Functional-coefficient Regression Models for Nonlinear Time Series

Zongwu Cai
Department of Mathematics
University of North Carolina
Charlotte, NC 28223, USA

Jianqing Fan*
Department of Statistics
University of California
Los Angeles, CA 90095, USA

Qiwei Yao†
Institute of Mathematics and Statistics
University of Kent at Canterbury
Canterbury, Kent CT2 7NF, UK

Abstract

We apply the local linear regression technique for estimation of functional-coefficient regression models for time series data. The models include threshold autoregressive models (Tong 1990) and functional-coefficient autoregressive models (Chen and Tsay 1993) as special cases but with the added advantages such as depicting finer structure of the underlying dynamics and better post-sample forecasting performance. We have also proposed a new bootstrap test for the goodness of fit of models and a bandwidth selector based on newly defined cross-validatory estimation for the expected forecasting errors. The proposed methodology is data-analytic and is of appreciable flexibility to analyze complex and multivariate nonlinear structures without suffering from the "curse of dimensionality". The asymptotic properties of the proposed estimators are investigated under the α-mixing condition. Both simulated and real data examples are used for illustration.

Keywords: α-mixing; Asymptotic normality; Bootstrap; Forecasting; Goodness-of-fit test; Local linear regression; Nonlinear time series; Varying-coefficient models.

*Partially supported by NSF Grant DMS-9803200 and NSA 96-1-0015.
†Partially supported by EPSRC Grant L16358 and BBSRC/EPSRC Grant 96/MMI09785.

1 Introduction

Until recently much of time series modeling has been confined to linear ARMA models (Box and Jenkins 1970). Although the original ARMA framework has been enlarged to include long range dependence with fractional ARMA (Granger and Joyeux 1980, and Dahlhaus 1989), multivariate VARMA and VARMAX models (Hannan and Deistler 1988) and random walk nonstationarities via cointegration (Engle and Granger 1987), there still exist so-called *nonlinear features* beyond the capacity of linear ARMA-modeling. For example, various "non-standard" phenomena such as non-normality, asymmetric cycles, bi-modality, nonlinear relationship between lagged variables, variation of prediction performance over the state-space, non-reversibility, and sensitivity to initial conditions have been well observed in many real time series data including some benchmark sets such as the sunspot, lynx and blowfly data. See Tong (1990, 1995) and Tjøstheim (1994) for further discussion in this aspect. Beyond linear domain, there are infinite many nonlinear forms to be explored. Early development of nonlinear time series analysis focused on various nonlinear (sometimes non-Gaussian) parametric forms (Tong 1990; Tjøstheim 1994, and references within). The successful examples include, among others, the ARCH-modeling of fluctuating structure for financial time series (Engel 1982, and Bollerslev 1986), the threshold modeling for biological and economic data (Tong 1990, and Tiao and Tsay 1994). On the other hand, recent development in nonparametric regression techniques provides an alternative to model nonlinear time series (Tjøstheim 1994; Yao and Tong 1995; Härdle, Lütkepohl and Chen 1997, and Masry and Fan 1997). The immediate advantage of this is that no prior information on model structure is assumed. Further, it may provide useful insight for further parametric fitting. However, an entire nonparametric approach is hampered by the requirement of large sample sizes and is often practically useful only for, for example, autoregressive models with order 1 or 2.

This paper adapts the functional-coefficient modeling technique to analyze nonlinear time series data. The approach allows appreciable flexibility on the structure of fitted model without suffering from the "curse of dimensionality". Let $\{Y_i,\, \mathbf{X}_i,\, \mathbf{U}_i\}_{i=-\infty}^{\infty}$ be jointly strictly stationary processes with $\mathbf{X}_i$ taking values in $\Re^p$ and $\mathbf{U}_i$ taking values in $\Re^k$. Typically k is small. Let $E(Y_1^2) < \infty$. We define the multivariate regression function

$$m(\mathbf{x},\ \mathbf{u}) = E\left(Y \,|\, \mathbf{X} = \mathbf{x},\, \mathbf{U} = \mathbf{u}\right), \tag{1.1}$$

where $(Y,\, \mathbf{X},\, \mathbf{U})$ has the same distribution as $(Y_i,\, \mathbf{X}_i,\, \mathbf{U}_i)$. In a pure time series context, both $\mathbf{X}_i$ and $\mathbf{U}_i$ consist of some lagged values of Y_i. The functional-coefficient regression model has the form

$$m(\mathbf{x},\, \mathbf{u}) = \sum_{j=1}^{p} a_j(\mathbf{u})\, x_j, \tag{1.2}$$

where $a_j(\cdot)$'s are measurable functions from $\Re^k$ to $\Re^1$ and $\mathbf{x} = (x_1,\, \ldots,\, x_p)^T$ with T denoting the transpose of a matrix or vector. The idea to model time series in such a form is not new; see, for example, Nicholls and Quinn (1982). In fact, many useful time series models may be regarded as special cases of model (1.2) (often with specified parametric forms for $a_j(\cdot)$'s; see §2 below). However, the potential of this modeling technique had not been fully explored until the seminal work of Cleveland *et al.* (1992) and Hastie and Tibshirani (1993) and Chen and Tsay (1993),

in which nonparametric techniques were developed for estimation of the functions $a_j(\cdot)$'s. In the context of independent samples, Fan and Zhang (1997) provided an innovative two-step method and insightful asymptotic results for the local polynomial estimation of $a_j(\cdot)$'s. They also pointed out that model (1.2) has strong connections with the functional linear models discussed in Ramsay and Silverman (1997) and Brumback and Rice (1998). Yet, few results are available in the time series context.

In this paper we adapt local linear regression technique to estimate the coefficient functions $a_j(\cdot)$'s. By smoothing $\mathbf{U}$ only, our method is particularly easy to implement. Within the framework of (1.2), the detailed form of model is determined by data, which will reduce the bias of fitting automatically. Since only k-dimensional functions are estimated, the difficulties associated with the "curse of dimensionality" will be substantially eased. Indeed, our data-analytic approach increases the modeling flexibility with little sacrifice of estimability (see Theorem 2 in §6 below). The specified form of (1.2) also facilitates the interpret-ability of the fitted model when k is small. This is particularly relevant in modeling longitudinal data where it is reasonable to assume that the coefficients change over time t. See Hoover *et al.* (1997) for a novel application of functional-coefficient models to longitudinal data. Model (1.2) is also important for modeling the population dynamics where it is reasonable to expect that animals behave differently based on its population size. Thus, using model (1.2) with $\mathbf{u}$ being the population size of a previous year captures such a kind of feature in the population dynamics. See Tong (1990, p.377) and (2.6) below for further discussions.

An important statistical question in fitting model (1.2) is if the coefficient function is really varying (namely, if a linear AR model is adequate) or more generally if a parametric model fits the given data. This amounts to testing if the coefficient functions are constant or in a certain parametric form. A new testing procedure is proposed based on the comparison of the residual sum of squares under the null and the alternative models. A bootstrap method is proposed for finding the null distribution of the test statistic. Our simulation shows that the resulting testing procedure is indeed powerful and the bootstrap method does give the right null distribution.

In §2, we list several familiar nonlinear times series models which can be regarded as special cases of model (1.2). Through the famous Canadian lynx data, we illustrate the advantages of the new approach over the existing parametric models on both modeling the underlying dynamics and post-sample forecasting. §3 presents the local linear regression estimators for functional-coefficient functions and a simple and fast algorithm for bandwidth selection. In §4, a bootstrap method is proposed for testing the goodness-of-fit of a parametric model against model (1.2). In §5, we use both simulated models and real data sets to illustrate the proposed methodology. The application with real data lends further support to use some well-known parametric models. The asymptotic properties of the proposed estimators are studied in §6. All technical proofs are relegated to the Appendix.

2 Models and an illustrative example

The general setting (1.2) includes many familiar time series models. We mention a few below. Some of them will be used in numerical illustration in §5.

FAR model. Chen and Tsay (1993) proposed the functional-coefficient autoregressive (FAR) model

$$x_t = a_1(\mathbf{X}^*_{t-1})\, x_{t-1} + \cdots + a_p(\mathbf{X}^*_{t-1})\, x_{t-p} + \varepsilon_t, \tag{2.1}$$

where $\mathbf{X}^*_{t-1} = (x_{t-i_1}, \ldots, x_{t-i_k})^T$, $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables, and ε_t is independent of $\{x_{t-i},\, i > 0\}$. Chen and Tsay (1993) studied probabilistic properties of FAR models and proposed an iterative algorithm to estimate the coefficient functions. In fact, their algorithm is in the spirit of local constant fitting, although they did not apply local regression technique directly. Instead, they constructed estimators based on an iterative recursive formula.

TAR model. One of the simplest nonlinear time series models is the threshold autoregressive (TAR) model

$$x_t = \phi_1^{(i)}\, x_{t-1} + \cdots + \phi_p^{(i)}\, x_{t-p} + \varepsilon_t^{(i)} \quad \text{if} \quad x_{t-d} \in \Omega_i, \qquad i = 1, \ldots, k, \tag{2.2}$$

where $\{\Omega_i\}$ form a (non-overlapping) partition of the real line. For both theoretical properties and practical implementations of TAR modeling, we refer to Tong (1990).

EXPAR model. The following generalized exponential autoregressive (EXPAR) model was proposed and studied by Haggan and Ozaki (1981) and Ozaki (1982)

$$x_t = \sum_{i=1}^{p} \Big\{ \alpha_i + (\beta_i + \gamma_i\, x_{t-d}) \exp(-\theta_1\, x_{t-d}^2) \Big\}\; x_{t-i} + \varepsilon_t, \tag{2.3}$$

where $\theta_i \geq 0$ for $i = 1, \ldots, p$.

Regression with random coefficients. Consider the model of Granger and Teräsvirta (1993):

$$Y_t = \boldsymbol{\beta}(t)^T\, \mathbf{X}_t + u_t, \tag{2.4}$$

where $\{u_t\}$ is a sequence of i.i.d. random variables with $E(u_t) = 0$ and $\mathrm{Var}(u_t) = \sigma^2$ and is independent of $\{\mathbf{X}_t\}$ and $\{\boldsymbol{\beta}(t)\}$. Further, $E(\boldsymbol{\beta}(t)) = \boldsymbol{\beta}$ and $\mathrm{Var}(\boldsymbol{\beta}(t)) = \Phi$, $\mathrm{Cov}(\boldsymbol{\beta}(s), \boldsymbol{\beta}(t)) = 0$ for $s \neq t$. The above random coefficient model has received considerable attention in econometrics; see Granger and Teräsvirta (1993). If $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p})^T$, then (2.4) is the random coefficient autoregressive model surveyed in Nicholls and Quinn (1982).

All the above models have been proved successful for modeling *some* nonlinear features. For example, the TAR model has received considerable attention due to its easy implementation and often nice interpretation. The application to Canadian lynx data (*i.e.* the annual fur returns of lynx at auction in 1982–1934) is arguably a showcase of the TAR modeling technique; see Tong (1990). The periodic fluctuation displayed in this time series has profoundly influenced ecological theory. The data set has been constantly used to examine the concepts as "balance-of-nature", predator and prey interaction, food web dynamics and etc; see Stenseth *et al.* (1999) and references within.

