Statistica Sinica **13**(2003), 143-162

USING LINEAR SMOOTHERS TO ASSESS THE STRUCTURAL DIMENSION OF REGRESSIONS

E. Bura

The George Washington University

Abstract: Sliced Inverse Regression (Li (1991)) is a simple nonparametric estimation method for the *structural dimension* of a regression, that is, for the dimension of the linear subspace spanned by projections of the multidimensional regressor vector $\mathbf{X}$ that contains part or all of the modelling information about the regression of a random variable Y on $\mathbf{X}$. In this paper, the nonparametric estimation method is extended to include the family of linear smoothers. No restrictions are placed on the distribution of the regressors except for the *linearity condition* and existence of second moments. An asymptotic chi-square test for dimension is obtained. Theoretical results are illustrated with a small comparative simulation study.

Key words and phrases: Asymptotic chi-square test, dimension reduction, inverse regression.

1. Introduction

High-dimensional data have become increasingly common in statistical applications. In a regression setting, a traditional way to cope with dimensionality is to impose assumptions for a specific structure on the mean regression function such as linearity or additivity. In general though, regression studies the nature of the relationship between Y and $\mathbf{X}$ at the conditional c.d.f. level.

In this framework, suppose $\mathbf{X}$ can be replaced by $k \leq p$ linear combinations of its components, $\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_k^T\mathbf{X}$, without losing information on $F(Y|\mathbf{X})$ so that, for all values of $\mathbf{X}$,

$$F(Y|\mathbf{X}) = F(Y|\boldsymbol{\eta}^T\mathbf{X}), \tag{1}$$

where $\boldsymbol{\eta}$ is the $p \times k$ matrix with columns $\boldsymbol{\eta}_j$. This formulation of the dependence of Y on $\mathbf{X}$ was introduced by Cook (1994a). A similar formulation was proposed by Li (1991). One has from (1) that the conditional c.d.f. of $Y|\mathbf{X}$ depends on $\mathbf{X}$ only through $\boldsymbol{\eta}^T\mathbf{X}$, the coordinates of a projection of $\mathbf{X}$ onto the k-dimensional linear subspace spanned by the columns of $\boldsymbol{\eta}$. Consequently, $\boldsymbol{\eta}^T\mathbf{X}$ contains equivalent or *sufficient*, in the statistical sense, information for the regression of Y on $\mathbf{X}$. Most importantly, if $k < p$, then sufficient reduction in the dimension of the regression is achieved, which in turn leads to *sufficient summary plots* of

Y versus $\boldsymbol{\eta}^T\mathbf{X}$ as graphical displays of all the necessary modelling information. These summary plots can subsequently guide the selection of parametric models for $F(Y|\mathbf{X})$ when deemed appropriate. Alternatively, the reduced data can be modelled nonparametrically more efficiently in that some part of the *curse of dimensionality* has been overcome.

When (1) holds then it also holds with $\boldsymbol{\eta}$ replaced by any basis for the range space $S(\boldsymbol{\eta})$ of $\boldsymbol{\eta}$. We follow Li ((1991), (1992) and call $S(\boldsymbol{\eta})$ a *dimension-reduction subspace* for $F(Y|\mathbf{X})$, or for the regression of Y on $\mathbf{X}$. The smallest dimension-reduction subspace provides the greatest dimension reduction in the predictor vector. There are several ways to define such a subspace. The focus here is on the *central dimension-reduction subspace*, denoted by $S_{Y|\mathbf{X}}$ (Cook (1994b), (1996), (1998a,b)). $S_{Y|\mathbf{X}}$ is the intersection of all dimension-reduction subspaces for $F(Y|\mathbf{X})$ and is trivially a subspace, but not necessarily a dimension-reduction one. The existence of central subspaces can be ascertained by placing fairly weak restrictions on aspects of the joint distribution of Y and $\mathbf{X}$ (Cook (1994a), (1996)).

Sliced inverse regression (SIR), introduced by Li (1991), is a simple *non-smooth nonparametric* estimation method for $S_{Y|\mathbf{X}}$ and its dimension. Aragon and Saracco (1997) developed a smoother version of SIR by pooling multiple slicings together, and Zhu and Fang (1996) used kernel smoothers to estimate the central subspace. In this article, the family of general linear smoothers is used to estimate the central dimension reduction subspace, and a new test for the dimension of $S_{Y|\mathbf{X}}$ is developed. The method imposes no restrictions on the predictors except for the *linearity condition* which will be discussed in the next section.

The regression context is presented and existing dimension estimation methods, with emphasis on SIR (Li (1991)), are reviewed in Section 2. The proposed estimation method based on linear smoothers is introduced and described in Section 3. In Section 4, a test statistic for dimension is derived and proved to be asymptotically chi-square distributed for both the homoskedastic and the heteroskedastic cases. In Section 5, a small simulation study compares the power of the two testing methods for dimension. Local linear smoothing is used for the proposed method. A concluding discussion is presented in Section 6. The lengthier proofs are given in the appendix.

2. Background: Inverse Regression and Dimension Reduction

Let $S_{\mathrm{E}(\mathbf{X}|Y)}$ denote the subspace spanned by $\{\,\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X}) : Y \in \Omega_Y\,\}$, where $\Omega_Y \subset \mathbb{R}$ is the marginal sample space of Y. Given (1), assume that the marginal distribution of the predictors $\mathbf{X}$ satisfy the following condition,

henceforth referred to as the *linearity condition*: for all $\mathbf{b} \in \mathbb{R}^p$, $\mathrm{E}(\mathbf{b}^T\mathbf{X}|\boldsymbol{\eta}^T\mathbf{X})$ is linear in $\boldsymbol{\eta}^T\mathbf{X}$.

Under this linearity condition on the regressor distribution, Li (1991, Thm. 3.1) showed that the centered inverse regression curve $\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X})$ satisfies $\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X}) \in S(\boldsymbol{\Sigma}_x\boldsymbol{\eta})$ or, equivalently,

$$S_{\mathrm{E}(\mathbf{X}|Y)} \subset S(\boldsymbol{\Sigma}_x\boldsymbol{\eta}) = \boldsymbol{\Sigma}_x S_{Y|\mathbf{X}}, \tag{2}$$

where $\boldsymbol{\Sigma}_x = \mathrm{Cov}(\mathbf{X})$. The linearity condition on $\mathrm{E}(\mathbf{b}^T\mathbf{X}|\boldsymbol{\eta}^T\mathbf{X})$ is required to hold only for the basis $\boldsymbol{\eta}$ of the central subspace. Since $\boldsymbol{\eta}$ is unknown, in practice one may require that it hold for all possible $\boldsymbol{\eta}$ and this is equivalent to elliptical symmetry of the distribution of $\mathbf{X}$ (Eaton (1986)). As Li (1991) pointed out, the linearity condition is not a severe restriction, since most low-dimensional projections of a high-dimensional data cloud are close to being normal (Diaconis and Freedman (1984), Hall and Li (1993)). In addition, there often exist transformations of the predictors that make them comply with the linearity condition.

Expression (2) leads to the use of inverse regression as an estimation tool for a fraction of, or the entire central dimension-reduction subspace. Sliced Inverse Regression (SIR), proposed by Li (1991), was the first dimension estimation method based on inverse regression. In SIR, the range of the one-dimensional variable Y is partitioned into a fixed number of slices and the p components of $\mathbf{Z}$, the standardized version of $\mathbf{X}$, are regressed on a discrete version of Y resulting from slicing its range. A very simple nonparametric estimate of the inverse regression curve $\mathrm{E}(\mathbf{Z}|Y)$ serves to estimate the central dimension-reduction subspace via estimating $\mathrm{Cov}(\mathrm{E}(\mathbf{Z}|Y))$, as follows: Within each slice the sample covariance matrix of $\mathbf{X}$ is computed and a weighted sum of the sample covariance matrices across slices serves as an estimate of $\mathrm{Cov}(\mathrm{E}(\mathbf{Z}|Y))$. Li (1991) computed a test statistic that is asymptotically chi-square distributed with $(p-d)(H-d-1)$ degrees of freedom, provided the regressors are normal, where $d = \dim(S_{\mathrm{E}(\mathbf{X}|Y)})$ and H is the fixed number of slices. The test statistic can be used to estimate the dimension of $S_{\mathrm{E}(\mathbf{X}|Y)}$ by performing tests of $d = j$ versus $d \geq j+1$, $j = 0, \ldots, p-1$.

SIR has been proven to be a simple and useful first method for reducing the dimension in a regression problem. Nonetheless, it requires normality of the regressor vector $\mathbf{X}$ for the chi-square asymptotic test for dimension to apply (Li (1991)) and the variance of the conditional distribution of $\mathbf{X}$ given Y has to be constant. Bura and Cook (2001b) studied and stated the minimal necessary conditions on both the regressor and the conditional distribution of $\mathbf{X}|Y$ for the SIR test statistic to be asymptotically chi-square. They also introduced the weighted chi-square test where the only distributional requirement is finite second moments. Additionally, several testing techniques based on inverse regression

that use the same simple nonparametric estimation method as Li (1991), and that try to lift the normality assumption in SIR, have been developed (Schott (1994), Velilla (1998), Ferré (1998)).

