The Canadian Journal of Statistics
Vol. 42, No. 3, 2014, Pages 423–435
La revue canadienne de statistique

Heteroscedastic modelling via the autoregressive conditional variance subspace

Jin-Hong PARK[1]* and S. Yaser SAMADI[2]

[1]*Department of Mathematics, College of Charleston, Charleston, SC, USA*
[2]*Department of Statistics, University of Georgia, Athens, GA, USA*

Key words and phrases: Autoregressive central variance subspace; autoregressive conditional heteroscedasticity; financial time series; Kernel method; modified information criterion.

MSC 2010: Primary 62M10; secondary 62G07

Abstract: The paper deals with nonparametric estimation of the conditional variance of a time series based on a nonlinear autoregressive model in the squared innovation time series, which does not require specification of a model. We introduce a notion called the autoregressive central variance subspace (ACVS) to obtain the information included in the conditional variance of time series data. We use the squared time series to identify the ACVS by a nonparametric kernel method. In addition, we simultaneously estimate the unknown dimension and lag of the ACVS by a modified information criterion. Finally, we investigate the performance of all the estimators including the ACVS through simulations and a real analysis, which suggests implementing a new dimension reduction approach to modelling time series data that exhibits volatility. *The Canadian Journal of Statistics* 42: 423–435; 2014

Résumé: Les auteu rs traitent de l'estimation non paramétrique de la variance conditionnelle d'une série chronologique basée sur un modèle autorégressif non linéaire appliqué à la série des innovations au carré et ne nécessitant pas la spécification d'un modèle. Ils présentent la notion de sous-espace de variance centrale autorégressive (SVCA) afin d'extraire l'information contenue dans la variance conditionnelle des données chronologiques. Ils utilisent la série chronologique au carré pour identifier le SVCA à l'aide d'une méthode non paramétrique par noyau. Les auteurs estiment simultanément la dimension inconnue et le lag du SVCA en se basant sur un critère d'information modifié. Ils étudient finalement la performance de tous les estimateurs contenant le SVCA à l'aide de simulations et d'une analyse de données qui suggèrent l'adoption d'une nouvelle approche de réduction de la dimension pour la modélisation de séries chronologiques qui exhibent de la volatilité. *La revue canadienne de statistique* 42: 423–435; 2014

1. INTRODUCTION

Many experiments and observational studies involve data sets in which the observations are collected sequentially over time. The most common concern of time series analysis has been to draw inference about the conditional mean of the distribution of the current observation given the past of the series. However, nonstandard features in time series data, such as conditional heteroscedasticity, have induced researchers to look beyond the scope of the conditional mean. Engle (1982) and Bollerslev (1986) focussed on the parametric forms of the variance function. Since then, a variety of articles have investigated methods that might be used to handle conditional heteroscedasticity. Masry & Tjøstheim (1995) developed an estimation method for the variance function based on a nonparametric framework. Fan & Yao (2003) discussed modern parametric and nonparametric methods for analysing nonlinear models for the description of time series. To

* *Author to whom correspondence may be addressed.*
E-mail: parkj@cofc.edu

tackle the issue of dimension reduction for volatile time series, Xia, Tong, & Li (2002) considered a single-index volatility model that avoids the curse of dimensionality.

In this article, we develop a multiple-index dimension reduction method for volatile time series, where our aim is to make inferences about the conditional variance of the series given the past. Recently, Park, Sriram, & Yin (2009) introduced a formal inference approach for the conditional mean of the response given the predictors through a notion called the time series central mean subspace (TSCMS). In this paper, we develop a notion called the autoregressive central variance subspace (ACVS) and propose a method to estimate it when the lag and the dimension are known. While the development of our notion of ACVS bears similarity to TSCMS, the ACVS method provides us with a specific procedure for analysing time series data with volatility. In addition, we investigate how to estimate the minimum dimension and lag of the variance function, simultaneously. Finally, we illustrate our method via simulations for nonlinear variance function models and through analysis of a financial data set.

The rest of the article is organized as follows. In Section 2, we develop the notion of ACVS. In Section 3, we discuss the estimation method for ACVS when the dimension and lag of the series are known. In addition, we introduce a data dependent method to estimate unknown lag and dimension simultaneously. In Section 4, we establish the consistency of the estimator. Monte Carlo simulations for nonlinear heteroscedastic time series models and a real data analysis are given in Section 5. A brief conclusion is given in Section 6.

2. THE CENTRAL VARIANCE SUBSPACE IN TIME SERIES

To build a parametric model for time series data exhibiting volatility, autoregressive conditional heteroscedasticity (ARCH) models are useful. These models require the following: (i) estimation of the mean function based on a serial dependence such as in an autoregressive model; (ii) use of the resulting residual series (after removing the impact of linear dependence) to test for ARCH effects; and (iii) the fitting of a final model if there are significant ARCH effects and estimation of the variance function. In some cases, we may have to employ an autoregressive (AR) model or several explanatory variables. Since this paper focuses on the nonparametric approach for the variance function model, we do not restrict the mean function of the series. That is, we assume that the time series considered in this article are neither autoregressive nor dependent on explanatory variables.

Let (z_t) denote the squared innovation time series. Let $\mathbf{Z}_{t-1} = (z_{t-1}, \ldots, z_{t-k})^T$, where k is known. When we are interested in the conditional variance of a series, dimension reduction hinges on finding a $k \times s$ matrix $\mathbf{\Gamma}$, where $s \leq k$, so that the $s \times 1$ vector $\mathbf{\Gamma}^T\mathbf{Z}_{t-1}$ includes all the information about z_t that is available from the random quantity $E(z_t|\mathbf{Z}_{t-1})$. Hence, we need a definition of dimension reduction for the conditional variance of a time series.

Definition 1. *Let z_t be the square of a time series. Consider $\mathbf{Z}_{t-1} = (z_{t-1}, \ldots, z_{t-k})^T$, where k is known. If*

$$z_t \perp E(z_t|\mathbf{Z}_{t-1})|\mathbf{\Gamma}^T\mathbf{Z}_{t-1}, \tag{1}$$

where $\perp$ stands for independence, then the subspace spanned by the column vectors of $\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_k)$, denoted by $S(\mathbf{\Gamma})$, is a variance dimension reduction subspace for the time series $\sqrt{z_t}$.

