The Pennsylvania State University

The Graduate School

DIMENSION REDUCTION FOR THE CONDITIONAL KTH MOMENT VIA CENTRAL SOLUTION SPACE

A Thesis in

Statistics

by

Yuexiao Dong

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

August 2008

The thesis of Yuexiao Dong was reviewed and approved* by the following:

Bing Li
Professor of Statistics
Thesis Advisor

Bruce G. Lindsay
Professor of Statistics
Head of the Department of Statistics

Runze Li
Professor of Statistics

*Signatures are on file in the Graduate School.

Abstract

The original aim of dimension reduction is to find linear combinations of predictor X, which contain all the regression information of Y versus X. Since the introduction of the very first dimension reduction methods such as OLS and SIR, various dimension reduction methods have been invented, such as SAVE and PHD. The invention of central mean subspace enriched the context of dimension reduction and brought more insight into existing dimension reduction methods. This idea is expanded later to central kth moment space. However, those methods all require stringent conditions on the joint distribution of the predictor. In this thesis, via the notion of central solution space, we want to relax the elliptical distribution assumption required by central kth moment space estimators. Central kth moment solution space is introduced and its estimators are compared with existing methods by simulation.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I want to express my thanks to my advisor, Professor Bing Li. He introduced me to the very interesting research of dimension reduction in autumn 2005. I am also grateful to Professor Runze Li for his valuable advice and encouragement throughout my study.

I appreciate the help from my friends in our department: Zhou Yu, Bo Kai, Michael Zhang, Anderson Liu, Daeyoung Kim and Andreas Artemiou. Thank you all very much for your support.

Chapter 1

Introduction

In this chapter, we are going to give a brief introduction of the development of dimension reduction. The notion of central space and common estimators of central space are introduced in the first two sections. In the following section, central mean space is introduced and we get a deeper understanding of those dimension reduction methods introduced earlier. The generalization of central mean space, central kth moment space, is introduced in the last section.

1.1 Central space

1.1.1 Basic notations

Let X be a p-dimensional random vector representing the predictor, and Y be a random variable representing the response. The goal of dimension reduction (K. C. Li, 1991, 1992; Cook and Weisberg, 1991; Cook, 1998) is to seek $\beta \in \mathbb{R}^{p\times d}(d < p)$, such that

$$Y \perp\!\!\!\perp X|\beta^T X.$$

For any $d \times d$ non-singular matrix A, $Y \perp\!\!\!\perp X|\beta^T X$ if and only if $Y \perp\!\!\!\perp X|(\beta A)^T X$. Thus it is the column space of β that really matters and this leads to the following definition of Dimension Reduction Space.

Definition 1.1.1. *A dimension reduction space (DRS) for (X, Y) is the column space of β, where β satisfies $Y \perp\!\!\!\perp X|\beta^T X$.*

We are not satisfied with just finding any dimension reduction space. In a certain sense, we want to find a minimum dimension reduction space. Under very mild conditions (Yin, Li, and Cook, 2007), the intersection of two dimension reduction spaces is still a dimension reduction space itself. This leads to the following definition of central space.

Definition 1.1.2. *The central space (CS) for (X, Y) is the intersection of all dimension reduction spaces for (X, Y). This space is written as $\mathcal{S}_{Y|X}$.*

In the literature, $\mathcal{S}_{Y|X}$ is also called central subspace, sufficient dimension reduction (SDR) central space or effective dimension reduction (e.d.r.) central space. Without ambiguity, we will just call $\mathcal{S}_{Y|X}$ central space throughout this thesis.

1.1.2 Invariance law of central space

The following invariance law of central space has very important implications both theoretically and in application.

Theorem 1.1.1. *Let $\mathcal{S}_{Y|X}$ be the central space for (X, Y). Let $Z = AX + b$, where A is a $p \times p$ non-singular matrix and $b \in \mathbb{R}^p$,. Then*

$$\mathcal{S}_{Y|Z} = A^{-T} \mathcal{S}_{Y|X}.$$

This invariance property guarantees that, instead of using the original predictors at the X-scale for the purpose of dimension reduction, we can always use standardized X, or the Z-scale predictor, by a simple transformation $Z = \Sigma_X^{-1/2}[X - E(X)]$, where $\Sigma_X = \mathrm{Var}(X)$ is the covariance matrix of X. Z-scale predictor satisfies $E(Z) = 0, \mathrm{Var}(Z) = I_p$ and is much easier to dealt with in theoretical derivation. It has also been shown by simulation that Z-scale predictor is more stable than the X-scale predictors for computational purposes. We thus make the following assumption.

Assumption 1.1.1. *We assume that for the Z-scale predictor, we have*

$$E(Z) = 0, \mathrm{Var}(Z) = I_p.$$

Throughout this thesis, we will always deal with Z-scale predictor unless specified otherwise.

1.2 Estimation of central space

1.2.1 Two common assumptions

Before we introduce various dimension reduction methods for the estimation of central space, we first present the two most common assumptions in dimension reduction literature. The first one is the linear conditional mean (LCM) assumption.

Assumption 1.2.1. *Let β be a $\mathbb{R}^{p\times d}$ matrix whose columns form a basis in $\mathcal{S}_{Y|Z}$. We will assume that $E(Z|\beta^T Z)$ is a linear function of $\beta^T Z$; that is, the conditional mean of Z given $\beta^T Z$ is linear in $\beta^T Z$.*

It can be shown that when LCM assumption holds, $E(Z|\beta^T Z) = P_\beta Z$, where $P_\beta = \beta(\beta^T\beta)^{-1}\beta^T$ is the projection matrix onto the column space of β. This actually tells us that L-2 projection coincides with the Euclidean projection under LCM assumption. Z must have elliptical distribution if LCM assumption is true for any β. This assumption seems to be stringent. However, due to the discovery of Hall and Li (P. Hall and Li, 1993), we learn that LCM assumption is reasonable when p is large.

The second assumption is known as the constant conditional variance (CCV) assumption.

Assumption 1.2.2. *Let* β *be a* $\mathbb{R}^{p\times d}$ *matrix whose columns form a basis in* $\mathcal{S}_{Y|Z}$*. We assume that the conditional variance* $\mathrm{Var}(Z|\beta^T Z)$ *is a non-random matrix.*

Let $Q_\beta = I_\beta - P_\beta$ be the projection onto the orthogonal complement of $span(\beta)$. The CCV assumption above then implies $\mathrm{Var}(Z|\beta^T Z) = Q_\beta$. For classical dimension reduction methods, CCV assumption is always used together with the LCM assumption. Predictor Z has to be normally distributed when both assumptions hold for any β.

1.2.2 Some classical dimension reduction methods

In this section, we are going to see some classical dimension reduction methods. Those methods are the very first dimension reduction methods. They play a very important role in the dimension reduction literature. We will focus on four methods, Ordinary Least Squares (OLS; Li and Duan, 1989), Sliced Inverse Regression (SIR; Li, 1991), Sliced Average Variance Estimator (SAVE; Cook and Weisberg, 1991) and Principle Hessian Direction (PHD; Li, 1992 and Cook, 1998), which are directly related to the main topic of this thesis.

When LCM assumption holds, OLS estimator is shown to fall in the central space. Thus we can use $E(YZ)$ as an estimator of central space $\mathcal{S}_{Y|Z}$. This result is given by the following theorem.

