ALMOST SURE CONVERGENCE UNDER ESTIMATING CONDITIONAL MEAN BASED ON DEPENDENT DATA

Kirill Chernyshov

Institute of Control Sciences
65 Profsoyuznaya, Moscow 117997, Russia
E-mail: myau@ipu.rssi.ru

Abstract. The paper is devoted to establishing strong consistency of estimates of nonlinear characteristics of dynamic stochastic systems. To describe the shape of the nonlinearities, regression functions, i.e. conditional expectations of a random variable with respect to another one, are used. In turn, the nonlinear regression functions are estimated by algorithms using the kernel-type approaches, which are suitable under fairly mild assumptions with respect to the system description. Within the approach, the key issue of the present paper is considering a case of mutually dependent observations in contrast to conventional nonparametric approaches based on regression estimates, which impose rather restrictive limitations on sampled data, e.g. mutual independence, various mixing conditions, etc., while such assumptions are not always acceptable within dynamic system considerations.

Keywords: Convergence analysis, Consistency, Nonlinear systems, Non-parametric identification, Recursive estimation, Sampled data.

1. PROBLEM STATEMENT

Let a dynamic stochastic discrete-time system be described by the following extended state-space equation

$$\mathrm{X}[t] = \Xi\left(t; \mathrm{X}[t-1], e[t]\right), \tag{1}$$

with $\mathrm{X}[t] = \left(\chi_1[t], \chi_2[t], \ldots, \chi_{\mathrm{N}}[t]\right)^T$ standing for the extended state vector involving both the system state variables as well as the system input and output variables; $e[t]$ standing for a white-noise process, i.e. $e[t]$ and $e[s]$ are mutually independent. The process $e[t]$ values are assumed to be bounded with probability 1 by a constant C, i.e. $\left|e[t]\right| < C$ almost surely, while the magnitude of C does not matter. The only requirement is that the constant C exists and is less than infinity. Also, the function $\left\| \Xi(\cdot;\cdot,\cdot) \right\|$ is assumed to be bounded in its arguments. These two assumptions thus imply the components of the state vector $\mathrm{X}[t]$ to be bounded in modulus almost surely. And, finally, system (1) is assumed to be exponentially stable.

The aim of the paper is to consider convergence properties of a recursive estimate of the regression function

$$m_{ij}(X) = \mathbf{M}\left\{ \chi_i[t] \Big/ \chi_j[t] = X \right\}. \tag{2}$$

Here $\mathbf{M}\left\{ \cdot \big/ \cdot \right\}$ stands for the conditional mathematical expectation. As such an estimate, the following kernel-type recursive algorithm is considered (represented by formulae (3) and (4))

$$\begin{aligned} m_{ijX}[t] &= m_{ijX}[t-1] + \\ R_{jX}^{-1}[t]&\left\{\chi_i[t] - m_{ijX}[t-1]\right\}\times, \\ &\times h^{-1}[t] K\left(\frac{X - \chi_j[t]}{h[t]}\right), \end{aligned} \tag{3}$$

$$R_{jX}[t] = R_{jX}[t-1] + h^{-1}[t] K\left(\frac{X - \chi_j[t]}{h[t]}\right). \quad (4)$$

The ratio $0/0$ is considered as 0. In the algorithms, $K(\cdot)$ is a Borel kernel function, $\{h[t]\}$ is a positive number sequence such as

$$\lim_{t\to\infty} h[t] = h > 0 . \quad (5)$$

The bounded and nonnegative Borel kernel $K(u)$ satisfies the following two conditions

$$0 < \int_{-\infty}^{\infty} K(u)du , \ \lim_{u\to\infty}\left\{|u| \cdot K(u)\right\} = 0 . \quad (6)$$

There exists a wide class of kernels satisfying conditions (6). In particular, the following functions belong to this class:

$$K(u) = \begin{cases} 1 & for \ |u| \le 1 \\ 0 & otherwise \end{cases},$$

$$K(u) = \begin{cases} 1-|u| & for \ |u| \le 1 \\ 0 & otherwise \end{cases},$$

$$K(u) = \begin{cases} 1-u^2 & for \ |u| \le 1 \\ 0 & otherwise \end{cases},$$

$$K(u) = e^{-u^2} , \ K(u) = e^{-|u|},$$

$$K(u) = \frac{1}{1+u^2} , \text{etc.} \quad (7)$$

Remark. For a conventional kernel estimate, $h[t]$ tends to zero,

$$\lim_{t\to\infty} h[t] \to 0 . \quad (8)$$

The main difference between conventional recursive kernel regression estimators and algorithm (3) to (5) just lies in using condition (5) rather than (8).

