Bilinear estimation of pollution source profiles in receptor models

Eun Sug Park   Clifford H Spiegelman   Ronald C. Henry

NRCSE

Technical Report Series

NRCSE-TRS No. 019

Bilinear estimation of pollution source profiles in receptor models

Eun Sug Park[1], Clifford H. Spiegelman[2], and Ronald C. Henry[3]

1. University of Washington

National Research Center for Statistics and the Environment

Seattle, WA 98195-1720

2. Texas A&M University

Department of Statistics

College Station, TX 77843-3143

3. University of Southern California

Department of Civil & Environmental Engineering

3620 Sourth Vermont Ave.

Los Angeles, CA 90089-2531

[+]*Address for correspondence*: NRCSE, University of Washington, Box 351720, Seattle, WA 98195-1720

E-mail: epark@stat.washington.edu

Summary

Receptor models aim to identify the pollution sources based on air pollution data. This article is concerned with estimation of the source profiles (pollution recipes) and their contributions (amounts of pollution). We take a constrained nonlinear least squares approach. To avoid having infinitely many solutions, we present new sets of model identifiability conditions, which are often reasonable in practice. The resulting estimators are shown to be consistent and asymptotically normal under appropriate identifiability conditions. Simulations and an application to real air pollution data illustrate the results.

Key words: Receptor model; Model identifiability; Constrained nonlinear least squares; Consistency; Asymptotic normality; VERTEX.

1. Introduction

Receptor modeling is a collection of methods used to model air pollution data. Air quality data typically consists of concentrations on fifty or sixty compounds of airborne gases or particles measured over time. The basic assumptions in receptor modeling is conservation of mass and chemical mass balance (see, e.g., Hopke, 1985, 1991). If there are q pollution sources, the i^{th} measurement from the receptor, $y_i = (y_{i1}, y_{i2}, \ldots ,y_{ip})$, can be represented as

$$y_i = \sum_{k=1}^{q} \alpha_{ik} P_k + \varepsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

where $P_k = (p_{k1}, p_{k2}, \ldots , p_{kp})$ is the k^{th} source profile which consists of the fractional amount of each species in the emissions from the k^{th} source, α_{ik} is the contribution from the k^{th} source on the i^{th} day, and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots , \varepsilon_{ip})$ is the measurement error on the i^{th} observation. For example, a profile for a refinery might look like Propane, 21%; n-Butane, 18%; i-Pentane, 17%; n-Pentane, 7%; 2-Methylpentane, 7%; other chemical species, 30%. The objectives in receptor modeling are to identify pollution sources and assess the contribution of each source based on this data. There have been two traditional approaches to receptor modeling, which are the chemical mass balance (CMB) receptor model and multivariate receptor model (Hopke, 1991). In CMB, the number of sources, q, and the source profiles, P_k's, are assumed known, and the main objective is to estimate the source contributions, α_{ik}'s. In that case, the problem reduces to the ordinary linear least squares regression. Several examples of CMB methods such as tracer element method, linear programming method, ordinary linear least squares method, efficient variance least-squares method, principal component regression method, and ridge regression method can be found in Henry *et al.* (1984) and Hopke (1985). The CMB methods are performed on one observation at a time. The CMB assumptions, however, on the known number of sources and the known profiles are often not useful in practice. Such limitations of CMB leads to the use of multivariate analysis in receptor modeling. In matrix terms, the model (1) can be written as

$$Y = AP + E \qquad (2)$$

where

A: $n \times q$ source contribution matrix

P: $q \times p$ source composition matrix

E: $n \times p$ *error matrix.*

Examples of those multivariate models include principal component analysis, factor analysis, target transformation factor analysis, self-modeling curve resolution, and so on (see, e.g., Henry 1991). The advantage of multivariate receptor modeling is that it does not require a priori knowledge of the source characteristics. Multivariate receptor modeling tries to get the estimates for the number of sources, q, their profiles, P, and contributions, A, all together from data. However, this goal cannot be easily achieved since there could be infinitely many solutions for A and P even with the known number of sources. There have been some attempts to avoid this problem by placing the constraints on the parameters (see Henry and Kim 1990; Yang 1994). Those constraints can be obtained from prior knowledge of the problem under study or from the data itself. This issue will be addressed in terms of model identifiability in more general in Section 2.

The first method developed by a statistician in receptor modeling field was Source Apportionment with one Source Unknown (SASU) by Bandeen-Roche and Ruppert (1991). They supposed that q = 2 and one source profile is known and one is unknown. They treated the source contributions as random quantities having a Dirichlet (Beta in the case of $q = 2$) distribution, and tried to estimate the unknown source profile and the parameters of the distribution of source contributions by assuming that in the limit the unknown source is observed. Spiegelman and Dattner (1993) tried a related estimate. They wrote each $P_k = (s_{1k}, ..., s_{pk})$ as p two dimensional probability mass functions, $(s_{jk},$

$1\text{-}s_{jk}$), $j=1,\ldots,p$. Then the ratio of probability masses given to a species on two different days is calculated. If that ratio is extreme (either big or small) then a candidate for a two dimensional source profile is found. Either the methods found in Bandeen-Roache and Ruppert (1991) or those found in Spiegelman and Dattner (1993) are examples of tracer methods looking for single species that is indicative of a single pollution source. The assumption of having tracer element for each source makes any possible source rotation or transformation impossible, and identifiability of model parameters is automatically achieved. Unfortunately, this assumption could be unrealistic in practice since big cities have a number of pollution species that do not occur by themselves. Yang (1994) tried confirmatory factor analysis model (see, e.g., Anderson 1984, sec. 14.2.2) under the assumptions that the number and types of contributing sources are known a priori. He treated the source contributions as the random vectors having a distribution with some unknown mean vector γ and covariance matrix Φ, and showed the estimators obtained by maximizing an objective function are consistent and asymptotically normal. His objective function is actually the log-likelihood function of the observations when they follow a multivariate normal distribution although he makes no normality assumptions about the observations. As a matter of fact, many environmental engineers want to view the source contributions as fixed parameters not random variables, and the assumption of prior knowledge of the types of all sources in the model is not a comfortable assumption.

2. Idenfifiability of the model parameters

The number of sources, q, needs to be determined. We are concerned with the number of major pollution sources not the number of all pollution sources since there could be millions of sources in nature, and it would be impossible or meaningless if we try to identify all of those sources. Therefore, q means the number of major pollution sources hereafter. Many air pollution datasets typically consist of the measurements on fifty or sixty variables (VOC chemical species). The data set is often too large to handle all at once. Furthermore, not all of the species are helpful in finding the major pollution products. In many environmental applications some species have a few common major sources and some have many more minuscule sources. If the species used in estimating q come from different sets of sources each with different number of sources the estimated number of sources is not likely to be interpretable. It is crucial to select an appropriate set of species to estimate the number of major sources, q. It could be done by environmental expert's judgements or in the lack of such source, by species selection algorithms such as SPECIESA or SPECIESB (see Park 1997).

In this section we assume that an appropriate set of species is selected and the number of major pollution sources, q, is correctly estimated. We also assume that in model (2) each row of matrix E has mean vector 0 and variance-covariance matrix Σ, and A and P are unknown constant matrices. We place physical constraints on A and P. The elements of A and the elements of P are nonnegative, and the row sum of P is 1. That is,

$$\alpha_{ik} \geq 0, \quad p_{kj} \geq 0, \quad \sum_{j=1}^{p} p_{kj} = 1, \tag{3}$$

where $i = 1, \ldots, n, \quad k = 1, \ldots, q, \quad j = 1, \ldots, p$. The constraint $\sum_{j=1}^{p} p_{kj} = 1$ indicates that only the relative amount of each species in a source is of our interest. Our sources have fixed ratios of the chemical species. As long as the relative amounts of species are given, we consider the source identified.

We first need to introduce the definition of the model identification.

Definition 1 Let Y be a matrix of the observable random variables, θ be a matrix of the parameters of interest, and let $F_Y(C;\theta)$ be the distribution function of Y for parameter θ evaluated at $Y = C$. The parameter θ is identified if, for any θ_1 and θ_2 in the parameter space,

$$F_Y(C;\theta_1) = F_Y(C;\theta_2) \text{ for all } C$$

implies that

$$\theta_1 = \theta_2.$$

If the the parameter θ is identified, we also say that the model is identified.

Under the model (2), the distribution of Y is determined by AP and Σ (in the normal error case). That is, $F_Y(C; A_1P_1,\Sigma_1) = F_Y(C; A_2P_2,\Sigma_2)$ implies that $A_1P_1 = A_2P_2$ and $\Sigma_1 = \Sigma_2$ and vice versa It does not, however, imply that $(P_1, A_1) = (P_2, A_2)$ which are the parameters of our interest. Thus, in our case, the definition 1 can be reduced to the following:

Definition 2. The parameter (P, A) is identified if, for any (P_1, A_1) and (P_2, A_2) in the parameter space,

$$A_1P_1 = A_2P_2$$

implies that

$$P_1 = P_2 \text{ and } A_1 = A_2.$$

We also define near identifiability of the model parameters.

Definition 3. The parameter (P, A) is nearly identified if, for any (P_1, A_1) and (P_2, A_2) in the parameter space,

$$A_1P_1 = A_2P_2$$

implies that

$$P_1 \approx P_2 \text{ and } A_1 \approx A_2.$$

Proposition 1. Assume $rank(A) = q$ and $rank(P) = q$. Then $A^*P^* = AP$ implies that $A^* = AR$ and $P^* = R^{-1}P$ for a nonsingular matrix $R = (A^tA)^{-1}A^tA^*$.

Since both A and P are unkown, our model (2) suffers from nonidentifiability of model parameters even without the error matrix, i.e., $Y = AP = ARR^{-1}P$ for any nonsigular matrix R. Even the reasonable constraints that we put, (3), do not remove this nonidentifiability. This type of nonidentifiability is often referred to as "factor indeterminacy" in the context of factor analysis. Since there are q^2 elements in the matrix R, we need to put q^2 independent conditions on P or A to rule out this indeterminacy. Preassigning zeros in specified positions of P is usually done in the confirmatory factor analysis. But, it requires some prior knowledge about the source profiles to be estimated. If information about the types of all the sources is available (as assumed in Yang 1994) one can get the idea of where to assign 0's in the matrix P and this indeterminacy would be taken out. Since in our case the source profiles are normalized to sum to 1, this puts q independent conditions on P. Thus the number of free parameters in R reduces to $q(q-1)$, and so we need only $q(q-1)$ more independent conditions. One set of such conditions are

C1. there are at least $q-1$ zero elements in each row of P,

C2. the rank of $P^{(k)}$ is $q-1$, where $P^{(k)}$ is the matrix composed of the columns containing the assigned 0's in the kth row with those assigned 0's deleted.

These conditions can be easily found in usual multivariate analysis textbook (see, e.g. Anderson 1984). Note that C1 and C2 are automatically satisfied if we have tracer element for each source.

A similar set of conditions can also be applied to the source contribution matrix A.

D1. There are at least $q-1$ zero elements in each column of A.

D2. The rank of $A^{(k)}$ is $q-1$, where $A^{(k)}$ is the matrix composed of the rows containing the assigned 0's in the ith column with those assigned 0's deleted.

These conditions are closely related to Henry's assumption that the data contains some points such that each source is missing (Henry 1997). He argued that if there are at least $(q-1)$ edge points (points that have one source missing) for each source and the edge points do not have any multicollinearities of dimension less than $q-1$ then the solution to the general mixture problem is unique. In no error case these conditions can be converted to model identifiability conditions. The condition D2 implies that no two sources have the same set of *q-1* edge points and the edge points (at least *q-1* of them) are linearly independent.

To help solve factor indeterminacy problem, here, we also present two new sets of assumptions for identifiability or near identifiability of *A* and *P* by modifying Henry's edge point assumption. We need only one set of assumptions to hold for *A* and *P* to be identifiable.

The first set of our basic assumptions are:

A1. Each source is missing on some days and we know when a source is missing.

A2. The average contribution of j^{th} source when k^{th} $(k \neq j)$ source is missing is equal to the average contribution of j^{th} source for all days.

A3. The source contribution matrix *A* is of full column rank and the source composition matrix *P* is of full row rank, *i.e.*, $rank(A) = q$ and $rank(P) = q$.

The second set of our basic assumptions are:

B1. Each source is missing on some days and we know when a source is missing.

B2. The difference between the average contribution of j^{th} source when k^{th} $(k \neq j)$ source is missing and the average contribution of j^{th} source for all days is small.

B3. The source contribution matrix *A* is of full column rank and the source composition matrix *P* is of full row rank, *i.e.*, $rank(A) = q$ and $rank(P) = q$.

Remark 1. The assumption A1 (or B1) is equivalent to preassigning zeros in a specified position of the source contribution matrix. This usually requires less prior information than the conditions based on the source composition matrix. Although Henry

(1997) assumed the existence of at least q-1 edge points for each of q sources, here, A1 (or B1) allows having less than q-1 edge points as long as the other assumptions are satisfied.

Defining $\bar{\alpha}_j^{(k)}$ as the average contribution of j^{th} source when the k^{th} source is missing and $\bar{\alpha}_j$ as the average contribution of j^{th} source for all days, we can reexpress the above assumptions as follows. Of course we require that $j \neq k$.

For A1-A3,

A1. $\alpha_{ik} = 0$ when $i \in I_k$, $k = 1,\ldots, q$.

Here I_k is defined to be a subset of $\{1, 2, \ldots, n\}$ for which the k^{th} source is missing.

A2. $\bar{\alpha}_j^{(k)} = \bar{\alpha}_j$, $j = 1,\ldots, q$, $j \neq k$.

A3. $rank(A) = q$, $rank(P) = q$.

For B1-B3,

B1. $\alpha_{ik} = 0$ when $i \in I_k$, $k = 1,\ldots, q$.

Here I_k is defined to be a subset of $\{1, 2, \ldots, n\}$ for which the k^{th} source is missing.

B2. $|\bar{\alpha}_j - \bar{\alpha}_j^{(k)}| \leq \varepsilon$, $j = 1,\ldots, q$.

B3. $rank(A) = q$, $rank(P) = q$.

The following results show that under each set of assumptions, A1-A3 and B1-B3, nonidentifiability of the model parameters can be removed. That is, $A^* = A$ and $P^* = P$ (or $A^* \approx A$ and $P^* \approx P$). The proofs are found in Appendix B.

Result 1. Let Assumptions A1-A3 hold. Then

$$R = \mathbf{I}$$

where **I** is the $q \times q$ Identity matrix and R is any nonsingular matrix satisfying $A^* = AR$ and $P^* = R^{-1}P$.

Result 2. Let Assumptions B1-B3 hold. Define $B = \frac{\varepsilon}{\bar{\alpha}_k}\left(1 + \frac{pq^2(q-1)}{\lambda_q}\right)$ where λ_q is the smallest eigenvalues of P^*P^{*t}. If B is small enough then the diagonal elements of R are close to 1, and the off-diagonal elements of R are close to 0.

Remark 2. We emphasize that all the conditions cited in this paper are sufficient conditions but not necessary conditions for model identifiability. On closer inspection if we knew that entries of the A matrix satisfy $a_{ij} = a_{i'j}$ we could difference the corresponding observations and create zeros. Thus by doing typical time series differencing we may create data that satisfies the identifiability conditions when the original data does not.

