Econometric Theory, **25**, 2009, 442–481. Printed in the United States of America.
doi:10.1017/S0266466608090142

NONPARAMETRIC ADDITIVE MODELS FOR PANELS OF TIME SERIES

ENNO MAMMEN
University of Mannheim
BÅRD STØVE
Norwegian School of Economics and Business Administration
DAG TJØSTHEIM
University of Bergen

This paper discusses nonparametric models for panels of time series. There is already a substantial literature on nonlinear models and nonparametric methods in a regression and time series setting. But almost without exception these developments have been limited to univariate and multivariate models of moderate dimensions. Very little has been done for panels, where the dimension, often corresponding to a number of individuals, typically is very large but where the number of observations for each individual may be small or moderate. It is the aim of this paper to start a systematic theoretical treatment of nonparametric models for panels of time series, in particular on additive models. Extending existing methodology to the panel situation is by no means trivial because already for the parametric case many problems are unsolved. Our estimation approach is based on backfitting methods.

1. INTRODUCTION

Over the last decades nonparametric smoothing has been one of the most active areas of statistical research. In nonparametric smoothing the aim is to avoid assumptions on the parametric form of the function of interest, often a density or a regression function. By using this approach, it may be possible to reveal structures in the data not captured by a predetermined parametric model.

There is already a substantial literature on nonlinear models and nonparametric methods in a regression and time series setting. But almost without exception these developments have been limited to univariate or multivariate models of moderate dimensions. Very little has been done for panels, where the dimension,

The comments from three anonymous referees and co-editor Oliver Linton have resulted in an improved presentation. This research was partly supported by grant 147231/432 of the Norwegian Research Council and grant MA1026/7-3 of the German Science Foundation. Address correspondence to Enno Mammen, Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany; e-mail: emammen@rumms.uni-mannheim.de.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

often corresponding to a number of individuals, typically is very large but where the number of observations for each individual may be small or moderate. One would expect that the practical need for nonlinear models would be no less in a panel situation, but a systematic theoretical foundation has been lacking. It is the aim of this paper to start establishing such a theory. The emphasis will be on a nonparametric approach, in particular on additive models.

Nonparametric additive models in a regression setting are widely used in both theoretical economics and empirical data analysis. In particular, these models are extremely useful for an economic analysis of production or matching functions; see Sperlich, Tjøstheim, and Yang (2002). However, extending existing methodology to the panel situation is by no means trivial (actually also in the linear case of interrelated time series in a panel there are many unsolved problems; see, e.g., Hjellvik and Tjøstheim, 1999; Fu, Li, and Fung, 2002; Wooldridge, 2005a; Chang, 2004).

To be more specific, let $\{Y^{it}\}, i = 1, \ldots, n;\ t = 1, \ldots, T$, be stochastic variables representing observations on a set of n individuals over T time periods. A rather general additive model for the panel $\{Y^{it}\}$ will be given by

$$Y^{it} = \sum_{j=1}^{p} m_j(X_j^{it}) + \eta_t + \lambda_i + \varepsilon^{it}. \tag{1.1}$$

Here $\{X_j^{it}\}$, $j = 1, \ldots, p$, is a set of explanatory variables (time lagged values of Y^{it} may be included among them), $\{\eta_t\}$ and $\{\lambda_i\}$ are temporary and individual effects, respectively, not captured by the explanatory variables, and $\{\varepsilon^{it}\}$ are error terms. The functions $\{m_j,\ j = 1, \ldots, p\}$ are unknown, and it is our task to estimate them. In the linear case $m_j(x) = \alpha_j x$ for some unknown parameters α_j, and our model reduces to the more familiar time series panel regression model treated in many papers and in the books by Hsiao (1986), Mátyás and Sevestre (1992), Baltagi (1995), and Arellano (2003).

The model (1.1) is very natural, and our contribution does not consist in proposing this model but rather in starting to analyze such a model in a systematic fashion. One of the difficulties arising in trying to establish an estimation theory for (1.1) is that the asymptotics can be imagined in several ways. We may let $n \to \infty$, $T \to \infty$, or both. Another problem is the presence of the individual effects λ_i and/or temporal effects η_t.

There are several ways of approaching the problem of estimating m_j. We have chosen to use backfitting (Hastie and Tibshirani, 1990) to take advantage of the recent developments in this area in Opsomer and Ruppert (1997) and Mammen, Linton, and Nielsen (1999). In particular, we will rely heavily on the smoothed backfitting approach introduced in Mammen et al. (1999). For a discussion of practical implementations of smooth backfitting, see also Mammen and Park (2005, 2006) and Nielsen and Sperlich (2005). An alternative would have been to use marginal integration (Newey, 1994; Tjøstheim and Auestad 1994; Linton and Nielsen, 1995). The latter method only works for low-dimensional covariates

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

and for well-behaved designs. But in case of deviations from the additive model (1.1) it is fairly robust and can be extended to include interaction terms (Sperlich et al., 2002).

Recent contributions to nonlinear and non/semiparametric modeling of panels are given in Hjellvik, Chen, and Tjøstheim (2004), Fan and Li (2004), Wooldridge (2005b), Cai and Li (2008), and Henderson, Carroll, and Li (2008). An interesting application with a nonparametric additive panel model is given in Profit and Sperlich (2004).

A somewhat related class of problems is emerging from describing a family of curves. This is the field of functional data analysis, which requires a rather different theoretical approach, as is apparent from, e.g., Ramsay and Silverman (1996) and Ferraty and Vieu (2006).

It will be seen that we do not at all manage to obtain a complete theory, and open problems will be pointed out as we proceed. An overview of the paper is as follows. In Section 2 we discuss Nadaraya–Watson smoothing in model (1.1). This will be done for models with and without individual effects ($\lambda_i \neq 0$ or $\lambda_i = 0$, respectively) and for two asymptotic settings. In the first asymptotic approach we fix T and let $n \to \infty$. In the second setting we let $T \to \infty$ and $n \to \infty$. Nadaraya–Watson smoothing in this model can be easily introduced and leads to an intuitive clear asymptotic theory. The more complicated theory for local linear smoothing will be developed in Section 3. Simulations and a real data example are presented in Section 4. The paper concludes in Section 5. In Section 6 we state a general theorem for the solution of empirical integral equations. This theorem is used when we discuss Nadaraya–Watson smoothing in an asymptotic framework with $n \to \infty$ and $T \to \infty$. Proofs are deferred to Section 7.

2. LOCAL CONSTANT SMOOTHING OF PANELS OF TIME SERIES

2.1. The Model and Its Basic Notation

We will start by looking at model (1.1) when there are no individual effects $\{\lambda_i\}$ present:

$$Y^{it} = \sum_{j=1}^{p} m_j(X_j^{it}) + \eta_t + \varepsilon^{it} \qquad (i = 1, \ldots, n; \qquad t = 1, \ldots, T). \tag{2.1}$$

The model with included individual effects will be postponed to Section 2.6. In model (2.1) one observes Y^{it} and covariables X_j^{it} with $j = 1, \ldots, p$. In particular, we think of the case where partially the covariables X_j^{it} are lagged observations $Y^{i,t-j}$ (for $j \leq p'$, say) and where the other covariables X_j^{it} ($p' < j \leq p$) are external covariables. We will make the assumption that the individual data (outcomes of the time series and covariables) are independent and identically distributed (i.i.d.) For the error variables ε^{it} we will assume that, given the covariables up to

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

time t and the responses up to time $t-1$, they have conditional mean zero. For more details, see Assumption (A1) in Section 2.4. The variables $\eta_1, \ldots, \eta_T$ are constants that are unknown. They model the influence of some additional external variables onto the development of the time series Y^{it}. Typically they could be considered nuisance parameters.

We assume that X^{it} has a density on $[a,b] = [a_1, b_1] \times \cdot \times [a_p, b_p]$. The conditional density of X^{it} given that X^{it} lies in $[a,b]$ is denoted by f^t. For simplicity of notation we put $a_1 = \cdots = a_p = 0$ and $b_1 = \cdots = b_p = 1$. Thus $[a,b] = [0,1]^p$. We will consider estimation for the functions $m_1, \ldots, m_p$ on $[0,1]$.

We denote by N the number of covariables X^{it} in $[0,1]^p$ for $1 \le i \le n$, $1 \le t \le T$. The number N_t is the number of X^{it} in $[0,1]^p$ for fixed t. By the law of large numbers we have that in probability $N_t/n \to d_t$ where $d_t = P(X^{it} \in [0,1]^p)$. We put $d = \sum_{t=1}^{T} d_t$. In Section 2.5 where T converges to infinity the sum d depends on T. Then we also write $d = d(T)$ to indicate this dependence in the notation. Then the typical case is that $d = d(T)$ grows to infinity with order $O(T)$. The one- and two-dimensional marginals of f^t are denoted by $f_j^t(x_j)$ or $f_{j,k}^t(x_j, x_k)$, respectively. Furthermore, we put $f(x) = \sum_{t=1}^{T} (d_t/d) f^t(x)$, $f_j(x_j) = \sum_{t=1}^{T} (d_t/d) f_j^t(x_j)$, and $f_{j,k}(x_j, x_k) = \sum_{t=1}^{T} (d_t/d) f_{j,k}^t(x_j, x_k)$.

We start our discussion in Section 2.2 with model (2.1) that only contains temporary effects $\{\eta_t\}$ but no individual effects $\{\lambda_i\}$. The presence of the temporary effects $\{\eta_t\}$ means that the time series of the panels will be interrelated (but not necessarily intercorrelated), even conditionally on the explanatory variables. A linear version similar to this case was considered in Hjellvik and Tjøstheim (1999) and Fu et al. (2002). Section 2.2 introduces a local constant backfitting estimator for model (2.1). The numerical implementation of this algorithm will be discussed in Section 2.3. As in Hjellvik and Tjøstheim (1999) and Fu et al. (2002) we first put the emphasis on asymptotics for the case $n \to \infty$ with T fixed. For this asymptotic framework convergence of the backfitting algorithm and the asymptotic distribution of our backfitting estimators are stated in Section 2.4. Asymptotics with $n \to \infty$ and $T \to \infty$ will be treated in Section 2.5. Section 2.6 contains a short discussion of models with included individual effects.

2.2. Definition of Local Constant Smoothers

We now define local constant smooth backfitting estimators $\widehat{m}_1, \ldots, \widehat{m}_p$ of the functions $m_1, \ldots, m_p$. The estimators are based on kernel smoothing. For the performance of our estimators we need that the kernel function integrates to one over the interval $[0,1]$. This is achieved by the following modification of a convolution kernel K:

$$K_h(u,v) = \begin{cases} \dfrac{K[h^{-1}(u-v)]}{\int_0^1 K[h^{-1}(w-v)]\, dw} & \text{if } u, v \in [0,1], \\ 0 & \text{else} \end{cases}.$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

Our estimators $\widehat{m}_1, \ldots, \widehat{m}_p, \widehat{\eta}_1, \ldots, \widehat{\eta}_T$ are defined as minimizer of a smoothed least squares criterion. They minimize the following criterion function:

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \int \left[Y^{it} - \sum_{j=1}^{p} \widehat{m}_j(u_j) - \widehat{\eta}_t \right]^2 \tag{2.2}$$

$$\cdot K_{h(1)}(u_1, X_1^{it}) \times \cdots \times K_{h(p)}(u_p, X_p^{it})\, du_1 \ldots du_p$$

under the constraints

$$\int \widehat{m}_j(x_j) \widehat{f}_j(x_j)\, dx_j = 0 \tag{2.3}$$

for $j = 1, \ldots, p$. Here $h(1), \ldots\ h(p)$ is a tuple of bandwidths. The constraint is motivated by the norming assumption (A4) in Section 2.4. The functions $\widehat{f}_j$ are the following kernel density estimators based on the covariables:

$$\widehat{f}_j(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it}).$$

The criterion function (2.2) is a smoothed version of a sum of squared residuals. Smoothing is done with respect to the covariables $X_1^{11}, \ldots, X_p^{nT}$ but not with respect to t. Smoothing with respect to t could be easily included. We will briefly comment on this in Section 5.

Our estimators only use response variables in the smoothing if the corresponding covariables lie in $[0,1]^p$. All other responses are thrown away and are not used in the construction of the estimator. In particular if p is large a relatively large portion of the observations may be lost. Clearly, one can use an arbitrarily large but fixed compact set to weaken this effect. Asymptotically, this could be more carefully analyzed by using sets that, depending on the sample size, converge to the whole space. Such an analysis would be technically very involved because one has to take care of the tail behavior of the covariables and of the regression functions. Another way out could be based on semiparametric modifications of the estimator. This could be done by fitting a constant value for $\widehat{m}_j(x_j)$ with x_j outside $[0,1]$ (or more generally to fit a parametric shape outside $[0,1]$). Then no observations are lost in the calculation of the estimator. We do not pursue the discussion of this estimator here.

In Section 2.3 we introduce a backfitting algorithm for the calculation of our estimator.

2.3. Smooth Backfitting Algorithm

Definition (2.2) is very similar to the definition of a backfitting estimator in an additive model in Mammen et al. (1999). As there, one gets by simple arguments using derivatives of the criterion function (2.2) that the estimators fulfill for $0 \leq x_j \leq 1$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$$\widehat{m}_j(x_j) = \widetilde{m}_j(x_j) - \sum_{t=1}^{T} \frac{N_t}{N} \widehat{\eta}_t \frac{\widehat{f}_j^t(x_j)}{\widehat{f}_j(x_j)} \tag{2.4}$$

$$- \sum_{l \neq j} \int \widehat{m}_l(x_l) \frac{\widehat{f}_{jl}(x_j, x_l)}{\widehat{f}_j(x_j)} \, dx_l \qquad (j = 1, \ldots, p),$$

$$\widehat{\eta}_t = \widetilde{\eta}_t - \sum_{j=1}^{p} \int \widehat{m}_j(x_j) \widehat{f}_j^t(x_j) \, dx_j \qquad (t = 1, \ldots, T), \tag{2.5}$$

where $\widetilde{m}_j, \widetilde{\eta}_t$ are the following "marginal" estimators:

$$\widetilde{m}_j(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) Y^{it} / \widehat{f}_j(x_j), \tag{2.6}$$

$$\widetilde{\eta}_t = N_t^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) Y^{it}.$$

The functions $\widehat{f}_{jk}, \widehat{f}_j^t$ are kernel density estimators based on the covariables:

$$\widehat{f}_{jk}(x_j, x_k) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it}) \, K_{h(k)}(x_k, X_k^{it}),$$

$$\widehat{f}_j^t(x_j) = N_t^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it}).$$

We now briefly comment on the numerical calculation of our estimator. Equations (2.4) and (2.5) suggest an iterative calculation of the estimators. Application of (2.4) for $j = 1, \ldots, p$ can be used for an update of $\widehat{m}_j$. In each application one plugs the current values of $\widehat{m}_l$ $(l \neq j)$ and of $\widehat{\eta}_t$ $(t = 1, \ldots, T)$ into the right-hand side of (2.4). Afterward one applies (2.5) for updates of $\widehat{\eta}_t$ for $t = 1, \ldots, T$. Again this is done by using the actual values of $\widehat{m}_j$ $(j = 1, \ldots, p)$ on the right-hand side of the equation. Let us call these iterative values $\widehat{m}_j^{[a]}$ and $\widehat{\eta}_t^{[a]}$ where a is the number of cycles of the algorithm that has been applied. The starting values are denoted by $\widehat{m}_j^{[0]}$ and $\widehat{\eta}_t^{[0]}$.

