Carroll, Raymond J.; Härdle, Wolfgang; Mammen, Enno

Working Paper

Estimation in an additive model when the components are linked parametrically

SFB 373 Discussion Paper, No. 1999,1

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Carroll, Raymond J.; Härdle, Wolfgang; Mammen, Enno (1999) : Estimation in an additive model when the components are linked parametrically, SFB 373 Discussion Paper, No. 1999,1, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10055992

This Version is available at:
https://hdl.handle.net/10419/61777

ESTIMATION IN AN ADDITIVE MODEL WHEN THE COMPONENTS ARE LINKED PARAMETRICALLY

Raymond J. Carroll, Wolfgang Härdle and Enno Mammen*

January 14, 1999

Abstract

Motivated by a nonparametric GARCH model we consider nonparametric additive regression and autoregression models in the special case that the additive components are linked parametrically. We show that the parameter can be estimated with parametric rate and give the normal limit. Our procedure is based on two steps. In the first step nonparametric smoothers are used for the estimation of each additive component without taking into account the parametric link of the functions. In a second step the parameter is estimated by using the parametric restriction between the additive components. Interestingly, our method needs no undersmoothing in the first step.

Key words and phrases: Additive Models; Asymptotics; Autoregression; Finance; GARCH Models; Measurement Error; Nonparametric Regression; Time Series.

Short title. Parametrically Linked Additive Models

*Raymond J. Carroll is University Distinguished Professor of Statistics, and Professor of Nutrition and Toxicology, Department of Statistics, Texas A&M University, College Station, TX 77843–3143. Carroll's research was supported by a grant from the National Cancer Institute (CA–57030) and through the Texas A&M Center for Environmental and Rural Health by a grant from the National Institute of Environmental Health Sciences (P30-ESO9106). Wolfgang Härdle is Professor of Econometrics, at the Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany. Enno Mammen is Associate Professor of Mathematical Statistics, Institut für Angewandte Mathematik, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany. Research of this paper was started when R. Carroll and E. Mammen were visiting the Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse", Humboldt-Universität zu Berlin. Research of Enno Mammen was supported by the Deutsche Forschungsgemeinschaft, project MA1026/6-1.

ESTIMATION IN AN ADDITIVE MODEL WHEN THE COMPONENTS ARE LINKED PARAMETRICALLY

Abstract

Motivated by a nonparametric GARCH model we consider nonparametric additive regression and autoregression models in the special case that the additive components are linked parametrically. We show that the parameter can be estimated with parametric rate and give the normal limit. Our procedure is based on two steps. In the first step nonparametric smoothers are used for the estimation of each additive component without taking into account the parametric link of the functions. In a second step the parameter is estimated by using the parametric restriction between the additive components. Interestingly, our method needs no undersmoothing in the first step.

Key words and phrases: Additive Models; Asymptotics; Autoregression; Finance; GARCH Models; Measurement Error; Nonparametric Regression; Time Series.

Short title. Parametrically Linked Additive Models

1 INTRODUCTION

Additive nonparametric regression models have found wide use in statistics (Hastie and Tibshirani, 1990), and remain an area of vigorous research (Opsomer and Ruppert,1997; Opsomer, 1998; Linton, Mammen and Nielsen, 1997; Linton, 1997; Fan, Härdle and Mammen, 1998). This paper explores a variant of the problem in which the components of the additive model are linked parametrically.

Let Y be a scalar response and $\mathbf{X} = (X_1, \ldots, X_J)^T$ a vector of regressors. In the ordinary additive model,

$$E(Y|\mathbf{X}) = E(Y) + \sum_{j=1}^{J} m_j(X_j), \tag{1}$$

where for identifiability the component functions satisfy $E\{m_j(X_j)\} = 0$.

Hafner (1998) describes a problem in finance where the component functions $m_1(\cdot), \ldots, m_J(\cdot)$ are linked by a parameter, see Section 4 for more details. Specifically, for a scale parameter β_0, and for $j \geq 2$,

$$m_j(x) = \beta_0^{j-1} m_1(x). \tag{2}$$

Our purpose here is to estimate both the parameter β_0 and the base function $m_1(x)$. Among the many possibilities, one stands out as relatively straightforward, namely to estimate the component functions in the general model (1) and somehow "shrink" them to the model (2). One method we pursue, which is based on considerations from the field of errors in variables, is computationally straightforward, with the estimate of β_0 having an easily estimated standard error. In addition, the estimator has the pleasing property that the fit to model (1) can be done in a standard fashion, without the need for any undersmoothing to insure that the estimate of β_0 converges at standard parametric rates. The analysis of this method leads to a second method which is equally simple to compute.

The paper is organized as follows. In Section 2, we define the methods used. Section 3 states the asymptotic properties when the additive component functions of (1) are estimated by a nonparametric smoother that allows a stochastic expansion of order $o_P(n^{-1/2})$. Section 4 provides details of the motivating example from a nonparametric GARCH model. This model can be approximated by a nonparametric additive autoregression model. Section 5 extends our results from regression to autoregression. In particular, it discusses the set up of the approximating model for our motivating example in Section 4. We discuss conditions under which estimates of the additive component allow a stochastic expansion of order $o_P(n^{-1/2})$. This is done for the integration estimates (see Linton

and Nielsen,1994; Tjøstheim and Auestad, 1994). Furthermore, the asymptotic theory of Section 3 is extended to autoregression. All proofs are in the appendix.

2 THE METHODS

The data are $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{iJ})^T$. We describe here the two methods used in this paper. Our methods rely on estimates $\{\widehat{m}_1(\cdot), \ldots, \widehat{m}_J(\cdot)\}$ from the model (1), specific examples of which are discussed later in this article.

2.1 A Method from Errors–in–Variables

The first method has a natural interpretation as a functional errors-in-variables method (Fuller, 1987). To explain this, first fix x. Then ignoring bias and other technical details, it is generally the case that for some constants $c_n \to 0$ and functions $w_j(x)$, the functions $\{\widehat{m}_j(x)\}_{j=1}^J$ form a set of nearly independent, nearly normal random variables: $\widehat{m}_j(x) \sim \text{Normal}\{\beta_0^{j-1} m_1(x),\ c_n/w_j(x)\}$. If one only had this single fixed x, then the unknowns would be β_0 and $m_1(x)$, and they could be estimated by minimizing in β and $m_1(x)$

$$\sum_{j=1}^{J} w_j(x) \left\{ \widehat{m}_j(x) - \beta^{j-1} m_1(x) \right\}^2 . \tag{3}$$

This is an errors-in-variables model because we have the multivariate "response" $\{\widehat{m}_2(x), \ldots, \widehat{m}_J(x)\}$ with means $\{\beta_0 m_1(x), \ldots, \beta_0^{J-1} m_1(x)\}$, but instead of observing $m_1(x)$ we only have the error-prone "predictor" $\widehat{m}_1(x)$. Equation (3) thus proposes estimating β_0 and $m_1(x)$ by an asymptotic form of maximum likelihood. This is exactly the classical functional errors-in-variables estimate.

Continuing with fixed x, for a given β the minimizer of (3) is $G\{\widehat{m}_1(x), \ldots, \widehat{m}_J(x), \beta\}$, where

$$G\{m_1(x), \ldots, m_J(x), \beta\} = \sum_{j=1}^{J} w_j(x) m_j(x) \beta^{j-1} / \sum_{j=1}^{J} w_j(x) \beta^{2j-2}.$$

Note that $G\{m_1(x), \ldots, m_J(x), \beta_0\} = m_1(x)$ since $m_j(x) = \beta_0^{j-1} m_1(x)$.

Hence, if we worked only with a fixed x, β_0 would be estimated to minimize

$$\sum_{j=1}^{J} w_j(x) \left[\widehat{m}_j(x) - \beta^{j-1} G\left\{ \widehat{m}_1(x), \ldots, \widehat{m}_J(x), \beta \right\} \right]^2 . \tag{4}$$

Summing (4) over all the data suggests that we estimate β by minimizing

$$\sum_{i=1}^{n} \sum_{k=1}^{J} \sum_{j=1}^{J} w_j(X_{ik}) \left[\widehat{m}_j(X_{ik}) - \beta^{j-1} G\left\{ \widehat{m}_1(X_{ik}), \ldots, \widehat{m}_J(X_{ik}), \beta \right\} \right]^2 . \tag{5}$$

2.2 A Method Based on Least Squares

A special case is illuminating and suggests a second method. Suppose that $J = 2$ and that the variances of $\widehat{m}_1(x)$ and $\widehat{m}_2(x)$ are asymptotically the same, so that we can set $w_1(\cdot) = w_2(\cdot) \equiv 1$. Then (5) reduces to minimizing in β

$$(1+\beta^2)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{2} \{\widehat{m}_2(X_{ij}) - \beta \widehat{m}_1(X_{ij})\}^2 . \tag{6}$$

The leading term $(1+\beta^2)^{-1}$ in (6) plays an important role in the usual errors–in–variables problem, but here we have a different situation, because the "errors–in–the–variables" are small asymptotically due to the fact that the "error–prone predictor $\widehat{m}_1(x)$" has an error which is asymptotically small. This suggests that one might find a reasonable estimate if one simply removes this leading term and minimizes instead

$$\sum_{i=1}^{n} \sum_{j=1}^{2} \{\widehat{m}_2(X_{ij}) - \beta \widehat{m}_1(X_{ij})\}^2 . \tag{7}$$

While there are numerical differences between our method (7) and the least squares method (6) (generally, the latter yields larger estimates for β_0), asymptotically the two lead to the same distribution for $\widehat{\beta}$, see Section 7.2 for a sketch.

