Single-index Diffusion Models and Their Estimation

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

[1]Department of Statistics, University of Hong Kong
Pokfulam, Hong Kong
[2]Institute of Mathematics and Statistics, University of Kent
Canterbury, Kent CT2 7NF, U.K.

Summary. We propose a single-index diffusion model in this paper. This model can avoid the 'curse of dimensionality' in estimating a multivariate nonparametric conditional variance. We adopt an absolute deviation estimation method to estimate the model. Comparing with the commonly used estimators, the absolute deviation estimator is more stable and efficient. Some simulations and applications to real data are reported.

Keywords: ARCH; Conditional variance; Local linear smoother; Strong mixing.

1. INTRODUCTION

Diffusion models, albeit commonly described in continuous time, has a long history in stochastic processes (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

*Corresponding author: Professor H. Tong, Department of Statistics, University of Hong Kong, Pokfulam, Hong Kong e-mail address: htong@hkustasa.hku.hk

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

$$\mathbf{y}_i = \mu(\mathbf{x}_i) + \sigma(\mathbf{x}_i)\varepsilon_i, \tag{1.1}$$

where $\{(\mathbf{y}_i, \mathbf{x}_i)\}$ is a two dimensional strictly stationary process having the same marginal distribution as $(\mathbf{y}, \mathbf{x})$ 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 'drift' and 'diffusion' functions. If we take $\mathbf{x}_i = \mathbf{y}_{i-1}$, then (1.1) is a nonparametric generalization of the ARCH model. We call it a drift-plus-diffusion model. There is much literature concerning the estimation of $\mu(x)$ (e.g. Tjøstheim, 1994 and the references therein). Here we are interested in the nonparametric estimation of the diffusion function $\sigma(x)$. Fan and Yao (1998) is a recent study in this area.

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

$$\mathbf{y}_i = \theta^T Z_i + (c_0 + \gamma^T X_i)^{1/2}\varepsilon_i, \tag{1.2}$$

where $Z_i = (\mathbf{y}_{i-1}, \cdots, \mathbf{y}_{i-p})^T$ and $X_i = (\xi_{i-1}^2,\ \cdots,\ \xi_{i-q}^2)^T$ with $\xi_i = \mathbf{y}_i - \theta^T Z_i$. Model (1.2) is useful for many practical situations including those which motivated Engle (*op. cit.*). To extend this model to a more flexible form, we follow the single-indexing idea (e.g. Ichimura, 1993) and propose a single-index drift-plus-diffusion model as follows

$$\mathbf{y}_i = \mu(\theta^T Z_i) + \sigma(\gamma^T X_i)\varepsilon_i, \tag{1.3}$$

where $\{(X_i,\ Z_i,\ \mathbf{y}_i)\}$ is a strictly stationary and strongly mixing sequence. If $\mu(\cdot)$ is piecewise linear and $\sigma(\gamma^T X_i) = (c_0 + c_1\gamma^T X_i)^{1/2}$ with $X_i = (\mathbf{y}_{i-1}^2, \cdots, \mathbf{y}_{i-q}^2)^T$, then (1.3) is the SETAR-ARCH model (Tong, p.116, 1990). Another obvious extension is the additive model (e.g. Linton and Härdle, 1997). However, we shall not investigate the latter in this paper.

Härdle *et al.* (1993) investigated the estimation of $\mu(\cdot)$ and θ in model (1.3) under an *i.i.d.* assumption. Their results can be extended to model (1.3). See Xia *et al.* (1997) for more details. In this paper, we mainly concentrate on the estimation of $\sigma(\cdot)$ and γ. First, we consider the univariate case (1.1). If $E\varepsilon^2 = 1$, then the usual estimate of $\sigma(x)$ is given by

$$\hat{\sigma}_0^2(x) = \frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)\mathbf{y}_i^2}{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)} - \hat{\mu}^2(x),$$

or

$$\hat{\sigma}_2^2(x) = \frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)(\mathbf{y}_i - \hat{\mu}(\mathbf{x}_i))^2}{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)}, \tag{1.4}$$

where $\hat{\mu}(x)$ is a kernel estimator of $\mu(x)$, $K(\cdot)$ is a kernel function, h is the bandwidth and

$$\begin{aligned}\mathcal{K}_{n.h}(\mathbf{x}_i - x) \quad = \quad & K_h(\mathbf{x}_i - x)\sum_{j=1}^n K_h(\mathbf{x}_j - x)(\mathbf{x}_j - x)^2 \\ & -K_h(\mathbf{x}_i - x)(\mathbf{x}_i - x)\sum_{j=1}^n K_h(\mathbf{x}_j - x)(\mathbf{x}_j - x),\end{aligned}$$

with $K_h(\cdot) = K(\cdot/h)$. Here, $\hat{\sigma}_0^2(x)$ and $\hat{\sigma}_2^2(x)$ are local linear smoothers. The corresponding local constant smoothers (i.e. N-W estimators) can be similarly given. Now, $\hat{\sigma}_0^2(x)$ or its N-W form has been investigated by Masry and Tjøstheim (1995) and Härdle and Tsybakov (1997). Further, $\hat{\sigma}_2^2(x)$ has been investigated by Fan and Yao (1998), who have further pointed out that $\hat{\sigma}_2^2(x)$ has some advantages over $\hat{\sigma}_0^2(x)$.

It is well known that the smoothers of the second order polynomial $\mathbf{y}^2$ or $(\mathbf{y} - \hat{\mu}(\mathbf{x}))^2$ are sensitive to aberrant observations. Furthermore, Masry and Tjøstheim (*op. cit.*) and Fan and Yao (*op. cit.*) have both proved that the asymptotic distributions of $\hat{\sigma}_0^2(x)$ and $\hat{\sigma}_2^2(x)$ depend on the fourth moment of ε, which may impact on their practical utility. Therefore, it is worthwhile considering a more robust estimation method.

Since $E(|\mathbf{y}_i - \mu(\mathbf{x}_i)|\big|\mathbf{x}_i = x) = \sigma(x)$ if $E|\varepsilon_i| = 1$, an alternative estimate of the diffusion function is

$$\hat{\sigma}_1(x) = \sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)|\mathbf{y}_i - \hat{\mu}(\mathbf{x}_i)|\Big/\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x). \tag{1.5}$$

We call $\hat{\sigma}_1(x)$ the absolute deviation estimate. It is easy to see that $\hat{\sigma}_1(x)$ needs weaker moment assumptions than $\hat{\sigma}_2(x)$ and is more robust.

In this paper, we first investigate the absolute deviation estimate of $\sigma(x)$ under model (1.1). Compared with the estimators $\hat{\sigma}_0^2(x)$ and $\hat{\sigma}_2^2(x)$, $\hat{\sigma}_1^2(x)$ is more stable and efficient. We then extend this idea to the single-index diffusion model and the single-index drift-plus-diffusion model (1.3). As applications of these models and estimation methods, some financial data sets will be analyzed.

2. NONPARAMETRIC PURE DIFFUSION MODEL

2.1. Univariate Pure Diffusion model

In this section, we consider first the simple case that $\mu(x)$ is known or simply the model $\mathbf{y}_i = \sigma(\mathbf{x}_i)\varepsilon_i$, where $\sigma(\cdot)$ (> 0) is unknown. This is a nonparametric pure diffusion model, also known as the volatility model in econometrics. See Examples 5 and 6 below. Notice that an alternative model can be written as

$$|\mathbf{y}_i| = \sigma_{10}\sigma(\mathbf{x}_i) + \sigma(\mathbf{x}_i)(|\varepsilon_i| - \sigma_{10}) \tag{2.1}$$

or

$$\mathbf{y}_i^2 = \sigma_{20}\sigma^2(\mathbf{x}_i) + \sigma^2(\mathbf{x}_i)(\varepsilon_i^2 - \sigma_{20}), \tag{2.2}$$

where $\sigma_{20} = E\varepsilon^2$ and $\sigma_{10} = E|\varepsilon|$. With these expressions, the estimation of the diffusion function is replaced by the estimation of a (conditional) mean function leading to the simpler forms

$$\hat{\sigma}_1(x) = \frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)|\mathbf{y}_i|}{\sigma_{10}\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)} \quad \text{and} \quad \hat{\sigma}_2^2(x) = \frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)\mathbf{y}_i^2}{\sigma_{20}\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)}.$$

Notice that we are only interested in the relative variation of $\sigma(x)$. Therefore we do not have to pay much attention to the estimation of the constants σ_{10} and σ_{20}. Usually, we may assume $\sigma_{10} = 1$ or $\sigma_{20} = 1$ for convenience.

With no loss of generality, we only consider those $x \in [a_1, a_2]$ and further assume that $\mathbf{x} \in [a_1,\ a_2]$ with density function $f(x)$. From (a.3) and (a.4) in the appendix, if $E\varepsilon^2 = 1$ and (C1)-(C6) hold, then we have

$$\begin{aligned}
E(\hat{\sigma}_1^2(x) - \sigma^2(x))^2 &= [\sigma(x)\sigma''(x)]^2h^4 + \frac{4}{nhf(x)}\sigma^4(x)(\sigma_{10}^{-2} - 1)\kappa_2 + o(h^4 + \frac{1}{nh}),\\
E(\hat{\sigma}_2^2(x) - \sigma^2(x))^2 &= [(\sigma'(x))^2 + \sigma(x)\sigma''(x)]^2h^4\\
&\qquad + \frac{1}{nhf(x)}\sigma^4(x)(\sigma_{40} - 1)\kappa_2 + o(h^4 + \frac{1}{nh}),
\end{aligned}$$

where $\sigma_{40} = E\varepsilon^4$ and $\kappa_2 = \int K^2(x)dx$. From these equations, the corresponding optimal variable bandwidths for $\hat{\sigma}_1^2(x)$ and $\hat{\sigma}_2^2(x)$ are respectively

$$h_1(x) = \Big\{\frac{\sigma^2(x)(\sigma_{10}^{-2} - 1)\kappa_2}{[\sigma''(x)]^2 f(x)}\Big\}^{\frac{1}{5}} n^{-1/5}$$

and

$$h_2(x) = \Big\{\frac{\sigma^4(x)(\sigma_{40} - 1)\kappa_2}{4[(\sigma'(x))^2 + \sigma(x)\sigma''(x)]^2 f(x)}\Big\}^{\frac{1}{5}} n^{-1/5}.$$

In practice, these bandwidths can be estimated by the method of Fan and Yao (1998).