3. Estimation of source profiles and contributions

The number of parameters in model (1) increases to infinity as the sample size increases. Kiefer and Wolfowitz (1956) addressed the issue of estimating the structural parameter consistently when there are infinitely many incidental parameters. They assumed that the incidental parameters were independently distributed chance variables with a common unknown distribution function. This assumption was made in Bandeen-Roche and Ruppert (1991) and Yang (1994). We do not make such assumption for our incidental parameters, the rows of A. Instead of treating them as chance variables, we just leave them as unknown parameters, which is the way that many scientists and in this application most environmental engineers want to view them. To achieve a consistent sequence of estimators we need to further restrict a parameter space for A, as well as utilizing the identifiability conditions in the fitting procedure. Two models, Quasi Random Functional Model (which is a generalization of the model used in Kiefer and Wolfowitz (1956)) and

Replicated Functional Model, are considered, and a set of algorithms, VERTEX, to find the least squares solution is introduced. Each of these algorithms can be selectively implemented according to the sets of identifiability conditions used. The resulting estimators are shown to be consistent, and also the uncertainties associated with them are provided.

3.1. Quasi random functional model

To overcome the difficulty of having infinitely many parameters we first restrict the parameter space of A by assuming that the first and the second sample moments of the rows of A converge to some fixed vector and matrix, respectively. This model is referred to as "quasi-random functional model" in Gleser (1983). We assume

$$y_i = \alpha_{0i} P_0 + \varepsilon_i,$$

where the ε_i are independent identically distributed p-dimensional random row vectors with zero mean vector, positive definite covariance matrix Σ_0, and $\{\alpha_{0i}\}$ is a fixed sequence satisfying

$$\bar{\alpha}_0 = n^{-1}\sum_{i=1}^{n}\alpha_{0i} \xrightarrow{n\to\infty} \alpha_0$$

and

$$\overline{\mathbf{K}}_0 = n^{-1}\sum_{i=1}^{n}(\alpha_{0i}-\bar{\alpha}_0)^t(\alpha_{0i}-\bar{\alpha}_0) \xrightarrow{n\to\infty} \mathbf{K}_0.$$

where α_0 is a q-dimensional vector and $\mathbf{K_0}$ is a $q\times q$ positive definite matrix.

We choose the estimators of A and P so as to minimize the sum of least squares,

$$Q_n(P, A) = n^{-1} tr[(Y - AP)^t(Y-AP)] = n^{-1}\sum_{i=1}^{n}\left\| y_i - \sum_{k=1}^{q}\alpha_{ik}P_k \right\|^2 \qquad (4)$$

subject to the constraints, (3), and identifiability conditions.

VERTEX 1

Since both of A and P are unknown parameters, our estimation procedure, VERTEX 1, consists of two steps:

1) Given P, A can be estimated by $\tilde{A} = YP^t(PP^t)^{-1}$.

2) Find $\hat{P}$ which minimizes

$$Q_n(P,\tilde{A}) = n^{-1}tr\left[(Y-\tilde{A}P)^t(Y-\tilde{A}P)\right]$$

$$= n^{-1}tr\left[(Y-YP^t(PP^t)^{-1}P)^t(Y-YP^t(PP^t)^{-1}P)\right]$$
$$= n^{-1}tr\left[(I_p-P^t(PP^t)^{-1}P)^tY^tY(I_p-P^t(PP^t)^{-1}P)\right]$$
$$= n^{-1}tr\left[Y^tY(I_p-P^t(PP^t)^{-1}P)\right]$$
$$= tr\left[(S+\bar{y}^t\bar{y})(I_p-P^t(PP^t)^{-1}P)\right]$$

where the vector $\bar{y} = n^{-1}\underline{1}^tY = n^{-1}\sum_{i=1}^{n} y_i$ and the matrix

$$S = n^{-1}(Y-\underline{1}\bar{y})^t(Y-\underline{1}\bar{y}) = n^{-1}\sum_{i=1}^{n}(y_i-\bar{y})^t(y_i-\bar{y})$$

where $\underline{1}$ is an n-dimensional column vector consisting of 1's

over the feasible set Ω for which the constraints on P and the identifiability conditions C1 and C2 are satisfied.

The consistency and the asymptotic normality of $\hat{P}$ can be proven by adapting the properties of least squares estimators in Fuller (1987). We state the asymptotic results for $\hat{P}$ in Theorem 1 and Theorem 2. The proofs of all our theorems are found in the appendix B.

Theorem 1 (Consistency of $\hat{P}$). Let A_0, P_0 and Σ_0 be the true values of A, P, and Σ respectively. Assume $\bar{\alpha}_0 = n^{-1}\sum_{i=1}^{n}\alpha_{0i} \xrightarrow{n\to\infty} \alpha_0$ where α_{0i} is the i^{th} row of A_0 and α_0 is a q-dimensional row vector and $\overline{\mathbf{K}}_0 = n^{-1}\sum_{i=1}^{n}(\alpha_{0i}-\bar{\alpha}_0)^t(\alpha_{0i}-\bar{\alpha}_0) \xrightarrow{n\to\infty} \mathbf{K}_0$ where $\mathbf{K}_0$ is a full rank matrix. Let the identifiability conditions C1-C2 hold. Let $\boldsymbol{W}$ be the subset of $p+p(p+1)/2$-dimensional Euclidean space and $\omega_0 = (\alpha_0 P_0, vech(\Sigma_0 + P_0^t \mathbf{K}_0 P_0)^t)^t$. Assume ω_0 is in the interior of $\boldsymbol{W}$ and $\Sigma_0 = \sigma^2 \mathbf{I_p}$. Then , when $n \rightarrow \infty$,

$$vec(\hat{P}) \xrightarrow{p} vec(P_0).$$

Remark 3. Note that $\dfrac{g(w_n;\hat{\theta})}{p-q}$ is a consistent estimator of σ^2, and that $\tilde{A}(\hat{P}) = Y\hat{P}^t(\hat{P}\hat{P}^t)^{-1}$ converges to usual least squares estimate of A when P is known.

Remark 4. The sample mean of the estimated daily source contributions, $n^{-1}\sum_{i=1}^{n}\hat{\alpha}_i$, where $\hat{\alpha}_i$ is the i^{th} row of $\tilde{A}(\hat{P}) = Y\hat{P}^t(\hat{P}\hat{P}^t)^{-1}$, can be shown to be consistent; $n^{-1}\sum_{i=1}^{n}\hat{\alpha}_i \xrightarrow{p} \alpha_0$.

Theorem 2 (Asymptotic Normality of $\hat{P}$). Let the assumptions for Theorem 1 hold. Let r be the number of free parameters in P and θ be the r-dimensional vector consisting of those free parameters. Assume that the true parameter value $\theta_0 \in int(\Theta)$, where Θ, the parameter space for θ, is a convex compact subset of r-dimensional Euclidean space. Assume the error covariance matrix $\Sigma_0 = \sigma^2\mathbf{I_p}$ and the errors have finite fourth moments. Let $\hat{\theta}$ be the value of θ that minimizes

$$g(\bar{y}, vechS; \theta) = tr\left[(S + \bar{y}^t\bar{y})(I_p - P^t(PP^t)^{-1}P)\right].$$

Then

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{H^{-1}BGB^tH^{-1}}),$$

where

G is the limiting covariance matrix of $\left[\bar{y},(vechS)^t\right]^t$, $\mathbf{B} = (\mathbf{B_2}, \mathbf{B_1})$,

$\mathbf{B_1}$= $\left[L^t\left\{\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\otimes(P_0P_0^t)^{-1}P_0\right\} + M^t\left\{(P_0P_0^t)^{-1}P_0\otimes\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\right\}\right]\Phi_p$,

$\mathbf{B_2} = -2L^t\left[\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\otimes\alpha_0^t\right]$,

$\mathbf{H} = 2L^t\left((I_p - P_0^t(P_0P_0^t)^{-1}P_0)\otimes\mathbf{K_0}\right)L + 2L^t\left[\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\otimes\alpha_0^t\alpha_0\right]L$,

$L = \dfrac{\partial vecP_0}{\partial\theta^t}$ is the matrix of partial derivatives of P with respect to θ evaluated at $\theta = \theta_0$,

$M = \dfrac{\partial vecP_0^t}{\partial\theta^t}$ is the matrix of partial derivatives of P^t with respect to θ evaluated at $\theta = \theta_0$,

and Φ_p is the $p^2\times\frac{1}{2}p(p+1)$ matrix such that $vecA = \Phi_p vechA$ for any $p\times p$ symmetric matrix A.

Remark 5. Note $\sigma_{ij} = 0$ for $i\neq j$ under our assumption that $\Sigma_0 = \sigma^2\mathbf{I_p}$.

Remark 6. If θ_0 is on the boundary of Θ, i.e., some of the elements of θ_0 is zero, then the limiting distribution of $\hat{\theta}$ would not be a normal distribution. It would be a mixture of point mass at zero and a normal distribution.

3.2. Replicated functional model

Consisder the model

$$Y = UAP + E, \tag{5}$$

where $U = \begin{bmatrix} \underline{1}_{m_1} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1}_{m_2} & \cdots & \underline{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{1}_{m_n} \end{bmatrix}$, $\sum_{i=1}^{n} m_i = N$, $\underline{1}_{m_i}$ is an m_i-dimensional column vector consisting of 1's, A is the $n\times q$ source contribution matrix, P is the $q\times p$ source composition matrix, and E is a $N\times p$ error matrix. The i^{th} observation in the j^{th} replication, $\underline{y}_{ij}$ with double script notation, is represented by

$$\underline{y}_{ij} = \alpha_i P + \varepsilon_{ij}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,m_i$$

where α_i is the q-dimensional row vector corresponding to the i^{th} source contribution, and $\underline{\varepsilon}_{ij}$ is a random error corresponding to the i^{th} observation in the j^{th} replication. We assume $\underline{\varepsilon}_{ij}$'s are independent and identically distributed with mean vector 0 and variance-covariance matrix Σ, and α_i's and P are unknown paramters. This model is recognizable as an "replicated functional model" (see, e.g., Gleser 1983). We have

$$E(Y) = UAP$$

and

$$Var(Y) = \mathbf{I}_N \otimes \Sigma.$$

Note that U is a known $N\times n$ matrix. Under the identifiability conditions, A1-A3 or B1-B3, described in section 2, this model is identified (or nearly identified). That is, $UA_1P_1 = UA_2P_2$ implies that $A_1 = A_2$ and $P_1 = P_2$ (or $A_1 \approx A_2$ and $P_1 \approx P_2$).

Let $\sum_{i=1}^{n} m_i = N$. Here, N is the total number of observations in the data.

The least squares estimators of A and P are obtained by minimizing the sum of squares,

$$Q_N(P, A) = N^{-1} tr\left[(Y - UAP)^t (Y - UAP)\right] \qquad (6)$$

$$= N^{-1}tr\left[(Y - U\overline{Y} + U\overline{Y} - UAP)^t (Y - U\overline{Y} + U\overline{Y} - UAP)\right]$$
$$= N^{-1}tr\left[(Y - U\overline{Y})^t (Y - U\overline{Y})\right] + N^{-1}tr\left[(U\overline{Y} - UAP)^t (U\overline{Y} - UAP)\right]$$
$$= N^{-1}tr\left[(Y - U\overline{Y})^t (Y - U\overline{Y})\right] + N^{-1}tr\left[(\overline{Y} - AP)^t U^t U(\overline{Y} - AP)\right]$$
$$= N^{-1}tr\left[(Y - U\overline{Y})^t (Y - U\overline{Y})\right] + N^{-1}tr\left[M(\overline{Y} - AP)(\overline{Y} - AP)^t\right]$$

over the feasible set Θ, where

$$\Theta = \left\{ (P, A) \middle| \alpha_{ik} \geq 0, \quad p_{kj} \geq 0, \quad \sum_{j=1}^{p} p_{kj} = 1, \quad i = 1, \ldots, n, \quad k = 1, \ldots, q, \quad j = 1, \ldots, p, \quad \mathrm{I}_{\mathrm{A}} \right\}$$

where I_{A} (= A1-A3 or B1-B3) is a set of the identifiability conditions defined in section 2. Since $N^{-1}tr\left[(Y - U\overline{Y})^t (Y - U\overline{Y})\right]$ does not depend on A or P, minimizing $Q_N(P, A)$ is equavelent to minimizing $Q_N^*(P, A) = N^{-1}tr\left[M(\overline{Y} - AP)(\overline{Y} - AP)^t\right]$ *w.r.t.* A and P.

A fitting algorithm for estimating A and P under this model is given below:

VERTEX 2

1) Given A, P can be estimated by

$$\tilde{P} = \left[(UA)^t (UA)\right]^{-1} (UA)^t U\overline{Y} = \left(A^t MA\right)^{-1} A^t M\overline{Y}$$

where $M = \begin{bmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & \ddots & \\ 0 & & & m_n \end{bmatrix}$ and $\overline{Y} = \begin{bmatrix} m_1^{-1}\sum_{j=1}^{m_1} \underline{y}_{1j} \\ m_2^{-1}\sum_{j=1}^{m_2} \underline{y}_{2j} \\ \vdots \\ m_n^{-1}\sum_{j=1}^{m_n} \underline{y}_{nj} \end{bmatrix}$.

2) Find $\hat{A}$ which minimizes

$$Q_N^*(P, A) = N^{-1}tr\left[M(\overline{Y} - A\tilde{P})(\overline{Y} - A\tilde{P})^t\right]$$

$$= N^{-1}tr\left[M\left(\bar{Y} - A\left(A^t MA\right)^{-1} A^t M\bar{Y}\right)\left(\bar{Y} - A\left(A^t MA\right)^{-1} A^t M\bar{Y}\right)^t\right]$$

$$= N^{-1}tr\left[M\left(I_n - A\left(A^t MA\right)^{-1} A^t M\right)\bar{Y}\bar{Y}^t\left(I_n - A\left(A^t MA\right)^{-1} A^t M\right)^t\right]$$

$$= N^{-1}tr\left[M\left(I_n - A\left(A^t MA\right)^{-1} A^t M\right)\bar{Y}\bar{Y}^t\left(I_n - A\left(A^t MA\right)^{-1} A^t M\right)^t\right]$$

$$= N^{-1}tr\left[\bar{Y}^t M\left(I_n - A\left(A^t MA\right)^{-1} A^t M\right)\bar{Y}\right]$$

over the feasible set Ω_A where

$$\Omega_A = \left\{A \middle| \alpha_{ik} \geq 0, \qquad i = 1,\ldots,n, \qquad k = 1,\ldots,q, \qquad \mathrm{I}_A\right\}$$

where I_A (= A1-A3 or B1-B3) is a set of the identifiability conditions defined in section 2.

Definition 4. Let A_n be a sequence of random matrices and A_0 be a constant matrix. Then $A_n \xrightarrow{p} A_0$ means $P(\|A_n - A_0\| > \varepsilon) \to 0$ as $n \to \infty$ for each ε > 0, where $\|A\| = \sum\sum a_{ij}^2$.

The asymptotic results for $\hat{P}$ are stated in Theorem 3 and Theorem 4, which are established based on the asymptotic results for $\hat{A}$. In the following propositions we first state the asymptotics for $\hat{A}$.