SIR utilizes simple non-smooth nonparametric estimates of the inverse regression curves which may miss important relevant information as the continuous nature of the data is ignored. To address this limitation, Aragon and Saracco (1997) combined many slicings to produce a pooled slicings based estimate of the dimension reduction subspace. The latter is consistent but the asymptotic properties were derived only for the one-dimensional case. Bura and Cook (2001a) developed parametric inverse regression (PIR) that fits continuous parametric curves to the p inverse regressions and requires no distributional assumptions except for finite second moments of $\mathbf{X}$ and $F(\mathbf{X}|Y)$.

Nevertheless, selecting appropriate parametric curves may be difficult. Zhu and Fang (1996) used kernel smoothers to estimate $\mathrm{Cov}\left(\mathrm{E}\left(\mathbf{Z}|Y\right)\right)$ without imposing distributional restrictions on $\mathbf{X}$ except for finite fourth moments. They proved that the kernel estimate is asymptotically normal and can, therefore, be used to provide a test for dimension. Fung, He, Liu and Shi (2000) replaced slicing by B-spline basis functions and estimated both the dimension and the central subspace using canonical correlations between the predictors and the basis functions of the response. An important aspect of their methodology is that it is also applicable to weakly dependent stationary sequences.

In this article, we generalize and extend SIR and the kernel method to the family of linear smoothers. Spline-basis smoothers belong to this family so the proposed methodology covers the development of Fung, He, Liu and Shi (2000) even though the estimation is not based on canonical correlation analysis. The resulting estimate of $S_{\mathrm{E}\left(\mathbf{X}|Y\right)}$ is shown to be asymptotically normal. The associated asymptotic chi-square test for dimension requires only finite second moments for both the regressor vector and the conditional distribution of $\mathbf{X}|Y$ at the expense of imposing regularity conditions on the smoother.

3. Nonparametric Inverse Regression

Let $\mathbf{X}^\star = (x_1^\star, \ldots, x_p^\star)^T$ be the centered regressor vector, i.e., $\mathbf{X}^\star = \mathbf{X} - \mathrm{E}\left(\mathbf{X}\right)$. In the inverse regression setting, Y is the explanatory variable. The regression relationship can be modeled as

$$\mathbf{X}^\star|Y = \boldsymbol{m}(Y) + \boldsymbol{\epsilon},$$

or, $X_j^\star|Y = m_j(Y) + \epsilon_j$, $j = 1, \ldots, p$, where $\boldsymbol{\epsilon}$ denotes the error vector and the regression function $\mathbf{m}(Y) = \mathrm{E}\left(\mathbf{X}^\star|Y\right)$ is unknown but centered so that the model is consistent with the fact that the expectation of $\mathbf{X}^\star$ equals zero. We assume

that the error vector has zero mean and a $p \times p$ positive definite covariance matrix $\mathbf{\Sigma}_{x|y}$.

If a random sample of size n is drawn from $(Y, \mathbf{X}^\star)$, the model becomes $X^\star_{ij}|Y_i = m_j(Y_i) + \epsilon_{ij}$ for $j = 1, \ldots, p$, $i = 1, \ldots, n$. Equivalently, in a matrix format, the model is written as

$$\mathbf{X}^\star_n|Y = \mathbf{M}(Y) + \mathbf{E}_n, \tag{3}$$

where $\mathbf{X}^\star_n = (X^\star_{ij})$, $\boldsymbol{M}(Y) = (m_j(Y_i) - \bar{m}_j)$, with $\bar{m}_j = \sum_{i=1}^n m_j(Y_i)/n$, and $\mathbf{E}_n = (\epsilon_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$. Define the nonparametric estimate of $m_j(Y_i)$ by $\hat{m}_j(Y_i) = \sum_{k=1}^n X^\star_{kj} \widetilde{W}_{nk}(Y_i)$, $j = 1, \ldots, p$, $i = 1, \ldots, n$, where $\{\widetilde{W}_{nk}(Y_i)\}_{k=1}^n$ is a sequence of weights which may depend on the whole vector $\mathbf{Y}^T = (Y_1, \ldots, Y_n)$. Then, the nonparametric estimate of the centered $m_j(Y_i)$ is given by

$$\hat{X}^\star_{ij} = \hat{m}_j(Y_i) - \bar{\hat{m}}_j = \sum_{k=1}^n X^\star_{kj} W_{nk}(Y_i), \tag{4}$$

where $W_{nk}(Y_i) = \widetilde{W}_{nk}(Y_i) - \sum_{i=1}^n \widetilde{W}_{nk}(Y_i)/n$. In a matrix format, we write

$$\widehat{\mathbf{X}}^\star_n = \begin{bmatrix} \hat{X}_1^{\star T} \\ \vdots \\ \hat{X}_n^{\star T} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n X^\star_{k1} W_{nk}(Y_1) & \sum_{k=1}^n X^\star_{k2} W_{nk}(Y_1) & \cdots & \sum_{k=1}^n X^\star_{kp} W_{nk}(Y_1) \\ \sum_{k=1}^n X^\star_{k1} W_{nk}(Y_2) & \sum_{k=1}^n X^\star_{k2} W_{nk}(Y_2) & \cdots & \sum_{k=1}^n X^\star_{kp} W_{nk}(Y_2) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n X^\star_{k1} W_{nk}(Y_n) & \sum_{k=1}^n X^\star_{k2} W_{nk}(Y_n) & \cdots & \sum_{k=1}^n X^\star_{kp} W_{nk}(Y_n) \end{bmatrix}$$

$$= \begin{bmatrix} W_{n1}(Y_1) & W_{n2}(Y_1) & \cdots & W_{nn}(Y_1) \\ W_{n1}(Y_2) & W_{n2}(Y_2) & \cdots & W_{nn}(Y_2) \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1}(Y_n) & W_{n2}(Y_n) & \cdots & W_{nn}(Y_n) \end{bmatrix} \begin{bmatrix} X^\star_{11} & X^\star_{12} & \cdots & X^\star_{1p} \\ X^\star_{21} & X^\star_{22} & \cdots & X^\star_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X^\star_{n1} & X^\star_{n2} & \cdots & X^\star_{np} \end{bmatrix},$$

$$\widehat{\mathbf{X}}^\star_n = \widehat{\mathbf{M}}(Y) = \mathbf{W}_n \mathbf{X}^\star_n. \tag{5}$$

The estimate of $\mathrm{E}\,(\mathbf{X}^\star_n|Y)$ is linear in $\mathbf{X}^\star_n$ and $\mathbf{W}_n$ is a matrix of weights that depends only on Y. Smoothers of this type are called linear and include many regression fitting techniques in the literature. For example, specific choices of weight sequences result in least squares, kernel, splines, nearest neighbor, orthogonal series and local polynomial smoothers.

Let $d = \dim(S_{\mathrm{E}\,(\mathbf{X}|Y)})$. Obviously, $d \leq p = \dim(\mathbf{X}) = \dim(\mathbf{X}^\star)$. Fix q so that $q \geq p$ and consider estimating $\mathrm{E}\,(\mathbf{X}^\star|Y)$ at q of the n Y-values using the same smoothers for all p $X^\star$-components. The resulting $\mathbf{W}_n$ is a $q \times n$ weight matrix, so that $\widehat{\mathbf{X}}^\star_n = \mathbf{W}_n \mathbf{X}^\star_n$ is a $q \times p$ matrix estimate of the conditional mean of $\mathbf{X}^\star$.

4. An Asymptotic Test for Dimension

Theorem (3) in the Appendix will be used to compute the asymptotic distribution of the linear smoother of the regression function. Let $e_i^{(n)}$ be the n-vector with one in the ith place and zeroes elsewhere. Also, let

$$\begin{aligned}\boldsymbol{\xi}_n &= \boldsymbol{H}_n^{-1/2}\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star - \mathbf{W}_n\mathbf{M}(Y)) = \boldsymbol{H}_n^{-1/2}(\boldsymbol{I}_p\otimes\mathbf{W}_n)\mathrm{vec}(\mathbf{E}_n)\\ &= \boldsymbol{H}_n^{-1/2}\sum_{i=1}^n(\boldsymbol{I}_p\otimes\mathbf{W}_n e_i^{(n)})\epsilon_i^{(n)} = \boldsymbol{H}_n^{-1/2}\sum_{i=1}^n \boldsymbol{A}_{in}\epsilon_i^{(n)},\end{aligned}$$

where $\boldsymbol{H}_n$ is the covariance matrix of $\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star)$, and

$$\mathbf{A}_{in} = \mathbf{I}_p\otimes\mathbf{W}_n e_i^{(n)} \qquad i=1,\ldots,n. \tag{6}$$

So far, we have made no other assumptions about $\boldsymbol{\Sigma}_{x|y}$ except that it is a $p\times p$ positive definite matrix. The development that follows depends on whether the covariance structure of the error vector depends on Y, that is, on whether the errors are homoscedastic or heteroscedastic. The two cases will be considered separately.