Condition (1) implies that z_t is independent of $E(z_t|\mathbf{Z}_{t-1})$ given $\mathbf{\Gamma}^T\mathbf{Z}_{t-1}$. Similar to Zhu & Zhu (2009), in a regression context, the following theorem gives equivalent conditions for conditional independence which clarifies Definition 1.

Theorem 1. *Suppose z_t denotes a squared innovation time series. The following three statements are equivalent:*

(I) $z_t \perp E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}$.
(II) $Cov(z_t, E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}) = 0$.
(III) $E(z_t|\mathbf{Z}_{t-1})$ *is a measurable function of* $\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}$.

Part (I) of Theorem 1 is exactly the same as Definition 1 and says there is no correlation between z_t and $E(z_t|\mathbf{Z}_{t-1})$ given $\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}$, which is indicated in Part (II). If $E(z_t|\mathbf{Z}_{t-1})$ is a measurable function of $\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}$, then $E(z_t|\mathbf{Z}_{t-1}) = E(z_t|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1})$. Hence, part (III) implies part (I) and part (II). Therefore, the proof of Theorem 1 can be concluded by showing that part (II) implies part (III).

Proof. To prove that part (II) implies part (III), after expanding the LHS of part (II), we obtain

$$\begin{aligned} E[z_t E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] &= E[E\{z_t E(z_t|\mathbf{Z}_{t-1})|\mathbf{Z}_{t-1}\}|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] \\ &= E[\{E(z_t|\mathbf{Z}_{t-1})\}^2|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] \end{aligned}$$

and

$$\begin{aligned} E(z_t|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1})E[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] &= E[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}]E[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] \\ &= \{E[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}]\}^2. \end{aligned}$$

By (II), we have

$$\mathrm{Var}[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] = E[\{E(z_t|\mathbf{Z}_{t-1})\}^2|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] - \{E[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}]\}^2 = 0;$$

$\mathrm{Var}[E(z_t|\mathbf{Z}_{t-1})|\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}] = 0$ implies given $\boldsymbol{\Gamma}^T\mathbf{Z}_{t-1}$, $E(z_t|\mathbf{Z}_{t-1})$ is a constant. Therefore, part (II) implies part (III), which leads to the desired result for Theorem 1. ■

The above theorem says that any of the three conditions could be taken as the definition of a variance dimension reduction subspace for time series. Next, we define the smallest variance dimension reduction subspace for the squared series.

Definition 2. *Suppose $S_{E(z_t|\mathbf{Z}_{t-1})} = \cap S_C$, the intersection is over all variance dimension reduction subspaces S_C for the time series. If $S_{E(z_t|\mathbf{Z}_{t-1})}$ is itself a variance dimension reduction subspace of a time series, it is called the autoregressive central variance subspace (ACVS).*

3. ESTIMATION

We estimate a basis $(\gamma_1, \ldots, \gamma_m)^T$ where the minimal dimension m and the lag k of the population are presumed known. In this section, we focus on estimating the conditional variance function of a time series. Most well-known sufficient dimension reduction approaches such as sliced inverse regression (SIR, Li, 1991) and the sliced average variance estimation (SAVE, Cook & Weisberg, 1991), pertain to various contexts of regression and are not easily applicable to the context of dependent data, such as time series. The average derivative estimation (ADE) proposed by Härdle & Stoker (1989) is a simple approach that estimates the directions in central mean subspace based on a single-index model. However, ADE has some limitations; it is an inappropriate method when the gradient function is symmetric with expectation is zero, and also it cannot recover more than one direction. The minimum average variance estimation (MAVE, Xia et al., 2002) and the time

series central mean subspace (TSCMS, Park, Sriram, & Yin, 2009) are useful ways of addressing the problem of sufficient dimension reduction in time series. However, these approaches focus only on estimation of dimensions in the mean function. Fan et al. (2013) generalized MAVE, and introduced the minimum average contrast estimate (MACE) technique to estimate the vector of the index parameter and the link function simultaneously in a single index model. While Xia et al. (2002) estimated a local linear estimation of the mean function, Park, Sriram, & Yin (2009) estimated the mean function directly using the Nadaraya–Watson estimator. The TSCMS developed by Park, Sriram, & Yin (2009) has similarities to this new class of volatility model. However, there are clear distinctions: (i) our method can be used to make inferences with sufficient dimension reduction in the variance function that captures the volatility often appearing in financial data; (ii) we use the modified information criterion (MIC) to estimate lags k and the minimal dimensions m simultaneously; and (iii) we show that the estimator in our new set-up is consistent. Finally, the ACVS approach proposed in this article is a short and clear way of dealing with volatility, motivated by the central variance subspace (CVS) of Zhu & Zhu (2009) in the regression context.

Based on the notion of ACVS in Section 2, we consider a general time series model which does not require specific assumptions for the variance function

$$\sqrt{z_t} = |\varepsilon_t| \sqrt{h(\mathbf{\Gamma}_s^T \mathbf{Z}_{t-1})}, \tag{2}$$

where ε_t is a sequence of independent and identically distributed random variables with zero mean and unit variance, h is an unknown smoothing link function, $\mathbf{\Gamma}_s$ is a $k \times s$ matrix, where $s \leq k$, and $\mathbf{Z}_{t-1} = (z_{t-1}, \ldots, z_{t-k})^T$.