Another suitable algorithm to estimate regression function (2) is even simpler and has the form

$$m_{ijX}[t] = m_{ijX}[t-1] +$$
$$R_{jX}^{-1}[t]\{\chi_i[t] - m_{ijX}[t-1]\}K\left(\frac{X - \chi_j[t]}{h[t]}\right), \quad (9)$$

$$R_{jX}[t] = R_{jX}[t-1] + K\left(\frac{X - \chi_j[t]}{h[t]}\right), \quad (10)$$

with Equations (9) and (10) to be complemented by condition (5).

2. MAIN RESULT

In general, the problem of estimating nonlinear regression function is well-known, and corresponding estimation method using kernel functions was originally proposed by Nadaraya (1964) and Watson (1964) for mutually independent observations. In sequel, such an estimate has been considerably developed by many authors. However, most of the papers as well as the original ones deal with the case of independent observations of variables, while such an assumption is not suitable within dynamic systems, at least when made with respect to the output processes, since for them dependence of observations is always inherent.

Further studies on nonlinear regression estimation by dependent observations (Andrews, 1995, Bianco and Boente, 1998, Bierens, 1983, Bosq, 1997a, 1997b, Cheze et al., 2003, Collomb and Hardle, 1986, Devroye and Wagner, 1980, Ferraty et al., 2002, Georgiev, 1984, Greblicki and Pawlak, 1989, Hall and Van Keilegom, 2003, Hasiewicz, 2001, Herbster, 2001, Juditsky et al., 1995, Krzyzak, 1993, 1996, 2001, Masry, 1997, Masry, and Tjostheim, 1997, Matzner-Loder et al., 1998, Peteret al., 2003, Rios, 1997, Robinson, 1986, Ruiz and Guillamon, 1996, Stenman and Gustafsson, 2001, Yakowitz, 1985) can be subdivided into the following two large classes. The first one involves deriving estimates converging in probability, i.e. in weak sense, under independent observations of the input variable. From a control system theory point of view, such a condition corresponds to using white-noise input process of systems.

The second class involves estimates converging with probability 1, i.e. in strong sense (strongly consistent estimates) under strong mixing condition imposed on input and output variables. The condition denotes asymptotic independence of the future and the past of the random processes and has the following form. Let $S(-\infty,t)$ and $S(t+k,\infty)$ be the σ-algebrae generated by a random process $u[s]$, i.e.

$$S(-\infty,t) = \sigma\{u[s], \ s < t\},$$
$$S(t+k,\infty) = \sigma\{u[s] \ge t+k\}.$$

Here $S(-\infty,t)$ is interpreted as the past of the process $u[s]$, and $S(t+k,\infty)$, as its future. Then, the random process $u[s]$ is said to be strongly mixing with the strong mixing coefficient $\alpha(k)$, if

$$\alpha(k) = \sup_{\substack{A\in S(-\infty,t) \\ B\in S(t+k,\infty)}} |P\{AB\} - P\{A\}P\{B\}| \to 0 .$$

Here $P\{\cdot\}$ is the probability of the event $\{\cdot\}$. Besides this, approaches within the second class of estimates assume the observed random processes to be strongly stationary.

In contrast to the above assumptions, the main subject of the present paper is the following *theorem*: *Let system (1) and its components meet the conditions described in Section 1. Then, for any initial approximation* $m_{ijX}[0]$, *algorithms (3) to (5) and (9), (10), (5) converge with probability 1 to a* *function* $\Phi_{ijh}(X)$ *depending on the parameter h from (5), with*

$$\Phi_{ijh}(X) \to \mathbf{M}\left\{ \chi_i[t] \middle/ \chi_j[t] = X \right\} \text{ as } h \to 0 .$$

Note that the magnitude of the parameter h in (5) may be chosen a priori and may be as small as required, i.e. it has only theoretical sense. Thus, in practice, the values of the function $\Phi_{ijh}(X)$ and regression $\mathbf{M}\left\{ \chi_i[t] \middle/ \chi_j[t] = X \right\}$ will be absolutely identical.