The marginal estimators $\widetilde{m}_j$, $\widetilde{\eta}_t$, $\widehat{f}_j$, $\widehat{f}_{jk}$, and $\widehat{f}_j^t$ must not be updated in the iteration. These estimators are calculated only once. Note further that it is enough to calculate the additive components only on the same grid as used for calculating the integrals in equation (2.4).

For clarity we give the algorithm for calculating the local constant smoother in a stepwise procedure. Assume that we want to estimate $\widehat{m}_j$ at some given points $x_i^0 = (x_{i1}^0, \ldots, x_{ip}^0)^T \in \mathbb{R}^p, i = 1, \ldots, M$. In practice we usually want to estimate the additive components on a grid over a certain range. In what follows we suppress the index i of x_{ij}^0, $j = 1, \ldots, p$.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

Step 1. Set $a = 0$ and calculate the smoothing weights around the point x_j^0.

Step 2. With these weights, obtain all $\widetilde{m}_j(x_j^0)$, $\widehat{f}_{jl}(x_j^0, x_l^0)\widehat{f}_j^{-1}$, $l \neq j$, and $\widetilde{\eta}_t$, $\widehat{f}_j^t(x_j^0)$, $t = 1, \ldots, T$. Then, set $\widehat{m}_j^{[a]} = \widetilde{m}_j$ and $\widehat{\eta}_t^{[a]} = \widetilde{\eta}_t$.

Step 3. For $j = 1, \ldots, p$ calculate for all points x_j^0

$$\widehat{m}_j^{[a+1]}(x_j^0) = \widetilde{m}_j(x_j^0) - \sum_{t=1}^{T} \frac{N_t}{N} \widehat{\eta}_t^{[a]} \frac{\widehat{f}_j^t(x_j^0)}{\widehat{f}_j(x_j^0)}$$

$$- \sum_{l<j} \int \widehat{m}_l^{[a+1]}(x_l) \frac{\widehat{f}_{jl}(x_j^0, x_l)}{\widehat{f}_j(x_j^0)} dx_l$$

$$- \sum_{l>j} \int \widehat{m}_l^{[a]}(x_l) \frac{\widehat{f}_{jl}(x_j^0, x_l)}{\widehat{f}_j(x_j^0)} dx_l.$$

Step 4. For $t = 1, \ldots, T$ calculate

$$\widehat{\eta}_t^{[a+1]} = \widetilde{\eta}_t - \sum_{j=1}^{p} \int \widehat{m}_j^{[a+1]}(x_j) \widehat{f}_j^t(x_j) dx_j.$$

Step 5. If a predetermined convergence criterion is fulfilled, stop. Otherwise set a to $a + 1$ and go to step 3.

One example of a convergence criterion is given in Section 4.

2.4. Asymptotics for $n \to \infty$ and T Fixed

This section contains two theoretical results for the case that $n \to \infty$ and that T is fixed. Our first result states the convergence of the backfitting algorithm that was introduced in Section 2.3. The second result gives the asymptotic distribution of the estimators. For our results we need the following assumptions.

For the error variables ε^{it} we assume the conditions that follow.

(A1) $(\mathbf{Y}^1, \mathbf{X}^1, \varepsilon^1), \ldots, (\mathbf{Y}^n, \mathbf{X}^n, \varepsilon^n)$ are i.i.d., where $\mathbf{Y}^i = (Y^{it} : 1 \leq t \leq T)$, $\mathbf{X}^i = (X_j^{it} : 1 \leq j \leq p, 1 \leq t \leq T)$, and $\varepsilon^i = (\varepsilon^{it} : 1 \leq t \leq T)$. Given $(X_j^{is} : 1 \leq j \leq p,\ 1 \leq s \leq t; \varepsilon^{is} : 1 \leq s \leq t-1)$, the variable ε^{it} has conditional mean zero and conditional variance σ^2, and they have a conditional absolute moment of order $5/2$ that is uniformly (in i and t) bounded by a constant.

We could treat more complicated data generating processes. We will briefly comment on this in Section 5.

As just mentioned, in this section we discuss asymptotics of a backfitting estimator in an asymptotic framework where we make the following assumption.

(A2) T is fixed and $n \to \infty$.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

This covers a large class of applications where one observes a large number of individual time series for a moderate number of time points. We will comment on the case $T \to \infty$, $n \to \infty$ in Section 2.5.

We make the following smoothness assumptions on the density of the covariables and the regression functions $m_1, \ldots, m_p$.

(A3) We assume that X^{it} has a density on $[0,1]^p$. The conditional density of X^{it} given that X^{it} lies in $[0,1]^p$ is denoted by f^t. We assume that f^t has continuous partial derivatives of order one and that it is bounded from below (on $[0,1]^p$). The regression functions m_j have continuous second derivatives on $[0,1]$.

We restrict the smoothness conditions here on the interval $[0,1]^p$ because we will consider estimation of $m_1, \ldots, m_p$ only on the compact intervals $[0,1]$. Assumption (A3) contains only standard smoothness conditions.

For identifiability of m_j we use the following norming conditions.

(A4) It holds that

$$\int_0^1 m_j(x_j) f_j(x_j)\, dx_j = 0$$

for $j = 1, \ldots, p$.

In our smoothing algorithms we use bandwidths $h(1), \ldots, h(p)$ that are of order $n^{-1/5}$. This is motivated by our smoothness conditions. Furthermore, we assume the following conditions for the smoothing kernel K.

(A5) The bandwidths $h(1), \ldots, h(p)$ fulfill $(dn)^{1/5} h(j) \to c_j$ for some constants $c_1, \ldots, c_p$. The kernel K is nonnegative, Lipschitz continuous, and symmetric around zero and has compact support ($[-1,1]$, say).

Our first theorem states convergence of the backfitting algorithm that was introduced in Section 2.3.

THEOREM 1. *Suppose that (A1)–(A5) hold. Then there exist constants $c > 0$ and $0 < \gamma < 1$ such that with probability tending to one*

$$\int \left[\widehat{m}_j^{[a]}(x_j) - \widehat{m}_j(x_j)\right]^2 f_j(x_j)\, dx_j \le c\gamma^{2a} R \quad \text{for } j = 1, \ldots, p$$

and

$$\left|\widehat{\eta}_t^{[a]} - \widehat{\eta}_t\right|^2 \le c\gamma^a R \quad \text{for } t = 1, \ldots, T,$$

where

$$R = 1 + \sum_{j=1}^{p} \int \widehat{m}_j^{[0]}(x_j)^2 f_j(x_j)\, dx_j + \sum_{t=1}^{T} (\widehat{\eta}_t^{[0]})^2 .$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

The following theorem describes the asymptotic behavior of $\widehat{m}_1, \ldots, \widehat{m}_p$. A uniform approximation of the estimators is stated. This immediately implies asymptotic normality of the estimators (see Corollary 1 later in this section). Furthermore, it can be used to construct uniform confidence bands. We will not state such a result here. It could be done along the lines of the treatment of standard classical nonparametric regression models.

THEOREM 2. *Assume (A1)–(A5). Put*

$$\gamma_{n,j} = -c_j^2 \int \left[m_j'(x_j) f_j'(x_j) + \frac{1}{2} m_j''(x_j) f_j(x_j) \right] dx_j \int u^2 K(u)\, du$$

and define $\beta^t, \beta_1, \ldots, \beta_p$ *as the minimizers of*

$$\int_{[0,1]^p} \sum_{t=1}^{T} [\beta(x,t) - \beta^t - \beta_1(x_1) - \cdots - \beta_p(x_p)]^2 \, \frac{d_t}{d} f^t(x)\, dx,$$

where the minimization runs over constants $\beta^1, \ldots, \beta^T$ *and functions* $\beta_1, \ldots, \beta_p : [0,1] \to \mathbb{R}$ *with* $\int \beta_j(x_j) f_j(x_j)\, dx_j = -\gamma_{n,j}$ *and where* $\beta(x,t)$ *is defined as*

$$\beta(x,t) = \sum_{j=1}^{p} c_j^2 \left[\frac{\partial f^t}{\partial x_j}(x)\, m_j'(x_j) f^t(x)^{-1} + \frac{1}{2} m_j''(x_j) \right] \int u^2 K(u)\, du\,.$$

Then the following uniform expansion holds:

$$\sup_{x_j \in [h_j, 1-h_j]} \left| \widehat{m}_j(x_j) - \left[\widetilde{m}_j^A(x_j) + (dn)^{-2/5} \beta_j(x_j) \right] \right| = o_p(n^{-2/5}),$$

$$\widehat{\eta}_t = \eta_t + (dn)^{-2/5} \beta^t + o_p(n^{-2/5}).$$

Here

$$\widetilde{m}_j^A(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) \varepsilon^{it} / \widehat{f}_j(x_j)\,. \tag{2.7}$$

Note that $\widetilde{m}_j^A$ is defined as $\widetilde{m}_j$ but with Y^{it} replaced by ε^{it}. It is easy to check under our conditions that $n^{2/5} \widetilde{m}_j^A$ has an asymptotic zero mean normal limit. Furthermore, $\beta_j(x_j)$ is a deterministic function. This shows that $\widehat{m}_j$ has the following asymptotic limit.

COROLLARY 1. *Under conditions (A1)–(A5) for* $x_1, \ldots, x_p \in (0,1)$ *the following convergence in distribution holds:*

$$(dn)^{2/5} \begin{bmatrix} \widehat{m}_1(x_1) - m_1(x_1) \\ \vdots \\ \widehat{m}_p(x_p) - m_p(x_p) \end{bmatrix} \to N\left(\begin{bmatrix} \beta_1(x_1) \\ \vdots \\ \beta_p(x_p) \end{bmatrix}, \begin{bmatrix} v_1(x_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & v_p(x_p) \end{bmatrix} \right),$$

where $v_j(x_j) = \int K(u)^2\, du\, \sigma^2 [f_j(x_j) c_j]^{-1}$.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

According to Theorem 2 in general the parameter η_t is only estimated with an order of $n^{-2/5}$. This is caused by a bias of the estimator that is of this order. We conjecture that the stochastic part of $\widehat{\eta}_t$ is of parametric order $n^{-1/2}$. As in other semiparametric problems use of smoothed likelihood functions leads to an estimator of the parametric component that does not achieve an $n^{-1/2}$ rate. Estimates of the parametric component with $n^{-1/2}$ rate can be achieved by application of a two-step procedure where an unsmoothed likelihood is used for the parametric component and a smoothed likelihood for the nonparametric part.

2.5. Asymptotics for $T \to \infty$, $n \to \infty$

The mathematical arguments of the last section make essential use of the assumption that T is fixed. In some applications of panels of time series this assumption may not be justified. We now discuss asymptotics for the estimators $\widehat{m}_j$ of the preceding sections for the case that $T \to \infty$. Again, the estimators are defined as minimizers of (2.2), and they fulfill equations (2.4) and (2.5) by the same arguments as in Section 2.3. By plugging (2.5) into (2.4) we get

$$\widehat{m}_j(x_j) = \widehat{m}_j^*(x_j) - \sum_{l \neq j} \int \widehat{m}_l(x_l) \frac{\widehat{f}_{jl}(x_j, x_l)}{\widehat{f}_j(x_j)}\, dx_l \tag{2.8}$$

$$+ \sum_{l=1}^{p} \sum_{t=1}^{T} \frac{N_t}{N} \int \widehat{m}_l(u_l) \frac{\widehat{f}_l^t(u_l)\, \widehat{f}_j^t(x_j)}{\widehat{f}_j(x_j)}\, du_l,$$

where

$$\widehat{m}_j^*(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})[Y^{it} - \bar{Y}^{\cdot t}] / \widehat{f}_j(x_j),$$

$$\bar{Y}^{\cdot t} = N_t^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) Y^{it}.$$

Again, the functions $\widehat{m}_j$ are chosen such that the constraints (2.3) are fulfilled.

