Determining the Dimensionality in Sliced Inverse Regression

James R. SCHOTT*

A general regression problem is one in which a response variable can be expressed as some function of one or more different linear combinations of a set of explanatory variables as well as a random error term. Sliced inverse regression is a method for determining these linear combinations. In this article we address the problem of determining how many linear combinations are involved. Procedures based on conditional means and conditional covariance matrices, as well as a procedure combining the two approaches, are considered. In each case we develop a test that has an asymptotic chi-squared distribution when the vector of explanatory variables is sampled from an elliptically symmetric distribution.

KEY WORDS: Eigenprojection; Elliptically symmetric distribution; General regression model; Projection matrix.

1. INTRODUCTION

This article deals with one aspect of a recently proposed analysis of a fairly general regression model involving a response variable y and a p-dimensional vector of predictors, $\mathbf{x}$. For an unspecified function, f, the model states that

$$y = f(\beta_1'\mathbf{x}, \ldots, \beta_k'\mathbf{x}, \varepsilon), \tag{1}$$

where $\beta_1, \ldots, \beta_k$ are unknown linearly independent vectors and ε is statistically independent of $\mathbf{x}$. Thus when $k < p$, y depends on $\mathbf{x}$ only through the k variables $x_1^* = \beta_1'\mathbf{x}, \ldots, x_k^* = \beta_k'\mathbf{x}$, so a reduction in dimensionality is possible.

Li (1991a) suggested a clever way of obtaining the β vectors. His method involves the use of inverse regression in which $\mathbf{x}$ is regressed on y. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean and covariance matrix of $\mathbf{x}$ so that $\mathbf{x}$ is standardized via the transformation $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$, where $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \mathbf{I}$. Then the standardized inverse regression curve $E(\mathbf{z}\,|\,y)$ will be contained in the standardized effective dimension-reduction (EDR) space (i.e., the space spanned by $\boldsymbol{\Sigma}^{1/2}\beta_1, \ldots, \boldsymbol{\Sigma}^{1/2}\beta_k$), if Li's condition (3.1) on the distribution of $\mathbf{x}$ is satisfied. This condition states that the conditional expectation $E(\mathbf{b}'\mathbf{x}\,|\,\beta_1'\mathbf{x}, \ldots, \beta_k'\mathbf{x})$ is linear in $\beta_1'\mathbf{x}, \ldots, \beta_k'\mathbf{x}$ for any p-dimensional vector, $\mathbf{b}$.

Some information about this space can be obtained from a sample by slicing the data according to the y values. For instance, if we have $N = nh$ observations $(y_j, \mathbf{x}_j)$ $(j = 1, \ldots, N)$, ordered so that $y_j \leq y_{j+1}$ for all j, then the mean of the ith slice $(i = 1, \ldots, h)$ would be given by $\bar{\mathbf{x}}_i = n^{-1} \sum_{j=1}^{n} \mathbf{x}_{(i-1)n+j}$. If $\mathbf{S} = \mathbf{S}^{1/2}\mathbf{S}^{1/2}$ is the sample covariance matrix for $\mathbf{x}$ and $\bar{\mathbf{x}}$ is the grand $\mathbf{x}$ mean, we then would perform an eigenanalysis of $\hat{V}_1 = S^{-1/2}\hat{\boldsymbol{\Delta}}\mathbf{S}^{-1/2}$, where $\hat{\boldsymbol{\Delta}} = \sum_{i=1}^{h} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'/h$. The latent vectors, $\hat{\boldsymbol{\eta}}_1, \ldots, \hat{\boldsymbol{\eta}}_k$, of $\hat{\mathbf{V}}_1$ corresponding to its k largest latent roots can then be used to estimate the β's via the equations $\hat{\beta}_i = \mathbf{S}^{-1/2}\hat{\boldsymbol{\eta}}_i$ $(i = 1, \ldots, k)$. More information concerning these estimates and the performance of the sliced inverse procedure can be found in Li's article.

Our interest here concerns k, the number of β vectors or, equivalently, the dimension of the ERD space. Obviously, in most applications k will be unknown and hence must be estimated from the data. One approach is to use the latent roots, $\hat{\theta}_1 \geq \cdots \geq \hat{\theta}_p$, of $\hat{\mathbf{V}}_1$. For instance, Li has shown that $N \sum_{i=k+1}^{p} \hat{\theta}_i$ is asymptotically chi-squared with $(p - k)(h - k - 1)$ degrees of freedom if $\mathbf{x}$ has a normal distribution. Thus we can estimate k by using this statistic repeatedly, starting with $k = 0$, and continually increasing k by 1 until the statistic does not exceed the $1 - \alpha$ quantile of the appropriate chi-squared distribution. This last value of k will then be our estimate.

The aforementioned procedure is valid when $\mathbf{x}$ is normal. Our purpose here is to develop a procedure with broader application. A procedure valid for any distribution satisfying Li's condition (3.1) would be ideal, but this appears to present a formidable task. As a result, in this article we will restrict attention to those situations in which $\mathbf{x}$ has an elliptically symmetric distribution. This differs from Li's condition (3.1), because there are nonelliptical distributions satisfying this condition. In this setting we develop an asymptotically chi-squared test for the dimensionality of the inverse regression curve.

As noted by Li and expanded on by Cook and Weisberg (1991) and in the rejoinder of Li (1991b), in some cases the inverse regression curve may fall within a proper subspace of the EDR space; that is, the sliced inverse regression procedure will be doomed to miss some of the β vectors. By looking at higher conditional moments of $\mathbf{x}$ given y in addition to $E(\mathbf{x}\,|\,y)$, we can reduce the chances of missing important directions. For instance, Cook and Weisberg suggested the eigenanalysis of a matrix formed from conditional covariance matrices, and Li (1991b) discussed the idea of combining first- and second-order moment matrices. In this article asymptotic tests of dimensionality are also obtained for some variations along these lines.

As discussed in detail by Kent (1991), after slicing the data one is essentially faced with a discriminant analysis problem involving h groups. In fact the methods developed in this article are motivated by some recent work in discriminant analysis (Schott, 1993). This work involved the determination of the dimensionality necessary to discriminate among h heterogeneous normal populations. In that case, because normal populations were involved, an investigation

* James R. Schott is Professor, Department of Statistics, University of Central Florida, Orlando, FL 32816-2370. The author thanks the associate editor and referees for comments that helped improve the presentation.

Journal of the American Statistical Association
March 1994, Vol. 89, No. 425, Theory and Methods

of differences in group means and group covariance matrices reveals all useful directions for discrimination. What we develop in this article essentially represents some generalizations of these results to elliptically symmetric populations.

2. DIMENSIONALITY MATRICES AND PROJECTION SPACES

Let the number of slices, h, be fixed and suppose that our slices will be chosen so that they contain the same (or nearly the same) number of observations. Let r_i ($i = 1, \ldots, h-1$) be the (i/h)th quantile of the distribution of y so that our h slices through y can be written as $I_1 = \{y: y \le r_1\}$, $I_2 = \{y: r_1 < y \le r_2\}, \ldots, I_h = \{y: y > r_{h-1}\}$. The conditional means and covariance matrices for $\mathbf{x}$ will then be given by $\boldsymbol{\mu}_i = E(\mathbf{x}|y \in I_i)$ and $\boldsymbol{\Omega}_i = \operatorname{cov}(\mathbf{x}|y \in I_i)$, whereas the average conditional mean and covariance matrix will be $\boldsymbol{\mu} = E(\mathbf{x}) = \sum_{i=1}^{h} \boldsymbol{\mu}_i/h$ and $\boldsymbol{\Omega} = E\{\operatorname{cov}(\mathbf{x}|y \in I_i)\} = \sum_{i=1}^{h} \boldsymbol{\Omega}_i/h$.

If $\mathbf{x}$ has been standardized by the transformation $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2} \times (\mathbf{x} - \boldsymbol{\mu})$, where $\boldsymbol{\Sigma}^{-1/2}$ is the symmetric square root of $\boldsymbol{\Sigma}^{-1}$, then model (1) implies that there are projection matrices $\mathbf{P}_1$, $\mathbf{P}_2 = \mathbf{I} - \mathbf{P}_1$, with $\mathbf{P}_1$ of rank k such that y is conditionally independent of $\mathbf{P}_2\mathbf{z}$ given $\mathbf{P}_1\mathbf{z}$. Thus if $\mathbf{P}_1 = \boldsymbol{\eta}_1\boldsymbol{\eta}_1' + \cdots + \boldsymbol{\eta}_k\boldsymbol{\eta}_k'$, then the β vectors in (1) can be defined as $\beta_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i$. Our goal then is to determine the number of $\boldsymbol{\eta}$'s or, equivalently, the rank of $\mathbf{P}_1$. One approach is to form a dimensionality matrix whose eigenprojection associated with its positive latent roots is the same as, or possibly a subspace of, the projection space of $\mathbf{P}_1$.

