Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates *

Tina Felber[1], Daniel Jones[1,†], Michael Kohler[1] and Harro Walk[2]

[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: tfelber@mathematik.tu-darmstadt.de, jones@mathematik.tu-darmstadt.de, kohler@mathematik.tu-darmstadt.de*

[2] *Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, email: walk@mathematik.uni-stuttgart.de.*

September 6, 2011

Abstract

Given a stationary and ergodic time series the problem of estimating the conditional expectation of the dependent variable at time zero given the infinite past is considered. It is shown that the mean squared error of a combination of suitably defined local averaging or least squares estimates converges to zero for all distributions whenever the dependent variable is square integrable.

AMS classification: Primary 62G05; secondary 62G20.

Key words and phrases: dependent data, static forecasting, mean squared error, time series, weak consistency.

1 Introduction

Let $((X_n, Y_n))_{n\in\mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\left\{Y_0^2\right\} < \infty$. In this paper we consider the following static forecasting problem: Given the dataset

$$\mathcal{D}_{-n}^{-1} = \{(X_{-n}, Y_{-n}), (X_{-n+1}, Y_{-n+1}), \dots, (X_{-1}, Y_{-1})\}$$

and X_0, construct estimates $m_n(X_0, \mathcal{D}_{-n}^{-1})$ of

$$\mathbf{E}\left\{Y_0 | X_{-\infty}^0, Y_{-\infty}^{-1}\right\} := \mathbf{E}\left\{Y_0 | X_0, (X_{-1}, Y_{-1}), (X_{-2}, Y_{-2}), \dots\right\}$$

which are weakly consistent in the sense that they satisfy

$$\mathbf{E}\left\{\left|m_n(X_0, \mathcal{D}_{-n}^{-1}) - \mathbf{E}\left\{Y_0 | X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right|^2\right\} \to 0 \quad (n \to \infty). \tag{1}$$

*Running title: *Weakly universally consistent forecasting*

†Corresponding author. Tel: +49-6151-16-2294, Fax: +49-6151-16-6822

In the sequel estimates m_n are constructed which are weakly universally consistent, i.e., which satisfy (1) for **all** stationary and ergodic time series with $\mathbf{E}\{Y_0^2\} < \infty$. The main aim is to construct the estimates in such a way that they are easy to compute and thus applicable in practice.

In the existing literature there are several results concerning consistency in various senses for nonparametric estimators under more or less strong mixing assumptions on the data [e.g., Collomb (1985), Marton and Shields (1994)]. The monograph by Györfi, Haerdle, Sarda and Vieu (1989) gives additional information. A big drawback of all these estimates is that mixing conditions are not verifiable in practice what makes them difficult to apply. Therefore it is desirable to construct estimates which are consistent in the above sense under considerably weaker conditions (e.g., stationarity and ergodicity of the data).

Several authors have investigated forecasting problems with stationarity and ergodicity as sole assumption on the data, usually in the context of autoregression (where there is no X_i only Y_i). Cover (1975) formulated two fundamental classes of prediction problems for stationary and ergodic time series, the static and the dynamic forecasting problem.

In the dynamic forecasting problem, specialized to autoregression, one wants to find an estimator $\hat{E}(Y_0^{n-1})$ of the value $\mathbf{E}\left\{Y_n \mid Y_0^{n-1}\right\}$ such that for all stationary and ergodic sequences $\{Y_i\}_{i\in\mathbb{N}}$ it holds

$$\lim_{n\to\infty} \left|\hat{E}(Y_0^{n-1}) - \mathbf{E}\left\{Y_n \mid Y_0^{n-1}\right\}\right| = 0 \qquad a.s.$$

Bailey (1976) showed that the dynamic forecasting problem cannot be solved in general, a much simpler proof of this result was given in Ryabko (1988). For related results see also Györfi, Morvai and Yakowitz (1998) and Weiss (2000). In case that the ergodic time series takes values from a finite alphabet Morvai and Weiss (2005) proposed a very simple estimator which is pointwise and weakly consistent for all such stationary and ergodic time series.

In the so-called static forecasting problem, specialized to autoregression, one wants to find an estimator $\hat{E}(Y_{-n}^{-1})$ of the value $\mathbf{E}\left\{Y_0 \mid Y_{-\infty}^{-1}\right\}$ such that for all stationary and ergodic sequences $\{Y_i\}_{i\in\mathbb{Z}}$ it holds

$$\lim_{n\to\infty} \hat{E}(Y_{-n}^{-1}) = \mathbf{E}\left\{Y_0 \mid Y_{-\infty}^{-1}\right\} \qquad a.s.$$

Ornstein (1978) gave such a strongly consistent estimator for the case that the time series values are finite. The algorithms of Algoet (1992) and the simpler algorithm of Morvai, Yakowitz and Györfi (1996) yield strong consistency for bounded stationary and ergodic sequences. For more results in view of static and dynamic forecasting and concerning connections and related problems we refer to Morvai, Yakowitz and Algoet (1997), Györfi, Lugosi and Morvai (1999), Györfi and Lugosi (2000), Györfi and Ottucsák (2007) and Morvai and Weiss (2011).

However, the estimates above are rather difficult to compute or consume data rapidly so that it is not clear whether they can be applied to any real or simulated data set. Moreover, it is known that a simple partitioning estimate which is strongly universally consistent in the case of mixing assumptions fails to be consistent when the data is only stationary and ergodic (cf., Györfi, Morvai and Yakowitz (1998)). Thus it is reasonable to conjecture that in case of stationary and ergodic data strongly universally consistent estimates, which are easy to compute, do not exist.

An application of weakly universally consistent estimates in mathematical finance is given in Kohler and Walk (2010). There the problem of exercising an American option in an optimal way is considered. An estimate of the optimal stopping time is constructed which achieves for sample size tending to infinity the optimal expected discounted payoff whenever the returns of the underlying asset are stationary and ergodic.

In Jones, Kohler and Walk (2011) it was shown that by applying the ideas of Kohler and Walk (2010) it is possible to construct weakly universally consistent localized least squares estimates. There techniques from the theory of prediction of individual sequences (cf., e.g., Cesa-Bianchi and Lugosi (2006)) are used to choose the smoothing parameters of the estimates. The same ideas have already been applied to sequential prediction problems where estimates are constructed which are universally consistent with respect to a normalized cumulative prediction error, see Györfi and Lugosi (2002), Györfi and Ottucsák (2007) and (in connection with portfolio optimization) Györfi and Schäfer (2003), Györfi, Lugosi and Udina (2006) as well as Györfi, Udina and Walk (2008).

The estimates in Jones, Kohler and Walk (2011) are easier to compute then the above mentioned strongly universally consistent estimates, in particular it is obvious that they can be applied to real or simulated data sets in practice. Nevertheless, through the fact that they combine ideas of local averaging and least squares it seems still to be a little bit challenging to apply them in practice. In this paper we focus on much simpler local averaging and least squares estimates and construct four estimates which are weakly universally consistent. They are based on simple kernel, partitioning and nearest neighbor regression estimates as well as least squares estimates. Again techniques from the theory of prediction of individual sequences are used to choose the smoothing parameters of the estimates, and a suitable averaging of the estimates is used to be able to ensure that (1) holds.

The estimates are defined in Section 2, Section 3 contains the main result, and the proofs are given in Section 4.

2 Definition of the estimates

In this section we introduce four estimates which are weakly universally consistent. They mainly distinguish in the choice of the so-called expert, which will be a kernel estimate for the first one, a partitioning estimate for the second one, a nearest neighbor estimate for the third one and a least squares estimate for the last one.

In order to simplify the notation we will use throughout this paper the abbreviations

$$Z_k^l = (Z_k, \cdots, Z_l), \qquad k \leq l$$

for arbitrary random variables $Z_j \quad (j \in \mathbb{Z})$,

$$\mathcal{D}_k^l = \{(X_k, Y_k), (X_{k+1}, Y_{k+1}), \ldots, (X_l, Y_l)\}, \qquad k \leq l,$$

and

$$d_k^l = \{x_k, y_k), (x_{k+1}, y_{k+1}), \ldots, (x_l, y_l)\}, \qquad k \le l.$$

$S_{x,r}$ denotes the closed sphere in a Euclidean space with center x and radius r.

We start with defining a parameter set

$$\mathcal{P} = \{(k, r, N) : k, r, N \in \mathbb{N}\}$$

and estimates $\tilde{m}^{(i)}_{n,(k,r,N)}(x^0_{-k}, y^{-1}_{-k}; d^{-1}_{-n})$ and $\hat{m}^{(i)}_{n,(k,r,N)}(x_0, d^{-1}_{-n})$ $(i \in \{1,2,3,4\})$ of

$$m_k(x^0_{-k}, y^{-1}_{-k}) := \mathbf{E}\left\{Y_0 \;\middle|\; X^0_{-k} = x^0_{-k}, Y^{-1}_{-k} = y^{-1}_{-k}\right\},$$

where the parameters of the estimates are determined by r and N, and where k indicates how far back the estimate will look. For the local averaging estimates, r will be used to determine how similar observations in the past to the current values have to be in order to be included in the estimation. For the least squares estimate, r determines the function space over which minimization takes place. N will be used to truncate the values of Y_i at level N.

For the first estimate assume that bandwidths $h_r > 0$ satisfying

$$h_r \to 0 \quad (r \to \infty)$$

and a nonincreasing and continuous function $H : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$H(0) > 0 \quad \text{and} \quad t \cdot H(t) \to 0 \quad (t \to \infty)$$

(e.g., $H(v) = e^{-v^2}$) are given. Using H we define a kernel function $K : \mathbb{R}^{(k+1)\cdot d+k} \to \mathbb{R}_+$ by

$$K(v) := H\left(\|v\|^{(k+1)\cdot d+k}\right),$$

where $\|v\|$ denotes the Euclidean norm of v.

With the convention $\frac{0}{0} := 0$ we define the kernel estimate $\tilde{m}^{(1)}_{n,(k,r,N)}$ by

$$\tilde{m}^{(1)}_{n,(k,r,N)}(u^0_{-k}, v^{-1}_{-k}; d^{-1}_{-n}) = \begin{cases} \dfrac{\sum_{i=-n+k-1}^{-2} T_N(y_{i+1}) K\left(\frac{(x^{i+1}_{i-k+1}, y^i_{i-k+1}) - (u^0_{-k}, v^{-1}_{-k})}{h_r}\right)}{\sum_{i=-n+k-1}^{-2} K\left(\frac{(x^{i+1}_{i-k+1}, y^i_{i-k+1}) - (u^0_{-k}, v^{-1}_{-k})}{h_r}\right)} & \text{if } n \ge k+1, \\ 0 & \text{else,} \end{cases}$$

where the truncation operator T_N is defined as

$$T_N(x) := x\ \mathbf{1}_{\{|x| \le N\}} + \operatorname{sign}(x)\ N\ \mathbf{1}_{\{|x| > N\}}.$$

For the second estimate, for every $k \in \mathbb{N}$ let $(\mathcal{P}_{k,r})_{r \in \mathbb{N}}$ be a sequence of finite or countably infinite partitions $\mathcal{P}_{k,r} = \{A_{k,r,1}, A_{k,r,2}, \cdots\}$ of $\left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$, where $A_{k,r,j} \subseteq$

$\left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$ are Borel sets. For $z \in \left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$ set $A_{k,r}(z) = A_{k,r,j}$ if $z \in A_{k,r,j}$. With the convention $\frac{0}{0} := 0$ we define our partitioning estimate $\tilde{m}^{(2)}_{n,(k,r,N)}$ as

$$\tilde{m}^{(2)}_{n,(k,r,N)}(u^0_{-k}, v^{-1}_{-k}; d^{-1}_{-n}) = \begin{cases} \dfrac{\sum_{i=-n+k-1}^{-2} T_N(y_{i+1}) \mathbf{1}_{\left\{(x^{i+1}_{i-k+1}, y^i_{i-k+1}) \in A_{k,r}\left(u^0_{-k}, v^{-1}_{-k}\right)\right\}}}{\sum_{i=-n+k-1}^{-2} \mathbf{1}_{\left\{(x^{i+1}_{i-k+1}, y^i_{i-k+1}) \in A_{k,r}\left(u^0_{-k}, v^{-1}_{-k}\right)\right\}}} & \text{if } n \geq k+1, \\ 0 & \text{else.} \end{cases}$$

For the third estimate let $p_r > 0$ be given such that

$$p_r \to 0 \quad (r \to \infty).$$

Let k, r be fixed and let n be so large that $n \geq k + \lfloor p_r n \rfloor \geq k+1$, where $\lfloor z \rfloor$ is the largest integer less than or equal to z. For each $\left(u^0_{-k}, v^{-1}_{-k}\right) \in \mathbb{R}^{(k+1)d+k}$ introduce the random set of $\lfloor p_r n \rfloor$ nearest neighbor (NN) matches

$$\begin{aligned} J^{(k,r)}_{n,(u^0_{-k}, v^{-1}_{-k})} := \Big\{ -n+k-1 \leq i \leq -2 \quad : \quad & \left(X^{i+1}_{i-k+1}, Y^i_{i-k+1}\right) \text{ is among the } \lfloor p_r n \rfloor \\ & \text{NNs of } \left(u^0_{-k}, v^{-1}_{-k}\right) \text{ in } \left(X^{-n+k}_{-n}, Y^{-n+k-1}_{-n}\right), \dots, \left(X^{-1}_{-k-1}, Y^{-2}_{-k-1}\right) \Big\} \end{aligned}$$

using the Euclidean norm $\|\cdot\|$ and define the $\lfloor p_r n \rfloor$−NN estimate $\tilde{m}^{(3)}_{n,(k,r,N)}$ via

$$\tilde{m}^{(3)}_{n,(k,r,N)}(u^0_{-k}, v^{-1}_{-k}; \mathcal{D}^{-1}_{-n}) = \begin{cases} \frac{1}{\left|J^{(k,r)}_{n,(u^0_{-k}, v^{-1}_{-k})}\right|} \sum_{i \in J^{(k,r)}_{n,(u^0_{-k}, v^{-1}_{-k})}} T_N(Y_{i+1}) & \text{if } n \geq k + \lfloor p_r n \rfloor \geq k+1 \\ 0 & \text{else} \end{cases}$$