Having incorporated biological evidence, Tong fitted the following TAR model with two regimes and the delay variable at lag 2 to the lynx data at the logarithmic scale with the base 10

$$x_t = \begin{cases} 0.62 + 1.25\, x_{t-1} - 0.43\, x_{t-2} + \varepsilon_t^{(1)}, & x_{t-2} \le 3.25, \\ 2.25 + 1.52\, x_{t-1} - 1.24\, x_{t-2} + \varepsilon_t^{(2)}, & x_{t-2} > 3.25. \end{cases} \tag{2.5}$$

See Tong (1990, p.377). This simple model admits nice biological interpretation. Indeed it can be viewed as derived from basic predator (lynx) and prey (hares) interaction model in ecology (see equation (2) in Stenseth *et al.* 1999). The lower regime corresponds *roughly* to the population increase phase, and the upper regime corresponds to the population decrease phase. Note that the coefficient of x_{t-1} in the model is significant positive, but less so during the increase phase. The coefficient of x_{t-2} is significantly negative, and more so during the decline phase. The signs of those coefficients reflect that lynx and hares relate with each other in a specified prey-predator interactive manner. The difference of the coefficients in increase and decrease phases reflects the so-called *phase-dependence* and *density-dependence* in ecology (Stenseth *et al.* 1999). The phase-dependence means that both lynx and hares behave differently (in hunting or escaping) when lynx population increases or decreases. The density-dependence implies that the reproduction rates of animals as well as their behavior depend on the abundance of the population. Clearly the above threshold model simplified the varying behavior into two states. With the new technique proposed in this paper, we fit the lynx data with the model

$$x_t = a_1(x_{t-2})\, x_{t-1} + a_2(x_{t-2})\, x_{t-2} + \varepsilon_t, \tag{2.6}$$

in which the coefficient $a_1(.)$ and $a_2(.)$ vary with respect to "threshold variable" x_{t-2}. Both $a_1(.)$ and $a_2(.)$ are estimated through a simple one-dimensional kernel regression. The estimators are plotted

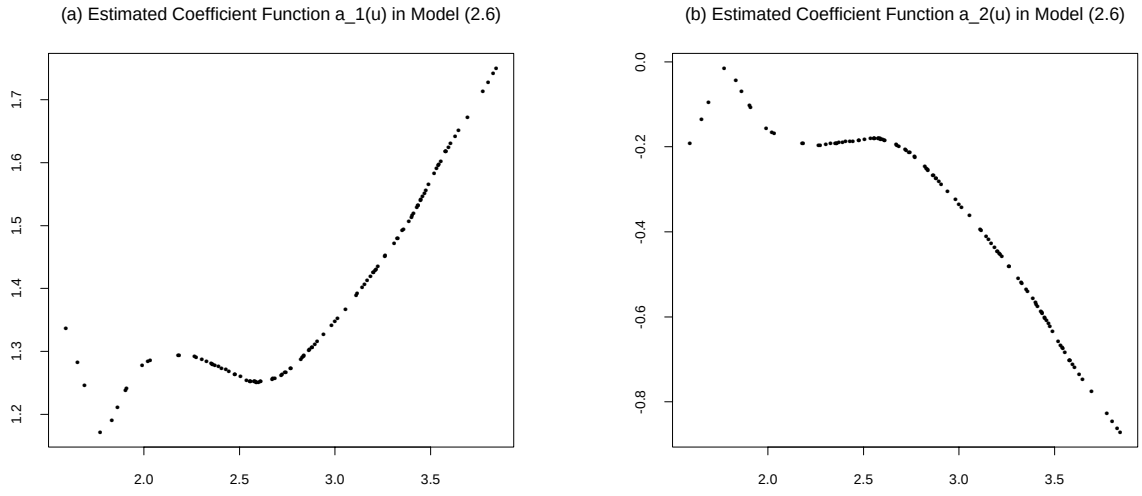

Figure 1: *Canadian Lynx data. The local linear estimator of (a)* $a_1(x_{t-2})$ *and (b)* $a_1(x_{t-2})$ *in model (2.6).*

in Figures 1(a) and (b). Except a few points near the low end, $a_1(.)$ is a positive increasing function, which depicts the *smooth* (rather than *radical*) density dependence. The function $a_2(.)$ is negative

and largely decreasing. This pictures a gradual change in animals' behavior in corresponding to the change of population abundance. By allowing the coefficient varying with respect to population density, the model presents the lynx and hares interaction in the manner which is one-step further closer to the reality than the TAR models. The advantages of the new technique on other aspects such as prediction will be reported in Section 5.

3 Estimation

For the simplicity we only consider the case $k = 1$ in (1.2). Extension to the case $k > 1$ involves no fundamentally new ideas. Note that models with large k are often not practically useful due to "curse of dimensionality".

3.1 Local linear regression estimation

The local linear fittings have several nice properties. They possess high statistical efficiency in an asymptotic minimax sense and are design-adaptive (Fan 1993). Further, they automatically correct edge effects (Fan and Gijbels 1992, Hastie and Loader 1993 and Ruppert and Wand 1994). We estimate the functions $a_j(\cdot)$'s using local linear regression method from observations $\{U_i, \mathbf{X}_i, Y_i\}_{i=1}^n$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$. We assume throughout the paper that $a_j(\cdot)$ has continuous second derivative. Note that we may approximate $a_j(\cdot)$ locally at u_0 by a linear function $a_j(u) \approx a_j + b_j\,(u - u_0)$. The local linear estimator is defined as $\widehat{a}_j(u_0) = \widehat{a}_j$, where $\{(\widehat{a}_j, \widehat{b}_j)\}$ minimize the sum of weighted squares

$$\sum_{i=1}^{n} \left[Y_i - \sum_{j=1}^{p} \{a_j + b_j\,(U_i - u_0)\}\, X_{ij} \right]^2 K_h(U_i - u_0). \tag{3.1}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function on $\Re^1$ and $h > 0$ is a bandwidth. It follows from the least squares theory that

$$\widehat{a}_j(u_0) = \sum_{k=1}^{n} K_{n,j}(\mathbf{X}_k,\, U_k - u_0)\, Y_k, \tag{3.2}$$

where

$$K_{n,j}(\mathbf{x},\, u) = e_{j,p}^T \left(\widetilde{\mathbf{X}}^T\, \mathbf{W}\, \widetilde{\mathbf{X}} \right)^{-1} \begin{pmatrix} \mathbf{x} \\ u\,\mathbf{x} \end{pmatrix} K_h(u). \tag{3.3}$$

In the above expression, $e_{j,p}$ is the $p \times 1$ unit vector with 1 at the j-th position, $\widetilde{\mathbf{X}}$ denotes an $n \times 2p$ matrix with $(\mathbf{X}_i^T, \mathbf{X}_i^T(U_i - u_0))$ as its i-th row, and $\mathbf{W} = \mathrm{diag}\,\{K_h(U_1 - u_0), \ldots, K_h(U_n - u_0)\}$.

3.2 Bandwidth selection

Now we propose a simple and quick method to select bandwidth for the above estimation. It can be regarded as a modified multi-fold cross-validation criterion which is attentive to the structure of time series data. Let m and Q be two given positive integers and $n > mQ$. The basic idea is first to use Q sub-series of lengths $n - qm$ ($q = 1,, \cdots, Q$) to estimate the unknown coefficient functions

and then to compute the one-step forecasting errors of the next section of the time series of length m based on the estimated models. More precisely, we choose h which minimizes

$$\mathrm{AMS}(h) = \sum_{q=1}^{Q} \mathrm{AMS}_q(h), \tag{3.4}$$

where for $q = 1, \cdots, Q$,

$$\mathrm{AMS}_q(h) = \frac{1}{m} \sum_{i=n-qm+1}^{n-qm+m} \left\{ Y_i - \sum_{j=1}^{p} \widehat{a}_{j,q}(U_i) X_{i,j} \right\}^2,$$

and $\{\widehat{a}_{j,q}(\cdot)\}$ are computed from the sample $\{(Y_i, U_i, \mathbf{X}_i), 1 \le i \le n - qm\}$ with bandwidth equal $h(\frac{n}{n-qm})^{1/5}$. Note that for different sample sizes, we re-scale bandwidth according to its optimal rate, i.e. $h \propto n^{-1/5}$. In the practical implementation, we may use $m = [0.1n]$ and $Q = 4$. We take $m = [0.1n]$ rather than $m = 1$ simply because of computation expediency.

3.3 Choosing smooth variable

It is important to choose an appropriate smooth variable U in applying functional-coefficient regression models. Knowledge on physical background of the data may be very helpful, as we have witnessed in modeling lynx data in Section 2. When no prior information is available, it is pertinent to choose U in terms of some data-driven methods such as AIC, cross-validation and other criteria. Ideally we would choose U as a linear function of given explanatory variables according to some optimal criterion, which is obviously beyond the scope of this paper and will be explored in a follow-up paper separately. Nevertheless, we propose here a simple and practical approach: let U be one of the given explanatory variables such that AMS defined in (3.4) obtains its minimum value. Obviously this idea can also be extended to select p as well. See Example 4 in Section 5.2 for practical implementation of this approach.

4 Goodness of fit test

To test whether model (1.2) holds with a specified parametric form such as TAR or EXPAR models (see §2), we propose a goodness-of-fit test based on the comparison of the Residual Sum of Squares (RSS) from both parametric and nonparametric fittings.

Consider the null hypothesis

$$H_0 : a_j(u) = \alpha_j(u, \boldsymbol{\theta}), \quad 1 \le j \le p, \tag{4.1}$$

where $\alpha_j(\cdot, \boldsymbol{\theta})$ is a given family of functions indexed by unknown parameter vector $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}$ be an estimatior of $\boldsymbol{\theta}$. The residual sum of squares under the null hypothesis is

$$\mathrm{RSS}_0 = n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \alpha_1(U_i, \widehat{\boldsymbol{\theta}}) X_{i1} - \cdots - \alpha_p(U_i, \widehat{\boldsymbol{\theta}}) X_{ip} \right\}^2.$$

Analogously, the residual sum of squares corresponding to model (1.2) is

$$\mathrm{RSS}_1 = n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \widehat{a}_1(U_i) X_{i1} - \cdots - \widehat{a}_p(U_i) X_{ip} \right\}^2.$$

The test statistic is defined as

$$T_n = (\mathrm{RSS}_0 - \mathrm{RSS}_1)/\mathrm{RSS}_1 = \mathrm{RSS}_0/\mathrm{RSS}_1 - 1$$

and we reject the null hypothesis (4.1) for large values of T_n. We use the following nonparametric bootstrap approach to evaluate p-value of the test.

1. Generate the bootstrap residuals $\{\varepsilon_i^*\}_{i=1}^n$ from the empirical distribution of the centered residuals $\{\widehat{\varepsilon}_i - \bar{\widehat{\varepsilon}}\}_{i=1}^n$, where

$$\widehat{\varepsilon}_i = Y_i - \widehat{a}_1(U_i)\, X_{i1} - \cdots - \widehat{a}_p(U_i)\, X_{ip}, \quad \bar{\widehat{\varepsilon}} = \frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i,$$

   and define

$$Y_i^* = \alpha_1(U_i, \widehat{\boldsymbol{\theta}})X_{i1} + \cdots + \alpha_p(U_i, \widehat{\boldsymbol{\theta}})X_{ip} + \varepsilon_i^*.$$

2. Calculate the bootstrap test statistic T_n^* based on the sample $\{U_i,\ X_{i1},\ \ldots,\ X_{ip},\ Y_i^*\}_{i=1}^n$.

3. Reject the null hypothesis H_0 when T_n is greater than the upper-α point of the conditional distribution of T_n^* given $\{U_i,\ X_{i1},\ \ldots,\ X_{ip},\ Y_i\}_{i=1}^n$.

The p-value of the test is simply the relative frequency of the event $\{T_n^* \geq T_n\}$ in the replications of the bootstrap sampling. For the sake of simplicity, we use the same bandwidth in calculating T_n^* as that in T_n. Note that we bootstrap the centralized residuals from the nonparametric fit instead of the parametric fit, because the nonparametric estimate of residuals is always consistent, no matter the null or the alternative hypothesis is correct. The method should provide a consistent estimator of the null distribution even when the null hypothesis does not hold. Kreiss, Neumann and Yao (1998) considered nonparametric bootstrap tests in a general nonparametric regression setting. They proved that asymptotically the conditional distribution of the bootstrap test statistic is indeed the distribution of the test statistic under the null hypothesis. It may be proved that the similar result holds here as long as $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ at the rate $n^{-1/2}$.