Proposition 2 (Consistency of $\hat{A}$). Let A_0, P_0 and Σ_0 be the true values of A, P, and Σ respectively. Let the parameter space for A, Ω_A be a compact subset of $n\times q$-dimensional Euclidean space containing A_0. Assume the identifiability conditions A1-A3 in Section 2 are satisfied. Also assume A_0P_0 is in the interior of a subset of $n\times p$-dimensional Euclidean space. Then, when $m_i \to \infty$, $\lim \frac{m_i}{N} = c_i > 0$, $i = 1, \ldots, n$,

$$\hat{A} \xrightarrow{p} A_0$$

Remark 7. Under the identifiability conditions B1-B3 in Section 2, we get the approximate consistency of $\hat{A}$, i.e., $P(\|A_n - A_0\| > \varepsilon) \to 0$ as $n \to \infty$ for practically small (not arbitrary) ε > 0.

Let $\hat{P} = \left(\hat{A}^t M \hat{A}\right)^{-1} \hat{A}^t M \bar{Y}$. As an immediate consequence of Proposition 2 and the fact that $\bar{Y} \xrightarrow{p} A_0 P_0$, and by using the continuous mapping theorem, we obtain the consistency of $\hat{P}$ as follows.

Theorem 3 (Consistency of $\hat{P}$). Under the definitions and the assumptions of Proposition 2, when $m_i \to \infty$, $\lim \frac{m_i}{N} = c_i > 0$, $i = 1, \ldots, n$,

$$\hat{P} \xrightarrow{p} P_0.$$

We now establish the asymptotic normality of $\hat{A}$.

Proposition 3 (Asymptotic Normality of $\hat{A}$). Let the assumptions for Proposition 2 hold. Let r be the number of free parameters in A and θ be the r-dimensional vector consisting of those free parameters. Assume that the true parameter value $\theta_0 \in int(\Theta)$, where Θ, the parameter space for θ, is a convex compact subset of r-dimensional Euclidean space. Assume the errors have finite fourth moments. Let $\hat{\theta}$ be the value of θ that minimizes

$$g(\bar{Y}, N^{-1}M; A) = tr\left[\bar{Y}^t N^{-1} M(\mathbf{I_n} - A(A^t N^{-1} M A)^{-1} A^t N^{-1} M)\bar{Y}\right].$$

Then

$$\sqrt{m^*}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H^{-1}B(}C^* \otimes \Sigma_p\mathbf{)B^tH^{-1}}\right),$$

where $m^* = \min_{1 \le i \le n}\{m_i\}$, $\mathbf{\Sigma_p}$ is error covariance matrix,

$\mathbf{B} = 2Q^t\left[C\left(I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right)\otimes P_0\right]$, $\mathbf{H} = Q^t\left[C\left(I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right)\otimes P_0P_0^t\right]Q$,

$C = \begin{bmatrix} c_1 & & & 0 \\ & c_2 & & \\ & & \ddots & \\ 0 & & & c_n \end{bmatrix}$ where $c_i = \lim_{m_i\to\infty}\frac{m_i}{N}$, $C^* = \begin{bmatrix} \frac{c^*}{c_1} & & & 0 \\ & \frac{c^*}{c_2} & & \\ & & \ddots & \\ 0 & & & \frac{c^*}{c_n} \end{bmatrix}$ where $c^* = \lim_{m^*\to\infty}\frac{m^*}{N}$, and

$Q = \frac{\partial vecA_0^t}{\partial\theta^t}$ is the matrix of partial derivatives of A with respect to θ evaluated at $\theta = \theta_0$.

Remark 8. The same comment as Remark 5 can be made here if θ_0 is on the boundary of Θ.

Remark 9. The asymptotic normality of $\hat{P} = \left(\hat{A}^tM\hat{A}\right)^{-1}\hat{A}^tM\overline{Y}$ can also be established by standard arguments for nonlinear least squares estimators. However, the resulting distrubution has an extremely complicated covariance matrix. Thus, we employ Bootstrap method to obtain the approximate covariance matrix. The estimates are aymptotically unbiased by theorem 3.

Under the replicated functional model (5), resampling can be done in two ways, Case resampling and Model based resampling. We adapt the algorithms in Davison and Hinkley (1997) for our model:

For $h = 1,\ldots,H$ (H: bootstrap size),

A) Case resampling

1. For each source contribution α_i ($i = 1,\ldots,n$), choose $y_{i1}^*,\ldots,y_{im_i}^*$ by randomly sampling with replacement from $y_{i1},\ldots,y_{im_i}$.

2. Combining $y_{i1}^*,\ldots,y_{im_i}^*$, $i = 1,\ldots,n$, leads to a bootstrap sample, Y^*.

B) Model based resampling

1. Find $\hat{A}$ and $\hat{P}$ based on the original data.

2. Compute residuals by $R = (I - DH)^{-\frac{1}{2}}\left(Y - U\hat{A}\hat{P}\right)$ where DH is a diagonal matrix consisting of the diagonal elements of $H = \left(U\hat{A}\right)\left(\left(U\hat{A}\right)^t\left(U\hat{A}\right)\right)^{-1}\left(U\hat{A}\right)^t$.

3. Randomly sample ε_j^* from $r_1 - \bar{r}, \ldots, r_N - \bar{r}$ where r_i $(i = 1, \ldots, N)$ is the i^{th} row of the matrix R and $\bar{r}$ is the residual mean vector, i.e., $\bar{r} = N^{-1}\left(\sum_{i=1}^{N} r_{i1}, \ldots, \sum_{i=1}^{N} r_{ip}\right)$.

4. Set $Y^* = U\hat{A}\hat{P} + E^*$ where $E^* = \left(\varepsilon_1^*, \ldots, \varepsilon_N^*\right)^t$.

Bootstrap estimators $\hat{P}^*$ are obtained for H bootstrap samples, and the sample covariance matrix of those $\hat{P}^*$ is used as approximate covariance matrix of $\hat{P}$.

Remark 10. It has been observed from the simulation study that case resampling leads to more stable bootstrap estimators than model based resampling. Accuracy of bootstrap estimators $\hat{P}^*$ from model based resampling depends heavily on how good the estimate $\hat{A}$ from the original data is since $\hat{A}$ is treated as A_0 in model based resampling.

4. Simulations

In this section we consider simulated examples to illustrate the proposed methods. For VERTEX 1, the data is generated by the model (2) where $n = 200$, $p = 9$, and $q = 3$. The errors are independently generated from the centered lognormal distribution so that they have mean 0 and variance-covariance matrix $\sigma^2\mathbf{I_p}$. The source composition matrix P (actually, P^t) is given in Table 1. The value of σ was chosen so that the proportions of the error standard deviations to the model standard deviations are mostly between 10~30%. The elements of source contribution matrix A are taken from the uniform random numbers (uniform(0,2)). The true mean source contributions are given in Table 2. The resulting data matrix Y consists of nonnegative numbers.

Table 1. True source composition profiles (P_0^t)

Species	*Source 1*	*Source 2*	*Source 3*
1	0.2242	0.2264	0
2	0	0.1678	0.1369
3	0.1932	0.1981	0.0438
4	0	0.1316	0.0385
5	0.0708	0	0.2067
6	0.0337	0.0387	0.0089
7	0.2266	0	0.2237
8	0.1452	0.1996	0.3414
9	0.1063	0.0378	0

Table 2. True mean source contributions

	Source 1	Source 2	Source 3
$\overline{\alpha}_0$	0.9919	1.0037	1.0158

To apply VERTEX 1, we assume that it is known beforehand that species 2 and 4 are missing in source 1, species 5 and 7 are missing in source 2, and species 1 and 9 are missing in source 3, i.e., some of the elements of the source composition matrix are prespecified. Table 3 shows the resulting estimates of the source compositions from VERTEX 1, and Table 4 shows the asymptotic standard deviations of the estimators.

Table 3. Estimated source composition profiles ($\hat{P}_{VERTEX1}^t$)

Species	*Source 1*	*Source 2*	*Source 3*
1	0.2224	0.2293	0
2	0	0.1637	0.1446
3	0.1953	0.2011	0.0392
4	0	0.1280	0.0456
5	0.0703	0	0.2045
6	0.0332	0.0368	0.0118
7	0.2304	0	0.2136
8	0.1449	0.2020	0.3406
9	0.1035	0.0392	0

Table 4. Asymptotic standard deviations of $\hat{P}^t_{VERTEX1}$

Species	*Source 1*	*Source 2*	*Source 3*
1	0.0257	0.0013	0
2	0	0.0012	0.0072
3	0.0134	0.0007	0.0015
4	0	0.0016	0.0047
5	0.0183	0	0.0024
6	0.0013	0.0006	0.0010
7	0.0162	0	0.0153
8	0.0144	0.0009	0.0012
9	0.0071	0.0062	0

The VERTEX 1 provides the estimates for all of *n* source contributions, but only the sample mean of those estimated source contributions is reported here. The estimated mean source contributions is given in Table 5.

Table 5. Estimated mean source contributions

	Source 1	Source 2	Source 3
$\bar{\hat{\alpha}}_{VERTEX1}$	1.0129	0.9774	1.0211

Note that the numbers in Table 5 are the estimates of the absolute source contributions in this case since the true source profiles are generated so that the sum of species in each profile is 1. In real situations, the sum could be any positive number, say c_k, *k=1, 2, 3*. In that case, $\left(\hat{\alpha}_{i1},\hat{\alpha}_{i2},\hat{\alpha}_{i3}\right)$ will be the estimate of $\left(c_1\alpha_{0i1},c_2\alpha_{0i2},c_3\alpha_{0i3}\right)$, $i=1,\ldots,n$, and $\left(\bar{\hat{\alpha}}_1,\bar{\hat{\alpha}}_2,\bar{\hat{\alpha}}_3\right)$ will be the estimate of $\left(c_1\bar{\alpha}_{01},c_2\bar{\alpha}_{02},c_3\bar{\alpha}_{03}\right)$.

Fig. 1 shows the princial component plot of the data, the estimated source profiles, and the true source profiles. It can be seen that the source profiles obtained from VERTEX 1 give very good approximation to the true source profiles.

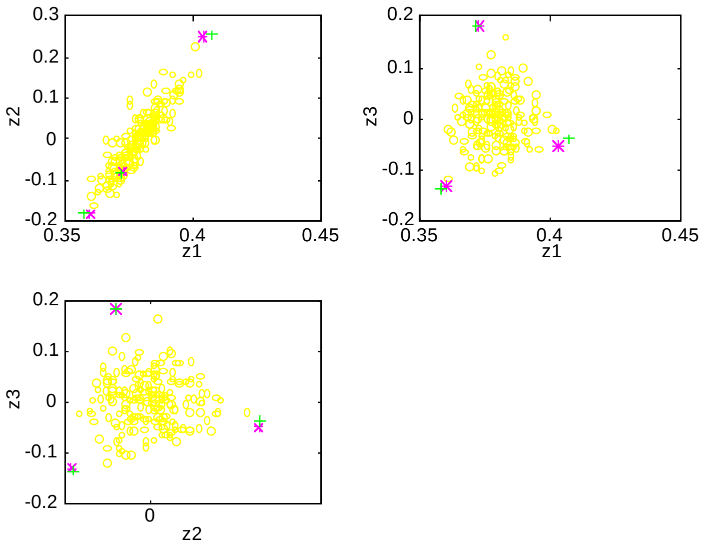

Fig. 1. Principal component plots of the data (o), the true sources (+), and the fitted sources by VERTEX 1 (*)

To illustrate VERTEX 2, the data is generated based on the model (5) where $N = 720$, $n = 24$ (assuming the source contributions are repeated every 24 hour), $p = 7$, and $q = 3$. Although the number of replications m_i need not be equal, for the sake of brevity, the same number of replications are used for the source contributions. Thus, $m_1 = m_2 = \ldots = m_{24} = 30$.

The source profiles normalized to sum to 1 are given in Table 6. The source contribution matrix A is generated to satisfy the conditions A1-A3 in section 2. It is assumed that source 1 is missing on 8th hour, source 2 is missing on 7th hour, and source 3 is missing on 6th hour, and when each source is missing, the average source contributions of the other sources stay the same. The errors associated with N observations are independently generated from the centered lognormal distribution so that the proportions of the error standard deviations to the model standard deviations (which can be defined as the squarerooted diagonal elements of $P_0{}^tK_0P_0$ in this case) are about 10~30%.

Table 6. True source composition profiles (P_0^t)

Species	*Source 1*	*Source 2*	*Source 3*
1	0.0852	0.2252	0.1627
2	0.2124	0.0417	0.0638
3	0.1155	0.1597	0.1902
4	0.0305	0.0185	0.0657
5	0.1981	0.1312	0.2486
6	0.1681	0.2132	0.1852
7	0.1902	0.2104	0.0838

Note that, in VERTEX 2, the constrained minimization is done with A. Once we get the estimated source contributions, $\hat{A}_{VERTEX2}$, the source compositions are estimated by ordinary least squares, i.e., $\hat{P}_{VERTEX2} = \left(\hat{A}^t_{VERTEX2}\hat{A}_{VERTEX2}\right)^{-1}\hat{A}_{VERTEX2}\overline{Y}$. Table 7 shows the estimated source profiles normalized to sum to 1. Although the nonnegativity constraints for the compositions were not used, the estimates of source profiles are all nonnegative. It is observed from the simulation that only when the true source composition matrix contains zeros, the corresponding estimates (of zeros) are negative. In that case, it would be a natural choice to replace the negative estimates with 0 and renormaliz each source profile. Fig. 2 shows the principal component plot of the data, the true source profiles, and $\hat{P}^t_{VERTEX2}$. It can be seen that $\hat{P}^t_{VERTEX2}$ gives a very good approximation to the true source composition matrix.

Table 7. Estimated source composition profiles ($\hat{P}^t_{VERTEX2}$)

Species	*Source 1*	*Source 2*	*Source 3*
1	0.0868	0.2265	0.1631
2	0.2107	0.0373	0.0617
3	0.1142	0.1627	0.1925
4	0.0315	0.0177	0.0688
5	0.1972	0.1339	0.2501
6	0.1680	0.2150	0.1852
7	0.1915	0.2070	0.0786

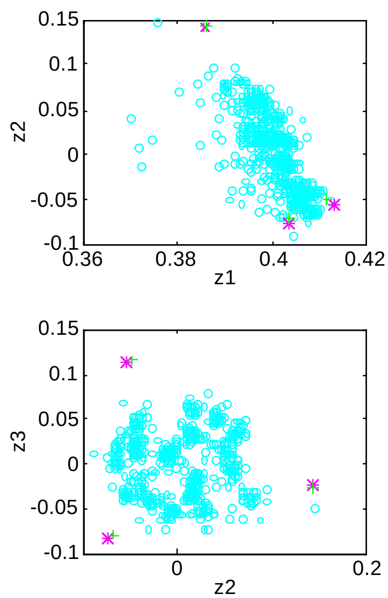

Fig. 2. Principal component plots of the data (o), the true sources (+), and the fitted sources by VERTEX 2 (*)

Table 8 and Table 9 show the bootstrap standard deviations of $\hat{P}^t_{VERTEX2}$ based on 200 bootstrap samples from Case resampling and Model based resampling, respectively.