The method (7) can be obtained alternatively by replacing $G(\cdot)$ in (5) by $\widehat{m}_1(\cdot)$. Thus the least squares method minimizes

$$\sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} w_j(X_{ik}) \left\{ \widehat{m}_j(X_{ik}) - \beta^{j-1} \widehat{m}_1(X_{ik}) \right\}^2 . \tag{8}$$

For $J \geq 3$, (8) leads to an estimator which is asymptotically different from the solution to (5). We explore the differences numerically in Section 4.

2.3 Alternatives

There are a host of possible alternative methods.

As we have described in section 2.1, the component functions can be looked at as if they came from a (nonlinear) errors–in–variables model. There is a huge literature on the topic of errors–in–variables, and at least in principle one can develop many alternatives to the two methods we have described, either based on small error considerations (Amemiya and Fuller, 1988; Carroll, Ruppert and Stefanski, 1995) or on simulation (Cook and Stefanski, 1995).

Alternatively, the method (8) can be looked upon as regressing $\widehat{m}_j(\cdot)$ for $j \geq 2$ on $\widehat{m}_1(\cdot)$. This could be expanded to doing all possible regressions of $\widehat{m}_j(\cdot)$ on $\widehat{m}_k(\cdot)$ for $j > k$.

We have not explored these alternatives, although our methods of argument can in principle be used to obtain limit distributions for them.

3 ASYMPTOTIC THEORY FOR REGRESSION DATA

It is possible to obtain the asymptotic distribution of the estimates of β for different estimates of the component functions $m_j(x)$. We assume that the estimates fulfill the stochastic expansion (9), given below. In Section 5 we will show that this expansion holds for the integration estimate.

Let $K(\cdot)$ be a symmetric density function chosen without loss of generality to have variance equal to one. Define $\epsilon = Y - E(Y) - \sum_{j=1}^{J} m_j(X_j)$ and let $f_j(x)$ be the density of the X_{ij}, $i = 1, \ldots, n; j = 1, ..J$. Let $\mathbf{X}_i$ be the vector $(X_{i1}, \ldots, X_{iJ})$. Let $h \to 0$ be the bandwidth, which is supposed to have the usual rate $h \sim n^{-1/5}$. One of the important features of our theoretical work is that we show that bandwidths of the usual rate can be used to estimate β, and no undersmoothing (i.e. $h \ll n^{-1/5}$) is necessary.

We suppose that the estimators have the asymptotic expansion

$$\begin{aligned} \widehat{m}_j(x) &= m_j(x) + (1/2)h^2\beta^{j-1}r(x) + n^{-1}\sum_{i=1}^{n} K_h(X_{ij} - x)u_j(\mathbf{X}_i, x)\epsilon_i \\ &\quad + n^{-1}\sum_{i=1}^{n} v_j(\mathbf{X}_i, x)\epsilon_i + n^{-1}\sum_{i=1}^{n} t_j(\mathbf{X}_i) + o_P(n^{-1/2}), \end{aligned} \tag{9}$$

where r, u_j, v_j and t_j are some functions, $K_h(v) = h^{-1}K(v/h)$. Furthermore, it is supposed that $E\{t_j(\mathbf{X}_i)\} = 0$.

We first consider the least squares estimator $\widehat{\beta}_{LS}$ minimizing (8). Make the definitions

$$\begin{aligned} \mathcal{H}_1(\mathbf{X}, \beta_0) &= \sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{\left[v_j^*(\mathbf{X}) - \beta_0^{j-1}v_1^*(\mathbf{X})\right]\right. \\ &\qquad + \sum_{k=1}^{J}[w_j(X_j)m_1(X_j)f_k(X_j)u_j(\mathbf{X}, X_j) \\ &\qquad\quad - \beta_0^{j-1}w_j(X_1)m_1(X_1)f_k(X_1)u_1(\mathbf{X}, X_1)\Big]\Big\}, \\ \mathcal{H}_2(\mathbf{X}, \beta_0) &= \sum_{j=1}^{J}(j-1)\beta_0^{j-2}\left\{s_j^*\left[t_j(\mathbf{X}) - \beta_0^{j-1}t_1(\mathbf{X})\right]\right\}; \\ D_{LS} &= \sum_{j=1}^{J}\sum_{k=1}^{J}\left\{(j-1)\beta_0^{j-2}\right\}^2 E\left\{w_j(X_k)m_1^2(X_k)\right\}; \\ v_j^*(\mathbf{x}) &= E\left[\sum_{k=1}^{J} w_j(X_k)m_1(X_k)v_j(\mathbf{x}, X_k)\right]; \end{aligned}$$

$$
\begin{aligned}
s_j^* &= E\left[\sum_{k=1}^{J} w_j(X_k) m_1(X_k)\right]; \\
\sigma^2(\mathbf{X}) &= \mathrm{Var}(\mathrm{Y}|\mathbf{X}).
\end{aligned}
$$

THEOREM 1 *Let $h \sim n^{-1/5}$. Then under conditions (A) listed in the appendix, $n^{1/2}(\widehat{\beta}_{LS} - \beta_0)$ is asymptotically normally distributed with mean zero and variance*

$$\Sigma_{LS} = E\left[\left\{\sigma^2(\mathbf{X})\mathcal{H}_1^2(\mathbf{X}, \beta_0) + \mathcal{H}_2^2(\mathbf{X}, \beta_0)\right\} / D_{LS}^2\right].$$

Now we turn to the errors-in-variables estimator $\widehat{\beta}_{EIV}$, which minimizes (5). Make the following definitions.

$$
\begin{aligned}
s(\beta, x) &= \{\sum_{j=1}^{J} w_j(x)\beta^{2j-2}\}^{-2}; \\
D_{EIV} &= \sum_{j=1}^{J}\sum_{k=1}^{J} E[w_j(X_k)s(\beta_0, X_k)m_1^2(X_k)\{\sum_{\ell=1}^{J}(\ell - j)w_\ell(X_k)\beta_0^{2\ell+j-4}t\}^2]; \\
R_j(x, \beta) &= \sum_{\ell=1}^{J} w_\ell(x)\left\{(2\ell-2)\beta^{2\ell-3}m_j(x) - (\ell + j - 2)\beta^{\ell+j-3}m_\ell(x)\right\}; \\
\mathcal{M}(\mathbf{x}, \beta) &= E\left[\sum_{j,k,\ell=1}^{J} w_j(X_k)s(\beta, X_k)R_j(X_k, \beta)w_\ell(X_k)\{\beta^{2\ell-2}v_j(\mathbf{x}, X_k)\right. \\
&\qquad \left. -\beta^{\ell+j-2}v_\ell(\mathbf{x}, X_k)\}\right]; \\
T_1(j, k, \ell, \beta, \mathbf{x}) &= \beta^{2\ell-2}w_j(x_j)w_\ell(x_j)s(\beta, x_j)R_j(x_j, \beta)f_k(x_j)u_j(\mathbf{x}, x_j); \\
T_2(j, k, \ell, \beta, \mathbf{x}) &= \beta^{\ell+j-2}w_j(x_\ell)w_\ell(x_\ell)s(\beta, x_\ell)R_j(x_\ell, \beta)f_k(x_\ell)u_\ell(\mathbf{x}, x_\ell); \\
\mathcal{D}_{j,\ell}(\beta) &= E\left\{\sum_{k=1}^{J} w_j(X_k)s(\beta, X_k)R_j(X_k)w_\ell(X_k)\right\}; \\
\mathcal{G}(\mathbf{X}, \beta) &= \sum_{j,\ell=1}^{J} \mathcal{D}_{j,\ell}(\beta)\left\{\beta^{2\ell-2}t_j(\mathbf{X}) - \beta^{\ell+j-2}t_\ell(\mathbf{X})\right\}; \\
\mathcal{H}_3(\mathbf{x}, \beta) &= \sum_{j,k,\ell=1}^{J} \{T_1(j, k, \ell, \beta, \mathbf{x}) - T_2(j, k, \ell, \beta, \mathbf{x})\}.
\end{aligned}
$$

THEOREM 2 *With $h \sim n^{-1/5}$, under the same regularity conditions as Theorem 1, $n^{1/2}(\widehat{\beta}_{EIV} - \beta_0)$ is asymptotically normally distributed with mean zero and variance*

$$\Sigma_{EIV} = E\left[\sigma^2(\mathbf{X})\left\{\mathcal{H}_3(\mathbf{X}, \beta_0) + \mathcal{M}(\mathbf{X}, \beta_0)\right\}^2 + \mathcal{G}^2(\mathbf{X}, \beta_0)\right] / D_{EIV}^2.$$

While the additive functions are naturally linked in our model through (2), it is conceivable to have other functions, e.g., $m_j(x) = F_j\{m_1(x), \beta_0\}$. Our methods can be used to analyze such models

should they arise. For example, the analogue of the least squares criterion (8) would minimize

$$\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})[\widehat{m}_j(X_{ik}) - F_j\{\widehat{m}_1(X_{ik}),\beta\}]^2.$$

Our methods can be used to analyze this estimator, but there is one catch. Unless $F_j\{m_1(x),\beta\} = e_j(\beta)m_1(x)$ for some function $e_j(\beta)$, the bandwidth condition $h \sim n^{-1/5}$ no longer suffices because the bias in the nonparametric regressions has a role. In our example the first order bias terms in $\widehat{m}_j(X_{ik}) - F_j\{\widehat{m}_1(X_{ik}),\beta\}$ cancel whereas for general F_j they do not. Cancelation of the bias terms can be forced by replacing F_j by an appropriately (data adaptively) chosen function $\widetilde{F}_j = F_j + O_P(h^2)$. Otherwise, without replacing F_j, we could require undersmoothing, so that $h = o(n^{-1/4})$. In fact, what basically happens in this case is that there exists $\beta_h^* = \beta_0 + O(h^2)$ such that $n^{1/2}(\widehat{\beta} - \beta_h^*)$ has a limit distribution similar to that described by Theorem 1. By imposing the condition $h = o(n^{-1/4})$ we can replace β_h^* by β_0.