The idea of the theorem proof is based on a consideration, that, for any *fixed X*, the regression $\mathbf{M}\left\{ \chi_i[t] \middle/ \chi_j[t] = X \right\}$ may just be interpreted as a "parameter" subject to identification. Hence, investigating strong consistency of recursive scheme (3) to (5) may be based on applying such a powerful tool as Ljung's Ordinary Differential Equation (ODE) method (Ljung, 1975). Its conditions relating to non-liner systems are presented in Appendix The entity of the method is substituting investigation of convergence properties of a recursive scheme by investigating of an ordinary differential equation solution associated with the scheme. Specifically, the ODE method part relating to nonlinear systems is to be used. To apply it, the corresponding ODE method conditions are to be verified with respect to system (1) and algorithm (3) to (5) (algorithm (9), (10), (5)) description.

3. THE THEOREM PROOF

To verify Ljung' ODE method conditions (presented in Appendix), let, at first, note estimate (3), (4) to be represented in the form

$$m_{ijX}[t] = \frac{\mu_{ijX}[t]}{\rho_{ijX}[t]}, \tag{11}$$

$$\mu_{ijX}[t] = \mu_{ijX}[t-1] + \\ + \frac{1}{t}\left(h^{-1}[t]\chi_i[t] K\left(\frac{X - \chi_j[t]}{h[t]} \right) - \mu_{ijX}[t-1] \right), \tag{12}$$

$$\rho_{jX}[t] = \rho_{jX}[t-1] + \\ + \frac{1}{t}\left(h^{-1}[t] K\left(\frac{X - \chi_j[t]}{h[t]} \right) - \rho_{jX}[t-1] \right). \tag{13}$$

Formally, in terms of the ODE method, relationships (12), (13) may be rewritten in the form of (A1) where θ, γ, and Q are as follows

$$\theta[t] = \begin{bmatrix} \mu_{ijX}[t] & \rho_{jX}[t] \end{bmatrix}^T, \tag{14}$$

$$\gamma[t] = \frac{1}{t}, \tag{15}$$

$$Q(t;\theta[t-1], \mathrm{X}[t]) = \widetilde{Q}(t;\mathrm{X}[t]) - \theta[t-1] \tag{16}$$

where

$$\widetilde{Q}(t;\mathrm{X}[t]) = \\ = \begin{pmatrix} h^{-1}[t] \cdot \mathbf{1}_i^T \mathrm{X}[t] \cdot K\left(\dfrac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h[t]} \right) \\ h^{-1}[t] \cdot K\left(\dfrac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h[t]} \right) \end{pmatrix},$$

and the following notation is used:

$$\mathbf{1}_l^T = (\underbrace{0,\ldots,0}_{l-1},1,\underbrace{0,\ldots,0}_{\mathrm{N}-l}) \text{ for any } l = 1,2,\ldots,\mathrm{N} .$$

Complemented by system description (1), Equations (14) to (16) are to be verified to meet the ODE presented in the Appendix. method conditions. This is straightforward: conditions (A1), (A2), (A7), (A15) are met by definition; conditions (A3) to (A6) are obvious; conditions (A9) to (A13) are also met by system description and since Equation (1) does not depend on θ at all. Condition (A8) is met under an appropriate choice of the kernel functions, i.e. these are to be continuously differentiable.

Note, that, strictly speaking, not all of functions (7) may formally be implemented within algorithm (3) to (5) since some of them do not meet condition (A8). However, smooth approximations which are as close to the "original" ones as required may be used instead of them. Thus, from a computational point of, there will be found no difference at all.

Again, based on Equation (16) and in accordance with condition (A14), it, formally, follows

$$\lim_{t\to\infty} \mathbf{M}\left\{ Q\left(t;\overline{\theta}, \overline{\mathrm{X}}(t,\overline{\theta}) \right) \right\} = \\ = \begin{pmatrix} h^{-1} \cdot \mathbf{M}\left\{ \mathbf{1}_i^T \mathrm{X}[t] \cdot K\left(\dfrac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\} \\ h^{-1} \cdot \mathbf{M}\left\{ K\left(\dfrac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\} \end{pmatrix} - \overline{\theta} . \tag{17}$$

Thus, the ODE method may be applied, what leads to the following differential equation

$$\frac{d}{d\tau}\theta^D(\tau) = \begin{pmatrix} \mathrm{M}_{ij} & \mathrm{M}_j \end{pmatrix}^T - \theta^D(\tau), \quad (18)$$

where

$$\mathrm{M}_{ij} = h^{-1} \cdot \mathbf{M}\left\{ \mathbf{1}_i^T \mathrm{X}[t] \cdot K\left(\frac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\},$$

$$\mathrm{M}_j = h^{-1} \cdot \mathbf{M}\left\{ K\left(\frac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\}.$$