Suppose the dimension of ACVS is $dim(S_{E[z_t|\mathbf{\Gamma}^T \mathbf{Z}_{t-1}]}) = m$. We define an objective function $\mathbf{B}(\mathbf{b}_m) = E[z_t - h(\mathbf{b}_m^T \mathbf{Z}_{t-1})]^2$, where $h(\mathbf{b}_m^T \mathbf{Z}_{t-1}) = E(z_t | \mathbf{b}_m^T \mathbf{Z}_{t-1})$, and we minimize $\mathbf{B}$ with respect to $\mathbf{b}_m$ such that $\mathbf{b}_m^T \mathbf{b}_m = I_m$. Here, we assume the existence of the ACVS. It follows that, after we combine this assumption with (III) in Theorem 1, $\mathbf{B}(\mathbf{b}_m) \geq \mathbf{B}(I_k) = \mathbf{B}(\mathbf{\Gamma}_m)$ for $\mathbf{\Gamma}_m = \arg\min_{\mathbf{b}_m} \mathbf{B}(\mathbf{b}_m)$ for any $\mathbf{b}_m$. We use the Nadaraya–Watson smoother to estimate the unknown variance function h:

$$\hat{h}_{\kappa_n}(\mathbf{b}_m^T \mathbf{Z}) = \sum_{i=1}^{n} W_i z_i, \tag{3}$$

where the weight $W_i = \{K[(\mathbf{b}_m^T \mathbf{Z} - \mathbf{b}_m^T \mathbf{Z}_{i-1})/\kappa_n]\} / \{\sum_{j=1}^{n} K[(\mathbf{b}_m^T \mathbf{Z} - \mathbf{b}_m{}^T \mathbf{Z}_{j-1})/\kappa_n]\}$, K is a kernel function, and κ_n is a sequence of bandwidths. Our ultimate goal is to find the directions by minimizing a sample objective function, $\hat{\mathbf{B}}_n(\mathbf{b}_m) = (1/n) \sum_{t=1}^{n} [z_t - \hat{h}_{\kappa_n}(\mathbf{b}_m^T \mathbf{Z}_{t-1})]^2$ for $\mathbf{b}_m$ such that $\mathbf{b}_m{}^T \mathbf{b}_m = I_m$. Here, we estimate all directions simultaneously without assuming any model. Hence, our approach differs from the projection pursuit approach which assumed a model and successively estimated the mean function through projection pursuit components. As defined in (3), we use a Gaussian kernel for one dimension and a product Gaussian kernel for multiple dimensions. If we denote $\omega = (\omega_1, \ldots, \omega_m)^T$ and $\nu = (\nu_1, \ldots, \nu_m)^T$, then the product Gaussian kernel is $G[(\omega - \nu)/\kappa_n] = [1/(n \prod_{j=1}^{m} \pi_{nj})] \prod_{j=1}^{m} G[(\omega_j - \nu_j)/\pi_{nj}]$, where $\pi_{nj} = a_m s_j n^{-[1/(m+4)]}$ for $j = 1, \ldots, m$, s_j is the corresponding sample standard deviation of ν_j and $a_m = [4/(m+2)]^{-[1/(m+4)]}$, which comes from Silverman (1986). The Nadaraya–Watson estimator in (3) including the estimates defined above, replaces the variance function in $\mathbf{B}(\mathbf{b}_m)$. Finally, we minimize $\hat{\mathbf{B}}_n(\mathbf{b}_m)$ for all $k \times m$ matrices $\mathbf{b}_m$ where $\mathbf{b}_m^T \mathbf{b}_m = I_m$.

Another issue with ACVS is how to estimate the lag k of $\{z_t\}$ and the minimal dimension m, because both are not known for real data. To estimate m and k simultaneously, we use the modified information criterion (MIC). Broman & Speed (2002) and Zhu, Miao, & Peng (2006)

propose modifications of BIC via the penalty term on $\ln(n)$, the coefficient before $\ln(n)$, or both. While Park, Sriram, & Yin (2009) performed two-step inference, this paper proposes simultaneous estimation of m and k. More specifically, we define MIC as

$$(\hat{m}, \hat{k}) = \arg\min_{(m,k)}\{n \log[\hat{\mathbf{B}}_n(\hat{\mathbf{\Gamma}}_{m,k})/n] + mkn^{0.75}\}, \tag{4}$$

where $(\hat{m}, \hat{k})$ are simultaneously obtained by minimizing MIC over m and k. This procedure provides the best estimated pair of dimension $\hat{m}$ and lag $\hat{k}$ of $\{z_t\}$. In Section 5, we will illustrate the performance of our ACVS using a simple simulation study. Then, two simulations and a financial data analysis are provided to illustrate the use of MIC in determining m and lag k, and the estimation of ACVS.

4. CONSISTENCY

We establish the consistency of the estimator of the basis $\mathbf{\Gamma}_m = (\gamma_1, \ldots, \gamma_m)^T$ of ACVS. Instead of proving the consistency of the estimator of the basis vectors of ACVS by contradiction, we use a direct proof to show that $\hat{\mathbf{B}}_n(\mathbf{b}_m)$ is uniformly consistent for $\mathbf{B}(\mathbf{b}_m)$. That is, if $\hat{\mathbf{\Gamma}}_{n,m}$ and $\Gamma_{\mathbf{m}}$ minimize $\hat{\mathbf{B}}_n(\mathbf{b}_m)$ and $\mathbf{B}(\mathbf{b}_m)$, respectively, then $\hat{\mathbf{\Gamma}}_{n,m} \to \Gamma_{\mathbf{m}}$. Moreover, we provide sufficient conditions on the structure of the ACVS model (2) to ensure that the time series (2) is strongly mixing.

Assumption 1. *We make the following assumptions to prove Lemma 1 and Theorem 2.*

(A1) Suppose that $\sqrt{z_t}$ is a strictly stationary and strongly mixing sequence, with a geometric mixing coefficient $\alpha(t) = O(\varrho^t)$ for some $0 < \varrho < 1$.

(A2) Let $\boldsymbol{A}_k$ be a bounded subset of $\boldsymbol{R}^k$ and convex. Let $f(\mathbf{z})$ denote the common joint density of $\sqrt{\mathbf{Z}_{t-1}}$, such that $m < f(\mathbf{z}) < M$ for $\mathbf{z} \in \boldsymbol{A}_k^{\varepsilon}$, where m and M are two positive constants, and

$$\boldsymbol{A}_k^{\varepsilon} \equiv \{\mathbf{z} \in \boldsymbol{R}^k : \text{for some } \mathbf{x} \in \boldsymbol{A}_k, \parallel \mathbf{z} - \mathbf{x} \parallel \leq \varepsilon\}.$$

(A3) The first three directional derivatives of the density $f(\mathbf{z})$ in (A2) and $h(\mathbf{v})$ in (2) exist and are continuous uniformly in $\mathbf{z} \in \boldsymbol{R}^k$ and $\mathbf{v} \in \boldsymbol{R}^m$, respectively, and in all directions; see Hall (1989) and Xia & An (1999).