($|\cdot|$ denoting cardinality). We use here the assumption $(A1)$ that, for each fixed k, for all vectors $(x^0_{-k}, y^{-1}_{-k}) \in \mathbb{R}^{(k+1)d+k}$ the random variables $\left\|\left(X^0_{-k}, Y^{-1}_{-k}\right) - \left(x^0_{-k}, y^{-1}_{-k}\right)\right\|$ have continuous distribution functions and that for each sphere S_{0,R^*} in $\mathbb{R}^{(k+1)\cdot d+k}$ the family of such distribution functions with "parameters" $(x^0_{-k}, y^{-1}_{-k}) \in S_{0,R^*}$ is equicontinuous, i.e., the distribution functions are uniformly continuous with respect to their argument $s \in \mathbb{R}$ and the "parameters" $(x^0_{-k}, y^{-1}_{-k}) \in S_{0,R^*}$. By the first part of $(A1)$ ties occur with probability zero. Thus

$$\left| J^{(k,r)}_{n,(u^0_{-k}, v^{-1}_{-k})} \right| = \lfloor p_r n \rfloor$$

with probability one. $(A1)$ can be assumed without loss of generality. Similarly to Györfi et al. (2002), pp. 86-87, we may add a component Z_i to the $(d+1)$−dimensional vector (X_i, Y_i), $i \in \mathbb{Z}$, where the Z_i's are i.i.d. uniform $[0,1]$ and also independent of the sequence $((X_n, Y_n))_{n \in \mathbb{Z}}$. In view of Theorem 1 below we mention that because of independence

$$\mathbf{E}\left(Y_0 \mid X^0_{-\infty}, Y^{-1}_{-\infty}, Z^0_{-\infty}\right) = \mathbf{E}\left(Y_0 \mid X^0_{-\infty}, Y^{-1}_{-\infty}\right) \quad \text{a.s.}$$

and that also the sequence $((X_i, Y_i, Z_i))_{i\in\mathbb{Z}}$ of $(d+2)-$dimensional random vectors is stationary and ergodic. For, because of Doob (1953), p. 457, for an independent pair $((U_i)_{i\in\mathbb{Z}}, (V_i)_{i\in\mathbb{Z}})$ of stationary and ergodic processes with state spaces $\prod_{-\infty}^{\infty} \mathbb{R}^{d'}$ and $\prod_{-\infty}^{\infty} \mathbb{R}^{d''}$, respectively, the stochastic process $((U_i, V_i))_{i\in\mathbb{Z}}$ with state space $\prod_{-\infty}^{\infty} \mathbb{R}^{d'+d''}$ is stationary and ergodic, too. In our construction, assumption $(A1)$ is fulfilled. To see this, observe that with abbreviations $\left(X_{-k}^0, Y_{-k}^{-1}\right) = W$, $Z_{-k}^0 = Z$, $\left(x_{-k}^0, y_{-k}^{-1}\right) = w \in S_{0,R^*}$, $z_{-k}^0 = z \in [0,1]^{k+1}$ and corresponding Euclidean norms $\|w\|$, $\|z\|_*$, the distribution functions are given by

$$F_{w,z}(s) \quad = \quad \mathbf{P}\left\{\|W-w\|^2 + \|Z-z\|_*^2 \le s^2\right\}, \quad s \ge 0.$$

Since

$$F_{w,z}(s) \ge \mathbf{P}\left\{\|W-w\|^2 \le s^2 - k - 1\right\} \to 1 \quad (s \to \infty)$$

uniformly in $w \in S_{0,R^*}$, it suffices to show that for arbitrary $s_{max} > 0$ the functions $F_{w,z}$ with parameters $w \in S_{0,R^*}$ and $z \in [0,1]^{k+1}$ are equicontinuous on $[0, s_{max}]$. Using

$$F_{w,z}(s) \quad = \quad \int \mathbf{P}\left\{Z \in S_{z,\sqrt{(s^2-\|t-w\|^2)_+}}\right\} \mathbf{P}_W(dt)$$

where $(a)_+ = \max\{a, 0\}$, we see that for $w \in S_{0,R^*}$, $z \in [0,1]^{k+1}$ and $s_1, s_2 \in [0, s_{max}]$ with $s_1 \le s_2$ we have

$$\begin{aligned}
&|F_{w,z}(s_1) - F_{w,z}(s_2)| = F_{w,z}(s_2) - F_{w,z}(s_1)\\
&\le \int_{S_{0,R^*}} \left(\mathbf{P}\left\{Z \in S_{z,\sqrt{(s_2^2-\|t-w\|^2)_+}}\right\} - \mathbf{P}\left\{Z \in S_{z,\sqrt{(s_1^2-\|t-w\|^2)_+}}\right\}\right) \mathbf{P}_W(dt)\\
&\quad + \mathbf{P}_W(S_{0,R^*}^c)\\
&\le \int_{S_{0,R^*}} c_{k+1} \cdot \left(\left(s_2^2 - \|t-w\|^2\right)_+^{\frac{k+1}{2}} - \left(s_1^2 - \|t-w\|^2\right)_+^{\frac{k+1}{2}}\right) \mathbf{P}_W(dt) + \mathbf{P}_W(S_{0,R^*}^c),
\end{aligned}$$

where c_{k+1} is the volume of the unit sphere in $\mathbb{R}^{k+1}$. By computing derivatives it is easy to see that the functions

$$s \mapsto \left(s^2 - \|t-w\|^2\right)_+^{\frac{k+1}{2}}$$

(with parameters $t \in S_{0,R}$ and $w \in S_{0,R^*}$) are Lipschitz continuous with uniformly bounded Lipschitz constants, which implies the assertion.

For the fourth estimate, let $B_1, \ldots, B_{K_r}$ be bounded and continuous functions $B_j : \left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k \to [-B, B]$ for some $B > 0$, and set

$$\mathcal{F}_{k,r} = \left\{\sum_{j=1}^{K_r} a_j \cdot B_j \quad : \quad a_j \in [-L_r, L_r] \quad (j = 1, \ldots, K_r)\right\}, \tag{2}$$

where $K_r, L_r > 0$. Define the corresponding least squares estimate by

$$\bar{m}_{n,(k,r,N)}^{(4)}(\cdot; d_{-n}^{-1}) = \arg\min_{f\in\mathcal{F}_{k,r}} \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|f(x_{i-k}^i, y_{i-k}^{i-1}) - T_N(y_i)\right|^2.$$

This definition only makes sense if $n \geq k+1$, so we set

$$\tilde{m}^{(4)}_{n,(k,r,N)}\left(u^0_{-k}, v^{-1}_{-k}; d^{-1}_{-n}\right) = \begin{cases} \bar{m}^{(4)}_{n,(k,r,N)}\left(u^0_{-k}, v^{-1}_{-k}; d^{-1}_{-n}\right) & \text{if } n \geq k+1, \\ 0 & \text{else.} \end{cases}$$

In the sequel let $i \in \{1,2,3,4\}$. Because we will need uniform boundedness of all experts for fixed sample size, we choose $0 < s < \frac{1}{2}$ and set

$$\hat{m}^{(i)}_{n,(k,r,N)}(x_0, d^{-1}_{-n}) := T_{n^s}\left(\tilde{m}^{(i)}_{n,(k,r,N)}(x^0_{-k}, y^{-1}_{-k}; d^{-1}_{-n})\right).$$

We will define our prediction strategy as a convex combination of these experts using weights, which are the higher the better the expert performed in the past. After $n-1$ rounds of play the normalized cumulative squared prediction error of $\hat{m}^{(i)}_{n,(k,r,N)}$ on the string d^{-1}_{-n} defined by

$$\begin{aligned} L^{(i)}_n(k,r,N) &:= L^{(i)}_n(k,r,N)(d^{-1}_{-n}) \\ &:= \frac{1}{n-1}\sum_{j=-n}^{-2}\left(T_{n^s}(y_{j+1}) - \hat{m}^{(i)}_{j+n+1,(k,r,N)}(x_{j+1}, d^{j}_{-n})\right)^2 \end{aligned}$$

quantizes the performance of the expert in the past. Let $(p_{(k,r,N)})_{(k,r,N)\in\mathcal{P}}$ be a probability distribution such that $p_{(k,r,N)} > 0$ for all $(k,r,N) \in \mathcal{P}$. Put $c_n = 8n^{2s}$ and define weights, which depend on this cumulative loss, by

$$w^{(i)}_{n,(k,r,N)} := p_{(k,r,N)} \cdot \exp\left(\frac{-(n-1)L^{(i)}_n(k,r,N)}{c_n}\right).$$

We define our prediction strategy $\hat{m}^{(i)}_n$ as a convex combination of $\hat{m}^{(i)}_{n,(k,r,N)}$ using the normalized values

$$v^{(i)}_{n,(k,r,N)} = \frac{w^{(i)}_{n,(k,r,N)}}{\sum_{(k,r,N)\in\mathcal{P}} w^{(i)}_{n,(k,r,N)}}$$

of $w^{(i)}_{n,(k,r,N)}$ as weights, i.e., $\hat{m}^{(i)}_n$ is defined by

$$\hat{m}^{(i)}_n(x_0, d^{-1}_{-n}) = \sum_{(k,r,N)\in\mathcal{P}} v^{(i)}_{n,(k,r,N)} \cdot \hat{m}^{(i)}_{n,(k,r,N)}(x_0, d^{-1}_{-n}).$$

In order to estimate

$$m(X^0_{-\infty}, Y^{-1}_{-\infty}) := \mathbf{E}\left\{Y_0 \mid X^0_{-\infty}, Y^{-1}_{-\infty}\right\},$$

we use the arithmetic mean of the first n estimates:

$$m^{(i)}_n(X_0, \mathcal{D}^{-1}_{-n}) = \frac{1}{n}\sum_{j=1}^{n}\hat{m}^{(i)}_j\left(X_0, \mathcal{D}^{-1}_{-j}\right).$$

3 Main Result

For fixed $k \in \mathbb{N}$ we call a sequence of partitions $(\mathcal{P}_{k,r})_{r\in\mathbb{N}}$ nested if the corresponding sequence of generated σ-algebras $\mathcal{F}(\mathcal{P}_{k,r})$ is increasing. Let $\mathcal{P}_{k,r} = \{A_{k,r,j}\}_j$ with Borel sets $A_{k,r,j}$ and denote by $A_{k,r}(z)$ the set $A_{k,r,j}$ which contains z.

Theorem 1. *Let $m_n^{(i)}$ be defined as in Section 2 ($i \in \{1,2,3,4\}$).*

a) Assume that in the definition of $m_n^{(1)}$ the sequence of bandwidths $h_r > 0$ satisfies

$$h_r \to 0 \quad (r \to \infty),$$

and that the kernel used in the definition of $\tilde{m}^{(1)}_{n,(k,r,N)}$ is defined by $K(v) := H\left(\|v\|^{(k+1)\cdot d+k}\right)$, where H is a nonincreasing nonnegative continuous function on $[0,\infty)$ satisfying

$$H(0) > 0 \quad \text{and} \quad t \cdot H(t) \to 0 \quad \text{as} \quad t \to \infty.$$

Then

$$\mathbf{E}\left\{\left|m_n^{(1)}(X_0, \mathcal{D}_{-n}^{-1}) - \mathbf{E}\left\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right|^2\right\} \to 0 \quad (n \to \infty)$$

for all stationary and ergodic sequences $((X_n, Y_n))_{n\in\mathbb{Z}}$ of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\{Y_0^2\} < \infty$.

b) Assume that the sequences $(\mathcal{P}_{k,r})_{r\in\mathbb{N}}$ of partitions used in the definition of $m_n^{(2)}$ satisfy:

- *(i) The sequences of partitions $(\mathcal{P}_{k,r})_{r\in\mathbb{N}}$ are nested ($k \in \mathbb{N}$).*
- *(ii) diam $A_{k,r}(z) := \sup_{u,v\in A_{k,r}(z)} \|u - v\| \to 0 \qquad (r \to \infty)$*
  for each $z \in \left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$ and every $k \in \mathbb{N}$.
- *(iii) $\left|\left\{A_{k,r,j} \in \mathcal{P}_{k,r} : A_{k,r,j} \cap [-L, L]^{d(k+1)+k} \neq \emptyset\right\}\right| < \infty$ for all $r \in \mathbb{N}$ and $L > 0$.*

Then

$$\mathbf{E}\left\{\left|m_n^{(2)}(X_0, \mathcal{D}_{-n}^{-1}) - \mathbf{E}\left\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right|^2\right\} \to 0 \quad (n \to \infty)$$

for all stationary and ergodic sequences $((X_n, Y_n))_{n\in\mathbb{Z}}$ of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\{Y_0^2\} < \infty$.

c) Assume $0 < p_r \to 0 \quad (r \to \infty)$ and (A1). Then

$$\mathbf{E}\left\{\left|m_n^{(3)}(X_0, \mathcal{D}_{-n}^{-1}) - \mathbf{E}\left\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right|^2\right\} \to 0 \quad (n \to \infty)$$

for all stationary and ergodic sequences $((X_n, Y_n))_{n\in\mathbb{Z}}$ of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\{Y_0^2\} < \infty$.

d) Assume that the function spaces $\mathcal{F}_{k,r}$ defined by (2) used in the construction of $m_n^{(3)}$ satisfy:

(i) $\limsup_{r\to\infty} K_r = \infty$.

