应用概率统计　第二十六卷
第五期　2010年10月

Chinese Journal of Applied Probability
and Statistics Vol.26 No.5 Oct. 2010

Marginal Coordinate Tests for Central Mean Subspace with Principal Hessian Directions *

Yu Zhou[1] Dong Yuexiao[2] Fang Yun[1★]

([1] *School of Finance and Statistics, East China Normal University, Shanghai, 200241*)

([2] *Department of Statistics, Temple University, Philadelphia, PA 19122, U.S.A.*)

Abstract

We provide marginal coordinate tests based on two competing Principal Hessian Directions (PHD) methods. Predictor contributions to central mean subspace can be effectively identified by our proposed testing procedures. PHD-based tests avoid choosing the number of slices, which is a well-known shortcoming of similar tests based on Sliced Inverse Regression (SIR) or Sliced Average Variance Estimation (SAVE). The asymptotic distributions of our test statistics under the null hypothesis are provided and the effectiveness of the new tests is illustrated by simulations.

Keywords: Marginal coordinate tests, Principal Hessian Directions.

AMS Subject Classification: 62H15.

§1. Introduction

Dimension reduction methods have gained increasing popularity in recent years due to an abundance of high-dimensional data. The theory of sufficient dimension reduction (Li, 1991; Cook, 1998) has been developed to reduce the predictor dimension prior to the model formulation, while preserving full regression information and imposing few probabilistic assumptions. Consider a univariate response Y and a p-dimensional predictor X. Sufficient dimension reduction aims to find a a subspace spanned by the column space of a $p \times d$ matrix η with $d \leq p$ such that

$$Y \perp\!\!\!\perp X | \eta^T X, \tag{1.1}$$

where $\perp\!\!\!\perp$ indicates independence. The column space of η is called a dimension reduction space. Under very mild conditions, such as those given in Cook (1998), Chiaromonte and Cook (2002), and recently further relaxed by Yin, Li and Cook (2007), the intersection of all such spaces is itself a dimension reduction space. We call this intersection the Central

*The first and corresponding authors were supported by National Social Science Foundation (08CTJ001).

★Corresponding author, E-mail: kellyfang0221@gmail.com.

Received September 30, 2009. Revised October 24, 2009.

Space (CS), and denote it by $\mathcal{S}_{Y|X}$ (Cook, 1998). If the mean response $\mathsf{E}(Y|X)$ is of primary interest, the objective is then tailored to find a $p \times d$ matrix η such that

$$Y \perp\!\!\!\perp \mathsf{E}(Y|X)|\eta^T X. \tag{1.2}$$

The smallest subspace satisfying (2) is called the central mean subspace (CMS), which is denoted by $\mathcal{S}_{\mathsf{E}(Y|X)}$ (Cook and Li, 2002). The central mean subspace contains all information in X about $\mathsf{E}(Y|X)$, and is always a subspace of the related CS. The dimension of the CS or CMS is usually referred to as the structure dimension, and denoted by d. Classic sufficient dimension reduction methods that seek to identify the CS include Sliced Inverse Regression (SIR; Li, 1991) and Sliced Average Variance Estimation (SAVE; Cook and Weisberg, 1991), while Principal Hessian Directions (PHD; Li, 1992, Cook, 1998a), as pointed out by Cook and Li (2002), is designed to target the CMS.

Testing the significance of subsets of predictors is frequently studied in the model-based regression. Cook (2004) first proposed such a test based on SIR in the context of sufficient dimension reduction, and named it marginal coordinate hypothesis test. Shao, Cook and Weisberg (2007) further extended the marginal coordinate test to SAVE. Tests with SIR and SAVE both target subset selection in $\mathcal{S}_{Y|X}$.