Another simple result from (a.3) and (a.4) is that $\hat{\sigma}_1(x)$ and $\hat{\sigma}_2(x)$ have the same uniform convergence rate. However, more detailed analysis in the following reveals some essential difference between them. By Theorem 3.3 of Masry and Tjøstheim (1995), the following asymptotic normality follows from (a.3) and (a.4) immediately.

Theorem 1. Suppose that conditions (C1)-(C6) hold and $E(\varepsilon^2) = 1$. Then

$$\sqrt{nh}\{\hat{\sigma}_1^2(x) - \sigma^2(x) - [\sigma(x)\sigma''(x)]^2 h^2\} \xrightarrow{D} N(0, 4\kappa_2(\sigma_{10}^{-2} - 1)f^{-1}(x)\sigma^4(x)),$$
$$\sqrt{nh}\{\hat{\sigma}_2^2(x) - \sigma^2(x) - [(\sigma'(x))^2 + \sigma(x)\sigma''(x)]^2 h^2\}$$
$$\xrightarrow{D} N(0, \kappa_2(\sigma_{40} - 1)f^{-1}(x)\sigma^4(x)).$$

The second part of Theorem 1 was earlier obtained by Fan and Yao (1998). If $E|\varepsilon| = 1$, then we have

$$\sqrt{nh}\{\hat{\sigma}_1(x) - \sigma(x) - \sigma''(x)h^2/2\} \xrightarrow{D} N(0,\ \kappa_2(\sigma_{20} - 1)f^{-1}(x)\sigma^2(x)).$$

Notice that the asymptotic distribution of $\hat{\sigma}_1(x)$ only depends on the second moment of ε. In contrast, the fourth moment of ε is needed for $\hat{\sigma}_2(x)$. From Theorem 1, we can list the biases and variances of $\hat{\sigma}_1^2(x)$ and $\hat{\sigma}_2^2(x)$ as follows,

$$\text{bias}\{\hat{\sigma}_1^2(x)\} : [\sigma(x)\sigma''(x)]^2 h^2, \qquad \text{var}\{\hat{\sigma}_1^2(x)\} : \frac{4\kappa_2}{nh}(\sigma_{10}^{-2} - 1)f^{-1}(x)\sigma^4(x),$$
$$\text{bias}\{\hat{\sigma}_2^2(x)\} : [\sigma(x)\sigma''(x) + (\sigma'(x))^2]^2 h^2,\ \text{var}\{\hat{\sigma}_2^2(x)\} : \frac{\kappa_2}{nh}(\sigma_{40} - 1)f^{-1}(x)\sigma^4(x).$$

Generally, $\sigma(x)$ is assumed to be $(\alpha_0 + \alpha_1 x^2)^{1/2}$ or $\exp(\alpha_0' + \alpha_1 x^2)$ with $\alpha_0, \alpha_1 > 0$. See, e.g., Tong (1990, p.116). In these cases, we have $\sigma''(x) > 0$ and therefore $\hat{\sigma}_2^2(x)$ has a bigger bias than $\hat{\sigma}_1^2(x)$. More generally, if $\sigma(x)$ is convex, then $\hat{\sigma}_2^2(x)$

has a bigger bias than $\hat{\sigma}_1^2(x)$. Let $b_1 = [\sigma(x)\sigma''(x)]^2$, $b_2 = [\sigma(x)\sigma''(x) + (\sigma'(x))^2]^2$ and $\mathcal{E} = (b_1/b_2)^{1/5}$. Note that $\mathcal{E} < 1$ in the convex case. Further, assume that both $\hat{\sigma}_1^2(x)$ and $\hat{\sigma}_2^2(x)$ are taken at the respective optimal bandwidths. Let $v_1 = 4(\sigma_{10}^{-2} - 1)$ and $v_2 = \sigma_{40} - 1$. Then the variance ratio of $\hat{\sigma}_1^2(x)$ relative to $\hat{\sigma}_2^2(x)$ is $\mathrm{var}(\hat{\sigma}_1^2(x))/\mathrm{var}(\hat{\sigma}_2^2(x)) = (v_1/v_2)^{4/5}\mathcal{E}$. We have listed the variance ratios for a number of distributions in Table 1. Generally, when the distribution of ε has a kurtosis greater than 3, then $\hat{\sigma}_1^2(x)$ tends to be more efficient than $\hat{\sigma}_2^2(x)$.

TABLE 1: The variance ratios of $\hat{\sigma}_1^2(x)$ with respect to $\hat{\sigma}_2^2(x)$ for different distributions

Distribution of ε	kurtosis	v_1/v_2	variance ratio
Uniform	1.8	1.6667	$1.5048\mathcal{E}$
Normal	3	1.1416	$1.1118\mathcal{E}$
t(5)	9	0.4359	$0.5146\mathcal{E}$
t(10)	4	0.8860	$0.9077\mathcal{E}$
t(15)	3.545	1.0082	$1.0066\mathcal{E}$
t(20)	3.375	1.0426	$1.0339\mathcal{E}$
Logistic	4.2	0.8898	$0.9108\mathcal{E}$
Laplace	6	0.8003	$0.8368\mathcal{E}$
$\chi^2(5)^*$	5.4	0.6218	$0.6838\mathcal{E}$
$\chi^2(10)^*$	4.2	0.7702	$0.8115\mathcal{E}$
$\chi^2(15)^*$	3.8	0.8709	$0.8953\mathcal{E}$
$\chi^2(20)^*$	3.6	0.8884	$0.9097\mathcal{E}$

* *The distribution has been centralized.*

Next, we give two examples to illustrate the robustness of $\hat{\sigma}_1(x)$ and compare $\hat{\sigma}_1^2(x)$ with $\hat{\sigma}_2^2(x)$.

Example 1. We first take a look at the robustness of the absolute deviation estimate. Consider the following model,

$$\mathbf{y} = \exp(2\mathbf{x})\varepsilon, \tag{2.3}$$

where $\varepsilon \sim N(0,1)$ and $\mathbf{x} \sim U(0,1)$. Here $\sigma(x) = \exp(2x)$. Independent samples are generated from (2.3) each with sample size $n = 200$ and $\hat{\sigma}_1(x)$ and $\hat{\sigma}_2(x)$ are calculated. Figures 1(a) and 1(c) are typical examples. Both estimates appear to be reasonable. Next, we add an outlier to the data set as shown in Figures 1(b) and 1(d). Figure 1(b) shows that the outlier does not significantly affect the estimate $\hat{\sigma}_1(x)$. However, Figure 1(d) shows that $\hat{\sigma}_2(x)$ is quite badly affected by the outlier.

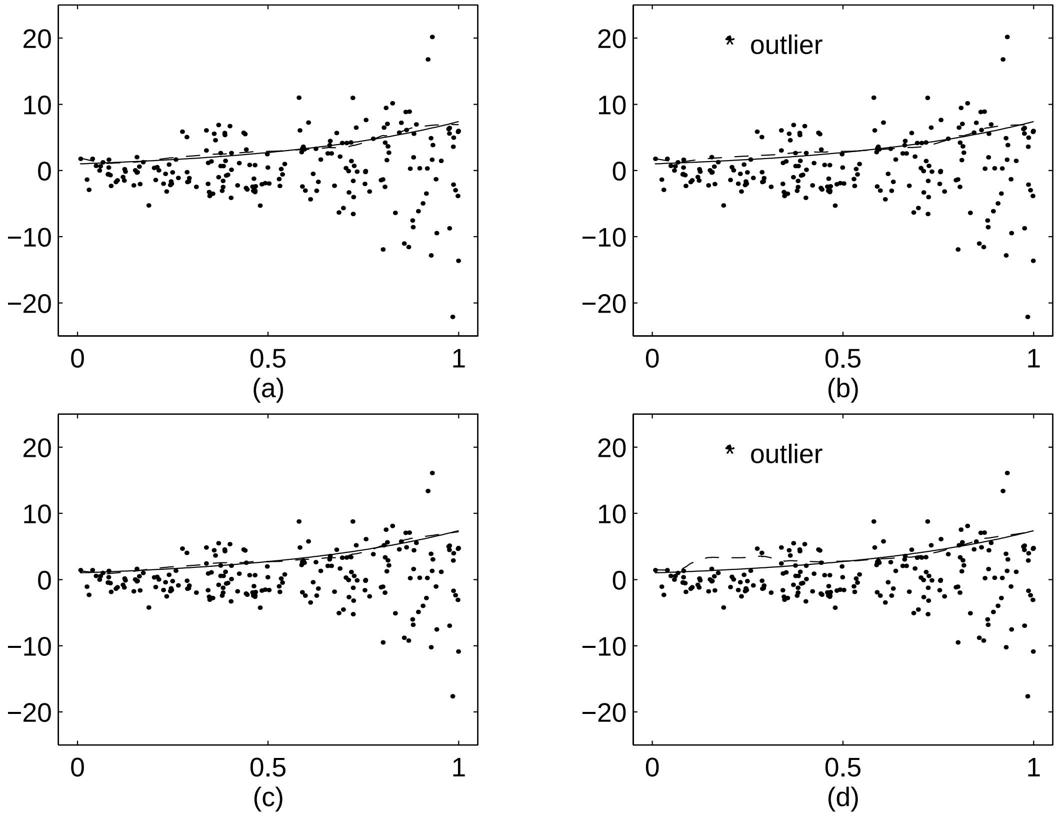

Figure 1: Simulation Results of Example 1. The solid lines denote the real $\sigma(x)$ and the dash lines in (a) and (b) denote $\hat{\sigma}_1(x)$ and those in (c) and (d) denote $\hat{\sigma}_2(x)$. The data in (a) and (c) are a typical data set from (2.3). The data in (b) and (d) are the same data set but contaminated by an outlier.