5 Numerical properties

We illustrate the proposed methods through two simulated and two real data examples. The estimators $\{\widehat{a}_j(\cdot)\}$ are assessed via the square-Root of Average Squared Errors (RASE):

$$\mathrm{RASE}_j = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} \{\widehat{a}_j(u_k) - a_j(u_k)\}^2 \right]^{1/2}, \qquad \text{and} \qquad \mathrm{RASE} = \sum_{j=1}^p \mathrm{RASE}_j, \tag{5.1}$$

where $\{u_k,\ k = 1,\ \ldots,\ n_{grid}\}$ are regular grid points. We also compare the post-sample forecasting performance of the new method with existing methods such as linear AR model, TAR model and FAR model (implemented by Chen and Tsay 1993). We consider three predictors based on functional-coefficient modeling: the one-step ahead predictor

$$\widehat{x}_{t+1} = \widehat{a}_1(\mathbf{X}_t^*)\, x_t + \cdots + \widehat{a}_p(\mathbf{X}_t^*)\, x_{t-p+1},$$

the iterative two-step ahead predictor

$$\widehat{x}_{t+2} = \widehat{a}_1(\mathbf{X}^*_{t+1})\widehat{x}_{t+1} + \widehat{a}_2(\mathbf{X}^*_{t+1})x_t + \cdots + \widehat{a}_p(\mathbf{X}^*_{t+1})\, x_{t-p+2}, \tag{5.2}$$

and the direct two-step ahead predictor based on the model

$$x_{t+2} = b_1(\mathbf{X}^*_t)\, x_t + \cdots + b_p(\mathbf{X}^*_t)\, x_{t-p} + \varepsilon'_t. \tag{5.3}$$

Note that model (2.1) does not necessarily imply (5.3). In this sense, the direct two-step ahead prediction explores the predictive power of the proposed modeling techniques when the model is misspecified. We always use the Epanechnikov kernel $K(u) = 0.75\ (1-u^2)_+$. For the two real data examples, we repeat bootstrap sampling 1000 times in goodness-of-fit tests and the bandwidths are selected by the method proposed in §3.2.

5.1 Simulated examples

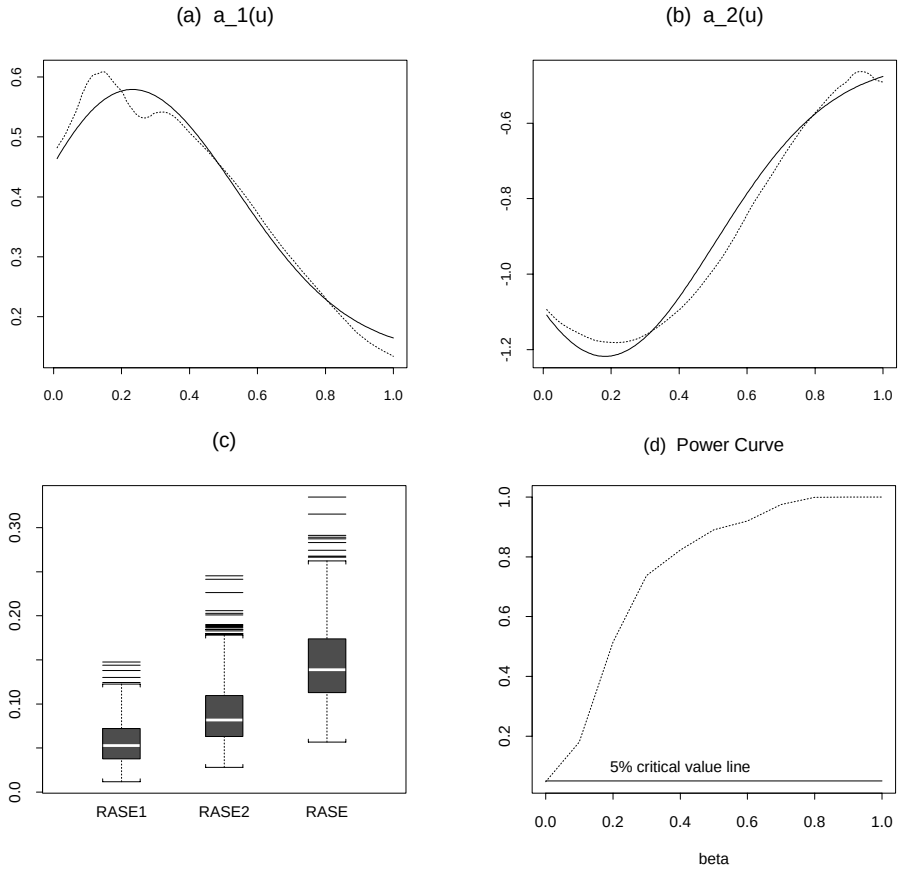

Figure 2: *Simulation results for Example 1. (a) The local linear estimator (dotted line) of the coefficient function $a_1(x_{t-1})$ (solid line). (b) The local linear estimator (dotted line) of $a_2(x_{t-1})$ (solid line). (c) The boxplots of the 400 RASE-values in estimation of $a_1(\cdot)$ and $a_2(\cdot)$. (d) The plot of power curve against β for the goodness-of-fit test.*

Example 1. We first consider an EXPAR model. We replicate simulation 400 times and each time we draw a time series with length 400 from the model

$$x_t = a_1(x_{t-1})\, x_{t-1} + a_2(x_{t-1})\, x_{t-2} + \varepsilon_t, \tag{5.4}$$

where $a_1(u) = 0.138 + (0.316 + 0.982\, u)\, e^{-3.89\, u^2}$, $a_2(u) = -0.437 - (0.659 + 1.260\, u)\, e^{-3.89\, u^2}$, and $\{\varepsilon_t\}$ are i.i.d. $N\,(0,\, 0.2^2)$. We choose the optimal bandwidth $h_n = 0.41$ which minimizes the sum of the integrated squared errors of estimators for $a_1(\cdot)$ and $a_2(\cdot)$. Figures 2(a)-(b) present the estimated $a_1(\cdot)$ and $a_2(\cdot)$ from a typical sample. The typical sample is selected in such a way that its RASE-value is equal to the median in the 400 replications. The boxplot for 400 RASE-values is presented in Figure 2(c). To gauge the performance of our procedure in terms of RASE, we computed the standard deviation of the time series $\{x_t\}$, denoted by σ_X. The mean and the standard deviation of the σ_X, in the simulation with 400 replications, are 0.5389 and 0.0480 respectively. Overall, the proposed modeling procedure performs fairly well.

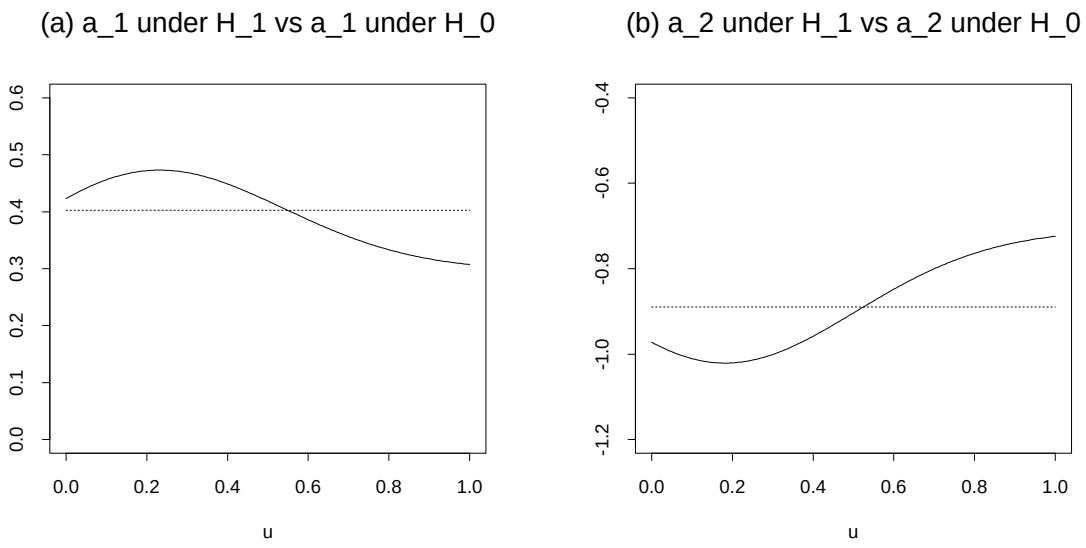

Figure 3: *The coefficient functions under the null hypothesis and a specific alternative hypothesis with* $\beta = 0.4$. *Solid curves are coefficient functions under* H_1 *and dashed lines are the coefficient functions under* H_0.

To demonstrate the power of the proposed bootstrap test, we consider the following null hypothesis:

$$H_0 : a_j(u) = \theta_j, \quad j = 1, 2,$$

namely a linear AR-model, against the alternative

$$H_1 : a_j(u) \neq \theta_j, \text{ for at least one } j.$$

The power function is evaluated under a sequence of the alternative models indexed by β:

$$H_1 : a_j(u) = \overline{a}_j^0 + \beta(a_j^0(u) - \overline{a}_j^0), \quad j = 1,\, 2 \quad (0 \leq \beta \leq 1),$$

where $\{a_j^0(u)\}$ are the solid curves given in Figures 2(a) and (b) and $\overline{a}_j^0$ is the average height of $a_j^0(u)$. We apply the goodness-of-fit test described in §4 in a simulation with 400 replication.

For each realization, we repeat bootstrap sampling 500 times. Figure 2(d) plots the simulated power function against β. When $\beta = 0$, the specified alternative hypothesis collapses into the null hypothesis. The power is 0.0470, which is close to the significant level 5%. This demonstrates that bootstrap estimate of the null distribution is approximately correct. The power function shows that our test is indeed powerful. To appreciate why, consider the specific alternative with $\beta = 0.4$. The functions $\{a_j(u)\}$ under H_1 are shown in Figure 3. The null hypothesis is essentially the constant curves in Figure 3. Even with such a small difference under our noise level, we can correctly detect the alternative over 80% of the 400 simulations. The power increases rapidly to one when $\beta = 0.8$. When $\beta = 1$, we test the constant functions in Figure 3 against the coefficient functions in Figures 2 (a) and (b).

Example 2. Now we consider a TAR model

$$x_t = a_1(x_{t-2})\, x_{t-1} + a_2(x_{t-2})\, x_{t-2} + \varepsilon_t, \tag{5.5}$$

where $a_1(u) = 0.4\, I(u \le 1) - 0.8\, I(u > 6)$, $a_2(u) = -0.6\, I(u \le 1) + 0.2\, I(u > 1)$, and $\{\varepsilon_t\}$ are i.i.d. $N(0,\, 1)$. With sample size $n = 500$, we replicate simulation 400 times. As in Example 1, the optimal bandwidth $h_n = 0.325$ is used. The boxplot for 400 RASE-values is presented in Figure 4(c). Further, the local linear estimators of $a_1(\cdot)$ and $a_2(\cdot)$ from a typical sample are plotted in Figures 4(a)-(b). The typical sample is selected in such a way that its RASE-value is equal to the median in the 400 replications.

To compare the prediction performance of the three predictors from functional-coefficient modeling with the best fitted linear AR(2) model

$$\widehat{x}_t = \widehat{\beta}_0 + \widehat{\beta}_1\, x_{t-1} + \widehat{\beta}_2\, x_{t-2},$$

we predict 10 post-sample points in each of the 400 replicated simulations. The mean and standard deviation (in parentheses) of average absolute predictive errors (AAPE) are recorded in Table 1. It is clear that the functional-coefficient autoregressive modeling while overparametrized provides more relevant predictors for the given model (5.5). Note that the direct predictor based on functional-coefficient model (5.3) performs reasonably well due to the flexibility of the functional-coefficient models.