With $\widehat{m}(x) = (\widehat{m}_1(x_1), \ldots, \widehat{m}_p(x_p))^T$, $\widehat{m}^*(x) = (\widehat{m}_1^*(x_1), \ldots, \widehat{m}_p^*(x_p))^T$ this can be rewritten as

$$\widehat{m}(x) = \widehat{m}^*(x) + \int \widehat{H}(x, y)\, \widehat{m}(y)\, \widehat{f}(y)\, dy, \tag{2.9}$$

where $\widehat{H}(x, y)$ is a $p \times p$ matrix with off-diagonal entries

$$\widehat{H}_{jl}(x, y) = -\frac{\widehat{f}_{jl}(x_j, y_l)}{\widehat{f}_j(x_j)\, \widehat{f}_l(y_l)} + \sum_{t=1}^{T} \frac{N_t\, \widehat{f}_l^t(y_l)\, \widehat{f}_j^t(x_j)}{N\, \widehat{f}_j(x_j)\, \widehat{f}_l(y_l)}$$

and diagonal elements

$$\widehat{H}_{jj}(x, y) = \sum_{t=1}^{T} \frac{N_t\, \widehat{f}_j^t(y_j)\, \widehat{f}_j^t(x_j)}{N\, \widehat{f}_j(x_j)\, \widehat{f}_j(y_j)}.$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

For (2.9) we use the following shorthand notation:

$$\widehat{m} = \widehat{m}^* + \widehat{H}\widehat{m}\,. \tag{2.10}$$

The operator $\widehat{H}$ is related to the operator $\widehat{S}$ that appears in the proof of Theorems 1 and 2. The operator $\widehat{S}$ consists of an iterative application of orthogonal projections (or more precisely of orthogonal projections that depend on n but converge to fixed orthogonal projections). This was the central argument to show that the operator norm of $\widehat{S}$ is strictly smaller than one. If the operator norm of $\widehat{H}$ were strictly smaller than one, then iterative application of (2.10) would give

$$\widehat{m} = \sum_{r=0}^{\infty} \widehat{H}^r \widehat{m}^*,$$

and we could treat asymptotics for $\widehat{m}$ as in Section 2.4. Unfortunately, we are not aware of an argument that implies that the norm of $\widehat{H}$ is bounded away from one (with probability tending to one for n large enough). Here, we will use a general simple theorem for the solution of empirical integral equations of the form (2.10). This theorem is a modification of Proposition 1 in Linton and Mammen (2005). The integral equation (2.10) is the empirical counterpart of the following equation:

$$m = m^* + Hm, \tag{2.11}$$

which now is the short form of $m(x) = m^*(x) + \int H(x,y)\,m(y)\,f(y)\,dy$ where $m(x) = (m_1(x_1), \ldots, m_p(x_p))^T$ and $m^*(x) = (m_1^*(x_1), \ldots, m_p^*(x_p))^T$ with

$$m_j^*(x_j) = m_j(x_j) + \sum_{l \neq j} \int \frac{f_{jl}(x_j, x_l)}{f_j(x_j)} m_l(x_l)\, dx_l$$

$$- \int \sum_{t=1}^{T} \sum_{l=1}^{p} \frac{d_t\, f_l^t(y_l)\, f_j^t(x_j)}{d f_j(x_j)} m_l(y_l)\, dy_l \tag{2.12}$$

and where $H(x,y)$ is a $p \times p$ matrix with off-diagonal entries

$$H_{jl}(x,y) = -\frac{f_{jl}(x_j, y_l)}{f_j(x_j)\, f_l(y_l)} + \sum_{t=1}^{T} \frac{d_t\, f_l^t(y_l)\, f_j^t(x_j)}{d f_j(x_j)\, f_l(y_l)}$$

and diagonal elements

$$H_{jj}(x,y) = \sum_{t=1}^{T} \frac{d_t\, f_j^t(y_j)\, f_j^t(x_j)}{d f_j(x_j)\, f_j(y_j)}\,.$$

Note that we now discuss a setting where T may tend to infinity for $n \to \infty$. For this reason the quantities d, f_j, and f_{jl} $(1 \le j, l \le p)$ depend on T. This implies

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

that also H and m_j^* depend on T. Only for $d = d(T)$ do we indicate this in our notation.

In allowing T tending to infinity, we assume that the rate is not too fast. Faster speeds could be allowed if more detailed assumptions were made on the time series structure of (X^{it}, Y^{it}). In particular, mixing type of conditions would be needed to show that averaging over t increases the speed of convergence of the estimators.

(A2′) T may depend on n, and it holds that $T^{3/2}n^{-1+\delta} \to 0$ for some $\delta > 0$.

We need the following uniform version of (A3).

(A3′) We assume that f^t has continuous partial derivatives of order one (uniformly for $1 \le t \le T$) and that it is bounded from below (on $[0,1]^p$, uniformly for $1 \le t \le T$). The regression functions m_j have continuous second derivatives on $[0,1]$.

We will show in the proof of the following theorem that in particular this assumption implies that the operator H is bounded (see condition (B1) of Theorem 5 in Section 6). Furthermore, in the proof we will see that the following assumption implies that $I - H$ is invertible and that its inverse is bounded (see condition (B2) of Theorem 5).

(A6) It holds that

$$\inf \sum_{t=1}^{T} \frac{d_t}{d(T)} \mathrm{E}\left[\left\{m_1(X_1^{it}) + \cdots + m_p(X_p^{it}) + \eta_t\right\}^2\right] > 0.$$

Here the infimum runs over $T \ge 1$, $\eta_1, \ldots, \eta_T \in \mathbb{R}$, and functions $m_1, \ldots, m_p$ with $\int m_1^2(x_1) f_1(x_1)\, dx_1 + \cdots + \int m_p^2(x_p) f_p(x_p)\, dx_p = 1$ and $\int m_1(x_1) f_1(x_1)\, dx_1 = \cdots = \int m_p(x_p) f_p(x_p)\, dx_p = 0$.

We are now in a position to state an expansion for $\widehat{m}$.

THEOREM 3. *Assume (A1), (A2′), (A3′), and (A4)–(A6). Then, as in Theorem 2, the following uniform expansion holds:*

$$\sup_{x_j \in [h_j, 1-h_j]} \left|\widehat{m}_j(x_j) - \left[\widetilde{m}_j^A(x_j) + (d(T)n)^{-2/5}\beta_j(x_j)\right]\right| = o_p((d(T)n)^{-2/5}).$$

Here $\widetilde{m}_j^A$, $\beta_1, \ldots, \beta_p$ are defined as in Theorem 2. For $x_1, \ldots, x_p \in (0,1)$ the following convergence in distribution holds:

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$$(d(T)n)^{2/5}\begin{bmatrix}\widehat{m}_1(x_1)-m_1(x_1)\\ \vdots\\ \widehat{m}_p(x_p)-m_p(x_p)\end{bmatrix}-\begin{bmatrix}\beta_1(x_1)\\ \vdots\\ \beta_p(x_p)\end{bmatrix}$$

$$\to N\left(\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix},\begin{bmatrix}v_1(x_1) & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & v_p(x_p)\end{bmatrix}\right),$$

where v_j is defined as in Corollary 1.

The typical order of $d(T)$ is $O(T)$. In particular, in case of a stationary setting we will have $d(T)/T \to d^*$ for a constant $d^* > 0$.

The estimator $\widehat{m}$ can be calculated as in Section 2.2. That means that the estimator of $\eta_1, \dots, \eta_T$ is also calculated and that it is iteratively updated together with the estimator of m. But it is not clear if the result of Theorem 1 still applies for this algorithm. The proof of Theorem 1 uses the fact that the operator S is fixed and does not depend on n. This does not hold anymore if T depends on n. In particular, we are not aware of an argument that shows that the operator norm of S is bounded away from one. It can be shown that the operator norm is smaller than one for each value of n, but not the uniform version of this statement. This implies that the algorithm converges for each n, but it does not imply that the speed of the algorithm is the same for all n, as stated in Theorem 1. In the simulations and in our data example we have used this algorithm.

2.6. Inclusion of Individual Effects into the Model

In a more general model individual effects are included. The model then becomes

$$Y^{it} = \sum_{j=1}^{p} m_j(X_j^{it}) + \eta_t + \lambda_i + \varepsilon^{it} \qquad (i = 1, \dots, n;\ \ t = 1, \dots, T).$$

Here $\lambda_1, \dots, \lambda_n$ are constants that are unknown. For identifiability, we assume $\sum_{i=1}^{n} N^i \lambda_i = 0$, where N^i is the number of covariables X^{it} that lie in $[0,1]^p$ for fixed i and for $t = 1, \dots, T$. In a first attempt one can again fit a model (2.12) by minimizing a smoothed least squares criterion. The estimators $\widehat{m}_1, \dots, \widehat{m}_p, \widehat{\eta}_1, \dots, \widehat{\eta}_T, \widehat{\lambda}_1, \dots, \widehat{\lambda}_n$ now are defined as minimizers of

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\int\left[Y^{it} - \sum_{j=1}^{p}\widehat{m}_j(u_j) - \widehat{\eta}_t - \widehat{\lambda}_i\right]^2$$

$$K_{h(1)}(u_1, X_1^{it}) \times \cdots \times K_{h(p)}(u_p, X_p^{it})\, du_1 \times \cdots \times du_p$$

under the constraints (2.3) and

$$\sum_{i=1}^{n} \frac{N^i}{N}\widehat{\lambda}_i = 0.$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

By taking derivatives one gets that the estimators fulfill linear equations that are similar to (2.4) and (2.5). By plugging the formulas of $\widehat{\eta}_t$ and $\widehat{\lambda}_i$ into the expression for $\widehat{m}_j$ one can show that $\widehat{m} = (\widehat{m}_1, \ldots, \widehat{m}_p)^T$ fulfills the following integral equation (compare (2.9)):

$$\widehat{m}(x) = m^*(x) + \int \widehat{H}(x, y)\widehat{m}(y)\, dy,$$

where

$$m_j^*(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) \left[Y^{it} - \bar{Y}^{\cdot t} - \bar{Y}^{i\cdot} + \bar{Y} \right],$$

$$\bar{Y}^{i\cdot} = (N^i)^{-1} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) Y^{it},$$

$$\widehat{H}_{jl}(x, y) = \frac{\widehat{f}_{jl}(x_j, y_l)}{\widehat{f}_j(x_j)} + \sum_{t=1}^{T} \frac{N^t \widehat{f}_l^t(y_l) \widehat{f}_j^t(x_j)}{N \widehat{f}_j(x_j)} + \sum_{i=1}^{n} \frac{N^i \widetilde{f}_l^i(y_l) \widetilde{f}_j^i(x_j)}{N \widehat{f}_j(x_j)} \quad \text{for} j \neq l,$$

$$\widehat{H}_{jj}(x, y) = \sum_{t=1}^{T} \frac{N^t \widehat{f}_l^t(y_l) \widehat{f}_j^t(x_j)}{N \widehat{f}_j(x_j)} + \sum_{i=1}^{n} \frac{N^i \widetilde{f}_l^i(y_l) \widetilde{f}_j^i(x_j)}{N \widehat{f}_j(x_j)},$$

$$\widetilde{f}_j^i(x_j) = \frac{1}{N^i} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p)\ K_{h(j)}(X_j^{it} - x_j).$$

Under appropriate conditions one can proceed for the discussion of $\widehat{m}_j(x_j)$ as in Section 2.5. For the "stochastic" term of $\widehat{m}_j(x_j)$ one now gets

$$N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) [\varepsilon^{it} - \bar{\varepsilon}^{\cdot t} - \bar{\varepsilon}^{i\cdot} + \bar{\varepsilon}]$$

with an obvious definition of $\bar{\varepsilon}^i$.

Compared with $\widetilde{m}_j^A$ this now contains the additional term $\bar{\varepsilon}^{i\cdot}$. Clearly, $\bar{\varepsilon}^{i\cdot}$ is not independent of X_j^{it} in the case that the covariables contain lagged observations, and for T small it is not asymptotically negligible (as is always the case for $\bar{\varepsilon}$). A more careful analysis shows that now the stochastic term only has asymptotic mean zero if T grows fast enough and appropriate mixing conditions apply (or if all covariables are external variables).

3. LOCAL LINEAR ESTIMATORS

The local constant estimators $\widehat{m}_j$ have a complicated bias; see Theorem 2 and Corollary 1. The bias of $\widehat{m}_j$ depends on the shape of the other regression

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

functions m_l ($l \neq j$). This complicates the statistical inference based on $\widehat{m}_j$. This disadvantage is known from backfitting estimators of additive models. There it can be avoided by using local linear instead of local constant smoothing. The same holds true for panels of time series. We will discuss this in this section. We will only do it for the setting of Section 2.2 (i.e., T constant, no individual effects). Local linear estimators

$$\widehat{m}_1, \ldots, \widehat{m}_p, \widehat{m}^1, \ldots, \widehat{m}^p, \widehat{\eta}_1, \ldots, \widehat{\eta}_T$$

are defined as minimizers of

$$\sum_{i=1}^{n}\sum_{t=1}^{n}\int \left[Y^{it} - \sum_{j=1}^{p} \widehat{m}_j(u_j) - \frac{X_j^{it} - u_j}{h(j)}\widehat{m}^j(u_j) - \widehat{\eta}_t\right]^2 \tag{3.1}$$

$$\cdot K_{h(1)}(u_1, X_1^{it}) \times \cdots \times K_{h(p)}(u_p, X_p^{it})\, du_1 \times \cdots \times du_p.$$

Instead of (2.3) we now use the following identifying constraints:

$$\int \widehat{m}_j(x_j)\widehat{f}_j(x_j)\, dx_j + \int \widehat{m}^j(x_j)\widehat{f}_j^1(x_j)\, dx_j = 0 \tag{3.2}$$

for $j = 1, \ldots, p$, where the kernel density estimators $\widehat{f}_j$ are defined as before and where

$$\widehat{f}_j^1(x_j) = N^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it})(X^{it} - x_j).$$

This constraint has also been used in Mammen and Park (2005) for smooth local linear backfitting. We will discuss it after the statement of the next theorem.

Again, $\widehat{m}_1, \ldots, \widehat{m}_p$ are estimators of $m_1, \ldots, m_p$. The estimators $\widehat{m}^1, \ldots, \widehat{m}^p$ fit the local slope and can be used as estimators of the derivatives $m_1', \ldots, m_p'$. By differentiation of the left-hand side of (3.1) (with respect to $\widehat{m}_j(u_j), \widehat{m}^j(u_j)$, and $\widehat{\eta}_t$) one gets the following linear equations for the estimators:

$$\widehat{M}_j(x_j)\begin{pmatrix}\widehat{m}_j(x_j)\\ \widehat{m}^j(x_j)\end{pmatrix} = \widehat{M}_j(x_j)\begin{pmatrix}\widetilde{m}_j(x_j)\\ \widetilde{m}^j(x_j)\end{pmatrix}$$

$$-\sum_{l\neq j}\int \widehat{S}_{l,j}(x_{l,j}(x_l, x_j)\begin{pmatrix}\widetilde{m}_l(x_l)\\ \widetilde{m}^l(x_l)\end{pmatrix} dx_l - \sum_{t=1}^{T}\widehat{\eta}_t\begin{pmatrix}\widehat{V}_{0,0}^{j,t}(x_j)\\ \widehat{V}_{j,0}^{j,t}(x_j)\end{pmatrix},$$

$$\widehat{\eta}_t = \widetilde{\eta}_t - \frac{N}{N_t}\sum_{j=1}^{p}\int \widehat{V}_{0,0}^{j,t}(u_j)\widehat{m}_j(u_j)\, du_j$$

$$-\frac{N}{N_t}\sum_{j=1}^{p}\int \widehat{V}_{0,1}^{j,t}(u_j)\widehat{m}^j(u_j)\, du_j,$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

where

$$\widehat{M}_j(x_j) = \begin{pmatrix} 1 & h(j)^{-1}[X_j^{it} - x_j] \\ h(j)^{-1}[X_j^{it} - x_j] & h(j)^{-2}[X_j^{it} - x_j]^2 \end{pmatrix},$$

$$\widehat{S}_{l,j}(x_l, x_j) = \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^i \in [0,1]^p) K_{h(j)}(x_j, X_j^{it})\ K_{h(l)}(x_l, X_l^{it})$$

$$\times \begin{pmatrix} 1 & h(l)^{-1}[X_l^{it} - x_l] \\ h(j)^{-1}[X_j^{it} - x_j] & h(j)^{-1}h(l)^{-1}[X_l^{it} - x_l][X_j^{it} - x] \end{pmatrix},$$

$$\widehat{V}_{0,0}^{j,t}(x_j) = \frac{1}{N} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it}),$$

$$\widehat{V}_{j,0}^{j,t}(x_j) = \frac{1}{N} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_{h(j)}(x_j, X_j^{it}) h(j)^{-1}(X_j^{it} - x_j).$$

We now state asymptotic normality of the local linear estimator.

