Statistica Sinica **12**(2002), 785-799

SINGLE-INDEX VOLATILITY MODELS AND ESTIMATION

Yingcun Xia[1,2], Howell Tong[3,4] and W. K. Li[3]

[1] *University of Cambridge,* [2] *Jinan University,*
[3] *University of Hong Kong and* [4] *London School of Economics*

Abstract: We develop a single-index volatility model in this paper. A new method is proposed to estimate the single-index coefficient and the link function. Unlike most existing estimation methods for semiparametric models, root-n consistency of the single-index coefficient can be achieved by our method without *under-smoothing* the unknown function. A Lagrange-multiplier type test is employed to determine the order of the model. Some simulations and applications to real data are included.

Key words and phrases: ARCH, conditional variance, local linear smoother, order determination.

1. Introduction

Conditional variance models within the context of diffusion models have a long history in stochastic processes, albeit commonly described in continuous time series (e.g., Doob (1953)). Recently, much attention has been paid to the study of diffusion in financial time series. Diffusion may exhibit itself in many different ways. One commonly adopted approach is to focus on the conditional variance. Its estimation may be considered either within the parametric framework as, e.g., the ARCH model of Engle (1982) or the nonparametric framework as, e.g., in Masry and Tjøstheim (1995). Consider the nonparametric model

$$\xi_i = \mu(x_i) + \sigma(x_i)\varepsilon_i, \tag{1.1}$$

where $\{(\xi_i, \mathbf{x}_i)\}$ is a two dimensional strictly stationary process having the same marginal distribution as $(\xi, \mathbf{x})$ and $\{\varepsilon_i\}$ are *i.i.d.* random variables having the same distribution as ε with $E\varepsilon = 0$; $\mu(x)$ and $\sigma(x)$ (> 0) are unknown regression and volatility functions. If we take $\mathbf{x}_i = \xi_{i-1}$, then (1.1) is a nonparametric generalization of the ARCH model. There is much literature concerning estimation of $\mu(x)$ and $\sigma(x)$. See, for example, Müller and Stadtmüller (1987), Ruppert, Wand, Holst and Hossjer (1997), Masry and Tjøstheim (1995) and Fan and Yao (1998).

If we extend (1.1) to the multivariate case, we encounter the problem of the "curse of dimensionality". Not surprisingly, existing estimation methods perform badly except in very low dimension. One approach is to restrict the functional form of the volatility function. In order to select a suitable form, we first take a look at the ARCH model of Eugle (1982):

$$\xi_i = \beta^T Z_i + (c_0 + \theta^T X_i)^{1/2}\varepsilon_i, \tag{1.2}$$

where $Z_i = (\xi_{i-1}, \ldots, \xi_{i-q})^T$ and $X_i = (\eta_{i-1}^2, \ \ldots, \ \eta_{i-p}^2)^T$ with $\eta_i = \xi_i - \beta^T Z_i$. To extend this model to a more flexible form, we follow the single-indexing idea (e.g., Ichimura (1993)) and propose a single-index volatility model:

$$\xi_i = \mu(\beta^T Z_i) + \sigma(\theta^T X_i)\varepsilon_i, \tag{1.3}$$

where $\{(X_i, \ Z_i, \ \xi_i)\}$ is a strictly stationary sequence. If $\mu(\cdot)$ is piecewise linear and $\sigma(\theta^T X_i) = (c_0 + c_1\theta^T X_i)^{1/2}$ with $X_i = (\xi_{i-1}^2, \ldots, \xi_{i-q}^2)^T$, then (1.3) is the SETAR-ARCH model (Tong (1990), p.116). Another obvious extension is the additive model (e.g., Linton and Härdle (1997)), which we do not investigate in this paper.

In this paper, we pay attention only to the estimation of σ and θ. By simple transformation to the model, several existing methods can be employed here. The disadvantages of these estimation methods are in four aspects. (1) Calculation burden. For example, Härdle, Hall and Ichimura (1993) and Carroll, Fan, Gijbels and Wang (1997) did not give an algorithm for their minimization problems. The problems are hard to implement. Actually, Härdle, Hall and Ichimura (1993) used the grid search algorithm in their simulations, which is very slow when the dimension is high. (2) Under-smoothing. See, for example, Weisberg and Welsh (1994) and Carroll *et al.* (1997). In order that the estimator of the single-index coefficient achieve root-n consistency, they had to undersmooth the unknown link function $\sigma(\cdot)$. As far as we know, there is no theoretical guidance as to how to select an appropriate data-driven bandwidth for such a under-smoothing. (3) Restriction on the design. As an application of the method proposed by Li (1991) and Duan and Li (1991) to this model, the design of X must be symmetric. This restriction is unreasonable since X_i is a positive random vector. (4) Inefficiency. The method proposed by Stoker (1986) uses a multivariate kernel method, which is not efficient. See also Härdle and Stoker (1989) and Powell, Stock and Stoker (1989). In this paper, a new method is proposed to estimate the single-index coefficient θ. Our method does not require under-smoothing the unknown link function. Therefore, the optimal convergence rates for the unknown link function and the single-index coefficient can be achieved simultaneously by using the optimal bandwidth in the sense of mean integrated squared errors (MISE). The bandwidth can be estimated by, e.g., cross-validation. The algorithm for our method is easy to implement. An important problem is the selection of the order

p. A Lagrange-multiplier type test is employed to determine the order. As an application of the model and estimation method, a financial data set is analyzed.

2. Estimation and Asymptotic Properties

In this section, we consider the estimation of the single-index volatility model $\xi_i = \sigma(\theta_0^T X_i)\varepsilon_i$ where $X_i = (x_{i1}, \ldots, x_{ip})^T$, θ_0 is an unknown parameter vector with $\|\theta_0\| = 1$ and $\sigma(\cdot)$ is an unknown link function. It follows that

$$|\xi_i|^\tau = \sigma_\tau \sigma^\tau(\theta_0^T X_i) + \sigma^\tau(\theta_0^T X_i)(|\varepsilon_i|^\tau - \sigma_\tau), \tag{2.1}$$

where $\tau > 0$ and $\sigma_\tau = E|\varepsilon_t|^\tau$. Note that most existing nonparametric estimations are based on $\tau = 2$. They are known to be sensitive to aberrant observations, see, Fan and Yao (1998), Masry and Tjøstheim (1995) and Härdle and Tsybakov (1997). A detailed discussion on how to choose τ can be found in Carroll and Ruppert (1988) and Mercurio and Spokoiny (2000). In this paper, we only consider the case $\tau = 1$, noting that other cases can be estimated similarly. For ease of exposition, we assume that $E|\varepsilon_t| = 1$ and let $y_i = |\xi_i|$ throughout the rest of the paper. Suppose $\{(X_i, y_i) : i = 1, \ldots, n\}$ is a realization from the model and with the marginal distribution of (X, y).