4 NONPARAMETRIC MODELS FOR FINANCIAL TIME SERIES

The recent development of nonlinear time series analysis is primarily due to the efforts to overcome the limitations of linear models such as autoregressive moving-average (ARMA) models of Box and Jenkins (1976) in real applications. It has long been recognized that financial time series models that incorporate clusters of volatilities are more appropriate than ARMA specifications. We consider here as a motivating example an application of nonlinear time series analysis to foreign exchange high frequency data.

For these data the autoregressive heteroscedastic models (ARCH) by Engle (1982) have been extensively studied. An ARCH model for time series $\{Y_t\}$ with ARCH error term of order q is defined through $Y_t = \sigma_t\xi_t$, where ξ_t are independent mean zero and variance one random variables and $\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 \ldots + \alpha_q Y_{t-q}^2$, with $\omega > 0, \alpha_i \geq 0, i = 1,\ldots,q$.

In foreign exchange data it has been found that the order q has to be selected quite high to fit the model well, see Bollerslev (1986). The reason are volatility clusters, i.e. the conditional variances are highly correlated. An ARMA like model for the squared observations was therefore proposed for σ_t^2 in Bollerslev (1986):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\alpha_i Y_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2. \tag{10}$$

Models of this type are called GARCH (p, q) models. For a general discussion of GARCH models see also Bollerslev, Engle and Nelson (1994). Although this model class showed better fitting properties it was soon criticized that the dependence of past observations is treated in a symmetric way: Positive and negative shocks of $Y_{t-1}, \ldots, Y_{t-q}$ have the same influence on the volatility of the current period. The forced symmetry of past shocks was one of the primary motivations for non- and semiparametric extensions of ARCH models. Based on the QTARCH model of Gourieroux and Monfort (1992), Härdle and Tsybakov (1997) considered the CHARN model $Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\xi_t$, which was applied to DEM/USD exchange rates by Bossaerts, Härdle and Hafner (1996) and extended to the multivariate case by Härdle, Tsybakov and Yang (1998). In their analysis volatility clusters and a strong asymmetry of the *news impact function* $\sigma(Y_{t-1})$ became apparent and motivated research on the following semiparametric extension of (10),

$$Y_t = \sigma_t \xi_t; \tag{11}$$

$$\sigma_t^2 = g(Y_{t-1}) + \beta \sigma_{t-1}^2. \tag{12}$$

The model we consider in this paper is motivated by (11) and (12). By repeated use of equation (12), we can write $\sigma_t^2 = \sum_{j=1}^{\infty} \beta^{j-1} g(Y_{t-j})$. An approximation of this model with a finite number J of lags reads

$$Y_t^2 = \sum_{j=1}^{J} \beta^{j-1} g(Y_{t-j}) + \epsilon_t, \tag{13}$$

with $\epsilon_t = Y_t^2 - \sigma_t^2$.

We now illustrate the application of model (13) to foreign exchange rates. The behavior of foreign exchange (FX) rates has been subject of many recent investigations. A correct understanding of the foreign exchange rate dynamics has important implications for international asset pricing theories, the pricing of contingent claims and policy–oriented questions. The foreign exchange market is decentralized with the main trading locations being the Far East (mainly Tokyo, Singapore and Hong Kong), Europe (London and Frankfurt) and North America (New York). It is an electronic market, active 24 hours a day. Banks act as market makers and place bid– and ask–quotes on the screen. Central information collectors such as Reuters provide the quotes for the market makers. Actual trade takes place over the phone. This is the reason why there is no information about actual prices and trading volume. High frequency financial data analysis has become a broad research field during the last decade. This is due to improved real–time information systems, relatively cheap data supply by banks and research institutions and improved storing facilities. The

data set HFDF93 on which the following analysis is based was acquired from Olsen and Associates, Zürich. It contains bid and ask quotes for the rates Deutsche Mark against US Dollar (DEM/USD), during the time Oct 1 1992 and February 16 1993. For each pair of bid– and ask-quotes, the time in GMT, the quoting bank and the location of the bank are recorded. The quotes are collected from the Reuters FXFX page, which is considered to be a broad but not 'complete' data supply. For more information about this data set, see Dacorogna, Müller, Nagler, Olsen and Pictet (1993) and, more generally for information about FX rate data suppliers and intra–daily FX data, Goodhart and Figliuoli (1991). Figure **??** shows a plot of the DEM/USD returns. Our data set contains 10000 data values. A kernel density estimate of the returns is shown in Figure **??**.

For the data set we calculated backfitting and integration estimates, see the next section for a description of these estimates. As discussed in the last section, in a first step estimation was done in an additive model (without assumed links on the components g_j) given by $Y_t^2 = \sum_{j=1}^J g_j(Y_{t-j}) + \epsilon_t$. In this model we chose $J = 5$ lags. Figures **??** and **??** show the resulting backfitting and integration estimates of the additive components. The integration estimate was calculated by fitting the full dimensional estimate on a grid of 26^5 points. This was done to save computation time. For all kernel estimates we choose the empirical standard deviation as bandwidth. Next we fitted the nonparametric GARCH model (13) with $m_j = \beta^{j-1} g$. For the estimation of β we used our method from errors-in-variables (see Section 2.1) and our least squares method (see Section 2.2). The resulting estimates were 0.793 and 0.736 (for the backfitting estimate) and 0.892 and 0.779 (for the integration estimate), respectively. There are some differences between these estimates, largely along the lines of what one would expect from Figures **??** and **??**. One would expect from these figures that the least squares estimate of β_0 would be smaller than the errors–in–variables estimate, since the latter compares $j = 2, 3, 4, 5$ to a weighted average of $j = 1, ..., 5$, which is closer to the results for $j = 2, 3, 4, 5$ than to the result for $j = 1$.

The estimates $\widehat{m}_1, \ldots, \widehat{m}_J$ and $\widehat{\beta}$ can be used to construct an estimate of m_1 that takes into account that the additive components are linked. This can be done by using the averaged estimate

$$\widehat{m}_1^*(x) = \sum_{j=1}^J \widehat{c}_j \widehat{\beta}^{-(j-1)} \widehat{m}_j(x) / \sum_{j=1}^J \widehat{c}_j, \tag{14}$$

where $\widehat{c}_j = \widehat{\beta}^{2(j-1)}$, see also (20) and the discussion following Theorem 7 in the next section. Here $\widehat{\beta}$ denotes our estimate based on the method from errors-in-variables or the least squares method, respectively. Figures **??**, **??**, **??** and **??** show plots of the estimates $\widehat{m}_1^*$ and $\widehat{m}_j^* = \widehat{\beta}^{j-1} \widehat{m}_1^*$. The

plots differ slightly for the different methods. This must be explained by the fact that model (13) only approximates the underlying model. In particular use of the method from errors-in-variables leads to more asymmetric news impact functions.

Figure **??** shows how the nonparametric estimates depend on the chosen number J of lags. It compares the backfitting estimates of $\widehat{m}_1^*$ for different number of lags ($J = 5, 10, 20, 30$). In these calculations, β was estimated by the method from errors-in-variables. The estimated values are 0.789 ($J = 5$), 0.799 ($J = 10$), 0.809 ($J = 20$), and 0.788 ($J = 30$). The nonparametric estimates (besides small differences of the estimate for $J = 5$) are nearly indistinguishable. So, we conclude that in this data example model (13) approximates the nonparametric GARCH model (11) and (12) reasonably well.

Of course, this analysis should be taken as illustrative, since it focuses on short–term dependencies. For modeling of long range dependencies like daily or weekly dependence, more complicated models may be needed.

The next section discusses asymptotics of these estimates in an autoregression model.