4.1. $\Sigma_{x|y}$ is constant

If $\boldsymbol{\Sigma}_{x|y}$ does not depend on Y, then $\mathrm{Cov}\,(\mathbf{E}_n) = \boldsymbol{\Sigma}_{x|y}\otimes\mathbf{I}_n$. Therefore

$$\begin{aligned}\mathbf{H}_n &= \mathrm{Cov}\,(\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star)) = \mathrm{Cov}\,((\mathbf{I}_p\otimes\mathbf{W}_n)\mathrm{vec}(\mathbf{X}_n^\star))\\ &= (\boldsymbol{I}_p\otimes\mathbf{W}_n)\,\mathrm{Cov}\,(\mathbf{X}_n^\star)(\boldsymbol{I}_p\otimes\mathbf{W}_n^T) = \boldsymbol{\Sigma}_{x|y}\otimes\mathbf{W}_n\mathbf{W}_n^T.\end{aligned} \tag{7}$$

Then, by Theorem 3 in the Appendix, we have that $\boldsymbol{\xi}_n \xrightarrow{\mathcal{D}} N_{pq}(0,\boldsymbol{I}_p\otimes\boldsymbol{I}_q)$ if $\max_{1\le i\le n}\mathrm{trace}[\boldsymbol{A}_{in}^T(\boldsymbol{A}_n\boldsymbol{A}_n^T)^{-1}\boldsymbol{A}_{in}] \xrightarrow[n\to\infty]{} 0$, where $\boldsymbol{A}_{in}$ is defined in (6). But,

$$\begin{aligned}&\max_{1\le i\le n}\mathrm{trace}[\boldsymbol{A}_{in}^T(\boldsymbol{A}_n\boldsymbol{A}_n^T)^{-1}\boldsymbol{A}_{in}]\\ &= \max_{1\le i\le n}\mathrm{trace}[(\boldsymbol{I}_p\otimes e_i^{(n)T}\mathbf{W}_n^T)(\boldsymbol{I}_p\otimes\mathbf{W}_n\mathbf{W}_n^T)^{-1}(\boldsymbol{I}_p\otimes\mathbf{W}_n e_i^{(n)})]\\ &= \max_{1\le i\le n}\mathrm{trace}(\boldsymbol{I}_p\otimes e_i^{(n)T}\mathbf{W}_n^T(\mathbf{W}_n\mathbf{W}_n^T)^{-1}\mathbf{W}_n e_i^{(n)})\\ &= p\max_{1\le i\le n}\mathrm{trace}(e_i^{(n)T}\mathbf{W}_n^T(\mathbf{W}_n\mathbf{W}_n^T)^{-1}\mathbf{W}_n e_i^{(n)}).\end{aligned}$$

Therefore, provided $\max_{1\le i\le n}\mathrm{trace}[e_i^{(n)T}\mathbf{W}_n^T(\mathbf{W}_n\mathbf{W}_n^T)^{-1}\mathbf{W}_n e_i^{(n)}] \xrightarrow[n\to\infty]{} 0$, we have $\boldsymbol{H}_n^{-1/2}\,\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0,\boldsymbol{I}_p\otimes\boldsymbol{I}_q)$. Consider the asymptotic distribution of $n^\lambda\,\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star - \mathbf{W}_n\mathbf{M}(Y))$, for some $\lambda>0$. The exponent λ signifies the rate of convergence of $\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star - \mathbf{W}_n\mathbf{M}(Y))$ and depends on the smoothers that are used in the estimating procedure. Let $\widetilde{\boldsymbol{H}}_n =$

$\mathrm{Cov}\,(n^{\lambda}\mathbf{W}_n\mathbf{X}_n^{\star}) = n^{2\lambda}\ (\mathbf{\Sigma}_{x|y} \otimes \mathbf{W}_n\mathbf{W}_n^T)$ and assume it has a positive definite limit matrix $\widetilde{\boldsymbol{H}}$. Then, $\widetilde{\boldsymbol{H}}_n^{-1/2}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \boldsymbol{I}_p \otimes \boldsymbol{I}_q)$ if and only if $n^{\lambda}\,\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \widetilde{\boldsymbol{H}})$. For any $m \times l$ matrix $\boldsymbol{A}$, define $\|\boldsymbol{A}\|_{\max} = \max|a_{ij}|$ over all $1 \le i \le m$, $1 \le j \le l$.

Lemma 1. *If* $\mathbf{G}_n = n^{2\lambda}\,\mathbf{W}_n\mathbf{W}_n^T \xrightarrow[n\to\infty]{} \boldsymbol{G}$, *where* $\mathbf{G}$ *is a* $q \times q$ *positive definite matrix, then* $n^{\lambda}\,vec(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \mathbf{\Sigma}_{x|y} \otimes \mathbf{G})$ *provided*

$$\|\mathbf{W}_n\|_{\max} = o(n^{-\lambda}) \tag{8}$$

and Conditions (II) *and* (III) *of Theorem* 3 *hold.*

Proof. By Theorem 3 we have that a sufficient condition for the result to hold is $\max_{1\le i\le n}\mathrm{trace}[\boldsymbol{A}_{in}^T(\boldsymbol{A}_n\boldsymbol{A}_n^T)^{-1}\boldsymbol{A}_{in}] \longrightarrow 0$ as $n \to \infty$ which is equivalent to $\max_{1\le i\le n}\|\boldsymbol{A}_{in}\|_{\max}^2 \longrightarrow 0$ by Lemma 2.4.2 in Bunke and Bunke ((1986), p.96). This is in turn equivalent to $\max_{1\le i\le n}\|n^{\lambda}\,\mathbf{W}_n e_i^{(n)}\|_{\max} \longrightarrow 0$ or, $\|\mathbf{W}_n\|_{\max} = o(n^{-\lambda})$.

Recall that $\mathbf{W}_n$ is a $q \times n$ matrix of rank q, which was assumed to be greater than or equal to p. Observe that $\mathrm{rank}(\mathbf{W}_n\mathbf{M}(Y)) \le \min(\mathrm{rank}(\mathbf{W}_n), \mathrm{rank}(\mathbf{M}(Y)) = \mathrm{rank}(\mathbf{M}(Y))$. Therefore, $d = \mathrm{rank}(\mathbf{M}(Y)) \ge \mathrm{rank}(\mathbf{W}_n\mathbf{M}(Y))$, and the estimate of $\mathbf{M}(Y)$, $\widehat{\mathbf{X}}_n^{\star} = \mathbf{W}_n\mathbf{X}_n^{\star}$, can be used to estimate a lower bound on d. Note that $\mathrm{rank}(\mathbf{W}_n\mathbf{M}(Y)) = \mathrm{rank}(\mathbf{M}(Y))$ if $S(\mathbf{M}(Y)) \cap N(\mathbf{W}_n) = \{0\}$, where $N(\mathbf{W}_n)$ is the null space of $\mathbf{W}_n$ (see Harville (1997, Thm 17.5.4)). Consequently, to ensure that the estimate of $\mathbf{M}(Y)$ yields an estimate of d, we may require that $N(\mathbf{W}_n) = \{0\}$. Furthermore, since $\mathbf{W}_n$ is a matrix that depends only on Y, in the inverse regression context where the conditioning is on Y, $\mathbf{W}_n\mathbf{M}(Y)$ is a fixed matrix and the Eaton and Tyler (1994) result can be used to obtain the asymptotic distribution of the singular values of $\mathbf{W}_n\mathbf{X}_n^{\star}$.

To ease the computation of the asymptotic distribution of the singular values of $\mathbf{W}_n\mathbf{X}_n^{\star}$, $n^{\lambda}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y))$ will be pre- and post-multiplied by convenient choices of nonsingular matrices. Assume that $\boldsymbol{G}_n = n^{2\lambda}\,\mathbf{W}_n\mathbf{W}_n^T \xrightarrow[n\to\infty]{} \boldsymbol{G}$, where $\mathbf{G}$ is positive definite and let $\widehat{\mathbf{\Sigma}}_{x|y}$ be a consistent estimate of $\mathbf{\Sigma}_{x|y}$. Then, the multivariate version of Slutsky's theorem and Lemma 1 yield

$$n^{\lambda}\,(\widehat{\mathbf{\Sigma}}_{x|y}^{-1/2} \otimes \boldsymbol{G}_n^{-1/2})\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \mathbf{I}_p \otimes \mathbf{I}_q). \tag{9}$$

The mean $\mathrm{E}\,(\mathbf{X})$ is usually unknown, so we let $\mathrm{vec}(\tilde{\mathbf{X}}_n) = (\widehat{\mathbf{\Sigma}}_{x|y}^{-1/2} \otimes \boldsymbol{G}_n^{-1/2})\times \mathrm{vec}(\mathbf{W}_n\hat{\mathbf{X}}_n)$, where $\hat{\mathbf{X}}_n = \mathbf{X}_n - \bar{\mathbf{X}}_n$, with $\bar{\mathbf{X}}_n = \mathbf{1}_n\bar{\mathbf{X}}^T$. Clearly, $\mathrm{rank}(\tilde{\mathbf{X}}_n) = \mathrm{rank}(\mathbf{W}_n\hat{\mathbf{X}}_n)$, and inference on d can be based on $\tilde{\mathbf{X}}_n$. The testing procedure

uses the test statistic

$$\Lambda_j = n^{2\lambda} \sum_{i=j+1}^{\min(p,q)} l_i^2, \tag{10}$$

where l_i, $i = 1, \ldots, \min(q, p)$, denote the singular values of $\tilde{\mathbf{X}}_n$. The choice of Λ_j is motivated by the fact that $\operatorname{rank}(\boldsymbol{G}_n^{-1/2}(\mathbf{W}_n\mathbf{M}(Y))\widehat{\boldsymbol{\Sigma}}_{x|y}^{-1/2}) \leq \operatorname{rank}(\mathbf{M}(Y)) = d$, and by the simplicity of the asymptotic covariance structure in (9). The asymptotic distribution of Λ_j is given in the following theorem which is proved in the appendix.