Theorem 1.2.1. *(OLS) If Assumption 1.2.1 (LCM) holds, then*

$$E(ZY) \in \mathcal{S}_{Y|Z}.$$

PROOF. Let β be a $p \times d$ matrix whose columns form a basis in $\mathcal{S}_{Y|Z}$. Note that

$$E(ZY) = E[E(ZY|Z)] = E[ZE(Y|Z)]. \tag{1.1}$$

However, because $Y \perp\!\!\!\perp Z|\beta^T Z$, we have

$$E(Y|Z) = E(Y|Z, \beta^T Z) = E(Y|\beta^T Z).$$

Therefore, the right hand side of (1.1) is $E[ZE(Y|\beta^T Z)]$. However, because conditional expectation is a self-joint operator, we have

$$E[ZE(Y|\beta^T Z)] = E[E(Z|\beta^T Z)Y].$$

Now recall that, under LCM assumption, the L-2 projection $E(Z|\beta^T Z)$ coincides with the Euclidean projection $P_\beta Z$. Thus we have

$$E[E(Z|\beta^T Z)Y] = E[(P_\beta Z)Y] = P_\beta E(ZY).$$

Thus

$$E(ZY) = P_\beta E(ZY).$$

In other words, $E(ZY)$ equals its projection onto $\text{span}(\beta) = \mathcal{S}_{Y|Z}$. Therefore $E(ZY) \in \mathcal{S}_{Y|Z}$. □

SIR is a very important dimension reduction method. The next Theorem tells us that under the LCM assumption, the inverse regression vector $E(Z|Y = y)$ belongs to the central space $\mathcal{S}_{Y|Z}$.

Theorem 1.2.2. *(SIR) If Assumption 1.2.1 (LCM) holds, then for any y, $E(Z|Y = y) \in \mathcal{S}_{Y|Z}$.*

PROOF. Let β be the $p \times d$ matrix whose columns form a basis in $\mathcal{S}_{Y|Z}$. Then

$$E(Z|Y) = E[E(Z|Y, \beta^T Z)|Y].$$

Because $Y \perp\!\!\!\perp Z|\beta^T Z$, we have that

$$E[E(Z|Y, \beta^T Z)|Y] = E[E(Z|\beta^T Z)|Y].$$

The L-2 projection $E(Z|\beta^T Z)$ is the same as the Euclidean projection $P_\beta(Z)$ and therefore,

$$E(Z|Y) = E(P_\beta Z|Y) = P_\beta E(Z|Y).$$

In other words $E(Z|Y)$ belongs to span(β), which is the central space. □

In practice, we will use the discretized version of the above result. Let $I_1, ..., I_k$ be k intervals that partition Υ, the space of Y. Let $\widetilde{Y}$ be the discretized Y, defined by

$$\widetilde{Y} = i, \text{ if } Y \in I_i, i = 1, ..., k.$$

Apply Theorem 1.2.2 to $\widetilde{Y}$. We know $E(Z|\widetilde{Y} = i) = P_\beta E(Z|\widetilde{Y} = i)$. Consequently $\text{Var}(Z|\widetilde{Y} = i) = P_\beta \text{Var}[E(Z|\widetilde{Y} = i)]P_\beta$. Thus we can use the column space $\text{Var}[E(Z|\widetilde{Y} = i)]$ to estimate the central space $\mathcal{S}_{Y|Z}$.

SAVE is another method of estimating the central subspace based on slicing the response Y. Instead of calculating the mean within each slice as in SIR, this time we compute the variance. SAVE requires both the LCM and the CCV assumptions.

Theorem 1.2.3. *(SAVE) If Assumption 1.2.1 (LCM) and Assumption 1.2.2 (CCV) hold, then for any value of y, the column space of the matrix $I_p - \mathrm{Var}(Z|Y = y)$ is a subspace of the central space. Consequently, the column space of the matrix $E[I_p - \mathrm{Var}(Z|Y = y)]^2$ is a subspace of the central space $\mathcal{S}_{Y|Z}$.*

PROOF. Let β be the $p \times d$ matrix whose columns form a basis in $\mathcal{S}_{Y|Z}$. The LCM assumption implies that $E(Z|\beta^T Z) = P_\beta Z$ and the CCV assumption implies that $\mathrm{Var}(Z|\beta^T Z) = Q_\beta$. Thus we have

$$\begin{aligned}\mathrm{Var}(Z|Y) =& E[\mathrm{Var}(Z|Y, \beta^T Z)|Y] + \mathrm{Var}[E(Z|Y, \beta^T Z)|Y] \\ =& E[\mathrm{Var}(Z|\beta^T Z)|Y] + \mathrm{Var}[E(Z|\beta^T Z)|Y] \\ =& Q_\beta + P_\beta \mathrm{Var}(Z|Y) P_\beta.\end{aligned}$$

Hence $I - \mathrm{Var}(Z|Y) = P_\beta[I - \mathrm{Var}(Z|Y)]P_\beta$. The column space of the matrix $I_p - \mathrm{Var}(Z|Y)$ thus belongs to the range of the projection operator P_β, which is $\mathcal{S}_{Y|Z}$. Consequently, the column space of the matrix $E[I_p - \mathrm{Var}(Z|Y)]^2$ is a subspace of the central space $\mathcal{S}_{Y|Z}$. □

Again, in practice, we discretize Y to be $\widetilde{Y}$ as in SIR and use the sample estimate of column space $E[I_p - \mathrm{Var}(Z|\widetilde{Y})]^2$ as the SAVE estimator for the central space $\mathcal{S}_{Y|Z}$.

PHD is another dimension reduction method that requires both LCM and CCV assumptions. Let α be the OLS vector $E(ZY)$. Let e be the residual from the simple linear regression; that is $e = Y - \alpha^T X$. The matrix $H_1 = E(YZZ^T)$ is then called the y-based Hessian matrix and the matrix $H_2 = E(eZZ^T)$ is called the e-based Hessian matrix.

As before, we work with the standardized Z-scale predictor. In addition, we can always work with $Y - E(Y)$ instead of the original Y and the central space will not change: $\mathcal{S}_{Y|Z} = \mathcal{S}_{[Y-E(Y)]|Z}$. Thus we assume $E(Y) = 0$.

Theorem 1.2.4. *(PHD; y-based) If Assumption 1.2.1 (LCM) and 1.2.2 (CCV) hold, then the column space of* $H_1 = E(YZZ^T)$ *is a subspace of* $\mathcal{S}_{Y|Z}$.

PROOF. Let β be the $p \times d$ matrix whose columns form a basis in $\mathcal{S}_{Y|Z}$. Then

$$\begin{aligned} E(YZZ^T) =& E[E(YZZ^T|Z)] = E[E(Y|Z)ZZ^T] \\ =& E[E(Y|\beta^T Z)ZZ^T] = E[YE(ZZ^T|\beta^T Z)]. \end{aligned}$$

By the LCM and CCV assumptions, we have

$$\begin{aligned} E(ZZ^T|\beta^T Z) =& \operatorname{Var}(Z|\beta^T Z) + E(Z|\beta^T Z)E(Z^T|\beta^T Z) \\ =& Q_\beta + P_\beta ZZ^T P_\beta. \end{aligned}$$

Thus

$$\begin{aligned} H_1 =& E[YE(ZZ^T|\beta^T Z)] \\ =& E[Y(Q_\beta + P_\beta ZZ^T P_\beta)] \\ =& E(Y)Q_\beta + P_\beta E(YZZ^T)P_\beta \\ =& P_\beta H_1 P_\beta. \end{aligned}$$

This means the column space of $H_1 = E(YZZ^T)$ is a subspace of the central space $\mathcal{S}_{Y|Z}$. $\square$

The same thing can be said about e-based PHD estimator $H_2 = E(eZZ^T)$. The proof is similar and thus omitted.

1.3 Central mean subspace

In many situations, regression analysis is mostly concerned about inferring the conditional mean of the response given the predictors. In some cases, all the regression information is actually contained in the conditional mean $E(Y|X)$. Central mean subspace (CMS; Cook and Li, 2002) is designed to address this problem. Parallel to the development of dimension reduction space and central space, we have the following definitions of mean dimension reduction subspace and central mean subspace. As before, X is a p-dimensional predictor, and Y is the response.

Definition 1.3.1. *If $Y \perp\!\!\!\perp E(Y|X)|\alpha^T X$, then $\mathcal{S}(\alpha) = \mathrm{span}(\alpha)$ is a mean dimension reduction subspace for the regression of Y versus X.*

It follows from this definition that a dimension-reduction subspace is necessarily a mean dimension reduction subspace, because $Y \perp\!\!\!\perp X|\alpha^T X$ implies $Y \perp\!\!\!\perp E(Y|X)|\alpha^T X$.