Example 2. Consider the following models

$$\mathbf{y} = \sqrt{a + b\mathbf{x}^2}\varepsilon \tag{2.4}$$

and

$$\mathbf{y} = (1 + \exp(-r\mathbf{x}^2))\varepsilon, \tag{2.5}$$

where $\mathbf{x}$ is uniformly distributed on $[-1,\ 1]$ and ε is independent of $\mathbf{x}$. We consider two cases: $\varepsilon \sim N(0,1)$, and $\varepsilon \sim L(0,1)$ with the density function $\exp(-|x|)/2$, i.e. a Laplace distribution. We use the mean integrated squared error (MISE) to compare $\hat{\sigma}_1^2(x)$ with $\hat{\sigma}_2^2(x)$. Define

$$\text{MISE}(\hat{\sigma}_i^2(x), h) = E\int_{-1}^{1}\left[\hat{\sigma}_i^2(x) - \sigma^2(x)\right]^2 dx, \quad i = 1, 2.$$

We generate 500 samples from these models each with sample size $n = 400$. Using the standardized Epanechnikov kernel, $K(x) = 3(1 - x^2/5)I(x^2 \leq 5)/(4\sqrt{5})$, the empirical MISE's are shown in Figure 2. The form of (2.4) is very common in practice. Let $r = b/a$; the shape of the MISE$(\cdot, h)$ only depends on r. From Figure 2(a), we can see that $\hat{\sigma}_2(x)$ shows no particular advantage over $\hat{\sigma}_1(x)$ even when ε is normally distributed. On the other hand, when ε has a long tail, e.g. a Laplace distribution, $\hat{\sigma}_2(x)$ is much worse than $\hat{\sigma}_1(x)$ in the sense of MISE. Although model (2.5) is uncommon in practice, it is instructive to see the effect of the degree of diffusion on the estimation of the diffusion function. Now, r controls the degree of the diffusion and Figure 2 shows that the larger is r, the better is $\hat{\sigma}_1(x)$ than $\hat{\sigma}_2(x)$.

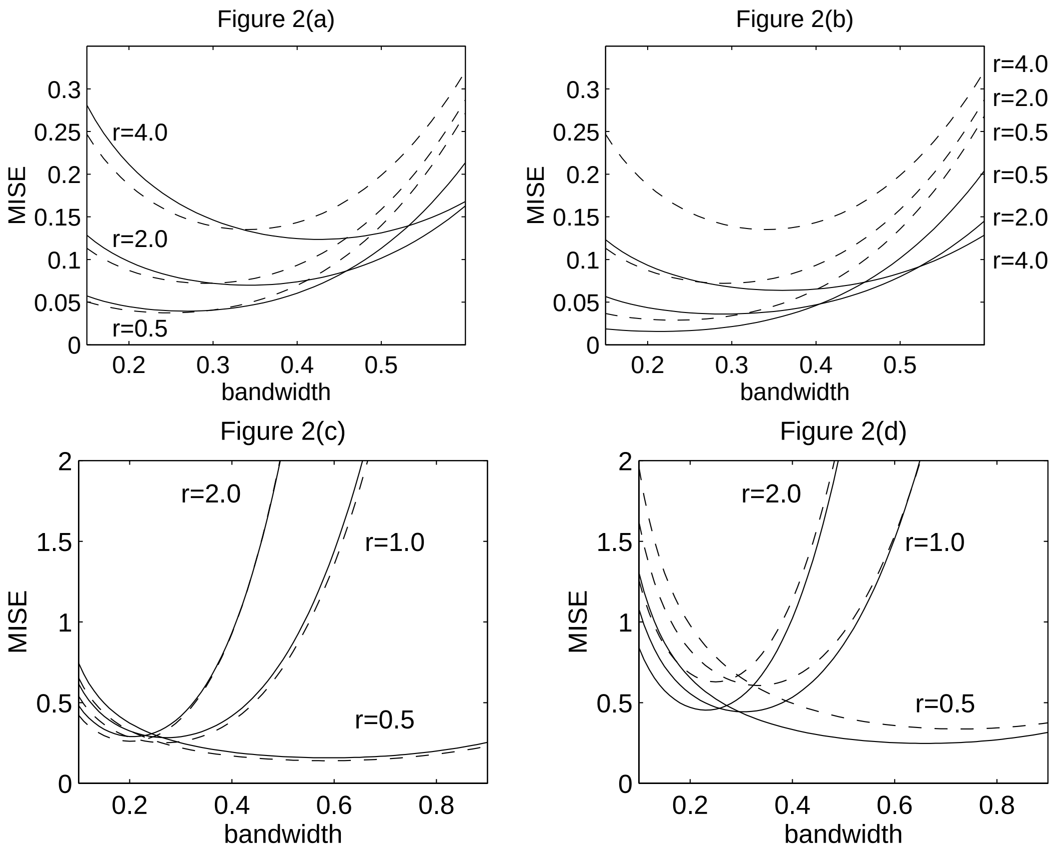

Figure 2: Results of Example 2 for models (2.4)-(2.5). The dash lines denote the MISE's of $\hat{\sigma}_2^2(x)$ and the solid lines denote the MISE's of $\hat{\sigma}_1^2(x)$. In (a) and (c), $\varepsilon \sim N(0,\ 1)$. In (b) and (d), $\varepsilon \sim L(0, 1)$.

2.2. Single-index Pure Diffusion model

Here we consider the single-index pure diffusion model, namely $\mathbf{y}_i = \sigma(\gamma_0^T X_i)\varepsilon_i$, from which

$$|\mathbf{y}_i| = \sigma_{10}\sigma(\gamma_0^T X_i) + \sigma(\gamma_0^T X_i)(|\varepsilon_i| - \sigma_{10}) \tag{2.6}$$

and

$$\mathbf{y}_i^2 = \sigma_{20}\sigma^2(\gamma_0^T X_i) + \sigma^2(\gamma_0^T X_i)(\varepsilon_i^2 - \sigma_{20}^2), \tag{2.7}$$

where $X_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{ip})^T$ and γ_0 is an unknown parameter vector with $\|\gamma_0\| = 1$. Let , be the parameter space. We can take (2.6) and (2.7) as extensions of (2.1) and (2.2) respectively. Now, (2.6) and (2.7) are two single-index regression models; the unknown function and the unknown parameter can be estimated following the method proposed by Härdle *et al.* (1993). Again, we call the estimate from expression (2.6) the absolute deviation estimate. Similarly, denote the estimators using (2.6) by $\hat{\gamma}_1$ and $\hat{\sigma}_1(\cdot)$ and the estimators using (2.7) by $\hat{\gamma}_2$ and $\hat{\sigma}_2(\cdot)$. The asymptotic properties of these estimators can be obtained using standard arguments. Here we consider only the case of (2.6) and list the results below. For detailed proofs, see Xia *et al.* (1997). Throughout the rest of this paper, we assume that the random sample $\{(X_i, \mathbf{y}_i), i = 1, \cdots n\}$ is strictly stationary having the same distribution as $(X, \mathbf{y})$, $\{\varepsilon_i\}$ are *i.i.d.* random variables defined as before. Let $f(\mathcal{X})$ denote the density function of X and $f_\gamma(v)$ that of $\gamma^T X$. Further, for simplicity, let $\mathcal{A} \subset \mathbb{R}^p$ be the union of a number of open convex sets. Given $\rho > 0$, let $\mathcal{A}^\rho$ denote the set of all points in $\mathbb{R}^p$ each at a distance no farther than ρ from $\mathcal{A}$. Assume $f(\mathcal{X}) > 0$ for $\mathcal{X} \in \mathcal{A}^\rho$. We will concentrate on the region $\mathcal{A}$. Let $\mathcal{U} = \{v = \gamma_0^T \mathcal{X} : \ \mathcal{X} \in \mathcal{A}\}$.

Let $\sigma_\gamma(v)$ denote

$$\arg\min_{\sigma} E\Big\{(|\mathbf{y}| - \sigma)^2 \Big| \gamma^T X = v\Big\}. \tag{2.8}$$

In particular, $\sigma(v)$ is just the solution of (2.8) at $\gamma = \gamma_0$. Following the idea of local linear smoothing, the estimator of $\sigma_\gamma(v)$ is the solution of the following minimization problem

$$\min_{c_1} \sum_{i=1}^{n} \Big[|\mathbf{y}_i| - c_1 - c_2(\gamma^T X_i - v)\Big]^2 K_h(\gamma^T X_i - v). \tag{2.9}$$

The solution of (2.9), i.e. the estimator of $\sigma_\gamma(v)$, is

$$\hat{\sigma}_\gamma(v) = \sum_{i=1}^{n} \mathcal{K}_{n.h}(\gamma^T X_i - v)|\mathbf{y}_i| \Big/ \sum_{i=1}^{n} \mathcal{K}_{n.h}(\gamma^T X_i - v). \tag{2.10}$$

In the case that $[\sum_{i=1}^n \mathcal{K}_{n.h}(\gamma^T X_i - v)]^{-1}$ does not exist or is too large, we may only consider the subset $\mathcal{A}$. Let

$$\hat{S}_n(\gamma, h) = \sum_{X_i \in \mathcal{A}} \Big[|\mathbf{y}_i| - \hat{\sigma}_{\gamma<i>}(\gamma^T X_i)\Big]^2,$$

where $\hat{\sigma}_{\gamma<i>}(v)$ is constructed by (2.10) using the data set $\{(X_j, \mathbf{y}_j),\ j \neq i\}$. See Härdle *et al.* (1993). We will estimate the parameter vector γ and the bandwidth h by minimizing $\hat{S}(\gamma, h)$. Let $(\hat{\gamma}, \hat{h})$ denote the pair of solutions. Finally, we obtain the absolute deviation estimates $\hat{\gamma}_1 = \hat{\gamma}$ and $\hat{\sigma}_1(\cdot) = \hat{\sigma}_{\hat{\gamma}}(\cdot)$ in (2.10) with h replaced by $\hat{h}$.