THEOREM 4. *Under conditions (A1)–(A5) the following convergence in distribution holds:*

$$(d(T)n)^{2/5} \begin{bmatrix} \widehat{m}_1(x_1) - m_1(x_1) \\ \vdots \\ \widehat{m}_p(x_p) - m_p(x_p) \end{bmatrix} \to N\left(\begin{bmatrix} \delta_1(x_1) \\ \vdots \\ \delta_p(x) \end{bmatrix}, \begin{bmatrix} v_1(x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & v_p(x_p) \end{bmatrix} \right),$$

where $v_j(x_j)$ is defined as in Corollary 1 and where

$$\delta_j(x_j) = c_j^2 \frac{1}{2} \int u^2 K(u)\, du\, m_j''(x_j).$$

Furthermore it holds that

$$\widehat{\eta}_t = \eta_t + o_p(n^{-2/5}).$$

The norming (3.2) could be replaced by other constraints. Other natural choices would be (2.3) or

$$N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) \widehat{m}_j(X_j^{it}) = 0. \tag{3.3}$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

We have chosen (3.2) because it leads to estimators of η_t that have bias terms of lower order $o_p(n^{-2/5})$. The normings (2.3) and (3.3) lead in the estimation of η_t to an additional bias term of order $n^{-2/5}$. Up to terms of order $o_p(n^{-2/5})$ this bias term is constant in t. Similar changes arise in the estimation of the nonparametric components m_j. Here, the replacement of (3.2) by (2.3) or (3.3) adds an additional bias term that is asymptotically constant (in x_j). For our choice (3.2) the asymptotic bias δ_j coincides with the asymptotic bias of the classical local linear estimator.

We now explain why the nonparametric bias term in the estimation of η_t does not depend on t. This can be seen by the following heuristics. For a full-dimensional local linear fit of the regression function $m(x,t) = \eta_t + m_1(x_1) + \cdots + m_p(x_p)$ one gets a bias term for the estimation of an additive component m_j that only depends on the second derivative of m_j but that does not depend on the density f_t. In particular, it does not depend on t. Therefore $\widehat{\eta}_t$ has only a nonparametric bias that does not depend on t and that could be removed by norming.

4. SIMULATIONS AND A REAL DATA APPLICATION

In this section we carry out a simulation study and use a real data set to see how one of the proposed estimation procedures behaves. The estimation algorithm in equations (2.4)–(2.6) has been implemented. In the implementation, however, we do not use the interval $[0,1]$, but take $N = nT$ and $N_t = n$. The convergence criterion we use in the simulations is as follows: if for all $j = 1, \ldots, p$

$$\frac{\sum[\widehat{m}_j^{[a+1]}(x_j) - \widehat{m}_j^{[a]}(x_j)]^2}{\sum[\widehat{m}_j^{[a]}(x_j)]^2 + 0.0001} < 0.0001, \tag{4.1}$$

then stop. Here a indicates the number of iterations performed. This convergence criterion is taken from Nielsen and Sperlich (2005). The Gaussian kernel has been used in the nonparametric estimator $\widetilde{m}(x_j)$ and in the density estimators. Using a compact kernel (cf. Assumption (A.5)) produces similar results. The choice of bandwidth is important in nonparametric estimation. For the smooth backfitting algorithm there exist some papers on how to choose the bandwidth, e.g., a proposed cross-validation technique in Nielsen and Sperlich (2005) and a bandwidth selector based on penalized sums of squared residuals in Mammen and Park (2005). However, for the sake of computing speed and simplicity for this panel data setting, we have choosen to use the "rule-of-thumb" bandwidth selector; see, e.g., Härdle (1990, p. 91). This bandwidth choice may not be optimal, especially for dependent data; thus another choice of bandwidth may improve the final estimates, as will be seen in our last simulation example. We leave the development of a more optimal bandwidth selector for the proposed algorithms for future research.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

4.1. Simulation Study

The first simulations are similar to the setup in Nielsen and Sperlich (2005), although there the authors only used cross-section data. We consider the model

$$Y^{it} = \sum_{j=1}^{p} m_j(X_j^{it}) + \eta_t + \varepsilon^{it}, \tag{4.2}$$

with $m_j(x_j^{it}) = \sin(\pi x_j^{it})$ and $\varepsilon^{it} \sim N(0,1)$. In the simulation, we first draw variables z_j^{it}, $j = 1, \ldots, p$, $t = 1, \ldots, T$, for i fixed, from $N(0,1)$ with given correlation ρ_{jk}, $j \neq k$ for each combination of covariables. That is, for each i, the variables z_j^{it} are correlated with z_k^{it}. This drawing of variables is repeated for all $i = 1, \ldots, n$. Then a transformation is made, $x_j^{it} = 2.5 \arctan(z_j^{it})/\pi$. Thus the variables are projected into the interval $[-1.25, 1.25]$.

In the first simulation, 100 realizations with $n = 100$, $T = 10$, $p = 4$, $\eta_t = 0$ for all t and $\rho_{jk} = 0.1$ for all $j \neq k$ are carried out. This simulation acts as a benchmark for the other simulations in this paper, where dependence between the series $\{Y^{it}\}$ is allowed.

Before we give the results from the first simulation, note that by using Corollary 1 we can construct confidence intervals of $\widehat{m}(x_j)$. Assuming that the bias of $\widehat{m}(x_j)$ is of negligible size compared with the variance, we set $\beta_1(x_1), \ldots, \beta_4(x_4)$ equal to zero in Corollary 1. Further we assume $c_j = 1$ for all j, where c_j is defined in cap Assumption (A5). This corresponds to the "rule-of-thumb" bandwidth used (cf. Härdle, 1990, p. 91). Then the confidence intervals for all j components are equal to

$$\left[\widehat{m}_j(x_j) - \frac{z_{\alpha/2} v_j(x_j)^{1/2}}{N^{2/5}}, \widehat{m}_j(x_j) + \frac{z_{\alpha/2} v_j(x_j)^{1/2}}{N^{2/5}} \right], \tag{4.3}$$

where $z_{\alpha/2}$ is a quantile in the standard normal distribution and

$$v_j(x_j) = \sigma^2 f_j(x_j)^{-1} \int K(u)^2 \, du. \tag{4.4}$$

Observe that we also have to estimate σ. This is done by calculating the empirical standard deviation of the estimated ε^{it}, where

$$\hat{\varepsilon}^{it} = Y^{it} - \sum_{j=1}^{p} \widehat{m}_j(X_j^{it}) - \widehat{\eta}_t. \tag{4.5}$$

Note that these confidence intervals may not be very helpful when only moderate sample sizes are available.

A typical example (of the 100 realizations) of the estimated functions $\widehat{m}_j(x_j)$ and $\widetilde{m}_j(x_j)$ from the first simulation is given as dashed and thin solid lines,

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

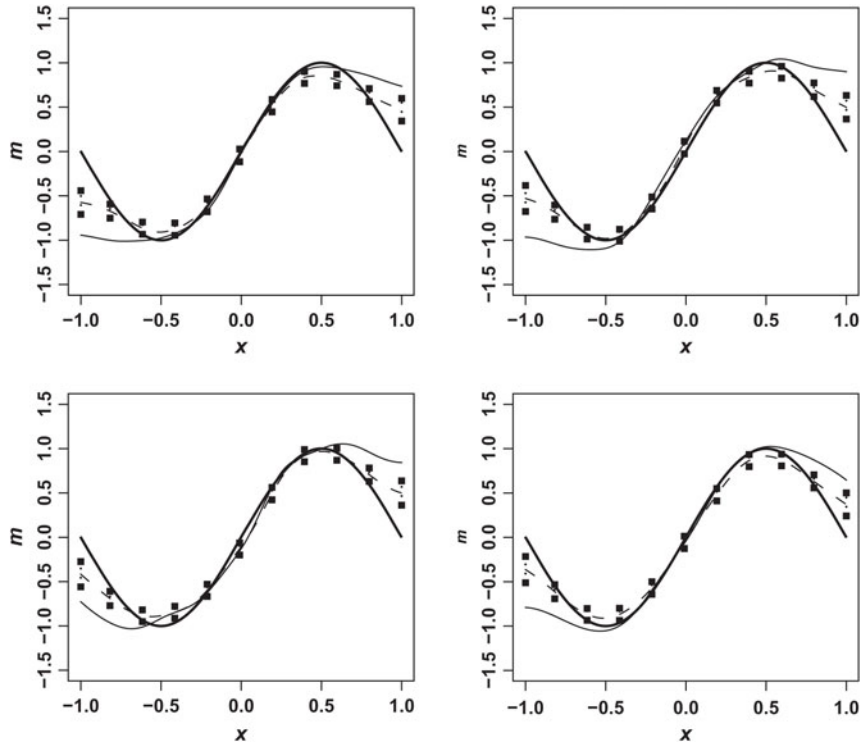


FIGURE 1. The estimated functions and confidence intervals for $m_j(x_j) = \sin(\pi x_j)$, $\rho_{jk} = 0.1$, and $\eta_t = 0$ (m_1 and m_2—upper plots; m_3 and m_4—lower plots).

respectively, in Figure 1. The thick solid lines are the true functions. The estimated functions are quite close to the true curves. Observe also that the iterated functions $\widehat{m}_j(x_j)$ are more accurate than the initial estimate $\widetilde{m}_j(x_j)$, as expected. The confidence intervals for $\alpha = 0.05$ are plotted in the figures as vertical dashed lines at each point. The confidence intervals are calculated from equation (4.3) by preestimation of σ (using (4.5)) and f_j. These confidence intervals agree reasonably well with the estimated plus or minus two standard deviations, based on the 100 realizations. Quite possibly a bootstrap approach for obtaining confidence intervals could be of interest. The average and the standard deviation of the estimated $\widehat{\eta}_t$ are given in the top panel of Table 1 (calculated from the 100 realizations).

The second simulation is identical to the first, except that the correlation $\rho_{jk} = 0.7$ for all $j \neq k$. In this case the initial estimates $\widetilde{m}_j(x_j)$ do not work at all, whereas the iterated estimates work well. This difference in behavior is expected because of the increased correlation ρ_{jk}. The average and standard deviations of the estimated η_t were similar to those in Table 1.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

TABLE 1. The estimated $\widehat{\eta}_t$'s for the first model

t	1	2	3	4	5	6	7	8	9	10
η_t	0	0	0	0	0	0	0	0	0	0
$\hat{\mathrm{E}}(\widehat{\eta}_t)$	−0.005	0.010	−0.023	−0.001	0.013	−0.006	0.028	−0.002	0.003	−0.017
$\hat{\mathrm{SD}}(\widehat{\eta}_t)$	0.099	0.102	0.100	0.102	0.094	0.105	0.111	0.119	0.097	0.095
η_t	−5	−4	−3	−2	−1	0	2	3	4	6
$\hat{\mathrm{E}}(\widehat{\eta}_t)$	−5.001	−4.013	−3.029	−1.989	−1.006	−0.017	2.004	3.003	3.985	5.983
$\hat{\mathrm{SD}}(\widehat{\eta}_t)$	0.122	0.130	0.109	0.133	0.116	0.095	0.107	0.117	0.124	0.119

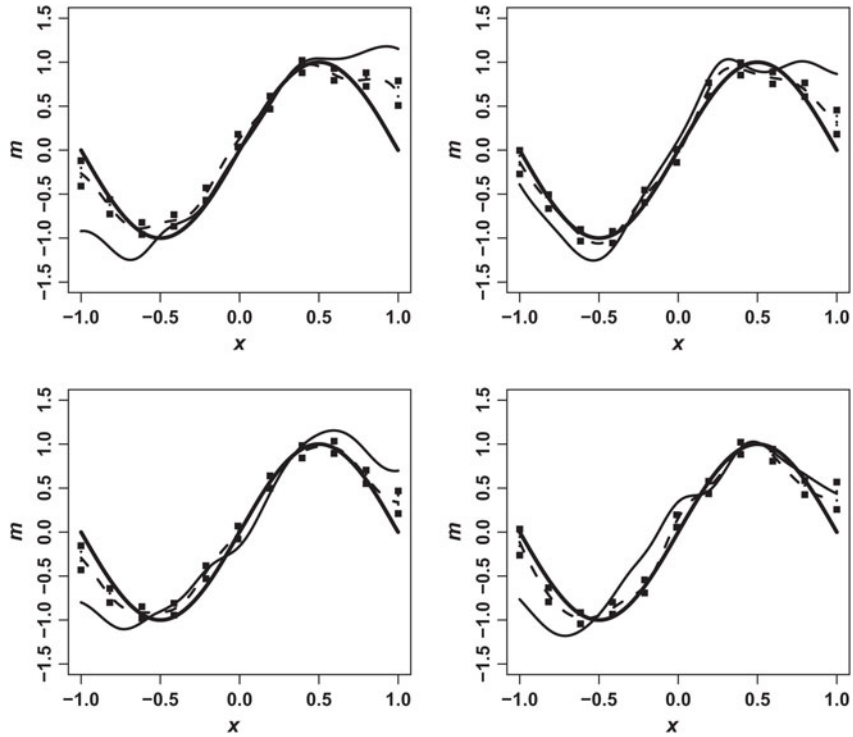


FIGURE 2. The estimated functions and confidence intervals for $m_j(x_j) = \sin(\pi x_j)$, $\rho_{jk} = 0.1$, and $\eta_t \neq 0$ (m_1 and m_2—upper plots; m_3 and m_4—lower plots).