Central mean subspace is the smallest mean dimension reduction subspace and is defined as follows.

Definition 1.3.2. *Let $\mathcal{S}_{E(Y|X)} = \cap \mathcal{S}_m$, where the intersection is over all mean dimension reduction subspaces $\mathcal{S}_m$. If $\mathcal{S}_{E(Y|X)}$ is itself a mean dimension reduction subspace, it is called the central mean subspace.*

Since any dimension-reduction subspace is a mean dimension reduction subspace, the central space must contain the central mean subspace. This is because the latter is the intersection of at least the same, if not more, subspaces. In other words, we always have $\mathcal{S}_{E(Y|X)} \subseteq \mathcal{S}_{Y|X}$. The central mean subspace also has the invariance property as does the central space.

Through the notion of central mean space, an apparent distinction among the four methods introduced in the previous section is discovered. It can be shown that OLS and PHD can only estimate the central mean space. On the other hand, SIR and SAVE estimators are in the central space but not necessarily in the central mean space. Those methods also require different set of assumptions. A summary is shown in Table 1.1:

	LCM	CCV
CMS	OLS	PHD
CS	SIR	SAVE

Table 1.1. Relationship between common CS and CMS estimators .

In a regression analysis, if we are mainly interested in the conditional mean and not the conditional distribution itself, then CMS is the parameter of interest and CS/CMS is the nuisance parameter.

1.4 Central kth moment DRS

Following the idea of the central mean space, central kth moment dimension reduction space (CKMS; Yin and Cook, 2002) is designed to aim dimension reduction at reducing the mean function, the variance function and up to the kth moment function, leaving the rest of regression Y versus X as the nuisance parameter.

We first define $M^{(k)}(Y|X) = E\{[Y - E(Y|X)]^k|X\}$ for $k \geq 2$ and $M^{(1)}(Y|X) = E(Y|X)$. Then we have the following set of definitions.

Definition 1.4.1. *If*

$$Y \perp\!\!\!\perp \{M^{(1)}(Y|X), \ldots, M^{(k)}(Y|X)\}|\eta^T X,$$

then $\mathcal{S}(\eta) = \mathrm{span}(\eta)$ *is a* k*th moment DRS for the regression of* Y *versus* X*.*

Definition 1.4.2. *Let* $\mathcal{S}_{Y|X}^{(k)} = \cap \mathcal{S}^{(k)}$*, where the intersection is over all* k*th moment DRSs* $\mathcal{S}^{(k)}$*. If* $\mathcal{S}_{Y|X}^{(k)}$ *is itself a* k*th moment DRS, it is called the central* k*th moment DRS, or CKMS for short.*

If $k = 1$ in the above definitions, they become exactly the same as the definitions of mean dimension reduction subspace and central mean subspace. Thus CKMS is a generalization of CMS.

We can also see from the definitions that a DRS is necessarily a kth moment DRS, which must be an ith moment DRS for any $i \leq k$. Just as CMS is always contained in the central space, the CKMS is also contained in the central space, because the former is the intersection of a larger collection of subspaces. These relationships can be summarized as below,

$$\mathcal{S}_{Y|X}^{(1)} \subseteq \cdots \subseteq \mathcal{S}_{Y|X}^{(k)} \subseteq \cdots \subseteq \mathcal{S}_{Y|X}.$$

We can see that if the conditional distribution of Y given X depends only on up to the kth moments of X, then $\mathcal{S}_{Y|X} = \mathcal{S}_{Y|X}^{(k)}$. Furthermore, when the moment-generating function of $Y|X$ exists, we have $\mathcal{S}_{Y|X} = \lim_{k \to \infty}(\mathcal{S}_{Y|X}^{(k)})$.

Chapter 2

Central Solution Space

In this chapter, the notion of central solution space is introduced. It is designed to loosen the linear conditional mean assumption, which is required by most existing dimension reduction methods. We will first study a set of estimators in the central kth moment space. These estimators will require LCM assumption to work properly.

2.1 Estimators in CKMS and their limitations

In this section, we are going to see how to find estimators in central kth moment dimension reduction space. We notice that the central kth moment space $\mathcal{S}_{Y|X}^{(k)}$ has invariance property, just like central space $\mathcal{S}_{Y|X}$ and central mean space $\mathcal{S}_{E(Y|X)}$. That is, if $W = A^T X$ for some invertible matrix A, then $\mathcal{S}_{Y|W}^{(k)} = A^{-1}\mathcal{S}_{Y|X}^{(k)}$. Thus we can always work with the Z-scale predictor and transform the estimated CKMS back to the original X-scale predictor.

The next theorem has the same flavor of Theorem 1.2.1, and it provides estimators comparable with OLS that falls in $\mathcal{S}_{Y|Z}^{(k)}$.

Theorem 2.1.1. *If Assumption 1.2.1 (LCM) holds, then*

$$E(ZY^k) \in \mathcal{S}^{(k)}_{Y|Z}.$$

Proof. Let γ be a basis for $\mathcal{S}^{(k)}_{Y|Z}$. Note that

$$E(ZY^k) = E[E(ZY^k|Z)] = E[ZE(Y^k|Z)].$$

The fact that γ is a basis for $\mathcal{S}^{(k)}_{Y|Z}$ implies $Y \perp\!\!\!\perp E(Y^k|Z)|\gamma^T Z$, which in turn implies

$$E(Y^k|Z) = E(Y^k|Z, \gamma^T Z) = E(Y^k|\gamma^T Z).$$

Hence

$$E(ZY^k) = E[ZE(Y^k|\gamma^T Z)] = E[E(Z|\gamma^T Z)Y^k] = E[(P_\gamma Z)Y^k] = P_\gamma E(ZY^k).$$

The third equation above holds because of the LCM assumption. Thus

$$E(ZY^k) = P_\gamma E(ZY^k).$$

In other words, $E(ZY^k)$ is the same as its projection onto span(γ). Therefore $E(ZY^k) \in \mathcal{S}^{(k)}_{Y|Z}$. □

The important role played by the LCM assumption is clear. This becomes the major limitation lies for those estimators. LCM assumption can not be made when we have non-elliptically distributed predictors, which are by no means uncommon in practice. Reweighting (Cook and Setodji, 1994) or transforming techniques can be used to counteract this limitation to a certain degree, but requires either intensive computation or only works on the marginal distribution of X. To cure this limitation when nonlinearity in the conditional mean $E(X|\beta^T X)$ does exist, we need to use the notion of central solution space (CSS; B. Li and Dong, 2008), which will be introduced in the next section.

2.2 Central solution space for SIR

Central solution space is a rather general idea that can be combined extensively with lots of existing dimension reduction methods. Instead of using the principal components of kernel matrixes to estimate the central space, CSS methods circumvent the linear conditional mean assumption by targeting directly at a set of solution equations.

Classical SIR estimator targets at estimating the column space of $\mathrm{Var}[E(Z|Y)]$. On the other hand, CSS-SIR method starts with the following SIR solution equation

$$E(Z|Y) = E[E(Z|\beta^T Z)|Y] \quad \text{a.s..} \tag{2.1}$$

Note that if $\beta \in \mathbb{R}^{p\times d}(d < p)$ solves this equation, then so does βA for any $d \times d$ nonsingular matrix A. Thus $\mathrm{span}(\beta)$, not β itself, is the parameter of interest. $\mathrm{Span}(\beta)$ is a *Solution Space* if β satisfies the solution equation (2.1) above.

It is easy to see that if β_1 satisfies (2.1) and β_2 is another matrix such that $\mathrm{span}(\beta_1) \subseteq \mathrm{span}(\beta_2)$, then β_2 also satisfies (2.1). For maximum dimension reduction we would like to seek β of the lowest rank. This leads to the following definition:

Definition 2.2.1. *If the intersection of any two solution spaces of (2.1) is itself a solution space of (2.1), then the intersection of all such spaces will be called the central solution space for SIR, or CSS-SIR for short, and is written as* $\mathcal{S}_{\mathrm{CSS-SIR}}$.

