{bmatrix} C_{11}^{(jl)} & C_{12}^{(jl)} & C_{13}^{(jl)} \\ C_{12}' & C_{22}^{(jl)} & C_{23}^{(jl)} \\ C_{13}' & C_{23}' & C_{33}^{(jl)} \end{bmatrix} .$$

For $u, v = 1, \ldots, 3$, the blocks $B_{uv}^{(j)}$ are the following:

$$B_{11}^{(j)} = \begin{bmatrix} p_1^{(j)} & -p_1^{(j)} & \ldots & -p_{H_j}^{(j)} \\ -p_2^{(j)} & p_2^{(j)} & \ldots & -p_{H_j}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ -p_1^{(j)} & -p_{H_j}^{(j)} & \ldots & -p_{H_j}^{(j)} \end{bmatrix} ,$$

$$B_{12}^{(j)} = \begin{bmatrix} 1 - p_1^{(j)} & -p_2^{(j)} & \ldots & -p_{H_j}^{(j)} \\ -p_1^{(j)} & 1 - p_2^{(j)} & \ldots & -p_{H_j}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ -p_1^{(j)} & -p_{H_j}^{(j)} & \ldots & 1 - p_{H_j}^{(j)} \end{bmatrix} .$$
\[ B_{12}^{(j)} = \begin{bmatrix}
(1 - p_1^{(j)} m_1^{(j)})' & -p_1^{(j)} m_2^{(j)}' & \ldots & -p_1^{(j)} m_H^{(j)}' \\
-p_2^{(j)} m_1^{(j)}' & (1 - p_2^{(j)} m_2^{(j)})' & \ldots & -p_2^{(j)} m_H^{(j)}' \\
\vdots & \vdots & \ddots & \vdots \\
-p_H^{(j)} m_1^{(j)}' & \ldots & \ldots & (1 - p_H^{(j)} m_H^{(j)})'
\end{bmatrix},
\]

\[ B_{13}^{(j)} = \begin{bmatrix}
(1 - p_1^{(j)}) \text{vec}(\tilde{V}_1^{(j)})' & -p_1^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & \ldots & -p_1^{(j)} \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
-p_2^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & (1 - p_2^{(j)}) \text{vec}(\tilde{V}_2^{(j)})' & \ldots & -p_2^{(j)} \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
\vdots & \vdots & \ddots & \vdots \\
-p_H^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & \ldots & \ldots & (1 - p_H^{(j)}) \text{vec}(\tilde{V}_{H_j}^{(j)})'
\end{bmatrix},
\]

\[ B_{22}^{(j)} = \begin{bmatrix}
\tilde{V}_1^{(j)} - m_1^{(j)} \tilde{m}_1^{(j)} & -m_1^{(j)} \tilde{m}_2^{(j)} & \ldots & -m_1^{(j)} \tilde{m}_H^{(j)} \\
-\tilde{m}_2^{(j)} \tilde{m}_1^{(j)} & \tilde{V}_2^{(j)} - m_2^{(j)} \tilde{m}_2^{(j)} & \ldots & -m_2^{(j)} \tilde{m}_H^{(j)} \\
\vdots & \vdots & \ddots & \vdots \\
-\tilde{m}_H^{(j)} \tilde{m}_1^{(j)} & \ldots & \ldots & \tilde{V}_{H_j}^{(j)} - m_H^{(j)} \tilde{m}_H^{(j)}
\end{bmatrix},
\]

\[ B_{23}^{(j)} = \begin{bmatrix}
\tilde{M}_1^{(j)} - m_1^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & -m_1^{(j)} \text{vec}(\tilde{V}_2^{(j)})' & \ldots & -m_1^{(j)} \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
-\tilde{m}_2^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & \tilde{M}_2^{(j)} - m_2^{(j)} \text{vec}(\tilde{V}_2^{(j)})' & \ldots & -m_2^{(j)} \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
\vdots & \vdots & \ddots & \vdots \\
-\tilde{m}_H^{(j)} \text{vec}(\tilde{V}_1^{(j)})' & \ldots & \ldots & \tilde{M}_{H_j}^{(j)} - m_H^{(j)} \text{vec}(\tilde{V}_{H_j}^{(j)})'
\end{bmatrix},
\]