The asymptotically stable solution of Equation (18) is, obviously, $\begin{pmatrix} \mathrm{M}_{ij} & \mathrm{M}_j \end{pmatrix}^T$. Hence, by the ODE method theorem,

$$\mu_{ijX}[t] \to h^{-1} \cdot \mathbf{M}\left\{ \mathbf{1}_i^T \mathrm{X}[t] \cdot K\left(\frac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\},$$

$$\rho_{jX}[t] \to h^{-1} \cdot \mathbf{M}\left\{ K\left(\frac{X - \mathbf{1}_j^T \mathrm{X}[t]}{h} \right) \right\}$$

with probability 1.

Then, the following lemma from Greblicki and Pawlak (1989) is to be applied:

Let the bounded Borel kernel satisfy conditions (6). Then, for a pair of random variables (Y, U) the following limit relationships hold

$$\lim_{h\to 0} h^{-1}\mathbf{M}\left\{ YK\left(\frac{u-U}{h} \right) \right\} =$$
$$= p(u)\mathbf{M}\left\{ Y\!\big/_{U=u} \right\} \int_{-\infty}^{\infty} K(u)du$$

and

$$\sup_{h>0} h^{-1}\mathbf{M}\left\{ YK\left(\frac{u-U}{h} \right) \right\} < \infty.$$

Simultaneously,

$$\lim_{h\to 0} h^{-1}\mathbf{M}\left\{ K\left(\frac{u-U}{h} \right) \right\} = p(u) \int_{-\infty}^{\infty} K(u)du,$$

where p(u) stands for the distribution density of U.

Thus, taking into account Equation (11) completes the proof.

Proving strong consistency of algorithm (9), (10), (5) is implemented in the same manner.

4. CONCLUSIONS

A nonparametric approach to system identification/estimation has been considered. The approach assumes the investigated dynamic system model to be nonlinear and to be identified by current estimation of nonlinear characteristics, with the characteristics not allowing finite parameterization. Within the approach, the recursive estimates are obtained which converge in strong rather than weak sense, with no specific conditions on the output and input processes of the investigated system being imposed. The only "specific" conditions used assume values of the input and output processes to be bounded by some constants, with magnitudes of the constants being not essential. The constants are not required to be known and only hypothetical existence of the bounds is assumed. In entity, such a condition is not a limitation at all, from a practical point of view, since real-world system processes are always bounded.

APPENDIX. LJUNG' ODE METHOD CONDITIONS

Following to Ljung (1975), consider the following "state space" representation of a stochastic approximation algorithm represented by conditions (A1) and (A2).

$$\theta[t] = \theta[t-1] + \gamma[t]Q(t;\theta[t-1],\mathrm{X}[t]), \quad (A1)$$
$$\mathrm{X}[t] = \Psi(t;\mathrm{X}[t-1],\theta[t-1],e[t]). \quad (A2)$$

In (A1), the sequence $\gamma(t)$ meets the following conventional conditions (A3) to (A6)

$$\sum_{t=1}^{\infty} \gamma[t] = \infty, \quad (A3)$$

$$\sum_{t=1}^{\infty} \gamma^p[t] < \infty \textit{ for some } p > 1, \quad (A4)$$

$$\gamma[t] \textit{ is a decreasing sequence } \gamma[t], \quad (A5)$$

$$\lim_{t\to\infty} \sup\left\{ \frac{1}{\gamma[t]} - \frac{1}{\gamma[t-1]} \right\} < \infty. \quad (A6)$$

Let, again, the function $\Psi(\cdot,\cdot,\cdot)$ be bounded for each $\theta[t]$, X[t], and $e[t]$, i.e.

$$\left\| \Psi(t;\mathrm{X}[t-1],\theta[t-1],e[t]) \right\| < C \quad (A7)$$
$$\forall \mathrm{X}, e \ \forall \theta \in D_R,$$

with C being able to depend on D_R, where D_R is a set.