Theorem 1. *Assume all conditions of Lemma* 1 *hold. Let* $\widehat{\boldsymbol{\Sigma}}_{x|y}$ *be a consistent estimate of* $\boldsymbol{\Sigma}_{x|y}$. *If* $k = \mathit{rank}(\widehat{\mathbf{G}}_n^{-1/2}(\mathbf{W}_n\mathbf{M}(Y))\widehat{\boldsymbol{\Sigma}}_{x|y}^{-1/2}) \leq \mathit{rank}(\mathbf{M}(Y)) = d$, *then* Λ_k, *as defined in* (10), *is asymptotically distributed as a* $\chi^2_{(p-k)(q-k)}$ *random variable.*

The asymptotic distribution of Λ_k can be used to estimate a lower bound on the rank d of $\mathbf{M}(Y)$, which is also the dimension of the subspace $S_{\mathrm{E}(\mathbf{X}|Y)}$. Furthermore, if the weights are selected so that $N(\mathbf{W}_n) = \{0\}$, the test yields an estimate of d. The computation of a consistent estimate of $\boldsymbol{\Sigma}_{x|y}$ is deferred to Section 4.3.

The d eigenvectors of the nonparametric estimate of $\mathbf{M}(Y)$ that correspond to its d largest eigenvalues, yield estimates of d of the basis vectors of $\boldsymbol{\Sigma}_x S_{\mathbf{Y}|\mathbf{X}}$. They can be scaled back to estimates of basis vectors of the central dimension-reduction subspace for the uncentered $\mathbf{X}$ through multiplication by $\hat{\boldsymbol{\Sigma}}_x^{-1}$ on the left.

4.2. $\boldsymbol{\Sigma}_{x|y}$ depends on Y

Assume that $\boldsymbol{\Sigma}_{x|y} = (\sigma_{ij}(Y))_{i,j=1}^p = \Sigma_{x|y}(Y)$ is a $p \times p$ matrix of continuous functions of Y. In this case the covariance structure of $\mathbf{X}_n|Y$, or equivalently the covariance structure of the error matrix, can no longer be represented by the Kronecker product of $\boldsymbol{\Sigma}_{x|y}$ and $\mathbf{I}_q$, since $\operatorname{Var}(X_{jj}|Y = Y_i) = \sigma_{jj}(Y_i)$, $\operatorname{Cov}(X_{ij}, X_{ik}|Y = Y_i) = \sigma_{jk}(Y_i)$ for $i = 1, \ldots, n$, $j, k = 1, \ldots, p$, $j \neq k$. The covariance matrix of $\mathbf{X}_n|Y$ is a $np \times np$ symmetric matrix of p^2 blocks of order $n \times n$, where the ijth block is given by

$$\operatorname{diag}(\sigma_{ij}(Y_1), \sigma_{ij}(Y_2), \ldots, \sigma_{ij}(Y_n)) = \begin{bmatrix} \sigma_{ij}(Y_1) & 0 & \cdots & 0 \\ 0 & \sigma_{ij}(Y_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij}(Y_n) \end{bmatrix}$$

for $i,\ j = 1, \ldots, p$. The covariance matrix $\widetilde{\mathbf{H}}_n$ of $n^{\lambda}\ \mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^{\star})$ is a $qp \times qp$ block matrix, whose ijth block is given by

$$n^{2\lambda}\ \mathbf{W}_n \mathrm{diag}(\sigma_{ij}(Y_1), \sigma_{ij}(Y_2), \ldots, \sigma_{ij}(Y_n))\mathbf{W}_n^T \tag{11}$$

for $i,\ j = 1, \ldots, p$. The following lemma is a direct analogue to Lemma 1 in the constant covariance case.

Lemma 2. *Suppose that $\mathbf{\Sigma}_{x|y}(Y_n) \to \mathbf{\Sigma}_{x|y}$, as $n \to \infty$, for all Y_n in the Y-sample space, where $\mathbf{\Sigma}_{x|y}$ is non-singular. If $n^{2\lambda}\mathbf{W}_n\mathbf{W}_n^T$ has a positive definite limit matrix $\mathbf{G} = (g_{lm})$, then $\widetilde{\mathbf{H}}_n$ has a positive definite limit matrix $\widetilde{\mathbf{H}}$, and n^{λ} vec$(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \widetilde{\mathbf{H}})$ provided $\|\mathbf{W}_n\|_{\max} = o(n^{-\lambda})$ and Conditions* (II) *and* (III) *of Theorem* 3 *hold.*

Proof. Since $\sigma_{ij}(Y_k) \to \sigma_{ij}$ as $k \to \infty$, we have that for all $\epsilon > 0$ there exists k_0 such that for all $k \geq k_0$, $\sigma_{ij} - \epsilon < \sigma_{ij}(Y_k) < \sigma_{ij} + \epsilon$. Without loss of generality, we can assume that $\sum_{k=1}^{\infty} nW_{lk}W_{mk} = g_{lm} > 0$ (the development is analogous for the case $g_{lm} < 0$; when $g_{lm} = 0$, or when $\sigma_{ij} = 0$, the limit is 0). Then, for a sufficiently large $k_1 \geq k_0$ we have

$$(\sigma_{ij} - \epsilon) \sum_{k \geq k_1} nW_{lk}W_{mk} < \sum_{k \geq k_1} \sigma_{ij}(\mathbf{Y}_k) nW_{lk}W_{mk} < (\sigma_{ij} + \epsilon) \sum_{k \geq k_1} nW_{lk}W_{mk}.$$

There are two cases: (i) $\sigma_{ij} > 0$ and (ii) $\sigma_{ij} < 0$. Case (i) is equivalent to $\sigma_{ij} - \epsilon > 0$, and similarly (ii) is equivalent to $\sigma_{ij} + \epsilon < 0$, for sufficiently small ϵ. Therefore, for case (i) and all $\epsilon > 0$ with $\min(g_{lm}, \sigma_{ij}) > \epsilon$, there exists $k_2 \geq k_1$ such that

$$(\sigma_{ij} - \epsilon)(g_{lm} - \epsilon) \leq \sum_{k \geq k_2} \sigma_{ij}(\mathbf{Y}_k) nW_{lk}W_{mk} \leq (\sigma_{ij} + \epsilon)(g_{lm} + \epsilon) \tag{12}$$

or, equivalently (since all products of ϵ are negligible), $\sum_k^n \sigma_{ij}(Y_k) nW_{lk}W_{mk} \longrightarrow \sigma_{ij}g_{lm}$, as $n \to \infty$. Case (ii) is analogous to case (i) with the inequalities in (12) reversed.

Therefore, the ijth block of $\widetilde{\mathbf{H}}_n$ satisfies $n\mathbf{W}_n \mathrm{diag}(\sigma_{ij}(Y_1), \ldots, \sigma_{ij}(Y_n))\mathbf{W}_n^T \to \sigma_{ij}\mathbf{G}$, which yields that $\widetilde{\boldsymbol{H}}_n \to \mathbf{\Sigma}_{x|y} \otimes \mathbf{G} = \widetilde{\boldsymbol{H}}$. Since the Kronecker product of two positive definite matrices is positive definite (see Harville (1997), p.369), $\widetilde{\boldsymbol{H}}$ is positive definite. By a direct application of Slutsky's theorem, we obtain that $\widetilde{\boldsymbol{H}}_n^{-1/2} \mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \boldsymbol{I}_p \otimes \mathbf{I}_q)$ if and only if $n^{\lambda}\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^{\star} - \mathbf{W}_n\mathbf{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \widetilde{\boldsymbol{H}})$ and the rest of the proof is the same as the proof of Lemma 1.