We expect that $\hat{\gamma} \to \gamma_0$ and $\hat{h}/h_0 \to 1$ *a.s.*, where h_0 is the theoretically optimal bandwidth which minimizes

$$J(h) = \int_{\mathcal{A}} E\Big[\hat{\sigma}(\gamma_0^T \mathcal{X}) - \sigma(\gamma_0^T \mathcal{X})\Big]^2 f(\mathcal{X}) d\mathcal{X},$$

where $\hat{\sigma}(v)$ is constructed by (2.10) with γ replaced by γ_0. It is not difficult to see that $h_0 \propto n^{-1/5}$. Next, we only consider the following case

$$\min_{\gamma \in \Gamma_n, h \in \mathcal{H}_n} \hat{S}_n(\gamma, h), \tag{2.11}$$

where $\Gamma_n = \{\gamma :\ \|\gamma - \gamma_0\| \leq C n^{-\delta}\}$ with $\delta > 0$, $\mathcal{H}_n = \{h :\ C_1 n^{-1/5} \leq h \leq C_2 n^{-1/5}\}$ for some constants C and $C_1 < C_2$. Similar to Härdle *et al.* (1993), we take $\delta = 1/2$. These regions are motivated by the following fact. Since we anticipate that $\hat{\gamma}$ is root-n consistent, and we expect $\hat{h}$ to be close to $h_0 \propto n^{-1/5}$, we should look for a minimum of $\hat{S}_n(\gamma, h)$ which involves γ differing from γ_0 by the order $n^{-1/2}$ and h is approximately equal to a constant multiple of $n^{-1/5}$. Our restriction of γ to the cone Γ_n does not exclude any minima of interest and is thus made without loss of generality. Following the proofs of Härdle *et al.* (1993) and Xia *et al.* (1997), we have the following results.

Theorem 2. Suppose that (C1$'$)-(C5$'$) and (C6) (in the appendix) hold and $E|\varepsilon| = 1$. Then

$$\hat{h} = h_0 + o_p(n^{-1/5}),$$
$$n^{1/2}(\hat{\gamma} - \gamma_0) \xrightarrow{D} N(0,\ V^{-}),$$

where $h_0 = \{A_1/(4A_2 n)\}^{1/5}$ with $A_1 = \sigma_0^2 \kappa_2 \int_{\mathcal{A}} f_{\gamma_0}(\gamma_0^T \mathcal{X})^{-1} \sigma(\gamma_0^T \mathcal{X})^2 f(\mathcal{X}) d\mathcal{X}$ and $A_2 = \frac{1}{4} \int_{\mathcal{A}} \sigma''(\gamma_0^T \mathcal{X})^2 f(\mathcal{X}) d\mathcal{X}$,

$$V = (\sigma_{20} - 1) \int_{\mathcal{A}} \{\mathcal{X} - E(X|\gamma_0^T X = \gamma_0^T \mathcal{X})\}\{\mathcal{X} - E(X|\gamma_0^T X = \gamma_0^T \mathcal{X})\}^T \times [\sigma'(\gamma_0^T \mathcal{X}) \sigma(\gamma_0^T \mathcal{X})]^2 d\mathcal{X}$$

and V^- denotes a generalized inverse of V. Consequently, we have

$$\hat{\sigma}_1(v) - \sigma(v) = O_p(n^{-2/5}(\log n)^{1/2})$$

uniformly for $v \in \{\gamma_0^T \mathcal{X} : \ \mathcal{X} \in \mathcal{A}\}$.

Example 4. We simulate 500 random samples of size n from the following model

$$\mathbf{y} = \exp(r(\gamma_{01}\mathbf{x}_1 + \gamma_{02}\mathbf{x}_2))\varepsilon, \quad (\gamma_{01} = 0.6, \ \gamma_{02} = 0.8), \tag{2.12}$$

for $r = 1, 2$ and 4, where $\mathbf{x}_1, \mathbf{x}_2 \overset{i.i.d.}{\sim} U(0,1)$ and $\varepsilon \sim N(0,1)$. We set $n = 100, 300$ and 500. The means and standard deviations of estimated γ_{01} and γ_{02} are shown in Table 2, from which, we find that the parameter estimates are quite reasonable and their variability depends on the curvature of the curve of the diffusion function $\sigma(\cdot)$. The bigger is the curvature the better is the estimates. This is intuitively obvious. Results from a typical data set with $r = 2$ and $n = 300$ are plotted in Figure 3.

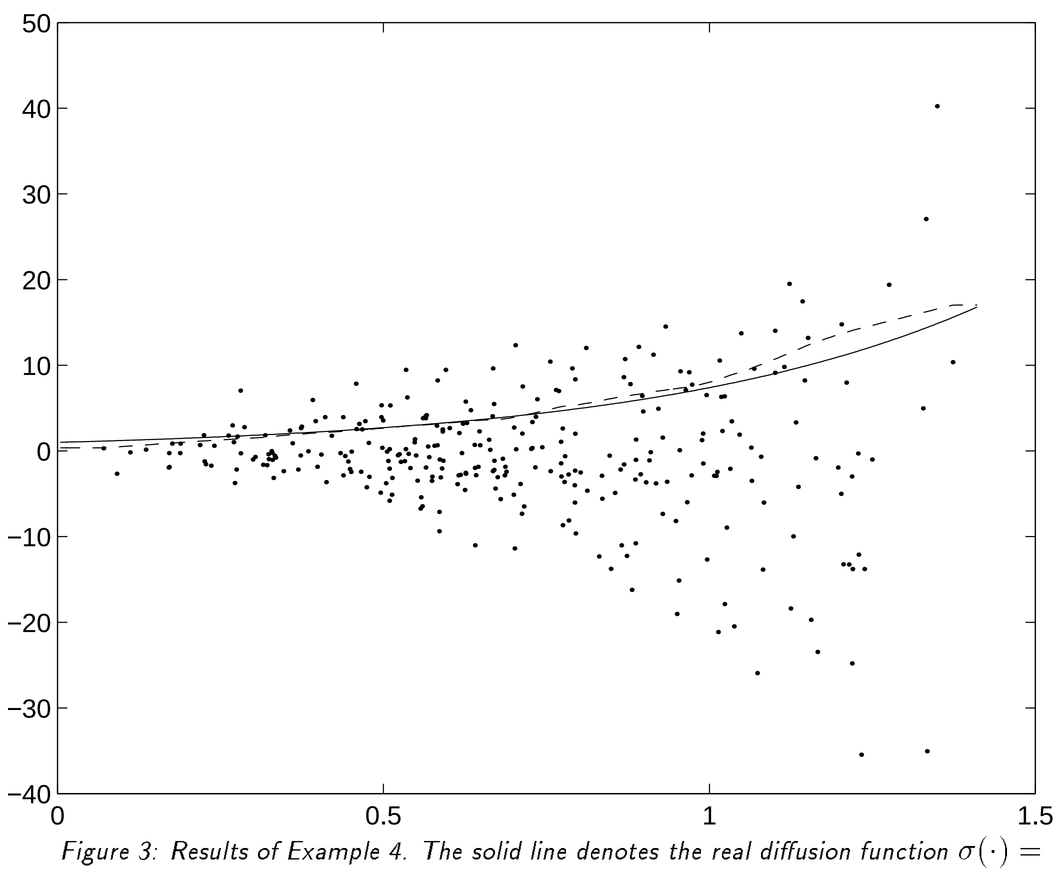

Figure 3: Results of Example 4. The solid line denotes the real diffusion function $\sigma(\cdot) = \exp(2\,\cdot)$. The dash line denotes the estimated diffusion function $\hat{\sigma}_1(\cdot)$. The dots denote $\mathbf{y}_i$ plotted against $\hat{\gamma}^T X_i$.

TABLE 2: Means and standard deviations (in the parentheses) of estimated γ_{01} and γ_{02} for model (2.12)

r	$n=100$		$n=300$		$n=500$	
	γ_{01}	γ_{02}	γ_{01}	γ_{02}	γ_{01}	γ_{02}
1	0.5677	0.7685	0.5883	0.7892	0.5936	0.7930
	(0.0487)	(0.0291)	(0.0199)	(0.0112)	(0.0119)	(0.0068)
2	0.5990	0.7817	0.5992	0.7942	0.6020	0.7948
	(0.0191)	(0.0110)	(0.0064)	(0.0038)	(0.0038)	(0.0021)
4	0.6073	0.7815	0.6088	0.7886	0.6088	0.7905
	(0.0126)	(0.0075)	(0.0046)	(0.0028)	(0.0028)	(0.0016)

Example 5. As an application to real data, we first investigate the daily closing Hang Seng index from 1987 to 1997. To induce approximate stationarity, we consider the first difference of the logarithmically transformed Hang Seng indices. The transformed data ($\mathbf{y}_t$) are plotted in Figure 4(a), which shows three possible 'outliers': the two largest crashes on 26/10/87 and 5/6/89, as well as the largest re-bound on 29/10/97. No trend in $\mathbf{y}_t$ is discernible and the sample autocorrelation function is not significantly different from the Kronecker delta function. Thus, we can assume that $\mu(x) = 0$. Consequently, we consider the following nonparametric pure diffusion model

$$\mathbf{y}_t = \sigma(\gamma^T X_t)\varepsilon_t,$$

where $X_t = (\mathbf{y}_{t-1}^2, \cdots, \mathbf{y}_{t-5}^2)^T$. The chosen lags are motivated by some background knowledge of the index. Using the above method and the transformed data set, we obtained estimates of γ as $\hat{\gamma}_1 = (0.221, 0.161, 0.556, 0.126, 0.775)^T$ and $\hat{\gamma}_2 = (0.049, 0.164, -0.138, -0.309, 0.783)^T$. Note that the estimate $\hat{\gamma}_1$ satisfies the non-negativity assumption of γ in Engle (1982). The estimated functions $\hat{\sigma}_1(\cdot)$ and $\hat{\sigma}_2(\cdot)$ are shown in Figures 4(b) and 4(e). Next, we removed the outliers by replacing each of them with a zero. We obtained the estimates $\hat{\gamma}_1 = (0.276, 0.291, 0.565, 0.412, 0.592)^T$ and $\hat{\gamma}_2 = (0.043, 0.117, 0.253, -0.016, 0.959)^T$. The estimate $\hat{\gamma}_1$ continues to satisfy the non-negativity assumption. The estimate $\hat{\gamma}_2$ now practically satisfies the non-negativity assumption. The estimated functions $\hat{\sigma}_1(\cdot)$ and $\hat{\sigma}_2(\cdot)$ are shown in Figures 4(c) and 4(f). Figures 4(d) and 4(g) show that the outliers have a strong effect on $\hat{\sigma}_2(\cdot)$. In contrast, the outliers show no significant effects on the absolute deviation estimates. Further, as expected from Theorem 1,