It is also assumed:

$$\begin{array}{l} \textit{the function } Q(t;\theta,\mathrm{X}) \textit{ to be continuously} \\ \textit{differentiable with respect to } \theta \textit{ and } \mathrm{X} \\ \textit{and the derivatives are bounded for} \\ \forall\theta\in D_R \end{array}, \quad \text{(A8)}$$

$$\begin{array}{l} \textit{the function } \Psi(t;\mathrm{X}[t-1],\theta[t-1],e[t]) \\ \textit{to be continuously differentiable} \\ \textit{with respect to } \theta \textit{ for } \forall\theta\in D_R \end{array}. \quad \text{(A9)}$$

Let, for any $\bar{\theta}$, the vector $\bar{\mathrm{X}}(t,\bar{\theta})$ be defined as

$$\bar{\mathrm{X}}(t,\bar{\theta}) = \Psi\left(t;\bar{\mathrm{X}}(t-1,\bar{\theta}),\bar{\theta},e[t]\right), \quad \text{(A10)}$$
$$\bar{\mathrm{X}}(0,\bar{\theta}) = 0,$$

and let also Ψ possesses the property

$$\left\|\bar{\mathrm{X}}(t,\bar{\theta}) - \mathrm{X}[t]\right\| < C\cdot\max_{n\le k\le t}\left\|\bar{\theta}-\theta[k]\right\| \quad \text{(A11)}$$

if

$$\bar{\mathrm{X}}(n,\bar{\theta}) = \mathrm{X}(n). \quad \text{(A12)}$$

This means that small variations in θ in (A2) are not amplified to a higher magnitude for the observations X.

Remark. One should be noted here that within the paper, conditions (A10) to (A12) are always valid since Ψ *does no depend on* $\bar{\theta}$.

Again, let $\bar{\mathrm{X}}_i(t,\bar{\theta})$ be solutions of (A10) with $\bar{\mathrm{X}}_i(s,\bar{\theta}) = \mathrm{X}_i^0$, $i=1,2$. Then let D_S be defined as the set of all $\bar{\theta}$ for which the following inequality holds

$$\left\|\bar{\mathrm{X}}_1(t,\bar{\theta}) - \bar{\mathrm{X}}_2(t,\bar{\theta})\right\| < C\left(\mathrm{X}_1^0,\mathrm{X}_2^0\right)\cdot\lambda^{t-s}(\bar{\theta}) \quad \text{(A13)}$$

where $t>s$ and $\lambda(\bar{\theta})<1$. This is the region of exponential stability of (A2).

Let, again, there exists the function Q expectation limit

$$\lim_{t\to\infty}\mathbf{M}\left\{Q\left(t;\bar{\theta},\bar{\mathrm{X}}(t,\bar{\theta})\right)\right\} = f(\bar{\theta}) \quad \text{(A14)}$$

for $\bar{\theta}\in D_R$, with the mathematical expectation being taken over $e[t]$.

Finally,

$$\begin{array}{l} e[t] \textit{ is a sequence of independent} \\ \textit{random variables} \end{array} \quad \text{(A15)}$$

Consider algorithm (A1) to (A6) under assumptions (A7) to (A15). Let D_R be an open connected subset of D_S. Then the following *theorem* holds for the algorithm:

Assume the ODE

$$\frac{d}{d\tau}\theta^D(\tau) = f\left(\theta^D(\tau)\right)$$

($\theta^D(\tau)$ will always refer to the solution of the ODE, while $\theta(t)$ are the estimates generated by algorithm (A1) to (A6)) to have a stationary point θ^* which is an asymptotically stable solution with domain of attraction $D_A\supset D_R$, i.e. for all initial values in D_A, the solution of the ODE tends to θ^* as $\tau\to\infty$). It is also assumed that D_R can be taken so that solutions of the ODE that start in D_R remain in there for $\tau>0$. Then $\theta(t)\to\theta^*$ with probability 1 as $t\to\infty$.

REFERENCES

Andrews, D.W.K. (1995). “Nonparametric kernel estimation for semiparametric models”, *Econ. Theory*, **11**, pp. 560-596.

Bianco, A., and G. Boente (1998). “Robust kernel estimators for additive models with dependent observations”, *Can. J. Statist.*, **26**, pp. 239-255.

Bierens, H.T. (1983). “Uniform consistency of kernel estimators of regression function under generalized conditions”, *J. Amer. Statist. Assoc.*, **78**, pp. 699-707.

Bosq, D. (1997a). “Parametric rates of nonparametric estimators and predictors for continuos time processes”, *Ann. Statist.*, vol. 25, pp. 982-1000.

Bosq, D. (1997b). “Nonparametric Estimation and Prediction for Continuous time Processes”, *Nonlinear Analysis*, **30**, no. 6, pp. 3547-3551.