Let $\widehat{\mathbf{\Sigma}}_{x|y}(Y)$ be a weakly consistent estimate of $\mathbf{\Sigma}_{x|y}(Y)$, for all Y in the relevant sample space. Let $\widetilde{\boldsymbol{H}}_n$ be a $qp \times qp$ matrix whose ijth block is $n\mathbf{W}_n \times$

$\mathrm{diag}(\hat{\sigma}_{ij}(Y_1), \hat{\sigma}_{ij}(Y_2), \ldots, \hat{\sigma}_{ij}(Y_n)\mathbf{W}_n^T$. It is easy to see that $\widehat{\boldsymbol{H}}_n$ is also weakly consistent for $\tilde{\boldsymbol{H}}$. By Slutsky's theorem we have $\sqrt{n}\,\widehat{\boldsymbol{H}}_n^{-1/2}\mathrm{vec}(\mathbf{W}_n\mathbf{X}_n^\star - \mathbf{W}_n\boldsymbol{M}(Y)) \xrightarrow{\mathcal{D}} N_{pq}(0, \boldsymbol{I}_{pq} = \boldsymbol{I}_p \otimes \boldsymbol{I}_q)$, and d was not affected by this transformation. Additionally, Lemma 3 in the Appendix states the conditions under which $\widehat{\boldsymbol{H}}_n$ is L_2-consistent for $\tilde{\boldsymbol{H}}$.

As in the constant covariance case, let $\tilde{\mathbf{X}}_n = \widehat{\boldsymbol{H}}_n^{-1/2}\mathrm{vec}(\mathbf{W}_n\hat{\mathbf{X}}_n)$. The following theorem summarizes the results of the present section.

Theorem 2. *Assume that all conditions of Lemma* 3 *hold. If* $k = rank(\widehat{\boldsymbol{G}}_n^{-1/2} \times (\mathbf{W}_n\mathbf{M}(Y))\widehat{\Sigma}_{x|y}^{-1/2}) \leq rank(\mathbf{M}(Y)) = d$, *then the test statistic* Λ_k, *defined in* (*10*), *is asymptotically distributed as a* $\chi^2_{(p-k)(q-k)}$ *random variable.*

The inferential procedure on d is the same as in the constant covariance case, provided $\boldsymbol{\Sigma}_{x|y}(Y)$ can be estimated consistently.

4.3. A consistent nonparametric estimate of $\boldsymbol{\Sigma_{x|y}}$

Stone (1977) presented a method of obtaining consistent nonparametric estimates of moments of the conditional distribution of $\mathbf{X}$ given Y, based on convenient choices of weights. Using Stone's terminology (1977), consistent estimates originate from consistent weight sequences: a sequence of weights $\{W_n\}$ is consistent if whenever $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are i.i.d., and $\mathrm{E}\,\|X\|^r < \infty$, then $\hat{\mathrm{E}}_n(X|Y) \to \mathrm{E}\,(X|Y)$ in L_r, as $n \to \infty$.

Assume that $\mathrm{E}\,(X_j^2) < \infty$, for all $j = 1, \ldots, p$. Let

$$\begin{aligned}\hat{\sigma}_{ii}(Y) &= \hat{\mathrm{E}}_n(X_i^{\star 2}|Y) - \hat{\mathrm{E}}_n(X_i^\star|Y)^2 \\ &= \sum_{j=1}^n W_{nj}(Y)X_{ij}^{\star 2} - (\sum_{j=1}^n W_{nj}(Y)X_{ij}^\star)^2, \qquad (13)\\ \hat{\sigma}_{ij}(Y) &= \hat{\mathrm{E}}_n(X_i^\star X_j^\star|Y) - \hat{\mathrm{E}}_n(X_i^\star|Y)\hat{\mathrm{E}}_n(X_j^\star|Y) \\ &= \sum_{k=1}^n W_{nk}(Y)X_{ik}^\star X_{jk}^\star - (\sum_{k=1}^n W_{nk}(Y)X_{ik}^\star)(\sum_{k=1}^n W_{nk}(Y)X_{jk}^\star) \qquad (14)\end{aligned}$$

for $i \neq j$, $i,\ j \in \{1, \ldots, p\}$, where $\{\mathbf{W}_n\}$ produces L_2 consistent estimates. An easy application of the Cauchy-Schwartz inequality shows that both (13) and (14) are L_1-consistent and therefore weakly consistent estimates of the corresponding population moments.

If $\boldsymbol{\Sigma}_{x|y}$ is constant, let $\widehat{\boldsymbol{\Sigma}}_{x|y}$ be the $p \times p$ matrix with entries given by (13) and (14), computed at any point in the Y-sample space. If $\boldsymbol{\Sigma}_{x|y}$ depends on Y, let $\widehat{\boldsymbol{\Sigma}}_{x|y}(Y_k)$ be the $p \times p$ matrix with entries given by (13) and (14), computed at Y_k for $k = 1, \ldots, n$. Then, by the above discussion, $\widehat{\boldsymbol{\Sigma}}_{x|y}$ is a weakly consistent

estimate of $\mathbf{\Sigma}_{x|y}$. Of course, in order to obtain consistent estimates of second moments we need to use consistent weight sequences. Stone (1977) shows how to construct *k nearest neighbor (k-NN)* and *local linear* consistent estimates. He proves that general kernel smoothers are not necessarily consistent, but they can be made so by imposing further regularity conditions on the kernel function. An important drawback of Stone's construction of consistent weights is that it requires significant computational effort.

In addition to Stone's results, if the weights are such that the resulting estimates $\hat{\mathrm{E}}_n(.|Y)$ are uniformly bounded functions of Y, then the second moment estimates are L_2 consistent as shown in Lemma 4 in the Appendix. Also, according to Lemma 3 in the appendix, if $\hat{\sigma}_{ij}(Y_k)$ are L_2 consistent for $\sigma_{ij}(Y_k)$ for all $i,\ j = 1,\ldots,p,\ k = 1,\ldots,n$, and $\mathrm{Cov}\,(\hat{\sigma}_{ij}(Y_k), \hat{\sigma}_{ij}(Y_l)) \xrightarrow[n\to\infty]{} 0$ for all k, $l = 1,\ldots,n,\ l \neq k$, then $\widehat{\mathbf{H}}_n$, as defined in (11), is a L_2-consistent estimate of the asymptotic covariance of $n^{\lambda}\ \mathbf{W}_n\mathbf{X}_n^{\star}$.

5. Application: Local Linear Smoothers

Local polynomial fitting (Fan and Gijbels (1996)) leads to a linear smoother of the form (5). Local polynomial smoothers are selected because they possess a number of attractive properties among linear smoothers: they are design adaptive; they have appealing bias and variance performance; they do not need modification at the boundaries; they have best minimax efficiency among all linear smoothers.

Local polynomial fitting is based on the Taylor expansion of the regression function which is modelled locally by a simple polynomial model. The latter is fitted locally using a weighted least squares regression. Each observation is assigned a kernel weight that downweighs observations far from the point where the regression curve is estimated. Here we focus attention to local linear fitting, that is, we require that the inverse regression curve is a twice differentiable function. Let

$$\mathbf{Y} = \begin{pmatrix} 1 & (Y_1 - Y) \\ \vdots & \vdots \\ 1 & (Y_n - Y) \end{pmatrix},$$

$\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{nj})^T$. The local linear estimate of $m_j(Y)$ is given by

$$\hat{m}_j(Y) = \sum_{i=1}^{n} W_n(\frac{Y_i - Y}{h})X_{ij}, \tag{15}$$

where $W_n((Y_i - Y)/h) = \tilde{W}_{ni}(Y)$ in the notation of Section 3. The weights are given by $W_n(t) = (1,0)S_n^{-1}(1, th)^T K(t)/h$, where $S_n = (S_{n,j+l})_{0\le j,l\le 1}$, $S_{n,j} = \sum_{i=1}^{n} K_h(Y_i - Y)(Y_i - Y)^j$ (Fan and Gijbels (1996), Chapter 3). Here $K_h(\cdot) =$

$K(\cdot/h)/h$, where h is the bandwidth controlling the size of the local neighborhood, and K is a kernel function that is assumed to be a symmetric probability density function with bounded support. Fan and Gijbels (1996) also show that $W_n(t) = (1,0)S_n^{-1}(1,t)^T K(t)\{1 + o_P(1)\}/nhf(Y)$. From this and since we are conditioning on the observed Y_i's, the weights W_n in (15) satisfy condition (8) in probability with $\lambda = 1/2$, provided $\sqrt{n}h \to \infty$ as $h \to 0$ and $n \to \infty$. Also, h and n should vary jointly in such a way that $n^{2\lambda}W_n(t)W_n^T(t) = nW_n(t)W_n^T(t)$ has a positive definite limit.

In the simulation study of the following section, the first step of the LOWESS (Cleveland (1979)) smoothing procedure is used to provide a local linear smoother with a nearest neighbor bandwidth and the tricube kernel. The robustification steps are dropped from the calculations.