Table 1: The mean and standard deviation of AAPE based on 400 replications

	One-step	Iterative two-step	Direct two-step
Model (5.5)	0.784(0.203)	0.904(0.273)	0.918(0.281)
Linear AR(2)	1.131(0.485)	1.117(0.496)	

5.2 Real-data examples

Example 3. We continue the discussion on Canadian lynx data in §2. To fit model (2.6), we select the bandwidth such that AMS(h) defined in (3.4) obtains its minimum. To this end, we let $Q = 4$ and $m = 11$. Figure 5(b) plots the AMS-values against h. The selected bandwidth is $h_n = 0.90$. The fitted values from both functional-coefficient model (2.6) and TAR model (2.5)

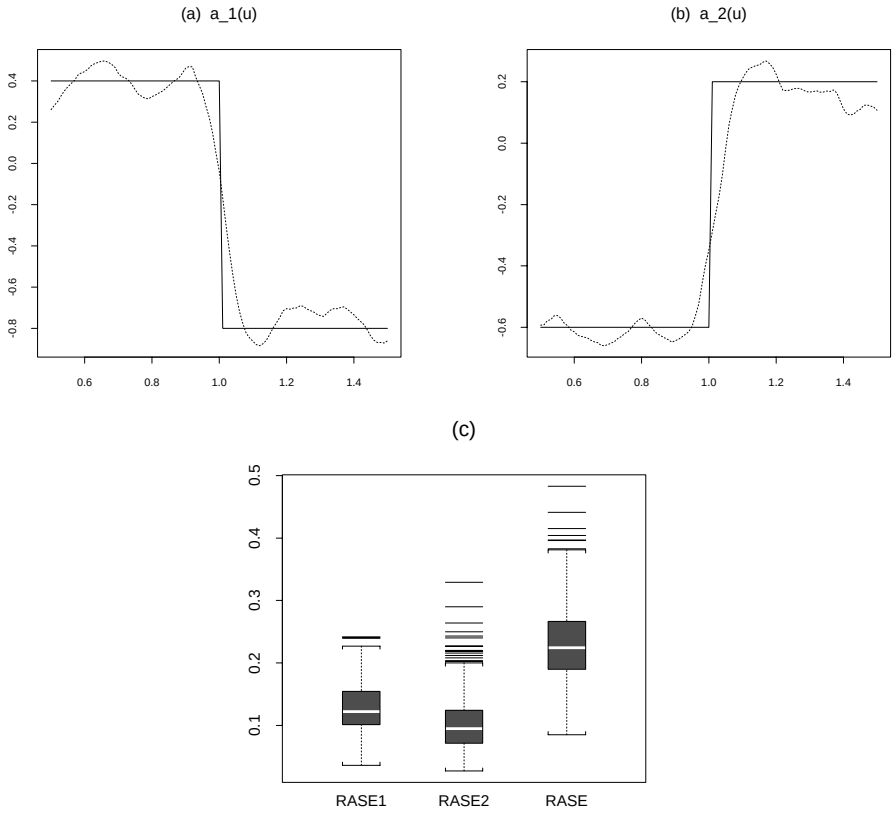

Figure 4: *Simulation results for Example 2. (a) The local linear estimator (dotted line) of the coefficient function $a_1(x_{t-2})$ (solid line). (b) The local linear estimator (dotted line) of $a_2(x_{t-2})$ (solid line). (c) The boxplot of the 400 RASE-values in estimation of $a_1(\cdot)$ and $a_2(\cdot)$.*

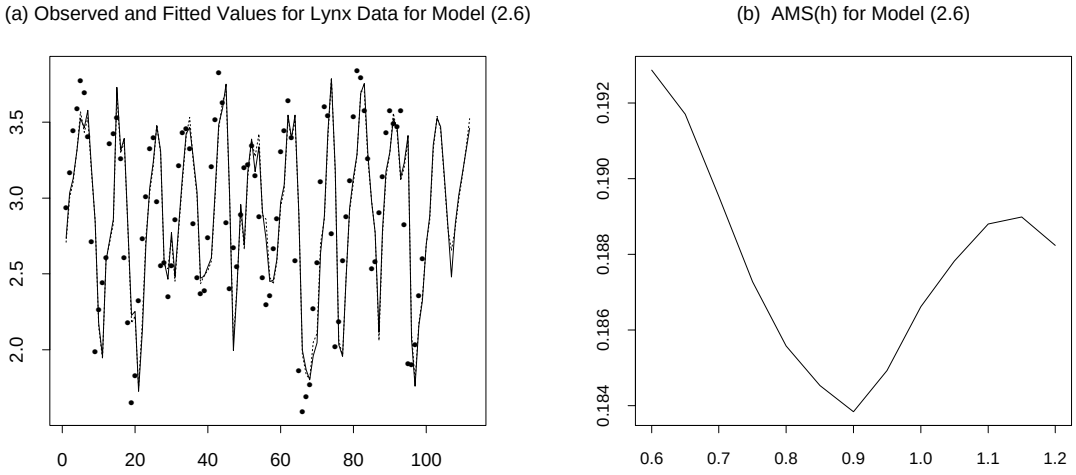

Figure 5: *Canadian Lynx data. (a) Time plots of the fitted values from TAR model (2.5) (solid line) and the fitted values of functional-coefficient model (2.6) (dashed line). The true values are indicated by "•". (b) The plot of the AMS against bandwidth.*

are very close to each other (see Figure 5(a)). Our goodness of fit test lends further support to the use of the TAR model. In fact, the residual sum of squares RSS_1 for model (2.6) is 0.0406, which is slightly smaller than $\mathrm{RSS}_0 = 0.0414$ for the TAR model (2.5). The p-value of the test is 0.714. Indeed, the TAR model (2.5) and the model (2.6) with coefficient functions given in Figure 1 are statistically indistinguishable for this data set. The difference lies in the interpretation of the two models (see §2). On the other hand, the p-value of the goodness-of-test to test for linear AR(2) model against the functional-coefficient model (2.6) is less than 0.001, which reinforces the existence of nonlinearity in lynx data.

To compare the prediction performance of various models, we estimate the functional-coefficient model (2.6), a TAR model and a linear AR(2) model using the first 102 data points only. We leave out last 12 points to check the prediction performance. The fitted TAR model is

$$\widehat{x}_t = \begin{cases} 0.424 + 1.255x_{t-1} - 0.348x_{t-2}, & x_{t-2} \le 2.981, \\ 1.882 + 1.516x_{t-1} - 1.126x_{t-2}, & x_{t-2} > 2.981. \end{cases} \tag{5.6}$$

The fitted linear AR(2) model is $\widehat{x}_t = 1.048 + 1.376x_{t-1} - 0.740x_{t-2}$. Both TAR and linear models are estimated using the least squares method. The threshold was searched among 60% inner sample points. The absolute prediction errors are reported in Table 2, which shows that the performance of functional-coefficient model is better than both TAR and linear AR(2) models. For example, for one-step ahead prediction, the average absolute predictive errors (AAPE) was reduced by 36% when the TAR model was used instead of linear AR(2) model. The AAPE was further reduced by 25% when the functional-coefficient model was used instead of the TAR model.

Table 2: The post-sample predictive errors for Canadian Lynx data

		Model (2.6)			TAR Model (5.6)		Linear AR(2)	
Year	x_t	One-step	Iterative	Direct	One-step	Iterative	One-step	Iterative
1923	3.054	0.157	0.156	0.209	0.187	0.090	0.173	0.087
1924	3.386	0.012	0.227	0.383	0.035	0.269	0.061	0.299
1925	3.553	0.021	0.035	0.195	0.014	0.038	0.106	0.189
1926	3.468	0.008	0.037	0.034	0.022	0.000	0.036	0.182
1927	3.187	0.085	0.101	0.295	0.059	0.092	0.003	0.046
1928	2.723	0.055	0.086	0.339	0.075	0.015	0.143	0.148
1929	2.686	0.135	0.061	0.055	0.273	0.160	0.248	0.051
1930	2.821	0.016	0.150	0.318	0.026	0.316	0.093	0.434
1931	3.000	0.017	0.037	0.111	0.030	0.062	0.058	0.185
1932	3.201	0.007	0.014	0.151	0.060	0.043	0.113	0.193
1933	3.424	0.089	0.098	0.209	0.076	0.067	0.191	0.347
1934	3.531	0.053	0.175	0.178	0.072	0.187	0.140	0.403
AAPE		0.055	0.095	0.206	0.073	0.112	0.114	0.214

Tong (1990) also suggested a further refined model involving 7 lagged variables

$$x_t = \begin{cases} 0.546 + 1.032\,x_{t-1} - 0.173\,x_{t-2} + 0.171\,x_{t-3} - 0.431\,x_{t-4} & \\ \quad +0.332\,x_{t-5} - 0.284\,x_{t-6} + 0.210\,x_{t-7} + \varepsilon_t^{(1)}, & \text{if } x_{t-2} \le 3.116, \\ 2.632 + 1.492\,x_{t-1} - 1.324\,x_{t-2} + \varepsilon_t^{(2)}, & \text{if } x_{t-2} > 3.116. \end{cases} \tag{5.7}$$

(See Tong, 1990, p.387). We fit the following more complex functional-coefficient model accordingly

$$x_t = \sum_{j=1}^{7} a_j(x_{t-2})\, x_{t-j} + \varepsilon_t. \tag{5.8}$$

The selected bandwidth is $h_n = 1.45$ (see Figure 6(c)). The estimated functions $a_j(\cdot)(1 \le j \le 7)$ are plotted in Figure 6(a), which shows that the dynamical change is predominantly dictated by $a_1(\cdot)$ and $a_2(\cdot)$. The fitted values of the two models are very close with each other, as shown in Figure 6(b). We apply the goodness-of-fit test to test for the TAR model (5.7) against the functional-coefficient model (5.8). The p-value is 0.883, which again supports the use of the TAR

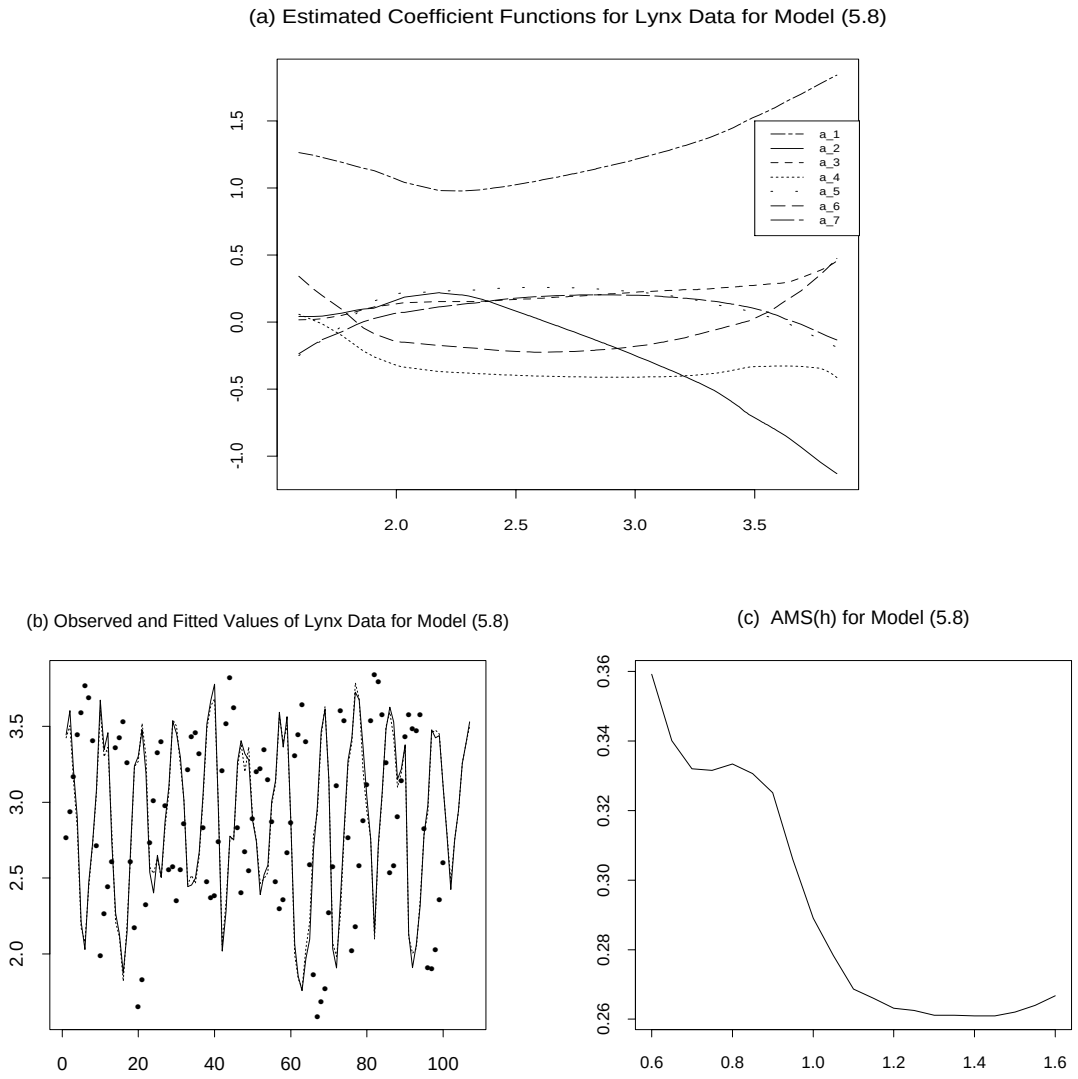

Figure 6: *Canadian Lynx data. (a) The estimated curves for functional coefficients* $a_i(x_{t-2})$ $(i = 1, \ldots, 7)$ *in model (5.9). (b) The time plots of the fitted values from model (5.8) (solid line) and the fitted values from model (5.9) (dashed line). The true values are indicated by "•". (c) The plot of the AMS against bandwidth.*

model for the lynx data.