If $\mathbf{x}$ is elliptically symmetric, then it follows (see Cook and Weisberg 1991, eq. 1) that for $i = 1, \ldots, h$,

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}) = \mathbf{P}_1\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}). \quad (2)$$

Further,

$$\begin{aligned}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Sigma}^{-1/2} &= \operatorname{var}(\mathbf{z}|y \in I_i)\\ &= E(\operatorname{var}(\mathbf{z}|\boldsymbol{\eta}_1'\mathbf{z}, \ldots, \boldsymbol{\eta}_k'\mathbf{z})|y \in I_i\}\\ &\quad + \operatorname{var}\{E(\mathbf{z}|\boldsymbol{\eta}_1'\mathbf{z}, \ldots, \boldsymbol{\eta}_k'\mathbf{z})|y \in I_i\}\\ &= E\{g(\boldsymbol{\eta}_1'\mathbf{z}, \ldots, \boldsymbol{\eta}_k'\mathbf{z})\mathbf{P}_2|y \in I_i\}\\ &\quad + \operatorname{var}\{\mathbf{P}_1\mathbf{z}|y \in I_i\}\\ &= \tau_i\mathbf{P}_2 + \mathbf{P}_1\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Sigma}^{-1/2}\mathbf{P}_1, \end{aligned} \quad (3)$$

where $\tau_i = E\{g(\boldsymbol{\eta}_1'\mathbf{z}, \ldots, \boldsymbol{\eta}_k'\mathbf{z})|y \in I_i\}$ and g is a function that depends on the particular elliptically symmetric distribution of $\mathbf{z}$.

Now, because $\boldsymbol{\Sigma} = \boldsymbol{\Omega} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta} = \sum_{i=1}^{h} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})'/h$, it follows that if for some vector $\boldsymbol{\beta}$, $\boldsymbol{\Delta}\boldsymbol{\beta} = \lambda\boldsymbol{\Sigma}\boldsymbol{\beta}$, then $\boldsymbol{\Delta}\boldsymbol{\beta} = \lambda(1-\lambda)^{-1}\boldsymbol{\Omega}\boldsymbol{\beta}$. Using this, we find that equations (2) and (3) are equivalent to the set of equations

$$\begin{aligned}\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}) &= \mathbf{P}_1\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu});\\ \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Omega}^{-1/2} &= \tau_i\mathbf{P}_2 + \mathbf{P}_1\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Omega}^{-1/2}\mathbf{P}_1. \end{aligned} \quad (4)$$

Note that $\mathbf{P}_1$ and $\mathbf{P}_2$ in (4) are not identical to those appearing in (2) and (3); now $\mathbf{P}_1$ and $\mathbf{P}_2$ would be expressed as $\mathbf{P}_1 = \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1' + \cdots + \boldsymbol{\gamma}_k\boldsymbol{\gamma}_k'$ and $\mathbf{P}_2 = \mathbf{I} - \mathbf{P}_1$, where $\boldsymbol{\gamma}_i$ is the standardized version of $\boldsymbol{\Omega}^{1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i$. In proceeding from here we could use equations (2) and (3) or, alternatively, equations (4); that is, we could work with the transformation $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ or the transformation $\mathbf{w} = \boldsymbol{\Omega}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$. We prefer the latter, because this approach will lead us to some statistics with simpler covariance structure than that obtained by the former. Consequently, any use of $\mathbf{P}_1$ or $\mathbf{P}_2$ in the remainder of this article refers to the quantities defined in (4).

Now it follows from (4) that if $\mathbf{P}_{11}$ is the eigenprojection matrix of $\mathbf{W}_1 = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1/2}$ associated with its k_1 positive latent roots, then $\mathbf{P}_1\mathbf{P}_{11} = \mathbf{P}_{11}$; that is, the projection space of $\mathbf{P}_{11}$ is a subspace of that of $\mathbf{P}_1$. Thus if $\boldsymbol{\Gamma}_{11}(p \times k_1)$ is such that $\mathbf{P}_{11} = \boldsymbol{\Gamma}_{11}\boldsymbol{\Gamma}_{11}'$, then the columns of $\mathbf{B}_1^* = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Gamma}_{11}$ span a subspace of the ERD space. If $k_1 < k$, then the columns of $\mathbf{B}_1^*$ span a proper subspace of the ERD space, and so we have missed some important directions. In an attempt to remedy this Cook and Weisberg (1991) and Li (1991b) suggested some second-moment dimensionality matrices. For instance, because $\bar{\tau} = \sum_{i=1}^{h} \tau_i/h = 1$, it follows from (4) that the matrix $\mathbf{W}_2^* = \sum_{i=1}^{h} (\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Omega}^{-1/2} - \mathbf{I})^2/h$ has projection matrix $\mathbf{P}_{22}$ associated with the multiple root $\sum_{i=1}^{h} (\tau_i - 1)^2/h$, which is repeated, say, $p - k_2$ times. In most cases this common root will be small (see Li 1990), and so if $\mathbf{P}_{21} = \mathbf{I} - \mathbf{P}_{22}$ is the eigenprojection matrix of $\mathbf{W}_2^*$ associated with its k_2 dominant roots, then $\mathbf{P}_1\mathbf{P}_{21} = \mathbf{P}_{21}$, and thus the projection space of $\mathbf{P}_{21}$ is also a subspace of that of $\mathbf{P}_1$. A variation of $\mathbf{W}_2^*$ with the common latent root equal to 0 can be obtained by noting from (4) that we must have $\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Omega}_i - \tau_i\boldsymbol{\Omega})\boldsymbol{\Omega}^{-1/2} = \mathbf{P}_1\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\Omega}_i - \tau_i\boldsymbol{\Omega})\boldsymbol{\Omega}^{-1/2}\mathbf{P}_1$. Consequently, the eigenprojection of

$$\mathbf{W}_2 = \sum_{i=1}^{h} (\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}_i\boldsymbol{\Omega}^{-1/2} - \tau_i\mathbf{I})^2/h$$

corresponding to its k_2 positive latent roots will also be $\mathbf{P}_{21}$.

As suggested by Li in his rejoinder (Li 1991b), instead of analyzing the first and second moments separately, we may combine them in some way into one analysis. In this article we will consider the analysis that simply combines $\mathbf{W}_1$ and $\mathbf{W}_2$ with equal weights; that is, our combined analysis involves the dimensionality matrix $\mathbf{W}_3 = \mathbf{W}_1 + \mathbf{W}_2$. Certainly better, more sophisticated combinations of $\mathbf{W}_1$ and $\mathbf{W}_2$ can be found, but we will not pursue that subject here. The eigenprojection matrix of $\mathbf{W}_3$ associated with its k_3 positive latent roots will be denoted by $\mathbf{P}_{31}$ and satisfies $\mathbf{P}_1\mathbf{P}_{31} = \mathbf{P}_{31}$. The primary advantage of a combined analysis over the two individual analyses is that it will properly distinguish between new and redundant directions. In other words k_1 and k_2 do not determine k_3 unless $k_1 = 0$ or $k_2 = 0$, because $\max(k_1, k_2) \le k_3 \le k_1 + k_2$.

If $k_3 < k$, then the analyses based on the first- and second-order moments will not be successful in finding all of the β_i's. One could proceed to higher moments and, in a fashion similar to that described in this section, exploit the structure of the moments to form a matrix whose eigenprojection matrix associated with its positive roots must be a function of the β's. Unfortunately, these matrices become increasingly complicated as the order increases, and so the inferential procedures developed in this article would be difficult to extend to these higher moments.

3. DETERMINING THE DIMENSION

We will develop a procedure for determining k_3. The process involves the sequence of tests

$$H^i_{0m}\colon k_i = m \quad \text{versus} \quad H^i_{1m}\colon k_i > m$$

for $i = 3$. The distribution of the test criteria that we construct for testing these hypotheses depends on k_2, so that before determining k_3 we will first have to determine k_2. Thus we would test H^2_{0m} versus H^2_{1m} starting with $m = 0$ and continuing until a null hypothesis, say $H^2_{0m^*}$, is not rejected; in this case we would take $k_2 = m^*$ and begin testing H^3_{0m} versus H^3_{1m} starting with $m = m^*$ and continuing until a null hypothesis is not rejected.

Let $\hat{\mathbf{\Omega}}_i$ be the sample covariance matrix computed from the ith slice and define $\hat{\mathbf{\Omega}} = (\hat{\mathbf{\Omega}}_1 + \cdots + \hat{\mathbf{\Omega}}_h)/h$. Then we define estimators of $\mathbf{W}_i (i = 1, 2, 3)$ as $\hat{\mathbf{W}}_1 = \hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Delta}}\hat{\mathbf{\Omega}}^{-1/2}$,

$$\hat{\mathbf{W}}_2 = \sum_{i=1}^{h} (\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Omega}}_i\hat{\mathbf{\Omega}}^{-1/2} - \hat{\tau}_i\mathbf{I})^2/h, \tag{5}$$