If we denote the column space of $\text{Var}[E(Z|Y)]$ by $\mathcal{S}_{\text{SIR}}$, then the next theorem reveals the relationship between SIR estimator, CSS-SIR estimator and the central space.

Theorem 2.2.1. *Suppose that Y and the elements of Z are square integrable. Then*

1. $\mathcal{S}_{\text{CSS}-\text{SIR}} \subseteq \mathcal{S}_{Y|Z}$.

2. *If, in addition, Assumption 1.2.1 (LCM) holds, then* $\mathcal{S}_{\text{SIR}} = \mathcal{S}_{\text{CSS}-\text{SIR}}$.

The proof of this Theorem will be skipped here. Later we are going to provide the proof in a similar case for the central kth moment solution space (CKMSS) estimators.

The key improvement of CSS-SIR estimators over classical SIR estimators is that the former does not require ellipticity of predictor X. Simulation has been done to verify the supremacy of CSS-SIR over SIR with non-elliptically distributed X. The same property holds for all other CSS based estimators. So far, first-moment based CSS methods such as CSS-SIR, CSS-PIR (parametric inverse regression; Bura and Cook, 2001) and CSS-KIR (kernel inverse regression; Zhu and Fang, 1996) has already been studied and second-moment based CSS methods such as CSS-SAVE and CSS-DR (directional regression; B. Li and Wang, 2007) are under development.

In the following chapters, I combine the idea of central solution space with the notion of central kth moment DRS. This will help cure the limitation of classical CKMS estimators, or their requirement of the LCM assumption.

Chapter 3

Central kth Moment Solution Space

Let X be a p-dimensional predictor and Y be a 1-dimensional response. We aim dimension reduction at obtaining information from the mean function, the variance function and up to the kth moment function. We also want to circumvent the limitation of linear conditional mean assumption. Population derivation of the central kth moment solution space (CKMSS) is provided in this chapter. Due to the invariance property of CKMS, we will standardize X and work with the Z-scale predictor.

3.1 Central mean solution space

As we have seen in Chapter 1, central mean space is a special case for the central kth moment space. In this section, we will first study a special case of central kth moment solution space with $k = 1$, which is the central mean solution space.

Definition 3.1.1. *A mean solution space for (Z, Y) is the column space of β, where β satisfies the following mean solution equation*

$$E(ZY) = E[E(Z|\beta^T Z)Y] \quad a.s.. \tag{3.1}$$

If the intersection of any two mean solution spaces is itself a mean solution space, then the intersection of all such spaces will be called the central mean solution space, and is written as $\mathcal{S}_{\text{CMSS}}$.

In the original paper about central solution space (B. Li and Dong, 2008), the space above was defined to be $\mathcal{S}_{\text{CSS}-\text{OLS}}$. Actually, it was shown that $\mathcal{S}_{\text{CMSS}} \subseteq \mathcal{S}_{Y|Z}$ and it can be used to estimate the central space. In addition, if LCM assumption holds, $\mathcal{S}_{\text{CMSS}}$ and the column space of OLS estimator $E(ZY)$ are the same.

More precisely, both OLS estimator and CMSS (CSS-OLS) estimator fall in the central mean space, or the central 1st moment space, $\mathcal{S}_{\text{CMSS}} \subseteq \mathcal{S}_{E(Y|X)} = \mathcal{S}^{(1)}_{Y|X}$.

3.2 Central kth moment solution space

3.2.1 Approach one: k separate solution equations

In this section, we are going to give k separate solution equations and base the definition of central kth moment solution space on those solution equations.

Definition 3.2.1. *The k-th order solution equation is defined to be*

$$E(ZY^k) = E[E(Z|\beta_k^T Z)Y^k] \quad a.s.. \tag{3.2}$$

For any β_k that satisfies equation (3.2), we call its column space k-th order solution space. Furthermore, if the intersection of any two k-th order solution spaces is itself k-th order solution space, then the intersection of all such spaces will be called the central k-th order solution space, and is denoted by $\mathcal{O}_k$. The central kth moment solution space is denoted by $\mathcal{S}^{(k)}_{\text{CKMSS}}$, and is the union of the first k central order solution space, or $\mathcal{S}^{(k)}_{\text{CKMSS}} = \cup_{i=1}^{k} \mathcal{O}_i$.

In this definition, we first define the k-th order solution equation, whose solution is the k-th order solution space. The intersection of such spaces is the central k-th order solution space $\mathcal{O}_k$. If we union $\mathcal{O}_1$, $\mathcal{O}_2$ up to $\mathcal{O}_k$, we will have the central kth moment solution space.

For $k = 1$, the central 1-st order solution space and the central 1-st moment solution space are the same, and they are both the central mean solution space defined in the previous section.

Before we move on to the next section, we will state the following Lemma. This Lemma is derived from Proposition 1 in the original CKMS (Yin and Cook, 2002) paper.

Lemma 3.2.1. *If γ is a basis for $\mathcal{S}^{(k)}_{Y|Z}$, then*

$$E(Y^j|Z) = E(Y^j|\gamma^T Z) \quad \textit{for} \;\; j\text{=}1,\ldots,k.$$

PROOF. Remember that in Chapter 1, we have $M^{(k)}(Y|Z) = E\{[Y - E(Y|Z)]^k|Z\}$ for $k \geq 2$ and $M^{(1)}(Y|Z) = E(Y|Z)$. As a basis for $\mathcal{S}^{(k)}_{Y|Z}$, γ satisfies

$$Y \perp\!\!\!\perp \{M^{(1)}(Y|Z), \ldots, M^{(k)}(Y|Z)\}|\gamma^T Z.$$

By Proposition 1 (Yin and Cook, 2002), this implies $E(Y^j|Z)$ is a function of $\gamma^T Z$ for $j = 1, \ldots, k$. Thus $E(Y^j|Z) = E(Y^j|Z, \gamma^T Z) = E(Y^j|\gamma^T Z)$ for $j = 1, \ldots, k$. □

3.2.2 Approach two: one solution equation for all

In the previous section, we have seen the first definition for the central kth moment solution space, which is based on k separate solution equations. In fact, we can define CKMSS in a more compact fashion, using only one solution equation.

Definition 3.2.2. *If γ satisfies*

$$E[Zf^k(Y)] = E[E(Z|\gamma^T Z)f^k(Y)] \quad a.s.$$

for any $f^k(Y)$, where $f^k(Y)$ is at most kth degree polynomial of Y, then we call the column space of γ a k-th moment solution space. Furthermore, if the intersection of any two k-th moment solution spaces is itself k-th moment solution space, then the intersection of all such spaces will be called the central k-th moment solution space, and is denoted by $\mathcal{S}^{(k)}_{\text{CKMSS}}$.

We now face the following question: Do Definition 3.2.2 above and Definition 3.2.1 from the previous Section 3.2.1 define the same central kth moment solution space? We need the following lemma first. This is a simple fact and its proof is omitted.

Lemma 3.2.2. *Let $f^k(Y)$ be any at most kth degree polynomial of Y and let $\mathbf{Y_k} = (Y, Y^2, \ldots, Y^k)^T$, then if γ satisfies any one of the following two equations, then it will also satisfy the other equation:*

1. $E[Zf^k(Y)] = E[E(Z|\gamma^T Z)f^k(Y)]$ *for any $f^k(Y)$;*
2. $E(Z\mathbf{Y_k}^T) = E[E(Z|\gamma^T Z)\mathbf{Y_k}^T]$.

Now we will start from a simple case with $k = 2$ and introduce one more Lemma.

Lemma 3.2.3. *The union of the central first order solution space $\mathcal{O}_1$ and the central second order solution space $\mathcal{O}_2$ is the same as the central second moment solution space defined by Definition 3.2.2.*

Proof. We have the first order solution equation

$$E(ZY) = E[E(Z|\beta_1^T Z)Y] \quad \text{a.s..} \tag{3.3}$$

$\mathcal{O}_1 = \cap\text{span}(\beta_1)$, where the intersection is over all β_1 that satisfies equation (3.3). Similarly, we have the second order solution equation

$$E(ZY^2) = E[E(Z|\beta_2^T Z)Y^2] \quad \text{a.s..} \tag{3.4}$$

$\mathcal{O}_2 = \cap\text{span}(\beta_2)$, where the intersection is over all β_2 that satisfies equation (3.4).