5 ASYMPTOTICS FOR AUTOREGRESSION

In this section we show that for the integration estimate the expansion (9) holds. The integration estimate has been introduced in Tjøstheim and Auestad (1994) and Linton and Nielsen (1995) for the estimation of additive nonparametric components $m_j(\cdot)$ in an additive model. We will do this for the time series setup of Section 4. We suppose that a stationary time series $X_0, \ldots, X_n$ is observed. We suppose that $E(X_{J+1}|X_J, \ldots, X_1) = \alpha + m_1(X_J) + \beta m_1(X_{J-1}) + \ldots \beta^{J-1} m_1(X_1)$, where for a weight function w the function m_1 satisfies $E\ w(X_i)m_1(X_i) = 0$. For this setup the expansion (9) is given by

$$\begin{aligned} \widehat{m}_j(x) &= m_j(x) + (1/2)h^2\beta^{j-1}r(x) + n^{-1}\sum_{i=J+1}^{n} K_h(X_{i-j} - x)u_j(\mathbf{X}_i, x)\epsilon_i \\ &\quad + n^{-1}\sum_{i=J+1}^{n} v_j(\mathbf{X}_i, x)\epsilon_i + n^{-1}\sum_{i=J+1}^{n} t_j(\mathbf{X}_i) + o_P(n^{-1/2}). \end{aligned} \tag{15}$$

where $\mathbf{X}_i$ is the vector $(X_{i-1}, \ldots, X_{i-J})$.

In this section we study the validity of (15) for the integration estimate (see Theorem 3). For simplicity of notation, we will do this only for the case $J = 2$. Then, we will consider errors-in-variables and least squares estimation of β using arbitrary estimates of m_j that fulfill (15). We

will show that for these estimates of β analogous asymptotic results apply as in the regression set up (compare Theorems 1 and 2 with Theorems 3 and 4). An improved estimate of m_1 can be constructed by use of the estimates of $\beta, m_1, \ldots, m_J$. Asymptotics for this estimate is described in Theorem 6.

We come now to the check of (15) for the integration estimate. In a first step this estimate uses a full dimensional local linear fit $\widehat{m}_{LL}$ of the data, i.e.the preliminary estimate $\widehat{m}_{LL}$ is defined as θ_0 where the vector $\theta = (\theta_0, \theta_1, \theta_2)^T$ is defined by

$$\sum_{i=3}^{n} K_{h_1}(X_{i-1} - x_1)\ K_{h_2}(X_{i-2} - x_2)[Y_i - \theta^T \xi_i(x)]\xi_i(x) = 0. \tag{16}$$

Here $\xi_i(x)$ denotes the vector $(1, \frac{X_{i-1}-x_1}{h_1}, \frac{X_{i-2}-x_2}{h_2})^T$. The integration estimate $\widehat{m}_1^I$ of m_1 is defined as

$$\widehat{m}_1^I(x_1) = \widetilde{m}_1^I(x_1) - n^{-1}\sum_{i=1}^{n} w(X_i)\widetilde{m}_1^I(X_i)/\frac{1}{n}\sum_{i=1}^{n} w(X_i). \tag{17}$$

Here, w is a weight function. The estimate $\widetilde{m}_1^I$ is achieved by summing out an argument of the full dimensional estimate $\widehat{m}_{LL}$

$$\widetilde{m}_1^I(x_1) = n^{-1}\sum_{i=1}^{n} w(X_i)\widehat{m}_{LL}(x_1, X_i)/n^{-1}\sum_{i=1}^{n} w(X_i). \tag{18}$$

For simplicity, here the same weight function has been used as in (17). The estimate $\widetilde{m}_2^I$ is achieved by summing out the other argument of the full dimensional estimate $\widehat{m}_{LL}$. Our first result shows that $\widehat{m}_1^I$ and $\widetilde{m}_2^I$ satisfy (15).

THEOREM 3 *Suppose $J = 2$. Under the regularity conditions (B) listed in the appendix, and if the bandwidths fulfill that $h_1 = h_2$, $h_2^4 = o(n^{-1/2})$ and $(\log n)^2[\sqrt{n}\ h_1 h_2]^{-1} \to 0$, the estimate $\widehat{m}_1^I$ has a stochastic expansion (15) with $h = h_1 = h_2$ and*

$$\begin{aligned}
r(x) &= m_1''(x) - \int w m_1'' f [\int wf]^{-1}, \\
u_1(\mathbf{X}_i, x) &= w(X_{i-2}) f(X_{i-2}) p(x, X_{i-2})^{-1}, \\
v_1(\mathbf{X}_i, x) &= -w(X_{i-1}) w(X_{i-2}) f(X_{i-1}) f(X_{i-2}) p(X_{i-1}, X_{i-2})^{-1} [\int wf]^{-1}, \\
t_1(\mathbf{X}_i) &= [w(X_{i-1}) m_1(X_{i-1})][\int wf]^{-1}.
\end{aligned}$$

Here, $\mathbf{X}_i$ denotes the vector $(X_{i-1}, X_{i-2})^T$. Furthermore, f denotes the density of X_i and p is the density of (X_{i-1}, X_i). For $\widehat{m}_2^I$ the expansion (15) holds with the same $r(x)$, the same $v_2(\mathbf{X}_i, x) =$

$v_1(\mathbf{X}_i, x)$ *and with* $t_2(\mathbf{X}_i) = \beta t_1(\mathbf{X}_i)$ *and with*

$$u_2(\mathbf{X}_i, x) = w(X_{i-1}) f(X_{i-1}) p(X_{i-1}, x)^{-1}.$$

In both cases, the expansion (15) holds uniformly for $x \in B$. *The set* B *was introduced in assumption B (iii).*

We conjecture that an expansion of the form (15) holds (uniformly) for the backfitting estimate. In Linton, Mammen and Nielsen (1997), for a version of the backfitting estimate $\widehat{m}_j^{BACK}$, a stochastic expansion has been given. Applied to our setup this expansion is

$$\begin{aligned} \widehat{m}_j^{BACK}(x) &= m_j(x) + (1/2)h^2 \beta^{j-1} r(x) + n^{-1} \sum_{i=J+1}^{n} K_h(X_{i-j} - x) u_j(\mathbf{X}_i, x) \epsilon_i \\ &\quad + O_P(n^{-1/2} \log n), \end{aligned} \tag{19}$$

with an appropriate choice of u_j and where $r(x)$ is as for $\widehat{m}_j^I$, see Theorem 3. The $O_P(n^{-1/2} \log n)$ term in (19) can be explicitly given by an infinite series, see Linton, Mammen and Nielsen (1997). However, it seems to be complicated to show that this term is of order $O_P(n^{-1/2})$ and that it has the form of the terms in (15). For another recent asymptotic treatment of (another version of) the backfitting estimate, see Opsomer (1998) and Opsomer and Ruppert (1997).

We suppose now that we have estimates of m_j that fulfill the expansion (15). These estimates can be used to construct an estimate of β. Asymptotics for this estimate is given in the next two theorems. The first theorem describes least squares estimation of β, see Section 2.2.

THEOREM 4 *Suppose that the regularity conditions (B) hold, and for some estimates* $\widehat{m}_j$ *assume that they fulfill (15) uniformly for* $x \in B$, *where* r, u_j, v_j *and* t_j *are bounded functions with* $E\ t_j(\mathbf{X}_i) = 0$ *and* $\sup_{x,z} |\frac{\partial^3}{(\partial x)^3} u_j(z, x)| < \infty$. *Then the estimate* $n^{1/2}(\widehat{\beta}_{LS} - \beta_0)$ *has an asymptotic normal distribution with mean* 0 *and variance* $D_{LS}^{-2} \sum_{k \in \mathbb{Z}} \mathrm{cov}(U_0, U_k)$ *where* $U_k = \mathcal{H}_1(\mathbf{X}_k, \beta_0)\epsilon_k + \mathcal{H}_2(\mathbf{X}_k, \beta_0)$.

The next theorem gives the asymptotic distribution of the errors-in-variables estimate of β, see Section 2.1.

THEOREM 5 *Suppose that the assumptions of Theorem 4 hold for some estimates* $\widehat{m}_j$. *Then the estimate* $n^{1/2}(\widehat{\beta}_{EIV} - \beta_0)$ *has an asymptotic normal distribution with mean* 0 *and variance* $D_{LS}^{-2} \sum_{k \in \mathbb{Z}} \mathrm{cov}(V_0, V_k)$, *where* $V_k = \{\mathcal{H}_3(\mathbf{X}_k, \beta_0) + \mathcal{M}(\beta_0, \mathbf{X}_k)\}\, \epsilon_k + \mathcal{G}(\mathbf{X}_k, \beta_0)$.

Under our model assumption that $m_j = \beta^{j-1} m_1$ an improved estimate $\widehat{m}_1^*$ of m_1 can be constructed by using the estimates $\widehat{m}_1, \ldots, \widehat{m}_J$ and an estimate $\widehat{\beta}$ of β. This can be done for example

by putting

$$\widehat{m}_1^*(x) = \sum_{j=1}^{J} \widehat{c}_j \widehat{\beta}^{-(j-1)} \widehat{m}_j(x) / \sum_{j=1}^{J} \widehat{c}_j, \tag{20}$$

where $\widehat{c}_j$ are some data adaptive choices of weights. The next theorem gives the asymptotic distribution $\widehat{m}_1^*(x)$.

