CentER

Discussion Paper

No. 2006–20

SMOOTHED L-ESTIMATION OF REGRESSION FUNCTION

By P. Čížek, J. Tamine, W. Härdle

March 2006

ISSN 0924-7815

TILBURG UNIVERSITY

Smoothed L-estimation of regression function[†]

P. ČÍŽEK,[‡,*] J. TAMINE,[§] and W. HÄRDLE[¶]

[‡] Tilburg University, Dept. of Econometrics & OR, P.O.Box 90153, 5000 LE Tilburg, The Netherlands

[§] Université de la Méditerranée, GREQAM Vieille Charité, 2 rue de la Charité, 13002 Marseille, France

[¶] Humboldt-Universität zu Berlin, Institut für Statistik und Ökonometrie,

Spandauer Str.1, D-10178 Berlin, Germany

(March 18, 2006)

The Nadaraya-Watson nonparametric estimator of regression is known to be highly sensitive to the presence of outliers in data. This sensitivity can be reduced, for example, by using local L-estimates of regression. Whereas the local L-estimation is traditionally done using an empirical conditional distribution function, we propose to use instead a smoothed conditional distribution function. The asymptotic distribution of the proposed estimator is derived under mild β-mixing conditions, and additionally, we show that the smoothed L-estimation approach provides computational as well as statistical finite-sample improvements. Finally, the proposed method is applied to the modelling of implied volatility.

Keywords: nonparametric regression, L-estimation, smoothed cumulative distribution function

JEL Classification: C13, C14

1 Introduction

The nonparametric estimation of regression functions has been receiving much attention in the literature (see Härdle, 1990, and Pagan and Ullah, 1999, for an overview) and one of the most widely used estimator is, without any doubt, the Nadaraya-Watson (NW) estimator by Nadaraya (1964) and Watson (1964). This estimator, being a local average of the response variable, is highly sensitive to the presence of outliers in data (see Barnett and Lewis, 1994, for a general discussion of the concept of an outlier). Possible outliers do not only increase the variance of the estimator, but can also create fictitious peaks and therefore structure in the

[†] The financial support of the German Science Foundation (grant SFB 373, A1) is gratefully acknowledged. The second author was also supported by Greqam, Université de la Méditerranée.

[*] Corresponding author. Email: P.Cizek@uvt.nl.

estimation. In order to robustify the NW estimator, Boente and Fraiman (1994) used a local L-estimate such as local α-trimmed means instead of a locally weighted average. Their procedure consists in using an empirical conditional distribution function (cdf) that allows us to estimate the amount of data to be discarded. We demonstrate that the empirical cdf is not an optimal choice and that the use of a smoothed cdf, which is shown to be a generalization of the empirical one, has substantial advantages. First, following theoretical arguments of Fernholz (1997) for unconditional L-estimates, we expect estimates based on a smoothed cdf to have better finite sample properties. Second, the proposed procedure does not require estimating of the local cdf at every point of the sample but only on an integration grid, and consequently, it will be shown to be less computationally intensive with an increasing sample size.

The paper is organized as follows. In Section 2, we describe both the empirical and smoothed estimators of cdf and explain how they can be used to estimate the conditional L-estimator of regression. In Section 3, asymptotic bounds for the smoothed cdf are given and the asymptotic distribution of the smoothed L-estimator is derived under mild β-mixing conditions. This allows for time-series applications, where outliers are particularly likely to appear as pointed out by Lucas (1995) and Sakata and White (1998), for instance. In Section 4, we show why the smoothed estimator is computationally less time-consuming than the empirical one. In Section 5, we briefly recall the arguments given by Fernholz (1997) in favor of smoothing (in the case of unconditional L-estimates) and perform a Monte Carlo study that demonstrates the superiority of the smoothed estimator in finite samples. Finally in Section 6, we give an example of an application that significantly benefits from the use of a robust nonparametric L-estimation. The proofs of the asymptotic results are given in Appendix A.

2 Robust L-estimates of regression

Here, the conditional L-estimator of regression is introduced in Section 2.1. As it depends on a generally unknown cdf, possible estimators are presented in Section 2.2 and plugged in the conditional L-estimator in Section 2.3. As the proposed procedure depends on bandwidths chosen by cross validation (CV), the choice of a CV criterion with respect to its robustness is discussed in Section 2.4.

2.1 *Model definition*

Let $(\boldsymbol{X}, Y)$ be a $(d+1)$-dimensional random vector. Following Boente and Fraiman (1994), we define the conditional L-estimate of regression by

$$m_L(\boldsymbol{x}) = \int yJ\{F_{\boldsymbol{x}}(y)\}\, dF_{\boldsymbol{x}}(y), \tag{1}$$

where $F_{\boldsymbol{x}}$ denotes the cumulative distribution function of the random variable Y conditional on the event $\{\boldsymbol{X} = \boldsymbol{x}\}$ and J is an L-score function assumed to be continuously differentiable with compact support $\langle a, b\rangle \subset (0, 1)$.

Definition (1) nests several useful statistical parameters. For example, if we consider $F_{\boldsymbol{x}}(y)$ symmetric around the conditional expectation $m(\boldsymbol{x}) = E(Y|\boldsymbol{X} = \boldsymbol{x})$ and the α-trimming score function $J_\alpha(u) = I_{\langle\alpha,1-\alpha\rangle}(u)/(1-2\alpha)$, $\alpha \in \langle 0, 1/2)$, the equality $m_L(\boldsymbol{x}) = m(\boldsymbol{x})$ holds. In such a case, the α-trimmed conditional expectation $m_L(\boldsymbol{x})$ can be used to remove outliers and to estimate conditional expectation in a robust way. In the limit case $\alpha \to 1/2$, $m_L(\boldsymbol{x})$ is equal to the conditional median.

As will be described in Section 2.3, a natural way of estimating $m_L(\boldsymbol{x})$ consists of plugging an estimator of cdf $F_{\boldsymbol{x}}(y)$ in expression (1).

2.2 *Estimation of conditional cumulative distribution function*

Let $f(\boldsymbol{x}, y)$ be the joint density of the random vector $(\boldsymbol{X}, Y)$ and $f(\boldsymbol{x})$ be the marginal density of the d-dimensional random vector $\boldsymbol{X}$. The cdf of the random variable Y conditional on the event $\{\boldsymbol{X} = \boldsymbol{x}\}$ is defined by

$$F_{\boldsymbol{x}}(y) = \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{f(\boldsymbol{x})} dv.$$

The common practice in literature (Härdle, 1990) consists in estimating this function using an empirical cdf defined by

$$\tilde{F}_{\boldsymbol{x}}(y) = \sum_{i=1}^{n} \frac{K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_{\boldsymbol{x}}}\right)}{\sum_{j=1}^{n} K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_{\boldsymbol{x}}}\right)} I(Y_i \le y), \tag{2}$$

where I denotes the indicator function and $K_{\boldsymbol{x}}$ is a d-dimensional product kernel. Further to avoid cumbersome notation, the same bandwith $h_{\boldsymbol{x}}$ is used in each $\boldsymbol{X}$ direction in the theoretical part of the paper; thus, $K_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{d} K_x(x_i/h_{\boldsymbol{x}})$.

The empirical cdf $\tilde{F}_{\boldsymbol{x}}$ is a step function. As pointed out by Fernholz (1997) in the case of unconditional cumulative distribution function, using an empirical cdf may not be the best choice for estimating quantiles, and more generally, computing L-estimates. Furthermore, the step structure of the empirical cdf and its weak regularity properties can be difficult to handle in a theoretical framework.

Therefore, we propose to apply an additional smoothing to the variable Y and to estimate $F_{\boldsymbol{x}}(y)$ by a smoothed cdf

$$\hat{F}_{\boldsymbol{x}}(y) = \sum_{i=1}^{n} \frac{K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_{\boldsymbol{x}}}\right)}{\sum\limits_{j=1}^{n} K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_{\boldsymbol{x}}}\right)} \int_{-\infty}^{y} \frac{1}{h_y} K\left(\frac{u-Y_i}{h_y}\right) du, \tag{3}$$

where h_y is an additional smoothing parameter, bandwidth in the Y direction. Estimator (3) inherits the regularity properties of the univariate kernel K and may thus be used, for instance, for estimating the derivatives of $F_{\boldsymbol{x}}(y)$.

Note that $\hat{F}_{\boldsymbol{x}}(y)$ is a generalization of the empirical cdf $\tilde{F}_{\boldsymbol{x}}(y)$. Under assumptions on kernel K given in Section 3, $\int_{-\infty}^{y} K\left\{(u-Y_i)/h_y\right\}/h_y du \to I(Y_i \leq y)$ for $h_y \to 0$. Hence, $\tilde{F}_{\boldsymbol{x}}(y)$ is the limit case of $\hat{F}_{\boldsymbol{x}}(y)$. As the asymptotic analysis of $\hat{F}_{\boldsymbol{x}}$ will classically assume that the bandwith h_y converges to zero with an increasing sample size n, one can "feel" that the asymptotic properties of $\hat{F}_{\boldsymbol{x}}(y)$ and $\tilde{F}_{\boldsymbol{x}}(y)$ will be identical (see Section 3 for details). Nevertheless, and this is the core of this work, their finite sample behavior and computational properties appear to be significantly different.