(ii) $\limsup_{r\to\infty} L_r = \infty$.

(iii) For all k and for any probability measure μ on $\left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$ and for every $g \in L_2\left(\left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k, \mu\right)$ it holds that

$$\liminf_{r\to\infty} \inf_{f\in\mathcal{F}_{k,r}} \int |g-f|^2 \, d\mu = 0.$$

Then

$$\mathbf{E}\left\{\left|m_n^{(4)}(X_0, \mathcal{D}_{-n}^{-1}) - \mathbf{E}\left\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\right\}\right|^2\right\} \to 0 \quad (n\to\infty)$$

for all stationary and ergodic sequences $((X_n, Y_n))_{n\in\mathbb{Z}}$ of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\left\{Y_0^2\right\} < \infty$.

Remarks.

a) The condition on the kernel is for example satisfied if we choose $H(t) = \exp(-t^2)$.

b) The proofs rely on pointwise consistency results for regression estimates. In order to apply these results in case of the partitioning estimate, we need the condition that the partitions are nested.

c) The conditions on the function spaces are for example met if we choose $\mathcal{F}_{k,r}$ as a suitably defined tensor product B-spline space (cf. Corollary 2 in Jones, Kohler and Walk (2011)).

4 Proofs

First we present some tools (Lemma 1-9) and then we prove Theorem 1. The following notations will be needed in Lemma 1:

In the time series problem at each time instant $i = 1, 2, \cdots$ the predictor is asked to guess the outcome y_i of a sequence of real numers $y_1, y_2, \cdots$ with knowledge of the past (x_1^i, y_1^{i-1}). A prediction strategy is a sequence $g = \{g_i\}_{i=1}^\infty$ of decision functions

$$g_i : (\mathbb{R}^d)^i \times \mathbb{R}^{i-1} \to \mathbb{R}$$

and the prediction formed at time i is $g_i(x_1^i, y_1^{i-1})$. After n rounds of play, the normalized cumulative prediction error on the string (x_1^n, y_1^n) is

$$L_n(g) = \frac{1}{n}\sum_{i=1}^n \left(g_i(x_1^i, y_1^{i-1}) - y_i\right)^2.$$

Lemma 1. *Let $\tilde{h}_1, \tilde{h}_2, \cdots$ be a sequence of prediction strategies (experts), and let $\{q_k\}$ be a probability distribution on the set of positive integers. Assume that $\tilde{h}_i(x_1^n, y_1^{n-1}) \in [-B, B]$ and $y_1^n \in [-B, B]^n$. Define*

$$w_{t,k} = q_k \cdot \exp\left(\frac{-(t-1)L_{t-1}(\tilde{h}_k)}{c}\right)$$

with $c \geq 8B^2$, and

$$v_{t,k} = \frac{w_{t,k}}{\sum_{i=1}^{\infty} w_{t,i}}.$$

If the prediction strategy $\tilde{g}$ is defined by

$$\tilde{g}_t(x_1^t, y_1^{t-1}) = \sum_{k=1}^{\infty} v_{t,k} \cdot \tilde{h}_k(x_1^t, y_1^{t-1})$$

then, for every $n \geq 1$,

$$L_n(\tilde{g}) \leq \inf_k \left(L_n(\tilde{h}_k) - \frac{c \ln q_k}{n}\right).$$

Here $-\ln 0$ is treated as ∞.

Proof. See proof of Lemma 27.3 in Györfi et al. (2002). □

Lemma 2. *Let $m \in L_2(\mu)$ and let m^* be the generalized Hardy-Littlewood maximal function of m defined by*

$$m^*(x) = \sup_{h>0} \frac{1}{\mu(S_{x,h})} \int_{S_{x,h}} |m| \, d\mu \quad (x \in \mathbb{R}^d).$$

Then $m^* \in L_2(\mu)$ and

$$\int m^*(x)^2 \mu(dx) \leq c^* \int m(x)^2 \mu(dx),$$

where $c^* < \infty$ depends only on d.

Proof. See proof of Lemma 24.7 in Györfi et al. (2002). □

Lemma 3. *Let $m \in L_2(\mu)$, let $(\mathcal{P}_n)_{n\in\mathbb{N}}$ be a sequence of partitions, and let m^* denote the corresponding generalized Hardy-Littlewood maximal function of m defined by*

$$m^*(x) = \sup_n \frac{1}{\mu(A_n(x))} \int_{A_n(x)} |m| \, d\mu \quad (x \in \mathbb{R}^d).$$

If the sequence of partitions $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is nested, then

$$m^* \in L_2(\mu) \qquad \text{and} \qquad \int m^*(x)^2 \mu(dx) \leq c^* \int m(x)^2 \mu(dx),$$

where $c^* < \infty$ depends only on d.

Proof. See Problem 24.4 in Györfi et al. (2002). □

Lemma 4. *Assume*

$$\begin{aligned}&c_1 H(\|x\|) \le K(x) \le c_2 H(\|x\|) \quad (x \in \mathbb{R}^d),\\ &H(+0) > 0,\\ &t^d H(t) \to 0 \quad as \quad t \to \infty,\end{aligned}$$

where H is a nonincreasing nonnegative Borel function on $[0, \infty)$ and $c_1, c_2 > 0$. Then, for all μ-integrable functions f,

$$\lim_{h \to 0} \frac{\int K((x-z)/h) f(z) \mu(dz)}{\int K((x-z)/h) \mu(dz)} = f(x)$$

for μ-almost all $x \in \mathbb{R}^d$.

Proof. See proof of Lemma 24.8 in Györfi et al. (2002). □

Lemma 5. *Assume that the sequence of partitions $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \cdots\}$ of $\mathbb{R}^d$ is nested and*

$$diam\ A_n(z) := \sup_{u,v \in A_n(z)} \|u - v\| \to 0 \qquad (n \to \infty)$$

for each $z \in \mathbb{R}^d$. Then

$$\lim_{n \to \infty} \frac{\int_{A_n(x)} f(z) \mu(dz)}{\mu(A_n(x))} = f(x)$$

for μ-almost all $x \in \mathbb{R}^d$.

Proof. See Problem 24.3 in Györfi et al. (2002). □

The proofs of the next three lemmas will be given after the proof of Theorem 1.

In Lemma 6 we will need the following notation: For an arbitrary function f defined on the image space of a random variable X define

$$\|f\|_{\infty, supp(\mathbf{P}_X)} := \sup_{x \in supp(\mathbf{P}_X)} |f(x)|.$$

Lemma 6. *Let $((X_n, Y_n))_{n \in \mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^d \times \mathbb{R}$-valued random variables.*

Let $B_1, \ldots, B_K$ be bounded and continuous functions $B_j : (\mathbb{R}^d)^{k+1} \times \mathbb{R}^k \to [-B, B]$ for some $k \in \mathbb{N}, B > 0$ and set

$$\mathcal{F} = \left\{ \sum_{j=1}^{K} a_j \cdot B_j \quad : \quad a_j \in [-L, L] \quad (j = 1, \ldots, K) \right\}$$

for some $L > 0$. Define the least squares estimate m_n by

$$m_n = \arg\min_{f\in\mathcal{F}} \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left| f(X_{i-k}^{i}, Y_{i-k}^{i-1}) - T_N\left(Y_i\right) \right|^2$$

for some $N > 0$, and assume that $n \geq k+1$. Then

$$f^* := \arg\min_{f\in\mathcal{F}} \mathbf{E}\left\{ \left| f(X_{-k}^{0}, Y_{-k}^{-1}) - T_N\left(Y_0\right) \right|^2 \right\}$$

exists and

$$\|m_n - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \to 0 \quad a.s.$$

Remark. It follows from the proof of Lemma 6 that $\min_{f\in\mathcal{F}} \mathbf{E}\left\{ \left| f(X_{-k}^{0}, Y_{-k}^{-1}) - Y_0 \right|^2 \right\}$ exists as well.

Lemma 7. *Let $((X_n, Y_n))_{n\in\mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^d \times \mathbb{R}$-valued random variables with $\mathbf{E}\left\{Y_0^2\right\} < \infty$.*

With the assumptions of Theorem 1 a) and

$$m_{k,r}^{(1)}(x_{-k}^0, y_{-k}^{-1}) := \frac{\mathbf{E}\left\{ Y_0 \cdot K\left(\frac{(X_{-k}^0, Y_{-k}^{-1}) - (x_{-k}^0, y_{-k}^{-1})}{h_r} \right) \right\}}{\mathbf{E}\left\{ K\left(\frac{(X_{-k}^0, Y_{-k}^{-1}) - (x_{-k}^0, y_{-k}^{-1})}{h_r} \right) \right\}}$$

or according to the assumptions of Theorem 1 b) and

$$m_{k,r}^{(2)}(x_{-k}^0, y_{-k}^{-1}) := \frac{\mathbf{E}\left\{ Y_0 \mathbf{1}_{\left\{(X_{-k}^0, Y_{-k}^{-1}) \in A_{k,r}\left(x_{-k}^0, y_{-k}^{-1}\right)\right\}} \right\}}{\mathbf{E}\left\{ \mathbf{1}_{\left\{(X_{-k}^0, Y_{-k}^{-1}) \in A_{k,r}\left(x_{-k}^0, y_{-k}^{-1}\right)\right\}} \right\}}$$

or according to the assumptions of Theorem 1 c) and

$$m_{k,r}^{(3)}(x_{-k}^0, y_{-k}^{-1}) := \frac{\mathbf{E}\left\{ Y_0 \mathbf{1}_{(X_{-k}^0, Y_{-k}^{-1}) \in S_{(x_{-k}^0, y_{-k}^{-1}), R_{k,r}(x_{-k}^0, y_{-k}^{-1})}} \right\}}{p_r}$$

(with arbitrary

$$R_{k,r}(x_{-k}^0, y_{-k}^{-1}) \in \left[R'_{k,r}(x_{-k}^0, y_{-k}^{-1}), R''_{k,r}(x_{-k}^0, y_{-k}^{-1})\right]$$

where $\left[R'_{k,r}(x_{-k}^0, y_{-k}^{-1}), R''_{k,r}(x_{-k}^0, y_{-k}^{-1})\right]$ is the set of values $R_{k,r}(x_{-k}^0, y_{-k}^{-1})$ such that $\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}\left(S_{(x_{-k}^0, y_{-k}^{-1}), R_{k,r}(x_{-k}^0, y_{-k}^{-1})} \right) = p_r$) or according to the assumptions of Theorem 1 d) and

$$m_{k,r}^{(4)} := \arg\min_{f\in\mathcal{F}_{k,r}} \mathbf{E}\left\{ \left| f(X_{-k}^0, Y_{-k}^{-1}) - Y_0 \right|^2 \right\}$$

it holds for $i \in \{1,2,3,4\}$:

$$\lim_{t\to\infty} \inf_{k,r\geq t} \mathbf{E}\left\{ \left| Y_0 - m_{k,r}^{(i)}(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\} \leq \mathbf{E}\left\{ \left| Y_0 - m(X_{-\infty}^0, Y_{-\infty}^{-1}) \right|^2 \right\}.$$

Lemma 8. *Let $((X_n,Y_n))_{n\in\mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^d\times\mathbb{R}$-valued random variables with $\mathbf{E}\left\{Y_0^2\right\}<\infty$, let $\tilde{m}^{(i)}_{n,(k,r,N)}$ be defined as in Section 2 and let $m^{(i)}_{k,r}$ be defined as in Lemma 7 for $i\in\{1,2,3,4\}$. With the assumptions of Theorem 1 a) or according to Theorem 1 b) or according to Theorem 1 c) or according to Theorem 1 d) it holds for arbitrary $k,r\in\mathbb{N}$ and $i\in\{1,2,3,4\}$:*

$$\limsup_{N\to\infty}\limsup_{n\to\infty}\mathbf{E}\left\{\left|m^{(i)}_{k,r}(X^0_{-k},Y^{-1}_{-k})-\tilde{m}^{(i)}_{n,(k,r,N)}(X^0_{-k},Y^{-1}_{-k};\mathcal{D}^{-1}_{-n})\right|^2\right\}=0.$$

4.1 Proof of Theorem 1

We recall that $m(X^0_{-\infty},Y^{-1}_{-\infty}):=\mathbf{E}\{Y_0|X^0_{-\infty},Y^{-1}_{-\infty}\}$. Using the definition of the estimate and the inequality of Jensen, we get

$$\begin{aligned}
&\mathbf{E}\left\{\left|m^{(i)}_n(X_0,\mathcal{D}^{-1}_{-n})-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\}\\
&=\mathbf{E}\left\{\left|\frac{1}{n}\sum_{j=1}^n\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\}\\
&\le\mathbf{E}\left\{\frac{1}{n}\sum_{j=1}^n\left|\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\}.
\end{aligned}$$