Inspired by previous work, we propose marginal coordinate tests based on PHD in this paper. Since $\mathcal{S}_{\mathsf{E}(Y|X)} \subseteq \mathcal{S}_{Y|X}$, when our objective is selecting active predictors in the CMS rather than the CS, PHD-based marginal coordinate tests would be more accurate than the tests based on SIR or SAVE. Moreover, SIR and SAVE both require slicing and might be sensitive to the number of slices. Choosing optimal slices is still an open problem in SIR and SAVE. Oftentimes different choices may effect the test performances severely. Our proposed tests with PHD do not have such restrictions. Under fairly general conditions, the test statistic converges in distribution to a weighted χ^2 distribution. If the predictors are normally distributed, the asymptotic null distribution of the modified test statistic further reduces to a central χ^2.

The rest of this paper is organized as follows. In Section 2 we briefly review PHD method and discuss under what kind of conditions PHD is an exhaustive estimator. In Section 3 we develop the marginal coordinate tests with PHD and asymptotic null distributions of the test statistics. Simulation results are reported in Section 4. Technical proofs are relegated to the Appendix.

§2. Principal Hessian Directions

Define the standardized predictor $Z = \Sigma^{-1/2}(X - \mu)$, where $\Sigma = \mathsf{Var}(X)$ and $\mu = \mathsf{E}(X)$. Due to an affine invariance law (Cook, 1998), we assume throughout this paper that the predictor X is standardized satisfying $\mathsf{E}(X)=0$, $\mathsf{Var}\,(X)=I_p$ and $\mathsf{E}(Y)=0$, where I_p is the $p\times p$ identity matrix. Classic sufficient dimension reduction can often be formulated as an eigen-decomposition problem:

$$M\beta_i=\lambda_i\beta_i, \qquad \text{for } \ i=1,\cdots,p,$$

where M is a method-specific symmetric kernel matrix with eigenvalues $\lambda_1\geq\cdots\geq\lambda_d>\lambda_{d+1}=\cdots=\lambda_p=0$. SIR and SAVE are based on matrix $M_{\rm SIR}=\mathsf{Var}\,[\mathsf{E}(Z|Y)]$ and $M_{\rm SAVE}=\mathsf{E}[I_p-\mathsf{Var}\,(Z|Y)]^2$. There are two variations for PHD, the response based PHD, denoted by y-PHD, and the residual based PHD, denoted by r-PHD. Their corresponding kernel matrices are $M_{y-\rm PHD}=\Sigma_{yzz}\Sigma_{yzz}$ and $M_{r-\rm PHD}=\Sigma_{rzz}\Sigma_{rzz}$, where $\Sigma_{yzz}=\mathsf{E}[(Y-\mathsf{E}(Y))ZZ^T]$, $\Sigma_{rzz}=\mathsf{E}(rZZ^T)$ with $r=Y-\mathsf{E}(Y)-\mathsf{E}(YZ^T)Z$. Under certain conditions, it can be shown that $(\eta_1,\cdots,\eta_d)=(\Sigma^{-1/2}\beta_1,\cdots,\Sigma^{-1/2}\beta_d)$, or the $\Sigma^{-1/2}$-transformed eigenvectors corresponding to the nonzero eigenvalues $\{\lambda_1\geq\cdots\geq\lambda_d>0\}$, form a basis for the CS or CMS under investigation. In order to produce consistent estimator of the CMS, PHD requires,

1. Linear Conditional Mean (LCM): $\mathsf{E}(X|\eta^TX)$ is a linear function of X.
2. Constant Conditional Variance (CCV): $\mathsf{Var}\,(X|\eta^TX)$ is nonrandom.

These conditions are rather common in sufficient dimension literature. SIR requires LCM for consistent estimation, and LCM and CCV conditions are both required for SAVE to work. In addition, we impose a coverage condition for PHD to estimate the CMS exhaustively.

3. Coverage condition: For any nonzero $p\times 1$ vector $\nu\in S_{\mathsf{E}(Y|Z)}$, $\nu^TM_{y-\rm PHD}\nu>0$ and $\nu^TM_{r-\rm PHD}\nu>0$.