THEOREM 6 *Suppose (B) and assume that $\widehat{m}_j$ are estimates with*

$$\widehat{m}_j(x) = m_j(x) + (1/2)h^2\beta^{j-1}r(x) + n^{-1}\sum_{i=J+1}^{n} K_h(X_{i-j} - x)u_j(\mathbf{X}_i, x)\epsilon_i + o_P(n^{-1/2}), \tag{21}$$

where u_j is a function with $\sup_z \sup_{|x-y|<\delta} |u_j(z,y) - u_j(z,x| \to 0$ for $\delta \to 0$. Furthermore suppose that $\widehat{\beta}$ is an estimate of β with $\widehat{\beta} = \beta + o_P(n^{-2/5})$ and that for some constants c_j it holds that $\widehat{c}_j = c_j + o_P(n^{-2/5})$. Then $n^{2/5}[\widehat{m}_1^*(x) - m(x)]$ has an asymptotic normal distribution with mean $(1/2)n^{2/5}h^2r(x)$ and variance $n^{1/5}h^{-1}\nu(K)f(x)\ \sum_{j=1}^{J} c_j^2\beta^{-2(j-1)}s_j^2(x)/(\sum_{j=1}^{J} c_j)^2$, where $\nu(K) = \int K^2(u)\,du$ and $s_j^2(x) = E\{\epsilon_{J+1}^2 u_j^2(X_J, \ldots, X_1, x)|X_{J-j+1} = x\}$. The variance is minimized by a choice $\widehat{c}_j$ with $\widehat{c}_j = c\beta^{2(j-1)}s_j^{-2}(x) + o_P(n^{-2/5})$, where c is some constant. In this case $n^{2/5}\{\widehat{m}_1^*(x) - m(x)\}$ has an asymptotic variance $n^{1/5}h^{-1}\nu(K)f(x)/\sum_{j=1}^{J}\beta^{2(j-1)}s_j^{-2}(x)$.

The asymptotic variance of $n^{2/5}[\widehat{\beta}^{-(j-1)}\widehat{m}_j(x) - m(x)]$ is equal to $n^{1/5}h^{-1}\nu(K)f(x)\beta^{-2(j-1)}s_j^2(x)$, see the proof of Theorem 6. Clearly, the asymptotic variance of $n^{2/5}[\widehat{m}_1^*(x) - m(x)]$ is strictly smaller for all j for an asymptotically optimal choice of $\widehat{c}_j$. Typically, application of asymptotically optimal weights requires estimation of $s_j^2(x)$. However, if the weight function w is chosen as indicator function of an interval $[-c, c]$ with c large enough we conjecture that for the backfitting estimate, $s_j^2(x)$ does not depend strongly on j. This motivates in these cases the choice $\widehat{c}_j = \widehat{\beta}^{2(j-1)}$ that leads to a nearly minimal asymptotic variance of $\widehat{m}_1^*(x)$ for all x.

It can be shown that the asymptotic result of Theorem 6 applies under the conditions of Theorems 4 and 5 for the choices $\widehat{\beta} = \widehat{\beta}_{EIV}$ and $\widehat{\beta} = \widehat{\beta}_{LS}$. In particular, this includes estimation of the additive components by the integration estimate or the backfitting estimate.

6 DISCUSSION

The key feature of our model (2) and (13) is that of an additive model with parametrically linked components. We have illustrated the use of the model in a financial time series context, and obtained asymptotic results for autoregression as well as for the usual independent error structure typical in additive models.

The methods are relatively simple. One first uses standard additive model techniques to obtain estimates of the components, and then estimates the linking parameter β_0 by combining the components. It is surprising and pleasing that standard additive model techniques can be used for the first stage without the need for undersmoothing which often occurs in semiparametric modeling.

We have illustrated the use of two such combinations of the component estimates, one an intuitive least squares approach (Section 2.2), and one motivated by errors–in–variables considerations (Section 2.1). At least in principle one would conjecture that the basic idea of estimating β_0 should generalize to such things as generalized linear models. Obtaining asymptotic distributions for such generalizations is likely to be challenging.

An interesting generalization of the model (13) would be to allow for the addition of other parametrically linked terms of the form $\theta^T Z_t$ based on covariates Z_t. In the context of the example, these covariates might include information about previous market behavior, e.g., yesterday's volatility. Again, while the ideas may seem straightforward, actually obtaining asymptotic results may well prove to be difficult.

REFERENCES

Amemiya, Y. and Fuller, W. A. (1988). Estimation for the nonlinear functional relationship. *Annals of Statistics*, 16, 147-160.

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

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models. In *Handbook of Econometrics. Vol. IV*, editors R.F. Engle and D. L. McFadden. *Elsevier Science*, 2959 - 3038.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Springer, New York, Berlin, Heidelberg.

Bossaerts, P., Härdle, W. and Hafner, C. (1996). Foreign exchange rates have surprising volatility. *In: Athens Conference on Applied Probability and Time Series,* Vol. 2, editor P. Robinson, *Lecture Notes in Statistics* 115. Springer, Heidelberg, Berlin and New York.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

Cook, J. and Stefanski, L. A. (1995). A simulation extrapolation method for parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314–1328.

Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B. and Pictet, O. V. (1993). A geographical model for daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12, 413–426.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50, 987–1008.

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low dimensional components in additive models. *Annals of Statistics*, 26, 943–971..

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

Goodhart, C. and Figliuoli, L. (1991). Every minute counts in financial markets. *Journal of International Money and Finance*, 10, 23–52.

Gouriéroux, C. and Monfort, A. (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, 52, 159–199.

Hafner, C. M. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility.* Physica, Heidelberg and New York.

Hastie, T. J. and Tibshirani, R. (1990), *Generalized Additive Models.* Chapman and Hall, London.

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

Härdle, W., Tsybakov, A. and Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning & Inference*, 68, 221-245.

Linton, O. B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, 84, 469–473.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82, 93-101.

Linton, O. B., Mammen, E. and Nielsen, J. (1997). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Preprint*.

Opsomer, J. D. (1998). On the existence and asymptotic properties of backfitting estimators. *Annals of Statistics*, to appear.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25, 186 - 211.

Tjøstheim, D. and Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, 89, 1398 - 1409.

7 APPENDIX

7.1 Assumptions

Condition A

(i) $(X_{11}, \ldots, X_{1J}, Y_1), \ldots, (X_{n1}, \ldots, X_{nJ}, Y_n)$ is an i.i.d. sequence with $E(Y_i|X_{i1}, \ldots, X_{iJ}) = \alpha + m_1(X_{i1}) + \ldots + \beta^{J-1} m_J(X_{iJ})$. For identifiability, it is assumed that $E(m_1(X_{ij})) = 0$ for $j = 1, \ldots, J$.

(ii) The weight functions w_j have a continuous derivative and a bounded support B_j, $j = 1, \ldots, J$.

(iii) The expansion (9) holds uniformly for $x \in B_j, j = 1, \ldots, J$ with bounded functions r, u_j, v_j and t_j. The function $u_j(z, x)$ has third partial derivative $\frac{\partial^3}{(\partial x)^3} u_j$ with respect to x, that are uniformly bounded $\sup_{x,z} |\frac{\partial^3}{(\partial x)^3} u_j(z, x)| < \infty$.

(iv) The density p of $(X_1, \ldots, X_J)$ is two times continuously differentiable and on $B_1 \times \ldots \times B_J$ it is bounded away from 0.

(v) The regression function m_1 is four times continuously differentiable.

(vi) The absolute moments $E|\epsilon_i|^q$ are finite for all q. Here, $\epsilon_i = Y_i - E(Y_i|X_{i1}, \ldots, X_{iJ})$.

(vi) The conditional variance $\sigma^2(x)$ is continuous.

(vii) The kernel K is a symmetric probability density with compact support. W.l.o.g. we assume that $\int u^2 K(u)\, du = 1$.

Condition B

(i) $X_1, X_2, \ldots$ is a stationary process that is geometrically strongly mixing, i.e., $\alpha(k) \leq c_0 \rho^k$ for some constants c_0 and $0 < \rho < 1$.

(ii) For all q there exists a constant c_q such that for all indices $i_1, \ldots, i_q$ the density of $(X_{i_1}, \ldots, X_{i_q})$ is bounded by c_q.

(iii) The weight function w has a continuous derivative and a bounded support B.

(iv) The density p of (X_2, X_1) is two times continuously differentiable and on $B \times B$ it is bounded away from 0.

(v) The regression function m_1 is four times continuously differentiable.

(vi) The variables ϵ_i have a finite Laplace transform, i.e. $E \exp(\gamma \epsilon_i) < \infty$ for $|\gamma|$ small enough.

(vii) The conditional variance $\sigma^2(x) = E(\epsilon_i^2 | \mathbf{X}_i = x)$ is continuous.

(viii) The kernel K is a symmetric probability density with compact support. W.l.o.g. we assume that $\int u^2 K(u)\, du = 1$.