For computational simplicity, $\mathbf{\Sigma}_{x|y}$ is estimated according to Cleveland's (1979) proposal and not by constructing consistent weights as suggested in Section 4.3:

$$\widehat{\mathbf{\Sigma}}_{x|y} = \frac{(\hat{\mathbf{X}}_n - \mathbf{W}_n\hat{\mathbf{X}}_n)^T(\hat{\mathbf{X}}_n - \mathbf{W}_n\hat{\mathbf{X}}_n)}{\operatorname{tr}(\mathbf{I} - \mathbf{W}_n)(\mathbf{I} - \mathbf{W}_n)^T}.$$

5.1. Simulation results

The power of the chi-square test based on slicing and the chi-square test based on local linear smoothing are compared via a small simulation study. For all regression models to be considered in this section, three sample sizes are used: $n = 100$, 200, 400. For each sample size, the p-values corresponding to the test statistics for selected dimensions for both tests were collected over 1,000 replications. The matrix $\mathbf{W}_n$ of weights has full rank, so that its null space is zero. The LOWESS chi-square test is, then, a test of dimension for $S_{\mathrm{E}(\mathbf{X}|Y)}$.

Two models will be considered. For the first, the response Y is generated according to the model

$$Y = X_1 + X_2 + X_3 + X_4 + 0.5\epsilon, \tag{16}$$

where $X_1, \ldots, X_5, \epsilon$ are i.i.d. standard normal variates. Model (16) is (6.1) in Li (1991), with central subspace the one-dimensional $S_{Y|\mathbf{X}} = S((1,1,1,1,0)^T)$.

The symbols L_j and Λ_j signify the SIR chi-square test statistic (Li (1991)) and the LOWESS test statistic for testing $d = j$ versus $d \geq j + 1$, respectively. The numerical entries of the rows of Table 1, corresponding to the test statistics indexed by 0, are empirical estimates of the power of the corresponding test. They represent the proportion of times the corresponding null hypothesis $d = 0$ is rejected, when the nominal significance level is 0.05 and 0.01, indicated parenthetically. The entries for the test statistics indexed by 1 are empirical estimates of the size of the test. The symbol H stands for number of slices, and

q for the number of randomly selected points at which the conditional inverse mean $\mathrm{E}(\mathbf{X}|Y)$ is estimated by the first step of the LOWESS smoother. The results for only one value of the smoothing parameter are reported. The value of the smoothing parameter signifies the proportion of the data used for the estimation at each data point. The values were selected by eyeballing the five inverse regression plots, without the help of any automatic bandwidth selection techniques. Computations were carried out in *Arc* (Cook and Weisberg (1999)), a regression package written in the *Xlisp-Stat* language (Tierney (1990)). The software can be obtained from the author upon request.

Table 1 indicates that both tests correctly estimate the dimension to be one across sample size, number of slices and q. Also, both chi-square tests have similar power with the SIR chi-square test performing better when the number of slices, or equivalently q, is larger. Additional non-reported simulations indicate the level of the LOWESS chi-square test depends on both the choice of smoothing parameter and q. Of course, this is not surprising as the LOWESS chi-square test for dimension requires the selection of both the bandwidth and number of points q where the inverse regression curves are estimated. SIR, on the other hand, requires the choice of only one tuning parameter, i.e., the number of slices.

In Table 2 the power and size results for the model

$$Y = (4 + X_1)(2 + X_2 + X_3) + 0.5\epsilon \tag{17}$$

Table 1. Empirical Power and Size for the SIR and the LOWESS chi-square tests applied to (16).

$\mathbf{X} \sim N_5(0, \mathbf{I}_5)$, bandwidth= 0.8

$n = 100$							
SIR chi-square				LOWESS chi-square			
H	6	10	15	q	6	10	15
L_0	1 (1)	1 (1)	1 (.997)	Λ_0	.994 (.959)	.967 (.790)	.529 (.191)
L_1	.054 (.01)	.036 (.008)	.028 (.002)	Λ_1	.047 (.009)	.037 (.004)	.031 (.006)
$n = 200$							
SIR chi-square				LOWESS chi-square			
H	10	20	30	q	10	20	30
L_0	1 (1)	1 (1)	1 (1)	Λ_0	1 (1)	1 (.993)	.941(.757)
L_1	.043(.007)	.039(.009)	.033(.007)	Λ_1	.056(.01)	.033(.01)	.021(.002)
$n = 400$							
SIR chi-square				LOWESS chi-square			
H	20	30	40	q	20	30	40
L_0	1 (1)	1 (1)	1 (1)	Λ_0	1 (1)	1 (1)	1(1)
L_1	.052 (.005)	.048 (.011)	.045 (.005)	Λ_1	.04 (.005)	.028 (.007)	.01 (.001)

are tabulated. The error term ϵ is a standard normal variate and

$$X_1 = W_1,\ X_2 = V_1 + \frac{W_2}{2},\ X_3 = -V_1 + \frac{W_2}{2},\ X_4 = V_2 + V_3,\ X_5 = V_2 - V_3. \quad (18)$$

The only restriction placed on $\boldsymbol{V}$ and $\boldsymbol{W}$ is that they be independent. The variables V_1, V_2, V_4 are i.i.d. $t_{(4)}$, $V_3 \sim t_{(3)}$, $V_5 \sim t_{(5)}$, and W_1, W_2 are i.i.d. Gamma(0.25) random variables. The row entries of Table 2 are to be interpreted in a way analogous to Table 1. Model (17) was used by Velilla (1998). The central subspace is two-dimensional and the joint predictor distribution (18) satisfies the linearity condition by construction (see Velilla (1998), p.1092-93), even though $\mathbf{X}$ does not have an elliptically contoured distribution. Also, the conditional variance of the predictors given the response is non-constant, as can be seen from a scatterplot matrix (not shown) of the simulated data.

In this case, the LOWESS chi-square test outperforms the SIR chi-square test both with respect to power and size. SIR tends to miss the second dimension much more often. These results are in accordance with the theory, as the predictor distribution is not normal and there is significant heteroskedasticity in the conditional distribution of $\mathbf{X}|Y$. Still, the benefits of continuous local linear fitting appear to yield as the sample size and q increase. This phenomenon was also observed in the first example.

Table 2. Empirical Power and Size for the SIR and the LOWESS chi-square tests applied to (17).

$\boldsymbol{X}$ distributed as in (18), bandwidth= 0.8

$n = 100$							
SIR chi-square				LOWESS chi-square			
H	6	10	15	q	6	10	15
L_0	1(1)	1(1)	1(1)	Λ_0	.989 (.953)	.967 (.873)	.849(.707)
L_1	.054 (.014)	.074 (.028)	.112 (.046)	Λ_1	.277 (.144)	.27 (.129)	.27 (.13)
L_2	.001 (0)	.004 (0)	.006 (.002)	Λ_2	.015 (.001)	.018 (.003)	.019 (.003)
$n = 200$							
SIR chi-square				LOWESS chi-square			
H	10	20	30	q	10	20	30
L_0	1 (1)	1 (1)	1(1)	Λ_0	1(.999)	.988(.978)	.915(.819)
L_1	.133 (.051)	.134 (.066)	.164 (.081)	Λ_1	.366 (.232)	.35 (.218)	.248 (.125)
L_2	.004 (0)	.007 (.001)	.008 (.002)	Λ_2	.02 (.005)	.016 (.004)	.008 (.002)
$n = 400$							
SIR chi-square				LOWESS chi-square			
H	20	30	40	q	20	30	40
L_0	1 (1)	1(1)	1(1)	Λ_0	1(1)	.999(.998)	.996(.99)
L_1	.24 (.136)	.265 (.166)	.236 (.129)	Λ_1	.454 (.324)	.279 (.179)	.192 (.121)
L_2	.008 (.002)	.012 (.002)	.011 (.001)	Λ_2	.011 (.005)	0(0)	0(0)

In addition to the results reported here, a number of other choices for bandwidth were considered. The conclusions were consistent with the above remarks. This limited simulation-based comparison of the two tests serves only to illustrate that both tests agree and compare well power-wise. There is more to be investigated with respect to optimal choices of bandwidth and q. In addition, the q points in this simulation study were randomly drawn from n available ones. It could be argued that the sampling scheme should try to reflect local trends in the data more accurately. For example, more densely data-populated areas should be sampled at a higher rate than others.

6. Discussion

Li (1991) proposed the SIR chi-square test procedure for multivariate regression analysis problems when $\mathbf{X}$ is normal. SIR is based on a regressogram-type nonparametric fitting of the inverse regression curves. In this paper, the continuous nature of the data is taken into account, and the nonparametric estimation technique is extended to include all linear smoothers. The regressor distribution is only required to have finite second moments. An asymptotic chi-square test for dimension is obtained, as well as an estimate of the central dimension reduction subspace, or a portion thereof. The result extends to heteroskedastic data.

The extension to smoothing comes at the price of increased complexity. The smoother must satisfy certain regularity conditions. Also, two tuning parameters need to be selected by the user: the bandwidth and the number of points where the estimation takes place. The latter is directly analogous to the number of slices in SIR. Optimal choice of number of points and their allocation, without affecting the estimation of dimension, is an open problem. Nevertheless, the simulation study seems to indicate that the estimation is fairly robust across different choices of number of points. On the other hand, optimal choice of bandwidth is an extensively explored issue in nonparametric curve fitting and many options are available.