2.3 *The empirical and smoothed L-estimators*

In order to estimate $m_L(\boldsymbol{x})$, one can plug estimators (2) and (3) of $F_{\boldsymbol{x}}(y)$ in expression (1). If one plugs in the empirical cdf (2) as proposed by Boente and Fraiman (1994), $m_L(\boldsymbol{x})$ is estimated by the local empirical L-estimator

$$\tilde{m}_L(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{Y_i K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_{\boldsymbol{x}}}\right)}{\sum\limits_{j=1}^{n} K_{\boldsymbol{x}}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_{\boldsymbol{x}}}\right)} J\left\{\tilde{F}_{\boldsymbol{x}}(Y_i)\right\}. \tag{4}$$

Instead employing the empirical cdf, we propose to plug in the smoothed cdf (3). The function $m_L(\boldsymbol{x})$ is then estimated by the local smoothed L-estimator

$$\hat{m}_L(\boldsymbol{x}) = \int yJ\left\{\hat{F}_{\boldsymbol{x}}(y)\right\} d\hat{F}_{\boldsymbol{x}}(y), \tag{5}$$

where the integral is approximated using classical numerical integration routines (Davis and Rabinowitz, 1984).

As will be shown in Section 3, both the local empirical and the local smoothed L-estimators have the same first-order asymptotic properties. However, we demonstrate that estimator (5) has superior computational and finite-sample statistical properties to (4); see Sections 4 and 5, respectively.

2.4 *Cross validation*

On the one hand, we study nonparametric regression estimators that are robust to outliers, contrary to the least-square based methods such as the NW estimator. On the other hand, nonparametric smoothing depends on the choice of smoothing parameters, primarily bandwidth, which are typically chosen by cross-validating some squared error (Park and Marron, 1990). By the same argument as in the case of NW, the standard CV based on squared-errors appears to be sensitive to data contamination and outliers in particular (Leung at al., 1993).

As a remedy, Wang and Scott (1994) proposed to cross-validate the absolute value of the error, the L_1 cross-validation. This and more general CV criteria have been recently analyzed by Leung (2005). Since we primarily use the α-trimming score J_α in this paper, we employ and compare the standard squared-errors cross-validation (CV2) and the L_1 cross-validation (CV1), which corresponds to the most robust choice of $\alpha \to 0.5$ (the median minimizes the least absolute deviation criterion). The CV1 criterion is not only simple to evaluate, but also proved its robust properties in the recent study by Čížek and Härdle (2006).

3 Asymptotic analysis

In order to establish asymptotic bounds on the smoothed cdf (3) and to derive the asymptotic distribution of $\hat{m}_L(\boldsymbol{x})$, the following assumptions are used. They are not the most general ones, but allow for an easy presentation of the main results and proofs.

A1: Random vectors $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ form a sequence of strictly stationary and β-mixing realizations of $(\boldsymbol{X}, Y)$ (see Davidson, 1994, for the concept of β-mixing).

A2: The density $f(\boldsymbol{x}, y)$ is compactly supported and admits continuous derivatives up to order r.

A3: The marginal density $f(\boldsymbol{x})$ admits a strictly positive lower bound b.

A4: The univariate kernel function K is a symmetric compactly supported kernel of order r_y (see Härdle, 1990) such that the integral $K_I(y) = \int_{-\infty}^{y} K(u)\, du$ is defined for all y.

A5: The multivariate kernel function $K_{\boldsymbol{x}}$ is a product kernel, $K_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{d} K_x(x_i/h_{\boldsymbol{x}})$, where K_x is a symmetric compactly supported kernel of order r_x, and $r = \min\{r_x, r_y\} \geq 2$.

A6: The bandwidths $h_{\boldsymbol{x}}$ and h_y satisfy $\lim_{n\to\infty} h_{\boldsymbol{x}} = 0$ and $\lim_{n\to\infty} h_y = 0$ in such a way that $\lim_{n\to\infty} n^{1/2} h_{\boldsymbol{x}}^{3d/2} \to \infty$ and $\lim_{n\to\infty} n^{1/2} h_{\boldsymbol{x}}^{d/2} (\max\{h_{\boldsymbol{x}}, h_y\})^r \to 0$ (the dependence of $h_{\boldsymbol{x}}$ and h_y on n is left implicit for the simplicity of notation).

A7: The L-score function $J : \langle 0, 1\rangle \to R$ is assumed to be non-negative, bounded, and having a bounded derivative almost everywhere.

First, the asymptotic bounds on the smoothed cumulative distribution function (3) are given.

LEMMA 3.1 *Under Assumptions A1 to A6, the following uniform bound holds for* $\hat{F}_{\boldsymbol{x}}(y)$

$$\sup_{x,y} \left| \hat{F}_{\boldsymbol{x}}(y) - F_{\boldsymbol{x}}(y) \right| = O_p\left[n^{-1/2} h_{\boldsymbol{x}}^{-d} + (\max\{h_{\boldsymbol{x}}, h_y\})^r \right].$$

Proof: See Appendix A.1.

We can now derive the asymptotic distribution of the smoothed L-estimator (5).

THEOREM 3.2 *Under Assumptions A1 to A7, the asymptotic distribution of* $\hat{m}_L(\boldsymbol{x})$ *is*

$$n^{1/2} h_{\boldsymbol{x}}^{d/2} \{\hat{m}_L(\boldsymbol{x}) - m_L(\boldsymbol{x})\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$$

for $n \to \infty$, *where*

$$V = \int \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left\{ I_{(-\infty,y\rangle}(w) - F_{\boldsymbol{x}}(y) \right\} dy \right]^2 f(\boldsymbol{x}, w) dw \cdot \int K_{h_{\boldsymbol{x}}}^2(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

Proof: See Appendix A.2.

It is interesting to note that the asymptotic distribution of the smoothed L-estimator (5) is equivalent to that of the empirical L-estimator (4), and therefore, the choice of bandwidth h_y does not asymptotically play any role (provided that Assumption A6 holds). This follows from the fact that the bandwidth h_y does not influence the number of observations used in (local) estimation: it is used to estimate a distribution function, which takes always into account all observations, rather then a density function, which is estimated locally.

4 Computational comparison

The asymptotic distribution of the smoothed L-estimator $\hat{m}_L(\boldsymbol{x})$ is identical to that of the empirical L-estimator $\tilde{m}_L(\boldsymbol{x})$ given by Boente and Fraiman (1994). On the other hand, the computational burden for its computation is much smaller than for the empirical L-estimator as the sample size increases. To show this, we will compare the computational costs for the empirical and smoothed L-estimators, both from a theoretical and an empirical points of view.

We assume that the computation of $K_{\boldsymbol{x}}\{(\boldsymbol{x}-\boldsymbol{X}_i)/h_{\boldsymbol{x}}\}$ requires $C_{\boldsymbol{x}}(d)$ operations, the computation of $K_I\{(y-Y_i)/h_y\}$ requires C_I operations, and the computation of the score function $J(u)$ requires C_J operations. Although the cost of computing $K_{\boldsymbol{x}}\{(\boldsymbol{x}-\boldsymbol{X}_i)/h_{\boldsymbol{x}}\}$ depends on the dimension d, we consider d to be fixed and thus $C_{\boldsymbol{x}}(d)$ to be a constant since d is determined only by the number of employed explanatory variables. Finally to determine the bandwidth parameters, we assume that CV is used on a finite grid of C_c points (for the sake of simplicity, identical for all variables).

Empirical estimation. The calculation of the empirical cdf $\tilde{F}_{\boldsymbol{x}}(\cdot)$ at one point y using (2) needs $n\,(C_{\boldsymbol{x}}(d)+3)$ operations. Since the computation of $\tilde{m}_L(\boldsymbol{x})$ requires the computation of $\tilde{F}_{\boldsymbol{x}}(\cdot)$ at every point Y_i, we get for the empirical L-smoother $\tilde{m}_L(\boldsymbol{x})$

$$n^2\{C_{\boldsymbol{x}}(d)+3\}+n\{C_J+C_{\boldsymbol{x}}(d)+3\}=O\left(n^2\right) \tag{6}$$

as the cost of operations. This is eventually multiplied by a constant, C_c or C_c^d, if bandwidth(s) $h_{\boldsymbol{x}}$ is determined by CV.

Smoothed estimation. Analogously, the computation of the smoothed cdf $\hat{F}_{\boldsymbol{x}}(\cdot)$ at one point y requires $n\,(C_{\boldsymbol{x}}(d)+C_I+2)$. We assume that the integral in expression (5) is approximated on a grid of k points by a numerical integration method of order $m \geq 2$, that is, with the integration

Table 1. Relative computational times of the empirical and smoothed L-estimators for sample size n and an integration grid of size k.