7.2 Sketch that (6) and (7) Lead to the Same Limit Distribution for $\widehat{\beta}$ When $J = 2$

It is easily shown that for $\widehat{\beta}_{LS}$, the minimizer of (7),

$$n^{1/2}(\widehat{\beta}_{LS} - \beta_0) \;=\; \frac{n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 m_1(X_{ij}) \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\}}{E\{m_1^2(X_1) + m_1^2(X_2)\}} + o_P(1).$$

For $\widehat{\beta}_{EIV}$, the minimizer of (6), a Taylor series shows that $n^{1/2}\left(\widehat{\beta}_{EIV} - \beta_0\right) = A_n/B_n + o_P(1)$, where

$$\begin{aligned} B_n &= n^{-1} \sum_{i=1}^n \sum_{j=1}^2 \Bigg[\widehat{m}_1^2(X_{ij}) + \frac{2\beta_0}{1+\beta_0^2} \widehat{m}_1(X_{ij}) \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\} \\ &+ \left\{ \frac{4\beta_0^2}{1-\beta_0^2)^2} - \frac{2}{1+\beta_0^2} \right\} \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\} \Bigg] \xrightarrow{p} E\left\{ m_1^2(X_1) + m_1^2(X_2) \right\}, \end{aligned}$$

and $A_n = A_{n1} + A_{n2}$, where

$$\begin{aligned} A_{n1} &= n^{1/2} \sum_{i=1}^n \sum_{j=1}^2 \widehat{m}_1(X_{ij}) \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\} \\ &= n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 m_1(X_{ij}) \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\} + o_P(1); \\ A_{n2} &= \left\{ 2\beta/(1+\beta^2) \right\} n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 \{\widehat{m}_2(X_{ij}) - \beta_0 \widehat{m}_1(X_{ij})\}^2. \end{aligned}$$

Since $A_{n2} = o_P(1)$, this completes the argument.

7.3 Sketch of Proof of Theorem 1

By a Taylor series expansion,

$$\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{J}(j-1)\beta_0^{j-2}w_j(X_{ik})\widehat{m}_1(X_{ik})\left\{\widehat{m}_j(X_{ik})-\beta_0^{j-1}\widehat{m}_1(X_{ik})\right\} \\
&+ n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{J}(j-1)(j-2)\beta_0^{j-3}w_j(X_{ik})\widehat{m}_1(X_{ik}) \\
&\quad \left\{\widehat{m}_j(X_{ik})-\beta_0^{j-1}\widehat{m}_1(X_{ik})\right\} n^{1/2}\left(\widehat{\beta}_{LS}-\beta_0\right) \\
&- n^{-1}\sum_{j=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{J}(j-1)^2\beta_0^{2j-4}w_j(X_{ik})\widehat{m}_1^2(X_{ik})n^{1/2}(\widehat{\beta}_{LS}-\beta_0)+o_P(1).
\end{aligned}$$

The middle term is easily seen to be $o_P(1)$. The last term is easily seen to be $D_{LS}n^{1/2}\left(\widehat{\beta}_{LS}-\beta_0\right)+o_P(1)$, where $D_{LS}=\sum_{j=1}^{J}\left\{(j-1)\beta_0^{j-2}\right\}^2 E\{\sum_{k=1}^{J}w_j(X_{ik})m_1^2(X_{ik})\}$. Finally, the first term has the same behavior as if the leading $\widehat{m}_1(X_{ik})$ were the same as $m_1(X_{ik})$. Making this substitution and invoking (9), since $m_j''(x)=\beta_0^{j-1}m_1''(x)$, we find that the result is asymptotically equivalent to R_1+R_2 where

$$\begin{aligned}
R_1 &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{J}(j-1)\beta_0^{j-2}w_j(X_{ik})m_1(X_{ik}) \\
&\quad \times\sum_{\ell=1}^{n}\epsilon_\ell\left\{K_h(X_{\ell j}-X_{ik})u_j(\mathbf{X}_l,X_{ik})-\beta_0^{j-1}K_h(X_{\ell 1}-X_{ik})u_1(\mathbf{X}_l,X_{ik})\right\}, \\
R_2 &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{J}(j-1)\beta_0^{j-2}w_j(X_{ik})m_1(X_{ik}) \\
&\quad \times\sum_{\ell=1}^{n}\left[\epsilon_\ell\left\{v_j(\mathbf{X}_l,X_{ik})-\beta_0^{j-1}v_1(\mathbf{X}_l,X_{ik})\right\}+\left\{t_j(\mathbf{X}_l)-\beta_0^{j-1}t_1(\mathbf{X}_l)\right\}\right].
\end{aligned}$$

Interchanging the indices i and ℓ and using the fact that $K(\cdot)$ is symmetric, the term R_1 equals

$$\begin{aligned}
n^{-1/2}\sum_{i=1}^{n}\epsilon_i\sum_{j=1}^{J}\sum_{k=1}^{J} &\quad (j-1)\beta_0^{j-2}n^{-1}\sum_{\ell=1}^{n}w_j(X_{\ell k})m_1(X_{\ell k}) \\
&\quad \times\left\{K_h(X_{\ell k}-X_{ij})u_j(\mathbf{X}_i,X_{\ell k})-\beta_0^{j-1}K_h(X_{\ell k}-X_{i1})u_1(\mathbf{X}_i,X_{\ell k})\right\}.
\end{aligned}$$

Using standard kernel theory, assumption A (iii) and $h^3n^{1/2}\to 0$ one shows that the last summation has the limit

$$w_j(X_{ij})m_1(X_{ij})f_k(X_{ij})u_j(\mathbf{X}_i,X_{ij})-\beta_0^{j-1}w_j(X_{i1})m_1(X_{i1})f_k(X_{i1})u_1(\mathbf{X}_i,X_{i1}).$$

Similarly, $R_2 = R_{21} + R_{22}$, where

$$\begin{aligned} R_{21} &= n^{-1/2} \sum_{i=1}^{n} \epsilon_i n^{-1} \sum_{j=1}^{J} \sum_{k=1}^{J} \sum_{\ell=1}^{n} (j-1)\beta_0^{j-2} w_j(X_{\ell k}) m_1(X_{\ell k}) \left\{ v_j(\mathbf{X}_i, X_{\ell k}) - \beta_0^{j-1} v_1(\mathbf{X}_i, X_{\ell k}) \right\} \\ &\approx n^{-1/2} \sum_{i=1}^{n} \epsilon_i \sum_{j=1}^{J} (j-1)\beta_0^{j-2} \left\{ v_j^*(\mathbf{X}_i) - \beta_0^{j-1} v_1^*(\mathbf{X}_i) \right\} \end{aligned}$$

and

$$\begin{aligned} R_{22} &= n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} (j-1)\beta_0^{j-2} w_j(X_{ik}) m_1(X_{ik}) \sum_{\ell=1}^{n} \left\{ t_j(\mathbf{X}_l) - \beta_0^{j-1} t_1(\mathbf{X}_l) \right\} \\ &\approx n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{J} (j-1)\beta_0^{j-2} s_j^* \left\{ t_j(\mathbf{X}_i) - \beta_0^{j-1} t_1(\mathbf{X}_i) \right\}. \end{aligned}$$

It thus follows that

$$n^{1/2}(\widehat{\beta}_{LS} - \beta_0) = D_{LS}^{-1} n^{-1/2} \sum_{i=1}^{n} \{\epsilon_i \mathcal{H}_1(\mathbf{X}_i, \beta_0) + \mathcal{H}_2(\mathbf{X}_i, \beta_0)\} + o_P(1). \tag{22}$$

This verifies Theorem 1.

7.4 Sketch of Proof of Theorem 2

The proof is facilitated by noting that

$$\begin{aligned} & m_j(x) - \beta^{j-1} G\{m_1(x), \ldots, m_J(x), \beta\} \\ &= \{\sum_{j=1}^{J} w_j(x)\beta^{2j-2}\}^{-1} \sum_{k=1}^{J} \{w_k(x)\beta^{2k-2} m_j(x) - w_k(x)\beta^{k+j-2} m_k(x)\} \\ &= \{\sum_{j=1}^{J} w_j(x)\beta^{2j-2}\}^{-1} \sum_{k=1}^{J} w_k(x) \left\{ \beta^{2k-2} m_j(x) - \beta^{k+j-2} m_k(x) \right\}. \end{aligned}$$

Hence, $\widehat{\beta}_{EIV}$ is formed by minimizing

$$n^{-1/2} \sum_{i=1}^{n} \sum_{j,k=1}^{J} w_j(X_{ik}) s(\beta, X_{ik}) \left[\sum_{\ell=1}^{J} w_\ell(X_{ik}) \left\{ \beta^{2\ell-2} \widehat{m}_j(X_{ik}) - \beta^{\ell-j-2} \widehat{m}_\ell(X_{ik}) \right\} \right]^2.$$

Define $s_\beta(\beta, x) = (\partial/\partial\beta) s(\beta, x)$ and $s_{\beta\beta}(\beta, x) = (\partial^2/\partial\beta^2) s(\beta, x)$. Further define

$$\begin{aligned} Q_j(x, \beta, m_1, \ldots, m_J) &= \sum_{\ell=1}^{J} w_\ell(x) \left\{ \beta^{2\ell-2} m_j(x) - \beta^{\ell+j-2} m_\ell(x) \right\}; \\ R_j(x, \beta, m_1, \ldots, m_J) &= \sum_{\ell=1}^{J} w_\ell(x) \left\{ (2\ell-2)\beta^{2\ell-3} m_j(x) - (\ell+j-2)\beta^{\ell+j-3} m_\ell(x) \right\}. \end{aligned}$$