Using

$$\begin{aligned}
&\mathbf{E}\left\{\frac{1}{n}\sum_{j=1}^n\left|Y_0-\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)\right|^2\right\}\\
&=\mathbf{E}\left\{\frac{1}{n}\sum_{j=1}^n\left|\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\}+\mathbf{E}\left\{\left|Y_0-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\},
\end{aligned}$$

which follows from

$$\begin{aligned}
&\mathbf{E}\left\{\left(\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\left(Y_0-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\right\}\\
&=\mathbf{E}\Big\{\left(\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\\
&\qquad\qquad\cdot\mathbf{E}\Big\{\left(Y_0-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\mid(X^0_{-\infty},Y^{-1}_{-\infty})\Big\}\Big\}\\
&=\mathbf{E}\left\{\left(\hat{m}^{(i)}_j\left(X_0,\mathcal{D}^{-1}_{-j}\right)-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\left(m(X^0_{-\infty},Y^{-1}_{-\infty})-m(X^0_{-\infty},Y^{-1}_{-\infty})\right)\right\}=0
\end{aligned}$$

for every $j\in\{1,\cdots,n\}$, we conclude

$$0\ \le\ \mathbf{E}\left\{\left|m^{(i)}_n(X_0,\mathcal{D}^{-1}_{-n})-m(X^0_{-\infty},Y^{-1}_{-\infty})\right|^2\right\}$$

$$\leq \quad \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|Y_0 - \hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right)\right|^2 \right\} - \mathbf{E}\left\{ \left|Y_0 - m(X_{-\infty}^0, Y_{-\infty}^{-1})\right|^2 \right\}.$$

Therefore it suffices to show

$$\limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|Y_0 - \hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right)\right|^2 \right\} \leq \mathbf{E}\left\{ \left|Y_0 - m(X_{-\infty}^0, Y_{-\infty}^{-1})\right|^2 \right\} =: L^*. \quad (3)$$

Using the inequality

$$(a+b)^2 \leq (1+\alpha)\ a^2 + (1+\frac{1}{\alpha})\ b^2, \quad (4)$$

for arbitrary $a, b \in \mathbb{R},\ \alpha > 0$, one has

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|Y_0 - \hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right)\right|^2 \right\} \\
&\leq \limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}(1+\alpha)\left|\hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right) - T_{n^s}(Y_0)\right|^2 \right\} \\
&\quad + \limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}(1+\frac{1}{\alpha})\left|T_{n^s}(Y_0) - Y_0\right|^2 \right\}.
\end{aligned}$$

With the assumption $\mathbf{E}\left\{Y_0^2\right\} < \infty$ and Lebesgue's dominated convergence theorem we get for the last summand:

$$(1+\frac{1}{\alpha})\limsup_{n\to\infty} \mathbf{E}\left\{ \left|T_{n^s}(Y_0) - Y_0\right|^2 \right\} \leq (1+\frac{1}{\alpha})\limsup_{n\to\infty} \mathbf{E}\left\{ Y_0^2 \cdot \mathbf{1}_{\{|Y_0| > n^s\}} \right\} = 0.$$

Since $\alpha > 0$ was arbitrary, (3) follows from

$$\limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|\hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right) - T_{n^s}(Y_0)\right|^2 \right\} \leq L^*. \quad (5)$$

Note that $\limsup_{n\to\infty} \frac{c_n \ln p_{(k,r,N)}}{n} = 0$. With the stationarity of the data and Lemma 1 we conclude

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|\hat{m}_j^{(i)}\left(X_0, \mathcal{D}_{-j}^{-1}\right) - T_{n^s}(Y_0)\right|^2 \right\} \\
&= \limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n}\left|\hat{m}_j^{(i)}\left(X_{j+1}, \mathcal{D}_1^{j}\right) - T_{n^s}(Y_{j+1})\right|^2 \right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \limsup_{n\to\infty} \mathbf{E}\Bigg\{ \inf_{(k,r,N)\in\mathcal{P}} \Big(\frac{1}{n}\sum_{j=1}^{n} \big|T_{n^s}(Y_{j+1}) - \hat{m}^{(i)}_{j,(k,r,N)}(X_{j+1},\mathcal{D}_1^j)\big|^2 \\
&\qquad\qquad - \frac{c_n \ln(p_{(k,r,N)})}{n} \Big) \Bigg\} \\
&\leq \inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{E}\Bigg\{ \frac{1}{n}\sum_{j=1}^{n} \big|T_{n^s}(Y_{j+1}) - \hat{m}^{(i)}_{j,(k,r,N)}(X_{j+1},\mathcal{D}_1^j)\big|^2 \Bigg\}.
\end{aligned}
$$

Using the inequality

$$
\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} a_j \leq \limsup_{n\to\infty} a_n
$$

for $a_n \in \mathbb{R}$, the stationarity of the data and

$$
\big|T_{n^s}(z) - T_{n^s}(y)\big| \leq \big|z-y\big|,
$$

we get

$$
\begin{aligned}
&\inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}\Big\{ \big|T_{n^s}(Y_{j+1}) - \hat{m}^{(i)}_{j,(k,r,N)}(X_{j+1},\mathcal{D}_1^j)\big|^2 \Big\} \\
&\leq \inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{E}\left\{ \big|T_{n^s}(Y_{n+1}) - \hat{m}^{(i)}_{n,(k,r,N)}(X_{n+1},\mathcal{D}_1^n)\big|^2 \right\} \\
&= \inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{E}\left\{ \big|T_{n^s}(Y_0) - \hat{m}^{(i)}_{n,(k,r,N)}(X_0,\mathcal{D}_{-n}^{-1})\big|^2 \right\} \\
&\leq \inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{E}\left\{ \big|Y_0 - \tilde{m}^{(i)}_{n,(k,r,N)}(X_{-k}^0, Y_{-k}^{-1};\mathcal{D}_{-n}^{-1})\big|^2 \right\}.
\end{aligned}
$$

Let $m^{(i)}_{k,r}$ be defined as in Lemma 7. Using

$$
\inf_{k,r,N}(a_{k,r} + b_{k,r,N}) \leq \lim_{t\to\infty} \inf_{k,r\geq t} a_{k,r} + \lim_{t\to\infty} \sup_{k,r\geq t} \limsup_{N\to\infty} b_{k,r,N},
$$

we can conclude from Lemma 7, Lemma 8 and inequality (4) that we have for arbitrary $\gamma > 0$

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbf{E}\left\{ \frac{1}{n}\sum_{j=1}^{n} \big|\hat{m}^{(i)}_j\left(X_0,\mathcal{D}_{-j}^{-1}\right) - T_{n^s}(Y_0)\big|^2 \right\} \\
&\leq \inf_{(k,r,N)\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{E}\left\{ \left|Y_0 - \tilde{m}^{(i)}_{n,(k,r,N)}(X_{-k}^0, Y_{-k}^{-1};\mathcal{D}_{-n}^{-1})\right|^2 \right\} \\
&\leq (1+\gamma) \lim_{t\to\infty} \inf_{k,r\geq t} \mathbf{E}\left\{ \left|Y_0 - m^{(i)}_{k,r}(X_{-k}^0, Y_{-k}^{-1})\right|^2 \right\} \\
&\quad + (1+\frac{1}{\gamma}) \lim_{t\to\infty} \sup_{k,r\geq t} \limsup_{N\to\infty} \limsup_{n\to\infty}
\end{aligned}
$$

$$\mathbf{E}\left\{\left|m_{k,r}^{(i)}(X_{-k}^0, Y_{-k}^{-1}) - \tilde{m}_{n,(k,r,N)}^{(i)}(X_{-k}^0, Y_{-k}^{-1}; \mathcal{D}_{-n}^{-1})\right|^2\right\}$$
$$\leq (1+\gamma)\mathbf{E}\left\{\left|Y_0 - m(X_{-\infty}^0, Y_{-\infty}^{-1})\right|^2\right\}.$$

With $\gamma \to 0$ this implies (5). The proof is complete. □

4.2 Proof of Lemma 6

Within the proof, we will need the notion of sup-norm covering numbers, which we introduce in the next definition.

Definition 1. *Let $\varepsilon > 0$ and let $\mathcal{G}$ be a set of functions $\mathbb{R}^d \to \mathbb{R}$. Every finite collection of functions $g_1, ..., g_N : \mathbb{R}^d \to \mathbb{R}$ with the property that for every $g \in \mathcal{G}$ there is a $j = j(g) \in \{1, ..., N\}$ such that*

$$\|g - g_j\|_\infty := \sup_z |g(z) - g_j(z)| < \varepsilon,$$

*is called an ε-cover of $\mathcal{G}$ with respect to $\|\cdot\|_\infty$. Let $\mathcal{N}(\varepsilon, \mathcal{G}, \|\cdot\|_\infty)$ be the size of the smallest ε-cover of $\mathcal{G}$ w.r.t. $\|\cdot\|_\infty$, take $\mathcal{N}(\varepsilon, \mathcal{G}, \|\cdot\|_\infty) = \infty$ if no finite ε-cover exists. Then $\mathcal{N}(\varepsilon, \mathcal{G}, \|\cdot\|_\infty)$ is called the ε-**covering number of** $\mathcal{G}$ w.r.t. $\|\cdot\|_\infty$ and will be abbreviated to $\mathcal{N}_\infty(\varepsilon, \mathcal{G})$.*

The proof will be divided into several steps. *In the first step of the proof* we show the existence of f^* as defined in the lemma and that whenever $f_1, f_2 \in \mathcal{F}$ satisfy

$$\mathbf{E}\left\{\left|f_1(X_{-k}^0, Y_{-k}^{-1}) - T_N(Y_0)\right|^2\right\} = \mathbf{E}\left\{\left|f_2(X_{-k}^0, Y_{-k}^{-1}) - T_N(Y_0)\right|^2\right\}$$
$$= \min_{f \in \mathcal{F}} \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N(Y_0)\right|^2\right\}, \tag{6}$$

then

$$\int |f_1(x_{-k}^0, y_{-k}^{-1}) - f_2(x_{-k}^0, y_{-k}^{-1})|^2 d\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(x_{-k}^0, y_{-k}^{-1}) = 0. \tag{7}$$

Moreover, we prove that for arbitrary continuous $f : \left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k \to \mathbb{R}$

$$\int |f(x_{-k}^0, y_{-k}^{-1})| d\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(x_{-k}^0, y_{-k}^{-1}) = 0 \tag{8}$$

implies

$$\|f\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} = 0. \tag{9}$$

As for the existence of f^* we observe that the minimization problem can be rewritten as follows:

$$\inf_{f \in \mathcal{F}} \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N(Y_0)\right|^2\right\} = \inf_{a_1, \ldots, a_K \in [-L, L]} g(a_1, ..., a_K),$$

where

$$g(a_1,...,a_K) = \mathbf{E}\left\{\left|\sum_{j=1}^{K} a_j \cdot B_j(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\}.$$

It can be easily shown that g is continuous, thus f^* exists.

Now suppose that $f_1, f_2 \in \mathcal{F}$ satisfy (6). The definition of $\mathcal{F}$ implies

$$\frac{f_1+f_2}{2} \in \mathcal{F}.$$

Using $(a+b)^2 = 2a^2+2b^2-(a-b)^2$ $(a,b \in \mathbb{R})$, we obtain

$$\begin{aligned}
&\mathbf{E}\left\{\left|\frac{f_1(X_{-k}^0, Y_{-k}^{-1}) + f_2(X_{-k}^0, Y_{-k}^{-1})}{2} - T_N\left(Y_0\right)\right|^2\right\} \\
&= \mathbf{E}\left\{\left|\frac{f_1(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)}{2} + \frac{f_2(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)}{2}\right|^2\right\} \\
&= \frac{1}{2}\cdot\mathbf{E}\left\{\left|f_1(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} + \frac{1}{2}\cdot\mathbf{E}\left\{\left|f_2(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \\
&\qquad - \frac{1}{4}\cdot\mathbf{E}\left\{\left|f_1(X_{-k}^0, Y_{-k}^{-1}) - f_2(X_{-k}^0, Y_{-k}^{-1})\right|^2\right\} \\
&= \min_{f\in\mathcal{F}} \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \\
&\qquad - \frac{1}{4}\mathbf{E}\left\{\left|f_1(X_{-k}^0, Y_{-k}^{-1}) - f_2(X_{-k}^0, Y_{-k}^{-1})\right|^2\right\}.
\end{aligned}$$

Because of the minimum property the last term vanishes and thus (7) holds.

In order to show the last result of the first step, let $f : \left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k \to \mathbb{R}$ be continuous and assume that $|f(x_{-k}^0, y_{-k}^{-1})| \geq \delta > 0$ for some $(x_{-k}^0, y_{-k}^{-1}) \in supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})$. Choose a ball S around x with radius greater than zero such that $|f|$ is greater than $\delta/2$ on S. Then

$$\int |f(x_{-k}^0, y_{-k}^{-1})| d\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(x_{-k}^0, y_{-k}^{-1}) \geq \frac{\delta}{2}\cdot\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(S) > 0.$$

Thus (8) implies (9).