Note that θ_0 satisfies

$$\theta_0 = \arg \inf_{\|\theta\|=1} E\Big\{\frac{y - E(y|\theta^T X)}{\sigma(\theta^T X)}\Big\}^2. \tag{2.2}$$

We have
$E\Big\{\sigma^{-1}(\theta^T X)[y - E(y|\theta^T X)]\Big\}^2 = E\Big\{\sigma^{-2}(\theta^T X)E\Big[\{y - E(y|\theta^T X)\}^2|\theta^T X\Big]\Big\}$. On here, $E[\{y - E(y|\theta^T X = \theta^T x)\}^2|\theta^T X = \theta^T x]$ is the conditional variance and can be estimated by $\min_{a,d} \sum_{i=1}^n \Big\{y_i - a - d\theta^T(X_i - x)\Big\}^2 w_{i0}$, where $\{w_{i0}, i = 1, \ldots, n\}$ is a set of weights centered about x. Therefore the minimization in (2.2) is equivalent to minimizing with respect to a_j, d_j and θ,

$$n^{-1}\sum_{j=1}^n \sigma^{-2}(\theta^T X_j)\sum_{i=1}^n \{y_i - a_j - d_j\theta^T(X_i - X_j)\}^2 w_{ij}, \tag{2.3}$$

where $\{w_{ij} : i = 1, 2, \ldots, n\}$ is a set of weights centered about X_j. Note that σ is still unknown in (2.3). But a_j is an estimate of it. We replace $\sigma(\theta^T X_j)$ by a_j. Based on (2.2) and (2.3), we estimate θ_0 by minimizing

$$n^{-1}\sum_{j=1}^n a_j^{-2}\sum_{i=1}^n \{y_i - a_j - d_j\theta^T(X_i - X_j)\}^2 w_{ij}, \tag{2.4}$$

iteratively with respect to (a_j, d_j) and θ. Let $V(\cdot)$ be a p-dimensional spherical density function and $V_{b,i}(x) = b^{-p}V\{(X_i - x)/b\}$, where b is the bandwidth.

At the first stage, we take $w_{ij} = V_{b,i}(X_j)/\sum_{i=1}^{n} V_{b,i}(X_j)$. Suppose that X is centralized and $\mathcal{D}$ is an open convex set about 0 such that $\inf_{x\in\mathcal{D}} f(x) > 0$, where $f(x)$ is the density function of X. To avoid the effect of the marginal points of the kernel estimators, we consider the observations in $\mathcal{D}$ as in Härdle, Hall and Ichimura (1993). Let $\sum_{j'}$ denote the summation over $\{j : X_j \in \mathcal{D}\}$. Based on (2.4), we can estimate an initial value of θ_0 as follows: choose a vector θ with norm 1; find a closed form for a_j and $d_j, j = 1,\ldots,n$ in terms of θ; with this a_j and d_j, calculate the optimizing θ; iterate to convergence. Denote the final value by θ_1.

Lemma 1. *Suppose* (C1)-(C6) *in the Appendix hold,* $nb^{p+2}/\log n \to \infty$ *and* $b \to 0$. *If the starting value of* θ *satisfies* $\theta^T\theta_0 \neq 0$, *then* $\theta_1 - \theta_0 = O(b^2 + \frac{\log n}{nb^{p+2}})$ *almost surely.*

A more detailed discussion can be found in Xia, Tong, Li and Zhu (2000). It is known that the inefficiency of using a high-dimensional kernel cannot be easily reduced. Note that the optimal bandwidth for the estimation of $\sigma(\cdot)$ in the sense of MISE is $b \sim n^{-1/(p+4)}$. Therefore, we have $\theta_1 - \theta_0 = O(n^{-2/(p+4)}\log n)$ almost surely. Next, we improve the efficiency by using a single-index kernel.

Let $K(\cdot)$ be a univariate density function, γ be a vector with norm 1, $K_{\gamma,h,i}(x) = h^{-1}K\{\gamma^T(X_i - x)/h\}$ where h is the bandwidth, and

$$S_{\gamma,n,0}(x) = n^{-1}\sum_{i=1}^{n} K_{\gamma,h,i}(x), \qquad S_{\gamma,n,1}(x) = n^{-1}\sum_{i=1}^{n} K_{\gamma,h,i}(x)(X_i - x),$$
$$S_{\gamma,n,2}(x) = n^{-1}\sum_{i=1}^{n} K_{\gamma,h,i}(x)(X_i - x)(X_i - x)^T.$$

Let $w_{\gamma,ij} = K_{\gamma,h,i}(\gamma^T X_j)/\sum_{\ell=1}^{n} K_{\gamma,h,\ell}(X_j)$, $W_{\gamma,n}(x) = S_{\gamma,n,2}(x)S_{\gamma,n,0}(x) - S_{\gamma,n,1}(x)S_{\gamma,n,1}^T(x)$ and

$$w_{\gamma,a,i}(x) = \{\gamma^T S_{\gamma,n,2}(x)\gamma\}K_{\gamma,h,i}(x) - \gamma^T S_{\gamma,n,1}(x)K_{\gamma,h,i}(x)\gamma^T(X_i - x),$$
$$w_{\gamma,d,i}(x) = S_{\gamma,n,0}(x)K_{\gamma,h,i}(x)\gamma^T(X_i - x) - \gamma^T S_{\gamma,n,1}(x)K_{\gamma,h,i}(x).$$

Based on (2.4), we can improve the estimator θ_1 as follows. We start with $k = 1$,

$$\tilde{a}_j = \{n\theta_k^T W_{\theta_k,n}(X_j)\theta_k\}^{-1}\sum_{i=1}^{n} w_{\theta_k,a,i}(X_j)y_i, \tag{2.5}$$

$$\tilde{d}_j = \{n\theta_k^T W_{\theta_k,n}(X_j)\theta_k\}^{-1}\sum_{i=1}^{n} w_{\theta_k,d,i}(X_j)y_i, \tag{2.6}$$

$$\theta_{k+1} = \Big\{\sum_{j'}(\tilde{d}_j/\tilde{a}_j)^2 S_{\theta_k,n,2}(X_j)\Big\}^{-1}\sum_{j'}\tilde{d}_j/\tilde{a}_j^2\sum_{i=1}^{n} w_{\theta_k,ij}(X_i - X_j)(y_i - \tilde{a}_j). \tag{2.7}$$

Replace θ_k in (2.5) and (2.6) by θ_{k+1} in (2.7), iterate to convergence. Denote the final value by $\hat{\theta}$.