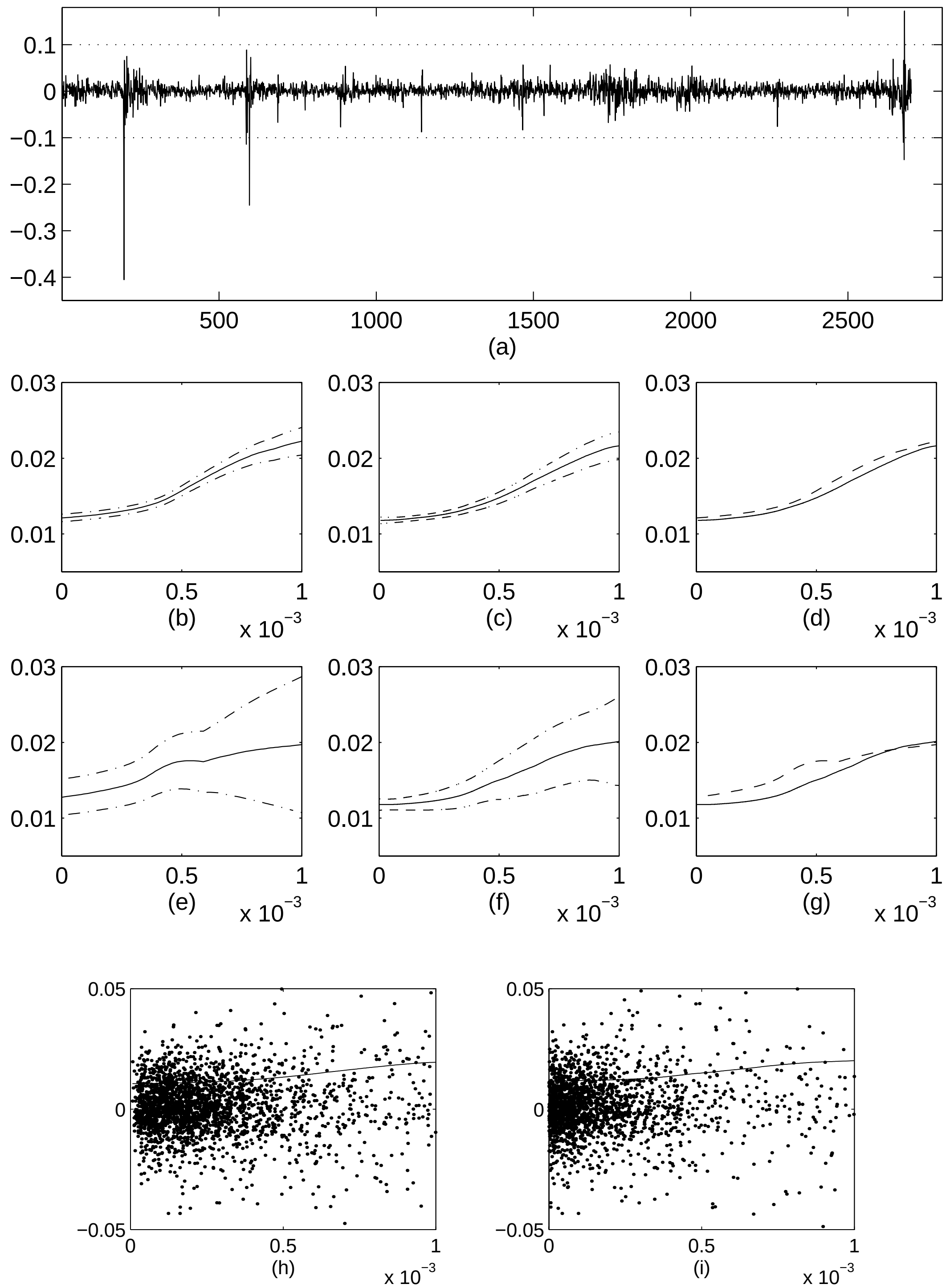

Figure 4: Results of Example 5. (a) is the transformed data. The solid line in (b) and the dash line in (d) denote $\hat{\sigma}_1(\cdot)$ using the transformed data. The solid line in (e) and the dash line in (g) denote $\hat{\sigma}_2(\cdot)$ using the transformed data. The solid lines in (c), (d) and (h) denote $\hat{\sigma}_1(\cdot)$ using the transformed data with the outliers removed. The solid lines in (f), (g) and (i) denote $\hat{\sigma}_2(\cdot)$ using the transformed data with the outliers removed. The dash-dot lines in (b), (c), (e) and (f) denote the approximate 95% pointwise confidence intervals for $\sigma(\cdot)$. The dots in (h) are the $\mathbf{y}_t$ plotted against $\hat{\gamma}_1^T X_t$. The dots in (i) are the $\mathbf{y}_t$ plotted against $\hat{\gamma}_2^T X_t$.

we can construct a narrower pointwise confidence interval for $\sigma(\cdot)$ by using $\hat{\sigma}_1(\cdot)$ than $\hat{\sigma}_2(\cdot)$. This is confirmed by Figures 4(b), 4(c), 4(e) and 4(f).

Example 6. Next, we consider the daily closing S&P 500 index from 29/7/94 to 29/7/97. The original data $(\mathbf{x}_t)$ are plotted in Figure 5. After the transformation $\mathbf{y}_t = \log(\mathbf{x}_t) - \log(\mathbf{x}_{t-1})$, again no trend in $\mathbf{y}_t$ is discernible and the sample autocorrelation function is not significantly different from the Kronecker delta function. Thus, we can assume that $\mu(x) = 0$. Consequently, we consider the same nonparametric pure diffusion model as in Example 5. Using the same method, we obtained $\hat{\gamma}_1 = (0.809,\ 0.054,\ 0.385,\ 0.438,\ -0.052)^T$. The results lend some support to the model proposed by Engle (1982). The function $\sigma(x)$ estimated by the absolute deviation estimation method is shown in Figures 5(b) and 5(c).

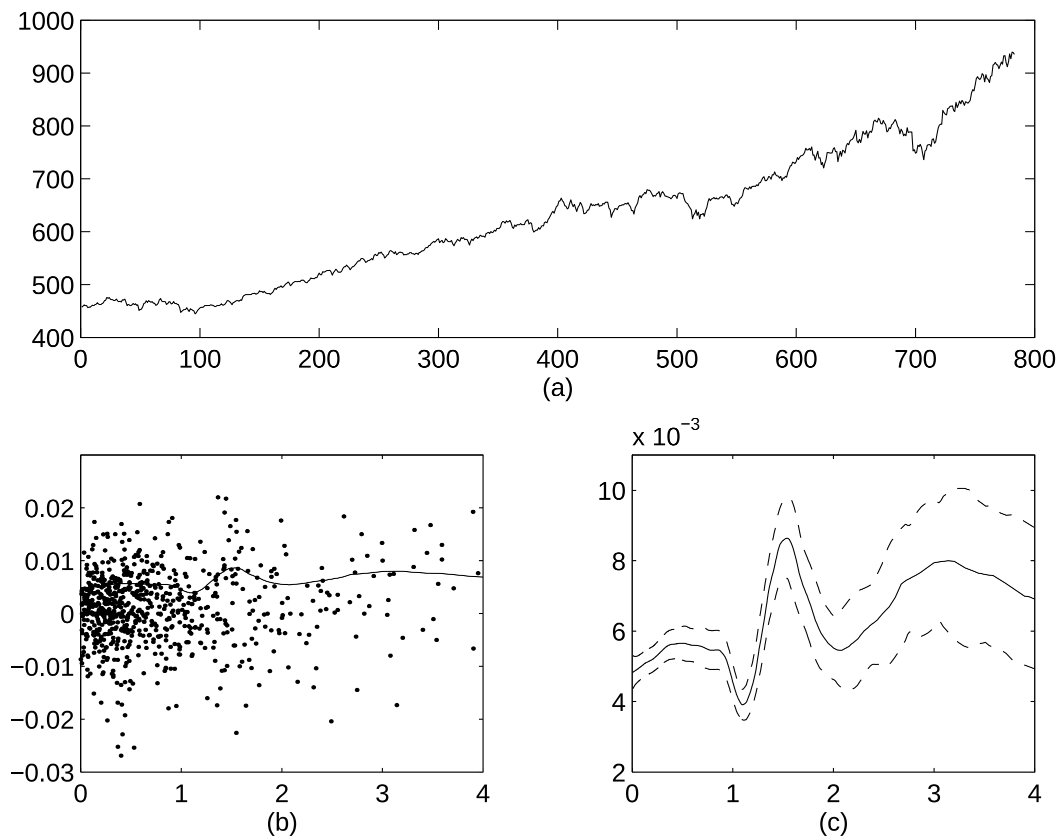

Figure 5: Results of Example 6. (a) is the original data. The solid lines in (b) and (c) denote the estimated diffusion function $\hat{\sigma}_1(\cdot)$. The dash lines in (c) denote the approximate 95% pointwise confidence intervals for $\sigma(\cdot)$. The dots in (b) are the $\mathbf{y}_t$ plotted against $1000 \times \hat{\gamma}_1^T X_t$.

From Figures 5(b) and 5(c), we find that $\sigma(\cdot)$ is not a simple increasing function of $\gamma^T X_t$ as is often assumed in the literature. One tentative explanation is that after a few days of high diffusion, the investors may become more conservative, thus reducing the diffusion. We have made some investigations to the other foreign exchange data sets and similar results are obtained. This kind of non-monotonicity was earlier noticed by Engle and Bollerslev (1986) and Higgins and Bera (1992). They proposed quadratic ARCH models and non-symmetric ARCH models to model the non-monotonicity. Here, we discern the non-monotonicity using a purely nonparametric setup.