(A4) $E|\sqrt{z_t}|^{2s} < \infty$ for some $s > 2$.

(A5) The conditional densities $f_{\mathbf{z}_0|z_1}(\mathbf{z}_0|x)$, $f_{(\mathbf{z}_0,\mathbf{z}_{l-1},\mathbf{z}_{s-1},\mathbf{z}_{s+1})|(z_1,z_l)}(\mathbf{z}_0, \mathbf{z}_{l-1}, \mathbf{z}_{s-1}, \mathbf{z}_{s+1}|x, y)$, and $f_{(\mathbf{z}_0,\mathbf{z}_{l-1})|(z_1,z_l)}(\mathbf{z}_0, \mathbf{z}_{l-1}|x, y)$ are bounded by a constant for all l, $s > 1$.

(A6) The kernel function K is Lipschitz continuous and compactly supported with a bounded second derivative such that $\int v_i K(\mathbf{v})\,\mathrm{d}\mathbf{v} = 0$, $i = 1, \ldots, m$, and $\int \|\mathbf{u}\|^2 K(\mathbf{v})\,\mathrm{d}\mathbf{v} < \infty$.

Lipschitz continuity means that, for some $s > 0$ and $M > 0$, and all real u, v, $|K(u) - K(v)| \leq M|u - v|^s$. A function f is said to be uniformly continuous on A if for each $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in A$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

Lemma 1. *Under the assumptions (A1)–(A6), suppose the sequence κ_n of bandwidths is such that $\kappa_n = O(n^{-1/(m+4)})$. Then, we have*

$$\sup_{\underline{\mathbf{t}}_m \in \boldsymbol{R}^{k\times m}, \mathbf{z} \in \boldsymbol{A}_k} \left|\hat{h}_{\kappa_n}(\underline{\mathbf{t}}_m^T \mathbf{z}) - h(\underline{\mathbf{t}}_m^T \mathbf{z})\right| \to 0$$

with probability one, as $n \to \infty$.

Proof. The proof of the lemma follows from similar arguments to those in Corollary 3.2 of Masry & Tjøstheim (1995), Lemma 6.2, and equation (6.7) in Xia and An (1999). ■

Let K_0 be a compact set in $\mathbf{R}^k$, with random functions $\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m)$ taking values in $C(K_0)$, the space of continuous functions on K_0. Define $\hat{\mathbf{B}}_n(\hat{\mathbf{\Gamma}}_{n,m}) = \inf_{\underline{\mathbf{t}}_m \in \mathbf{R}^{k\times m}} \hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m)$, and $\mathbf{B}(\Gamma_\mathbf{m}) = \inf_{\underline{\mathbf{t}}_\mathbf{m} \in \mathbf{R}^{\mathbf{k}\times\mathbf{m}}} \mathbf{B}(\underline{\mathbf{t}}_\mathbf{m})$. To prove the consistency of $\hat{\mathbf{\Gamma}}_{n,m}$, we first show that $\hat{\mathbf{B}}_n$ is close to $\mathbf{B}$ in sup norm.

Corollary 1. *Under the assumptions of Lemma 1, we have*

$$\sup_{\underline{\mathbf{t}}_m \in \boldsymbol{R}^{k\times m}} \left|\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) - \mathbf{B}(\underline{\mathbf{t}}_m)\right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty.$$

Proof. Write

$$\begin{aligned}\sup_{\underline{\mathbf{t}}_m \in \mathbf{R}^{k\times m}} \left|\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) - \mathbf{B}(\underline{\mathbf{t}}_m)\right| \le & \sup_{\underline{\mathbf{t}}_m \in \mathbf{R}^{k\times m}} \frac{1}{n}\Big|\sum_{t=1}^{n}[z_t - \hat{h}_{\kappa_n}(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})]^2 - \sum_{t=1}^{n}[z_t - h(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})]^2\Big| \\ & + \sup_{\mathbf{t}_m \in \mathbf{R}^{k\times m}} \Big|\frac{1}{n}\sum_{t=1}^{n}[z_t - h(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})]^2 - \mathbf{B}(\underline{\mathbf{t}}_m)\Big| \\ = & (I) + (II).\end{aligned}$$

By assumption (A1) and the ergodic theorem, $(II) \to 0$ with probability one. For (I), we use the identity $a^2 - b^2 = (a-b)^2 + 2b(a-b)$, Lemma 1, the Cauchy–Schwarz inequality for the cross product, and the fact that $(II) \to 0$ to show that $(I) \to 0$ with probability one as $n \to \infty$. Therefore, the nonparametric estimator $\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m)$ is consistent. ■

Theorem 2. *Let* $\hat{\mathbf{\Gamma}}_{n,m} = \arg\min_{\underline{\mathbf{t}}_m} \hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m)$ *and* $\Gamma_\mathbf{m} = \arg\min_{\underline{\mathbf{t}}_\mathbf{m}} \mathbf{B}(\underline{\mathbf{t}}_\mathbf{m})$, *where* $\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) = (1/n)\sum_{t=1}^{n}[z_t - \hat{h}_{\kappa_n}(\mathbf{b}_m^T\mathbf{Z}_{t-1})]^2$, $\mathbf{B}(\underline{\mathbf{t}}_m) = E[z_t - h(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})]^2$, $h(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1}) = E(z_t|\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})$, *and* $\hat{h}_{\kappa_n}(\underline{\mathbf{t}}_m^T\mathbf{Z}_{t-1})$ *is defined in* (3) *for all* $k \times m$ *matrices* $\underline{\mathbf{t}}_m$ *such that* $\underline{\mathbf{t}}_m^T\underline{\mathbf{t}}_m = I_m$. *Then under the assumptions of Lemma* 1, $\hat{\mathbf{\Gamma}}_{n,m}$ *converges to* $\Gamma_\mathbf{m}$ *with probability one as* $n \to \infty$, *where the distance of convergence is* $\|(I - \hat{\mathbf{\Gamma}}_{n,m}\hat{\mathbf{\Gamma}}_{n,m}^T)\Gamma_\mathbf{m}\|$.

