SOME EXTENSIONS OF NONPARAMETRIC REGRESSION MODELS

Zdeněk Hlávka

HABILITATION THESIS
Mathematics—Probability and Mathematical Statistics

Charles University in Prague, Faculty of Mathematics and Physics

October 2011

Preface

The main topic of this habilitation thesis are nonparametric regression methods, briefly introduced in Chapter 1. The most important author's results related to nonparametric regression may be found in the papers attached in the Appendix. Apart of these attached papers, the same results are summarized in a unified manner in Chapters 2–4 where we also propose some extensions that were necessary in order to unite these papers smoothly into a single publication.

The main body of this habilitation thesis is divided into three chapters corresponding, respectively, to the three most important stages of a data-analytical process:

1. design of the experiment (Chapter 2),

2. estimation (Chapter 3),

3. verification of assumptions (Chapter 4).

The Appendix contains five papers written by the author and various coauthors during the years 2006–2011. These papers discuss successively the three stages of the data-analytical process and represent most important author's results related to nonparametric regression models during this period:

- The first paper, **Hlávka (2011): On nonparametric estimators of location of maximum**, concerns the planning of a nonparametric regression experiment and it proposes an experimental design that is optimal for the estimation of a location of maximum of an unknown regression function. In Chapter 2, we extend this result also to estimation of zeros under more general assumptions.

- The second paper, **Härdle and Hlávka (2009): Dynamics of state price densities**, concerns a complex application in option pricing, where the nonparametric regression estimator must be adapted in order to account for correlated data and constrained regression function. The third and the fourth paper, **Hlávka (2006a): Fast algorithm for nonparametric arbitrage-free SPD estimation** and **Hlávka and Svojík (2009): Application of extended Kalman filter to SPD estimation**, discuss some computational aspects of the proposed estimator. The contents of these three papers is summarized in Chapter 3.

- Finally, the fifth paper, **Hlávka et al. (2011): Tests for independence in non-parametric heteroscedastic regression models**, proposes a new approach to a verification of assumptions of a nonparametric regression model. In Chapter 4, this approach is modified, compared to other tests proposed in the literature, and used to test the validity of assumptions of the constrained nonparametric regression model described in Chapter 3.

Prague, October 2011 *Zdeněk Hlávka*

Contents

Chapter 1

Introduction

Regression analysis is certainly one of the most useful and widely applicable statistical techniques. In general, it allows to investigate the dependency of the so-called *dependent variable* on one (or more) *explanatory* or *independent variables*.

Linear model The regression analysis comes in many flavors. In the classical linear model, the symbol Y denotes the dependent variable such that its conditional expectation is a linear function of the observed value x of the explanatory variable or its arbitrary transformation, i.e.,

$$E(Y|X = x) = f(x)^\top \beta,$$

where $f(x) = (f_1(x), \dots, f_p(x))^\top$ is a known function of x, β is a p-dimensional vector of unknown parameters, and $\mathrm{Var}(Y|X = x) = \sigma^2 > 0$. Often, we introduce an artificial random variable ε, the so-called *random error*, in order to write the linear model in the most popular way:

$$Y = f(x)^\top \beta + \varepsilon, \tag{1.1}$$

where the unobservable random error ε is centered and has variance σ^2, i.e., $E\varepsilon = 0$ and $E\varepsilon^2 = \sigma^2$.

In the linear model (1.1), we assume that the functional form of the dependency is known in advance and it remains to estimate only the vector of unknown parameters β. For example, by setting $f(x) = (1, x)^\top$ we obtain a straight line with intercept given by β_0 and slope equal to β_1. Parabolic and cubic function may be obtained by choosing $f(x) = (1, x, x^2)^\top$ or $f(x) = (1, x, x^2, x^3)^\top$, respectively.

An estimator $\hat{\beta}$ of the unknown parameter β is usually calculated from n observations of pairs (Y_i, x_i) satisfying (1.1), i.e.,

$$Y_i = f(x_i)^\top \beta + \varepsilon_i, \quad i = 1, \dots, n, \tag{1.2}$$

assuming that the random errors ε_i, $i = 1, \dots, n$, are independent and identically distributed. The n equations (1.2) are often rewritten in a matrix notation:

$$\mathcal{Y} = \mathcal{X}_F \beta + \boldsymbol{\varepsilon}, \tag{1.3}$$

where $\mathcal{Y} = (Y_1, \dots, Y_n)^\top$ is the response vector, $\mathcal{X}_F$ is the so-called *design matrix* with rows $f(x_i)$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ is a centered random vector with variance matrix $\sigma^2 \mathcal{I}_n$.

Assuming that the design matrix $\mathcal{X}_F$ has full rank, equation (1.3) allows to express the *Least Squares* (LS) estimator of β as:

$$\hat{\beta}^{LS} = (\mathcal{X}_F^\top \mathcal{X}_F)^{-1} \mathcal{X}_F^\top \mathcal{Y}, \tag{1.4}$$

see, e.g., Zvára (2008).

An illustration of a linear, parabolic and cubic function fitted by the Least Squares estimator (1.4) is given in Figure 1.1. Notice that these parametric estimators do not fully capture the shape of the true regression function. In such situation, standard regression diagnostic tools (Belsley et al.; 1980; Zvára; 2008) may be used to diagnose the lack-of-fit but it may not be possible to improve the model if the true shape of the regression function is not known or if the function $f(.)$ is misspecified.
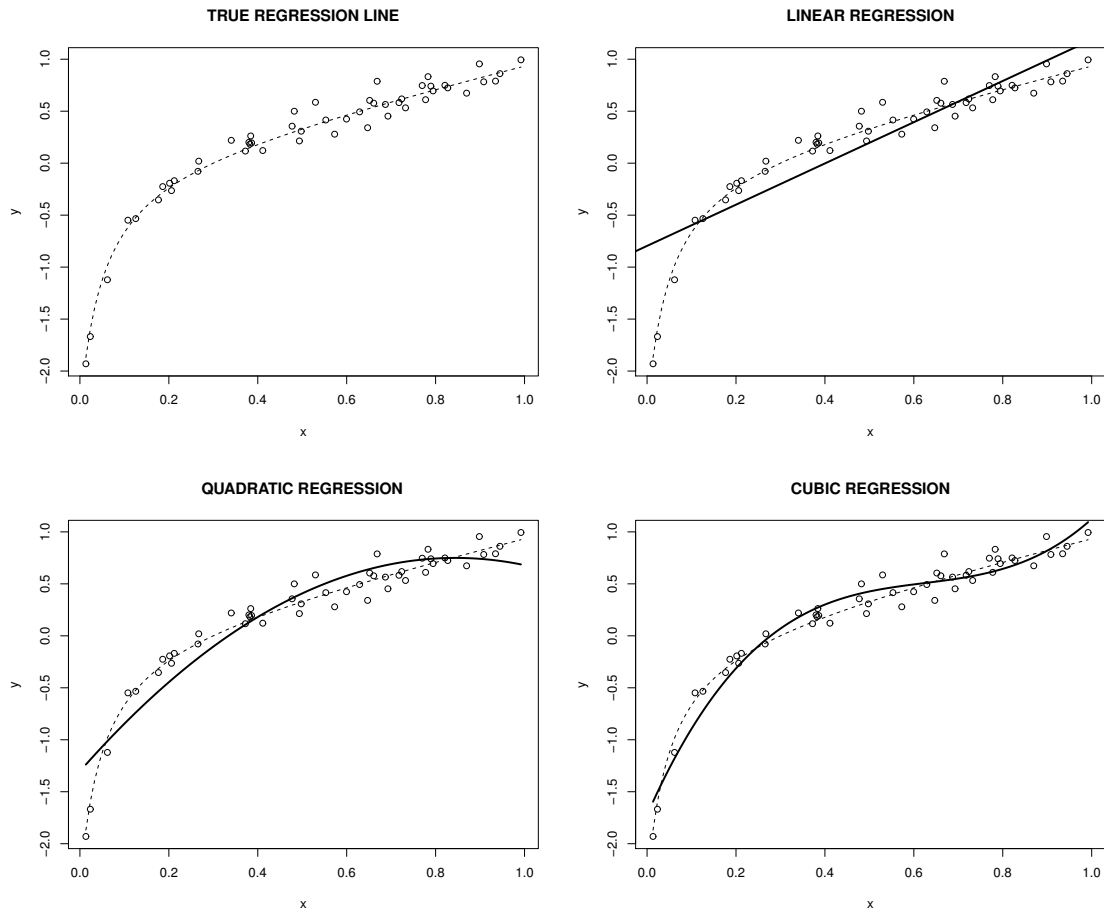
Figure 1.1: Simulated example (50 observations): the dashed line is the true regression function, thick lines denote various parametric linear regression estimators.
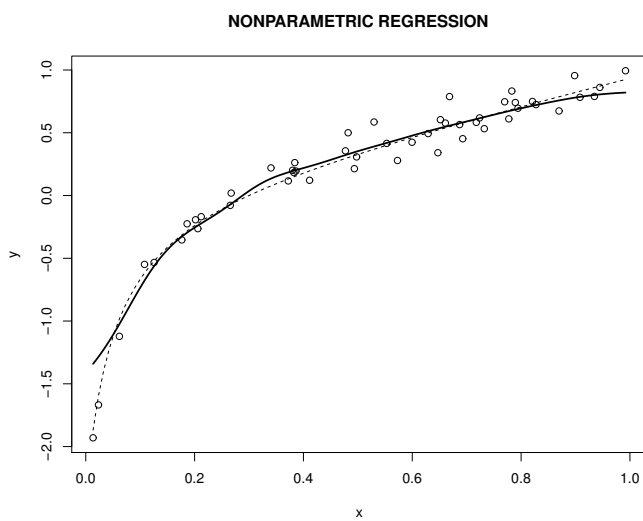

Figure 1.2: Simulated example (50 observations): the dashed line is the true regression function, thick line denotes the nonparametric regression estimator.

Nonparametric regression   Removing the parametric assumptions from (1.1), we obtain the *non-parametric regression model*:

$$Y = E(Y|X = x) + \varepsilon = m(x) + \varepsilon, \tag{1.5}$$

where Y is the response, ε the random error, and $m(.)$ denotes the unknown regression function. An example of a nonparametric regression estimator of $m(.)$ is plotted in Figure 1.2.

Compared to the linear model (1.1), the nonparametric regression model (1.5) is more flexible. On the other hand, the nonparametric regression estimator is not as easily interpretable and it is often used only as a graphical tool. Additionally, one also has to choose an appropriate value of some *smoothing parameters* that typically control the smoothness of the estimator.

An overview of nonparametric regression (or smoothing) methods may be found, e.g., in Härdle (1990); Simonoff (1996); Fan and Gijbels (1996); Härdle et al. (2004). The standard "smoothing" approaches include splines, wavelets, moving averages, running medians, local polynomials, regression trees, neural networks, and other methods. From now on, we concentrate on the kernel method: the kernel regression estimators are defined as locally weighted averages and its properties may be derived quite easily.

1.1   Kernel regression

Let us now concentrate on the kernel approach. In this section, we introduce basic notation and provide a short review of some well-known results concerning the kernel regression estimator based on the nonparametric regression model (1.5).

In practice, it is important to distinguish fixed and random design experiments. In a fixed design experiment, we choose the values of the explanatory variable according to a certain rule, e.g., as a quantiles of certain probability distribution. In a random design experiment, we may control only the probability distribution of the explanatory variable but the observed values are random.

Random design   We assume that model (1.5) holds and that we observe pairs of random variables (X_i, Y_i), $i = 1, \ldots, n$, such that:

$$Y_i = m(X_i) + \varepsilon_i, \tag{1.6}$$

where $E\varepsilon_i = 0$ and $\operatorname{Var}\varepsilon_i = \sigma^2$. The unknown regression function $m(.)$ may be estimated, for example, by using the classical *Nadaraya-Watson estimator* (Nadaraya; 1964; Watson; 1964):

$$m_b^{NW}(x) = \frac{\sum_{i=1}^{n} K_b(x - X_i)Y_i}{\sum_{j=1}^{n} K_b(x - X_j)} = \sum_{i=1}^{n} \frac{K_b(x - X_i)}{\sum_{j=1}^{n} K_b(x - X_j)} Y_i, \tag{1.7}$$

where $K_b(x) = K(x/b)/b$, $K(.)$ is a *kernel function*, and $b > 0$ is a *bandwidth*.

Under some assumptions, it may be shown that the Nadaraya-Watson estimator $m_b^{NW}(x)$ is asymptotically normally distributed (Härdle; 1990, Theorem 4.2.1) with variance depending on σ^2, $\int K^2(u)du$, and $f_X(.)$, the density of the explanatory variable X. The bias of the Nadaraya-Watson estimator is proportional to the second moment of the kernel, i.e., $\int u^2 K^2(u)du$, and a measure of local curvature of the regression function. Interestingly, the variance is a decreasing function and the bias an increasing function of the bandwidth and, therefore, the choice of the optimal bandwidth always involves a certain *bias-variance trade-off*. Some further remarks concerning the choice of the bandwidth and the kernel function may be found at the end of this section.

The Nadaraya-Watson estimator (1.7) may be written as a ratio with a kernel estimator of the probability density $f_X(x)$ in the denominator:

$$m_b^{NW}(x) = \sum_{i=1}^{n} \frac{n^{-1} K_b(x - X_i)}{n^{-1} \sum_{j=1}^{n} K_b(x - X_j)} Y_i = \frac{\sum_{i=1}^{n} n^{-1} K_b(x - X_i)Y_i}{\widehat{f}_X(x)} \tag{1.8}$$

and, clearly, it may be significantly simplified if the density $f_X(.)$ is known.

Fixed design   In some real life applications, we may have the opportunity to choose all values of the explanatory variable in advance. More precisely, in an experiment with *fixed design* given by a

probability density function $f_X(x)$, we observe the response Y_i, $i = 1, \ldots, n$, in fixed design points $x_1 < \cdots < x_n$ such that $\int_{x_i}^{x_{i+1}} f_X(u)du \doteq 1/n$, i.e.,

$$Y_i = m(x_i) + \varepsilon_i. \tag{1.9}$$

Replacing the weights $K_b(x - X_i)/\sum K_b(x - X_j)$ in (1.8) by $\int_{s_{i-1}}^{s_i} K_b(x-u)$, we avoid the estimator $\widehat{f}_X(x)$ in the denominator of (1.8) and obtain the fixed design *Gasser-Müller estimator* (Gasser and Müller; 1984):

$$m_b^{GM}(x) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_b(x - u)du\, Y_i, \tag{1.10}$$

where $s_i = (x_i + x_{i+1})/2$. Notice that, similarly as in (1.8), the sum of the weights is equal to one because $\sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_b(x - u)du = \int K(x)dx = 1$.

The properties of the Gasser-Müller estimator are investigated in Gasser and Müller (1984) and we will summarize it under more general assumptions in Section 2.1.2. Notice that if the design points are uniformly distributed, then $\int_{s_{i-1}}^{s_i} K_b(x - u)du \doteq n^{-1}K_b(x - x_{(i)})$ and the Nadaraya-Watson estimator (1.8) may be interpreted as an approximation of the Gasser-Müller estimator (1.10) (or the other way around).

Similarly as in the random design situation, the bias and the variance of the Gasser-Müller estimator depend on the kernel function and on the bandwidth. In order to choose appropriate values of these tuning parameters, we need a simple statistics (measure of accuracy) that would allow us to compare two or more competing nonparametric regression estimators.

Asymptotic MSE   The most often used measure of accuracy of a nonparametric regression estimator $m_b(x)$ is the *Mean Squared Error* (MSE):

$$\mathrm{MSE}\{m_b(x)\} = E\{m_b(x) - m(x)\}^2.$$

If a random variable X_i is uniformly distributed on $\langle 0, 1 \rangle$ then "under certain assumptions" (Härdle et al.; 2004, Theorem 4.3) we have for the random design Nadaraya-Watson estimator:

$$\mathrm{MSE}\{m_b^{NW}(x)\} \doteq \frac{1}{nb}\sigma^2 \int K^2(s)ds + \frac{1}{4}b^4\{m''(x)\}^2.$$

Interestingly, exactly the same asymptotic result holds, "under certain assumptions" (Härdle et al.; 2004, Theorem 4.2), also for the fixed design Gasser-Müller estimator, i.e.,

$$\mathrm{MSE}\{m_b^{GM}(x)\} \doteq \frac{1}{nb}\sigma^2 \int K^2(s)ds + \frac{1}{4}b^4\{m''(x)\}^2.$$

It follows that the precision of the random and fixed design estimators is very similar. In practice, the type of design is usually implied by the setup of the experiment and it cannot be changed.

Choice of tuning parameters   In order to calculate the nonparametric regression estimator (1.7) or (1.10), we have to choose the kernel function $K(.)$ and the bandwidth parameter b. Both theoretical findings and practical recommendations may be summarized by saying that the choice of the kernel function is much less important than the choice of the bandwidth parameter (Härdle et al.; 2004, Section 4.3). By minimizing the asymptotic MSE, it may be shown that the asymptotically optimal value of the bandwidth parameter is $b \propto n^{-1/5}$. Unfortunately, the optimal value of the bandwidth involves also a constant depending on some unknown parameters, e.g., the second derivative of the unknown regression function and the unknown variance of the unobserved random errors. Therefore, a more practical recommendation is to use the so-called *leave-one-out* (or *cross-validation*) approach. The choice of the bandwidth parameter is crucial and it is thoroughly discussed in the relevant literature, see, e.g., Nadaraya (1989); Härdle (1990); Ruppert and Wand (1994); Wand and Jones (1995); Simonoff (1996); Fan and Gijbels (1996); Efromovich (1999); Härdle et al. (2004); Tsybakov (2009) among many others.
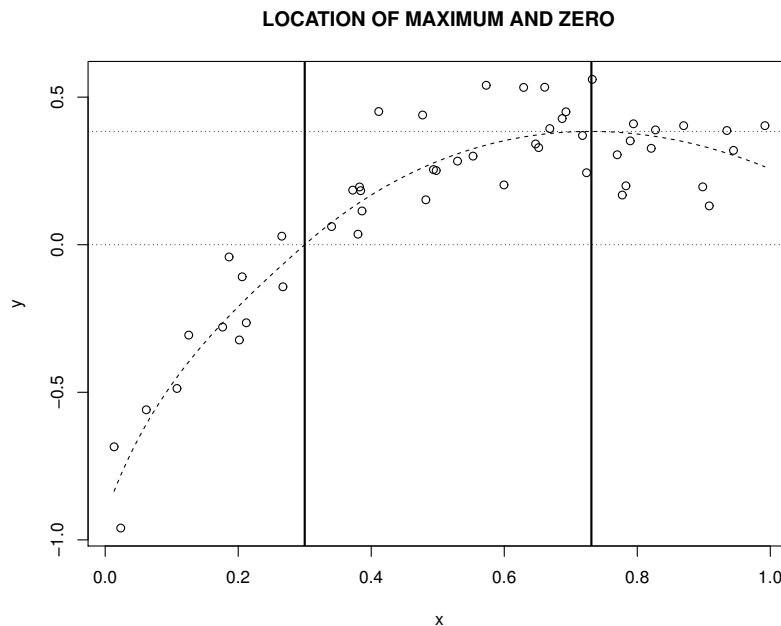
LOCATION OF MAXIMUM AND ZERO

Figure 1.3: Simulated example (50 observations): the dashed line is the true regression function, dotted horizontal lines denote the horizontal axis and the maximum of the regression function, thick vertical lines denote the zero and the location of maximum of the regression function.

1.2 Some non-standard problems

The statistical analysis should consist of careful planning of each experiment, of proper statistical analysis of collected data, and of verification of assumptions of the statistical methodology. In some situations, the standard methodology has to be modified in order to accommodate for some additional constraints or requirements. In Chapters 2–4, we will present some modifications of standard approaches concerning each stage of a nonparametric regression experiment.

Optimal design Optimality of a nonparametric regression experiment is most often discussed from the point of view of proper choice of the bandwidth, the kernel function of the type of nonparametric regression estimator. Concerning the problem of the choice of the density $f_X(.)$ of the design points, Müller (1984b) derived the distribution of design points minimizing the *Asymptotic Integrated Mean Squared Error* (AIMSE), defined as the limit of IMSE $= E[\int \{\hat{m}(x) - m(x)\} dH(x)]$ for $n \to \infty$.

Unfortunately, the results of Müller (1984b) lack clear interpretation because the definition of AIMSE involves a probability measure $H(.)$ that does not have any clear interpretation. In Chapter 2, we will try to overcome this obstacle by looking at a different criterion of optimality. More precisely, we will try to find designs minimizing the variability of estimators of location of some interesting points on the unknown regression curve, see Figure 1.3 for an illustration of a zero (i.e., the x-coordinate of the point, where the regression curve meets the horizontal axis) and the location of maximum.

In Chapter 2, we investigate asymptotic properties of estimators of zero and location of maximum under general conditions. The optimal design density, minimizing MSE or *Mean Absolute Deviation* (MAD) of the nonparametric regression estimator of the zero and the location of maximum, is then derived by applying standard calculus of variations. Chapter 2 generalizes results from Hlávka (2011) that were obtained only for the location of maximum with constant bandwidth and independent and identically distributed random errors.

Constraints Chapter 3 is devoted mainly to a constrained estimation with motivation coming from option pricing: we are interested in nonparametric estimation of the *state price density* (SPD) that may be expressed as the second derivative of the option pricing function, i.e., the prices of European

Call or Put options expressed as a function of the strike price. This setup naturally implies some no-arbitrage constraints on the nonparametric regression function describing this relationship and a covariance structure of the observed option prices (Härdle and Hlávka; 2009). In Chapter 3, we discuss this application as a special case of nonparametric regression. Some computational aspects of the proposed SPD estimator are discussed in two related papers (Hlávka; 2006a; Hlávka and Svojík; 2009).

Specification tests   After collecting and analyzing data, it is a good statistical practice to check the validity of the model assumptions. In the nonparametric regression model (1.5), the assumptions are usually expressed in terms of the random errors ε_i, $i = 1, \ldots, n$. Typically, we assume that these random errors are independent and identically distributed but, in certain situations, the nonparametric regression model may require that the random errors are heteroscedastic or correlated.

In Chapter 4, we modify some recently proposed specification tests (Hlávka et al.; 2011; Neumeyer; 2009) in order to test the appropriateness of the correlation structure of the observed option prices proposed in Chapter 3.

1.3   Results

The Appendix contains the most important author's papers related to the above topics. The paper Hlávka (2011) concerning the optimal design for nonparametric estimation of location of maximum may be found on page **??**. Some problems related to SPD estimation are solved in papers Hlávka (2006a); Härdle and Hlávka (2009); Hlávka and Svojík (2009) included, respectively, on pages **??**, **??**, and **??**. A specification test usable for homo- and heteroscedastic nonparametric regression models is described in the paper Hlávka et al. (2011) starting on page **??**.

All these papers concern nonparametric regression models. However, some extensions and modifications, mainly in Chapters 2 and 4, were necessary in order to unite these results in this habilitation thesis.

References

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*, John Wiley & Sons. Wiley Series in Probability and Mathematical Statistics.

Efromovich, S. (1999). *Nonparametric curve estimation*, Springer Series in Statistics, Springer-Verlag, New York.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Vol. 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics* **11**: 171–185.

Härdle, W. (1990). *Applied nonparametric regression*, Vol. 19 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities, *Journal of Econometrics* **150**(1): 1–15.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and semiparametric models*, Springer Series in Statistics, Springer-Verlag, New York.

Hlávka, Z. (2006a). Fast algorithm for nonparametric arbitrage-free SPD estimation, *Computational Statistics & Data Analysis* **51**(4): 2339–2349.

Hlávka, Z., Hušková, M. and Meintanis, S. G. (2011). Tests for independence in non-parametric heteroscedastic regression models, *Journal of Multivariate Analysis* **102**(4): 816–827.

Hlávka, Z. and Svojík, M. (2009). Application of extended Kalman filter to SPD estimation, *in* W. K. Härdle, N. Hautsch and L. Overbeck (eds), *Applied Quantitative Finance*, Springer, Berlin, pp. 233–247.

Hlávka, Z. (2011). On nonparametric estimators of location of maximum, *Acta Universitatis Carolinae - Mathematica et Physica* **52**(1): 5–13.

Müller, H.-G. (1984b). Optimal designs for nonparametric kernel regression, *Statistics & Probability Letters* **2**: 285–290.

Nadaraya, E. (1964). On estimating regression, *Theory of Probability and its Applications* **9**: 157–159.

Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves*, Vol. 20 of *Mathematics and its Applications (Soviet Series)*, Kluwer Academic Publishers Group, Dordrecht. Translated from the Russian by Samuel Kotz.

Neumeyer, N. (2009). Testing independence in nonparametric regression, *Journal of Multivariate Analysis* **100**: 1551–1566.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression, *Annals of Statistics* **22**(3): 1346–1370.

Simonoff, J. S. (1996). *Smoothing methods in statistics*, Springer Series in Statistics, Springer-Verlag, New York.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*, Springer Series in Statistics, Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, Vol. 60 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Watson, G. (1964). Smooth regression analysis., *Sankhyā, Ser. A* **26**: 359–372.

Zvára, K. (2008). *Regression (in Czech)*, Matfyzpress, Prague.

Chapter 2

Nonparametric estimators of location

In this chapter, we investigate a nonparametric estimator of the zero and the location of maximum of a regression function (or its derivative) and we derive the distribution of design points minimizing the expected variance of the location estimator. We concentrate on fixed-design nonparametric regression models that allow a natural interpretation of the proposed experiment design.

The problem of nonparametric estimation of the location of maximum is somewhat related to nonparametric estimation of the mode of a probability distribution (i.e., the value at which the probability density function attains its maximum), see Parzen (1962), Eddy (1980), Eddy (1982), Romano (1988b). Asymptotic normality of the random-design Nadaraya-Watson nonparametric estimator of the location of maximum of an unknown regression function (regression mode) is derived in Ziegler (2003). The mean squared error (MSE) of both random and fixed design estimators (Nadaraya-Watson, Gasser-Müller, Priestly-Chao) of the location of maximum under α-mixing assumptions is investigated in Wieczorek and Ziegler (2010). In this chapter, we complement these results by proving asymptotic normality of both the empirical zero and of the empirical location of maximum using the Gasser-Müller estimator and assuming that the random errors are strongly mixing. In addition, we review and generalize the asymptotically optimal distribution of design points obtained in Hlávka (2011) to estimators of zeros and locations of extrema of derivatives of regression functions.

In Sections 2.1–2.3, we extend the results of Müller (1985) and establish asymptotic normality of the *empirical zero* (the value at which the nonparametric regression estimator meets the horizontal axis) and the *empirical location of maximum* (the value at which the nonparametric regression estimator attains its maximum) under strongly mixing assumptions. In Section 2.4, we discuss the bandwidth choice. The asymptotically optimal distribution of design points is derived in Section 2.5, see also Müller (1984c); Hlávka (2011). The small sample properties of the proposed experiment design are investigated in a simulation study in Section 2.6.

2.1 Introduction

In this chapter, the basis of our investigation is the fixed design nonparametric regression model (1.9), i.e.,

$$Y_i = m(x_i) + \varepsilon_i, \text{ for } i = 1, \ldots, n, \tag{2.1}$$

where Y_i's are observations of the response variable, $0 \le x_1 < \cdots < x_n \le 1$ are fixed values of the explanatory variable defined by a probability density function $f_X(.)$ (i.e., $F_X(x_i) = \int_0^{x_i} f_X(u)du = i/n + O(n^{-2})$), and ε_i's are random errors.

Recall that the Gasser–Müller (GM) kernel regression estimator (1.10) of the regression function $m(x)$ is defined as:

$$\hat{m}_{n,b_n}(x) = \frac{1}{b_n} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{b_n}\right) du\, Y_i, \tag{2.2}$$

where $s_{i-1} = \frac{1}{2}(x_i + x_{i-1})$, b_n is the bandwidth and $K(.)$ a kernel function.

Similarly, we may estimate derivatives of the regression function:

$$\hat{m}_{n,b_n,\nu}(x) = \frac{1}{b_n^{\nu+1}} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du Y_i = \sum_{i=1}^{n} a_{n,i,b_n,\nu}(x) Y_i, \qquad (2.3)$$

where $a_{n,i,b_n,\nu}(x)$ are (non-random) weights, b_n is the bandwidth, and $K_\nu(.)$ is a kernel function of order (ν, k) (Gasser and Müller; 1984), i.e., the support of the Lipschitz continuous function $K_\nu(.)$ is the interval $\langle -1, 1 \rangle$ and

$$\int_{-1}^{1} K_\nu(x) x^j dx = \begin{cases} (-1)^\nu \nu! & \text{for } j = \nu, \\ 0 & 0 \le j < k, j \neq \nu, \\ (-1)^k k! B_{k,k} & \text{for } j = k, \\ (-1)^{k+1}(k+1)! B_{k,k+1} & \text{for } j = k+1. \end{cases} \qquad (2.4)$$

For example, the quartic kernel $K(x) = I(|x| \le 1)15(1-x^2)^2/16$ is of order $(0,2)$ with $B_{2,2} = 1/14$ because $\int K(x)dx = 1$, $\int K(x)xdx = 0$, and $\int K(x)x^2dx = 1/7$.

The properties of $\hat{m}_{n,b_n}(x)$ and $\hat{m}_{n,b_n,\nu}(x)$ are described in detail in Gasser and Müller (1984). We review and extend some of these results in Section 2.1.2.

Empirical location of maximum   The natural estimator of the location of maximum is the so-called *empirical location of maximum* defined as the x-coordinate of the point in which the nonparametric regression estimator reaches its maximum. In the following, the symbol $\theta_{n,\nu}$ denotes the (unknown) true location of maximum of the ν-th derivative of the regression function $m(.)$, i.e., $\theta_\nu = \arg\max_{x\in[0,1]} m^{(\nu)}(x)$ and $\hat{\theta}_{n,b_n,\nu}$ denotes the location of maximum of the nonparametric regression estimator $\hat{m}^{(\nu)}(.)$, i.e., $\hat{\theta}_{n,b_n,\nu} = \arg\max_{x\in[0,1]} \hat{m}_{n,b_n,\nu}(x)$. The value of the nonparametric regression estimator in the empirical location of maximum, $\hat{m}_{n,b_n,\nu}(\hat{\theta}_{n,b_n,\nu})$ is an estimator of the maximum of the ν-th derivative of the regression function $m(.)$, i.e., $m^{(\nu)}(\theta_\nu) = \max_{x\in[0,1]} m^{(\nu)}(x)$

Empirical zero   Along the same lines, we define $\hat{\xi}_{n,b_n,\nu}$, the *empirical zero* of ν-th derivative of $m(.)$ as a solution of the equation $\hat{m}_{n,b_n,\nu}(\hat{\xi}_{n,b_n,\nu}) = 0$. The empirical zero estimates the zero, ξ_ν, of $m^{(\nu)}(.)$ defined as a solution of the equation $m^{(\nu)}(\xi_\nu) = 0$. The zero ξ is sometimes also called a root of the equation $m^{(\nu)}(\xi_\nu) = 0$.

2.1.1   Notation and assumptions

Let us fix the parameters ν and k and the kernel function $K_\nu(.)$ of order (ν, k). We start by establishing asymptotic properties of the nonparametric regression estimator $\hat{m}_{n,b_n,\nu}(.)$ defined by (2.3). Later on, in Sections 2.2 and 2.3, these properties are used to establish the asymptotic normality of $\hat{\xi}_{n,b_n,\nu}$, the nonparametric estimator of the zero of $m^{(\nu)}(.)$ and $\hat{\theta}_{n,b_n,\nu}$, the nonparametric estimator of the location of maximum of $m^{(\nu)}(.)$. For simplicity, the indices k, ν, or b_n may be omitted whenever it is convenient. For example, we may write $a_{n,i}$ or $a_{n,i}(x)$ instead of $a_{n,i,b_n,\nu}(x)$.

The following assumptions are used in order to derive the asymptotic distribution of the nonparametric regression function estimator:

A1. **Regression function:** $m^{(k)}(.)$ is Lipschitz continuous.

A2. **Kernel:** The Lipschitz continuous kernel function $K_\nu(.)$ with support $\langle -1, 1 \rangle$ is of order (ν, k), where $\nu \ge 0$, $k > \nu + 1$ and the difference $k - \nu$ is even.

A3. **Bandwidth:** $b_n \to 0$, $nb_n^2 \to \infty$, $nb_n^{2\nu+1} \to \infty$.

A4. **Design:** The density $f(x) > \delta_f > 0$ is Lipschitz (γ_f)-continuous with $0 < \gamma_f \le 1$, i.e., $|f(u) - f(v)| \le L_f|u - v|^{\gamma_f}$ for all u and $v \in [0, 1]$ and for some $L_f > 0$.

A5. **Random errors:** Denoting $\varepsilon_i = \varepsilon_n(x_i)$, where $\varepsilon_n(x)$ is defined for all $x \in [0, 1]$, we assume:

$$E\varepsilon_n(x) = O(n^{-1}) \text{ uniformly in } x \in [0, 1],$$
$$\text{Cov}\{\varepsilon_n(x_i), \varepsilon_n(x_j)\} = \sigma^2(x_i)I(i = j) + O(n^{-1}) \text{ uniformly in } \{x_i, x_j\} \in [0, 1]^2,$$

the function $\sigma^2(.)$ is Lipschitz continuous, there exist $0 < \delta_\sigma$ and $D_\sigma < +\infty$ such that $\delta_\sigma < \sigma^2(x) < D_\sigma$ for all $x \in [0,1]$. Assume that $E|\varepsilon_i|^r < \infty$ for some $r > 2$ and that $\xi_{n,i} = (\varepsilon_i - E\varepsilon_i)/(\mathrm{Var}\,\varepsilon_i)^{1/2}$ is a strictly stationary strongly mixing sequence of random variables such that $\{\xi_{n,i}^2\}$ is a uniformly integrable family and $\lim_{k\to\infty} \bar{\rho}_k^* = \lim_{k\to\infty}[\sup_n\{\sup_{\mathrm{dist}(T,S)\geq k} \rho(\sigma(\xi_{n,i}, i \in T), \sigma(\xi_{n,j}, j \in S))\}] < 1$, where the symbol $\rho(.,.)$ denotes the maximal coefficient of correlation, i.e., $\rho(\mathcal{A}, \mathcal{B}) = \sup_{f \in L_2(\mathcal{A}), g \in L_2(\mathcal{B})} |corr(f,g)|$. For each n, let $\alpha_n(.)$ denote the mixing coefficients of the sequence $\xi I(|\xi_i| < n^{1/r})$ and assume that $\lim \sum_{i=1}^{\infty} \alpha_n(i) = \alpha_n^* < \infty$, $\lim_{\alpha_n^*} = \alpha^*$, and $\limsup_{i\to\infty} \left[1 + 6e^{1/2}\{\alpha_n(i^{1/r})\}^{2i^{r/(r-1)}}\right]^{i(r-1)/(2r)} < M < \infty$.