Example 4. In many respects, Wolf's annual sunspot numbers are known to be challenging (see, for example, Tong 1990). Following the convention in the literature, we apply the transform $x_t = 2\,(\sqrt{1+y_t}-1)$ to the 288 annual sunspot numbers in 1700 - 1987 first. (See, for example, Ghaddar and Tong 1981, and Chen and Tsay 1993.) We apply the technique proposed in Section 3.3 to select the *optimum* functional-coefficient models among the class of models $x_t = \sum_{j=1}^{p} a_j(x_{t-d})\, x_{t-j} + \varepsilon_t$ with $1 \le d \le p$ and $2 \le p \le 11$. We let $m = 28$ and $Q = 4$ in AMS defined as in (3.4). Table 3

Table 3: The selected functional-coefficient models for Sunspot Data

p	2	3	4	5	6	7	8	9	10	11
d	1	3	3	2	2	3	3	5	3	5
AMS	18.69	13.46	13.90	12.26	13.93	11.68	11.95	14.06	14.26	13.91

records the best model with each value of p between 2 and 11. The overall optimum model should be of the order $p = 7$ or 8, and the smooth variable at lag $d = 3$.

Note that the FAR model proposed by Chen and Tsay (1993, p.305) is

$$x_t = \begin{cases} 1.23 + (1.75 - 0.17\,|x_{t-3} - 6.6|)\, x_{t-1} + (-1.28 + 0.27\,|x_{t-3} - 6.6|)\, x_{t-2} \\ \qquad +0.20\, x_{t-8} + \varepsilon_t^{(1)}, & \text{if } x_{t-3} < 10.3, \\ 0.92 - 0.24\, x_{t-3} + 0.87\, x_{t-1} + 0.17\, x_{t-2} - 0.06\, x_{t-6} + 0.04\, x_{t-8} + \varepsilon_t^{(2)}, & \text{if } x_{t-3} \ge 10.3. \end{cases} \tag{5.9}$$

Combining this with the aforementioned result from the model selection, we fit the data with the functional-coefficient model

$$x_t = a_1(x_{t-3})\, x_{t-1} + a_2(x_{t-3})\, x_{t-2} + a_3(x_{t-3})\, x_{t-3} + a_6(x_{t-3})\, x_{t-6} + a_8(x_{t-3})\, x_{t-8} + \varepsilon_t. \tag{5.10}$$

The estimated coefficient functions are plotted in Figures 7(a)–(e). The selected bandwidth is $h_n = 4.75$ (see Figure 7(f)), which minimizes the AMS defined as in (3.4).

To compare the prediction performance, the first 280 data points (in 1700-1979) are used to estimate the coefficient functions in (5.10). Table 4 reports the absolute errors in predicting the

Table 4: The post-sample predictive errors for Sunspot data

		Model (5.10)			FAR Model (5.9)		TAR Model (5.11)	
Year	x_t	One-step	Iterative	Direct	Error	Iterative	Error	Iterative
1980	154.7	1.4	1.4	1.4	13.8	13.8	5.5	5.5
1981	140.5	11.4	10.4	3.7	0.0	3.8	1.3	0.0
1982	115.9	15.7	20.7	12.9	10.0	16.4	19.5	22.1
1983	66.6	10.3	0.7	11.0	3.3	0.8	4.8	6.5
1984	45.9	1.0	1.5	4.3	3.8	5.6	14.8	15.9
1985	17.9	2.6	3.4	7.8	4.6	1.7	0.2	2.7
1986	13.4	3.1	0.7	1.9	1.3	2.5	5.5	5.4
1987	29.2	12.3	13.1	18.9	21.7	23.6	0.7	17.5
AAPE		7.2	6.5	7.7	7.3	8.3	6.6	9.5

sunspot numbers in 1980 – 1987 from the newly estimated model (5.10) as well as those from the

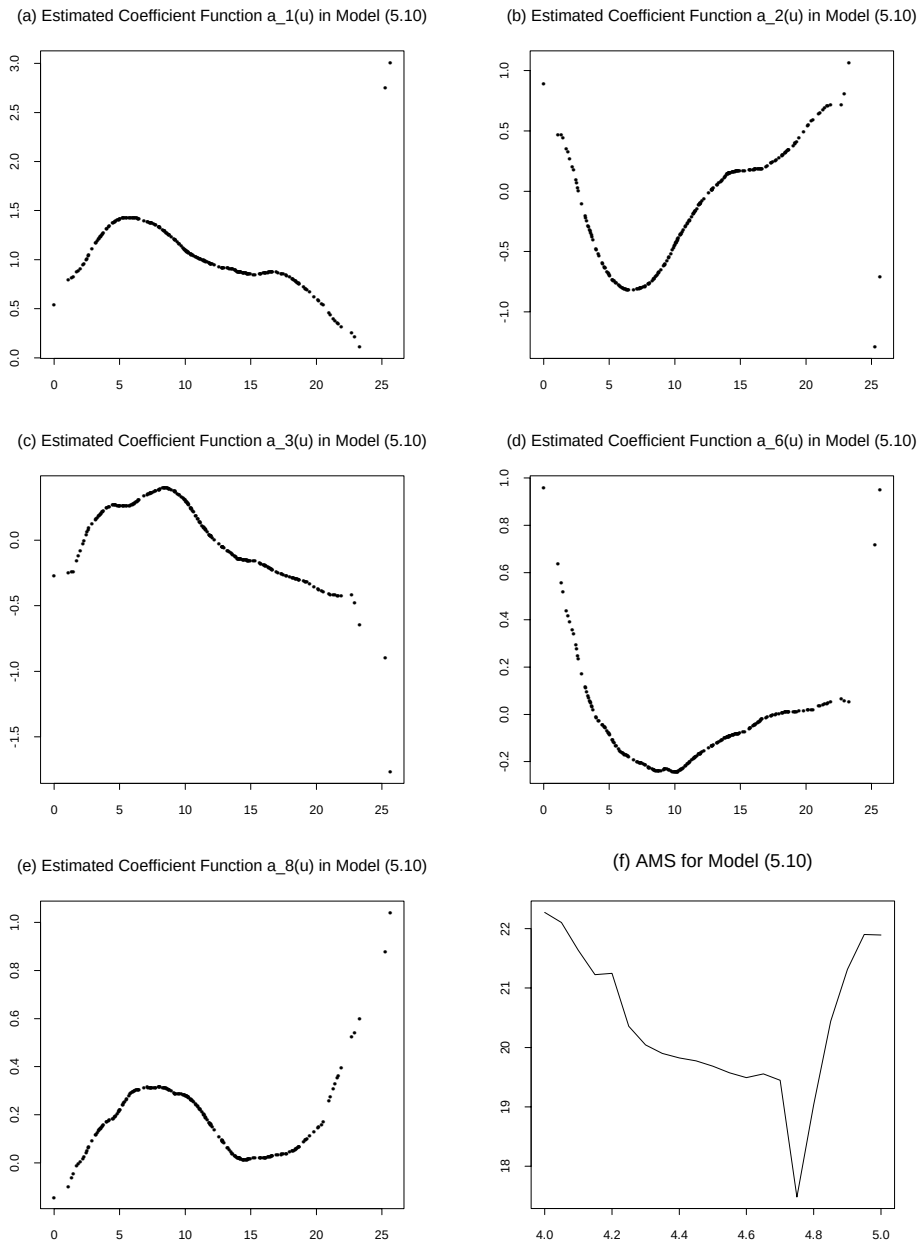

Figure 7: *Wolf's sunspot numbers. (a)-(e) The estimated functional coefficients in model (5.10). The x-axis is x_{t-3}. (f) The plot of the AMS against bandwidth (for estimation of model (5.10)).*

FAR model (5.9) and the following TAR model (Tong 1990, p.420)

$$x_t = \begin{cases} 1.92 + 0.84\,x_{t-1} + 0.07\,x_{t-2} - 0.32\,x_{t-3} + 0.15\,x_{t-4} - 0.20\,x_{t-5} - 0.00\,x_{t-6} & \\ +0.19\,x_{t-7} - 0.27\,x_{t-8} + 0.21\,x_{t-9} + 0.01\,x_{t-10} + 0.09\,x_{t-11} + \varepsilon_t^{(1)}, & \text{if } x_{t-8} \le 11.93, \\ 4.27 + 1.44\,x_{t-1} - 0.84\,x_{t-2} + 0.06\,x_{t-3} + \varepsilon_t^{(2)}, & \text{if } x_{t-8} > 11.93. \end{cases} \tag{5.11}$$

Note that both models (5.9) and (5.11) were estimated using the first 280 sample points too (see Tong 1990, p.420 and Chen and Tsay 1993, p.304). According to the average absolute predictive

errors, the functional-coefficient model performs as well as both TAR and FAR models in one-step ahead prediction. Furthermore, it performs better in two-step prediction with both iterative and direct methods.

Finally we apply the goodness-of-fit test to test the hypothesis of the FAR model (5.9) against the nonparametric model (5.10). The residual sum of squares RSS_1 for model (5.10) is 2.932 in contrast to $\mathrm{RSS}_0 = 3.277$ for the FAR model. The p-value for this test is 0.454, which lends further support to using the FAR model in this context. We also apply the goodness-of-fit test to test the hypothesis of the TAR model (5.11) against the functional-coefficient model

$$x_t = \sum_{j=1}^{11} a_j(x_{t-8})\, x_{t-j} + \varepsilon_t. \tag{5.12}$$

The residual sum of squares RSS_1 for model (5.12) is 2.077, which is about 43.64% smaller than $\mathrm{RSS}_0 = 3.685$ for the TAR model. The p-value of the test is 0.101.

6 Asymptotic Results

Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{(U_j,\, \mathbf{X}_j,\, Y_j);\, a \le j \le b\}$. Denote by

$$\alpha(k) = \sup_{\substack{A \in \mathcal{F}_{-\infty}^{0} \\ B \in \mathcal{F}_{k}^{\infty}}} |P(AB) - P(A)P(B)|$$

The coefficient $\alpha(k)$ is called the strong mixing coefficient of the stationary processes $\{U_j,\, \mathbf{X}_j,\, Y_j\}_{j=-\infty}^{\infty}$. If $\alpha(k) \to 0$ as $k \to \infty$, the processes $\{U_j,\, \mathbf{X}_j,\, Y_j\}_{j=-\infty}^{\infty}$ are called strongly mixing.