Proof. We use Corollary 1 and arguments similar to the second assertion of Theorem 9.4 from Keener (2010). Note, however, that the constraint $\underline{\mathbf{t}}_m^T\underline{\mathbf{t}}_m = I_m$ does not guarantee that the function $\mathbf{B}(\underline{\mathbf{t}}_m)$ has a unique minimum, but the subspace corresponding to it is unique. For identification of the parameters, we may consider any basis matrix that minimizes the objective function by its orthogonal projection matrix, which is unique. Therefore, without loss of generality, we assume that the matrix solution is unique.

The goal is to show that $\hat{\mathbf{\Gamma}}_{n,m} \to \Gamma_\mathbf{m}$. To this end, fix ε and let $K_0^\varepsilon = K_0 - B_\varepsilon(\Gamma_\mathbf{m})$, where $B_\varepsilon(\mathbf{\Gamma}_m) = \bigcap_{i=1}^{m} B_{d_i}(\gamma_i)$, and $B_{d_i}(\gamma_i)$ is defined as

$$B_{d_i}(\gamma_i) \equiv \left\{\gamma^* \in \mathbf{R}^k;\ \sqrt{\sum_{j=1}^{k}(\gamma_j^* - \gamma_{ij})^2} < d_i\right\},$$

where γ_i is the ith column of the perpendicular $k \times m$ matrix $\Gamma_\mathbf{m}$. Then, the set K_0^ε is compact; it is bounded because K_0 is bounded, and it is closed because it is the intersection of two closed sets,

K_0 and the complement of $B_\varepsilon(\Gamma_{\mathbf{m}})$. Let $M = \inf_{K_0} \mathbf{B}(\underline{\mathbf{t}}_m) = \mathbf{B}(\Gamma_{\mathbf{m}})$ and let $M_\varepsilon = \inf_{K_0^\varepsilon} \mathbf{B}(\underline{\mathbf{t}}_m)$. Since K_0^ε is compact, $M_\varepsilon = \mathbf{B}(\underline{\mathbf{t}}_m^\varepsilon)$ for some $\underline{\mathbf{t}}_m^\varepsilon \in K_0^\varepsilon$, and since $\mathbf{B}(\underline{\mathbf{t}}_m)$ has a unique minimum over K_0, $M < M_\varepsilon$. Define $\delta = M_\varepsilon - M > 0$. Under Corollary 1, we have

$$\mathbf{B}(\underline{\mathbf{t}}_m) - \delta/2 < \hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) < \mathbf{B}(\underline{\mathbf{t}}_m) + \delta/2.$$

Then,

$$\inf_{K_0^\varepsilon} \hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) > \inf_{K_0^\varepsilon} \mathbf{B}(\underline{\mathbf{t}}_m) - \delta/2 = M + \delta/2,$$

and

$$\hat{\mathbf{B}}_n(\hat{\mathbf{\Gamma}}_{n,m}) = \inf_{K_0} \hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) \le \inf_{K_0} \hat{\mathbf{B}}_n(\mathbf{\Gamma}_m) < \mathbf{B}(\Gamma_{\mathbf{m}}) + \delta/2 = M + \delta/2,$$

and $\hat{\mathbf{\Gamma}}_{n,m}$ must lie in $B_\varepsilon(\Gamma_{\mathbf{m}})$. Thus,

$$P\big(\sup|\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) - \mathbf{B}(\underline{\mathbf{t}}_m)| < \delta/2\big) \le P\left(\left|\sqrt{\sum_{j=1}^k \hat{\gamma}_{ij}^2} - \sqrt{\sum_{j=1}^k \gamma_{ij}^2}\right| < \varepsilon,\ \forall i = 1, \ldots, m\right),$$

where $\hat{\gamma}_{ij}$ is the ijth element of $\hat{\mathbf{\Gamma}}_{n,m}$. Taking complements for every $i = 1, \ldots, m$, gives

$$P\left(\left|\sqrt{\sum_{j=1}^k \hat{\gamma}_{ij}^2} - \sqrt{\sum_{j=1}^k \gamma_{ij}^2}\right| \ge \varepsilon\right) \le P\big(\sup|\hat{\mathbf{B}}_n(\underline{\mathbf{t}}_m) - \mathbf{B}(\underline{\mathbf{t}}_m)| \ge \delta/2\big) \to 0$$

and so $\hat{\mathbf{\Gamma}}_{n,m}$ converges to $\Gamma_{\mathbf{m}}$ with probability one as $n \to \infty$. ■

The following Lemma gives sufficient conditions to verify (A1) above.

Lemma 2. *Suppose the following conditions are satisfied for the ACVS model (2).*

(B1) The i.i.d. random variables $|\varepsilon_t|$ have a probability density function which is positive over $[0, \infty]$.

(B2) The function h is positive, nonperiodic, and bounded on compact sets for all $\mathbf{v} \in \boldsymbol{R}^m$.

(B3) There exists a vector $\mathbf{d} = (d_1, \ldots, d_m)$ with $d_i \ge 0$ and $\max_i d_i < 1$, $i = 1, \ldots, m$, such that, as $\sum v_i \to \infty$,

$$h(\mathbf{v}) = \sum_{j=1}^m d_j v_j + o\left(\left(\sum v_i\right)^2\right).$$

Then, the nonlinear ACVS model $\sqrt{z_t}$ in (2) is strongly mixing with geometric mixing coefficient $\alpha(t) = O(\varrho^t)$ for some $0 < \varrho < 1$.