In the second step of the proof we show

$$\begin{aligned}
&\int \left|m_n\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\
&\qquad \to \min_{f\in\mathcal{F}} \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \quad a.s.
\end{aligned}$$

Since $m_n \in \mathcal{F}$ for all n, it suffices to verify

$$\limsup_{n\to\infty} \int \left|m_n\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right)$$

$$\le \mathbf{E}\left\{\left|f^*(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \quad a.s.$$

But this follows from

$$\begin{aligned}
&\int \left|m_n\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\
&\le \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|m_n(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \\
&\qquad + \sup_{f\in\mathcal{F}} \left| \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \right. \\
&\qquad\qquad \left. - \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|f(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \right| \\
&\le \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|f^*(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \\
&\qquad + \sup_{f\in\mathcal{F}} \left| \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} - \right. \\
&\qquad\qquad \left. \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|f(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \right| \\
&\le \mathbf{E}\left\{\left|f^*(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \\
&\qquad + 2\cdot \sup_{f\in\mathcal{F}} \left| \mathbf{E}\left\{\left|f(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \right. \\
&\qquad\qquad \left. - \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|f(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \right|,
\end{aligned}$$

where the second inequality is implied by the definition of m_n. By discretizing the coefficients a_j accordingly, it is easy to see that $\mathcal{N}_\infty\left(\varepsilon, \mathcal{F}\right) < \infty$ for arbitrary $\varepsilon > 0$. By using the ε-cover $\mathcal{G}_\varepsilon$ and the boundedness of functions in $\mathcal{F}$, the convergence of the last term can be reduced to the convergence of

$$\begin{aligned}
&\sup_{g\in\mathcal{G}_\varepsilon} \left| \mathbf{E}\left\{\left|g(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \right. \\
&\qquad \left. - \frac{1}{n-k} \sum_{i=-n+k}^{-1} \left|g(X_{i-k}^i, Y_{i-k}^{i-1}) - T_N\left(Y_i\right)\right|^2 \right| \to 0 \quad a.s.,
\end{aligned}$$

for arbitrary $\varepsilon > 0$. As $\mathcal{G}_\varepsilon$ is finite, this in turn follows from the ergodic theorem.

In the third and last step of the proof we show that for $f_n \in \mathcal{F}$ with

$$\int \left|f_n\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right)$$

$$\to \mathbf{E}\left\{\left|f^*(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \quad (n \to \infty), \tag{10}$$

we have

$$\|f_n - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \to 0 \quad (n \to \infty).$$

Assume that the assertion is not true. Then we can find functions $f_n \in \mathcal{F}$ such that

$$\begin{aligned}\int \left|f_n\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\ \to \mathbf{E}\left\{\left|f^*(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\}\end{aligned}$$

and

$$\|f_n - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \geq \delta \quad (n \in \mathbb{N})$$

for some $\delta > 0$. By applying the theorem of Bolzano-Weierstrass successively to the coefficients of these functions, we can construct a function $\bar{f} \in \mathcal{F}$ such that a subsequence of $(f_n)_{n \in \mathbb{N}}$ satisfies

$$\|f_{n_k} - \bar{f}\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \to 0 \quad (n \to \infty). \tag{11}$$

But this implies

$$\begin{aligned}&\|\bar{f} - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} + \|f_{n_k} - \bar{f}\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \\ &\geq \|f_{n_k} - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \geq \delta\end{aligned}$$

for arbitrary k and thus by (11)

$$\|\bar{f} - f^*\|_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})} \geq \delta. \tag{12}$$

Let $\gamma > 0$ be arbitrary. By inequality (4) we have

$$\begin{aligned}&\mathbf{E}\left\{\left|\bar{f}(X_{-k}^0, Y_{-k}^{-1}) - T_N\left(Y_0\right)\right|^2\right\} \\ &\leq (1+\gamma) \int \left|f_{n_k}\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\ &\quad + \left(1 + \frac{1}{\gamma}\right) \int \left|f_{n_k}\left(x_{-k}^0, y_{-k}^{-1}\right) - \bar{f}\left(x_{-k}^0, y_{-k}^{-1}\right)\right|^2 \\ &\qquad\qquad d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\ &\leq (1+\gamma) \int \left|f_{n_k}\left(x_{-k}^0, y_{-k}^{-1}\right) - T_N\left(y_0\right)\right|^2 d\mathbf{P}_{\left(X_{-k}^0, Y_{-k}^0\right)}\left(x_{-k}^0, y_{-k}^0\right) \\ &\quad + \left(1 + \frac{1}{\gamma}\right) \|f_{n_k} - \bar{f}\|^2_{\infty, supp(\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})})}.\end{aligned}$$

Because of (10) and (11), this implies

$$\mathbf{E}\left\{\left|\bar{f}(X_{-k}^{0},Y_{-k}^{-1})-T_N\left(Y_0\right)\right|^2\right\}\leq(1+\gamma)\,\mathbf{E}\left\{\left|f^{*}(X_{-k}^{0},Y_{-k}^{-1})-T_N\left(Y_0\right)\right|^2\right\}$$

for arbitrary $\gamma>0$ and thus, by definition of f^*,

$$\mathbf{E}\left\{\left|\bar{f}(X_{-k}^{0},Y_{-k}^{-1})-T_N\left(Y_0\right)\right|^2\right\}=\mathbf{E}\left\{\left|f^{*}(X_{-k}^{0},Y_{-k}^{-1})-T_N\left(Y_0\right)\right|^2\right\},$$

in contradiction to (12) and what we have shown in the first part of the proof. This completes step three. Combining the results of steps two and three, we see that Lemma 6 indeed holds. □

Remark. If $|Y_0|\leq\beta$ almost surely for some $\beta>0$, we can drop the truncation operator throughout Lemma 6 by choosing N accordingly.

4.3 Proof of Lemma 7

By Lemma 4 and Lemma 5, respectively, we get for $i\in\{1,2\}$

$$\lim_{r\to\infty}m_{k,r}^{(i)}(x_{-k}^{0},y_{-k}^{-1})=\mathbf{E}\left\{Y_0\mid X_{-k}^{0}=x_{-k}^{0},Y_{-k}^{-1}=y_{-k}^{-1}\right\}\tag{13}$$

for $\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}$-almost all $(x_{-k}^{0},y_{-k}^{-1})\in(\mathbb{R}^d)^{k+1}\times\mathbb{R}^k$. (13) also holds for $i=3$, because $p_r\to0\quad(r\to\infty)$ implies $R''_{k,r}(x_{-k}^{0},y_{-k}^{-1})\to0$ $(r\to\infty)$ for $\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}$-almost all $(x_{-k}^{0},y_{-k}^{-1})\in(\mathbb{R}^d)^{k+1}\times\mathbb{R}^k$ which allows to apply Lemma 4.

Arguing as in the proof of Theorem 24.2 in Györfi et al. (2002), we get for $m_{k,r}^{(1)}$

$$\begin{aligned}m_{k,r}^{(1)}(x_{-k}^{0},y_{-k}^{-1})&\leq\sup_{r>0}\frac{\mathbf{E}\left\{Y_0\cdot K\left(\frac{(X_{-k}^{0},Y_{-k}^{-1})-(x_{-k}^{0},y_{-k}^{-1})}{h_r}\right)\right\}}{\mathbf{E}\left\{K\left(\frac{(X_{-k}^{0},Y_{-k}^{-1})-(x_{-k}^{0},y_{-k}^{-1})}{h_r}\right)\right\}}\\&\leq\frac{c_2}{c_1}\sup_{h>0}\frac{\int_{S_{(x_{-k}^{0},y_{-k}^{-1}),h}}\left|m_k(u_{-k}^{0},v_{-k}^{-1})\right|d\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}(u_{-k}^{0},v_{-k}^{-1})}{\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}\left(S_{(x_{-k}^{0},y_{-k}^{-1}),h}\right)},\end{aligned}$$

and an analogous result (with $c_1=c_2=1$) for $m_{k,r}^{(3)}$. Obviously, we have in addition

$$\begin{aligned}&m_{k,r}^{(2)}(x_{-k}^{0},y_{-k}^{-1})\\&\leq\sup_{r>0}\frac{\int_{A_{k,r}\left((x_{-k}^{0},y_{-k}^{-1})\right)}\left|m_k(u_{-k}^{0},v_{-k}^{-1})\right|d\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}(u_{-k}^{0},v_{-k}^{-1})}{\mathbf{P}_{(\mathbf{X_{-k}^{0}},\mathbf{Y_{-k}^{-1}})}\left(A_{k,r}\left((x_{-k}^{0},y_{-k}^{-1})\right)\right)}.\end{aligned}$$

Using these relations, Lemma 2 and Lemma 3, respectively, and the dominated convergence theorem, we see that (13) implies for $i\in\{1,2,3\}$

$$\lim_{t\to\infty}\inf_{k,r\geq t}\mathbf{E}\left\{\left|Y_0-m_{k,r}^{(i)}(X_{-k}^{0},Y_{-k}^{-1})\right|^2\right\}$$

$$
\begin{aligned}
&\le \limsup_{k\to\infty} \limsup_{r\to\infty} \mathbf{E}\left\{ \left| Y_0 - m_{k,r}^{(i)}(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\} \\
&= \limsup_{k\to\infty} \mathbf{E}\left\{ \left| Y_0 - \mathbf{E}\left\{ Y_0 \mid X_{-k}^0, Y_{-k}^{-1} \right\} \right|^2 \right\}.
\end{aligned}
$$

The sequence $(M_k)_{k\in\mathbb{N}} := \left(\mathbf{E}\left\{Y_0 \mid X_{-k}^0, Y_{-k}^{-1}\right\}\right)_{k\in\mathbb{N}}$ is a martingale with the property $\sup_{k\in\mathbb{N}} \mathbf{E}\left\{M_k^2\right\} \le \mathbf{E}\left\{Y_0^2\right\} < \infty$. By Loève (1977), 32.4.A, we know that M_k converges almost surely and in L_2 to a square integrable random variable and the limit is $\mathbf{E}\left\{Y_0 \mid X_{-\infty}^0, Y_{-\infty}^{-1}\right\}$. Now we can conclude

$$
\limsup_{k\to\infty} \mathbf{E}\left\{ \left| Y_0 - \mathbf{E}\left\{ Y_0 \mid X_{-k}^0, Y_{-k}^{-1} \right\} \right|^2 \right\} = \mathbf{E}\left\{ \left| Y_0 - m(X_{-\infty}^0, Y_{-\infty}^{-1}) \right|^2 \right\}
$$

and the proof is complete for $i \in \{1, 2, 3\}$.

For $i = 4$ set $L^* := \mathbf{E}\left\{ \left| Y_0 - \mathbf{E}\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\} \right|^2 \right\}$. Straightforward calculation leads to

$$
\begin{aligned}
\mathbf{E}\left\{ \left| f(X_{-k}^0, Y_{-k}^{-1}) - Y_0 \right|^2 \right\} = L^* &+ \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\} - \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} \right|^2 \right\} \\
&+ \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} - f(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\}.
\end{aligned}
$$

for arbitrary $f \in \mathcal{F}_{k,r}$. Thus by definition of $m_{k,r}^{(4)}$

$$
\begin{aligned}
&\lim_{t\to\infty} \inf_{k,r\ge t} \mathbf{E}\left\{ \left| Y_0 - m_{k,r}^{(4)}(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\} \\
&= \lim_{t\to\infty} \inf_{k,r\ge t} \inf_{f\in\mathcal{F}_{k,r}} \mathbf{E}\left\{ \left| f(X_{-k}^0, Y_{-k}^{-1}) - Y_0 \right|^2 \right\} \\
&= L^* + \lim_{t\to\infty} \inf_{k,r\ge t} \left(\inf_{f\in\mathcal{F}_{k,r}} \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} - f(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\} \right. \\
&\qquad\qquad \left. + \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\} - \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} \right|^2 \right\} \right) \\
&\le L^* + \limsup_{k\to\infty} \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-\infty}^0, Y_{-\infty}^{-1}\} - \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} \right|^2 \right\} \\
&\qquad + \lim_{t\to\infty} \inf_{k,r\ge t} \inf_{f\in\mathcal{F}_{k,r}} \mathbf{E}\left\{ \left| \mathbf{E}\{Y_0|X_{-k}^0, Y_{-k}^{-1}\} - f(X_{-k}^0, Y_{-k}^{-1}) \right|^2 \right\}.
\end{aligned}
$$

As seen before, the second term of the right-hand side above equals 0. For the remaining term set $l_k = d \cdot (k+1) + k$, identify $\left(\mathbb{R}^d\right)^{k+1} \times \mathbb{R}^k$ and $\mathbb{R}^{l_k}$ and put

$$
g_k(x, y) = \mathbf{E}\left\{ Y_0 | X_{-k}^0 = x, Y_{-k}^{-1} = y \right\}.
$$

Then, by the inequality of Jensen, $g_k \in L_2\left(\mathbb{R}^{l_k}, \mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}\right)$ and

$$
\mathbf{E}\left\{ \left| \mathbf{E}\left\{ Y_0 | X_{-k}^0, Y_{-k}^{-1} \right\} - f\left(X_{-k}^0, Y_{-k}^{-1}\right) \right|^2 \right\} = \int_{\mathbb{R}^{l_k}} |g_k(x, y) - f(x, y)|^2 \, d\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(x, y).
$$

This allows us to conclude from assumption (iii) that

$$\begin{aligned}
&\lim_{t\to\infty}\inf_{k,r\geq t}\inf_{f\in\mathcal{F}_{k,r}}\mathbf{E}\left\{\left|\mathbf{E}\{Y_0|X_{-k}^0,Y_{-k}^{-1}\}-f(X_{-k}^0,Y_{-k}^{-1})\right|^2\right\}\\
&\leq\liminf_{k\to\infty}\liminf_{r\to\infty}\inf_{f\in\mathcal{F}_{k,r}}\mathbf{E}\left\{\left|\mathbf{E}\{Y_0|X_{-k}^0,Y_{-k}^{-1}\}-f(X_{-k}^0,Y_{-k}^{-1})\right|^2\right\}=0.
\end{aligned}$$

The proof is complete. □

4.4 Proof of Lemma 8

There is no loss of generality in assuming $n\geq k+1$ throughout the proof. In order to bound

$$\mathbf{E}\left\{\left|m_{k,r}^{(i)}(X_{-k}^0,Y_{-k}^{-1})-\tilde{m}_{n,(k,r,N)}^{(i)}(X_{-n}^0,Y_{-n}^{-1};\mathcal{D}_{-n}^{-1})\right|^2\right\}$$

we add an auxiliary quantity defined by

$$m_{(k,r,N)}^{(1)}(x_{-k}^0,y_{-k}^{-1}):=\frac{\mathbf{E}\left\{T_N(Y_0)K\left(\frac{(X_{-k}^0,Y_{-k}^{-1})-(x_{-k}^0,y_{-k}^{-1})}{h_r}\right)\right\}}{\mathbf{E}\left\{K\left(\frac{(X_{-k}^0,Y_{-k}^{-1})-(x_{-k}^0,y_{-k}^{-1})}{h_r}\right)\right\}}$$

for $i=1$, by

$$m_{(k,r,N)}^{(2)}(x_{-k}^0,y_{-k}^{-1}):=\frac{\mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\left\{(X_{-k}^0,Y_{-k}^{-1})\in A_{k,r}\left((x_{-k}^0,y_{-k}^{-1})\right)\right\}}\right\}}{\mathbf{E}\left\{\mathbf{1}_{\left\{(X_{-k}^0,Y_{-k}^{-1})\in A_{k,r}\left((x_{-k}^0,y_{-k}^{-1})\right)\right\}}\right\}}$$

for $i=2$, by

$$m_{(k,r,N)}^{(3)}(x_{-k}^0,y_{-k}^{-1}):=\frac{1}{p_r}\mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\{(X_{-k}^0,Y_{-k}^{-1})\in S_{(x_{-k}^0,y_{-k}^{-1}),R_{k,r}(x_{-k}^0,y_{-k}^{-1})}\}}\right\}$$

with arbitrary $R_{k,r}(x_{-k}^0,y_{-k}^{-1})\in\left[R'_{k,r}(x_{-k}^0,y_{-k}^{-1}),R''_{k,r}(x_{-k}^0,y_{-k}^{-1})\right]$ for $i=3$ (for notation see Lemma 7) and by

$$m_{(k,r,N)}^{(4)}:=\arg\min_{f\in\mathcal{F}_{k,r}}\mathbf{E}\left\{\left|f(X_{-k}^0,Y_{-k}^{-1})-T_N\left(Y_0\right)\right|^2\right\}$$

for $i=4$.