and $\hat{\mathbf{W}}_3 = \hat{\mathbf{W}}_1 + \hat{\mathbf{W}}_2$, where $\hat{\tau}_i$ is an estimator of τ_i to be specified later. The latent vectors of $\mathbf{W}_i$ and $\hat{\mathbf{W}}_i$ corresponding to roots in descending order will be denoted by $\boldsymbol{\gamma}_{i1}, \ldots, \boldsymbol{\gamma}_{ip}$ and $\hat{\boldsymbol{\gamma}}_{i1}, \ldots, \hat{\boldsymbol{\gamma}}_{ip}$; projection matrices and their estimators can then be expressed as $\mathbf{P}_{ij} = \mathbf{\Gamma}_{ij}\mathbf{\Gamma}'_{ij}$ and $\hat{\mathbf{P}}_{ij} = \hat{\mathbf{\Gamma}}_{ij}\hat{\mathbf{\Gamma}}'_{ij}$ $(j = 1, 2)$, where $\mathbf{\Gamma}_{i1} = (\boldsymbol{\gamma}_{i1}, \ldots, \boldsymbol{\gamma}_{ik_i})$ and $\mathbf{\Gamma}_{i2} = (\boldsymbol{\gamma}_{ik_i+1}, \ldots, \boldsymbol{\gamma}_{ip})$ and similarly for $\hat{\mathbf{\Gamma}}_{i1}$ and $\hat{\mathbf{\Gamma}}_{i2}$. The latent roots of $\mathbf{W}_i$ will be denoted by $\lambda_{i1} \geq \cdots \geq \lambda_{ip}$. One natural choice of a statistic for testing H^i_{0m} would be the average of the $p - m$ smallest latent roots of $\hat{\mathbf{W}}_i$. This is the statistic suggested by Li for testing H^1_{0m} when the distribution of $\mathbf{x}$ is normal. Unfortunately, this statistic, as well as its generalizations for testing H^2_{0m} and H^3_{0m}, does not have a scalar multiple of a chi-squared distribution as its asymptotic null distribution for general elliptically symmetric distributions. Instead these asymptotic null distributions will generally be that of linear combinations of independent chi-squared random variables, and so practical use of these would require an approximation such as $c\chi^2_d$, where c and d are computed by fitting moments. We prefer to develop alternative statistics that have known and tabulated asymptotic null distributions. Our approach involves the construction of Wald-type statistics; that is, we begin with an asymptotically normal estimator that has an asymptotic zero mean vector if and only if the appropriate null hypothesis is true. Thus by using an estimate of the asymptotic covariance matrix of this estimator, we can construct a statistic that has, asymptotically, a central chi-squared distribution if and only if the null hypothesis is satisfied.

We begin by providing a test of $H^2_{0m}\colon k_2 = m$ versus $H^2_{1m}\colon k_2 > m$. The construction of our statistic is motivated by a basic identity involving the projection matrix $\mathbf{P}_{22} = \mathbf{I} - \mathbf{P}_{21}$ associated with the $p - m$ zero latent roots of $\mathbf{W}_2$. It follows from (4) and the construction of $\mathbf{W}_2$ that

$$\mathbf{P}_{22}\mathbf{\Omega}^{-1/2}(\mathbf{\Omega}_i - \tau_i\mathbf{\Omega})\mathbf{\Omega}^{-1/2}\mathbf{P}_{22} = (0). \tag{6}$$

The sample version of the left side of (6) multiplied by $n^{1/2}$ is

$$\mathbf{B}_i = \mathbf{B}_i(\hat{\mathbf{W}}_2) = n^{1/2}\hat{\mathbf{P}}_{22}\hat{\mathbf{\Omega}}^{-1/2}(\hat{\mathbf{\Omega}}_i - \hat{\tau}_i\hat{\mathbf{\Omega}})\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{P}}_{22},$$

where $\hat{\mathbf{P}}_{22} = \hat{\mathbf{\Gamma}}_{22}\hat{\mathbf{\Gamma}}'_{22}$ is the eigenprojection of $\hat{\mathbf{W}}_2$ corresponding to its $p - m$ smallest latent roots. Now an estimate, $\hat{\tau}_i$, of τ_i can be obtained by making use of the fact that $\mathbf{P}_{22}\mathbf{\Omega}^{-1/2}\mathbf{\Omega}_i\mathbf{\Omega}^{-1/2}\mathbf{P}_{22} = \tau_i\mathbf{P}_{22}$, from which we get $\tau_i = \operatorname{tr}(\mathbf{P}_{22}\mathbf{\Omega}^{-1/2}\mathbf{\Omega}_i\mathbf{\Omega}^{-1/2}\mathbf{P}_{22})/(p - m)$. For estimating $\mathbf{P}_{22}$ in this equation, note that we cannot use $\hat{\mathbf{P}}_{22}$ because this comes from $\hat{W}_2$, which depends on $\hat{\tau}_i$. Thus we need some other preliminary estimate of $\mathbf{P}_{22}$ that does not depend on $\hat{\tau}_i$. We will use $\hat{\mathbf{P}}^*_{22}$, the eigenprojection of $\mathbf{W}^*_2$ corresponding to its $p - m$ smallest latent roots. Thus

$$\hat{\tau}_i = \hat{\tau}_i(\hat{\mathbf{W}}^*_2) = \operatorname{tr}(\hat{\mathbf{P}}^*_{22}\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Omega}}_i\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{P}}^*_{22})/(p - m) \tag{7}$$

is used in the equation for $\mathbf{B}_i$ as well as in (5).

To form a Wald statistic based on the $\mathbf{B}_i$'s, we need the variances and covariances of these $\mathbf{B}_i$'s. These and other details are presented in the Appendix; the resulting test statistic can be summarized as follows.

Theorem 1. Let $\mathbf{Z}_i = \mathbf{Z}_i(\hat{\mathbf{W}}_2) = n^{1/2}\hat{\mathbf{\Gamma}}'_{22}\hat{\mathbf{\Omega}}^{-1/2}(\hat{\mathbf{\Omega}}_i - \hat{\tau}_i\hat{\mathbf{\Omega}})\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Gamma}}_{22}$ and let $\hat{g}^{ij}_*$ be the (i, j)th element of the matrix $\hat{\mathbf{G}}^+_* = \{\hat{\mathbf{G}}_*(\hat{\mathbf{W}}_2)\}^+$ specified in Appendix Section A.1. Then if H^2_{0m} is true, the statistic

$$T^{(2)}_{2,m} = \frac{1}{2}\sum_{i=1}^{h}\sum_{j=1}^{h} \hat{g}^{ij}_*\operatorname{tr}(\mathbf{B}_i\mathbf{B}_j) = \frac{1}{2}\sum_{i=1}^{h}\sum_{j=1}^{h} \hat{g}^{ij}_*\operatorname{tr}(\mathbf{Z}_i\mathbf{Z}_j)$$

has an asymptotic chi-squared distribution with degrees of freedom $\nu_2 = (h - 1)\{(p - m)(p - m + 1)/2 - 1\}$.

Next, in a similar fashion we develop a test of $H^3_{0m}\colon k_3 = m$ versus $H^3_{1m}\colon k_3 > m$. If H^3_{0m} is true and $\mathbf{P}_{32} = \mathbf{I} - \mathbf{P}_{31}$ is the eigenprojection corresponding to the $p - m$ zero latent roots of $\mathbf{W}_3$, then it follows, from (4) and the construction of $\mathbf{W}_3$, that for each i

$$\mathbf{P}_{32}\mathbf{\Omega}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}) = \mathbf{0},\ \mathbf{P}_{32}\mathbf{\Omega}^{-1/2}(\mathbf{\Omega}_i - \tau_i\mathbf{\Omega})\mathbf{\Omega}^{-1/2}\mathbf{P}_{32} = (0).$$

We consider the corresponding sample statistics

$$\mathbf{a}_i = \mathbf{a}_i(\hat{\mathbf{W}}_3) = n^{1/2}\hat{\mathbf{P}}_{32}\hat{\mathbf{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}),$$

$$\mathbf{B}_i = \mathbf{B}_i(\hat{\mathbf{W}}_3) = n^{1/2}\hat{\mathbf{P}}_{32}\hat{\mathbf{\Omega}}^{-1/2}(\hat{\mathbf{\Omega}}_i - \hat{\tau}_i\hat{\mathbf{\Omega}})\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{P}}_{32},$$

where $\hat{\mathbf{P}}_{32}$ is the eigenprojection of $\hat{\mathbf{W}}_3$ corresponding to its $p - m$ smallest latent roots. The estimator $\hat{\tau}_i = \hat{\tau}_i(\hat{\mathbf{W}}^*_3)$ can be computed by using (7) after replacing $\hat{\mathbf{P}}^*_{22}$ by $\hat{\mathbf{P}}^*_{32}$, the eigenprojection of $\hat{\mathbf{W}}^*_3 = \hat{\mathbf{W}}_1 + \hat{\mathbf{W}}^*_2$ corresponding to its $p - m$ smallest latent roots. The following Wald statistic based on the $\mathbf{a}_i$'s and $\mathbf{B}_i$'s can then be used in a test of H^3_{0m}.

Theorem 2. Let $\mathbf{y}_i = \mathbf{y}_i(\hat{\mathbf{W}}_3) = \hat{\mathbf{\Gamma}}'_{32}\hat{\mathbf{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$ and let $\hat{c}^{ij}$ be the (i, j)th element of the matrix $\hat{\mathbf{C}}^+ = \{\hat{\mathbf{C}}(\hat{\mathbf{W}}_3)\}^+$ given in Appendix Section A.4. Also let $\mathbf{Z}_i = \mathbf{Z}_i(\hat{\mathbf{W}}_3) = n^{1/2}\hat{\mathbf{\Gamma}}'_{32}\hat{\mathbf{\Omega}}^{-1/2}(\hat{\mathbf{\Omega}}_i - \hat{\tau}_i\hat{\mathbf{\Omega}})\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Gamma}}_{32}$ and let $\hat{g}^{ij}_*$ be the (i, j)th element of the matrix $\hat{\mathbf{G}}^+_* = \{\hat{\mathbf{G}}_*(\hat{\mathbf{W}}_3)\}^+$, which can also be found in the Appendix. Define the statistics

$$T^{(3)}_{1,m} = \sum_{i=1}^{h}\sum_{j=1}^{h} \hat{c}^{ij}\mathbf{y}'_i\mathbf{y}_j, \qquad T^{(3)}_{2,m} = \frac{1}{2}\sum_{i=1}^{h}\sum_{j=1}^{h} \hat{g}^{ij}_*\operatorname{tr}(\mathbf{Z}_i\mathbf{Z}_j).$$