Among various mixing conditions used in literature, α-mixing is reasonably weak, and is known to be fulfilled for many stochastic processes including many time series models. Gorodetskii (1977) and Withers (1981) have derived the conditions under which a linear process is α-mixing. In fact, under very mild assumptions linear autoregressive and more generally bilinear time series models are strongly mixing with mixing coefficients decaying exponentially. Auestad and Tjøstheim (1990) provided illuminating discussions on the role of α-mixing (including geometric ergodicity) for model identification in nonlinear time series analysis. Chen and Tsay (1993) showed that the FAR process defined in (2.1) is geometrically ergodic under certain conditions. Further, Masry and Tjøstheim (1995, 1997) showed that under some mild conditions, both autoregressive conditional heteroscedastic (ARCH) process and additive autoregressive process with exogenous variables (NAARX), which are particularly popular in finance, are stationary and α-mixing.

We first present a result on mean square convergence which serves as a building block to our main result and is also of independent interest. We introduce some notation now. Let

$$\mathbf{S}_n = \mathbf{S}_n(u_0) = \begin{pmatrix} \mathbf{S}_{n,0} & \mathbf{S}_{n,1} \\ \mathbf{S}_{n,1} & \mathbf{S}_{n,2} \end{pmatrix}, \qquad \text{and} \quad T_n = T_n(u_0) = \begin{pmatrix} T_{n,0}(u_0) \\ T_{n,1}(u_0) \end{pmatrix} \tag{6.1}$$

with

$$\mathbf{S}_{n,j} = \mathbf{S}_{n,j}(u_0) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i\, \mathbf{X}_i^T \left(\frac{U_i - u_0}{h}\right)^j K_h(U_i - u_0), \tag{6.2}$$

and

$$T_{n,j}(u_0) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \left(\frac{U_i - u_0}{h}\right)^j K_h(U_i - u_0)\, Y_i. \tag{6.3}$$

Then, the solution to (3.1) can be expressed as

$$\widehat{\boldsymbol{\beta}} = \mathbf{H}^{-1}\, \mathbf{S}_n^{-1}\, T_n, \tag{6.4}$$

where $\mathbf{H} = \mathrm{diag}\,(1, \ldots, 1, h, \ldots, h)$ with p-diagonal elements ones and p diagonal elements h's. To facilitate the notation, we denote

$$\mu_j = \int_{-\infty}^{\infty} u^j\, K(u)\, du, \qquad \nu_j = \int_{-\infty}^{\infty} u^j\, K^2(u)\, du,$$

and

$$\Omega = (\omega_{l,m})_{p\times p} = E\left(\mathbf{X}\,\mathbf{X}^T \,|\, U = u_0\right). \tag{6.5}$$

Also, let $f(\mathbf{x},\, u)$ denote the joint density of $(\mathbf{X},\, U)$ and $f_{\mathrm{U}}(u)$ be the marginal density of U. We use the following convention: if $U = X_{j_0}$ for some $1 \le j_0 \le p$, then $f(\mathbf{x},\, u)$ becomes $f(\mathbf{x})$ the joint density of $\mathbf{X}$.

Theorem 1. Let Condition 1 in the Appendix hold, and $f(\mathbf{x},\, u)$ be continuous at the point u_0. Let $h_n \to 0$ and $n\,h_n \to \infty$, as $n \to \infty$. Then it holds that

$$E(\mathbf{S}_{n,j}(u_0)) \to f_{\mathrm{U}}(u_0)\, \Omega(u_0)\, \mu_j, \qquad \text{and} \qquad n\,h_n\, \mathrm{Var}(\mathbf{S}_{n,j}(u_0)_{l,m}) \to f_{\mathrm{U}}(u_0)\, \nu_{2j}\, \omega_{l,m}$$

for each $0 \le j \le 3$ and $1 \le l,\ m \le p$.

As a consequence of Theorem 1, we have

$$\mathbf{S}_n \quad \xrightarrow{\mathcal{P}} \quad f_{\mathrm{U}}(u_0)\, \mathbf{S}, \qquad \text{and} \qquad \mathbf{S}_{n,3} \quad \xrightarrow{\mathcal{P}} \quad \mu_3\, f_{\mathrm{U}}(u_0)\, \Omega$$

in the sense that each element converges in probability, where

$$\mathbf{S} = \begin{pmatrix} \Omega & \mu_1\, \Omega \\ \mu_1\, \Omega & \mu_2\, \Omega \end{pmatrix}.$$

Put

$$\sigma^2(\mathbf{x},\, u) = \mathrm{Var}(Y \,|\, \mathbf{X} = \mathbf{x},\, U = u), \tag{6.6}$$

and

$$\Omega^*(u_0) = E\left[\mathbf{X}\,\mathbf{X}^T\, \sigma^2(\mathbf{X},\, U) \,|\, U = u_0\right]. \tag{6.7}$$

Denote by $c_0 = \mu_2/\left(\mu_2 - \mu_1^2\right)$ and $c_1 = -\mu_1/\left(\mu_2 - \mu_1^2\right)$.

Theorem 2. Let $\sigma^2(\mathbf{x},\, u)$ and $f(\mathbf{x},\, u)$ be continuous at the point u_0. Then under Conditions 1 and 2 in the Appendix,

$$\sqrt{n\,h_n}\left[\widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{h^2}{2}\, \frac{\mu_2^2 - \mu_1\,\mu_3}{\mu_2 - \mu_1^2}\, \mathbf{a}''(u_0)\right] \quad \xrightarrow{\mathcal{D}} \quad N\left(0,\ \Theta^2(u_0)\right), \tag{6.8}$$

provided that $f_{\rm U}(u_0) \neq 0$, where

$$\Theta^2(u_0) = \frac{c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2}{f_{\rm U}(u_0)}\,\Omega^{-1}(u_0)\,\Omega^*(u_0)\,\Omega^{-1}(u_0). \tag{6.9}$$

Theorem 2 indicates that the asymptotic bias of $\widehat{a}_j(u_0)$ is

$$\frac{h^2}{2}\,\frac{\mu_2^2 - \mu_1\,\mu_3}{\mu_2 - \mu_1^2}\,a_j''(u_0),$$

and the asymptotic variance is $(n\,h_n)^{-1}\,\theta_j^2(u_0)$, where

$$\theta_j^2(u_0) = \frac{c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2}{f_{\rm U}(u_0)}\,e_{j,p}^T\,\Omega^{-1}(u_0)\,\Omega^*(u_0)\,\Omega^{-1}(u_0)\,e_{j,p}.$$

When $\mu_1 = 0$, the bias and variance expressions can be simplified respectively as $\frac{h^2}{2}\mu_2 a_j''(u_0)$ and

$$\theta_j^2(u_0) = \frac{\nu_0}{f_{\rm U}(u_0)}\,e_{j,p}^T\,\Omega^{-1}(u_0)\,\Omega^*(u_0)\,\Omega^{-1}(u_0)\,e_{j,p}.$$

The optimal bandwidth for estimating $a_j(\cdot)$ can be defined to be the one which minimizes the squared bias plus variance. The optimal bandwidth is given by

$$h_{j,\rm opt} = \left[\frac{\mu_2^2\,\nu_0 - 2\,\mu_1\,\mu_2\,\nu_1 + \mu_1^2\,\nu_2}{f_{\rm U}(u_0)\;(\mu_2^2 - \mu_1\,\mu_3)^2}\;\frac{e_{j,p}^T\,\Omega^{-1}(u_0)\,\Omega^*(u_0)\,\Omega^{-1}(u_0)\,e_{j,p}}{\left\{a_j''(u_0)\right\}^2}\right]^{1/5} n^{-1/5}. \tag{6.10}$$

Recently, Fan and Gijbels (1995) and Ruppert *et al.* (1995) developed data-driven bandwidth selection schemes based on asymptotic formulas for the optimal bandwidths, which are less variable and more effective than the conventional data-driven bandwidth selectors such as cross-validation bandwidth rule. The similar algorithms can be developed for estimation of functional-coefficient models based on (6.10), which is however beyond the scope of this paper.

References

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika*, **77**, 669–687.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis, Forecasting, and Control.* Holden Day, San Francisco.

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

Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association*, **93**, 961–976.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**, 298–308.

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1992). Local regression models. In *Statistical Models in S* (Chambers, J.M. and Hastie, T.J., eds), 309–376. Wadsworth & Brooks, Pacific Grove.

Dahlhaus, R. (1989). Efficient parameter estimation for self similar processes. *The Annals of Statistics*, **17**, 1749-1766.

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

Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica*, **55**, 251-276.

Fan, J. (1993). Local linear regression smoothers and their minimax. *The Annals of Statistics*, **21**, 196–216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, **20**, 2008-2036.

Fan, J. and Gijbels, I. (1995). Data driven bandwidth selection in local polynomial fitting: Variable bandwidth spatial adaptation. *Journal of the Royal Statistical Society, Series B*, **57**, 371–394.

Fan, J. and Zhang, W. (1997). Statistical estimation in varying coefficient models. Submitted to *The Annals of Statistics*, revised.

Ghaddar, D.K. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *Journal of the Royal Statistical Society, Series C*, **30**, 238–248.

Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. *Theory of Probability and Its Applications*, **22**, 411–413.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15-29.

Granger, C.W.J. and Teräsvirta, T. (1993). *Modeling Nonlinear Dynamic Relationships.* Oxford University Press, Oxford.

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

Haggan, V. and Ozaki, T. (1981). Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, **68**, 189–196.

Hannan, E.J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems.* Wiley, New York.

Härdle, W., Lütkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, **65**, 49–72.

Hastie, T.J. and Loader, C. (1993). Local regression: automatic kernel carpentry (with discussion). *Statistical Science*, **8**, 120-43.

Hastie, T.J. and Tibshirani, R.J. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B*, **55**, 757-796.

Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1997). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, to appear.

Ibragimov, I.A. and Linnik, Yu.V. (1971). *Independent and Stationary Sequences of Random Variables.* Walters-Noordhoff, Groningen, the Netherlands.

Kreiss, J.P., Neumann, M. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. (Submitted.)

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165–179.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory*, **11**, 258–289.

Masry, E. and Tjøstheim, D. (1997). Additive nonlinear ARX time series and projection estimates. *Econometric Theory*, **13**, 214–252.

Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statistics, No. 11. Springer-Verlag, New York.

Ozaki, T. (1982). The statistical analysis of perturbed limit cycle processes using nonlinear time series models. *Journal of Time Series Analysis*, **3**, 29–41.

Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis.* Springer-Verlag, Berlin.

Ruppert, D. and Wand, M.P. (1994). Multivariate weighted least squares estimation. *The Annals of Statistics*, **22**, 1346–1370.

Shao, Q. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability*, **24**, 2098–2127.

Stenseth, N.C., Falck, W., Chan, K.S., Bjørnstad, O.N., Tong, H., O'Donoghue, M., Boonstra, R., Boutin, S., Krebs, C.J., and Yoccoz, N.G. (1999). From patterns to processes: phase- and density-dependencies in Canadian lynx cycle. *Proceedings of National Academy of Science, Washington*, to appear.

Sun, Z. (1984). Asymptotic unbiased and strong consistency for density function estimator. *Acta Mathematica Sinica*, **27**, 769–782.

Tiao, G.C. and Tsay, R.S. (1994). Some advances in nonlinear and adaptive modeling in time series. *Journal of Forecasting*, **13**, 109-131.

Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scandinavian Journal of Statistics*, **21**, 97-130.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective (with discussion). *Scandinavian Journal of Statistics*, **22**, 399-445.

Volkonskii, V.A. and Rozanov, Yu.A. (1959). Some limit theorems for random functions. I. *Theory of Probability and Its Applications*, **4**, 178–197.

Withers, C.S. (1981). Conditions for linear processes to be strong mixing. *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, **57**, 477–480.

Yao, Q. and Tong, H. (1995). On initial-condition sensitivity and prediction in nonlinear stochastic systems. *Bulletin of International Statistical Institute*, **IP 10.3**, 395-412.

Acknowledgement

We thank the Editor, the Associate Editor and two referees for their insightful comments, which let a substantial improvement of the paper.

Appendix

We first impose some conditions on the regression model. They are not the weakest possible.

Condition 1

(i) The kernel function $K(\cdot)$ is a bounded density with a bounded support $[-1,\ 1]$.