With inequality (4) for $\delta=1$ we get

$$\begin{aligned}
&\limsup_{N\to\infty}\limsup_{n\to\infty}\mathbf{E}\left\{\left|m_{k,r}^{(i)}(X_{-k}^0,Y_{-k}^{-1})-\tilde{m}_{n,(k,r,N)}^{(i)}(X_{-n}^0,Y_{-n}^{-1};\mathcal{D}_{-n}^{-1})\right|^2\right\}\\
&\leq 2\limsup_{N\to\infty}\mathbf{E}\left\{\left|m_{k,r}^{(i)}(X_{-k}^0,Y_{-k}^{-1})-m_{(k,r,N)}^{(i)}(X_{-k}^0,Y_{-k}^{-1})\right|^2\right\}\\
&\quad+2\limsup_{N\to\infty}\limsup_{n\to\infty}\mathbf{E}\left\{\left|m_{(k,r,N)}^{(i)}(X_{-k}^0,Y_{-k}^{-1})-\tilde{m}_{n,(k,r,N)}^{(i)}(X_{-k}^0,Y_{-k}^{-1};\mathcal{D}_{-n}^{-1})\right|^2\right\}. \qquad (14)
\end{aligned}$$

Considering the first summand of (14), we have for arbitrary $k, N \in \mathbb{N}$

$$\left|m_{k,r}^{(1)}(x_{-k}^{0}, y_{-k}^{-1}) - m_{(k,r,N)}^{(1)}(x_{-k}^{0}, y_{-k}^{-1})\right| \leq \frac{\mathbf{E}\left\{|Y_0 - T_N(Y_0)|\, K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}}{\mathbf{E}\left\{K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}}$$
$$\leq \frac{\mathbf{E}\left\{|Y_0|\, K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}}{\mathbf{E}\left\{K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}}.$$

With the same argument as in the proof of Lemma 7 we have that this upper bound is square integrable.

Because of $K \geq c \cdot \mathbf{1}_{S_{0,R}}$ for suitable $c > 0, R > 0$, where $S_{0,R}$ is the ball in $(\mathbb{R}^d)^{k+1} \times \mathbb{R}^k$ centered at 0 with radius R, we have

$$\mathbf{E}\left\{K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\} \geq c \cdot \mathbf{P}_{(\mathbf{X_{-k}^{0}}, \mathbf{Y_{-k}^{-1}})}\left(S_{0,R\cdot h_r} + (x_{-k}^{0}, y_{-k}^{-1})\right) > 0 \quad (15)$$

$\mathbf{P}_{(\mathbf{X_{-k}^{0}}, \mathbf{Y_{-k}^{-1}})}$ – almost everywhere (cf., e.g., Györfi et al. (2002), pp. 499, 500). Thus we obtain by applying the dominated convergence theorem

$$\lim_{N\to\infty} \frac{\mathbf{E}\left\{|Y_0 - T_N(Y_0)|\, K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}}{\mathbf{E}\left\{K\left(\frac{(X_{-k}^{0}, Y_{-k}^{-1}) - (x_{-k}^{0}, y_{-k}^{-1})}{h_r}\right)\right\}} = 0$$

$\mathbf{P}_{(\mathbf{X_{-k}^{0}}, \mathbf{Y_{-k}^{-1}})}$ – almost everywhere.

Another application of the dominated convergence theorem yields for the first summand of (14) in case $i = 1$

$$\lim_{N\to\infty} \mathbf{E}\left\{\left|m_{k,r}^{(1)}(X_{-k}^{0}, Y_{-k}^{-1}) - m_{(k,r,N)}^{(1)}(X_{-k}^{0}, Y_{-k}^{-1})\right|^2\right\}$$
$$= \mathbf{E}\left\{\lim_{N\to\infty}\left|m_{k,r}^{(1)}(X_{-k}^{0}, Y_{-k}^{-1}) - m_{(k,r,N)}^{(1)}(X_{-k}^{0}, Y_{-k}^{-1})\right|^2\right\} = 0.$$

Similarly to the proof for $i = 1$ we can conclude

$$\limsup_{N\to\infty} \mathbf{E}\left\{\left|m_{k,r}^{(2)}(X_{-k}^{0}, Y_{-k}^{-1}) - m_{(k,r,N)}^{(2)}(X_{-k}^{0}, Y_{-k}^{-1})\right|^2\right\} = 0.$$

Further

$$\left|m_{k,r}^{(3)}(x_{-k}^{0}, y_{-k}^{-1}) - m_{(k,r,N)}^{(3)}(x_{-k}^{0}, y_{-k}^{-1})\right|$$
$$\leq \frac{\mathbf{E}\left\{|Y_0 - T_N(Y_0)|\, \mathbf{1}_{\left\{(X_{-k}^{0}, Y_{-k}^{-1}) \in S_{(x_{-k}^{0}, y_{-k}^{-1}), R_{k,r}(x_{-k}^{0}, y_{-k}^{-1})}\right\}}\right\}}{p_r}$$

for all (x^0_{-k}, y^{-1}_{-k}), thus by the dominated convergence theorem

$$\lim_{N\to\infty} \mathbf{E}\left\{\left|m^{(3)}_{k,r}\left(X^0_{-k}, Y^{-1}_{-k}\right) - m^{(3)}_{k,r,N}\left(X^0_{-k}, Y^{-1}_{-k}\right)\right|^2\right\} = 0.$$

By definition of $m^{(4)}_{k,r,N}$, $m^{(4)}_{k,r}$, respectively, it holds for arbitrary $\delta > 0$ that

$$\begin{aligned}
&\mathbf{E}\left\{\left|m^{(4)}_{k,r}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\}\\
&\leq \mathbf{E}\left\{\left|m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\}\\
&\leq (1+\delta)\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k}) - T_N\left(Y_0\right)\right|^2\right\} + \left(1+\frac{1}{\delta}\right)\cdot \mathbf{E}\left\{\left|Y_0 - T_N\left(Y_0\right)\right|^2\right\}\\
&\leq (1+\delta)\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - T_N\left(Y_0\right)\right|^2\right\} + \left(1+\frac{1}{\delta}\right)\cdot \mathbf{E}\left\{\left|Y_0 - T_N\left(Y_0\right)\right|^2\right\}\\
&\leq (1+\delta)^2\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\} + (2+\delta)\cdot\left(1+\frac{1}{\delta}\right)\cdot \mathbf{E}\left\{\left|Y_0 - T_N\left(Y_0\right)\right|^2\right\}.
\end{aligned}$$

As $\delta > 0$ was arbitrary, we obtain

$$\lim_{N\to\infty} \mathbf{E}\left\{\left|m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\} = \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\} \tag{16}$$

As in the first step of the proof of Lemma 6 we have for arbitrary N

$$\begin{aligned}
&\mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\}\\
&\leq \mathbf{E}\left\{\left|\frac{m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) + m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k})}{2} - Y_0\right|^2\right\}\\
&= \frac{1}{2}\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\} + \frac{1}{2}\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k}) - Y_0\right|^2\right\}\\
&\qquad - \frac{1}{4}\cdot \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k})\right|^2\right\}.
\end{aligned}$$

Because of (16) this implies

$$\limsup_{N\to\infty} \mathbf{E}\left\{\left|m^{(4)}_{(k,r)}(X^0_{-k}, Y^{-1}_{-k}) - m^{(4)}_{(k,r,N)}(X^0_{-k}, Y^{-1}_{-k})\right|^2\right\} = 0.$$

Considering the second summand of decomposition (14) for $i = 1$ we have by the ergodic theorem:

$$\frac{1}{n-k}\sum_{i=-n+k-1}^{-2} T_N(Y_{i+1}) K\left(\frac{(X^{i+1}_{i-k+1}, Y^{i}_{i-k+1}) - (x^0_{-k}, y^{-1}_{-k})}{h_r}\right)$$

$$\to \mathbf{E}\left\{T_N(Y_0)K\left(\frac{(X_{-k}^0, Y_{-k}^{-1}) - (x_{-k}^0, y_{-k}^{-1})}{h_r}\right)\right\} \quad a.s. \tag{17}$$

and

$$\frac{1}{n-k}\sum_{i=-n+k-1}^{-2} K\left(\frac{(X_{i-k}^{i+1}, Y_{i-k}^{i}) - (x_{-k}^0, y_{-k}^{-1})}{h_r}\right)$$
$$\to \mathbf{E}\left\{K\left(\frac{(X_{-k}^0, Y_{-k}^{-1}) - (x_{-k}^0, y_{-k}^{-1})}{h_r}\right)\right\} \quad a.s. \tag{18}$$

Because of the continuity of the kernel function K and $K(v) \to 0$ ($\|v\| \to \infty$) we can use an ergodic theorem applied to random variables with values in the separable Banach space of continuous functions vanishing at infinity with supremum norm and get that the almost sure convergence above is uniformly with respect to (x_{-k}^0, y_{-k}^{-1}) (cf., e.g., Krengel (1985), Chapter 4, Theorem 2.1).

Let $\epsilon > 0$ be arbitrary and define

$$S_\epsilon := \left\{(x_{-k}^0, y_{-k}^{-1}) \in (\mathbb{R}^d)^{k+1} \times \mathbb{R}^k : \mathbf{E}\left\{K\left(\frac{(X_{-k}^0, Y_{-k}^{-1}) - (x_{-k}^0, y_{-k}^{-1})}{h_r}\right)\right\} \geq \epsilon\right\}.$$

From (15) we know

$$\mathbf{P}_{(\mathbf{X_{-k}^0}, \mathbf{Y_{-k}^{-1}})}(S_\epsilon) \to 1 \qquad (\epsilon \to 0).$$

In addition we can conclude

$$\sup_{(x_{-k}^0, y_{-k}^{-1}) \in (\mathbb{R}^d)^{k+1} \times \mathbb{R}^k,\ (x_{-k}^0, y_{-k}^{-1}) \in S_\epsilon} \left|\tilde{m}_{n,(k,r,N)}^{(1)}(x_{-k}^0, y_{-k}^{-1}; \mathcal{D}_{-n}^{-1}) - m_{(k,r,N)}^{(1)}(x_{-k}^0, y_{-k}^{-1})\right| \to 0 \quad a.s.$$

By the boundedness of $m_{(k,r,N)}^{(1)}(\cdot)$ and $\tilde{m}_{n,(k,r,N)}^{(1)}(\cdot)$ by N and the dominated convergence theorem we get for arbitrary $k, r, N \in \mathbb{N}$ and $\epsilon > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{E}\left\{\left|m_{(k,r,N)}^{(1)}(X_{-k}^0, Y_{-k}^{-1}) - \tilde{m}_{n,(k,r,N)}^{(1)}(X_{-k}^0, Y_{-k}^{-1}; \mathcal{D}_{-n}^{-1})\right|^2\right\} \\
&\leq \limsup_{\epsilon\to 0}\limsup_{n\to\infty}\Bigg(4N^2 \cdot \mathbf{E}\left\{\mathbf{1}_{S_\epsilon^c}(X_{-k}^0, Y_{-k}^{-1})\right\} \\
&\quad + \mathbf{E}\Big\{\sup_{(x_{-k}^0, y_{-k}^{-1}) \in \mathbb{R}^{d(k+1)+k}:\ (x_{-k}^0, y_{-k}^{-1}) \in S_\epsilon} \Big|\tilde{m}_{n,(k,r,N)}^{(1)}(x_{-k}^0, y_{-k}^{-1}; \mathcal{D}_{-n}^{-1}) - \\
&\qquad\qquad m_{(k,r,N)}^{(1)}(x_{-k}^0, y_{-k}^{-1})\Big|^2\Big\}\Bigg) \\
&\leq \limsup_{\epsilon\to 0} 4N^2 \mathbf{P}_{(\mathbf{X_{-k}^0}, \mathbf{Y_{-k}^{-1}})}(S_\epsilon^c) = 0.
\end{aligned}$$