Sample size n	50	100	200	500	1000	2000
Method (k)						
Empirical	0.4	0.8	1.4	3.5	7.0	14.0
Smoothed (25)	0.3	0.2	0.2	0.3	0.3	0.3
Smoothed (50)	0.5	0.4	0.5	0.5	0.5	0.5
Smoothed (100)	1.0	1.0	1.0	1.0	1.0	1.0
Smoothed (150)	2.0	1.4	1.5	1.5	1.5	1.5
Smoothed (250)	3.0	2.5	2.4	2.5	2.5	2.5
Smoothed (500)	5.5	4.9	4.9	5.0	5.0	5.0

error is proportional to k^{-m}.[1] Thus, its computation needs $O(k)$ operations. Consequently, the computation of $\hat{m}_L(\boldsymbol{x})$ requires

$$n\{C_{\boldsymbol{x}}(d) + C_I + 2\}\, O(k) = O(nk) \tag{7}$$

operations. This is again multiplied by a constant, C_c^2 or $C_c^{(d+1)}$, if bandwidths $h_{\boldsymbol{x}}$ and h_y are determined by CV.

Thus, the crucial factor for the computational speed is how quickly k should grow with n. Since the errors made due to the numerical integration should not be larger than the errors (biases and variance) of nonparametric estimation, $k^{-m} < (nh_{\boldsymbol{x}}^d)^{-1/2}$ and hence $k > (nh_{\boldsymbol{x}}^d)^{1/(2m)}$. As $h_{\boldsymbol{x}}^d < n^{1/3}$ by Assumption A6, it follows that $k > n^{2/(3m)}$ in the worst case. For the simplest and least precise methods with $m = 2$, we should choose $k \sim n^{1/3}$. Consequently, the computational complexity of the smoothed L-estimator is at most $O(n^{4/3})$ and further decreases when the optimal or a smaller bandwidth is employed and when better integration methods are used.

Computation time results. For a large n, the smoothed L-estimator requires substantially smaller computational time than the empirical L-estimator. To corroborate this theoretical result, we performed a set of simulations in the univariate case and measure the computational

[1]For example, the well-known trapezoidal and Simpson rules have precisions of orders $m = 2$ and $m = 4$, respectively, see Davis and Rabinowitz (1984).

times of both methods. We used the data generating process

$$Y = m(X) + \varepsilon, \tag{8}$$

where the regressor X is univariate and uniformly distributed on the interval $\langle 0, 1\rangle$, $X \sim U(0, 1)$ and the regression function is given by $m(x) = -1 + \sqrt{x} - x^2$. The error term ε has a symmetric distribution so that the regression function $m(x)$ and the α-trimmed expectation $m_L(x)$ are equal. The estimations are performed using the α-trimming L-score J_α with $\alpha = 0.25$.

Table 1 contains the relative time necessary for the estimation of the empirical L-estimator (4) and of the smoothed one (5) for different sample sizes n and different numbers k of points on the numerical integration grid. Increasing the number of integration points decreases the error of approximation in the integral (4) . In our opinion, choosing 250–500 points is a good choice for applications with samples with up to several thousand observations. Table 1 presents times relative to the smoothed L-estimation using $k = 100$ points integration grid. Results in Table 1 confirm our theoretical findings that the smoothed estimator will be faster to compute for large samples. Already when sample size n is twice the size k of the integration grid, $n > 2k$, the smoothed L-estimator is faster to compute.

5 Finite sample comparison

As noticed in Section 3, the empirical L-estimator and its smoothed counterpart need not be compared from an asymptotic point of view since they share the same asymptotic distribution. However, relying on arguments established by Fernholz (1997) in an unconditional setting, we can hypothesize that the smoothed L-estimator has better finite sample properties than the empirical one. This might seem surprising since the additional smoothing involved in estimating the smoothed cdf $\hat{F}_{\boldsymbol{x}}(\cdot)$ may cause an additional bias (asymptotically negligible but sensible in finite sample). Nevertheless, as demonstrated by Fernholz (1997), this additional bias goes along with a decrease of the variance of the estimator. Because this decrease surpasses the additional bias, the smoothing will result in a gain in terms of the mean squared error of the smoothed conditional L-estimator.

To verify this claim, we perform a set of Monte Carlo simulations and compare the Nadaraya-Watson (NW), the empirical and smoothed L-estimators, where the α-trimming score function

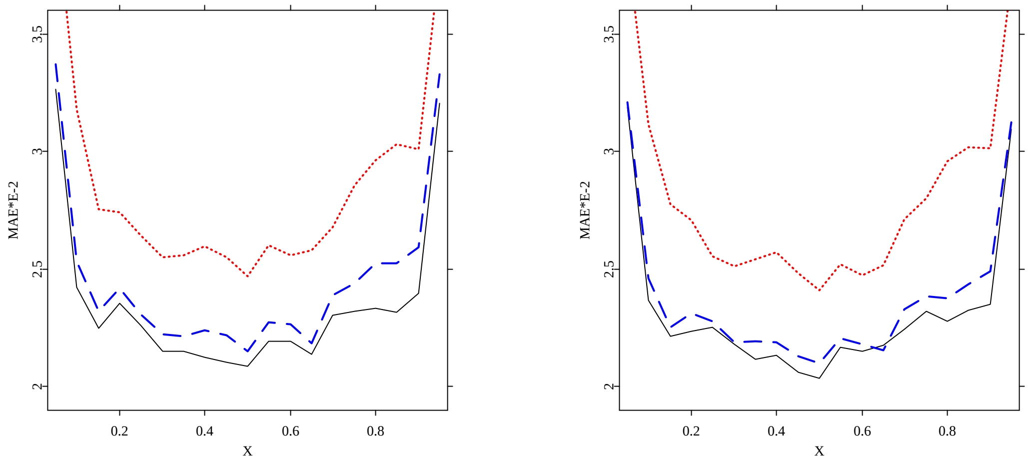


Figure 1. The mean absolute errors of the Nadaraya-Watson (solid line), empirical L-estimator (dotted line), and smoothed L-estimator (dashed line) for the L-score J_α with $\alpha = 0.50$ and different cross-validation rules. The left panel corresponds to CV1 and the right panel to CV2.

J_α is used for $\alpha = 0.25$ and $\alpha = 0.50$ (the choice $\alpha = 0$ would correspond to NW). We use again the data generating process (8), that is $Y = -1 + \sqrt{X} - X^2 + \varepsilon$, where $X \sim U(0,1)$, and study performance of all methods at different error distributions and sample sizes. The presented results are based on 500 replications. All estimates are computed using the quartic kernel and both bandwidths $h_{\boldsymbol{x}}$ and h_y are chosen by L_1 cross validation (CV1) on interval $(0, 0.2)$. Hence to match the use of CV1, the estimation results are compared by means of the mean absolute errors (MAE). One can of course argue that CV1, used to guarantee robust bandwidth choice, could possibly lead to worse performance of all methods, and NW in particular, for Gaussian data. To address this issue, results for the standard squared-error cross validation (CV2) and CV1 are compared in Gausian samples $\varepsilon \sim N(0, 0.1)$ of size $n = 100$, see Figure 1. Even though CV1 is slightly worse than CV2, the differences are negligible for the comparison of the estimation methods.

In the rest of this section, we attempt to compare the finite sample properties of the empirical and smoothed L-estimators for different sample sizes (Section 5.1), for different error-term distributions (Section 5.2), and for data contaminated by outliers (Section 5.3).

5.1 *Influence of sample size*

Let us now compare the empirical and smoothed L-estimators for Gaussian data, $\varepsilon \sim N(0, 0.1)$, at various sample sizes: $n = 50, 100, 200$. The results for J_α with $\alpha = 0.25$ and $\alpha = 0.50$ are