Then under $H^{(3)}_{0m}$, $T^{(3)}_{3,m} = T^{(3)}_{1,m} + T^{(3)}_{2,m}$ is asymptotically chi-

Table 1. Estimated Significance Levels When α = .05 and h = 5

Hypothesis tested	n	$p = 5$						$p = 10$					
		N_5	$t_{5,10}$	$t_{5,5}$	N_5	$t_{5,10}$	$t_{5,5}$	N_{10}	$t_{10,10}$	$t_{10,5}$	N_{10}	$t_{10,10}$	$t_{10,5}$
		Model 1			Model 3			Model 1			Model 3		
H_{0k}^1	10	.062	.062	.057	.016	.016	.012	.036	.049	.050	.008	.003	.004
	20	.047	.072	.065	.030	.042	.031	.057	.048	.043	.014	.024	.015
	40	.062	.055	.053	.042	.052	.031	.052	.045	.043	.041	.032	.029
	60	.042	.047	.041	.048	.045	.039	.054	.048	.049	.051	.028	.035
	80	.061	.048	.050	.043	.048	.051	.052	.064	.052	.049	.039	.044
	100	.057	.048	.044	.040	.043	.048	.043	.040	.049	.045	.038	.050
		Model 2			Model 4			Model 2			Model 4		
H_{0k}^2	10	.095	.111	.145	.066	.070	.122	.179	.144	.110	.116	.125	.152
	20	.079	.062	.093	.049	.066	.060	.085	.064	.092	.076	.073	.081
	40	.062	.043	.077	.058	.062	.071	.061	.058	.066	.069	.066	.068
	60	.058	.061	.070	.051	.054	.062	.062	.059	.088	.060	.075	.060
	80	.049	.054	.049	.055	.052	.051	.057	.035	.067	.066	.066	.058
	100	.052	.060	.070	.042	.053	.059	.060	.045	.080	.052	.053	.067
		Model 2			Model 4			Model 2			Model 4		
H_{0k}^3	10	.378	.353	.485	.243	.236	.310	.856	.857	.854	.754	.770	.828
	20	.149	.140	.185	.109	.120	.157	.375	.314	.374	.281	.294	.358
	40	.078	.077	.119	.079	.090	.083	.149	.119	.164	.128	.149	.167
	60	.072	.072	.102	.082	.070	.077	.128	.088	.119	.120	.091	.109
	80	.064	.083	.101	.066	.074	.069	.096	.063	.105	.096	.061	.107
	100	.070	.057	.068	.070	.062	.066	.088	.075	.095	.091	.078	.083

squared with $v_1 + v_2$ degrees of freedom, where $v_1 = (p - m)(h - 1 - m + k_2)$ and v_2 is as defined previously.

If one desires a separate test for k_1, the statistic $T_{1,m}^{(3)}$ can be adapted for this purpose. If $k_1 = m$, then it follows that $\mathbf{P}_{12}\boldsymbol{\Omega}^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}) = \mathbf{0}$, and thus a Wald test can be developed from the $\mathbf{a}_i$'s, where now $\mathbf{a}_i = \mathbf{a}_i(\hat{\mathbf{W}}_1) = n^{1/2}\hat{\mathbf{P}}_{12}\hat{\boldsymbol{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$. The following corollary, which follows from the proof of Theorem 2, summarizes the result.

Corollary 1. Let $\mathbf{y}_i = \mathbf{y}_i(\hat{\mathbf{W}}_1) = \hat{\boldsymbol{\Gamma}}'_{12}\hat{\boldsymbol{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$ and let $\hat{c}_1^{ij}$ be the (i, j)th element of the matrix $\hat{\mathbf{C}}_1^+ = \{\hat{\mathbf{C}}_1(\hat{\mathbf{W}}_1)\}^+$ specified in Appendix Section A.4. Then if H_{0m}^1 is true, the statistic

$$T_{1,m}^{(1)} = \sum_{i=1}^{h} \sum_{j=1}^{h} \hat{c}_1^{ij}\mathbf{y}'_i\mathbf{y}_j$$

has an asymptotic chi-squared distribution with $(p - m)(h - 1 - m)$ degrees of freedom.

When $\mathbf{x}$ is normally distributed, the problem of determining the dimensionality of sliced inverse regression simplifies somewhat. First, because the normal distribution is characterized by its first two moments, we are guaranteed that $\mathbf{W}_1$ and $\mathbf{W}_2$ will reveal all useful directions; that is, k_3 must be the same as k. In addition, there is substantial simplification of the test statistics as a result of the fact that, under normality, $\tau = 1$ and the covariance structures of $\mathbf{a}_i$ and $\mathbf{B}_i$ simplify. For example, the matrix $\mathbf{G}$ in Appendix Section A.1 reduces to $h\mathbf{I} - \mathbf{1}'$, where $\mathbf{1}$ denotes the vector that has each element equal to 1. More details on these simplified statistics can be found in Schott (1993).

4. SIMULATION RESULTS

Some simulation results were obtained to get some idea of the sample sizes necessary for adequate performance of the chi-squared distributions in approximating the null distributions of the statistics developed in the previous section. For each case considered the actual significance level was estimated from 1,000 independent simulations when the

Table 2. Estimated Significance Levels for Testing H_{0k}^3 When α = .05 and $h = k + 2$

n	$p = 5$						$p = 10$					
	N_5	$t_{5,10}$	$t_{5,5}$	N_5	$t_{5,10}$	$t_{5,5}$	N_{10}	$t_{10,10}$	$t_{10,5}$	N_{10}	$t_{10,10}$	$t_{10,5}$
	Model 2			Model 4			Model 2			Model 4		
10	.258	.216	.277	.190	.171	.244	.597	.561	.478	.619	.609	.673
20	.127	.108	.133	.102	.101	.138	.216	.182	.173	.195	.220	.255
40	.088	.078	.087	.066	.066	.088	.119	.094	.085	.107	.115	.121
60	.065	.065	.079	.071	.086	.092	.090	.081	.064	.085	.091	.101
80	.052	.054	.061	.052	.058	.063	.077	.075	.075	.077	.079	.102
100	.064	.060	.070	.066	.058	.058	.070	.067	.056	.072	.082	.085

Table 3. Estimated Significance Levels for Testing H_{02}^{1} on Model 5 When α = .05 and h = 5

	p = 5			p = 10		
n	N_5	$t_{5,10}$	$t_{5,5}$	N_{10}	$N_{10,10}$	$t_{10,5}$
10	.002	.001	.000	.001	.000	.001
20	.000	.002	.003	.002	.000	.001
40	.005	.005	.002	.002	.002	.001
60	.013	.011	.005	.003	.005	.004
80	.005	.015	.012	.010	.003	.004
100	.019	.011	.011	.014	.010	.010

nominal significance level was .05. Thus if the actual significance level is .05, then the estimated size would have probability .99 of being contained in the interval .05 $\pm z_{.995}\{(.05)(.95)/1{,}000\}^{1/2} = (.032, .068)$. We considered models of the form $y = f(\beta_1'\mathbf{x}, \ldots, \beta_k'\mathbf{x}) + .5\varepsilon$, where ε has the standard normal distribution and is independently distributed of $\mathbf{x}$. For the function f we included models for which $k = 1$:

- Model 1: $x_1 + x_2$
- Model 2: $(x_1 + x_2)^{-2}$

and models for which $k = 2$:

- Model 3: $x_1(x_1 + x_2 + 2)$
- Model 4: $(x_1 + x_2)(x_2 + x_3 + x_4 + x_5)$
- Model 5: $x_1(x_1 + x_2 + .5)$.

The random variables $x_1, \ldots, x_p$ were generated independently from the standard normal distribution. To get some idea of the performance for nonnormal elliptical distributions, we also generated samples from the multivariate t distribution with degrees of freedom 5 and 10. This was done by dividing p independent standard normal variates each by the same quantity—the square root of the ratio of an independent chi-squared random variable to its degrees of freedom.

Table 1 reports some of the results obtained when $h = 5$ slices were used in testing H_{0k}^{i}. Here we use the notation N_p and $t_{p,\nu}$ to represent the p-variate multivariate normal distribution and the p-variate multivariate t distribution with ν degrees of freedom. The tests involving $T_{1,m}^{(1)}$ and $T_{2,m}^{(2)}$ generally yield reasonable significance levels for within-slice sample sizes as small as 20 or 40, whereas the test involving $T_{3,m}^{(3)}$ requires larger samples of 80 or 100. One way of improving the fit to the approximating null distributions is to reduce h, the number of slices, thereby increasing the within-slice sample size for fixed total sample size. For instance, in testing H_{0k}^{3} we will need at the very least $h = k + 2$ slices to be able to detect $k + 1$ directional differences in the means. Table 2 lists some estimated significance levels for testing H_{01}^{3} with model 2 and H_{02}^{3} with model 4 when this minimum number of slices was used.

It is not surprising that the procedures become very conservative when one or more directions are only weakly detected by the appropriate moment matrix. An example of this is illustrated in Table 3. Here model 5, which is very similar to model 3, has $k_1 = 2$. But the distribution of $T_{1,2}^{(1)}$ converges to the asymptotic null distribution more slowly for model 5. A comparison of the latent roots of $\mathbf{W}_1$ for the two models reveals that $\mathbf{W}_1$ is more successful at detecting the two dimensions of model 3 than those of model 5. For instance, from the simulations for $p = 5$ and the normal distribution, estimates of the two nonzero roots are approximately .53 and .16 for model 3 but only .08 and .03 for model 5.