Acknowledgements

This work was completed during my stay at the Department of Statistics, Stanford University, in the summer of 2000. I would like to thank Ingram Olkin and the National Science Foundation (DMS-9631278) for giving me the opportunity to visit the department. Also, I thank two anonymous referees for their helpful comments.

Appendix

Let $\mathcal{F}$ be a non-empty set of p-dimensional distribution functions with 0 mean and positive definite covariance matrix. Let $\mathcal{M}_p^{>}$ be the space of all $p \times p$

positive definite matrices and $\mathcal{M}(\mathcal{F}) = \{ \int_{R^p} xx^T dF(x) : F \in \mathcal{F}\} \subset \mathcal{M}_p^>$. Let $\{\boldsymbol{\epsilon}_i^{(n)}\}_{i=1,\ldots,n,\ n=1,\ldots}$ be an array of random p-vectors whose distribution functions belong to $\mathcal{F}$. For all n, let $\boldsymbol{\epsilon}_1^{(n)}, \boldsymbol{\epsilon}_2^{(n)}, \ldots, \boldsymbol{\epsilon}_n^{(n)}$ be independent and set $\mathbf{E}_n = (\boldsymbol{\epsilon}_1^{(n)}, \ldots, \boldsymbol{\epsilon}_n^{(n)})^T$.

Assume that $\{\boldsymbol{A}_n\} = \{(\boldsymbol{A}_{1n}, \ldots, \boldsymbol{A}_{nn})\}$, $n = 1, \ldots$ is a sequence of non-stochastic $s \times pn$-matrices with $\boldsymbol{A}_{in} \in \mathcal{M}_{s\times p}$, and $\operatorname{rank}(\boldsymbol{A}_{in}) = s$, $i = 1, \ldots, n$. Let $\mathbf{H}_n = \sum_{i=1}^n \boldsymbol{A}_{in} \operatorname{Cov}(\epsilon_i^{(n)})\, \boldsymbol{A}_{in}^T$ denote the positive definite covariance matrix of $\sum_{i=1}^n \boldsymbol{A}_{in}\epsilon_i^{(n)}$ and $\boldsymbol{\xi}_n = \mathbf{H}_n^{-1/2} \sum_{i=1}^n \boldsymbol{A}_{in}\epsilon_i^{(n)}$.

The following is a central limit theorem providing conditions under which $\boldsymbol{\xi}_n$ is asymptotically normal (Bunke and Bunke (1986), Theorem 2.4.3). It is a generalization of a theorem proved by Eicker (1966). Let $\lambda_{min}[\boldsymbol{\Sigma}]$ denote the minimum eigenvalue of a symmetric matrix $\boldsymbol{\Sigma}$. The notation $\|\boldsymbol{x}\|$, for $\boldsymbol{x} \in \mathbb{R}^p$, is used to denote the Euclidean norm on $\mathbb{R}^p$.

Theorem 3. [Bunke and Bunke, 1986] *The conditions*

(I) : $\max_{1\le i\le n} \operatorname{tr}[\boldsymbol{A}_{in}^T(\boldsymbol{A}_n\boldsymbol{A}_n^T)^{-1}\boldsymbol{A}_{in}] \longrightarrow 0$ *as* $n \to \infty$,

(II) : $\sup_{F\in\mathcal{F}} \int_{\|x\|>c} \|x\|^2 dF(x) \longrightarrow 0$ *as* $c \to \infty$,

(III) : $\inf_{\Sigma\in\mathcal{M}(\mathcal{F})} \lambda_{min}[\boldsymbol{\Sigma}] \ge r > 0$,

are sufficient for all sequences $\{\epsilon_i^{(n)}\}$, *with* $\epsilon_i^{(n)} \sim \mathcal{F}$, *to fulfill*

A: $\boldsymbol{\xi}_n \xrightarrow{\mathcal{D}} N_s(0, I_s)$ *as* $n \to \infty$,

B: $\max_{1\le i\le k_n} P(\|\mathbf{H}_n^{-1/2}\boldsymbol{A}_{in}\epsilon_i^{(n)}\| > \delta) \longrightarrow 0$ *for all* $\delta > 0$.

Proof of Theorem 1. Consider the singular value decomposition

$$\mathbf{G}^{-1/2}\mathbf{W}_n\mathbf{M}(Y)\boldsymbol{\Sigma}_{x|y}^{-1/2} = \boldsymbol{\Gamma}_1^T \begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\Gamma}_2^T,$$

$\boldsymbol{D}$ a $k \times k$ diagonal matrix with the positive singular values of $\mathbf{G}^{-1/2}\mathbf{W}_n\mathbf{M}(Y)\times \boldsymbol{\Sigma}_{z|y}^{-1/2}$ along its main diagonal. Partition the $q\times q$ matrix $\boldsymbol{\Gamma}_1^T$ as $(\boldsymbol{\Gamma}_{11}, \boldsymbol{\Gamma}_{12})$, where $\boldsymbol{\Gamma}_{11}$ is $q \times k$ and $\boldsymbol{\Gamma}_{12}$ is $q \times (q-k)$; partition the $p \times p$ matrix $\boldsymbol{\Gamma}_2^T$ as $(\boldsymbol{\Gamma}_{21}, \boldsymbol{\Gamma}_{22})^T$, where $\boldsymbol{\Gamma}_{21}^T$ is $k \times p$, and $\boldsymbol{\Gamma}_{22}^T$ is $(p-k) \times p$. By the Eaton-Tyler result (Eaton and Tyler (1994)) about the asymptotic distribution of the singular values of a matrix, the limiting distribution of the smallest $\min(q-k, p-k)$ singular values of $\sqrt{n}\ (\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{X}_n\hat{\boldsymbol{\Sigma}}_{z|y}^{-1/2})$ is the same as the limiting distribution of the singular

values of the $(q-k)\times(p-k)$ matrix $\sqrt{n}\ (\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{X}_n\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\mathbf{\Gamma}_{22})$. By (9) we have that

$$\sqrt{n}\ \mathrm{vec}(\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{X}_n\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\mathbf{\Gamma}_{22}) \xrightarrow{\mathcal{D}} N_{(p-d)(q-d)}(0, \mathbf{I}_{p-k}\otimes \boldsymbol{I}_{q-k}). \tag{19}$$

Observe that

$$\begin{aligned}&\sqrt{n}\ \mathrm{vec}(\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\hat{\mathbf{X}}_n\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\\ &=\sqrt{n}\ \mathrm{vec}(\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}(\mathbf{W}_n\hat{\mathbf{X}}_n-\mathbf{W}_n\mathbf{M}(Y))\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\mathbf{\Gamma}_{22})\\ &\quad+\sqrt{n}\ \mathrm{vec}(\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{M}(Y)\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\mathbf{\Gamma}_{22}).\end{aligned} \tag{20}$$

The second term of (20) is going to 0 by Slutsky's theorem and the singular value decomposition of $\mathbf{W}_n\mathbf{M}(Y)$, since $\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{M}(Y)\xrightarrow[n\to\infty]{}\mathbf{\Gamma}_{12}^T\mathbf{G}^{-1/2}\mathbf{W}_n\mathbf{M}(Y)$ $=0$. Also, observe that

$$\begin{aligned}\sqrt{n}(\mathbf{W}_n\hat{\mathbf{X}}_n-\mathbf{W}_n\mathbf{M}(Y))&=\sqrt{n}(\mathbf{W}_n(\mathbf{X}_n+\mathrm{E}\,(\mathbf{X}_n)-\bar{\mathbf{X}})-\mathbf{W}_n\mathbf{M}(Y))\\ &=\sqrt{n}(\mathbf{W}_n\mathbf{X}_n-\mathbf{W}_n\mathbf{M}(Y))-\sqrt{n}\mathbf{W}_n(\bar{\mathbf{X}}-\mathrm{E}\,(\mathbf{X}_n)).\end{aligned} \tag{21}$$

By (8) and the fact that the sample mean is strongly consistent for the distribution mean, the second term in (21) goes to zero. Therefore, the first term of (20) is going to the distribution indicated in (19). Consequently, Λ_k has the same asymptotic distribution as the sum of the squares of the singular values of $\sqrt{n}\ (\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\mathbf{W}_n\mathbf{X}_n\hat{\mathbf{\Sigma}}_{z|y}^{-1/2}\mathbf{\Gamma}_{22})$, which is $\chi^2_{(p-k)\times(q-k)}$ by (19).

Lemma 3. *Suppose that*

$$n\,\mathbf{W}_n\mathbf{W}_n^T\xrightarrow[n\to\infty]{}\boldsymbol{G}\in\mathcal{M}_q^{>}. \tag{22}$$

If $\hat{\sigma}_{ij}(Y)$ converges to $\sigma_{ij}(Y)$ in quadratic mean for all i, $j=1,\ldots,p$, and all Y in the relevant sample space, and if $\mathrm{Cov}\,(\hat{\sigma}_{ij}(Y_k),\hat{\sigma}_{ij}(Y_l))\longrightarrow 0$ as $n\to\infty$ for all i, $j=1,\ldots,p$, k, $l=1,\ldots,n$, $k\neq l$, then $\widehat{\mathbf{H}}_n$ is a L_2-consistent estimate of $\tilde{\boldsymbol{H}}$.