3. SINGLE-INDEX DRIFT-PLUS-DIFFUSION MODELS

In this section, we consider the drift-plus-diffusion model (1.3). We assume that $\{(X_i, Z_i, \mathbf{y}_i)\}$ is a strictly stationary and strongly mixing sequence and they have the same marginal distribution as $(X, Z, \mathbf{y})$. Denote the density function of Z by $g(\mathcal{Z})$ and the density function of $\theta^T Z$ by $g_\theta(v)$. We first use the local linear smoother to estimate the unknown drift function $\mu(\cdot)$ and parameter θ_0 following the method of Härdle *et al.* (1993). Following the procedure of (2.9)-(2.10), we define

$$\hat{\mu}_\theta(v) = \frac{\sum_{i=1}^n \mathcal{V}_{n.b}(\theta^T Z_i - v)\mathbf{y}_i}{\sum_{i=1}^n \mathcal{V}_{n.b}(\theta^T Z_i - v)}, \tag{3.1}$$

where

$$\begin{aligned}\mathcal{V}_{n.b}(\theta^T Z_i - v) = {} & V_b(\theta^T Z_i - v)\sum_{j=1}^n V_b(\theta^T Z_j - v)(\theta^T Z_j - v)^2 \\ & - V_b(\theta^T Z_i - v)(\theta^T Z_i - v)\sum_{j=1}^n V_b(\theta^T Z_j - v)(\theta^T Z_j - v),\end{aligned}$$

$V_b(\cdot) = V(\cdot/b)$, $V(\cdot)$ being another kernel function and b another bandwidth. Similar to $\mathcal{A}$, let $\mathcal{D} \subset \mathbb{R}^q$ be the union of a number of open convex sets. Given $\rho > 0$, let $\mathcal{D}^\rho$ denote the set of all points in $\mathbb{R}^q$ each at a distance no farther than ρ from $\mathcal{D}$. We assume $g(\mathcal{Z}) > 0$ for $\mathcal{Z} \in \mathcal{D}^\rho$. We will concentrate on the region $\mathcal{D}$. Let $\mathcal{V} = \{v = \theta_0^T \mathcal{Z} : \ \mathcal{Z} \in \mathcal{D}\}$ and

$$\hat{M}_n(\theta, b) = \sum_{Z_i \in \mathcal{D}} \left[\mathbf{y}_i - \hat{\mu}_{\theta<i>}(\theta^T Z_i)\right]^2,$$

where $\hat{\mu}_{\theta<i>}(v)$ is constructed by (3.1) using data set $\{(Z_j, \mathbf{y}_j),\ j \neq i\}$. We estimate θ and b, $\hat{\theta}$ and $\hat{b}$ say, by

$$\min_{\theta\in\Theta_n, b\in\mathcal{B}_n} \hat{M}_n(\theta, b), \tag{3.2}$$

where $\Theta_n = \{\theta :\ \|\theta-\theta_0\| \le Cn^{-1/2}\}$, $\mathcal{B}_n = \{b :\ C_1 n^{-1/5} \le b \le C_2 n^{-1/5}\}$ for some constants C and $C_1 < C_2$. See section 2.2. Let

$$\begin{aligned} A_1' &= \kappa_2(\sigma_{20}-1)E\Big[g_{\theta_0}(\theta_0^T Z)^{-1}\sigma(\gamma_0^T X)^2 I(Z\in\mathcal{D})\Big], \\ A_2' &= \frac{1}{4}E\Big[\mu''(\theta_0^T Z)^2 I(Z\in\mathcal{D})\Big]^2 \end{aligned}$$

and

$$W = E\Big\{\Big[Z - E(Z|\theta_0^T Z)\Big]\Big[Z - E(Z|\theta_0^T Z)\Big]^T \mu'(\theta_0^T Z)^2\sigma(\gamma_0^T X)^2 I(Z\in\mathcal{D})\Big\}.$$

From Xia and Li (1997) (see also Härdle *et al*, 1993), we have the following theorem.

Theorem 3. Suppose that (C1′)-(C5′) and (C6) (in the appendix) hold, we have

$$\begin{aligned} &\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{D} N(0, W^-), \\ &\hat{b} = b_0 + o_p(n^{-1/5}), \\ &\sup_{Z\in\mathcal{D}}\Big|\hat{\mu}_{\hat{\theta}}(\hat{\theta}^T Z) - \mu(\theta_0^T Z)\Big| = O_p(n^{-2/5}(\log n)^{1/2}), \end{aligned}$$

where W^- denotes a generalized inverse of W and $b_0 = \{A_1'/(4A_2' n)\}^{1/5}$.

Given γ, let

$$\sigma_\gamma(v) = E(|\mathbf{y} - \mu(\theta_0^T Z)|\Big|\gamma^T X = v).$$

We estimate $\sigma_\gamma(v)$ by

$$\tilde{\sigma}_\gamma(v) = \sum_{i=1}^n \mathcal{K}_{n.h}(\gamma^T X_i - v)|\mathbf{y}_i - \hat{\mu}_{\hat{\theta}}(\hat{\theta}^T Z_i)|\Big/\sum_{i=1}^n \mathcal{K}_{n.h}(\gamma^T X_i - v). \tag{3.3}$$

Next, we obtain the estimates of γ and h by minimizing (2.11) with $|\mathbf{y}_i|$ replaced by $|\mathbf{y}_i - \hat{\mu}_{\hat{\theta}}(\hat{\theta}^T Z_i)|$ and $\hat{\sigma}_{\gamma<i>}(\cdot)$ replaced by $\tilde{\sigma}_{\gamma<i>}(\cdot)$, where $\tilde{\sigma}_{\gamma<i>}(\cdot)$ is constructed by (3.3) using data set $\{(X_j, Z_j, \mathbf{y}_j),\ j\neq i\}$.

Following the same procedure as in section 2, we can obtain the estimators of $\tilde{\gamma}$ and therefore $\tilde{\sigma}_{\tilde{\gamma}}(v)$. For these estimators, we have the following asymptotic properties. The proof is tedious but can be obtained by following the method of Xia and An (1998).

Theorem 4. Suppose that (C1′)-(C5′) and (C6) (in the appendix) hold and $E|\varepsilon| = 1$. If $\delta = 3/10$ (see Hall, 1989) in $,_n$ (defined after (2.11)), then $\|\tilde{\gamma} - \gamma_0\| = O_p(n^{-2/5})$ and

$$\tilde{\sigma}_{\tilde{\gamma}}(v) - \sigma(v) = O_p(n^{-2/5}(\log n)^{1/2})$$

uniformly for $v \in \{\gamma_0^T \mathcal{X} : \mathcal{X} \in \mathcal{A}\}$.

Example 7. We simulate the following models

$$\mathbf{y}_t = \sin(2\pi(\theta_1 \mathbf{y}_{t-1} + \theta_2 \mathbf{y}_{t-2})) + e^{-8(\gamma_1 \mathbf{y}_{t-1} + \gamma_2 \mathbf{y}_{t-2})^2} \varepsilon_t, \tag{3.4}$$

where $\theta_1 = 0.6$, $\theta_2 = 0.8$, $\gamma_1 = 0.707$, $\gamma_2 = 0.707$ and ε_t's are independent $N(0, \frac{\pi}{2})$. We generate 400 independent samples each of size $n = 200$ and 500. The simulated results are listed in Table 3. From Table 3, we see that the estimates of the parameters are quit reasonable. Therefore accurate estimates of the unknown functions can be obtained. Figure 6 gives the estimated functions from a typical data set with sample size $n = 200$. The fit appears to be good.

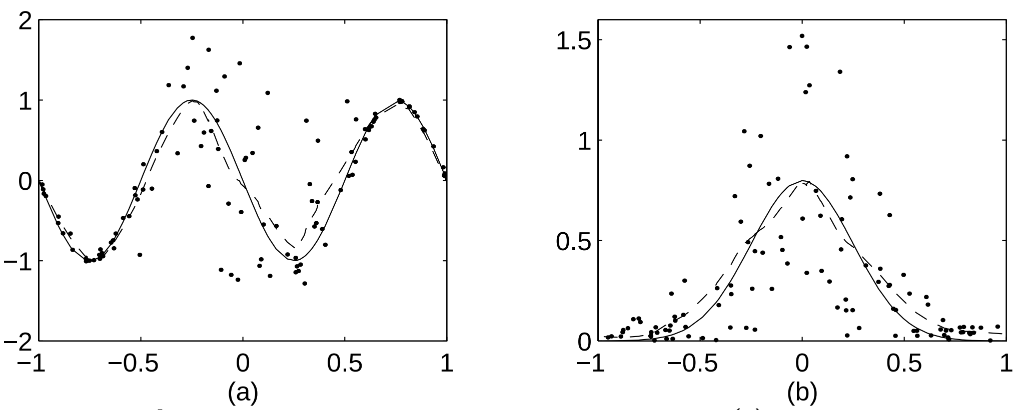

Figure 6: Simulation results of Example 7. The solid lines in (a) denotes the real mean function and that in (b) denotes the real diffusion function. The dash line in (a) denotes the estimated mean function and that in (b) denotes the estimated diffusion function. The dots in (a) are the $\mathbf{y}_i$ plotted against $\hat{\theta}_1 \mathbf{y}_{t-1} + \hat{\theta}_2 \mathbf{y}_{t-2}$. The dots in (b) are absolute values of residuals plotted against $\tilde{\gamma}_1 \mathbf{y}_{t-1} + \tilde{\gamma}_2 \mathbf{y}_{t-2}$.