2.1.2 Asymptotic properties of the Gasser-Müller estimator

In this section, we summarize some basic asymptotic properties of the nonparametric regression estimator $\hat{m}_{n,b_n,\nu}(x)$ that are needed later on.

Assuming independent and identically distributed (iid) random errors, the asymptotic distribution of the Gasser-Müller estimator is given, e.g., in Gasser and Müller (1984). The consistency of the fixed-design nonparametric regression estimator under various mixing assumptions is investigated in Roussas (1989), the uniform strong consistency of the Nadaraya-Watson regression estimator is established in Roussas (1990b), and the asymptotic normality of the fixed-design nonparametric regression estimator under mixing assumptions is investigated in Roussas et al. (1992).

Let us now establish the asymptotic normality of the Gasser-Müller estimator for weakly dependent sequences of random errors (Lin and Lu; 1996) by using a Central Limit Theorem (CLT) for nonstationary weakly dependent triangular arrays of random variables (Peligrad; 1996).

Lemma 2.1. *Assume A1–A5. Then*

$$E\hat{m}_{n,b_n,\nu}(x) = m^{(\nu)}(x) + b_n^{k-\nu}\{B_{k,k}m^{(k)}(x) + o(1)\} + O(n^{-1}b_n^{-\nu}), \qquad (2.5)$$

$$\mathrm{Var}\,\hat{m}_{n,b_n,\nu}(x) = \frac{1}{nb_n^{2\nu+1}}\frac{\sigma^2(x)}{f(x)}\{V + O(b_n) + O(n^{-1}b_n^{-2})\}, \qquad (2.6)$$

where $V = \int_{-1}^{1} K_\nu^2(t)dt$ and

$$\frac{\hat{m}_{n,b_n,\nu}(x) - E\hat{m}_{n,b_n,\nu}(x)}{\{\mathrm{Var}\,\hat{m}_{n,b_n,\nu}(x)\}^{1/2}} \xrightarrow{\mathcal{D}} N(0,1). \qquad (2.7)$$

If $m^{(k+1)}(.)$ is continuous then

$$E\hat{m}_{n,b_n,\nu}(x) = m^{(\nu)}(x) + b_n^{k-\nu}\{B_{k,k}m^{(k)}(x) + b_n B_{k,k+1}m^{(k+1)} + o(b_n)\} + O(n^{-1}b_n^{-\nu}). \qquad (2.8)$$

Proof. Equations (2.5) and (2.8) follow exactly as in Gasser and Müller (1984); Müller (1984b, 1985) using Taylor expansion of $m(.)$ and

$$\sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right)du = \int_0^1 K_\nu\left(\frac{x-u}{b_n}\right)du = \int_{\max\left(-1,\frac{x-1}{b_n}\right)}^{\min\left(\frac{x}{b_n},1\right)} K_\nu(t)b_n dt = O(b_n). \qquad (2.9)$$

Next, similarly as in Müller (1985), we show that for the design points x_i defined in Section 2.1.1 it holds:

$$\begin{aligned}
s_i - s_{i-1} &= \frac{x_{i+1} - x_{i-1}}{2} = \frac{F^{-1}\left(\frac{i+1}{n}\right) - F^{-1}\left(\frac{i-1}{n}\right)}{2} + O(n^{-2}) \\
&= \frac{1}{n}\frac{F^{-1}\left(\frac{i+1}{n}\right) - F^{-1}\left(\frac{i-1}{n}\right)}{2/n} + O(n^{-2}) = \frac{1}{nf\{F^{-1}(u_i^*)\}} + O(n^{-2}) \\
&= \frac{1}{nf(x_i^*)} + O(n^{-2}) = \frac{1}{nf(x_i)} + O(n^{-2})
\end{aligned} \qquad (2.10)$$

for some $u_i^* \in ((i-1)/n, (i+1)/n)$ and $F^{-1}(u_i^*) = x_i^* = x_i + O(n^{-2})$. Similarly, for $a_{n,i} = a_{n,i,b_n,\nu}(x)$ defined in (2.3), we have that:

$$
\begin{aligned}
a_{n,i} &= \frac{1}{b_n^{\nu+1}} \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du = \frac{1}{b_n^{\nu+1}} \int_{s_{i-1}}^{s_i} \left\{ K_\nu\left(\frac{x-x_i}{b_n}\right) + O(n^{-1}b_n^{-1}) \right\} du \\
&= \frac{1}{b_n^{\nu+1}} (s_i - s_{i-1}) \left\{ K_\nu\left(\frac{x-x_i}{b_n}\right) + O(n^{-1}b_n^{-1}) \right\} \\
&= \frac{1}{b_n^{\nu+1}} \left\{ \frac{1}{nf(x_i)} + O(n^{-2}) \right\} \left\{ K_\nu\left(\frac{x-x_i}{b_n}\right) + O(n^{-1}b_n^{-1}) \right\} \\
&= \frac{1}{b_n^{\nu+1}} \left[\frac{K_\nu\{(x-x_i)/b_n\}}{nf(x_i)} + O(n^{-2}b_n^{-1}) \right] = \frac{K_\nu\{(x-x_i)/b_n\}}{nf(x_i)b_n^{\nu+1}} + O(n^{-2}b_n^{-\nu-2}). \quad (2.11)
\end{aligned}
$$

By the Lipschitz continuity of $K_\nu(.)$ and (2.10) we have that:

$$
\begin{aligned}
\sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du \right\}^2 &= \sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-x_i}{b_n}\right) du + O(n^{-2}b_n^{-1}) \right\}^2 \\
&= \sum_{i=1}^n \left\{ \frac{1}{nf(x_i)} K_\nu\left(\frac{x-x_i}{b_n}\right) + O(n^{-2}b_n^{-1}) \right\}^2 = \sum_{i=1}^n \left[\frac{1}{\{nf(x_i)\}^2} K_\nu^2\left(\frac{x-x_i}{b_n}\right) + O(n^{-3}b_n^{-1}) \right] \\
&= \sum_{i=1}^n \int_{s_{i-1}}^{s_i} \frac{1}{nf(u)} K_\nu^2\left(\frac{x-u}{b_n}\right) du + O(n^{-2}b_n^{-1}) = \int_0^1 \frac{1}{nf(u)} K_\nu^2\left(\frac{x-u}{b_n}\right) du + O(n^{-2}b_n^{-1}) \\
&= \int_{\max\left(-1,\frac{x-1}{b_n}\right)}^{\min\left(\frac{x}{b_n},1\right)} \frac{1}{nf(x-tb_n)} K_\nu^2(t) b_n dt + O(n^{-2}b_n^{-1}) = O(n^{-1}b_n) \quad (2.12)
\end{aligned}
$$

because by our assumptions $nb_n^2 \longrightarrow \infty$.

Using A5, (2.9), and (2.12) and proceeding similarly as in (2.12), we obtain that:

$$
\begin{aligned}
\operatorname{Var} \hat{m}_{n,b_n,\nu}(x) &= \frac{1}{b_n^{2\nu+2}} \sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du \right\}^2 \{\sigma^2(x_i) + O(n^{-1})\} \\
&\quad + \frac{1}{b_n^{2\nu+2}} \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du \right\} \left\{ \int_{s_{j-1}}^{s_j} K_\nu\left(\frac{x-u}{b_n}\right) du \right\} O(n^{-1}) \\
&= \frac{1}{b_n^{2\nu+2}} \sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du \right\}^2 \{\sigma^2(x_i)\} + O(n^{-2}b_n^{-2\nu-1}) + O(n^{-1}b_n^{-2\nu}) \\
&= \frac{1}{b_n^{2\nu+2}} \left\{ \int_{\max\left(-1,\frac{x-1}{b_n}\right)}^{\min\left(\frac{x}{b_n},1\right)} \frac{\sigma^2(x-tb_n)}{nf(x-tb_n)} K_\nu^2(t) b_n dt + O(n^{-2}b_n^{-1}) \right\} + O(n^{-2}b_n^{-2\nu-1}) + O(n^{-1}b_n^{-2\nu}) \\
&= \frac{1}{nb_n^{2\nu+1}} \frac{\sigma^2(x)}{f(x)} \int_{-1}^1 K_\nu^2(t) dt + O(n^{-1}b_n^{-2\nu}) + O(n^{-2}b_n^{-2\nu-3}) + O(n^{-2}b_n^{-2\nu-1}) + O(n^{-1}b_n^{-2\nu}) \\
&= \frac{1}{nb_n^{2\nu+1}} \frac{\sigma^2(x)}{f(x)} \int_{-1}^1 K_\nu^2(t) dt + O(n^{-1}b_n^{-2\nu}) + O(n^{-2}b_n^{-2\nu-3}) \\
&= \frac{1}{nb_n^{2\nu+1}} \frac{\sigma^2(x)}{f(x)} \{V + O(b_n) + O(n^{-1}b_n^{-2})\}.
\end{aligned}
$$

Defining:

$$
\tilde{m}_{n,b_n,\nu}(x) = \sum_{i=1}^n a_{n,i}\{m(x_i) + (\operatorname{Var}\varepsilon_i)^{1/2}\xi_i\} = \sum_{i=1}^n a_{n,i}\{m(x_i) + \varepsilon_i - E\varepsilon_i\} = \hat{m}_{n,b_n,\nu}(x) - \sum_{i=1}^n a_{n,i} E\varepsilon_i,
$$

it is easy to see that $\tilde{m}_{n,b_n,\nu}(x) = \hat{m}_{n,b_n,\nu}(x) + O(1/n)$, $E\tilde{m}_{n,b_n,\nu}(x) = E\hat{m}_{n,b_n,\nu}(x) + O(1/n)$, and $\text{Var}\,\tilde{m}_{n,b_n,\nu}(x) = \text{Var}\,\hat{m}_{n,b_n,\nu}(x)$. Hence, in order to establish the asymptotic normality result (2.7), it suffices to show the asymptotic normality of

$$\frac{\tilde{m}_{n,b_n,\nu}(x) - E\tilde{m}_{n,b_n,\nu}(x)}{\{\text{Var}\,\tilde{m}_{n,b_n,\nu}(x)\}^{1/2}}$$

which immediately follows by Peligrad (1996, Corollary 2.1) from assumption A5 because:

$$\max_{1 \le i \le n} \frac{|a_{ni}(\text{Var}\,\varepsilon_i)^{1/2}|}{\{\text{Var}\,\tilde{m}_{n,b_n,\nu}(x)\}^{1/2}} = O(n^{-1/2}b_n^{-1/2}) \overset{n\to\infty}{\longrightarrow} 0$$

by (2.11) and (2.6), and

$$\sup_n \frac{\sum_{i=1}^n a_{ni}^2 \text{Var}\,\varepsilon_i}{\text{Var}\,\tilde{m}_{n,b_n,\nu}(x)} \quad \le \quad \sup_n \frac{b_n^{-2\nu-2}D_\sigma \sum_{i=1}^n \left\{ \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du \right\}^2}{n^{-1}b_n^{-2\nu-1}\sigma^2(x)\{V + O(b_n) + O(n^{-1}b_n^{-2})\}/f(x)}$$

$$= \quad \sup_n \frac{\int_{\max\{-1,(x-1)/b_n\}}^{\min(x/b_n,1)} \{D_\sigma K_\nu^2(t)/f(x-tb_n)\}dt + O(n^{-2}b_n^{-1})}{\sigma^2(x)\{V + O(b_n) + O(n^{-1}b_n^{-2})\}/f(x)} < +\infty$$

by (2.12) and (2.6). $\qquad\square$

After establishing the pointwise asymptotic normality of $\hat{m}_{n,b_n,\nu}(.)$, we investigate the uniform convergence rate.

Lemma 2.2. *Assume A1–A5. Then*

$$\sup_{x\in[\delta,1-\delta]} \left| E\hat{m}_{n,b_n,\nu}(x) - m^{(\nu)}(x) \right| = O(b_n^{k-\nu}) + O(n^{-1}b_n^{-\nu})$$

for any fixed $\delta \in (0, 1/2)$.

Proof. By the Taylor expansion:

$$m(x - tb_n) = \sum_{j=0}^{k-1} \frac{m^{(\nu)}(x)}{j!}(-tb_n)^j + \int_0^{-tb_n} \frac{m^{(k)}(x+u)}{(k-1)!}(-tb_n - u)^{k-1}du.$$

Recalling that $m^{(k)}(.)$ is Lipschitz continuous with constant L_k by A1, we obtain:

$$\sup_{|t|\le 1} \left| m(x-tb_n) - \sum_{j=0}^{k-1} \frac{m^{(\nu)}(x)}{j!}(-tb_n)^j \right|$$

$$= \quad \sup_{|t|\le 1} \left| \int_0^{-tb_n} \frac{m^{(k)}(x+u)}{(k-1)!}(-tb_n - u)^{k-1}du \right| \le \frac{b_n}{(k-1)!} \sup_{|t|\le 1} \left| m^{(k)}(x+u)(tb_n)^{k-1} \right|$$

$$\le \quad \frac{b_n^k}{(k-1)!} \sup_{|t|\le 1} \left| m^{(k)}(x) + L_k tb_n \right| \le \frac{b_n^k|m^{(k)}(x)| + L_k b_n^{k+1}}{(k-1)!} = O(b_n^k).$$

Similarly as in the proof of Lemma 2.1, using the assumption that $K_\nu(.)$ is a kernel function of order (ν, k), we have uniformly in $[\delta, 1-\delta]$ that:

$$E\hat{m}_{n,b_n,\nu}(x) \quad = \quad \frac{1}{b_n^{\nu+1}} \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_\nu\left(\frac{x-u}{b_n}\right) du\{m(x) + E\varepsilon_i\}$$

$$= \quad \frac{1}{b_n^\nu} \int_{\max\{-1,(x-1)/b_n\}}^{\min(x/b_n,1)} m(x-tb_n)K_\nu(t)\,dt + O(n^{-1}b_n^{-\nu})$$

$$= \quad \frac{1}{b_n^\nu} \int_{\max\{-1,(x-1)/b_n\}}^{\min(x/b_n,1)} \left\{ \sum_{j=0}^{k-1} \frac{m^{(\nu)}(x)}{j!}(-tb_n)^j + O(b_n^k) \right\} K_\nu(t)\,dt + O(n^{-1}b_n^{-\nu})$$

$$= \quad m^{(\nu)}(x) + O(b_n^{k-\nu}) + O(n^{-1}b_n^{-\nu}). \tag{2.13}$$

$\qquad\square$

Corollary 1. *The optimum rate of convergence in Lemma 2.2 is $O(b_n^{k-\nu})$ and it is achieved for $b_n = n^{-1/k}$.*

Proof. The statement follows directly by comparing the rates in Lemma 2.2. $\square$

In order to deal with correlated random variables, we apply a Hoeffding type exponential inequality for strongly mixing sequences of random variables, see Roussas (1996).

Lemma 2.3. *Assume A1–A5. Then*

$$\sup_{x \in [\delta, 1-\delta]} |\hat{m}_{n,b_n,\nu}(x) - m^{(\nu)}(x)| = O\left\{ b_n^{(r-1)/2} + b_n^{k-\nu} + \left(\frac{\log n}{nb_n^{2\nu+1}} \right)^{1/2} \right\} \quad a.s.$$

for any fixed $\delta \in (0, 1/2)$.

Proof. Without loss of generality, we assume that $b_n < \delta$. Another possible approaches to the boundary effects would be to introduce circularity assumptions or to use a boundary kernel (Müller; 1984a).

The proof is organized similarly as the proof of Lemma 2.2 in Müller (1985) or the proof of the second Lemma in Müller (1984a), see also Cheng and Lin (1981), Mack and Silverman (1982), or Roussas (1990b).

Similarly as in the proof of Lemma 2.1 we define $\tilde{m}_{n,b_n,\nu}(x) = \sum_{i=1}^n a_{n,i}(x)\{m(x_i) + (\text{Var}\,\varepsilon_i)^{1/2}\xi_i\}$ and recall that:

$$\sup_{x \in [\delta, 1-\delta]} |\hat{m}_{n,b_n,\nu}(x) - \tilde{m}_{n,b_n,\nu}(x)| = O(n^{-1}). \tag{2.14}$$

In order to establish the uniform convergence rate of $\tilde{m}_{n,b_n,\nu}(.)$, we define estimators $\tilde{h}_{n,b_n,\nu}(x) = \sum_{i=1}^n a_{n,i}(x)(\text{Var}\,\varepsilon_i)^{1/2}\xi_i$, and $\bar{h}_{n,b_n,\nu}(x) = \sum_{i=1}^n a_{n,i}(x)(\text{Var}\,\varepsilon_i)^{1/2}\xi_i I(|\xi_i| < B_n)$.

Let $R_n \subset (0,1)$ denote a set with cardinality $O(n^2)$ such that for all $x \in (0,1)$ there exists $\tau(x) \in R_n$ satisfying $\sup_{x \in (0,1)} |x - \tau(x)| = \|x - \tau(x)\| = O(n^{-2})$. Then

$$\sup_{x \in [\delta, 1-\delta]} |\tilde{m}_{n,b_n,\nu}(x) - E\tilde{m}_{n,b_n,\nu}(x)|$$

$$= \sup_{x \in [\delta, 1-\delta]} |\tilde{h}_{n,b_n,\nu}(x) - E\tilde{h}_{n,b_n,\nu}(x)|$$

$$\leq \sup_{x \in [\delta, 1-\delta]} |\tilde{h}_{n,b_n,\nu}(x) - \bar{h}_{n,b_n,\nu}(x)| + \sup_{x \in [\delta, 1-\delta]} |\bar{h}_{n,b_n,\nu}(x) - \bar{h}_{n,b_n,\nu}\{\tau(x)\}|$$

$$+ \sup_{x \in [\delta, 1-\delta]} |\bar{h}_{n,b_n,\nu}\{\tau(x)\} - E\bar{h}_{n,b_n,\nu}\{\tau(x)\}| + \sup_{x \in [\delta, 1-\delta]} |E\bar{h}_{n,b_n,\nu}\{\tau(x)\} - E\bar{h}_{n,b_n,\nu}\{x\}|$$

$$+ \sup_{x \in [\delta, 1-\delta]} |E\bar{h}_{n,b_n,\nu}\{x\} - E\tilde{h}_{n,b_n,\nu}\{x\}| = A + B + C + D + E. \tag{2.15}$$

Concerning A, Markov's inequality implies that $P(|\xi_i| > B_i) \leq E|\xi_i|^r/B_i^r$. This implies that $\sum_{i=1}^{+\infty} P(|\xi_i| > B_i) \leq \sum_{i=1}^{+\infty} \{E|\xi_i|^r/B_i^r\} < +\infty$, and the Borel-Cantelli lemma immediately leads that the probability of the event $\{|\xi_i| > B_i\}$ occurring infinitely many times is equal to zero. Hence, there exists N_ω such that $\tilde{h}_{n,b_n,\nu}(x) = \bar{h}_{n,b_n,\nu}(x)$ for all x and for all $n > N_\omega$.

Concerning E, we have by Jensen's, Hölder's, and Markov's inequality uniformly in x that:

$$|E\bar{h}_{n,b_n,\nu}\{x\} - E\tilde{h}_{n,b_n,\nu}\{x\}| \leq E|\bar{h}_{n,b_n,\nu}\{x\} - \tilde{h}_{n,b_n,\nu}\{x\}|$$

$$\leq E\sum_{i=1}^n |a_{n,i}(x)|\,(\text{Var}\,\varepsilon_i)^{1/2}\,|\xi_i I(|\xi_i| \geq B_n)| \leq D_\sigma^{1/2}\left\{\sum_{i=1}^n |a_{n,i}(x)|\right\} E\{|\xi_1|\,I(|\xi_1| \geq B_n)\}$$

$$\leq D_\sigma^{1/2}\left\{\sum_{i=1}^n |a_{n,i}(x)|\right\} (E\,|\xi_1|^r)^{1/r}\{P(|\xi_1| \geq B_n)\}^{1-1/r}$$

$$\leq D_\sigma^{1/2}\left\{\sum_{i=1}^n |a_{n,i}(x)|\right\} (E\,|\xi_1|^r)^{1/r}\{E\,|\xi_1|^r/B_n^r\}^{1-1/r} = O(B_n^{1-r}). \tag{2.16}$$

Concerning the terms B and D, we use the Lipschitz continuity of the kernel function to obtain uniformly in x that:

$$|\bar{h}_{n,b_n,\nu}(x) - \bar{h}_{n,b_n,\nu}\{\tau(x)\}|$$

$$= \left| \sum_{i=1}^{n} a_{n,i}(x)(\operatorname{Var}\varepsilon_i)^{1/2}\xi_i I(|\xi_i| < B_n) - \sum_{i=1}^{n} a_{n,i}\{\tau(x)\}(\operatorname{Var}\varepsilon_i)^{1/2}\xi_i I(|\xi_i| < B_n) \right|$$

$$= \left| \sum_{i=1}^{n} [a_{n,i}(x) - a_{n,i}\{\tau(x)\}](\operatorname{Var}\varepsilon_i)^{1/2}\xi_i I(|\xi_i| < B_n) \right|$$

$$\leq \frac{1}{b_n^{\nu+1}} \sum_{i=1}^{n} \left| \int_{s_{i-1}}^{s_i} \left\{ K_\nu\left(\frac{x-u}{b_n}\right) - K_\nu\left(\frac{\tau(x)-u}{b_n}\right) \right\} du\, \xi_i I(|\xi_i| < B_n) \right|$$

$$\leq \frac{1}{b_n^{\nu+1}} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} L\left| \frac{\tau(x)-x}{b_n} \right| du\, |\xi_i| I(|\xi_i| < B_n)$$

$$\leq \frac{1}{b_n^{\nu+2}} \sum_{i=1}^{n} L(s_i - s_{i-1})\, |\tau(x) - x|\, du\, |B_n = O(b_n^{-\nu-2}n^{-2}B_n). \tag{2.17}$$

It remains to investigate the term C:

$$\sup_{x \in [\delta, 1-\delta]} |\bar{h}_{n,b_n,\nu}\{\tau(x)\} - E\bar{h}_{n,b_n,\nu}\{\tau(x)\}| = \max_{\tau \in R_n} |\bar{h}_{n,b_n,\nu}\{\tau\} - E\bar{h}_{n,b_n,\nu}\{\tau\}| \tag{2.18}$$

and, obviously,

$$P(\max_{\tau \in R_n} |\bar{h}_{n,b_n,\nu}(\tau) - E\bar{h}_{n,b_n,\nu}(\tau)| > \varepsilon_n) \leq \sum_{\tau \in R_n} P(|\bar{h}_{n,b_n,\nu}(\tau) - E\bar{h}_{n,b_n,\nu}(\tau)| > \varepsilon_n). \tag{2.19}$$

Denoting $\xi_i^{*,n} = \xi_i I(|\xi_i| < B_n)$, η_n a sequence of positive integers tending to ∞, and $\mu_n = \lfloor n/2\eta_n \rfloor$ (so that $n \approx 2\mu_n\eta_n$), we apply a Hoeffding-type inequality for an α-mixing sequence of random variables (Roussas; 1996, Theorem 4.1) to show that:

$$P\left\{ \left(\frac{\beta_n}{\log n}\right)^{1/2} |\bar{h}_{n,b_n,\nu}(\tau) - E\bar{h}_{n,b_n,\nu}(\tau)| > \varepsilon \right\}$$

$$= P\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} n a_{n,i}(\tau)(\operatorname{Var}\varepsilon_i)^{1/2}(\xi_i^{*,n} - E\xi_i^{*,n}) \right| > \varepsilon(\log n)^{1/2}\beta_n^{-1/2} \right\}$$

$$\leq P\left(\left| \frac{1}{n} \sum_{i=1}^{n} K_0(\xi_i^{*,n} - E\xi_i^{*,n}) \right| > \varepsilon b_n^{\nu+1}(\log n)^{1/2}\beta_n^{-1/2} \right)$$

$$\leq K_1 \left[1 + 6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n} \right]^{\mu_n} \exp(-K_2 n\varepsilon^2 b_n^{2\nu+2}\beta_n^{-1}B_n^{-2}\log n),$$

$$= K_1 \left[1 + 6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n} \right]^{\mu_n} \exp\left(\log n^{-K_2 n\varepsilon^2 b_n^{2\nu+2}\beta_n^{-1}B_n^{-2}} \right)$$

$$= K_1 \left[1 + 6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n} \right]^{\mu_n} n^{-K_2 n\varepsilon^2 b_n^{2\nu+2}\beta_n^{-1}B_n^{-2}} \tag{2.20}$$

where the sequence $\beta_n \to \infty$ will be specified later, $\varepsilon b_n^{\nu+1}(\log n)^{1/2}\beta_n^{-1/2} < K_3\eta_n^{-1}$, and K_0, K_1, K_2, and K_3 are some constants.

Combining (2.18)–(2.20), we obtain for any $\varepsilon > 0$ that:

$$P\left\{ \left(\frac{\beta_n}{\log n}\right)^{1/2} \sup_{x \in [\delta, 1-\delta]} |\bar{h}_{n,b_n,\nu}\{\tau(x)\} - E\bar{h}_{n,b_n,\nu}\{\tau(x)\}| > \varepsilon \right\}$$

$$\leq \sum_{\tau \in R_n} \left\{ \left(\frac{\beta_n}{\log n}\right)^{1/2} |\bar{h}_{n,b_n,\nu}\{\tau\} - E\bar{h}_{n,b_n,\nu}\{\tau\}| > \varepsilon \right\}$$

$$\leq K_1 \left[1 + 6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n} \right]^{\mu_n} n^{2-K_2 n\varepsilon^2 b_n^{2\nu+2}\beta_n^{-1}B_n^{-2}} \xrightarrow{n \to \infty} 0,$$

assuming that $\liminf_{n\to\infty} nb_n^{2\nu+2}\beta_n^{-1}B_n^{-2} > 2K_2^{-1}\varepsilon^{-2}$, $\limsup_{n\to\infty}\left[1+6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n}\right]^{\mu_n} < \infty$, and $\eta_n b_n^{\nu+1}(\log n)^{1/2}\beta_n^{-1/2} \to 0$.

Using (2.18)–(2.19), we finally obtain for the term C:

$$\sum_{n=1}^{\infty} P\left\{\left(\frac{\beta_n}{\log n}\right)^{1/2} \sup_{x\in[\delta,1-\delta]} |\bar{h}_{n,b_n,\nu}\{\tau(x)\} - E\bar{h}_{n,b_n,\nu}\{\tau(x)\}| > \epsilon\right\}$$

$$\leq \sum_{n=1}^{\infty} K_1 \left[1+6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n}\right]^{\mu_n} n^{2-K_2 n\varepsilon^2 b_n^{2\nu+2}\beta_n^{-1}B_n^{-2}} \qquad (2.21)$$

and, if $\liminf_{n\to\infty} nb_n^{2\nu+2}\beta_n^{-1}B_n^{-2} > 3K_2^{-1}\varepsilon^{-2}$, the term on the right hand side of (2.21) is finite and Borel-Cantelli lemma implies that $(\beta_n/\log n)^{1/2} C = O(1)$ a.s.

Assuming that $\liminf_{n\to\infty} nb_n^{2\nu+2}\beta_n^{-1}B_n^{-2} > 3K_2^{-1}\varepsilon^{-2}$, $\limsup_{n\to\infty}\left[1+6e^{1/2}\{\alpha(\eta_n)\}^{1/\mu_n}\right]^{\mu_n} < \infty$, and $\eta_n = o(b_n^{-\nu-1}\beta_n^{1/2}\log n)$ and using (2.16) and (2.17), we thus obtain for (2.15):

$$\sup_{x\in[\delta,1-\delta]} |\tilde{m}_{n,b_n,\nu}(x) - E\tilde{m}_{n,b_n,\nu}(x)| = O\{B_n^{1-r} + n^{-2}b_n^{-\nu-2}B_n + (\log n/\beta_n)^{1/2}\}, \quad \text{a.s.} \qquad (2.22)$$

Clearly, we may now choose constant c_3 large enough so that $B_n = c_3 b_n^{-1/2} = O(b_n^{-1/2})$ and the sequence $\beta_n = nb_n^{2\nu+1}$ still satisfies the assumption $\liminf_{n\to\infty} nb_n^{2\nu+2}\beta_n^{-1}B_n^{-2} > 3K_2^{-1}\varepsilon^{-2}$. Plugging $B_n = c_3 b_n^{-1/2}$ and $\beta_n = O(b_n^{-1/2})$ into (2.22) and using (2.14), we get:

$$\sup_{x\in[\delta,1-\delta]} |\hat{m}_{n,b_n,\nu}(x) - E\hat{m}_{n,b_n,\nu}(x)| = O\left\{b_n^{(r-1)/2} + \left(\frac{\log n}{nb_n^{2\nu+1}}\right)^{1/2}\right\}, \quad \text{a.s.} \qquad (2.23)$$

Applying Lemma 2.2, we finally obtain:

$$\sup_{x\in[\delta,1-\delta]} |\hat{m}_{n,b_n,\nu}(x) - m^{\nu}(x)|$$

$$\leq \sup_{x\in[\delta,1-\delta]} |\hat{m}_{n,b_n,\nu}(x) - E\hat{m}_{n,b_n,\nu}(x)| + \sup_{x\in[\delta,1-\delta]} |E\hat{m}_{n,b_n,\nu}(x) - m^{\nu}(x)|$$

$$= O\left\{b_n^{(r-1)/2} + b_n^{k-\nu} + \left(\frac{\log n}{nb_n^{2\nu+1}}\right)^{1/2}\right\}, \quad \text{a.s.}$$

$\square$

Remark 1. *The optimal rate of uniform convergence in (2.23) is*

$$\left(\frac{n}{\log n}\right)^{-\frac{r-1}{2r+4\nu}}$$

and it is achieved for the bandwidth

$$b_n^* = \left(\frac{n}{\log n}\right)^{-1/(r+2\nu)}.$$

Proof. The optimal rate of uniform convergence follows directly from the comparison of the terms $b_n^{(r-1)/2}$ and $\left(\frac{\log n}{nb_n^{2\nu+1}}\right)^{1/2}$. Notice that b_n^* satisfies assumption A3 because $r > 2$ by A5 and

$$n(b_n^*)^{2\nu+1} = n^{1-\frac{2\nu+1}{2\nu+r}}\log^{\frac{2\nu+1}{2\nu+r}} = n^{1-\frac{2\nu+r}{2\nu+r}+\frac{r-1}{2\nu+r}}\log^{\frac{2\nu+1}{2\nu+r}} = n^{\frac{r-1}{2\nu+r}}\log^{\frac{2\nu+1}{2\nu+r}} \to +\infty.$$

$\square$

Remark 2. *Assuming that $k - \nu \leq (r-1)/2$, the optimal rate of uniform convergence in Lemma 2.3 is*

$$\left(\frac{n}{\log n}\right)^{-\frac{k-\nu}{2k+1}}$$

and it is achieved for the bandwidth

$$b_n^* = \left(\frac{n}{\log n}\right)^{-1/(2k+1)}.$$

Proof. The rate of convergence follows directly from the comparison of the terms $b_n^{k-\nu}$ and $\left(\frac{\log n}{nb_n^{2\nu+1}}\right)^{1/2}$. Notice that b_n^* satisfies assumption A3 because $k > \nu$ by assumption A2 and

$$n(b_n^*)^{2\nu+1} = n^{1-\frac{2\nu+1}{2k+1}} \log^{\frac{2\nu+1}{2k+1}} = n^{\frac{2k-2\nu}{2k+1}} \log^{\frac{2\nu+1}{2\nu+r}} =\to +\infty.$$

$\square$

Remark 3. *The optimal rate of uniform convergence in Lemma 2.3 is*

$$\left(\frac{n}{\log n}\right)^{-\min\left(\frac{k-\nu}{2k+1}, \frac{r-1}{2r+4\nu}\right)}$$

and it is achieved for the bandwidth

$$b_n^* = \left(\frac{n}{\log n}\right)^{-1/\min(2k+1, r+2\nu)}.$$

Proof. The proof follows from Remarks 1 and 2. $\square$

2.2 Nonparametric estimators of zeros

Let us recall that the symbol $\hat{\xi}_{n,b_n,\nu}$ denotes the empirical zero of $m^\nu(.)$, see Section 2.1.

In order to establish the asymptotic distribution of the estimator $\hat{\xi}_{n,b_n,\nu}$, we need some additional assumptions concerning the geometry of the regression function close to the true zero ξ_ν:

A6. **Geometry at zero:** Assume that there exist a, b, $c > 0$, and $\tau \geq 1$ such that $0 < a < \xi_\nu < b < 1$, $m^{(\nu)}$ is strictly monotonous on $\langle a, b \rangle$ and $|m^{(\nu)}(t)| \geq c|t - \xi_\nu|^\tau$ for $t \in \langle a, b \rangle$.

We start with a lemma establishing the rate of convergence of the estimators $\hat{\xi}_{n,b_n,\nu}$ by using the geometry of the regression function and the uniform convergence rate derived in Lemma 2.3.

Lemma 2.4. *Assuming A1–A6, then*

$$|\hat{\xi}_{n,b_n,\nu} - \xi_\nu| = O\left\{b_n^{(r-1)/(2\tau)} + b_n^{(k-\nu)/\tau} + \left(\frac{\log n}{nb_n^{2\nu+1}}\right)^{1/(2\tau)}\right\} \quad a.s.$$

Proof. It follows from Lemma 2.3 and Müller (1985, Lemma 2.4). $\square$

The proof of the following Theorems 2.1 follows very closely the proof of Theorem 3.1 in Müller (1985) but our assumptions concerning random errors are less restrictive.

Theorem 2.1. *Assume that assumptions A1–A6 hold, $nb_n^{2\nu+3}/\log n \to \infty$, the kernel $K_\nu(.)$ is differentiable, $K_\nu(-1) = K_\nu(1) = 0$, the derivative $K_\nu'(.)$ is Lipschitz continuous, the regression function $m(.)$ is $(k+1)$-times continuously differentiable and $m^{(\nu+1)}(\xi_\nu) \neq 0$. If $n^{1/2}b_n^{k+1/2} \to d \geq 0$, then*

$$(nb_n^{2\nu+1})^{1/2}(\hat{\xi}_{n,b_n,\nu} - \xi_\nu) \xrightarrow{\mathcal{D}} N\left(-\frac{dB_{k,k}m^{(k)}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)}, \frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)}\frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2}\right). \tag{2.24}$$

Proof. Notice that $\hat{m}'_{n,b_n\nu}(.)$, the first derivative of the GM kernel regression estimator $\hat{m}_{n,b_n\nu}(.)$, is also a GM kernel regression estimator with the kernel function $K_\nu'(.)$.