After obtaining $\hat{\theta}$, we can estimate the unknown function by $\hat{\sigma}(v) = \sum_{i=1}^n w_{\hat{\theta},n}(X_i, v)y_i / \sum_{i=1}^n w_{\hat{\theta},n}(X_i, v)$, where

$$w_{\hat{\theta},n}(X_i, v) = \sum_{j=1}^n K\{(\hat{\theta}^T X_j - v)/h\}(\hat{\theta}^T X_j - v)^2 K(\hat{\theta}^T X_i - v) \\ - \sum_{j=1}^n K\{(\hat{\theta}^T X_j - v)/h\}(\hat{\theta}^T X_j - v)^2 K\{(\hat{\theta}^T X_i - v)\}(\hat{\theta}^T X_i - v).$$

Let $\mu(x|\theta) = E(X|\theta^T X = \theta^T x)$, $\sigma_2 = E\varepsilon^2$ and $W_0 = (\sigma_2 - 1)^{-1} \int_{\mathcal{D}} \{x - \mu(x|\theta_0)\}\{x - \mu(x|\theta_0)\}^T (\sigma'(\theta_0^T x)/\sigma(\theta_0^T x))^2 f(x)dx$.

Theorem 1. *Suppose that* (C1)-(C6) *in the Appendix hold. For a bandwidth* $h \sim n^{-\delta}$ *with* $1/6 < \delta < 1/4$, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, W_0^{-1})$. *If the density function* $f_{\theta_0}(v)$ *of* $\theta_0^T X$ *is positive at* v, *then* $\sqrt{nh}\{\hat{\sigma}(v) - \sigma(v) - \sigma''(v)h^2/2\} \xrightarrow{D} N(0, (\sigma_2 - 1) f_{\theta_0}^{-1}(v)\sigma^2(v) \int K^2(u)du)$.

Remark 1. Note that the optimal bandwidth for estimation of the unknown volatility function in the sense of MISE is $h_{opt} \sim n^{-1/5}$, which satisfies the condition of Theorem 1. Therefore, our estimators of the single-index coefficients and the unknown function can achieve optimal convergence rates simultaneously by using the same bandwidth. Under-smoothing is unnecessary for our method.

Remark 2. In practice we suggest, after iteration (2.5)-(2.7), the standardization $\theta_{k+1} \triangleq \theta_{k+1}/\|\theta_{k+1}\|$. Then root-$n$ consistency also holds, although the asymptotic distribution is difficult to derive.

Remark 3. Sometimes there is a nonnegative constraint on the parameter θ. Based on (2.4), the constraint entails a simple quadratic programming problem, which can be solved easily. Thus we need only replace θ_{k+1} in (2.7) with the solution of the corresponding quadratic programming.

Remark 4. Note that $n^{-1} \sum_{j'} (\tilde{d}_j/\tilde{a}_j)^2 S_{\hat{\theta},n,2}(X_j) \to W_0(\sigma_2 - 1)$ in probability as $n \to \infty$. Thus the limiting distribution can be approximated easily. By the results of Carroll *et al.* (1997), W_0 is the information matrix for model (2.1). Our estimation method is asymptotically efficient.

Order selection of the single-index volatility model is an important problem. Here we extend Lagrange-multiplier type testing to the single-index volatility specification. The basic idea of the procedure is to test the specification of a statistical model by overfitting: the null hypothesis that the model is correct is tested against a suitable alternative hypothesis of which the null hypothesis is a special case. See, for example, Hosking (1980) and Li and Mak (1994). Consider

$$\xi_i = \sigma(\theta_0^T X_i)\varepsilon_i, \tag{2.8}$$

where $X_i = (\xi_{i-1}^2, \ldots, \xi_{i-p}^2)^T$. We investigate an alternative model

$$\xi_i = \sigma(\theta_0^T X_i + \lambda^T \tilde{X}_i)\varepsilon_i, \tag{2.9}$$

where $\lambda=(\lambda_1, \ldots, \lambda_m)^T$ are constant parameters and $\tilde{X}_i=(\xi_{i-p-1}^2, \ldots, \xi_{i-p-m}^2)^T$. The alternative is also a single-index volatility model but, for moderate or large m, is capable of representing a wide variety of volatility features in time series. Therefore, our problem may be changed to that of testing $\lambda_k = 0$, $k = 1, \ldots, m$. See also, for example, Hosking (1980). Let $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)^T$ be the corresponding estimators using the proposed method.

Corollary 1. *Under the assumptions of Theorem* 1, *if* $\lambda_k = 0$, $k = 1, \ldots, m$, *then* $n\hat{\lambda}^T\tilde{W}\hat{\lambda} \to \chi^2(m)$, *where* $\tilde{W}$ *is the submatrix corresponding to the last* m *rows and columns of the matrix* $(\sigma_2-1)^{-1}\int_{\tilde{D}}\{\tilde{x}-\mu(\tilde{x}|\theta_0)\}\{\tilde{x}-\mu(\tilde{x}|\theta_0)\}^T(\sigma'(\theta_0^T x)/\sigma(\theta_0^T x))^2$ $\tilde{f}(\tilde{x})d\tilde{x}$ *with* $\tilde{x} = (x_1, \ldots, x_{p+m})^T$, $\tilde{f}(\tilde{x})$ *is the density function of* $(X_i, \tilde{X}_i)$.

3. Simulations and Applications

In this section, we carry out simulations to check our estimation method for some finite data sets, then we apply the single-index volatility model and estimation methods to a real data set.