Table 8. Estimated standard deviations of $\hat{P}^t_{VERTEX2}$ from Case resampling

Species	*Source 1*	*Source 2*	*Source 3*
1	0.0022	0.0037	0.0019
2	0.0033	0.0053	0.0031
3	0.0019	0.0023	0.0015
4	0.0016	0.0014	0.0011
5	0.0015	0.0020	0.0015
6	0.0014	0.0017	0.0012
7	0.0019	0.0025	0.0017

Table 9. Estimated standard deviations of $\hat{P}^{t}_{VERTEX2}$ from Model based resampling

Species	*Source 1*	*Source 2*	*Source 3*
1	0.0022	0.0020	0.0019
2	0.0025	0.0022	0.0021
3	0.0014	0.0015	0.0013
4	0.0010	0.0012	0.0010
5	0.0019	0.0021	0.0016
6	0.0013	0.0012	0.0012
7	0.0021	0.0023	0.0018

5. Applications

In this section we present some applications of our methods to real air pollution datasets.

5.1 Example 1 (Air pollution composition)

As part of a large air quality study, hourly concentrations of hydrocarbon gases were determined by automated gas chromatography at two sites in Houston Texas from June to November 1993. The original data consists of 2,541 hourly observations (after initial screening of the outliers) on 54 volatile organic compounds (VOC) and total nonmethane organic carbon (TNMOC). The wind data consisting of hourly average wind direction, standard deviation of the wind direction, and resultant wind direction and speed were also provided. These data were used in Henry, Spiegleman, Collins, and Park (1997). The 12 important species were selected by examination of the scatterplots, the correlation matrix, and environmental engineer's judgement for further analysis in Henry *et al.* (1997). We use the same set of species. According to Henry *et al.* (1997) Industrial 2 and Industrial 3 show especially high emissions for the wind direction 180^o to 190^o. To the dataset consisting of 183 observations with the 180-190 wind direction, VERTEX 1 is applied again with $\hat{q} = 3$. Here, information needed for prespecification of zero element is obtained from the SAFER result in Herny *et al.* (1997). Table 10 shows the estimated source profiles. The fitted sources are very close in their compositions to Industrial 2, Industrial 3, and Industrial 5 from SAFER. The R^2 (r^2) values are given in Table 11. Fig. 3 shows the principal component plot of the data and the fitted sources by VERTEX 1.

Table 10. Estimated source composition profiles ($\hat{P}_{VERTEX1}$)

Species	*Source 1*	*Source 2*	*Source 3*
Ethane	0.0173	0.0106	0.5152
N_prop	0.0000	0.0000	0.3254
Acetyl	0.0074	0.0000	0.0362
T2pene	0.0422	0.0000	0.0000
Bu23dm	0.1093	0.0033	0.0115
Pena2m	0.6316	0.0000	0.0000
Pen23m	0.0000	0.0297	0.0378
Hexa3m	0.1127	0.0385	0.0496
Etbz	0.0598	0.2330	0.0003
Mp_xyl	0.0000	0.6421	0.0000
Bz135m	0.0000	0.0234	0.0000
Bz124m	0.0195	0.0194	0.0242

Table 11. R^2 values between $\hat{P}_{VERTEX1}$ and $\hat{P}_{SAFER}$

	Source 1	*Source 2*	*Source 3*
Industrial 2	0.9880	0.0271	0.0435
Industrial 3	0.0268	0.8925	0.0179
Industrial 5	0.0425	0.0097	0.9269

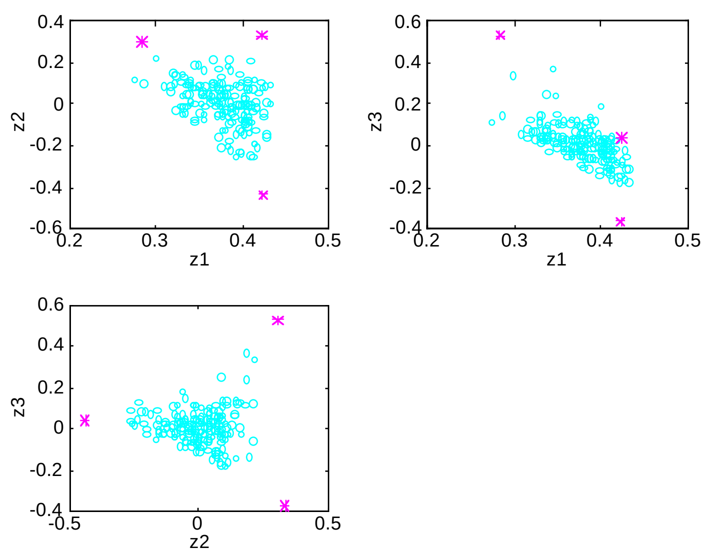

Fig. 3. Principal component plots of the data (o) and the fitted sources by VERTEX 2 (*)

5.2 Example 2 (Air pollution spatial)

As the second example we consider measurements on sulfur dioxide gas collected from 11 receptor sites in the nearby Grand Canyon National Park (Henry, 1992). The resulting data set consists of 53 observations on 11 variables (here receptor sites). The number of sources, q, is estimated to be 3 by the NUMFACT algorithm (Henry, Park, and Spiegelman, 1997). Physically, there are three known source regions of sulfur dioxide gases in the region. These sources are believed to correspond to pollution sources in southern California, copper smelters in southern Arizona and northern Mexico, and electric power plants. Not all the source profiles have the required number of zeros to apply VERTEX 1, so we apply VERTEX 2 with identifiability conditions A1-A3. Source 1 is assumed to be missing on Day 1, Source 2 is missing on Day 12, and Source 3 is assumed to be missing on Day 44. These edge points (obs. 1, 12, and 44) are taken by the principal component plots of the data and the SAFER fit results (Henry and Kim, 1990). The estimated source profiles normalized to sum to 1 and standard errors based on 8 bootstrap samples from Model based resampling appear in Table 12. Fig. 4 shows the principal component plot of the data and the fitted sources. From the plot it can be seen that the estimated source profiles give a reasonable fit to the data.

Table 12. Estimated source composition profiles ($\hat{P}^{t}_{VERTEX2}$)$^{+}$

Variables	*Source 1*	*Source 2*	*Source 3*
1	0.1295 (0.0030)	0.0707 (0.0044)	0.0686 (0.0046)
2	0.0521 (0.0024)	0.1138 (0.0051)	0.0299 (0.0029)
3	0.1235 (0.0032)	0.0684 (0.0042)	0.0708 (0.0041)
4	0.0886 (0.0064)	0.0476 (0.0086)	0.1275 (0.0078)
5	0.0635 (0.0060)	0.1100 (0.0119)	0.2178 (0.0098)
6	0.1272 (0.0042)	0.0686 (0.0042)	0.0686 (0.0059)
7	0.0586 (0.0037)	0.1064 (0.0056)	0.0427 (0.0046)
8	0.0278 (0.0044)	0.1270 (0.0043)	0.0317 (0.0029)
9	0.0857 (0.0035)	0.1010 (0.0082)	0.2167 (0.0077)
10	0.1346 (0.0065)	0.0994 (0.0071)	0.0731 (0.0038)
11	0.1090 (0.0021)	0.0872 (0.0067)	0.0525 (0.0060)

$^{+}$ Standard errors are given in parentheses.

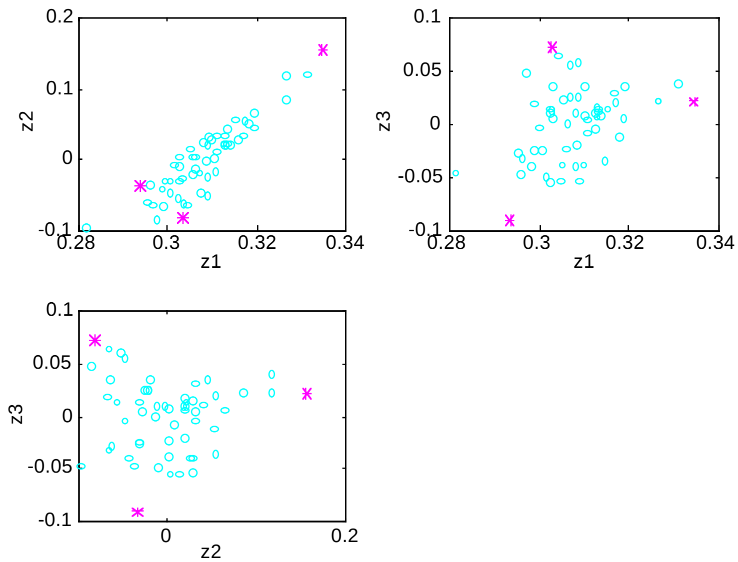

Fig. 4. Principal component plots of the data (o) and the fitted sources by VERTEX 2 (*)

6. Conclusions

This article has been concerned with consistent estimation of source profiles and uncertainty estimation. To eliminate model nonidentifiability problem, new sets of identifiability conditions based on the source contribution matrix were proposed in addition to a set of traditional identification conditions that use preassigned 0's in the source composition matrix. These new conditions usually require less prior information than the conditions based on the source composition matrix. As a method of estimating the source compositions and the contributions, simultaneously, the constrained nonlinear least squares approach was suggested. Two algorithms to find the least squares solution, VERTEX 1 and VERTEX 2 were presented. Each of these can selectively be implemented according to the identifiability conditions that can be achieved in the problem under study. The estimators from VERTEX were shown to be consistent and asymptotically normal under appropriate identifiability conditions.

Acknowledgements

This Research was partially supported by the Texas A&M Center for Environmental and Rural Health through Grant RF9700608 from the National Institute of Environmental Health Sciences and by Grant DMS-9523878 from the Chemistry and Statistics and Probability programs at the National Science Foundation (C.H. Spiegelman and E.S. Park).

References

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis* (2nd ed.), New York: Wiley.

Bandeen-Roche, K., and Ruppert, D. (1991) Source apportionment with one source unknown, *Chemometrics and Intelligent Laboratory Systems*, **10**, 169-184.

Davison, A.C., and Hinkely, D.V. (1997) *Bootstrap Methods and their Application*, New York, Cambridge University Press.

Fuller, W.A. (1987) *Measurement Error Models*, New York: Wiley.

Gleser, L.J. (1983) Functional, structural and ultrastructural errors-in-variable models, *proceedings of the business and economics section*, American Statistical Association, Washington, DC, 57-66.

Henry, R.C. (1991) Multivariate receptor models. In *Receptor Modeling for Air Quality Management* (ed. P. Hopke), pp.117-147. Amsterdam: Elsevier.

———— (1992) Personal communication

———— (1997) History and fundamentals of multivariate air quality receptor models, *Chemometrics and Intelligent Laboratory Systems*, **37**, 37-42.

Henry, R.C., Park, E.S., and Spiegelman, C.H. (1997) Estimating the number of factors to include in a multivariate mixture model, Technical Report No.279.

Henry, R.C., and Kim, B.M. (1990) Extension of self-modeling curve resolution to mixtures of more than three components. part 1. Finding the basic feasible region, *Chemometrics and Intelligent Laboratory Systems*, **8**, 205-216.

Henry, R.C., Lewis, C.W., and Hopke, P.K. (1984) Review of receptor model fundamentals, *Atmospheric Environment*, **18**, 1507-1515.

Henry, R.C., Lewis, C.W., and Collins, J.F. (1994) Vehicle-related hydrocarbon source compositions from ambient data: the grace/safer method, *Environmental Science and Technology*, **28**, 823-832.

Henry, R.C., Spiegelman, C.H., Collins, J.F., and Park, E.S. (1997) Reported emissions of organic gases are not consistent with observations, *Proceedings of National Academy of Science*, **94**, 6596-6599.

Hopke, P.K. (1985) *Receptor Modeling in Environmental Chemistry*, New York: Wiley.

——— (1991) An introduction to receptor modeling, *Chemometrics and Intelligent Laboratory Systems*, **10**, 21-43.

Kiefer, J. and Wolfowitz, J. (1956) Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Annals of Mathematical Statistics*, **27**, 887-906.

Park, E. S. (1997) Multivariate receptor modeling from a statistical science viewpoint, unpublished Ph.D. dissertation, Texas A&M University, Dept. of Statistics.

Rao, C. R. (1973) *Linear Statistical Inference and its Applications* (2nd ed.), New York: Wiley.

Searle, S. R. (1982) *Matrix Algebra Useful for Statistics*, New York: Wiley.

Spiegelman, D.H., and Dattner, S. (1993) Multivariate chemometrics, a case study: applying and developing receptor models for the 1990 El Paso winter PM_{10} receptor modeling scoping study. In *Multivariate Evironmental Statistics* (ed. C.R. Rao), pp. 509-524. Amsterdam: Elsevier.

Yang, H. (1994) Confirmatory factor analysis and its application to receptor modeling, unpublished Ph.D. dissertation, University of Pittsburgh, Dept. of Mathematics and Statistics.

Appendix A: Notations

A : $n \times q$ source contribution matrix

P : $q \times p$ source composition matrix

E : $N \times p$ error matrix

m_i : # of replications for i^{th} source contribution

$\sum_{i=1}^{n} m_i = N$: total number of observations

$\underline{1}_{m_i}$: m_i-dimensional column vector consisting of 1's

$$U = \begin{bmatrix} \underline{1}_{m_1} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1}_{m_2} & \cdots & \underline{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{1}_{m_n} \end{bmatrix}, \; M = \begin{bmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & \ddots & \\ 0 & & & m_n \end{bmatrix}, \; \overline{Y} = \begin{bmatrix} m_1^{-1} \sum_{j=1}^{m_1} \underline{y}_{1j} \\ m_2^{-1} \sum_{j=1}^{m_2} \underline{y}_{2j} \\ \vdots \\ m_n^{-1} \sum_{j=1}^{m_n} \underline{y}_{nj} \end{bmatrix}$$

$$c_i = \lim_{m_i \to \infty} \frac{m_i}{N}, \; m^* = \min_{1 \le i \le n} \{m_i\}, \; c^* = \lim_{m^* \to \infty} \frac{m^*}{N}$$

$$C = \begin{bmatrix} c_1 & & & 0 \\ & c_2 & & \\ & & \ddots & \\ 0 & & & c_n \end{bmatrix}, \quad C^* = \begin{bmatrix} \frac{c^*}{c_1} & & & 0 \\ & \frac{c^*}{c_2} & & \\ & & \ddots & \\ 0 & & & \frac{c^*}{c_n} \end{bmatrix}$$

Σ_p : variance-covariance matrix of error vectors, i.e., $\text{Var}(\underline{\varepsilon}_{ij}) = \Sigma_p$

$$X = AP, \; W = \frac{\partial vec X_0}{\partial \theta^t}, \; F = \frac{\partial vec X_0^t}{\partial \theta^t}.$$

$$Q = \frac{\partial vec A_0}{\partial \theta^t}, \; R = \frac{\partial vec A_0^t}{\partial \theta^t}, \; T = \frac{\partial vec P_0}{\partial \theta^t}, \; U = \frac{\partial vec P_0^t}{\partial \theta^t},$$

$$\alpha_0 = \sum_{i=1}^{n} c_i \alpha_{0i}, \; K_0 = \sum_{i=1}^{n} c_i \left(\alpha_{0i} - \sum_{i=1}^{n} c_i \alpha_{0i} \right)^t \left(\alpha_{0i} - \sum_{i=1}^{n} c_i \alpha_{0i} \right).$$

Appendix B: Proofs

B.1. Proof of proposition 1

Let's call a new source contribution matrix and a new source composition matrix obtained by a linear transformation A^* and P^*, respectively. And, assume $A^*P^* = AP$.