Clearly, $\int_{-1}^1 K_\nu' x^0 dx = \int_{-1}^1 K_\nu' dx = [K_\nu(x)]_{-1}^1 = 0$ and integrating per partes we obtain:

$$\int_{-1}^1 K_\nu' x^j dx = [K_\nu(x)x^j]_{-1}^1 - j\int_{-1}^1 K_\nu(x)x^{j-1}dx = -j\int_{-1}^1 K_\nu(x)x^{j-1}dx.$$

From (2.4) it follows that:

$$\int_{-1}^1 K_\nu'(x)x^j dx = -j\int_{-1}^1 K_\nu(x)x^{j-1}dx = \begin{cases} (-1)^{\nu+1}(\nu+1)! & \text{for } j = \nu+1, \\ 0 & 1 \leq j < k+1, j \neq \nu+1, \\ (-1)^{k+1}(k+1)!B_{k,k} & \text{for } j = k+1, \end{cases}$$

and it is easy to see that the derivative $K'_\nu(.)$ is actually a kernel function of order $(\nu + 1, k + 1)$. Therefore, Lemma 2.3 applies and we have for any $\delta \in (0, 1/2)$ that:

$$\sup_{x \in [\delta, 1-\delta]} |\hat{m}'_{n,b_n,\nu}(x) - m^{(\nu+1)}(x)| = O\left\{ b_n^{(r-1)/2} + b_n^{k-\nu} + \left(\frac{\log n}{n b_n^{2\nu+3}} \right)^{1/2} \right\} \to 0 \quad \text{a.s.} \qquad (2.25)$$

if $n b_n^{2\nu+3} / \log n \to \infty$.

By Taylor expansion of $\hat{m}_{n,b_n,\nu}$, we obtain that there exists $\xi^*_{n,b_n,\nu}$ between $\hat{\xi}_{n,b_n,\nu}$ and ξ_ν such that:

$$m^{(\nu)}(\xi_\nu) = 0 = \hat{m}_{n,b_n,\nu}(\hat{\xi}_{n,b_n,\nu}) = \hat{m}_{n,b_n,\nu}(\xi_\nu) + (\hat{\xi}_{n,b_n,\nu} - \xi_\nu) \hat{m}'_{n,b_n,\nu}(\xi^*_{n,b_n,\nu})$$

implying that:

$$\hat{\xi}_{n,b_n,\nu} - \xi_\nu = \frac{m^{(\nu)}(\xi_\nu) - \hat{m}_{n,b_n,\nu}(\xi_\nu)}{\hat{m}'_{n,b_n,\nu}(\xi^*_{n,\nu})}. \qquad (2.26)$$

For the denominator of (2.26) we have:

$$|\hat{m}'_{n,b_n,\nu}(\xi^*_{n,\nu}) - m^{(\nu+1)}(\xi_\nu)| \le |\hat{m}'_{n,b_n,\nu}(\xi^*_{n,\nu}) - m^{(\nu+1)}(\xi^*_{n,\nu})| + |m^{(\nu+1)}(\xi^*_{n,\nu}) - m^{(\nu+1)}(\xi_\nu)| \to 0 \quad \text{a.s.}$$

by (2.25) and Lemma 2.4. Hence, the asymptotic distribution of $\hat{\xi}_{n,b_n,\nu} - \xi_\nu$ is, up to the multiplicative constant $m^{(\nu+1)}(\xi_\nu)$, equal to the asymptotic distribution of $m^{(\nu)}(\xi_\nu) - \hat{m}_{n,b_n,\nu}(\xi_\nu)$. Applying Lemma 2.1, we obtain:

$$\frac{E\hat{m}_{n,b_n,\nu}(\xi_\nu) - \hat{m}_{n,b_n,\nu}(\xi_\nu)}{\{\text{Var}\,\hat{m}_{n,b_n,\nu}(x)\}^{1/2} m^{(\nu+1)}(\xi_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\{m^{(\nu+1)}(\xi_\nu)\}^2} \right)$$

and, if $b_n \to 0$, $n b_n^2 \to \infty$, and $n b_n^\nu \to \infty$,

$$n^{1/2} b_n^{\nu+1/2} \frac{m^{(\nu)}(\xi_\nu) + b_n^{k-\nu} B_{k,k} m^{(k)}(\xi_\nu) - \hat{m}_{n,b_n,\nu}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)} \frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2} \right),$$

implying that:

$$n^{1/2} b_n^{\nu+1/2} (\hat{\xi}_{n,b_n,\nu} - \xi_\nu) + n^{1/2} b_n^{k+1/2} \frac{B_{k,k} m^{(k)}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)} \frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2} \right).$$

$\square$

2.3   Nonparametric estimators of location of maximum

Let us recall that the symbol $\hat{\theta}_{n,b_n,\nu}$ denotes the empirical location of the maximum of $m^{(\nu)}(.)$, see Section 2.1. Similarly as in the previous section, we need an additional assumption concerning the geometry of the regression function close to the true location of maximum θ_ν:

A7. **Geometry at maximum:** Assume that there exist a, b, $c > 0$, and $\rho \ge 1$ such that $0 < a < \theta_\nu < b < 1$, $m^{(\nu)}$ is monotonously increasing on $\langle a, \theta_\nu \rangle$ and monotonously decreasing on $\langle \theta_\nu, b \rangle$, and $|m^{(\nu)}(t) - m^{(\nu)}(\theta_\nu)|$ for $t \in \langle a, b \rangle$.

We start with a lemma establishing the rate of convergence of the estimator $\hat{\theta}_{n,b_n,\nu}$ by using the geometry of the regression function and the uniform convergence rate derived in Lemma 2.3.

Lemma 2.5. *Assuming A1–A5 and A7, then*

$$|\hat{\theta}_{n,b_n,\nu} - \theta_\nu| = O\left\{ b_n^{(r-1)/(2\rho)} + b_n^{(k-\nu)/\rho} + \left(\frac{\log n}{n b_n^{2\nu+1}} \right)^{1/(2\rho)} \right\} \quad a.s.$$

and

$$|\hat{m}_{n,b_n,\nu}(\hat{\theta}_{n,b_n,\nu}) - m^{(\nu)}(\theta_\nu)| = O\left\{ b_n^{(r-1)/2} + b_n^{(k-\nu)} + \left(\frac{\log n}{n b_n^{2\nu+1}} \right)^{1/2} \right\} \quad a.s.$$

Proof. It follows from Lemma 2.3 and Müller (1985, Lemma 2.3). $\qquad\square$

The proof of the following Theorem 2.2 follows very closely the proof of Theorem 3.1 in Müller (1985) but our assumptions concerning random errors are less restrictive.

Theorem 2.2. *Assume A1–A5 and A7, $nb_n^{2\nu+5}/\log n \to \infty$, the kernel $K_\nu(.)$ is twice differentiable, $K_\nu'(-1) = K_\nu'(1) = 0$, the second derivative $K_\nu^{(2)}(.)$ is Lipschitz continuous, the regression function $m(.)$ is $(k+1)$-times continuously differentiable and $m^{(\nu+2)}(\theta_\nu) \neq 0$. If $n^{1/2}b_n^{k+3/2} \to d' \geq 0$, then*

$$(nb_n^{2\nu+3})^{1/2}(\hat{\theta}_{n,b_n,\nu} - \theta_\nu) \xrightarrow{\mathcal{D}} N\left(\frac{d'B_{k+1,k+1}m^{(k+1)}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)}, \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)}\frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2}\right), \qquad (2.27)$$

where $V' = \int_{-1}^{1}\{K_\nu'(t)\}^2 dt$.

Proof. Similarly as in the proof of Theorem 2.1, we notice that the second derivative of the GM kernel regression estimator, $\hat{m}_{n,b_n\nu}^{(2)}(.)$, is a GM kernel regression estimator with the kernel function $K_\nu^{(2)}(.)$.

Clearly, $\int_{-1}^{1}K_\nu^{(2)}x^0dx = \int_{-1}^{1}K_\nu^{(2)}dx = [K_\nu'(x)]_{-1}^1 = 0$ and integrating per partes we obtain:

$$\int_{-1}^{1}K_\nu^{(2)}x^j dx = [K_\nu'(x)x^j]_{-1}^1 - j\int_{-1}^{1}K_\nu'(x)x^{j-1}dx = -j\int_{-1}^{1}K_\nu'(x)x^{j-1}dx.$$

Similarly as in the proof of Theorem 2.1, it follows that:

$$\int_{-1}^{1}K_\nu^{(2)}(x)x^j dx = -j\int_{-1}^{1}K_\nu'(x)x^{j-1}dx$$

$$= j(j-1)\int_{-1}^{1}K_\nu'(x)x^{j-2}dx = \begin{cases} (-1)^{\nu+2}(\nu+2)! & \text{for } j = \nu+2, \\ 0 & 1 \leq j < k+2, j \neq \nu+2, \\ (-1)^{k+2}(k+2)!B_{k,k} & \text{for } j = k+2, \end{cases}$$

and the second derivative $K_\nu^{(2)}(.)$ is obviously a kernel function of order $(\nu+2, k+2)$. Therefore, Lemma 2.3 applies and we have for any $\delta \in (0, 1/2)$ that:

$$\sup_{x \in [\delta, 1-\delta]} |\hat{m}_{n,b_n,\nu}^{(2)}(x) - m^{(\nu+2)}(x)| = O\left\{b_n^{(r-1)/2} + b_n^{k-\nu} + \left(\frac{\log n}{nb_n^{2\nu+5}}\right)^{1/2}\right\} \to 0 \quad \text{a.s.} \qquad (2.28)$$

if $nb_n^{2\nu+5}/\log n \to \infty$. Equation (2.28) and Lemma 2.5 together with the continuity of $m^{(\nu+2)}(.)$ imply that:

$$|\hat{m}_{n,b_n,\nu}^{(2)}(\hat{\theta}_{n,b_n,\nu}) - m^{(\nu+2)}(\theta_\nu)|$$
$$\leq |\hat{m}_{n,b_n,\nu}^{(2)}(\hat{\theta}_{n,b_n,\nu}) - m^{(2)}(\hat{\theta}_{n,b_n,\nu})| + |m^{(\nu+2)}(\hat{\theta}_{n,b_n,\nu}) - m^{(\nu+2)}(\theta_\nu)| \xrightarrow{n\to\infty} 0 \quad \text{a.s.} \qquad (2.29)$$

and, therefore, $\hat{m}_{n,b_n,\nu}^{(2)}(\hat{\theta}_{n,b_n,\nu}) \neq 0$ for n large enough.

Observing that, by our assumptions, $\hat{m}_{n,b_n,\nu}'(\hat{\theta}_{n,b_n,\nu}) = m^{(\nu+1)}(\theta_\nu)$, we obtain by Taylor expansion:

$$m^{(\nu+1)}(\theta_\nu) = \hat{m}_{n,b_n,\nu}'(\hat{\theta}_{n,b_n,\nu}) = \hat{m}_{n,b_n,\nu}'(\theta_\nu) + (\hat{\theta}_{n,b_n,\nu} - \theta_\nu)\hat{m}_{n,b_n,\nu}^{(2)}(\theta_\nu^*),$$

where θ_ν^* is between $\hat{\theta}_{n,b_n,\nu}$ and θ_ν. Therefore,

$$\hat{\theta}_{n,b_n,\nu} - \theta_\nu$$
$$= \frac{m^{(\nu+1)}(\theta_\nu) - \hat{m}_{n,b_n,\nu}'(\theta_\nu)}{\hat{m}_{n,b_n,\nu}^{(2)}(\theta_\nu^*)}$$
$$= \frac{m^{(\nu+1)}(\theta_\nu) - \hat{m}_{n,b_n,\nu}'(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)}\left(1 + \frac{m^{(\nu+2)}(\theta_\nu) - \hat{m}_{n,b_n,\nu}^{(2)}(\theta_\nu^*)}{\hat{m}_{n,b_n,\nu}^{(2)}(\theta_\nu^*)}\right)$$
$$= \frac{m^{(\nu+1)}(\theta_\nu) - \hat{m}_{n,b_n,\nu}'(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)}(1 + R_n), \qquad (2.30)$$

where, similarly as in (2.29), it is easy to see that $R_n \to 0$ a.s. Hence, the asymptotic distribution of $\hat{\theta}_{n,b_n,\nu} - \theta_\nu$ is the same as the distribution $\{m^{(\nu+1)}(\theta_\nu) - \hat{m}'_{n,b_n,\nu}(\theta_\nu)\}/m^{(\nu+2)}(\theta_\nu)$. Using Lemma 2.1 with the kernel function $K'\nu(.)$ of order $(\nu+1, k+1)$, we obtain:

$$\frac{E\hat{m}'_{n,b_n,\nu}(\theta_\nu) - \hat{m}'_{n,b_n,\nu}(\theta_\nu)}{\{\text{Var}\,\hat{m}'_{n,b_n,\nu}(\theta_\nu)\}^{1/2}m^{(\nu+2)}(\theta_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\{m^{(\nu+2)}(\theta_\nu)\}^2}\right)$$

and, if $b_n \to 0$, $nb_n^2 \to \infty$, and $nb_n^{\nu+1} \to \infty$,

$$n^{1/2}b_n^{\nu+3/2}\frac{m^{(\nu+1)}(\theta_\nu) + b_n^{k-\nu}B_{k+1,k+1}m^{(k)}(\theta_\nu) - \hat{m}'_{n,b_n,\nu}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)}\frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2}\right),$$

implying that:

$$n^{1/2}b_n^{\nu+3/2}(\hat{\theta}_{n,b_n,\nu} - \theta_\nu) + n^{1/2}b_n^{k+3/2}\frac{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)} \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)}\frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2}\right).$$

<div align="right">□</div>

2.4   Mean squared error and bandwidth

Looking at the mean and the variance of the asymptotic distribution of the empirical zero given in Theorem 2.1 and replacing d by nb_n^{2k+1}, it is easy to express the Mean Squared Error:

$$\begin{aligned}
MSE(\hat{\xi}_{n,b_n,\nu}) &= \text{bias}^2 + \text{variance} \\
&= \frac{nb_n^{2k+1}}{nb_n^{2\nu+1}}\left(\frac{B_{k,k}m^{(k)}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)}\right)^2 + \frac{1}{nb_n^{2\nu+1}}\frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)}\frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2} \\
&= b_n^{2k-2\nu}\left(\frac{B_{k,k}m^{(k)}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)}\right)^2 + \frac{1}{nb_n^{2\nu+1}}\frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)}\frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2}.
\end{aligned}$$

Assuming that the bias term is not equal to zero, i.e., assuming that the k-th derivative of the regression function $m^{(k)}(\xi_\nu) \neq 0$, we may calculate the bandwidth that minimizes the MSE. Setting the derivative of $MSE(\hat{\xi}_{n,b_n,\nu})$ with respect to b_n equal to zero, we obtain:

$$\begin{aligned}
0 &= (2k-2\nu)b_n^{2k-2\nu-1}\left(\frac{B_{k,k}m^{(k)}(\xi_\nu)}{m^{(\nu+1)}(\xi_\nu)}\right)^2 - (2\nu+1)\frac{1}{nb_n^{2\nu+2}}\frac{\sigma^2(\xi_\nu)}{f(\xi_\nu)}\frac{V}{\{m^{(\nu+1)}(\xi_\nu)\}^2} \\
0 &= (2k-2\nu)b_n^{2k+1}\{B_{k,k}m^{(k)}(\xi_\nu)\}^2 - (2\nu+1)\frac{1}{n}\frac{\sigma^2(\xi_\nu)V}{f(\xi_\nu)}.
\end{aligned}$$

It is easy to see that the second derivative of $MSE(\hat{\xi}_{n,b_n,\nu})$ is positive because it is a sum of two positive terms. Hence, we obtain that:

$$b_{0,n} = n^{-1/(2k+1)}\left[\frac{2\nu+1}{2k-2\nu}\frac{\sigma^2(\xi_\nu)V}{f(\xi_\nu)\{B_{k,k}m^{(k)}(\xi_\nu)\}^2}\right]^{1/(2k+1)} = O(n^{-1/(2k+1)}) \qquad (2.31)$$

is the optimal bandwidth for the estimation of the zero.

The bandwidth parameter does not necessarily have to be constant and we may consider also *local bandwidths* depending on the known density of design points, $f(.)$, and possibly also on a variance function known up to a multiplicative constant, i.e., $\sigma^2(.) = \sigma^2 w(.)$. In principle, we could also consider dependency of the bandwidth on the term $m^{(\nu+1)}(.)$ but, in practice, the derivatives of the regression function are rarely known and, therefore, we simplify the problem by assuming that $m^{(\nu+1)}(.) = m_1$ is an unknown constant. Looking at (2.31), we define the local bandwidth:

$$b_{0,n,f}(x) \propto \left(\frac{1}{nf(x)}\right)^{1/(2k+1)} \qquad (2.32)$$

in a homoscedastic situation ($\sigma^2(.) = \sigma^2$) and

$$b_{0,n,f,w}(x) \propto \left(\frac{w(x)}{nf(x)} \right)^{1/(2k+1)}, \tag{2.33}$$

in a heteroscedastic setup (assuming that $\sigma^2(.) = \sigma^2 w(.)$).

The MSE of the empirical location of maximum may be derived similarly. From Theorem 2.2, we obtain:

$$
\begin{aligned}
MSE(\hat{\theta}_{n,b_n,\nu}) &= \frac{nb_n^{2k+3}}{nb_n^{2\nu+3}} \left(\frac{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)} \right)^2 + \frac{1}{nb_n^{2\nu+3}} \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)} \frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2} \\
&= b_n^{2k-2\nu} \left(\frac{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)} \right)^2 + \frac{1}{nb_n^{2\nu+3}} \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)} \frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2}.
\end{aligned}
$$

Assuming that $m^{(k+1)}(\theta_\nu) \neq 0$ and setting the derivative of the MSE equal to zero, we get:

$$
\begin{aligned}
0 &= (2k-2\nu)b_n^{2k-2\nu-1} \left(\frac{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)}{m^{(\nu+2)}(\theta_\nu)} \right)^2 - (2\nu+3)\frac{1}{nb_n^{2\nu+4}} \frac{\sigma^2(\theta_\nu)}{f(\theta_\nu)} \frac{V'}{\{m^{(\nu+2)}(\theta_\nu)\}^2} \\
0 &= (2k-2\nu)b_n^{2k+3}\{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)\}^2 - (2\nu+3)\frac{1}{n}\frac{\sigma^2(\theta_\nu)V'}{f(\theta_\nu)}
\end{aligned}
$$

implying that the optimal bandwidth for the estimation of the location of maximum is:

$$b_{e,n} = n^{-1/(2k+3)} \left[\frac{2\nu+3}{2k-2\nu} \frac{\sigma^2(\theta_\nu)V'}{f(\theta_\nu)\{B_{k+1,k+1}m^{(k+1)}(\theta_\nu)\}^2} \right]^{1/(2k+3)} = O(n^{-1/(2k+3)}) \tag{2.34}$$

since the second derivative of $MSE(\hat{\theta}_{n,b_n,\nu})$ is again just a sum of two positive terms.

Similarly as in (2.32) and (2.33), we define local bandwidths:

$$b_{e,n,f}(x) \propto \left(\frac{1}{nf(x)} \right)^{1/(2k+3)} \tag{2.35}$$

and

$$b_{e,n,f,w}(x) \propto \left(\frac{w(x)}{nf(x)} \right)^{1/(2k+3)} \tag{2.36}$$

depending on the design $f(.)$ and on the variance function $w(.)$ if $\sigma^2(.) = \sigma^2 w(.)$.

2.5 Optimal distribution of design points

The problem of finding the optimal distribution of design points in nonparametric kernel regression has been previously addressed in Müller (1984b) from the point of view of the *integrated mean squared error* (IMSE) of the GM kernel regression estimator $\hat{m}(x)$. Choosing a probability measure H with a positive and continuous density $h(.)$ on $\langle 0, 1 \rangle$ and considering:

$$IMSE = E \int \{\hat{m}(x) - m(x)\} dH(x) \approx \frac{1}{nb_n} \int K^2(s) ds \int \frac{h(x)}{f(x)} dx,$$

the AIMSE (asymptotic IMSE) optimal density of the design points $f_X^*(x) = h(x)^{1/2}/\int h(u)^{1/2}du \propto h(x)^{1/2}$ has been derived in Müller (1984b).

Unfortunately, the probability measure H lacks any clear interpretation and, therefore, the AIMSE optimal design is not well applicable in practice. Similarly as in Hlávka (2011), we overcome this obstacle by obtaining designs minimizing the variability of the empirical zero and of the empirical location of maximum.

Let the symbol A denote a probability measure describing the prior distribution of the zero or the location of maximum and let us assume that A has a positive and continuous density $a(.)$ such that:

A8. **Prior location density:** There exists $\delta > 0$ such that $a(x) > \delta$, for all $x \in \langle 0, 1 \rangle$.

The optimal density of the design points in a heteroscedastic situation is investigated in Section 2.5.1. The solution in a homoscedastic situation is given in Section 2.5.2.

2.5.1   Heteroscedastic random errors

Often, the precision of the measurements depends on the location and the shape of the variance function is known in advance. In this section, we assume that $\sigma^2(.) = \sigma^2 w(.)$, where $w(.) > 0$ is a known function and $\sigma^2 > 0$ an unknown constant, and we describe the distribution of design points minimizing the variability of nonparametric estimators of the zero and the location of maximum in Theorems 2.3 and 2.4.

Theorem 2.3. *Assume that the assumptions of Theorem 2.1 and A8 hold, $\sigma^2(.) = \sigma^2 w(.)$, where $w(.)$ is a known function and $0 < \sigma^2 < \infty$, and that $m^{(\nu+1)}(\xi_\nu) = m_1$ does not depend on the value of ξ_ν.*

1. *Assuming that the product $w(.)a(.)$ satisfies assumption A4, the density of design points $f_{V,w}(x) \propto \{w(x)a(x)\}^{1/2}$ minimizes the expectation of the asymptotic variance of the empirical zero, i.e., $\int Var(\hat{\xi}_{n,b_n,\nu}|\xi_\nu = u)a(u)du$, with respect to the prior density $a(.)$.*

2. *Assuming that $\{w(.)\}^{2/3}\{a(.)\}^{4/3}$ satisfies assumption A4, the density of design points $f_{L,w}(x) \propto \{w(x)\}^{1/3}\{a(x)\}^{2/3}$ minimizes the expected length of confidence intervals with respect to the prior density $a(.)$.*

Proof. We prove only the first part because the proof of the second part is very similar. At first, we recall that for a density of design points $f_X(.)$ satisfying A4 it follows from Theorem 2.1 that:

$$Var(\hat{\xi}|\xi = x) = cw(x)f_X^{-1}(x),$$

where c is a constant depending on n, b_n, $K(.)$, σ^2, and m_1 and we notice that the resulting minimization problem:

$$f_V = \arg\min_{f_X} \int_0^1 Var(\hat{\xi}_{n,b_n,\nu}|\xi_n = x)a(x)dx = \arg\min_{f_X} \int_0^1 f_X^{-1}(x)w(x)a(x)dx$$

belongs to the classical calculus of variations.

Denoting $F(x,y,y') = F(x,F_X,f_x) = f_X^{-1}(x)w(x)a(x)$, the necessary condition for an extreme of $I(f_X) = I(y') = \int_0^1 F(x,y,y')dx$ is $F_y' - \frac{d}{dx}F_{y'}' = 0$, see e.g. Nožička (1998); Smirnow (1958). In our setup, $F_y' = 0$ and $F_{y'}'(x) = -f_X^{-2}(x)w(x)a(x)$ and the above condition thus implies that the optimal density of design points $f_V(.)$ has to satisfy $\frac{d}{dx}\{f_V^{-2}(x)w(x)a(x)\} = 0$, i.e., $f_V^{-2}(x)w(x)a(x) = \text{constant}$.

Next, let $f_{V,w}^\star(x) \propto w^{1/2}(x)a^{1/2}(x)$ denote the candidate solution. Recall that our assumptions imply that $w(.) > \delta_\sigma/\sigma^2 > 0$ and $\{w(.)a(.)\}^{1/2} > (\delta_w\delta_\sigma)^{1/2}/\sigma$. We obtain:

$$w^{1/2}(x_1)a^{1/2}(x_1) - w^{1/2}(x_2)a^{1/2}(x_2)$$
$$< \{w^{1/2}(x_1)a^{1/2}(x_1) - w^{1/2}(x_2)a^{1/2}(x_2)\}\sigma\frac{w^{1/2}(x_1)a^{1/2}(x_1) + w^{1/2}(x_2)a^{1/2}(x_2)}{(\delta_w\delta)^{1/2}}$$
$$= \sigma\{w(x_1)a(x_1) - w(x_2)a(x_2)\}/(\delta_w\delta)^{1/2}.$$

Hence, the Lipschitz continuity of the function $w(.)a(.)$ (see assumption A4) implies that also the candidate solution $f_{V,w}^\star(.)$ is Lipschitz continuous and satisfies assumption A4.

It remains to verify that the candidate solution $f_{V,w}^\star(.)$ minimizes the expected variance. Considering another probability density functions f_Y and $f_Z = \alpha f_Y + (1-\alpha)f_{V,w}^\star$ for $\alpha \in \langle 0,1\rangle$ and defining the integration constant $k = \{\int_0^1 w^{1/2}(u)a^{1/2}(u)du\}^{-1}$ and a function:

$$Z(\alpha) = \int_0^1 f_Z^{-1}(x)w(x)a(x)dx = \int_0^1 \frac{1}{\alpha\{f_Y(x) - kw^{1/2}(x)a^{1/2}(x)\} + kw^{1/2}(x)a^{1/2}(x)}w(x)a(x)dx,$$

it is easy to verify that the function $Z(\alpha)$ is continuously differentiable, $Z'(0) = 0$ and, if f_Y and $f_{V,w}^\star$ are not equal A-a.e., $Z^{(2)}(\alpha) > 0$, for $\alpha \in \langle 0,1\rangle$. This implies that $Z'(\alpha) > 0$ for $\alpha \in (0,1\rangle$ and, therefore, $I(f_Y) = \int f_Y^{-1}(x)a(x)dx = Z(1) > Z(0) = \int\{f_{V,w}^\star(x)\}^{-1}a(x)dx = I(f_V^*)$ and the assertion follows. $\square$

The optimal design for the empirical location of maximum $\hat{\theta}_{n,b_n,\nu}$ is established in the following Theorem 2.4.

Theorem 2.4. *Assume that the assumptions of Theorem 2.2 and A8 hold, $0 < \sigma^2 < \infty$, and that $m^{(\nu+2)}(\theta_\nu) = m_2$ does not depend on the location of maximum θ_ν.*

1. *Assuming that the product $w(.)a(.)$ satisfies assumption A4, the density of design points $f_V(x) \propto \{w(x)a(x)\}^{1/2}$ minimizes the expectation of the asymptotic variance of the empirical location of maximum, $\int Var(\hat{\theta}_{n,b_n,\nu}|\theta_\nu = u)a(u)du$, with respect to the prior density $a(.)$.*

2. *Assuming that $\{w(.)\}^{2/3}\{a(.)\}^{4/3}$ satisfies assumption A4, the density of design points $f_L(x) \propto \{w(x)\}^{1/3}\{a(x)\}^{2/3}$ minimizes the expected length of the confidence interval for the true location of maximum with respect to the prior density $a(.)$.*

Proof. The proof proceeds exactly as the proof of Theorem 2.3. $\qquad\square$

2.5.2 Homoscedastic random errors

In practice, the variance of the random errors is often constant, i.e., $\sigma^2(x) = \sigma^2$. The optimal distribution of the design points in this situation is a simple corollary of Theorems 2.3 and 2.4.

Corollary 2. *Under the assumptions of Theorem 2.3 with $w(.) \equiv 1$, the density of design points $f_V(x) \propto a^{1/2}(x)$ minimizes the expectation of the asymptotic variance of the empirical zero and the density of design points $f_L(x) \propto a^{2/3}(x)$ minimizes the expected length of confidence intervals for the true zero.*

Proof. The assertion follows immediately from Theorem 2.3 with $w(.) \equiv 1$. $\qquad\square$

Corollary 3. *Under the assumptions of Theorem 2.4 with $w(.) \equiv 1$, the density of design points $f_V(x) \propto a^{1/2}(x)$ minimizes the expectation of the asymptotic variance of the empirical location of maximum and the density of design points $f_L(x) \propto a^{2/3}(x)$ minimizes the expected length of confidence intervals for the location of maximum.*

Proof. The assertion follows immediately from Theorem 2.4 with $w(.) \equiv 1$. $\qquad\square$

2.5.3 Local bandwidth

In this section, we establish the optimal distribution of design points in a heteroscedastic situation using the local bandwidth derived in Section 2.4. The optimal design for the empirical zero is given in the following Theorem 2.5.

Theorem 2.5. *Assume that the assumptions of Theorem 2.1 and A8 hold, $\sigma^2(x) = \sigma^2 w(x)$, where $w(x)$ is a known function and $0 < \sigma^2 < \infty$, $m^{(\nu+1)}(\xi_\nu) = m_1$ does not depend on the value of the true zero ξ_ν, and that the local bandwidth $b_{0,n,f,w}(.)$ is given by (2.33).*

1. *Assuming that the product $\{w(x)\}^{(4k-4\nu)/(4k-2\nu+1)}\{a(x)\}^{(4k+2)/(4k-2\nu+1)}$ satisfies A4, the density of design points $f_{0,V,w,l}(x) \propto \{w(x)\}^{(2k-2\nu)/(4k-2\nu+1)}\{a(x)\}^{(2k+1)/(4k-2\nu+1)}$ minimizes the expectation of the asymptotic variance of the empirical zero, $\int Var(\hat{\xi}_{n,b_n,\nu}|\xi_\nu = u)a(u)du$, with respect to the prior density $a(.)$.*

2. *Assuming that $\{w(x)\}^{(4k-4\nu)/(3k-\nu+1)}\{a(x)\}^{(4k+2)/(3k-\nu+1)}$ satisfies assumption A4, the density of design points $f_{0,L,w,l}(x) \propto \{w(x)\}^{(k-\nu)/(3k-\nu+1)}\{a(x)\}^{(2k+1)/(3k-\nu+1)}$ minimizes the expected length of confidence intervals for the true zero with respect to the prior density $a(.)$.*

Proof. Plugging the local bandwidth $b_{0,n,f,w}(.)$ into the asymptotic variance of the estimator provided by Theorem 2.1, we obtain that:

$$Var(\hat{\xi}_{n,b_n,\nu}) \propto \frac{w(\xi_\nu)}{f(\xi_\nu)} \frac{1}{\{b_{0,n,f,w}(\xi_\nu)\}^{2\nu+1}} \propto \frac{w(\xi_\nu)}{f(\xi_\nu)} \left\{\frac{f(\xi_\nu)}{w(\xi_\nu)}\right\}^{(2\nu+1)/(2k+1)} = \left\{\frac{w(\xi_\nu)}{f(\xi_\nu)}\right\}^{(2k-2\nu)/(2k+1)}.$$

Proceeding similarly as in the proof of Theorem 2.3, we solve the minimization problem:

$$\begin{aligned}
f_{0,V,w,l} &= \arg\min_{f_X} \int_0^1 Var(\hat{\xi}_{n,b_{0,n,f,w}(x),\nu}|\xi_\nu = x)a(x)dx \\
&= \arg\min_{f_X} \int_0^1 \{f(x)\}^{-(2k-2\nu)/(2k+1)}\{w(x)\}^{(2k-2\nu)/(2k+1)}a(x)dx
\end{aligned}$$

and we obtain that the solution has to satisfy:

$$\{f_{0,V,w,l}(x)\}^{-(4k-2\nu+1)/(2k+1)}\{w(x)\}^{(2k-2\nu)/(2k+1)}a(x) = \text{constant},$$

i.e.,

$$f_{0,V,w,l}(x) \propto \{w(x)\}^{(2k-2\nu)/(4k-2\nu+1)}\{a(x)\}^{(2k+1)/(4k-2\nu+1)}. \tag{2.37}$$

Concerning the second part of the theorem, we solve the minimization problem:

$$f_{0,L,w,l} = \arg\min_{f_X} \int_0^1 \{f(x)\}^{-(k-\nu)/(2k+1)}\{w(x)\}^{(k-\nu)/(2k+1)}a(x)dx$$

leading in the same way that:

$$\{f_{L,w,l}(x)\}^{-(3k-\nu+1)/(2k+1)}\{w(x)\}^{(k-\nu)/(2k+1)}a(x) = \text{constant}$$

and

$$f_{0,L,w,l}(x) \propto \{w(x)\}^{(k-\nu)/(3k-\nu+1)}\{a(x)\}^{(2k+1)/(3k-\nu+1)}. \tag{2.38}$$

The proof may now be finished in the same way as the proof of Theorem 2.3. □

The optimal design for the empirical location of maximum in a heteroscedastic situation with local bandwidth is given in the following Theorem 2.6.