Next we consider the second term for $i = 2$ for fixed $k, r \in \mathbb{N}$. By the ergodic theorem we have

$$\frac{1}{n-k}\sum_{i=-n+k-1}^{-2} T_N(Y_{i+1})\mathbf{1}_{\left\{(X_{i-k+1}^{i+1},Y_{i-k+1}^{i})\in A_{k,r}\left(x_{-k}^{0},y_{-k}^{-1}\right)\right\}}$$

$$\to \mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\left\{(X_{-k}^{0},Y_{-k}^{-1})\in A_{k,r}\left(x_{-k}^{0},y_{-k}^{-1}\right)\right\}}\right\} \quad a.s. \quad (n\to\infty)$$

and

$$\frac{1}{n-k}\sum_{i=-n+k-1}^{-2} \mathbf{1}_{\left\{(X_{i-k+1}^{i+1},Y_{i-k+1}^{i})\in A_{k,r}\left(x_{-k}^{0},y_{-k}^{-1}\right)\right\}}$$

$$\to \mathbf{E}\left\{\mathbf{1}_{\left\{(X_{-k}^{0},Y_{-k}^{-1})\in A_{k,r}\left(x_{-k}^{0},y_{-k}^{-1}\right)\right\}}\right\} \quad a.s. \quad (n\to\infty).$$

With the third assumption of Theorem 1 b) we get that for k, r fixed the almost sure convergence above is uniformly with respect to (x_{-k}^{0},y_{-k}^{-1}) on any compact set (since for any compact set C we have that $\left|\{A_{k,r}(x_{-k}^{0},y_{-k}^{-1}),\ (x_{-k}^{0},y_{-k}^{-1})\in C\}\right|$ is finite). Let $L>0$ and $\epsilon>0$ be arbitrary and set

$$S_{L,r,\epsilon} := \left\{(x_{-k}^{0},y_{-k}^{-1})\in[-L,L]^{d\cdot(k+1)+k} : \mathbf{E}\left\{\mathbf{1}_{\left\{(X_{-k}^{0},Y_{-k}^{-1})\in A_{k,r}(x_{-k}^{0},y_{-k}^{-1})\right\}}\right\}\geq\epsilon\right\}.$$

Because $S_{L,r,\epsilon_1}\subseteq S_{L,r,\epsilon_2}$ for $\epsilon_1\geq\epsilon_2$ it holds $S_{L,r,\epsilon_1}^c\supseteq S_{L,r,\epsilon_2}^c$ and

$$\bigcap_{\epsilon>0} S_{L,r,\epsilon}^c\cap[-L,L]^{d\cdot(k+1)+k}$$
$$=\left\{(x_{-k}^{0},y_{-k}^{-1})\in[-L,L]^{d\cdot(k+1)+k} : \mathbf{E}\left\{\mathbf{1}_{\left\{(X_{-k}^{0},Y_{-k}^{-1})\in A_{k,r}(x_{-k}^{0},y_{-k}^{-1})\right\}}\right\}=0\right\}.$$

By assumption (iii) we know that $\bigcap_{\epsilon>0} S_{L,r,\epsilon}^c\cap[-L,L]^{d\cdot(k+1)+k}$ is contained in the union of finitely many sets from $\mathcal{P}_{k,r}$ with $\mathbf{P}_{(X_{-k}^{0},Y_{-k}^{-1})}$-measure zero. As a consequence we have

$$\mathbf{P}_{(X_{-k}^{0},Y_{-k}^{-1})}\left(S_{L,r,\epsilon}^c\cap[-L,L]^{d\cdot(k+1)+k}\right)\to 0 \qquad (\epsilon\to 0).$$

From the relations above we can conclude in addition

$$\sup_{\substack{(x_{-n}^{0},y_{-n}^{0})\in\mathbb{R}^{d(n+1)}\times\mathbb{R}^{n}:\\ (x_{-k}^{0},y_{-k}^{-1})\in S_{L,r,\epsilon}}}\left|\tilde{m}_{n,(k,r,N)}^{(2)}(x_{-n}^{0},y_{-n}^{-1};\mathcal{D}_{-n}^{-1})-m_{(k,r,N)}^{(2)}(x_{-n}^{0},y_{-n}^{-1})\right|\to 0 \quad a.s.$$

With the boundedness of $m_{(k,r,N)}^{(2)}(\cdot)$ and $\tilde{m}_{n,(k,r,N)}^{(2)}(\cdot)$ by N and the dominated convergence theorem we get for arbitrary $k,r,N\in\mathbb{N}$,

$$\limsup_{n\to\infty}\mathbf{E}\left\{\left|m_{(k,r,N)}^{(2)}(X_{-k}^{0},Y_{-k}^{-1})-\tilde{m}_{n,(k,r,N)}^{(2)}(X_{-n}^{0},Y_{-n}^{-1};X_0,\mathcal{D}_{-n}^{-1})\right|^2\right\}$$

$$
\begin{aligned}
&\leq \limsup_{L\to\infty}\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\Bigg(4N^2\cdot\mathbf{E}\left\{\mathbf{1}_{S^c_{L,r,\epsilon}\cap[-L,L]^{d\cdot(k+1)+k}}(X^0_{-k},Y^{-1}_{-k})\right\}\\
&\qquad\qquad +4N^2\cdot\mathbf{E}\left\{\mathbf{1}_{\left([-L,L]^{d\cdot(k+1)+k}\right)^c}(X^0_{-k},Y^{-1}_{-k})\right\}\\
&\quad+\mathbf{E}\Big\{\sup_{\substack{(x^0_{-n},y^{-1}_{-n})\in\mathbb{R}^{d(n+1)}\times\mathbb{R}^n:\\ (x^0_{-k},y^{-1}_{-k})\in S_{L,r,\epsilon}}}\left|\tilde{m}^{(2)}_{n,(k,r,N)}(x^0_{-n},y^{-1}_{-n};\mathcal{D}^{-1}_{-n})-m^{(2)}_{(k,r,N)}(x^0_{-n},y^{-1}_{-n})\right|^2\Big\}\Bigg)\\
&=0.
\end{aligned}
$$

Now we consider the second term of decomposition (14) for $i=3$. We shall use arguments of Györfi, Udina and Walk (2008). It suffices to show

$$
\mathbf{E}\left\{\left|\tilde{m}^{(3)}_{n,(k,r,N)}(X^0_{-k},Y^{-1}_{-k};\mathcal{D}_n)-m^{(3)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\right\}\to 0 \quad (n\to\infty) \tag{19}
$$

for each $N>0$. For $\epsilon>0$ arbitrary and $R^*=R^*(\epsilon)>0$ sufficiently large the left-hand side above is bounded from above by

$$
\mathbf{E}\left\{\left|\tilde{m}^{(3)}_{n,(k,r,N)}(X^0_{-k},Y^{-1}_{-k};\mathcal{D}_n)-m^{(3)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\cdot 1_{S_{0,R^*}}((X^0_{-k},Y^{-1}_{-k}))\right\}+\epsilon,
$$

and thus because of boundedness and the dominated convergence theorem it suffices to show

$$
\left|\tilde{m}^{(3)}_{n,(k,r,N)}(X^0_{-k},Y^{-1}_{-k};\mathcal{D}_n)-m^{(3)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\cdot 1_{S_{0,R^*}}((X^0_{-k},Y^{-1}_{-k}))\to 0 \quad a.s.
$$

for each $R^*>0$. The latter is obtained if for each $R^*>0$ one can show

$$
\tilde{m}^{(3)}_{n,(k,r,N)}(x^0_{-k},y^{-1}_{-k};\mathcal{D}_n)\to m^{(3)}_{(k,r,N)}(x^0_{-k},y^{-1}_{-k}) \quad a.s. \tag{20}
$$

uniformly with respect to $(x^0_{-k},y^{-1}_{-k})\in S_{0,R^*}$.

As an auxiliary result we state a.s.

$$
\begin{aligned}
&\frac{1}{n-k}\sum_{i=1}^{n-k}\mathbf{1}_{S_{0,1}}\left(\frac{\left(X^{-i}_{-i-k},Y^{-i-1}_{-i-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)}{R}\right)\\
&\to\mathbf{E}\left\{\mathbf{1}_{S_{0,1}}\left(\frac{\left(X^0_{-k},Y^{-1}_{-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)}{R}\right)\right\}=\mathbf{P}_{(X^0_{-k},Y^{-1}_{-k})}\left(S_{(x^0_{-k},y^{-1}_{-k}),R}\right)
\end{aligned} \tag{21}
$$

uniformly with respect to $(x^0_{-k},y^{-1}_{-k})\in S_{0,R^*}$ and $R\in(0,\infty)$.

First we show (21) for $R\in[R_1,R_2]$ with arbitrary $R_2>R_1>0$. Let $K_{1,\epsilon},K_{2,\epsilon}$ $(0<\epsilon<\frac{1}{2})$ be continuous kernel functions satisfying $\mathbf{1}_{S_{0,1-\epsilon}}\leq K_{1,\epsilon}\leq\mathbf{1}_{S_{0,1}}$, $\mathbf{1}_{S_{0,1}}\leq K_{2,\epsilon}\leq\mathbf{1}_{S_{0,1+\epsilon}}$. Let $\mathbb{R}^{(k+1)\cdot d+k}$ be endowed with the Euclidean norm $\|\cdot\|$ and $R=\mathbb{R}^{(k+1)\cdot d+k}\times[R_1,R_2]$ be endowed with the Euclidean metric. Further let $C^*(R)$ be the separable Banach space of continuous real-valued functions g on R satisfying $g(x,y) \to 0$ ($\|x\| \to \infty$) uniformly with respect to $y \in [R_1, R_2]$, endowed with the maximum norm and the corresponding Borel σ-algebra. Then as in the context of (17) and (18), by the ergodic theorem in Krengel (1985), Chapter 4, Theorem 2.1, one obtains for $l \in \{1,2\}$

$$\frac{1}{n-k}\sum_{i=1}^{n-k} K_{l,\epsilon}\left(\frac{\left(X_{-i-k}^{-i}, Y_{-i-k}^{i-1}\right) - \left(x_{-k}^{0}, y_{-k}^{-1}\right)}{R}\right)$$
$$\to \mathbf{E}\left\{K_{l,\epsilon}\left(\frac{\left(X_{-k}^{0}, Y_{-k}^{-1}\right) - \left(x_{-k}^{0}, y_{-k}^{-1}\right)}{R}\right)\right\} \quad a.s.$$

uniformly with respect to $(x_{-k}^0, y_{-k}^{-1}) \in \mathbb{R}^{(k+1)\cdot d+k}$, especially $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$, and $R_1 \le R \le R_2$.

But

$$\mathbf{E}\left\{K_{l,\epsilon}\left(\frac{\left(X_{-k}^{0}, Y_{-k}^{-1}\right) - \left(x_{-k}^{0}, y_{-k}^{-1}\right)}{R}\right)\right\} \to \mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}\left(S_{(x_{-k}^0, y_{-k}^{-1}),R}\right) \quad (\epsilon \to 0)$$

uniformly with respect to $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$ and $R_1 \le R \le R_2$ by the bounds of K and assumption $(A1)$. This yields (21) for $R \in [R_1, R_2]$.

Next we show that it suffices to show (21) for $R \in [R_1, \infty)$ with arbitrary $R_1 > 0$. For, in this situation, choosing an arbitrary $\epsilon > 0$, because of $(A1)$ an $R' > 0$ exists such that

$$\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(S_{(x_{-k}^0, y_{-k}^{-1}),R}) \le \epsilon \quad \text{for all } (x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*} \text{ and all } R \in (0, R'].$$

Further the left-hand side of (21) is majorized by the left-hand side of (21) for $R = R'$ which $a.s.$ converges to

$$\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(S_{(x_{-k}^0, y_{-k}^{-1}),R'}) \le \epsilon$$

uniformly with respect to $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$. Thus $a.s.$ for n sufficiently large the left-hand side of (21) and the right-hand side of (21) differ at most by $2 \cdot \epsilon$ for $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$. Similarly, by a minorization argument, it even suffices to show (21) for $R \in [R_1, R_2]$ with arbitrary fixed $R_2 < \infty$, which completes the proof of (21).