\[ B_{33}^{(j)} = \begin{bmatrix}
\tilde{N}_1^{(j)} - \text{vec}(\tilde{V}_1^{(j)})' \text{vec}(\tilde{V}_1^{(j)})' & -\text{vec}(\tilde{V}_1^{(j)})' \text{vec}(\tilde{V}_2^{(j)})' & \ldots & -\text{vec}(\tilde{V}_1^{(j)})' \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
-\text{vec}(\tilde{V}_2^{(j)})' \text{vec}(\tilde{V}_1^{(j)})' & \tilde{N}_2^{(j)} - \text{vec}(\tilde{V}_2^{(j)})' \text{vec}(\tilde{V}_2^{(j)})' & \ldots & -\text{vec}(\tilde{V}_2^{(j)})' \text{vec}(\tilde{V}_{H_j}^{(j)})' \\
\vdots & \vdots & \ddots & \vdots \\
-\text{vec}(\tilde{V}_{H_j}^{(j)})' \text{vec}(\tilde{V}_1^{(j)})' & \ldots & \ldots & \tilde{N}_{H_j}^{(j)} - \text{vec}(\tilde{V}_{H_j}^{(j)})' \text{vec}(\tilde{V}_{H_j}^{(j)})'
\end{bmatrix}.
\]

For \( u, v = 1, \ldots, 3 \), the \((h, k)\) elements of the blocks \( C_{uv}^{(jl)} \) are the following:

\[
[C_{11}^{(jl)}]_{hk} = \tilde{p}_{hk}^{(j)} - p_h^{(j)} m_k^{(l)}' = \tilde{m}_{hk}^{(j)} - p_h^{(j)} m_k^{(l)}',
\]

\[
[C_{22}^{(jl)}]_{hk} = \tilde{V}_{hk}^{(j)} - m_h^{(j)} m_k^{(l)}',
\]

\[
[C_{23}^{(jl)}]_{hk} = \tilde{M}_{hk}^{(j)} - m_h^{(j)} \text{vec}(\tilde{V}_k^{(j)})',
\]

\[
[C_{33}^{(jl)}]_{hk} = \tilde{Q}_{hk}^{(j)} - \text{vec}(\tilde{V}_h^{(j)}) \text{vec}(\tilde{V}_k^{(j)})',
\]

\[
A_{12} = \begin{bmatrix}
B_1^{(1)} \\
\vdots \\
B_{21}^{(1)} \\
B_{22}^{(1)} \\
B_{31}^{(1)} \\
B_{32}^{(1)}
\end{bmatrix}, \text{ where } B^{(j)} = \begin{bmatrix}
B_{11}^{(j)} & B_{12}^{(j)} \\
B_{21}^{(j)} & B_{22}^{(j)} \\
B_{31}^{(j)} & B_{32}^{(j)}
\end{bmatrix} \text{ with}
\]
\[
B_{11}^{(j)} = \begin{bmatrix} \tilde{m}_1^{(j)} - p_1^{(j)} \mu' \\ \vdots \\ \tilde{m}_{H_j}^{(j)} - p_{H_j}^{(j)} \mu' \\ \end{bmatrix}, \\
B_{12}^{(j)} = \begin{bmatrix} \text{vec}(\tilde{V}_1^{(j)})' - p_1^{(j)} \text{vec}(\Sigma + \mu \mu') \\ \vdots \\ \text{vec}(\tilde{V}_{H_j}^{(j)})' - p_{H_j}^{(j)} \text{vec}(\Sigma + \mu \mu') \\ \end{bmatrix}, \\
B_{21}^{(j)} = \begin{bmatrix} \tilde{V}_1^{(j)} - \tilde{m}_1^{(j)} \mu' \\ \vdots \\ \tilde{V}_{H_j}^{(j)} - \tilde{m}_{H_j}^{(j)} \mu' \\ \end{bmatrix}, \\
B_{22}^{(j)} = \begin{bmatrix} \tilde{M}_1^{(j)} - \tilde{m}_1^{(j)} \text{vec}(\Sigma + \mu \mu') \\ \vdots \\
\tilde{M}_{H_j}^{(j)} - \tilde{m}_{H_j}^{(j)} \text{vec}(\Sigma + \mu \mu') \\ \end{bmatrix}, \\
B_{31}^{(j)} = \left[ \tilde{M}_1^{(j)} - \mu \text{vec}(\tilde{V}_1^{(j)})' \ldots \tilde{M}_{H_j}^{(j)} - \mu \text{vec}(\tilde{V}_{H_j}^{(j)})' \right], \\
B_{32}^{(j)} = \left[ \tilde{N}_1^{(j)} - \text{vec}(\Sigma + \mu \mu') \text{vec}(\tilde{V}_1^{(j)})' \ldots \tilde{N}_{H_j}^{(j)} - \text{vec}(\Sigma + \mu \mu') \text{vec}(\tilde{V}_{H_j}^{(j)})' \right]. \\
\]

- \[ A_{22} = \begin{bmatrix} \Sigma & M - \mu \text{vec}(\Sigma + \mu \mu') \\ (M - \mu \text{vec}(\Sigma + \mu \mu'))' & N - \text{vec}(\Sigma + \mu \mu') \text{vec}(\Sigma + \mu \mu')' \\ \end{bmatrix}. \]

From the Central Limit Theorem, \( \sqrt{n} \left( \bar{U} - \mu_U \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_U) \), where \( \bar{U} = n^{-1} \sum_{i=1}^{n} U_i \).