An important problem for the application of the kernel smooth method is the selection of the bandwidth. As mentioned, the main difference between our estimation method and others is that our method allows the bandwidth to take its optimal value in the sense of MISE. Therefore we may use common bandwidth selection methods. We give the details for the cross-validation bandwidth. Write

$$S_{h,\gamma,n,0}(X_j) = n^{-1}\sum_{i=1,i\neq j}^{n} K_{\gamma,h,i}(X_j),$$

$$S_{h,\gamma,n,1}(x) = n^{-1}\sum_{i=1,i\neq j}^{n} K_{\gamma,h,i}(x)(X_i - X_j),$$

$$S_{h,\gamma,n,2}(X_j) = n^{-1}\sum_{i=1,i\neq j}^{n} K_{\gamma,h,i}(X_j)(X_i - X_j)(X_i - X_j)^T,$$

$$W_{h,\gamma,n}(X_j) = S_{h,\gamma,n,2}(X_j)S_{h,\gamma,n,0}(X_j) - S_{h,\gamma,n,1}(X_j)S_{h,\gamma,n,1}^T(X_j),$$

$$w_{h,\gamma,a,i}(x) = \{\gamma^T S_{h,\gamma,n,2}(x)\gamma\}K_{\gamma,h,i}(x) - \gamma^T S_{h,\gamma,n,1}(x)K_{\gamma,h,i}(x)\gamma^T(X_i - x).$$

We estimate a_j with observation (X_j, y_j) deleted as

$$\breve{a}_j = \{\theta_k^T W_{h,\theta_k,n}(X_j)\theta_k\}^{-1}\sum_{i=1,i\neq j}^{n} w_{h,\theta_k,a,i}(X_j)y_i.$$

Then the bandwidth at the $(k+1)$'th iteration is given by $h = \arg\min_h n^{-1}\sum_{j=1}^{n}\{y_j - \breve{a}_j\}^2$.

Example 1. We simulate 200 random realizations of size n from

$$\xi_i = \sqrt{h_i}\varepsilon_i, \quad h_i = 0.1 + 0.5\theta_0^T X_i, \tag{3.1}$$

where $X_i = (\xi_{i-1}^2, \xi_{i-2}^2, \xi_{i-3}^2, \xi_{i-4}^2)^T$ and $\theta_0 = (0.8, 0, 0.6, 0)^T$, and from

$$\xi_i = \sqrt{h_i}\varepsilon_i, \quad h_i = 0.1 + 1.5\theta^T X_i \sin^2(\frac{5}{3}\theta^T X_i), \tag{3.2}$$

where $X_i = (|\xi_{i-1}|, |\xi_{i-2}|, |\xi_{i-3}|, |\xi_{i-4}|)$ and $\theta = (0.6, 0.3, 0, 0)^T$, i.e., $\theta_0 = (0.8944, 0.4472, 0, 0)$. In both models, $\varepsilon \sim N(0,1)$. We set $n = 400$ and 800. To avoid denominators of estimators a_j and d_j close to zero, we choose $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \mathcal{D}_3 \times \mathcal{D}_4$ to be the region bounded by the 95% quantile of ξ_k^2 (or $|\xi_k|$). The means and standard deviations of estimates of θ_0 are listed in Table 1, from which we find that the parameter estimates are quite reasonable. Results from a typical data set with $n = 500$ from model (3.2), where $\sigma(v) = \{0.1 + 1.5v\sin^2(5v/3)\}^{1/2}$, are plotted in Figure 1. The volatility function is estimated well.

Table 1. Means and standard deviations (in parentheses) of estimated θ_0 for model (3.1) and (3.2).

Model (3.1)		Model (3.2)	
$n = 400$	$n = 800$	$n = 400$	$n = 800$
0.7517 (0.1164)	0.7765 (0.0771)	0.8769 (0.0630)	0.8805 (0.0457)
0.0684 (0.1057)	0.0558 (0.0836)	0.4561 (0.1132)	0.4617 (0.0806)
0.6114 (0.1398)	0.6166 (0.0994)	0.0321 (0.0499)	0.0255 (0.0376)
0.0606 (0.0923)	0.0484 (0.0633)	0.0240 (0.0466)	0.0154 (0.0248)

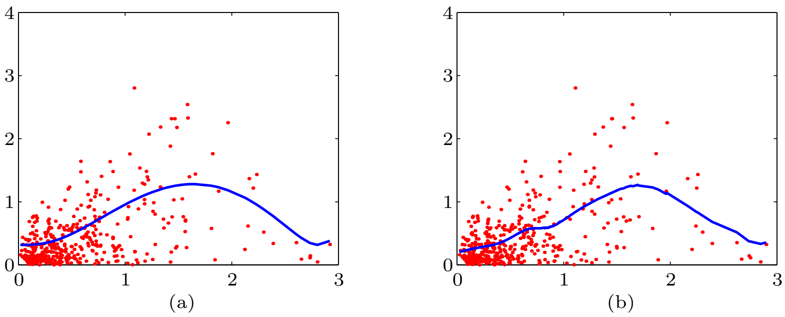

Figure 1. Results for model (3.2): the solid lines in (a) and (b) denote the true volatility function $\sigma(\cdot)$ and its estimate, respectively; the dots in (a) and (b) denote $|\xi_i|$ plotted against $\theta_0^T X_i$ and $\hat{\theta}^T X_i$, respectively.

We make a further comparison between the initial estimates based on the multidimensional kernel and the refined estimates based on the single-index kernel in model (3.2). Boxplots are employed for this purpose, as shown in Figure 2. As we expect, the single-index kernel can improve the estimates substantially. Finally, we use these two models to check our order selection method. We extend the models to order 10 and 5, respectively, with the structures unchanged. Table 2 lists the frequencies of lack of fit at the 0.10 significance level out of 500 replications. The simulation results are reasonable. However, our simulations show that when the link function fluctuates a lot, the proposed method works less well.

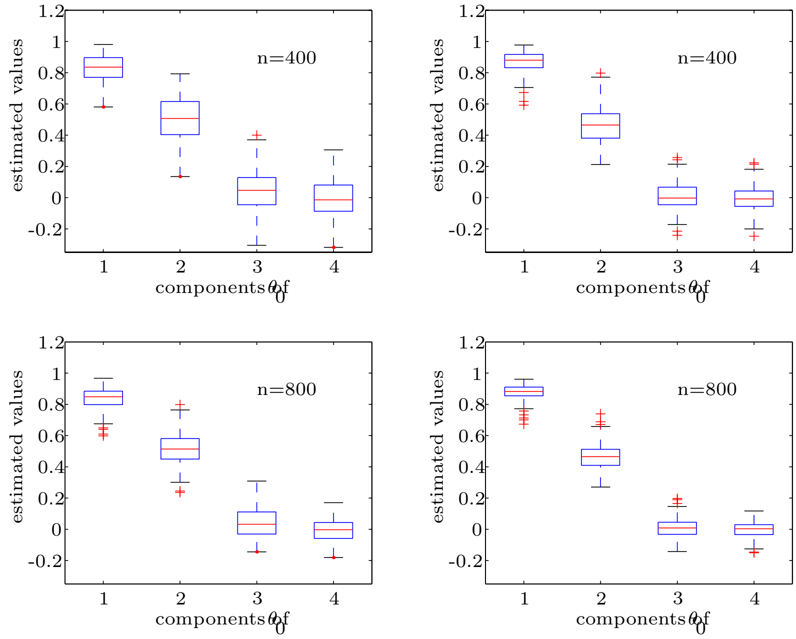

Figure 2. The Boxplots of the simulation results for model (3.2). In each panel, the box-plots are for θ_{01}, θ_{02}, θ_{03} and θ_{04}, respectively, where $\theta_0 = (\theta_{01}, \theta_{02}, \theta_{03}, \theta_{04})^T$. Panels 1 and 3 are the initial estimates using multivariate kernel. Panels 2 and 4 are the final estimates using single-index kernels.