By Definition 3.2.2 and Lemma 3.2.2, with $\mathbf{Y_2} = (Y, Y^2)^T$, we have the central second moment solution space $\mathcal{S}^{(2)}_{\text{CKMSS}}$, which is the intersection of span(γ) over all γ that satisfies

$$E(Z\mathbf{Y_2}^T) = E[E(Z|\gamma^T Z)f^2\mathbf{Y_2}^T] \quad \text{a.s..} \tag{3.5}$$

Now we want to show

$$\cap\text{span}(\gamma) = [\cap\text{span}(\beta_1)] \cup [\cap\text{span}(\beta_2)].$$

On one hand, if β_1 satisfies equation (3.3), then $\cap\text{span}(\beta_1)$ satisfies equation (3.3). In addition, if β_2 satisfies equation (3.4), then $[\cap\text{span}(\beta_1)] \cup \beta_2$ satisfies both equation (3.3) and equation (3.4). In another word, $[\cap\text{span}(\beta_1)] \cup \beta_2$ satisfies equation (3.5). Thus

$$\cap\text{span}(\gamma) \subseteq \cap\{[\cap\text{span}(\beta_1)] \cup \beta_2\} = [\cap\text{span}(\beta_1)] \cup [\cap\text{span}(\beta_2)].$$

The inclusion $\cap\text{span}(\gamma) \subseteq \cap\{[\cap\text{span}(\beta_1)] \cup \beta_2\}$ above is true because the latter is an intersection of a smaller collection of spaces. The equation $\cap\{[\cap\text{span}(\beta_1)] \cup \beta_2\} = [\cap\text{span}(\beta_1)] \cup [\cap\text{span}(\beta_2)]$ above is the distributive laws of intersection and union.

On the other hand, for any γ which satisfies equation (3.5), it will satisfy equation (3.3) and equation (3.4) separately. Thus $\cap\text{span}(\beta_1) \subseteq \cap\text{span}(\gamma)$ and $\cap\text{span}(\beta_2) \subseteq$

$\cap\text{span}(\gamma)$, which will lead to $[\cap\text{span}(\beta_1)] \cup [\cap\text{span}(\beta_2)] \subseteq \cap\text{span}(\gamma)$. □

From Lemma 3.2.3 and mathematical induction, we can easily show the equivalence of two seemingly different versions of CKMSS from Definition 3.2.2 and Definition 3.2.1. This fact is stated in the following Theorem.

Theorem 3.2.1. *(Equivalence) If γ satisfies*

$$E(Zf^k(Y)) = E[E(Z|\gamma^T Z)f^k(Y)] \quad a.s. \tag{3.6}$$

for any $f^k(Y)$, where $f^k(Y)$ is at most kth degree polynomial of Y, then

$$\mathcal{S}^{(k)}_{\text{CKMSS}} = \cap\text{span}(\gamma) = \cup_{i=1}^{k}\mathcal{O}_i,$$

where $\mathcal{O}_i$ is the central i-th order solution space given in Definition 3.2.1.

PROOF. By Lemma 3.2.2, we know whenever γ satisfies equation (3.6), it will equivalently satisfy $E(Z\mathbf{Y_k}^T) = E[E(Z|\gamma^T Z)\mathbf{Y_k}^T]$, with $\mathbf{Y_k} = (Y, Y^2, \ldots, Y^k)^T$, and vice versa.

From Lemma 3.2.3, we know when $k = 2$, $\mathcal{S}^{(2)}_{\text{CKMSS}} = \mathcal{O}_1 \cup \mathcal{O}_2$ is true.

By mathematical induction, we only need to show that if $\mathcal{S}^{(k)}_{\text{CKMSS}} = \cup_{i=1}^{k}\mathcal{O}_i$ holds for k, $\mathcal{S}^{(k+1)}_{\text{CKMSS}} = \cup_{i=1}^{k+1}\mathcal{O}_i$ should hold for $k+1$.

Let $\mathbf{Y_k} = (Y, Y^2, \ldots, Y^k)^T$ and set

$$E(Z\mathbf{Y_k}^T) = E[E(Z|\beta_1^T Z)\mathbf{Y_k}^T] \quad \text{a.s.},$$

which is parallel to equation (3.3). We also have

$$E(ZY^{k+1}) = E[E(Z|\beta_2^T Z)Y^{k+1}] \quad \text{a.s.},$$

which is parallel to equation (3.4).

Use exactly the same reasoning in Lemma 3.2.3, we have

$$\mathcal{S}^{(k+1)}_{\text{CKMSS}} = [\cap \text{span}(\beta_1)] \cup [\cap \text{span}(\beta_2)].$$

By the assumption of the mathematical induction, we have

$$\cap \text{span}(\beta_1) = \mathcal{S}^{(k)}_{\text{CKMSS}} = \cup_{i=1}^{k} \mathcal{O}_i.$$

By the definition of the central k-th order solution space, $\cap \text{span}(\beta_2) = \mathcal{O}_{k+1}$. Thus $\mathcal{S}^{(k+1)}_{\text{CKMSS}} = (\cup_{i=1}^{k} \mathcal{O}_i) \cup \mathcal{O}_{k+1} = \cup_{i=1}^{k+1} \mathcal{O}_i$. □

So far we have seen the equivalence of two different versions of the central kth moment solution space. The next theorem tells us that the central kth moment solution space belongs to the central kth moment DRS, which in turn belongs to the central space.

Theorem 3.2.2. *(Inclusion)*

$$\mathcal{S}^{(k)}_{\text{CKMSS}} \subseteq \mathcal{S}^{(k)}_{Y|Z} \subseteq \mathcal{S}_{Y|Z}.$$

Proof. The second part $\mathcal{S}^{(k)}_{Y|Z} \subseteq \mathcal{S}_{Y|Z}$ is already shown in Chapter 1. We want to show that $\mathcal{S}^{(k)}_{\text{CKMSS}} \subseteq \mathcal{S}^{(k)}_{Y|Z}$.

Let γ be a basis for $\mathcal{S}^{(k)}_{Y|Z}$, then by Lemma 3.2.1, we know $E(Y^j|Z) = E(Y^j|\gamma^T Z)$ for $j = 1, \ldots, k$. Consequently, $E[f^k(Y)|Z] = E[f^k(Y)|\gamma^T Z]$, where $f^k(Y)$ is any at most kth degree polynomial of Y. Thus we have

$$E[f^k(Y)Z] = E\{E[f^k(Y)Z]|Z\}$$

$$
\begin{aligned}
&= E\{E[f^k(Y)|Z]Z\} \\
&= E\{E[f^k(Y)|\gamma^T Z]Z\} \\
&= E[f^k(Y)E(Z|\gamma^T Z)],
\end{aligned}
$$

which means spanγ is a kth moment solution space by Definition 3.2.2.

Hence $\mathcal{S}^{(k)}_{\text{CKMSS}} = \cap \text{span}\gamma \subseteq \mathcal{S}^{(k)}_{Y|Z}$. □

In this chapter, we have two sets of definitions for CKMSS and show their equivalency. The inclusion theorem shows that at the population level, the central k-th moment solution space falls in the central space without the LCM assumption.

Chapter 4

Sample Estimation of CKMSS

In this chapter, an existing estimation method (covariance subspace; Yin and Cook, 2002) of the central kth moment DRS is introduced first. We then show the population relationships between this covariance subspace and the central kth moment solution space. Parallel to the sample estimation algorithm of the covariance subspace, we provide the sample estimation algorithm for the CKMSS in the last section.