The above three conditions guarantee that $\operatorname{span}(M_{y-\rm PHD})=S_{\mathsf{E}(Y|Z)}$ and $\operatorname{span}(M_{r-\rm PHD})=S_{\mathsf{E}(Y|Z)}$, where $\operatorname{span}(A)$ is defined as the subspace spanned by the columns of A. Such a coverage condition is similar to that of SIR (Cook, 2004) and SAVE (Shao, Cook and Weisberg, 2007), where they want to exhaustively estimate the CS.

The sampling estimation of PHD is straightforward. Let $\{(x_i,y_i),i=1,\cdots,n\}$ be an i.i.d. sample of (X,Y). Note that $\{(x_i,y_i),i=1,\cdots,n\}$ is an also i.i.d. sample of (Z,Y) since $\mathsf{E}(X)=0$ and $\mathsf{Var}\,(X)=I_p$ by our assumptions. Thus we can define $z_i=x_i$, $i=1,\cdots,n$. Denote $\overline{x}=\mathsf{E}_n(x)$ and $\widehat{\Sigma}=\mathsf{E}_n[(x-\overline{x})(x-\overline{x})^T]$ to be the sample mean and sample covariance of X, where $\mathsf{E}_nf(x)$ stands for $n^{-1}\sum\limits_{i=1}^n f(x_i)$. Let $\widehat{z}_i=\widehat{\Sigma}^{-1/2}(x_i-\overline{x})$, $i=1,\cdots,n$ and $\overline{y}=\mathsf{E}_n(y)$. Then we can estimate $M_{y-\rm PHD}$ by $\widehat{M}_{y-\rm PHD}=\widehat{\Sigma}_{yzz}\widehat{\Sigma}_{yzz}$, where $\widehat{\Sigma}_{yzz}=(1/n)\sum\limits_{i=1}^n(y_i-\overline{y})\widehat{z}_i\widehat{z}_i^T$. To estimate the population regression error $r_i=y_i-\mathsf{E}(YZ)^Tz_i$, we use its sample version $\widehat{r}_i=y_i-\overline{y}-\widehat{\gamma}^T\widehat{z}_i$, $i=1,\cdots,n$, where $\widehat{\gamma}=\mathsf{E}_n(\widehat{z}(y-\overline{y}))$. The kernel matrix for the r-PHD is then estimated by $\widehat{M}_{r-\rm PHD}=\widehat{\Sigma}_{rzz}\widehat{\Sigma}_{rzz}$, where $\widehat{\Sigma}_{rzz} = (1/n)\sum_{i=1}^{n}\widehat{r}_i\widehat{z}_i\widehat{z}_i^T$. Then we estimate $S_{\mathsf{E}(Y|X)}$ consistently by the span of $(\widehat{\eta}_1,\cdots,\widehat{\eta}_d) = (\widehat{\Sigma}^{-1/2}\widehat{\beta}_1,\cdots,\widehat{\Sigma}^{-1/2}\widehat{\beta}_d)$, where $(\widehat{\beta}_1,\cdots,\widehat{\beta}_d)$ are eigenvectors associated with the d-largest eigenvalues of $\widehat{M}_{y-\mathrm{PHD}}$ or $\widehat{M}_{r-\mathrm{PHD}}$.

§3. Marginal Coordinate Hypotheses

Marginal coordinate hypotheses are statements of the form $S_{\mathsf{E}(Y|X)} \subseteq \mathcal{V}$, where $\mathcal{V}$ is a user-selected subspace of the predictor space that specifies the hypothesis. For example, suppose $Y \perp\!\!\!\perp \mathsf{E}(Y|X)|X_1$ for a given partition $X^T = (X_1^T, X_2^T)$, which means the $p-m$ selected predictors X_2 do not contribute to the regression mean function $\mathsf{E}(Y|X)$. In this case, $\mathcal{V}$ can be selected as the subspace of $\mathbf{R}^p$ corresponding to the coordinates of X_1, that is, $\mathcal{V} = \mathrm{span}((I_m, 0)^T)$. We formulate the following marginal coordinate hypothesis:

$$H_0 : S_{\mathsf{E}(Y|X)} \subseteq \mathcal{V} \qquad \text{versus} \qquad H_1 : S_{\mathsf{E}(Y|X)} \nsubseteq \mathcal{V}, \tag{3.1}$$

where $\dim(\mathcal{V}) = m < p$.