Theorem 2.6. *Assume that the assumptions of Theorem 2.2 and A8 hold, $0 < \sigma^2 < \infty$, $m^{(\nu+2)}(\theta_\nu) = m_2$ does not depend on the location of maximum θ_ν, and the local bandwidth $b_{e,f,w}(.)$ is given by (2.36).*

1. *Assuming that $\{w(x)\}^{(4k-4\nu)/(4k-2\nu+3)}\{a(x)\}^{(4k+6)/(4k-2\nu+3)}$ satisfies A4, the density of design points $f_{e,V,w,l}(x) \propto \{w(x)\}^{(2k-2\nu)/(4k-2\nu+3)}\{a(x)\}^{(2k+3)/(4k-2\nu+3)}$ minimizes the expectation of the asymptotic variance of the empirical location of maximum, $\int Var(\hat{\theta}_{n,b_n,\nu}|\theta_\nu = u)a(u)du$, with respect to the prior density $a(.)$.*

2. *Assuming that $\{w(x)\}^{(4k-4\nu)/(3k-\nu+3)}\{a(x)\}^{(4k+6)/(3k-\nu+6)}$ satisfies assumption A4, the density of design points $f_{e,L,w,l}(x) \propto \{w(x)\}^{(k-\nu)/(3k-\nu+3)}\{a(x)\}^{(2k+3)/(3k-\nu+3)}$ minimizes the expected length of confidence intervals for the true location of maximum with respect to the prior density $a(.)$.*

Proof. The proof proceeds similarly as the proof of Theorem 2.5. Plugging the local bandwidth $b_{e,n,f,w}(.)$ given in (2.36) into the asymptotic variance of the estimator provided by Theorem 2.1, we obtain that:

$$Var(\hat{\theta}_{n,b_n,\nu}) \propto \frac{w(\theta_\nu)}{f(\theta_\nu)}\frac{1}{\{b_{0,n,f,w}(\theta_\nu)\}^{2\nu+3}} \propto \frac{w(\theta_\nu)}{f(\theta_\nu)}\left\{\frac{f(\theta_\nu)}{w(\theta_\nu)}\right\}^{(2\nu+3)/(2k+3)} = \left\{\frac{w(\theta_\nu)}{f(\theta_\nu)}\right\}^{(2k-2\nu)/(2k+3)}.$$

Once again, similarly as in the proof of Theorems 2.3 and 2.5, we solve the minimization problem:

$$\begin{aligned}
f_{e,V,w,l} &= \arg\min_{f_X} \int_0^1 Var(\hat{\xi}_{n,b_{0,n,f,w}(x),\nu}|\xi_\nu = x)a(x)dx \\
&= \arg\min_{f_X} \int_0^1 \{f(x)\}^{-(2k-2\nu)/(2k+3)}\{w(x)\}^{(2k-2\nu)/(2k+3)}a(x)dx
\end{aligned}$$

and we obtain that the solution has to satisfy:

$$\{f_{e,V,w,l}(x)\}^{-(4k-2\nu+3)/(2k+3)}\{w(x)\}^{(2k-2\nu)/(2k+3)}a(x) = \text{constant},$$

i.e.,

$$f_{e,V,w,l}(x) \propto \{w(x)\}^{(2k-2\nu)/(4k-2\nu+3)}\{a(x)\}^{(2k+3)/(4k-2\nu+3)}. \tag{2.39}$$

In the second part of the theorem, we solve the minimization problem:

$$f_{e,L,w,l} = \arg\min_{f_X} \int_0^1 \{f(x)\}^{-(k-\nu)/(2k+3)}\{w(x)\}^{(k-\nu)/(2k+3)}a(x)dx$$

leading in the same way that:

$$\{f_{e,L,w,l}(x)\}^{-(3k-\nu+3)/(2k+3)}\{w(x)\}^{(k-\nu)/(2k+3)}a(x) = \text{constant}$$

and

$$f_{e,L,w,l}(x) \propto \{w(x)\}^{(k-\nu)/(3k-\nu+3)}\{a(x)\}^{(2k+3)/(3k-\nu+3)}. \tag{2.40}$$

The proof may be now finished similarly as the proof of Theorem 2.3. □

2.6 Simulation study

In this section, we investigate the finite sample properties of the method in a short simulation study. The simulation study was implemented in the statistical computing environment R (R Development Core Team; 2011). All simulation results are based on the Gasser–Müller (GM) kernel regression estimator using the quartic kernel and 1000 simulations.

	constant bandwidth				local bandwidth			
	$m(.)$		$m'(.)$		$m(.)$		$m'(.)$	
	zero	max.	zero	max.	zero	max.	zero	max.
$r_{opt}(MSE)$	1/2	1/2	1/2	1/2	$5/9 \doteq 0.56$	$7/11 \doteq 0.63$	$7/11 \doteq 0.63$	$9/13 \doteq 0.69$
$r_{opt}(MAD)$	2/3	2/3	2/3	2/3	$5/7 \doteq 0.71$	$7/9 \doteq 0.78$	$7/9 \doteq 0.78$	$9/11 \doteq 0.82$

Table 2.1: Powers of the prior density defining the optimal experiment design for all estimators related to the regression function and its first derivative.

In Table 2.1, we summarize the optimal nonparametric regression designs concerning the regression function and its first derivative under assumptions of homoscedasticity, see Corollaries 2 and 3 and Theorems 2.5 and 2.6. It is interesting that the design is "more concentrated" for MAD and for local bandwidth. As one might expect, the optimal design for estimation of the location of maximum of a regression function is exactly the same as the optimal design for estimation of a zero of its derivative.

2.6.1 Zeros

We start by investigating the optimality properties of the experimental design proposed for estimation of zeros.

Linear regression function

As a most simple example, let us first investigate a linear function:

$$m_1(x) = 4(x - \theta),$$

where the design points x_i are uniformly distributed in $\langle 0, 1 \rangle$ and we set $n = 50$, $\sigma = 0.5$, and the bandwidth $b_n = 50$. Both the true regression line and the resulting GM estimator calculated using the quartic kernel, i.e.,

$$K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1),$$

with bandwidth $b_n = 0.07$ are plotted in Figure 2.1.

Apart of the estimators, it is very useful to display also the precision of the estimator. The standard approach is to plot asymptotic confidence intervals derived from the asymptotic distribution given in Theorem 2.1. For the linear regression function $m_1(.)$, the bias term in Theorem 2.1 is clearly equal to zero because the quartic kernel is a kernel function of order $(0, 2)$ and the second derivative $m_1^{(2)}(\xi_0) = 0$ implying that:

$$P\left\{ \left(\hat{\xi}_{n.b_n,0} \pm u_{1-\alpha/2} \frac{\sigma(\xi_0)}{\{f(\xi_0)\}^{1/2}} \frac{V^{1/2}}{m'(\xi_0)} \right) \ni \xi_0 \right\} = 0.95.$$

The resulting 95% confidence interval $\left(\hat{\xi}_{n.b_n,0} \pm u_{0.975} \frac{\sigma(\xi_0)}{\{f(\xi_0)\}^{1/2}} \frac{V^{1/2}}{m'(\xi_0)} \right)$ obtained for the bandwidth $b_n = 0.07$ is plotted in Figure 2.2.

In Figure 2.3, we display both the estimator and the corresponding 95% confidence intervals in order to illustrate the link between the bandwidth and the length of the confidence interval. Figure 2.3 suggests that longer bandwidths lead shorter 95% confidence intervals without decreasing the coverage probability. However, confidence intervals based on oversmoothed regression estimators should be used with caution and only if the bias term in Theorem 2.1 is equal to zero (i.e., when the second derivative of the regression function is zero). In the last plot in Figure 2.3, the estimator already suffers from boundary effects increasing the bias also in the empirical zero.
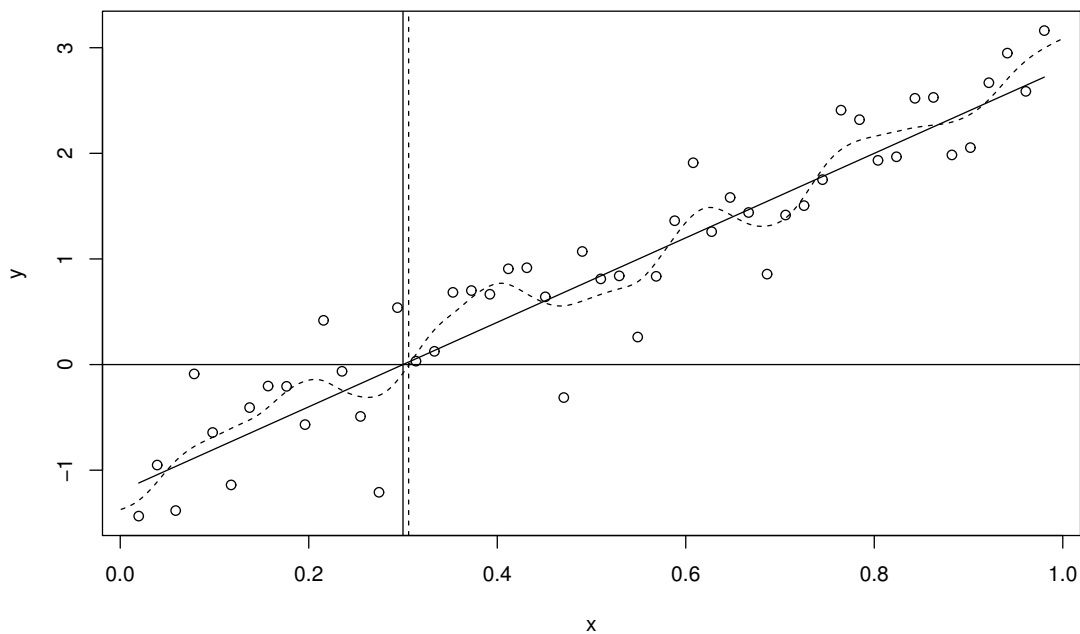
Figure 2.1: Simulated example: empirical zero for a linear regression function $m_1(.)$. The dashed lines denote the empirical zero. The full lines denote the true regression line and the true zero.
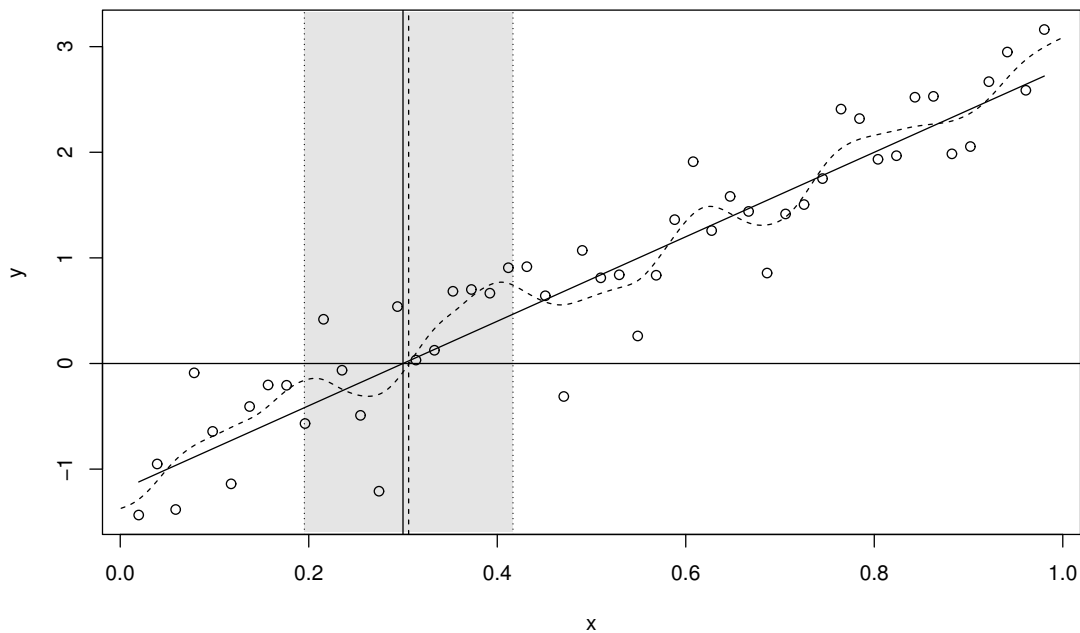
Figure 2.2: Simulated example: asymptotic 95% confidence intervals for the true zero, regression function $m_1(.)$, bandwidth $b_n = 0.07$.
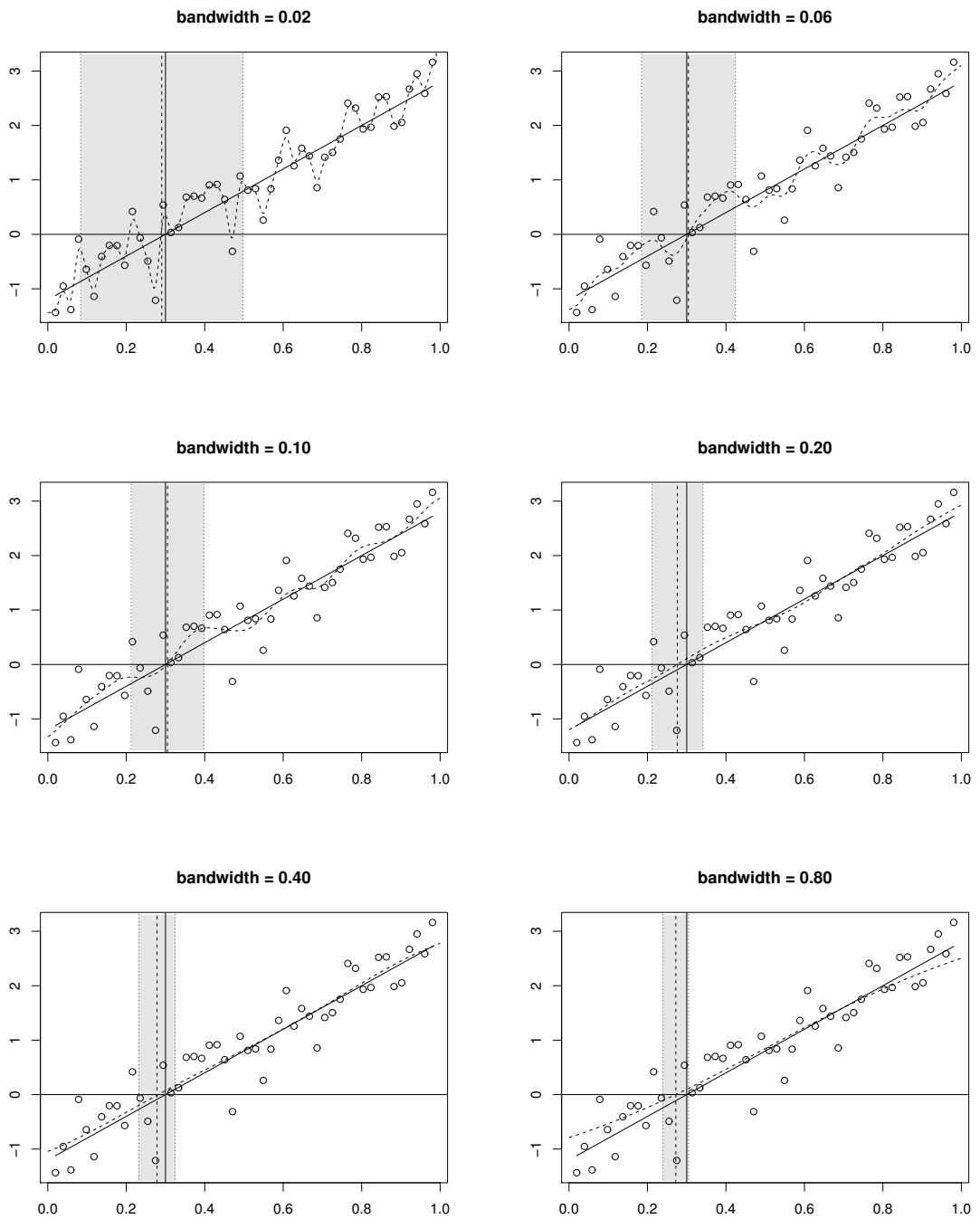
Figure 2.3: Simulated example: asymptotic 95% confidence intervals for the zero of $m_1(.)$ calculated for several bandwidths.

Nonlinear regression function

In order to investigate the effect of the design in a more interesting situation, we use the following
nonlinear regression function:

$$m(x) = \frac{(x + 0.05)^2 - (\xi_0 + 0.05)^2}{(x + 0.05)}.$$

Note that the second derivative $m^{(2)}(\xi_0) \neq 0$ and that the first derivative $m^{(1)}(\xi_0) = 2$ does not
depend on ξ_0.

Bandwidth parameter   Preliminary simulation study, not included in this report, showed some
rather unclear behavior of the estimators when constant bandwidth was used. Therefore, we use only
the local bandwidth, $b_{0,n,f}(.)$, defined in (2.32). For simplicity, we rewrite the local bandwidth as
$b_n(x) = b\{f(x)\}^{-1/(2k+1)}$ and b is the only bandwidth parameter used in the remaining part of this
section.

Prior distribution of zero   Similarly as in Hlávka (2011), we choose the prior distribution of the
zero as a mixture of Uniform, $U(0,1)$, and Normal distribution, $N(\mu_\theta, \sigma_\theta^2)$, restricted to the interval
$\langle 0, 1 \rangle$. More precisely, the prior density of the zero in our simulation study is:

$$a(\theta) \propto (1 - p)\phi_{\mu_\theta, \sigma_\theta^2}(\theta | \theta \in \langle 0, 1 \rangle) + p,$$

where $\phi(.|\langle 0, 1 \rangle)$ denotes the density of a $N(\mu_\theta, \sigma_\theta^2)$ distribution restricted to the interval $\langle 0, 1 \rangle$. In
the following, we set $\mu_\theta = 0.4$, $\sigma_\theta^2 = 0.01$, and $p = 0.1$.

Design density   The density of the design points is controlled by a parameter r such that for a
fixed value of r, the density of the design points, $f_{X,r}(.)$, is proportional to the r-th power of the prior
density $a(.)$, i.e., $f_{X,r}(x) \propto a^r(x)$. An example of one step of the simulation study including the prior
density and the design density with $r = 1/2$ is plotted in Figure 2.4. The meaning of the parameter
r is very simple. For example, the value $r = 0$ corresponds to uniformly distributed design points,
$f_{X,0}(x) = I(x \in \langle 0, 1 \rangle)$. The value $r = 1$ would mean that the density of design points is equal to
the prior density of the location of maximum, i.e., $f_{X,1}(x) = a(x)$. Higher values of the parameter
r mean that the design points are more concentrated in the neighborhood of the mode of the prior
distribution $a(\theta)$, see Figure 2.5 for an illustration.

Setup of the simulation study   In each step of the simulation, for a fixed sample size n, the
standard deviation $\sigma \in \{0.1, 0.5\}$, the bandwidth parameter $b \in \langle 0.005, 0.5 \rangle$, and the parameter
controlling the density of design points $r \in \langle 0, 1.2 \rangle$, we:

1. calculate the design points according to the density $f_{X,r}(x) \propto a^r(x)$,

2. simulate the responses $Y_i = m(x_i) + \sigma \varepsilon_i$, for $i = 1, \ldots, n$, where ε_i are iid $N(0, 1)$ pseudo-random
   variables and $m(.)$ is the regression function with zero drawn from the prior distribution,

3. calculate the empirical zero using the GM estimator with bandwidth parameter b and the func-
   tion `uniroot()` in the statistical computing environment R (R Development Core Team; 2011).

Optimality criteria   For each sample size n, standard deviation σ, the bandwidth parameter b, and
each value of the parameter r, we calculate the Mean Squared Error (MSE) and the Mean Absolute
Deviation (MAD) of the empirical zero from 1000 simulations. In all tables, the MSE and MAD are
presented only for the best value of the parameter r (denoted as r_{opt}) for each bandwidth and the
bandwidths with the smallest MSE or MAD are denoted by the symbol $\star$. For example, in Table 2.2,
for $n = 20$ observations and $\sigma = 0.1$, the MSE is minimized for local bandwidth parameter $b = 0.2$
(i.e., for the local bandwidth $b_n(.) = 0.2\{f(.)\}^{-1/(2k+1)}$) and the density of design points proportional
to $a^{0.6}(.)$ (since $r_{opt} = 0.6$).
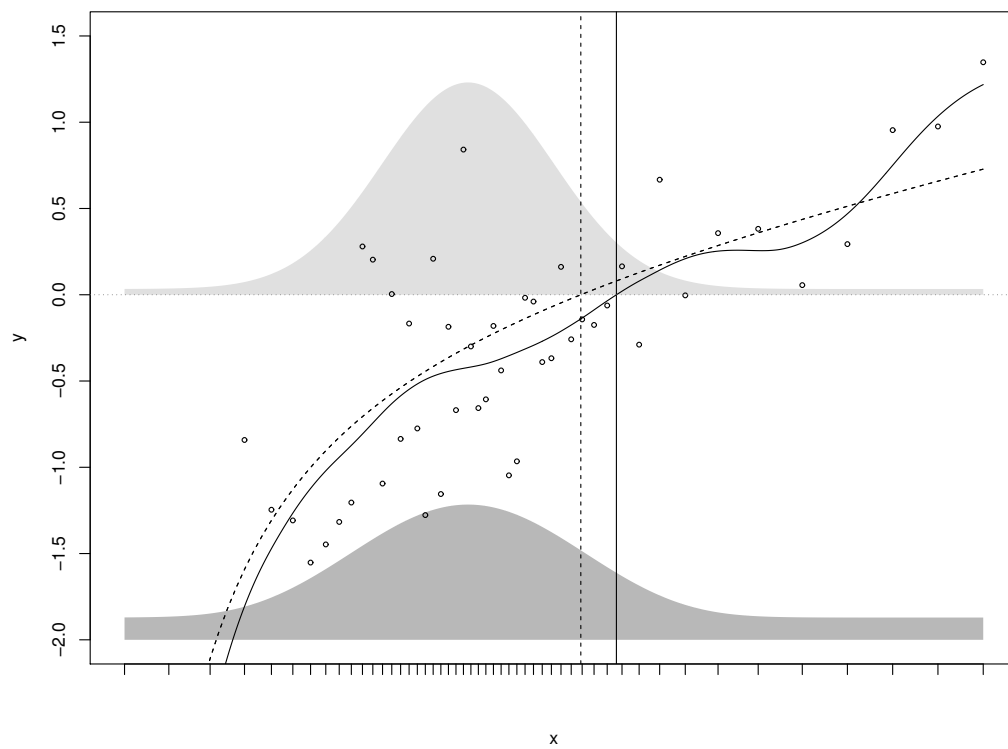
Figure 2.4: An example of a prior density and the corresponding design density with $r = 0.5$. The vertical dashed line denotes the true zero generated from the prior distribution and the corresponding regression function $m(.)$. The full vertical line denotes the estimator.

Figure 2.5: Vertical lines denote the designs points for $n = 50$ defined by three design densities $f_{X,r} \propto a^2(x)$ with $r \in \{0, 1/2, 1\}$. The dots denote the mode and symmetrized 95% prediction interval for the prior distribution of the true zero. The three curves denote the corresponding regression lines.

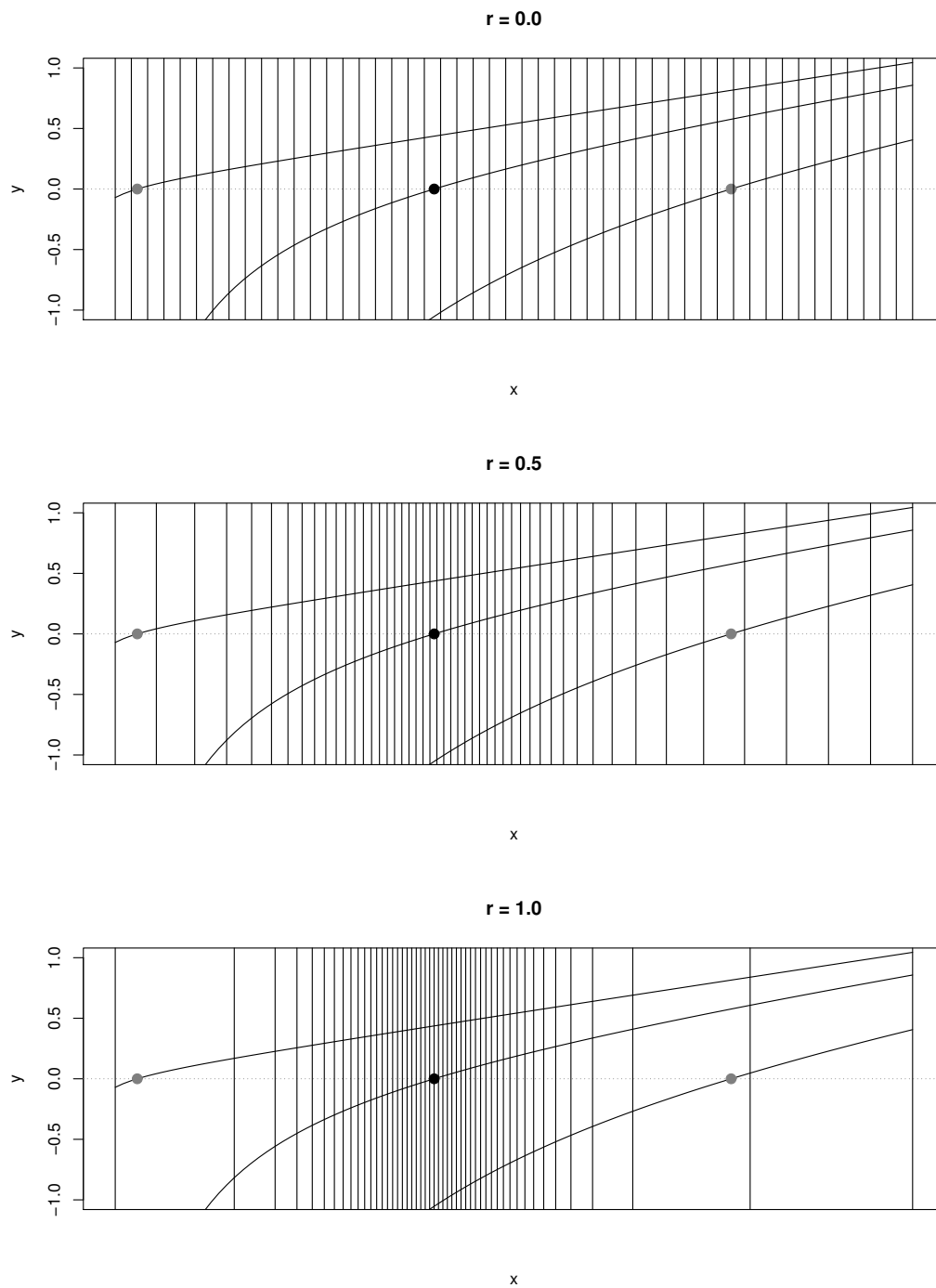| $n = 20$	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
b	$\sigma = 0.1$		$\sigma = 0.5$	
0.06	0.650 (2.1868)	0.600 (0.0814)	0.600 (9.3079)	0.550 (1.6594)
0.08	0.650 (1.9761)	0.600 (0.0664)	0.750 (9.5271)	0.950 (1.8105)
0.10	0.650 (1.8203)	0.600 (0.0564)	0.950 (9.4956)	0.050 (1.8109)
0.15	0.650 (1.5236)	0.600 (0.0401)	0.350 (8.5319)	0.250 (1.2938)
0.20	0.650 (1.4215)$\star$	0.600 (0.0349)$\star$	0.250 (7.4125)	0.250 (0.9347)
0.30	0.650 (1.6826)	0.350 (0.0460)	0.250 (6.0326)	0.250 (0.6270)
0.40	0.650 (2.5834)	0.250 (0.0936)	0.250 (5.3373)$\star$	0.250 (0.4951)$\star$
0.50	0.650 (4.1449)	0.650 (0.2094)	0.250 (5.7023)	0.250 (0.5335)
$n = 50$	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
b	$\sigma = 0.1$		$\sigma = 0.5$	
0.06	0.650 (1.4830)	0.750 (0.0372)	0.100 (7.0341)	0.100 (0.9000)
0.08	0.650 (1.3109)	0.650 (0.0294)	0.250 (6.3327)	0.150 (0.7083)
0.10	0.750 (1.1854)	0.750 (0.0241)	0.250 (5.9188)	0.100 (0.6170)
0.15	0.750 (1.0088)	0.750 (0.0179)	0.250 (5.1974)	0.250 (0.4587)
0.20	0.650 (1.0024)$\star$	0.750 (0.0176)$\star$	0.400 (4.5714)	0.400 (0.3618)
0.30	0.850 (1.4056)	0.500 (0.0305)	0.400 (3.9342)$\star$	0.150 (0.2713)
0.40	0.750 (2.3862)	0.750 (0.0758)	0.350 (3.9557)	0.450 (0.2685)$\star$
0.50	0.750 (3.8955)	0.600 (0.1795)	0.450 (4.8090)	0.350 (0.3675)
$n = 200$	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
b	$\sigma = 0.1$		$\sigma = 0.5$	
0.06	0.850 (0.7596)	0.500 (0.0095)	0.500 (3.5015)	0.350 (0.1999)
0.08	0.850 (0.6637)	0.550 (0.0075)	0.500 (3.1637)	0.500 (0.1679)
0.10	0.850 (0.6060)	0.550 (0.0062)	0.500 (2.9274)	0.500 (0.1439)
0.15	0.650 (0.5678)$\star$	0.550 (0.0055)$\star$	0.500 (2.5180)	0.500 (0.1089)
0.20	0.850 (0.6663)	0.550 (0.0075)	0.550 (2.2664)	0.550 (0.0874)
0.30	0.750 (1.2714)	0.550 (0.0221)	0.550 (2.1450)$\star$	0.550 (0.0768)$\star$
0.40	0.750 (2.2988)	0.550 (0.0636)	0.500 (2.7543)	0.500 (0.1187)
0.50	0.750 (3.7597)	0.750 (0.1588)	0.500 (4.0340)	0.500 (0.2165)
$n = 800$	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
b	$\sigma = 0.1$		$\sigma = 0.5$	
0.06	0.550 (0.3969)	0.500 (0.0026)	0.500 (1.8691)	0.500 (0.0563)
0.08	0.500 (0.3492)	0.500 (0.0020)	0.500 (1.6666)	0.500 (0.0453)
0.10	0.500 (0.3298)$\star$	0.500 (0.0019)$\star$	0.500 (1.5161)	0.500 (0.0379)
0.15	0.700 (0.3781)	0.500 (0.0027)	0.750 (1.2935)	0.550 (0.0297)
0.20	0.700 (0.5812)	0.700 (0.0054)	0.700 (1.2205)$\star$	0.550 (0.0271)$\star$
0.30	0.850 (1.2881)	0.700 (0.0210)	0.700 (1.5245)	0.700 (0.0426)
0.40	0.850 (2.3169)	0.700 (0.0640)	0.700 (2.4420)	0.700 (0.0859)
0.50	0.850 (3.7740)	0.750 (0.1611)	0.700 (3.9144)	0.700 (0.1947)
$n = 2500$	$r_{opt}(MAD)$	$r_{opt}(MSE)$	$r_{opt}(MAD)$	$r_{opt}(MSE)$
b	$\sigma = 0.1$		$\sigma = 0.5$	
0.06	0.800 (0.2223)	0.400 (0.0009)	0.950 (1.1019)	0.650 (0.0205)
0.08	0.550 (0.2091)$\star$	0.550 (0.0008)$\star$	0.550 (0.9572)	0.550 (0.0157)
0.10	0.550 (0.2147)	0.550 (0.0009)	0.550 (0.8580)	0.550 (0.0126)
0.15	0.550 (0.3300)	0.550 (0.0018)	0.550 (0.7696)$\star$	0.350 (0.0115)$\star$
0.20	1.000 (0.5678)	0.550 (0.0047)	0.350 (0.8508)	0.350 (0.0136)
0.30	1.000 (1.2477)	0.550 (0.0216)	1.000 (1.3807)	0.450 (0.0319)
0.40	1.000 (2.2324)	1.000 (0.0634)	1.000 (2.3276)	0.450 (0.0829)
0.50	0.900 (3.7032)	0.900 (0.1602)	0.900 (3.8772)	0.450 (0.1962)

Table 2.2: Results of simulations using the regression function $m(.)$ and $p = 0.1$: the powers r_{opt} defining the distribution of design points with the smallest MSE and MAD for various sample sizes n, standard deviations $\sigma = 0.1$ and 0.5, and bandwidth parameters $b \in (0.06, 0.50)$. The star $\star$ denotes best bandwidths.

Results Results obtained using the local bandwidth are summarized in Table 2.2. It seems that the optimal bandwidth decreases with the sample size and that the values of r_{opt} are quite close to the theoretical values. The optimal distribution of design points observed in the simulation study seems to be somewhat different from the optimal values. This could be caused by a too sparse grid of bandwidths chosen in this simulation study combined with a rather small effect of the design in the neighborhood of the optimal values.

In order to obtain more reliable and informative result, we now rerun the simulations for $n = 800$ and $\sigma = 0.5$ on a finer grid for the bandwidth parameter b and display the dependency of the MAD and MSE of the empirical zero in contour- and heatplots in Figures 2.6 and 2.7. In this case, the optimal bandwidth parameter seems to be $b = 0.2$ and, for this bandwidth, the simulation agrees very well with the theoretically optimal values.

The improvement of using the optimal density of design points is displayed in Figures 2.9 and 2.8 for $n = 800$ and $b = 0.2$ (these plots may be interpreted as "cuts" of the heatplots 2.6 and 2.7 at the level $b = 0.2$). In this case, the results of the simulation agree very well with the theoretical result. On the other hand, both MSE and MAD do not change much for the parameter $r \in (0.4, 0.8)$ and even the uniformly distributed design points, i.e., $r = 0$, do not perform much worse than the optimal design.

2.6.2 Location of maximum

A small simulation study concerning the location of the maximum may be found in Hlávka (2011), where the regression functions $m_3(x) = \cos\{2\pi(x - \theta)\}$ and $m_4(x) = \theta^{-2}\cos\{2\pi(x^2 - \theta^2)\}$ were investigated. Notice that the regression function $m_3(.)$ leads to an asymptotically unbiased estimator of the location of maximum because $m_3^{(3)}(\theta) = 0$. For $m_4(.)$, we have that $m_4^{(3)}(\theta) \neq 0$ and the empirical location of maximum is an asymptotically biased estimator.

The dependency of MSE and MAD on the design density is investigated similarly as in in the previous Section 2.6.1. The simulation results are summarized in Hlávka (2011, Tables 1 and 2) (note that this paper is enclosed in Appendix on page **??**).

The conclusions of the simulation study for location of maximum are very similar to the results obtained in Section 2.6.1 and also do not contradict the theoretical findings derived in Section 2.5.

2.7 Conclusion

Both the theoretical results in Section 2.5 and the simulation study in Section 2.6 show that the proposed design of experiment increases the precision of the nonparametric kernel estimator of the zero of a regression function.

Comparison with Müller (1984b) shows that, with constant bandwidth and with the probability measure H replaced by the prior density $a(.)$, the AIMSE optimal design is optimal also for the MSE of empirical zero and location of extremes. Using local bandwidth, the AIMSE optimal design proposed in Müller (1984b) is optimal also from the point of view of the MSE of the empirical zero.

Our approach to nonparametric regression experiment design is applicable also to other nonparametric regression estimators. For example, one might be interested in local polynomial estimators that do not suffer from the boundary effect. Actually, the optimal designs for local polynomials should be the same because the variance of a local polynomial estimator is also inversely proportional to the density of design points (Simonoff; 1996, Section 5.2.2, p. 140) but some work is still needed to derive this result in a mathematically correct way.

References

Cheng, K. F. and Lin, P. E. (1981). Nonparametric estimation of a regression function, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**(2): 223–233.

Eddy, W. F. (1980). Optimum kernel estimators of the mode, *The Annals of Statistics* **8**(4): 870–882.

Eddy, W. F. (1982). The asymptotic distributions of kernel estimators of the mode, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **59**(3): 279–290.

MAD

log(MAD)

Figure 2.6: Simulations for local bandwidth: $n = 800$, $\sigma = 0.5$, thick vertical line denotes the MAD optimal value $r_{opt}(MAD) = 7/9 \doteq 0.71$.

Figure 2.7: Simulations for local bandwidth: $n = 800$, $\sigma = 0.5$, thick vertical line denotes the MSE optimal value $r_{opt}(MSE) = 5/9 \doteq 0.55$.