Cheze, N., J.-M. Poggi and B. Porter (2003). “Partial and Recombined Estimators for Nonlinear Additive Models”, *Statistical Inference for Stochastic Processes*, **6**, no. 2, pp. 155-197.

Collomb, G. and W. Hardle (1986). “Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations”, *Stochastic Processes Appl.*, **23**, pp. 77-89.

Devroye, L. and T.J. Wagner (1980). “On the L_1 convergence of the kernel estimators of regression function with applications in discrimination”, *Z. Wahrscheinlichkeitstherie ver. Gebiete.*, **51**, pp. 18-25.

Ferraty, F., A. Goia, and Ph. Vieu (2002). “Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeants”, *C. R. Acad. Sci. Ser. 1*, **334**, no. 3, pp. 217-220. (in French)

Georgiev, A. (1984). "Nonparametric system identification by kernel methods", *IEEE Trans. Automatic Control*, **AC-29**, pp. 356-358.

Greblicki, W. and M. Pawlak (1989). "Recursive nonparametric identification of Hammerstein systems", *J. Franklin Institute*, **326**, pp. 461-481.

Hall, P. and I. Van Keilegom (2003). "Using difference-based methods for inference in nonparametric regression with time series errors", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, no. 2, pp. 443-456.

Hasiewicz, Z. (2001). "Non-parametric estimation of non-linearity in a cascade time-series system by multiscale approximation", *Signal Processing*, vol. 81, no. 4, pp., 791-807.

Herbster, M. (2001). "Learning Additive Models Online with Fast Evaluating Kernels", *Lecture Notes in Computer Science*, **2111**.

Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjoberg, and Z. Qinghua (1995) "Nonlinear Black-box Models in System Identification: Mathematical Foundations", *Automatica*, **31**, no. 12, pp. 1725-1750.

Krzyzak, A. (1993). "Identification of nonlinear block-oriented systems by the recursive kernel estimate", *J. Franklin Institute*, **330**, pp. 605-627.

Krzyzak, A. (1996). "On nonparametric estimation of nonlinear dynamic systems by the Fourier series estimate", *Signal Processing*, **52**, no. 3 pp. 299-321.

Krzyzak, A. (2001). "Nonlinear Function Learning and Classification Using Optimal Radial Basis Function Networks", *Lecture Notes in Computer Science*, **2123**.

Ljung L. (1975). "Theorems for the asymptotic analysis of recursive stochastic algorithms", *Report 7522, December 1975, Department of Automatic Control, Lund Institute of Technology*.

Masry, E. (1997). "Multivariate regression estimation: local polynomial fitting for time series", *Nonlinear Analysis*, **30**, no. 6, pp. 3575-3581.

Masry, E., and D. Tjostheim (1997). "Additive nonlinear ARX time series and projection estimates", *Econ. Theory*, **13**, pp. 214-252.

Matzner-Loder E., A. Gannoun and J.G. De Gooijer (1998). "A comparison of three kernel-based methods", *Commun. Statist. Theory and Meth*., **27**, pp. 1563-1617.

Nadaraya, E.A. (1964). "On estimating regressions", *Theor. Probability Appl.*, **9**, pp. 141-142.

Peterson D.R., Hongwei Zhao, and S. Eapen (2003). "Using Local Correlation in Kernel-Based Smoothers For Dependent Data", *Biometrics*, **59**, no. 4, pp. 984-991.

Rios, R. (1997). "On the bias in local polynomial regression estimation for dependent data", *C. R. Acad. Sci. Ser. 1*, **324**, pp. 117-122.

Robinson, P.M. (1986). "On the consistency and finite-sample properties of nonparametric kernel time-series regression, auto-regression and density estimators", *Ann. Inst. Statist. Math. Part A*, **38**. pp. 539-549.

Ruiz, J.M. and A. Guillamon (1996). "Nonparametric recursive estimator for mean residual life and vitality function under dependence conditions", *Commun. Statist. Theory and Meth*., **25**, pp. 1997-2011.

Stenman, A. and F. Gustafsson (2001). "Adaptive smoothing methods for frequency-function estimation", *Automatica*, **37**, no. 5, pp. 675-685.

Watson, G.S. (1964). "Smooth regression analysis", *Sankhya. Ser. A*, **26**, pp. 355-372.

Yakowitz, S. (1985). "Nonparametric density estimation, prediction and regression for Markov sequences", *J. Amer. Statist. Assoc*., vol. 80, pp. 215-221.