**Step 2:** Asymptotic distribution of some intermediate random variables.

In order to use the Delta method, we stack the variables comprising the matrices \( \hat{M}_1^{(j)}, \hat{M}_{II}^{(j)} \) and \( \hat{\Sigma} \) into the vector

\[
\bar{U}_1 = \left( \bar{U}^{(1)}, \ldots, \bar{U}^{(q)}, \bar{x}', \text{vec}(xx') \right)',
\]

where

\[
\bar{U}^{(j)} = \left( \bar{\gamma}_1^{(j)}, \ldots, \bar{\gamma}_{H_j}^{(j)}, (xx')_{1}^{(j)}/\bar{\gamma}_1^{(j)}', \ldots, (xx')_{H_j}^{(j)}/\bar{\gamma}_{H_j}^{(j)}' \right)'.
\]

We define the function \( f_1 \) from \( \mathbb{R}^{d_1 + p + p^2} \) to \( \mathbb{R}^{d_1 + p + p^2} \) by

\[
f_1 \left( (u^{(1)}', \ldots, u^{(q)}', d', e')' \right) = \left( u_1^{(1)}', \ldots, u_1^{(q)}', d', e' \right)',
\]

where

\[
\begin{align*}
\bar{\gamma}_1^{(j)} & = \left( \bar{\gamma}_1^{(j)}, \ldots, \bar{\gamma}_{H_j}^{(j)}, (xx')_{1}^{(j)}/\bar{\gamma}_1^{(j)}', \ldots, (xx')_{H_j}^{(j)}/\bar{\gamma}_{H_j}^{(j)}' \right)'. \\
\end{align*}
\]
where
\[ u^{(j)} = (a^{(j)} r, b_{1}^{(j)} r, \ldots, b_{H_{j}}^{(j)} r, c_{1}^{(j)} r, \ldots, c_{H_{j}}^{(j)} r)^{\prime} \]
and
\[ u_{1}^{(j)} = (a^{(j)} r, b_{1}^{(j)} r / a_{1}, \ldots, b^{(j)} H_{j} r / a_{1}, c_{1}^{(j)} r / a_{1}, \ldots, c^{(j)} H_{j} r / a_{1})^{\prime} \]
with \( a^{(j)} = (a_{1}, \ldots, a_{H_{j}}) \in \mathbb{R}^{H_{j}} \) (assumed non-null), \( b^{(j)} \in \mathbb{R}^{p} \) and \( c^{(j)} \in \mathbb{R}^{p^{2} H_{j}} \) are column vectors.

Under (A2), it is clear that \( U_{1} = f_{1} (U) \).

Let us define
\[ \mu_{1} = f_{1} (\mu_{U}) = \left( \mu_{1}^{(1)}, \ldots, \mu_{1}^{(q)}, \mu^{\prime}, \text{vec}(\Sigma + \mu \mu^{\prime}) \right)^{\prime} \]
where
\[ \mu_{1}^{(j)} = \left( p_{1}^{(j)}, \ldots, p_{H_{j}}^{(j)}, m_{1}^{(j)} r, \ldots, m_{H_{j}}^{(j)} r, \text{vec}(\tilde{V}_{H_{j}}^{(j)}) / p_{1}^{(j)}, \ldots, \text{vec}(\tilde{V}_{H_{j}}^{(j)}) / p_{H_{j}}^{(j)} \right)^{\prime} \]

Let us also define \( \Sigma_{1} = F_{1}^{\prime} \Sigma_{U} F_{1} \), where \( F_{1} = \partial f_{1} / \partial u |_{E} \). Here and subsequently the notation \( g |_{E} \) is the evaluation of \( g \) at the expectation of its argument. After straightforward calculations, we get

\[
F_{1} = \begin{bmatrix}
\tilde{F}_{1}^{(1)} & 0_{d_{q},p} & 0_{d_{q},p^{2}} \\
\vdots & \ddots & \vdots \\
\tilde{F}_{1}^{(q)} & 0_{d_{q},p} & 0_{d_{q},p^{2}} \\
0_{p,d_{q}} & I_{p} & 0_{p,p^{2}} \\
0_{p^{2},d_{q}} & 0_{p^{2},p} & I_{p^{2}}
\end{bmatrix},
\]