TABLE 3: Means and standard deviations (in the parentheses) of the estimated θ and γ for models (3.4) with different sample size n

	$n = 200$		$n = 500$	
θ	0.6070(0.0109)	0.7803(0.0117)	0.6039(0.0012)	0.7950(0.0021)
γ	0.6812(0.0086)	0.7217(0.0065)	0.6848(0.0053)	0.7237(0.0019)

An obvious approach to improve the estimation is to iterate the procedure and use the weighted local linear smoother as follows. First, use the above method to obtain $\hat{\mu}_{\hat{\theta}}(\hat{\theta} Z_i)$ and $\tilde{\sigma}_{\tilde{\gamma}}(\tilde{\gamma}^T X_i)$. Then, re-estimate $\mu_\theta(v)$ by minimizing

$$\sum_{Z_i \in \mathcal{D}} V_b(\theta^T Z_i - v)\Big[(\mathbf{y}_i - c_1 - c_2(\theta^T Z_i - v))/\tilde{\sigma}_{\tilde{\gamma}}(\tilde{\gamma}^T X_i)\Big]^2.$$

Denote the solution of c_1 to the above minimization by $\tilde{\mu}_\theta(v)$. We estimate θ by minimizing (3.2) with $\hat{\mu}_{\theta<i>}(v)$ replaced by $\tilde{\mu}_{\theta<i>}(v)$. Denote the estimate by $\tilde{\mu}_{\tilde{\theta}}(\cdot)$. Finally, replace $\hat{\mu}_{\hat{\theta}}(\hat{\theta}^T Z_i)$ in (3.3) by $\tilde{\mu}_{\tilde{\theta}}(\tilde{\theta}^T Z_i)$ and obtain the estimate of $\sigma(\cdot)$. Repeat this procedure until the estimates become stable. The following example is an application of the idea to a linear autoregression model with a diffusion function which is unknown.

Example 8. We simulated the following models

$$
\begin{aligned}
(i) \qquad & \mathbf{y}_t = \alpha_1 \mathbf{y}_{t-1} + \alpha_2 \mathbf{y}_{t-2} + \exp(-8(\mathbf{y}_{t-1} + \mathbf{y}_{t-2})^2)\varepsilon_t, \\
(ii) \qquad & \mathbf{y}_t = \alpha_1 \mathbf{y}_{t-1} + \alpha_2 \mathbf{y}_{t-2} + (0.1 + \sin^2(\frac{\pi}{2}(\mathbf{y}_{t-1}^2 + \mathbf{y}_{t-2}^2)))\varepsilon_t, \\
(ii) \qquad & \mathbf{y}_t = \alpha_1 \mathbf{y}_{t-1} + \alpha_2 \mathbf{y}_{t-2} + (0.1 + \sin^2(\frac{\pi}{2}\mathbf{y}_{t-1}\mathbf{y}_{t-2}))\varepsilon_t,
\end{aligned}
$$

where $\alpha_1 = 0.5$, $\alpha_2 = 0.4$ and $\varepsilon_t \overset{i.i.d.}{\sim} N(0,\ 1)$. We may estimate $\alpha = (\alpha_1,\ \alpha_2)^T$ by simple least squares method (LS). Denote them by $\hat{\alpha} = (\hat{\alpha}_1,\ \hat{\alpha}_2)^T$. We can also estimate α iteratively as follows. First, use $\{\mathbf{y}_t - \hat{\alpha}_1 \mathbf{y}_{t-1} - \hat{\alpha}_2 \mathbf{y}_{t-2}\}$ and the above single-indexing method to estimate the diffusion functions with $X_i = (\mathbf{y}_{i-1},\ \mathbf{y}_{i-2})^T$. Then, use the estimated diffusion functions as weight functions to obtain weighted least squares estimates of α, $\tilde{\alpha}^{(1)}$ say. Replace $\hat{\alpha}$ in the first step with $\tilde{\alpha}^{(1)}$ and repeat the above procedure to obtain $\tilde{\alpha}^{(2)}$, $\tilde{\alpha}^{(3)}$ and so on. We stop when the $\tilde{\alpha}^{(i)}$'s stabilize. We call this procedure nonparametric re-weighted least square method (NRWLS).

For the above models, we generate 200 samples with sample size $n = 50$ and 100. Table 4 shows the simulation results. Table 4 tells us that the NRWLS method can improve estimation of an autoregression model even if the diffusion function is not in a single-index form (models (ii) and (iii)).

TABLE 4: Means and standard deviations (in the parentheses) of estimated $\alpha = (\alpha_1\ \alpha_2)^T$ for models (i)-(iii)

model	$n = 50$		$n = 100$	
	LS method	NRWLS method	LS method	NRWLS method
(i)	0.5298	0.5191	0.5165	0.5075
	(0.0215)	(0.0131)	(0.0123)	(0.0029)
	0.3742	0.3817	0.3819	0.3933
	(0.0208)	(0.0105)	(0.0123)	(0.0029)
(ii)	0.4598	0.5108	0.4656	0.4781
	(0.0334)	(0.0392)	(0.0243)	(0.0163)
	0.3108	0.3573	0.3287	0.3779
	(0.0324)	(0.0479)	(0.0251)	(0.0150)
(iii)	0.4341	0.4676	0.4616	0.4733
	(0.0219)	(0.0454)	(0.0154)	(0.0178)
	0.3348	0.3558	0.3352	0.3607
	(0.0301)	(0.0408)	(0.0225)	(0.0136)

Example 9. We here report the USD/MARK daily closing exchange rate ($\mathbf{x}_t$) from 5/1/75 to 19/9/89. See Figure 7(a). We first make the transformation $\mathbf{y}_t = \log(\mathbf{x}_t) - \log(\mathbf{x}_{t-1})$. We then use the single-index model to fit the mean function and the diffusion function as

$$\mathbf{y}_t = \mu(\theta^T Z_t) + \xi_t$$

with

$$\xi_t = \sigma(\gamma^T X_t)\varepsilon_t,$$

where $Z = (\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-5})^T$, $X = (\xi^2_{t-1}, \cdots, \xi^2_{t-5})^T$. Using the estimation method of Härdle *et al.* (1993) and the method of this paper, we obtain $\hat{\theta} = (0.182\ 0.581\ 0.177\ {-0.754}\ {-0.174})^T$ and $\hat{\gamma}_1 = (0.767\ 0.512\ 0.303\ {-0.059}\ 0.234)^T$. The estimated $\mu(\cdot)$ and $\sigma(\cdot)$ using the absolute deviation estimation method are shown in Figures 7(b)-7(e). From Figures 7(d) and 7(e), we can draw similar conclusions as in Example 6.

Conclusions: In this paper, we use the single-index model to approximate the unknown diffusion functions and propose the single-index diffusion model. This approximation is helpful for other purposes, e.g. Example 7. The idea of the absolute deviation estimation method is simple but has many advantages over the existing

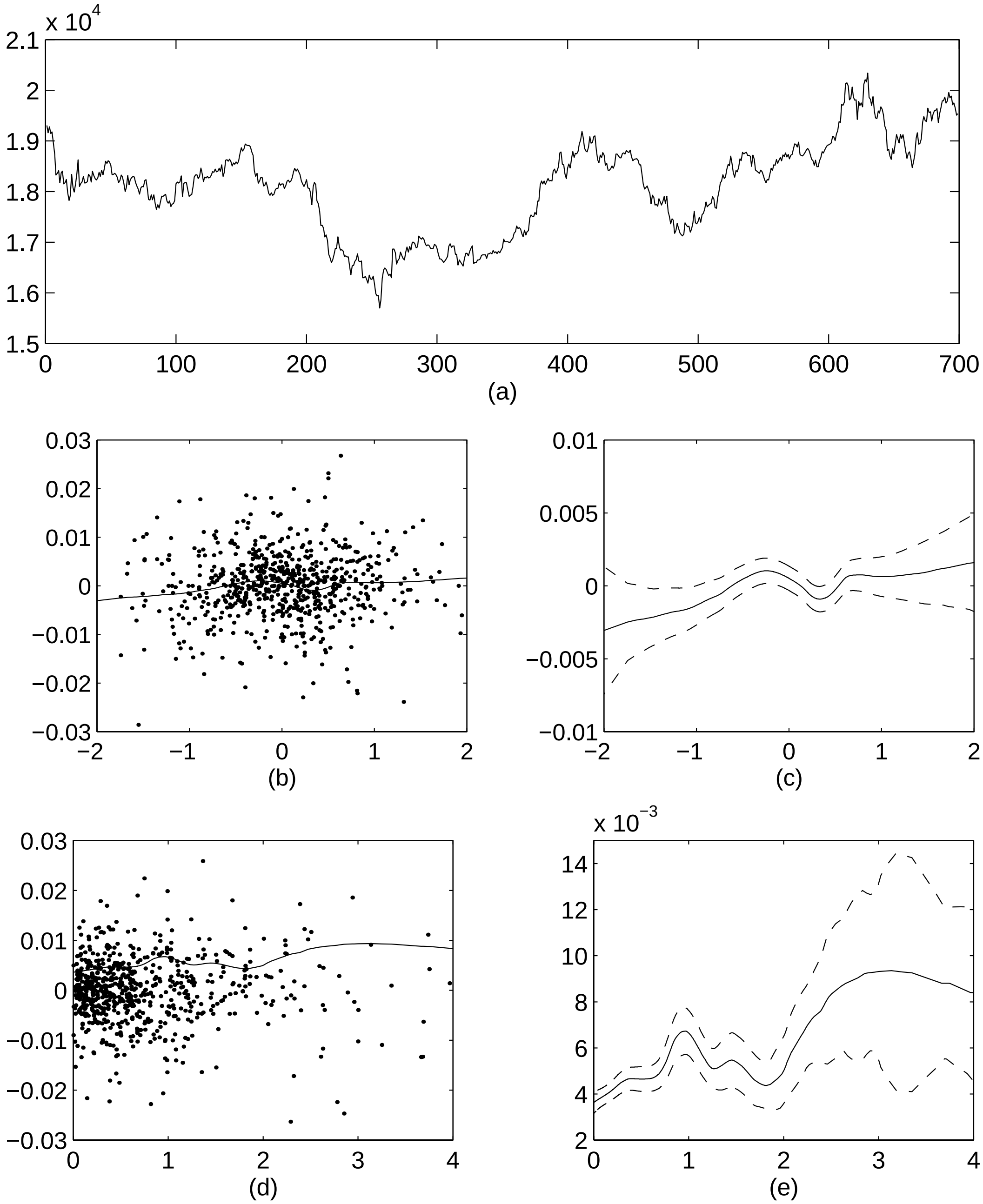

Figure 7: Results for Example 9. (a) is the original data. The solid lines in (b) and (c) denote the estimated mean function and those in (d) and (e) denote the estimated diffusion function $\hat{\sigma}_1(\cdot)$. The dash lines in (d) and (e) denote the approximate 95% pointwise confidence intervals for $\mu(\cdot)$ and $\sigma(\cdot)$ respectively. The dots in (b) are the $\mathbf{y}_t$ plotted against $1000 \times \hat{\theta}^T X_t$. The dots in (d) are the residual $\mathbf{y}_t - \hat{\mu}(\hat{\theta}^T X_t)$ plotted against $1000 \times \hat{\gamma}_1^T X_t$.

methods as shown in the paper. Following the idea of Friedman and Stuetzle (1981), we can further use projection pursuit regression method to approximate the diffusion functions. The calculations can be easily carried out with the module 'PPREG' in S-plus. From $\mathbf{y}_i = \sigma(X_i)\varepsilon$, it follows that $|\mathbf{y}_i| = \sigma(X_i) + \sigma(X_i)(|\varepsilon_i| - 1)$, which is a regression model. We conjecture that other semiparametric regression models can be extended to provide suitable diffusion functions. Since the estimation of the diffusion function is very important for financial data sets, these extensions are of interest.