Table 2. Frequencies of lack of fit at 0.10 significance level for models (3.1) and (3.2).

model	n	order p 1	2	3	4	5	6	7	8	9
(3.1)	400	421	421	75	76	76	79	78	60	71
	800	467	469	63	67	65	62	66	48	63
(3.2)	400	490	145	145	100					
	800	500	110	114	70					

Example 2. As an application, we investigate the daily closing Hang Seng in-

dex from 1987 to 1997. To induce approximate stationarity, we take the first difference of logarithmically transformed Hang Seng indices. The transformed data (ξ_t) are plotted in Figure 3(a), which shows three possible 'outliers': the two largest crashes on 26/10/87 and 5/6/89, and the rebound of 29/10/97. With these outliers removed, no trend in ξ_i is discernible and the sample autocorrelation function is not significantly different from the Kronecker delta function. Thus, we assume $\mu(\cdot) = 0$ and consider the nonparametric conditional variance model

$$\xi_i = \sigma(\theta^T X_i)\varepsilon_i, \tag{3.3}$$

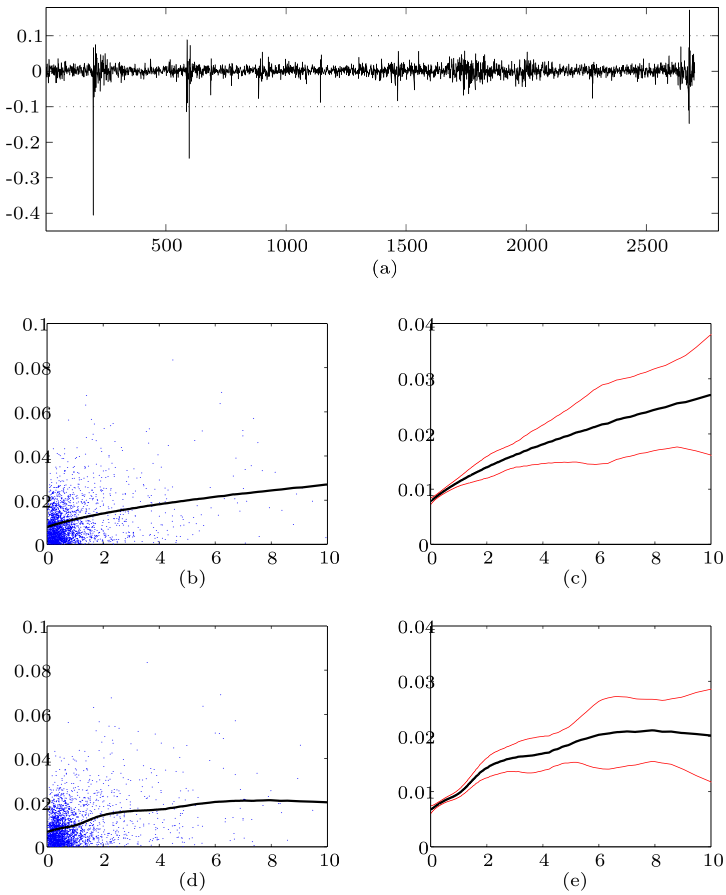

Figure 3. Results of Example 2. (a) is the transformed data; the thick lines in (b)-(e) denote $\hat{\sigma}(\cdot)$ using the transformed data; the dots in (b) and (d) are the ξ_i plotted against $\hat{\theta}^T X_i$; the thin lines in (c) and (e) denote the approximate 95% pointwise confidence intervals for $\sigma(\cdot)$; (b) and (c) are based on the parametric ARCH model of order 9; (d) and (e) are based on the single-index model (3.3).

where $X_i = (\xi_{i-1}^2, \ldots, \xi_{i-p}^2)^T$. The order is chosen to be 9 using the proposed method (the order of alternative model is 15, i.e., 3 weeks). Using the proposed estimation method and choosing the region $\mathcal{D}$ in the same way as in the simulations, we obtained $\hat{\theta} = ($ 0.618, 0.324, 0.488, 0.152, 0.207, 0.061, 0.242, 0.180, $0.239)^T$ based on model (2.1) with $\tau = 1$. The volatility function $\sigma(\cdot)$ is shown in Figures 3(d)-(e). Note that the estimate $\hat{\theta}$ satisfies the non-negativity assumption of θ in Engle (1982). We further fit the data using the parametric ARCH model $\xi_i = (\alpha_0 + \alpha_1\xi_{i-1}^2 + \ldots + \alpha_9\xi_{i-9}^2)^{1/2}\varepsilon_i$. Using the method proposed Engle (1982), we obtain $(\hat{\alpha}_0, \ldots, \hat{\alpha}_9) = \vartheta 10^{-4}(1.860, 0.754, 0.521, 0.392, 0.069, 0.158, 0.051, 0.115, 0.028, 0.224)$. The estimated volatility function is shown in Figure 3(b)-(c) (after transformation). Compared with the parametric ARCH model, the single-index model is more informative.

4. Conclusions

In this paper, we use the single-index model to approximate the unknown conditional variance functions and propose the single-index conditional variance model. The flexibility of the model allows us to capture more features of volatility in real data sets. A method is proposed for the estimation of the single-index parameters, which requires no under-smoothing of the unknown function. The estimators of both the parameters and the unknown function can achieve their respective optimal consistency rates by using the optimal bandwidth in the sense of MISE. Therefore, a data-driven bandwidth can be used. Because of its simplicity, constraints on the parameters can also be imposed as discussed under Remark 3.