Some simulated powers are given in Table 4, in which the hypotheses H_{0i}^{j} ($i = 0, 1$; $j = 1, 2, 3$) are tested for some

Table 4. Estimated Power When k = 2, α = .05, and h = 5

		p = 5			p = 10		
Hypothesis Tested	n	N_5	$t_{5,10}$	$t_{5,5}$	N_5	$t_{5,10}$	$t_{5,5}$
		Model 3					
H_{00}^{1}	10	.973	.925	.836	.855	.758	.675
	20	1.00	1.00	.990	1.00	.986	.954
	40	1.00	1.00	1.00	1.00	1.00	1.00
	60	1.00	1.00	1.00	1.00	1.00	1.00
	80	1.00	1.00	1.00	1.00	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00
		Model 4					
H_{00}^{2}	10	.836	.825	.580	.523	.453	.190
	20	.997	.993	.889	.900	.866	.354
	40	1.00	1.00	.996	1.00	1.00	.786
	60	1.00	1.00	.996	1.00	1.00	.940
	80	1.00	1.00	1.00	1.00	1.00	.977
	100	1.00	1.00	1.00	1.00	1.00	.992
		Model 4					
H_{00}^{3}	10	.968	.969	.899	.993	.979	.913
	20	.999	1.00	.940	.989	.985	.765
	40	1.00	1.00	.993	1.00	1.00	.874
	60	1.00	1.00	.999	1.00	1.00	.947
	80	1.00	1.00	1.00	1.00	1.00	.981
	100	1.00	1.00	.999	1.00	1.00	.990

		p = 5			p = 10		
Hypothesis tested	n	N_{10}	$t_{10,10}$	$t_{10,5}$	N_{10}	$t_{10,10}$	$t_{10,5}$
		Model 3					
H_{01}^{1}	10	.319	.306	.220	.168	.122	.106
	20	.712	.701	.485	.458	.470	.317
	40	.984	.987	.891	.913	.919	.724
	60	.997	1.00	.985	.993	.997	.918
	80	1.00	1.00	.998	1.00	1.00	.986
	100	1.00	1.00	1.00	1.00	1.00	1.00
		Model 4					
H_{01}^{2}	10	.326	.359	.294	.262	.267	.173
	20	.613	.638	.421	.346	.341	.165
	40	.944	.964	.716	.729	.723	.328
	60	.998	.999	.895	.946	.955	.514
	80	1.00	1.00	.950	.995	.997	.713
	100	1.00	1.00	.979	1.00	1.00	.842
		Model 4					
H_{01}^{3}	10	.633	.652	.654	.932	.941	.904
	20	.630	.708	.546	.708	.718	.567
	40	.936	.935	.726	.812	.846	.490
	60	.994	.997	.863	.954	.955	.611
	80	.999	1.00	.936	.996	.995	.733
	100	1.00	1.00	.965	.999	1.00	.834

models for which $k_j = 2$. These values should be used in conjunction with the significance levels given in Table 1 to properly assess the power. For example, the relatively large powers for testing H_{0i}^3 when n is small are a consequence of very inflated type I error probabilities, whereas the relatively low powers for testing H_{01}^1 when $n = 10$ and $p = 10$ are due to the overly conservative significance levels. As would be expected, because of the nonrobustness of the sample mean and covariance matrix, our statistics do not perform quite as well for the heavier-tailed t distribution with 5 degrees of freedom. Thus some potential further work in this area would be the development of more robust versions of our statistics based on, for example, M estimators of location and scatter (Maronna 1976).

APPENDIX: PROOFS

A.1 Proof of Theorem 1

We will begin by finding the asymptotic null distribution of $\mathbf{b}_i = \text{vec}(\mathbf{B}_i)$ ($i = 1, \ldots, h$), where $\text{vec}(\mathbf{B}_i)$ is the p^2-dimensional vector formed by stacking the columns of $\mathbf{B}_i$. Because $\mathbf{W}_i$ and $\hat{\mathbf{W}}_i$ are unaffected by nonsingular transformations on $\mathbf{x}$ of the form $A\mathbf{x}$, we may assume without loss of generality that $\boldsymbol{\Omega} = \mathbf{I}$. Let $\mathbf{U}_i = n^{1/2}(\hat{\boldsymbol{\Omega}}_i - \boldsymbol{\Omega}_i)$ so that $\hat{\boldsymbol{\Omega}} = \sum_{i=1}^{h} (\boldsymbol{\Omega}_i + n^{-1/2}\mathbf{U}_i)/h = \mathbf{I} + n^{-1/2}\mathbf{U}$ and $\hat{\boldsymbol{\Omega}}^{-1/2} = \mathbf{I} - \frac{1}{2}n^{-1/2}\mathbf{U} + O_p(n^{-1})$, where $\mathbf{U} = \sum_{i=1}^{h} \mathbf{U}_i/h$. Now if we write $\hat{\mathbf{W}}_2$ in the form $\hat{\mathbf{W}}_2 = \mathbf{W}_2 + n^{-1/2}\mathbf{A}_2 + O_p(n^{-1})$, then its eigenprojection $\hat{\mathbf{P}}_{22}$ can be expressed as (see, for example, Tyler 1981) $\hat{\mathbf{P}}_{22} = \mathbf{P}_{22} + n^{-1/2}\mathbf{R}_2 + O_p(n^{-1})$, where

$$\mathbf{R}_2 = -\sum_{j=m+1}^{p} \{\boldsymbol{\gamma}_{2j}\boldsymbol{\gamma}_{2j}'\mathbf{A}_2(\mathbf{W}_2 - \lambda_{2j}\mathbf{I})^+ + (\mathbf{W}_2 - \lambda_{2j}\mathbf{I})^+\mathbf{A}_2\boldsymbol{\gamma}_{2j}\boldsymbol{\gamma}_{2j}'\}$$

$$= -\sum_{j=1}^{m} \{\lambda_{2j}^{-1}(\mathbf{P}_{22}\mathbf{A}_2\boldsymbol{\gamma}_{2j}\boldsymbol{\gamma}_{2j}' + \boldsymbol{\gamma}_{2j}\boldsymbol{\gamma}_{2j}'\mathbf{A}_2\mathbf{P}_{22})\},$$

and $(\mathbf{W}_2 - \lambda_{2j}\mathbf{I})^+$ denotes the Moore–Penrose generalized inverse of $(\mathbf{W}_2 - \lambda_{2j}\mathbf{I})$. Note that $\text{tr}(\mathbf{P}_{22}\mathbf{R}_2) = 0$. By using this fact in a similar expansion for $\hat{\mathbf{P}}_{22}^*$, it can be easily shown that

$$\hat{\tau}_i = \tau_i + \frac{1}{(p-m)} n^{-1/2}\{\text{tr}(\mathbf{P}_{22}\mathbf{U}_i\mathbf{P}_{22}) - \tau_i\text{tr}(\mathbf{P}_{22}\mathbf{U}\mathbf{P}_{22})\} + O_p(n^{-1}).$$

The first-order expansion for $\mathbf{B}_i$ does not involve $\mathbf{R}_2$, because

$$\mathbf{B}_i = n^{1/2}\mathbf{P}_{22}(\hat{\boldsymbol{\Omega}}_i - \hat{\tau}_i\hat{\boldsymbol{\Omega}})\mathbf{P}_{22} + O_p(n^{-1/2})$$

$$= \mathbf{P}_{22}\left[\mathbf{U}_i - \frac{1}{(p-m)}\text{tr}(\mathbf{P}_{22}\mathbf{U}_i\mathbf{P}_{22})\mathbf{I} - \tau_i\left\{\mathbf{U} - \frac{1}{(p-m)}\text{tr}(\mathbf{P}_{22}\mathbf{U}\mathbf{P}_{22})\mathbf{I}\right\}\right]\mathbf{P}_{22} + O_p(n^{-1/2}),$$

so that

$$\mathbf{b}_i = \text{vec}(\mathbf{B}_i) = \mathbf{t}_i - \frac{\tau_i}{h}\sum_{j=1}^{h} \mathbf{t}_j + O_p(n^{-1/2}),$$

where

$$\mathbf{t}_j = (\boldsymbol{\Gamma}_{22} \otimes \boldsymbol{\Gamma}_{22})\text{vec}\left\{\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22} - \frac{1}{(p-m)}\text{tr}(\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22})\mathbf{I}\right\}$$

$$= (\boldsymbol{\Gamma}_{22} \otimes \boldsymbol{\Gamma}_{22})\left\{\mathbf{I} - \frac{1}{(p-m)}\text{vec}(\mathbf{I})\text{vec}(\mathbf{I})'\right\}\text{vec}(\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22}).$$

Thus, due to the asymptotic normality of sample covariance matrices, $\text{vec}(\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22})$ is asymptotically normal, and hence so is $\mathbf{b}_i$.