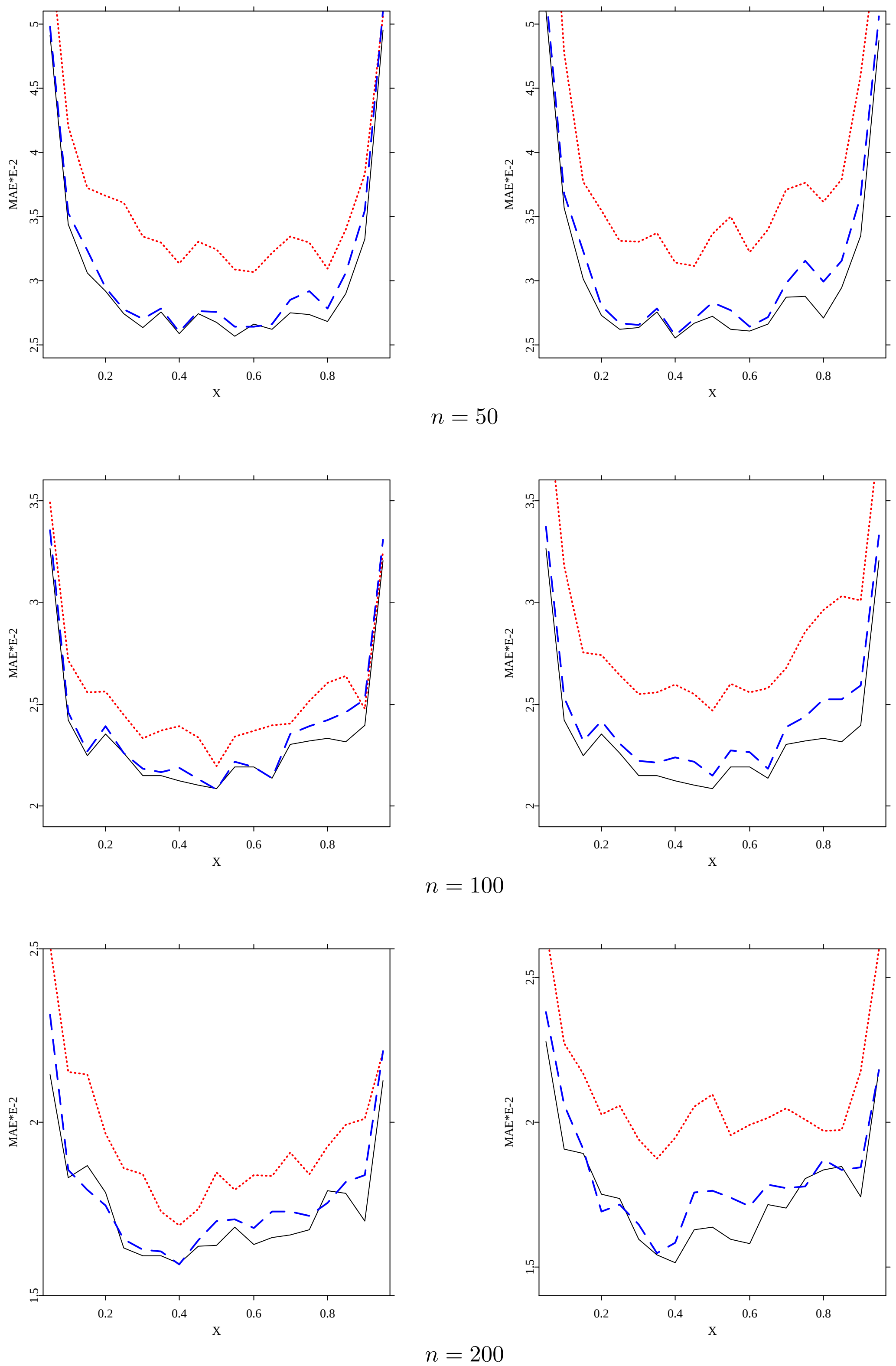


Figure 2. The mean absolute errors of the Nadaraya-Watson (solid line), empirical L-estimator (dotted line), and smoothed L-estimator (dashed line) for different sample sizes n and the L-score J_α. The left panels contain estimators for $\alpha = 0.25$ and the right panels contain estimators for $\alpha = 0.50$.

summarized in Figure 2. The first observation is that the smoothed L-estimator performs in all cases better than the empirical L-estimator and is generally very close to the NW estimator. The relative difference between the empirical and smoothed L-estimators seems to be roughly the same at various sample sizes, but, since the estimation errors are decreasing as n increases, the difference decreases in absolute terms. Moreover, it seems that the difference is smaller for $\alpha = 0.25$ than for $\alpha = 0.50$. This indicates that the smoothed cdf helps more if there are more observations close the discontinuity of the L-score function J_α.

5.2 *Effect of distributional model*

Next, we look at the performance of all methods under various distributional models: for Gaussian $\varepsilon \sim N(0, 0.1)$, Student $\varepsilon \sim \sigma_t t_3$, and double exponential $\varepsilon \sim DExp(\sigma_e)$ errors, where $\sigma_t \doteq 0.057$ and $\sigma_e \doteq 0.071$ are chosen so that the standard deviation of errors is always equal to 0.1. Results for all error distributions and $n = 100$ are summarized in Figure 3. We can again observe that the smoothed L-estimator generally outperforms the empirical L-estimator, although the difference is pronounced only for $\alpha = 0.50$ and is largest for the Gaussian data. Looking at the results for the Student and double-exponential errors, both the empirical and smoothed L-estimators are preferable to the NW estimator.

5.3 *Sensitivity to outliers*

Finally, let us compare all methods for Gaussian samples, $\varepsilon \sim N(0, 0.1)$, of $n = 100$ observations contaminated by three (relatively small) outliers at $x = 0.25, 0.50$, and 0.75 with the values of the dependent variable y uniformly distributed on $(-1.5, 1.5)$, $y \sim U(-1.5, 1.5)$. The results using CV1 are in the upper row of Figure 4. Obviously, NW is heavily influenced by the outliers, but this is not the case of the L-estimators. The smoothed L-estimator again performs better than the empirical L-estimator, but due to the smoothing in the y-direction and its dependence on bandwidth h_y, the smoothed L-estimator exhibits a limited sensitivity to the outliers. For comparison, the bottom row of Figure 4 presents results when CV2 is used. One can observe that, in spite of using locally robust α-trimmed means, a non-robust CV criterion can significantly worsen the estimation results. The only, although unexplained exception in this respect is the empirical L-estimator for $\alpha = 0.25$.

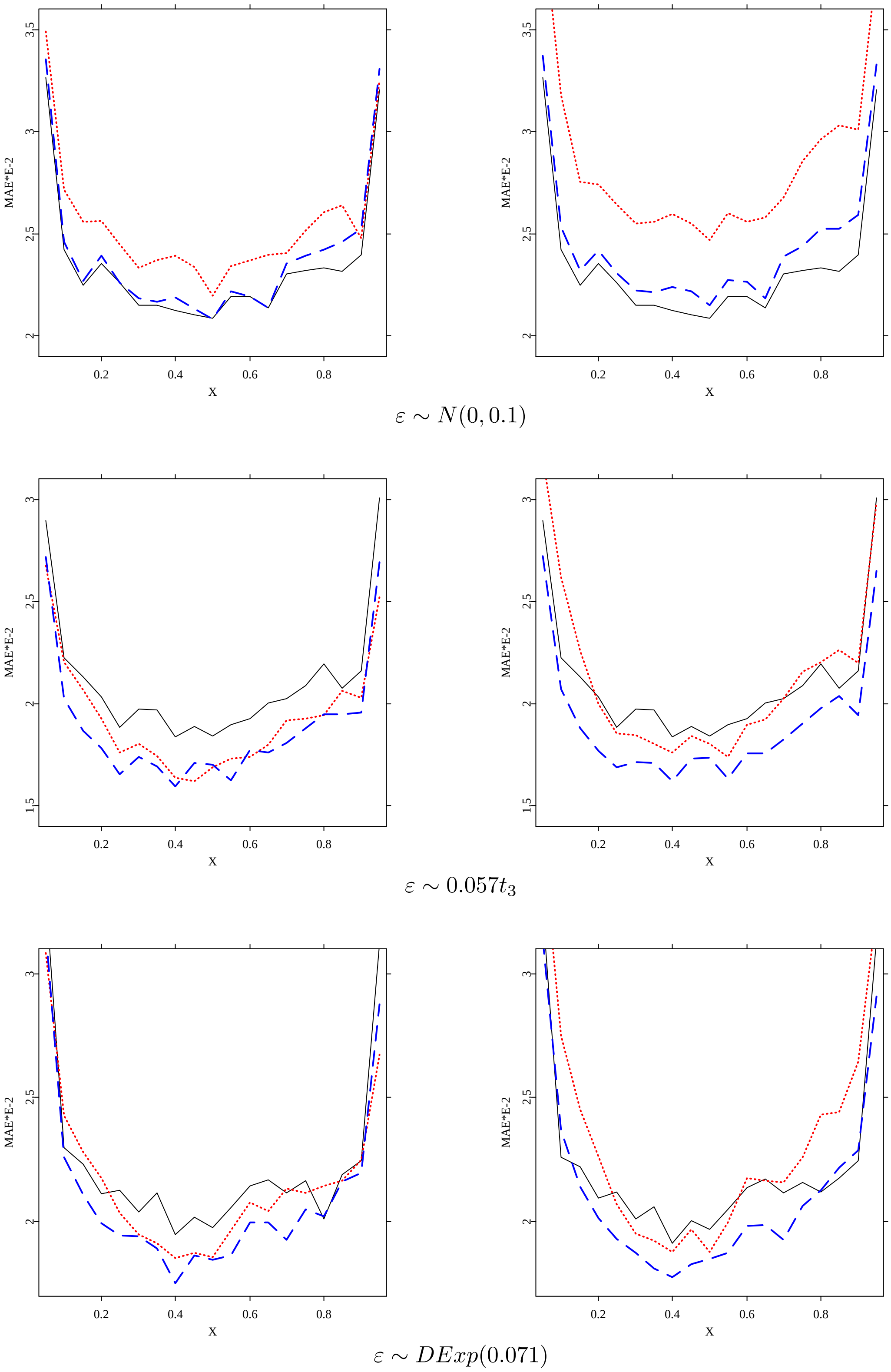


Figure 3. The mean absolute errors of the Nadaraya-Watson (solid line), empirical L-estimator (dotted line), and smoothed L-estimator (dashed line) for sample size $n = 100$, the L-score J_α, and various error distributions. The left panels contain estimators for $\alpha = 0.25$ and the right panels contain estimators for $\alpha = 0.50$.