Proof. The proof of the Lemma essentially follows from arguments similar to that for Lemma 3.1 in Masry & Tjøstheim (1995). ■

Remark. *Note that, there is a slight difference in notation between this paper and the paper by Masry & Tjøstheim (1995). In the Masry and Tjøstheim paper, the response variable (Y_t) can take any values on the real line whereas here the response variable ($\sqrt{z_t}$) is nonnegative; however the proof still works with this notation.*

TABLE 1: Model 1: Average values of accuracy measure ρ based on 100 Monte Carlo replications.

n	k	m	ρ_1	ρ_2
100	4	1	0.8554	0.8328
200	4	1	0.9052	0.8374
300	4	1	0.9033	0.8333

Here, ρ_1 and ρ_2 are the results from the standard normal and $t(8)$ innovations, respectively.

5. ILLUSTRATIVE RESULTS

In this section, we will perform three simulation studies and a real example to demonstrate the performance of our method. One simulated example will focus only on the variance function for the accuracy of ACVS, and the other two simulations perform the simultaneous estimation of the dimension and the lag discussed in Section 2. In addition, we analyse the *CBOE S&P 500 3-Month Volatility Index* data to illustrate the estimation procedure of the variance function in the financial time series.

5.1. Accuracy of ACVS

The true values are known in the simulated data sets. Hence, we can measure the distance between the real and estimated subspaces. This will assess the accuracy of the ACVS estimate. Similar to Yoo (2011) (or Ye & Weiss, 2003), we use the vector correlation coefficient (Hotelling, 1936) to measure the distance between the estimated ACVS $S_{z_t|\mathbf{Z}_{t-1}}(\hat{\mathbf{\Gamma}}_m)$ and the true ACVS $S_{z_t|\mathbf{Z}_{t-1}}(\mathbf{\Gamma}_m)$. This distance measure is $\rho = \sqrt{|\prod_{j=1}^{n} \varphi_j^2|}$, where $0 \le \varphi_n^2 \le \cdots \le \varphi_1^2 \le 1$ and φ is the eigenvalue of $\hat{\mathbf{\Gamma}}_{n,m}^T \mathbf{\Gamma}_m \mathbf{\Gamma}_m^T \hat{\mathbf{\Gamma}}_{n,m}$. Then, ρ is defined as the correlation between the two subspaces where $0 \le \rho \le 1$. Finally, for $\rho = 1$, $S_{z_t|\mathbf{Z}_{t-1}}(\hat{\mathbf{\Gamma}}_m) = S_{z_t|\mathbf{Z}_{t-1}}(\mathbf{\Gamma}_m)$. That is, the higher value of ρ means that our estimates are more accurate estimates, closer to the real variance dimension reduction subspace. The simulation model below is used to assess the accuracy of our ACVS estimates. Samples of size n = 100, 200, and 300 are considered, and the accuracy of estimates is based on 100 Monte Carlo replications.

Model 1: Consider the linear variance function of squared series

$$\sqrt{z_t} = 0.5|\varepsilon_t|\sqrt{1 + (1/\sqrt{0.1^2 + 4^2})(0.1z_{t-1} + 4z_{t-4})},$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically normal random variables with zero mean and unit variance. In addition, we check the accuracy of this model with a Student-t innovation such as $t(8)$ because many financial time series have heavier tails than for normal error innovations. In this simulation, we first randomly select 100 initial values for model parameters and, for each initial set of values, we minimize our sample version of the objective function $\hat{\mathbf{B}}_n(\hat{\mathbf{\Gamma}})$ and choose the initial value set for which the value is the smallest. We then estimate $\mathbf{\Gamma}_1$ and measure the accuracy of the estimate using the method mentioned above. Table 1 gives the average values of accuracy measures ρ for fixed $m = 1$ and $k = 4$. It demonstrates that the average values of ρ are in general close to 1, implying that our estimates of $\mathbf{\Gamma}_1$ are accurate in both normal and t innovations. Hence, this result suggests that the accuracy of our ACVS is reasonably good for different sample sizes.

TABLE 2: Model 2: The results are based on 100 Monte Carlo replications.

	m/k	2	3	4	5
100	1	14	0	*68	6
	2	0	0	12	0
	3	0	0	0	0
	4	0	0	0	0
200	1	7	0	*70	8
	2	0	0	14	1
	3	0	0	0	0
	4	0	0	0	0
300	1	4	0	*72	8
	2	0	0	15	1
	3	0	0	0	0
	4	0	0	0	0

Frequencies of the estimated dimensions and lags are reported using the MIC in (4). True dimension and lag are $m = 1$ and $k = 4$, respectively. * The frequency when $(\hat{m}, \hat{k}) = (1, 4)$.

5.2. Detection of Dimension and Lag of ACVS

In the following simulations, we detect the minimal dimension m and the lag k simultaneously using linear and nonlinear variance function models.

Model 2: Consider the linear variance function of squared series

$$\sqrt{z_t} = 0.5|\varepsilon_t|\sqrt{3 + z_{t-4}},$$

where $k = 4$, $m = 1$, and $\{\varepsilon_t\}$ is a sequence of independent and identically normal random variables with zero mean and unit variance. After randomly selecting 10 initial values, we minimize our sample version of the objective function $\hat{\mathbf{B}}_n(\hat{\mathbf{\Gamma}})$ for each initial set of values and choose the initial value set for which the value is the smallest. We have done 100 Monte Carlo replications for the sample size n = 100, 200, and 300. Based on the MIC in (4), Table 2 gives the frequencies over several dimension and lag combinations. Here, we used dimension $m = 1, 2, 3, 4$ and lags $k = 2, 3, 4, 5$ to find which combination of m and k has the largest frequency. The results of Table 2 show that the relative frequencies for the estimated dimension $\hat{m} = 1$ and lag $\hat{k} = 4$ for $n = 100$, 200, and 300 are 68%, 70%, and 72%, respectively. The results indicate that our procedure works well for this linear variance function model of the squared series.