4.1 Covariance subspace estimation

Let X be a p-dimensional predictor, and Y be a 1-dimensional response. In Chapter 2, we have seen that when LCM assumption holds, $E(ZY^k) \in \mathcal{S}_{Y|Z}^{(k)}$. Thus $E(ZY^k)$ can be used as an estimate for the central kth moment DRS. As in the original CKMS (Yin and Cook, 2002) paper, we define the population kernel matrix

$$\mathbf{K} = (E(ZY), \ldots, E(ZY^k)).$$

The column space of this kernel matrix $\mathbf{K}$ is called the *kth order covariance subspace*:

$$\mathcal{S}_{cov}^{(k)} = \text{span}(\mathbf{K}).$$

Theorem 2.1.1 tells us that $\mathcal{S}_{cov}^{(k)} \subseteq \mathcal{S}_{Y|Z}^{(k)}$. Thus we can use the subspace spanned by the left singular vectors of $\mathbf{K}$ corresponding to its non-zero singular values to span an estimator of $\mathcal{S}_{Y|Z}^{(k)}$.

At the sample level, suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is an iid sample of (X, Y). $\bar{X}$ is the sample mean of the X vector. $\hat{\Sigma}_X$ is the sample covariance matrix estimate of X . Let $Z_i = \hat{\Sigma}_X^{-1/2}(X_i - \bar{X})$ for $i = 1, \ldots, n$. Then the sample estimate of the kernel matrix corresponding to $\mathbf{K}$ is:

$$\hat{\mathbf{K}} = (\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i, \ldots, \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i^k).$$

Assume $d = dim\{\text{span}(\mathbf{K})\}$ is known. Let $\hat{\eta}_1, \ldots, \hat{\eta}_d$ be the left singular vectors of $\hat{\mathbf{K}}$ corresponding to its d largest singular values $\hat{\lambda}_1, \ldots, \hat{\lambda}_d$. Then we have the estimate of span($\mathbf{K}$): $\hat{\mathcal{S}}(\mathbf{K}) = \text{span}(\hat{\eta}_1, \ldots, \hat{\eta}_d)$. Thus we get the sample covariance subspace estimates, which is also an estimate for $\mathcal{S}_{Y|Z}^{(k)}$.

4.2 Covariance subspace and CKMSS

In the original paper about central solution space (B. Li and Dong, 2008), it was shown that when the linear conditional mean assumption holds, the classical OLS estimator and the CSS-OLS estimator are the same. As a generalization with $k \geq 2$, we can show that kth order covariance subspace and the central kth moment solution space are the same when the LCM assumption holds.

First, we are going to present a lemma. Its proof is standard in central solution space literature.

Lemma 4.2.1. *Suppose that Y and the elements of Z are square integrable and Assumption 1.2.1 (LCM) holds. Then for $i = 1, \ldots, k$, the central i-th order solution space $\mathcal{O}_i$ equals the column space of $E(ZY^i)$.*

PROOF. On one hand, we want to show for $i = 1, \ldots, k$, $\mathcal{O}_i \subseteq \mathrm{span}[E(ZY^i)]$. Let β be a basis for $\mathrm{span}[E(ZY^i)]$. Then we have

$$E(ZY^i) = P_\beta E(ZY^i) = E[(P_\beta Z)Y^i].$$

LCM assumption implies $P_\beta Z = E(Z|\beta^T Z)$. Thus β satisfies

$$E(ZY^i) = E[E(Z|\beta^T Z)Y^i],$$

which is the k-th order solution equation (3.2). Thus

$$\mathcal{O}_i \subseteq \mathrm{span}(\beta) = \mathrm{span}[E(ZY^i)].$$

On the other hand, we want to show for $i = 1, \ldots, k$, $\mathrm{span}[E(ZY^i)] \subseteq \mathcal{O}_i$. Let γ be a basis for $\mathcal{O}_i$. Then we have

$$E(ZY^i) = E[E(Z|\gamma^T Z)Y^i],$$

which, in addition to the LCM assumption, leads to

$$E(ZY^i) = E[(P_\gamma Z)Y^i] = P_\gamma E(ZY^i) = \gamma(\gamma^T\gamma)^{-1}\gamma^T E(ZY^i).$$

This means γ is a basis for $\mathrm{span}[E(ZY^i)]$. Thus we have $\mathrm{span}[E(ZY^i)] \subseteq \mathcal{O}_i$. □

Theorem 4.2.1. *(Equation) If Assumption 1.2.1 (LCM) holds, then*

$$\mathcal{S}^{(k)}_{\text{CKMSS}} = \mathcal{S}^{(k)}_{cov}.$$

Proof. Notice we have two sets of definitions for CKMSS. The *Inclusion Theorem 3.2.2* is easier to prove by using Definition 3.2.2, where we have a single solution equation. However, this *Equation Theorem* here is easier to prove by using Definition 3.2.1, where we have multiple solution equations.

Notice that $\mathcal{S}^{(k)}_{cov} = \text{span}(\mathbf{K})$ is the column space of $\mathbf{K} = (E(ZY), \ldots, E(ZY^k))$ and CKMSS is the union of the first k central order solution space $\mathcal{S}^{(k)}_{\text{CKMSS}} = \cup_{i=1}^{k}\mathcal{O}_i$. By Lemma 4.2.1, we have $\text{span}[E(ZY^i)] = \mathcal{O}_i$ for $i = 1, \ldots, k$. Then we have $\mathcal{S}^{(k)}_{cov} = \text{span}(\mathbf{K}) = \mathcal{S}^{(k)}_{\text{CKMSS}}$ accordingly. □

4.3 Sample estimation of CKMSS

In this section, we will first introduce a population-level objective function whose minimizer yields the solution to

$$E(Z\mathbf{Y_k}^T) = E[E(Z|\gamma^T Z)\mathbf{Y_k}^T],$$

where$\mathbf{Y_k} = (Y, Y^2, \ldots, Y^k)^T$. Based on this objective function, we then provide a sample estimation algorithm of CKMSS.

4.3.1 Objective function

We will describe how to estimate the CKMSS with an iid sample of (X, Y). This estimation method is based on minimizing an objective function, the rationale of which is shown in the next theorem.

Theorem 4.3.1. *Suppose that* $\mathcal{S}^{(k)}_{\text{CKMSS}}$ *has dimension* $d \leq p$ *and let* β *be a* $p \times d$ *matrix whose columns form a basis in* $\mathcal{S}^{(k)}_{\text{CKMSS}}$. *Let* $\mathbf{Y_k} = (Y, Y^2, \ldots, Y^k)^T$. *Let* $f(\eta^T X)$ *be a square-integrable function such that, whenever* $\text{span}(\eta) = \text{span}(\beta)$, $f(\eta^T X) = E(X|\beta^T X)$, *and whenever* $\text{span}(\eta) \neq \text{span}(\beta)$,

$$P\left\{E[f(\eta^T X)\mathbf{Y_k}^T] \neq E[f(\beta^T X)\mathbf{Y_k}^T]\right\} > 0. \tag{4.1}$$

Let $\eta_0 \in \mathbb{R}^{p\times d}$ *be the minimizer of*

$$L(\eta) = E\left\|E\{[X - f(\eta^T X)]\mathbf{Y_k}^T\}\right\|^2 \tag{4.2}$$

over $\mathbb{R}^{p\times d}$. *Then* $\text{span}(\eta_0) = \mathcal{S}^{(k)}_{\text{CKMSS}}$.