By postmultiplying $A^*P^* = AP$ by P^t, we get

$$A^*P^*P^t = APP^t$$

$$A^*P^*P^t(PP^t)^{-1} = A$$

since PP^t is of full rank by the assumption. Letting $S = P^*P^t(PP^t)^{-1}$,

$$A = A^*S . \qquad \text{(A.1)}$$

Similarly, premultiplying $A^*P^* = AP$ by A^t, we get

$$A^tA^*P^* = A^tAP$$

$$(A^tA)^{-1}A^tA^*P^* = P$$

since A^tA is of full rank by the assumption. Letting $R = (A^tA)^{-1}A^tA^*$,

$$P = RP^* \qquad \text{(A.2)}$$

By (A.1), (A.2), and the assumption that $A^*P^* = AP$, we have

$$AP = A^*S\ RP^* = A^*P^*.$$

Since A^* is of full column rank and P^* is of full row rank, we get from (A.2),

$$(A^{*t}A^*)^{-1}A^*A^*S\ RP^*P^{*t}(P^*P^{*t})^{-1*} = (A^{*t}A^*)^{-1}A^{*t}A^*P^*P^{*t}(P^*P^{*t})^{-1}$$

and hence

$$SR = \mathbf{I}.$$

Note that both of S and R are $q\times q$ full rank matrices. Hence,

$$S = R^{-1}.$$

Using this and (A.1), we get

$$A^* = AR ,$$

and from (A.2),

$$P^* = R^{-1}P .$$

Thus, if $A^*P^* = AP$, then $A^* = AR$ and $P^* = R^{-1}P$ for a nonsingular matrix

$R = (A^tA)^{-1}A^tA^*$. Note that R can have different expressions.

B.2. Proof of result 1.

We need the following lemmas to prove result 1 and result 2. Assume the physical constraints for the source composition matrix and the source contribution matrix, (3), hold throughout.

Lemma A1. Let r_{kj} denote the $(k,j)^{th}$ element of R where $k = 1, \ldots, q$, $j = 1, \ldots, q$. Then

$$\sum_{j=1}^{q} r_{kj} = 1, \quad k = 1, \ldots, q.$$

Proof. We have $P = RP^*$ from $P^* = R^{-1}P$. Thus, $(k,j)^{th}$ element of the matrix P can be expressed as

$$\sum_{i=1}^{q} r_{ki} p_{ij}^* = p_{kj}.$$

Due to the constraint that row sum of P is 1,

$$\sum_{j=1}^{p} \sum_{i=1}^{q} r_{ki} p_{ij}^* = \sum_{j=1}^{p} p_{kj} = 1,$$

and by interchanging the summations,

$$\sum_{i=1}^{q} r_{ki} \sum_{j=1}^{p} p_{ij}^* = 1.$$

It follows from the constraint $\sum_{j=1}^{p} p_{ij}^* = 1$ that

$$\sum_{i=1}^{q} r_{ki} = 1.$$

Lemma A2. Under the assumptions A1-A3,

$$r_{kj} = 0, \quad k = 1, \ldots q, \quad j = 1, \ldots, q, \quad j \neq k.$$

Proof. From $A^* = AR$, the $(i,j)^{th}$ element of the matrix A^* can be expressed as

$$\alpha_{i1}r_{1j} + \alpha_{i2}r_{2j} + \cdots + \alpha_{iq}r_{qj} = \alpha_{ij}^{*}, \quad i = 1,\cdots,n, \quad j = 1,\cdots,q$$

and hence

$$\overline{\alpha}_1 r_{1j} + \overline{\alpha}_2 r_{2j} + \cdots + \overline{\alpha}_q r_{qj} = \overline{\alpha}_j^{*}. \tag{A.3}$$

Say kth ($k = 1,\cdots,q$) source is missing on some days. Then

$$\overline{\alpha}_1^{(k)} r_{1j} + \cdots + \overline{\alpha}_{k-1}^{(k)} r_{k-1,j} + \overline{\alpha}_{k+1}^{(k)} r_{k+1,j} + \cdots + \overline{\alpha}_q^{(k)} r_{qj} = \overline{\alpha}_j^{*(k)}, \tag{A.4}$$

$j = 1,\cdots,q$ since $\overline{\alpha}_k^{(k)} = 0$. Here $\overline{\alpha}_j^{*(k)}$ is defined in the similar way as $\overline{\alpha}_j^{(k)}$.

Substracting (A.4) from (A.3),

$$(\overline{\alpha}_1 - \overline{\alpha}_1^{(k)})r_{1j} + \cdots + (\overline{\alpha}_{k-1} - \overline{\alpha}_{k-1}^{(k)})r_{k-1,j} + \overline{\alpha}_k r_{kj} + (\overline{\alpha}_{k+1} - \overline{\alpha}_{k+1}^{(k)})r_{k+1,j} + \cdots + (\overline{\alpha}_q - \overline{\alpha}_1^{(k)})r_{qj}$$

$$= \overline{\alpha}_j^{*} - \overline{\alpha}_j^{*(k)}, \tag{A.5}$$

By applying A2, we have

$$\overline{\alpha}_k r_{kj} = 0$$

for $j \neq k$.

This implies $r_{kj} = 0$ since $\overline{\alpha}_k \neq 0$ by A3.

Proof of resutl. From lemma A2,

$$r_{kj} = 0, \quad k = 1,\cdots q, \quad j = 1,\cdots,q, \quad j \neq k.$$

Using this and lemma A1 together, we get

$$\sum_{j=1}^{q} r_{kj} = r_{kk} = 1, \quad k = 1,\cdots q.$$

This completes the proof.

B.3. Proof of result 2.

We need the following lemmas additionaly to prove result 2.

Lemma A3. Under the assumptions B1 and B3, for $k = 1,\cdots,q$, $j = 1,\cdots,q$,

$$\left| r_{kj} \right| \leq \frac{pq^2}{\lambda_q}$$

where λ_q is the smallest eigenvalue of P^*P^{*t}.

Proof. Recall that $P = RP^*$. Postmultiplying this by $P^{*t}(P^*P^{*t})^{-1}$, we get

$$R = PP^{*t}(P^*P^{*t})^{-1}.$$

Let $C = PP^{*t}$ and $D = \left(P^*P^{*t}\right)^{-1}$. The *(k,j)*th element of *C* can be expressed by

$$c_{kj} = \sum_{i=1}^{p} p_{ki} p_{ji}^* .$$

Each summand, $p_{ki}p_{ji}^*$, in the above equation is bounded by 1 due to the constraints that $0 \le p_{ki} \le 1$ and $0 \le p_{ji}^* \le 1$. Hence,

$$0 \le c_{kj} = \sum_{i=1}^{p} p_{ki} p_{ji}^* \le p, \qquad \text{(A.6)}$$

$k = 1, \ldots, q$, $j = 1, \ldots, q$. Note that the matrix P^*P^{*t} is symmetric. By the diagonability of symmetric matrices (see, e.g., Searle 1982, sec. 11.6b), we have the expression that

$$P^*P^{*t} = U\Lambda U^t$$

where Λ is the diagonal matrix of the eigenvalues of P^*P^{*t} and *U* is the orthogonal matrix consisting of the eigenvectors of P^*P^{*t}, and hence

$$D = \left(P^*P^{*t}\right)^{-1} = \left(U\Lambda U^t\right)^{-1} = U\Lambda^{-1}U^t \qquad \text{(A.7)}$$

Using (A.7), the *(i,j)*th element of *D* can be expressed by

$$d_{ij} = \sum_{s=1}^{q} \frac{1}{\lambda_s} u_{is} u_{js}, \quad i = 1, \ldots, q, \qquad j = 1, \ldots, q.$$

Note that $|u_{is}| \le 1$, $|u_{js}| \le 1$ since *U* is a orthogonal matrix. Letting λ_q be the smallest eigenvalue of P^*P^{*t}, we obtain for $k = 1, \ldots, q$, $j = 1, \ldots, q$,

$$\left|d_{ij}\right| = \left|\sum_{s=1}^{q} \frac{1}{\lambda_s} u_{is} u_{js}\right| \le \sum_{s=1}^{q} \frac{1}{\lambda_s} \left|u_{is} u_{js}\right| \le \frac{q}{\lambda_q} \qquad \text{(A.8)}$$

by the triangle inequality and the nonnegativity of the eigenvalues of a nonnegative definite matrix. Note that $\lambda_q > 0$ due to the assumption B3.

The $(k,j)^{th}$ element of the matrix $R = CD$ is written by

$$r_{kj} = \sum_{i=1}^{q} c_{ki} d_{ij} .$$

By the triangle inequality, (A.6), and (A.8),

$$\left|r_{kj}\right| = \left|\sum_{i=1}^{q} c_{ki} d_{ij}\right| \le \sum_{i=1}^{q} c_{ki}\left|d_{ij}\right| \le \frac{q}{\lambda_q}\sum_{i=1}^{q} c_{ki} \le \frac{pq^2}{\lambda_q},$$

$k = 1,\ldots,q,\ \ j = 1,\ldots,q.$

Lemma A4. Under the assumptions B1 - B3, for $k = 1,\ldots,q,\ \ j = 1,\ldots,q,\ k \ne j,$

$$\left|r_{kj}\right| \le B$$

where B is defined by $B = \dfrac{\varepsilon}{\overline{\alpha}_k}\left\{1 + \dfrac{pq^2(q-1)}{\lambda_q}\right\}.$

Proof. From (A.5) in the proof of lemma A2, we get

$$(\overline{\alpha}_1 - \overline{\alpha}_1^{(k)})r_{1j} + \ldots + (\overline{\alpha}_{k-1} - \overline{\alpha}_{k-1}^{(k)})r_{k-1,j} + \overline{\alpha}_k r_{kj} + (\overline{\alpha}_{k+1} - \overline{\alpha}_{k+1}^{(k)})r_{k+1,j} + \ldots + (\overline{\alpha}_q - \overline{\alpha}_1^{(k)})r_{qj}$$

$$= \overline{\alpha}_j^{*} - \overline{\alpha}_j^{*(k)},$$

That is,

$$\overline{\alpha}_k r_{kj} = (\overline{\alpha}_j^{*} - \overline{\alpha}_j^{*(k)}) - (\overline{\alpha}_1 - \overline{\alpha}_1^{(k)})r_{1j} - \ldots - (\overline{\alpha}_{k-1} - \overline{\alpha}_{k-1}^{(k)})r_{k-1,j} - (\overline{\alpha}_{k+1} - \overline{\alpha}_{k+1}^{(k)})r_{k+1,j} - \ldots - (\overline{\alpha}_q - \overline{\alpha}_q^{(k)})r_{qj}$$

and hence for $k = 1,\ldots,q,\ \ j = 1,\ldots,q,\ k \ne j,$

$$\begin{aligned}\left|\overline{\alpha}_k r_{kj}\right| &= \left|(\overline{\alpha}_j^{*} - \overline{\alpha}_j^{*(k)}) - (\overline{\alpha}_1 - \overline{\alpha}_1^{(k)})r_{1j} - \ldots - (\overline{\alpha}_{k-1} - \overline{\alpha}_{k-1}^{(k)})r_{k-1,j} - (\overline{\alpha}_{k+1} - \overline{\alpha}_{k+1}^{(k)})r_{k+1,j} - \ldots - (\overline{\alpha}_q - \overline{\alpha}_q^{(k)})r_{qj}\right| \\ &\le \left|\overline{\alpha}_j^{*} - \overline{\alpha}_j^{*(k)}\right| + \left|\overline{\alpha}_1 - \overline{\alpha}_1^{(k)}\right|\left|r_{1j}\right| + \ldots + \left|\overline{\alpha}_{k-1} - \overline{\alpha}_{k-1}^{(k)}\right|\left|r_{k-1,j}\right| + \left|\overline{\alpha}_{k+1} - \overline{\alpha}_{k+1}^{(k)}\right|\left|r_{k+1,j}\right| + \ldots + \left|\overline{\alpha}_q - \overline{\alpha}_q^{(k)}\right|\left|r_{qj}\right|\end{aligned}$$

$$\leq \varepsilon + \varepsilon\left(\frac{pq^2}{\lambda_q}\right) + \cdots + \varepsilon\left(\frac{pq^2}{\lambda_q}\right) = \varepsilon + \varepsilon\left(\frac{pq^2}{\lambda_q}\right)(q-1).$$

by the triangle inequality, B2, and lemma A3.

Thus

$$\left|\overline{\alpha}_k r_{kj}\right| \leq \varepsilon\left\{1 + \frac{pq^2(q-1)}{\lambda_q}\right\}.$$

It follows from B3 that

$$\left|r_{kj}\right| \leq \frac{\varepsilon}{\overline{\alpha}_k}\left\{1 + \frac{pq^2(q-1)}{\lambda_q}\right\} = B,$$

where $k = 1, \cdots, q$, $j = 1, \cdots, q$, $k \neq j$.

Proof of result. From lemma A4, $\left|r_{kj}\right| \leq B$, $k = 1, \cdots q$, $j = 1, \cdots, q$, $k \neq j$. Using this and lemma A1 and by the assumption that B is small enough, we get

$$1 = \sum_{j=1}^{q} r_{kj} \approx r_{kk}, \quad k = 1, \cdots q.$$

This completes the proof.

B.4. Proof of theorem 1.

We need the following lemmas to prove the theorem.

Lemma A5. Let $y_i = \alpha_{0i}P_0 + \varepsilon_i$, where the ε_i are independent identically distributed p-dimensional random row vectors with zero mean vector, positive definite covariance matrix Σ_0. Let $\{\alpha_{0i}\}$ be a fixed sequence satisfying

$$\overline{\alpha}_0 = n^{-1}\sum_{i=1}^{n}\alpha_{0i} \xrightarrow{n\to\infty} \alpha_0$$

and

$$\overline{\mathbf{K}}_0 = n^{-1}\sum_{i=1}^{n}(\alpha_{0i} - \overline{\alpha}_0)^t(\alpha_{0i} - \overline{\alpha}_0) \xrightarrow{n\to\infty} \mathbf{K}_0.$$

Then, when $n \to \infty$,

$$\bar{y} = n^{-1}\sum_{i=1}^{n} y_i \xrightarrow{p} \alpha_0 P_0,$$

and

$$S = n^{-1}\sum_{i=1}^{n} (y_i - \bar{y})^t (y_i - \bar{y}) \xrightarrow{p} \Sigma_0 + P_0^t \mathbf{K}_0 P_0.$$

Proof. This result follows from *WLLN*. A detailed proof can be found in Park (1997).

Lemma A6. Let $g(x,y)$ be a continuous real valued function defined on the Cartesian product $A \times B$, where A is a subset of p-dimensional Euclidean space and B is a compact subset of q-dimensional Euclidean space. Let x_0 be an interior point of A. Assume that the point y_0 is the unique point for which $Min_{y \in B} g(x_0, y)$ is attained. Let $y_m(x)$ be a point in B such that

$$g(x, y_m(x)) = Min_{y \in B} g(x, y).$$

Then $y_m(x)$ is a continuous function of x at $x = x_0$.

Proof. Appendix 4.B of Fuller (1987).