Figure 2.8: Simulations for local bandwidth: $n = 800$, $\sigma = 0.5$, $b = 0.2$, thick vertical line denotes the MAD optimal value $r_{opt}(MAD) \doteq 0.71$.

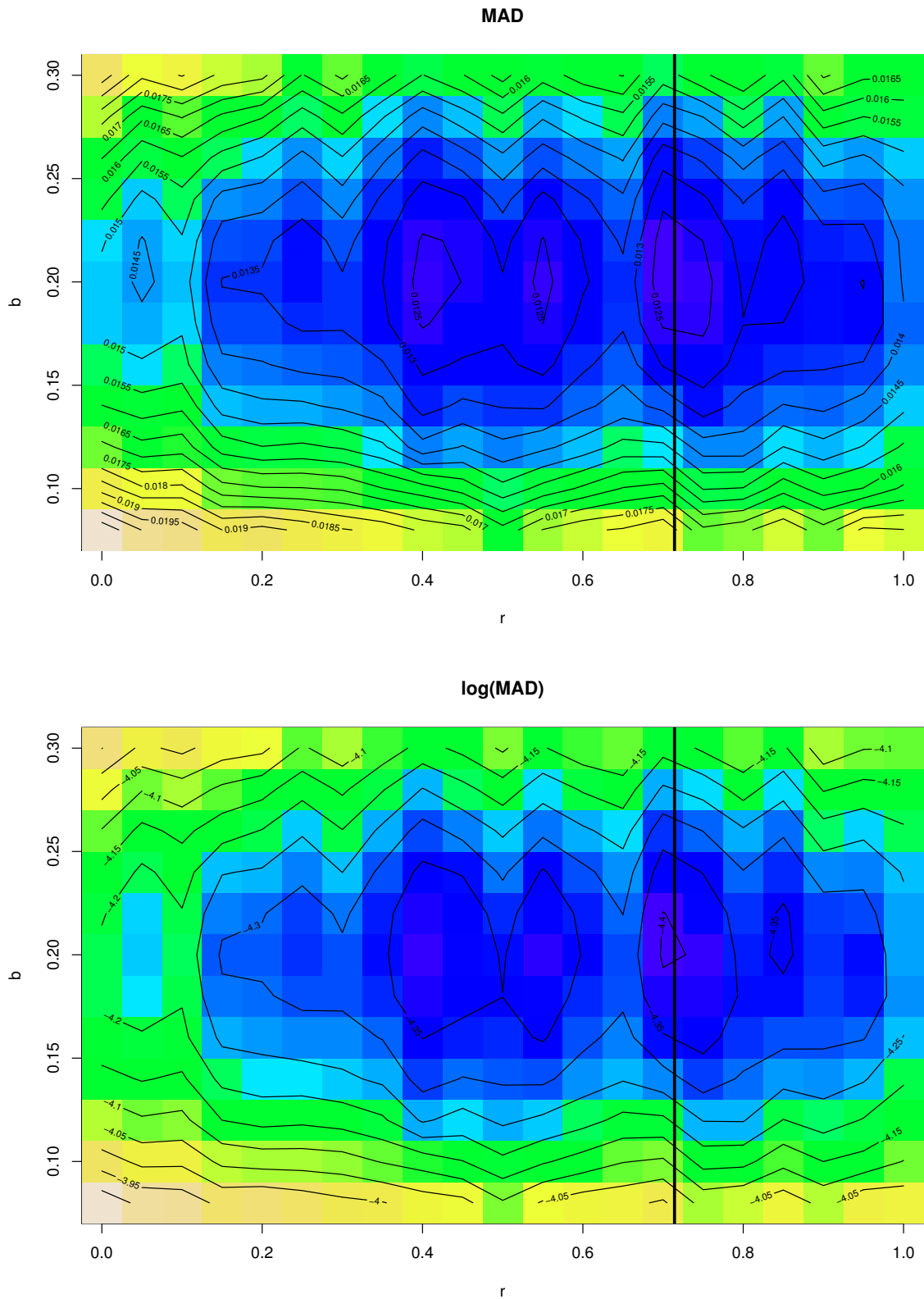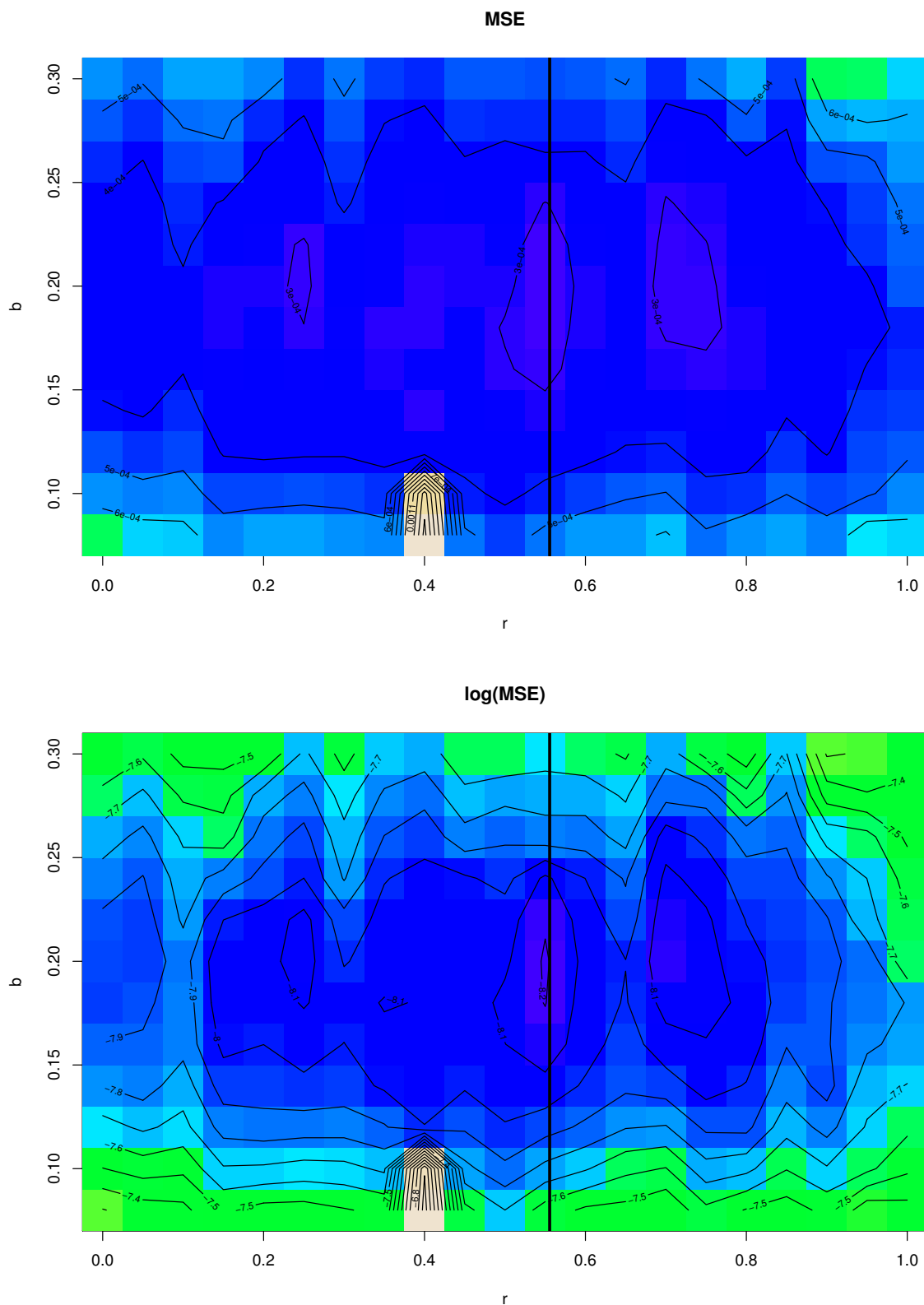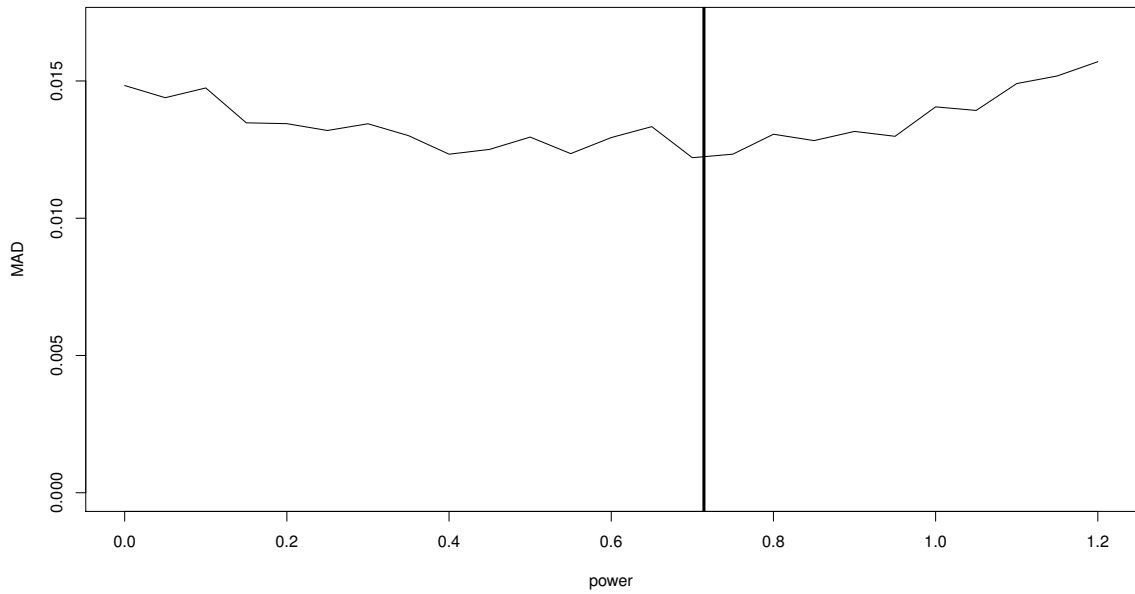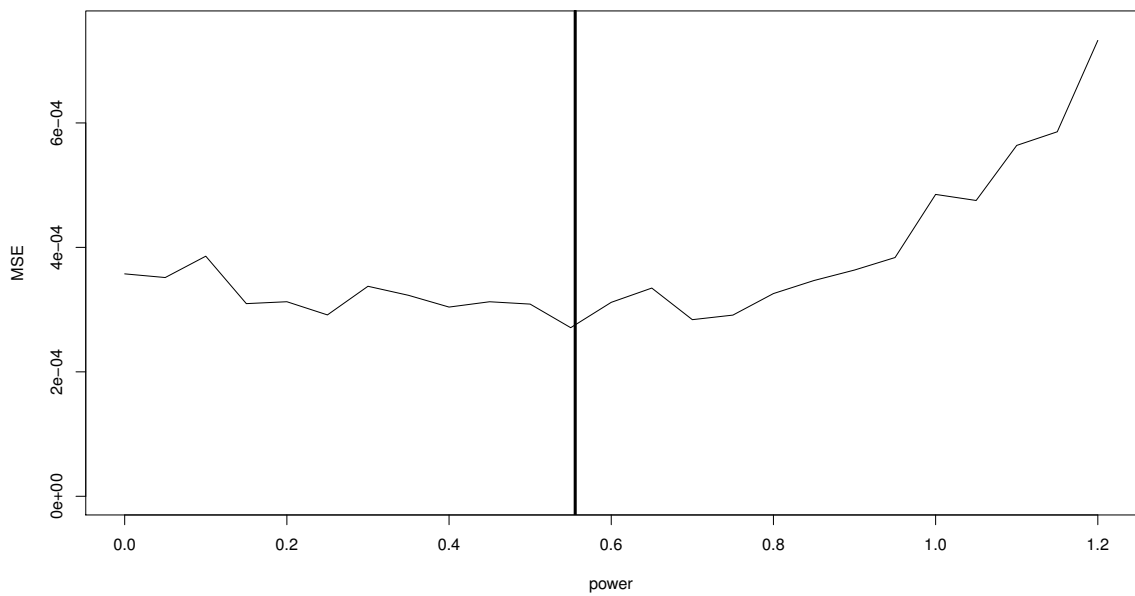Figure 2.9: Simulations for local bandwidth: $n = 800$, $\sigma = 0.5$, $b = 0.2$, thick vertical line denotes the MSE optimal value $r_{opt}(MSE) \doteq 0.55$.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics* **11**: 171–185.

Hlávka, Z. (2011). On nonparametric estimators of location of maximum, *Acta Universitatis Carolinae - Mathematica et Physica* **52**(1): 5–13.

Lin, Z. and Lu, C. (1996). *Limit theory for mixing dependent random variables*, Kluwer.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**(3): 405–415.

Müller, H.-G. (1984a). Boundary effects in nonparametric curve estimation models, *in* T. Havránek, Z. Šidák and M. Novák (eds), *COMPSTAT 1984: Proceedings in Computational Statistics: 6th Symposium held in Prague, 1984*, Physica-Verlag, pp. 84–89.

Müller, H.-G. (1984b). Optimal designs for nonparametric kernel regression, *Statistics & Probability Letters* **2**: 285–290.

Müller, H.-G. (1984c). Smooth optimum kernel estimators of regression curves, densities and modes, *The Annals of Statistics* **12**: 766–774.

Müller, H.-G. (1985). Kernel estimators of zeros and of location and size of extrema or regression functions, *Scandinavian Journal of Statistics* **12**: 221–232.

Nožička, F. (1998). Calculus of variations [in Czech], *in* K. Rektorys (ed.), *Přehled užité matematiky*, 5 edn, SNTL, Prague, pp. 841–858.

Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**: 1065–1076.

Peligrad, M. (1996). On the asymptotic normality of sequences of weak dependent random variables, *Journal of Theoretical Probability* **9**(3): 703–715.

R Development Core Team (2011). R: A language and environment for statistical computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Romano, J. P. (1988b). On weak convergence and optimality of kernel density estimates of the mode, *The Annals of Statistics* **16**(2): 629–647.

Roussas, G. G. (1989). Consistent regression estimation with fixed design points under dependence conditions, *Statistics & Probability Letters* **8**(1): 41–50.

Roussas, G. G. (1990b). Nonparametric regression estimation under mixing conditions, *Stochastic Processes and their Applications* **36**(1): 107–116.

Roussas, G. G. (1996). Exponential probability inequalities with some applications, *Statistics, probability and game theory*, Vol. 30 of *IMS Lecture Notes Monogr. Ser.*, Institute of Mathematical Statistics, Hayward, CA, pp. 303–319.

Roussas, G. G., Tran, L. T. and Ioannides, D. A. (1992). Fixed design regression for time series: asymptotic normality, *Journal of Multivariate Analysis* **40**(2): 262–291.

Simonoff, J. S. (1996). *Smoothing methods in statistics*, Springer Series in Statistics, Springer-Verlag, New York.

Smirnow, W. (1958). *Lehrgang der höheren Mathematik, Teil IV*, VEB Deutscher Verlag der Wissenschaften, Berlin.

Wieczorek, B. and Ziegler, K. (2010). On optimal estimation of a non-smooth mode in a nonparametric regression model with α-mixing errors, *Journal of Statistical Planning and Inference* **140**(2): 406–418.

Ziegler, K. (2003). On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model, *Journal of Statistical Planning and Inference* **115**(1): 123–144.

Chapter 3

Constrained nonparametric estimators

In this chapter, we present the content of the papers Hlávka (2006a); Härdle and Hlávka (2009); Hlávka and Svojík (2009) in a unified manner and we comment some non-standard properties of the proposed nonparametric regression estimator. Härdle and Hlávka (2009) apply nonparametric regression methods in order to estimate the so-called *State Price Density* (SPD) from observed option prices but the standard nonparametric regression estimator has to be modified in order to accommodate for the theoretical no-arbitrage requirements in a real-life situation. This leads to an estimator of a common second derivative of two regression curves assuming that this second derivative is a probability density function, i.e., it is positive and it integrates to 1.

Section 3.1 describes the SPD and its basic properties. In Section 3.2, we will show that the SPD may be estimated from the observed European Call and Put option prices via a nonparametric estimator of the second derivative of the pricing function. In order to improve the proposed estimator, we take into account the no-arbitrage constraints and use a simple approximation of the covariance structure of the observed option prices (Härdle and Hlávka; 2009). Some computational aspects (Hlávka; 2006a; Hlávka and Svojík; 2009) are reviewed in Sections 3.3 and 3.4. Section 3.4 also explains the difference between regression smoothing and nonparametric regression.

3.1   State price density

The probability density function of interest, the SPD, may be interpreted as a probability density function describing the expectations of the market. Therefore, it is very important in applied quantitative finance because the existence of a unique (risk neutral) SPD implies the absence of arbitrage (Harrison and Pliska; 1981) and its knowledge allows pricing of complicated (exotic) options.

The SPD cannot be observed directly but it is related to European Call and Put option prices in a simple way: let the symbol $C_d(K, D)$ denote the price of a European Call option with payoff $(S_D - K)_+ = \max(S_D - K, 0)$, where S_D denotes the price of the underlying stock at time D, d is the current time, and K is the strike price (recall that the buyer of a European Call option with strike prices K expiring at time D has the right to buy the underlying share on day D for the fixed price K and, therefore, he earns the difference $S_D - K$ if $S_D > K$; the buyer of a European Put option can sell the underlying share at time D for price K and earns $K - S_D$ if $S_D < K$). Clearly, the fair price of a European Call option should be equal to the expected value of the gain, i.e,

$$C_d(K, D) = \exp\{-r(D - d)\} \int_0^{+\infty} (S_D - K)_+ f(S_D) dS_D, \qquad (3.1)$$

where $f(.)$ is the probability density function of S_D. If the option prices are "fair" and the investors are "risk-neutral" then the density $f(.)$ is equal to SPD. Differentiating (3.1) with respect to the strike price K allows to express the SPD in terms of the European Call option pricing function (Breeden and Litzenberger; 1978):

$$f(K) = \exp\{r(D - d)\} \frac{\partial^2 C_d(K, D)}{\partial K^2}. \qquad (3.2)$$

37

S_d	K	r	τ	P_d	P/C	M	IV_d	time	volume	date
2079.23083	1925	0.05153	0.05	1.5	0	0.91667	0.20430	36369.67	100	19950102
2079.23083	1925	0.05153	0.05	1.5	0	0.91667	0.20430	38483.60	100	19950102
2079.23083	1925	0.05153	0.05	1.5	0	0.91667	0.20430	38794.47	100	19950102
2079.72548	1925	0.05153	0.05	1.4	0	0.91645	0.20217	40751.65	100	19950102
2079.23083	1925	0.05153	0.05	1.3	0	0.91667	0.19896	44073.10	100	19950102
2078.73618	1925	0.05153	0.05	1.8	0	0.91688	0.21110	45841.68	20	19950102
2083.68272	1925	0.05153	0.05	1.4	0	0.91471	0.20605	54200.51	30	19950102
2083.18806	1925	0.05153	0.05	1.3	0	0.91492	0.20280	54609.59	90	19950102
2081.20945	1925	0.05153	0.05	1.5	0	0.91579	0.20626	57224.43	1	19950102
2081.20945	1925	0.05153	0.05	1.2	0	0.91579	0.19801	57224.88	1	19950102
2079.23083	1925	0.05153	0.05	1.2	0	0.91667	0.19611	57555.44	49	19950102
2079.23083	1950	0.05153	0.05	2.6	0	0.92857	0.19862	45867.00	40	19950102
2074.28429	1975	0.05153	0.05	4.8	0	0.94272	0.19150	36827.09	2	19950102
2076.75756	1975	0.05153	0.05	4.3	0	0.94160	0.18869	44673.96	30	19950102
2075.27360	1975	0.05153	0.05	4.3	0	0.94227	0.18685	46724.45	50	19950102
2083.68272	1975	0.05153	0.05	4.3	0	0.93847	0.19716	54543.59	10	19950102
2078.24152	2000	0.05153	0.05	7.5	0	0.95283	0.18810	36450.55	100	19950102
2079.23083	2000	0.05153	0.05	7.2	0	0.95238	0.18682	38786.55	100	19950102
2082.19876	2000	0.05153	0.05	7.0	0	0.95102	0.18913	42075.81	10	19950102
2082.19876	2000	0.05153	0.05	6.3	0	0.95102	0.18258	42077.89	50	19950102
2077.74687	2000	0.05153	0.05	7.0	0	0.95306	0.18294	44219.41	40	19950102
2076.75756	2000	0.05153	0.05	7.2	0	0.95352	0.18334	44601.38	100	19950102
2074.77894	2000	0.05153	0.05	7.5	0	0.95443	0.18318	44856.34	100	19950102
2075.27360	2000	0.05153	0.05	7.5	0	0.95420	0.18389	46747.75	3	19950102
2084.17737	2000	0.05153	0.05	6.0	0	0.95012	0.18235	54276.01	75	19950102
2082.19876	2000	0.05153	0.05	6.4	0	0.95102	0.18353	54887.14	100	19950102
2089.12392	2025	0.05153	0.05	9.5	0	0.95972	0.18170	34442.91	10	19950102

Table 3.1: Beginning of a data set containing intra-day European option prices. The columns are option prices S_d, strike price K, risk-free interest rate r, time to maturity $\tau = D - d$ ($0.05 \times 360 = 18$ days), European put option price P_d, call indicator P/C (1 =Call, 0 =Put), moneyness M, implied volatility IV_d, time (in sec. after midnight, $36000 = 10 : 00$), volume, and date (yyyymmdd, $19950102 = 2^{\mathrm{nd}}$ January 1995).

Equation (3.2) has been previously used to estimate the state price density by means of nonparametric regression (Aït-Sahalia and Lo; 2000; Aït-Sahalia and Duarte; 2003; Bondarenko; 2003; Yatchew and Härdle; 2006). Considering the vector of the observed intra-day option prices and an approximation for its covariance matrix, Härdle and Hlávka (2009) proposed a constrained nonparametric estimator of the SPD and demonstrated that the proposed covariance structure leads to more flexible SPD estimators.

3.2　Constrained nonparametric SPD estimation

An example of a data set containing option prices is given in Table 3.1. Considering (3.2), we need to derive a nonparametric regression estimator for the option price as a function of the strike price such that its second derivative (up to known multiplicative constant) is a probability density function. Hence, we start by developing notation that will enable us:

1. describe the structure of the data set,

2. obtain a simple SPD estimator,

3. introduce the constraints,

4. consider correlated observations.

In order to introduce a simple model linking the SPD, i.e., the second derivative of the option pricing function with respect to the strike price, and the data set from Table 3.1, we now choose a subset of the data set corresponding only to one trading day and a single time to maturity, i.e., we fix the parameters d and $\tau = D - d$ and, in the following, these symbols may be omitted at our convenience.

Call options Let $\mathcal{C} = (C_1, \ldots, C_n)^\top$ denote the vector of the intra-day Call option prices observed during the selected day d and reordered in such a way that the corresponding vector of the strike prices has the following structure:

$$\mathcal{K} = \begin{pmatrix} K_1 \\ K_2 \\ \vdots \\ K_n \end{pmatrix} = \begin{pmatrix} k_1 \mathbf{1}_{n_1} \\ k_2 \mathbf{1}_{n_2} \\ \vdots \\ k_p \mathbf{1}_{n_p} \end{pmatrix},$$

where $k_1 < k_2 < \cdots < k_p$ are the distinct values of the observed strike prices, $n_j = \sum_{i=1}^{n} I(K_i = k_j)$ with $I(.)$ denoting the indicator function and $\mathbf{1}_m$ a vector of ones of length m.

We assume that, on the fixed trading day d with time to maturity $\tau = D - d$, the i-th observed Call option price (corresponding to strike price K_i) follows the model:

$$C_{d,i}(K_i, D) = C_i(K_i) = \mu(K_i) + \varepsilon_i, \tag{3.3}$$

where ε_i are correlated random errors, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, $\mathrm{Var}\,\varepsilon = V\sigma^2 > 0$.

The model (3.3) is a nonparametric regression model because we do not assume that the function $\mu(.)$ is a known function depending only on a few unknown parameters. On the other hand, this nonparametric regression model is non-standard because the explanatory variable is observed on a regular grid of only a few distinct strike prices. For example, the distinct strike prices observed in Table 3.1 are 1925, 1950, ..., 2025.

Due to the discrete nature of the explanatory variable, the nonparametric regression estimator of the option pricing function eventually degenerates to a p-dimensional vector containing the fitted values $\hat{\mu}(k_i)$. However, this is not a disadvantage because we may rewrite the degenerated nonparametric regression model (3.3) as a linear model that will allow an easy implementation of the no-arbitrage constraints on $\mu(.)$, i.e.,

(A) positivity,

(B) monotonicity,

(C) convexity,

(D) second derivative of $\mu(.)$ is a probability density function.

The shape of the Call option pricing function (or surface) implied by constraints (A)–(D) is nicely visible in Figure 3.1 displaying the observed prices of European Call options written on the DAX on January 16th, 1995. The left panel shows the ensemble of call option prices for different strikes and maturities as a free structure together with a smooth surface. The typical shape of dependency of the option price on the strike price can be observed on the right panel containing the option prices only for the shortest time to expiry, $\tau = D - d = 4$ days.

We have already explained that the option prices are observed only for few distinct strike prices $k_1 < \cdots < k_p$, see also Figure 3.1 and Table 3.1. Hence, the above assumptions (A)–(D) have to be reformulated in terms of the estimator $\hat{\mu}(.)$ defined only by its function values $\hat{\mu}(k_j)$ for $j = 1, \ldots, p$. In this way, we obtain the constraints:

(C1) $\hat{\mu}(k_i) \geq 0$, $i = 1, \ldots, p$,

(C2) $k_i < k_j$ implies $\hat{\mu}(k_i) \geq \hat{\mu}(k_j)$,

(C3) $k_i < k_j < k_l$ implies $-1 \leq \hat{\mu}^{(1)}_{k_i, k_j} \leq \hat{\mu}^{(1)}_{k_j, k_l} \leq 0$,
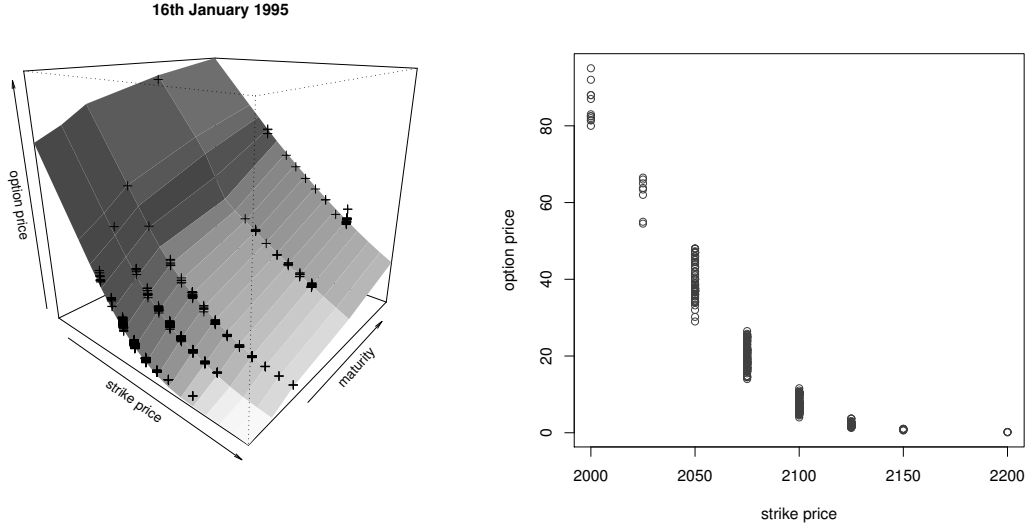
16th January 1995

Figure 3.1: Call option prices plotted against strike price and time to maturity with the fitted two-dimensional kernel regression surface (left) and the ensemble of the call option prices with shortest time to expiry against strike price (right) on 16th January 1995. SFB and CASE data base: sfb649.wiwi.hu-berlin.de. Figure reprinted from Härdle and Hlávka (2009).

where $\hat{\mu}^{(1)}_{k_i,k_j} = \{\hat{\mu}(k_i) - \hat{\mu}(k_j)\}/\{k_i - k_j\}$ denotes the natural estimator of the first derivative of the function $\mu(.)$. The regression function $\hat{\mu}(.)$ is defined by its value in points $k_1, \ldots, k_p$ and its second differences may be used to estimate the SPD (Härdle and Hlávka; 2009).

In Härdle and Hlávka (2009), the model (3.3) has been reparameterized in terms of parameters $\beta = (\beta_0, \ldots, \beta_p)^\top$ so that an estimator of β can be interpreted as an estimator of the SPD. This is achieved by modeling the vector of conditional means $\mu = (\mu_1, \ldots, \mu_p)^\top = (\mu(k_1), \ldots, \mu(k_p))^\top$ as $\mu = \Delta\beta$, with:

$$\Delta = \begin{pmatrix} 1 & \Delta^1_p & \Delta^1_{p-1} & \Delta^1_{p-2} & \cdots & \Delta^1_3 & \Delta^1_2 \\ 1 & \Delta^2_p & \Delta^2_{p-1} & \Delta^2_{p-2} & \cdots & \Delta^2_3 & 0 \\ \vdots & & & & & & \\ 1 & \Delta^{p-1}_p & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{3.4}$$

where $\Delta^i_j = \max(k_j - k_i, 0)$ denotes the positive part of the distance between k_i and k_j, the i-th and the j-th ($1 \le i \le j \le p$) sorted distinct strike price.

The model for the observed intra-day call option prices can now be written as:

$$C(\mathcal{K}) = \mathcal{X}_\Delta \beta + \varepsilon, \tag{3.5}$$

where $\mathcal{X}_\Delta$ is the design matrix obtained by repeating each row of matrix Δ n_i-times, $i = 1, \ldots, p$.

Using the coefficients $\beta_0, \ldots, \beta_{p-1}$, the constraints (C1)–(C3) are equivalent to $\beta_i > 0$, $i = 0, \ldots, p-1$ a $\sum_{j=2}^{p-1} \beta_j \le 1$. Next, we try to explain the meaning of the parameters $\beta_0, \ldots, \beta_{p-1}$ in the following Example 3.1.

Example 3.1. *Let $\mu_j = EC(k_j)$ denote the mean option prices for a given strike price k_j. Assuming that $\Delta^i_{i+1} = 1$ for all $i = 1, \ldots, p-1$, we may express that conditional expectations μ_j, $j = 1, \ldots, p$,*
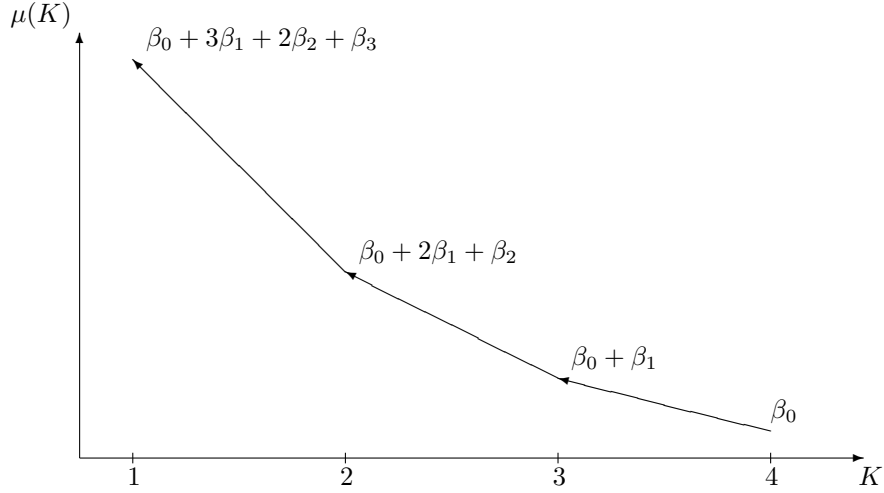
Figure 3.2: Four points example: illustration of the dummy variables for Call options with $\Delta_{i+1}^i = 1$: $\beta_3 = \mu''(2)$, $\beta_2 = \mu''(3)$.

in terms of the parameters β_i, $i = 1, \ldots, (p-1)$ in this way:

$$
\begin{aligned}
\mu_p &= \beta_0, \\
\mu_{p-1} &= \beta_0 + \beta_1, \\
\mu_{p-2} &= \beta_0 + 2\beta_1 + \beta_2, \\
\mu_{p-3} &= \beta_0 + 3\beta_1 + 2\beta_2 + \beta_3, \\
&\vdots \\
\mu_1 &= \beta_0 + (p-1)\beta_1 + (p-2)\beta_2 + \cdots + \beta_{p-1}.
\end{aligned}
$$

An illustration for $p = 4$ is plotted in Figure 3.2.

The meaning of parameters β_j may be now easily explained using Figure 3.2. The parameter β_0 is the mean option price in point 4. According to assumption (C1), β_0 has to be positive. The coefficient β_1 is the difference between option prices in points 4 and 3 and, by assumption (C2), it must be positive as well. The next coefficient, β_2, could be described as a change of the first derivative in point 3 and, hence, it is an estimator of the second derivative of the regression curve in this point. Similarly, β_3 may be interpreted as an estimator of the second derivative of the function $\mu(.)$ in point 2. Assumption (C3) implies that both β_2 and β_3 are positive. Assumption (C3) also implies that $\beta_1 + \beta_2 + \beta_3 \leq 1$.

The interpretation of coefficients $\beta_0, \ldots, \beta_3$ plotted in Figure 3.2 is simplified because the distance between neighboring strike prices is set to one. In practice, this is not a problem because we obtain the same interpretation of the parameters β_j even in the non-equidistant situation simply by using the matrix (3.4).

Put and Call options Apart of the Call options prices, we observe also the prices of the closely related Put options. The standard practice is to use the so-called *Put-Call parity* (Stoll; 1960), i.e., to use the relationship:

$$C_d(K, D) = P_d(K, D) + S_d - Ke^{-r\tau} \tag{3.6}$$

to transform all observed Put option prices $P_d(K, D)$ into Call option prices. However, we notice that (3.6) implies that the second derivatives of the pricing functions $C_d(K, D)$ and $P_d(K, D)$ w.r.t. K are equal and, therefore, we may include the Put option prices by extending the linear model (3.5) and estimate both option pricing function simultaneously.