Proof. If $\text{span}(\eta) = \text{span}(\beta)$, then

$$E[f(\eta^T X)\mathbf{Y_k}^T] = E[E(X|\beta^T X)\mathbf{Y_k}^T] = E(X\mathbf{Y_k}^T) \quad \text{a.s..}$$

Hence $L(\eta) = 0$. If $\text{span}(\eta) \neq \text{span}(\beta)$, then by assumption (4.1),

$$E\|E\{[f(\eta^T X) - f(\beta^T X)]\mathbf{Y_k}^T\}\|^2 > 0.$$

In the meantime,

$$
\begin{aligned}
&L(\eta) = E\|E\{[X - f(\beta^T X)]\mathbf{Y_k}^T\}\|^2 \\
&+ E\|E\{[f(\beta^T X) - f(\eta^T X)]\mathbf{Y_k}^T\}\|^2 \\
&+ 2E\left(E\{[X - f(\beta^T X)]\mathbf{Y_k}^T\}^T E\{[f(\beta^T X) - f(\eta^T X)]\mathbf{Y_k}^T\}\right).
\end{aligned}
$$

Because $\text{span}(\beta) = \mathcal{S}^{(k)}_{\text{CKMSS}}$, the last term is 0. Therefore

$$
L(\eta) \geq E\|E\{[f(\beta^T X) - f(\eta^T X)]\mathbf{Y_k}^T\}\|^2 > 0.
$$

Hence the minimizer of $L(\eta)$ must satisfy $\text{span}(\eta) = \text{span}(\beta) = \mathcal{S}^{(k)}_{\text{CKMSS}}$. □

4.3.2 Sample estimation algorithm

Let $f_1, \ldots, f_l$ be functions from $\mathbb{R}^d$ to $\mathbb{R}$. We will assume that $E(X|\beta^T X)$ lies in the space spanned by $f_1(\beta^T X), \ldots, f_l(\beta^T X)$. That is, each component of $E(X|\beta^T X)$ is a linear combination of $f_1(\beta^T X), \ldots, f_l(\beta^T X)$. Under this assumption the conditional expectation $E(X|\beta^T X)$ can be expressed explicitly as

$$
\begin{aligned}
E(X|\beta^T X) =& E[XG^T(\beta^T X)]\left\{E\left[G(\beta^T X)G^T(\beta^T X)\right]\right\}^{-1} G(\beta^T X), \quad \text{where} \\
&G(\beta^T X) = (f_1(\beta^T X), \ldots, f_l(\beta^T X))^T.
\end{aligned}
$$

Note that we are not assuming — and we do not need to assume — that $E(X|\eta^T X)$ is a linear function of $f_1(\eta^T X), \ldots, f_l(\eta^T X)$ for any η in $\mathbb{R}^{p\times d}$. All we need is that this holds at the true β. We use the function

$$
E[XG^T(\eta^T X)]\left\{E\left[G(\eta^T X)G^T(\eta^T X)\right]\right\}^{-1} G(\eta^T X) \tag{4.3}
$$

as the $f(\eta^T X)$ in the definition (4.2) of the objective function $L(\eta)$.

We now construct the sample estimate $L_n(\eta)$ of $L(\eta)$. Suppose that we have an iid sample of (X, Y): $(X_1, Y_1), \ldots, (X_n, Y_n)$. For a function $r(X, Y)$, let $E_n r(X, Y)$ denote the sample average $n^{-1} \sum_{i=1}^n r(X_i, Y_i)$.

1. Center $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_n$ as

$$\hat{Y}_i = Y_i - E_n(Y), \quad \hat{X}_i = X_i - E_n(X).$$

2. Select $\{f_1, \ldots, f_l\}$ that we deem sufficiently flexible to describe the conditional mean $E(X|\beta^T X)$. For example, based on our experience it often suffices to include linear and quadratic functions of $\beta^T X$. In this case, the set $\{f_1, \ldots, f_l\}$ includes the following $d(d+3)/2 + 1$ functions

$$\{1\} \cup \{\eta_i^T X : i = 1, \ldots, d\} \cup \{\eta_j^T X \, \eta_k^T X : \; 1 \leq j \leq k \leq d\},$$

where $\eta_1, \ldots, \eta_d$ are columns η. Let

$$\hat{f}(\eta^T \hat{X}) = E_n[\hat{X} G^T(\eta^T \hat{X})]\{E_n[G(\eta^T \hat{X}) G^T(\eta^T \hat{X})]\}^{-1} G(\eta^T \hat{X}).$$

3. Define $L_n(\eta) = E_n \|[\hat{X} - \hat{f}(\eta^T \hat{X})] \hat{\mathbf{Y}}_{\mathbf{k}}^T\|^2$, where $\hat{\mathbf{Y}}_{\mathbf{k}} = (\hat{Y}, \hat{Y}^2, \ldots, \hat{Y}^k)^T$.

4. Find $\eta_0 \in \mathbb{R}^{p \times d}$, the minimizer of $L_n(\eta)$. Then $\mathrm{span}(\eta_0) = \hat{\mathcal{S}}^{(k)}_{\mathrm{CKMSS}}$ is our sample estimator of the central kth moment solution space.

Chapter 5

Simulation Study

In this chapter, simulation study is done to compare the CKMSS estimators proposed in Chapter 4 and the existing covariance subspace estimators for CKMS. Two models are used for our simulation. For each model, two different scenarios are considered. We either have elliptical distribution of predictor X or have non-elliptically distributed X. Simulation results shows the supremacy of CKMSS estimators over CKMS estimators when the LCM assumption fails.

5.1 Model description

We consider the following models. The first one is a 1-dimensional model favorable for traditional OLS. The second model is borrowed from the CKMS (Yin and Cook, 2002) paper, where they show the effectiveness of CKMS on the population level.

$$\text{Model I:} \quad Y = X_1/4 + X_2 + \epsilon,$$

$$\text{Model II:} \quad Y = X_1 + X_1 X_2 + 0.1\epsilon,$$

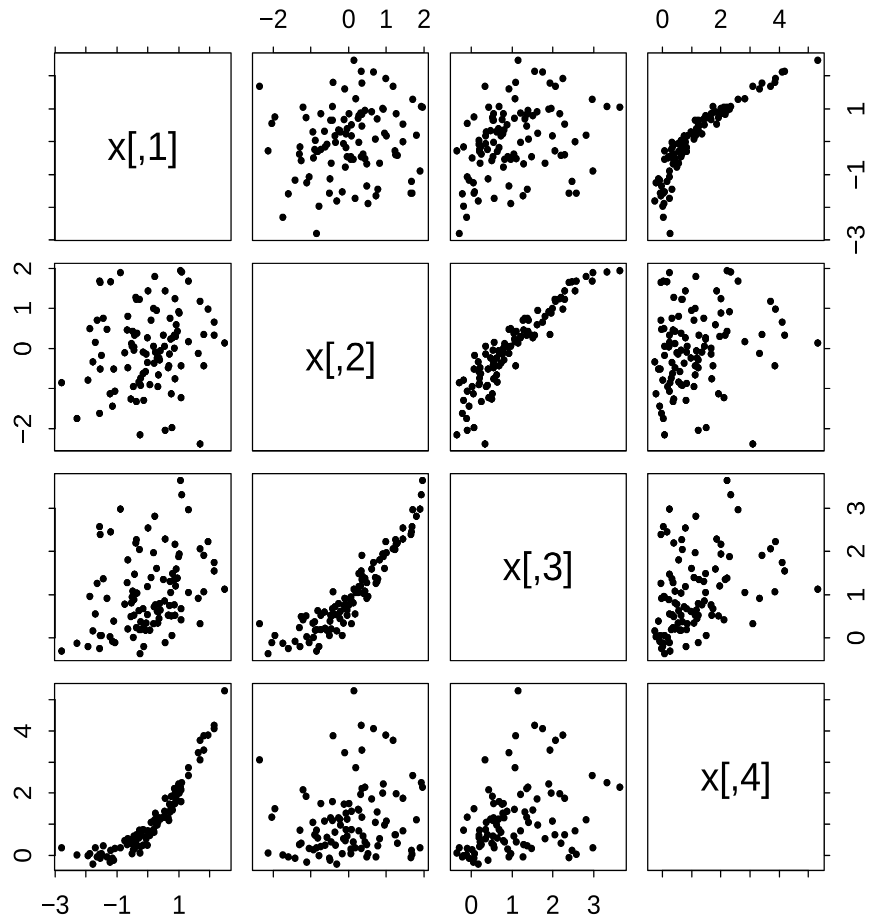

Figure 5.1. Scatter plot matrix for the 4-dimensional nonelliptically distributed predictor X.

where $\epsilon \sim N(0,1)$ and $\epsilon \perp\!\!\!\perp X$. The sample size is taken to be $n = 100$. The dimensions of X are chosen to be $p = 4, 5, 6$.