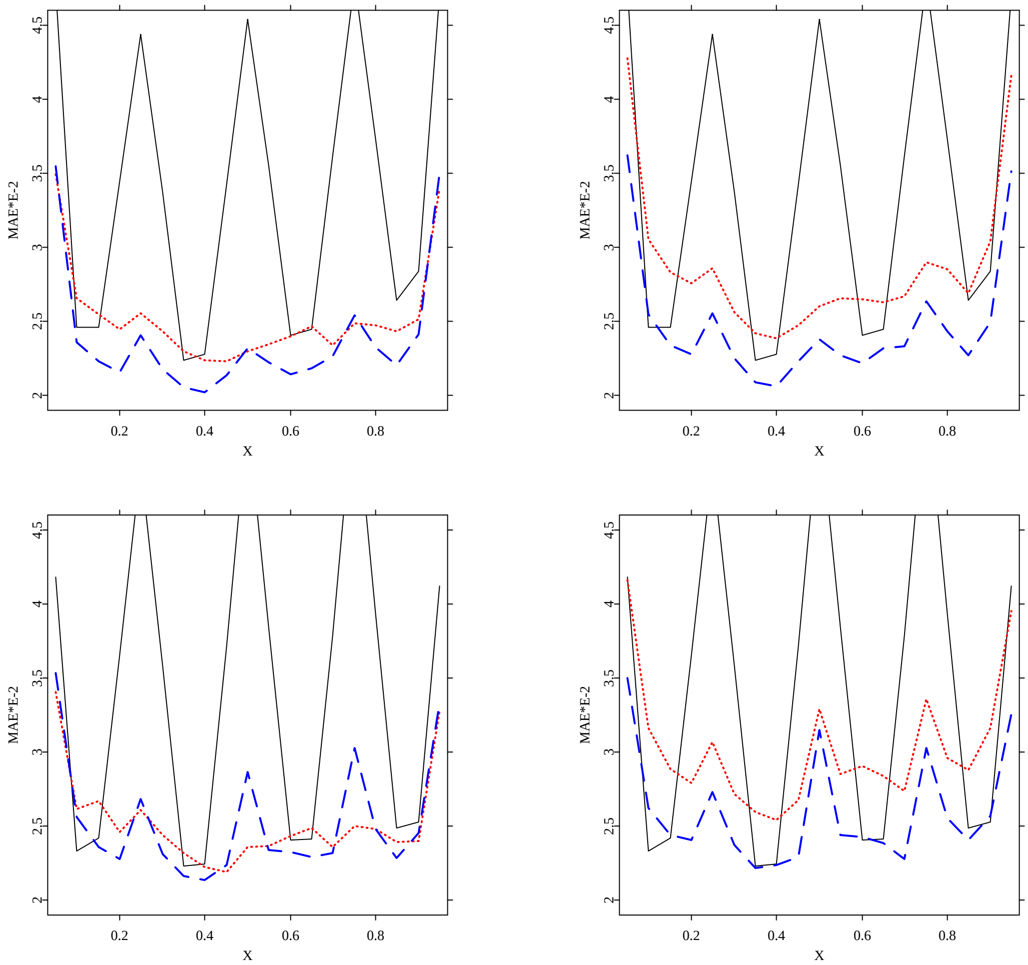


Figure 4. The mean absolute errors of the Nadaraya-Watson (solid line), empirical L-estimator (dotted line), and smoothed L-estimator (dashed line) for sample size $n = 100$ for the L-score J_α under contamination by 3 outliers The left panels contain estimators for $\alpha = 0.25$ and the right panels contain estimators for $\alpha = 0.50$. The upper panels corresponds to CV1 and the bottom panels to CV2.

All presented simulations support two important observations. First, the smoothed L-estimator performs in all simulations better than the empirical one, although the difference does not have to be very large in some cases. Second, the performance of the smoothed L-estimator seems to be robust with respect to the error distribution and its contamination in the sense that the corresponding mean absolute errors are more or less similar under all tested distribution models. Additionally, since the smoothed L-estimator performs as well as the least-squares based NW under Gaussian errors, using the smoothed L-estimator estimator seems to be a good strategy for nonparametric regression estimation even if the presence of outliers in the data or heavier-tailed error distribution are only hypothetical. The only advantage of the classical NW estimator lies in the speed of computation, which the discussed L-estimators cannot match.

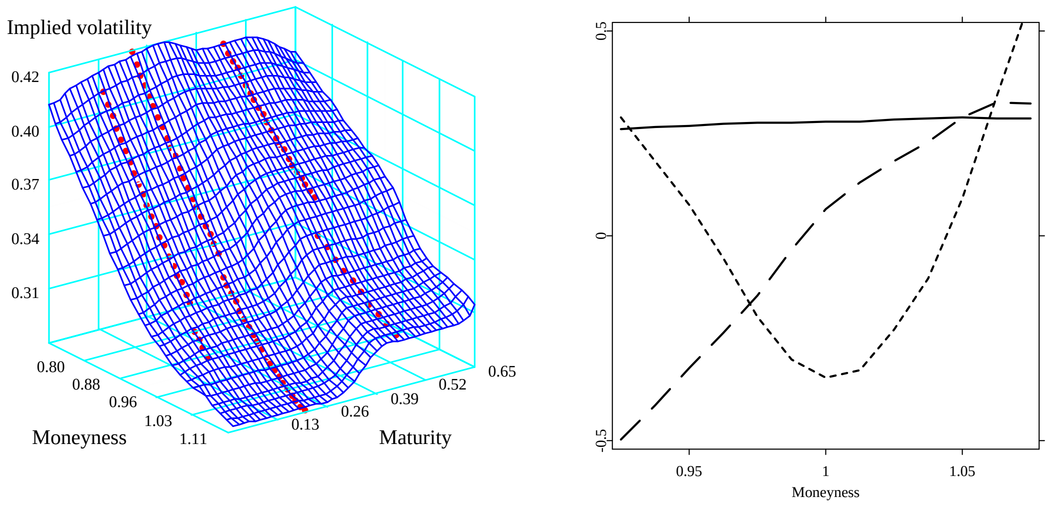


Figure 5. Example of an implied volatility surface of the DAX index from January 4, 1999, (left panel) and the first three common principle components of DAX implied volatility surfaces (March 2001, to May, 2001).

6 Application

An important application of robust smoothing, such as the local L-estimators, could be in the analysis of volatility, that is, the conditional variance of financial time series (e.g., exchange rates or stock prices). One of approaches, the so-called implied volatility (IV) approach, derives the implied volatility of a stock or a stock index from the currently observed prices of related financial instruments (options) on the market. Having data on individual transactions concerning options, one can estimate IV as a function of their maturity (time to expiry) and their strike price, which results in the so-called implied volatility surface (see Härdle, Kleinow, and Stahl, 2002, Chapter 6). The series of such surfaces are then usually analyzed in time to reveal the dynamics of IV and to predict future development (e.g., Fengler, Härdle, and Villa, 2003; Kim and Kim, 2003). Thus, the nonparametric estimators of IV surfaces form an important input of further statistical analysis.

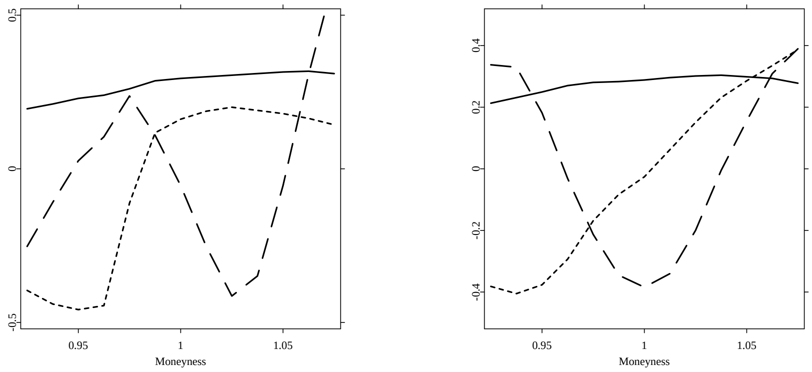
An IV surface on a specified grid is typically estimated by a nonparametric smoothing method; often, the NW estimator is employed (Aït-Sahalia and Lo, 1998, 2000). An example of an IV surface estimated by NW is on Figure 5. Such estimation could greatly benefit from using a robust nonparametric estimator such as the smoothed L-estimator. There are several reasons: data used to estimate IV are often scarce or of low quality for some combinations of maturities and strike prices (data concern individual transactions on the market); the structure of data often enforces oversmoothing (see Fengler, Härdle, and Villa, 2002), which further increases the impact area of wrong observations (outliers); and transaction data are difficult to check and clear, especially because of their size.