We need the following notations to construct the test statistics. Let α_x be a $p \times m$ matrix of column rank m and $\mathrm{span}(\alpha_x) = \mathcal{V}$. At the Z scale, without loss of generality we take the columns of $\alpha = \Sigma^{1/2}\alpha_x(\alpha_x^T\Sigma\alpha_x)^{-1/2}$ to be an orthonormal basis. Denote its sample estimator by $\widehat{\alpha} = \widehat{\Sigma}^{1/2}\alpha_x(\alpha_x^T\widehat{\Sigma}\alpha_x)^{-1/2}$. Finally, let columns of $p \times (p-m)$ matrices H and $\widehat{H}$ be orthonormal bases for the orthogonal complements of $\mathrm{span}(\alpha)$ and $\mathrm{span}(\widehat{\alpha})$ respectively. In fact, H is chosen to select $p-m$ appropriate rows of η such that $H^T\Sigma^{1/2}\eta = 0$. The test statistics based on y-PHD and r-PHD are proposed as following:

$$T_1(\widehat{H}) = n\mathrm{tr}(\widehat{H}^T\widehat{\Sigma}_{yzz}\widehat{H})^2/2\widehat{\mathsf{Var}}\,(Y), \tag{3.2}$$

$$T_2(\widehat{H}) = n\mathrm{tr}(\widehat{H}^T\widehat{\Sigma}_{rzz}\widehat{H})^2/2\widehat{\mathsf{Var}}\,(r), \tag{3.3}$$

where tr stands for the trace operator, $\widehat{\mathsf{Var}}\,(Y)$ and $\widehat{\mathsf{Var}}\,(r)$ are consistent sample estimates of $\mathsf{Var}\,(Y)$ and $\mathsf{Var}\,(r)$ respectively. Let $V = H^TZ$, then for the test statistics T_1 and T_2 given by (3.2) and (3.3), we have the following theorem.

Theorem 3.1 Assume that LCM, CCV and coverage condition hold. Then, under the marginal coordinate hypothesis $S_{\mathsf{E}(Y|X)} \subseteq \mathcal{V}$,

$$2T_1 \longrightarrow \sum_j \delta_j\chi_j^2(1) \qquad \text{and} \qquad 2T_2 \longrightarrow \sum_j \tau_j\chi_j^2(1),$$

where the convergence is in distribution as $n \to \infty$. For $j = 1,\cdots,(p-m)(p-m+1)/2$, $\chi_j^2(1)$ denotes independent chi-squared random variable with one degree of freedom, δ_j and τ_j are the jth largest eigenvalues of the $(p-m)\times(p-m)$ matrices $\mathsf{Var}\,(\mathrm{vec}(Y(VV^T - I_{p-m})))/\mathsf{Var}\,(Y)$ and $\mathsf{Var}\,(\mathrm{vec}(r(VV^T - I_{p-m})))/\mathsf{Var}\,(r)$ respectively, in which $\mathrm{vec}(\cdot)$ denotes the operator that stacking the columns of a matrix.

The variance matrices $\mathsf{Var}\,(\mathrm{vec}(Y(VV^T-I_{p-m})))$ and $\mathsf{Var}\,(\mathrm{vec}(r(VV^T-I_{p-m})))$ can be consistently estimated by the sample covariance matrices of $y_i(\widehat{H}\widehat{z}_i\widehat{z}_i^T\widehat{H}-I_{p-m})$ and $r_i(\widehat{H}\widehat{z}_i\widehat{z}_i^T\widehat{H}-I_{p-m})$, $i=1,\cdots,n$. Their largest $(p-m)(p-m+1)/2$ eigenvalues can thus be consistently estimated correspondingly.

When X is normally distributed, LCM and CCV condition naturally hold and we have $H^TZ\sim N(0,I_{p-m})$. With the additional coverage condition, the asymptotic null distributions of test statistics T_1 and T_2 degenerates to central χ^2.