Model 3: Consider the nonlinear variance function of squared series

$$\sqrt{z_t} = 0.5|\varepsilon_t|\sqrt{1 + (1/\sqrt{2})(z_{t-1} + z_{t-4}) + \cos\{0.1 + (1/\sqrt{2})(z_{t-2} + z_{t-3})\}},$$

where $k = 4$, $m = 2$, and $\{\varepsilon_t\}$ is a sequence of independent and identically normal random variables with zero mean and unit variance. The replications, sample sizes, and experimental dimensions/lags are all the same as those of Model 2. Table 3 demonstrates the frequencies for sixteen combinations of $m = 1, 2, 3, 4$ and $k = 2, 3, 4, 5$, based on the MIC in (4). Finally, the results of Table 3 show that the relative frequencies for the estimated dimension $\hat{m} = 2$ and

TABLE 3: Model 3: The results are based on 100 Monte Carlo replications.

	m/k	2	3	4	5
100	1	16	1	6	3
	2	2	15	*29	10
	3	0	0	5	8
	4	1	1	1	2
200	1	4	1	0	2
	2	3	9	*44	11
	3	0	1	15	9
	4	0	0	0	1
300	1	1	1	1	0
	2	0	3	*49	12
	3	0	1	22	9
	4	0	1	0	0

Frequencies of the estimated dimensions and lags are reported using the MIC in (4). True dimension and lag are $m = 2$ and $k = 4$, respectively. * The frequency when $(\hat{m}, \hat{k}) = (2, 4)$.

lag $\hat{k} = 4$ for $n = 100$, 200, and 300 are 29%, 44%, and 49%, respectively. Due to the higher dimension and the more complicated model of the nonlinear variance function, the frequencies are comparatively lower than those of Model 2. However, the accuracy for detecting $\hat{m} = 2$ and $\hat{k} = 4$ increases as the sample size increases. Hence, this result attests to the fact that our estimation procedure performs reasonably well for the nonlinear variance function model of the squared series.

6. REAL DATA

To examine whether our proposed method works well in a real data example, we apply it to the data set, CBOE S&P 500 3-Month Volatility Index, released by CBOE Market Statistics. The unit of this index is percent, and the source is the Chicago Board Options Exchange. The data set under discussion gives the daily series from August 29, 2008, to June 30, 2009 (210 observations), collected from the Federal Reserve Bank of St. Louis. (http://research.stlouisfed.org/fred2/series/VXVCLS?rid=200). As displayed in Figure 1, the selected data are affected by the financial crisis, which induces volatility.

The goal of this real data analysis is to assess the performance of our ACVS method. We use the first differenced CBOE S&P 500 3-Month Volatility Index to compare our method with the conventional approach. Our process begins with estimation of m and k for the squared series, followed by the estimation of the ACVS. To estimate m and k simultaneously, we compute MIC with respect to $k = 1, 2, 3, 4$, and $m = 1, 2, 3, 4$. Table 4 lists the MIC values for these lags and dimensions. We found the smallest MIC value when $\hat{m} = 1$ and $\hat{k} = 2$. That is, we conclude that the minimal dimension $m = 1$ and the lag $k = 2$ for the series. Setting $m = 1$ and $k = 2$, we use our estimation method to obtain an estimate, $\hat{\mathbf{\Gamma}}_1 = \hat{\gamma}_1$, of the 2×1 basis matrix $\mathbf{\Gamma}_1$ in $S_{E(z_t|\mathbf{Z}_{t-1})}(\mathbf{\Gamma}_1)$.

Next, we examined the plot of z_t against $d_t(= \hat{\gamma}_1 \mathbf{Z}_{t-1})$ to build a variance function model. Because the one dimensional plot of d_t versus z_t appears as linear and nonlinear patterns, we regress the predictors d_t, d_t^2 and cos_t where $cos_t = \cos(1.0034 d_{1t} + 1.3876)$. Finally, this led us

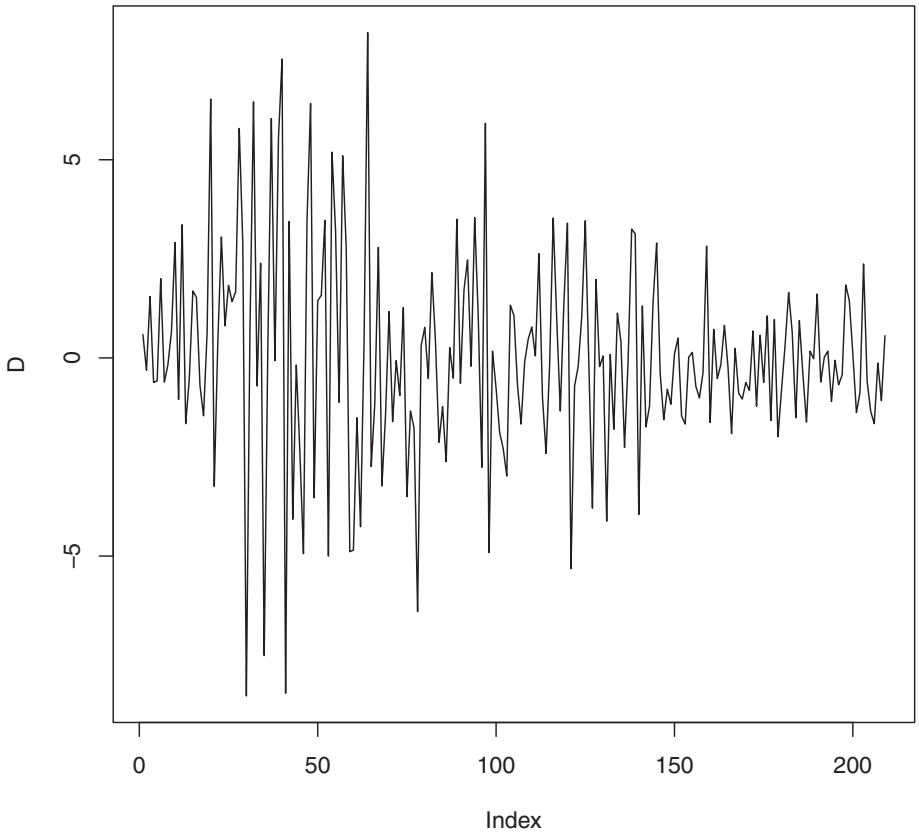

FIGURE 1: *CBOE S&P 500 3-Month Volatility Index :* Differenced values from August 29, 2008 to June 30, 2009.

to fit a variance function model of the squared series given by

$$\hat{\sigma}_t^2 = 2.135 + 0.719d_{1t} - 0.006d_t^2 - 3.856cos_t. \tag{5}$$

The coefficients were significant with standard errors, 1.220, 0.180, 0.003 and 1.532, respectively.