Then $\widehat{\beta}_{EIV}$ minimizes $n^{-1/2}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta,X_{ik})Q_j^2\,(X_{ik},\beta,\widehat{m}_1,\ldots,\widehat{m}_j)$, and hence necessarily solves

$$\begin{aligned} 0 &= n^{-1/2}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s_\beta(\beta,X_{ik})Q_j^2(X_{ik},\beta,\widehat{m}_1,\ldots,\widehat{m}_J) \\ &+ 2n^{-1/2}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta,X_{ik})Q_j(X_{ik},\beta,\widehat{m}_1,\ldots,\widehat{m}_J)R_j(X_{ik},\beta,\widehat{m}_1,\ldots,\widehat{m}_J). \end{aligned}$$

This is an estimating equation, and with an admitted lack of rigor we proceed to analyze it in a standard fashion. Indeed, $n^{-1/2}$ times its first derivative evaluated at $(\beta_0,m_1,\ldots,m_J)$ is

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s_{\beta\beta}(\beta_0,X_{ik})Q_j^2(X_{ik},\beta_0,m_1,\ldots,m_J) \\ + \;& 4n^{-1}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s_\beta(\beta_0 X_{ik})Q_j(X_{ik},\beta_0,m_1,\ldots,m_J)R_j(X_{ik},\beta_0,m_1,\ldots,m_J) \\ + \;& 2n^{-1}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta_0,X_{ik})R_j^2(X_{ik},\beta_0,m_1,\ldots,m_J) \\ + \;& 2n^{-1}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta_0,X_{ik})Q_j(X_{ik},\beta_0,m_1,\ldots,m_J)(\partial/\partial\beta)R_j(X_{ik},\beta_0,m,\ldots,m_J). \end{aligned}$$

However, note that $Q_j(x,\beta_0,m_1,\ldots,m_J)=0$, so that the first derivative of the estimating equation when normalized and evaluated at $(\beta_0,m_1,\ldots,m_J)$ is

$$\begin{aligned} & 2n^{-1}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta_0,X_{ik})R_j^2(X_{ik},\beta_0,m_1,\ldots,m_J) \\ & \xrightarrow{p} 2\sum_{j,k=1}^{J} Ew_j(X_k)s(\beta_0,X_k)m_1^2(X_k)\{\sum_{\ell}^{J} w_\ell(X_k)\beta_0^{2\ell+j-4}(\ell-j)\}^2 = 2D_{EIV}. \end{aligned}$$

It is immediately obvious that since $Q_j^2(x,\beta_0,\widehat{m}_1,\ldots,\widehat{m}_J)=o_P(n^{-1/2})$, then $n^{1/2}(\widehat{\beta}_{EIV}-\beta_0)$ is asymptotically equivalent to

$$D_{EIV}^{-1}n^{-1/2}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta_0,X_{ik})Q_j(X_{ik},\beta_0,\widehat{m}_1,\ldots,\widehat{m}_J)R_j(X_{ik},\beta_0,\widehat{m}_1,\ldots,\widehat{m}_J).$$

Since $Q_j(x,\beta_0,m_1,\ldots,m_J)=0$, $n^{1/2}(\widehat{\beta}_{EIV}-\beta_0)$ is asymptotically equivalent to

$$\begin{aligned} & D_{EIV}^{-1}n^{-1/2}\sum_{i=1}^{n}\sum_{j,k=1}^{J} w_j(X_{ik})s(\beta_0,X_{ik})R_j(X_{ik},\beta_0,m_1,\ldots,m_J) \\ & \quad\times\sum_{\ell=1}^{J} w_\ell(X_{ik})\Big[\beta_0^{2\ell-2}\left\{\widehat{m}_j(X_{ik})-m_j(X_{ik})\right\}-\beta_0^{\ell-j-2}\left\{\widehat{m}_\ell(X_{ik})-m_\ell(X_{ik})\right\}\Big]. \quad (23) \end{aligned}$$

The proof is completed by using expansion (9) and the fact that $m_j''(x) = \beta_0^{j-1} m_1''(x)$, in a manner similar to that of Theorem 1. After some algebra one arrives at

$$n^{1/2}(\widehat{\beta}_{EIV} - \beta_0) = D_{LS}^{-1} n^{-1/2} \sum_{i=1}^{n} \{\epsilon_i \{\mathcal{H}_3(\mathbf{X}_i, \beta_0) + \mathcal{M}(\mathbf{X}_i, \beta_0)\} + \mathcal{G}(\mathbf{X}_i, \beta_0)\} + o_P(1). \tag{24}$$

This verifies Theorem 2.

7.5 Proof of Theorem 3

Note that $\widehat{m}^{LL}(x) = e_1^T \{n^{-1} \sum_{i=1}^n A_i(x)\xi_i(x)\xi_i(x)^T\}^{-1} n^{-1} \sum_{i=1}^n A_i(x) Y_i \xi_i(x)$, where $A_i(x) = K_{h_1}(X_{i-1} - x_1) K_{h_2}(X_{i-2} - x_2)$ and where $e_1^T = (1, 0, 0)$. Note that $E(X_i | X_{i-1}, X_{i-2}) = \alpha + m_1(X_{i-1}) + \beta m_1(X_{i-2})$. Put $\epsilon_i = X_i - \alpha - m_1(X_{i-1}) - \beta m_1(X_{i-2})$, $m(x) = \alpha + m_1(x_1) + \beta m_1(x_2)$ and $\widehat{r}_i(x) = m(X_{i-1}, X_{i-2}) - m(x) - (X_{i-1} - x_1) m_1'(x_1) - (X_{i-2} - x_2)\beta m_1'(x_2)$. We can write $\widehat{m}^{LL}(x) - m(x) = e_1^T \{n^{-1} \sum_{i=1}^n A_i(x)\xi_i(x)\xi_i(x)^T\}^{-1} n^{-1} \sum_{i=1}^n A_i(x)\{\epsilon_i + \widehat{r}_i(x)\}\xi_i(x)$.

Now note that

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} A_i(x)\xi_i(x)\xi_i(x)^T &= p(x) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + h_1 p^{(1,0)}(x) \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &\quad + h_2 p^{(0,1)}(x) \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + O_P(h_1^2 + h_2^2 + \frac{\log n}{\sqrt{n h_1 h_2}}) \\ &= p(x) I + h_1 p^{(1,0)}(x) I_1 + h_2 p^{(0,1)}(x) I_2 \\ &\quad + O_P(h_1^2 + h_2^2 + \frac{\log n}{\sqrt{n h_1 h_2}}). \end{aligned} \tag{25}$$

Here $p^{(0,1)}$ and $p^{(1,0)}$ denote the partial derivative of p with respect to x_1 or x_2, respectively. The expansion in (25) holds uniformly for $x \in B \times B$. For a proof of (25) one proceeds as in Bosq (1996) where uniform rates are shown for kernel density estimates of strongly mixing observations by using exponential inequalities for mixing sequences, see Theorem 2.2 in Bosq (1996).

Equation (25) shows

$$\begin{aligned} [n^{-1} \sum_{i=1}^{n} A_i(x)\xi_i(x)\xi_i(x)^T]^{-1} &= p(x)^{-1} \Big\{ I - h_1 p^{(1,0)}(x) p(x)^{-1} I_1 - h_2 p^{(0,1)}(x) p(x)^{-1} I_2 \\ &\quad + O_P(h_1^2 + h_2^2 + \frac{\log n}{\sqrt{n h_1 h_2}}) \Big\}. \end{aligned} \tag{26}$$

Similarly as (25) one shows uniformly for $x \in B \times B$

$$n^{-1} \sum_{i=1}^{n} A_i(x)\epsilon_i \xi_i(x) = O_P(\frac{\log n}{\sqrt{n h_1 h_2}}) \tag{27}$$

and

$$n^{-1}\sum_{i=1}^{n} A_i(x)\widehat{r}_i(x)\xi_i(x) = O_P(h_1^2 + h_2^2). \tag{28}$$

Equations (26) - (28) give uniformly for $x \in B \times B$:

$$\widehat{m}^{LL}(x) - m(x) = e_1^T p(x)^{-1}\Big\{I - h_1 p^{(1,0)}(x)p(x)^{-1}I_1 - h_2 p^{(0,1)}(x)p(x)^{-1}I_2$$
$$+O_P(h_1^2 + h_2^2 + \frac{\log n}{\sqrt{nh_1h_2}})\Big\} n^{-1}\sum_{i=1}^{n} A_i(x)[\epsilon_i + \widehat{r}_i(x)]\xi_i(x) + o_P(n^{-1/2}). \tag{29}$$

We use now that uniformly in $x \in B \times B$:

$$n^{-1}\sum_{i=1}^{n} A_i(x)\widehat{r}_i(x)\xi_i(x) = (1/2)h_1^2 m_1''(x_1)e_1 + (1/2)h_2^2\beta m_1''(x_2)e_1 + o_P(n^{-1/2}). \tag{30}$$

For a proof of (30) one uses the fact that m_1 is four times continuously differentiable and for the treatment of the resulting sum of strongly mixing summands one proceeds as in the proof of (25).