Theorem 3.2 Assume that X is normally distributed and the coverage condition is enhanced to $S_{\mathsf{E}(Y|X)}=S_{Y|X}$. Then, under the coordinate hypothesis $S_{\mathsf{E}(Y|X)}\subseteq\mathcal{V}$,

$$T_1\longrightarrow\chi^2\{(p-m)(p-m+1)/2\},$$

where the convergence is in distribution as $n\to\infty$. If in addition we assume $\mathsf{E}(r|Z)=\mathsf{E}(r|\beta^TZ)$ and $\mathsf{Var}\,(r|Z)=\mathsf{Var}\,(r|\beta^TZ)$, where β is a $p\times d$ matrix composed of the eigenvectors corresponding the nonzero eigenvalues of $M_{r-\mathrm{PHD}}$, then

$$T_2\longrightarrow\chi^2\{(p-m)(p-m+1)/2\},$$

where the convergence is in distribution as $n\to\infty$.

$\mathsf{E}(r|Z)=\mathsf{E}(r|\beta^TZ)$ and $\mathsf{Var}\,(r|Z)=\mathsf{Var}\,(r|\beta^TZ)$ are commonly used conditions in marginal dimension test for r-PHD (Cook, 1998a). The marginal coordinate tests with PHD enable us to perform model-free variable selection by examining the possibility of excluding a subset of predictors, as we are going to see in the next section.

§4. Simulation Studies

To check the performance of the PHD-based marginal coordinate tests, we focus on the following two models:

$$\text{Model I:}\ \ Y=\cos(X_1)+0.2\varepsilon, \tag{4.1}$$

$$\text{Model II:}\ \ Y=\cos(2X_1)-\cos(X_2)+0.2\varepsilon, \tag{4.2}$$

where the error ε is standard normal and independent of X. Set p to be 5 and 10. We also try different sample sizes with $n=50,100,150,200$. Let $e_1=(1,0,\cdots,0)^T$ and $e_2=(0,1,0,\cdots,0)^T$. $\mathcal{V}$ can be chosen as $\mathrm{span}(e_1)$ for Model I or as $\mathrm{span}(e_1,e_2)$ for Model II. For illustration purpose, we set $X\sim N(0,I_p)$. As a result, the test statistics for y-PHD and r-PHD are central χ^2 distribution as stated in Theorem 3.2. We genarate 1000 replications at each sample configuration. The estimated significance levels are reported by counting the number of p-values that are less than or equal to a nominal level 5%. We also include the marginal coordinate test with SAVE (Shao, Cook and Weisberg, 2007) for comparison. To the best of our knowledge, there is no data-driven method in selecting the optimal number of slices for SAVE. A wide range of number of slices are included and we set $n_s = 2, 5, 10$ and 25 for SAVE.

Tables 1 Estimated levels, as percentages of nominal 5% tests, based on Model I

		PHD		SAVE			
p	n	y-PHD	r-PHD	$n_s=2$	$n_s=5$	$n_s=10$	$n_s=25$
$p=5$	$n=50$	5.10	3.70	2.20	2.50	0.20	0.10
	$n=100$	5.60	5.80	3.50	3.60	2.40	0
	$n=200$	5.00	4.90	4.00	3.20	2.60	0.70
$p=10$	$n=50$	6.60	4.50	1.30	0.80	0.10	0
	$n=100$	6.40	4.90	3.40	2.40	0.50	0
	$n=200$	5.40	5.30	3.70	3.10	2.00	0.50

Tables 2 Estimated levels, as percentages of nominal 5% tests, based on Model II

		PHD		SAVE			
p	n	y-PHD	r-PHD	$n_s=2$	$n_s=5$	$n_s=10$	$n_s=25$
$p=5$	$n=50$	5.30	4.60	3.30	1.70	1.10	0
	$n=100$	4.80	4.50	4.10	2.80	1.30	0.40
	$n=200$	5.20	5.40	4.40	4.10	3.50	1.80
$p=10$	$n=50$	4.40	3.60	1.30	1.00	0	0
	$n=100$	4.60	4.80	3.10	2.30	0.40	0
	$n=200$	4.80	4.90	2.90	3.00	1.80	0