In the last simulation for this model nonzero values of η_t were allowed, as given in the bottom panel of Table 1. Now ρ_{jk} is 0.1. A typical example of the estimated functions with confidence intervals (again, these intervals are estimated using **(4.3)**) is given in Figure 2, and the average and standard deviations for the estimated η_t's are given in the lower part of Table 1.

Next, we look at a panel of time series. The time series lag dependence is taken from Fan and Yao (2003, p. 356). Now,

$$X^{it} = \sum_{j=1}^{2} m_j(X^{i,t-j}) + \eta_t + \varepsilon^{it}, \tag{4.6}$$

with

$$m_1(x) = 4x/(1+0.8x^2), \tag{4.7}$$

$$m_2(x) = \exp\{3(x-2)\}/[1+\exp\{3(x-2)\}]. \tag{4.8}$$

In the first simulation, we ran 100 realizations with $n = 200$, $T = 10$, $\eta_t = 0$ for all t and ε uniform on $[-1, 1]$. An example of the estimated $\widehat{m}_j(x)$ functions is

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

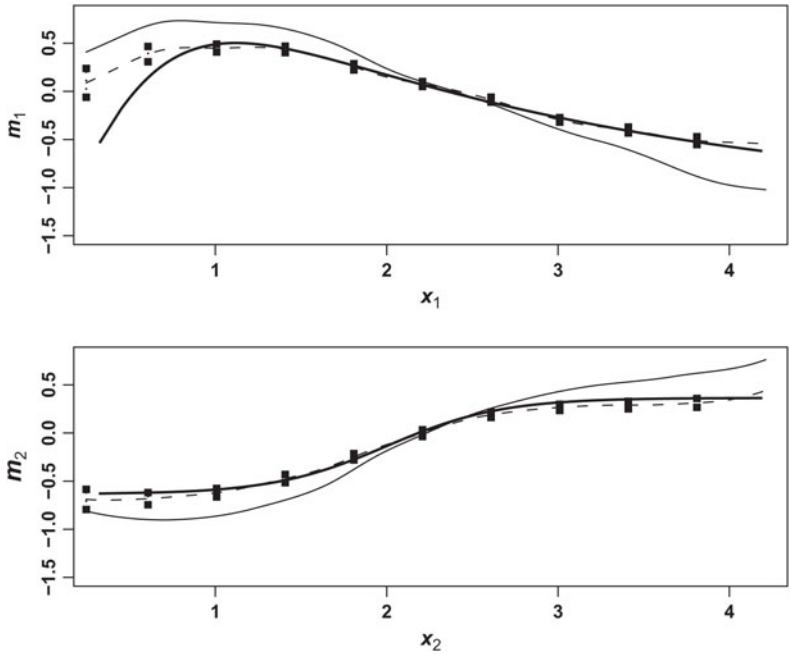


FIGURE 3. The estimated functions and empirical standard deviations in the time series model with $\eta_t = 0$.

given in Figure 3 as the dashed lines, and they give a good approximation to the true functions, thick solid lines. Again, $\widetilde{m}_j(x_j)$ are given as thin solid lines. Note that the additive components can only be identified up to a constant; thus the true lines in this figure and in Figure 4 are the recentered versions of the functions m_j, i.e., $m_j - \bar{m}_j$, plotted against $X^{i,t-j}$, where $\bar{m}_j$ indicates averaging. The standard deviations of $\widehat{m}_j(x_j)$ are estimated based on the 100 realizations in 11 different points, and plus or minus one standard deviation is plotted in the figures as vertical dashed lines at each point. As expected, the standard deviations are quite large at both boundaries for all components but are small at interior points.

The averages of the estimated η_t and their standard deviations are given in the top panel of Table 2. These are the empirical estimates based on the 100 realizations.

The last simulation is identical to the preceding one, except that η_t is now nonzero, as given in the bottom panel of Table 2. This introduces a quite strong relationship between the time series, especially at the ends. Now the estimated functions deteriorate in quality, and the estimates are more wiggly. The estimated η_t however, are quite good, with relatively small standard deviations; see the bottom panel of Table 2. Because the estimates of the two components are more wiggly, one might believe that a larger bandwidth in the estimates $\widetilde{m_j}(x_j)$ might

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

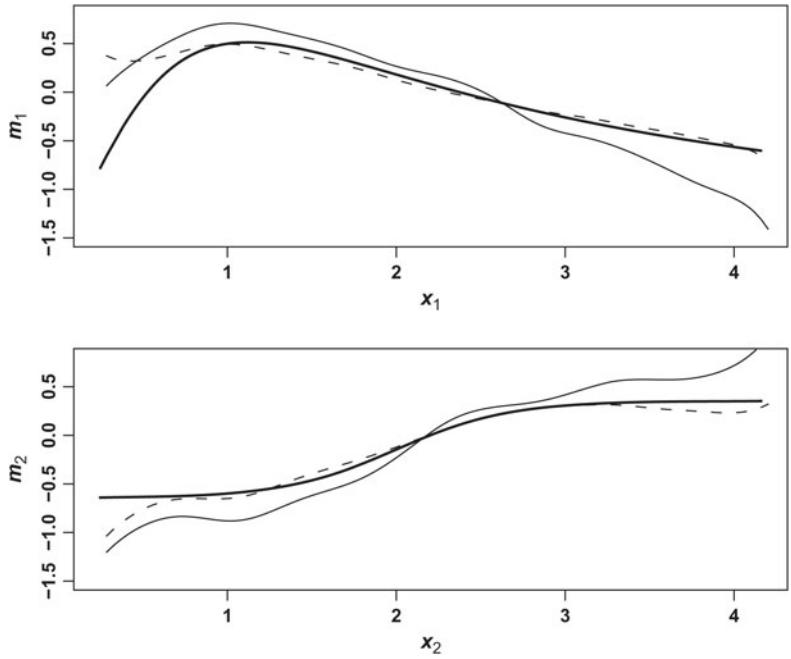


FIGURE 4. The estimated functions in the time series model, $\eta_t \neq 0$, with a larger bandwidth.

improve the estimates. And indeed that is the case. In Figure 4 the estimated components are given, based on one realization, but now using a bandwidth 1.5 times larger.

4.2. Application to a Real Data Set

The real data set is from Baltagi, Griffin, and Xiong (2000). They estimate a dynamic demand model for cigarettes based on panel data from 46 American states over the period 1963–1992. The estimated equation is

$$\ln Y^{it} = \alpha + \beta_1 \ln Y^{i,t-1} + \beta_2 \ln X_2^{it} + \beta_3 \ln X_3^{it} + \beta_4 \ln X_4^{it} + u^{it}, \tag{4.9}$$

where $i = 1, \ldots, 46$ denotes the ith state and $t = 1, \ldots, 29$ denotes the tth year. The term Y^{it} is real per capita sales of cigarettes, X_2^{it} is the average retail price of a pack of cigarettes measured in real terms, X_3^{it} is real per capita disposable income, and X_4^{it} denotes the minimum real price of cigarettes in any neighboring state. The disturbance term in general could be specified as

$$u^{it} = \lambda_i + \eta_t + \varepsilon^{it}. \tag{4.10}$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

TABLE 2. The estimated $\widehat{\eta}_t$'s for the second model

t	1	2	3	4	5	6	7	8	9	10
η_t	0	0	0	0	0	0	0	0	0	0
$\hat{\mathrm{E}}(\widehat{\eta}_t)$	0.017	0.018	0.024	0.024	0.018	0.026	0.022	0.024	0.020	0.019
$\hat{\mathrm{SD}}(\widehat{\eta}_t)$	0.047	0.040	0.045	0.046	0.041	0.043	0.047	0.044	0.038	0.040
η_t	−5	−4	−3	−2	−1	0	2	3	4	6
$\hat{\mathrm{E}}(\widehat{\eta}_t)$	−4.989	−3.994	−2.994	−1.996	−0.998	0.007	2.005	2.994	4.002	6.007
$\hat{\mathrm{SD}}(\widehat{\eta}_t)$	0.323	0.337	0.329	0.325	0.333	0.332	0.307	0.349	0.317	0.322

Baltagi et al. (2000) use different techniques to estimate the unknown β's and also in some cases λ_i and η_t. We assume that the model is

$$\ln Y^{it} = m_1(\ln Y^{it-1}) + m_2(\ln X_2^{it}) + m_3(\ln X_3^{it}) + m_4(\ln X_4^{it}) + \eta_t + \varepsilon^{it}, \quad \textbf{(4.11)}$$

and we would like to estimate the unknown m_j and η_t for all j and t. The presence of a λ_i term would require a large T as indicated in Section 2.6.

Because both the response $\ln Y^{it}$ and all the log-transformed covariates exhibit a trend in t, we detrend our observations as follows:

$$Y_d^{it} = \ln Y^{it} - g_Y(t), \quad \textbf{(4.12)}$$

where $g_Y(t)$ is a nonparametric estimator based on all observations of $\ln Y^{it}$. Likewise the different log-transformed covariates were detrended using

$$X_{jd}^{it} = \ln X_j^{it} - g_{X_j}(t). \quad \textbf{(4.13)}$$

The detrended Y_d^{it} observations are plotted against the different detrended covariates, X_{jd}^{it}, in Figure 5. Clearly the first covariate (the lagged variable), in the upper plot, is linear. The other covariates do not display strong nonlinearities. These detrended observations are then used as input to the implemented algorithm. Because we have rather few observations, we relax our convergence criteria.

The estimated functions $\widehat{m}_j$ are given in Figure 6. Here the thick solid lines correspond to the estimated linear model, ordinary least squares (OLS) with time-period dummies, in Baltagi et al. (2000). The estimated $\widehat{m}_j(x_j)$ and $\widetilde{m}_j(x_j)$ are again given as dashed and thin solid lines, respectively (note that the scale is very different from that of Figure 5). As could be expected from Figure 5, the estimated components using our algorithm only display weak nonlinearities, and the $\widehat{m}_j(x_j)$ estimates are also quite close to the estimated linear model. The initial $\widetilde{m}_j(x_j)$ estimates seem rather unstable in this situation, and it is encouraging that the $\widehat{m}_j(x_j)$ estimates do better in this respect. Because of the relatively small sample size we, do not plot the confidence intervals for the estimated components.

We briefly comment on the estimates and compare them to the results of the OLS with time-period dummies in Baltagi et al. (2000). The lagged consumption, i.e., Y^{t-1}, has a coefficient of 0.95 in the OLS. Note that this estimate is biased, because it is correlated with the omitted state-specific effects. Our estimate of m_1 seems to be relatively similar to this linear model. The price, income, and neighboring price elasticity (X_2, X_3, and X_4, respectively) are low in the OLS (-0.137, -0.01, and 0.037), whereas the nonparametric estimates seem slightly more significant. The outcome of the nonparametric analysis is that it supports the parametric specification.

The detrending of the responses, $\ln Y^{it}$, implies that the estimated η_t's should be close to zero, because $g_Y(t)$ can be thought of as an adjusted η_t for all t. And the estimation gives that the interval $[\min(\widehat{\eta}_t), \max(\widehat{\eta}_t)]$ is equal to $[-0.033, 0.025]$. We have also used the algorithm assuming $\eta_t = 0$ for all t, i.e., not estimating the η_t's, and similar results as in Figure 6 were obtained.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

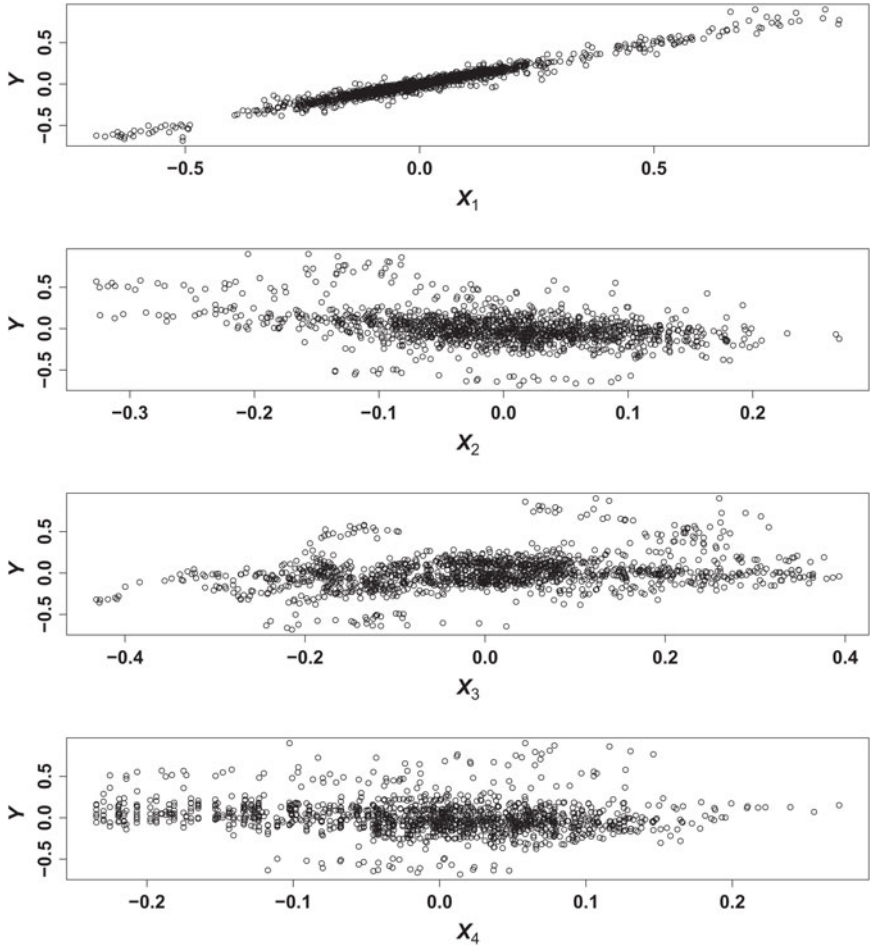


FIGURE 5. The detrended observations.

5. SUMMARY AND EXTENSIONS

The greatest part of the literature examining panels of time series considers parametric models. In this paper we have introduced iterative nonparametric estimators for additive nonlinear panel data models. The estimators presented are based on the smoothed backfitting approach of Mammen et al. (1999) for estimation in additive regression models.