(ii) $|f(\mathbf{x}_0,\, u;\ \mathbf{x}_1,\, v;\ l) - f(\mathbf{x}_0,\, u)\, f(\mathbf{x}_1,\, v)| \le M$, for all $l \ge 1$, where $f(\mathbf{x}_0,\, u,\ \mathbf{x}_1,\, v;\ l)$ is the joint density of $(\mathbf{X}_0,\, U_0)$ and $(\mathbf{X}_l,\, U_l)$.

(iii) The processes $\{U_i,\, \mathbf{X}_i,\, Y_i\}$ are α-mixing with $\sum k^c [\alpha(k)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.

(iv) $E|\mathbf{X}|^{2\,\delta} < \infty$, where δ is given in Condition 1(iii).

Condition 2

(i) Assume that

$$f(\mathbf{x}_0,\, u;\ \mathbf{x}_1,\, v;\ l) \le M_1, \quad \text{and} \quad E\left\{Y_0^2 + Y_l^2 \,|\, \mathbf{X}_0 = \mathbf{x}_0,\, U_0 = u;\ \mathbf{X}_l = \mathbf{x}_1,\, U_l = v\right\} \le M_2, \tag{A.1}$$

for all $l \ge 1,\ \mathbf{x}_0,\ \mathbf{x}_1 \in \Re^p,\ u$ and v in a neighborhood of u_0.

(ii) Assume that $h_n \to$ and $n\,h_n \to \infty$. Further, assume that there exists a sequence of positive integers s_n such that $s_n \to \infty,\ s_n = o\left((n\,h_n)^{1/2}\right)$, and $(n/h_n)^{1/2}\,\alpha(s_n) \to 0$, as $n \to \infty$.

(iii) There exists $\delta^* > \delta$, where δ is given in Condition 1(iii), such that

$$E\left\{|Y|^{\delta^*}\,|\mathbf{X}=\mathbf{x},\,U=u\right\} \le M_4 < \infty \tag{A.2}$$

for all $\mathbf{x} \in \Re^p$ and u in a neighborhood of u_0, and

$$\alpha(n) = O\left(n^{-\theta^*}\right), \tag{A.3}$$

where $\theta^* \ge \delta\,\delta^*/\{2(\delta^*-\delta)\}$.

(iv) $E|\mathbf{X}|^{2\,\delta^*} < \infty$, and $n^{1/2-\delta/4}\,h^{\delta/\delta^*-1} = O(1)$.

Remark 1. We provide a sufficient condition for the mixing coefficient $\alpha(n)$ to satisfy Conditions 1(iii) and 2(ii). Suppose that $h_n = A\,n^{-\rho}(0 < \rho < 1,\ A > 0)$, $s_n = (n\,h_n/\log n)^{1/2}$ and $\alpha(n) = O\left(n^{-d}\right)$ for some $d > 0$. Then Condition 1(iii) is satisfied for $d > 2(1-1/\delta)/(1-2/\delta)$ and Condition 2(ii) is satisfied if $d > (1+\rho)/(1-\rho)$. Hence both conditions are satisfied if

$$\alpha(n) = O\left(n^{-d}\right), \qquad d > \max\left\{\frac{1+\rho}{1-\rho},\ \frac{2(1-1/\delta)}{1-2/\delta}\right\}.$$

Note that this is a trade off between the order δ of the moment of Y and the rate of decay of the mixing coefficient; the larger the order δ, the weaker is the decay rate of $\alpha(n)$.

To study the joint asymptotic normality of $\widehat{\mathbf{a}}(u_0)$, we need to center the vector $T_n(u_0)$ by replacing Y_i with $Y_i - m(\mathbf{X}_i,\,U_i)$ in the expression (6.3) of $T_{n,j}(u_0)$. Let

$$T^*_{n,j}(u_0) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\left(\frac{U_i-u_0}{h}\right)^j K_h(U_i-u_0)\,[Y_i - m(\mathbf{X}_i,\,U_i)], \qquad \text{and} \qquad T^*_n = \begin{pmatrix} T^*_{n,0} \\ T^*_{n,1}\end{pmatrix}.$$

Since the coefficient functions $a_j(u)$ are conducted in the neighborhood of $|U_i - u_0| < h$, by Taylor's expansion,

$$m(\mathbf{X}_i,\,U_i) = \mathbf{X}_i^T\,\mathbf{a}(u_0) + (U_i-u_0)\,\mathbf{X}_i^T\,\mathbf{a}'(u_0) + \frac{h^2}{2}\left(\frac{U_i-u_0}{h}\right)^2\mathbf{X}_i^T\,\mathbf{a}''(u_0) + o_p(h^2),$$

where $\mathbf{a}'(u_0)$ and $\mathbf{a}''(u_0)$ are the vectors consisting of the first and the second derivative of the functions $a_j(\cdot)$. Then,

$$T_{n,0} - T^*_{n,0} = \mathbf{S}_{n,0}\,\mathbf{a}(u_0) + h\,\mathbf{S}_{n,1}\,\mathbf{a}'(u_0) + \frac{h^2}{2}\,\mathbf{S}_{n,2}\,\mathbf{a}''(u_0) + o_p(h^2),$$

and

$$T_{n,1} - T^*_{n,1} = \mathbf{S}_{n,1}\,\mathbf{a}(u_0) + h\,\mathbf{S}_{n,2}\,\mathbf{a}'(u_0) + \frac{h^2}{2}\,\mathbf{S}_{n,3}\,\mathbf{a}''(u_0) + o_p(h^2),$$

so that

$$T_n - T^*_n = \mathbf{S}_n\,\mathbf{H}\,\boldsymbol{\beta} + \frac{h^2}{2}\begin{pmatrix}\mathbf{S}_{n,2}\\ \mathbf{S}_{n,3}\end{pmatrix}\mathbf{a}''(u_0) + o_p(h^2), \tag{A.4}$$

where $\boldsymbol{\beta} = (\mathbf{a}(u_0)^T,\,\mathbf{a}'(u_0)^T)^T$. Therefore, it follows from (6.4), (A.4) and Theorem 1 that

$$\mathbf{H}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = f_{\mathrm{U}}^{-1}(u_0)\,\mathbf{S}^{-1}\,T^*_n + \frac{h^2}{2}\,\mathbf{S}^{-1}\begin{pmatrix}\mu_2\,\Omega\\ \mu_3\Omega\end{pmatrix}\mathbf{a}''(u_0) + o_p(h^2) \tag{A.5}$$

from which the bias term of $\widehat{\boldsymbol{\beta}}(u_0)$ is evident. Clearly,

$$\widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) = \frac{\Omega^{-1}}{f_{\mathrm{U}}(u_0)\ (\mu_2 - \mu_1^2)}\ \left[\mu_2\, T^*_{n,0} - \mu_1\, T^*_{n,1}\right] + \frac{h^2}{2}\,\frac{\mu_2^2 - \mu_1\,\mu_3}{\mu_2 - \mu_1^2}\,\mathbf{a}''(u_0) + o_p(h^2). \qquad \text{(A.6)}$$

Thus, (A.6) indicates that the asymptotic bias of $\widehat{\mathbf{a}}(u_0)$ is

$$\frac{h^2}{2}\,\frac{\mu_2^2 - \mu_1\,\mu_3}{\mu_2 - \mu_1^2}\,\mathbf{a}''(u_0).$$

Let

$$\mathbf{Q}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i, \qquad \text{(A.7)}$$

where

$$\mathbf{Z}_i = \mathbf{X}_i\ \left[c_0 + c_1\ \left(\frac{U_i - u_0}{h}\right)\right] K_h(U_i - u_0)\,[Y_i - m(\mathbf{X}_i,\, U_i)] \qquad \text{(A.8)}$$

with $c_0 = \mu_2/\ (\mu_2 - \mu_1^2)$ and $c_1 = -\mu_1/\ (\mu_2 - \mu_1^2)$. It follows from (A.6) and (A.7) that

$$\sqrt{n\,h_n}\left[\widehat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{h^2}{2}\,\frac{\mu_2^2 - \mu_1\,\mu_3}{\mu_2 - \mu_1^2}\,\mathbf{a}''(u_0)\right] = \frac{\Omega^{-1}}{f_{\mathrm{U}}(u_0)}\,\sqrt{n\,h_n}\,\mathbf{Q}_n + o_p(1). \qquad \text{(A.9)}$$

We need the following lemma, whose proof is more involved than that for Theorem 1. Therefore, we only prove this lemma. Throughout this appendix, we denote by C a generic constant, which may take different values at different places.

Lemma 1. Under Conditions 1 and 2 and the assumption that $h_n \to 0$ and $n\,h_n \to \infty$, as $n \to \infty$, if $\sigma^2(\mathbf{x},\, u)$ and $f(\mathbf{x},\, u)$ are continuous at the point u_0, then, we have

(a) $h_n\,\mathrm{Var}(\mathbf{Z}_1) \to f_{\mathrm{U}}(u_0)\,\Omega^*(u_0)\ [c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2]$;

(b) $h_n\ \sum_{l=1}^{n-1|}|\mathrm{Cov}(\mathbf{Z}_1,\, \mathbf{Z}_{l+1})| = o(1)$;

(c) $n\,h_n\,\mathrm{Var}(\mathbf{Q}_n) \to f_{\mathrm{U}}(u_0)\,\Omega^*(u_0)\ [c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2]$.

Proof: First of all, by conditioning on $(\mathbf{X}_1,\, U_1)$ and using Theorem 1 in Sun (1984), we have

$$\begin{aligned}\mathrm{Var}(\mathbf{Z}_1) &= E\left[\mathbf{X}_1\,\mathbf{X}_1^T\,\sigma^2(\mathbf{X}_1,\, U_1)\left\{c_0 + c_1\ \left(\frac{U_1 - u_0}{h}\right)\right\}^2\ K_h^2(U_1 - u_0)\right] \\ &= \frac{1}{h}\left[f_{\mathrm{U}}(u_0)\,\Omega^*(u_0)\ \left\{c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2\right\} + o(1)\right]. \qquad \text{(A.10)}\end{aligned}$$

The result (c) follows in an obvious manner from (a) and (b) along with

$$\mathrm{Var}(\mathbf{Q}_n) = \frac{1}{n}\,\mathrm{Var}(\mathbf{Z}_1) + \frac{2}{n}\,\sum_{l=1}^{n-1}\left(1 - \frac{l}{n}\right)\ \mathrm{Cov}(\mathbf{Z}_1,\, \mathbf{Z}_{l+1}). \qquad \text{(A.11)}$$

Therefore, it remains to prove part (b). To this end, let $d_n \to \infty$ be a sequence of positive integers such that $d_n\, h_n \to 0$. Define

$$J_1 = \sum_{l=1}^{d_n - 1} |\mathrm{Cov}(\mathbf{Z}_1,\, \mathbf{Z}_{l+1})|, \qquad \text{and} \qquad J_2 = \sum_{l=d_n}^{n-1} |\mathrm{Cov}(\mathbf{Z}_1,\, \mathbf{Z}_{l+1})|.$$

It remains to show that $J_1 = o\left(h^{-1}\right)$ and $J_2 = o\left(h^{-1}\right)$.