Tables 1 and 2 contain results from marginal coordinate tests for Model I and II respectively. The results of our proposed marginal coordinate tests with PHD are very satisfying. The estimated levels of the PHD tests give a uniformly closer approximation to the true nominal level than the tests based on SAVE, regardless of the choice of number of slices. We confirm the findings of Li and Zhu (2007) that marginal coordinate test with SAVE depends heavily on the choice of number of slices. For fixed p, all methods improve with increasing sample size n. It is a nice surprise that with moderate sample size $n = 50$, when the SAVE-based tests are meaningless, tests based on PHD can still yield very good results sometimes. The y-PHD test seems to be more sensitive to the change of p. It deteriorates more when p increases from 5 to 10 with fixed n. The dimension of predictors p has little effect on the results of r-PHD tests. Free from the issue of choosing optimal number of slices, marginal coordinate tests with PHD are effective and more efficient in performing model free subset selection and hence are strongly recommended.

§5. Appendix

Before proceeding to the proof of Theorem 3.1, we first give two useful lemmas.

Lemma 5.1 Let $\xi = zz^T - I$, $\gamma = \mathsf{E}(YZ)$, then we have

$$\widehat{\Sigma}_{yzz} = \Sigma_{yzz} + \mathsf{E}_n(y\xi - \Sigma_{yzz}) - \mathsf{E}_n(\xi)\Sigma_{yzz}/2 - \Sigma_{yzz}\mathsf{E}_n(\xi)/2 - \overline{z}\gamma^T - \gamma\overline{z}^T + O_p(n^{-1}),$$
$$\widehat{\Sigma}_{rzz} = \Sigma_{rzz} + \mathsf{E}_n(r\xi - \Sigma_{yzz}) - \mathsf{E}_n(\xi)\Sigma_{rzz}/2 - \Sigma_{rzz}\mathsf{E}_n(\xi)/2 + O_p(n^{-1}).$$

Proof It is easy to verify that

$$\begin{aligned}\widehat{\Sigma}_{yxx} &= \mathsf{E}_n(yxx^T) - \overline{y}\mathsf{E}_n(xx^T) - \overline{x}\mathsf{E}_n(yx^T) - \mathsf{E}_n(yx)\overline{x}^T + O_p(n^{-1})\\ &= \mathsf{E}_n(yxx^T) - \overline{y}I_p - \overline{x}\mathsf{E}(YX)^T - \mathsf{E}(YX)\overline{x}^T + O_p(n^{-1})\\ &= \Sigma_{yxx} + \mathsf{E}_n[y(xx^T - I) - \Sigma_{yxx}] - \overline{x}\mathsf{E}(YX)^T - \mathsf{E}(YX)\overline{x}^T + O_p(n^{-1}).\end{aligned}$$

According to Cook (1998), we have $\widehat{\Sigma}^{-1/2} = I_p - \mathsf{E}_n(xx^T - I_p)/2 + O_p(n^{-1})$. Then the conclusion for $\widehat{\Sigma}_{yzz}$ can be easily derived since $\widehat{\Sigma}_{yzz} = \widehat{\Sigma}^{-1/2}\widehat{\Sigma}_{yxx}\widehat{\Sigma}^{-1/2}$. Similar techniques can be applied to expand $\widehat{\Sigma}_{rzz}$ and hence we omit the details here. Just keep in mind that $\mathsf{E}(rX) = 0$ but $\mathsf{E}(YX)$ may not be zero. □

Lemma 5.2 Suppose that $\widehat{H}$ is a $p\times(p-m)$ matrix such that $\widehat{H}^T\widehat{H} = I_{p-m}$ and $\widehat{H} = H + O_p(n^{-1/2})$. Then under null hypothesis, $T_i(\widehat{H}) = T_i(H) + o_p(1)$, $i = 1, 2$.