where
\[
\tilde{F}_{1}^{(j)} = \begin{bmatrix}
I_{H_{j}} & -\frac{\tilde{m}^{(j)}}{(p_{1}^{(j)})^{2}} & -\frac{\text{vec}(\tilde{V}_{H_{j}}^{(j)})}{(p_{1}^{(j)})^{2}} \\
0_{p H_{j},H_{j}} & I_{p} / p_{1}^{(j)} & 0_{p H_{j},p^{2} H_{j}} \\
0_{p^{2} H_{j},H_{j}} & I_{p / p_{H_{j}}} & I_{p^{2} / p_{1}^{(j)}}
\end{bmatrix}.
\]
Since $f_1$ satisfies the required conditions of the Delta method theorem and from Step 1, we get
\[ \sqrt{n} \left( \bar{U}_1 - \mu_1 \right) \longrightarrow_d N(0, \Sigma_1). \]

**Step 3:** Asymptotic distribution of the vector stacking the matrices $\hat{M}_1^{(j)}, \hat{K}_h^{(j)}$ and $\hat{\Sigma}$. Let us define $f_2$ on $\mathbb{R}^{dq+p+p^2}$ to $\mathbb{R}^{qp^2+p^2 \sum_{j=1}^q H_j+p^2}$ as follows:
\[ f_2 \left( (v^{(1)}', \ldots, v^{(q)}', d', e')' \right) = \left( v_1^{(1)}', \ldots, v_1^{(q)}', (d - \text{vec}(ee'))' \right)', \]
where $v^{(j)} = (a^{(j)}', b_1^{(j)}, \ldots, b_{H_j}^{(j)}, c_1^{(j)}, \ldots, c_{H_j}^{(j)})'$ and
\[
 v_1^{(j)} = \begin{pmatrix}
 \frac{H_j}{\sqrt{a_H^{(j)}}} \left( e_1^{(j)} - \text{vec}(b_1^{(j)} b_1^{(j)'} - \sum_{j=1}^{H_j} a_j^{(j)} [c_j^{(j)} - \text{vec}(b_j^{(j)} b_j^{(j)'}))] \right) \\
 \vdots \\
 \frac{1}{\sqrt{a_H^{(j)}}} \left( e_H^{(j)} - \text{vec}(b_H^{(j)} b_H^{(j)'} - \sum_{j=1}^{H_j} a_j^{(j)} [c_j^{(j)} - \text{vec}(b_j^{(j)} b_j^{(j)'}))] \right)
\end{pmatrix}.
\]

It is clear that
\[ f_2 \left( \bar{U}_1 \right) = \left( \text{vec}(\hat{M}_1^{(1)}'), \text{vec}(\hat{K}_1^{(1)}'), \ldots, \text{vec}(\hat{K}_{H_1}^{(1)}'), \ldots, \right. \]
\[ \text{vec}(\hat{M}_1^{(q)}'), \text{vec}(\hat{K}_1^{(q)}'), \ldots, \text{vec}(\hat{K}_{H_1}^{(q)}'), \text{vec}(\hat{\Sigma})' \left. \right) \]
and
\[ f_2(\mu_1) = \left( \text{vec}(M_1^{(1)}'), \text{vec}(K_1^{(1)}'), \ldots, \text{vec}(K_{H_1}^{(1)}'), \ldots, \text{vec}(M_1^{(q)}'), \right. \]
\[ \left. \text{vec}(K_1^{(q)}'), \ldots, \text{vec}(K_{H_1}^{(q)}'), \text{vec}(\Sigma)' \right). \]

Let us also define $\Sigma_2 = F_2' \Sigma_1 F_2$

where $F_2 = \left. \frac{\partial f_2'}{\partial u} \right|_E = \begin{bmatrix}
 \hat{F}_2^{(1)} & \cdots & 0_{dq, p^2} \\
 \hat{E}_2^{(1)} & \hat{F}_2^{(q)} & 0_p \\
 0_{p^2, p^2+q^{\sum_{j=1}^q H_j}} & I_{p^2} & \hat{F}_2^{(j)} & \cdots & \hat{F}_2^{(j)} \\
 \end{bmatrix}$, $\hat{F}_2^{(j)} = \begin{bmatrix}
 E_1^{(j)} & E_2^{(j)} \\
 E_3^{(j)} & E_4^{(j)} \\
 0_{p^2 H_1, p^2} & E_5^{(j)} \\
 \end{bmatrix}$.
and 

\[ F_2^{(j)} = \left[ E_6^{(j)} \ 0_{p_p^2H_j} \right], \]

with

\[
E_1^{(j)} = \begin{bmatrix}
(m_1^{(j)} - \mu)' \otimes (m_1^{(j)} - \mu)'