Proof of theorem. Let $w_n = (\bar{y}, vech(S)^t)^t$, and $\theta = vec(P)$. Let Θ be the compact subset of pq-dimensional Euclidean space. Define $g(w_n;\theta) = tr\left[(S + \bar{y}^t\bar{y})\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$. Note that $g(w_n;\theta)$ is a continuous real valued function defined on $\boldsymbol{W} \times \Theta$. By lemma A5, as $n \to \infty$,

$$w_n \xrightarrow{p} \omega_0$$

We show that $\underset{\theta \in \Theta}{Min}\, g(\omega_0;\theta) = \underset{\theta \in \Theta}{Min}\, tr\left[(\Sigma_0 + P_0^t \mathbf{K}_0 P_0 + P_0^t \alpha_0^t \alpha_0 P_0)\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$ is uniquely attained at $\theta = \theta_0$ if $\Sigma_0 = \sigma^2 \mathbf{I_p}$. Assuming $\Sigma_0 = \sigma^2 \mathbf{I_p}$,

$$g(\omega,\theta) = tr\left[(\sigma^2 I_p + P_0^t \mathbf{K}_0 P_0 + P_0^t \alpha_0^t \alpha_0 P_0)\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$$

$$= \sigma^2 tr\left[I_p - P^t(PP^t)^{-1}P\right] + tr\left[\mathbf{K}_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t\right]$$
$$+ tr\left[\alpha_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t\alpha_0^t\right]$$
$$= \sigma^2(p-q) + tr\left[\mathbf{K}_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t\right] + tr\left[\alpha_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t\alpha_0^t\right].$$

Note that $I_p - P^t(PP^t)^{-1}P$ is a projection matrix and so columns of $I_p - P^t(PP^t)^{-1}P$ are orthogonal to and linearly independent of columns of P^t (rows of P). Since K_0 is of full rank, K_0P_0 also span the row space of P_0. Since the projection matrix is unique (see, e.g., Rao 1973, sec 1c.4), $tr\left[K_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t\right]$ has a unique minimum when

$P^t(PP^t)^{-1}P = P_0^t(P_0P_0^t)^{-1}P_0$, which is true for $P = RP_0$ for any $q\times q$ nonsingular matrix R.

By the identifiability conditions on P, C1-C2, R should be an identity matrix, and hence, $tr\left[K_0 P_0\left\{I_p - P^t(PP^t)^{-1}P\right\}P_0^t K_0^t\right]$ has a unique minimum 0 at $P = P_0$. Thus,

$$\underset{\theta\in\Theta}{Min}\, g(\omega_0;\theta)$$
$$= \sigma^2(p-q) + \underset{\theta\in\Theta}{Min}\, tr\left\{\mathbf{K}_0 P_0\left(I_p - P^t(PP^t)^{-1}P\right)P_0^t\right\} + tr\left\{\alpha_0 P_0\left(I_p - P^t(PP^t)^{-1}P\right)P_0^t\alpha_0^t\right\}$$
$$= \sigma^2(p-q) + tr\left[\mathbf{K_0} P_0\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0^t\right\}P_0^t\right] + tr\left[\alpha_0 P_0\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0^t\right\}P_0^t\alpha_0^t\right]$$
$$= \sigma^2(p-q)$$

is uniquely attained at $P = P_0$, i.e., $\theta = \theta_0$. By lemma A6, $vec(\hat{P})$ which is the value of $vec(P)$ such that $g(w_n;vec(\hat{P})) = \underset{\theta\in\Theta}{Min}\, g(w_n;vec(P))$ is a continuous function of w_n and the result follows from the continuous mapping theorem.

B.5. Proof of theorem 2

We will need the following lemmas in the sequel to prove theorem 2.

Lemma A7. Let $y_i = \alpha_{0i}P_0 + \varepsilon_i$, where the ε_i are independent identically distributed p-dimensional random row vectors with zero mean vector, positive definite covariance matrix Σ_0, and finite fourth moments. Let $\{\alpha_{0i}\}$ be a fixed sequence satisfying

$$\bar{\alpha}_0 = n^{-1}\sum_{i=1}^{n}\alpha_{0i} \xrightarrow{n\to\infty} \alpha_0$$

and

$$\bar{\mathbf{K}}_0 = n^{-1}\sum_{i=1}^{n}(\alpha_{0i}-\bar{\alpha}_0)^t(\alpha_{0i}-\bar{\alpha}_0) \xrightarrow{n\to\infty} \mathbf{K}_0.$$

Let $\hat{\gamma} = \left[\bar{y},(vechS)^t\right]^t$ and $\gamma_n = \left[\bar{\alpha}_0 P_0,\left(vech\bar{\mathbf{m}} + vech\Sigma_0\right)^t\right]^t$ where $\bar{\mathbf{m}} = P_0^t\bar{\mathbf{K}}_0P_0$. Then

$$\mathbf{G}_n^{-1/2}\left(\hat{\gamma}-\gamma_n\right) \xrightarrow{d} N(\mathbf{0},\mathbf{I}),$$

where the elements of $\mathbf{G}_n$ are the covariances of the elements of $\hat{\gamma}$,

$$Cov\left(\bar{y}_i,\bar{y}_j\right) = n^{-1}\sigma_{ij},$$
$$Cov\left(\bar{y}_i,S_{jk}\right) = \frac{n-1}{n^2}\tau_{ijk},$$
$$C\left(S_{ij},S_{kl}\right) = \frac{n-1}{n^2}\left(\bar{m}_{ik}\sigma_{jl} + \bar{m}_{il}\sigma_{jk} + \bar{m}_{jk}\sigma_{il} + \bar{m}_{jl}\sigma_{ik} + \kappa_{ij,kl}\right) + O(n^{-2}),$$
$$\kappa_{ij,kl} = E\left\{\left(\varepsilon_i\varepsilon_j - \sigma_{ij}\right)\left(\varepsilon_k\varepsilon_l - \sigma_{kl}\right)\right\}$$

$\sigma_{ij} = E\left(\varepsilon_i\varepsilon_j\right)$, and $\tau_{ijk} = E\left(\varepsilon_i\varepsilon_j\varepsilon_k\right)$.

Proof. It is a direct adaptation of theorem 1.C.2 of Fuller (1987).

Lemma A8. Under the definitions and the assumptions of lemma A7,

$$\sqrt{n}\left(\hat{\gamma}-\gamma_n\right) \xrightarrow{d} N(\mathbf{0},\mathbf{G}),$$

where the elements of ***G*** are the limiting values of the elements of $n\boldsymbol{G}_n$,

$$\lim_{n\to\infty} nCov\left(\bar{y}_i,\bar{y}_j\right) = \sigma_{ij},$$
$$\lim_{n\to\infty} nCov\left(\bar{y}_i,S_{jk}\right) = \tau_{ijk}$$
$$\lim_{n\to\infty} nCov\left(S_{ij},S_{kl}\right) = \left(m_{ik}\sigma_{jl} + m_{il}\sigma_{jk} + m_{jk}\sigma_{il} + m_{jl}\sigma_{ik} + \kappa_{ij,kl}\right),$$

and m_{ij}'s are the elements of $\mathbf{m} = P_0^t\mathbf{K}_0P_0$.

Proof. It follows from the fact that $\sqrt{n}\mathbf{G}_n^{1/2} \xrightarrow{n\to\infty} \mathbf{G}^{1/2}$ in probability and Slutsky theorem.

Lemma A9. Let $g_1(vechS,\theta) = tr\left[S\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$. Then

(a) $\frac{\partial}{\partial\theta} g_1(vechS,\theta_0) = -\Big[L^t\Big\{\Big(I_p - P_0^t(P_0P_0^t)^{-1}P_0\Big)\otimes(P_0P_0^t)^{-1}P_0\Big\}$

$$+M^t\Big\{(P_0P_0^t)^{-1}P_0\otimes\Big(I_p - P_0^t(P_0P_0^t)^{-1}P_0\Big)\Big\}\Big]\Phi_p vech\Big[S-\Big(\Sigma_0 + P_0^t\overline{K}_0P_0\Big)\Big]$$

where $L = \frac{\partial vecP_0}{\partial\theta^t} = \frac{\partial vecP}{\partial\theta^t}\Big|_{P=P_0}$, $M = \frac{\partial vecP_0^t}{\partial\theta^t} = \frac{\partial vecP^t}{\partial\theta^t}\Big|_{P=P_0}$, and Φ_p is the $p^2\times\frac{1}{2}p(p+1)$ matrix such that $vec\,\Big(S-\Big(I_p + P_0^t\overline{K}_0P_0\Big)\Big) = \Phi_p\, vech\,\Big(S-\Big(I_p + P_0^t\overline{K}_0P_0\Big)\Big)$.

(b) $\frac{\partial^2}{\partial\theta\partial\theta^t} g_1(vechS,\theta_0) = 2L^t\Big[\Big\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\Big\}\otimes\overline{\mathbf{K}}_\mathbf{0}\Big]L + o_p(1)$

where $L = \frac{\partial vecP_0}{\partial\theta^t} = \frac{\partial vecP}{\partial\theta^t}\Big|_{P=P_0}$.

Proof. A detailed proof can be found in Park (1997).

Lemma A10. Let $g_2(\bar{y},\theta) = \bar{y}(I_p - P^t(PP^t)^{-1}P)\bar{y}^t$. Then,

(a) $\frac{\partial g_2(\bar{y},\theta_0)}{\partial\theta} = -2L^t\Big[\Big\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\Big\}\otimes\overline{\alpha}_0^t\Big]\Big(\bar{y}-\overline{\alpha}_0P_0\Big) + o_p(n^{-\frac{1}{2}})$.

(b) $\frac{\partial^2 g_2(\bar{y},\theta_0)}{\partial\theta\partial\theta^t} = 2L^t\Big[\Big\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\Big\}\otimes\overline{\alpha}_0^t\overline{\alpha}_0\Big]L + O_p(n^{-\frac{1}{2}})$.

Proof. A detailed proof can be found in Park (1997).

Proof of theorem. Let $\Theta = \{0 \le \theta_i \le 1,\ i=1,\ldots,r\}$. First, we show with probability approaching one $\hat{\theta}$ is in the interior of the parameter space. That is, we need to show that there is an open r-ball with center $\hat{\theta}$, all of whose points belong to Θ. Let's denote the set of all points θ in R^r such that $\|\theta - a\| < d$, which is an open r-ball of radius d and center a, by $B(a;d)$. By the assumption, $\theta_0 \in Int\,(\Theta)$, and hence there exists a $\delta_0 > 0$ such that $B(\theta_0;\ \delta_0) \subset \Theta$. By theorem 1 $\hat{\theta}$ is consistent for θ_0. Therefore, with probability approaching one as n increases, $\|\hat{\theta}-\theta_0\| < \varepsilon$. Setting $\varepsilon = \delta_0/2$, we have

$$B(\hat{\theta};\ \delta_0/2) \subset B(\theta_0;\ \delta_0) \subset \Theta$$

with probability approaching one as n increases. Thus $\hat{\theta} \in Int(\Theta)$ with probability approaching one as n increases.

The rest of proof is based on Taylor's theorem and the asymptotic normality of $\bar{y}$ and S. The same argument in the proof of theorem 4.B.2 in Fuller (1987) can be used for our case too. We only need to calculate the first and second derivatives of the objective function $g(\bar{y}, S; \theta)$ with respect to θ since the objective function which we minimize is different from that of Fuller. From (4.B.19) in Fuller (1987), we obtain with probability approaching one as $n \to \infty$,

$$\hat{\theta} - \theta = -\left\{ \frac{\partial^2 g(\bar{y}, vechS; \theta^*)}{\partial\theta\partial\theta^t} \right\}^{-1} \frac{\partial g(\bar{y}, vechS; \theta_0)}{\partial\theta} \qquad \text{(A.9)}$$

where the elements of θ^* are evaluated at points on the line segment joining θ_0 and $\hat{\theta}$.

Note that our objective function $g(\bar{y}, vechS; \theta)$ can be reexpressed as follows:

$$g(\bar{y}, vechS; \theta) = tr\left[(S + \bar{y}^t\bar{y})\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$$

$$= tr\left[S\left\{I_p - P^t(PP^t)^{-1}P\right\}\right] + \bar{y}\left\{I_p - P^t(PP^t)^{-1}P\right\}\bar{y}^t.$$

Let $g_1(vechS, \theta) = tr\left[S\left\{I_p - P^t(PP^t)^{-1}P\right\}\right]$ and $g_2(\bar{y}, \theta) = \bar{y}\left\{I_p - P^t(PP^t)^{-1}P\right\}\bar{y}^t$. Then,

$$\frac{\partial g(\bar{y}, vechS; \theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left\{g_1(vechS, \theta) + g_2(\bar{y}, \theta)\right\}$$

$$= \frac{\partial}{\partial\theta} g_1(vechS, \theta) + \frac{\partial}{\partial\theta} g_2(\bar{y}, \theta). \qquad \text{(A.10)}$$

By lemma A9 (a),

$$\frac{\partial}{\partial\theta} g_1(vechS, \theta_0) = -\Big[L^t\left\{\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right) \otimes (P_0P_0^t)^{-1}P_0\right\}$$

$$+ M^t\left\{(P_0P_0^t)^{-1}P_0 \otimes \left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\right\}\Big]\Phi_p vech\left[S - \left(\Sigma_0 + P_0^t\overline{K}_0P_0\right)\right].$$

By lemma A10 (a),

$$\frac{\partial g_2(\bar{y},\theta_0)}{\partial\theta} = -2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\bar{\alpha}_0^t\right]\left(\bar{y}-\bar{\alpha}_0P_0\right)+o_p(n^{-\frac{1}{2}}).$$

Let

$$\mathbf{B_1} = \left[L^t\left\{\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\otimes(P_0P_0^t)^{-1}P_0\right\}+M^t\left\{(P_0P_0^t)^{-1}P_0\otimes\left(I_p - P_0^t(P_0P_0^t)^{-1}P_0\right)\right\}\right]\mathbf{\Phi}_p$$

and

$$\overline{\mathbf{B}}_2 = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\bar{\alpha}_0^t\right].$$

Then it follows from (A.10) that

$$\frac{\partial g(\bar{y},S;\theta_0)}{\partial\theta} = -\mathbf{B}_1 vech\left[S-\left(\Sigma_0+P_0^t\overline{K}_0P_0\right)\right]-\overline{\mathbf{B}}_2\left(\bar{y}-\bar{\alpha}_0P_0\right)^t+o_p(n^{-\frac{1}{2}}). \qquad \text{(A.11)}$$

Note that

$$\frac{\partial^2 g(\bar{y},vechS;\theta)}{\partial\theta\partial\theta^t} = \frac{\partial^2}{\partial\theta\partial\theta^t}g_1(vechS,\theta)+\frac{\partial^2}{\partial\theta\partial\theta^t}g_2(\bar{y},\theta). \qquad \text{(A.12)}$$

By lemma A9 (b),

$$\frac{\partial^2}{\partial\theta\partial\theta^t}g_1(vechS,\theta_0) = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\overline{\mathbf{K}}_\mathbf{0}\right]L+o_p(1).$$

By lemma A10 (b),

$$\frac{\partial^2 g_2(\bar{y},\theta_0)}{\partial\theta\partial\theta^t} = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\bar{\alpha}_0^t\bar{\alpha}_0\right]L+O_p(n^{-\frac{1}{2}}).$$

Let $\overline{\mathbf{H}}_\mathbf{1} = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\overline{\mathbf{K}}_\mathbf{0}\right]L$ and $\overline{\mathbf{H}}_2 = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\}\otimes\bar{\alpha}_0^t\bar{\alpha}_0\right]L$.