Proof We only deal with $T_1(\widehat{H})$ since derivation for $T_2(\widehat{H})$ is almost the same. Note that $H^T\Sigma_{yzz} = 0$ under null hypothesis, then

$$\begin{aligned}\widehat{H}^T\widehat{\Sigma}_{yzz}\widehat{H} &= H^T\widehat{\Sigma}_{yzz}H + (\widehat{H} - H)^T\widehat{\Sigma}_{yzz}H + H^T\widehat{\Sigma}_{yzz}(\widehat{H} - H) + O_p(n^{-1})\\ &= H^T\widehat{\Sigma}_{yzz}H + (\widehat{H} - H)^T\Sigma_{yzz}H + H^T\Sigma_{yzz}(\widehat{H} - H) + O_p(n^{-1})\\ &= H^T\widehat{\Sigma}_{yzz}H + O_p(n^{-1}).\end{aligned}$$

The conclusion is then straightforward by invoking slutsky theorem. □

Proof of Theorem 3.1 Under null hypothesis, $H^T\Sigma_{yzz} = 0$ and $H^T\Sigma_{rzz} = 0$. Moreover, from Cook and Li (2002), the OLS estimator $\gamma = \mathsf{E}(YZ) \subseteq S_{\mathsf{E}(Y|Z)}$, then $H^T\gamma = 0$. By invoking Lemma 5.1, we can derive that

$$H^T\widehat{\Sigma}_{yzz}H = \frac{1}{n}\sum_{i=1}^n y_i(H^Tz_iz_i^TH - I_{p-m}) + O_p(n^{-1}),$$
$$H^T\widehat{\Sigma}_{yzz}H = \frac{1}{n}\sum_{i=1}^n y_i(H^Tr_ir_i^TH - I_{p-m}) + O_p(n^{-1}).$$

By central limit theorem, we can conclude that $n^{1/2}\mathrm{vec}(H^T\widehat{\Sigma}_{yzz}H)$ and $n^{1/2}\mathrm{vec}(H^T\widehat{\Sigma}_{yzz}H)$ converge in distribution to $N\big(0, \mathsf{Var}\,(\mathrm{vec}(Y(VV^T - I_{p-m})))\big)$ and $N\big(0, \mathsf{Var}\,(\mathrm{vec}(r(VV^T - I_{p-m})))\big)$ respectively. The conclusion can then be easily derived by invoking Lemma 5.2 and slutsky theorem. $\square$

Proof of Theorem 3.2 Define $U = VV^T - I_{p-m}$. By the EV-VE formula, we have $\mathsf{Var}\,(V) = \mathsf{E}[(\mathsf{Var}\,V|\alpha^TZ)] + \mathsf{Var}\,[\mathsf{E}(V|\alpha^TZ)]$. Since Z is normally distributed, we have $\mathsf{Var}\,(Z|\alpha^TZ)$ and $\mathsf{Var}\,(\mathrm{vec}(U)|\alpha^TZ)$ are nonrandom, and $\mathsf{E}(V|\alpha^TZ) = H^T\mathsf{E}(Z|\alpha^TZ) = H^TP_\alpha Z = 0$, where P_α denotes the projection operator for $\mathrm{span}(\alpha)$ with respect to the identity inner product. Then $\mathsf{Var}\,(V|\alpha^TZ) = \mathsf{Var}\,(V) = I_{p-m}$. Similarly, $\mathsf{E}(U|\alpha^TZ) = \mathsf{Var}\,(V|\alpha^TZ) + \mathsf{E}(V|\alpha^TZ)\mathsf{E}(V^T|\alpha^TZ) - I_{p-m} = 0$ and hence $\mathsf{Var}\,(\mathrm{vec}(U)|\alpha^TZ) = \mathsf{Var}\,(\mathrm{vec}(U))$. Moreover, from Proposition 1 in Cook (2004), we know that $Y \perp\!\!\!\perp V|\alpha^TZ$, which yields to $\mathsf{Var}\,[\mathsf{E}(Y\mathrm{vec}(U)|\alpha^TZ)] = \mathsf{Var}\,[\mathsf{E}(Y|\alpha^TZ)\mathsf{E}(\mathrm{vec}(U)|\alpha^TZ)] = 0$. We then can further derive that