The asymptotic behavior of the proposed estimators has been studied, mainly in the case where the number of individuals goes to infinity and the time periods are fixed. The presence of individual effects and/or temporal effects complicates the analysis.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

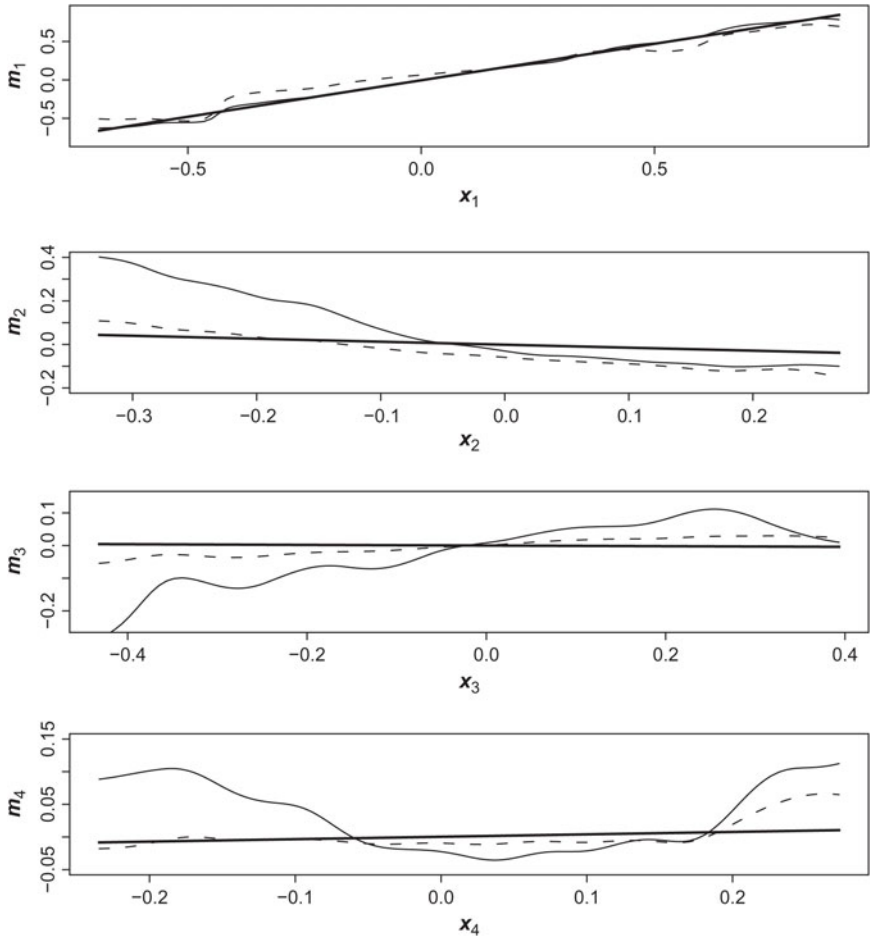


FIGURE 6. The estimated components in the real data set.

Several simulation experiments are done, and they indicate that the proposed estimators perform well. A real data set is also studied, with a reasonable result.

We could treat more complicated data generating processes. In particular, one could allow for conditional heteroskedasticity, and one could allow for weak conditional dependence of $\varepsilon^{1t}, \ldots, \varepsilon^{nt}$. We did not treat such general cases mostly for simplicity of notation. Essentially, for the proof of our results we only need asymptotic normality of $\sqrt{nTh}\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(x_j, X_j^{it})\varepsilon^{it}$ and rates of convergence for terms $(Tn)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}g_n(X_j^{it})\varepsilon^{it}$ for certain sequences of functions g_n. Clearly, this could be studied in a much more general setup. Furthermore, one could treat non-i.i.d. $\mathbf{X}^i$. In fact, we only make use of some ergodicity properties of the covariable vectors that are needed to get asymptotic formulas for bias and

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

variance expressions. Other generalizations include semiparametric extensions of the model. In particular, we think of partial additive specifications.

The criterion function (2.2) is a smoothed version of a sum of squared residuals. Smoothing is done with respect to the covariables $X_1^{11}, \ldots, X_p^{nT}$ but not with respect to t. Smoothing with respect to t could be easily included by minimizing

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\int \sum_{s=1}^{T}[Y^{it} - \sum_{j=1}^{p}\widehat{m}_j(u_j) - \widehat{\eta}_s]^2$$

$$\cdot\, \pi_{st}^{g} K_{h(1)}(u_1, X_1^{it}) \times \cdots \times K_{h(p)}(u_p, X_p^{it})\, du_1 \ldots du_p,$$

where π_{st}^{g} are smoothing weights depending on a "bandwidth" g and fulfilling $\sum_{s=1}^{T}\pi_{st}^{g} = 1$ for $t = 1, \ldots, T$. In particular, such an approach could be used in an asymptotic setting with $T \to \infty$. Discrete smoothing could also be applied to discrete covariables. This approach could be treated along the lines of the paper.

6. A GENERAL THEOREM FOR SOLUTIONS OF EMPIRICAL INTEGRAL EQUATIONS

In this section we state a general theorem for the solution of empirical integral equations of the form (2.10).

For linear spaces $\mathcal{M}_n$ of functions $m : \mathbb{R}^p \to \mathbb{R}^q$ endowed with pseudo norms $\|m\|_2^2 = \int \|m(x)\|^2 w_n(x)\, dx$ and $\|m\|_\infty = \sup_{x \in S_n} \|m(x)\|$ we consider linear operators $H_n : \mathcal{M}_n \to \mathcal{M}_n$ and solutions m_n of the equations $m_n = m_n^* + H_n m_n$. Here m_n^* is an element of $\mathcal{M}_n$. Furthermore, $w_n : \mathbb{R}^p \to \mathbb{R}$ is a positive function ($w_n(x) \geq 0$ for $x \in \mathbb{R}^p$), and S_n is a sequence of subsets of $\mathbb{R}^p$. We suppose that there exist estimators $\widehat{m}_n^*$ and $\widehat{H}_n$ of m_n^* and H_n, respectively.

THEOREM 5. *For the function m_n^* and the operator H_n and their estimators $\widehat{m}_n^*$ and $\widehat{H}_n$ we make the following assumptions. For two sequences $\delta_n, \Delta_n \geq 0$ we suppose (B1)–(B4).*

(B1) It holds that $\sup_{\|m\|_2 \leq 1} \|H_n m\|_\infty = O(1)$.

(B2) For all $n \geq 1$ the operator $I - H_n$ is invertible, and it holds that $\sup_{\|m\|_2 \leq 1} \|(I - H_n)^{-1} m\|_2 = O(1)$.

(B3) It holds that $\sup_{\|m\|_2 \leq 1} \|(\widehat{H}_n - H_n)m\|_\infty = o_P(\delta_n)$.

(B4) It holds that

$$\widehat{m}_n^* + \widehat{H}_n m_n - m_n = \widehat{m}_n^* - m_n^* + \widehat{H}_n m_n - H_n m_n$$

$$= \widehat{m}_n^A + \widehat{m}_n^B + \widehat{m}_n^C + \widehat{m}_n^D$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

with random functions $\widehat{m}_n^A$, $\widehat{m}_n^B$, *and* $\widehat{m}_n^C$ *that fulfill*

$$\|\widehat{m}_n^A\|_2 = O_P(\Delta_n \delta_n^{-1}), \tag{6.1}$$

$$\|H_n(I - H_n)^{-1}\widehat{m}_n^A\|_\infty = o_P(\Delta_n), \tag{6.2}$$

$$\|\widehat{m}_n^B\|_2 = O_P(\Delta_n \delta_n^{-1}), \tag{6.3}$$

$$\|\widehat{m}_n^D\|_\infty = o_P(\Delta_n). \tag{6.4}$$

Then with probability tending to one, there exists a solution $\widehat{m}_n$ *of the equation* $\widehat{m}_n = \widehat{m}_n^* + \widehat{H}_n\widehat{m}_n$, *and it holds that*

$$\|\widehat{m}_n - m_n - \widehat{m}_n^A - (I - H_n)^{-1}\widehat{m}_n^B - (I - \widehat{H}_n)^{-1}\widehat{m}_n^C\|_\infty = o_P(\Delta_n).$$

We now briefly discuss the statement of the theorem. The theorem states an expansion of the estimator $\widehat{m}_n$ that is of order $o_P(\Delta_n)$. Typically, the sequence Δ_n is the (nonparametric) rate of convergence of $\widehat{m}_n$. The main assumption is an expansion of $\widehat{m}_n^* - m_n^* + \widehat{H}_n m_n - H_n m_n$ into four terms $\widehat{m}_n^A, \ldots, \widehat{m}_n^D$. In many applications expansions for $\widehat{m}_n^* - m_n^*$ and $\widehat{H}_n m_n - H_n m_n$ can be given by using standard smoothing theory. We have stated the theorem for the following application where the terms $\widehat{m}_n^A, \ldots, \widehat{m}_n^D$ have the following meaning. The term $\widehat{m}_n^A$ is the "stochastic part," i.e., it is a weighted local average of mean-zero error variables. The weights depend on the covariables. Typically it holds that for an integral operator L the statistic $L\widehat{m}_n^A$ is a global average of mean-zero error variables. Thus $L\widehat{m}_n^A$ is of order $n^{-1/2} = o(\Delta_n)$. Condition (6.2) assumes this property for the integral operator $L = H_n(I - H_n)^{-1}$. The terms $\widehat{m}_n^B$ and $\widehat{m}_n^C$ are bias terms. The first term $\widehat{m}_n^B$ is a bias term that is of order Δ_n. Typically, it is asymptotically deterministic and does not vanish for most parts of $[0, 1]$. The term $\widehat{m}_n^C$ collects boundary bias terms. They are of larger order than Δ_n, but they vanish outside boundary regions. The term $\widehat{m}_n^D$ collects remainder terms in the expansion of (B4); see (6.4). The sequence δ_n describes the accuracy of the estimation of the operator H_n by $\widehat{H}_n$. This estimation requires higher dimensional smoothing, and therefore it is typically slower than Δ_n. The conditions (6.1) and (6.3) are technical assumptions that allow us to replace $\widehat{H}_n$ by H_n in some steps of the proofs. Typically these conditions can be easily verified.

7. PROOFS

Proof of Theorems 1 and 2. We proceed similarly as in the proofs of Theorems 1–4 in Mammen et al. (1999). We now define the space of additive functions:

$$\mathcal{H} = \{m : m(x,t) = m_1(x_1) + \cdots + m_p(x_p) + m_0(t) \quad \text{with } m_j \in L_2(f_j) \text{ for } 1 \le j \le p\}.$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

For $j = 1, \ldots, p$ we consider the following operator: $\widehat{\psi}_j : \mathcal{H} \to \mathcal{H}$. For a function $m(x,t) = m_1(x_1) + \cdots + m_p(x_p) + m_0(t)$ the operator does not change m_r for $r \neq j$, and it replaces m_j by

$$-\sum_{t=1}^{T} \frac{N_t}{N} m_0(t) \frac{\widehat{f}_j^t(x_j)}{\widehat{f}_j(x_j)} - \sum_{l \neq j} \int m_l(x_l) \frac{\widehat{f}_{jl}(x_j, x_l)}{\widehat{f}_j(x_j)}\, dx_l.$$

Note that for the index j equation (2.4) can be rewritten as

$$\widehat{m}(x,t) = \widetilde{m}_j(x_j) + \widehat{\psi}_j \widehat{m}(x,t), \tag{7.1}$$

where $\widehat{m}(x,t) = \widehat{m}_1(x_1) + \cdots + \widehat{m}_p(x_p) + \widehat{\eta}_t$. Furthermore, (2.5) can be rewritten as

$$\widehat{m}(x,t) = \widetilde{\eta}_t + \widehat{\psi}_0 \widetilde{m}(x,t), \tag{7.2}$$

where for a function $m(x,t) = m_1(x_1) + \cdots + m_p(x_p) + m_0(t)$ the function $r(x,t) = \widehat{\psi}_0 m(x,t)$ is defined as $r_1(x_1) + \cdots + r_p(x_p) + r_0(t)$ with $r_j(x_j) = m_j(x_j)$ and

$$r_0(t) = -\sum_{j=1}^{p} \int m_j(x_j) \widehat{f}_j^t(x_j)\, dx_j.$$

The algorithm studied in Theorem 1 is based on iterative application of the operator $\widehat{S} = \widehat{\psi}_0 \widehat{\psi}_p, \ldots, \widehat{\psi}_1$. We compare this operator with $S = \psi_0 \psi_p \ldots \psi_1$ where ψ_j is defined as $\widehat{\psi}_j$ but with $\widehat{f}_{jk}, \widehat{f}_j, \widehat{f}_j^t$, and N/N_t replaced by f_{jk}, f_j, f_j^t, and d/d_t, respectively. This operator does not depend on n, and it can be easily checked that ψ_j is the orthogonal projection in $\mathcal{H} \subset L_2(f)$ onto the orthogonal complement of $\mathcal{H}_j = \{m \in \mathcal{H} : m(x,t) = m_j(x_j)$ for a function m_j with $\int m_j(x_j) f_j(x_j)\, dx_j = 0$ (if $j \neq 0$) or $m(x,t) = m_0(t)$ for a function m_t with $\sum_{t=1}^{T} d_t m_0(t) = 0$ (if $j = 0)\}$. The operator S consists of an iterative application of orthogonal projections. As in Mammen et al. (1999) this can be used to show that the operator norm

$$\rho := \sup\{\|S(f_0 + \cdots + f_p)\|_2 : f_j \in \mathcal{H}_j, \quad \|f_0 + \cdots + f_p\|_2 \leq 1\}$$

is strictly less than one. Here $\|\ \|_2$ is the norm of the space $L_2(f)$. Furthermore, using uniform consistency of $\widehat{f}_{jk}$ and $\widehat{f}_j$ we get that

$$\widehat{\rho} := \sup\{\|\widehat{S}(f_0 + \cdots + f_p)\|_2 : f_j \in \mathcal{H}_j^n, \|f_0 + \cdots + f_p\|_2 \leq 1\} = \rho + o_p(1),$$

where $\mathcal{H}_j^n$ is defined as $\mathcal{H}_j$ but with the restriction $\int m_j(x_j)\ f_j(x_j)\ dx_j = 0$ replaced by $\int m_j(x_j) \widehat{f}_j(x_j)\ dx_j = 0$ for $1 \leq j \leq p$ and with $\sum_{j=1}^{T} d_t m_0(t) = 0$ replaced by $\sum_{j=1}^{T} N_t m_0(t) = 0$ for $j = 0$. (Compare Lemma 2 in Mammen et al., 1999.) In particular, there exists a ρ' with $\rho' < 1$ such that

$$\widehat{\rho} < \rho' \tag{7.3}$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

with probability tending to one. By iterative application of (7.1) and (7.2) (or, equivalently, of (2.4) and (2.5)) we get that (with probability tending to one)

$$\widehat{m} = \sum_{a=0}^{\infty} \widehat{S}^a \widehat{\tau}$$

with $\widehat{\tau} = \widehat{\psi}_0 \widehat{\psi}_p \dots \widehat{\psi}_2 \widetilde{m}_1 + \dots + \widehat{\psi}_0 \widetilde{m}_p + \widetilde{m}_0$, where (in abuse of notation) $\widetilde{m}_j(x,t) = \widetilde{m}_j(x_j)$ and $\widetilde{m}_0(x,t) = \widetilde{\eta}_t$; see also Lemma 3 in Mammen et al. (1999).