Let $\mathcal{P} = (P_1, \ldots, P_m)^\top$ denote the vector of the intra-day Put option prices. Proceeding similarly as with the Call option prices, we define a linear model for the observed Put option prices in terms of parameters $\alpha_0, \ldots, \alpha_{p-1}$. Once again, it is easiest to explain the proposed model using Figure 3.3 and a simplified example.
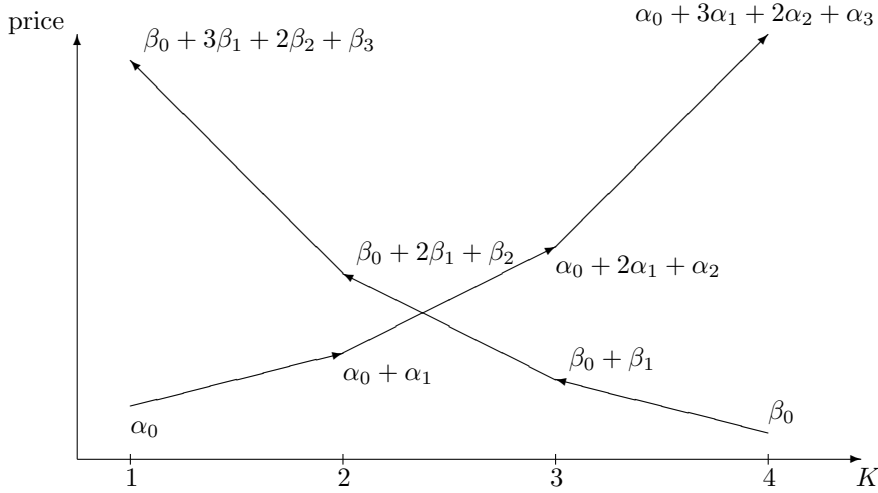
Figure 3.3: Dummy variables for both Call (β) and Put (α) options with $\Delta_i^{i+1} = 1$: $\beta_3 = \alpha_2 = \mu''(2)$, $\beta_2 = \alpha_3 = \mu''(3)$.

Example 3.2. *The typical shape of the increasing Put option pricing curve is displayed in Figure 3.3 together with the decreasing Call option pricing curve described in Example 3.1.   The Put option pricing curve is described in terms of parameters $\alpha_0, \ldots, \alpha_{p-1}$ so that, if $\nu_j = EP(k_j)$, we obtain:*

$$
\begin{aligned}
\nu_1 &= \alpha_0, \\
\nu_2 &= \alpha_0 + \alpha_1, \\
\nu_3 &= \alpha_0 + 2\alpha_1 + \alpha_2, \\
\nu_4 &= \alpha_0 + 3\alpha_1 + 2\alpha_2 + \alpha_3, \\
&\vdots \\
\nu_p &= \alpha_0 + (p-1)\alpha_1 + (p-2)\alpha_2 + \cdots + \alpha_{p-1}.
\end{aligned}
$$

It follows that, in the nonequidistant situation, we may write:

$$
\nu = \Delta^P \alpha = \begin{pmatrix}
1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
1 & \Delta_1^2 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & & & & & & \\
1 & \Delta_1^{p-1} & \Delta_2^{p-1} & \Delta_3^{p-1} & \cdots & \Delta_{p-2}^{p-1} & 0 \\
1 & \Delta_1^p & \Delta_2^p & \Delta_3^p & \cdots & \Delta_{p-2}^p & \Delta_{p-1}^p
\end{pmatrix}
\begin{pmatrix}
\alpha_0 \\
\alpha_1 \\
\vdots \\
\alpha_{p-2} \\
\alpha_{p-1}
\end{pmatrix},
\tag{3.7}
$$

where the symbol Δ_i^j is defined in (3.4). The linear model for the Put option prices can be written as:

$$
P(\mathcal{K}) = \mathcal{X}_\Delta^P \beta + \varepsilon,
\tag{3.8}
$$

where $\mathcal{X}_\Delta^P$ is the design matrix obtained by repeating each row of the matrix Δ^P m_i-times, $i = 1, \ldots, p$.
   The interpretation of the coefficients α_i, $i = 0, \ldots, p-1$ is very similar to the interpretation of coefficients β_i, $i = 0, \ldots, p-1$ explained in Example 3.1 and it is easy to show that in Figure 3.3, the coefficient α_2 estimates the value of the SPD for $K = 2$ and the coefficient α_3 estimates SPD for $K = 3$. Comparing this interpretation with the interpretation of the coefficients β_i and recalling that, using Put-Call parity (3.6), the SPD estimator calculated from the Call and Put options should be identical, we obtain:

$$
\alpha_i = \beta_{p-i+1}, \quad for \ i = 2, \ldots, p-1
\tag{3.9}
$$

and

$$
\alpha_1 = 1 - \sum_{i=1}^{p-1} \beta_i,
\tag{3.10}
$$

i.e., in matrix notation:

$$\alpha_{[-1]} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{p-2} \\ \alpha_{p-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & -1 & -1 & \cdots & -1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \vdots & & & & & \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-2} \\ \beta_{p-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \Gamma \beta_{[-1]}. \quad (3.11)$$

Equation (3.11) may be used to fit a vector containing both Put and Call option prices within a single linear model. The joint vector of parameters is $\gamma = (\alpha_0, \beta_0, \beta_1, \ldots, \beta_{p-1})^\top$ and the corresponding design matrix $\mathcal{X}_\Delta^{PC}$ is obtained by repeating each row of the matrix:

$$\Delta^{PC} = \begin{pmatrix} 0_p & \Delta \\ 1_p & \left(0_p & \Delta_{[-1]}^P \Gamma \right) \end{pmatrix}$$

n_i or m_i times, where $\Delta_{[-1]}^P$ denotes the matrix Δ^P without its first column. The resulting linear model for both Call and Put option prices:

$$\begin{pmatrix} \mathcal{C} \\ \mathcal{P} \end{pmatrix} = \mathcal{X}_\Delta^{PC} \gamma + \varepsilon \qquad (3.12)$$

allows to estimate the SPD by means of a simple linear regression. An advantage of this estimator is its simplicity but, on the other hand, the resulting estimator does not have to satisfy the constraints (C1)–(C2) and the resulting SPD estimator does not have to be a probability density function.

Constraints Härdle and Hlávka (2009) showed, applying ideas from Robertson et al. (1988), that the least squares estimator satisfying constraints (C1)–(C2) exists and that it is unique.

In order to calculate the least squares estimator in practice, we consider a simple reparameterization of the linear model (3.12) in terms of an unconstrained vector of parameters ξ:

$$\begin{aligned} \beta_0(\xi) &= \exp(\xi_0), \\ \beta_1(\xi) &= \frac{\exp(\xi_1)}{\sum_{j=1}^{p-1} \exp(\xi_j)}, \\ &\vdots \\ \beta_{p-1}(\xi) &= \frac{\exp(\xi_{p-1})}{\sum_{j=1}^{p-1} \exp(\xi_j)}, \\ \alpha_0(\xi) &= \exp(\xi_p). \end{aligned} \qquad (3.13)$$

Notice that arbitrary value of ξ guarantees that $\beta_i > 0$ and that the $\sum_{i=1}^{p-1} \beta_i = 1$. The parameter ξ can be estimated by using nonlinear least squares. The asymptotic distribution of the estimator $\hat{\xi}$ is described in literature (Seber and Wild; 1989).

Algorithm The following numerical algorithm was proposed in Härdle and Hlávka (2009):

1. Obtain an initial estimate $\hat{\beta}$, e.g., by running the Pool-Adjacent-Violators algorithm (Robertson et al.; 1988, Chapter 1) on the unconstrained least squares estimates of the first derivative or by using Hlávka (2006a).

2. Transform the initial estimates $\hat{\beta}$ into the estimates $\hat{\xi}$ using the method described in Härdle and Hlávka (2009, Section 3.2).

3. Estimate the parameters of the model (3.12) by minimizing the sum of squares in terms of parameters ξ using nonlinear least squares (Seber and Wild; 1989).

Pointwise confidence intervals for the SPD estimator can be obtained by calculating the asymptotic variance matrix of the transformation of the nonlinear least squares estimator $\hat{\xi}$, see Härdle and Hlávka (2009, Section 5.2).

Covariance implied by transaction time   Until now, we did not specify any assumptions concerning the vector of random errors ε in (3.5), (3.8), and (3.12) but, obviously, it would not be very reasonable to use the standard iid random errors assumptions. In order to approximate the covariance structure of the observed intra-day option prices, we assume that the SPD develops dynamically during the trading day and rewrite the linear model in terms of the "most recent value of the SPD".

To specify the covariance between any two option prices, we have to calculate the covariance between two Call option prices, the covariance between Put and Call option price and the covariance between two Put option prices. The covariances proposed in Härdle and Hlávka (2009) depend on the strike prices and on times of the transactions.

Let t_i denote the time of the i-th transaction and let $\widetilde{\beta}_i$ and $\widetilde{\alpha}_i$ denote the vectors containing the "current states" of the unknown parameters α_j and β_j, $j = 0, \ldots, p-1$, at time t_i. Let δ_i denote the difference between the transaction times t_i and t_{i-1}. Then it is straightforward to calculate the covariances:

$$
\begin{aligned}
\mathrm{Cov}\{C_{i-u}(k_j), C_{i-v}(k_i)\} &= \mathrm{Cov}(\Delta_j \widetilde{\beta}_{i-u}, \Delta_i \widetilde{\beta}_{i-v}) \\
&= \sigma^2 \Delta_j \Delta_i^\top \sum_{l=1}^{\min(u,v)} \delta_{i+1-l}, \quad\quad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}\{P_{i-u}(k_j), P_{i-v}(k_i)\} &= \mathrm{Cov}(\Delta_j^P \widetilde{\alpha}_{i-u}, \Delta_i^P \widetilde{\alpha}_{i-v}) \\
&= \sigma^2 \Delta_j^P (\Delta_i^P)^\top \sum_{l=1}^{\min(u,v)} \delta_{i+1-l}, \quad\quad (3.15)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}\{C_{i-u}(k_j), P_{i-v}(k_i)\} &= \mathrm{Cov}(\Delta_j \widetilde{\beta}_{i-u}, \Delta_i^P \widetilde{\alpha}_{i-v}) \\
&= \sigma^2 \sum_{l=1}^{\min(u,v)} \delta_{i+1-l} \sum_{k=2}^{p-1} \Delta_{p+1-k}^j \Delta_i^{p+1-k}. \quad\quad (3.16)
\end{aligned}
$$

Härdle and Hlávka (2009, Section 4.4) recommend to add an additional *microstructure noise* η_i $(0, \omega^2)$, such that η_i are iid and independent from the SPD dynamics.

Denoting by $\sigma^2 W$ the covariance matrix containing the covariances of the observed option prices given by (3.14)–(3.16) and $V = \sigma^2 W + \omega^2 \mathcal{I}$, we may finally say that we assume that $E\varepsilon = 0_n$ and $\mathrm{Var}\,\varepsilon = V$ in linear models (3.5), (3.8), and (3.12). A constrained estimate of the SPD $\widetilde{\beta}_1$ and of the variances σ^2 and ω^2 may be obtained by applying a standard iterative minimization procedure (Härdle and Hlávka; 2009, Sections 4.4 and 5.4).

Estimate with a fixed time to expiry   The proposed algorithm may be applied within each trading day in our data set. This means that the resulting SPD estimators will correspond to a linearly decreasing time to maturity with jumps occurring whenever the option with the shortest time to maturity expires. In our data set, these jumps occur every month, i.e., approximately every 20 trading days. In order to make the SPD estimators comparable, we combine two centered SPD estimates $\hat{f}_{\tau_1}(.)$ and $\hat{f}_{\tau_2}(.)$ corresponding to the times of expiry $\tau_1 < \tau_2$ and we construct an estimate $\hat{f}_\tau(.)$ for an arbitrary $\tau \in (\tau_1, \tau_2)$ as:

$$
\hat{f}_\tau(.) = \frac{(\tau_2 - \tau)\hat{f}_{\tau_1}(.) + (\tau - \tau_1)\hat{f}_{\tau_2}(.)}{\tau_2 - \tau_1}. \quad\quad (3.17)
$$

In this way, the variance of the price of the underlying asset implied by the observed SPD, $\hat{V}_\tau = \int x^2 \hat{f}_\tau(x) dx$, can be expressed as:

$$
\begin{aligned}
\hat{V}_\tau &= \int x^2 \hat{f}_\tau(x) dx \\
&= \int x^2 \frac{(\tau_2 - \tau)\hat{f}_{\tau_1}(x) + (\tau - \tau_1)\hat{f}_{\tau_2}(x)}{\tau_2 - \tau_1} dx \\
&= \frac{(\tau_2 - \tau)\hat{V}_{\tau_1} + (\tau - \tau_1)\hat{V}_{\tau_2}}{\tau_2 - \tau_1}.
\end{aligned}
$$

Härdle and Hlávka (2009, Section 6.1) argue that this estimator with a fixed time to expiry is reasonable because the SPD variances $\hat{V}_\tau$ are a linearly decreasing function of the time to expiry τ.

The main advantage of the SPD estimator with the fixed time to expiry is that the dynamics of such estimator has a very clear and straightforward interpretation, see Härdle and Hlávka (2009, Section 6.2 and 6.3) for a detailed analysis of the SPDs implied by DAX/EUREX option prices from January 1995 until March 2003 or Hlávka (2006b) for a functional principal component analysis of the estimated SPDs with a fixed time to maturity.

3.3 Speed of computation

In the previous section, we have mentioned that Härdle and Hlávka (2009) analyzed options on DAX from 1995 until 2003. During this period, the number of trades increased from few hundreds in 1995 to several thousands per day in 2003. In the same period, the number of traded distinct strike prices increased from approximately 10 to 100. This implies that we also need much more computer time to evaluate the estimators. A short simulation (Hlávka; 2006a, Table 1) shows that approximately 0.6 seconds are needed to calculate the estimator for $n = 200$ option prices but the same computer needs more than 1 minute to calculate the estimator for $n = 2000$, and even more than 12 minutes to calculate the estimator for $n = 5000$.

The structure of the data set suggests that many of the observations contribute only a small amount of information because most of the observed strike prices are concentrated close to the current value of the underlying stock. In Hlávka (2006a), the structure of the covariance matrix V is exploited in order to reduce the sample size without losing too much of the information.

The estimator In this section, we will use the general linear model (1.3) with the usual symbols for the response and explanatory variables, i.e.,

$$\mathcal{Y} = \mathcal{X}\beta + \varepsilon, \tag{3.18}$$

where $\mathcal{Y} = (Y_1, \ldots, Y_n)^\top$ is the vector of responses, $\mathcal{X}$ denotes the $n \times p$ design matrix, β is the vector of unknown parameters, and $\operatorname{Var} \varepsilon = V\sigma^2$ with a known $n \times n$ matrix $V > 0$ and unknown parameter $\sigma^2 > 0$. We assume that the structure of the design matrix $\mathcal{X}$ can be written as:

$$\mathcal{X} = M\mathcal{X}_M = \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & & & \\ 0_{n_p} & 0_{n_p} & \cdots & 1_{n_p} \end{pmatrix} \mathcal{X}_M, \tag{3.19}$$

where 1_n and 0_n denote, respectively, column vectors of ones and zeros of length n. In practice, the computation of the *general least squares* (GLS) estimator:

$$\widehat{\beta} = \left(\mathcal{X}^\top V^{-1} \mathcal{X} \right)^{-1} \mathcal{X}^\top V^{-1} \mathcal{Y} \tag{3.20}$$

can be very time consuming. Obviously, the most computationally demanding step is the evaluation of the inverse of the $(n \times n)$ variance matrix V. Therefore, in order to avoid the inversion of this variance matrix, Hlávka (2006a) proposed to replace the n-dimensional vector $\mathcal{Y}$ in (3.18) by a p-dimensional vector $\mathcal{Y}_M = M^\top \operatorname{diag}(w)\mathcal{Y}$ and to use the linear model:

$$\mathcal{Y}_M = \mathcal{X}_M \beta + \eta, \tag{3.21}$$

where w is a column vector of weights such that $M^\top \operatorname{diag}(w)M = \mathbf{I}_p$, $E\eta = \mathbf{0}_p$, and $\operatorname{Var} \eta = \sigma^2 V_M$ with $V_M = M^\top \operatorname{diag}(w)V \operatorname{diag}(w)M$. Using the simplified linear model (3.21) instead of (3.18), we obtain a "faster" least squares estimator:

$$\widehat{\beta}_M(w) = \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M \right)^{-1} \mathcal{X}_M^\top V_M^{-1} \mathcal{Y}_M. \tag{3.22}$$

It is very easy to see that the estimator $\widehat{\beta}_M(w)$ is unbiased:

$$
\begin{aligned}
E\widehat{\beta}_M(w) &= E\left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} \mathcal{Y}_M = \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} E\mathcal{Y}_M \\
&= \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} M^\top \operatorname{diag}(w) E\mathcal{Y} \\
&= \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} M^\top \operatorname{diag}(w) \mathcal{X}\beta \\
&= \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} M^\top \operatorname{diag}(w) M \mathcal{X}_M\beta \\
&= \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1} \mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\beta = \beta
\end{aligned}
$$

and that its variance is:

$$
\operatorname{Var}\widehat{\beta}_M(w) = \sigma^2 \left(\mathcal{X}_M^\top V_M^{-1} \mathcal{X}_M\right)^{-1}.
$$

Some theory   Let us now introduce notation for elements and submatrices of both the variance matrix V and its inverse V^{-1} corresponding to the blocks defined in (3.19): $V = (V_{ij})_{i,j=1,\ldots,p} = (v_{ij})_{i,j=1,\ldots,n}$ and $V^{-1} = (V^{ij})_{i,j=1,\ldots,p} = (v^{ij})_{i,j=1,\ldots,n}$, where V_{ij} are matrices of dimension $(n_i \times n_j)$ and, similarly, $V_M = (m_{ij})_{i,j=1,\ldots,p}$ and $V_M^{-1} = (m^{ij})_{i,j=1,\ldots,p}$.

For an arbitrary variance matrix V, it is very complicated to find the vector of weights w minimizing the variance of the estimator $\widehat{\beta}_M(w)$. Hlávka (2006a) describes minimizers of the trace and determinant of the matrix V_M.

Theorem 3.1. *Assume that the matrices V_{ii}, $i = 1, \ldots, p$, are positive definite. The vector, w_1 : $M^\top w_1 M = \mathcal{I}_p$, minimizing $\operatorname{tr}(V_M)$ is:*

$$
\tilde{w}_{1,i} = \frac{V_{ii}^{-1} 1_{n_i}}{1_{n_i}^\top V_{ii}^{-1} 1_{n_i}}, \ \text{for } i = 1, \ldots, p,
$$

where $\tilde{w}_{1,i} = (w_{1,n_{i-1}+1}, \ldots, w_{1,n_i})^\top$, for $i = 1, \ldots, p$.

Proof. See Hlávka (2006a, Theorem 2). □

Theorem 3.2. *Assume that the matrix V is positive definite. The vector, w_2 : $M^\top w_2 M = \mathbf{I}_p$, of weights minimizing $\det(V_M)$ has to satisfy the equations:*

$$
\begin{aligned}
\textstyle\sum_{j=1}^{p} V_{ij}(\tilde{w}_{2,j} m^{ij}) &= 0_{n_i}, \ \text{for } i = 1, \ldots, p, \\
1_{n_i}^\top \tilde{w}_{2,i} &= 1, \ \text{for } i = 1, \ldots, p,
\end{aligned}
\tag{3.23}
$$

where $\tilde{w}_{2,i} = (w_{2,n_{i-1}+1}, \ldots, w_{2,n_i})^\top$.

Proof. See Hlávka (2006a, Theorem 3). □

It is interesting to note that the weights minimizing the trace of V_M, derived in Theorem 3.1, depend only on the diagonal submatrices V_{ii}, $i = 1, \ldots, p$. Theorem 3.2 states that the weights minimizing the determinant of V_M have more complicated structure and depend on all submatrices V_{ij}, $i, j = 1, \ldots, p$. In practice, the weights w_2 from Theorem 3.2 may be calculated using an iterative numerical algorithm (Hlávka; 2006a, Remark 5).

Application to SPD estimation   In Section 3.2, we have described the covariance matrix $V = \sigma^2 W + \omega^2 \operatorname{diag} I$ proposed for the observed European Put and Call option prices in (3.14)–(3.16). Notice that, apart of two unknown parameters σ^2 and ω^2, the variance matrix depends only on W that in turn depends only on the known strike prices and transaction times.

For simplicity, we now consider a model without the microstructure noise, i.e., we set $\omega^2 = 0$. Under this assumption, it is very simple to calculate the weights w_1 proposed in Theorem 3.1: the weights minimizing the trace of V_M assign weight 1 to the most recent observation within each block (distinct strike price) and weight 0 to all other observations. The resulting estimator may be interpreted as an estimator using only the most recent observed market price for each strike price.

The application of Theorem 3.2 to SPD estimation is more complicated. A numerical study in (Hlávka; 2006a, Section 3.2) suggests that the largest weight is always assigned to the most recent observation within each block. Moreover, observation from other blocks that are "close in time" to observations with nonzero weight from neighboring blocks also seem to have large weights.

Using this observation, it is straightforward to understand the principle of the proposed computationally efficient estimators:

1. take one (or more) most recent observations in each block,

2. add the closest observations to already selected observations in adjacent blocks (neighboring strike prices),

3. calculate the estimator using the selected observations.

Several modifications of this algorithm are investigated in a simulation study in Hlávka (2006a, Section 4.1). The proposed estimator has slightly higher variance but, in some situations, the time of calculation was reduced from more than one minute to less than one second.

Time permitting, it is always better to use all available observations. These algorithms can be recommended, e.g., as pilot estimates for the computationally intensive iterative algorithms described in Section 3.2 or if large data sets have to be evaluated in real-time or before a fast approaching deadline.

3.4 Kernel smoothing of a nonlinear Kalman filter

In this section, the constrained nonparametric SPD estimator from Section 3.2 is implemented as a nonlinear Kalman filter (Svojík; 2007). We apply the Nadaraya-Watson estimator (1.7) to smooth the resulting nonparametric SPD estimator and obtain pointwise asymptotic confidence bands (Hlávka and Svojík; 2009). The proposed algorithm could be used for on-line monitoring of option prices.

Kalman filter   The Kalman filter is a recursive estimator of a true (but unobserved) *state* of a system based on continuously arriving *measurements*. The Kalman filter has many applications in engineering and technology but it is interesting also from a statistical point of view as a powerful tool of structural time series analysis (Harvey; 1989).

In order to rewrite the linear model (3.5) for the call option prices in the usual *state-space form*, we write for the i-th observation on a fixed day d:

$$C_i = \Delta_i \widetilde{\beta}_i + \varepsilon_i, \tag{3.24}$$
$$\widetilde{\beta}_i = \widetilde{\beta}_{i-1} + \eta_i, \tag{3.25}$$

where we use the same symbols as in (3.14), Δ_i denotes the i-th row of the design matrix $\mathcal{X}_\Delta$ (i.e., the row corresponding to the i-th observed strike price K_i), and we assume that the random variable $\varepsilon_i \sim N\left(0, \sigma^2\right)$ and the random vector $\eta_i \sim N\left(0_p, \omega^2 \delta_i \mathcal{I}_p\right)$ are uncorrelated. Similarly as in (3.14), the symbol t_i denotes the time of the i-th trade and $\delta_i = t_i - t_{i-1}$.

In the setup of SPD estimation, the standard Kalman filter algorithm, consisting of a prediction and updating step, has to be only slightly modified as in every step i we observe only one option price, C_i, corresponding to only one strike price K_i.

The Kalman filter starts in time t_0 in the *initial state* $\widetilde{\beta}_{0|0}$, interpretable either as an estimator of $\widetilde{\beta}_0$ or a predictor of $\widetilde{\beta}_1$, and its variance matrix $\Sigma_{0|0}$. Next, in every *prediction step*, in time t_{i-1}, we forecast the future value of the state vector in time t_i and calculate the variance matrix of the predictor:

$$\widetilde{\beta}_{i|i-1} = E(\widetilde{\beta}_i | \mathcal{F}_{i-1}) = \widetilde{\beta}_{i-1|i-1}, \tag{3.26}$$
$$\Sigma_{i|i-1} = \Sigma_{i-1|i-1} + \omega^2 \delta_i \mathcal{I}_p. \tag{3.27}$$

In the *updating step*, we use the new information available at time t_i to obtain the estimator:

$$\widetilde{\beta}_{i|i} = \widetilde{\beta}_{i|i-1} + K_i I_i, \tag{3.28}$$
$$\Sigma_{i|i} = \left(\mathcal{I}_p - K_i \Delta_i\right) \Sigma_{i|i-1}, \tag{3.29}$$

where $I_i = C_i(K_i) - C_{i|i-1}(K_i) = C_i(K_i) - \Delta_i \widetilde{\beta}_{i|i-1}$ is the prediction error with variance $F_{i|i-1} = \text{Var}(I_i | \mathcal{F}_{i-1}) = \sigma^2 + \Delta_i \Sigma_{i|i-1} \Delta_i^\top$ and where the symbol $K_i = \Sigma_{i|i-1} \Delta_i^\top F_{i|i-1}^{-1}$ denotes the so-called *Kalman gain*.

Similarly as in Section 3.2, the linear Kalman filter (3.26)–(3.29) does not guarantee that the resulting SPD estimator is a probability density function. Therefore, similarly as in Section 3.2, we use a transformation in order to obtain a nonlinear modification of the linear Kalman filter.

Extended Kalman filter   In Section 3.2, we have used a reparameterization of the linear model in terms of parameters $\widetilde{\xi} = (\xi_0, \ldots, \xi_{p-1})^\top$ by setting $\widetilde{\beta} = g(\widetilde{\xi})$ so that:

$$
\begin{align}
\beta_0 &= g_0(\xi_0) = \exp(\xi_0), \tag{3.30} \\
\beta_k &= g_k(\widetilde{\xi}) = \exp(\xi_k)/S(\widetilde{\xi}), \quad \text{for } k = 1, \ldots, p-1, \tag{3.31}
\end{align}
$$

where $S(\widetilde{\xi}) = \sum_{j=1}^{p-1} \exp(\xi_j)$. In this way, we obtain a nonlinear state space model:

$$
\begin{align}
C_i &= \Delta_i g(\widetilde{\xi}_i) + \varepsilon_i, \tag{3.32} \\
\widetilde{\xi}_i &= \widetilde{\xi}_{i-1} + \eta_i, \tag{3.33}
\end{align}
$$

where $\widetilde{\xi}_i$ and $g(\widetilde{\xi}_i) = \widetilde{\beta}_i$ denote the values of the unknown parameters at time t_i and the random variable $\varepsilon_i \sim N(0, \sigma^2)$ and the random vector $\eta_i \sim N\left(0_p, \omega^2 \delta_i \mathcal{I}_p\right)$ are uncorrelated.

The extended Kalman filter is obtained by a linearization using the Jacobian matrix:

$$
B_{i|i-1} = \left. \frac{\partial g(\widetilde{\xi}_i)}{\partial \widetilde{\xi}_i^\top} \right|_{\widetilde{\xi}_i = \widetilde{\xi}_{i|i-1}} \tag{3.34}
$$

that is derived explicitely in Svojík (2007) or Hlávka and Svojík (2009) and leads the extended prediction equations:

$$
\begin{align}
\widetilde{\xi}_{i|i-1} &= \widetilde{\xi}_{i-1|i-1}, \tag{3.35} \\
\Sigma_{i|i-1} &= \Sigma_{i-1|i-1} + \omega^2 \delta_i \mathcal{I}_p, \tag{3.36}
\end{align}
$$

and the extended updating equations:

$$
\begin{align}
\widetilde{\xi}_{i|i} &= \widetilde{\xi}_{i|i-1} + K_i I_i, \tag{3.37} \\
\Sigma_{i|i} &= \left(\mathcal{I}_p - K_i \Delta_i B_{i|i-1}\right) \Sigma_{i|i-1}, \tag{3.38}
\end{align}
$$

where $I_i = C_i - \Delta_i g(\widetilde{\xi}_{i|i-1})$ is the prediction error, the symbol $F_{i|i-1} = \mathrm{Var}(I_i | \mathcal{F}_{i-1}) = \sigma^2 + \Delta_i B_{i|i-1} \Sigma_{i|i-1} B_{i|i-1}^\top \Delta_i^\top$ denotes its variance, and $K_i = \Sigma_{i|i-1} B_{i|i-1}^\top \Delta_i^\top F_{i|i-1}^{-1}$ is the Kalman gain. The above equations (3.35)–(3.35) constitute the recursion of the extended Kalman filter and, for each transaction time t_i, the vector $g(\widetilde{\xi}_i) = \widetilde{\beta}_i$ is a recursive estimator of the current value of the SPD.

Finally, we remark that the unknown parameters σ^2 and ν^2 may be estimated by the *prediction error decomposition* of the likelihood function described in Kellerhals (2001, Chapter 5) with the resulting log-likelihood being maximized numerically. Similarly as in Section 3.2, the extended Kalman filter may use the additional information from the observed Put options prices, see Tlustý (2010) for a detailed description of the resulting nonlinear Kalman filter.

Kernel smoothing   In Figure 3.4, the Call option prices observed on January 15th, 1995, are plotted as a function of their strike price on the left-hand side plot in Figure 3.5.

Two of the resulting filtered SPD estimates are plotted in the graphics on the right-hand side of Figure 3.4. At the end of this trading day, for $i_2 = n = 410$ (15:59:52.14), the estimate is shifted a bit to the left and more concentrated. The shift to the left corresponds very well to a decrease in the value of the DAX from 2089.377 to 2075.989 observed in the market. The estimates of the variance of the error terms based on the prediction error decomposition were $\hat{\sigma}^2 = 0.0496$ and $\hat{\omega}^2 = 0.621$.

In Figure 3.5, we plot both Call option prices observed on February 25th, 2003, and the corresponding SPD estimates for times $i_1 = n/2 = 732$ and $i_2 = n = 1464$. Here, the parameter estimates were $\hat{\sigma}^2 = 0.0324$ and $\hat{\omega}^2 = 3.1953$. Unfortunately, the estimates obtained in year 2003 do not look at all like a smooth and unimodal probability density that we would like to obtain. Instead, we observe a lot of spikes and valleys. This is due to a larger number of distinct strike prices and the fact that the algorithm does not penalize non-smoothness and guarantees only that the resulting SPD estimates are positive and integrate to one.

In order to obtain better looking and more easily interpretable results, the resulting estimates may be smoothed using, e.g., the Nadaraya-Watson kernel regression estimator (1.7). This additional
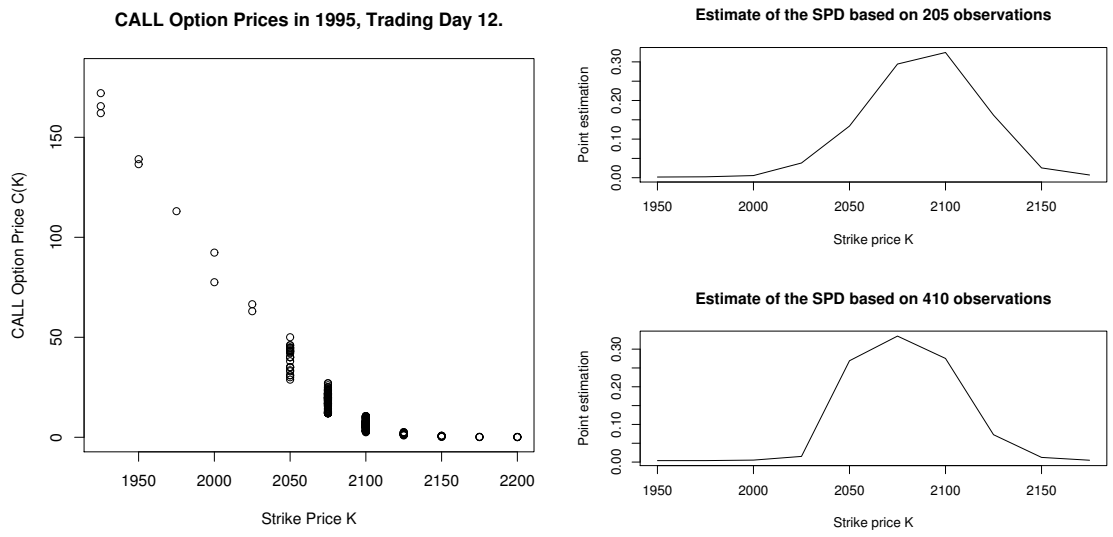
Figure 3.4: European call option prices with the shortest time to expiry plotted against strike price K (left) and two of the filtered SPD estimates (right) on January 15$^{\text{th}}$, 1995, $n = 410$, $p = 12$. Figure reprinted from Hlávka and Svojík (2009).
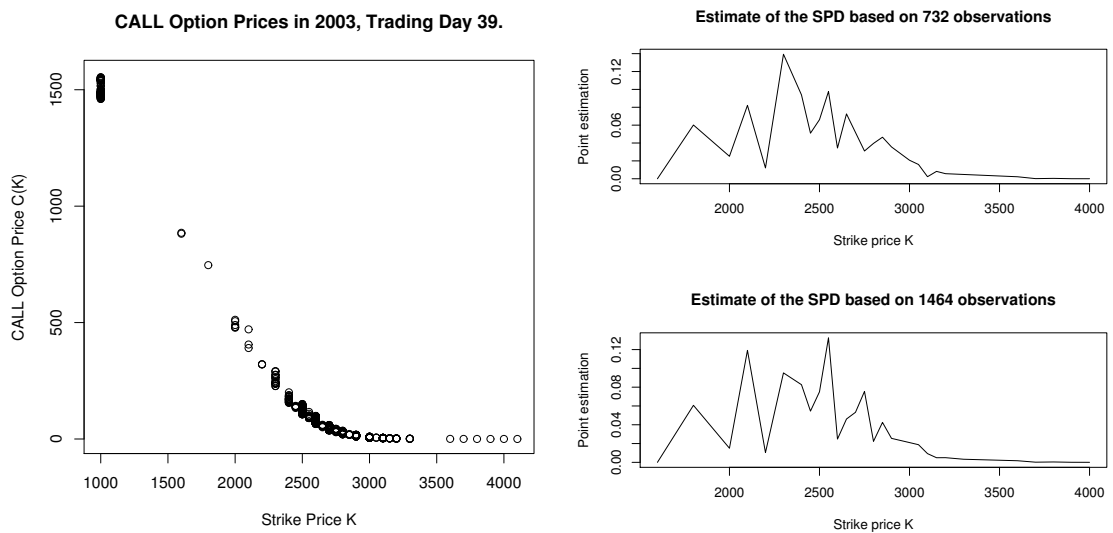

Figure 3.5: European call option prices with shortest time to expiry plotted against strike price K on February 25$^{\text{th}}$, 2003, $p = 30$, $n = 1464$, observed prices (left) and the resulting SPD estimates after 10 iterations (right). Figure reprinted from Hlávka and Svojík (2009). Figure reprinted from Hlávka and Svojík (2009).
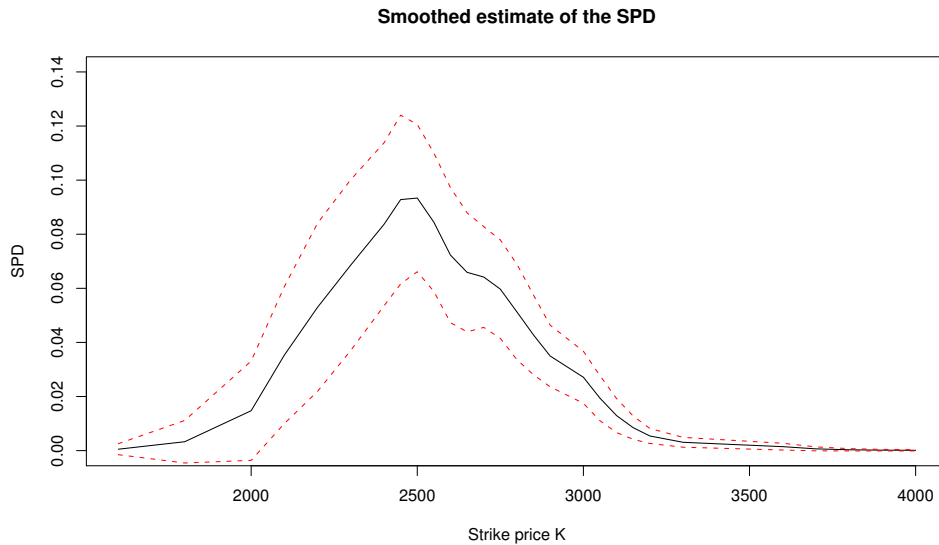
Figure 3.6: Smoothed SPD estimate on FEB-25-2003, $n = 1464$, $p = 30$, with pointwise asymptotic confidence intervals. Figure reprinted from Hlávka and Svojík (2009).

smoothing step can be easily implemented after the Kalman filtering, because the kernel smoothing of the vector $\widetilde{\beta}_{n|n}$ may be expressed as a matrix multiplication of $\widetilde{\beta}_{n|n}$ and a smoothing matrix, say S. Using the variance matrix $\Sigma_{n|n}$ from the filtering step of the extended Kalman filtering algorithm and the Jacobian matrix (3.34), Hlávka and Svojík (2009) obtain the asymptotic variance matrix of the smooth SPD estimator:

$$\operatorname{Var} \widetilde{\beta}_{n|n}^{smooth} = B_{n|n} S \Sigma_{n|n} S^{\top} B_{n|n}^{\top}. \tag{3.39}$$

The variance matrix (3.39) is used to calculate the asymptotic pointwise 95% confidence intervals plotted in Figure 3.6 together with the smoothed SPD estimate, see also Tlustý (2010) for a more detailed analysis.