\vdots

(m_{H_j}^{(j)} - \mu)' \otimes (m_{H_j}^{(j)} - \mu)'
\end{bmatrix},
\]

\[
E_2^{(j)} = \frac{1}{2p_1^{(j)}} [ (1 - 3p_1^{(j)}) \text{vec}(V_1^{(j)\prime} - \xi_1^{(j)}) - \sqrt{p_2^{(j)}} \text{vec}(V_1^{(j)\prime}) \ldots - \sqrt{p_{H_j}^{(j)}} \text{vec}(V_1^{(j)\prime})],
\]

\[
E_3^{(j)} = \begin{bmatrix}
p_1^{(j)} [ I_p \otimes (m_1^{(j)} - \mu)' + (m_1^{(j)} - \mu)' \otimes I_p]

\vdots

p_{H_j}^{(j)} [ I_p \otimes (m_{H_j}^{(j)} - \mu)' + (m_{H_j}^{(j)} - \mu)' \otimes I_p]
\end{bmatrix},
\]

\[
E_4^{(j)} = \begin{bmatrix}
\sqrt{p_1^{(j)}} (p_1^{(j)} - 1) \xi_1^{(j)}

\sqrt{p_2^{(j)}} p_1^{(j)} \xi_1^{(j)}

\vdots

\sqrt{p_{H_j}^{(j)}} p_1^{(j)} \xi_1^{(j)}
\end{bmatrix},
\]

\[
E_5^{(j)} = \begin{bmatrix}
\sqrt{p_1^{(j)}} (p_1^{(j)} - 1) I_{p_2} 

\sqrt{p_2^{(j)}} p_1^{(j)} I_{p_2} 

\vdots

\sqrt{p_{H_j}^{(j)}} (p_1^{(j)} - 1) I_{p_2}
\end{bmatrix},
\]

\[
E_6^{(j)} = \sum_{h=1}^{H_j} p_h^{(j)} \left( I_p \otimes (\mu - m_h^{(j)})' + (\mu - m_h^{(j)})' \otimes I_p \right),
\]

\[
E_7 = -(I_p \otimes \mu' + \mu' \otimes I_p)
\]

and, for \( h = 1, \ldots, H_j \), \( \xi_h^{(j)} = \sum_{k \neq h} p_k^{(j)} \text{vec}(V_k^{(j)\prime})' \) and \( \xi_h^{(j)} = (I_p \otimes m_h^{(j)\prime} + m_h^{(j)\prime} \otimes I_p) \).

Since \( f_2 \) satisfies the required conditions of the Delta method theorem and from Step 2, we get

\[ \sqrt{n} \left( f_2(U_1) - f_2(\mu_1) \right) \longrightarrow_d N(0, \Sigma_2). \]

**Step 4:** Asymptotic distribution of the vector stacking the matrices \( \hat{M}_1^{(j)}, \hat{K}_h^{(j)} \) and \( \hat{\Sigma}^{-1} \). Let \( R \) be the matrix
\[
R = \begin{bmatrix}
I_{p^2+p^2 \sum_{j=1}^{q} H_j} & 0_{p^2+p^2 \sum_{j=1}^{q} H_j, p^2} \\
0_{p^2+p^2 \sum_{j=1}^{q} H_j, p^2} & (\Sigma^{-1} \otimes \Sigma^{-1})
\end{bmatrix}.
\]

Under (A3), and using the first order approximation \(\hat{\Sigma}^{-1} = \Sigma^{-1} - \Sigma^{-1} (\hat{\Sigma} - \Sigma) \Sigma^{-1}\) and Step 3, we get

\[
\sqrt{n} \begin{bmatrix}
\text{vec}(\hat{M}_1^{(1)}) \\
\text{vec}(\hat{K}_1^{(1)}) \\
\vdots \\
\text{vec}(\hat{M}_1^{(q)}) \\
\text{vec}(\hat{K}_1^{(q)}) \\
\vdots \\
\text{vec}(\hat{M}_H^{(q)}) \\
\text{vec}(\hat{K}_H^{(q)}) \\
\vdots \\
\text{vec}(\hat{\Sigma})
\end{bmatrix} - \begin{bmatrix}
\text{vec}(M_1^{(1)}) \\
\text{vec}(K_1^{(1)}) \\
\vdots \\
\text{vec}(M_1^{(q)}) \\
\text{vec}(K_1^{(q)}) \\
\vdots \\
\text{vec}(M_H^{(q)}) \\
\text{vec}(K_H^{(q)}) \\
\vdots \\
\text{vec}(\Sigma)
\end{bmatrix},
\]

which converges in distribution to \(N(0, C_R)\) where \(C_R = R \Sigma_2 R'\).