To determine the asymptotic covariance structure of the $\mathbf{b}_i$'s, we will first determine $\text{cov}(\mathbf{t}_j)$. Because $\mathbf{w}$ has an elliptically symmetric distribution, it follows that under the null hypothesis, the distribution of $\boldsymbol{\Gamma}_{22}'\mathbf{w}$ given $\boldsymbol{\Gamma}_{21}'\mathbf{w}$ is also elliptically symmetric with mean vector $\mathbf{0}$ and covariance matrix $g(\boldsymbol{\Gamma}_{21}'\mathbf{w})\mathbf{I}$, where $g(\boldsymbol{\Gamma}_{21}'\mathbf{w})$ is defined as in Section 2 except that now we write it as a function of $\mathbf{w}$. In addition,

$$\text{cov}(\boldsymbol{\Gamma}_{22}'\mathbf{w}\mathbf{w}'\boldsymbol{\Gamma}_{22} \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}) = g^2(\boldsymbol{\Gamma}_{21}'\mathbf{w})[\{1 + \kappa(\boldsymbol{\Gamma}_{21}'\mathbf{w})\} \times (\mathbf{I} + \mathbf{K}_{pp})(\mathbf{I} \otimes \mathbf{I}) + \kappa(\boldsymbol{\Gamma}_{21}'\mathbf{w})\text{vec}(\mathbf{I})\text{vec}(\mathbf{I})'],$$

where $\mathbf{K}_{pp}$ is the commutation matrix defined by

$$\mathbf{K}_{pp} = \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbf{J}_{ij} \otimes \mathbf{J}_{ij}',$$

with $\mathbf{J}_{ij}$ the $p \times p$ matrix with 1 in the (i, j)th position and 0s elsewhere, and $\kappa(\boldsymbol{\Gamma}_{21}'\mathbf{w})$ is a kurtosis parameter satisfying

$$g^2(\boldsymbol{\Gamma}_{21}'\mathbf{w})\{1 + \kappa(\boldsymbol{\Gamma}_{21}'\mathbf{w})\} = E\{(\boldsymbol{\Gamma}_{22}'\mathbf{w})_l^4 \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}\}/3,$$

where $(\boldsymbol{\Gamma}_{21}'\mathbf{w})_l$ denotes the lth component of $\boldsymbol{\Gamma}_{21}'\mathbf{w}$, $l = 1, \ldots, p - m$. Now

$$E(\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22} \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}_{(j-1)n+1}, \ldots, \boldsymbol{\Gamma}_{21}'\mathbf{w}_{jn}) = n^{1/2}\left\{\frac{1}{n}\sum_{l=1}^{n} g(\boldsymbol{\Gamma}_{21}'\mathbf{w}_{(j-1)n+l}) - \tau_j\right\}\mathbf{I},$$

from which it follows that $E(\mathbf{t}_j \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}_{(j-1)n+1}, \ldots, \boldsymbol{\Gamma}_{21}'\mathbf{w}_{jn}) = \mathbf{0}$. The conditional covariance matrix is

$$\text{cov}\{\text{vec}(\boldsymbol{\Gamma}_{22}'\mathbf{U}_j\boldsymbol{\Gamma}_{22} \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}_{j(n-1)+1}, \ldots, \boldsymbol{\Gamma}_{21}'\mathbf{w}_{jn})\} = \frac{1}{n}\sum_{l=1}^{n} \text{cov}\{\text{vec}(\boldsymbol{\Gamma}_{22}'\mathbf{w}_{j(n-1)+l}\mathbf{w}_{j(n-1)+l}'\boldsymbol{\Gamma}_{22} \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}_{j(n-1)+1})\},$$

so that

$$\text{cov}(\mathbf{t}_j) = E\{\text{var}(\mathbf{t}_j \mid \boldsymbol{\Gamma}_{21}'\mathbf{w}_{(j-1)n+1}, \ldots, \boldsymbol{\Gamma}_{21}'\mathbf{w}_{jn})\} = \alpha_j\left\{(\mathbf{I} + \mathbf{K}_{pp})(\mathbf{P}_{22} \otimes \mathbf{P}_{22}) - \frac{2}{p-m}\text{vec}(\mathbf{P}_{22})\text{vec}(\mathbf{P}_{22})'\right\},$$

where $\alpha_j = \alpha_j(\mathbf{W}_2) = E[g^2(\boldsymbol{\Gamma}_{21}'\mathbf{w})\{1 + \kappa(\boldsymbol{\Gamma}_{21}'\mathbf{w})\} \mid y \in I_j]$. Thus if we let $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_h')'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_h)'$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_h)'$, $\mathbf{D}_\alpha = \text{diag}(\alpha_1, \ldots, \alpha_h)$, and $\mathbf{G} = \mathbf{G}(\mathbf{W}_2) = \mathbf{D}_\alpha - h^{-1}(\boldsymbol{\alpha}\boldsymbol{\tau}' + \boldsymbol{\tau}\boldsymbol{\alpha}') + h^{-2}(\sum_{i=1}^{h} \alpha_i)\boldsymbol{\tau}\boldsymbol{\tau}'$, then the asymptotic covariance matrix of $\mathbf{b}$ is given by

$$\boldsymbol{\Phi} = \mathbf{G} \otimes \left\{(\mathbf{I} + \mathbf{K}_{pp})(\mathbf{P}_{22} \otimes \mathbf{P}_{22}) - \frac{2}{(p-m)}\text{vec}(\mathbf{P}_{22})\text{vec}(\mathbf{P}_{22})'\right\}.$$

Because $\mathbf{G}$ has a single singularity associated with the latent vector $\mathbf{1}$, $\nu_2 = \text{rank}(\boldsymbol{\Phi}) = (h - 1)\{(p - m)(p - m + 1)/2 - 1\}$, and so, due to the asymptotic normality of $\mathbf{b}$, $\mathbf{b}'\boldsymbol{\Phi}^+\mathbf{b}$ is asymptotically chi-squared with ν_2 degrees of freedom. Note that because $\text{vec}(\mathbf{P}_{22})'\mathbf{b} = O_p(n^{-1/2})$, $\mathbf{b}'\boldsymbol{\Phi}_*^+\mathbf{b}$ has this same asymptotic distribution, where $\boldsymbol{\Phi}_* = \mathbf{G} \otimes (\mathbf{I} + \mathbf{K}_{pp})(\mathbf{P}_{22} \otimes \mathbf{P}_{22})$.

All that remains is to obtain a consistent estimator for $\boldsymbol{\Phi}_*^+$. Now $\hat{\mathbf{P}}_{22}$ will be a consistent estimator for $\mathbf{P}_{22}$ when $\lambda_{2m} > \lambda_{2,m+1}$, so that $\hat{\boldsymbol{\Phi}}_*^+ = \hat{\mathbf{G}}_*^+ \otimes \frac{1}{4}(\mathbf{I} + \mathbf{K}_{pp})(\hat{\mathbf{P}}_{22} \otimes \hat{\mathbf{P}}_{22})$ will be consistent for $\boldsymbol{\Phi}_*^+$ as long as $\hat{\mathbf{G}}_*^+$ is consistent for $\mathbf{G}^+$. Similarly, $\hat{\mathbf{P}}_{22}^*$ is also consistent for $\mathbf{P}_{22}$, so that the consistency of the $\hat{\boldsymbol{\Omega}}_j$'s guarantees the consistency of the $\hat{\tau}_j$'s, whereas a simple consistent moment estimator for α_j is given by

$$\hat{\alpha}_j = \hat{\alpha}_j(\hat{\mathbf{W}}_2) = \frac{1}{3n(p-m)} \sum_{i=1}^{n} \sum_{l=1}^{p-m} (\hat{\mathbf{\Gamma}}'_{22}\mathbf{w}_{(j-1)n+i})_l^4.$$

Consequently, substitution of $\hat{\tau}_j$ and $\hat{\alpha}_j$ in the formula for $\mathbf{G}$ yields a consistent estimator, $\hat{\mathbf{G}} = \hat{\mathbf{G}}(\hat{\mathbf{W}}_2)$, of $\mathbf{G}$. If $\mathbf{P}_{\hat{G}}$ is the eigenprojection of $\hat{\mathbf{G}}$ corresponding to its $h-1$ largest latent roots and $\hat{\mathbf{G}}_* = \hat{\mathbf{G}}_*(\hat{\mathbf{W}}_2) = \mathbf{P}_{\hat{G}}\hat{\mathbf{G}}\mathbf{P}_{\hat{G}}$, then the consistency of $\hat{\mathbf{G}}_*^+ = \{\hat{\mathbf{G}}_*(\hat{\mathbf{W}}_2)\}^+$ as an estimator of $\mathbf{G}^+$ follows from the consistency of $\hat{\mathbf{G}}_*$ and from the continuity of the Moore–Penrose generalized inverse, because $\text{rank}(\hat{\mathbf{G}}_*) = \text{rank}(\mathbf{G})$. Thus in testing H_{0m}^2 we may use

$$T_{2,m}^{(2)} = \frac{1}{4} b'\{\hat{\mathbf{G}}_*^+ \otimes (\mathbf{I} + \mathbf{K}_{pp})(\hat{\mathbf{P}}_{22} \otimes \hat{\mathbf{P}}_{22})\}\mathbf{b} = \frac{1}{2}\sum_{i=1}^{h}\sum_{j=1}^{h} \hat{g}_*^{ij}\text{tr}(\mathbf{Z}_i\mathbf{Z}_j), \tag{A.1}$$

which also has an asymptotic chi-squared distribution with ν_2 degrees of freedom under the null. Here $\hat{g}_*^{ij}$ denotes the (i,j)th element of $\hat{\mathbf{G}}_*^+$, and $\mathbf{Z}_i = \hat{\mathbf{\Gamma}}'_{22}\hat{\mathbf{\Omega}}^{-1/2}(\hat{\mathbf{\Omega}}_i - \hat{\tau}_i\hat{\mathbf{\Omega}})\hat{\mathbf{\Omega}}^{-1/2}\hat{\mathbf{\Gamma}}_{22}$. Note that although we used the symmetric square root matrix $\hat{\mathbf{\Omega}}^{-1/2}$ in our derivation, any square root matrix can be used because $\hat{\mathbf{\Gamma}}'_{22}\hat{\Omega}^{-1/2}$, and hence $T_{2,m}^{(2)}$ does not depend on this choice.