For the proof of Theorem 1 one now uses that $\widehat{m}^{[a+1]} = \widehat{S}\widehat{m}^{[a]} + \widehat{\tau}$ and therefore $\widehat{m}^{[a+1]} = \sum_{s=0}^{a} \widehat{S}^s \widehat{\tau} + \widehat{S}^{a+1} \widetilde{m}^{[0]}$. Because of (7.3) this shows Theorem 1. For the proof of Theorem 2 one decomposes $\widetilde{m}$ into $\widetilde{m} = \widetilde{m}^A + \widetilde{m}^B$, where $\widetilde{m}^A$ was defined in (2.7). Now put $\widehat{\tau}^j = \widehat{\psi}_0 \widehat{\psi}_p \dots \widehat{\psi}_2 \widetilde{m}_1^j + \dots + \widetilde{m}_0^j$ (for $j = A, B$). Then we have $\widehat{m} = \widehat{m}^A + \widehat{m}^B$ with $\widehat{m}^j = \sum_{a=0}^{\infty} \widehat{S}^a \widehat{\tau}^j$ (for $j = A, B$). The term $\widehat{m}^A$ is now analyzed as in Lemma 4 in Mammen et al. (1999). The basic fact that was used there was that an integral operator that is applied to a *local* average of ε^i (as is $\widetilde{m}^A$) leads to a *global* average of ε^i. A global average is of order $n^{-1/2}$ and therefore of lower order than $n^{-2/5}$. We now give a more detailed discussion of this argument. The basic step is to show that for $1 \le j, k \le p$

$$\sup_{0 \le x_k \le 1} \left| \int_0^1 \frac{\widehat{f}_{j,k}(x_j, x_k)}{\widehat{f}_k(x_k)} \widetilde{m}_j^A(x_j)\, dx_j \right| = o_P(n^{-2/5}). \tag{7.4}$$

Similarly as in Lemma 4 in Mammen et al. (1999) equation (7.4) implies that also iterative applications of the integral operator are of order $o_P(n^{-2/5})$. This gives that

$$\widehat{m}_j^A(x_j) - \widetilde{m}_j^A(x_j) = o_P(n^{-2/5}). \tag{7.5}$$

For (7.5) it remains to check (7.4). For the proof we use that

$$\widetilde{m}_j^A(x_j) = \sum_{t=1}^{T} \widetilde{r}_{j,t}^A(x_j) \Big/ \widehat{f}_j(x_j),$$

where

$$\widetilde{r}_{j,t}^A(x_j) = N^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) \varepsilon^{it}. \tag{7.6}$$

Using Assumption (A1) we get that

$$\sup_{0 \le x_j \le 1} \left| \widetilde{r}_{j,t}^A(x_j) \right| = O_P(\log(n)^{1/2} n^{-2/5}).$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

By application of (A1) and standard smoothing theory we get that

$$\sup_{0\le x_k\le 1}\left|\int_0^1 \frac{\widehat{f}_{j,k}(x_j,x_k)}{\widehat{f}_k(x_k)}\widetilde{m}_j^A(x_j)\,dx_j - \sum_{t=1}^T \int_0^1 \frac{\mathrm{E}[\widehat{f}_{j,k}(x_j,x_k)]}{\mathrm{E}[\widehat{f}_j(x_j)]\mathrm{E}[\widehat{f}_k(x_k)]}\widetilde{r}_{j,t}^A(x_j)\,dx_j\right|$$

$$= o_P(n^{-2/5}).$$

Claim (7.4) now follows from

$$\int_0^1 \frac{\mathrm{E}[\widehat{f}_{j,k}(x_j,x_k)]}{\mathrm{E}[\widehat{f}_j(x_j)]\mathrm{E}[\widehat{f}_k(x_k)]}\widetilde{r}_{j,t}^A(x_j)\,dx_j$$

$$= N^{-1}\sum_{i=1}^n \mathbf{1}(X^{it}\in[0,1]^p)g_h(X_j^{it})\varepsilon^{it} = o_P(n^{-2/5}),$$

where

$$g_h(u_j) = \int_0^1 \frac{\mathrm{E}[\widehat{f}_{j,k}(x_j,x_k)]}{\mathrm{E}[\widehat{f}_j(x_j)]\mathrm{E}[\widehat{f}_k(x_k)]}K_h(x_j,u_j)\,dx_j.$$

For the statement of Theorem 1 it remains to prove that

$$\sup_{x_j\in[h_j,1-h_j]}\left|\widehat{m}_j^B(x_j) - n^{-2/5}\beta_j(x_j)\right| = o_p(n^{-2/5}). \tag{7.7}$$

For a proof of (7.7) one can apply a generalization of Theorem 3 in Mammen et al. (1999) to our model. As in the application of Theorem 3 in Theorem 4 in Mammen et al. (1999) the generalized version of Theorem 3 could be used with $\alpha_{n,j}(x_j) = m_j(x_j) + m_j'(x_j)\int K_{h(j)}(x_j,u)(u-x_j)\,du\,[\int K_{h(j)}(x_j,v)\,dv]^{-1}$. The conditions of the generalized version of Theorem 3 can be verified as in the proof of Theorem 4 in Mammen et al. (1999). Note that in this theorem an incorrect value was given for $\gamma_{n,j}$. This was also remarked in Mammen and Park (2005), where a typo appears in the definition of $\gamma_{n,j}$. Compare our definition in the statement of Theorem 2 with the definition of $\gamma_{n,j}$ at the end of the statement of Theorem 6.1 of Mammen and Park (2005), where erroneously an additional factor $\frac{1}{2}$ appears.

Proof of Corollary 1. It can be easily checked that the conditional distribution of $\widetilde{r}_{j,t}^A(x_j)/\widehat{f}_j(x_j)$ (see (7.6)), given $(X_j^{is} : 1\le j\le p, 1\le s\le t, 1\le i\le n, \varepsilon^{is} : 1\le s\le t-1, 1\le i\le n)$, has a limiting distribution that is nonrandom and that therefore does not depend on the conditioning set. This implies that $\widetilde{m}_j^A(x_j) = \sum_{t=1}^T \widetilde{r}_{j,t}^A(x_j)/\widehat{f}_j(x_j)$ converges in distribution to the convolution of these limiting distributions. The corollary then follows by application of (7.7).

Proof of Theorem 3. For the proof of the theorem we apply Theorem 5 with $H_n = H$, with $\Delta_n = (dn)^{-2/5}$, $\delta_n = n^{-3/10(1+\delta)}d^{1/5}$, and with the linear space

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$\mathcal{M}_n$ of tuples of functions $(m_1, \ldots, m_p)^T$ with $m_j : [0,1] \to \mathbb{R}$ and $\int_0^1 m_j(u) f_j(u)\, du = 0$ for $j = 1, \ldots, p$. The constant δ was introduced in Assumption (A2′). Note that d and T are of the same order because of (A3′). Thus, (A2′) implies that $\delta_n \to 0$. We endow the space $\mathcal{M}_n$ with the pseudo norms $\|m\|_2^2 = \int_{[0,1]^p} \|m(x)\|^2 f(x)\, dx$ and $\|m\|_\infty = \sup_{x \in [0,1]^p} \|m(x)\|$. The space $\mathcal{M}_n$ depends on the density f that is not known. In the construction of $\widehat{m}_n$ the density f is used to apply the standardization $\int \widehat{m}_{n,j}(x_j) f_j(x_j)\, dx_j = 0$. We use the standardization (2.3) where f_j is replaced by an estimate. This results in fits of the components of m_n that differ by additive constants from the theoretical estimator for which the expansion is stated in Theorem 5.

We now check the conditions of Theorem 5. Condition (B1) follows directly from our uniform smoothness conditions (A3′). The inequality in (A6) implies that for a constant $c > 0$

$$\sum_{t=1}^{T} \frac{d_t}{d} \mathrm{E}[\{m_1(X_1^{it}) + \cdots + m_p(X_p^{it}) - \mathrm{E}[m_1(X_1^{it}) + \cdots + m_p(X_p^{it})]\}^2] > c$$

for all functions $m_1, \ldots, m_p$ with $\int m_1^2(x_1) f_1(x_1)\, dx_1 + \cdots + \int m_p^2(x_p) f_p(x_p)\, dx_p = 1$ and $\int m_1(x_1) f_1(x_1)\, dx_1 = \cdots = \int m_p(x_p) f_p(x_p)\, dx_p = 0$. This can be rewritten as

$$1 - \langle m, Hm \rangle > c.$$

Here for two elements $m, r \in \mathcal{M}_n$ the scalar product $\langle m, r \rangle$ is defined as $\langle m, r \rangle = \int_{[0,1]^p} \langle m(x), r(x) \rangle f(x)\, dx$. This inequality implies that in $\mathcal{M}_n$ all eigenvalues of H are bounded away from one. This shows (B2). For the check of (B3) one uses expansions of $\widehat{f}_j^{\,t}(x_j)$ and $\widehat{f}_{l,j}^{\,t}(x_l, x_j)$ that hold uniformly over $1 \le t \le T$, $1 \le l, j \le p$, and $x_j, x_l \in [0,1]$. The rates of convergence of these expansions carry over to an expansion of the averages $\widehat{f}_j(x_j)$ and $\widehat{f}_{l,j}(x_l, x_j)$. Under our assumptions on the growth of T (see (A2′)) this implies (B3).

For a check of (B4) we make use of the following decomposition:

$$\widehat{m}_j^*(x_j) + (\widehat{H}m)_j(x_j) - m_j(x_j) = T_1 + \cdots + T_4 \tag{7.8}$$

with

$$T_1 = \widetilde{m}_j^{*,A}(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})(\varepsilon^{it} - \varepsilon^{\cdot t}) / \widehat{f}_j(x_j),$$

$$T_2 = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})[m_j(X_j^{it}) - m_j(x_j)] / \widehat{f}_j(x_j),$$

$$T_3 = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$$\cdot \left[\sum_{l \neq j} m_l(X_l^{it}) - \sum_{l \neq j} \int m_l(y_l) K_h(y_l, X_l^{it})\, dy_l \right] \Big/ \widehat{f}_j(x_j),$$

$$T_4 = -N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})$$

$$\cdot \left[\frac{1}{N_t} \sum_{k=1}^{n} \sum_{l=1}^{p} m_l(X_l^{kt}) - \frac{1}{N_t} \sum_{k=1}^{n} \sum_{l=1}^{p} \int K_h(y_l, X_l^{kt}) m_l(y_l)\, dy_l \right] \Big/ \widehat{f}_j(x_j).$$

First note that

$$T_1 \asymp \widetilde{m}_j^A(x_j), \tag{7.9}$$

where we use the notation $\mu_n \asymp \nu_n$ for two random functions μ_n and ν_n if $\sup_x |\mu_n(x) - \nu_n(x)| = o_P((Tnh)^{-1/2})$. Claim (7.9) follows from

$$N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) \varepsilon^{\cdot t} / \widehat{f}_j(x_j) = \sum_{t=1}^{T} \frac{N_t}{N} \frac{\widehat{f}_j^{\,t}(x_j)}{\widehat{f}_j(x_j)} \varepsilon^{\cdot t} \asymp 0. \tag{7.10}$$

The expansion (7.10) follows by application of an exponential inequality for martingales.

Note now that by Taylor expansion

$$T_2 \asymp m_j'(x_j) \frac{\widehat{f}_j^*(x_j)}{\widehat{f}_j(x_j)} + \frac{1}{2} m_j''(x_j) \frac{\widehat{f}_j^{**}(x_j)}{\widehat{f}_j(x_j)},$$

$$T_3 \asymp \sum_{l \neq j} \int m_l'(y_l) \frac{\widehat{f}_{l,j}^*(y_l, x_j)}{\widehat{f}_j(x_j)}\, dy_l + \frac{1}{2} \int m_l''(y_l) \frac{\widehat{f}_{l,j}^{**}(y_l, x_j)}{\widehat{f}_j(x_j)}\, dy_l,$$

$$T_4 \asymp -\sum_{t=1}^{T} \frac{N_t}{N} \frac{\widehat{f}_j^{\,t}(x_j)}{\widehat{f}_j(x_j)} \sum_{l=1}^{p} \left[\int \widehat{f}_l^{\,t,*}(y_l) m_l'(y_l)\, dy_l + \frac{1}{2} \int \widehat{f}_l^{\,t,**}(y_l) m_l''(y_l)\, dy_l \right],$$

where

$$\widehat{f}_j^*(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})(X_j^{it} - x_j),$$

$$\widehat{f}_j^{**}(x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it})(X_j^{it} - x_j)^2,$$

$$\widehat{f}_{l,j}^*(x_l, x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) K_h(x_l, X_l^{it})(X_l^{it} - x_l),$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$$\widehat{f}^{**}_{l,j}(x_l, x_j) = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) K_h(x_l, X_l^{it}) (X_l^{it} - x_l)^2,$$

$$\widehat{f}^{t,*}_{j}(x_j) = N^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) (X_j^{it} - x_j),$$

$$\widehat{f}^{t,**}_{j}(x_j) = N^{-1} \sum_{i=1}^{n} \mathbf{1}(X^{it} \in [0,1]^p) K_h(x_j, X_j^{it}) (X_j^{it} - x_j)^2.$$