To demonstrate the impact of replacing the NW estimator by the smoothed L-estimator, we estimate the IV surfaces of DAX (a German stock index) from January 1999 to May 2001 both by NW and the smoothed L-estimator with the L-score $J_\alpha, \alpha = 0.25$ (the quartic kernel is used in both cases). Instead of plainly comparing them as in Čížek (2004), we use them in further analysis. More specifically, to reduce the dimension of IV surfaces, several authors analyzed the IV surfaces cut at several maturities as a set of vectors (functions) and tried to find their common principle components (CPC); see Fengler, Härdle, and Villa (2003) and Benko and Härdle (2005), for instance. The shapes of the first three CPC are well established and correspond to the main stylized facts concerning IV; see the right panel of Figure 5. These shapes represent level (solid flat line), skew (dashed sloped line), and turn (U-shaped finely dashed line) of the IV surface.

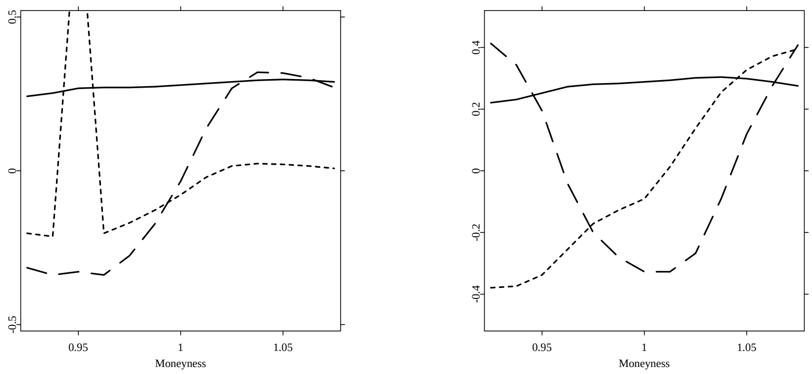
Now, we replicate the analysis of Fengler, Härdle, and Villa (2003) and estimate the CPC of the IV surfaces produced by the NW and smoothed L-estimator for different periods in years 1999 and 2000. Results based on both nonparametric estimator are presented in Figure 6. The three estimated CPC based the IV surfaces obtained by the smoothed L-estimator are stable: they are similar irrespective of the time period considered. On the other hand, the estimated CPC based on the NW estimation are somewhat similar to those on Figure 5 and those based on the L-smoothing when the whole period of two years is considered, but greatly differ for various one-year subperiods. This documents a high sensitivity of NW to various data errors.

7 Conclusion

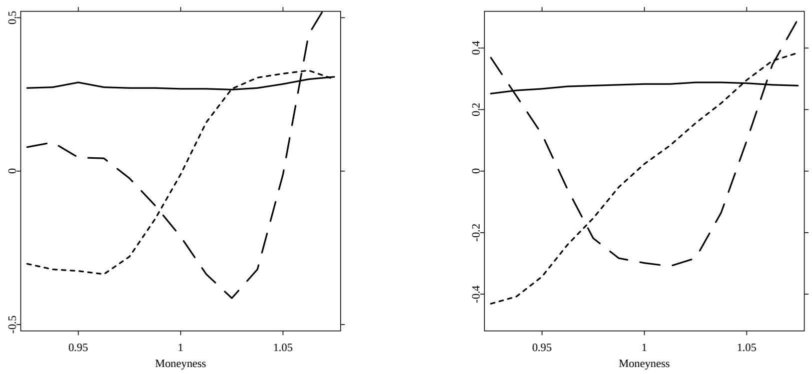
Our theoretical and empirical results clearly point out the superiority of the smothed L-estimator over the empirical L-estimator both for computational and finite-sample properties. Although one might argue that, with an increasing sample size, the difference between the two estimators disappears as suggested by the asymptotic results, it is necessary to keep in mind that this is implied by the first-order asymptotics. Additionally, the computational burden of using the empirical L-estimator becomes very pronounced in large samples. This is where one can benefit from the lower computational demands of the smoothed L-estimator, growing much slower with



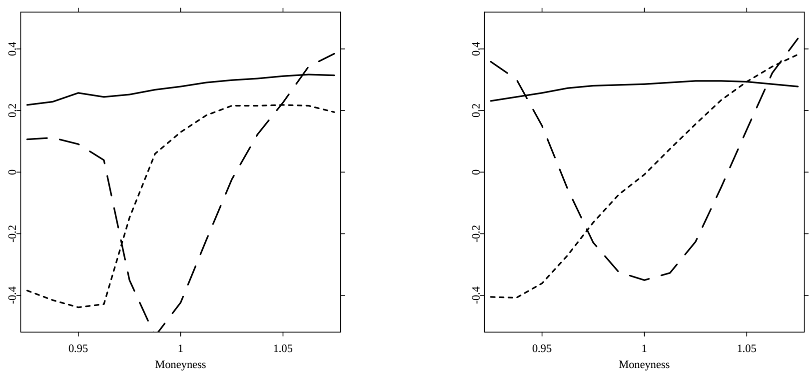
January 1999 – December 1999



July 1999 – June 2000



January 2000 – December 2000



January 1999 – December 2000

Figure 6. The first three common principle components of implied volatility surfaces for the DAX index estimated by the Nadaraya-Watson (left panels) and smoothed local L-estimators (right panels) in different (overlapping) time periods.

an increasing sample size than in the case of the empirical L-estimator. The use of the smoothed L-estimator is thus indicated in all cases.

Appendix A: Proofs

A.1 *Proof of Lemma 3.1*

Let $\hat{f}(\boldsymbol{x}, y)$ denote the Parzen-Rosenblatt density estimator of the density $f(\boldsymbol{x}, y)$,

$$\hat{f}(\boldsymbol{x}, v) = \frac{1}{n} \sum_{i=1}^{n} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i) K_{h_y}(y - Y_i), \tag{A1}$$

and $\hat{f}(\boldsymbol{x})$ denote the Parzen-Rosenblatt density estimator of the density $f(\boldsymbol{x})$,

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i), \tag{A2}$$

where $K_{h_{\boldsymbol{x}}}(\boldsymbol{x}) = \prod_{i=1}^{d} \frac{1}{h_{\boldsymbol{x}}} K_x\left(\frac{x_i}{h_{\boldsymbol{x}}}\right)$ and $K_{h_y}(y) = \frac{1}{h_y} K\left(\frac{y}{h_y}\right)$. Given Assumption A4, $\hat{F}_{\boldsymbol{x}}(y)$ can be written as $\hat{F}_{\boldsymbol{x}}(y) = \int_{-\infty}^{y} \hat{f}(\boldsymbol{x}, v)/\hat{f}(\boldsymbol{x}) dv$. Then

$$\begin{aligned}\left|\hat{F}_{\boldsymbol{x}}(y) - F_{\boldsymbol{x}}(y)\right| &= \left| \int_{-\infty}^{y} \frac{\hat{f}(\boldsymbol{x}, v)}{\hat{f}(\boldsymbol{x})} dv - \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{f(\boldsymbol{x})} dv \right| \\ &\leq \left| \frac{1}{\hat{f}(\boldsymbol{x})} \right| \left| \int_{-\infty}^{y} \hat{f}(\boldsymbol{x}, v) dv - \int_{-\infty}^{y} f(\boldsymbol{x}, v) dv \right| + \left| \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{\hat{f}(\boldsymbol{x})} dv - \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{f(\boldsymbol{x})} dv \right|\end{aligned}$$

For n large enough, we have almost surely $\inf_{\boldsymbol{x}} \left|\hat{f}(\boldsymbol{x})\right| \geq \frac{b}{2}$ (which comes from the almost sure uniform convergence of $\hat{f}(\boldsymbol{x})$ towards $f(\boldsymbol{x})$ under Assumptions A1 to A6 and $\inf_{\boldsymbol{x}} |f(\boldsymbol{x})| \geq b$, Assumption A3). Hence, we obtain first

$$\sup_{\boldsymbol{x}, y} \left| \frac{1}{\hat{f}(\boldsymbol{x})} \right| \left| \int_{-\infty}^{y} \hat{f}(\boldsymbol{x}, v) dv - \int_{-\infty}^{y} f(\boldsymbol{x}, v) dv \right| \leq 2b^{-1} \sup_{\boldsymbol{x}, y} \left| \int_{-\infty}^{y} \hat{f}(\boldsymbol{x}, v) dv - \int_{-\infty}^{y} f(\boldsymbol{x}, v) dv \right|.$$

Using the expression

$$\sup_{\boldsymbol{x}, y} \left| \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{\hat{f}(\boldsymbol{x})} dv - \int_{-\infty}^{y} \frac{f(\boldsymbol{x}, v)}{f(\boldsymbol{x})} dv \right| \leq \sup_{\boldsymbol{x}, y} \left\{ \left| \frac{f(\boldsymbol{x}) - \hat{f}(\boldsymbol{x})}{\hat{f}(\boldsymbol{x}) f(\boldsymbol{x})} \right| \left| \int_{-\infty}^{y} f(\boldsymbol{x}, v) dv \right| \right\}$$

and Assumption A3, we also obtain

$$\sup_{\boldsymbol{x},y}\left|\int_{-\infty}^{y}\frac{f(\boldsymbol{x},v)}{\hat{f}(\boldsymbol{x})}dv-\int_{-\infty}^{y}\frac{f(\boldsymbol{x},v)}{f(\boldsymbol{x})}dv\right|\leq 2b^{-2}\sup_{\boldsymbol{x},v}|f(\boldsymbol{x},v)|\sup_{\boldsymbol{x}}\left|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})\right|.$$