APPENDIX. ASSUMPTIONS AND PROOFS

To discuss the asymptotic properties of the univariate model (1.1), we need the following assumptions.

(C1) $\sigma(x)$ has a bounded and continuous third order derivative on $[a_1,\ a_2]$;

(C2) $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ is a strictly stationary and strongly mixing sequence with mixing coefficient $\alpha(k) = O(c^k)$ for some $0 < c < 1$;

(C3) $M < f(x) < M'$ for some positive constants M and M' and has bounded derivative on $[a_1,\ a_2]$;

(C4) The conditional density functions $f_{\mathbf{x}_1|\mathbf{y}_1}(x|y)$ and $f_{(\mathbf{x}_1,\mathbf{x}_l)|(\mathbf{y}_1,\mathbf{y}_l)}(x_1,\ x_l|y_1,\ y_l)$ are bounded for all $l > 1$;

(C5) For each i, ε_i is independent of $\{\mathbf{x}_{j+1}, \mathbf{y}_j,\ j < i\}$ and $E|\varepsilon|^l < \infty$, $E|\mathbf{y}|^l < \infty$ for some large enough $l > 0$;

(C6) $K(v)$ and $V(v)$ are two symmetric probability density functions supported on the interval $(-c_0, c_0)$ with bounded derivatives and $\int v^2 K(v)dv = \int v^2 V(v)dv = 1$. Furthermore, the Fourier transforms of $K(v)$ and $V(v)$ are absolutely integrable.

These assumptions are satisfied by most time series models. For detailed discussions, see Härdle and Tsybakov (1997) and Fan and Yao (1998).

To discuss the single-index diffusion model in section 2.2, we need the following assumptions.

(C1$'$) $\sigma_\gamma(v)$ and $f_\gamma(v)$ ($\mu_\theta(v)$ and $g_\theta(v)$) have bounded continuous third order derivatives in $\mathcal{U}^\rho$ ($\mathcal{V}^\rho$) for all $\gamma \in$, $(\theta \in \Theta)$;

(C2$'$) $\{(Z_i, X_i, \mathbf{y}_i)\}$ is a strictly stationary and strongly mixing sequence with mixing coefficient $\alpha(k) = O(c^k)$ for some $0 < c < 1$;

(C3$'$) $M < f(\mathcal{X}),\ g(\mathcal{Z}) < M'$ for some positive constants M and M' and have bounded second derivatives in $\mathcal{A}^\rho$ (or $\mathcal{D}^\rho$);

(C4$'$) The conditional density functions $f_{\gamma_0^T X_1|\mathbf{y}_1}(v|y)$, $f_{(\gamma_0^T X_1, \gamma_0^T X_l)|(\mathbf{y}_1,\mathbf{y}_l)}(v_1, v_l|y_1, y_l)$, $g_{\theta_0^T Z_1|\mathbf{y}_1}(v|y)$ and $g_{(\theta_0^T Z_1, \theta_0^T Z_l)|(\mathbf{y}_1,\mathbf{y}_l)}(v_1, v_l|y_1, y_l)$ are bounded for all $l > 1$;

(C5$'$) For each i, ε_i is independent of $\{X_{j+1},\ Z_{j+1},\ \mathbf{y}_j,\ j < i\}$ and $E|\varepsilon|^l < \infty$, $E|\mathbf{y}|^l < \infty$ for some large enough $l > 0$.

Following the proofs of Theorems 5 and 6 of Masry (1996), we can easily show that

$$\frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)}{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)} = \sigma(x) + \frac{1}{2}\sigma''(x)h^2 + o(h^2) \quad a.s.\ ,$$

$$\frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)(|\varepsilon_i| - \sigma_{10})}{\sum_{i=1}^n K_{n.h}(\mathbf{x}_i - x)} = \frac{1}{nhf(x)}\sum_{i=1}^n K_h(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)(|\varepsilon_i| - \sigma_{10}) + o(h^2) \quad a.s.$$

uniformly for $x \in [a_1,\ a_2]$. Therefore

$$\begin{aligned}\hat\sigma_1(x) &= \sigma(x) + \frac{1}{2}\sigma''(x)h^2 \\ &\quad + \frac{1}{nhf(x)}\sum_{i=1}^n K_h(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)(\sigma_{10}^{-1}|\varepsilon_i| - 1) + o(h^2) \quad a.s.\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$. Thus,

$$\begin{aligned}\hat\sigma_1^2(x) &= \sigma^2(x) + \sigma(x)\sigma''(x)h^2 \\ &\quad + 2\sigma(x)\frac{1}{nhf(x)}\sum_{i=1}^n K_h(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)(\sigma_{10}^{-1}|\varepsilon_i| - 1) + o(h^2) \quad a.s. \qquad \text{(a.1)}\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$. Similarly,

$$\begin{aligned}\hat\sigma_2^2(x) &= \sigma^2(x) + [(\sigma'(x))^2 + \sigma''(x)\sigma(x)]h^2 \\ &\quad + \frac{1}{nhf(x)}\sum_{i=1}^n K_h(\mathbf{x}_i - x)\sigma^2(\mathbf{x}_i)(\sigma_{20}^{-1}\varepsilon_i^2 - 1) + o(h^2) \quad a.s. \qquad \text{(a.2)}\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$.

Following the proofs of Theorems 5 and 6 of Masry (1996), we can easily show that

$$\frac{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)\sigma(\mathbf{x}_i)}{\sum_{i=1}^n \mathcal{K}_{n.h}(\mathbf{x}_i - x)} = \sigma(x) + \frac{1}{2}\sigma''(x)h^2 + o(h^2) \quad a.s.\ ,$$

$$\frac{\sum_{i=1}^{n}\mathcal{K}_{n.h}(\mathbf{x}_i-x)\sigma(\mathbf{x}_i)(|\varepsilon_i|-\sigma_{10})}{\sum_{i=1}^{n}K_{n.h}(\mathbf{x}_i-x)} = \frac{1}{nhf(x)}\sum_{i=1}^{n}K_h(\mathbf{x}_i-x)\sigma(\mathbf{x}_i)(|\varepsilon_i|-\sigma_{10})+o(h^2) \quad a.s.$$

uniformly for $x \in [a_1,\ a_2]$. Therefore

$$\begin{aligned}\hat{\sigma}_1(x) &= \sigma(x)+\frac{1}{2}\sigma''(x)h^2 \\ &\quad +\frac{1}{nhf(x)}\sum_{i=1}^{n}K_h(\mathbf{x}_i-x)\sigma(\mathbf{x}_i)(\sigma_{10}^{-1}|\varepsilon_i|-1)+o(h^2) \quad a.s.\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$. Thus,

$$\begin{aligned}\hat{\sigma}_1^2(x) &= \sigma^2(x)+\sigma(x)\sigma''(x)h^2 \\ &\quad +2\sigma(x)\frac{1}{nhf(x)}\sum_{i=1}^{n}K_h(\mathbf{x}_i-x)\sigma(\mathbf{x}_i)(\sigma_{10}^{-1}|\varepsilon_i|-1)+o(h^2) \quad a.s. \quad \text{(a.3)}\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$. Similarly,

$$\begin{aligned}\hat{\sigma}_2^2(x) &= \sigma^2(x)+[(\sigma'(x))^2+\sigma''(x)\sigma(x)]h^2 \\ &\quad +\frac{1}{nhf(x)}\sum_{i=1}^{n}K_h(\mathbf{x}_i-x)\sigma^2(\mathbf{x}_i)(\sigma_{20}^{-1}\varepsilon_i^2-1)+o(h^2) \quad a.s. \quad \text{(a.4)}\end{aligned}$$

uniformly for $x \in [a_1,\ a_2]$.

ACKNOWLEDGMENTS

We thank Professor J. Fan and Dr. Q. Yao for making their manuscript available to us before its publication. H. Tong was partially supported by EPSRC (Grant no. L16385) and the EC Human Mobility Programme (No. ERBCHRX-CT-94-0693). W. K. Li was supported by the Hong Kong Research Grant Council.

REFERENCES

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**, 618-622.

Doob, J. L. (1953). *Stochastic Processes*, John Wiley & Sons.

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

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Reviews*, **5**, 51-87.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in Stochastic regression. (To appear in *Biometrika*).

Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression, *J. Amer. Statist. Ass.*, **76**, 817-823.

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

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, **21**, 157-178.

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

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

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

Kim, T. Y. and Cox, D. D. (1995). Asymptotic behavior of some measures of accuracy in nonparametric curve estimation with dependent observations. *Journal of Multivariate Analysis*, **53**, 67-93.

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

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, **17**, 571-599.

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.

Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scandinvian Journal of Statistics*, **21**, 97-130.

Tong, H. (1990). *Nonlinear Time Series Analysis: a Dynamic Approach.* Oxford University Press, London.

Xia, Y. and H-Z., An (1998). Projection pursuit autoregression in time series. (to appear in *Journal of Time Series Analysis*)

Xia, Y., Tong, H. and Li, W. K. (1997). On extended partially linear single-index models. Research Report #167, Dept. of Statistics, The University of Hong Kong.