Acknowledgement

We thank the BBSRC/EPSRC of UK, the Research Grants Council of Hong Kong, the CRCG of the University of Hong Kong, the Friends of London School of Economics (Hong Kong) and the Wellcome Trust for partial support. We are most grateful to the two referees for their thorough comments and constructive suggestions.

Appendix. Assumptions and Proofs

To obtain the asymptotic properties of the single-index conditional variance model in section 2, we need the following assumptions.

(C1) $\{(X_i, y_i)\}$ is a strictly stationary and strongly mixing sequence with mixing coefficient $\alpha(k) = O(c^k)$ for some $0 < c < 1$.

(C2) The density function $f_\theta(v)$ of $\theta^T X$ has bounded continuous second order derivatives and is bounded away from 0 on $\{v = \theta^T x : \ x \in \mathcal{D}, \ c_0 \le \|\theta\| \le c_0'\}$ for some $0 < c_0 < 1 < c_0'$.

(C3) $M < f(x) < M'$ for some positive constants M and M', and has bounded second derivatives in $\mathcal{D}$. The joint densities of X_1 and X_k for all $k \ge 2$ are bounded.

(C4) For each i, ε_i is independent of $\{X_{j+1}, y_j, j < i\}$, $E|\varepsilon_j|^l < \infty$ and $E|y_j|^l < \infty$ for some $l > 2$.

(C5) $\sigma(v)$ has bounded rth order derivatives with some $r > \max(3,(p+4)/10)$, and $\inf_{v\geq 0}\sigma(v) > 0$.

(C6) $K(v)$ is a symmetric density function with moments of all orders; $V(x)$ is a spherical symmetric density function and $\int u_1^{k_1}\ldots u_p^{k_p}V(U)dU < \infty$ for all $k_1+\cdots+k_p > 1$, $k_1 > 0,\ldots,k_p > 0$. The Fourier transforms of $K(v)$ and $V(U)$ are absolutely integrable. (For ease of exposition, we further assume that $\int UU^TV(U)dU = I$.)

The basic results are Lemmas A.1 and A.2 below, other lemmas can be derived by simple algebraic calculations. Let $\delta_\theta = \|\theta-\theta_0\|$, $\delta_n = \{\log n/(nh)\}^{1/2}$ and $\delta_{0n} = (\log n/n)^{1/2}$. Let $\Theta = \{\theta : c_0 \leq \|\theta\| \leq c_0'\}$, where $0 < c_0 < 1 < c_0'$ are constants. Suppose A_n is a matrix. For ease of exposition, $A_n = O_{a.s.}(a_n)$ means every element in A_n is $O_{a.s.}(a_n)$ almost surely. Let $\eta_i = \sigma(\theta_0^TX_i)(|\varepsilon_i|-1)$.

Lemma A.1. *Suppose that $m(X)$ and $m(x,\theta)$ are bounded measurable functions, $m(x,\theta)$ has bounded derivative with respect to θ. Under* (C1)-C(6), *we have*

$$\sup_{x\in\mathcal{D},\gamma\in\Theta}\Big|n^{-1}\sum_{i=1}^n\Big[K_{\gamma,h,i}(x)m(X_i)-E\{K_{\gamma,h,i}(x)m(X_i)\}\Big]\Big| = O_{a.s.}(\delta_n),$$

$$\sup_{x\in\mathcal{D},\gamma\in\Theta}\Big|n^{-1}\sum_{i=1}^n K_{\gamma,h,i}(x)m(X_i)\eta_i\Big| = O_{a.s.}(\delta_n),$$

$$\sup_{\|\gamma-\theta_0\|<a_n}\Big|n^{-1}\sum_{i=1}^n\Big[m(X_i,\theta_0)-m(X_i,\gamma)\Big]\eta_i\Big| = O_{a.s.}(a_n\delta_{0n}).$$

This lemma can be proved following Xia and An (1999), noting that $\mathcal{D}\otimes\Theta$ is a compact set and that $K_{\gamma,h,i}(x)$ is continuous in both x and γ.

Lemma A.2. *Suppose $m(u)$ is any bounded measurable function. Then*

$$n^{-2}\sum_{j'}\sum_{i=1}^n\Big\{V_{b,i}(X_j)m(X_j)-\int V_{b,i}(x)m(x)f(x)dx\Big\}\eta_i = O_{a.s.}(\delta_{pn}^2),$$

$$\sup_{\gamma\in\Theta}\Big|n^{-2}\sum_{j'}\sum_{i=1}^n\Big\{K_{\gamma,h,i}(X_j)m(X_j)-\int K_{\gamma,h,i}(x)m(x)f(x)dx\Big\}\eta_i\Big| = O_{a.s.}(\delta_n^2).$$

Proof. We only prove the second part. Let $\bar\delta_n = \delta_n^2h$ and

$$\Delta_n(\gamma) = n^{-2}\sum_{j'}\sum_{i=1}^n\Big\{K_{\gamma,h,i}(X_j)m(X_j)-\int K_{\gamma,h,i}(x)m(x)f(x)dx\Big\}\eta_i.$$

By the continuity of $K_{\gamma,h,i}(x)$ in γ, there are $n_1(< cn^{p+2}$, where c is a constant) points $\gamma_{n,1},\ldots,\gamma_{n,n_1}$ in Θ such that $\cup_{k=1}^{n_1}\{\gamma : \|\gamma-\gamma_{n,k}\| < \bar{\delta}_n\} \supset \Theta$ and

$$\max_{1\le k\le n_1} \sup_{\|\gamma-\gamma_{n,k}\|<\bar{\delta}_n} \Big|\Delta_n(\gamma) - \Delta_n(\gamma_{n,k})\Big| = O_{a.s.}(\delta_n^2). \tag{A.1}$$

By the Fourier inversion formula, $K(v) = \int \exp(-\imath sv)\phi(s)ds$, where $\imath$ is the imaginary unit and $\phi(s)$ is the Fourier transformation of K. We have

$$\begin{aligned}\Delta_n(\gamma_{n,k}) &= n^{-2}h^{-1}\sum_{j'}\sum_{i'}\int\Big[\exp\{-\imath s\gamma_{n,k}^T(X_i-X_j)/h\}m(X_j)\\&\qquad -\int\exp\{-\imath s\gamma_{n,k}^T(X_i-x)/h\}dx\Big]\phi(s)ds\eta_i\\&= h^{-1}\int n^{-1}\sum_{i'}\exp(-\imath s\gamma_{n,k}^T X_i/h)\eta_i\cdot n^{-1}\sum_{j'}\Big[\exp(\imath s\gamma_{n,k}^T X_j/h)m(X_j)\\&\qquad -\int\exp(\imath s\gamma_{n,k}^T x/h)m(x)f(x)dx\Big]\phi(s)ds.\end{aligned}$$