Combining these two results, we can finally write

$$\begin{aligned}\sup_{\boldsymbol{x},y}\left|\hat{F}_{\boldsymbol{x}}(y)-F_{\boldsymbol{x}}(y)\right|\leq 2b^{-1}\sup_{\boldsymbol{x},y}&\left|\int_{-\infty}^{y}\hat{f}(\boldsymbol{x},v)dv-\int_{-\infty}^{y}f(\boldsymbol{x},v)dv\right|\\&+2b^{-2}\sup_{\boldsymbol{x},v}|f(\boldsymbol{x},v)|\sup_{\boldsymbol{x}}\left|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})\right|.\end{aligned}$$

From Aït-Sahalia (1995), we have under Assumptions A1 to A6

$$\sup_{\boldsymbol{x},y}\left|\int_{-\infty}^{y}\hat{f}(\boldsymbol{x},v)dv-\int_{-\infty}^{y}f(\boldsymbol{x},v)dv\right|=O_p\left(n^{-\frac{1}{2}}h_{\boldsymbol{x}}^{-d}+(\max\{h_{\boldsymbol{x}},h_y\})^r\right)$$

and

$$\sup_{\boldsymbol{x}}\left|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})\right|=O_p\left[n^{-\frac{1}{2}}h_{\boldsymbol{x}}^{-d}+h_{\boldsymbol{x}}^{r}\right]=O_p\left[n^{-\frac{1}{2}}h_{\boldsymbol{x}}^{-d}+(\max\{h_{\boldsymbol{x}},h_y\})^r\right]$$

so that $\sup_{\boldsymbol{x},y}\left|\hat{F}_{\boldsymbol{x}}(y)-F_{\boldsymbol{x}}(y)\right|=O_p\left[n^{-\frac{1}{2}}h_{\boldsymbol{x}}^{-d}+(\max\{h_{\boldsymbol{x}},h_y\})^r\right]$.

A.2 *Proof of Theorem 3.2*

Let us denote by T the functional $T(F_{\boldsymbol{x}})=\int_{-\infty}^{+\infty}yJ\{F_{\boldsymbol{x}}(y)\}dF_{\boldsymbol{x}}(y)$ and by τ the function $\tau(t)=T(F_{\boldsymbol{x}}+tH_{\boldsymbol{x}})$, where $H_{\boldsymbol{x}}:R\to R$ is a continuously differentiable function with a compactly supported derivative and satisfying $\sup_{\boldsymbol{x},y}|H_{\boldsymbol{x}}(y)|<\infty$.

It follows that

$$\tau'(t)=\int_{-\infty}^{+\infty}yH_{\boldsymbol{x}}(y)J'\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}d\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}+\int_{-\infty}^{+\infty}yJ\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}dH_{\boldsymbol{x}}(y).$$

An integration by parts gives us

$$\begin{aligned}\int_{-\infty}^{+\infty}yH_{\boldsymbol{x}}(y)J'\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}d\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}&=[yH_{\boldsymbol{x}}(y)J\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}]_{+\infty}^{-\infty}\\&-\int_{-\infty}^{+\infty}H_{\boldsymbol{x}}(y)J\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}dy-\int_{-\infty}^{+\infty}yJ\{F_{\boldsymbol{x}}(y)+tH_{\boldsymbol{x}}(y)\}dH_{\boldsymbol{x}}(y)\end{aligned}$$

so that

$$\tau'(t) = -\int_{-\infty}^{+\infty} H_{\boldsymbol{x}}(y) J\left\{F_{\boldsymbol{x}}(y) + tH_{\boldsymbol{x}}(y)\right\} dy.$$

In particular for $t = 0$, we obtain $\tau'(0) = -\int_{-\infty}^{+\infty} H_{\boldsymbol{x}}(y) J\{F_{\boldsymbol{x}}(y)\} dy$. The second derivative of τ is then

$$\tau''(t) = -\int_{-\infty}^{+\infty} H_{\boldsymbol{x}}^2(y) J'\left\{F_{\boldsymbol{x}}(y) + tH_{\boldsymbol{x}}(y)\right\} dy.$$

Hence, Assumption A7 implies $|\tau''(t)| = O\left(\sup_{\boldsymbol{x},y} |H_{\boldsymbol{x}}(y)|^2\right)$ for all $t \in \langle 0, 1\rangle$. Now, the Taylor expansion of τ between 0 and 1 gives us

$$T(F_{\boldsymbol{x}} + H_{\boldsymbol{x}}) = T(F_{\boldsymbol{x}}) - \int_{-\infty}^{+\infty} H_{\boldsymbol{x}}(y) J\{F_{\boldsymbol{x}}(y)\} dy + O\left(\sup_{\boldsymbol{x},y} |H_{\boldsymbol{x}}(y)|^2\right) \tag{A3}$$

Taking $H_{\boldsymbol{x}}(y) = \hat{F}_{\boldsymbol{x}}(y) - F_{\boldsymbol{x}}(y)$ in expression (A3) and using Lemma 3.1, we obtain

$$\hat{m}_L(\boldsymbol{x}) - m_L(\boldsymbol{x}) = -\int_{-\infty}^{+\infty} \left[\hat{F}_{\boldsymbol{x}}(y) - F_{\boldsymbol{x}}(y)\right] J\left\{F_{\boldsymbol{x}}(y)\right\} dy + O_p\left[n^{-1} h_{\boldsymbol{x}}^{-2d} + \left(\max\left\{h_{\boldsymbol{x}}, h_y\right\}\right)^{2r}\right]$$

Let us study the leading order term

$$\begin{aligned} L_n &= -\int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{y} \frac{\hat{f}(\boldsymbol{x}, v)}{\hat{f}(\boldsymbol{x})} - \frac{f(\boldsymbol{x}, v)}{f(\boldsymbol{x})} dv \right\} J\{F_{\boldsymbol{x}}(y)\} dy \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{y} \frac{\left\{\hat{f}(\boldsymbol{x}, v) - f(\boldsymbol{x}, v)\right\} f(v) - f(\boldsymbol{x}, v)\left\{\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right\}}{\hat{f}(\boldsymbol{x}) f(\boldsymbol{x})} dv \right] J\{F_{\boldsymbol{x}}(y)\} dy. \end{aligned}$$

Using $\sup_x \left|\frac{\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})}{\hat{f}(\boldsymbol{x})}\right| = o_p(1)$ (which holds under Assumptions A1 to A6, see the proof of Lemma 3.1), we get

$$\begin{aligned} L_n = &\left\{ \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{y} \frac{\left\{\hat{f}(\boldsymbol{x}, v) - f(\boldsymbol{x}, v)\right\} f(\boldsymbol{x}) - f(\boldsymbol{x}, v)\left\{\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right\}}{f^2(\boldsymbol{x})} dv \right] J\{F_{\boldsymbol{x}}(y)\} dy \right\} \times \\ &\times\{1 + o_p(1)\} = \tilde{L}_n\{1 + o_p(1)\} \end{aligned}$$

Thus by the Slutsky lemma, we only have to study the asymptotic distribution of $\tilde{L}_n$. Using expressions (A1) and (A2) for $\hat{f}(\boldsymbol{x}, v)$ and $\hat{f}(\boldsymbol{x})$, we obtain

$$\tilde{L}_n = \frac{1}{n} \sum_{i=1}^{n} \int \left\{ \int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left[\begin{array}{c} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i) \int\limits_{-\infty}^{y} K_{h_y}(v - Y_i) dv - \\ - \int\limits_{-\infty}^{y} f(\boldsymbol{x}, v) dv - F_{\boldsymbol{x}}(y) \left\{ K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i) - f(\boldsymbol{x}) \right\} \end{array} \right] dy \right\}.$$

Next, we are going to split the study of $\tilde{L}_n$ into analyzing a deterministic 'bias' term $\tilde{B}_n = E(\tilde{L}_n)$ and a stochastic 'variance' term $\tilde{V}_n = \tilde{L}_n - E(\tilde{L}_n)$, where

$$\tilde{B}_n = \int \left\{ \int_{-\infty}^{+\infty} \left[\frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{u}) \left\{ \int_{-\infty}^{y} K_{h_y}(v - w) dv - F_{\boldsymbol{x}}(y) \right\} \right] dy \right\} f(\boldsymbol{u}, w) d\boldsymbol{u} dw$$
$$- \int_{-\infty}^{+\infty} \left[\frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left\{ \int_{-\infty}^{y} f(\boldsymbol{x}, v) dv - F_{\boldsymbol{x}}(y) f(\boldsymbol{x}) \right\} \right] dy = \tilde{B}_{1n} - \tilde{B}_{2n}$$

and $\tilde{B}_{2n}$ represents the latter, constant term. Changing variable $\boldsymbol{\xi} = (\boldsymbol{x} - \boldsymbol{u})/h_{\boldsymbol{x}}$ in the integration with respect to $\boldsymbol{u}$ and $\varsigma = (y - w)/h_y$ in the integration with respect to w, we obtain for the non-constant term $\tilde{B}_{1n}$

$$\tilde{B}_{1n} = \int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left[\int_{-\infty}^{y} \int \int K_{\boldsymbol{x}}(\boldsymbol{\xi}) \left\{ K(\varsigma) - F_{\boldsymbol{x}}(y) \right\} f(\boldsymbol{x} - h_{\boldsymbol{x}} \boldsymbol{\xi}, v - h\varsigma) d\boldsymbol{\xi} d\varsigma dv \right] dy.$$

Using the Taylor expansion of $f(\boldsymbol{x} - h_{\boldsymbol{x}}\boldsymbol{\xi}, v - h_y\varsigma)$ at $f(\boldsymbol{x}, v)$ up to order r (see Assumptions A2, A4, and A5), we obtain that $\tilde{B}_n = \tilde{B}_{1n} - \tilde{B}_{2n} = O\left[(\max\{h_{\boldsymbol{x}}, h_y\})^r\right]$. Finally, under Assumption A6, we have $n^{1/2} h_{\boldsymbol{x}}^{d/2} \tilde{B}_n = o_p(1)$.