A.2 PROOF OF THEOREM 2

First we show that $\mathbf{a}_i$ is asymptotically normal. Using the expansion $\hat{\Delta} = \Delta + n^{-1/2}\mathbf{F} + O_p(n^{-1})$, along with the expansions introduced in Section A.1, we can write $\hat{\mathbf{W}}_3 = \mathbf{W}_3 + n^{-1/2}\mathbf{A}_3 + O_p(n^{-1})$, where $\mathbf{A}_3$ satisfies

$$\mathbf{P}_{32}\mathbf{A}_3 = \mathbf{P}_{32}\left\{\mathbf{F} - \frac{1}{2}\mathbf{U}\mathbf{W}_3 + \frac{1}{h}\sum_{j=1}^{h}(\mathbf{U}_j - \tau_j\mathbf{U})(\mathbf{\Omega}_j - \tau_j\mathbf{I})\right\}.$$

An expansion for $\hat{\mathbf{P}}_{32}$, similar to that of $\hat{\mathbf{P}}_{22}$ in Section 3.2, is given by $\hat{\mathbf{P}}_{32} = \mathbf{P}_{32} + n^{-1/2}\mathbf{R}_3 + O_p(n^{-1})$, where

$$\begin{aligned}\mathbf{R}_3 &= -\sum_{j=1}^{m}\{\lambda_{3j}^{-1}(\mathbf{P}_{32}\mathbf{A}_3\gamma_{3j}\gamma'_{3j} + \gamma_{3j}\gamma'_{3j}\mathbf{A}_3\mathbf{P}_{32})\} \\ &= -(\mathbf{P}_{32}\mathbf{A}_3\mathbf{W}_3^+ + \mathbf{W}_3^+\mathbf{A}_3\mathbf{P}_{32}).\end{aligned}$$

Letting $n^{-1/2}\mathbf{v}_i = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) - (\boldsymbol{\mu}_i - \boldsymbol{\mu})$, it is straightforward to show that

$$\begin{aligned}\mathbf{a}_i &= \mathbf{R}_3(\boldsymbol{\mu}_i - \boldsymbol{\mu}) - \frac{1}{2}\mathbf{P}_{32}\mathbf{U}(\boldsymbol{\mu}_i - \boldsymbol{\mu}) + \mathbf{P}_{32}\mathbf{v}_i + O_p(n^{-1/2}) \\ &= -\{(\boldsymbol{\mu}_i - \boldsymbol{\mu})'\mathbf{W}_3^+ \otimes \mathbf{P}_{32}\}\text{vec}(\mathbf{F}) \\ &\quad - \frac{1}{h}\sum_{j=1}^{h}\left[(\boldsymbol{\mu}_i - \boldsymbol{\mu})'\mathbf{W}_3^+\left\{(\mathbf{\Omega}_j - \tau_j\mathbf{I}) - \frac{1}{h}\sum_{l=1}^{h}\tau_l(\mathbf{\Omega}_l - \tau_l\mathbf{I})\right\} \otimes \mathbf{P}_{32}\right] \\ &\quad \times \text{vec}(\mathbf{U}_j) + \mathbf{P}_{32}\mathbf{v}_i + O_p(n^{-1/2}).\end{aligned}$$

The asymptotic normality of $\mathbf{a}_i$ then follows from the asymptotic normality of $\text{vec}(\mathbf{F})$, $\text{vec}(\mathbf{U}_i)$, and $\mathbf{v}_i$. To determine the covariance structure of the $\mathbf{a}_i$'s, note that

$$\begin{aligned}&\text{var}\{\text{vec}(\mathbf{P}_{31}\mathbf{U}_i\mathbf{P}_{32})\} \\ &\quad = E[\text{var}\{\text{vec}(\mathbf{P}_{31}\mathbf{U}_i\mathbf{P}_{32}) \mid \mathbf{\Gamma}'_{31}\mathbf{w}_{(i-1)n+1}, \ldots, \mathbf{\Gamma}'_{31}\mathbf{w}_{in}\}] \\ &\quad = \mathbf{H}_i \otimes \mathbf{P}_{32},\end{aligned}$$

where $\mathbf{H}_i = E\{g(\mathbf{\Gamma}'_{31}\mathbf{w})\mathbf{P}_{31}\mathbf{w}\mathbf{w}'\mathbf{P}_{31} \mid y \in I_i\}$. Using this along with the variances and covariances of $\text{vec}(\mathbf{F})$ and $\mathbf{v}_i$, we find that the asymptotic covariance matrix of $\mathbf{a} = (\mathbf{a}'_1, \ldots, \mathbf{a}'_h)'$ is $\mathbf{\Psi} = \mathbf{C} \otimes \mathbf{P}_{32}$, where $\mathbf{C} = \mathbf{C}(\mathbf{W}_3) = \mathbf{C}_1 + \mathbf{C}_2$,

$$\begin{aligned}\mathbf{C}_1 = \mathbf{C}_1(\mathbf{W}_3) &= \{\mathbf{I} - h^{-1}(\mathbf{1}\mathbf{1}' + \mathbf{M}'\mathbf{W}_3^+\mathbf{M})\} \\ &\quad \times \mathbf{D}_\tau\{\mathbf{I} - h^{-1}(\mathbf{1}\mathbf{1}' + \mathbf{M}'\mathbf{W}_3^+\mathbf{M})\}, \end{aligned} \tag{A.2}$$

$$\begin{aligned}\mathbf{C}_2 = \mathbf{C}_2(\mathbf{W}_3) &= \mathbf{M}'\mathbf{W}_3^+\left[\sum_{j=1}^{h}\left\{(\mathbf{\Omega}_j - \tau_j\mathbf{I}) - \frac{1}{h}\sum_{l=1}^{h}\tau_l(\mathbf{\Omega}_l - \tau_l\mathbf{I})\right\}\right. \\ &\quad \left.\times \mathbf{H}_j\left\{(\mathbf{\Omega}_j - \tau_j\mathbf{I}) - \frac{1}{h}\sum_{l=1}^{h}\tau_l(\mathbf{\Omega}_l - \tau_l\mathbf{I})\right\}\right]\mathbf{W}_3^+\mathbf{M},\end{aligned}$$

$\mathbf{D}_\tau = \text{diag}(\tau_1, \ldots, \tau_h)$, and $\mathbf{M} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}, \ldots, \boldsymbol{\mu}_h - \boldsymbol{\mu})$. It is shown in Section A.3 that $\text{rank}(\mathbf{\Psi}) = (p-m)(h-1-m+k_2)$. As in the previous section, we can obtain an expansion for $\mathbf{b}_i = \text{vec}(\mathbf{B}_i)$, from which we find that $\text{cov}(\mathbf{a}_i, \mathbf{b}_j) = O(n^{-1/2})$ for all i, j and the asymptotic covariance matrix of $\mathbf{b}_i$ is now

$$\mathbf{\Phi} = \mathbf{G}(\mathbf{W}_3) \otimes \left\{(\mathbf{I} + \mathbf{K}_{pp})(\mathbf{P}_{32} \otimes \mathbf{P}_{32}) - \frac{2}{h}\text{vec}(\mathbf{P}_{32})\text{vec}(\mathbf{P}_{32})'\right\}.$$

Thus if $T_{2,m}^{(3)}$ is evaluated using (A.1), except that $\hat{\mathbf{P}}_{32}$, $\hat{\mathbf{P}}_{32}^*$, and $\hat{\mathbf{\Gamma}}_{32}$ are used in place of $\hat{\mathbf{P}}_{22}$, $\hat{\mathbf{P}}_{22}^*$, and $\hat{\mathbf{\Gamma}}_{22}$, then we find that asymptotically $T_{2,m}^{(3)}$ is chi-squared with ν_2 degrees of freedom. Returning to $\mathbf{a}$, a consistent estimator of $\mathbf{\Psi}^+$ is $\hat{\mathbf{\Psi}}^+ = \hat{\mathbf{C}}^+ \otimes \hat{\mathbf{P}}_{32}$, where $\hat{\mathbf{C}}^+ = \{\hat{\mathbf{C}}(\hat{\mathbf{W}}_3)\}^+$ is given in Section A.4. Using this we form

$$T_{1,m}^{(3)} = \mathbf{a}'\hat{\mathbf{\Psi}}^+\mathbf{a} = \sum_{i=1}^{h}\sum_{j=1}^{h}\hat{c}^{ij}\mathbf{y}'_i\mathbf{y}_j,$$

where $\hat{c}^{ij}$ is the (i,j)th element of $\hat{\mathbf{C}}^+$ and $\mathbf{y}_i = \mathbf{y}_i(\hat{\mathbf{W}}_3) = \hat{\mathbf{\Gamma}}'_{32}\hat{\mathbf{\Omega}}^{-1/2} \times (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$. Under H_{0m}^3, $T_{1,m}^{(3)}$ is asymptotically chi-squared with $\nu_1 = (p-m)(h-1-m+k_2)$ degrees of freedom. Thus, because $\mathbf{a}$ and $\mathbf{b}$ are asymptotically independent, H_{0m}^3 can be tested by comparing $T_{3,m}^{(3)} = T_{1,m}^{(3)} + T_{2,m}^{(3)}$ to the chi-squared distribution with $\nu_1 + \nu_2$ degrees of freedom.