To compare the performance of our model with the parametric model, we simply fit the ARCH model for the differenced series. A convenient way to perform this parametric modelling is to use *R* software. By relying on *R*, we computed the Akaike information criterion (AIC) for the first six lags. They are 4.71, 4.59, 4.55, 4.54, 4.50, 4.51 which show the smallest AIC value is achieved

TABLE 4: *CBOE S&P 500 3-Month Volatility Index*: The table gives MIC values for the pairs (m, k).

k	m=1	$m = 2$	$m = 3$	$m = 4$
1	5.1969			
2	*5.1830	5.2805		
3	5.3086	5.4359	5.9738	
4	5.4322	5.6480	6.4821	7.4701

* Denotes $(\hat{m}, \hat{k})$ = (1,2) using (4).

at the fifth lag. However, coefficients in the ARCH(3), ARCH(4) and ARCH(5) models is not significant, we arrived at the following ARCH(2) model

$$\hat{\sigma}_t^2 = 2.835 + 0.267 z_{t-1} + 0.363 z_{t-2}. \tag{6}$$

The coefficients were significant with standard errors, 0.563, 0.119, and 0.122, respectively. In addition, since we are interested in comparing the standard benchmark volatility model such as GARCH (Generalized ARCH), the following GARCH(1,1) model is fitted again by *R* software:

$$\hat{\sigma}_t^2 = 0.026 + 0.099 z_{t-1} + 0.897 \hat{\sigma}_{t-1}^2. \tag{7}$$

The coefficients were significant with standard errors, 0.046, 0.038, and 0.034, respectively. Note that the intercept of this model is not significant.

Finally, to compare the three models above, we calculate the mean absolute residual (MAR), $|n^{-1} \sum_{t=1}^{n} (\sqrt{z_t}/\hat{\sigma}_t)|$ for the entire data values, where $\sqrt{z_t}$ is the differenced series, $\hat{\sigma}_t$ is its predicted standard deviation series and n is the number of data values. If the model eliminates the heteroscedasticity, it should attain a value of 1. Hence, a good criterion would be the deviation of the MAR from 1. The values of our ACVS, ARCH, and GARCH models are 0.2248, 0.2548 and 0.2283, respectively. This result implies that our ACVS model produces better precision than those of the ARCH/GARCH model. This demonstrates that our ACVS approach is a promising methodology for estimating time series models exhibiting volatility.

7. CONCLUSION

In this paper, we develop a new approach for dimension reduction in time series which focuses only on the variance function of time series. After introducing the notion of ACVS, we propose a kernel method to estimate the variance function, and thereby estimate ACVS when the lag k and the minimal dimension m are known. Next we use a data dependent method such as MIC in order to estimate the minimal dimension m and lag k simultaneously. In addition, we show that the estimator is consistent for known m and k. Overall, the variance dimension reduction subspace in time series poses many challenges, but the variety of encouraging results presented through our simulations and the real financial data set suggest that our method is useful in analyzing time series that exhibit volatility.

In this article, for dealing with volatility clearly, we consider the estimation of the variance function model only. However, financial time series typically involve more complicated aspects such as the fact that a mean function can be either autoregressive or dependent on explanatory variables as well as having a residual variance that has a heavier tail than that for normal innovations. Hence, further extensions may focus on a simultaneous estimation of the mean and variance function of time series, which is much more challenging in terms of developing theory as well as for practical implementation.

ACKNOWLEDGEMENTS

The authors wish to thank the editor, an associate editor, and two referees for their constructive comments that improved the paper. Park is supported in part by the College of Charleston.

BIBLIOGRAPHY

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

Broman, K. W. & Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *Journal of the Royal Statistical Society B*, 64, 641–656.

Cook, R. D. & Weisberg, S. (1991). Discussion of 'sliced inverse regression' by K. C. Li. *Journal of American Statistical Association*, 86, 328–332.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.

Fan, J. & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.

Fan, Y., Härdle, W., Wang, W., & Zhu, L.-X. (2013). Composite quantile regression for the single-index model. *SFB 649 Discussion Paper*, 010.

Hall, P. (1989). On projection pursuit regression. *Annals of Statistics*, 17, 573–588.

Härdle, W. & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of American Statistical Association*, 84, 986–995.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321–377.

Keener, R. W. (2010). *Theoretical Statistics: Topics for a Core Course*, Springer, New York.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 86, 316–42.

Masry, E. & Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory*, 11, 258–289.

Park, J.-H., Sriram, T. N., & Yin, X. (2009). Central mean subspace in time series. *Journal of Computational and Graphical Statistics*, 18(3), 717–730.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London and New York.

Xia, Y. & An, H. Z.(1999). Projection pursuit autoregressive in time series. *Journal of Time Series Analysis*, 20, 693–714.

Xia, Y., Tong, H., & Li, W. K. (2002). Single-index volatility models and estimation. *Statistica Sinica*, 12, 785–799.

Xia, Y., Tong, H., Li, W. K., & Zhu, L.-X. (2002). An adaptive estimation of dimension reduction. *Journal of Royal Statistical Society B*, 64, 363–410.

Ye, Z. & Weiss, R. E. (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association*, 98, 968–979.

Yoo, J. (2011). Unified predictor hypothesis tests in sufficient dimension reduction: A bootstrap approach. *Journal of Korean Statistical Society*, 40, 217–225.

Zhu, L.-X., Miao, B. Q., & Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*, 101, 630–643.

Zhu, L.-P. & Zhu, L.-X. (2009). Dimension reduction for conditional variance in regression. *Statistica Sinica*, 19, 869–883.

Received 1 May 2013
Accepted 22 April 2014