$$\begin{aligned}\mathsf{Var}\,(Y\mathrm{vec}(U)) &= \mathsf{E}[\mathsf{Var}\,(Y\mathrm{vec}(U)|\alpha^TZ)] = \mathsf{E}(\mathsf{Var}\,(Y|\alpha^TZ)\mathsf{Var}\,(\mathrm{vec}(U)|\alpha^TZ)]\\ &= \mathsf{E}[\mathsf{E}(YY^T|\alpha^TZ)\mathsf{Var}\,(\mathrm{vec}(U))] = \mathsf{Var}\,(Y)\mathsf{Var}\,(\mathrm{vec}(VV^T)).\end{aligned}$$

Since $V \sim N(0, I_{p-m})$, $\mathsf{Var}\,(\mathrm{vec}(VV^T))/2$ is a projection matrix with rank $(p-m)(p-m+1)/2$ (Schott, 1997) and hence all its nonzero eigenvalues are all equal to 1. Then the asymptotic distribution of T_1 can be easily derived by invoking slutsky theorem. Similar procedures can be taken to derive that $\mathsf{Var}\,(r\mathrm{vec}(U)) = \mathsf{Var}\,(r)\mathsf{Var}\,(\mathrm{vec}(U))$ and then the asymptotic distribution of T_2 under null hypotheis. We omit the details here. $\square$

References

[1] Chiaromonte, F. and Cook, R.D., Sufficient dimension reduction and graphics in regression, *Annals of the Institute of Statistical Mathematics*, **54**(2002), 768–795.

[2] Cook, R.D., *Regression Graphics: Ideas for Studying Regressions Through Graphics*, New York: Wiley, 1998.

[3] Cook, R.D., Principal Hessian directions revisited, *Journal of the American Statistical Association*, **93**(1998a), 84–94.

[4] Cook, R.D., Testing predictor contributions in sufficient dimension reduction, *Annals of Statististics*, **32**(2004), 1062–1092.

[5] Cook, R.D. and Li, B., Dimension reduction for the conditional mean, *Annals of Statististics*, **30**(2002), 455–474.

[6] Cook, R.D. and Weisberg, S., Discussion of paper by K.C. Li, *Journal of the American Statistical Association*, **86**(1991), 318–332.

[7] Li, K.C., Sliced inverse regression for dimension reduction (with discussion), *Journal of the American Statistical Association*, **86**(1991), 316–342.

[8] Li, K.C., On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma, *Journal of the American Statistical Association*, **87**(1992), 1025–1039.

[9] Li, Y.X. and Zhu, L.X., Asymptotics for sliced average variance estimation, *Annals of Statistics*, **35**(2007), 41–69.

[10] Schott, J.R., *Matrix Analysis for Statistics*, John Wiley, New York, 1997.

[11] Shao, Y.W., Cook, R.D. and Weisberg, S., Marginal tests with sliced average variance estimation, *Biometrika*, **94**(2007), 285–296.

[12] Yin, X., Li, B. and Cook, R.D., Successive direction extraction for estimating the central subspace in a multiple-index regression, *Journal of Multivariate Analysis*, **99**(2007), 1733–1757.

基于主Hessian方向的中央均值子空间边际坐标检验

於 州[1] 董玥潇[2] 房 云[1]

([1]华东师范大学金融与统计学院, 上海, 200241; [2]天普大学统计系, 美国宾州费城, 19122)

本文给出了基于两种相近的主Hessian方向方法的边际坐标检验. 这种检验方法能够非常有效的识别自变量对于回归均值中央子空间的贡献. 此外, 与利用切片逆回归和切片平均方差估计的检验方法不同的是, 本文中主Hessian方向的检验方法可以避免对切片数目的选择. 我们证明了检验统计量在原假设下的渐近分布, 并且通过模拟, 证实了检验的有效性.

关键词: 边际坐标检验, 主Hessian方向.

学科分类号: O212.