**Step 5:** Asymptotic distribution of \(\sqrt{n} \text{vec} \left( \Sigma^{-1} \hat{M}_{\alpha, p} - \Sigma^{-1} M_{\alpha, p} \right)\).

For the \(p \times p\) matrices \(A^{(j)}, B_1^{(j)}, \ldots, B_{H_j}^{(j)}\) and \(C\), let \(f_3\) be defined from \(\mathbb{R}^{p^2(2+\sum_{j=1}^{q} H_j)}\) to \(\mathbb{R}^{p^2}\) by

\[
f_3 \left( \left(\text{vec}(A^{(1)})', \text{vec}(B_1^{(1)})', \ldots, \text{vec}(B_{H_1}^{(1)})', \ldots, \text{vec}(A^{(q)})', \text{vec}(B_1^{(q)})', \ldots, \right. \right.
\]

\[
\left. \left. \text{vec}(B_{H_q}^{(q)})', \text{vec}(C)' \right) \right) = \sum_{j=1}^{q} w_j \text{vec} \left( C \left[ (1 - \alpha_j) A^{(j)} C A^{(j)} + \alpha_j \sum_{h=1}^{H_j} B_h^{(j)} C B_h^{(j)} \right] \right).
\]
It is clear that
\[
f_3 \left( \left( \text{vec}(\hat{M}(1)_1)', \text{vec}(\hat{K}(1)_1)', \ldots, \text{vec}(\hat{M}(q)_1)', \text{vec}(\hat{K}(1)_1)', \ldots, \text{vec}(\hat{K}(1)_{H_1})', \text{vec}(\hat{\Sigma}^{-1})' \right) \right) = \text{vec} \left( \hat{\Sigma}^{-1} \hat{M}_{x,p} \right)
\]
and
\[
f_3 \left( \left( \text{vec}(M(1)_1)', \text{vec}(K(1)_1)', \ldots, \text{vec}(M(q)_1)', \text{vec}(K(1)_1)', \ldots, \text{vec}(K(1)_{H_1})', \text{vec}(\Sigma^{-1})' \right) \right) = \text{vec} \left( \Sigma^{-1} M_{x,p} \right).
\]

Let us also define
\[
\mathcal{C} = F_3' C R F_3,
\]
where
\[
F_3 = \frac{\hat{c} f_3'}{\hat{u}} \bigg|_E = \begin{bmatrix} \hat{F}_3(1) \\ \vdots \\ \hat{F}_3(q) \\ E_8 \end{bmatrix}
\]
with
\[
\hat{F}_3(j) = w_j \begin{bmatrix} (1 - \alpha_j)(I_p \otimes \Sigma^{-1} M_\alpha(j) + M_\alpha(j) \Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \\ \alpha_j(I_p \otimes \Sigma^{-1} K(j) + K(j) \Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \\ \vdots \\ \alpha_j(I_p \otimes \Sigma^{-1} K(j)_H + K(j)_H \Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \end{bmatrix}
\]
and
\[
E_8 = \sum_{j=1}^q w_j \left( (1 - \alpha_j)(I_p \otimes \Sigma^{-1} M_\alpha(j) + M_\alpha(j) \Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) + \alpha_j \sum_{h=1}^H (I_p \otimes \Sigma^{-1} K(j)_h + K(j)_h \Sigma^{-1} \otimes I_p)(I_p \otimes \Sigma^{-1}) \right).
\]

Since the Delta method applies to $f_3$ and, from Step 4, a final application of the Delta method leads to
\[
\sqrt{n} \text{vec} \left( \hat{\Sigma}^{-1} \hat{M}_{x,p} - \Sigma^{-1} M_{x,p} \right) \to_d \mathcal{N}(0, \mathcal{C}).
\]

**A.2. Proof of Theorem 2**

Let $M$ be a square matrix of order $p$. We set $||M|| = \left[ \max \text{ eigenvalue of } (\Sigma^{-1} M' \Sigma M) \right]^{1/2}$, as in [22]. The Moore–Penrose generalized inverse of $M$ is denoted by $M^+$. 
Thus with probability 1, for $n$ sufficiently large, we have
\[
\left\| \hat{\Sigma}^{-1} \hat{M}_{z,p} - \Sigma^{-1} M_{z,p} \right\| \leq \lambda_K / 2.
\] (9)