Then it follows from (A.12) that

$$\frac{\partial^2 g_1(vechS,\theta_0)}{\partial\theta\partial\theta^t} = \overline{\mathbf{H}}_\mathbf{1}+\overline{\mathbf{H}}_\mathbf{2}+o_p(1)\cdot \qquad \text{(A.13)}$$

By theorem 1,

$$\hat{\theta}\xrightarrow{p}\theta_0 \text{ as } n\to\infty,$$

and so θ^* also converges to θ_0 since θ^* is between $\hat{\theta}$ and θ_0. Because the partial derivatives are continuous,

$$\frac{\partial^2 g(\bar{y}, vechS; \theta^*)}{\partial\theta\partial\theta^t} = \frac{\partial^2 g(\bar{y}, vechS; \theta_0)}{\partial\theta\partial\theta^t} + o_p(1)$$

$$= \overline{\mathbf{H}}_\mathbf{1} + \overline{\mathbf{H}}_\mathbf{2} + o_p(1)$$

(A.14)

where we have used (A.13). Letting $\overline{\mathbf{H}} = \overline{\mathbf{H}}_\mathbf{1} + \overline{\mathbf{H}}_\mathbf{2}$ and $\overline{\mathbf{B}} = (\overline{\mathbf{B}}_2, \mathbf{B}_\mathbf{1})$, we obtain that

$$\hat{\theta} - \theta_0 = \overline{\mathbf{H}}^{-1}\overline{\mathbf{B}}\left[\bar{y} - \bar{\alpha}_0 P_0, \left\{vech\mathbf{S} - vech\left(\Sigma_0 + P_0^t \overline{K}_0 P_0\right)\right\}^t\right]^t + o_p(n^{-\frac{1}{2}}), \qquad \text{(A.15)}$$

where we have used (A.9), (A.11), and (A.14).

Let $\mathbf{B}_2 = -2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\} \otimes \alpha_0^t\right]$, $\mathbf{B} = (\mathbf{B}_2, \mathbf{B}_\mathbf{1})$ and

$\mathbf{H} = 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\} \otimes \mathbf{K_0}\right]L + 2L^t\left[\left\{I_p - P_0^t(P_0P_0^t)^{-1}P_0\right\} \otimes \alpha_0^t\alpha_0\right]L$. Then, by the assumptions $\bar{\alpha}_0 \xrightarrow{n\to\infty} \alpha_0$, $\overline{\mathbf{K}}_\mathbf{0} \xrightarrow{n\to\infty} \mathbf{K_0}$, and hence by the continuous mapping theorem,

$$\overline{\mathbf{B}} \xrightarrow{n\to\infty} \mathbf{B}, \quad \overline{\mathbf{H}} \xrightarrow{n\to\infty} \mathbf{H}. \qquad \text{(A.16)}$$

It follows from (A.15), (A.16), lemma A8, and the continuous mapping theorem that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}^{-1}\mathbf{B}\mathbf{G}\mathbf{B}^\mathbf{t}\mathbf{H}^{-1}\right)$$

where **G** is defined in lemma A8.

B.6. Proof of proposition 2

By the *WLLN*, as $m_i \to \infty$, $i = 1, ..., n$,

$$\overline{Y} \xrightarrow{p} A_0 P_0.$$

Let $g(\overline{Y}, N^{-1}M; A) = \frac{1}{N} tr\left[\overline{Y}^t M\left\{\mathbf{I_n} - A(A^tMA)^{-1}A^tM\right\}\overline{Y}\right]$. Then by the continuous mapping theorem,

$$g(\overline{Y}, N^{-1}M; A) \xrightarrow{p} g(A_0P_0, C; A)$$

where $C = \begin{bmatrix} c_1 & & & 0 \\ & c_2 & & \\ & & \mathbf{0} & \\ 0 & & & c_n \end{bmatrix}$ and $c_i = \lim_{m_i \to \infty} \frac{m_i}{N} > 0.$

Note that

$$\underset{\theta\in\Theta}{Min}\, g(A_0P_0, C; A) = \underset{\theta\in\Theta}{Min}\left[tr\left\{P_0^tA_0^tC\left(\mathbf{I_n} - A(A^tCA)^{-1}A^tC\right)A_0P_0\right\}\right]$$

$$= \underset{\theta\in\Theta}{Min}\left[tr\left\{P_0^tA_0^tC^{\frac{1}{2}}\left(\mathbf{I_n} - C^{\frac{1}{2}}A(A^tCA)^{-1}A^tC^{\frac{1}{2}}\right)C^{\frac{1}{2}}A_0P_0\right\}\right]$$

is uniquely attained when $C^{\frac{1}{2}}A(A^tCA)^{-1}A^tC^{\frac{1}{2}} = C^{\frac{1}{2}}A_0(A_0^tCA_0)^{-1}A_0^tC^{\frac{1}{2}}$ since the projection matrix is unique. By the identifiability of the model parameters discussed in section 2, this implies that $A = A_0$. Thus, $\underset{\theta\in\Theta}{Min}\, g(A_0P_0, C; A)$ is uniquely attained at $A = A_0$. By lemma A6, $\hat{A}$ which is the value of A such that $g(\overline{Y}, N^{-1}M; \hat{A}) = \underset{A}{Min}\, g(\overline{Y}, N^{-1}M; A)$ is a continuous function of $\left(\overline{Y}, N^{-1}M\right)$ and the result follows from the continuous mapping theorem.

B.7. Proof of proposition 3.

We need the following lemmas in the sequel to prove the proposition.

Lemma A11. Let $M = \begin{bmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & \mathbf{0} & \\ 0 & & & m_n \end{bmatrix}$ and $m^* = \min_{1\le i\le n}\{m_i\}$. Then

(a) $vec\left(M^{\frac{1}{2}}(\overline{Y} - A_0P_0)\right)^t \xrightarrow{d} N\left(\mathbf{0_{np}}, \mathbf{I_n} \otimes \Sigma_p\right)$

(b) $vec\left(\sqrt{m^*}(\overline{Y} - A_0P_0)\right)^t \xrightarrow{d} N\left(\mathbf{0_{np}}, C^* \otimes \Sigma_p\right)$

where $C^* = \begin{bmatrix} \frac{c^*}{c_1} & & & 0 \\ & \frac{c^*}{c_2} & & \\ & & \mathbf{0} & \\ 0 & & & \frac{c^*}{c_n} \end{bmatrix}$, $c^* = \lim_{m^* \to \infty} \frac{m^*}{N}$, and $c_i = \lim_{m_i \to \infty} \frac{m_i}{N}$.

Proof. (a) $M^{\frac{1}{2}}(\bar{Y} - A_0 P_0) = \begin{bmatrix} \sqrt{m_1}\left(\bar{y}_{1\bullet} - a_{01} P_p\right) \\ \sqrt{m_2}\left(\bar{y}_{2\bullet} - a_{02} P_p\right) \\ \vdots \\ \sqrt{m_n}\left(\bar{y}_{n\bullet} - a_{0n} P_p\right) \end{bmatrix}$.

By the multivariate CLT, $\sqrt{m_i}\left(\bar{y}_{i\bullet} - a_{0i} P_0\right)^t \xrightarrow{d} N\left(\mathbf{0_p}, \Sigma_p\right)$, $i = 1, \ldots, n$. The conclusion follows because $\bar{y}_{1\bullet}, \ldots, \bar{y}_{n\bullet}$ are independent.

(b) The conclusion is immediate because $\sqrt{m^*}\,(\bar{Y} - A_0 P_0) = \sqrt{m^*}\, M^{-\frac{1}{2}} M^{\frac{1}{2}}(\bar{Y} - A_0 P_0)$ and

$$\sqrt{m^*}\, M^{-\frac{1}{2}} = \begin{bmatrix} \frac{m^*}{m_1} & & 0 \\ & \frac{m^*}{m_2} & \\ & \ddots & \\ 0 & & \frac{m^*}{m_n} \end{bmatrix}^{\frac{1}{2}} \xrightarrow{m_i \to \infty,\, i = 1, \ldots, n} \begin{bmatrix} \frac{c^*}{c_1} & & 0 \\ & \frac{c^*}{c_2} & \\ & \ddots & \\ 0 & & \frac{c^*}{c_n} \end{bmatrix}^{\frac{1}{2}} = C^{*\frac{1}{2}}.$$

Lemma A12. Let M be an $n \times n$ diagonal matrix and A be an $n \times q$ matrix, and $\theta = vecA$. Then

(a) $M \dfrac{\partial}{\partial \theta_i}\left\{I_n - A(A^t M A)^{-1} A^t M\right\} A = -M\left\{I_n - A(A^t M A)^{-1} A^t M\right\} \dfrac{\partial A}{\partial \theta_i}$.

(b) $A^t M \dfrac{\partial}{\partial \theta_i}\left\{I_n - A(A^t M A)^{-1} A^t M\right\} = -\dfrac{\partial A^t}{\partial \theta_i} M\left\{I_n - A(A^t M A)^{-1} A^t M\right\}$.

(c) $A^t M \dfrac{\partial}{\partial \theta_i}\left\{I_n - A(A^t M A)^{-1} A^t M\right\} A = \mathbf{0_{q \times q}}$.

(d) $A^t M \dfrac{\partial^2}{\partial \theta_i \partial \theta_j}\left\{I_n - A(A^t M A)^{-1} A^t M\right\} A$

$$= \frac{\partial A^t}{\partial \theta_j} M\left\{I_n - A(A^t M A)^{-1} A^t M\right\} \frac{\partial A}{\partial \theta_i} + \frac{\partial A^t}{\partial \theta_i} M\left\{I_n - A(A^t M A)^{-1} A^t M\right\} \frac{\partial A}{\partial \theta_j}.$$

Proof. (a) For an element θ_i of θ,

$$\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^t M A)^{-1} A^t M\right\}$$

$$= \frac{\partial A}{\partial \theta_i}\left(A^t MA\right)^{-1} A^t M - A\frac{\partial\left(A^t MA\right)^{-1}}{\partial \theta_i} A^t M - A\left(A^t MA\right)^{-1}\frac{\partial A^t}{\partial \theta_i} M$$

$$= \frac{\partial A}{\partial \theta_i}\left(A^t MA\right)^{-1} A^t M - A\left\{\left(A^t MA\right)^{-1}\frac{\partial\left(A^t MA\right)}{\partial \theta_i}\left(A^t MA\right)^{-1}\right\} A^t M - A\left(A^t MA\right)^{-1}\frac{\partial A^t}{\partial \theta_i} M$$

$$= \frac{\partial A}{\partial \theta_i}\left(A^t MA\right)^{-1} A^t M - A\left(A^t MA\right)^{-1}\left\{\frac{\partial A^t}{\partial \theta_i} MA + A^t M\frac{\partial A}{\partial \theta_i}\right\}\left(A^t MA\right)^{-1} A^t M - A\left(A^t MA\right)^{-1}\frac{\partial A^t}{\partial \theta_i} M$$

$$= \frac{\partial A}{\partial \theta_i}\left(A^t MA\right)^{-1} A^t M - A\left(A^t MA\right)^{-1}\frac{\partial A^t}{\partial \theta_i} MA\left(A^t MA\right)^{-1} A^t M$$

$$-A\left(A^t MA\right)^{-1} A^t M\frac{\partial A}{\partial \theta_i}\left(A^t MA\right)^{-1} A^t M - A\left(A^t MA\right)^{-1}\frac{\partial A^t}{\partial \theta_i} M$$

$$= -\left\{I_n - A(A^t MA)^{-1} A^t M\right\}\frac{\partial A}{\partial \theta_i}(A^t MA)^{-1} A^t M - A(A^t MA)^{-1}\frac{\partial A^t}{\partial \theta_i} M\left\{I_n - A(A^t MA)^{-1} A^t M\right\} \tag{A.17}$$

Postmultiplying both sides of the above equation by *A*, we get

$$M\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^t MA)^{-1} A^t M\right\} A$$

$$= -M\left\{I_n - A(A^t MA)^{-1} A^t M\right\}\frac{\partial A}{\partial \theta_i}(A^t MA)^{-1} A^t MA - MA(A^t MA)^{-1}\frac{\partial A^t}{\partial \theta_i} M\left\{I_n - A(A^t MA)^{-1} A^t M\right\} A$$

$$= -M\left\{I_n - A(A^t MA)^{-1} A^t M\right\}\frac{\partial A}{\partial \theta_i} - MA(A^t MA)^{-1}\frac{\partial A^t}{\partial \theta_i}(MA - MA)$$

$$= -M\left\{I_n - A(A^t MA)^{-1} A^t M\right\}\frac{\partial A}{\partial \theta_i}.$$

(b) Premultiplying both sides of the equation (A.17) by A^t, we get

$$A^t M\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^t MA)^{-1} A^t M\right\}$$

$$= -A^t M\left\{I_n - A(A^t MA)^{-1} A^t M\right\}\frac{\partial A}{\partial \theta_i}(A^t MA)^{-1} A^t M - A^t MA(A^t MA)^{-1}\frac{\partial A^t}{\partial \theta_i} M\left\{I_n - A(A^t MA)^{-1} A^t M\right\}$$

$$= -\left(A^tM - A^tM\right)\frac{\partial A}{\partial \theta_i}(A^tMA)^{-1}A^tM - \frac{\partial A^t}{\partial \theta_i}M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}$$

$$= -\frac{\partial A^t}{\partial \theta_i}M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}.$$

(c) Postmultiplying both sides of (b) by A

$$A^tM\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}A = -\frac{\partial A^t}{\partial \theta_i}M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}A = -\frac{\partial A^t}{\partial \theta_i}(MA - MA) = \mathbf{0}_{\mathbf{q}\times\mathbf{q}}.$$

(d) Differentiating both sides of (a) by θ_j, we get

$$M\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\right]A + M\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_j}$$

$$= -M\frac{\partial}{\partial \theta_j}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_i} - M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial^2 A}{\partial \theta_i \partial \theta_j}.$$

Premultiplying both sides of the above equation by A^t, we get

$$A^tM\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\right]A + A^tM\frac{\partial}{\partial \theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_j}$$

$$= -A^tM\frac{\partial}{\partial \theta_j}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_i} - A^tM\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial^2 A}{\partial \theta_i \partial \theta_j}.$$

It follows that

$$A^tM\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\right]A$$

$$= \frac{\partial A^t}{\partial \theta_j}M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_i} + \frac{\partial A^t}{\partial \theta_i}M\left\{I_n - A(A^tMA)^{-1}A^tM\right\}\frac{\partial A}{\partial \theta_j}$$

where we have used (b).

Lemma A13. Let $g(\overline{Y}, N^{-1}M; A) = tr\left[\overline{Y}^tN^{-1}M\left\{\mathbf{I_n} - A(A^tN^{-1}MA)^{-1}A^tN^{-1}M\right\}\overline{Y}\right]$ and $Q = \frac{\partial vecA_0}{\partial \theta^t}$. Assume that $\left|\frac{m_i}{N} - c_i\right| = o(m^{*-1})$, $i = 1, \ldots, n$. Then

(a) $\frac{\partial g(\overline{Y}, N^{-1}M, \theta_0)}{\partial \theta} = -2Q^t\left[C\left\{I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right\} \otimes P_0\right]vec\left(\overline{Y} - A_0P_0\right)^t + o_p(m^{*-\frac{1}{2}})$.