References

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions, *Journal of Econometrics* **116**(1-2): 9–47.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion, *Journal of Econometrics* **94**(1-2): 9–51.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation, *Journal of Econometrics* **116**(1-2): 85–112.

Breeden, D. and Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices, *Journal of Business* **51**: 621–651.

Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities, *Journal of Econometrics* **150**(1): 1–15.

Harrison, J. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes Appl.* **11**: 215–260.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

Hlávka, Z. (2006a). Fast algorithm for nonparametric arbitrage-free SPD estimation, *Computational Statistics & Data Analysis* **51**(4): 2339–2349.

Hlávka, Z. (2006b). Functional principal components for estimated SPDs (in Czech), *in* J. Antoch and G. Dohnal (eds), *Proceedings of 14th Winter School JČMF Robust 2006*, JČMF, Prague, pp. 91–98.

Hlávka, Z. and Svojík, M. (2009). Application of extended Kalman filter to SPD estimation, *in* W. K. Härdle, N. Hautsch and L. Overbeck (eds), *Applied Quantitative Finance*, Springer, Berlin, pp. 233–247.

Kellerhals, B. P. (2001). *Financial pricing models in continuous time and Kalman filtering.*, Springer, Berlin.

Robertson, T., Wright, F. and Dykstra, R. (1988). *Order restricted statistical inference*, John Wiley & Sons, Chichester.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear regression*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York.

Stoll, H. R. (1960). The relationship between put and call option prices, *Journal of Finance* **24**: 801–824.

Svojík, M. (2007). Application of Kalman Filtering, Master thesis, Charles University in Prague, Faculty of Mathematics and Physics.

Tlustý, P. (2010). Extension of Kalman Filter [in Czech], Master thesis, Charles University in Prague, Faculty of Mathematics and Physics.

Yatchew, A. and Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap, *Journal of Econometrics* **133**(2): 579–599.

Chapter 4

Specification tests in nonparametric regression

In Chapter 1, we have introduced the nonparametric regression model. Usually, statistical inference concerning the nonparametric regression estimator rests on the assumption that the unobserved random errors ε_i, $i = 1, \ldots, n$, are independent and identically distributed. In practice, the validity of this assumption may be verified (tested) by applying so-called *specification tests*.

In this chapter, we describe some recently proposed specification tests applicable in nonparametric regression (Hlávka et al.; 2011; Einmahl and Van Keilegom; 2008a; Neumeyer; 2009) and we investigate the applicability of these specification tests in the framework of SPD estimation.

In Section 4.1, we describe specification tests based either on the joint characteristic function (Hlávka et al.; 2011) or on the joint distribution function (Einmahl and Van Keilegom; 2008a; Neumeyer; 2009) of the explanatory variable and the random error. In Section 4.2, we show that modified versions of these specification tests may be used to test hypotheses concerning the covariance matrix of random errors proposed in Section 3.2. The significance level and the power against various alternatives occurring in SPD estimation are investigated in a simulation study in Section 4.2.1. The asymptotic null distributions of two well-behaving test statistics are derived in Sections 4.2.2 and 4.2.3. In Section 4.3, we apply the resulting modified specification tests and investigate the validity of assumptions used in Chapter 3.

4.1 Tests of independence in nonparametric regression

Let us recall the random design nonparametric regression model (1.6) introduced in Section 1.1:

$$Y_i = m(X_i) + \varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{4.1}$$

where ε_i are iid centered random errors with finite variance, i.e., $E\varepsilon_i = 0$ and $\operatorname{Var} \varepsilon_i = \sigma^2$. The statistical inference is often based on the assumption that the unobservable random errors $\varepsilon_i = Y_i - m(X_i)$ are independent and identically distributed but, in practice, this assumption may not be always justified. Therefore, one should apply tests of independence described in this section in order to indicate a possible violation of model assumptions.

The joint distribution function of a random vector consisting of two independent random variables is equal to the product of the marginal distribution functions. Therefore, the hypothesis of independence between the explanatory variable X and the random errors ε may be expressed as:

$$H_0 : F_{X,\varepsilon} \equiv F_X F_\varepsilon, \tag{4.2}$$

where $F_{X,\varepsilon}(.,.)$ denotes the joint distribution function of the random vector $(X, \varepsilon)^\top$, and $F_X(.)$ and $F_\varepsilon(.)$ are the marginal distribution functions.

Tests based on characteristic function In terms of characteristic functions, the null hypothesis (4.2) may be equivalently rewritten as:

$$H_0 : \varphi_{X,\varepsilon} \equiv \varphi_X \varphi_\varepsilon, \tag{4.3}$$

where $\varphi_{X,\varepsilon}(.,.)$ denotes the joint characteristic function of the random vector $(X,\varepsilon)^\top$ and $\varphi_X(.)$ and $\varphi_\varepsilon(.)$ denote the marginal characteristic functions of random variables X and ε.

Hlávka et al. (2011) proposed a test statistic based on a difference $D_n(t_1,t_2) = \hat{\varphi}(t_1,t_2) - \hat{\varphi}_X(t_1)\hat{\varphi}_{\hat{\varepsilon}}(t_2)$ between the joint empirical characteristic function of the explanatory variable X and the estimated residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$:

$$\hat{\varphi}(t_1,t_2) = \frac{1}{n}\sum_{j=1}^n e^{it_1 X_j + it_2 \hat{\varepsilon}_j},$$

and the product of the marginal empirical characteristic functions:

$$\hat{\varphi}_X(t) = \frac{1}{n}\sum_{j=1}^n e^{itX_j} \quad \text{and} \quad \hat{\varphi}_{\hat{\varepsilon}}(t) = \frac{1}{n}\sum_{j=1}^n e^{it\hat{\varepsilon}_j}.$$

More precisely, Hlávka et al. (2011) suggest to reject the null hypothesis (4.3) for large values of the test statistic:

$$T_{n,W} = n\int_{-\infty}^\infty \int_{-\infty}^\infty |D_n(t_1,t_2)|^2 W(t_1,t_2)dt_1 dt_2, \tag{4.4}$$

where $W(t_1,t_2)$ is a suitable weight function.

Under some assumptions, Hlávka et al. (2011) derive that:

$$T_{n,W} \xrightarrow{d} \int_{\mathbb{R}^2} |Z(t_1,t_2)|^2 W(t_1,t_2)dt_1 dt_2, \tag{4.5}$$

where $\{Z(t_1,t_2), t \in \mathbb{R}^2\}$ is a Gaussian process with zero mean function and the covariance structure as the process $\{Z_0(t_1,t_2), (t_1,t_2) \in \mathbb{R}^2\}$ defined as:

$$\begin{aligned}
Z_0(t_1,t_2) &= \{\cos(t_2\varepsilon) - C_\varepsilon(t_2) + \varepsilon S_\varepsilon(t_2) + (\varepsilon^2-1)C_\varepsilon'(t_2)/2\} \\
&\quad \times \{\cos(t_1 X) + \sin(t_1 X) - C_X(t_1) - S_X(t_1)\} \\
&\quad + \{\sin(t_2\varepsilon) - S_\varepsilon(t_2) - \varepsilon C_\varepsilon(t_2) - (\varepsilon^2-1)S_\varepsilon'(t_2)/2\} \\
&\quad \times \{\cos(t_1 X) - \sin(t_1 X) - C_X(t_1) + S_X(t_1)\},
\end{aligned}$$

where $C_\varepsilon(.)$ and $S_\varepsilon(.)$ are the real and the imaginary part of the characteristic function of ε_j and $C_\varepsilon'(.)$ and $S_\varepsilon'(.)$ are respective derivatives. Similarly, $C_X(.)$ and $S_X(.)$ denote the real and the imaginary part of the characteristic function of X_j.

We remark that, compared to the papers by Neumeyer (2009) and Einmahl and Van Keilegom (2008a), no assumptions concerning the smoothness and boundedness of the probability density function of the random errors are needed in Hlávka et al. (2011, Section 2).

The asymptotic distribution of $T_{n,W}$ depends on the hypothetical distribution of the error terms and, therefore, the limit distribution does not provide an applicable approximation for the critical values. Therefore, both in Hlávka et al. (2011) and in the simulation study in Section 4.2.1, the critical values are obtained by bootstrap.

Computation of the test statistic and choice of the weight function In order to calculate the test statistic (4.4) from observations $(X_i, \hat{\varepsilon}_i)$, $i = 1,\ldots,n$, it is useful to rewrite it, using a straightforward calculation, as:

$$T_{n,W} = \frac{1}{n}\sum_{j,k=1}^n I_W(X_j - X_k, \hat{\varepsilon}_j - \hat{\varepsilon}_k) + \frac{1}{n^3}\sum_{j,k,l,m=1}^n I_W(X_j - X_l, \hat{\varepsilon}_k - \hat{\varepsilon}_m) - \frac{2}{n^2}\sum_{j,k,l=1}^n I_W(X_j - X_l, \hat{\varepsilon}_k - \hat{\varepsilon}_l),$$
(4.6)

where

$$I_W(x,y) = \int_{-\infty}^\infty \int_{-\infty}^\infty \cos(t_1 x + t_2 y)W(t_1,t_2)dt_1 dt_2.$$

In order to simplify the calculation of the test statistic (4.6), we consider the weight functions:

$$\begin{aligned}
W_1(t_1,t_2) &= e^{-(\gamma_{1,1}|t_1| + \gamma_{1,2}|t_2|)} & (4.7) \\
W_2(t_1,t_2) &= e^{-(\gamma_{2,1}t_1^2 + \gamma_{2,2}t_2^2)} & (4.8)
\end{aligned}$$

leading to a simple closed form expression for the test statistic (4.6).

The parameters $\gamma_{i,j}$ of the weight functions $W_i(t_1, t_2)$ in (4.7) and (4.8) are set proportionally to the observed standard deviations $sd(X) = \sqrt{n^{-1} \sum_{k=1}^{n} (X_k - \overline{X})^2}$ and $sd(\hat{\varepsilon}) = \sqrt{n^{-1} \sum_{k=1}^{n} (\hat{\varepsilon}_k - \overline{\hat{\varepsilon}})^2}$, where $\overline{\hat{\varepsilon}} = n^{-1} \sum_{j=1}^{n} \hat{\varepsilon}_j$ denotes the sample mean of the residuals. More precisely, for $i = 1, 2$, $\gamma_{i,1} = \gamma_i sd(X)$ and $\gamma_{i,2} = \gamma_i sd(\hat{e})$. In a homoscedastic situation, Hlávka et al. (2011) propose $\gamma_1 = 2$ and $\gamma_2 = 1$. In a more complicated heteroscedastic setup, the recommended values are $\gamma_1 = 0.75$ and $\gamma_2 = 0.5$.

Tests based on distribution function A family of tests based on measures of the difference between the product of the marginal empirical distribution functions and the joint empirical distribution function of X and the estimated residuals $\hat{\varepsilon}_i$ is proposed in Einmahl and Van Keilegom (2008a). These tests are straightforward generalizations of the well known Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests:

$$T_{n,KS} = \sqrt{n} \sup_{x,y} |F_n(x, \varepsilon) - F_{n,X}(x) F_{n,\hat{\varepsilon}}(\varepsilon)| \tag{4.9}$$

$$T_{n,CM} = n \int \int \{F_n(x, \varepsilon) - F_{n,X}(x) F_{n,\hat{\varepsilon}}(\varepsilon)\}^2 dF_{n,X}(x) dF_{n,\varepsilon}(\varepsilon), \tag{4.10}$$

$$T_{n,AD} = n \int \int \frac{\{F_n(x, \varepsilon) - F_{n,X}(x) F_{n,\hat{\varepsilon}}(\varepsilon)\}^2}{F_{n,X}(x) F_{n,\hat{\varepsilon}}(\varepsilon)\{1 - F_{n,X-}(x)\}\{1 - F_{n,\hat{\varepsilon}-}(\varepsilon)\}} dF_{n,X}(x) dF_{n,\hat{\varepsilon}}(\varepsilon), \tag{4.11}$$

where $F_n(.,.)$ denotes the joint empirical distribution function calculated from $(X_i, \hat{\varepsilon}_i)^{\top}$, $i = 1, \ldots, n$, with marginals $F_{n,X}(.) = F_n(., \infty)$ and $F_{n,\hat{\varepsilon}}(\infty, .)$ and where, for a distribution function $F(.)$, the symbol $F_-(.)$ denotes its left continuous version. Einmahl and Van Keilegom (2008a) derive the asymptotic distribution of the test statistics (4.9)–(4.11) but note simulations show that the resulting test does not achieve the prescribed size of the test and, in practice, a bootstrap procedure is preferable. A similar difference-based test of independence of X and $\varepsilon = Y - m(X)$ is proposed in Einmahl and Van Keilegom (2008b).

Kernel based test statistic Another interesting possibility of testing the hypothesis (4.2), using a kernel based test statistic proposed by Zheng (1997) for testing independence in an iid sample of paired observations, has been introduced in Neumeyer (2009):

$$T_{n,NZ} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{h_n}\right) \int \{I(X_i \leq x) - F_{X,n}(x)\} \{I(X_j \leq x) - F_{X,n}(x)\} w(x) dx,$$

$$\tag{4.12}$$

where $K(.)$ is a kernel function with bandwidth h_n and $w(.)$ is a weight function. This test statistic will be investigated more closely in the framework of SPD estimation in Section 4.2.3.

Results of a simulation study Hlávka et al. (2011) study the empirical significance level of the test based on the characteristic function using the weight functions $W_1(t_1, t_2)$ (denoted as CF1) and $W_2(t_1, t_2)$ (CF2), the Kolmogorov-Smirnov (KS), Cramér-von Mises (CM), and Anderson-Darling (AD) tests proposed by Einmahl and Van Keilegom (2008a), and the Neumeyer-Zheng test (NZ) proposed by Neumeyer (2009) for several distributions of the random errors:

$$\begin{aligned}
H_A &: \quad (\varepsilon | X = x) \sim N(0, 1), \\
H_{B(d)} &: \quad (\varepsilon | X = x) \sim (\chi_d^2 - d)/\sqrt{2d}, \\
H_{C(d)} &: \quad (\varepsilon | X = x) \sim t_d / \sqrt{d/(d-2)}.
\end{aligned}$$

The power is investigated against the following alternatives:

$$\begin{aligned}
H_{Aalt} &: \quad (\varepsilon | X = x) \sim N(0, 1 + ax), \\
H_{Balt} &: \quad (\varepsilon | X = x) \sim (1 + ax)^{1/2}(W_x - r_x)/\sqrt{2r_x}, \text{ where } W_x \sim \chi_{r_x}^2, r_x = 1/(bx), \\
H_{Calt} &: \quad (\varepsilon | X = x) \sim \sqrt{(1 + ax)\{1 - (cx)^{1/4}\}} T_x, \text{ where } T_x \sim t_{2/(cx)^{1/4}}.
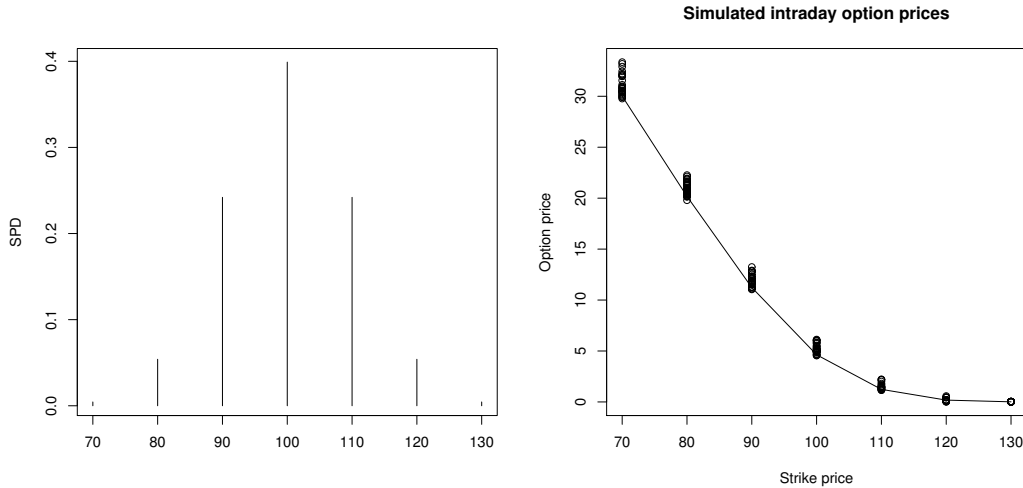\end{aligned}$$

Figure 4.1: An example SPD is plotted on the left hand side, the line on the right hand side shows the corresponding European call pricing function, the circles are simulated intra-day option prices ($p = 7$, $n = 200$).

The parameters $a > 0$, $b > 0$, and $c \in [0, 1]$ control, respectively, the dependency of the variance, skewness, and kurtosis of the random error ε on the design variable X.

The critical values are obtained by a bootstrap approximation utilizing standardized residuals and oversmoothing of the regression function (Härdle and Marron; 1991), see (Hlávka et al.; 2011) for a complete description of the bootstrap algorithm.

In a homoscedastic nonparametric regression model, the observed empirical levels for CF1($\gamma_1 = 2$), CF2($\gamma_2 = 1$) and NZ statistics are reasonably close to the nominal size $\alpha = 0.05$, with the more classical procedures KS, CM, and AD failing to capture the nominal size (Hlávka et al.; 2011, Section 4.1.4). Comparing the CF and NZ tests, it seems that CF tests perform slightly better against alternatives H_{Aalt} and H_{Balt} and the NZ test performs slightly better against the alternative H_{Calt}.

In a heteroscedastic setup, the NZ test seems to provide the most stable results (Hlávka et al.; 2011, Section 4.2.2).

4.2   Specification tests in SPD estimation

In this section, we investigate the appropriateness of the covariance structure proposed in Chapter 3 using tests of independence in nonparametric regression described in the previous Section 4.1.

Recall that in Chapter 3 we have estimated the SPD by using a linear model (3.5), i.e.:

$$\mathcal{C}(\mathcal{K}) = \mathcal{X}_\Delta(\mathcal{K})\beta + \varepsilon, \tag{4.13}$$

where $\mathcal{C}$ denotes the vector of observed Call option prices, $\mathcal{K}$ is the vector of corresponding strike prices, $\mathcal{X}_\Delta(\mathcal{K})$ is the design matrix (3.4), and ε are correlated random errors such that $E\varepsilon = 0_n$ and $\text{Var}\,\varepsilon = \sigma^2 V_0(\mathcal{K}, \mathcal{T})$. In Chapter 3, the matrix $V_0(\mathcal{K}, \mathcal{T})$ is calculated from the observed strike prices $\mathcal{K}$ and the transaction times $\mathcal{T}$, see (3.14).

As an illustration, we plot an example of a very simple artificial SPD in the left plot in Figure 4.1. The corresponding European Call pricing function and $n = 200$ simulated intra-day option prices are displayed in the right plot in Figure 4.1. A detailed description of the simulation algorithm based on (3.5) is given in Section 4.2.1. In the right plot in Figure 4.1, we observe strong heteroscedasticity of the simulated observations. Moreover, the simulated data feature also strong dependency of the option prices on the simulated transaction times. Notice that the simulated data set looks very similar to real DAX/EUREX European Call option prices plotted in Figure 3.1.

Assuming that the matrix $V_0 = V_0(\mathcal{K}, \mathcal{T})$ is positive definite, the linear model (4.13) can be written as:

$$V_0^{-1/2}\mathcal{C}(\mathcal{K}) = V_0^{-1/2}X_\Delta(\mathcal{K})\beta + V_0^{-1/2}\varepsilon, \tag{4.14}$$

where $\eta = V_0^{-1/2}\varepsilon$ is a vector of iid random errors. This suggests that the model assumptions could be verified by testing the hypothesis of independence between the strike prices $\mathcal{K}$ and the random errors η.

Test statistic Similarly as in Chapter 3, we assume that the observed intra-day option prices follow the linear model (4.13), where $\text{Var}\,\varepsilon = V\sigma^2$ for some (unknown) positive definite $(n \times n)$ matrix V. We are interested in testing the null hypothesis:

$$H_0 : V = V_{H_0} \tag{4.15}$$

against the general alternative $H_1 : V \neq V_{H_0}$, where $V_{H_0} > 0$ is a prespecified variance matrix, e.g., the covariance matrix V_0 defined by (3.14). More precisely, we apply the tests of independence from Section 4.1 to test the hypothesis:

$H_0 :$ " the distribution of the standardized random errors $\eta = V_{H_0}^{-1/2}\varepsilon$ does not depend on K"

against general alternatives.

In order to define the test statistic, we proceed similarly as in Section 4.1:

S1. We calculate the estimate $\hat{\beta}$ of the vector of parameters β from (4.14) with V_0 replaced by the hypothetical V_{H_0}, and the vector of residuals $\hat{\varepsilon} = \mathcal{C} - \hat{\mathcal{C}} = \mathcal{C} - \mathcal{X}_\Delta\hat{\beta}$.

S2. We calculate the vector of standardized residuals $\hat{\eta} = V_{H_0}^{-1/2}\hat{\varepsilon}$.

S3. We test the independence of $\eta = V_{H_0}^{-1/2}\varepsilon$ and $\mathcal{K}$ using the test statistics $T_n = T(\hat{\eta}, \mathcal{K})$: we will denote by $T_{n,KS}$, $T_{n,CM}$, and $T_{n,AD}$ respectively the Kolmogorov-Smirnov, the Cramér-von Mises, and the Anderson-Darling test statistics (4.9)–(4.11) (Einmahl and Van Keilegom; 2008a), by $T_{n,NZ}$ the Neumeyer-Zheng test statistic (4.12) (Neumeyer; 2009), and by $T_{n,CF1(\gamma_1)}$ and $T_{n,CF2(\gamma_2)}$ the test statistic (4.4) with weight functions (4.7)–(4.8) (Hlávka et al.; 2011).

The critical values of all tests are obtained by bootstrap. We fix the number of bootstrap replicates B and proceed as follows:

B1. For each $b = 1, \ldots, B$, we generate a bootstrap sample $\eta^{*b} = (\eta_1^{*b}, \ldots, \eta_n^{*b})^\top$ as an iid sample from the the distribution given by the empirical distribution function of the residuals $\hat{\eta}_1, \ldots, \hat{\eta}_n$ or the centered residuals $\hat{\eta}_i^c = \hat{\eta}_i - \sum_{j=1}^n \eta_j/n$, $i = 1, \ldots, n$.

B2. Define the vector of bootstrap option prices $\mathcal{C}^{*b} = X_\Delta\hat{\beta} + V_{H_0}^{1/2}\eta^{*b}$.

B3. Denote by T^{*b} the test statistic obtained from the b-th bootstrap sample, i.e., the test statistic calculated from the option prices $\mathcal{C}^{*b}$, the strike prices $\mathcal{K}$, and the transaction times $\mathcal{T}$.

B4. The null hypothesis is rejected if the test statistics T_n exceeds the critical value obtained as the $1 - \alpha$ empirical quantile of the sample $T^{*1}, \ldots, T^{*B}$.

We note that oversmoothing of the regression function cannot be used because the explanatory variable has discrete distribution.

4.2.1 Simulation study

The following short simulation study is based on the SPD plotted in Figure 4.1. We consider three types of null and alternative hypotheses concerning the variance matrix V of normally distributed random errors ε:

$H_{iid} :$ $V = \mathcal{I}_n$, i.e., iid random errors ε,

$H_{het} :$ $V = \text{diag}(V_0)$,, i.e., independent heteroscedastic random errors ε,

$H_{V_0} :$ $V = V_0$, i.e., iid random errors $\eta = V_0^{-1/2}\varepsilon$ with $V_0 = V_0(\mathcal{K}, \mathcal{T})$ given by (3.14).

In each step of the simulation, the vector of strike prices $\mathcal{K}$ is generated from Uniform distribution on $\{k_1, \ldots, k_p\}$ and the vector of the transaction times $\mathcal{T}$ is generated from Uniform distribution on $(0, 1)$ (from $\mathcal{K}$ and $\mathcal{T}$, we may already calculate the variance matrix $V_0(\mathcal{K}, \mathcal{T})$ given by (3.14)). Finally, we simulate normally distributed iid random errors $\eta \sim \mathcal{N}_n(0_n, \sigma^2\mathcal{I}_n)$ and, depending on the chosen variance matrix V, we obtain the vector of the simulated intra-day option prices as $\mathcal{C} = X_\Delta\widetilde{\beta}_1 + V^{1/2}\eta$. An example of such simulated data set with $\sigma = 0.04$ is plotted on the right-hand side in Figure 4.1.

	H_0	n	CF1(2)	CF2(1)	CF1($\frac{3}{4}$)	CF2($\frac{1}{2}$)	NZ	KS	CM	AD
		50	5.8	5.8	3.0	4.2	4.6	2.2	3.4	3.2
	iid	100	4.6	4.6	4.2	4.0	5.2	1.8	4.0	2.6
		200	5.4	5.4	3.6	4.2	2.8	2.2	4.8	4.2
		50	4.4	4.4	3.2	4.2	5.8	3.8	5.6	6.0
η^*	het	100	4.2	4.2	5.2	4.8	5.0	2.0	4.0	2.6
		200	5.6	5.0	4.8	4.6	6.2	4.0	6.6	6.6
		50	4.8	5.2	4.6	4.8	7.6	3.2	7.0	8.4
	V_0	100	5.8	5.4	5.6	6.0	7.0	2.6	5.6	4.0
		200	5.6	4.6	5.6	5.8	4.6	3.4	4.8	6.0
		50	5.8	5.8	3.0	4.2	4.6	2.2	3.4	3.2
	iid	100	4.6	4.6	4.2	4.0	5.2	1.8	4.0	2.6
		200	5.4	5.4	3.6	4.2	2.8	2.2	4.8	4.2
		50	4.8	5.0	3.2	4.2	6.4	3.8	5.6	5.2
η_c^*	het	100	4.6	4.2	5.2	4.6	5.0	2.4	3.8	2.2
		200	5.2	5.0	4.6	4.4	6.4	4.2	6.0	6.0
		50	5.0	5.2	4.8	5.2	7.8	3.2	6.2	6.8
	V_0	100	5.8	5.8	5.8	6.4	7.0	2.2	5.0	4.6
		200	5.4	4.4	5.6	6.0	5.6	3.4	4.6	6.0

Table 4.1: Empirical significance level (in %, unconstrained estimator, 500 simulations, $B = 250$ bootstrap replicates, $\alpha = 0.05$, sample size n) of the specification tests using ordinary η^* and centered η_c^* bootstrap residuals.

Empirical significance level and power for the unconstrained estimate In the simulation study, we choose the variance matrix of the random errors $V = \text{Var}(\varepsilon)$ and we test the null hypothesis $H_0 : \text{Var}(\varepsilon) = V_{H_0}$ for several choices of V and V_{H_0}. In each step of the simulation, we simulate a new data set with the variance of the random errors given by V. The test statistics and critical value for testing $H_0 : V = V_{H_0}$ are calculated according to algorithms S1–S3 and B1–B4, respectively.

In order to calculate the NZ test statistics, we use the bandwidth proposed in Neumeyer (2009), i.e., $h_n = \left[\sum_{j=2}^n (\hat{\eta}_{[i]} - \hat{\eta}_{[i-1]})^2 / \{2n(n-1)\} \right]^{1/5}$, where $\hat{\eta}_{[i]}$'s are sorted according to increasing covariates, Epanechnikov kernel $K(.)$, and function $w(x) = 1$. The parameters of the weight function needed for CF1 and CF2 test statistics were chosen according to Hlávka et al. (2011). No tuning parameters are required to calculate the KS, CM, and AD test statistics.

In Tables 4.1–4.3, we investigate the behavior of the independence tests using the unconstrained estimator of $m(.)$.

	Alt.	H_0	n	CF1(2)	CF2(1)	CF1($\frac{3}{4}$)	CF2($\frac{1}{2}$)	NZ	KS	CM	AD
			50	99.0	98.4	97.8	98.2	97.4	0.4	12.6	1.8
		iid	100	100.0	100.0	100.0	100.0	100.0	0.8	24.4	1.8
			200	100.0	100.0	100.0	100.0	100.0	2.0	33.0	1.4
η^*	V_0		50	61.0	59.6	63.8	60.8	38.8	19.8	37.8	31.6
		het	100	78.8	74.8	84.2	79.0	69.2	26.8	37.0	32.4
			200	91.4	90.0	97.4	96.0	91.0	40.6	48.2	48.2
			50	99.0	98.4	97.8	98.2	97.4	0.4	12.6	1.8
		iid	100	100.0	100.0	100.0	100.0	100.0	0.8	24.4	1.8
			200	100.0	100.0	100.0	100.0	100.0	2.0	33.0	1.4
η_c^*	V_0		50	61.2	60.4	63.4	61.4	39.6	19.4	36.8	31.2
		het	100	79.6	75.6	84.6	79.4	70.2	24.6	36.4	30.8
			200	93.2	91.2	97.8	96.2	93.0	39.8	47.8	47.0

Table 4.2: Empirical power (in %, unconstrained estimator, 500 simulations with $B = 250$ bootstrap replicates, $\alpha = 0.05$, centered bootstrap residuals, sample size n) of the specification tests against fixed alternative $V = V_0$.

	H_0	Alt.	n	CF1(2)	CF2(1)	CF1($\frac{3}{4}$)	CF2($\frac{1}{2}$)	NZ	KS	CM	AD
η^*	V_0	iid	50	67.8	67.8	43.8	46.2	44.2	57.8	38.0	52.4
			100	90.6	89.2	78.0	80.4	80.2	90.6	66.2	76.4
			200	97.0	96.8	95.2	96.0	97.0	98.8	91.2	93.4
		het	50	3.2	3.6	3.8	4.6	4.6	0.6	0.0	0.0
			100	3.8	3.6	5.8	6.6	6.6	0.4	0.0	0.0
			200	6.2	6.4	11.0	10.0	11.4	0.8	0.0	0.0
η_c^*	V_0	iid	50	67.4	67.4	43.8	46.0	44.0	58.2	37.6	53.0
			100	90.6	89.2	77.8	80.4	80.6	90.8	66.0	76.4
			200	97.0	96.8	95.2	96.0	97.0	98.8	91.4	93.4
		het	50	3.4	3.6	4.0	4.4	4.8	0.6	0.0	0.0
			100	3.8	3.6	5.8	6.6	6.6	0.4	0.0	0.0
			200	6.2	6.4	11.0	10.0	11.4	0.8	0.0	0.0

Table 4.3: Empirical power (in %, unconstrained estimator, 500 simulations with $B = 250$ bootstrap replicates, $\alpha = 0.05$, centered bootstrap residuals, sample size n) of the specification tests for fixed null hypothesis $H_0 : V = V_0$ against two alternatives.

The empirical significance levels in Table 4.1 seem to be reasonably close to the nominal value $\alpha = 0.05$. For the first two hypotheses, the significance level of the KS test is a bit smaller than 0.05. Looking at the third and most interesting null hypothesis, $H_0 : V = V_0(\mathcal{K}, \mathcal{T})$, the significance levels of all tests are improving with increasing sample size. The behavior of the test seems to be slightly more stable for the centered bootstrap residuals.

The empirical power for several combinations of null and alternative hypotheses is summarized in Tables 4.2–4.3. For the null hypothesis $H_0 : V = V_0$ and the alternative $H_1 : V = \mathcal{I}_n$, the best results are given by the KS test. In all other cases, the best power is achieved by the NZ test statistic. In general, the worst power is achieved by the tests of $H_0 : V = V_0$ against the alternative $H_1 : V = \text{diag}(V_0)$, where the KS, CM, and AD tests completely fail.

Empirical significance level and power for the constrained estimate In practice, one is interested in using the constrained estimator $\hat{m}_c(.)$ satisfying the no-arbitrage conditions (A)–(D) given in Section 3.2. The empirical significance level and power for the constrained estimator are summarized in Tables 4.4–4.5, where the constrained estimator $\hat{m}_c(.)$ has been used both for the evaluation of the test statistic in step S1 and within the bootstrap algorithm in step B3.

The empirical significance levels in Table 4.4 are acceptable only for the null hypothesis $H_0 : V = V_0$. The empirical significance levels for the hypothesis $H_0 : V = \text{diag}(V_0)$ are either too large (for the CF and NZ test statistics) or too small (all tests based on the empirical distribution function). The worst results are obtained for the null hypothesis $H_0 : V = \mathcal{I}_n$ suggesting that the constrained estimator works well only when the covariances between random errors ε are given by the covariance matrix V_0.