(i) From Theorem 1 and (9), we are now in position to apply Lemma 4.1 of Tyler [22], so
\[
\hat{p} = p - \sum_{\lambda_k \in \Omega} \left[ P_{\lambda_k} \left( \hat{\Sigma}^{-1} \hat{M}_{z,p} - \Sigma^{-1} M_{z,p} \right) \left( \Sigma^{-1} M_{z,p} - \lambda_k I_p \right)^+ \right.
\]
\[
+ \left( \Sigma^{-1} M_{z,p} - \lambda_k I_p \right)^+ \left( \hat{\Sigma}^{-1} \hat{M}_{z,p} - \Sigma^{-1} M_{z,p} \right) P_{\lambda_k} \right] + \hat{E}_o,
\]
where
\[
\left\| \hat{E}_o \right\| \leq \left( 1 + \frac{\lambda_1 - \lambda_K}{\lambda_K} \right) \left( \frac{2}{\lambda_K} \left\| \hat{\Sigma}^{-1} \hat{M}_{z,p} - \Sigma^{-1} M_{z,p} \right\| \right)^2
\]
\[
\times \left( 1 - \frac{2}{\lambda_K} \left\| \hat{\Sigma}^{-1} \hat{M}_{z,p} - \Sigma^{-1} M_{z,p} \right\| \right)^{-1}.
\]
Let $\Phi_p = -\sum_{\lambda_k \in \Omega} \left[ P_{\lambda_k} (\hat{\Sigma}^{-1} M_{z,p} - \lambda_k I_p)^+ + (\Sigma^{-1} M_{z,p} - \lambda_k I_p)^+ P_{\lambda_k} \right]$. From the above, it follows that $\sqrt{n} \left( \hat{p} - p \right) \rightarrow_d \Phi_p$. We remark that $\text{vec}(\Phi_p) = C \text{vec}(\Phi)$ where $C = -\sum_{\lambda_k \in \Omega} \left[ (M_{z,p} \Sigma^{-1} - \lambda_k I_p)^+ \otimes P_{\lambda_k} + P_{\lambda_k}' \otimes (\Sigma^{-1} M_{z,p} - \lambda_k I_p)^+ \right]$. Note that $(\Sigma^{-1} M_{z,p} - \lambda_k I_p)^+$ can be replaced by $S_{\lambda_k} = \sum_{\lambda_l \neq \lambda_k} \frac{1}{\lambda_l - \lambda_k} P_{\lambda_l}$. Then $\text{vec}(\Phi_p)$ follows the normal distribution $N(0, C_p)$ where
\[
C_p = C \Omega C' \Omega.
\] (10)

(ii) This result is a straightforward application of the Lemma 2 of Saracco [18]. We first show that
\[
\sqrt{n} \left( \begin{bmatrix} \text{vec} (\hat{\Sigma}^{-1} \hat{M}_{z,p}) \\ \text{vec} (\Sigma) \end{bmatrix} - \begin{bmatrix} \text{vec} (\Sigma^{-1} M_{z,p}) \\ \text{vec}(\Sigma) \end{bmatrix} \right) \rightarrow_d \Phi^*,
\] (11)
where
\[
\Phi^* = \begin{pmatrix} \text{vec}(\Phi) \\ \text{vec}(\Phi)' \end{pmatrix}
\]
follows the normal distribution $N(0, C^*)$ with $C^*$ given by (13). The proof of this results is based on slight modifications of Step 4 and Step 5 in the proof of Theorem 1.

Take
\[
R^* = \begin{bmatrix} R \\ 0_{p^2, p^2} (1 + \sum_{j=1}^q H_j) \end{bmatrix} I_p^2,
\]
where \( R \) is defined in (6). Then, by the use of Step 3 in the proof of Theorem 1,
\[
\sqrt{n} \left( \vec{\text{vec}}(\hat{M}^{(1)}_1), \vec{\text{vec}}(\hat{K}^{(1)}_1), \ldots, \vec{\text{vec}}(\hat{K}^{(q)}_H), \ldots, \vec{\text{vec}}(\hat{K}^{(1)}_1), \ldots, \right.
\]
\[
\vec{\text{vec}}(\hat{K}^{(q)}_H), \vec{(\Sigma^{-1})'}, \vec{(\Sigma)'}, - \left( \vec{\text{vec}}(M^{(1)}_1), \vec{\text{vec}}(K^{(1)}_1), \ldots, \vec{\text{vec}}(K^{(1)}_H), \ldots, \right.
\]
\[
\vec{\text{vec}}(M^{(q)}_1), \vec{\text{vec}}(K^{(q)}_1), \ldots, \vec{\text{vec}}(K^{(q)}_H), \vec{(\Sigma^{-1})'}, \vec{(\Sigma)'}, \right) \]
\[
\equiv R^* \sqrt{n} \left( \vec{\text{vec}}(\hat{M}^{(1)}_1), \vec{\text{vec}}(\hat{K}^{(1)}_1), \ldots, \vec{\text{vec}}(\hat{K}^{(q)}_H), \ldots, \vec{\text{vec}}(\hat{K}^{(1)}_1), \ldots, \right.
\]
\[
\vec{\text{vec}}(M^{(q)}_1), \vec{\text{vec}}(K^{(q)}_1), \ldots, \vec{\text{vec}}(K^{(q)}_H), \vec{(\Sigma^{-1})'}, \vec{(\Sigma)'}, \right) \]
\[
\longrightarrow d \mathcal{N}(0, C_{R^*}), \quad \text{(12)}
\]
where \( C_{R^*} = R^* \Sigma_2 R^* \).