We treat now the estimate $\widetilde{m}_1^I(x_1) = n^{-1}\sum_{i=1}^{n} w(X_{i-2})\widehat{m}^{LL}(x, X_{i-2})/[n^{-1}\sum_{i=1}^{n} w(X_{i-2})]$.

Note now that for all q with $\gamma_1 > 0$ arbitrarily small and $\gamma_2 > 0$ arbitrarily large:

$$\sup_{x_1 \in B} E\left[h_1 n^{-1}\sum_{j=1}^{n} w(X_{j-2})\frac{p^{(1,0)}(x_1, X_{j-2})}{p(x_1, X_{j-2})^2} n^{-1}\sum_{i=1}^{n} A_i(x_1, X_{j-2})\epsilon_i^*\right]^{2q}$$
$$= O(h_1^{2q}(nh_1)^{-q}n^{\gamma_1 q}), \tag{31}$$

where $\epsilon_i^* = \epsilon_i \mathbf{1}[|\epsilon_i| \le \gamma_2 \log n] - E\{\epsilon_i \mathbf{1}[|\epsilon_i| \le \gamma_2 \log n]\}$. Claim (31) can be shown by application of Davydov's inequality, see Corollary 1.1 in Bosq (1996). For this purpose one writes the left hand side of (31) as $\sum_{j_1,...,j_{2q}}\sum_{i_1,..,i_{2q}} h_1^{2q} n^{-4q} E Z_{i_1,j_1}\cdot....\cdot Z_{i_{2q},j_{2q}}$, where $Z_{i,j} = w(X_{j-2})p^{(1,0)}(x_1, X_{j-2})$ $A_i(x_1, X_{j-2})$ $\epsilon_i^*/p(x_1, X_{j-2})^2$. Davydov's inequality and our mixing conditions imply that for arbitrarily large C there exist constants C' and C'' such that

$$\left|E Z_{i_1,j_1}\cdot....\cdot Z_{i_{2q},j_{2q}}\right| \le C'' n^{-C}, \tag{32}$$

for all indices $i_1, .., j_{2q}$ such that there exists an $1 \le l \le 2q$ with $|i_l - i_k| \ge C' \log n$ for all $k \ne l$ and $|i_l - j_k| \ge C' \log n$ for all k. For the proof of (32) one makes use of B (ii) -(iv) and of the fact that $|\epsilon_i^*| \le 2\gamma_2 \log n$. Claim (31) follows by a bound on the remaining terms.

With the help of (31) and using the fact that, for c large enough,

$$\left|\frac{\partial}{\partial x_1} h_1 n^{-1}\sum_{j=1}^{n} w(X_{j-2})\frac{p^{(1,0)}(x_1, X_{j-2})}{p(x_1, X_{j-2})^2} n^{-1}\sum_{i=1}^{n} A_i(x_1, X_{j-2})\epsilon_i^*\right| \le n^c \tag{33}$$

we get that uniformly for $x_1 \in B$

$$h_1 n^{-1} \sum_{j=1}^{n} w(X_{j-2}) \frac{p^{(1,0)}(x_1, X_{j-2})}{p(x_1, X_{j-2})^2} n^{-1} \sum_{i=1}^{n} A_i(x_1, X_{j-2}) \epsilon_i^* = o_P(n^{-1/2}). \tag{34}$$

For a proof that (34) holds uniformly for $x_1 \in B$ one shows first that it holds uniformly for $x_1 \in B_n$ where B_n is a subset of B with $n^{c'}$ equidistant points and where c' is an arbitrary positive constant. Then the claim follows from (33) if c' has been chosen large enough.

Note now that for all C''' there exists γ_2 such that $\max_{1 \le i \le n} |\epsilon_i| \le \gamma_2 \log n$ [with probability tending to 1] and $|E\{\epsilon_i \mathbf{1}[|\epsilon_i| \le \gamma_2 \log n]\}| \le n^{-C'''}$. Remember that ϵ_i has a finite Laplace transform, see B (vi). Therefore (34) with γ_2 large enough implies

$$h_1 n^{-1} \sum_{j=1}^{n} w(X_{j-2}) \frac{p^{(1,0)}(x_1, X_{j-2})}{p(x_1, X_{j-2})^2} n^{-1} \sum_{i=1}^{n} A_i(x_1, X_{j-2}) \epsilon_i = o_P(n^{-1/2}). \tag{35}$$

Again, this expansion holds uniformly for $x_1 \in B$.

With (35) and with similar expansions for other terms we arrive at:

$$\begin{aligned} \widetilde{m}_1^I(x_1) &= \alpha + m_1(x_1) + (1/2) h_1^2 m_1''(x_1) + (1/2) h_2^2 \beta \int m_1'' w f \\ &\quad + \frac{1}{n^2} \sum_{i,j=1}^{n} w(X_{j-2}) p(x_1, X_{j-2})^{-1} A_i(x_1, X_{j-2}) \epsilon_i + o_P(n^{-1/2}). \end{aligned}$$

This expansion holds uniformly for $x_1 \in B$. Using again Davydov's inequality one shows that

$$\begin{aligned} &E[n^{-2} \sum_{i,j=1}^{n} \{w(X_{j-2}) \frac{K_{h_2}(X_{i-2} - X_{j-2})}{p(x_1, X_{j-2})^2} - w(X_{i-2}) \frac{f(X_{i-2})}{p(x_1, X_{i-2})^2}\} K_{h_1}(X_{i-1} - x_1) \epsilon_i^*]^{2q} \\ &= O(h_2^{4q} (n h_1)^{-q} n^{\gamma q}). \end{aligned}$$

As above this shows that uniformly for $x_1 \in B$

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \left[w(X_{j-2}) \frac{K_{h_2}(X_{i-2} - X_{j-2})}{p(x_1, X_{j-2})^2} - w(X_{i-2}) \frac{f(X_{i-2})}{p(x_1, X_{i-2})^2} \right] K_{h_1}(X_{i-1} - x_1) \epsilon_i = o_P(n^{-1/2}).$$

This gives uniformly for $x_1 \in B$

$$\begin{aligned} \widetilde{m}_1^I(x_1) &= \alpha + m_1(x_1) + (1/2) h_1^2 m_1''(x_1) + (1/2) h_2^2 \beta \int m_1'' w f \\ &\quad + n^{-1} \sum_{i=1}^{n} w(X_{i-2}) p(x_1, X_{i-2})^{-1} f(X_{i-2}) K_{h_1}(X_{i-1} - x_1) \epsilon_i + o_P(n^{-1/2}). \end{aligned}$$

The expansion for $\widehat{m}_1^I$ stated in the Theorem follows by some straight forward calculations. For the treatment of $\widehat{m}_2^I$ one proceeds similarly.

7.6 Sketch of Proof of Theorem 4

Along the lines of the proof of Theorem 1 one verifies that the stochastic expansion (22) remains valid for autoregression. For doing this one proceeds as in the proof of Theorem 3. Again, the main tools are exponential inequalities and Davydov's inequality for mixing sequences. Asymptotic normality of $n^{1/2}(\widehat{\beta}_{LS} - \beta_0)$ follows by application of a central limit theorem for strongly mixing sequences, see e.g. Theorem 1.7 in Bosq(1996). We have to verify that for some $\gamma > 2$ $E\,|\epsilon_i \mathcal{H}_1(\mathbf{X}_i, \beta_0) + \mathcal{H}_2(\mathbf{X}_i, \beta_0)|^\gamma < \infty$. This follows easily from B (vi) and from the assumption that the functions $w_j, m_1, v_j, f_j, u_j, t_j$ are bounded for $j = 1, \ldots, J$.

7.7 Sketch of Proof of Theorem 5

One shows first that the stochastic expansion (24) remains valid for autoregression and then one proceeds as in the last proof.

7.8 Sketch of Proof of Theorem 6

Proceeding as in the proof of central limit theorems for kernel estimates of strongly mixing sequences [see e.g. Theorems 2.3 and 3.4 in Bosq (1996)] and using the stochastic expansion (21), one shows that the vector $n^{2/5}(\widehat{m}_1(x) - m_1(x), \ldots, \widehat{m}_J(x) - \beta^{J-1} m_1(x))^T$ has a normal limit with mean $(1/2)n^{2/5}h^2 r(x)(1, \ldots, \beta^{J-1})^T$ and covariance matrix equal to a diagonal matrix with diagonal elements $n^{1/5}h^{-1}\nu(K)f(x)\ s_j^2(x)$. Note that $\widehat{m}_j(x)$ and $\widehat{m}_k(x)$ are asymptotically independent for $j \neq k$. The first statement of Theorem 6 follows from the fact that $\widehat{m}_1^*(x) - \sum_{j=1}^J c_j \beta^{-(j-1)} \widehat{m}_j(x) / \sum_{j=1}^J c_j = o_P(n^{-2/5})$. For the second statement of Theorem 6 note that $\sum_{j=1}^J c_j^2 \beta^{-2(j-1)} s_j^2(x) / \sum_{j=1}^J c_j^2$ is minimized by $c_j = \beta^{2(j-1)} s_j^{-2}(x)$.