We now use that these random functions are asymptotically equivalent to its expectations. This gives

$$T_2 \asymp m_j'(x_j)\alpha_j(x_j) + h^2 c_K \left[\frac{f_j'(x_j) m_j'(x_j)}{f_j(x_j)} + \frac{1}{2} m_j''(x_j) \right],$$

$$T_3 \asymp \sum_{l \neq j} \int m_l'(y_l) \alpha_l(y_l) \frac{\widehat{f}_{l,j}(y_l, x_j)}{\widehat{f}_j(x_j)} \, dy_l$$

$$+ h^2 c_K \int \frac{f_{l,j}(y_l, x_j)}{f_j(x_j)} \left[\frac{\frac{\partial}{\partial y_l} f_{l,j}(y_l, x_j) m_l'(y_l)}{f_{l,j}(y_l, x_j)} + \frac{1}{2} m_l''(y_l) \right] dy_l,$$

$$T_4 \asymp - \sum_{t=1}^{T} \sum_{l=1}^{p} \frac{N_t}{N} \int \frac{\widehat{f}_j^t(x_j) \widehat{f}_l^t(y_l)}{\widehat{f}_j(x_j)} m_l'(y_l) \alpha_l(y_l) \, dy_l$$

$$+ h^2 c_K \sum_{t=1}^{T} \sum_{l=1}^{p} \frac{N_t}{N} \int \frac{f_j^t(x_j)}{f_j(x_j)} \left[\frac{\partial}{\partial y_l} f_l^t(y_l) m_l'(y_l) + \frac{1}{2} f_l^t(y_l) m_l''(y_l) \right] dy_l,$$

where

$$\alpha_j(x_j) = \frac{\int K_h(x_j, u)(u - x_j) \, du}{\int K_h(x_j, u) \, du}.$$

Now introduce the operator H^* by

$$[(I - H^*)\mu]_j(x_j) = \int \sum_{t=1}^{T} \frac{d_t}{d} \frac{f^t(x)}{f_j(x_j)} \mu(x,t) \, dx_{-j}$$

$$- \sum_{t=1}^{T} \frac{d_t}{d} \frac{f_j^t(x_j)}{f_j(x_j)} \int \mu(y,t) f^t(y) \, dy,$$

where $x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p)$ and where μ is a function mapping $[0,1]^p \times \{1, \ldots, T\}$ to $\mathbb{R}$. With this notation we can rewrite our last result as

$$T_2 + T_3 + T_4 \asymp [(I - \widehat{H})\mu)_j(x_j) + [(I - H^*)\mu^*]_j(x_j)$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

with $\mu_j(x_j) = m_j'(x_j)\alpha_j(x_j)$ and

$$\mu^*(x,t) = c_K h^2 \sum_{l=1}^{p} \left[\frac{\frac{\partial}{\partial x_l} f^t(x) m_l'(x_l)}{f^t(x)} + \frac{1}{2} m_l''(x_l) \right].$$

We apply Theorem 5 with $\widehat{m}_n^A = \widetilde{m}_j^A - \mathrm{E}_j[m_j^A]$, $\widehat{m}_n^B = (I - H^*)\mu^* - \mathrm{E}_j[(I - H^*)\mu^*]$, and $\widehat{m}_n^C = (I - \widehat{H})\mu - \mathrm{E}_j[(I - \widehat{H})\mu]$. Here we use the notation $\mathrm{E}_j[g] = \int_0^1 g(u) f_j(u)\, du$ for a function $g : [0,1] \to \mathbb{R}$. By application of Theorem 5 we get with (random) constants C_j that

$$\sup_{0 \le x_j \le 1} \Big| \widehat{m}_j(x_j) - m_j(x_j) - \widetilde{m}_j^A(x_j) - [(I - H)^{-1}(I - H^*)\mu^*]_j(x_j) - \mu_j(x_j) - C_j \Big| = o_P(\Delta_n). \tag{7.11}$$

Here, for a function r the function $g = (I - H)^{-1} r$ is defined as the solution of $(I - H)g = r$ with $\int g_j(x_j) f_j(x_j)\, dx_j = 0$. We now show that

$$C_j = \gamma_{n,j} + o_P(\Delta_n). \tag{7.12}$$

Because of (2.3) this follows from

$$\int_0^1 [m_j(x_j) + \widetilde{m}_j^A(x_j) + [(I - H)^{-1}(I - H^*)\mu^*]_j(x_j) + \mu_j(x_j)]\widehat{f}_j(x_j)\, dx_j = \gamma_{n,j} + o_P(\Delta_n). \tag{7.13}$$

For the proof of (7.13) note first that by standard smoothing theory

$$\int_0^1 \widetilde{m}_j^A(x_j)\widehat{f}_j(x_j)\, dx_j = o_P(\Delta_n),$$

$$\int_0^1 [(I - H)^{-1}(I - H^*)\mu^*]_j(x_j)\widehat{f}_j(x_j)\, dx_j = o_P(\Delta_n).$$

Thus for (7.13) it remains to show that

$$\int_0^1 \big[m_j(x_j) + \mu_j(x_j)\big] f_j(x_j)\, dx_j = \gamma_{n,j} + o_P(\Delta_n). \tag{7.14}$$

Claim (7.14) follows from

$$\int_0^1 [m_j(x_j) + \mu_j(x_j)]\widehat{f}_j(x_j)\, dx_j = \int_0^1 \int_0^1 m_j(u) K_h(u,v) f_j(v)\, du\, dv$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

$$
\begin{aligned}
&+\int_0^1\int_0^1 m_j'(u)\alpha_j(u)K_h(u,v)f_j(v)\,du\,dv+o_P(\Delta_n)\\
&=\int_0^1\int_0^1 f_j(v)\left[m_j(v)+m_j'(v)(u-v)+\frac{1}{2}m_j''(v)(u-v)^2\right]K_h(u,v)\,du\,dv\\
&\quad+\int_0^1\int_0^1\int_0^1 m_j'(u)K_h(u,w)(w-u)K_h(u,v)f_j(v)\\
&\qquad\cdot\left[\int_0^1 K_h(u,z)\,dz\right]^{-1}du\,dv\,dw+o_P(\Delta_n)\\
&=\int_0^1 f_j(v)m_j(v)\,dv+\int_0^1 m_j'(v)(u-v)K_h(u,v)f_j(v)\,du\,dv\\
&\quad+\frac{1}{2}h_j^2\int u^2K(u)\,du\int_0^1 m_j''(v)f_j(v)\,dv\\
&\quad+\int_0^1\int_0^1\int_0^1 m_j'(u)K_h(u,w)(w-u)K_h(u,v)f_j(u)\\
&\qquad\cdot\left[\int_0^1 K_h(u,z)\,dz\right]^{-1}du\,dv\,dw+o_P(\Delta_n)\\
&=\int_0^1 m_j'(v)K_h(u,v)(u-v)f_j(u)\,du\,dv\\
&\quad+\frac{1}{2}h_j^2\int u^2K(u)du\int_0^1 m_j''(v)f_j(v)\,dv\\
&\quad+\int_0^1\int_0^1 m_j'(u)K_h(u,w)(w-u)f_j(u)\,du\,dw+o_P(\Delta_n)\\
&=-h_j^2\int u^2K(u)\,du\int_0^1\left[m_j'(v)f_j'(v)+\frac{1}{2}m_j''(v)f_j(v)\right]dv+o_P(\Delta_n)\\
&=\gamma_{n,j}+o_P(\Delta_n).
\end{aligned}
$$

Because μ_j vanishes at the boundary we get from (7.11) and (7.12)

$$
\begin{aligned}
&\sup_{h_j\le x_j\le 1-h_j}\left|\widehat{m}_j(x_j)-m_j(x_j)-\widetilde{m}_j^A(x_j)-[(I-H)^{-1}(I-H^*)\mu^*]_j(x_j)-\gamma_{n,j}\right|\\
&\quad=o_P(\Delta_n).
\end{aligned}
$$

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

The expansion of the theorem now follows from the fact that a tuple $g = (g_1, \dots, g_p)$ of functions $g_1, \dots, g_p$ minimizes the criterion

$$\min_{\eta_1,\dots,\eta_T \in \mathbb{R}} \sum_{t=1}^{T} \frac{d_t}{d} \int [g_1(x_1) + \cdots + g_p(x_p) + \eta_t - \mu^*(x,t)]^2 f^t(x)\, dx$$

if and only if

$$g = (I - H^*)\mu^* + Hg$$

or equivalently

$$g = (I - H)^{-1}(I - H^*)\mu^*.$$

Proof of Theorem 4. The proof follows along the same lines of the proofs of Theorems 1 and 2. Compare also the proofs of Theorems 1′–4′ in Mammen et al. (1999).

Proof of Theorem 5. By Assumption (B4) we have that

$$(I - \widehat{H}_n)(\widehat{m}_n - m_n) = \widehat{m}_n^* - (I - \widehat{H}_n)m_n = \widehat{m}_n^A + \widehat{m}_n^B + \widehat{m}_n^C + \widehat{m}_n^D.$$

Thus, for the theorem it suffices to show that

$$\|(I - \widehat{H}_n)^{-1}\widehat{m}_n^A - \widehat{m}_n^A\|_\infty = o_P(\Delta_n), \tag{7.15}$$

$$\|(I - \widehat{H}_n)^{-1}\widehat{m}_n^B - (I - H_n)^{-1}\widehat{m}_n^B\|_\infty = o_P(\Delta_n), \tag{7.16}$$

$$\|(I - \widehat{H}_n)^{-1}\widehat{m}_n^D\|_\infty = o_P(\Delta_n). \tag{7.17}$$

For the proof of (7.15) it is sufficient to show

$$\|(I - \widehat{H}_n)^{-1}\widehat{m}_n^A - (I - H_n)^{-1}\widehat{m}_n^A\|_\infty = o_P(\Delta_n), \tag{7.18}$$

$$\|(I - H_n)^{-1}\widehat{m}_n^A - \widehat{m}_n^A\|_\infty = o_P(\Delta_n). \tag{7.19}$$

Claim (7.18) follows from (6.1), (B1), (B2), and

$$(I - \widehat{H}_n)^{-1} - (I - H_n)^{-1} = (I - H_n)^{-1}\{[I - (\widehat{H}_n - H_n)(I - H_n)^{-1}]^{-1} - I\}$$

$$= [I + H_n(I - H_n)^{-1}] \sum_{j=1}^{\infty} [(\widehat{H}_n - H_n)(I - H_n)^{-1}]^j. \tag{7.20}$$

Claim (7.19) follows from

$$(I - H_n)^{-1} - I = H_n(I - H_n)^{-1}.$$

For the proof of (7.16) one uses (7.20), (B1), (B2), and $\|\widehat{m}_n^B\|_2 = O_P(\Delta_n \delta_n^{-1})$.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

For the proof of (7.17) one uses $(I - \widehat{H}_n)^{-1}\widehat{m}_n^C = [(I - \widehat{H}_n)^{-1}\widehat{m}_n^C - (I - H_n)^{-1}\widehat{m}_n^C] + [(I - H_n)^{-1}\widehat{m}_n^C - \widehat{m}_n^C] + \widehat{m}_n^C$. Now, one can proceed as in the proof of (7.16) to show that the first term has a sup norm of order $o_P(\Delta_n)$. The second term has a sup norm of order $o_P(\Delta_n)$ because of $(I - H_n)^{-1} - I = H_n(I - H_n)^{-1}$, (B1), (B2), and $\|\widehat{m}_n^C\|_2 \leq \|\widehat{m}_n^C\|_\infty = o_P(\Delta_n)$. Furthermore the third term has a sup norm of order $o_P(\Delta_n)$ by assumption.

REFERENCES

Arellano, M. (2003) *Panel Data Econometrics*. Oxford University Press.

Baltagi, B.H. (1995) *Econometric Analysis of Panel Data*. Wiley.

Baltagi, B.H., J.M. Griffin, & W. Xiong (2000) To pool or not to pool: Homogenous versus heterogenous estimators applied to cigarette demand. *Review of Economics and Statistics* 82, 117–126.

Cai, Z. & Q. Li (2008) Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24, 1321–1342.

Chang, Y. (2004) Bootstrap unit root test in panels with cross-sectional dependency. *Journal of Econometrics* 120, 263–293.

Fan, J. & R. Li, (2004) New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association* 99, 710–723.

Fan, J. & Q. Yao, (2003) *Nonlinear Time Series*. Springer-Verlag.

Ferraty, F. & P. Vieu, (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. Springer-Verlag.

Fu, B., W.-K. Li, & W.-K. Fung (2002) Testing model adequacy for dynamic panel data with intercorrelation. *Biometrika* 89, 591–601.

Härdle, W. (1990) *Smoothing Techniques: With Implementation in S*. Springer-Verlag.

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

Henderson, D.J, R.J. Carroll, & Q. Li (2008) Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144, 257–275.

Hjellvik, V., R., Chen & D. Tjøstheim (2004) Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis* 25, 831–872.

Hjellvik, V. & D. Tjøstheim (1999) Modelling panels of intercorrelated autoregressive time series. *Biometrika* 86, 573–590.

Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge University Press.

Linton, O. & E. Mammen (2005) Estimating semiparametric ARCH (∞) models by kernel smoothing methods. *Econometrica* 73, 771–836.

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

Mammen, E., O. Linton, & J.P. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443–1490.

Mammen, E. & B. Park (2005) Bandwidth selection for smooth backfitting in additive models. *Annals of Statistics* 33, 1260–1294.

Mammen, E. & B. Park (2006) A simple smooth backfitting method for additive models. *Annals of Statistics* 34, 2252–2271.

Mátyás, L. & Sevestre, P. (1992) *The Econometrics of Panel Data*. Kluwer Academic.

Newey, W.K. (1994) Kernel estimation of partial means. *Econometric Theory* 10, 233–253.

Nielsen, J.P. & Sperlich, S. (2005) Smooth backfitting in practice. *Journal of the Royal Statistical Society, Series B* 67, 43–61.

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

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press

Profit, S. and S. Sperlich (2004) Non-uniformity of job-matching in a transition economy—A nonparametric analysis for the Czech Republic. *Applied Economics* 36, 695–714.

Ramsay, J.O. and B.W. Silverman (1996) *Functional Data Analysis*. Springer-Verlag.

Sperlich, S., D. Tjøstheim & L. Yang (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* 18, 197–251.

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

Wooldridge, J. (2005a) Fixed effects and related estimators for correlated random-coefficient and treatment effect panel data models. *Review of Economics and Statistics* 87, 385–390.

Wooldridge, J. (2005b) Simple solutions to the initial conditions problem for dynamic nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20, 39–54.

https://doi.org/10.1017/S0266466608090142 Published online by Cambridge University Press