For the \( p \times p \) matrices \( A_1^{(1)}, \ldots, B_1^{(1)}, C, D \), let \( f_3^* \) be defined from \( \mathbb{R}^{p^2(2+q+\sum_{j=1}^q H_j)} \) to \( \mathbb{R}^{p^2+p^2} \) by
\[
f_3^* \left( \begin{bmatrix} \vec{\text{vec}}(A^{(1)}), \vec{\text{vec}}(B^{(1)}_1), \ldots, \vec{\text{vec}}(B^{(q)}_H), \ldots, \vec{\text{vec}}(A^{(q)}), \vec{\text{vec}}(B^{(q)}_1), \ldots, \vec{\text{vec}}(C), \vec{\text{vec}}(D) \end{bmatrix} \right)
\]
\[
\quad = \frac{q}{\sum_{j=1}^q \text{vec}(C)((1 - \alpha_j)A^{(j)}CA^{(j)} + \alpha_j \sum_{h=1}^{H_j} B^{(j)}_hCB^{(j)}_h)} \left( \frac{\text{vec}(\Sigma^{-1}\hat{M}_{2,p})}{\text{vec}(\Sigma)} \right).
\]

It is clear that
\[
f_3^* \left( \begin{bmatrix} \vec{\text{vec}}(\hat{M}^{(1)}_1), \vec{\text{vec}}(\hat{K}^{(1)}_1), \ldots, \vec{\text{vec}}(\hat{K}^{(q)}_H), \ldots, \vec{\text{vec}}(\hat{M}^{(1)}_1), \ldots, \right.
\]
\[
\vec{\text{vec}}(\hat{K}^{(q)}_H), \vec{(\Sigma^{-1})'}, \vec{(\Sigma)'}, \right) = \left( \begin{bmatrix} \vec{\text{vec}}(\Sigma^{-1}\hat{M}_{2,p}) \end{bmatrix} \right).
\]

and
\[
f_3^* \left( \begin{bmatrix} \vec{\text{vec}}(M^{(1)}_1), \vec{\text{vec}}(K^{(1)}_1), \ldots, \vec{\text{vec}}(K^{(q)}_H), \ldots, \vec{\text{vec}}(M^{(q)}_1), \ldots, \right.
\]
\[
\vec{\text{vec}}(K^{(q)}_H), \vec{(\Sigma^{-1})'}, \vec{(\Sigma)'}, \right) = \left( \begin{bmatrix} \vec{\text{vec}}(\Sigma^{-1}\hat{M}_{2,p}) \end{bmatrix} \right).
\]

Let us also define
\[
C^* = F_3^{*'}C_{R^*}F_3^*, \quad \text{(13)}
\]
where
\[
F_3^* = \left[ \frac{\partial f_3^{*'}}{\partial u} \right]_E = \begin{bmatrix} F_3 & 0_{p^2(1+q+\sum_{j=1}^q H_j), p^2} \\ 0_{p^2, p^2} & I_{p^2} \end{bmatrix}
\]
with $F_3$ defined in (8). Since the Delta method applies to $f_{3}^{*}$ and, from (12), a final application of the Delta method leads to (11).

Therefore, we get 

$$
\sqrt{n}(\hat{b}_k - b_k) \xrightarrow{d} \Phi b_k
$$

where

$$
\Phi b_k = M_k^* \Phi^* = (\Sigma^{-1} M_{d,p} - \hat{\lambda}_k I_p)^+ \Phi b_k - \frac{1}{2} (b_k' \Phi \Sigma b_k) b_k
$$

with $M_k^* = \left[ b_k' \otimes (\Sigma^{-1} M_{d,p} - \hat{\lambda}_k I_p)^+ - \frac{1}{2} b_k (b_k' \otimes b_k') \right]$. Then $\Phi b_k$ follows the normal distribution $\mathcal{N}(0, C_{b_k})$ where

$$
C_{b_k} = M_k^* \Phi^* M_k'^*.
$$

(iii) Using Theorem 1, Corollary 4 of Saracco [18] gives $\Phi \lambda_k = b_k' \Sigma b_k$. Then we use only the fact that $\Phi \lambda_k = (b_k' \otimes b_k') \text{vec}(\Phi)$ to complete the proof.

References