Proof. Since $\tilde{\boldsymbol{H}}$ is the limit matrix of $\tilde{\boldsymbol{H}}_n$, it suffices to show that

$$\widehat{\boldsymbol{H}}_n-\tilde{\boldsymbol{H}}_n\xrightarrow[n\to\infty]{}0\quad\text{in } L_2, \tag{23}$$

for then it follows that $\widehat{\mathbf{H}}_n$ is a consistent estimate of $\tilde{\boldsymbol{H}}$ from the triangle inequality.

Consider the ijth blocks of $\widehat{\boldsymbol{H}}_n$ and $\tilde{\boldsymbol{H}}_n$. We find that $\sum_k^n\hat{\sigma}_{ij}(Y_k)(nW_{nk}(Y_l)\times W_{nk}(Y_m))$ and $\sum_k^n\sigma_{ij}(Y_k)(nW_{nk}(Y_l)W_{nk}(Y_m))$ are the lmth entries of the ijth

blocks of $\widehat{\mathbf{H}}_n$ and $\tilde{\boldsymbol{H}}_n$, respectively, for $m,\ n = 1,\dots,q$, and $i,\ j = 1,\dots,p$. Then (23) is true if and only if

$$\sum_k^n \hat{\sigma}_{ij}(Y_k)(nW_{nk}(Y_l)W_{nk}(Y_m)) - \sum_k^n \sigma_{ij}(Y_k)(nW_{nk}(Y_l)W_{nk}(Y_m)) \xrightarrow[n\to\infty]{} 0$$

in L_2, for all $m,\ n = 1,\dots,q$, and $i,\ j = 1,\dots,p$. Now,

$$\begin{aligned}
&\mathrm{E}\Big[\sum_{k=1}^n \hat{\sigma}_{ij}(Y_k)(nW_{nk}(Y_l)W_{nk}(Y_m)) - \sum_{k=1}^n \sigma_{ij}(Y_k)(nW_{nk}(Y_l)W_{nk}(Y_m))\Big]^2 \\
&= \mathrm{E}\Big[\sum_{k=1}^n (\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))(nW_{nk}(Y_l)W_{nk}(Y_m))\Big]^2 \\
&= \sum_{k=1}^n \mathrm{E}\Big[(\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))^2 (nW_{nk}(Y_l)W_{nk}(Y_m))^2\Big] \\
&\quad + \sum_{k=1}^n \sum_{r\neq k}^n \mathrm{E}\Big[(\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))(\hat{\sigma}_{ij}(Y_r) - \sigma_{ij}(Y_r)) \\
&\qquad\qquad \times (nW_{nk}(Y_l)W_{nk}(Y_m))(nW_{nr}(Y_l)W_{nr}(Y_m))\Big].
\end{aligned}$$

The integration can be brought inside the sum by the Bounded Convergence Theorem (see Billingsley (1986), p.214). Since (22) holds by assumption and $\hat{\sigma}_{ij}(Y_k)$ is consistent in quadratic mean for $\sigma_{ij}(Y_k)$, it follows that $\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k)$ is L_2 bounded for all $k = 1,\dots,n$. But then, we also have that $\mathrm{E}\,[(\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))^2] \to 0$, as $n \to \infty$, by the L_2 consistency of $\hat{\sigma}_{ij}(Y_k)$. Also, by assumption $\mathrm{E}\,[(\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))(\hat{\sigma}_{ij}(Y_r) - \sigma_{ij}(Y_r))] \xrightarrow[n\to\infty]{} 0$. Therefore, $\mathrm{E}\,[\sum_{k=1}^n (\hat{\sigma}_{ij}(Y_k) - \sigma_{ij}(Y_k))(nW_{nk}(Y_l)W_{nk}(Y_m))]^2 \xrightarrow[n\to\infty]{} 0$ for all $l,\ m = 1,\dots,q,\ i,\ j = 1,\dots,p$.

Lemma 4. *Let* $f_n(Y) = \hat{\mathrm{E}}_n(g(X)|Y) \xrightarrow[n\to\infty]{} f(Y) = \mathrm{E}\,(g(X)|Y)$ *in* L_2, *and* $f_n'(Y) = \hat{\mathrm{E}}_n(h(X)|Y) \xrightarrow[n\to\infty]{} f'(Y) = \mathrm{E}\,(h(X)|Y)$ *in* L_2. *If* $f_n(Y)$ *and* $f_n'(Y)$ *are uniformly bounded functions of* Y, *then*

$$\widehat{\mathrm{Cov}}_n(g(X), h(X)|Y) \longrightarrow \mathrm{Cov}\,(g(X), h(X)|Y) \quad \text{in } L_2, \tag{24}$$

$$\widehat{\mathrm{Var}}_n(g(X)|Y) \longrightarrow \mathrm{Var}\,(g(X)|Y) \quad \text{in } L_2. \tag{25}$$

Proof. Let $\|\cdot\|_2$ denote the L_2 norm. By the triangle inequality we have

$$\|f_n f_n' - f f'\|_2 \le \|(f_n - f)f'\|_2 + \|f_n(f_n' - f')\|_2. \tag{26}$$

Since both sequences of functions are uniformly bounded, there exists a positive number M such that $|f_n| \le M$, $|f_n'| \le M$. Therefore,

$$\|f_n(f_n' - f')\|_2 \le M\|f_n' - f'\|_2. \tag{27}$$

Now, since $f_n'(Y) \xrightarrow[n\to\infty]{} f'(Y)$ in L_2, there exists a subsequence of f_n' that converges to f' almost everywhere. Thus, $|f'| \leq M$ a.e., which implies that

$$\|(f_n - f)f'\|_2 \leq M\|f_n - f\|_2 \quad \text{a.e.} \tag{28}$$

From (26), (27), (28) we conclude that $f_n(Y)f_n'(Y) \xrightarrow[n\to\infty]{} f(Y)f'(Y)$ in L_2. Hence, (24) holds, of which (25) is a trivial consequence.

References

Aragon, Y. and Saracco, J. (1997). Sliced inverse regression (SIR): An appraisal of small sample alternatives to slicing. *Comput. Statist.* **12**, 109-130.

Bunke, H. and Bunke, O. (1986). *Statistical Inference in Linear Models. Statistical Methods of Model Building*, Vol I. Wiley, Chichester.

Bura, E. and Cook, R. D. (2001a). Estimating the structural dimension of regressions via parametric inverse regression. *J. Roy. Statist. Soc. Ser. B* **63**, 393-410.

Bura, E. and Cook, R. D. (2001b). Extending SIR: the weighted chi-squared test. *J. Amer. Statist. Assoc.* To appear.

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **7**, 368-385.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

Cook, R. D. (1998a). Principal Hessian directions revisited (with discussion). *J. Amer. Statist. Assoc.* **91**, 983-992.

Cook, R. D. (1998b). *Regression Graphics*. Wiley, New York.

Cook, R. D. (1996). Graphics for regressions with a binary response. *J. Amer. Statist. Assoc.* **91**, 983-992.

Cook, R. D. (1994a). On the interpretation of regression plots. *J. Amer. Statist. Assoc.* **89**, 177-189.

Cook, R. D. (1994b). Using dimension-reduction subspaces to identify important inputs in models of physical systems in *1994 Proceedings of the Section on Physical and Engineering Sciences*, 18-25. American Statistical Association. Alexandria, VA.

Cook, R. D. and Weisberg, S. (1999). *Applied Regression including Computing and Graphics*. Wiley, New York.

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12**, 793-815.

Eaton, M. L.(1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20**, 272-276.

Eaton, M. L. and Tyler, D. E. (1994). The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. *J. Multivariate Anal.* **34**, 439-446.

Eicker, F. (1966). A multivariate central limit theorem for random linear vector forms. *Ann. Math. Statist.* **37**, 1825-1828.

Ferré, L. (1998). Determining the Dimension in Sliced Inverse Regression and Related Methods. *J. Amer. Statist. Assoc.* **93**, 132-140.

Fung, W. K., He, X., Liu, L. and Shi, P. D. (2000). Dimension reduction based on canonical correlation. Tentatively accepted by *Statist. Sinica*.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *Ann. Statist.* **21**, 867-889.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag, New York.

Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20**, 1040-1061.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316-342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87**, 1025-1039.

Tierney, L. (1990). *Lisp - Stat: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley, New York.

Velilla, S. (1998). Assessing the number of linear components in a general regression problem. *J. Amer. Statist. Assoc.* **93**, 1088-1098.

Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statist. Sinica* **5**, 727-736.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **24**, 1053-1068.

Department of Statistics, The George Washington University, Washington, DC 20052, U.S.A.

E-mail: ebura@gwu.edu

(Received August 2001; accepted July 2002)