For each model, we either set X to be jointly standard multivariate normal (when LCM assumption is satisfied), or we introduce nonlinearity in the predictor as follows: $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$,

$$X_3 = 0.2X_1 + 0.2(X_2 + 2)^2 + 0.2\delta,$$

$$X_4 = 0.1 + 0.1(X_1 + X_2) + 0.3(X_1 + 1.5)^2 + 0.2\delta,$$

where $\delta \perp\!\!\!\perp (X, Y)$ and $\delta \sim N(0, 1)$. When $p = 5, 6$, X_5 and X_6 are taken to be independent $N(0, 1)$, and to be independent of $(X_1, \ldots, X_4)$. Figure 5.1 shows the scatter plot matrix of $X_1, \ldots, X_4$. We introduce nonlinearity among predictors like this according to the original CSS (B. Li and Dong, 2008) paper. Predictors of this type are common in practice.

5.2 Simulation result and interpretation

For each model, we do 100 simulations to compare the performance of CKMS (covariance subspace) estimators and the performance of CKMSS estimators we proposed.

Model I	Method	$p = 4$	$p = 5$	$p = 6$
Elliptical	CKMS	0.985 (0.0013)	0.984 (0.0010)	0.977 (0.0015)
X	CKMSS	0.993 (0.0004)	0.990 (0.0008)	0.986 (0.0012)
Non-elliptical	CKMS	0.965 (0.0020)	0.964 (0.0017)	0.964 (0.0017)
X	CKMSS	0.988 (0.0024)	0.988 (0.0010)	0.985 (0.0010)

Table 5.1. Model I: Relationship between CKMS and CKMSS estimators .

The simulation result of Model I is listed in Table 5.1. This is a 1-dimensional model. In the case of Model I, $\mathcal{S}_{Y|X} = \mathcal{S}^{(k)}_{Y|Z}$ with $k = 1$ and $dim(\mathcal{S}^{(k)}_{Y|Z}) = 1$. We measure the the estimation error by the absolute correlation coefficient between $\beta^T X$ and $\hat{\beta}^T X$, where β is the basis for the central kth moment space and $\hat{\beta}$ is its estimate. The closer this correlation is to 1, the better estimation we have got.

Each entry of Table 5.1 is formatted as $a(b)$, where a is the average of the above criterion across the 100 simulated samples, and b is the standard error of this average. From the table we see that in the absence of LCM assumption, or when we have non-elliptically distributed X, the CSS-based methods are more accurate than their classical counterparts across all 3 different values of p. Even when we do have elliptically distributed X and LCM assumption holds true, CKMSS estimators still outperform CKMS estimators. The CKMSS estimators result in larger correlation between $\hat{\beta}^T X$ and $\beta^T X$, and is more precise with smaller standard error.

CKMS in this case is actually the same as OLS, which is known to work well when the underling relationship between response Y and predictor X is linear. The simulation result of Model I shows the competency of CKMSS even when the model is favorable for CKMS.

For Model II, we use the squared multiple correlation coefficient to assess the accuracy of each method. Specifically, suppose U and V are d dimensional random vectors, and Σ_{UV}, Σ_U and Σ_V are the covariance matrix between U and V, the covariance matrix of U, and the covariance matrix of V, respectively. Then the square multiple correlation coefficient is defined by

$$\rho^2 = \text{trace}\left[\Sigma_U^{-1/2}\Sigma_{UV}\Sigma_V^{-1}\Sigma_{VU}\Sigma^{-1/2}\right] = \text{trace}\left[\Sigma_V^{-1/2}\Sigma_{VU}\Sigma_U^{-1}\Sigma_{UV}\Sigma^{-1/2}\right]. \quad (5.1)$$

See reference paper (W. J. Hall and Mathiason, 1990). This measure takes maximum value d if U and V have a linear relation, and takes minimum value 0 if the components of U and V are uncorrelated. At the sample level, given an estimator $\hat{\beta}$ of β, we use the sample version of the above measure based on

$$\{\hat{\beta}^T X_1, \ldots, \hat{\beta}^T X_n\} \text{ and } \{\beta^T X_1, \ldots, \beta^T X_n\}.$$

Note that the larger value of this criterion corresponds to a better dimension reduction estimate.

The simulation result of Model II is listed in Table 5.2. This is a 2-dimensional model. In this case, $\mathcal{S}_{Y|X} = \mathcal{S}_{Y|Z}^{(k)}$ with $k = 1$ and $dim(\mathcal{S}_{Y|Z}^{(k)}) = 2$. Thus the closer

Model II	Method	$p = 4$	$p = 5$	$p = 6$
Elliptical	CKMS	1.869 (0.0099)	1.808 (0.0129)	1.739 (0.0172)
X	CKMSS	1.932 (0.0082)	1.902 (0.0096)	1.869 (0.0129)
Non-elliptical	CKMS	1.750 (0.0049)	1.745 (0.0053)	1.741(0.0053)
X	CKMSS	1.964 (0.0047)	1.915(0.0092)	1.822 (0.0174)

Table 5.2. Model II: Relationship between CKMS and CKMSS estimators .

the squared multiple correlation is to 2, the better estimation we have got. Again, either in elliptical or non-elliptical predictor cases, compared with CKMS estimators, CKMSS estimators result in larger correlation between $\hat{\beta}^T X$ and $\beta^T X$. Meanwhile, the estimation standard errors of CKMSS estimators are comparable with CKMS estimators.

There is an important feature in both Table 5.1 and Table 5.2: when the distribution of predictor X changes from being elliptical to being non-elliptical, the accuracy of CKMS estimators degrades while the CKMSS estimators keep good performance.

Chapter 6

Conclusion

In this thesis, we have introduced the notion of central kth moment solution space. This is a method that combines the idea of central kth moment DRS and the idea of central solution space. It shares the properties of both CKMS and CSS estimators. It provides an estimation of CKMS $\mathcal{S}_{Y|X}^{(k)}$, but it does not require ellipticity of predictor X. A sample estimation algorithm is introduced based on minimization of a carefully constructed objective function. Simulation has been done to show that CKMSS estimators perform competitively with existing CKMSS (covariance subspace) estimators for a variety of different models.

Throughout this thesis, the working dimension d of $\mathcal{S}_{Y|X}^{(k)}$ is supposed to be already known. In practice, we need a good estimate of d before we can use this CKMSS estimation method. However, this is not an easy problem and is beyond the scope of this thesis. Typically, asymptotic distribution of CKMSS estimators need to be derived first. Then a sequence of hypothesis testing based on the known distribution of certain statistics is used to determine the dimension d of $\mathcal{S}_{Y|X}^{(k)}$.

References

Bura, E., and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, *63*, 393-410.

Cook, R. D. (1998). *Regression graphics: Ideas for studying regressions through graphics* (First ed.). New York: Wiley.

Cook, R. D., and Li, B. (2002). Dimension reduction for the conditional mean. *The Annals of Statistics*, *30*, 455-474.

Cook, R. D., and Setodji, C. M. (1994). Re-weighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association*, *89*, 592-599.

Cook, R. D., and Weisberg, S. (1991). Discussion of "sliced inverse regression for dimension reduction". *Journal of the American Statistical Association*, *86*, 316-342.

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

Hall, W. J., and Mathiason, D. J. (1990). On large-sample estimation and testing in parametric models. *International Statistical Review*, *58*, 77-79.

Li, B., and Dong, Y. (2008). Dimension reduction for non-elliptically distributed predictors. *The Annals of Statistics. To appear.*

Li, B., and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, *102*, 997-1008.

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

Li, K. C. (1992). On principal hessian directions for data visualization and dimension reduction: another application of stein's lemma. *Journal of the American Statistical Association*, *87*, 1025-1039.

Yin, X., and Cook, R. D. (2002). Dimension reduction for the conditional k-th moment in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, *64*, 159-175.

Yin, X., Li, B., and Cook, R. D. (2007). Successive direction extraction for estimating the central subspace in a multiple-index regression. *Journal of Multivariate Analysis. To appear.*

Zhu, L. X., and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics*, *3*, 1053-1068.