A.3 THE RANK OF $\mathbf{\Psi}$

Note that $\mathbf{a}$ can be expressed asymptotically as

$$\begin{aligned}\mathbf{a} &= \{(\mathbf{I} - \mathbf{M}'\mathbf{W}_3^+\mathbf{M}) \otimes \mathbf{P}_{32}\}\mathbf{v} - (\mathbf{M}'\mathbf{W}_3^+ \otimes \mathbf{P}_{32}) \\ &\quad \times \left(\frac{1}{h}\sum_{i=1}^{h}\left[\left\{(\mathbf{\Omega}_i - \tau_i\mathbf{I}) - \frac{1}{h}\sum_{l=1}^{h}\tau_l(\mathbf{\Omega}_l - \tau_l\mathbf{I})\right\} \otimes \mathbf{I}\right]\text{vec}(\mathbf{U}_i)\right) \\ &= \mathbf{v}_* - \mathbf{u}_*,\end{aligned}$$

where $\mathbf{v} = (\mathbf{v}'_1, \ldots, \mathbf{v}'_h)'$. Define $\mathbf{Q}_1(p \times m - k_2)$, $\mathbf{Q}_2(p \times k_1 + k_2 - m)$, and $\mathbf{Q}_3(p \times m - k_1)$ so that the columns of $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ form an orthonormal set and the columns of $(\mathbf{Q}_1, \mathbf{Q}_2)$ and $(\mathbf{Q}_2, \mathbf{Q}_3)$ form orthonormal bases for the projection spaces of $\mathbf{W}_1$ and $\mathbf{W}_2$. Thus we can write

$$\mathbf{M} = \mathbf{Q}_1\mathbf{A}_1 + \mathbf{Q}_2\mathbf{A}_2, \qquad \mathbf{W}_2 = (\mathbf{Q}_2, \mathbf{Q}_3)\mathbf{E}\begin{pmatrix}\mathbf{Q}'_2 \\ \mathbf{Q}'_3\end{pmatrix},$$

where $\mathbf{A}_1$ is $(m-k_1) \times h$, $\mathbf{A}_2$ is $(k_1 + k_2 - m) \times h$, $\mathbf{A} = (\mathbf{A}'_1, \mathbf{A}'_2)$ is full rank, and $\mathbf{E}$ is a nonsingular $k_2 \times k_2$ matrix. Because $(\mathbf{1}' \otimes \mathbf{I})\mathbf{a} = \mathbf{0}$ and $(\mathbf{I} \otimes \mathbf{P}_{31})\mathbf{a} = \mathbf{0}$, it follows that $\text{rank}(\mathbf{\Psi}) \le (h-1)(p-m)$. Clearly, $(\mathbf{c}' \otimes \mathbf{P}_{32})\mathbf{v}_* \ne \mathbf{0}$ for any $\mathbf{c}$ orthogonal to $\mathbf{1}$ and to the rows of $\mathbf{A}_1$ and $\mathbf{A}_2$, so $\text{rank}(\mathbf{\Psi}) \ge (h-1-k_1)(p-m)$. In addition, it can be verified that $(\mathbf{Q}_1\mathbf{A}_1 \otimes \mathbf{P}_{32})\mathbf{a} = \mathbf{0}$, whereas the distribution of $(\mathbf{Q}_2\mathbf{A}_2 \otimes \mathbf{P}_{32})\mathbf{u}_*$ is nonsingular except for the singularities from $\mathbf{P}_{32}$, so the result follows.

A.4 ESTIMATORS FOR $\mathbf{\Psi}^+$ AND $\mathbf{\Psi}_1^+$

When forming estimators it must be remembered that $\mathbf{\Psi}$ and $\mathbf{\Psi}_1$ were derived under the assumption that the data were transformed so that $\boldsymbol{\Omega} = \mathbf{I}$. Due to the continuity of the Moore–Penrose inverse, the estimator $\hat{\mathbf{C}}^{+} \otimes \hat{\mathbf{P}}_{32}$ will be a consistent estimator of $\boldsymbol{\Psi}^{+} = \mathbf{C}^{+} \otimes \mathbf{P}_{32}$ if $\hat{\mathbf{C}} = \hat{\mathbf{C}}(\hat{\mathbf{W}}_3)$ is a consistent estimator of $\mathbf{C} = \mathbf{C}(\mathbf{W}_3)$ and $\operatorname{rank}(\hat{\mathbf{C}}) = \operatorname{rank}(\mathbf{C}) = (h - 1 - m + k_2)$. The previously defined estimator $\hat{\tau}_i = \hat{\tau}_i(\hat{\mathbf{W}}_3^{*})$ is consistent for τ_i, and a simple consistent estimator of $\mathbf{H}_i$ is given by the moment estimator

$$\hat{\mathbf{H}}_i = \frac{1}{n} \sum_{j=1}^{n} \tilde{g}_{ij} \hat{\mathbf{P}}_{31} \hat{\boldsymbol{\Omega}}^{-1/2} (\mathbf{x}_{(i-1)n+j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{(i-1)n+j} - \bar{\mathbf{x}}_i)' \hat{\boldsymbol{\Omega}}^{-1/2} \hat{\mathbf{P}}_{31},$$

where

$$\tilde{g}_{ij} = \frac{1}{p - m} (\mathbf{x}_{(i-1)n+j} - \bar{\mathbf{x}}_i) \hat{\boldsymbol{\Omega}}^{-1/2} \hat{\mathbf{P}}_{32} \hat{\boldsymbol{\Omega}}^{-1/2} (\mathbf{x}_{(i-1)n+j} - \bar{\mathbf{x}}_i).$$

A consistent estimator of $\mathbf{W}_3^{+}$ is given by $(\hat{\mathbf{P}}_{31}\hat{\mathbf{W}}_3\hat{\mathbf{P}}_{31})^{+}$. Substitution of all these estimators in the expression for $\mathbf{C}$, along with $\hat{\boldsymbol{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$ and $\hat{\boldsymbol{\Omega}}^{-1/2}\hat{\boldsymbol{\Omega}}_i\hat{\boldsymbol{\Omega}}^{-1/2}$ for $(\boldsymbol{\mu}_i - \boldsymbol{\mu})$ and $\boldsymbol{\Omega}_i$, yields a consistent estimator, $\tilde{\mathbf{C}}$, of $\mathbf{C}$. Finally, if we define $\hat{\mathbf{C}} = \mathbf{P}_{\tilde{C}}\tilde{\mathbf{C}}\mathbf{P}_{\tilde{C}}$, where $\mathbf{P}_{\tilde{C}}$ is the eigenprojection of $\tilde{\mathbf{C}}$ corresponding to its $(h - 1 - m + k_2)$ largest latent roots, then $\hat{\mathbf{C}}^{+}$ will be consistent for $\mathbf{C}^{+}$.

Similarly, our estimator $\hat{\boldsymbol{\Psi}}_1^{+} = \hat{\mathbf{C}}_1^{+} \otimes \hat{\mathbf{P}}_{22}$ will consistently estimate $\boldsymbol{\Psi}_1^{+} = \mathbf{C}_1^{+} \otimes \mathbf{P}_{22}$ if $\hat{\mathbf{C}}_1^{+} = \{\hat{\mathbf{C}}_1(\hat{\mathbf{W}}_1)\}^{+}$ is consistent for $\mathbf{C}_1^{+} = \{\mathbf{C}_1(\mathbf{W}_1)\}^{+}$. First, an estimator $\tilde{\mathbf{C}}_1$ is obtained by replacing $(\boldsymbol{\mu}_i - \boldsymbol{\mu})$, $\mathbf{W}_3^{+}$, and τ_i in (A.2) by $\hat{\boldsymbol{\Omega}}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$, $(\hat{\mathbf{P}}_{11}\hat{\mathbf{W}}_1\hat{\mathbf{P}}_{11})^{+}$, and $\hat{\tau}_i = \hat{\tau}_i(\hat{\mathbf{W}}_1)$. Then our estimator of $\mathbf{C}_1^{+}$ is given by $\hat{\mathbf{C}}_1^{+} = (\mathbf{P}_{\tilde{C}_1}\tilde{\mathbf{C}}_1\mathbf{P}_{\tilde{C}_1})^{+}$, where $\mathbf{P}_{\tilde{C}_1}$ is the eigenprojection of $\tilde{\mathbf{C}}_1$ corresponding to its $(h - 1 - m)$ largest latent roots.

[*Received April 1992. Revised April 1993.*]

REFERENCES

Cook, R. D., and Weisberg, S. (1991), Comment on "Sliced Inverse Regression for Dimension Reduction," by K. C. Li, *Journal of the American Statistical Association,* 86, 328–332.

Li, K. C. (1990), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," technical report, UCLA, Dept. of Mathematics.

——— (1991a), "Sliced Inverse Regression for Dimension Reduction" (with comments), *Journal of the American Statistical Association,* 86, 316–327.

——— (1991b), Rejoinder to "Sliced Inverse Regression for Dimension Reduction," by K. C. Li, *Journal of the American Statistical Association,* 86, 337–342.

Kent, J. T. (1991), Comment on "Sliced Inverse Regression for Dimension Reduction," by K. C. Li, *Journal of the American Statistical Association,* 86, 336–337.

Maronna, R. A. (1976), "Robust *M*-Estimators of Multivariate Location and Scatter," *The Annals of Statistics,* 4, 51–67.

Schott, J. R. (1993), "Dimensionality Reduction in Quadratic Discriminant Analysis," *Computational Statistics & Data Analysis,* 16, 161–174.

Tyler, D. E. (1981), "Asymptotic Inference for Eigenvectors," *The Annals of Statistics,* 9, 725–736.