Following the steps of Xia and An (1999), we have

$$\begin{aligned}&\max_{1\le k\le n_1}|n^{-1}\sum_{i'}\exp(-\imath s\gamma_{n,k}^T X_i/h)\eta_i| \le c_1\delta_{0n},\\&\max_{1\le k\le n_1}\Big|n^{-1}\sum_{j'}\Big[\exp(\imath s\gamma_{n,k}^T X_j/h)m(X_j) - \int\exp(\imath s\gamma_{n,k}^T x/h)m(x)f(x)\Big]\Big| \le c_2\delta_{0n}\end{aligned}$$

almost surely, where c_1 and c_2 are constants which do not depend on s. Hence

$$\max_{1\le k\le n_1}\Big|\Delta_n(\gamma_{n,k})\Big| \le h^{-1}\int c_1\delta_{0n}c_2\delta_{0n}|\phi(s)|ds = O_{a.s.}(\delta_n^2). \tag{A.2}$$

Note that

$$\sup_\gamma|\Delta_n(\gamma)| \le \max_{1\le k\le n_1}\Big|\Delta_n(\gamma_{n,k})\Big| + \max_{1\le k\le n_1}\sup_{\|\gamma-\gamma_{n,k}\|<\bar{\delta}_n}\Big|\Delta_n(\gamma)-\Delta_n(\gamma_{n,k})\Big|. \tag{A.3}$$

Lemma A.2. follows from (A.1), (A.2) and (A.3).

Corresponding to (2.5) and (2.6), let

$$\tilde{a} = \{n\gamma^T W_{\gamma,n}(x)\gamma\}^{-1}\sum_{i=1}^n w_{\gamma,a,i}(x)y_i, \quad \tilde{d} = \{n\gamma^T W_{\gamma,n}(x)\gamma\}^{-1}\sum_{i=1}^n w_{\gamma,d,i}(x)y_i.$$

Lemma A.3. *Under* (C1)-(C6), *we have*

$$\tilde{a} = \sigma(x) + \sigma'(\theta_0^T x)\{\mu(x|\gamma) - x\}^T(\theta_0-\gamma) + \frac{1}{2}\sigma''(\theta_0^T x)h^2$$

$$+f_\gamma^{-1}(\gamma^T x)n^{-1}\sum_{i=1}^n K_{\gamma,h,i}(x)\eta_i+O_{a.s.}\{\delta_\gamma^2+(h^2+h^{-1}\delta_n)\delta_\gamma+h^3+h\delta_n+\delta_n^2\},$$

$$\begin{aligned}\tilde d &= \{\gamma^T\theta_0\}\sigma'(\theta_0^T x)+\{\mu'(x|\gamma)\}^T(I-\gamma\gamma^T)\theta_0\sigma'(\theta_0^T x)\\ &\quad+\{\mu(x|\gamma)-x\}^T(\theta_0-\gamma)\sigma''(\theta_0^T x)+\frac{1}{2}f_\gamma^{-1}(\gamma^T x)f_\gamma'(\gamma^T x)h^2(\kappa_4-1)\sigma''(\theta_0^T x)\\ &\quad+R_n(x)+O_{a.s.}(h^{-1}\delta_\gamma^r+\delta_\gamma^2+h^{-1}\delta_n\delta_\gamma+h^2+h^{-1}\delta_n^2),\end{aligned}$$

where $\kappa_4=\int K(v)v^4dv$, $\mu'(x|\gamma)=dE(X|\gamma^T X=v)/dv|_{v=\gamma^T x}$ *and*

$$R_n(x)=n^{-1}\sum_{i=1}^n K_{\gamma,h,i}(x)\Big\{f_\gamma^{-1}(\gamma^T x)h^{-2}\gamma^T(X_i-x)-f_\gamma^{-2}(\gamma^T x)f_\gamma'(\gamma^T x)\Big\}\eta_i.$$

Let

$$C_n=n^{-2}\sum_{j'}(\tilde d_j/\tilde a_j)^2\sum_{i=1}^n K_{\gamma,h,i}(X_j)(X_i-X_j)(X_i-X_j)^T/S_{\gamma,n,0}(X_j),$$

$$B_n=n^{-2}\sum_{j'}\tilde d_j/\tilde a_j^2\sum_{i=1}^n K_{\gamma,h,i}(X_j)(X_i-X_j)(y_i-\hat a_j-\hat d_j(X_i-X_j)^T\theta_0)/S_{\gamma,n,0}(X_j),$$

$$\hat\gamma=C_n^{-1}n^{-2}\sum_{j'}\tilde d_j/\tilde a_j^2\sum_{i=1}^n K_{\gamma,h,i}(X_j)(X_i-x)(y_i-\hat a_j)/S_{\gamma,n,0}(X_j).$$

Then

$$\hat\gamma=\theta_0+C_n^{-1}B_n. \tag{A.4}$$

Lemma A.4. *Under assumptions (C1)-(C6), we have*

$$\begin{aligned}&n^{-2}\sum_{j'}\tilde d_j/\tilde a_j^2\sum_{i=1}^n K_{\gamma,h,i}(X_j)\{X_i-X_j\}(y_i-\hat a_j-\hat d_j(X_i-X_j)^T\theta_0)/S_{\gamma,n,0}(X_j)\\ &=W_0(\gamma-\theta_0)+n^{-1}\sum_{i=1}^n\{\sigma'(\theta_0^T X_i)/\sigma(\theta_0^T X_i)\}\{\mu(X_i|\theta_0)-X_i\}\eta_i\\ &\quad+O_{a.s.}\{(h^{-1}\delta_\gamma^r+\delta_\gamma+h^2+h^{-1}\delta_n)\delta_\gamma+h^3+\delta_n^2+h\delta_n\},\\ C_n&=2W_0+O_{a.s.}(\delta_\gamma+h^2+h^{-1}\delta_n).\end{aligned}$$

Lemma A.5. *Assume* (C1)-(C6). *Let* $\hat\gamma$ *be the value on the right side of* (2.7) *with* θ *replaced by* γ. *Then* $\hat\gamma-\theta_0=\frac{1}{2}(\gamma-\theta_0)+\frac{1}{2}N_n+O_{a.s.}\{(h^{-1}\delta_\gamma^r+\delta_\gamma+h^2+h^{-1}\delta_n+\delta_0)\delta_\gamma+\tilde\delta_n\}$, *where* $\tilde\delta_n=h^3+h\delta_n+h^{-1}\delta_n^2$ *and* $N_n=W_0^{-1}n^{-1}\sum_{j'}\{\sigma'(\theta_0^T X_j)/\sigma(\theta_0^T X_j)\}$ $\{\mu(X_j|\theta_0)-X_j\}\eta_j$.