In Table 4.5, we report the empirical power only for the null hypothesis $H_0 : V = V_0$. Comparing Tables 4.3 and 4.5, we observe that the power of the CF and NZ tests is actually higher for the procedure using the constrained estimator and that the empirical power for the KS, CM, and AD tests is very similar for both the constrained and the unconstrained version of the estimator.

We conclude that the constrained estimator combined with NZ or CF test statistic may be used to test the null hypothesis $H_0 : V = V_0$ but it should not be used to test any other null hypothesis.

4.2.2 Asymptotic null distribution of $T_{n,W}$

In the simulation study in Section 4.2.1, we have found out that the best statistics are $T_{n,W}$ and $T_{n,NZ}$ corresponding to CF (Hlávka et al.; 2011) and NZ (Neumeyer; 2009) specification tests. In Sections 4.2.2 and 4.2.3, we shortly comment the validity of the asymptotic null distribution of these two test statistics in the setup of SPD estimation. For simplicity, we concentrate on the unconstrained estimator that, in the simulation study, seemed to be more robust.

Assuming a nonparametric regression model and a continuous explanatory variable, the asymptotic null distribution of $T_{n,W}$ has been derived in Hlávka et al. (2011). Unfortunately, this result does not

	H_0	n	CF1(2)	CF2(1)	CF1($\frac{3}{4}$)	CF2($\frac{1}{2}$)	NZ	KS	CM	AD
		50	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0
	iid	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		50	28.0	28.6	12.2	13.8	30.2	0.4	0.0	0.0
η^*	het	100	49.2	47.4	29.4	27.6	52.4	0.8	0.0	0.0
		200	81.4	80.4	60.2	55.8	85.8	1.0	0.0	0.0
		50	4.8	4.8	4.4	5.0	6.2	2.6	6.0	5.4
	V_0	100	5.4	5.0	5.6	6.2	5.6	2.6	5.6	3.6
		200	5.6	4.6	5.8	6.0	5.0	3.4	4.4	5.6
		50	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0
	iid	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		50	34.0	33.8	14.0	14.2	36.6	0.6	0.0	0.0
η_c^*	het	100	54.6	54.0	33.0	30.6	60.6	0.8	0.0	0.0
		200	86.0	84.6	65.4	61.4	89.4	1.0	0.0	0.0
		50	5.0	5.0	4.6	4.8	6.6	2.8	5.2	4.8
	V_0	100	5.2	5.4	5.6	6.4	5.4	2.2	5.0	3.8
		200	5.6	4.6	5.8	6.0	5.4	3.4	4.0	5.0

Table 4.4: Empirical significance level (in %, constrained estimator, 500 simulations, $B = 250$ bootstrap replicates, $\alpha = 0.05$, sample size n) of the specification tests using ordinary η^* and centered η_c^* bootstrap residuals.

	H_0	Alt.	n	CF1(2)	CF2(1)	CF1($\frac{3}{4}$)	CF2($\frac{1}{2}$)	NZ	KS	CM	AD
			50	85.6	82.4	72.6	72.2	72.6	58.0	27.8	40.0
		iid	100	95.8	95.6	96.8	95.2	98.2	92.2	57.4	75.6
			200	99.4	99.2	99.4	99.2	100.0	99.0	91.2	98.8
η^*	V_0		50	2.4	2.6	3.4	3.4	4.2	0.4	0.0	0.0
		het	100	3.4	3.2	5.8	6.6	6.2	0.2	0.0	0.0
			200	6.0	5.6	10.8	11.8	11.6	0.8	0.0	0.0
			50	85.4	82.6	72.6	73.2	74.2	57.6	29.2	41.6
		iid	100	95.8	95.6	97.2	95.4	98.4	92.4	58.4	76.0
			200	99.4	99.2	99.4	99.2	100.0	99.0	91.4	98.8
η_c^*	V_0		50	2.4	2.6	3.4	3.4	4.2	0.4	0.0	0.0
		het	100	3.4	3.2	5.8	6.6	6.2	0.2	0.0	0.0
			200	6.0	5.6	10.8	11.8	11.6	0.8	0.0	0.0

Table 4.5: Empirical power (in %, constrained estimator, 500 simulations with $B = 250$ bootstrap replicates, $\alpha = 0.05$, centered bootstrap residuals, sample size n) of the specification tests for fixed null hypothesis $H_0 : V = V_0$ against two alternatives.

automatically hold for model (4.13) and, in order to derive the asymptotic null distribution of $T_{n,W}$ for discretely distributed explanatory variable occurring in Chapter 3, we have to proceed similarly as Hlávka et al. (2011) with the residuals from the nonparametric regression model replaced by residuals from the linear model (4.13) similarly as in Hušková and Meintanis (2009, Theorem 1), see also Van Keilegom et al. (2008, Lemma 4.2).

In order to derive the asymptotic null distribution of $T_{n,W}$, we need to use assumptions allowing discretely distributed explanatory variable:

H1. Let the random errors $\varepsilon_1, \ldots, \varepsilon_n$ be iid random variables with zero mean and $E\varepsilon_i^4 < \infty$.

H2. $X_1, \ldots, X_n$ are iid random variables, independent with $\varepsilon_1, \ldots, \varepsilon_n$, having a discrete distribution on $\mathcal{X} = \{x_1, \ldots, x_p\}$.

H3. Let $m(.)$ be a regression function with Lipschitz first derivative satisfying the no-arbitrage assumptions (A)–(D) given in Chapter 3 and let $\hat{m}(.)$ denote its unconstrained estimate, i.e., $\hat{m}(x_j) = \sum_{i=1}^{n} Y_i I(X_i = x_j) / \sum_{i=1}^{n} I(X_i = x_j)$.

H4. The weight function $W(., .)$ is nonnegative and such that $W(t_1, t_2) = W(-t_1, t_2) = W(t_1, -t_2)$, $t_j \in \mathbf{R}$, $j = 1, 2$, and $\int_{\mathbf{R}^2} (t_1^4 + t_2^4) W(t_1, t_2) dt_1 dt_2 < \infty$.

The following Remark 4 says that the asymptotic distribution (4.5) derived in Hlávka et al. (2011) does not change if the explanatory variable has discrete distribution.

Remark 4. *Assuming H1–H4, under the null hypothesis* (4.15)*, the test statistic* $T_{n,W}$ *defined in* (4.4) *has the asymptotic distribution given in* (4.5)*.*

Proof. The proof is very similar to the proof of Theorem 1 in Hlávka et al. (2011). The only difference is in the proof of Lemma 2, where the remainder terms $Q_{nj}^c(t)$ and $Q_{nj}^s(t)$ in the Taylor expansions:

$$\cos(t\hat{\varepsilon}_j) = \cos(t\varepsilon_j) - t\sin(t\varepsilon_j)\{m(X_j) - \hat{m}_n(X_j)\} + t^2 Q_{nj}^c(t)$$

and

$$\sin(t\hat{\varepsilon}_j) = \sin(t\varepsilon_j) + t\cos(t\varepsilon_j)\{m(X_j) - \hat{m}_n(X_j)\} + t^2 Q_{nj}^s(t)$$

have to be treated similarly as the remainders in the proof of Theorem 1 in Hušková and Meintanis (2009), see also Lemma 4.2 in Van Keilegom et al. (2008). It follows that it suffices to verify that the estimator $\hat{m}(.)$ satisfies assumptions (A.10)–(A.13) in Hušková and Meintanis (2009). It is easy to see that the assumption (A.10) is satisfied because we assume that the true regression function (true parameter) satisfies the no-arbitrage constraints (A)–(D) given in Chapter 3. The assumption (A.11) is satisfied because the unknown parameter is equal to the value of the regression function. Finally, denoting $\theta = (m(x_1), \ldots, m(x_n))^\top$, and noticing that

$$\frac{\partial m(x_i; \theta)}{\partial \theta} = (\underbrace{0, \ldots, 0}_{(i-1)\times}, 1, 0, \ldots, 0)^\top$$

it is easy to see that also (A.12) and (A.13) are satisfied. □

In Remark 4, we assume that the random errors are iid and, strictly speaking, the result applies only when we test the null hypothesis $H_{iid} : V = \mathcal{I}_n$. However, further generalization for the null hypotheses H_{het} is very simple if the estimator $\hat{m}(.)$ is defined as a weighted average and if the residuals are divided by the known standard deviations of the random errors. Similarly, it is easy to see that Remark 4 applies also to the estimator (3.22) with the weights proposed in Theorem 3.1. The generalization for the null hypothesis H_{V_0} is also not very complicated because the residuals $\hat{\varepsilon}_i$ in the definition of (4.4) only have to be replaced by the residuals $\hat{\eta}_i$ obtained from the linear model (4.14).

Similarly as in Hlávka et al. (2011), the asymptotic null distribution (4.5) is mostly a theoretical consideration and, in practice, the critical values are obtained by bootstrap.

4.2.3   Asymptotic null distribution of $T_{n,NZ}$

In order to show that the asymptotic null distribution from Zheng (1997), Dette and Neumeyer (2000), and Neumeyer (2009) applies also in the setup of SPD estimation, we derive the asymptotic null distribution of the test statistic $T_{n,NZ}$ assuming that:

N1. The explanatory variable X has a discrete distribution on $\mathcal{X} = \{x_1, \ldots, x_p\}$ and assume that $F_{X|\varepsilon}(.|y) = P(X \leq x|\varepsilon = y)$ has uniformly bounded continuous derivations with respect to y.

N2. The random errors ε_i, $i = 1, \ldots, n$, are iid and centered with a bounded density $f_\varepsilon(.)$ with two continuous and bounded derivatives.

N3. Let $\hat{m}(x)$ denote the unconstrained estimate of the regression function $m(.)$, i.e., $\hat{m}(x_j) = \sum_{i=1}^n Y_i I(X_i = x_j) / \sum_{i=1}^n I(X_i = x_j)$.

N4. Let $K(.)$ be a symmetric and three times continuously differentiable kernel function with bounded derivatives. Assume that $\int K(u)du = 1$ and $\int K^2(u)u^2 du < \infty$. Let h_n be a sequence of bandwidths such that $h_n n^{1/7} \to 0$ and $h_n n^{1/2} \to \infty$ for $n \to \infty$.

N5. Let $w(.)$ be a positive and integrable weight function.

The following Theorem 4.1 gives the asymptotic null distribution of the test statistic $T_{n,NZ}$.

Theorem 4.1. *Under assumptions N1–N5 and under the null hypothesis, $nh_n^{1/2} T_{n,NZ} \xrightarrow{d} N(0, \sigma^2)$, where $\sigma^2 = 2 \int K^2(u)du E f_\varepsilon(X) \int\int F^2\{\min(x,y)\}[1 - F\{\max(x,y)\}]^2 w(x)w(y)dx$.*

Proof. The proof is postponed to Section 4.5.   □

In Theorem 4.1, we have derived the asymptotic null distribution only for iid random errors. Similarly as in Section 4.2.2, the generalization for heteroscedastic random errors is straightforward, see also Neumeyer (2009)[Section 5] for a generalization of this test statistic to a heteroscedastic nonparametric regression model with unknown variance function. Concerning the null hypothesis H_{V_0}, we only note that the proof of Theorem 4.1 under H_{V_0} works also with the correlated random errors ε_i from (4.13) replaced by the iid random errors η_i from (4.14).

The asymptotic distribution of $T_{n,NZ}$ is simpler than the asymptotic distribution of $T_{n,W}$ but, in practice, Neumeyer (2009) recommends the bootstrap approximation of critical values.

Comparing assumptions H1–H4 with assumptions N1–N5, we note that $T_{n,NZ}$ is preferable for heavy tailed random errors with smooth and bounded densities and that $T_{n,W}$ requires that $E\varepsilon_i^4$ is finite but it does not need any assumptions concerning the smoothness or boundedness of the density of the random errors.

4.3   Application

In Figure 4.2, we plot altogether $n = 577$ observed intra-day DAX European Call option prices from January 24th, 1995, the 17th trading (working) day in 1995. The line on the left-hand side of Figure 4.2 denotes the constrained estimate, $\hat{m}_c(.)$ of the unknown true function $m(.)$. The darkness of the points denotes the time of each trade and the fitted option pricing curve $\hat{m}_c(.)$ lies indeed closer to the darker, i.e., more recent, observations. On the right-hand side of Figure 4.2, we plot the standardized residuals $\hat{\eta} = V_0^{-1/2}\{\hat{m}_c(\mathcal{K}) - \mathcal{C}\}$ assuming that the variance of the random errors is given by the matrix $V_0(\mathcal{K}, \mathcal{T})$ defined by (3.14).

We consider only tests based on the unconstrained estimator of $m(.)$ which seem to be working quite well for all three null hypotheses considered in Section 4.2.1. We use 250 bootstrap replications with centered residuals.

Both the CF and NZ tests reject the hypotheses $H_0 : V = \mathcal{I}_n$ (p-value 0.000 for all tests) and $H_0 : V = \mathrm{diag}(V_0)$ (p-values 0.028, 0.040, 0.016, respectively for CF1, CF2, and NZ) whereas the hypothesis $H_0 : V = V_0$ is not rejected (p-values 0.556, 0.468, and 0.556, respectively).

On the contrary, the distribution function based tests that did not perform well in the simulation study in Section 4.2.1, do not reject any of the hypotheses (all p-values are 1.000 for $H_0 : V = \mathcal{I}_n$, the p-values 0.304, 0.640, and 0.892 were obtained for $H_0 : V = \mathrm{diag}(V_0)$ respectively for KS, CM, and AD tests, and the p-values 1.000, 0.420, and 0.236, respectively, for $H_0 : V = V_0$.

These results seem to be with good agreement with the simulation study in Section 4.2.1 according to which the CF and NZ specification tests provide the most reliable and most powerful results.
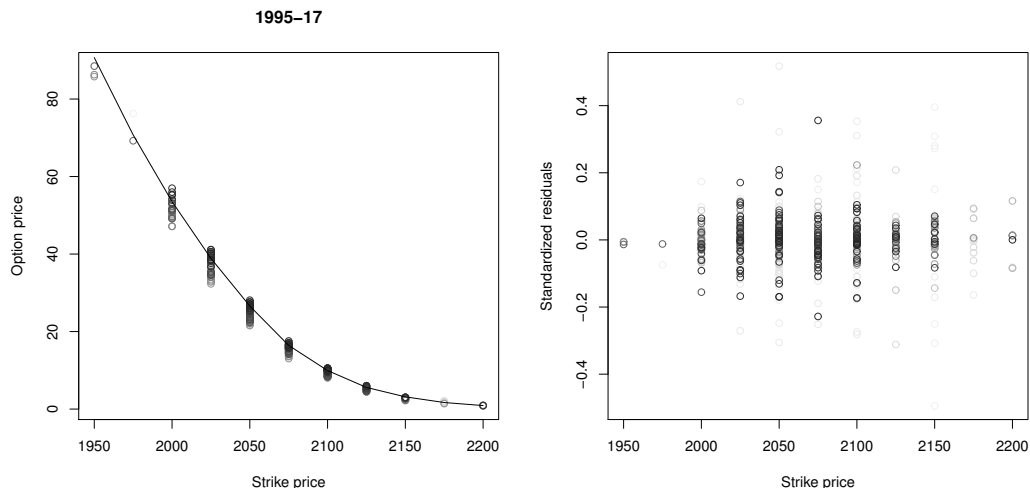
Figure 4.2: Observed intra-day European Call option prices on January 24th, 1995, the constrained estimate, and the residuals standardized by the covariance matrix V_0. Option prices corresponding to later transaction times are marked by darker color.

4.4 Summary

In this chapter, we have investigated the validity of the covariance structure proposed in the framework of the SPD estimation in Chapter 3 by using appropriate modifications of some recently proposed specification tests based either on the empirical joint characteristic function or on the empirical joint distribution function of the observed strike prices and the standardized residuals.

We have demonstrated in a simulation study that any specification test should not be applied together with the constrained estimator of the SPD.

Using the unconstrained SPD estimator, best results are achieved by the NZ test and the CF tests. The application of the constrained SPD estimator may be recommended only with the NZ and CF test and only for the null hypothesis $H_0 : V = V_0$. The simulation study also suggests that centering of the bootstrap residuals leads slightly better power.

The asymptotic null distribution of the NZ and CF test statistics in the setup of SPD estimation is investigated in Sections 4.2.2 and 4.2.3.

Finally, an application of the proposed methodology to a real data set shows that the proposed specification tests reject diagonal variance matrices and do not reject the covariance matrix $V_0(\mathcal{K}, \mathcal{T})$ proposed in Chapter 3.

4.5 Proof of Theorem 4.1

The proof of Theorem 4.1 is a minor modification of a part of the proof of Theorem 3.1 in Neumeyer (2009) and the proof of Theorem 1 in Zheng (1997). Compared to these papers, we assume that the residuals are estimated similarly as in Neumeyer (2009) and the distribution of the explanatory variable is discrete as in Dette and Neumeyer (2000).

For simplicity, we denote $T_n = T_{n,NZ}$ and consider:

$$\tilde{T}_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int a_i(x) a_j(x) w(x) dx, \tag{4.16}$$

where $\varepsilon_i = Y_i - m(X_i)$ and

$$a_i(x) = I(X_i \leq x) - F_{X,n}(x) = \frac{1}{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \{I(X_i \leq x) - I(X_k \leq x)\} = \frac{1}{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \eta_{i,k}(x).$$

As in Neumeyer (2009), we have by Taylor expansion:

$$
\begin{aligned}
T_n &= \tilde{T}_n + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} \left\{ K\left(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{h_n}\right) - K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \right\} \int a_i(x) a_j(x) w(x) dx \\
&= \tilde{T}_n + \sum_{l=1}^{3} \frac{1}{l!} V_n^{(l)},
\end{aligned}
$$

where

$$
\begin{aligned}
V_n^{(l)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n^{l+1}} K^{(l)}\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \{ m(X_i) - \hat{m}(X_i) - m(X_j) + \hat{m}(X_j) \}^l \\
&\quad \times \int a_i(x) a_j(x) w(x) dx, \quad \text{for } l = 1, 2, \\
V_n^{(3)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n^4} K^{(3)}\left(\xi_{i,j,n}\right) \{ m(X_i) - \hat{m}(X_i) - m(X_j) + \hat{m}(X_j) \}^\lambda \\
&\quad \times \int a_i(x) a_j(x) w(x) dx,
\end{aligned}
$$

with $\xi_{i,j,n}$ between $(\varepsilon_i - \varepsilon_j)/h_n$ and $(\hat{\varepsilon}_i - \hat{\varepsilon}_j)/h_n$.

By Lemma 4.1, we have that:

$$
\sum_{l=1}^{3} \frac{1}{l!} V_n^{(l)} = o_P(n^{-1} h_n^{-1/2})
$$

and it remains to investigate the term $\tilde{T}_n$.

We may closely follow the proof of Theorem 1 in Zheng (1997). We note that assumptions allowing discrete distribution of one of the variables were already used in Dette and Neumeyer (2000) who investigated the behavior of Zheng's independence test for fixed alternatives.

The term $\tilde{T}_n$ may be split into three parts:

$$
\begin{aligned}
\tilde{T}_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \{ I(X_i \leq x) - F_{X,n}(x) \} \{ I(X_j \leq x) - F_{X,n}(x) \} w(x) dx \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \{ I(X_i \leq x) - F_X(x) \} \{ I(X_j \leq x) - F_X(x) \} w(x) dx \\
&\quad - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \{ I(X_i \leq x) - F_X(x) \} \{ F_{n,X}(x) - F_X(x) \} w(x) dx \\
&\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h_n} K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \{ F_{n,X}(x) - F_X(x) \}^2 w(x) dx \\
&= T_{1n} - 2 T_{2n} + T_{3n}.
\end{aligned}
$$

Zheng (1997) showed that T_{2n} and T_{3n} are $o_P(n^{-1} h_n^{-1/2})$ and noticed that T_{1n} is a U-statistic with kernel $H_n(Z_i, Z_j) = K\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \xi_i(x) \xi_j(x) w(x) dx / h_n$, where $\xi_i(x) = I(X_i \leq x) - F_X(x)$ and $Z_i = (X_i, \varepsilon_i)^\top$. Moreover, under the null hypothesis, T_{1n} is a degenerate U-statistic because $E\{H_n(Z_1, Z_2) | Z_1 = z_1\} = 0$ and its asymptotic distribution may be obtained by Hall (1984, Theorem 1). Towards this end, we calculate for $x < y$:

$$
\begin{aligned}
E\{\xi_i(x) \xi_i(y)\} &= E[\{ I(X_i \leq x) - F_X(x) \} \{ I(X_i \leq y) - F_X(y) \}] \\
&= F_X(x)\{1 - F_X(x)\}\{1 - F_X(y)\} - \{F_X(y) - F_X(x)\} F_X(x)\{1 - F_X(y)\} \\
&\quad + \{1 - F_X(y)\} F_X(x) F_X(y) = F_X(x)\{1 - F_X(y)\}.
\end{aligned}
$$

In order to derive the asymptotic variance of T_{1n}, we calculate:

$$
\begin{aligned}
E\{H_n^2(Z_1, Z_2)\} &= \frac{1}{h_n^2} E\left\{ K^2\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int \xi_i(x)\xi_j(x)w(x)dx \right\}^2 \\
&= \frac{1}{h_n^2} EK^2\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \int\int E\{\xi_i(x)\xi_j(x)\xi_i(y)\xi_j(y)\}w(x)w(y)dxdy \\
&= \frac{1}{h_n} \int\int K^2(u)f_\varepsilon(x)f_\varepsilon(x - h_n u)dudx \int\int F^2\{\min(x,y)\}[1 - F\{\max(x,y)\}]^2 w(x)w(y)dxdy \\
&= \frac{1}{h_n} \left\{ \int K^2(u)du \int f_\varepsilon^2(x)dx + O(h_n^2) \right\} \int\int F^2\{\min(x,y)\}[1 - F\{\max(x,y)\}]^2 w(x)w(y)dxdy \\
&= \frac{1}{h_n} \int K^2(u)du E f_\varepsilon(X) \int\int F^2\{\min(x,y)\}[1 - F\{\max(x,y)\}]^2 w(x)w(y)dxdy + O(h_n) \\
&= \sigma^2/2h_n + O(h_n).
\end{aligned}
$$

It follows that $nh_n^{1/2}T_{1n} \overset{d}{\to} N(0, \sigma^2)$ because the condition:

$$
\left[E\{G_n^2(X_1, X_2)\} + n^{-1}E\{H_n^4(X_1, X_2)\}\right]/\left[E\{H_n^2(X_1, X_2)\}\right]^2 \to 0,
$$

where $G_n(x,y) = E\{H_n(Z_1, x)H_n(Z_2, y)\}$ (Hall; 1984, Theorem 1) is now easy to verify as we have already seen that $E\{H_n^2(X_1, X_2)\} = O(h_n^{-1})$ and similarly we obtain also that $E\{H_n^4(X_1, X_2)\} = O(h_n^{-3})$ and $E\{G_n^2(X_1, X_2)\} = O(1)$.

Lemma 4.1. *Under the assumptions of Theorem 4.1, we have that $V_n^{(l)} = o_P(n^{-1}h_n^{-1/2})$ for $l = 1, 2, 3$.*

Proof. Using the symmetry of $K(.)$ we obtain:

$$
V_n^{(1)} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{h_n^2} K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \{m(X_j) - \hat{m}(X_j)\} \int a_i(x)a_j(x)w(x)dx.
$$

Let $p_X(x) = P(X = x)$ and $\hat{p}_X(x) = \sum_{i=1}^n I(X_i = x)/n$. For the difference $m(X_j) - \hat{m}(X_j)$ we have:

$$
\begin{aligned}
m(X_j) &- \hat{m}(X_j) \\
&= \frac{\sum_{i=1}^n Y_i I(X_i = X_j)}{\sum_{i=1}^n I(X_i = X_j)} - m(X_j) \\
&= \frac{\sum_{i=1}^n \{m(X_i) + \varepsilon_i\} I(X_i = X_j)}{\sum_{i=1}^n I(X_i = X_j)} - m(X_j) \\
&= \frac{\sum_{i=1}^n \{m(X_j) + \varepsilon_i\} I(X_i = X_j)}{\sum_{i=1}^n I(X_i = X_j)} - m(X_j) \\
&= \frac{\sum_{i=1}^n \varepsilon_i I(X_i = X_j)/n}{\sum_{i=1}^n I(X_i = X_j)/n} \\
&= \frac{1}{n}\sum_{i=1}^n \varepsilon_i I(X_i = X_j)\left\{ \frac{1}{p_X(X_j)} + \frac{1}{\hat{p}_X(X_j)} - \frac{1}{p_X(X_j)} \right\} \\
&= \frac{1}{n}\sum_{i=1}^n \varepsilon_i I(X_i = X_j)\frac{1}{p_X(X_j)} + \frac{1}{n}\sum_{i=1}^n \varepsilon_i I(X_i = X_j)\left\{ \frac{1}{\hat{p}_X(X_j)} - \frac{1}{p_X(X_j)} \right\} \\
&= \mu_{j,n}^{(1)} + \mu_{j,n}^{(2)}.
\end{aligned}
$$

It follows that $V_n^{(1)} = U_n^{(1)} + U_n^{(2)}$, where

$$
U_n^{(k)} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{h_n^2} K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \mu_{j,n}^{(k)} \int a_i(x)a_j(x)w(x)dx, \quad \text{for } k = 1, 2.
$$

For $U_n^{(1)}$ we have:

$$
\begin{aligned}
U_n^{(1)} &= \frac{2}{n^2(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{v=1}^{n} \frac{1}{h_n^2} K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \varepsilon_v I(X_v = X_j) \frac{1}{p_X(X_j)} \int a_i(x) a_j(x) w(x) dx \\
&= \frac{2}{n^4(n-1)h_n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{v=1}^{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{\substack{l=1 \\ l \neq j}}^{n} K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \frac{\varepsilon_v I(X_v = X_j)}{p_X(X_j)} \int \eta_{i,k}(x) \eta_{j,l}(x) w(x) dx.
\end{aligned}
$$

As $E\varepsilon_v = 0$, only terms with $v \in \{i, j\}$ are not zero in $EU_n^{(1)}$. For $v = i$, we obtain:

$$
\begin{aligned}
&E\left\{ K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \frac{\varepsilon_i I(X_i = X_j)}{p_X(X_j)} \int \eta_{i,k}(x) \eta_{j,l}(x) w(x) dx \right\} \\
&= \int\int\int\int\int\int K'\left(\frac{y - z}{h_n}\right) \frac{y I(s = t)}{p_X(t)} \int \{I(s \leq x) - I(v \leq x)\} \\
&\qquad \times \{I(t \leq x) - I(w \leq x)\} w(x) dx \, dF_{X|\varepsilon=y}(s) dF_{X|\varepsilon=z}(t) dF_\varepsilon(y) dF_\varepsilon(z) dF_X(v) dF_X(w) \\
&= \int\int\int\int\int\int K'(u) \frac{y I(s = t)}{p_X(t)} f_\varepsilon(y - h_n u) f_\varepsilon(z) \int \{I(s \leq x) - I(v \leq x)\} \\
&\qquad \times \{I(t \leq x) - I(w \leq x)\} w(x) dx \, dF_{X|\varepsilon=y-h_n u}(s) dF_{X|\varepsilon=z}(t) dy dz dF_X(v) dF_X(w) h_n \\
&= O(h_n^2),
\end{aligned}
$$

where the last equation follows by Taylor expansion of $F_{X|\varepsilon=y-h_n u}(s)$ and $f_\varepsilon(y - h_n u)$ and from $\int K'(u) du = 0$. The same rate is obtained for $v = j$. This gives

$$
E\{U_n^{(1)}\} = \frac{2}{n^4(n-1)h_n^2} O(n^4) O(h_n^2) = O(1/n).
$$

Next, we investigate $E\{U_n^{(1)}\}^2$ under the null hypothesis:

$$
\begin{aligned}
E\{U_n^{(1)}\}^2 &= O(n^{-10} h_n^{-4}) \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{v=1}^{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{\substack{l=1 \\ l \neq j}}^{n} \sum_{i'=1}^{n} \sum_{\substack{j'=1 \\ j' \neq i'}}^{n} \sum_{v'=1}^{n} \sum_{\substack{k'=1 \\ k' \neq i'}}^{n} \sum_{\substack{l'=1 \\ l' \neq j'}}^{n} \\
&\quad \times E\left\{ K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) K'\left(\frac{\varepsilon_{i'} - \varepsilon_{j'}}{h_n}\right) \varepsilon_v \varepsilon_{v'} \frac{I(X_v = X_j) I(X_{v'} = X_{j'})}{p_X(X_j) p_X(X_{j'})} \right. \\
&\quad \left. \times \int \eta_{i,k}(x) \eta_{j,l}(x) w(x) dx \int \eta_{i',k'}(x) \eta_{j',l'}(x) w(x) dx \right\}.
\end{aligned}
$$

The expectation $E\left\{ K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) K'\left(\frac{\varepsilon_{i'} - \varepsilon_{j'}}{h_n}\right) \varepsilon_v \varepsilon_{v'} \right\}$ is not zero only if at least two pairs of the indices $\{i, j(\neq i), v, i', j'(\neq i'), v'\}$ are equal. When the remaining indices are distinct, the term is of order $O(h_n^4)$ and there are at most $O(n^8)$ of such pairs. If there are more indices equal to each other, the term is of order $o(h_n)$ and there are at most $O(n^7)$ of such combinations. Thus we have

$$
E\{U_n^{(1)}\}^2 = O(n^{-10} h_n^{-4}) \{O(n^8 h_n^4) + o(n^7 h_n)\} = O(n^{-2}) + o(n^{-3} h_n^{-3}) = o(n^{-2} h_n^{-1})
$$

and by Chebyshev's inequality we obtain that $U_n^{(1)}$ is of order $o_P(n^{-1} h_n^{-1/2})$.

Similarly, we may show that, under the null hypothesis, the terms

$$
\begin{aligned}
U_n^{(2)} &= \frac{2}{n^4(n-1)h_n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{v=1}^{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{\substack{l=1 \\ l \neq j}}^{n} K'\left(\frac{\varepsilon_i - \varepsilon_j}{h_n}\right) \varepsilon_v I(X_v = X_j) \\
&\quad \times \left\{ \frac{1}{\hat{p}_X(X_j)} - \frac{1}{p_X(X_j)} \right\} \int \eta_{i,k}(x) \eta_{j,l}(x) w(x) dx.
\end{aligned}
$$

and $V_n^{(2)}$ are also of order $o_P(n^{-1} h_n^{-1/2})$ (in order to show this, one has to use that $E\eta_{i,k}(x) = 0$).

Finally, for the term $V_n^{(3)}$, we have by assumption N4:

$$
|V_n^{(3)}| \leq O\left(\frac{1}{h_n^4}\right) \sup_{x \in \{x_1, \ldots, x_p\}} |m(x) - \hat{m}(x)|^\lambda = O_p\left(h_n^{-4} n^{-3/2}\right) = o_P\left(n^{-1} h_n^{-1/2}\right),
$$

because in our setup $\hat{m}(x) = m(x) + O_P(n^{-1/2})$. $\qquad \square$

References

Dette, H. and Neumeyer, N. (2000). A note on a specification test of independence, *Metrika* **51**: 133–144.

Einmahl, J. H. J. and Van Keilegom, I. (2008a). Specification tests in nonparametric regression, *Journal of Econometrics* **143**: 88–102.

Einmahl, J. H. J. and Van Keilegom, I. (2008b). Tests for independence in nonparametric regression, *Statistica Sinica* **18**: 601–615.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis* **14**: 1–16.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *The Annals of Statistics* **19**(2): 778–796.

Hlávka, Z., Hušková, M. and Meintanis, S. G. (2011). Tests for independence in non-parametric heteroscedastic regression models, *Journal of Multivariate Analysis* **102**(4): 816–827.

Hušková, M. and Meintanis, S. G. (2009). Goodness-of-fit tests for parametric regression models based on empirical characteristic functions, *Kybernetika* **45**(6): 960–971.

Neumeyer, N. (2009). Testing independence in nonparametric regression, *Journal of Multivariate Analysis* **100**: 1551–1566.

Van Keilegom, I., González Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution, *TEST* **17**(2): 401–415.

Zheng, J. X. (1997). A consistent specification test of independence, *Journal of Nonparametric Statistics* **7**: 297–306.

Chapter 5

Conclusion

Statistics is sometimes wrongly seen only as a collection of standard methods but, in real-life applications, standard statistical methods must be tweaked and modified. Therefore, proper statistical data analysis can not be carried out without a deep understanding of the theoretical background and, on the other hand, new statistical applications motivate new directions in theoretical research.

Apart of introducing and summarizing the content of the five papers included in the Appendix, this habilitation thesis contains several extensions of standard nonparametric regression methods concerning all stages of the data analytical process. I believe that these papers and its extensions are interesting both from practical and theoretical point of view.

From the practical point of view, the optimal designs from Chapter 2 improve the planning of nonparametric regression experiments, the Kalman filtering techniques from Section 3.4 may be used for an online monitoring of option markets, and specification tests from Chapter 4 may be used to test correct specification of Analysis of Variance (ANOVA) models.

From the theoretical point of view, we have described an interesting special case of a degenerated constrained nonparametric regression estimator in Chapter 3 and the optimization and Fourier techniques from Chapters 2 and 4 will hopefully inspire further generalizations using, e.g., a different nonparametric regression estimator or more explanatory variables.

C. R. Rao, in a preface to his *Linear methods of statistical inference and its applications (Wiley, 1965)*, described statistics as the "new technology" of the 20th century. Almost 50 years later, statistics can not be characterized as an entirely new technology but there still continuously appear new real-life problems assuring that mathematical statistics remains a "somewhat old but up-to-date" branch of mathematics capable of adapting to new challenges and possibilities.

Attached papers

Hlávka, Z. (2011). On nonparametric estimators of location of maximum, *Acta Universitatis Carolinae - Mathematica et Physica* **52**(1): 5–13.

Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities, *Journal of Econometrics* **150**(1): 1–15.

Hlávka, Z. (2006a). Fast algorithm for nonparametric arbitrage-free SPD estimation, *Computational Statistics & Data Analysis* **51**(4): 2339–2349.

Hlávka, Z. and Svojík, M. (2009). Application of extended Kalman filter to SPD estimation, *in* W. K. Härdle, N. Hautsch and L. Overbeck (eds), *Applied quantitative finance*, Springer, Berlin, pp. 233–247.

Hlávka, Z., Hušková, M. and Meintanis, S. G. (2011). Tests for independence in non-parametric heteroscedastic regression models, *Journal of Multivariate Analysis* **102**(4): 816–827.