We remark that since $K(\cdot)$ has a bounded support $[-1,\ 1]$, $a_j(u)$ is bounded in the neighborhood of $u \in [u_0 - h,\ u_0 + h]$. Let $B = \max_{1 \le j \le p} \sup_{|u-u_0|<h} |a_j(u)|$ and $g(\mathbf{x}) = \sum_{j=1}^{p} |x_j|$. Then $\sup_{|u-u_0|<h} |m(\mathbf{x},\, u)| \le B\, g(\mathbf{x})$. By conditioning on $(\mathbf{X}_1,\, U_1)$ and $(\mathbf{X}_{l+1},\, U_{l+1})$, and using (A.1) and Condition 2(iii), we have, for all $l \ge 1$,

$$\begin{aligned}
&|\mathrm{Cov}(\mathbf{Z}_1,\, \mathbf{Z}_{l+1})| \\
&\le C\, E\left[|\mathbf{X}_1\, \mathbf{X}_{l+1}^T|\, \{|Y_1| + B\, g(\mathbf{X}_1)\}\{|Y_{l+1}| + B\, g(\mathbf{X}_{l+1})\}\, K_h(U_1 - u_0)\, K_h(U_{l+1} - u_0)\right] \\
&\le C\, E\left[|\mathbf{X}_1 \mathbf{X}_{l+1}^T|\, \left\{M_2 + B^2 g^2(\mathbf{X}_1)\right\}^{1/2} \left\{M_2 + B^2 g^2(\mathbf{X}_{l+1})\right\}^{1/2} K_h(U_1 - u_0) K_h(U_{l+1} - u_0)\right] \\
&\le C\, E\left[|\mathbf{X}_1\, \mathbf{X}_{l+1}^T|\, \{1 + g(\mathbf{X}_1)\}\, \{1 + g(\mathbf{X}_{l+1})\}\right] \\
&\le C. \qquad (A.12)
\end{aligned}$$

It follows that

$$J_1 \le C\, d_n = o\left(h^{-1}\right)$$

by the choice of d_n. Next we consider the upper bound of J_2. To this end, by using Davydov's inequality (see Hall and Heyde 1980, Corollary A.2), we obtain, for all $1 \le j,\ m \le p$ and $l \ge 1$,

$$|\mathrm{Cov}(Z_{1j},\, Z_{l+1,m})| \le C\, [\alpha(l)]^{1-2/\delta}\, \left[E|Z_j|^\delta\right]^{1/\delta}\, \left[E|Z_m|^\delta\right]^{1/\delta}. \qquad (A.13)$$

By conditioning on $(\mathbf{X},\, U)$ and using Condition 2(iii), one has

$$\begin{aligned}
E\left[|Z_j|^\delta\right] &\le C\, E\left[|X_j|^\delta\, K_h^\delta(U - u_0)\left\{|Y|^\delta + B^\delta\, g^\delta(\mathbf{X})\right\}\right] \\
&\le C\, E\left[|X_j|^\delta\, K_h^\delta(U - u_0)\left\{M_3 + B^\delta\, g^\delta(\mathbf{X})\right\}\right] \\
&\le C\, h^{1-\delta}\, E\left[|X_j|^\delta\, \left\{M_3 + B^\delta\, g^\delta(\mathbf{X})\right\}\right] \\
&\le C\, h^{1-\delta}. \qquad (A.14)
\end{aligned}$$

A combination of (A.13) and (A.14) leads to

$$J_2 \le C\, h^{2/\delta - 2}\, \sum_{l=d_n}^{\infty} [\alpha(l)]^{1-2/\delta} \le C\, h^{2/\delta-2}\, d_n^{-c}\, \sum_{l=d_n}^{\infty} l^c\, [\alpha(l)]^{1-2/\delta} = o\left(h^{-1}\right) \qquad (A.15)$$

by choosing d_n such that $h^{1-2/\delta}\, d_n^c = C$, so that the requirement that $d_n\, h_n \to 0$ is satisfied. ∎

Proof of Theorem 2:

We employ the small-block and large-block technique. Namely, partition $\{1, \ldots, n\}$ into $2\,q_n + 1$ subsets with large-block of size $r = r_n$ and small-block of size $s = s_n$. Set

$$q = q_n = \left\lfloor \frac{n}{r_n + s_n} \right\rfloor. \tag{A.16}$$

We now employ the Cramér-Wold device to derive the asymptotic normality of $\mathbf{Q}_n$. For any unit vector $\mathbf{d} \in \Re^p$, let $Z_{n,i} = \sqrt{h}\,\mathbf{d}^T\,\mathbf{Z}_{i+1},\ \ i = 0, \ldots, n-1$. Then

$$\sqrt{n\,h}\,\mathbf{d}^T\,\mathbf{Q}_n = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} Z_{n,i}$$

and by Lemma 1

$$\mathrm{Var}(Z_{n,0}) = f_{\mathrm{U}}(u_0)\,\mathbf{d}^T\,\Omega^*(u_0)\,\mathbf{d}\ \left[c_0^2\,\nu_0 + 2\,c_0\,c_1\,\nu_1 + c_1^2\,\nu_2\right]\ (1 + o(1)) \equiv \theta^2(u_0)(1 + o(1), \tag{A.17}$$

and

$$\sum_{l=0}^{n-1} |\mathrm{Cov}(Z_{n,0},\ Z_{n,l})| = o(1). \tag{A.18}$$

Define the random variables, for $0 \le j \le q-1$,

$$\eta_j = \sum_{i=j(r+s)}^{j(r+s)+r-1} Z_{n,i}, \qquad \xi_j = \sum_{i=j(r+s)+r}^{(j+1)(r+s)} Z_{n,i}, \qquad \text{and} \qquad \zeta_q = \sum_{i=q(r+s)}^{n-1} Z_{n,i}.$$

Then,

$$\sqrt{n\,h}\,\mathbf{d}^T\,\mathbf{Q}_n = \frac{1}{\sqrt{n}} \left\{ \sum_{j=0}^{q-1} \eta_j + \sum_{j=0}^{q-1} \xi + \zeta_q \right\} \equiv \frac{1}{\sqrt{n}} \left\{ Q_{n,1} + Q_{n,2} + Q_{n,3} \right\}. \tag{A.19}$$

We will show that, as $n \to \infty$,

$$\frac{1}{n}\,E\left[Q_{n,2}\right]^2 \to 0, \qquad \frac{1}{n}\,E\left[Q_{n,3}\right]^2 \to 0, \tag{A.20}$$

$$\left| E\left[\exp(i\,t\,Q_{n,1})\right] - \prod_{j=0}^{q-1} E\left[\exp(i\,t\,\eta_j)\right] \right| \to 0, \tag{A.21}$$

$$\frac{1}{n} \sum_{j=0}^{q-1} E\left(\eta_j^2\right) \to \theta^2(u_0), \tag{A.22}$$

and

$$\frac{1}{n} \sum_{j=0}^{q-1} E\left[\eta_j^2 I\left\{|\eta_j| \ge \varepsilon\,\theta(u_0)\,\sqrt{n}\right\}\right] \to 0 \tag{A.23}$$

for every $\varepsilon > 0$. (A.20) implies that $Q_{n,2}$ and $Q_{n,3}$ are asymptotically negligible in probability; (A.21) shows that the summands η_j in $Q_{n,1}$ are asymptotically independent; and (A.22) and (A.23) are the standard Lindeberg-Feller conditions for asymptotic normality of $Q_{n,1}$ for the independent setup.

Let us first establish (A.20). For this purpose, we choose the large-block size. Condition 2(ii) implies that there is a sequence of positive constants $\gamma_n \to \infty$ such that

$$\gamma_n \, s_n = o\left(\sqrt{n\, h_n}\right), \qquad \text{and} \qquad \gamma_n (n/h_n)^{1/2}\, \alpha(s_n) \to 0. \tag{A.24}$$

Define the large-block size r_n by $r_n = \lfloor (n\, h_n)^{1/2}/\gamma_n \rfloor$ and the small-block size s_n. Then, it can easily be shown from (A.24) that, as $n \to \infty$,

$$s_n/r_n \to 0, \qquad r_n/n \to 0, \qquad r_n\, (n\, h_n)^{-1/2} \to 0, \qquad \text{and} \qquad (n/r_n)\, \alpha(s_n) \to 0. \tag{A.25}$$

Observe that

$$E\,[Q_{n,2}]^2 = \sum_{j=0}^{q-1} \mathrm{Var}(\xi_j) + 2 \sum_{0 \le i < j \le q-1} \mathrm{Cov}(\xi_i,\, \xi_j) \equiv I_1 + I_2. \tag{A.26}$$

It follows from stationarity and Lemma 1 that

$$I_1 = q_n\, \mathrm{Var}(\xi_1) = q_n\, \mathrm{Var}\left(\sum_{j=1}^{s_n} Z_{n,j}\right) = q_n\, s_n\, [\theta^2(u_0) + o(1)]. \tag{A.27}$$

Next consider the second term I_2 in the right hand side of (A.26). Let $r_j^* = j(r_n + s_n)$, then $r_j^* - r_i^* \ge r_n$ for all $j > i$, we therefore have

$$|I_2| \le 2 \sum_{0 \le i < j \le q-1} \sum_{j_1=1}^{s_n} \sum_{j_2=1}^{s_n} |\mathrm{Cov}(Z_{n, r_i^* + r_n + j_1},\, Z_{n, r_j^* + r_n + j_2})| \le 2 \sum_{j_1=1}^{n - r_n} \sum_{j_2 = j_1 + r_n}^{n} |\mathrm{Cov}(Z_{n,j_1},\, Z_{n,j_2})|.$$

By stationarity and Lemma 1, one obtains

$$|I_2| \le 2n \sum_{j=r_n+1}^{n} |\mathrm{Cov}(Z_{n,1},\, Z_{n,j})| = o(n). \tag{A.28}$$

Hence, by (A.25)-(A.28), we have

$$\frac{1}{n}\, E[Q_{n,2}]^2 = O\left(q_n\, s_n\, n^{-1}\right) + o(1) = o(1). \tag{A.29}$$

It follows from stationarity, (A.25) and Lemma 1 that

$$\mathrm{Var}\,[Q_{n,3}] = \mathrm{Var}\left(\sum_{j=1}^{n - q_n(r_n + s_n)} Z_{n,j}\right) = O(n - q_n(r_n + s_n)) = o(n). \tag{A.30}$$

Combining (A.25), (A.29) and (A.30), we establish (A.20). As for (A.22), by stationarity, (A.25) and Lemma 1, it is easily seen that

$$\frac{1}{n} \sum_{j=0}^{q_n - 1} E\left(\eta_j^2\right) = \frac{q_n}{n} E\left(\eta_1^2\right) = \frac{q_n\, r_n}{n} \cdot \frac{1}{r_n} \mathrm{Var}\left(\sum_{j=1}^{r_n} Z_{n,j}\right) \to \theta^2(u_0).$$

In order to establish (A.21), we make use of Lemma 1.1 in Volkonskii and Rozanov (1959) (see also Ibragimov and Linnik 1971, p.338) to obtain

$$\left| E\,[\exp(i\, t\, Q_{n,1})] - \prod_{j=0}^{q_n - 1} E\,[\exp(i\, t\, \eta_j)] \right| \le 16\, (n/r_n)\, \alpha(s_n)$$

tending to zero by (A.25).

It remains to establish (A.23). For this purpose, we employ Theorem 4.1 in Shao and Yu (1996) and Condition 2 to obtain

$$E\left[\eta_1^2\, I\left\{|\eta_1| \geq \varepsilon\,\theta(u_0)\,\sqrt{n}\right\}\right] \leq C\, n^{1-\delta/2}\, E\left(|\eta_1|^{\delta}\right) \leq C\, n^{1-\delta/2}\, r_n^{\delta/2}\, \left\{E\left(|Z_{n,0}|^{\delta^*}\right)\right\}^{\delta/\delta^*}. \tag{A.31}$$

As in (A.14),

$$E\left(|Z_{n,0}|^{\delta^*}\right) \leq C\, h^{1-\delta^*/2}. \tag{A.32}$$

Therefore, by (A.31) and (A.32),

$$E\left[\eta_1^2\, I\left\{|\eta_1| \geq \varepsilon\,\theta(u_0)\,\sqrt{n}\right\}\right] \leq C\, n^{1-\delta/2}\, r_n^{\delta/2}\, h^{(2-\delta^*)\delta/(2\,\delta^*)}. \tag{A.33}$$

Thus, by (A.16) and the definition of r_n, and using Conditions 2(iii) and (iv), we obtain

$$\frac{1}{n}\sum_{j=0}^{q-1} E\left[\eta_j^2 I\left\{|\eta_j| \geq \varepsilon\,\theta(u_0)\,\sqrt{n}\right\}\right] \leq C\,\gamma_n^{1-\delta/2}\, n^{1/2-\delta/4}\, h_n^{\delta/\delta^*-1} \to 0 \tag{A.34}$$

since $\gamma_n \to \infty$. This completes the proof of the theorem. ∎