Proof of Theorem 1. We only prove the first part, the second part is straightforward. By Lemma A.5, for any step k we have

$$\theta_{k+1}-\theta_0=\frac{1}{2}(\theta_k-\theta_0)+\frac{1}{2}N_n+\Delta_k, \tag{A.5}$$

where $|\Delta_k| < M\{(h^{-1}\delta^r_{\theta_k}+\delta_{\theta_k}+h^2+h^{-1}\delta_n+\delta_0)\delta_{\theta_k}+\tilde{\delta}_n\}$ a.s., M is a constant. For ease of exposition, we take $M > 1$ and $h < 1$ when n is sufficiently large. From (A.5), $\delta_{\theta_{k+1}} \le \{\frac{1}{2}+M(h^{-1}\delta^r_{\theta_k}+\delta_{\theta_k}+h^2+h^{-1}\delta_n+\delta_0)\}\delta_{\theta_k}+M\tilde{\delta}_n+\frac{1}{2}\delta_0$. Note that $\delta_0 h^{-1} \to 0$, $h^{-1}\delta_n \to 0$ and $\tilde{\delta}_n h^{-1} \to 0$. We can assume that $h^2 + h^{-1}\delta_n + \delta_0 \le (8M)^{-1}$, $(M\tilde{\delta}_n + \frac{1}{2}\delta_0)h^{-1/r} < (64M)^{-1}$, $(M\tilde{\delta}_n + \frac{1}{2}\delta_0) < (64M)^{-1}$. If

$$\delta^r_{\theta_k} h^{-1} \le (8M)^{-1} \quad \text{and} \quad \delta_{\theta_k} \le (8M)^{-1} \quad \text{a.s.,} \tag{A.6}$$

we have $\frac{1}{2} + M(h^{-1}\delta^r_{\theta_k} + \delta_{\theta_k} + h^2 + h^{-1}\delta_n + \delta_0) \le \frac{7}{8}$ a.s. Therefore

$$\delta_{\theta_{k+1}} \le \frac{7}{8}\delta_{\theta_k} + M\tilde{\delta}_n + \frac{1}{2}\delta_0 \le (8M)^{-1} \quad \text{a.s.} \tag{A.7}$$

and $\delta_{\theta_{k+1}} h^{-1/r} \le (\frac{7}{8}\delta_{\theta_k} + M\tilde{\delta}_n + \frac{1}{2}\delta_0)h^{-1/r} \le \frac{7}{8}\delta_{\theta_k} h^{-1/r} + (64M)^{-1} \le (8M)^{-1/r}$ a.s. Therefore

$$\delta^r_{\theta_{k+1}} h^{-1} \le (8M)^{-1} \quad \text{a.s.} \tag{A.8}$$

By Lemma 1 and Assumption (C5), (A.6) is true for $k = 1$. Therefore, (A.7) and (A.8) hold for all k. It follows that $\frac{1}{2}+M(h^{-1}\delta^r_{\theta_{k+1}}+\delta_{\theta_{k+1}}+h^2+h^{-1}\delta_n+\delta_0) \le \frac{7}{8}$ a.s. Therefore $\delta_{\theta_{k+1}} \le \frac{7}{8}\delta_{\theta_k} + M\tilde{\delta}_n + \frac{1}{2}\delta_0$ a.s. and

$$\delta_{\theta_{k+1}} \le \left(\frac{7}{8}\right)^k \delta_{\theta_1} + \{1 + \frac{7}{8} + \cdots + \left(\frac{7}{8}\right)^k\}(M\tilde{\delta}_n + \frac{1}{2}\delta_0) \quad \text{a.s.} \tag{A.9}$$

It follows that $\delta_{\hat{\theta}} = O_{a.s.}(\tilde{\delta}_n + \delta_0)$. The first part of Theorem 2 follows from (A.9) and (A.5).

References

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York.

Carroll, R. J., Fan, J., Gijbels, I. and Wang, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Doob, J. L. (1953). *Stochastic Processes.* John Wiley, New York.

Duan, N. and Li, K. C. (1991) Slicing regresion: a link-free regession method. *Ann. Statist.* **19**, 505-530.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 24-34.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645-660.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157-178.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the diffusion function in nonparametric autoregression. *J. Econometrics* **81**, 223-242.

Higgins, M. L. and Bera, A. K. (1992). A class of nonlinear ARCH models. *International Economic Rev.* **33**, 137-158.

Hosking, J. R. M. (1980) Lagrange-multiplier tests of time series model. *J. Roy. Statist. Soc. Ser. B* **42**, 170-181.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Econometrics* **58**, 71-120.

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

Li, W. K. and Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskecasticity. *J. Time Ser. Anal.* **15**, 627-636.

Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika* **83**, 529-540.

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

Mercurio, D. and Sponkoiny, V. (2000). Statistical inference for time-inhomo geneous volatility models. Humboldt-University of Berlin, Manuscript.

Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610-625.

Powell, L, J., Stock, J. M. and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403-1430.

Ruppert, D., Wand, M. P., Holst, U. and Hossjer, O. (1997). Local polynomial variance-function estimation. *Technometrics* **39**, 262-273.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54**, 1461-1481.

Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*. Oxford University Press, London.

Weisberg, S. and Welsh, A. H. (1994). Adapting for the Missing link. *Ann. Statist.* **22**, 1674-1700.

Xia, Y. and An, H-Z. (1999). Projection pursuit autoregression in time series. *J. Time Ser. Anal.* **20**, 693-714.

Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275-1285.

Xia, Y., Tong, H. Li, W. K. and Zhu, L.-X. (2000). A unified approach to dimension reduction. Technical report #275, Department of Statistis, University of Hong Kong.

Department of Zoology, University of Cambridge, Downing street, Cambridge, CB2 3EJ, U.K.

E-mail: ycxia@zoo.cam.ac.uk

Department of Statistics, Room B611, Lolumbia House, London School of Economics, Houghton street, London, WC2A 2AE, U.K.

E-mail: n_tong@lse.ac.uk

Department of Statistics and Acturial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: hrntlwk@hkucc.hku.hk

(Received Septmber 2000; accepted November 2001)