Set

$$\begin{aligned} Z_n\left(x_{-k}^0, y_{-k}^{-1}\right) \;=\; & \Big\|\big(\text{the } \lfloor p_r n \rfloor\text{-th NN of } \left(x_{-k}^0, y_{-k}^{-1}\right) \text{ among} \\ & (X_{-n}^{-n+k}, Y_{-n}^{-n+k-1}), \ldots, (X_{-k-1}^{-1}, Y_{-k-1}^{-2})\big) - \left(x_{-k}^0, y_{-k}^{-1}\right)\Big\|. \end{aligned}$$

We notice

$$\left\{Z_n\left(x_{-k}^0, y_{-k}^{-1}\right) > R\right\} = \left\{\frac{1}{n-k}\sum_{i=1}^{n-k} \mathbf{1}_{\left\{\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) \in S_{(x_{-k}^0, y_{-k}^{-1}),R}\right\}} < \frac{\lfloor p_r n \rfloor}{n-k}\right\},$$

$0 < R < \infty$, and $\frac{\lfloor p_r n \rfloor}{n} \to p_r$. Let $0 < \epsilon < \min\{p_r, 1-p_r\}$. Choose $R_{k,r}^{-\epsilon}(x_{-k}^0, y_{-k}^{-1})$ as the largest R satisfying $\mathbf{P}_{(X_{-k}^0, Y_{-k}^{-1})}(S_{(x_{-k}^0, y_{-k}^{-1}),R}) = p_r - \epsilon$ and $R_{k,r}^{\epsilon}(x_{-k}^0, y_{-k}^{-1})$ as the smallest R satisfying $\mathbf{P}_{(X_{-k}^0,Y_{-k}^{-1})}(S_{(x_{-k}^0,y_{-k}^{-1}),R}) = p_r + \epsilon$. Then by (21) a random index N_ϵ independent of $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$ exists such that almost surely for all $n \geq N_\epsilon$ the above event appears if $R \leq R_{k,r}^{-\epsilon}(x_{-k}^0, y_{-k}^{-1})$ and does not appear if $R \geq R_{k,r}^{\epsilon}(x_{-k}^0, y_{-k}^{-1})$. Because of $(A1)$ we have

$$R_{k,r}^{-\epsilon}(x_{-k}^0, y_{-k}^{-1}) \uparrow R'_{k,r}(x_{-k}^0, y_{-k}^{-1}) \quad \text{and} \quad R_{k,r}^{\epsilon}(x_{-k}^0, y_{-k}^{-1}) \downarrow R''_{k,r}(x_{-k}^0, y_{-k}^{-1})$$

uniformly with respect to $(x_{-k}^0, y_{-k}^{-1}) \in S_{0,R^*}$ for $\epsilon \to 0$. Therefore the distance between $Z_n\left(x_{-k}^0, y_{-k}^{-1}\right)$ and the interval

$$\left[R'_{k,l}\left(x_{-k}^0, y_{-k}^{-1}\right), R''_{k,l}\left(x_{-k}^0, y_{-k}^{-1}\right)\right]$$

converges to zero uniformly with respect to (x_{-k}^0, y_{-k}^{-1}). We notice that $\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right)$ is among the $\lfloor p_r n \rfloor$ NNs of $\left(x_{-k}^0, y_{-k}^{-1}\right)$ in $\left(X_{-n}^{-n+k}, Y_{-n}^{-n+k-1}\right), \ldots, \left(X_{-k+1}^{-1}, Y_{-k+1}^{-2}\right)$ if and only if

$$\left\|\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq Z_n(x_{-k}^0, y_{-k}^{-1}).$$

Then a.s. for $n \geq N_\epsilon$ we have the implications

$$\begin{aligned}
&\left\|\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,l}^{(-\epsilon)}(x_{-k}^0, y_{-k}^{-1}) \\
&\quad \Rightarrow \left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) \text{ is among the } \lfloor p_r n \rfloor \text{ NNs of } (x_{-k}^0, y_{-k}^{-1}) \text{ in} \\
&\qquad\qquad\qquad\qquad \left(X_{-n}^{-n+k}, Y_{-n}^{-n+k-1}\right), \ldots, \left(X_{-k-1}^{-1}, Y_{-k-1}^{-2}\right) \\
&\quad \Rightarrow \left\|\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,l}^{(\epsilon)}(x_{-k}^0, y_{-k}^{-1}).
\end{aligned}$$

We introduce the sets

$$\begin{aligned}
&\bar{J}_{n,(x_{-k}^0, y_{-k}^{-1})}^{(k,r)} \\
&= \left\{i:\ -n+k \leq i \leq -1,\ \left\|\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,r}^{(-\epsilon)}(x_{-k}^0, y_{-k}^{-1})\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\tilde{J}_{n,(x_{-k}^0, y_{-k}^{-1})}^{(k,r)} \\
&= \left\{i:\ -n+k \leq i \leq -1,\ \left\|\left(X_{-i-k}^{-i}, Y_{-i-k}^{-i-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,r}^{(\epsilon)}(x_{-k}^0, y_{-k}^{-1})\right\}.
\end{aligned}$$

Without loss of generality assume $Y_i \geq 0$. The auxiliary result (21) also holds if in the left-hand side and in the right-hand side the uniformly bounded random variables $T_N(Y_i)$ and $T_N(Y_0)$, respectively, are inserted as factors. By (21) in both versions, we obtain a.s.

$$\frac{\frac{1}{n-k}\sum_{i \in \bar{J}_{n,(x_{-k}^0, y_{-k}^{-1})}^{(k,r)}} T_N(Y_i)}{\frac{1}{n-k}\left|\tilde{J}_{n,(x_{-k}^0, y_{-k}^{-1})}^{(k,r)}\right|} \to \frac{\mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\left\{\left\|\left(X_{-k}^0, Y_{-k}^{-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,r}^{(-\epsilon)}(x_{-k}^0, y_{-k}^{-1})\right\}}\right\}}{\mathbf{P}\left\{\left\|\left(X_{-k}^0, Y_{-k}^{-1}\right) - \left(x_{-k}^0, y_{-k}^{-1}\right)\right\| \leq R_{k,r}^{(\epsilon)}(x_{-k}^0, y_{-k}^{-1})\right\}}$$

(22)

and

$$\frac{\frac{1}{n-k}\sum_{i\in \tilde{J}^{(k,r)}_{n,(x^0_{-k},y^{-1}_{-k})}} T_N(Y_i)}{\frac{1}{n-k}\left|\tilde{J}^{(k,r)}_{n,(x^0_{-k},y^{-1}_{-k})}\right|} \to \frac{\mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\left\{\left\|\left(X^0_{-k},Y^{-1}_{-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)\right\|\le R^{(\epsilon)}_{k,r}(x^0_{-k},y^{-1}_{-k})\right\}}\right\}}{\mathbf{P}\left\{\left\|\left(X^0_{-k},Y^{-1}_{-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)\right\|\le R^{(-\epsilon)}_{k,r}(x^0_{-k},y^{-1}_{-k})\right\}} \tag{23}$$

uniformly with respect to $(x^0_{-k}, y^{-1}_{-k}) \in S_{0,R^*}$. For $\epsilon \to 0$ the right-hand sides of (22) and (23) converge to

$$\frac{\mathbf{E}\left\{T_N(Y_0)\mathbf{1}_{\left\{\left\|\left(X^0_{-k},Y^{-1}_{-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)\right\|\le R_{k,r}(x^0_{-k},y^{-1}_{-k})\right\}}\right\}}{\mathbf{P}\left\{\left\|\left(X^0_{-k},Y^{-1}_{-k}\right)-\left(x^0_{-k},y^{-1}_{-k}\right)\right\|\le R_{k,r}(x^0_{-k},y^{-1}_{-k})\right\}} = m^{(3)}_{(k,r,N)}(x^0_{-k},y^{-1}_{-k}) \tag{24}$$

uniformly with respect to $\left(x^0_{-k}, y^{-1}_{-k}\right)$, because the denominators in the right-hand sides of (22) and (23) and in the left-hand side of (24) are $p_r+\epsilon$, $p_r-\epsilon$ and p_r, respectively, and because the corresponding numerators in (22) and (23) differ from the numerator in (24) at most by $N\cdot\epsilon$. Since $\tilde{m}^{(3)}_{n,(k,r,N)}(x^0_{-k}, y^{-1}_{-k}; \mathcal{D}_n)$ is included between the left-hand sides from (22) and (23) we can conclude from (22) and (23) that (20) and thus also (19) holds.

For $i=4$, denote by A the support of $\mathbf{P}_{(X^0_{-k},Y^{-1}_{-k})}$. As $n \ge k+1$, by definition

$$\begin{aligned}
&\mathbf{E}\left\{\left|\tilde{m}^{(4)}_{n,(k,r,N)}\left(X^0_{-k},Y^{-1}_{-k};\mathcal{D}^{-1}_{-n}\right)-m^{(4)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\right\}\\
&=\mathbf{E}\left\{\left|\bar{m}^{(4)}_{n,(k,r,N)}\left(X^0_{-k},Y^{-1}_{-k};\mathcal{D}^{-1}_{-n}\right)-m^{(4)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\right\}\\
&=\mathbf{E}\left\{\left|\bar{m}^{(4)}_{n,(k,r,N)}\left(X^0_{-k},Y^{-1}_{-k};\mathcal{D}^{-1}_{-n}\right)-m^{(4)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\mathbb{1}_A(X^0_{-k},Y^{-1}_{-k})\right\}\\
&\le\mathbf{E}\left\{\|\bar{m}^{(4)}_{n,(k,r,N)}-m^{(4)}_{(k,r,N)}\|^2_{\infty,supp(\mathbf{P}_{(X^0_{-k},Y^{-1}_{-k})})}\right\}.
\end{aligned}$$

By definition of $\bar{m}^{(4)}_{n,(k,r,N)}$ we can apply Lemma 6 (where $f^* = m^{(4)}_{(k,r,N)}$). Moreover, the boundedness of functions in $\mathcal{F}_{k,r}$ allows us to apply Lebesgues dominated convergence theorem which yields

$$\limsup_{n\to\infty}\mathbf{E}\left\{\left|\tilde{m}^{(4)}_{n,(k,r,N)}\left(X^0_{-k},Y^{-1}_{-k};\mathcal{D}^{-1}_{-n}\right)-m^{(4)}_{(k,r,N)}(X^0_{-k},Y^{-1}_{-k})\right|^2\right\}=0.$$

The proof is complete. □

References

[1] Algoet, P. (1992). Universal schemes for predicition, gambling and portfolio selection. *Ann. Probab.* **20**, pp. 901–941.

[2] Bailey, D. H. (1976). Sequential schemes for classifying and predicting ergodic processes. PhD thesis, Stanford University.

[3] Cesa-Bianchi, N., and Lugosi, G. (2006). *Prediction, Learning, and Games.* Cambridge University Press, New York.

[4] Collomb, G. (1985). Nonparametric time series analysis and prediction: uniform almost sure convergence. *Statistics* **2**, pp. 197–307.

[5] Cover, T. (1975). Open problems in information theory. *IEEE Joint Workshop on Information Theory*, pp. 35–36. IEEE Press, New York.

[6] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[7] Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric curve estimation from time series.* Springer Verlag, New York.

[8] Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics, Springer, New York.

[9] Györfi, L., and Lugosi, G. (2000). Strategies for sequential prediction of stationary time series. *Internat. Ser. Oper. Res. Management Sci.* **46**, pp. 225–248.

[10] Györfi, L., and Lugosi, G. (2002). Strategies for sequential prediction of stationary time series. In *Modelling Uncertainty: An Examination of Its Theory, Methods and Applications*, M. Dror, P. L'Ecuyer, and F. Szidarovsky, eds. Kluwer Academic Publishers, Norwell, MA, 2002, pp. 225–248.

[11] Györfi, L., Lugosi, G., and Morvai, G. (1999). A simple randomized algorithm for sequential prediction of ergodic time series. *IEEE Trans. Inform. Theory* **45**, pp. 2642–2650.

[12] Györfi, L., Lugosi, G., and Udina, F. (2006). Nonparametric kernel-based sequential investment strategies. *Mathematical Finance* **16**, pp. 337–357.

[13] Györfi, L., Morvai, G., and Yakowitz, S. (1998). Limits to consistent on-line forecasting for ergodic time series. *IEEE Trans. Inform. Theory* **44**, pp. 886–892.

[14] Györfi, L., and Ottucsak, G. (2007). Sequential prediction of unbounded stationary time series. *IEEE Trans. Inform. Theory* **53**, pp. 1866–1872.

[15] Györfi, L., and Schäfer, D. (2003). Nonparametric prediction. In *Advances of Learning Theory: Methods, Models and Applications.* J. A. Suykens, G, Horváth, S. Basu, C. Micchelli, and J. Vandevalle, eds. IOS Press, NATO Science Series, 2003, pp. 339-354.

[16] Györfi, L., Udina, F., and Walk, H. (2008). Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics & Decisions* **26**, pp. 145–157.

[17] Jones, D., Kohler, M., and Walk, H. (2011). Weakly universally consistent forecasting of stationary and ergodic time series. To appear in *IEEE Trans. Inform. Theory.*

[18] Kohler, M., and Walk, H. (2010). On optimal exercising of American options in discrete time for stationary and ergodic data. Submitted for publication.

[19] Krengel, U. (1985). *Ergodic Theorems.* de Gruyter, Berlin, New York.

[20] Loève, M. (1977). *Probability Theory I*, *4th* ed. Springer Verlag, New York, Berlin.

[21] Marton, K., and Shields, P.C. (1994). Entropy and the consistent estimation of joint distributions. *Ann. Prob.* **22**, pp. 960–977.

[22] Morvai, G., and Weiss, B. (2005). Forward estimation for ergodic time series. *Ann. Inst. H. Poincare Probab. Statist.* **41**, pp. 859–870.

[23] Morvai, G., and Weiss, B. (2011). Nonparametric sequential prediction for stationary processes. *Ann. Prob.* **39**, pp. 1137-1160.

[24] Morvai, G., Yakowitz, S., and Algoet, P. (1997). Weakly convergent nonparametric forecasting of stationary time series. *IEEE Trans. Inform. Theory* **43**, pp. 483–498.

[25] Morvai, G., Yakowitz, S., and Györfi, L. (1996). Nonparametric inference for ergodic stationary time series. *Ann. Statist.* **24**, pp. 370–379.

[26] Ornstein, D. (1978). Guessing the next output of a stationary process. *Israel J. Math.* **30**, pp. 292–296.

[27] Ryabko, B. Y. (1988). Prediction of random sequences and universal coding. *Problems Inform. Transmission* **24**, pp. 87–96.

[28] Weiss, B. (2000). Single Orbit Dynamics. CBMS, American Mathematical Society.