We now have to study the 'variance' term normalized at rate $n^{1/2} h_{\boldsymbol{x}}^{d/2}$

$$n^{1/2} h_{\boldsymbol{x}}^{d/2} \tilde{V}_n = \sum_{i=1}^{n} \sqrt{\frac{h_{\boldsymbol{x}}^d}{n}} \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i) \left\{ \int_{-\infty}^{y} K_{h_y}(v - Y_i) dv - F_{\boldsymbol{x}}(y) \right\} dy - \tilde{B}_n \right]$$

Let us define the σ-field $\mathcal{F}_{n,i}$ generated by $\{\boldsymbol{X}_j, Y_j\}_{j=1,\dots i}$ and let us define the random variable

$$Z_{n,i} = \sqrt{\frac{h_{\boldsymbol{x}}^d}{n}} \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{X}_i) \left\{ \int_{-\infty}^{y} K_{h_y}(v - Y_i) dv - F_{\boldsymbol{x}}(y) \right\} dy - \tilde{B}_n \right]$$

so that $n^{1/2} h_{\boldsymbol{x}}^{d/2} \tilde{V}_n = \sum_{j=1}^{n} Z_{n,j}$. The array $\left\{ \sum_{j=1}^{i} Z_{n,j}, \mathcal{F}_{n,i}; 1 \le i \le n, n \ge 1 \right\}$ is a zero-mean, square integrable martingale array. For $k \leq n$, we have

$$\sum_{j=1}^{k} E(Z_{n,j}^2) = h_{\boldsymbol{x}}^d \int\int \left\{ \int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} K_{h_{\boldsymbol{x}}}(\boldsymbol{x}-\boldsymbol{u}) \left[\int_{-\infty}^{y} K_{h_y}(v-w)dv - F_{\boldsymbol{x}}(y) \right] dy \right\}^2 \times$$
$$\times f(\boldsymbol{u}, w) d\boldsymbol{u} dw - h_{\boldsymbol{x}}^d \tilde{B}_n^2.$$

With the change of variable $\boldsymbol{\xi} = (\boldsymbol{x}-\boldsymbol{u})/h_{\boldsymbol{x}}$ in the integration with respect to $\boldsymbol{u}$, we obtain

$$\sum_{j=1}^{k} E(Z_{n,j}^2) = \int \left[\begin{array}{c} \left\{ \int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left[\int_{-\infty}^{y} K_{h_y}(v-w)dv - F_{\boldsymbol{x}}(y) \right] dy \right\}^2 \times \\ \times \int K_{h_{\boldsymbol{x}}}^2(\boldsymbol{\xi}) f(\boldsymbol{\xi} - h_{\boldsymbol{x}}\boldsymbol{\xi}, w) d\boldsymbol{\xi} \end{array} \right] dw - h_{\boldsymbol{x}}^d \tilde{B}_n^2$$

Obviously, $\lim_{n\to\infty} \int_{-\infty}^{y} K_{h_y}(v-w)dv = 1$ if $y \geq w$ and 0 otherwise. Recalling that $\tilde{B}_n = O\left[(\max\{h_{\boldsymbol{x}}, h_y\})^r\right]$, the Taylor expansion of $f(\boldsymbol{x} - h_{\boldsymbol{x}}\boldsymbol{\xi}, w)$ implies that the only term that is not asymptotically negligible in probability, $n \to \infty$, is

$$\lim_{n\to\infty} E\left(\sum_{j=1}^{n} Z_{n,j}^2\right) = \int \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left\{ I_{(-\infty,y\rangle}(w) - F_{\boldsymbol{x}}(y) \right\} dy \right]^2 f(\boldsymbol{x}, w) dw \cdot \int K_{h_{\boldsymbol{x}}}^2(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

so that

$$\sum_{j=1}^{n} E\left(Z_{n,j}^2 | \mathcal{F}_{n,j-1}\right) \xrightarrow{P} \int \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left\{ I_{(-\infty,y\rangle}(w) - F_{\boldsymbol{x}}(y) \right\} dy \right]^2 f(\boldsymbol{x}, w) dw \cdot \int K_{h_{\boldsymbol{x}}}^2(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

Under Assumptions A1 to A4, conditions 3.19 and 3.20 of Corollary 3.1 of Hall and Heyde (1981, pp. 58) are satisfied, so that we obtain $n^{1/2} h_{\boldsymbol{x}}^{d/2} \tilde{V}_n \to \mathcal{N}(0, V)$, where

$$V = \int \left[\int_{-\infty}^{+\infty} \frac{J\{F_{\boldsymbol{x}}(y)\}}{f(\boldsymbol{x})} \left\{ I_{(-\infty,y\rangle}(w) - F_{\boldsymbol{x}}(y) \right\} dy \right]^2 f(\boldsymbol{x}, w) dw \cdot \int K_{h_{\boldsymbol{x}}}^2(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

References

[1] Y. Aït-Sahalia, 1995, The delta method for nonlinear kernel functionals. Ph. D. dissertation, University of Chicago, USA.

[2] Y. Aït-Sahalia and A. W. Lo, 1998, Nonparametric estimation of state-price densities implicit in financial assets. *Journal of Finance*, **53**, 499–547.

[3] Y. Aït-Sahalia and A. W. Lo, 2000, Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, **94**, 9–51.

[4] V. Barnett and T. Lewis, 1994, *Outliers in Statistical Data*. Wiley, New York.

[5] M. Benko and W. Härdle, 2005, Common functional implied volatility analysis. In: P. Čížek, W. Härdle, and R. Weron (Eds.) *Statistical Tools for Finance and Insurance* (Springer, Heidelberg), 115–134.

[6] G. Boente and R. Fraiman, 1994, Local L-estimators for nonparametric regression under dependence. *Journal of Nonparametric Statistics*, **4**, 91–101.

[7] P. Čížek, 2004, Smoothed local L-estimation with an application. In: M. Hubert, G. Pison, A. Struyf, and S. van Aelst (Eds.), *Theory and Applications of Recent Robust Methods* (Birkhäuser, Basel), 59–70.

[8] P. Čížek and W. Härdle, 2006, Robust estimation of dimension reduction space. *Computational Statistics & Data Analysis*, in press.

[9] J. Davidson, 1994, *Stochastic Limit Theory.* Oxford University Press, New York.

[10] P. J. Davis and P. Rabinowitz, 1984, *Methods of Numerical Integration.* Academic Press, New York.

[11] M. Fengler, W. Härdle, and P. Villa, 2003, The dynamics of implied volatilities: a common principle components approach. *Review of Derivative Research*, **6**, 179–202.

[12] L. T. Fernholz, 1997, Reducing the variance by smoothing. *Journal of Statistical Planning and Inference*, **57**, 29–38.

[13] P. Hall and C. C. Heyde, 1981, *Martingale Limit Theory and its Applications.* Academic Press, New York.

[14] W. Härdle, 1990, *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

[15] W. Härdle, T. Kleinow, and G. Stahl, 2002, *Applied Quantitative Finance.* Springer, Heidelberg.

[16] Lucas, A., 1995, An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics*, **66**, 153–173.

[17] D. Leung, 2005, Cross-validation in nonparametric regression with outliers. *The Annals of Statistics*, **33**, 2291–2310.

[18] D. Leung, F. Marriott, and E. Wu, 1993, Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, **2**, 333–339.

[19] E. A. Nadaraya, 1964, On estimating regression. *Theory. Prob. App*, **10**, 186–190.

[20] A. Pagan and A. Ullah, 1999, *Nonparametric Econometrics.* Cambridge University Press, Cambridge.

[21] B. U. Park and J. S. Marron, 1990, Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, **85**, 66–72.

[22] S. Sakata and H. White, 1998, High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica*, **66**, 529–567.

[23] F. Wang and D. Scott, 1994, The L_1 method for robust nonparametric regression. *Journal of the American Statistical Association*, **89**, 65–76.

[24] G. S. Watson, 1964, Smooth regression analysis. *Sankhya*, Ser. A, **26**, 359-372.