(b) $\dfrac{\partial^2 g(\overline{Y}, N^{-1}M, \theta_0)}{\partial\theta\partial\theta^t} = Q^t\left[C\left\{I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right\} \otimes P_0P_0^t\right]Q + o_p(1).$

Proof. (a) For an element θ_i of θ,

$$\frac{\partial}{\partial\theta_i} g(\overline{Y}, N^{-1}M, \theta) = N^{-1}tr\left[(\overline{Y} - A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}(\overline{Y} - A_0P_0)\right]$$
$$+N^{-1}tr\left[(\overline{Y} - A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}A_0P_0\right]$$
$$+N^{-1}tr\left[(A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}(\overline{Y} - A_0P_0)\right]$$
$$+N^{-1}tr\left[(A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}(A_0P_0)\right]. \qquad \text{(A.18)}$$

By the property of *trace,* we have

$$N^{-1}tr\left[(\overline{Y} - A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}(\overline{Y} - A_0P_0)\right]$$
$$=\left(vec(\overline{Y} - A_0P_0)^t\right)^t\left[N^{-1}M\frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\} \otimes \mathbf{I_p}\right]vec(\overline{Y} - A_0P_0)^t$$
$$=\left\{vec(\overline{Y} - A_0P_0)^t\right\}^t\left[C\frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tCA)^{-1}A^tC\right\} \otimes \mathbf{I_p} + o(1)\right]vec(\overline{Y} - A_0P_0)^t.$$

By lemma A11 (b), $vec(\overline{Y} - A_0P_0)^t = O_p(m^{*-\frac{1}{2}})$. Since $I_n - A(A^tCA)^{-1}A^tC$ is the projection matrix, we have

$$N^{-1}tr\left[(\overline{Y} - A_0P_0)^t M \frac{\partial}{\partial\theta_i}\left\{I_n - A(A^tMA)^{-1}A^tM\right\}(\overline{Y} - A_0P_0)\right] = O_p(m^{*-1}).$$

Thus the first term of the equation (A.18) is negligible. For the remaining terms of the equation (A.18) when $\theta = \theta_0$, we get

$$N^{-1}tr\left[(\overline{Y} - A_0P_0)^t M\left\{\frac{\partial}{\partial\theta_i}\left(I_n - A(A^tMA)^{-1}A^tM\right)\Big|_{A=A_0}\right\}A_0P_0\right]$$

$$= -N^{-1} tr\left[(\bar{Y} - A_0 P_0)^t M \left\{ \left(I_n - A_0 (A_0^t M A_0)^{-1} A_0^t M \right) \frac{\partial A}{\partial \theta_i}\bigg|_{A=A_0} \right\} P_0 \right]$$

by lemma A12 (a),

$$N^{-1} tr\left[(A_0 P_0)^t M \left\{ \frac{\partial}{\partial \theta_i} \left(I_n - A(A^t M A)^{-1} A^t M \right)\bigg|_{A=A_0} \right\} (\bar{Y} - A_0 P_0) \right]$$

$$= -N^{-1} tr\left[P_0^t \left(\frac{\partial A^t}{\partial \theta_i}\bigg|_{A=A_0} \right) M \left\{ I_n - A_0 (A_0^t M A_0)^{-1} A_0^t M \right\} (\bar{Y} - A_0 P_0) \right]$$

by lemma A12 (b), and

$$N^{-1} tr\left[(A_0 P_0)^t M \left\{ \frac{\partial}{\partial \theta_i} \left(I_n - A(A^t M A)^{-1} A^t M \right)\bigg|_{A=A_0} \right\} (A_0 P_0) \right] = P_0^t \mathbf{0}_{\mathbf{q}\times\mathbf{q}} P_0 = \mathbf{0}_{\mathbf{p}\times\mathbf{p}}$$

by lemma A12 (c). It follows that

$$\frac{\partial}{\partial \theta_i} g(\bar{Y}, N^{-1}M, \theta_0) = -N^{-1} tr\left[(\bar{Y} - A_0 P_0)^t M \left\{ I_n - A_0 (A_0^t M A_0)^{-1} A_0^t M \right\} \left(\frac{\partial A}{\partial \theta_i}\bigg|_{A=A_0} \right) P_0 \right]$$

$$-N^{-1} tr\left[P_0^t \left(\frac{\partial A}{\partial \theta_i}\bigg|_{A=A_0} \right) M \left(I_n - A_0 (A_0^t M A_0)^{-1} A_0^t M \right) (\bar{Y} - A_0 P_0) \right] + O_p(m^{*-1}). \qquad \text{(A.19)}$$

Using the property of *trace*, we get for the first term of (A.19),

$$N^{-1} tr\left[(\bar{Y} - A_0 P_0)^t M \left\{ I_n - A_0 (A_0^t M A_0)^{-1} A_0^t M \right\} \left(\frac{\partial A}{\partial \theta_i}\bigg|_{A=A_0} \right) P_0 \right]$$

$$= N^{-1} tr\left[P_0^t \left(\frac{\partial A^t}{\partial \theta_i}\bigg|_{A=A_0} \right) \left\{ I_n - M A_0 (A_0^t M A_0)^{-1} A_0 \right\} M (\bar{Y} - A_0 P_0) \right]$$

$$= N^{-1} tr\left[\left(\frac{\partial A^t}{\partial \theta_i}\bigg|_{A=A_0} \right) \left\{ I_n - M A_0 (A_0^t M A_0)^{-1} A_0 \right\} M (\bar{Y} - A_0 P_0) P_0^t \right]$$

$$= N^{-1} \left(vec \frac{\partial A^t}{\partial \theta_i}\bigg|_{A=A_0} \right)^t \left[\left\{ I_n - M A_0 (A_0^t M A_0)^{-1} A_0 \right\} M \otimes P_0 \right] vec(\bar{Y} - A_0 P_0)^t,$$

and for the second term of (A.19),

$$N^{-1}tr\left[P_0^t\left(\frac{\partial A}{\partial\theta_i}\bigg|_{A=A_0}\right)M\left\{I_n - A_0(A_0^tMA_0)^{-1}A_0^tM\right\}(\bar{Y}-A_0P_0)\right]$$

$$= N^{-1}tr\left[\left(\frac{\partial A}{\partial\theta_i}\bigg|_{A=A_0}\right)M\left\{I_n - A_0(A_0^tMA_0)^{-1}A_0^tM\right\}(\bar{Y}-A_0P_0)P_0^t\right]$$

$$= N^{-1}\left(vec\frac{\partial A^t}{\partial\theta_i}\bigg|_{A=A_0}\right)^t\left[M\left\{I_n - A_0(A_0^tMA_0)^{-1}A_0^tM\right\}\otimes P_0\right]vec(\bar{Y}-A_0P_0)^t.$$

Thus,

$$\frac{\partial}{\partial\theta_i}g(\bar{Y},N^{-1}M,\theta_0)$$

$$= -2N^{-1}\left(vec\frac{\partial A^t}{\partial\theta_i}\bigg|_{A=A_0}\right)^t\left[M\left(I_n - A_0(A_0^tMA_0)^{-1}A_0^tM\right)\otimes P_0\right]vec(\bar{Y}-A_0P_0)^t + O_p(m^{*-1})$$

$$= -2\left(vec\frac{\partial A^t}{\partial\theta_i}\bigg|_{A=A_0}\right)^t\left[C\left(I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right)\otimes P_0\right]vec(\bar{Y}-A_0P_0)^t + o_p(m^{*-\frac{1}{2}})$$

by the assumption that $\left|\frac{m_i}{N}-c_i\right| = o(m^{*-1}),\ i=1,\ldots,n$.

It follows that

$$\frac{\partial g(\bar{Y},N^{-1}M,\theta_0)}{\partial\theta} = -2Q^t\left[C\left(I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right)\otimes P_0\right]vec\left(\bar{Y}-A_0P_0\right)^t + o_p(m^{*-\frac{1}{2}})$$

where $Q = \frac{\partial vecA_0^t}{\partial\theta^t}$.

(b) For elements θ_i and θ_j of θ,

$$\frac{\partial^2 g(\bar{Y},N^{-1}M,\theta)}{\partial\theta_i\partial\theta_j} = N^{-1}tr\left\{\left(\bar{Y}-A_0P_0\right)^tM\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(I_n - A(A^tMA)^{-1}A^tM\right)\left(\bar{Y}-A_0P_0\right)\right\}$$

$$+ N^{-1}tr\left\{\left(\bar{Y}-A_0P_0\right)^tM\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(I_n - A(A^tMA)^{-1}A^tM\right)A_0P_0\right\}$$

$$+ N^{-1}tr\left\{\left(A_0P_0\right)^tM\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(I_n - A(A^tMA)^{-1}A^tM\right)\left(\bar{Y}-A_0P_0\right)\right\}$$

$$+ N^{-1} tr\left\{ \left(A_0 P_0\right)^t M \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(I_n - A(A^t M A)^{-1} A^t M\right)\left(A_0 P_0\right) \right\}. \qquad \text{(A.20)}$$

Note that $\frac{\partial^2}{\partial\theta_i \partial\theta_j}\left\{I_n - A(A^t C A)^{-1} A^t C\right\}$ is bounded over Θ since $I_n - A(A^t C A)^{-1} A^t C$ is the projection matrix. We have for the first term of the equation (A.20),

$$N^{-1} tr\left[\left(\overline{Y} - A_0 P_0\right)^t M \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t M A)^{-1} A^t M\right\}\left(\overline{Y} - A_0 P_0\right) \right]$$

$$= tr\left[\left(\overline{Y} - A_0 P_0\right)^t C \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t C A)^{-1} A^t C\right\}\left(\overline{Y} - A_0 P_0\right) \right] + O_p(m^{*-1})$$

$$= O_p(m^{*-1}),$$

for the second term of (A.20),

$$N^{-1} tr\left[\left(\overline{Y} - A_0 P_0\right)^t M \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t M A)^{-1} A^t M\right\} A_0 P_0 \right]$$

$$= tr\left[\left(\overline{Y} - A_0 P_0\right)^t C \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t C A)^{-1} A^t C\right\} A_0 P_0 \right] + O_p(m^{*-\frac{1}{2}})$$

$$= O_p(m^{*-\frac{1}{2}}),$$

and for the third term of (A.20),

$$N^{-1} tr\left[\left(A_0 P_0\right)^t M \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t M A)^{-1} A^t M\right\}\left(\overline{Y} - A_0 P_0\right) \right] = O_p(m^{*-\frac{1}{2}}).$$

Thus

$$\frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta)}{\partial\theta_i \partial\theta_j} = N^{-1} tr\left[\left(A_0 P_0\right)^t M \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{I_n - A(A^t M A)^{-1} A^t M\right\}\left(A_0 P_0\right) \right] + O_p(m^{*-\frac{1}{2}}).$$

For $\theta = \theta_0$,

$$\frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta_0)}{\partial\theta_i \partial\theta_j} = N^{-1} tr\left[\left(A_0 P_0\right)^t M \left\{ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left(I_n - A(A^t M A)^{-1} A^t M\right)\Big|_{A=A_0} \right\} \left(A_0 P_0\right) \right] + O_p(m^{*-\frac{1}{2}})$$

$$= tr\left[P_0^t \frac{\partial A_0^t}{\partial \theta_j} C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \frac{\partial A_0}{\partial \theta_i} P_0 + P_0^t \frac{\partial A_0^t}{\partial \theta_i} C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \frac{\partial A_0}{\partial \theta_j} P_0 \right] + o_p(1)$$

$$= tr\left[\frac{\partial A_0^t}{\partial \theta_j} C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \frac{\partial A_0}{\partial \theta_i} P_0 P_0^t \right] + tr\left[\frac{\partial A_0^t}{\partial \theta_i} C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \frac{\partial A_0}{\partial \theta_j} P_0 P_0^t \right] + o_p(1)$$

$$= \left(vec \frac{\partial A_0^t}{\partial \theta_j} \right)^t \left[C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \otimes P_0 P_0^t \right] vec \frac{\partial A_0^t}{\partial \theta_i}$$

$$+ \left(vec \frac{\partial A_0^t}{\partial \theta_i} \right)^t \left[C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \otimes P_0 P_0^t \right] vec \frac{\partial A_0^t}{\partial \theta_j} + o_p(1)$$

where we have used lemma A12 (d) and the properties of *trace.* It follows that

$$\frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta_0)}{\partial \theta \partial \theta^t} = 2\, Q^t \left[C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \otimes P_0 P_0^t \right] Q + o_p(1).$$

Proof of proposition. By the same argument as that in theorem 2, it can be shown that $\hat{\theta} \in Int(\Theta)$ with probability approaching one as m^* increases. The rest of proof is based on Taylor's theorem and the asymptotic normality of $\overline{Y}$.

The same argument in the proof of theorem 4.B.2 in Fuller (1987) can be used for our case too. We only need to calculate the first and second derivatives of the objective function $g(\overline{Y}, N^{-1}M; \theta)$ with respect to θ since the objective function which we minimize is different from that of Fuller. From (4.B.19) in Fuller (1987), we obtain with probability approaching one as $m^* \to \infty$,

$$\hat{\theta} - \theta_0 = -\left\{ \frac{\partial^2 g(\overline{Y}, N^{-1}M; \theta^*)}{\partial \theta \partial \theta^t} \right\}^{-1} \frac{\partial g(\overline{Y}, N^{-1}M; \theta_0)}{\partial \theta}$$

where the elements of θ^* are evaluated at points on the line segment joining θ_0 and $\hat{\theta}$.

By lemma A13 (a),

$$\frac{\partial g(\overline{Y}, N^{-1}M, \theta_0)}{\partial \theta} = -2Q^t \left[C\left\{ I_n - A_0 (A_0^t C A_0)^{-1} A_0^t C \right\} \otimes P_0 \right] vec\left(\overline{Y} - A_0 P_0 \right)^t + o_p(m^{*-\frac{1}{2}}).$$

By lemma A13 (b),

$$\frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta_0)}{\partial\theta\partial\theta^t} = 2\,Q^t\left[C\left\{I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right\} \otimes P_0P_0^t\right]Q + o_p(1).$$

By proposition 2,

$$\hat{\theta} \xrightarrow{p} \theta_0 \text{ as } m^* \to \infty,$$

and so θ^* also converges to θ_0 since θ^* is between $\hat{\theta}$ and θ_0. Because the partial derivatives are continuous,

$$\frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta^*)}{\partial\theta\partial\theta^t} = \frac{\partial^2 g(\overline{Y}, N^{-1}M, \theta_0)}{\partial\theta\partial\theta^t} + o_p(1).$$

Letting $\mathbf{B} = Q^t\left[C\left\{I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right\} \otimes P_0\right]$ and

$\mathbf{H} = Q^t\left[C\left\{I_n - A_0(A_0^tCA_0)^{-1}A_0^tC\right\} \otimes P_0P_0^t\right]Q$, we obtain that

$$\hat{\theta} - \theta_0 = \mathbf{H}^{-1}\mathbf{B}vec\left(\overline{Y} - A_0P_0\right)^t + o_p(m^{*-\frac{1}{2}}). \qquad \text{(A.21)}$$

It follows from (A.21), lemma A11 (b), and the continuous mapping theorem that

$$\sqrt{m^*}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}^{-1}\mathbf{B}(C^* \otimes \Sigma_p)\mathbf{B}^t\mathbf{H}^{-1}\right).$$